EMBRY-RIDDLE Aeronautical University
SCHOLARLY COMMONS

Publications

7-2016

Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016

Robert Carver
Stonehill College

Michelle Everson
The Ohio State University

John Gabrosek
Grand Valley State University

Nicholas Horton
Amherst College

Robin Lock
Saint Lawrence University

See next page for additional authors

Follow this and additional works at: https://commons.erau.edu/publication

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Higher Education Commons, and the Statistics and Probability Commons

Scholarly Commons Citation

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., Witmer, J., & Wood, B. (2016). Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016. *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016,* (). Retrieved from https://commons.erau.edu/publication/1083

This Report is brought to you for free and open access by Scholarly Commons. It has been accepted for inclusion in Publications by an authorized administrator of Scholarly Commons. For more information, please contact commons@erau.edu.

Authors
Robert Carver, Michelle Everson, John Gabrosek, Nicholas Horton, Robin Lock, Megan Mocko, Allan Rossman, Ginger Holmes Roswell, Paul Velleman, Jeffrey Witmer, and Beverly Wood

This report is available at Scholarly Commons: https://commons.erau.edu/publication/1083

Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016

Committee:

Robert Carver (Stonehill College), Michelle Everson, co-chair (The Ohio State University), John Gabrosek (Grand Valley State University), Nicholas Horton (Amherst College), Robin Lock (St. Lawrence University), Megan Mocko, co-chair (University of Florida), Allan Rossman (Cal Poly – San Luis Obispo), Ginger Holmes Rowell (Middle Tennessee State University), Paul Velleman (Cornell University), Jeffrey Witmer (Oberlin College), and Beverly Wood (Embry-Riddle Aeronautical University)

Citation: GAISE College Report ASA Revision Committee, "Guidelines for Assessment and Instruction in Statistics Education College Report 2016," http://www.amstat.org/education/gaise.

Endorsed by the American Statistical Association

July 2016

Contents

Executive Summary

In 2005 the American Statistical Association (ASA) endorsed the *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) College Report. This report has had a profound impact on the teaching of introductory statistics in two- and four-year institutions, and the six recommendations put forward in the report have stood the test of time. Much has happened within the statistics education community and beyond in the intervening 10 years, making it critical to re-evaluate and update this important report.

For readers who are unfamiliar with the original GAISE College Report or who are new to the statistics education community, the full version of the 2005 report can be found at http://www.amstat.org/education/gaise/GaiseCollege_full.pdf and a brief history of statistics education can be found in Appendix A of this new report.

The revised GAISE College Report takes into account the many changes in the world of statistics education and statistical practice since 2005 and suggests a direction for the future of introductory statistics courses. Our work has been informed by outreach to the statistics education community and by reference to the statistics education literature.

We continue to endorse the six recommendations outlined in the original GAISE College Report. We have simplified the language within some of these recommendations and re-ordered other recommendations so as to focus first on *what* to teach in introductory courses and then on *how* to teach those courses. We have also added two new emphases to the first recommendation. The revised recommendations are:

1. Teach statistical thinking.
    - Teach statistics as an investigative process of problem-solving and decision-making.
    - Give students experience with multivariable thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

This report includes an updated list of learning objectives for students in introductory courses, along with suggested topics that might be omitted from or de-emphasized in an introductory course. In response to feedback from statistics educators, we have substantially expanded and updated some appendices. We also created some new appendices to provide details about the evolution of introductory statistics courses; examples involving multivariable thinking; and ideas for implementing the GAISE recommendations in a variety of different learning environments.

Introduction

Background

Much has changed since the ASA endorsed the *Guidelines for Assessment and Instruction in Statistics Education* College Report (hereafter called the GAISE College Report) in 2005. Some highlights include:

- **More students are studying statistics.** According to the Conference Board on Mathematical Sciences (CBMS) survey, 508,000 students took an introductory statistics course in a two- or four-year college/university in the fall of 2010, a 34.7% increase from 2005. More than a quarter (27.0%) of these enrollments were at two-year colleges[1]. Nearly 200,000 students took the Advanced Placement (AP) Statistics exam in 2015, an increase of more than 150% over 2005. In addition, many high school students took the AP course without taking the exam or took a non-AP statistics course. At the undergraduate level, the number of students completing an undergraduate major in Statistics grew by more than 140% between 2003 and 2013 and continues to grow rapidly[2].

- **Many students are exposed to statistical thinking in grades 6 – 12**, because more state standards include a considerable number of statistical concepts and methods. Many of these standards have been influenced by the GAISE PreK – 12 report developed and endorsed by the ASA[3]. In particular, the Common Core[4] includes standards on interpreting categorical and quantitative data and on making inferences and justifying conclusions.

- **The rapid increase in available data has made the field of statistics more salient.** Many have heralded the flood of information now available. *The Economist* published a special report on the "data deluge" in 2010[5]. Statisticians such as Hans Rosling and Nate Silver have achieved celebrity status by demonstrating how to garner insights from data[6].

- **The discipline of Data Science has emerged as a field that encompasses elements of statistics, computer science, and domain-specific knowledge[7].** Data science has been described as the interplay between computational and inferential thinking[8]. It includes the analysis of data types such as text, audio, and video, which are becoming more

[1] http://www.ams.org/profession/data/cbms-survey/cbms2010-Report.pdf
[2] http://www.amstat.org/education/curriculumguidelines.cfm
[3] http://www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf
[4] http://www.corestandards.org/
[5] http://www.economist.com/printedition/2010-02-27
[6] Silver, N. (2012) *The Signal and the Noise: Why So Many Predictions Fail—but Some Don't.* New York: Penguin Books. Also see http://www.fivethirtyeight.com and http://www.gapminder.org.
[7] http://magazine.amstat.org/blog/2015/10/01/asa-statement-on-the-role-of-statistics-in-data-science/
[8] Jordan, M. (2016) Computational Thinking and Inferential Thinking: Foundations of Data Science eCOTS 2016. Also see at https://www.causeweb.org/cause/ecots/ecots16/keynotes/jordan.

prevalent. There has been a parallel development of "analytics" as the study of extracting information from big data—particularly with business and governmental applications.

- **More and better technology options for education have become widely available**. These include course management systems, automated homework systems, technology for facilitating discussion and engagement, audience response systems, and videos now used in many courses. Applets and other applications, such as Shiny apps coded in the R programming language, that are designed to explore statistical concepts have come into widespread use. Many general-purpose statistical packages have developed functions specifically for teaching and learning.

- **Alternative learning environments have become more popular.** These include online courses, hybrid courses, flipped classrooms, and Massively Open Online Courses (MOOCs). Many of these environments may be particularly helpful for supporting faculty development.

- **Some have called for an update to the consensus introductory statistics curriculum** to account for the rich data that are available to answer important statistical questions.

- **Innovative ways to teach the logic of statistical inference have received increasing attention.** Among these are greater use of computer-based simulations and the use of resampling methods (randomization tests and bootstrapping) to teach concepts of inference[9].

Concurrent with these changes, the ASA has promoted effective and innovative activities in statistics education on several fronts, including the development and release of the following reports:

- ***Curriculum Guidelines for Undergraduate Programs in Statistical Science***[10], which identifies the increased importance of teaching data science, real applications, more diverse models and approaches, and the ability to communicate;

- ***Statistical Education of Teachers***[11], a report that provides recommendations for preparing teachers of statistics at elementary, middle, and high school levels, and is meant to accompany the influential *Mathematical Education of Teachers* report[12];

- ***Qualifications for Teaching Introductory Statistics***[13], a statement produced by a joint committee of the ASA and Mathematical Association of America, which recommends that statistics teachers have at least the equivalent of two courses in statistical methods and some experience with data analysis beyond the material taught in introductory courses;

---

[9] G.W. Cobb's plenary at USCOTS 2005 presentation and later article http://escholarship.org/uc/item/6hb3k0nz .
[10] http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf
[11] http://www.amstat.org/education/SET/SET.pdf
[12] http://www.cbmsweb.org/MET2/met2.pdf
[13] http://www.amstat.org/education/pdfs/TeachingIntroStats-Qualifications.pdf

- ***ASA's Statement on* p*-Values*[14]**, which puts forward several important principles about hypothesis testing based on consensus among those in the statistical community, in an effort to improve the ways in which the statistical results of scientific studies are reported and interpreted.

Recommendations

We are gratified by how well the GAISE recommendations from 2005 and those of the 1992 Cobb report have held up over time. We attribute this to the broad, general, useful, and universal nature of the framework for instruction.

We continue to endorse the six GAISE recommendations put forth in 2005. We have reordered the recommendations so the first two address *what* to teach and the next four concern *how* to teach. We have also simplified and clarified some of the recommendations. The revised recommendations are

1. Teach statistical thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and a purpose.
4. Foster active learning.
5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

In addition to these six recommendations, which remain central, we suggest two new emphases for the first recommendation (teach statistical thinking) that reflect modern practice and take advantage of widely available technologies:

a. **Teach statistics as an investigative process of problem-solving and decision-making**. Students should not leave their introductory statistics course with the mistaken impression that statistics consists of an unrelated collection of formulas and methods. Rather, students should understand that statistics is a problem-solving and decision-making *process* that is fundamental to scientific inquiry and essential for making sound decisions.

b. **Give students experience with multivariable thinking.** We live in a complex world in which the answer to a question often depends on many factors. Students will encounter such situations within their own fields of study and everyday lives. We must prepare our students to answer challenging questions that require them to investigate and explore relationships among many variables. Doing so will help them to appreciate the value of statistical thinking and methods.

---

[14] Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on *p*-Values: Context, Process, and Purpose," *The American Statistician,* 70. http://dx.doi.org/10.1080/00031305.2016.1154108.

Scope

There is **no single** introductory statistics course. The variety of courses reflects a wide range of needs.

- Some introductory courses address statistical *literacy*, while others focus on statistical *methods*. This distinction is sometimes referred to as courses for *consumers* versus those for *producers* of analyses.
- Introductory statistics courses target many different *student audiences*. There are different needs at different institutions, from elite universities to community colleges, with varying access to technology and support. Some statistics courses aim for a general student audience, while others are targeted at students in the life sciences, at business students, at future engineers, or at mathematics majors. There is a demand in some of these fields for examples and topical coverage tailored to the specific needs of the applied discipline.
- *Prerequisites* differ among introductory statistics courses, the primary distinction being that some require calculus but most require no more than high school algebra.
- *Class sizes* range from small courses of a dozen students that might be taught in a computer lab to large lecture courses for hundreds of students to massively open online courses (MOOCs) taught to thousands asynchronously.

We believe that the six GAISE recommendations apply to the many variations of introductory statistics courses, although the specifics of how they are implemented in these courses will vary to suit the situation. Despite the fact that this report focuses on introductory courses, we believe that the GAISE recommendations also apply to statistics courses **beyond the introductory level**. We urge instructors to consider applying these recommendations throughout undergraduate statistics courses, including courses in statistical practice, statistical computing, and statistical theory.

Support for Implementation

Throughout the development of this report, we have tried to maintain *realistic* expectations while setting *aspirational* goals. We hope that this report can help instructors of introductory statistics improve their courses. We recognize that instructors face constraints that can make innovation challenging, but we believe that any statistics course can benefit from incremental changes that produce closer alignment with the six recommendations.

To facilitate innovation, this report includes substantially expanded and revised appendices that provide many examples of

- **activities and datasets** to illustrate active learning of statistical thinking,
- **assessment** items, instruments, assignments, and rubrics,
- **technology** tools for exploring concepts and analyzing data, and

- suggestions by which the guidelines can be fulfilled in various **learning environments** (face-to-face, flipped, online, etc.).

Goals for Students in Introductory Statistics Courses

The desired result of all introductory statistics courses is to produce statistically educated students, which means that students should develop the ability to think statistically.

The following goals reflect major strands in the collective thinking expressed in the statistics education literature. They summarize what a student should know and understand at the conclusion of a first course in statistics. Achieving this knowledge will require learning some statistical techniques, but mastering specific techniques is not as important as understanding the statistical concepts and principles that underlie such techniques. Therefore, we are not recommending specific topical coverage.

1. Students should become *critical consumers* of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.
2. Students should be able to recognize questions for which the *investigative process* in statistics would be useful and should be able to answer questions using the investigative process.
3. Students should be able to produce *graphical displays and numerical summaries* and interpret what graphs do and do not reveal.
4. Students should recognize and be able to explain the central role of *variability* in the field of statistics.
5. Students should recognize and be able to explain the central role of *randomness* in designing studies and drawing conclusions.
6. Students should gain experience with how *statistical models*, including multivariable models, are used.
7. Students should demonstrate an understanding of, and ability to use, basic ideas of *statistical inference*, both hypothesis tests and interval estimation, in a variety of settings.
8. Students should be able to interpret and draw conclusions from standard output from *statistical software packages*.
9. Students should demonstrate an awareness of *ethical issues* associated with sound statistical practice.

Goal 1: Students should become *critical consumers* of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.

To be a critical consumer of statistically-based results, it is necessary to understand the components that produced them: the design of the investigation, the data, its analysis, and its interpretation. Identifying the variables in a study, which includes consideration of the measurement units, is a necessary step to inform judgments or comparisons. Identifying the subjects (cases, observational units) of a study and the population to which the results of an analysis can be generalized helps the consumer to recognize whether the reported results can reasonably support the conclusions claimed for an analysis. Being able to interpret displays of data (tables, graphs, and visualizations) and statistical analyses also informs the consumer about the reasonableness of the claims being presented.

Goal 2: Students should be able to recognize questions for which the *investigative process* in statistics would be useful and should be able to answer questions using the investigative process.

The investigative process begins with a question that can be translated into one or more statistical questions – questions that can be investigated using data. While many questions do not have simple yes or no answers, knowing how to obtain or generate data that are relevant to the goals of a study is crucial to providing useful information that supports decision-making in the sciences, business, healthcare, law, the humanities, etc. Understanding and applying the principles of representative sampling for an observational study or designing an experiment is critical to the investigative process. Understanding and, when possible, controlling for the impact of other variables is important.

Once high quality data have been collected, meaningful graphs and numerical summaries (generally created using technology) shed light on the question under study. These summaries help to identify statistical inference procedures that are appropriate to the question. The results of the data analysis, and any limitations, need to be clearly communicated.

Goal 3: Students should be able to produce *graphical displays and numerical summaries* and interpret what these do and do not reveal.

Data analysis involves much more than constructing a confidence interval or finding a p-value. Graphical displays of data provide information on the distribution of data values, relationships among variables, and outliers. With the advent of large datasets – often from observational studies that may not be a random sample from a defined population, making standard inferential techniques inappropriate – the proper use of graphical displays is critical. Using software to produce graphical displays makes visualization of large data sets relatively easy. Important univariate graphical displays include histograms, boxplots, dotplots, and bar charts. Bivariate graphical displays include scatterplots, clustered and stacked bar charts, and comparative histograms and boxplots. Additional variables can often be added to a graphical display (for example, separately colored points and regression lines for males and females can be included in a scatterplot that relates age to height for children 3 years to 18 years).

Goal 4: Students should recognize and be able to explain the central role of *variability* in the field of statistics.

Variability is a key characteristic of data that underlies statistical associations and inference. Identifying the sources of variability in a statistical study is an important consideration. Graphical displays and numerical summaries help to illustrate and describe distributions of data (shape, center, variability, and unusual observations) and to select appropriate inference techniques. The role of sampling variability is the bridge to making comparisons and drawing inferences. At the introductory level, this includes an understanding of univariate (and perhaps bivariate) sampling distribution and/or randomization distribution models, and the role of features such as sample size, variability in the statistics, and distributional shape in these models. Understanding how results vary from sample to sample is a challenging topic for many students.

Goal 5: Students should recognize and be able to explain the central role of *randomness* in designing studies and drawing conclusions.

The mathematical understanding of “random” (*not* synonymous with haphazard or unplanned) is fundamental to the role that randomness plays in statistical studies. Distinction of probabilistic sampling techniques from non-probabilistic ones help to recognize when it is appropriate for the results of surveys and experiments to be generalized to the population from which the sample was taken. Similarly, random assignment in comparative experiments allows direct cause-and-effect conclusions to be drawn while other data collection methods usually do not.

Goal 6: Students should gain experience with how *statistical models*, including multivariable models, are used.

Understanding the role of models in statistics is a critical skill for being able to investigate the distribution of data values and the relationships between variables. The first recommendation of the GAISE report is to *teach statistical thinking*. One of the key features of statistical thinking is to understand that variables have distributions. Models help us describe the distribution of variables, especially the distribution of one or more variables conditional upon the values of one or more other variables.

It is important to understand that two variables may be associated and that statistical models can be used to assess the strength and direction of the association. Bivariate models that relate two variables – such as the regression model relating a dependent quantitative response variable to an independent quantitative explanatory variable – are building blocks for more complicated multivariable models. While the details of these more complicated models may be beyond most introductory courses, it is important that students have an appreciation that the relationship between two variables may depend on other variables. Multivariable relationships, illustrating Simpson’s Paradox or investigated via multiple regression, help students discover that a two-way table or a simple regression line does not necessarily tell the entire (or even an accurate) story of the relationship between two variables.

Goal 7: Students should demonstrate an understanding of, and ability to use, basic ideas of *statistical inference*, both hypothesis tests and interval estimation, in a variety of settings.

Statistical inference involves drawing conclusions about a population from the information contained in a sample. Often this involves calculation of sample statistics to make inferences about population parameters either through estimation (for example, a confidence interval to estimate the proportion of voters who have a favorable impression of the President of the United States) or testing (for example, a hypothesis test to determine if the mean time to headache relief is less for a new drug than a current drug). At least as important as calculating confidence intervals and *p*-values is understanding the concepts underlying statistical inference. Understanding the limitations of inferential procedures, including checking assumptions, and the effect of sample size and other factors, are important to assessing the practical significance of results and that if you conduct multiple tests, some results might be significant just by chance. Being able to identify which inferential methods are appropriate for common one-sample and two-sample parameter problems helps develop statistical thinking skills. Providing ample opportunity to practice drawing and communicating appropriate conclusions from inferential procedures allows students to demonstrate understanding of statistical inference.

Goal 8: Students should be able to interpret and draw conclusions from standard output from *statistical software*.

Modern data analysis involves the use of statistical software to store and analyze (potentially large) datasets. While there may be value to performing some calculations by hand, it is unrealistic to analyze data without the aid of software for all but the smallest datasets. At a minimum, students should interpret output from software. Ideally, students should be given numerous opportunities to analyze data with the best available technology (preferably, statistical software).

Goal 9: Students should demonstrate an awareness of *ethical issues* associated with sound statistical practice.

As data collection becomes more ubiquitous, the potential misuse of statistics becomes more prevalent. Application of proper data collection principles, including human subjects review and the importance of informed consent, are central to the effective and ethical use of statistical methods. Relying on statistical methods to inform decisions should not be confused with abusing data to justify foregone conclusions. With large datasets containing many variables, especially from observational studies, understanding of confounding and multiple testing false positive rates becomes even more relevant.

Recommendations

The American Statistical Association continues to endorse the six recommendations put forward in the original GAISE College Report. The intent of these recommendations is to help students attain the learning goals described previously. Much has changed since 2005; therefore, we have reworded some recommendations to highlight new emphases. We have also reordered the recommendations so that the first two focus on *what* to teach in the introductory courses and the next four focus on *how* to teach the courses. In the sections below, we provide additional explanations and suggestions regarding the recommendations; these have been updated and expanded from the 2005 report.

Recommendation 1: Teach statistical thinking.

An introductory course is also a terminal course for many students. As such, it is important that we think carefully about what our focus should be in this course: what do we want to teach, what skills do we want our students to have when they leave the course? Will they use statistics in follow-up courses and careers, and will they be consumers of statistical information presented in the news and abounding in everyday life?

We propose that it is essential to work on the development of skills that will allow students to think critically about statistical issues and recognize the need for data, the importance of data production, the omnipresence of variability, and the quantification and explanation of variability. In other words, statistical thinking – the type of thinking that statisticians use when approaching or solving statistical problems – should be taught and emphasized in introductory courses (see Wild and Pfannkuch 1999 and Chance 2003 for more discussion of statistical thinking). As part of the development of statistical thinking skills, it is crucial to focus on helping students become better educated consumers of statistical information by introducing them to the basic language and the fundamental ideas of statistics, and by emphasizing the use and interpretation of statistics in everyday life. We want our students to become statistically literate (for more on statistical literacy, see Utts 2003, 2010, 2015).

We urge instructors of statistics to emphasize the practical problem-solving skills that are necessary to answer statistical questions. We should model statistical thinking for our students throughout the course, rather than present students with a set of isolated tools, skills, and procedures. Effective statistical thinking requires seeing connections among statistical ideas and recognizing that most statistical questions can be solved with a variety of procedures and that there is often more than one acceptable solution.

Expanding upon the metaphor introduced by Shoenfeld (1998), Garfield, delMas, and Zieffler (2012) proposed the need to rethink the teaching of the introductory courses so that students leave the course with an understanding not just of routine procedures but of the "big picture of the statistical process that will allow them to solve unfamiliar problems and to articulate and apply their understanding" (p.885). These authors argue that we should refrain from teaching our students merely how to follow recipes and should instead teach them how to really cook. Using a "cooking analogy," they explain that someone who can really "cook" not only understands how to follow recipes but can make easy adjustments at a moment's notice by knowing exactly what to focus on and look for when attempting to assemble not just a single dish but a full meal.

The following carpentry story provides another illustration of statistical thinking and an alternative to the "cooking analogy:"

> In week 1 of the carpentry (statistics) course, we learned to use various kinds of planes (summary statistics). In week 2, we learned to use different kinds of saws (graphs). Then, we learned about using hammers (confidence intervals). Later, we learned about the characteristics of different types of wood (tests). By the end of the course, we had covered many aspects of carpentry (statistics). But I wanted to learn how to build a table (collect and analyze data to answer a question) and I never learned how to do that. *We should teach students that the practical operation of statistics is to collect and analyze data to answer questions.*

As a part of the overarching emphasis to teach statistical thinking, we propose that the introductory course teach statistics as an investigative process of problem-solving and decision-making. We also propose that all students be given experience with multivariable thinking in the introductory course. We expand on each of these ideas below.

Teach statistics as an investigative process of problem-solving and decision-making.

We urge instructors to emphasize the investigative nature of statistics throughout their courses. We hope that doing so can avoid the unfortunate, but not uncommon, reality that many students leave their introductory course thinking of statistics only as a disconnected collection of methods and tools.

In their early work on statistical thinking, Wild and Pfannkuch (1999) summarize the investigative cycle with the acronym PPDAC (Problem, Plan, Data, Analysis, Conclusion). A nice illustration of this cycle appears in classrooms throughout New Zealand: http://new.censusatschool.org.nz/wp-content/uploads/2012/11/data-detective1.pdf.

Another way of thinking about the statistical investigative cycle is provided in the GAISE PreK-12 Report (Franklin et al. 2007), where this process is laid out in four stages:

1. Formulate questions.
2. Collect data.
3. Analyze data.
4. Interpret results.

We do not advocate one particular conception of the investigative process, nor do we recommend a specific number of stages or steps in this process; we do strongly recommend that instructors emphasize the investigative nature of the field of statistics throughout their introductory course. Mentioning the investigative process at the beginning of the course but then treating various course topics in a compartmentalized manner does not help students to see the big picture. We recommend that throughout the entire introductory course, instructors illustrate the complete investigative cycle with every example/exercise presented, starting with the motivating question that led to the data collection and ending with the scope of conclusions and directions for future work.

As we think about engaging students in the investigative process, we hope to create mental habits such as the six mental habits described by Chance (2002):

1. Understand the statistical process as a whole.
2. Always be skeptical.
3. Think about the variables involved.
4. Always relate the data to the context.
5. Understand (and believe) the relevance of statistics.
6. Think beyond the textbook.

De Veaux and Velleman (2008) reiterate this approach in their suggestion that introductory statistics courses should involve students in the process of proposing questions, testing assumptions, and drawing conclusions from data. Statistics involves an investigative process of problem-solving and decision-making, which makes it a fundamental discipline in advancing both scientific discoveries and business and personal decisions.

One way of incorporating the investigative process into a first statistics course is to ask students to complete projects that involve study design, data collection, data analysis, and interpretation. We can also attempt to include activities in our courses that involve students in different parts of the investigative cycle. We might further share examples of real studies that are reported in the news or in journal articles and engage our students in discussion of the conclusions drawn from these studies and whether such conclusions are valid in light of the methods used to gather and explore the data.

Give students experience with multivariable thinking.

When students leave an introductory course, they will likely encounter situations within their own fields of study in which multiple variables relate to one another in intricate ways. We should prepare our students for challenging questions that require investigating and exploring relationships among more than two variables.

Kaplan (2012) has criticized the tendency for the introductory course to focus on simple questions about how two groups differ or about how two variables are correlated. Such questions, while interesting, do not necessarily prepare students to tackle more complicated real-world questions that involve more than one or two variables. Horton (2015) writes that "the lack of application for simple multivariable methods is a major limitation of too many of our courses" (p. 141). To illustrate the power of multivariable thinking and modeling, consider an example that shows how accounting for the percentage of students taking the SAT exam in a state completely changes the conclusion that would be drawn about the relationship between average SAT score and average teacher salary in the state (see Appendix B for more details). This example illustrates that helping students to think about three or more variables does not necessarily require introducing multiple regression; simple graphical displays or techniques such as stratification can suffice.

De Veaux (2015) has also challenged statistics educators to think about how to improve introductory courses by, among other things, emphasizing the multivariate nature of the discipline. He calls for the motivation of univariate questions to arise from more complex models, and he illustrates how this can be done with examples that highlight (a) the relationship between diamond price and color, and how this relationship changes when carat weight is taken into account, and (b) the relationship between the presence or absence of a fireplace and the price of a home in New England, and how this relationship also changes markedly when square footage is taken into account.

Kaplan's, Horton's, and De Veaux's examples illustrate that instructors do not need to go into detail about multivariable modeling in order to provide students with an appreciation for the need to consider how multiple variables interact. Students can explore and investigate such relationships by being presented with interesting questions from rich datasets and then producing appropriate graphical displays. These examples also give rise to discussions of how confounding plays an important role in determining the appropriate scope of conclusions to be drawn from such data. (See http://community.amstat.org/stats101/home for the STAT101 toolkit for instructors of introductory statistics courses.)

Suggestions for teachers:

- Model statistical thinking for students by working examples and explaining the questions and processes involved in solving statistical problems from conception to conclusion.
- Give students practice with developing and using statistical thinking. This should include open-ended problems and projects, in addition to real-life scenarios with multiple variables that can also help students appreciate the role that statistics plays in everyday life. Provide students with examples of real studies and, within each study, discuss the research questions that guided the study, the collection of the data, the analysis of the results, the conclusions that were reached, and the scope of the conclusions.
- Begin most examples throughout the course by considering basic issues such as identifying observational units and variables, classifying variables as categorical or quantitative, and considering whether the study made use of random sampling, random assignment, both, or neither.

- Offer students considerable practice with selecting an appropriate technique to address a particular research question, rather than telling them which technique to use and merely having them implement it.
- Use technology and show students how to use technology effectively to manage data, explore and visualize data, perform inference, and check conditions that underlie inference procedures.
- Assess and give feedback on students' statistical thinking (also see Recommendation 6 below) as they progress through the course. In the appendices to this report, we present examples of projects, activities, and assessment instruments and questions that can be used to develop and evaluate statistical thinking.

Recommendation 2: Focus on conceptual understanding.

Earlier, we highlighted important learning objectives that an instructor hopes their students will achieve. It can be challenging to present material in a way that facilitates students' development of more than just a surface level understanding of important concepts and ideas.

Certainly, an introductory course will involve some computation, though most should be facilitated by technology. It is desirable for students to be able to make decisions about the most appropriate ways to visualize, explore, and, ultimately, analyze a set of data. It will not be helpful for students to know about the tools and procedures that can be used to analyze data if students don't first understand the underlying concepts. Having a good understanding of the concepts will make it easier for students to use necessary tools and procedures to answer particular questions about a dataset.

Procedural steps too often claim students' attention that an effective teacher could otherwise direct toward concepts. Students with a good conceptual foundation from an introductory course will be well-prepared to study additional statistical techniques in a second course.

Suggestions for teachers:

- View the primary goal as to discover and apply concepts.
- Focus on students' understanding of key concepts, illustrated by a few techniques, rather than covering a multitude of techniques with minimal focus on underlying ideas.
- Pare down content of an introductory course to focus on core concepts in more depth.
- Perform most computations using technology to allow greater emphasis on understanding concepts and interpreting results.
- Although the language of mathematics provides compact expression of key ideas, use formulas that enhance the understanding of concepts, and avoid computations that are divorced from understanding.

Recommendation 3: Integrate real data with a context and a purpose.

Using real data in context is crucial in teaching and learning statistics, both to give students experience with analyzing genuine data and to illustrate the usefulness and fascination of our discipline. Statistics can be thought of as the science of learning from data, so the context of the data becomes an integral part of the problem-solving experience. The introduction of a data set should include a context that explains how and why the data were produced or collected. Students should practice formulating good questions and answering them appropriately based on how the data were produced and analyzed.

Using real data sets of interest to students is a good way to engage students in thinking about the data and relevant statistical concepts. Neumann, Hood and Neumann (2013) explored reflections of students who used real data in a statistics course and found the use of real data was associated with students' appreciating the relevance of the course material to everyday life. Further, students indicated that they felt the use of real data made the course more interesting.

Suggestions for teachers:

- Use real data from studies to enliven your class, motivate students, and increase the relevance of the course to the real world.
- Use data with a context as the catalyst for exploration, generating the questions, and informing interpretations to conclusions.
- Make sure questions used with data sets are of interest to students so they can be easily motivated. Take the time to explain why we are interested in this type of data and what it represents. Note: Few data sets interest all students, so instructors should use data from a variety of contexts.
- Use class-generated data to formulate statistical questions and plan uses for the data before developing the questionnaire and collecting the data. For example, ask questions likely to produce different shaped histograms, or use interesting categorical variables to investigate relationships. It is important that data gathered from students in class does not contain information that could be embarrassing to students and that students' privacy is maintained.
- If data entry is a part of the course, get students to practice entering raw data using a small data set or a subset of data, rather than spending time entering a large data set.
- Use statistical software to analyze larger datasets that are available electronically.
- Use subsets of variables in different parts of the course, but integrate the same data sets throughout. (Example: Use side-by-side boxplots to compare two groups, then use two-sample t-tests on the same data. Use histograms to investigate shape, and then later in the course to verify conditions for hypothesis tests. Encourage students to explore how multiple variables in the data set relate to one another.)
- Minimize the use of hypothetical data sets to illustrate a particular point or to assess a specific concept.
- See the Appendices C, D, and E for examples of good ways to use data in class activities, homework, projects, tests, etc.

- Search web data repositories, textbooks, journal articles, software packages, and websites with surveys/polls for good raw data or summarized data to use in class activities. Expect new sources of data to become available each year. Appendix C includes a list of useful websites with data repositories.
- Expose students to data that they interact with on a regular basis, such as data generated by online social networks or data tracked regularly on mobile smart devices (Gould 2010).
- Be alert to the messiness of much real data before using it in a course; better still expose students to typical issues such as missing observations, inconsistent identifiers, and the challenges of merging data from multiple sources (Carver and Stephens, 2014).
- Introduce students to interactive data visualization websites (Ridgway, in press), such as the Gapminder software of Hans Rosling (http://www.gapminder.org) or the website provided by the Office of National Statistics in the UK to explore commuting patterns (http://www.neighbourhood.statistics.gov.uk/HTMLDocs/dvc193/).
- Consider opportunities to align the data sources you select to institutional objectives at your school. For example, you may want to seek out datasets related to expanding students' global awareness, focusing on social justice concerns, or exploring issues of local importance.
- Seek out real data directly from a practicing research scientist through a journal or at one's home institution.

Recommendation 4: Foster active learning.

Active learning has been described as a set of approaches that involve students in *doing* things and *thinking* about the things they are doing (Bonwell and Eison 1991). Using active learning methods in class allows students to discover, construct, and understand important statistical ideas as well as to engage in statistical thinking. Other benefits include the practice students get communicating in statistical language and learning to work in teams to solve problems. Activities provide teachers with a method of assessing student learning and provide feedback to the instructor on how well students are learning. A recent meta-analysis (Freeman et al. 2014) concludes that there are distinct advantages in terms of course outcomes when active learning is employed in STEM courses.

Instructors should not underestimate the learning gains that can be achieved with activities or overestimate the value of lectures to convey information. Embedding even brief activities within lectures can break the natural occasional dips in attention associated with passive or minimally-engaged listeners.

Whereas some rich activities can take an entire class session, many valuable activities need not take much time. A *think-pair-share* discussion or prediction exercise may take only 2-3 minutes, which might otherwise be spent in redundant lecturing due to audience inattention. Collecting on-the-spot data may take more time but reaps benefits beyond the single activity that prompted the collection (see Recommendation 3). Appendix C contains many activities that may replace (or drastically reduce) some lectures and Appendix F has additional suggestions specifically geared to implementing this recommendation in large classes.

Suggestions for teachers:

- Ground activities in the context of real data with a motivating question. Do not "collect data to collect data" for its own sake.
- Consider the student need for physical explorations (*e.g.*, die rolling, card drawing) prior to the use of computer simulations to illustrate or practice concepts.
- Encourage predictions from students about the results of a study that provides the data for an activity before analyzing the data. This motivates the need for statistical methods. (If all results were predictable, we would not need either data or statistics.)
- Avoid activities that lead students step-by-step through a list of procedures. Instead, allow students to discuss and think about the data and the problem.
- When planning activities, be sure there is enough time to explain the problem, let the students work through the problem, and wrap-up the activity during the same class period.
- Consider low-/no-stakes peer assessment (where students comment on or rate a classmate's work) within class to provide quick feedback and to improve the quality of final assessments.

Recommendation 5: Use technology to explore concepts and analyze data.

Technology has changed the practice of statistics and hence should change what and how we teach. By technologies, we refer to a range of hardware and software that can do far more than handle the computational burden of analysis. By adopting the best available tools (subject to institutional constraints), we allow students to *do* analysis more easily and therefore open up time to focus on interpretation of results and testing of conditions, rather than on computational mechanics. Technology should aid students in learning to think statistically and to discover concepts. It should also facilitate access to real (and often large) datasets, foster active learning, and embed assessment into course activities.

Statistics is practiced with computers and usually with specially designed computer software. Students should learn to use a statistical software package if possible. Calculators can provide some limited functionality for smaller datasets, but their use should be supplemented with experience reading typical computer results. **Regardless of the tools used, it is important to view the use of technology not just as a way to generate statistical output but as a way to explore conceptual ideas and enhance student learning.** We caution against using technology merely for the sake of using technology or for pseudo-accuracy (carrying out results to many decimal places). Not all technology tools will have all desired features.

When computers are not available to all students at all times, experience with computers could include one or more of the following:

- A brief introduction to a statistical software package, for example in a computer lab.
- Watching an instructor demonstrate the use of a statistical software package in the context of a statistical investigation.
- Reading generic "computer output" designed to resemble computer package results, but not specifically reproducing any of the major packages. This can be coupled with questions that probe student understanding. (*e.g.*, what is the regression equation?)

For example, an instructor might demonstrate how to estimate a regression equation using a statistical package, then provide students with copies of the resulting regression table and residual plots, and ask students to summarize the results and assess model conditions. Alternatively, an instructor might create an exploratory graph, elicit questions or suggestions from the class, modify the graph in real time, and share the results from the final analysis.

Technology tools should also be used to help students visualize concepts and develop an understanding of abstract ideas by simulations. Some tools offer both types of uses, while in other cases, a statistical software package may be supplemented by web applets.

We note that technology continues to evolve rapidly. Many smart phones or tablets can provide access to online statistical software when sufficient internet access is available. We also note that institutions and courses vary widely in funding and the resources necessary to support this recommendation. The catchphrase should be "use the best available technology."

Some technologies available:

(See Appendix D for in-depth discussion and examples.)

- Interactive applets
- Statistical software
- Web-based resources, including
  - sources of experimental, survey, and observational data
  - online texts
  - data analysis routines
- Games and other virtual environments
- Spreadsheets
- Graphing calculators

Suggestions for teachers:

- Perform routine computations using technology to allow greater emphasis on interpretation of results.

- View the primary goal as discovering concepts rather than covering methods.
- Implement computer-intensive methods to find *p*-values and de-emphasize *t*-, normal and other probability tables. Analyze large, real, data sets.
- Generate and modify appropriate statistical graphics, including relatively recent innovations like motion charts and maps.
- Perform simulations to illustrate abstract concepts.
- Explore "what happens if... " questions.
- Create reports.
- Harness the impact of interactive, real-time visualizations to engage students in the investigative process and in multivariable thinking.
- Use real-time response systems for formative assessment.
- Use games and virtual environments to engage students, teach concepts and gather data.

Considerations for teachers when selecting technology tools:

- Ease of data entry, ability to import data in multiple formats
- Interactive capabilities
- Support of specific pedagogical goals
- Dynamic linking between data, graphical, and numerical analyses
- Ease of use for particular audiences (including those with visual or hearing impairments)
- Availability to students, portability
- Support for reproducible analysis and integration with word-processing and presentation software
- Support for merging data from multiple sources and data management
- Functional consistency across platforms (*i.e.* consistency for Mac and Windows users where students have laptops)
- Tablet and mobile support

Recommendation 6: Use assessments to improve and evaluate student learning.

Students will value what you assess; therefore, assessments need to be aligned with learning goals. Assessments need to focus on understanding key ideas, and not just on skills, procedures, and computed answers. Being able to calculate a *p*-value is not enough; students need to be able to draw conclusions about the research question from a *p*-value and also explain the reasoning process that leads from the *p*-value to the conclusion.

Useful and timely feedback is essential for assessments to lead to learning. There are two types of assessment. *Formative assessment* aims to monitor and improve student learning by providing students with ongoing feedback about their learning during the learning process. Such feedback can also help instructors to improve their teaching by focusing on ideas and concepts that are most challenging for students. Examples of formative assessments include quizzes, homework assignments, and minute papers. *Summative assessment*, in contrast, focuses on evaluating student learning at the end of instruction. Examples of summative assessments include exams (such as a midterm or a final) and final course projects. Formative and summative assessments are certainly not mutually exclusive categories in that good assessments can both promote and evaluate student learning. We encourage instructors to maximize opportunities to include formative assessments into their courses rather than focus exclusively on summative assessments.

Types of assessment:

The practicality of any given type of assessment will vary with each type of course. However, it is possible, even in large classes, to implement good assessments. Below, we list several possible assessment methods that can be used in a course, and Appendix E includes a variety of different assessment items.

- Homework questions
- Quizzes and exams
- Projects
- Activities
- Oral presentations
- Written reports
- Minute papers
- Article critiques

Suggestions for teachers:

- Integrate assessment as an essential component of the course. Assessment tasks that are well-coordinated with what the teacher is doing in class are more effective than tasks that focus on what happened in class two weeks earlier.
- Written assignments such as minute papers, lab reports or even semester long projects can help students strengthen their knowledge of statistical concepts and practice good communication skills.
- Use a variety of assessment methods to provide a more complete evaluation of student learning.
- Use items that focus on choosing good interpretations of graphs or selecting appropriate statistical procedures.
- Have students interpret or critique articles in the news and graphs in media.
- Encourage students to work in groups on some low-stakes assessments (*e.g.*, quizzes) to promote learning from each other.
- Collaborative projects have been identified by the Association of American Colleges and Universities (AAC&U) as a high impact practice (AAC&U, 2008).

- Consider assessing statistical thinking using student projects and open-ended investigative tasks.

Suggestions for student assessment in large classes:

- Use small group projects instead of individual projects.
- Use peer review of projects to provide feedback and improve projects before grading.
- Use discussion sections for student presentations.
- Incorporate real-time response systems (*e.g.*, clickers) in the classroom in order to provide students with opportunities to demonstrate their understanding of course material and instructors with feedback about possible misconceptions or misunderstanding about course material.

Resources on assessment:

- http://www.amstat.org/publications/jse/v2n1/garfield.html
- http://www.amstat.org/publications/jse/v5n3/chance.html
- https://apps3.cehd.umn.edu/artist/index.html

Suggestions for Topics that Might be Omitted from Introductory Statistics Courses

While there is an impressive growth in the number of students taking more advanced courses in statistics, many of our students take only a single course in statistics. This has led to a tendency to cram as much material into the syllabus as possible. The natural question then is what to minimize or diminish. We offer these topics as candidates for reconsideration in the traditional course.

Our guide for these suggestions is to keep in mind *why* the course is required for so many of our students (and elected by so many others). We believe that students need to learn to think scientifically and to deal with statistics in their own disciplines. Students should be able to read research literature with a critical eye. They should be able to understand what was studied, what was concluded, and how as (eventual) professionals and citizens they should judge the conclusions in the context of their own discipline.

The goals set out in this document address concepts and methods that support the development of such a student. Here are some thoughts on topics that might be reconsidered:

- *Probability theory*. The original GAISE report recommended less emphasis on probability in the introductory course and we continue to endorse that recommendation. For many students, an introductory course may be the only statistics course that they take; therefore some instructors will want to teach basic probability and rules about random variables, with perhaps the binomial as a special case. However, the GAISE goals and recommendations can be met without these topics.

- *Constructing plots by hand*. Data displays are now made by computers. Students need to know how to read and interpret them. Instead of spending lots of time creating histograms by hand, use some of that time instead to develop a deeper understanding and ask more challenging questions about what the plots tell us about the data.
- *Basic statistics*. Histograms, pie charts, scatterplots, means, and medians are now taught in middle and high school and are a prominent part of the Common Core State Standards in Mathematics. Classes taught to adults continuing their education or to students with a different high school background may need to spend a bit more time on basic statistics. No matter the audience, instructors will want to be sure that students truly understand these concepts, but should not dwell on them more than is necessary. Instructors may want to briefly review them to be sure terminology and notation are consistent, but this should take little time.
- *Drills with z-, t-, χ^2, and F-tables*. These skills are no longer necessary and do not reflect modern statistical practice. Apps that perform the lookup (and are not limited to a finite list of *df* values) are available in general purpose statistical software packages, web pages, smartphones, or (soon) watches. Since statistical software produces a *p*-value as part of performing a hypothesis test, a shift from finding *p*-values to interpreting *p*-values in context is appropriate (see also the ASA statement on *p*-values: Wasserstein, R. L., and Lazar, N. A., 2016). This shift makes it unnecessary to examine students on their ability to use these tables, so they can usually be dispensed with on exams.
- *Advanced training on a statistical software program*. SAS certification, non-introductory R programming, and other more extensive programming topics belong in subsequent courses. Modern students have grown up with computers and know how to search for support online. The basic computer package skills needed to undertake analyses for the introductory statistics course can often be taught throughout the course or developed using online training. Some instructors may train students in using a specific software package, but mastery of advanced programming skills should not be allowed to crowd out data analysis skills or statistical thinking.

References

Association of American Colleges and Universities (2008), High Impact Educational Practices. https://www.aacu.org/leap/hips.

Bonwell, C. C. and Eison, J. A. (1991), *Active Learning: Creating Excitement in the Classroom* [Monograph]. Retrieved from http://eric.ed.gov/.

Carver, R. and Stephens, M. (2014) "It is Time to Include Data Management in Introductory Statistics", In K. Makar, B. de Sousa, & R. Gould (Eds.), Sustainability in Statistics Education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA. Voorburg, The Netherlands: International Statistical Institute, available at http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_C134_CARVER.pdf.

Chance, B. (2002), "Components of Statistical Thinking and Implications for Instruction and Assessment," *Journal of Statistics Education,* 10. Available at http://www.amstat.org/publications/jse/v10n3/chance.html.

Cobb, G.W. (2005), "Introductory Statistics: A Saber Tooth Curriculum?" Plenary talk given at the United States Conference on Teaching Statistics (USCOTS) . Available at https://www.causeweb.org/cause/sites/default/files/uscots/uscots05/downloads/USCOTS_Dinner_Cobb.ppt.

Cobb, G. W. (2007), "The Introductory Statistics Course: A Ptolemaic Curriculum?" *Technology Innovations in Statistics Education*, *1*. Available at http://escholarship.org/uc/item/6hb3k0nz.

De Veaux, R. (2015), "What's wrong with Stat 101?" Presentation given at the United States Conference on Teaching Statistics (USCOTS). Available at https://www.causeweb.org/cause/sites/default/files/uscots/uscots15/presentations/DickOpening.pptx .

De Veaux, R. and Velleman P. (2008), "Math is Music; Statistics is Literature (Or, Why Are There No Six Year Old Novelists?)," *Amstat News*, 375, 56-58.

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*, Alexandria, VA: American Statistical Association.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., and Wenderoth, M.P. (2014), "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proceedings of the National Academy of Sciences of the United States of America,* 111, 8410–8415.

Garfield, J., delMas, R. and Zieffler, A. (2012), "Developing Statistical Modelers and Thinkers in an Introductory, Tertiary-level Statistics Course," *ZDM: The International Journal on Mathematics Education,* 44, 883-898.

Gould, R. (2010), "Statistics and the Modern Student," *International Statistical Review*, 78, 297–315.

Horton, N.J. (2015), "Challenges and Opportunities for Statistics and Statistics Education: Looking Back, Looking Forward," *The American Statistician,* 69, 138-145.

Jordan, M. (2016) "Computational Thinking and Inferential Thinking: Foundations of Data Science", eCOTS 2016. Available at https://www.causeweb.org/cause/ecots/ecots16/keynotes/jordan.

Kaplan, D. (2012), *Statistical Modeling: A Fresh Approach* (2nd ed.), Charleston, SC: CreateSpace Independent Publishing Platform. Available at http://www.mosaic-web.org/go/StatisticalModeling.

Neumann, D., Hood, M., and Neumann, M. (2013), "Using Real-Life Data when Teaching Statistics: Student Perceptions of this Strategy in an Introductory Statistics Course," *Statistics Education Research Journal,* 12, 59-70. Available at http://iase-web.org/documents/SERJ/SERJ12%282%29_Neumann.pdf

Ridgway, J. (in press), "Implications of the Data Revolution for Statistics Education," *International Statistical Review.* Available at http://onlinelibrary.wiley.com/doi/10.1111/insr.12110/abstract

Schoenfeld, A. H. (1998), "Making Mathematics and Making Pasta: From Cookbook Procedures to Really Cooking," in J. G. Greeno and S.V. Goldman (Eds.), *Thinking Practices in Mathematics and Science Learning*, Mahwah, NJ: Lawrence Erlbaum.

Utts, J. (2003), "What Educated Citizens Should Know about Statistics and Probability," *The American Statistician,* 57, 74–79.

Utts, J. (2010), "Unintentional Lies in the Media: Don't Blame Journalists for What We Don't Teach," Invited paper in C. Reading (Ed.), *Data and context in statistics education: Proceedings of the Eighth International Conference on Teaching Statistics*, Voorsburg, The Netherlands: International Statistical Institute.

Utts, J. (2015), "The Many Facets of Statistics Education: 175 Years of Common Themes," *The American Statistician,* 69, 100-107.

Wasserstein, R. L., and Lazar, N. A. (2016), "The ASA's Statement on *p*-Values: Context, Process, and Purpose," *The American Statistician,* 70. http://dx.doi.org/10.1080/00031305.2016.1154108, Wild, C. J., and Pfannkuch, M. (1999), "Statistical Thinking in Empirical Enquiry," *International Statistical Review*, 67, 223-265.

APPENDIX A: Evolution of Introductory Statistics and Emergence of Statistics Education Resources

Transformation of the Introductory Course

The modern introductory statistics course has roots that go back a long way, to early books about statistical methods. R. A. Fisher's *Statistical Methods for Research Workers*, which first appeared in 1925, was aimed at practicing scientists. A dozen years later, the first edition of George Snedecor's *Statistical Methods* presented an expanded version of the same content, but there was a shift in audience to prospective scientists who were still completing their degrees. By 1961, with the publication of *Probability with Statistical Applications* by Fred Mosteller, Robert Rourke, and George Thomas, statistics had begun to make its way into the broader academic curriculum, but statistics still had to lean heavily on probability for its legitimacy.

During the late 1960s and early 1970s, John Tukey's ideas of exploratory data analysis launched the "data revolution" in the beginning statistics curriculum, freeing certain kinds of data analysis from ties to probability-based models. Analysis of data began to acquire status as an independent intellectual activity that did not require hours chained to a bulky mechanical calculator. Computers later expanded the types of analysis that could be completed by learners.

Two influential books appeared in 1978: *Statistics*, by David Freedman, Robert Pisani, and Roger Purves, and *Statistics: Concepts and Controversies*, by David S. Moore. These textbooks were distinctive in focusing almost exclusively on statistical concepts rather than statistical methods. They were aimed at a broad audience of consumers of statistical information rather than for students who needed to learn to conduct statistical analyses. Then in the 1980s, more and more introductory textbooks on statistical methods included a focus on concepts and real data[15].

The evolution of content has been paralleled by other trends. One of these is a striking and sustained growth in enrollments. Statistics from three groups of students illustrate the growth:

- At two-year colleges, according to the Conference Board of the Mathematical Sciences (CBMS)[16], statistics enrollments grew from 27% the size of calculus enrollments in 1970 to 74% in 2000 and exceeded calculus by 2010.
- Also from the CBMS survey, enrollments in elementary statistics courses at four-year institutions were up 56% in math departments and 50% in statistics departments from 2005 to 2010.

[15] Cobb, G.W. (1987), "Introductory Textbooks: A Framework for Evaluation," *Journal of the American Statistical Association*, 82, 321-339.
[16] http://www.ams.org/profession/data/cbms-survey/cbms-survey

- The Advanced Placement exam in statistics was first offered in 1997 when 7,500 students took it, more than in the first offering of an AP exam in any subject up to that time. More than four times as many students were taking the exam by 2015 when nearly 200,000 students took the test[17].

The democratization of introductory statistics has broadened and diversified the backgrounds, interests, and motivations of those who take the course. Statistics is no longer reserved for future scientists in narrow fields but is now a family of courses, taught to students at many levels, from pre-high school to post-baccalaureate, with very diverse interests and goals. A teacher of today's beginning statistics courses can no longer assume that students are quantitatively skilled and adequately motivated by their career plans.

Not only have the "what, why, who, and when" of introductory statistics been changing, but so has the "how." The last few decades have seen an extraordinary level of activity focused on how students learn statistics and on how teachers can effectively help them learn.

Influential Documents on the Teaching of Statistics

As part of the Curriculum Action Project of the Mathematics Association of America (MAA), George Cobb coordinated a focus group about important issues in statistics education. The 1992 report was published in the MAA volume *Heeding the Call for Change*[18]. It included the following recommendations for teaching introductory courses:

> *Emphasize Statistical Thinking*
>
> Any introductory course should take as its main goal helping students to learn the basic elements of statistical thinking. Many advanced courses would be improved by a more explicit emphasis on those same basic elements, namely:
>
> - The need for data. The importance of data production.
> - The omnipresence of variability.
> - The quantification and explanation of variability.
>
> *More Data and Concepts, Less Theory and Fewer Recipes*
>
> Almost any course in statistics can be improved by more emphasis on data and concepts, and less emphasis on theory and recipes. To the maximum extent feasible, automate calculations and graphics.
>
> *Foster Active Learning*

---

[17] http://research.collegeboard.org/programs/ap/data/participation/ap-2015
[18] Cobb, G.W. (1992), "Teaching Statistics," in *Heeding the Call for Change: Suggestions for Curricular Action* (MAA Notes No. 22), 3–43. Washington, D.C.: The Mathematical Association of America.

> As a rule, teachers of statistics should rely much less on lecturing and much more on alternatives such as projects, lab exercises, and group problem-solving and discussion activities. Even within the traditional lecture setting, it is possible to get students more actively involved.

The three recommendations were intended to apply quite broadly (*e.g.*, whether or not a course has a calculus prerequisite and regardless of the extent to which students are expected to learn specific statistical methods). Cobb's focus group evolved into the joint ASA/MAA Committee on Undergraduate Statistics. A growing body of statistics educators were implementing the recommendations and actively sharing their experiences with peers, often through projects and workshops funded by the National Science Foundation (NSF).

In the late 1990s, Joan Garfield led an NSF-funded survey[19] to explore the impact of this educational reform movement. A large number of statistics instructors from mathematics and statistics departments and a smaller number of statistics instructors from departments of psychology, sociology, business, and economics were included. The responses were encouraging: many reported increased use of technology, diversification of assessment methods, and successful implementation of active learning strategies.

The American Statistical Association funded a strategic initiative to create a set of Guidelines for Assessment and Instruction in Statistics Education (GAISE) at the outset of the 21st century. This was a two-part project that resulted in the publication[20] of *A Pre-K–12 Curriculum Framework*[21] and the original 2005 *College Report* that expanded upon the recommendations from the Cobb Report to address technology and assessment. These two reports have had a profound effect on the practice of teaching statistics and on the training of statistics educators at all levels.

Since the GAISE publications, the widespread adoption of the *Common Core State Standards*[22] has both strengthened the status of statistics as an academic necessity and challenged the content of a first collegiate course in statistics. The arrival of students from high school already exposed to topics formerly taught only in a college course (*e.g.*, probability, exploratory data analysis, measures of center and variability, basic ideas of inference) is a shift as profound as the 1970s arrival of students without strong quantitative skills. New technology allowed the restructuring of the 20th century curriculum to include focus on concepts rather than computation as an adaptation to the new type of student. Forty years later, the 21st century curriculum has the opportunity to build on a broader foundation of prior knowledge that leaves room to delve deeper and farther than ever before possible.

---

[19] Garfield, J. (2000), *An Evaluation of the Impact of Statistics Reform: Final Report.* National Science Foundation (REC-9732404).
[20] amstat.org/education/gaise
[21] Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007), *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K-12 Curriculum Framework*, Alexandria, VA: American Statistical Association.
[22] http://www.corestandards.org/

The Emergence of Statistics Education Research and Resources

Even before the publication of the original GAISE College Report, distinctions between mathematics education and statistics education were being made[23]. The connection between the disciplines, however, remains important and interesting to both mathematicians and statisticians. The American Statistical Association (ASA) maintains joint committees with the Mathematical Association of America (MAA), the American Mathematical Association of Two-Year Colleges (AMATYC), and the National Council of Teachers of Mathematics (NCTM).

The Statistical Education Section is one of the oldest sections within the ASA, founded in 1948, originally focused on the education of professional statisticians. The current mission statement of the Section includes advising the Association on educational elements in communication with non-statistical audiences, promoting reach and practice in statistical education; supporting the dissemination of development/funding opportunities, teaching resources, and research findings in statistical education; and improving the pipeline from K-12 through colleges to statistics professionals[24].

The American Statistical Association recently updated their guidelines for undergraduate programs in statistical science[25, 26] to ensure that minors and majors in statistics provide sufficient background in core skill areas: statistical methods and theory, data manipulation, computation, mathematical foundations, and statistical practice.

In 2014 the ASA and MAA jointly endorsed a set of guidelines[27] for those teaching an introductory statistics course. These guidelines stipulate that instructors of statistics ideally meet the following qualifications:

- Experience with data and appropriate use of technology to support data analyses
- Deep knowledge of statistics and appreciation for the differences between statistical thinking and mathematical thinking
- Understanding the ways statisticians work with real data and approach problems and experiencing the joys of making discoveries using statistical reasoning
- Mentoring by an experienced statistics instructor for instructors unfamiliar with the data-driven techniques used in modern introductory statistics courses

These guidelines recommend minimum qualifications for teaching introductory statistics as consisting of the following:

---

[23] Ben-Zvi, D., and Garfield, J. (2008), "Introducing the Emerging Discipline of Statistics Education," *School Science & Mathematics*, 108, 355-361.
[24] http://community.amstat.org/statisticaleducationsection/home
[25] http://www.amstat.org/education/curriculumguidelines.cfm
[26] Horton NJ and Hardin J. "Teaching the Next Generation of Statistics Students to 'Think with Data': Special Issue on Statistics and the Undergraduate Curriculum", The American Statistician, 2015; 69(4):259-265, http://www.tandfonline.com/doi/full/10.1080/00031305.2015.1094283.
[27] http://magazine.amstat.org/blog/2014/04/01/asamaaguidelines/

- two statistical methods courses, including content knowledge of data-collection methods, study design, and statistical inference, and
- experience with data analysis beyond material taught in the introductory class (*e.g.*, advanced courses, projects, consulting, or research).

In 2000, the MAA founded a special interest group (SIGMAA) on Statistics Education. Their purpose is also four-fold: facilitate the exchange of ideas about teaching statistics, the undergraduate statistics curriculum, and other issues related to providing effective/engaging encounters for students; foster increased understanding of statistics through publication; promote the discipline of statistics among students; and work cooperatively with other organizations to encourage effective teaching and learning[28].

The AMATYC Committee on Statistics was founded in 2010 to provide a forum for the exchange of ideas, the sharing of resources, and the discussion of issues of interest to the statistics community. The committee strives to provide professional development opportunities that support the teaching and learning of statistics and that foster the use of the GAISE College Report recommendations in the first two years of college. It also serves as a liaison with faculty at four-year institutions and with other professional organizations for the purpose of resource sharing (see the AMATYC Statistics Resources Page[29]).

A 2006 charter established the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) which had grown out of a 2002 strategic initiative within the ASA. The mission of CAUSE is to support and advance undergraduate statistics education through resources, professional development, outreach and research. CAUSEweb.org serves as a repository for all of those areas. CAUSE also coordinates the US Conference on Teaching Statistics (USCOTS) which has been held in the spring of odd-numbered years since 2005. Since 2012, the electronic Conference on Teaching Statistics (eCOTS) has provided a virtual conference experience on even-numbered years.

The oldest conference for statistics educators, however, is sponsored by the International Association of Statistical Education (IASE)[30], a section of the International Statistical Institute. The International Conference on Teaching Statistics (ICOTS) has been held every four years since 1982 at various global locations. The IASE also supports the *Statistics Education Research Journal* (SERJ), a peer-reviewed electronic journal in publication since 2002.

Other refereed journals of interest to statistics educators include *Teaching Statistics*[31], the *Journal of Statistics Education* (JSE)[32], and *Technology Innovations in Statistics Education* (TISE)[33].

---

[28] http://sigmaa.maa.org/stat-ed/art1.html
[29] http://www.amatyc.org/?page=StatsResources
[30] http://iase-web.org/
[31] http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1467-9639
[32] http://www.amstat.org/publications/jse/
[33] http://escholarship.org/uc/uclastat_cts_tise

APPENDIX B: Multivariable Thinking

The 2014 ASA guidelines for undergraduate programs in statistics recommend that students obtain a clear understanding of principles of statistical design and tools to assess and account for the possible impact of other measured and unmeasured confounding variables (ASA 2014)[34]. An introductory statistics course cannot cover these topics in depth, but it is important to expose students to them even in their first course (Meng 2011). Perhaps the best place to start is to consider how a third variable can change our understanding of the relationship between two variables.

In this appendix we describe simple examples where a third factor clouds the association between two other variables. Simple approaches (such as stratification) can help to discern the true associations. Stratification requires no advanced methods, nor even any inference, though some instructors may incorporate other related concepts and approaches such as multiple regression. These examples can help to introduce students to techniques for assessing relationships between more than two variables.

Including one or more multivariable examples early in an introductory statistics course may help to prepare students to deal with more than one or two variables at a time and examples of observational (or "found") data that arise more commonly than results from randomized comparisons.

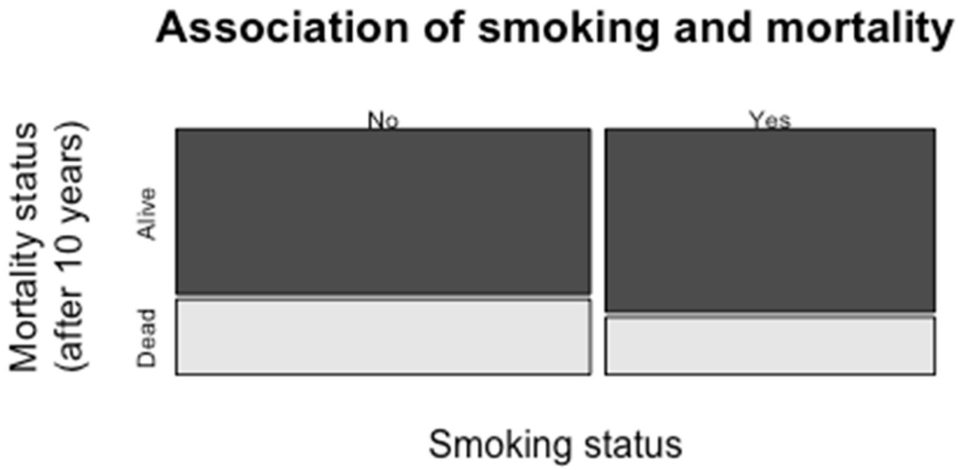
Smoking in Whickham

A follow-up study of 1,314 people in Whickham, England characterized smoking status at baseline, then mortality after 10 years (Appleton et al. 1996). The summary data are provided in the following table:

```
SMOKER        Alive          Dead
  No      502 (68.6%)     230 (31.4%)
  Yes     443 (76.1%)     139 (23.9%)
```

We see that the risk of dying is lower for smokers than for non-smokers, since 31.4% of the non-smokers died, but only 23.9% of the smokers did not survive over the ten year period. A graphical representation using a mosaicplot (also known as an *Eikosogram*) represents the cell probabilities as a function of area.

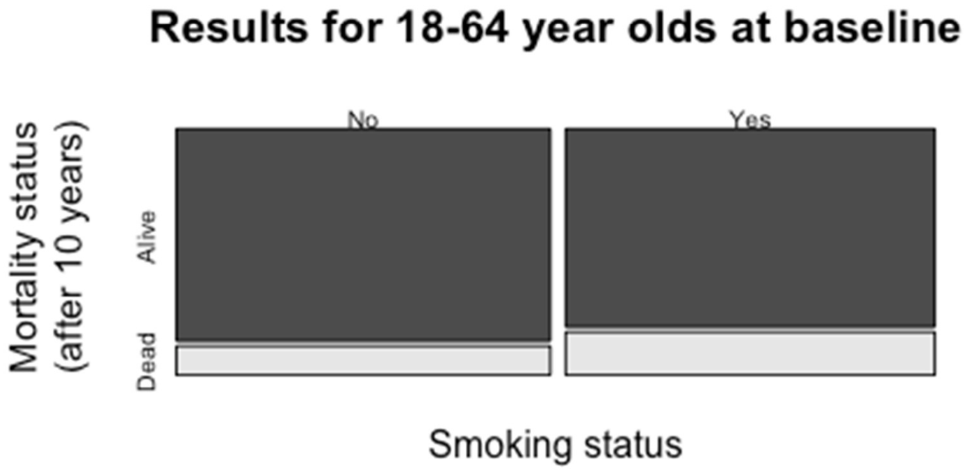
[34] See also Wild's "On Locating Statistics in the World of Finding Out," http://arxiv.org/abs/1507.05982.

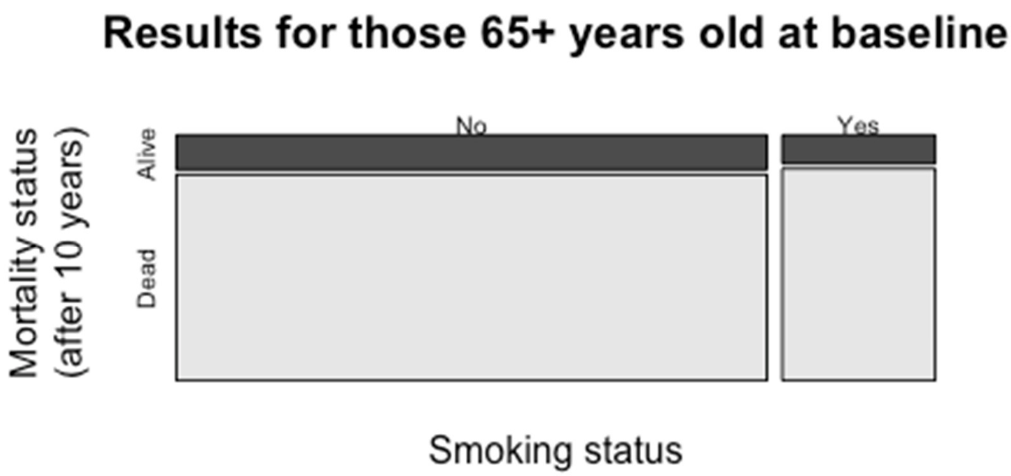


We note that the majority of subjects have survived, but that the number of the smokers who are still alive is greater than we would expect if there were no association between these variables. What could explain this result?

Let's consider stratification by age of the participants (older vs. younger). The following table and figure display the relationship between smoking and mortality over a 10-year period for two groups: those age 18-64 and subjects that were 65 or older at baseline.

```
Baseline age  SMOKER        Alive          Dead
  18-64         No       474 (87.9%)      65 (12.1%)
  18-64         Yes      437 (82.1%)      95 (17.9%)
  65+           No        28 (14.5%)     165 (85.5%)
  65+           Yes        6 (12.0%)      44 (88.0%)
```



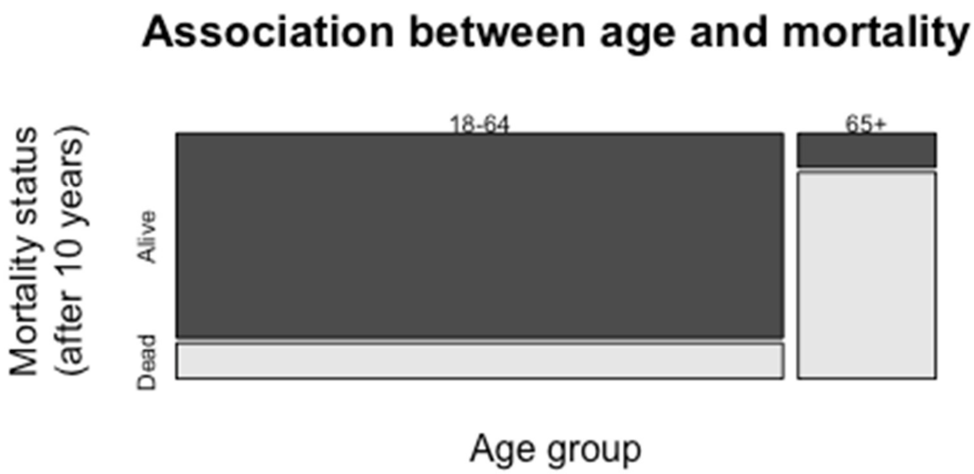


We see that mortality rates are low for the younger group, but the mortality rate is slightly higher for smokers than non-smokers (17.9% for smokers vs 12.1% for the non-smokers).

Almost all of the participants who were 65 or older at baseline died during the follow-up period, but the probability of dying was also slightly higher for smokers than non-smokers.

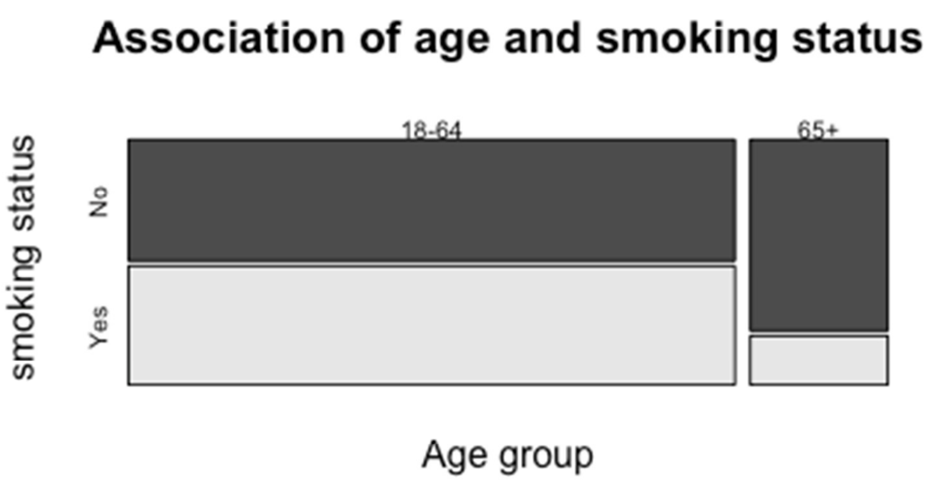
This example represents a classic example of *Simpson's paradox* (Simpson 1951; Norton and Divine 2015). For all of the subjects, smoking appears to be "protective," but within each age group smokers have a higher probability of dying than non-smokers.

How can this be happening? The following figure and table us to disentangle these relationships.



Not surprisingly, we see that mortality rates are highest for the oldest subjects.

We also observe that there is an association between age group and smoking status, as displayed in the following figure and table.



Age roup	Non-smoker	Smoker
18-64	539 (50.3%)	532 (49.7%)
65+	193 (79.4%)	50 (20.6%)

Smoking is associated with age, with younger subjects more likely to have been smokers at baseline.

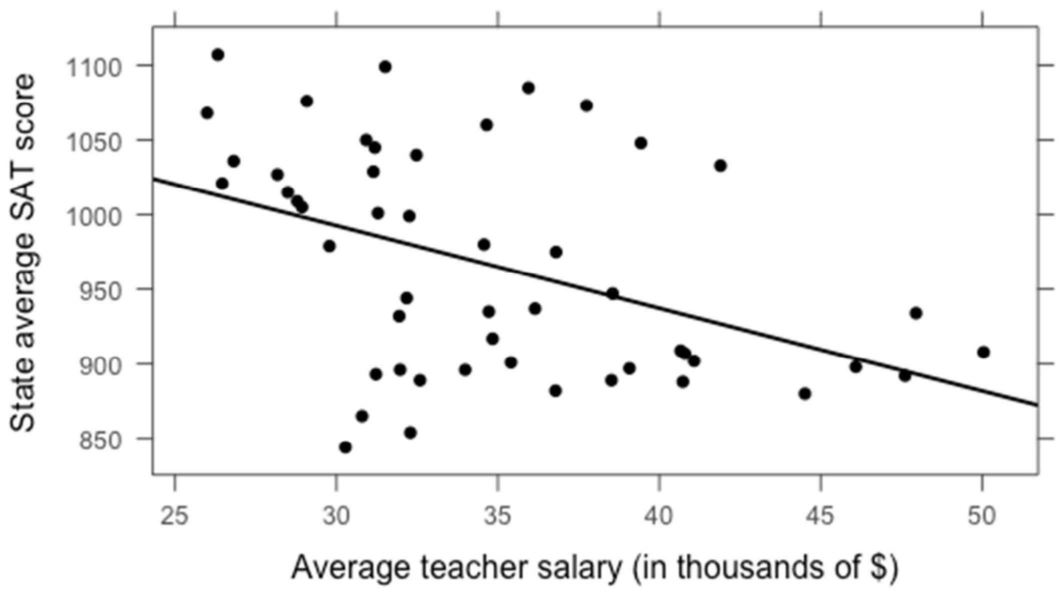
What should we conclude? After controlling for age, smokers have a higher rate of mortality than non-smokers in this study. This other factor is important when considering the association between smoking and mortality.

Simple methods such as stratification can allow students to think beyond two dimensions and reveal effects of confounding variables. Introducing this thought process early on helps students easily transition to analyses involving multiple explanatory variables.

SAT Scores and Teacher Salaries

Consider an example where statewide data from the mid-1990s are used to assess the association between average teacher salary in the state and average SAT (Scholastic Aptitude Test) scores for students (Guber 1999; Horton 2015). These high stakes high school exams are sometimes used as a proxy for educational quality.

The following figure displays the (unconditional) association between these variables. There is a statistically significant negative relationship (β_1 hat = -5.54 points, $p = 0.001$). The model predicts that a state with an average salary that is one thousand dollars higher than another would have SAT scores that are on average 5.54 points lower.

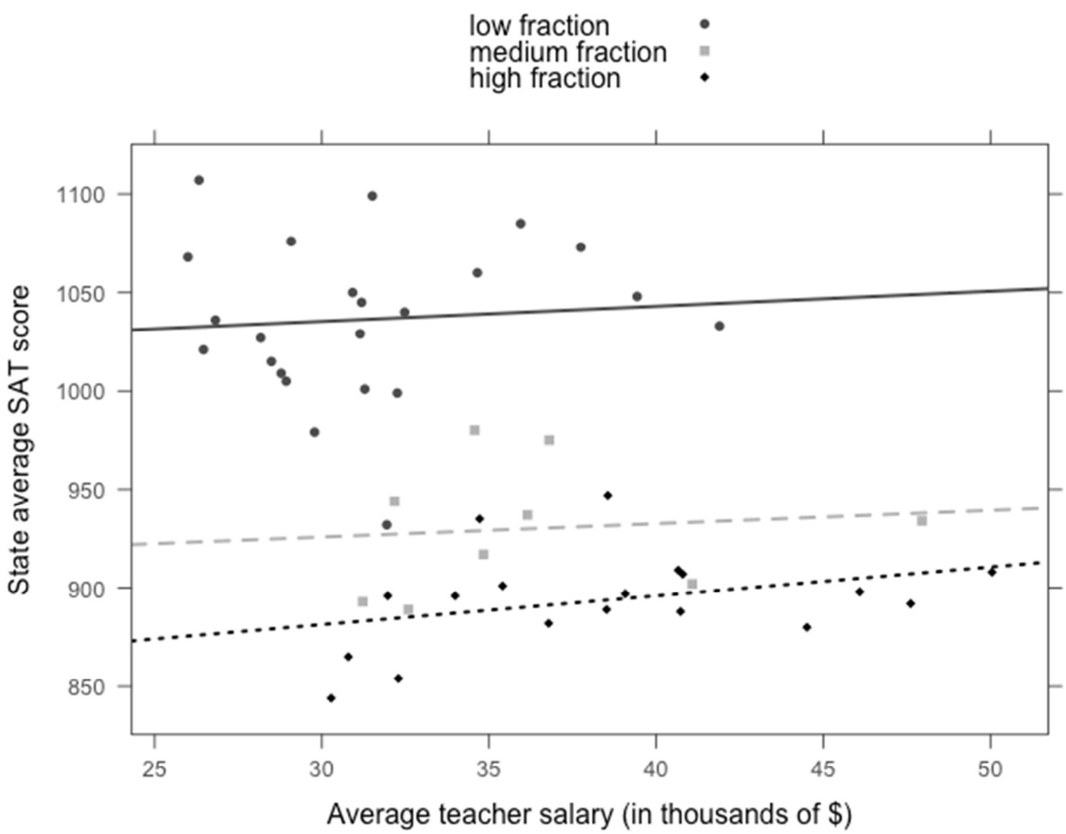


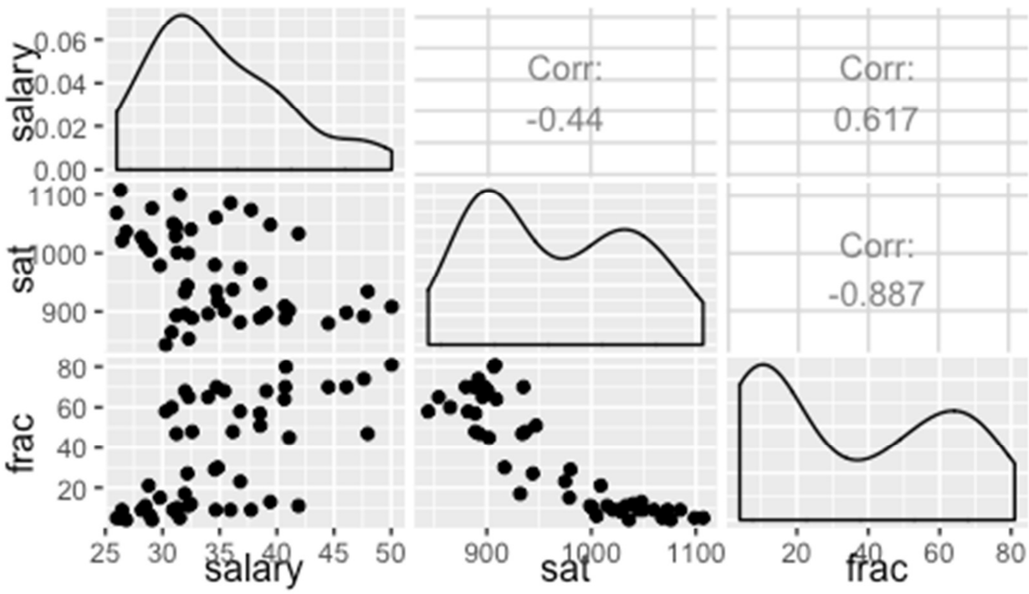
But the real story is hidden behind one of the "other factors" that we warn students about but do not generally teach how to address! The proportion of students taking the SAT varies dramatically between states, as do teacher salaries. In the Midwest and Plains states, where teacher salaries tend to be lower, relatively few high school students take the SAT. Those that do are typically the top students who are planning to attend college out of state, while many others take the alternative standardized ACT test that is required for their state. For each of the three groups of states defined by the fraction taking the SAT, the association is non-negative. The net result is that the fraction taking the SAT is a confounding factor.

This problem is a continuous example of Simpson's paradox. Statistical thinking with an appreciation of Simpson's paradox would alert a student to *look for* the hidden confounding variables. To tackle this problem, students need to know that multivariable modeling exists but not all aspects of how it can be utilized.

Within an introductory statistics course, the use of stratification by a potential confounder is easy to implement. By splitting states up into groups based on the fraction of students taking the SAT it is possible to account for this confounder and use bivariate methods to assess the relationship for each of the groups.

The scatterplot in the next figure displays a grouping of states with 0-22% of students ("low fraction," top line), 23-49% of students ("medium fraction," middle line), and 50-81% ("high fraction," bottom line). The story is clear: there is a positive or flat relationship between teacher salary and SAT score for each of these groups, but when we average over the groups, we observe a negative relationship.





Further light is shed via a matrix of scatterplots (see the above figure): we see that the fraction of students taking the SAT is negatively associated with the average statewide SAT scores and positively associated with statewide teacher salary.

Recall that in a multiple regression model that controls for the fraction of students taking the SAT variable, the sign of the slope parameter for teacher salary flips from negative to positive.

It's important to have students look for possible confounding factors when the relationship isn't what they expect, but it is also important when the relationship is what is expected. It's not always possible to stratify by factors (particularly if important confounders are not collected).

Multiple Regression

The most common multivariable model is a multiple regression. Regression can be introduced as soon as students have seen scatterplots and thought about the patterns we look for in them. When students have access to a statistics program on a computer, they can fit regression analyses themselves. But even without computer access, they can learn about typical regression output. The point is to show students a model involving three (or more) variables and discuss some of the subtleties of such models. Here is one example.

Scottish hill races are scheduled throughout the year and throughout the country of Scotland (http://www.scottishhillracing.co.uk). The official site gives the current records (in seconds) for men and women in these races along with facts about each race including the distance covered (in km) and the total amount of hill climbing (in meters). Naturally, both the distance and the climb affect the record times. So a simple regression to predict time from either one would miss an important aspect of the races.

For example, the simple regression of time versus climb for women's records looks like this:

```
Response variable is: Women's Record
R squared = 85.2%     R squared (adjusted) = 84.9%
s = 1126  with  70-2 = 68  degrees of freedom

Variable        Coefficient     SE(Coeff)     t-ratio    P-value
Intercept       320.528         222.2          1.44        0.1537
Climb             1.755           0.088       19.8       < 0.0001
```

We see that the time is greater, on average, by 1.76 seconds per meter of climb. The R^2 value of 85.2% assures us that the fit of the model is good with 85.2% of the variance in women's records accounted for by a regression on the climb.

But surely that isn't all there is to these races. Longer races should take more time to run. And although an R^2 of 0.852 is good, the model fails to account for almost 15% of the variance.

It is straightforward for students to learn that multiple regression models work the same way as simple regression models but include two or more predictors. Statistics programs fit multiple regressions in the same way as simple ones. Here is the regression with both `Climb` and `Distance` as predictors:

```
Response variable is:   Women's Record
R squared = 97.5%     R squared (adjusted) = 97.4%
s = 468.0  with  70 - 3 = 67  degrees of freedom
Variable      Coefficient  SE(Coeff)     t-ratio   P-value
Intercept     -497.656     102.8         -4.84     < 0.0001
Distance       387.628      21.45        18.1      < 0.0001
Climb            0.852       0.0621      13.7      < 0.0001
```

This regression model shows both the distance and the climb as predictors and has an R^2 of 0.975; a substantial improvement. More interestingly, the coefficient of `Climb` has changed from 1.76 to 0.85. That's because in a multiple regression, we interpret each coefficient as the effect of its variable on y after allowing for the effects of the other predictors.

Closing Thoughts

Multivariable thinking is critical to make sense of the observational data around us.
This type of thinking might be introduced in stages:

1. learn to identify observational studies,
2. explain why randomized assignment to treatment improves the situation,
3. learn to be wary of cause-and-effect conclusions from observational studies,
4. learn to consider potential confounding factors and explain why they might be confounding factors, and
5. use simple approaches (such as stratification) to address confounding.

Multivariable models are necessary when we want to model many aspects of the world more realistically. The real world is complex and can’t be described well by one or two variables. If students do not have exposure to simple tools for disentangling complex relationships, they may dismiss statistics as an old-school discipline only suitable for small sample inference of randomized studies.

Simple examples are valuable for introducing concepts, but when we don't demonstrate realistic models students are left with the impression that statistics is trivial and not really useful. This report recommends that students be introduced to multivariable thinking, preferably early in the introductory course and not as an afterthought at the end of the course.

References

Appleton, D. R., French, J. M., and Vanderpump, M.P. (1996), "Ignoring a Covariate: An Example of Simpson's Paradox," *The American Statistician*, 50, 340-341.

American Statistical Association (2014), *2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science*, Alexandria, VA: Author. Available at http://www.amstat.org/education/curriculumguidelines.cfm.

Guber, D. L. (1999), "Getting What You Pay for: The Debate over Equity in Public School Expenditures," *Journal of Statistics Education*, 7. Available online at http://www.amstat.org/publications/jse/secure/v7n2/datasets.guber.cfm.

Horton, N.J. (2015), "Challenges and Opportunities for Statistics and Statistical Education: Looking Back, Looking Forward," *The American Statistician*, 69, 138–145.

Meng, X.L. (2011), "Statistics: Your Chance for Happiness (or Misery)," *The Harvard Undergraduate Research Journal*, *2*. Available at http://thurj.org/as/2011/01/1259.

Norton, H. J. and Divine, G. (2015), "Simpson's Paradox, and How to Avoid It," *Significance*, 40-43.

Simpson, E. H. (1951), "The Interpretation of Interaction in Contingency Tables," *Journal of the Royal Statistical Society, Series B*, 13, 238-241.

APPENDIX C: Activities, Projects, and Datasets

The GAISE College Report emphasizes the importance of students being actively engaged in their own learning. Activities, projects, and interesting datasets can help instructors engage students. In this appendix, we begin with a description of desirable characteristics of class activities. We provide examples of activities that illustrate a simple two quantitative variable data collection, a randomization test for the difference in two means, experimental design in a matched pairs study, and multivariable thinking. We conclude with examples of datasets and websites that house data.

Desirable Characteristics of Class Activities

In this appendix we focus on activities to be conducted in the classroom. Many of the desirable characteristics described are also applicable to unsupervised activities conducted outside the traditional classroom setting.

Structure and timing…

- Learning Goals – An activity should have clear and attainable learning goals. The activity should build upon what students already know and lead students to discover or explore a statistical concept. Ideally, activities completed early in the course become scaffolding for concepts explored later in the course.

- Self-Contained and Complete – An activity should include all the important statistical concepts, necessary materials, and information from past class activities to complete the activity in a timely fashion.

- Beginning and Ending an Activity – The activity should begin with an overview and end with a summary of what is being done and why. This should include connections that build upon and extend statistical conceptual and methodological knowledge and application, how the statistical analysis helps to answer questions specific to the context of the activity, and what students are expected to learn from the activity.

Choosing data…

- Relevance – The activity should involve data about topics that interest students. Using real data makes data relevant to a wide variety of student majors. If real data are not used, then the activity should mimic a real-world situation. It should not seem like "busywork" to students. For example, if you use coins or cards to conduct a binomial experiment, explain real-world binomial experiments they could represent.

Note: Student interests vary such that a dataset that might be interesting and relevant for one student may not be as interesting or relevant for another student. It is important to use a variety of datasets that speak to students from diverse backgrounds, majors, and interests. One way to gauge student interest is to give the class an option of what dataset to work with in an activity. The choice could be made by student vote or even by using a poll during the first week of class to judge the students' interests and majors.

- Contextual Background – Students should read and be asked questions about the background that informs the context of the data. For example, if the data involves the number of friends a person has on Facebook, then students should read a brief background on some pertinent aspects about Facebook.

If the activity involves collecting/generating data…

- Design Decisions and Data Collection – Activities can include those that require class input into design and data collection and those that are more prescriptive. It is desirable that the class be involved in some of the decisions about how to conduct the activity when time permits and when class learning objectives are advanced.

  Note: When students are involved in construction and implementation of design and data collection decisions, it is important that they invoke good design and data collection principles taught in the class. For instance, when designing an experiment, students should consider principles of good experimental design including randomization, replication, controlling outside factors, etc., rather than "intuitively" deciding how to conduct the experiment.

- Human Subjects Review – Most classroom data collection activities are exempt from the need for review by an Institutional Review Board (IRB). However, students should be made aware of the importance of review when collecting data, especially data on human subjects. Many students will work with an IRB in research methods courses in their own disciplines.

Working in groups…

- Teamwork – Students can learn effectively from each other. While many students are drawn to working in teams, whether formal or informal, some students may resist working with peers. Because working effectively in teams is a highly valued skill in government, industry, and academia that can be practiced in the classroom, instructors should consider requiring some degree of teamwork in activities and projects.

  Note: Appendix F on Learning Environments includes a discussion of the use of and value of cooperative groups in the classroom.

- The Role of Groups in Design Decisions – It is sometimes better to have students work in teams to discuss how to design a statistical investigation and then reconvene the class to discuss how it will be done, but it is sometimes better to have the class work together for the initial design decisions. It depends on how difficult the issues to be discussed are and whether each team will need to carry out data collection in exactly the same way.

Sharing activities…

- Resources for Activities – There are numerous sources for activities including the *Journal of Statistics Education* (http://www.amstat.org/publications/jse/), STatistics Education Web (http://www.amstat.org/education/stew/), the Consortium for Undergraduate Statistics Education (http://www.causeweb.org) , and *Teaching Statistics* (http://www.teachingstatistics.co.uk/).
- Sharing Activities with Other Instructors – For an activity to be easily usable and modifiable, the following characteristics are desirable: (1) quick data collection with low cost in time and resources, (2) available in a file format (such as Word) that makes modification by the instructor easy, and (3) includes a sample answer key for instructors.

Final thoughts…

- In our experience, students enjoy seeing their own data amongst their classmates' data. Activities that collect non-sensitive data from students, either inside class or outside class, perhaps through an online survey, provide this opportunity.

- The activity should be substantive, compelling, and, when possible, fun!

This is a list of desirable characteristics of class activities. This does not imply that an activity that does not meet every characteristic on this list is a poor activity. These characteristics are items to consider when creating, adapting, or using an activity.

Example Activities/Datasets

In-Class Data Collection/Analysis Activities

Example Activity 1: Leg Length and Stride Length

(based on an activity used by John Gabrosek in his classroom)

Exploring bivariate relationships is an important part of an introductory statistics course. In this activity, students investigate whether there is a relationship between the length of a person's leg and the number of steps required to walk a specified distance.

Materials Needed:

- Tape Measures

Procedure:

- Split class into groups of 3 or 4 students
- Have groups measure each student's leg length from outside hip bone to floor
- Simultaneously, have each student walk a specified route. Same route for each student. Instruct students to silently count the number of steps they take as they walk the route.
- Have students enter data into a simple data collection form

Name	Sex (M or F)	Leg Length (inches) – measure right leg	Count of Steps

- Compile class data and use to illustrate scatterplots, correlation, and regression. It is likely that the relationship will be weak and negative.

Instructor Notes:

- The weak relationship lends itself to a discussion of what other variables might impact step count. Students usually identify that different people have different gaits (though they are unlikely to use the term gait). Gait analysis can be used to assess deviations from normal, especially if a person's baseline gait has been analyzed prior to an injury.
- A short distance (no more than a few hundred steps) is sufficient; data collection takes about 5 minutes. Data entry can be done on the spot or by the teacher between class sessions.

Example Activity 2: Randomization Test for a Difference in Means - Cola and Calcium

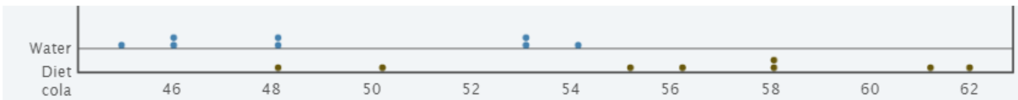
(adapted from Larson, 2010 and Kahn and Laflamme 2015)

Setting: A study by Larson et al. (2010) examined the effect of diet cola consumption on calcium levels in women. A sample of 16 healthy women aged 18-40 were randomly assigned (eight to each group) to drink 24 ounces of either diet cola or water. Their urine was collected for three hours after ingestion of the beverage and calcium excretion (in mg) was measured. The researchers were investigating whether diet cola leaches calcium out of the system, which would

Drink	Calcium
Diet Cola	48
Diet Cola	50
Diet Cola	55
Diet Cola	56
Diet Cola	58
Diet Cola	58
Diet Cola	61
Diet Cola	62
Water	45
Water	46
Water	46
Water	48
Water	48
Water	53
Water	53
Water	54

increase the amount of calcium in the urine for diet cola drinkers. Low calcium levels are associated with increased risk of osteoporosis (Kahn and Laflamme 2015).

Results:



Diet Cola mean $\bar{x}_D = 56.000$ Water mean: $\bar{x}_W = 49.125$

The difference in means is: $\bar{x}_D - \bar{x}_W = 56.000 - 49.125 = 6.875$.

> **Key Question**: Does this difference (6.875) provide *convincing* evidence that the mean amount of calcium excreted after drinking diet cola is higher than after water OR could this difference be just due to *random chance* (in assigning volunteers to the two groups)?

Approach: Simulate new samples generated by random chance and see how often we get a difference as large as (or larger than) what was observed in the original sample (6.875). We will do this first using a physical simulation (by hand), then switch to computer technology to automate the process.

Physical Simulation

1. Start with a sheet of paper that has the 16 calcium amounts from the experiment (such as the table above) and cut/tear the paper so that the numbers are separated from the diet cola/water groups and each value is on its own slip of paper. [Instructor alternative: Put the 16 calcium amounts on individual cards.]

2. Shuffle the slips/cards with calcium amounts and "deal" them randomly into two groups with 8 going to the diet cola group and 8 going to the water group.

3. Find the mean for each group and the difference in the two means.

$\bar{x}_D =$ ________	$\bar{x}_W =$ ________	$\bar{x}_D - \bar{x}_W =$ ________
(Diet cola)	(Water)	(Difference)

Is this difference bigger than the 6.875 from the original sample? ____

4. Look at some of the other simulated differences from your classmates How many of them are bigger than 6.875? [Instructor note: Perhaps draw a class dotplot of differences].

Simulation via technology

[*Instructor note:* Specific instructions below will depend on your technology. See the technology notes below for several options including StatKey (http://lock5stat.com/statkey), a Rossman/Chance applet (http://www.rossmanchance.com/applets), or the R package.]

5. Use technology to simulate the process you just did by hand – scrambling the 16 calcium values and reassigning them to diet cola & water groups.

   Which group got the smallest amount (45)? Diet Cola Water

   Which group got the largest amount (62)? Diet Cola Water

6. Put the difference in means for your simulated sample in the table below, then repeat to generate four more simulations and record the difference in means for each simulation.

   Simulated
   $\bar{x}_D - \bar{x}_W$: ______________ ______________ ______________ ______________

7. Now use the technology to generate a thousand or more simulations. Look at a dotplot (or histogram) of the differences in means for all of these simulations. This is called a *randomization distribution* of the differences and shows what we might expect to see if there really is no difference in calcium excretion between the two groups.

   Where is your randomization distribution centered? _____

   Why does this make sense?

   Does it look like the difference from the actual sample (6.875) is in an unusual place in your randomization distribution?

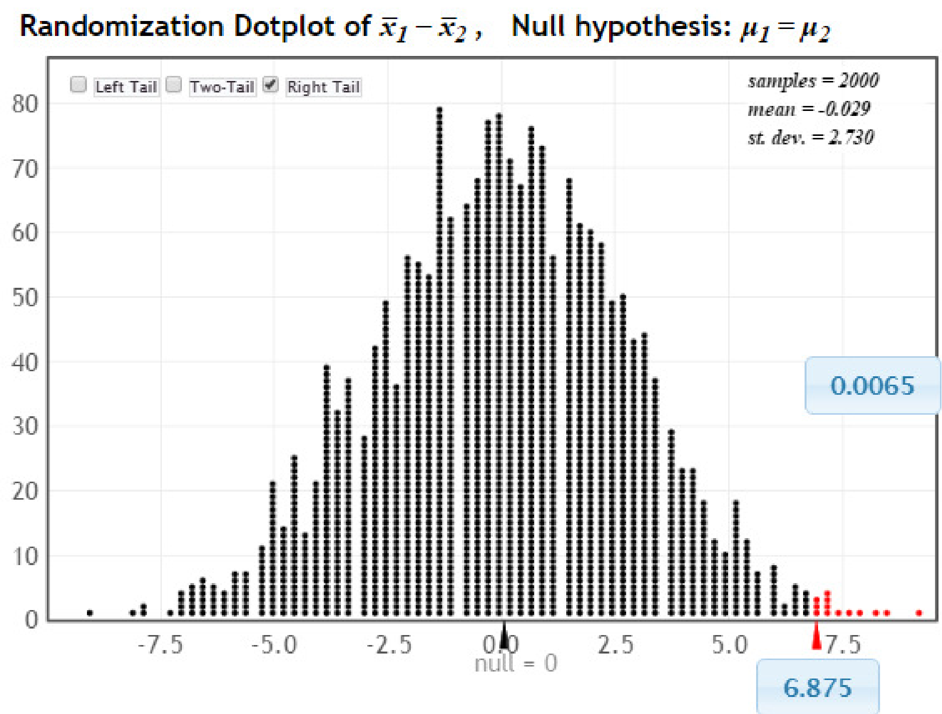
8. To quantify the last question, we will estimate a **p-value** as the proportion of all those random chance samples that have a difference in means as large as (or larger than) the original difference of 6.875. Use technology to estimate the p-value for your randomization distribution.

   What proportion of your randomization differences are 6.875 or larger?

   p-value = ____________________

9. Interpretation: What does this p-value tell you about the "significance" of the difference in the original sample? Does the difference look unusually large (indicating strong evidence that mean calcium excretion tends to be higher after drinking diet cola) or does the difference look more typical of what you would expect to see by random chance alone?

Instructor note: Here is a typical example (from StatKey) of a randomization distribution students might produce in this activity:



Technology notes for the Randomization Activity:
We provide three different technology options (each freely available) for doing the randomizations needed for parts 5-9 of the activity above.

StatKey (available at *http://lock5stat.com/statkey*)

- From the main StatKey page, choose the Randomization “Test for Difference in Means.”
- This dataset is already included in StatKey, so click on the drop down menu (labeled “Leniency and Smiles,” just below the StatKey icon) to bring up a list of datasets and choose “Cola and Calcium excretion.”
- Check that the data and summary statistics shown in the “Original Sample” graph match the data for this activity. Note: For data not already in StatKey, you can use the “Edit Data” button to copy/paste or enter your own data.
- Click on “Generate 1 Sample” to do a single randomization (Step 5 in the activity). The randomized data is displayed and summarized in the bottom right and the difference in means is plotted in the main dotplot at the left. Repeat this for several more randomizations (Step 6).
- Click on the “Generate 1000 Samples” a few times to get a better picture of the randomization distribution (Step 7).

- To find what proportion of the randomizations gave differences as large as the original difference (Step 9) choose the "Right Tail" option, click on the blue box that appears on the horizontal axis, and change the endpoint to 6.875 (the difference in the original sample). The p-value is shown in the box about the right tail.

RossmanChance Applet (available at *http://www.rossmanchance.com/applets*)

- Under "Statistical Inference," choose the option for "two means" (under "Randomization test for quantitative response").
- You need to copy/paste or enter the data from the table in the activity above to replace the default data in the applet. Also, the applet wants the group identifiers to be single words so delete the spaces to change "Diet Cola" to "DietCola" for the first 8 cases.
- Click on "Use Data" and check that the plot and summary statistics match the original sample.
- Click the box next to "Show Shuffle Options" to bring up the controls for the randomizations.
- Leave the number of Shuffles at 1 and click on 'Shuffle Responses" to generate one randomization (Step 5). The shuffled difference is shown below the summary statistics and plotted to the right. Repeat for Step 6.
- Change the number of shuffles to a larger number (like 3000) and "Shuffle Responses" again to generate a histogram of the randomization distribution (Step 7).
- To find what proportion of the randomizations gave differences as large as the original difference (Step 9), fill in that value (6.875) in the box after "Count Samples" leaving the "Greater than ≥" alone. Click on the "Count" button to see the count and proportion.

R (downloadable from *http://www.r-project.org*)

Here is an R script for creating the randomization differences and seeing what proportion are as extreme as the 6.875 difference in the original sample. It uses the nice do( ) function from the *mosaic* package to generate repeated samples without needing a formal loop.

```
#Randomization test to compare two means

library(mosaic)       # Load the mosaic package
library(Lock5Data)    # Load a package with the ColaCalcium dataset
data("ColaCalcium")   # Load the dataset

mean(Calcium~Drink, data=ColaCalcium)   #Compare means for two groups

#Compare means when the Drink values have been randomly permuted
mean(Calcium~shuffle(Drink), data=ColaCalcium)

#collect such simulated means for both groups for 2000 simulations
manymeans=do(2000) * mean(Calcium~shuffle(Drink), data=ColaCalcium)

head(manymeans)       #See some of what the do( ) function collects

#Find the difference in means for each simulation
randomdiffs=manymeans$Diet.cola - manymeans$Water
```

```
dotPlot(randomdiffs,width=0.1)   #get a plot of the random differences

#find the proportion of simulations with differences as large as 6.875
sum(randomdiffs>=6.875)/2000
```

Example Activity 3: Comparing Manual Dexterity under Two Conditions

(Adapted from Project 12.2, Utts and Heckard 2007)

Included in this activity description is

- An Overview of the Activity
- Suggestions for Design and Analysis
- Project Team Form

Overview of Activity

These instructions are for the teacher. Instructions for students are on the "Project Team Form." (below)

Goal: Provide students with experience in designing, conducting and analyzing an experiment.

Supplies: (N = number of students, T = number of teams)

- *T bowls filled with about 30 of each of two distinct colors of dried beans*
- *2T empty paper cups or bowls*
- *T stop watches or watches with second hand*

Instructor Note: A variation is to have students do the task both with and without wearing a "surgical" or "food-service" glove instead of with the dominant and non-dominant hand. In that case you will need *N pairs of gloves*.

The Story: A company has many workers whose job is to sort two types of small parts. Workers are prone to get repetitive strain injury, so the company wonders if there would be a big loss in productivity if the workers switch hands, sometimes using their dominant hand and sometimes using their non-dominant hand. (Or, if you are using gloves, the story can be that for health reasons they might want to require gloves.) Therefore, you are going to design, conduct, and analyze an experiment making this comparison. Students will be timed to see how long it takes to separate the two colors of beans by moving them from the bowl into the two paper cups, with one color in each cup. (To add some context, you can state that each color bean represents an automotive part of a slightly different size – for example, a front door bolt and a back door bolt.) A comparison will be done after using dominant and non-dominant hands. (An alternative is to time students for a fixed time, such as 30 seconds, and see how many beans can be moved in that amount of time.)

Design and Analysis

Step 1: As a class, discuss how the experiment will be done. This could be done in teams first. See below for suggestions.

1. What are the treatments? What are the experimental units?
2. Principles of experimental design to consider are as follows. Use as many of them as possible in designing and conducting this experiment. Discuss why each one is used.
   a. Blocking or creating matched-pairs
   b. Randomization of treatments to experimental units, or randomization of order of treatments
   c. Blinding or double blinding
   d. Control group
   e. Placebo
   f. Learning effect or getting tired
3. What is the parameter of interest?
4. What type of analysis is appropriate – hypothesis test, confidence interval or both? What numerical and graphical analyses are appropriate?

The class should decide that each student will complete the task once with each hand. Why is this preferable to randomly assigning half of the class to use their dominant hand and the other half to use their non-dominant hand? How will the order be decided? Should it be the same for all students? Will practice be allowed? Is it possible to use a single or double blind procedure?

Note: Example 4 below deals with multivariable thinking in data analysis. Study design is an example of multivariable thinking where different variables are controlled so that the relationship between variables of interest can be isolated. For the bean sorting experiment, the matched pairs design controls for student-to-student variability and randomizing the order of dominant/non-dominant hand controls for the learning effect.

Step 2: Divide into teams and carry out the experiment.

The Project Team Form shows one way to assign tasks to team members.

Step 3: Descriptive statistics and preparation for inference

Convene the class and create a plot of the differences. Discuss whether the necessary conditions for any inferential analysis are met. Were there any outliers? If so, can they be explained? Compute the mean and standard deviation for the differences.

Step 4: Inference

Have each team find a confidence interval for the mean difference and conduct the hypothesis test.

Step 5: Reconvene the class and discuss conclusions

Instructor Notes on Design:

- On Step 1

  a. Blocking or creating matched-pairs - Each student should be used as a matched pair, doing the task once with each hand.
  b. Randomization of treatments to experimental units, or randomization of order of treatments - Randomize the order of which hand to use for each student.
  c. Blinding or double blinding - Obviously the student knows which hand is being used, but the time-keeper doesn't need to know.
  d. Control group - Not relevant for this experiment.
  e. Placebo - Not relevant for this experiment.
  f. Learning effect or getting tired - There is likely to be a learning effect, so you may want to build in a few practice rounds. Also, randomizing the order of the two hands for each student will help with this.

- One possible design: Have each student flip a coin. Heads, start with dominant hand. Tails, start with non-dominant hand. Time students to see how long it takes to separate the beans. The person timing can be blinded to the condition by not watching.

Instructor Notes on Analysis:

- *What is the parameter of interest?*
  Define the random variable of interest for each person to be a "manual dexterity difference" of

  d = *number of extra seconds required with non-dominant hand*
  = *time with non-dominant hand − time with dominant hand.*

  Define μ_d = population mean manual dexterity difference.

- *What are the null and alternative hypotheses?*
  $H_0 : \mu_d = 0$ and Ha: $\mu_d > 0$ (faster with dominant hand)
  To carry out the test, compute $t = \frac{\bar{d} - 0}{s_d / \sqrt{n}}$ then compare to the *t*-table or use technology to find the *p*-value.

- *Is a confidence interval appropriate?*
  Yes, a confidence interval will provide information about how much faster workers can accomplish the task with their dominant hands. The formula for the confidence interval is: $\bar{d} \pm t^* \frac{s_d}{\sqrt{n}}$, where t^* is from the *t*-table with df = n -1, and s_d is the standard deviation of the difference scores.

Project Team Form

TEAM MEMBERS:

1. ______________________	4. ______________________
2. ______________________	5. ______________________
3. ______________________	6. ______________________

INSTRUCTIONS:
You will work in teams. Each team should take a bowl of beans and two empty cups. You are each going to separate the beans by moving them from the bowl to the empty cups, with one color to each cup. You will be timed to see how long it takes. You will each do this twice, once with each hand, with order randomly determined.

1. Designate these jobs. You can trade jobs for each round if you wish.
   Coordinator – runs the show.
   Randomizer – flips a coin to determine which hand each person will start with, separately for each person.
   Time Keeper – must have watch with second hand or cell phone timer. Times each person for the task.
   Recorder – records the results in the table below.
2. Choose who will go first. The *Randomizer* tells the person which hand to use first. Each person should complete the task once before moving to the 2nd hand for the first person. That gives everyone a chance to rest between hands.
3. The *Time Keeper* times the person, while they move the beans **one at a time** from the bowl to the cups, separating colors.
4. The *Recorder* notes the time and records it in the table below.
5. Repeat this for each team member.
6. Each person then goes a second time, with the hand *not* used the first time.
7. Calculate the difference for each person.

NAME:	Time for *non-dominant* hand	Time for *dominant* hand	*d* = difference = non-dominant – dominant hand

RESULTS FOR THE CLASS:
Record the data here____
Parameter to be tested and estimated is ________________________________
Confidence interval____
Hypothesis test – hypotheses and results_______________________________

Example Activity 4: Multivariable Thinking in the Analysis of Diamond Pricing

(adapted from De Veaux, 2015)

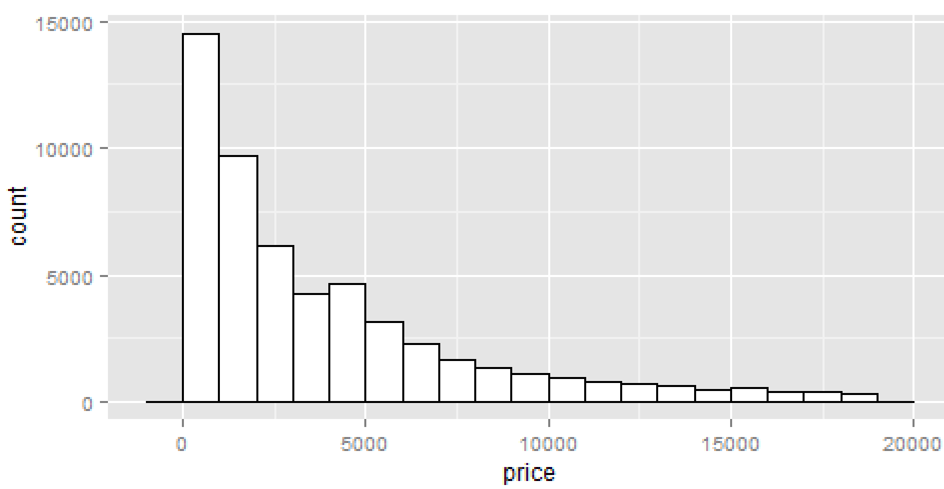
Goal: Provide students with experience investigating a dataset where the relationship between variables is conditional upon other variables. In this write-up, we use the R program and the ggplot2 package to analyze the data. We place any graphs that students should create in the body of the report. The same analysis could be done in SAS, SPSS or any other statistical software program.

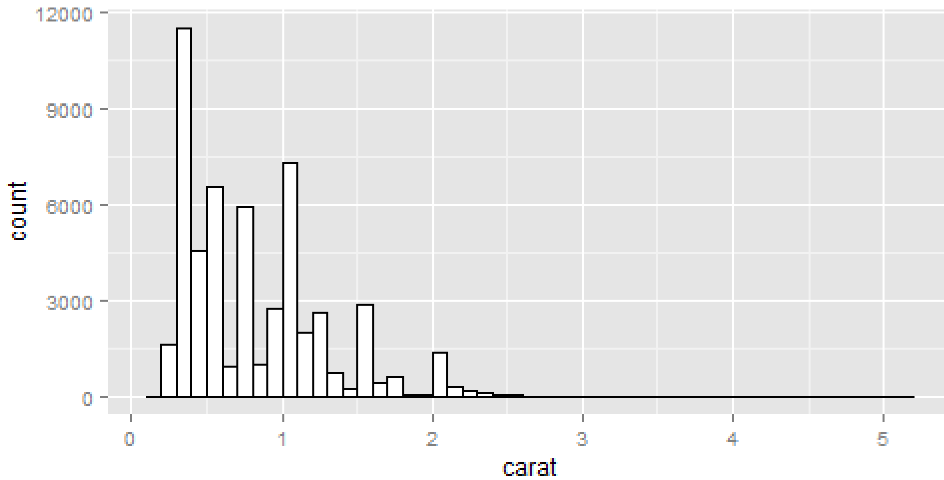
Data: This activity uses the diamonds dataset that is part of the ggplot2 R package (*http://www.inside-r.org/packages/cran/ggplot2/docs/diamonds*). The dataset includes information on 53,940 diamonds. There are ten variables measured on each diamond including price (in U.S. Dollars), cut (quality of the cut), clarity (a measurement of how clear the diamond is), color, and carat (weight from 0.2-5.01 carats). For a full description of the dataset open R and then enter code: help(diamonds).

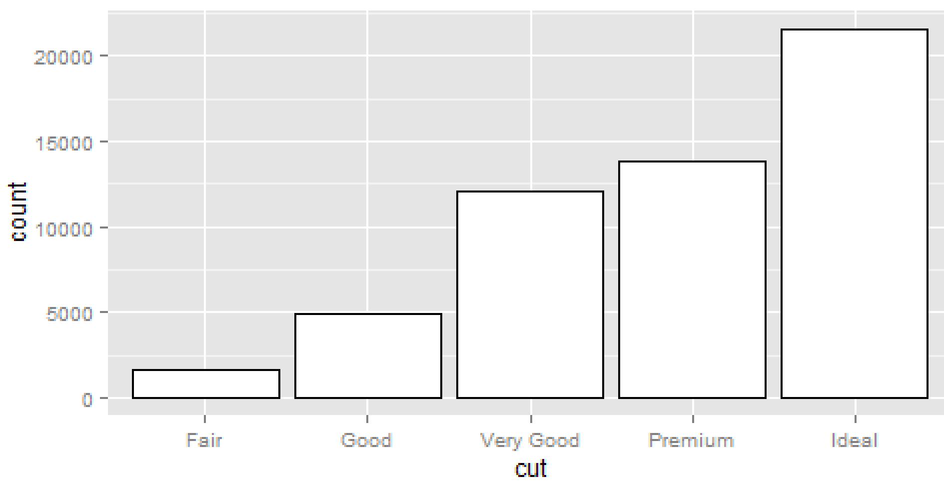
The Story: Diamond prices depend on the four C's of a diamond; cut, clarity, color, carat. It is pretty obvious that bigger diamonds cost more, or is it? In this activity you investigate the relationship between cut and price of diamonds using a dataset that includes 53,940 diamonds.

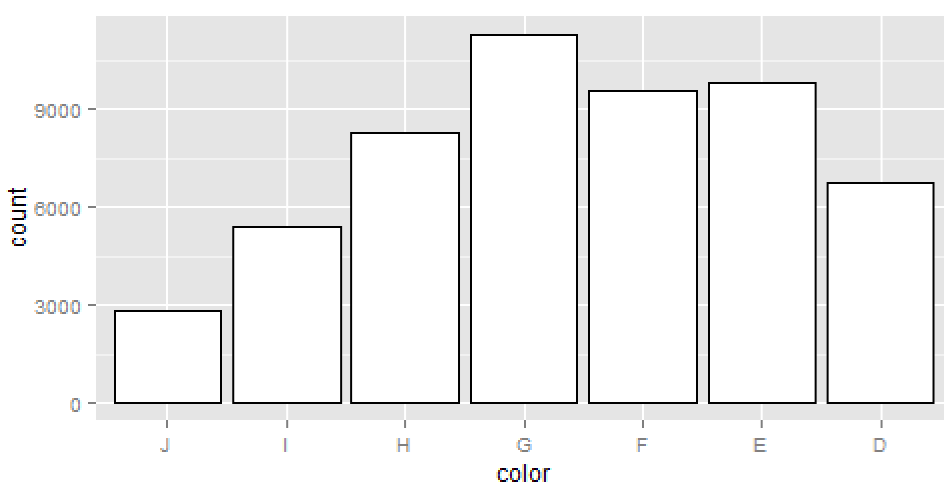
Part 1: Univariate Analysis

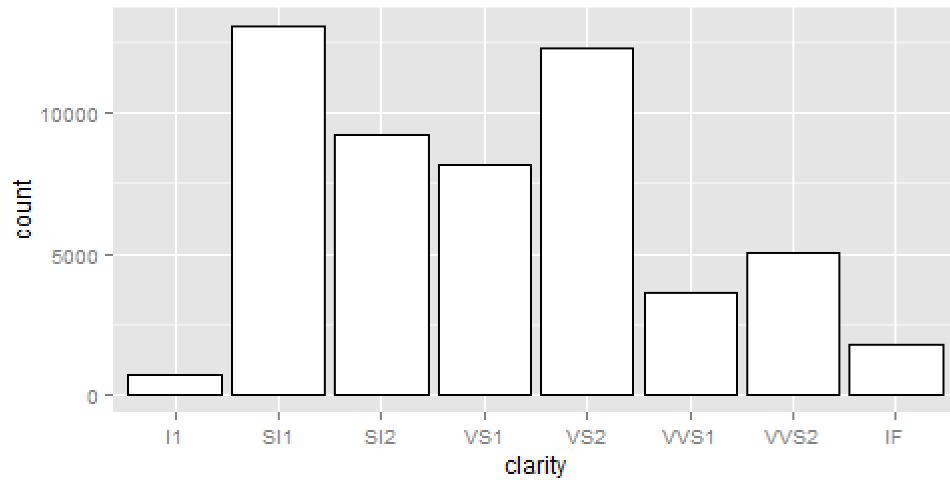
1. Make an appropriate graph for each of the five variables; price, cut, clarity, color, and carat. For the categorical variables, be sure that the categories are placed in a logical order from worst to best.




count
carat
0
3000
6000
9000
12000
0
1
2
3
4
5


count
cut
0
5000
10000
15000
20000
Fair
Good
Very Good
Premium
Ideal


count
color
0
3000
6000
9000
J
I
H
G
F
E
D

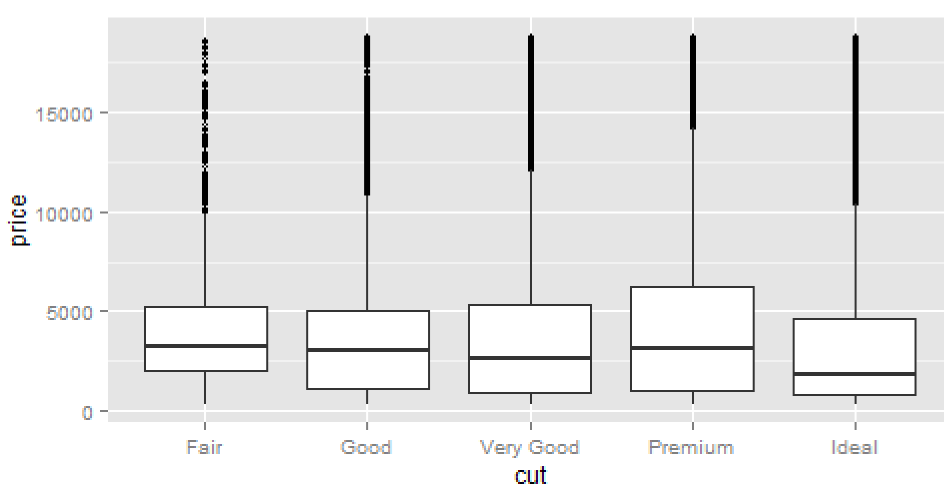


2. Describe any interesting features of the distribution of each variable.

   Students should point out that: (1) price is unimodal, peak from \$0-\$1000 and very skewed right; (2) carat is very choppy with a peak around 0.1-0.2 carats and then a smaller peak at around 1 carat and is skewed right; (3) most diamonds are of at least very good cut; (4) color has large variability with numerous diamonds of a lower color quality (left of G) and numerous diamonds of a higher color quality; and, (5) relatively few diamonds are of very high clarity (far right side of graph).

Part 2: Bivariate Analysis – Your goal is to investigate the relationship between a diamond's cut and the price of the diamond.

3. Make an appropriate graph to investigate the relationship between price and cut. Describe what you see. Is there anything surprising?



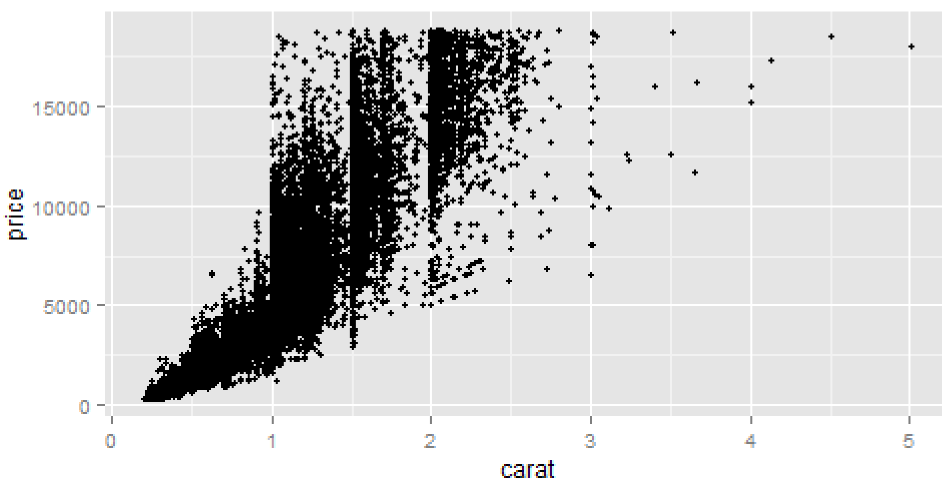
Students should point out that the median price for ideal cut diamonds is less than any of the other cuts of diamonds. This does not make sense because ideal cut is the highest quality cut possible.

Part 3: Multivariable Thinking

4. You should have noticed that ideal cut diamonds tend to have lower prices than any other cut. The median price for ideal cut diamonds is $1810, while fair cut diamonds have median price $3282. Brainstorm with a partner some ideas on why this might be true.

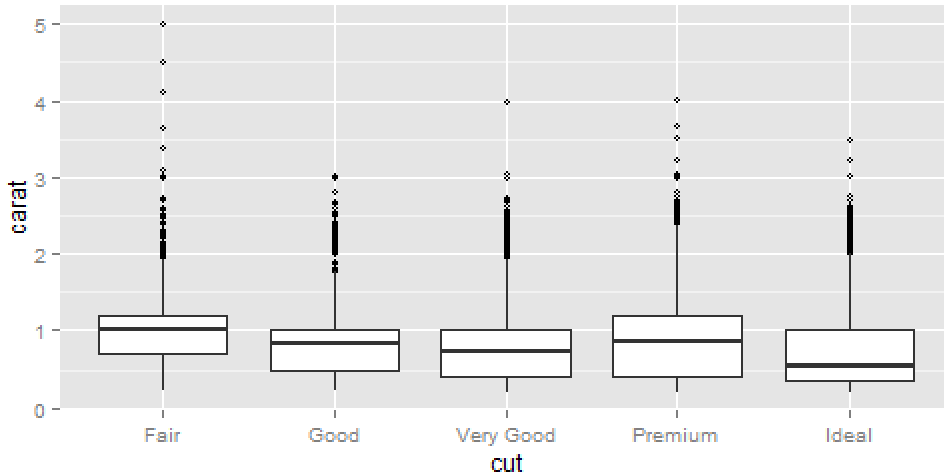
5. Now that you have brainstormed some ideas, let us see if the data can help us.

   a. First, make a scatterplot of price against carat. Describe what you see.



As expected there is a positive relationship between carat and price, with higher carat generally associated with higher price.

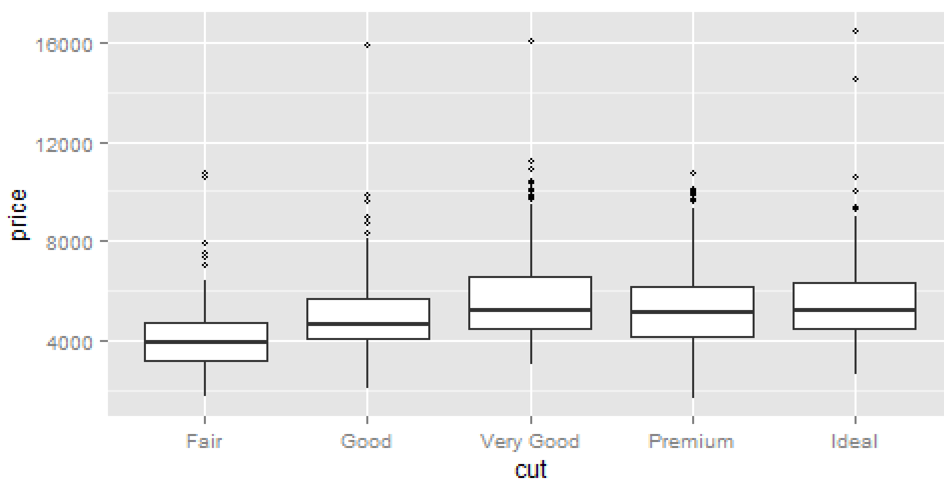
   b. Make an appropriate graph to investigate the relationship between carat and cut. Describe how cut is related to carat.



Fair cut diamonds tend to be much larger than ideal cut diamonds. Basically, it is very difficult to find a large diamond that can be cut as perfectly as necessary for an ideal cut designation.

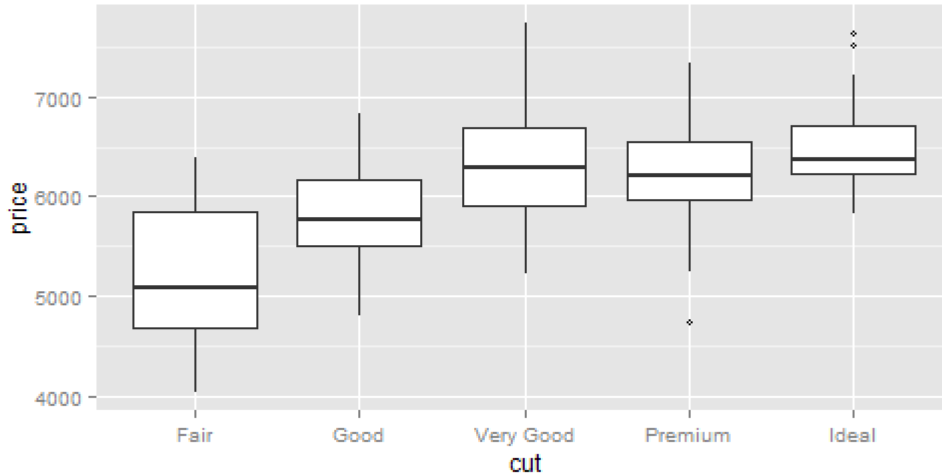
6. Now, let us look at only diamonds of size 1 carat.

   a. Below is a plot of price broken down by cut for these diamonds. What do you see?



For 1 carat diamonds, the price tends to increase as the cut quality improves. But, we have not accounted for the other variables color and clarity.

   b. Now let us take our 1 carat diamonds and only look at those of color = G or H and clarity = VS1 or VS2. There are 22 Fair, 53 Good, 64 Very Good, 82 Premium, and 27 Ideal diamonds meeting these conditions. What do you see in the plot?



When you control for carat, color, and clarity, then, as expected, fair cut diamonds are priced much less than ideal cut diamonds. The price of a diamond now seems to follow the cut.

7. What does this activity tell you about investigating the relationship between two variables?

Note that this example and others like it are included in the STAT101 toolkit for instructors available at http://community.amstat.org/stats101/home).

Examples of Naked, Realistic and Real data

One of the core recommendations of this report and its predecessor is to "*Use real data.*" The next few small examples illustrate a continuum along a spectrum of "reality," starting with data having no context at all and progressing to data from an actual study designed to address a question of interest in a particular field. The task at hand (fit a least squares line) is the same in each case, and to help illustrate the distinctions, we have kept the number of data cases small in each situation. In practice, electronic access to data and technology for doing graphics and analysis frees us from restrictions of using such small datasets.

Naked data (not recommended)

Find the least squares line for the data below. Use it to predict Y when X=5.

X	1	2	3	4	6	8
Y	3	4	6	7	14	20

Critique: Made-up data with no context (not recommended). The exercise is purely computational with no possibility of meaningful interpretation.

Realistic data (better, but not recommended)

The data below show the number of customers in each of six tables at a restaurant and the size of the tip left at each table at the end of the meal. Use the data to find the least squares line predicting the size of the tip from the number of diners at the table. Use your result to predict the size of the tip at a table that has five diners.

Diners	1	2	3	4	6	8
Tip	$3	$4	$6	$7	$14	$20

Critique: A context has been added which makes the exercise slightly more appealing and shows students a practical use of statistics. The actual data values are made-up and example feels somewhat contrived.

Real data (better but not recommended)

The data below show the quiz scores (out of 20) and the grades on the midterm exam (out of 100) for a sample of eight students who took this course last semester. Use these data to find a least squares line for predicting the midterm score from the quiz score.

Assuming that the quiz and midterm are of equal difficulty this semester and the same linear relationship applies this year, what is the predicted score on the midterm for a student who got a score of 17 on the quiz?

Quiz	20	15	13	18	18	20	14	16
Midterm	92	72	72	95	88	98	65	77

Critique: While data are from a real situation that should be of interest to students taking the course, and the question asked may be relevant for the situation, the data do not provide a compelling application of statistics.

Real Data, from a Real Study (preferable)

In a study of honeybees, Seeley (2010) observed that scout bees do a "waggle dance" to help communicate the distance to a new nest site to bees back in the original nest. The table below shows the distance to the new site (in meters) and duration of the dance (in seconds) recorded for seven different scout honeybees. Use the data to find the least squares regression line and predict the distance to a new nest when a honeybee dances for 1.5 seconds.

Duration (seconds)	0.40	0.45	0.95	1.30	2.00	3.10	4.10
Distance (meters)	200	250	500	950	1950	3500	4300

Critique: While the dataset is very small and only contains two variables (could there be other factors at play?) the data arise from a real study (with reference) and address a real and compelling research question.

Data (with Background and Stories) Available on the Web

EXAMPLE 1: AMES, IOWA REAL ESTATE DATA

The paper and dataset by DeCock (2011) describes sale of residential properties in Ames, Iowa from 2006 to 2010. The dataset contains 2930 observations of home sales and 80 variables. The data lends itself to a variety of analyses that can be done at the introductory statistics level (regressing y = sales price on x = square footage is one example) and at a more advanced modelling level.

While the paper does not contain a complete, ready-to-go activity, the author describes in detail potential uses of the dataset. He provides helpful hints to employ and potential pitfalls to avoid when using the dataset. It is quite easy to construct a simple activity that utilizes the data to illustrate concepts in regression.

The dataset meets many of the desirable characteristics listed previously, including:

- Data Relevance - The dataset is real data that represents actual real estate sales.
- Contextual Background - Understanding contextual background is important to understanding the data. Students need to be made aware of what different variables mean (the paper includes a documentation file with detailed variable descriptions) to be able to realistically model sales prices.
- Team Work - The richness of the dataset lends itself to use in a semester-long project that can best be done in teams working together. This is especially true if teams are tasked with developing a best regression model from the more than 70 potential predictor variables.

Teacher Hints:

1. The dataset is rich enough for many uses. In a regression modeling course students could be given the dataset and asked to find an appropriate model to predict sales price. In an introductory statistics course Sales Price can be summarized using basic numerical and graphical techniques. Pairs of variables can be used to discuss correlation, two-way tables, etc.
2. The introductory statistics instructor may want to work with a smaller subset of the variables so as not to overwhelm the students.

3. Instructors might use the Ames dataset and article as a template for collecting (or having students collect) similar data from the local area.

Note: Appendix D on Technology includes a further discussion of the Ames, Iowa real estate dataset.

Example 2: U.S. Road Location Data

The paper by Stoudt et al. (2014) describes a lesson to randomly sample points in the continental United States, determine whether or not each point is within one mile of a road, and use the sample data to infer the proportion of the continental United States that is within one mile of a road. The paper requires use of the R programming language and Internet access.

As with all papers published on the STEW website, the paper includes a complete, ready-to-go activity.

The dataset meets many of the desirable characteristics listed previously, including:

- Data Relevance - The dataset is real data collected in real-time by the students to answer a question of importance to biologists, natural resource managers, and others concerned with providing habitat for plants and animals.
- Design and Data Collection – Students collect data to answer an important question. Using simple tools available on the Internet students are able to quickly and accurately collect data to address the question.
- Team Work – The data collection allows for students to collect roughly 20 data points in a class period. Students see that by pooling their data collection results they are rewarded with a more precise interval estimate of the proportion of the U.S. within one mile of a road.

Teacher Hints:

1. Students use the latitude, longitude coordinates of a point to do two things: (i) determine if the point falls within the continental U.S. and (ii) assuming the point is within the continental U.S., determine whether or not the point is within one mile of a road.
2. Students will generate points that do not fall in the continental U.S. (points in the Pacific Ocean, Atlantic Ocean, and Mexico are common). Because of this students will have unequal sample sizes unless instructed to continue generating points until 20 are within the continental U.S.
3. Students can edit their data file in Excel and then read back into R for the analysis if desired.

Websites with Data

There are numerous websites that have freely available data. Data formats vary, but it is usually simple to convert one of the datasets found at these sites to work with software available to the instructor. The complexity of the data and the amount of processing (*i.e.*, data cleaning) to get the data ready for classroom use varies greatly. The list below provides a few places where an instructor can get data.

- Journal of Statistics Education (JSE) - http://www.tandfonline.com/loi/ujse20 The Data Sets and Stories section of the journal includes data (often in several formats and a documentation file explaining variable names).
- Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) - https://www.causeweb.org/ The CAUSE website includes links to hundreds of locations for data. On the Home page in the upper right corner type "Datasets" in the Search field.
- New York City Open Data - https://data.cityofnewyork.us/ More than 1000 datasets on various aspects of life in the Big Apple. You can search a specific term or click to view all available datasets.
- Winner data - http://www.stat.ufl.edu/~winner/datasets.html Larry Winner from the Department of Statistics at the University of Florida has amassed hundreds of datasets. Each dataset includes a description.
- OzDASL - http://www.statsci.org/data/multiple.html Datasets categorized by statistical topic.
- Revolution Analytics R Data - http://mran.revolutionanalytics.com/documents/data/ Links to various data sources.
- Kaggle - https://www.kaggle.com/ Website that hosts data analysis competitions. Many datasets here are quite large and very messy. Establishing a free account is necessary for access to the data.

Note: Appendix D on Technology includes other sources of data available on the web.

An additional source of data is from published papers. You can contact the author and journal asking for permission to use a dataset in teaching. Many authors and journals will grant permission for educational purposes and provide you with the dataset.

References

DeCock, D. (2011), "Ames, Iowa: Alternative to the Boston Housing Data as an End of Semester Regression Project," *Journal of Statistics Education*, 19. Available at http://www.amstat.org/publications/jse/v19n3/decock.pdf

De Veaux, Dick (2015) "What Makes Diamonds so Expensive?" *Stats 101 Public Library.* Available at http://community.amstat.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=b13124e7-09aa-a775-fb51-6639d1d457ca&forceDialog=0

Kahn, A. and Laflamme, M. R. (2015), “Calcium Deficiency Disease,” *Heathline*. Available at www.healthline.com/health/calcium-deficiency-disease.

Larson, N. S., Amin, R., Olsen, C., and Poth, M. A. (2010), “Effect of Diet Cola on Urine Calcium Excretion,” *Endocrine Reviews – Endo 2010 Abstracts*, 31, S1070.

Seeley, T. (2010), *Honeybee Democracy*, Princeton University Press, Princeton, NJ, p. 128.

Stoudt, S., Cao, Y., Udwin, D., and Horton, N. J. (2014), “What Percent of the Continental US is Within One Mile of a Road?” *STatistics Education Web*. Available at http://www.amstat.org/education/stew/pdfs/PercentWithinMileofRoad.pdf

Utts, J. and Heckard, R. (2007), *Mind on Statistics – Instructor’s Resource Manual*, Duxbury Press, Belmont, CA.

APPENDIX D: Examples of Using Technology

This appendix introduces different forms of technology that can be used to fulfill the GAISE recommendation to "Use technology to explore concepts and analyze data." Additionally, these forms of technology can help us to achieve some of the other GAISE College Report recommendations, such as stressing conceptual understanding, gathering and using real data, and fostering active learning.

Because technology changes so quickly and access to forms of technology varies from instructor to instructor, we will be highlighting how certain methods can be used to meet the recommendation; specific guidelines for particular brands or forms of technology will be avoided. A special effort has been made to keep the material current and relevant in light of constant advancements in technology.

When considering the use of technology in the classroom, the instructor should first start with the learning goal or learning objective and then carefully consider what forms of technology could be used to best meet that goal or objective. Next, the students in the classroom must be considered. What form of technology would students learn most quickly (if needed)? How much training would be necessary to allow students to seamlessly engage with the technology? It is important to pick technology that does not become an additional burden for students or that hinders them further from meeting goals or objectives.

In this appendix, we will focus on the following:

1. Interactive Applets
2. Statistical Software
3. Accessing Real Data online (Observational, Experimental, Survey)
4. Using Games and Other Virtual Environments
5. Real Time Response Systems

Using Interactive Applets

Interactive applets can be used to emphasize important statistical concepts without being encumbered by lots of calculations. [35] It's important to note, however, that free applets vary widely in terms of support, maintenance, and compatibility with evolving technology platforms. We recommend that instructors test applets on classroom systems each time they plan to use them in the classroom.

Applets are available that focus on a wide array of topics. To list just a few, there are applets that are available for using randomization and bootstrap techniques to conduct inference, for

[35] There are many places on the web that house statistical applets, and an internet search on a statistical concept can often generate a few openly available applets. Often times, applets are also available with certain textbooks.

discussing sampling distributions of the sample proportion and the sample mean, and for demonstrating the concept of "confidence" with confidence intervals. Additional applets can help students recognize the effect of outliers on the simple linear regression equation as well as the effect of outliers on the values of measures of center and variability. There are also applets that can simulate probabilities taken over the long run.

Applets can be used in many ways. As an example, applets can be the focus of a class demonstration, or they can be used by students as part of a homework assignment, a computer lab activity, a class project, a quiz, or an exam. Applets can used by a single student at a time, as a team/partner activity, or by the whole class at once.

Best Practices and Ideas found in Statistics Education Literature

- Applets work well with the query first method. This means that the students try to answer the conceptual questions first on their own and then again after using the applet.
    - To see more information, see the following article:
      delMas, R., Garfield, J., and Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistical Education*, *7*. Available at http://www.amstat.org/publications/JSE/secure/v7n3/delmas.cfm.

- In the case of an applet that uses the concept of repeated sampling for randomization tests or bootstrapping techniques, first sample one at a time, and then stop to explain what is being illustrated. You may need to take another sample and explain the process again. After the students appear to understand, you can then increase the number of samples to 1000 or a higher value.
    - For an example of this process, see the following article:
      Lock Morgan, K., Lock, R., Lock, P.F., Lock, E., and Lock, D. (2014), "StatKey: Online Tools for Bootstrap Intervals and Randomization Tests," *ICOTS9 Proceedings* [online]. Available at http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9B2_LOCKMORGAN.pdf

- Pick applets that make it easier to focus on the concepts and to help introductory students experience the entire investigative process. For example, the simulation should be similar to a physical method students could use to illustrate a concept, for example, by using cards or coins. Additionally, the simulation should allow for easy transition to multiple types of inference (e.g., from inference about difference between two independent proportions to difference between two independent means).
    - To see more about this and how simulation-based inference is changing the modern curriculum, see the following articles:
      Rossman, A., and Chance, B. (2014), "Using Simulation-based Inference for Learning Introductory Statistics," *WIREs Computational Stat*, *6*, 211-221.

    - Tintle, N., Chance, B., Cobb, G., Roy, S., Swanson, T., and VanderStoep, J. (2015), "Combating Anti-statistical Thinking using Simulation-based Methods through the Undergraduate Curriculum," *The American Statistician, 69(4)*, 362-370.

Future Direction of Applets and Interactive Visualizations

In the past, the statistics education community has mostly relied on a handful of people and organizations to provide applets to help students build conceptual understanding. However, it is becoming easier and easier to design one's own statistical applets and other interactive visualizations, and soon, instructors will be able to use readily available open software to create their own public interactive visualizations. For example, Shiny, a web application framework for R, allows the user to turn displays and analysis into interactive web applications (some level of familiarity with R is needed).

- For an overview of using R and Shiny in the classroom, see Doi, J., Potter, G., Wong, J., Alcaraz, I., and Chi, P. (2016), "Web Application Teaching Tools for Statistics using R and Shiny," *Technology Innovations in Statistics Education* 9(1), available at http://escholarship.org/uc/item/00d4q8cp
- To see examples of some interactive visualizations: see http://shiny.rstudio.com/gallery/
- For more information about writing code for these visualizations: see http://www.r-bloggers.com/interactive-visualizations-with-r-a-minireview/

Even instructors who do not have the time or desire to create their own visualizations might find the list of example visualizations under the Shiny gallery a way to bridge some of the gap between what students may traditionally see in an introductory course and the real data they may see outside of the classroom (*e.g.*, movie reviews, airline data, bus route data, etc.).

- For further discussion of data for the modern student, see the following article: Gould, R., (2010), "Statistics and the Modern Student," *International Statistical Review, 78*, 297–315.

Example 1: Using Statistical Applets to perform Randomization and Bootstrapping Techniques

Since the first GAISE College Report, more and more introductory courses have been incorporating randomization and bootstrapping techniques into the curriculum, and one way in which these techniques can be incorporated into a course is by using statistical applets[36]. To see an example of this type of activity, please see Appendix C in this report.

Example 2: Creating a Story Board, Video or Cartoon about Findings from an Applet

The following is a sample handout that can be used or modified in conjunction with an activity involving the use of an applet.

Student Handout

[36] Two free websites that can be used to do this are the Rossman and Chance website (http://www.rossmanchance.com/applets/) as well as the Lock 5 website (http://www.lock5stat.com/statkey/index.html). Some statistical software packages do this as well, such as Statcrunch, R and JMP.

In class, we have been talking about confidence intervals for the population mean. What does the term "confidence" mean? A link to an applet about confidence intervals has been provided by your instructor. With your teammates, explore the applet and determine what it is trying to demonstrate.

You should be able to answer the following three questions:

1.) If you were to take 100 different random samples and construct 100 95% confidence intervals for the population mean, would each of the intervals be exactly the same -- having exactly the same upper and lower bound? Explain your reasoning. Why would or wouldn't they be the same?

2.) For these intervals, what do we know about the population mean in relation to those 100 confidence intervals?

3.) What does it mean to be 95% confident?

Now that you understand the simulation, do one of the following activities:

- Create a script for a two-minute educational video that explains what is happening in the applet. The audience of the educational video should be people who have not taken a statistics course.
- Imagine that you have been given the opportunity to create a cartoon about statistics for the college newspaper. Create a cartoon demonstrating the concept of "confidence."
- Create a quick two minute video using a free online recording program[37] that explains what is being demonstrated in the applet.

Inspiration:

Fewster, R. (2014), "Teaching Statistics to Real People: Adventures in Social Stochastics," *ICOTS 9 Proceedings.* Available at http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_PL3_FEWSTER.pdf .

Turner, S., and Dabney, A. (2015), "A Story Based Simulation for Teaching Sampling Distributions," *Teaching Statistics, 37*, 23-25.

Teaching Note:

It's important to think carefully about how long students should spend on this task. You might want to give them a timeline of one class period so they can focus more on the statistical concepts and less, for example, on perfecting a two-minute recording.

Example 3: Exploring Misconceptions about Sampling Distributions by using an Applet

Here is an example student handout that can be used or modified in conjunction with an applet that demonstrates the sampling distribution.

[37] For example, Jing! (www.techsmith.com/jing.html) or Screencast-O-Matic (www.screencast-o-matic.com).

Before using the applet, answer the following questions. For these questions, write down what you think is the best answer. Please write these down in pen, so that you can't change these answers. As you complete this activity you might find out that these ideas have been confirmed or are incorrect, that is okay. If they are incorrect, it is important to see why they are incorrect and to identify them correctly later on. Seeing mistakes and misconceptions is important so that you remember them later on.

PART ONE

Sketch the graph of each of these.

Normal	**Sampling Distribution of the sample mean with *n* = 10, where the original population was Normal**	**Sampling Distribution of the sample mean with *n* = 100 where the original population was Normal**

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?
4.) Very briefly explain your thinking: *Why* do you anticipate these specific descriptions?

Sketch the graph of each of these.

Right Skewed	**Sampling Distribution of the sample mean with *n* = 10, where the original population was Right Skewed**	**Sampling Distribution of the sample mean with *n* = 100 where the original population was Right Skewed**

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?
4.) Very briefly explain your thinking: *Why* do you anticipate these specific descriptions?

Sketch the graph of each of these.

Uniform	**Sampling Distribution of the sample mean with *n* = 10, where the original population was Uniform**	**Sampling Distribution of the sample mean with *n* = 100 where the original population was Uniform**

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?
4.) Very briefly explain your thinking: *Why* do you anticipate these specific descriptions?

PART TWO

Now go to the interactive applet website[38] given to you by your instructor. Complete the table below.

	Shape	Mean	Standard deviation

[38] There are many options for applets that may work. Some textbook publishers include applets with their textbooks. Applets can also be found within some statistical packages like StatCrunch or even openly available on the internet. Since instructor resources will vary and websites change rapidly, specific websites are not given here. When choosing an applet, make sure that you pick an applet that allows students to easily change the sample size and the population distribution. It should also allow the instructor to show the results one sample at a time, a few samples at a time, and then many samples at a time.

Normal Distribution			
Sampling Distribution of the sample mean with n = 10 where the original population was Normal.			
Sampling Distribution of the sample mean with n = 100 where the original population was Normal			

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?

Complete the table below.

	Shape	Mean	Standard deviation
Right Skewed Distribution			
Sampling Distribution of the			

sample mean with *n* = 10 where the original population was Right Skewed			
Sampling Distribution of the sample mean with *n* = 100 where the original population was Right Skewed			

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?

Complete the table below.

	Shape	Mean	Standard deviation
Uniform Distribution			
Sampling Distribution of the sample mean with *n* = 10 where the original population was Uniform.			

Sampling Distribution of the sample mean with *n* = 100 where the original population was Uniform.			

1.) Describe the centers of the distributions. Are they the same? Different? In what way?
2.) Describe the variability of the distributions. Are they the same? Different? In what way?
3.) Describe the shape of the distributions. Are they the same? Different? In what way?

REFLECTIONS

Where did you see that you were initially correct?
Where did you see that you were initially incorrect?
What did you learn during this experience about what happens to the sampling distribution of the sample mean as n increases?

‘

Inspiration:

delMas, R., Garfield, J., and Chance, B. (1999), "A Model of Classroom Research in Action: Developing Simulation Activities to Improve Students' Statistical Reasoning," *Journal of Statistical Education, 7.* Available at http://www.amstat.org/publications/jse/secure/v7n3/delmas.cfm.

Teaching Notes:

- It is important that students understand they don't have to be correct on the first attempt. They should do their best to think about the issues involved, but they are not required to be correct. In fact, if they discover a misconception, it gives them something to write about in the reflection part of the experience. During the lab activity, the instructor should walk around the room and help students discover their own misconceptions. Perhaps even a few students can share their misconceptions so that other students might realize they also had those misconceptions.

- It is helpful if you give out this activity in two parts. First, give out part one and have students complete it in pen; then, give out part two. It is also helpful, when possible, if part one and part two can be copied onto different colored sheets of paper.

Teaching Concepts and Analyzing Data with Statistical Software

It is self-evident that statistical software can eliminate or reduce the computational burdens in a statistics course, especially when using large real data sets. One of the key advantages of incorporating a package like SPSS, Minitab, JMP, R, StatCrunch, Stata, any of the many Excel add-ins, or on-line tools into the course is that they can relieve both students and teachers of the drudgery of computational tasks. Software can considerably reduce the amount of class and homework time devoted to calculation, and can free up cognitive and time resources for other ends. Most introductory-level textbooks now include examples and/or instruction in the use of software and provide datasets to accompany the text. Statistical software allows us to easily show an example data set with thousands of observations and explore a potential multivariate relationship within that data set. Other examples that involve having students perform common analytical tasks and explore data using statistical software can be found in Appendix C.

Using Statistical Software to Teach Concepts

Educators who view a statistical package only as a computational engine may want to consider the considerable potential of these packages for helping students build deep understanding of fundamental abstract **concepts**. The statistics education literature contains numerous articles that both advocate for, and demonstrate the efficacy of, using software to improve student learning (Chance et al. 2007; West 2009). Software simplifies and expedites the process of constructing and modifying graphs and also allows for replicating operations. For example, in the past, we might have been reluctant to ask students to make multiple histograms of the same data to illustrate the effects of bin width. With software, the task becomes easy. As such, software affords instructors the chance to create in-class demonstrations and homework assignments that guide learners to the "aha! moment" – that moment when a concept is no longer a technical textbook definition but an insight the student genuinely owns. One common example is the concept of a **distribution**. By using software to make dotplots, boxplots and histograms to visualize how individual data points vary along a number line, students gradually construct the idea that observations spread out across a certain range, while also concentrating in certain regions.

For a more subtle example, consider **sampling variability** among simple random samples. This is a concept that is particularly elusive for many students, even for those with considerable prior exposure. Students may read lucid explanations, hear a clear lecture, view or interact with a Central Limit Theorem applet, and yet still not really be able to write or speak confidently about sampling variability or sampling error. Students often have trouble reconciling the images of "all possible samples" with the knowledge that we typically draw a single random sample. Software may provide an additional avenue to build understanding.

EXAMPLE ACTIVITY

To use this activity in class, students need to have access to computers with the school's favored statistical package available. In classes where this is not feasible, the activity can be modified to be a homework assignment to be completed before class. As suggested elsewhere in this report, if technology is not available to support a homework assignment, instructors might still demonstrate software in class or share images of relevant software output.

Select a set of microdata from a population, such as live births in the United States for one month (full years available for download as a text file at http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Births). The entire dataset contains nearly four million rows, exceeding the limits of Excel and Notepad (among other programs), so it requires some instructor pre-work. It may be wise to select a single month to provide the class with a smaller, but still quite large, table of data. Choose one continuous variable, such as birthweight. Have each individual student open the data set within the software, and explain to the class that this really is population data: every child born in the U.S. for that period. The purpose of this demonstration is to explore the extent to which simple random sampling reliably produces "good" estimates of various population parameters. For example, we might consider the mean, median, and standard deviation.

The instructor uses her/his computer to find the parameters of the population. For dramatic effect, one might even write them on the board and then conceal them, as a metaphor for the true but unknown parameter.

Then, in the first stage, each student uses the software to select a random sample of, perhaps n = 50 rows of data, and saves this subset as the student's personal, single random sample. Indicate to the class that each student is a separate, independent investigator gleaning information from a large population. In a small class, students might take multiple samples to achieve the desired result.

The instructor would then ask each student to use the software to find the sample mean and construct both a 95% and a 90% confidence interval for the mean of the population.

Once the students have constructed their confidence intervals, they would be reminded of the value of the actual population mean. The instructor could then say, "First look at your 95% confidence interval, and see if it contains/brackets the actual value of μ. Remember, ordinarily we don't know mu, and our only knowledge about the population would come from our one sample. If your only knowledge of this population had been your sample, how many of you would have 'missed' μ?"

Students could be asked to stand in place and count off. Naturally, this should be roughly 5% of the class.

Once the students are seated, the instructor could conduct a very brief discussion to inquire why their results were "wrong," leading to the conclusion that sampling error is inherent in the practice of random sampling, rather than any kind of mistake by the investigator.

While these students remain standing, the instructor could ask the same question again, this time referencing the 90% interval, and ask unlucky students to stand. The instructor could point out to the class that (a) the same students are standing again, but (b) they are now joined by an approximately equal number.

Depending on time and the type of software, one might also construct confidence intervals for the median and standard deviation.

If students do not have software in class, one might simply have them take the random samples and construct CIs for homework, bringing their results to class and/or submitting the results online prior to class. At that point, the instructor could create a graph summarizing the distribution of sample means and sample 95% and 90% confidence intervals.

Inspiration:

Centers for Disease Control and Prevention, National Center for Health Statistics (2016), "Vital Statistics Data Available Online." Available at http://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Births

Chance, B., Ben-Zvi, D., Garfield, J., and Medina, E. (2007), "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovation in Statistics Education, 1*. Available at http://escholarship.org/uc/item/8sd2t4rr.

Diez, D.M., Barr, C.D., and Cetinkaya-Rundel, M. (2012), *OpenIntro Stats* (2nd Ed.). Available at https://www.openintro.org/stat/textbook.php?stat_book=os.

Scheaffer, R., Gnanadesikan, M., Watkins, A., and Witmer, J. (1996), *Activity-Based Statistics: Instructor Resources*. New York: Springer Verlag.

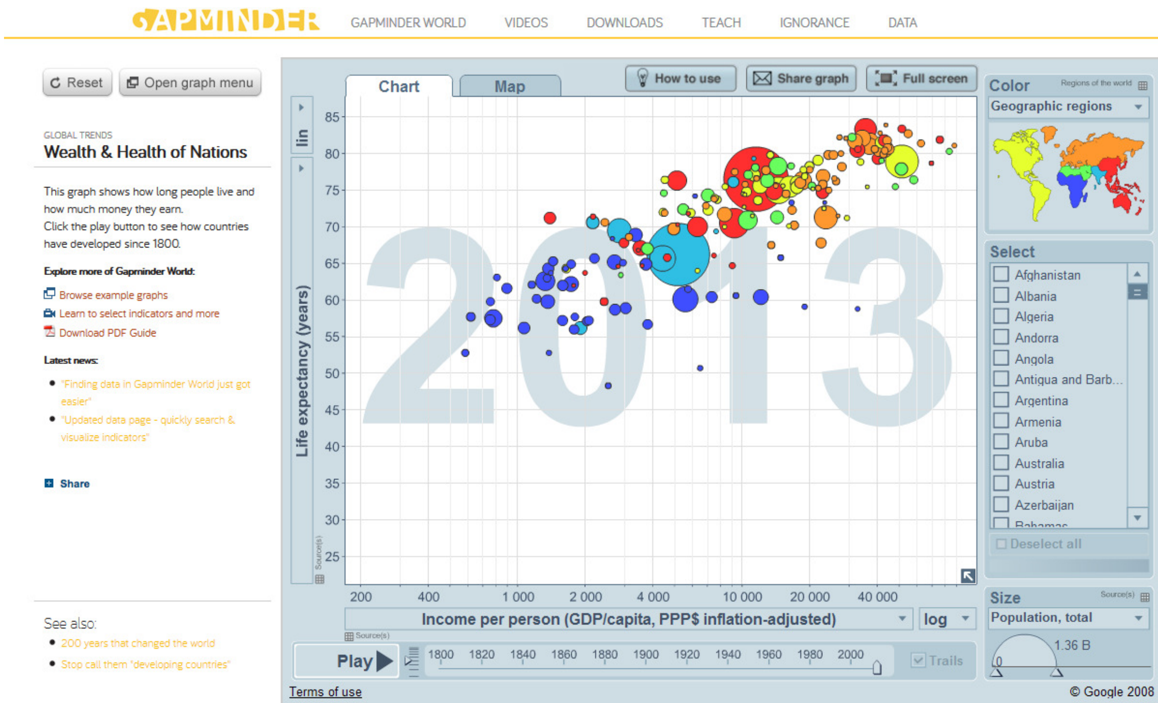
West, W. (2009), "Social Data Analysis with StatCrunch: Potential Benefits to Statistical Education," *Technology Innovations in Statistics Education*, *3*. Available at http://escholarship.org/uc/item/67j8j18s.

Using Software to Create a Wider Variety of Visualizations

Software not only facilitates the creation and manipulation of traditional statistical graphs, it has also introduced new methods of visualizing large data sets in ways that can stimulate insight and curiosity among undergraduates (and their teachers). Through the use of interactive controls and visual primitives like color, shape, and size, a user can quickly learn to create, modify, or manipulate multidimensional graphics. Such graphing tools are engaging and fun, and they invite the kind of exploration that lies at the heart of statistical thinking.

One innovative graph type is the *bubble chart*, which might be thought of as a super-charged scatterplot. In a bubble plot, there are two quantitative variables on the X and Y axes. Additionally, by replacing dots with "bubbles" of varying sizes and colors, one can represent two additional categorical or quantitative variables. Finally, one can include a time dimension and animate the scatterplot.

As a first illustration, visit the http://www.gapminder.org/world to see a vivid interactive five-dimensional graph that is interactive and animated (see image below). At the same site, one can download the data (originally from the World Bank's World Development Indicators) and the software needed to create the graph.

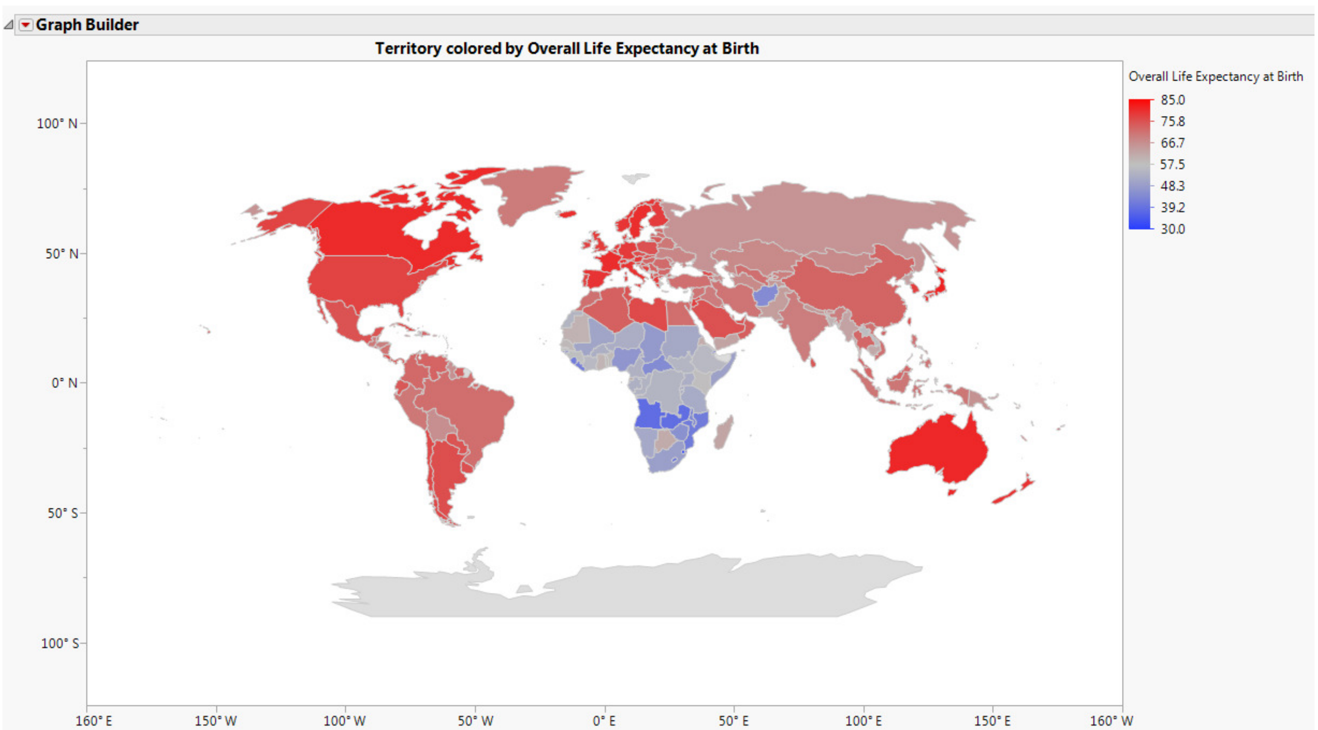


In the default graph as shown, we have data from more than 200 countries covering the period from 1800 through 2013. The vertical axis is Life Expectancy at Birth (in years) and the horizontal axis is the log of Income per Person (GDP per capita, in inflation-adjusted purchasing-power parity in US dollars). Bubble areas correspond to the population of each country annually, and the bubble colors indicate the region of the world for each country.

By pressing the Play button, one sets the graph into motion, tracing the changes in the variables starting in 1800 and progressing through 2013. As the animation continues, patterns and deviations from those patterns appear quite vividly. For example, in the years from roughly 1913 through 1919, life expectancy in much of the world plummets and then rebounds; this time period corresponds to World War I and the Spanish flu epidemic. In our classroom experience, students recognize the dramatic shift in the numbers and raise questions about it. In other words, students engage in statistical thinking as a consequence of viewing this particular visualization. Moreover, one does not need extensive instruction in how to interpret such a graph.

Mapping is another increasingly common visualization that is intuitive and insight-provoking, though statistics textbooks have been slow to add maps to the canon of basic statistical graphing. The Gapminder site includes a world map, as do some other software packages. For this illustration, we'll look at the mapping feature that is standard in JMP, along with a dataset that ships with JMP. This example does not provide complete step by step instructions, but merely illustrates the capability of the software.

JMP's "World Demographics" data set contains observations of 32 variables for 238 countries of the world in 2009. After opening the data table, the user invokes JMP's Graph Builder platform which presents a list of available variables and a blank "canvas" for graph creation. To make a map, one selects a geographic identifier variable and another variable to determine a color gradient. Below is a world map showing the 2009 life expectancy at birth, by country.



Student users can create this graph in a few steps with a point-and-click interface, and can explore different variables in similar fashion. Here again, both the construction and the interpretation of the visualization can occur with minimal instruction—in contrast to, for example, a histogram or box-and-whiskers plot of the same data. Given their visual impact, simple maps like this provide an excellent opportunity to tell a story from data, a key goal of undergraduate statistics education.

Inspiration:

JMP Software, Sample Data Files.

Rosling, Hans. *Gapminder World.* Accessed June 9, 2015 at www.gapminder.org/world.

Using Software for Reproducibility and Better, Clearer Student Assignments

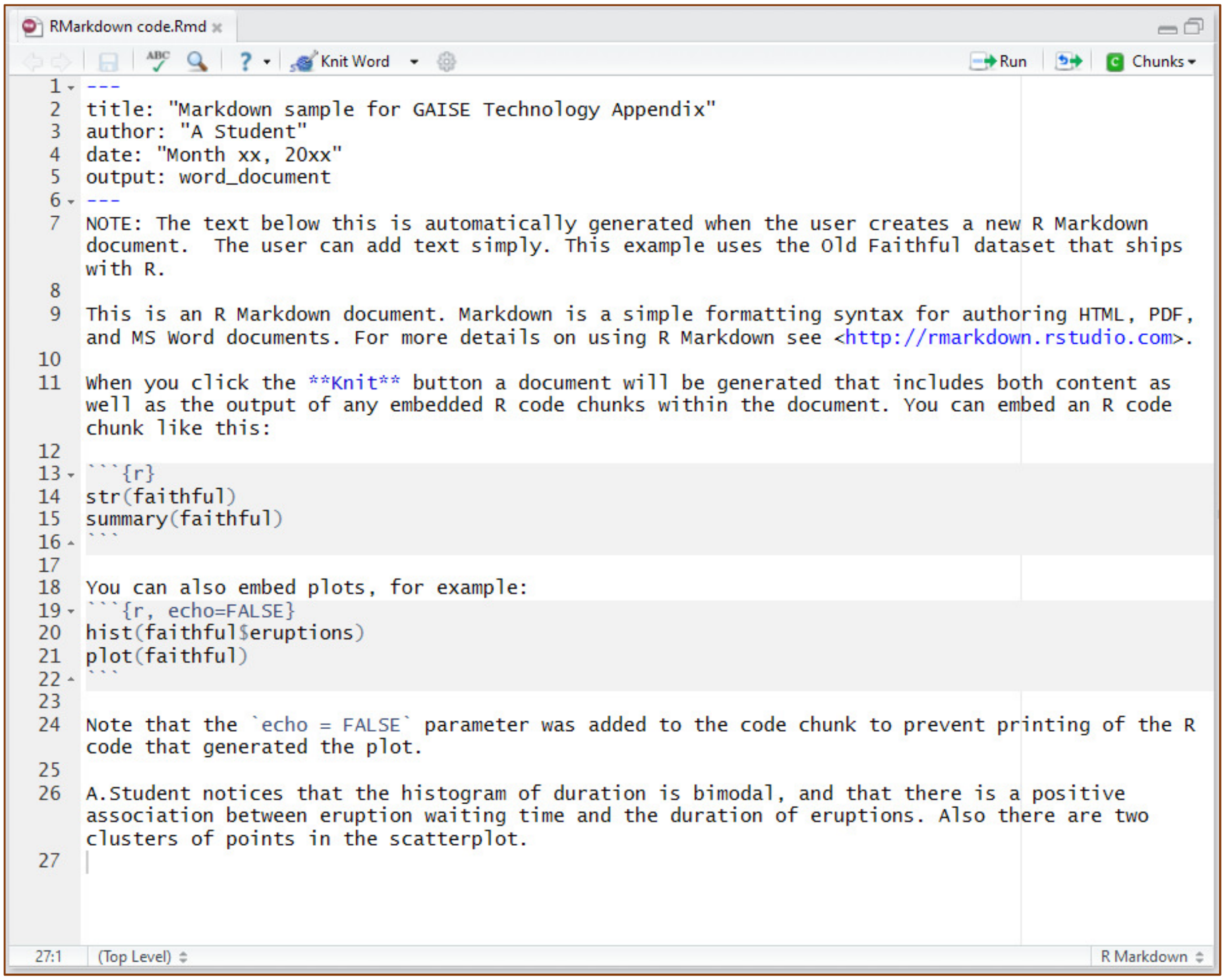
One recent trend in the scientific community is an emerging consensus on the value and importance of reproducibility in scientific publications (see, for example, the editorial in *Nature*, 2014). In 2014, at a gathering convened by the US National Institutes of Health and the journals *Nature* and *Science,* attendees adopted a set of guidelines calling for, among other provisions, publication of statistical procedures, method of randomization and other detailed information to allow for others to reproduce the published work.

In a statistics classroom, instructors may seek reproducibility as well; in addition to asking students to report final results of an analysis, we may also wish to see the code or dialog choices that generated the output, as well as reading the conclusions that student authors drew. Here again, technology can simplify and expedite the process.

One freely available tool is R Markdown. For this example, we'll demonstrate the ease of use and the instructional benefits of using R Markdown as implemented in RStudio. In this example, we use the Old Faithful dataset to illustrate how the R Markdown environment integrates a student's code with output and with whatever commentary or responses a student might add. Full instructions are beyond the scope of this example; we include it to indicate how this particular technology can overcome a common obstacle in the preparation of coherent complete lab reports.

In R Markdown, a user can combine text with "chunks" of R code, and by "knitting" the R Markdown document, produce a fully-editable Word document (or HTML or pdf object) for submission. The rendered document will contain the student's code, output, and written work. Presumably, the student would be advised to develop and test the code in R prior to transferring chunks to the R Markdown document.

Below is a screen shot of an R Markdown document, followed by the Word document created by it.




```
---
title: "Markdown sample for GAISE Technology Appendix"
author: "A Student"
date: "Month xx, 20xx"
output: word_document
---
NOTE: The text below this is automatically generated when the user creates a new R Markdown
document.  The user can add text simply. This example uses the Old Faithful dataset that ships
with R.

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF,
and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as
well as the output of any embedded R code chunks within the document. You can embed an R code
chunk like this:

```{r}
str(faithful)
summary(faithful)
```

You can also embed plots, for example:
```{r, echo=FALSE}
hist(faithful$eruptions)
plot(faithful)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R
code that generated the plot.

A.Student notices that the histogram of duration is bimodal, and that there is a positive
association between eruption waiting time and the duration of eruptions. Also there are two
clusters of points in the scatterplot.
```

Markdown sample for GAISE Technology Appendix

A Student

Month xx, 20xx

NOTE: The text below this is automatically generated when the user creates a new R Markdown document. The user can add text simply. This example uses the Old Faithful dataset that ships with R.

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
str(faithful)

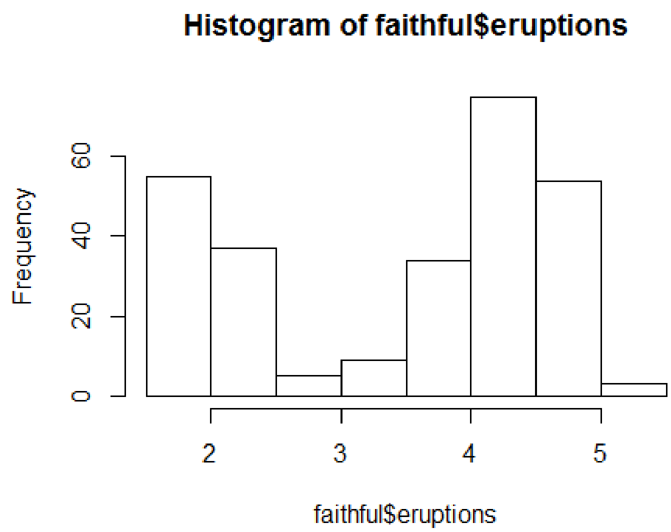
## 'data.frame':    272 obs. of  2 variables:
##  $ eruptions: num  3.6 1.8 3.33 2.28 4.53 ...
```

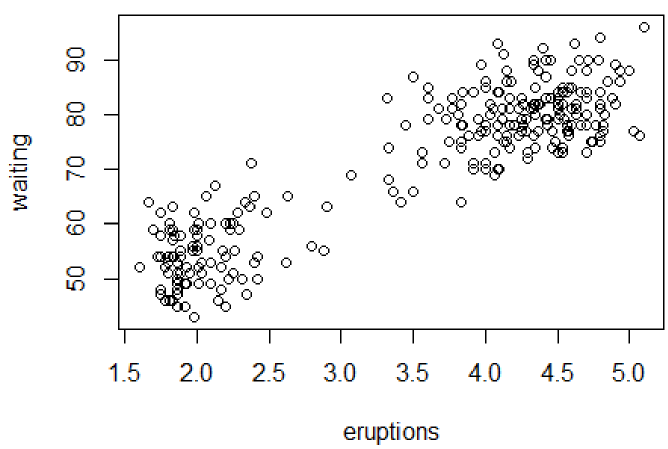
```
## $ waiting : num 79 54 74 62 85 55 88 85 51 85 ...
```

```
summary(faithful)
```

```
##    eruptions        waiting
##  Min.   :1.600   Min.   :43.0
##  1st Qu.:2.163   1st Qu.:58.0
##  Median :4.000   Median :76.0
##  Mean   :3.488   Mean   :70.9
##  3rd Qu.:4.454   3rd Qu.:82.0
##  Max.   :5.100   Max.   :96.0
```

You can also embed plots, for example:





Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.

A.Student notices that the histogram of duration is bimodal, and that there is a positive association between eruption waiting time and the duration of eruptions. Also there are two clusters of points in the scatterplot.

Inspiration:

Baumer, B., Cetinkaya-Rundel, M, Bray, A., Loi, L., and Horton, N. J. (2014), "R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics," *Technology Innovations in Statistics Education, 8*. Available at http://escholarship.org/uc/item/90b2f5xh.

Editorial (2014), "Journals Unite for Reproducibility," *Nature*, *515* (7). 06 November. Available at http://www.nature.com/news/journals-unite-for-reproducibility-1.16259.

Accessing Real Raw Data online

The third GAISE recommendation is to "Integrate real data with a context and a purpose." Appendix C suggests several ways to generate or acquire real data. This section augments those methods by highlighting web-based avenues to obtain discipline-specific observational, experimental, and survey data. Instructors seeking real data are advised to begin with visits to curated data repositories such as:

- Data and Story Library (DASL), http://dasl.datadesk.com/
- Nationmaster – http://www.nationmaster.com/ portal to international economic, demographic and social data.
- Awesome Public Data Sets - https://github.com/caesar0301/awesome-public-datasets by Xiaming Chen and other contributors
- WISE (Web Interface for Statistics Education) has a collection of demonstrations and tutorials. In addition, a list of data sources posted on their website under helpful links http://wise.cgu.edu/helpful-links/data-sources/

Accessing Observational Data

Massive volumes of real-time, automatically generated and/or captured data are now available publicly and for free across disciplines, making it possible to find and use raw data attuned to the needs of courses and the interests of students. This short section briefly describes how one might locate and extract such data, using two examples that illustrate both ends of the "ease of use" spectrum. Because on-line observational data can allow instructors flexibility in choosing and creating assignments, the classroom illustrations shown in this section are framed as templates that instructors should tailor to their students and courses.

USER-FRIENDLY ACCESS ILLUSTRATION:

Many federal agencies in the U.S. provide user interfaces to build custom queries for transactional databases. The example here comes from the Bureau of Transportation Statistics and the On-Time Performance database. According to the BTS website, the database "contains on-time arrival data for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, cancelled or diverted flights, taxi-out and taxi-in times, air time, and non-stop distance." Users can download 109 light-level variables for a selected month and year, and can filter geographically. Hence, an instructor can obtain data for nearby airports and recent time periods, selecting variables of particular interest.

The query screen presents a set of drop-down filter selectors and variable checkboxes to specify which variable fields a user wants to download. After making the selections, the site generates and delivers a zipped comma-separated-values (csv) file of raw data.

In late 2015, the search screen looked like this:

Prototype Assignment—Student Prompt

Have you ever been on a flight that left the gate on time, only to be frustrated by a long wait before takeoff? How often do such delays occur? The U.S. Department of Transportation maintains a database with information about the departures and arrivals of every domestic commercial flight in the United States.

We have a dataset that can help us answer the question, and to compare our local airport with some of the busiest airports in the country. In the airline industry, the time elapsed between departing the gate and "wheels up" is known as "Taxi out" time. For this exercise, I have downloaded from the Bureau of Transportation Statistics and placed it in a file called {filename}. The file contains several variables about individual flights departing from our nearest airport (airport code XXX), as well as flights from three airports with very heavy traffic: Atlanta (ATL), Los Angeles (LAX), and Chicago (ORD) during {MONTH, YEAR}.

For this assignment, the three variables of interest are:

- DayofWeek numeric code for day (1 = Sunday, 2 = Monday, etc.)
- Origin three-letter airport ID code
- Carrier airline identification code
- DepTimeBlk standard departure time intervals from the Computerized Reservations System (CRS)
- TaxiOut taxi out time, in minutes

For the month of {MONTH}, use appropriate graphical and descriptive summaries to investigate these questions and prompts:

1. On a typical flight that month at our airport, how long did it take for flights to take off after leaving the gate?
2. How did our taxi out times compare to ATL, LAX and ORD?
3. At our airport, how did the different airlines compare in terms of taxi out times?
4. How (if at all) did taxi out times vary by the day of the week at our airport? Is the variability similar at the other large airports in the dataset?
5. How (if at all) did taxi out times vary by time of day at our airport?
6. Suppose you are planning to fly out of our airport next month. Would you be inclined to take this analysis into account in preparing for your flight? Why or why not?

COMPUTING-INTENSIVE DATA ACCESS ILLUSTRATION:

Professional sports have been among the early adopters of the managerial and strategic use of statistical analysis due in some measure to the technologies that automatically capture data at a very granular level. Major League Baseball (MLB) has led the way in this regard. With the types of data available to date, one can investigate questions related to teams, players, plays, and even individual pitches.

For example, since 2007, MLB has been tracking and publishing measurements of every pitch thrown in every game of the season (see Inspiration and References below). The technology behind the data collection is called PITCHf/x, and the relevant MLB website contains XML files within a hierarchical directory tree structure organized by date. More specifically, to access

PITCHf/x data for a single game, one first identifies the year, month, date and "game id" for the game of interest.

The data are freely available in the sense that there is no payment required, but in contrast to the prior example, actually accessing and preparing the data for analysis is best done by writing code to automatically scrape dates, games and innings of interest. There are some third-party sites that help with access, but the MLB site exemplifies the challenges and rewards of the "big data" era: the available data open the doors to previously unimaginable analysis, but the doors are complicated to navigate.

This illustration is inspired by a 2010 article in the *Journal of Statistics Education* by Jim Albert, a leader in using sports data for instruction. We are aware that sports data examples excite some students and can alienate others. The point here is to illustrate the availability and challenges of accessing some web sources, not to advocate the use of this particular type of baseball data.

In Albert's article, we have a dataset with the following variables:

Table 4. Variables and description of pitch-by-pitch data frame **pitchdata** using the text file "pitchfx.dat".

Variable	Description	Variable	Description
pitcher	name of pitcher	start_speed	starting speed of pitch
game	game number	end_speed	speed of pitch crossing plate
id	pitcher id number	sz_top	top of strike zone
inning	inning of game	sz_bot	bottom of strike zone
num	number of batter	pfx_x	deviation in horizontal direction
batter	batter id number	pfx_z	deviation in vertical location
stand	hitting side of batter	px	pitch location in x direction
b_height	height of batter	pz	pitch location in z direction
p_throws	throwing side of pitcher	pitch_type	pitch classification
des	play description	count	current pitch count
event	result of plate appearance	new.count	new pitch count
brief_event	brief description of result	value	pitch value
des2	pitch outcome	new.count.type	PA event or new count
type	ball, strike, or in-play?	count.adv	pitcher or batter or neutral count

Prototype Assignment—Student Prompt

In Major League Baseball, successful pitchers combine athletic skill and tactical judgment to outwit opposing batters. Pitchers vary in many respects, including the variety of pitches they use (fastballs, curveballs, etc.), the way they sequence pitches, as well as the speed and movement of their different pitches.

Since the 2007 season, the MLB has used a digital recording system to measure particular parameters of every pitch thrown in every game. The technology involved is known as "PITCHf/x" and for this assignment we have a data file called {FILENAME}. This file contains all of the games played by our local team, the {TEAM} during the {YEAR} season. In this assignment, we'll focus on the pitching performance of our ace pitcher {PITCHER}. Your tasks:

1. Because teams have several pitchers who rotate from game to game, our first task is to isolate only the data rows involving {PITCHER}. Use software to create a subset of {FILENAME} that contains only pitches thrown by PITCHER.
2. What types of pitches does {PITCHER} tend to throw, and how often does he throw each?
3. What were the outcomes of his pitches as the end of the plate appearance?
4. How do the outcomes vary by type of pitch?
5. Are particular pitches more successful at inducing batters to swing and miss?

Of the different pitch varieties, some are distinguished by their "movement," or the extent to which they deviate from an imaginary straight line between the pitcher's hand and home plate. In the PITCHf/x data, pfx_x and pfx_z represent the movement in the horizontal and vertical directions respectively. The horizontal break is calculated from a perspective behind the point of home plate, so that negative values of pfx_x move towards a right-handed batter and away from a left-handed batter.

Pitches also vary in speed, and PITCHf/x records two speeds: when the ball leaves the pitcher's hand and when it crosses the plate.

6. Use software to create a scatterplot of the horizontal and vertical movement of {PITCHER's} 4-seam fastballs, curveballs, and change-ups. In the plot, use different symbols for the three pitch types and vary the intensity of the color by the end_speed of the pitch (speed arriving at the batter).
7. In plain English, explain what you see in this scatterplot.

Inspiration and References

Albert, J. (2010), "Baseball Data at Season, Play-by-play, and Pitch-by-pitch levels," *Journal of Statistics Education, 18*. www.amstat.org/publications/jse/v18n3/albert.pdf.

Major League Baseball (2010), *Pitch F/X data files*. Available at http://gd2.mlb.com/components/game/mlb.

Silver, N. (2015), "The Best and Worst Airlines, Airports and Flights, Summer 2015 Update." Downloaded June 17, 2015 from http://fivethirtyeight.com/features/the-best-and-worst-airlines-airports-and-flights-summer-2015-update/.

U.S. Department of Transportation, Bureaus of Transportation Statistics (2015), *Database Profile.* Available at http://www.transtats.bts.gov/TableInfo.asp?Table_ID=236.

U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology (2015), *Airline On-Time Performance Data database*. Available at http://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airline%20On-Time%20Performance%20Data&DB_Short_Name=On-Time.

Accessing Experimental Data

Traditionally, raw experimental data has been proprietary and therefore difficult to obtain. More recently, we are seeing contemporaneous phenomena including the Open Science, Science Commons, and Citizen Science movements which all make use of the Internet to advance the sharing of data. As Dawson (2012) has stated,

> "Taking inspiration from the open source software and open access movements, some scientists are now sharing their lab notebooks and raw experimental data openly online. Open science is a broad concept that includes these closely related areas of open notebook science and open data. Advocates of open science believe that there should be no insider information, and all protocols and results -- even those of failed experiments -- should be made visible and open to reuse as soon as possible in open lab notebooks and data repositories."

Recently, in the United States, the National Science Foundation has, as a matter of policy, committed itself to pressing NSF-funded researchers to publish not only results and reports of research, but to publish raw data as well (NSF 2013).

These developments offer great promise for statistics education to use web browsers and statistical software to obtain and access experimental data, including the ability to access both significant and non-significant results drawn for a wide variety of client disciplines.

The following example uses a subset of data contributed to the University of California Irvine Machine Learning Repository, and it illustrates a common practice of A-B testing in web-based environments. The data were provided by YouTube, as described in the Student Handout that follows. This example deals with a substantive domain that should be familiar to most undergraduates.

Student Handout

YouTube Comedy Slam was a video discovery experiment running on YouTube's version of labs (called TestTube) for a few months in 2011 and 2012. In this experiment, a pair of videos was shown to each user, and users were asked to vote for the video they found to be funnier. Left/right positions of the videos were randomly selected before being presented to users to eliminate position bias. Videos were selected from a large pool of weekly updated sets of videos. Users were self-selected visitors to YouTube.

We have provided you a dataset (n = 3,545 observations), drawn from the original sample of more than 1.1 million preference votes. Each line in this dataset corresponds to one vote over a pair of YouTube videos. One of the videos was a compilation of amusing footage of cats in various settings, and the other was a practical joke played on a co-worker by a friend.

Each video is represented by its YouTube video ID (see references for the URLs of each video). There are three columns in the dataset: Left, Right, and Choice. The first two columns just indicate which video ID was shown in which position, and the third column represents the user's choice.

Shetty (2012) provides a more complete description of the experimental setting:

Quantifying comedy on YouTube: why the number of o's in your LOL matter

Posted by Sanketh Shetty, YouTube Slam Team, Google Research

In a previous post, we talked about quantification of musical talent using machine learning on acoustic features for YouTube Music Slam. We wondered if we could do the same for funny videos, i.e. answer questions such as: is a video funny, how funny do viewers think it is, and why is it funny? We noticed a few audiovisual patterns across comedy videos on YouTube, such as shaky camera motion or audible laughter, which we can automatically detect. While content-based features worked well for music, identifying humor based on just such features is AI-Complete. Humor preference is subjective, perhaps even more so than musical taste.

Fortunately, at YouTube, we have more to work with. We focused on videos uploaded in the comedy category. We captured the uploader's belief in the funniness of their video via features based on title, description and tags. Viewers' reactions, in the form of comments, further validate a video's comedic value. To this end we computed more text features based on words associated with amusement in comments. These included (a) sounds associated with laughter such as hahaha, with culture-dependent variants such as hehehe, jajaja, kekeke, (b) web acronyms such as lol, lmao, rofl, (c) funny and synonyms of funny, and (d) emoticons such as :), ;-), xP. We then trained classifiers to identify funny videos and then tell us why they are funny by categorizing them into genres such as “funny pets”, “spoofs or parodies”, “standup”, “pranks”, and “funny commercials”.

Next we needed an algorithm to rank these funny videos by comedic potential, e.g. is “Charlie bit my finger” funnier than “David after dentist”? Raw viewcount on its own is insufficient as a ranking metric since it is biased by video age and exposure. We noticed that viewers emphasize their reaction to funny videos in several ways: e.g. capitalization (LOL), elongation (loooooool), repetition (lolololol), exclamation (lolllll!!!!!), and combinations thereof. If a user uses an “loooooool” vs an “loool”, does it mean they were more amused? We designed features to quantify the degree of emphasis on words associated with amusement in viewer comments. We then trained a passive-aggressive ranking algorithm using human-annotated pairwise ground truth and a combination of text and audiovisual features. Similar to Music Slam, we used this ranker to populate candidates for human voting for our Comedy Slam.

So far, more than 75,000 people have cast more than 700,000 votes, making comedy our most popular slam category. Give it a try!

Further reading:

"Opinion Mining and Sentiment Analysis," by Bo Pang and Lillian Lee.

"A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews," by Oren Tsur, Dmitry Davidov, and Ari Rappoport.

"That's What She Said: Double Entendre Identification," by Chloe Kiddon and

1. After reading the background material and exploring the data set, briefly describe this experimental design. To what extent will it be reasonable to generalize from any conclusions drawn? Explain your thinking.
2. What is the response (dependent) variable in this experiment?
3. What is/are the explanatory (independent) variable(s) in this experiment?
4. Discuss how you intend to use your statistical software to prepare the data for analysis and then analyze the data from this experiment.
5. Run an appropriate analysis of the data and report on your conclusions.

Teaching Notes:

- Technology is present in three forms in this exercise. First, the experiment is inherently technologically-based and exemplifies a widely-adopted practice in the design of web interfaces. Second, the data were obtained on-line and required some manipulation to create a student-friendly dataset. Finally, students should be expected to perform the analysis. Depending on the statistical software available to them, instructors may wish to reformat the data.
- The questions listed above should be tailored depending upon course emphasis.

Suggestions to Instructors for further investigation

This example is based on a dataset available at the Machine Learning Repository at the University of California at Irvine (see references below). Instructors who want to find other experimental datasets to create an original assignment might consult other sites.

- There are compendia of web resources like http://www.statsci.org/datasets.html or CAUSEWeb, which both curate datasets as well as links to other data compilations.
- Those who teach students in health-related disciplines should visit US government sites such as http://clinicaltrials.gov/ for a clearer sense of current trends in making shared data available.

Inspiration and References

Dawson, D. (2012), "Open Science and Crowd Science: Selected Sites and Resources," *Issues in Science and Technology Librarianship, Spring*. Available at http://www.istl.org/12-spring/internet2.html#online-tools.

"Hilarious Cats" (n.d.) Available at https://www.youtube.com/watch?v=7zCIRPQ8qWc.

National Science Foundation (2013), "National Science Foundation Collaborates with Federal Partners to Plan for Comprehensive Public Access to Research Results," Press

Release 13-030, February 22, 2013. Available at http://www.nsf.gov/news/news_summ.jsp?cntn_id=127043.

"O susto do gordinho" (n.d.) Available at https://www.youtube.com/watch?v=W9y6nwBwwyQ.

Shetty, S. (2012), "Quantifying Comedy on YouTube: Why the Number of o's in your LOL matter," *Google Research Blog*, February 9. Available at http://googleresearch.blogspot.com/2012/02/quantifying-comedy-on-youtube-why.html.

UC Irvine Machine Learning Repository (2012). Available at https://archive.ics.uci.edu/ml/datasets/YouTube+Comedy+Slam+Preference+Data#.

Accessing Real Survey Data

In their quest to use real data from client disciplines, instructors have ready access to an enormous variety of large-scale survey data collected by reputable agencies. Typically, such datasets are accessible via user-friendly web interfaces that include codebooks and background information, and many such sites permit selection by variable. This example uses the General Social Survey (GSS), an annual full-probability, personal interview survey designed to monitor changes in both social characteristics and attitudes in the United States (NORC 2014b). The GSS is a project of the National Opinion Research Center (NORC) at the University of Chicago, and has been administered since 1972.

The NORC website (2014a) provides this brief overview of the survey: "The GSS contains a standard 'core' of demographic, behavioral, and attitudinal questions, plus topics of special interest. Many of the core questions have remained unchanged since 1972 to facilitate time-trend studies as well as replication of earlier findings. The GSS takes the pulse of America, and is a unique and valuable resource. It has tracked the opinions of Americans over the last four decades." Subject areas cover a wide range of social and cultural issues, including attitudes about government, religion, the workplace, equality, and popular culture. Annually, the sample size is approximately 2,000 respondents, aged 18 years and above.

Most variables are categorical, using Likert-type scales. Downloads typically include respondent identifiers, basic demographics (*e.g.*, age, region, gender), as well as interview dates and sampling weights. The GSS covers a wide variety of topics, so it is suitable for many introductory courses, but should be of particular value for any instructor teaching an introductory statistics course with a social science focus.

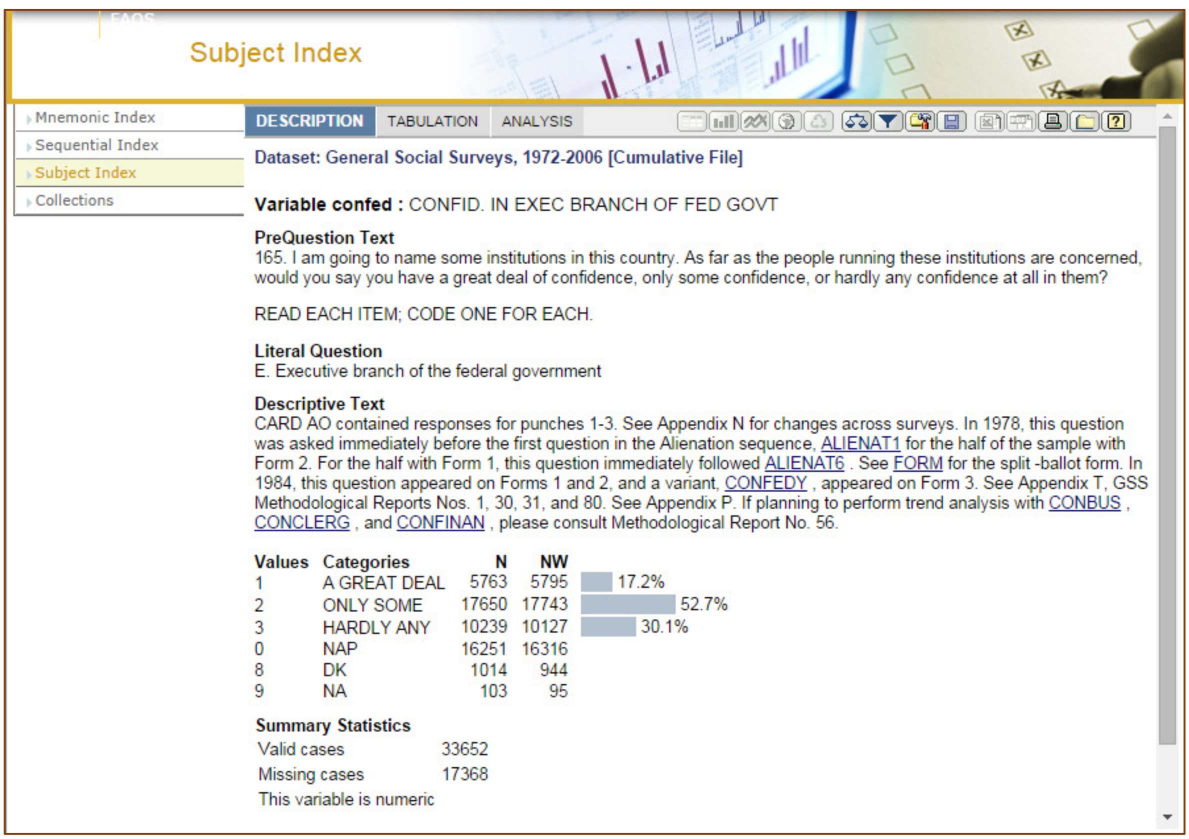
Due to the changing design and content of websites, this discussion minimizes references to specific URLs or menus, but a web search for NORC or the General Social Survey should suffice. At the time of this writing, one should navigate to the main NORC site: http://www3.norc.org/GSS+Website/

- From the NORC site, users choose between SPSS or STATA formats. These are available freely to any user.

- Alternatively, the NORC download site also provides links to the GSS Data Explorer, to the Roper Center at the University of Connecticut and to ICPSR (Inter-university Consortium for Political and Social Research at the University of Michigan), including ASCII, SAS, delimited, and R. At the latter two sites, membership and/or fees may be required.

Users can download multi-year cumulative files or annual complete datasets. Alternatively, one is able to browse variables by subject, by variable name, or in other ways. Hence, instructors seeking a dataset suitable for a particular course have considerable flexibility of access.

Questions at the core of the GSS appear annually in the instrument, while others may be included just once or periodically. For example, a nearly-annual question (NORC 2014c) asks "I am going to name some institutions in this country. As far as the people running these institutions are concerned, would you say you have a great deal of confidence, only some confidence, or hardly any confidence at all in them?... Executive branch of the federal government." The NORC site provides detailed documentation for each variable, as well as summary statistics for the cumulative period 1972–2006. The figure below is a screen capture at the time of this writing (Fall 2014) of the Subject Index page for this particular question, showing the question text and a descriptive summary of responses aggregated over the cumulative time period.



Given the wide-ranging scope of the GSS, instructors can make use of the data for a variety of descriptive and inferential assignments and examples, and highlight important concepts in survey construction and/or interpretation of raw data. Here are some possible activities related to this particular question. Note that the frequency table provided above reports both the raw number of responses (N) and the *weighted* number of responses (NW). The reported percentages use the weighted counts divided by the number of valid cases (33,652). The difference between weighted and unweighted counts is probably beyond the scope of most introductory courses, but instructors should preface these questions by noting that survey researchers typically use weighting as a way to compensate for the fact that some demographic groups are over- or under-represented by a particular sampling method.

1. Trained employees of NORC administer the General Social Survey. The "PreQuestion Text" and "Literal Question" and the instruction "READ EACH ITEM; CODE ONE FOR EACH" are worded quite precisely. Why do you think the General Social Survey administrators are so particular about the wording of questions? What difference would it make if an interviewer asked the question using different wording?

2. Before analyzing the responses to this question, look closely at the **Categories** listed. The first three are straightforward enough to interpret, but "NAP," "DK," and "NA" are ambiguous. Visit the NORC website and search for these three abbreviations. Report briefly on what you find.

3. Below the heading "Summary Statistics," notice the phrase "This variable is numeric." Is it? What do you think the General Social Survey folks mean by this?

4. In the United States, the Federal government consists of three branches: executive, legislative, and judicial. In the aggregate from 1972 through 2006, approximately 17.2% of respondents expressed "a great deal" of confidence in the executive branch. How does that compare to confidence in the other two branches? Visit the NORC site (or use data provided by your instructor) to locate the corresponding variables for the legislative and judicial branches. Write a few short sentences comparing respondents' confidence in the three branches.

5. Looking at these aggregated responses over a period of more than 30 years may raise some questions in your mind. Write down one or two questions about confidence in the institutions of government that you would like to investigate further using GSS data.

Inspiration:

National Opinion Research Center (2014a), *General Social Survey*. Available at http://www3.norc.org/GSS+Website/.

National Opinion Research Center (2014b), *National Data Program for Social Sciences*. Available at http://www3.norc.org/GSS+Website/About+GSS/National+Data+Program+for+Social+Sciences/.

National Opinion Research Center (2014c), *Browse GSS Variables Subject Index*. Available at http://www3.norc.org/GSS+Website/Browse+GSS+Variables/Subject+Index/.

Using Games and Other Virtual Environments

Computer gaming has become a large source of entertainment for many people, including college students. In the past few years, energy has been spent to design online games for use in statistics classrooms. The hope is that by using computer games in the statistics classroom, a higher level of engagement can occur. There are multiple ways that we can use games in the classroom.

Real Data: Students can have personal experiences with games. Students can put together jigsaw puzzles, complete crossword puzzles or play some other quick online game. The students can then analyze the completion times or other variables from these games.[39]

Gathering Virtual Data: Sometimes data from virtual reality environments can also be used to engage students. For example, students can experience trying to collect data from an endangered species[40] or conduct a health survey across an entire Island or group of Islands. [41]

Experimental Design: Another way to incorporate the use of games into the classroom is by having students think about what factors affect the time to win a game or the points earned in a game. For example, does gender, the amount of hints, color of the pieces, and/or seeing a preview of a game affect the chance of winning the game?

Statistical Concept: With some games, players can only advance or move ahead in the game if they have mastered a statistical concept. By learning and applying a statistical concept, you are able to win. For example, perhaps by analyzing a set of past attempts, a more successful method of getting to the goal can be discovered. Or, perhaps studying the conditional probabilities of moves or studying a scatterplot may increase the number of points a player has in the game.[42]

Inspiration:

---

[39] Games and written lab activities can be found at this website http://web.grinnell.edu/individuals/kuipers/stat2labs/. Other statistical games and puzzles can be found online through a google search.

[40] For example, Shonda Kuiper has designed TigerStat that can be used to gather information about tigers: http://statgames.tietronix.com/TigerSTAT/.

[41] Michael Bulmer has also designed the Islands for data collection across a virtual population: http://islands.smp.uq.edu.au/login.php

[42] Some of these games can be found at http://play.ccssgames.com/

Erikson, T. (2013), “Designing Games for Understanding in a Data Analysis Environment,” *Technology Innovations in Statistics Education, 7*. Available at http://escholarship.org/uc/item/31t469kg.

Erikson, T., and Triggs, M. (2014), “An Early Look at Rich Learning Analytics: Statistics Students Playing ‘Markov’,” *ICOTS 9 Proceedings*. Available at http://icots.info/9/proceedings/pdfs/ICOTS9_9G1_ERICKSON.pdf .

Kuiper, S., and Sturdivant, R. (2014), “Using Fun and Games to Engage Real-world Learning,” *eCOTS 14 Workshop*. Available at https://www.causeweb.org/ecots/ecots14/34/.

Sturdivant, R., Jackson, J., and Cummiskey, K. (2013), “TigerStat: An Immersive 3-D Game for Statistics Classes,” *CAUSE webinar.* Available at https://www.causeweb.org/webinar/activity/2013-04/.

Bulmer, M., and Haladyn, J.K. (2011), “Life on an Island: A Simulated Population to Support Student Projects in Statistics,” *Technology Innovations in Statistics Education, 5***.** **Available** at http://escholarship.org/uc/item/2q0740hv.

Baglin, J., Bedford, A., and Bulmer, M. (2013), “Students’ Experiences and Perceptions of using a Virtual Environment for Project-based Assessment in an Online Introductory Statistics Course,” *Technology Innovations in Statistics Education, 7***.** Available at http://escholarship.org/uc/item/137120mt.

Student Handout

Your instructor will present you with a game.[43] Take a few minutes to familiarize yourself with the game.

Part One: As you familiarize yourself with the game, what are you interested in discovering? List a few of these questions here:

1.)

2.)

3.)

Part Two: Pick one of your above questions and think about how you would gather evidence to answer it. Here are some issues that you should consider.

1. What type of data would you need to collect?

[43] There are many online games where data can be collected. You can even use a traditional puzzle or board puzzle.

2. Would these variables be categorical or quantitative?
3. What type of lurking variables would you need to be concerned about? Why?
4. How would you include random sampling and/or random allocation?
5. What statistical method have we discussed in class could be used to analyze that data?

Part Three: Write up a study protocol. Be very detailed. Write the instructions for how an experimenter would go about conducting this experiment.

Teaching Note:

After this activity has been completed, one possible next step is to have the class pick one of the study ideas. The class can then collect data and analyze it. It might be a good idea for the students to complete the game outside of class time. Some students may take longer to complete certain puzzles and you don't want to accidently embarrass someone who takes twice as long as other students. If you decide to have students complete the puzzle outside of class, you can then also discuss what lurking variables this might introduce. If you decide to complete the puzzle in class, another option would be to give the students one minute to complete as much of the puzzle as they can and then measure percentage of completion.

Real Time Response

While not limited to statistics classrooms, real time responses systems (clickers) can be an asset in achieving the GAISE recommendations. The GAISE recommendations encourage instructors to use technology for computation and emphasizing concepts. Real time responses allow us to explore concepts, as well as gather data to analyze in the classroom.

Initially, real time response systems,[44] or audience response systems, started as devices similar to TV remotes. The device, commonly known as a clicker, could only be used to respond in class to a teacher's multiple choice question posted on a projection screen. Over the past decade or more, these devices have evolved. Devices that only allow for multiple choice entry are still available, and some also allow for numeric entry. Moreover, now there are systems that allow the students to enter information from their computers, tablets or even phones. Questions sometimes even

[44] Some of these devices include: specific devices: "clickers" such as iClicker or H~ITT; web based services such as Learning Catalytics, TopHat, b.socrative.com, Google Forms, Survey Monkey, Qualtrics, Poll Anywhere, Course Management Software

appear directly on the devices, and students are no longer limited to just multiple choice questions. Students can enter numerical values, equations, and even draw on the screen. These responses can then be combined and shared with the class. There is a vast array of options in how anonymous the responses from the students can be. The student's responses could be made anonymous from both other students and even from the instructor (in some cases). Having various options for anonymity allows for flexibility in teaching methods.

For example, some systems post a summary of responses provided by the class – the percentage who answered A, B, C, etc. This allows students to be able to judge where they stand in the class. Since this data is summarized, it provides a layer of protection for a shy student who might not ordinarily participate.

Real time response systems allow us to receive responses synchronously in a face-to-face class from all students. However, these systems can also be used for completion of assignments outside of class. The questions can be assigned for homework and recorded on a real time response system outside of class.

Some of the new systems even help enable team-based answering. For example, the system first asks a question or several questions which are answered by the individual student. Then, the students form teams. The team could be created by the instructor, self-selected by the students, or created by the real time response system based on individual student answers. Then, the team discusses the questions, and answers them again as a team. Some systems even show what each member of the team initially answered, enabling better team discussion.

The real time class responses can be used in multiple ways in the face-to-face classroom environment. Some of these examples include:

- Review the previous day's assignment or ask questions about an assigned reading
- Expose misconceptions and use them as talking points
- Illustrate a concept
- Collect data to analyze
- Allow instructors to determine if they are going too fast/ too slow
- Supplement applets by helping to focus the students' explorations of the applet. For example, first ask the students to answer a series of questions, then have them play with an applet and re-answer the same series of questions.

Future Direction of Real Time Response

With the presence of cloud-based computing, more opportunities will become available for students to collaborate in real time but separate spaces. For example, students might work together to create one document[45] or annotate an existing document[46].

[45] For example, using Goggle Docs or etherpad.
[46] For example, using ClassroomSalon (www.classroomSalon.com) or https://oeit.mit.edu/gallery/projects/nb-pdf-annotation-tool

Best Practices and Ideas from Statistics Education Literature

- Use a small number of clicker questions with a clear defined reason for each question and its placement in the lesson.
    - From:
      McGowan, H., and Gunderson, B. (2010), "A Randomized Experiment Exploring How Certain Features of Clicker Use Effect Undergraduate Student's Engagement and Learning in Statistics," *Technology Innovations in Statistics Education*, *4*. Available at http://escholarship.org/uc/item/2503w2np .

- Use clickers to promote understanding of concepts, not just calculation, for topics such as inference and applets.
    - See the following for a compilation of ideas:
      Kaplan, J. (2011), "Innovative Activities: How Clickers can Facilitate the use of Simulations in Large Lecture Classes," *Technology Innovations in Statistics Education*, *5*. Available at http://escholarship.org/uc/item/1jg0274b/.

- Ask concept questions or questions that investigate common misconceptions.
    - See the following for more information:
      Perez, C, and Lane, D. (2012), "Simulations, Audience Response Systems and the Classroom: Engaging the Modern Student," *eCOTS 12 breakout session*. Available at https://www.causeweb.org/ecots/ecots12/breakouts/8/

Best Practices and Ideas from Education Literature from Other Disciplines

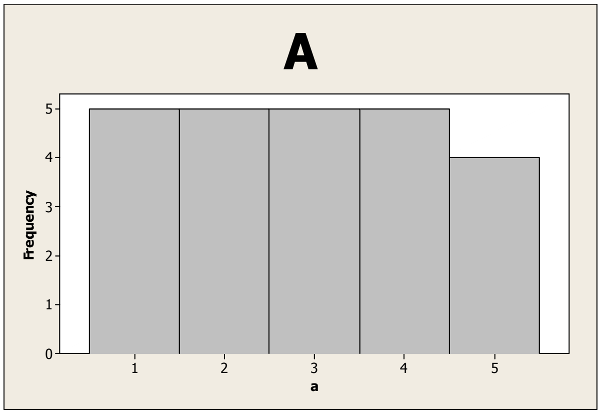
- For a summary of best practices from various academic disciplines, including life sciences and physics see the following:
  Caldwell, J. (2007), "Clickers in Large Classrooms: Current Research and Best Practice Tips," *CBE Life Sciences Education, Spring, 6*. Available at http://ncbi.nlm.nih.gov/pmc/articles/PMC1810212/.

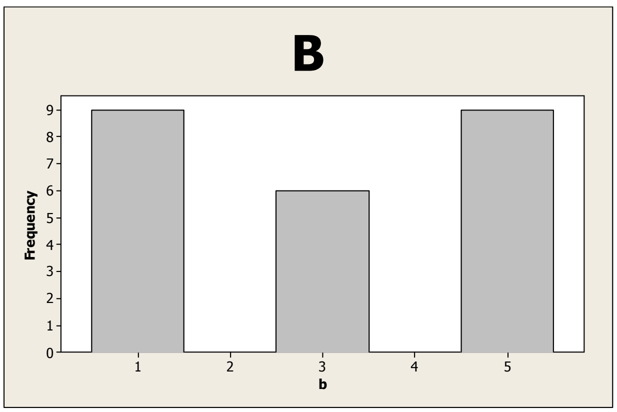
- For a summary of experiences with Peer Instruction:
  Lasry, N., Mazur, E., and Watkins, J. (2008), "Peer Instruction: From Harvard to Community Colleges," *American Journal of Physics*, *76*, 1066-1069. Available at http://dx.doi.org/10.1119/1.2978182 .

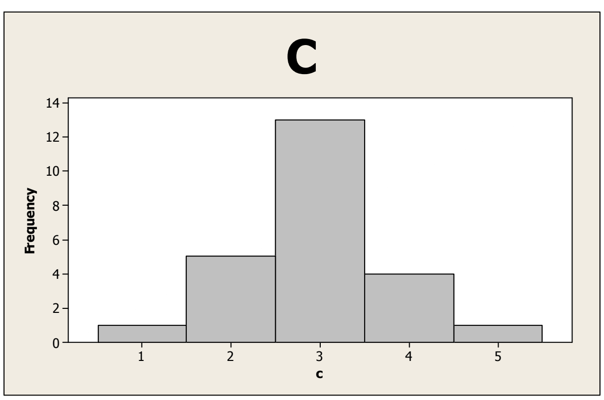
EXAMPLE ASSIGNMENT 1: USING REAL TIME RESPONSE AS AN EXAMPLE OF HOW TO BRING ATTENTION TO A MISCONCEPTION.

Teacher Resources: Slides Posing Questions

Which of the following graphs has the smallest standard deviation?







1. First have the students respond on their own using their own device.

2. After the students have responded, have the students re-answer the question after a team discussion.

Inspiration:

- Kaplan, J., Gabrosek, J., Curtiss, P., and Malone, C. (2014), "Investigating Student Understanding of a Histogram," *Journal of Statistical Education*, *22*. Available at http://www.amstat.org/publications/jse/v22n2/kaplan.pdf.
- Cooper, L.L., and Shore, F.S. (2010), "The Effects of Data and Graph Type on Concepts and Visualizations of Variability," *Journal of Statistical Education*, *18*. Available at http://www.amstat.org/publications/jse/v16n2/cooper.html.

Teaching Notes:

- A known misconception is that students seem to assume that flatter histograms mean less variation.
- The correct answer is C. In histogram C, the average distance of points from the mean is less than in the other histograms. In histograms A and B, more of the points are a further distance from the mean.

EXAMPLE ASSIGNMENT 2: USING A REAL TIME RESPONSE AS A WAY TO GUIDE A STUDENT'S EXPERIENCE WITH AN APPLET.

Teacher Resources: Slides Posing Questions

1. Do outliers affect the value of the standard deviation?
   a.) No
   b.) Yes

2. Suppose that your data set has a point that is much lower than the rest. What type of effect (if any) would this have on the value of the standard deviation?
   a.) It would make it larger.
   b.) It would make is smaller.
   c.) It would stay the same.
   d.) Unable to be determined.

3. Suppose that your data set has a point that is much higher than the rest. What type of effect (if any) would this have on the value of the standard deviation?
   a.) It would make it larger.
   b.) It would make is smaller.
   c.) It would stay the same.
   d.) Unable to be determined.

Inspiration:

Perez, C., and Lane, D. (2012), "Simulations, Audience Response Systems and the Classroom: Engaging the Modern Student," *eCOTS 12 breakout session*. Available at https://www.causeweb.org/ecots/ecots12/breakouts/8/ .

Teaching Notes:

- Correct Answers: 1.b, 2. a 3. a
- In order to focus students' attention on an applet, it is advised that the students be aware of the questions they are investigating before being exposed to the applet. The real time response systems help with this by focusing students' thoughts before their experience using the applet and then having them re-evaluate their answers at the end of the experience.
- Sometimes, students think that adding a lower point to a data set will cause the standard deviation to get smaller, similar to what would happen with the mean.
- You could also ask the students to explore this concept with their calculators or statistical software if you do not have internet access to an applet.

EXAMPLE ASSIGNMENT 3: ILLUSTRATING THE SAMPLING DISTRIBUTION OF THE SAMPLE PROPORTION

Teacher Resources: Slides Posing Questions[47]

When you flipped the coin 1 time, what proportion of heads did you get? ________

When you flipped the coin 5 times, what proportion of heads did you get? _________

When you flipped the coin 25 times, what proportion of heads did you get? _________

Student Handout

Flip a fair coin one time, five times and then twenty-five times. Enter your responses below and enter the responses in the survey mechanism. After all students have entered their data, you will be given the data from all students to graph. Make a graph of the results for n = 1, n = 5, and n = 25.

	Your Response	Sketch of results
n = 1		
n = 5		

[47] This set of questions is designed for response systems that allow for quantitative responses; however, with a little work, an instructor can create multiple responses that would work. After the responses have been received, the instructor will need to download the results and distribute them to the students. The instructor can also demonstrate making the histograms.

n = 25		

Based on your observations above . . .

- What happens to the center as *n* increases?

- What happens to the standard deviation as *n* increases?

- What happens to the shape as *n* increases?

- In your own words, why does increasing the number of flips lead to these results?

Inspiration:

Taylor, L., and Doehler, K. (2014), "Using Online Surveys to Promote and Assess Learning," *Teaching Statistics, 36*, 34- 40.

Teaching Notes:

- The objective of this activity is to explore the concepts of the sampling distribution of the sample proportion.
- Have the students flip a coin and record the proportion of heads for their series of trials. Once they have completed the flips, they should enter the data into the "Real Time" response system.

- The students should notice that as *n* increases, the bell shaped distribution starts to appear, the amount of variation decreases, but the center stays at 0.5[48].
  Kaplan, J. (2011), "Innovative Activities: How Clickers can Facilitate the use of Simulations in Large Lecture Classes," *Technology Innovations in Statistics Education, 5*. Available at http://escholarship.org/uc/item/1jg0274b/.

[48] This activity focuses on flipping a fair coin because coins are easily accessible to most students and teachers; however, you could alter the activity so that the probability of success was something other than 0.5.

APPENDIX E: Examples of Assessment Items

Well-designed assessment items help to determine whether students understand key statistical concepts. Since the original GAISE report was written in 2005, there have been many improvements in the ways that instructors and institutions determine whether students have met the learning outcomes for introductory statistics courses.

Students value that which is assessed[49], so it is important that we assess student learning in a manner consistent with our stated goals. Good items assess the development of statistical thinking and conceptual understanding, preferably using technology and real data.

Below, we present exemplary assessment items, some of which include commentary. We also present a few items that are not strong, with suggestions on how they can be improved. Finally, we present advice on constructing a rubric when assessing a project report or presentation.

Examples of Exemplary Assessment Items

We begin by providing examples of exemplary assessment items with commentary about the items. We regard these as exemplary because they reflect the GAISE recommendations of setting problems in realistic, meaningful contexts; they are data-based; and they go beyond calculation to probe deeper understanding of concepts.

ITEM 1

Scientists use metal bands to tag penguins. Do the bands harm the birds?

Researchers investigated this question with a sample of 100 penguins near Antarctica. All of these penguins had already been tagged with RFID chips, and the researchers randomly assigned 50 of them to receive a metal band on their flippers in addition to the RFID chip. The other 50 penguins did not receive a metal band. Researchers then kept track of which penguins survived for the 4.5-year study and which did not. They found that 16 of the 50 penguins with a metal band survived, compared to 31 of the 50 penguins without a metal band.

1. Calculate the difference in the proportions who survived between the two groups.

2. The p-value for comparing the two group's survival proportions turns out to be 0.005. Explain (as if to someone who has not studied statistics) what this p-value means: This is the probability of…

3. Summarize your conclusion from this p-value. Do bands hurt the penguins? Be sure to address the issue of causation as well as the issue of significance. Also justify your conclusion.

---

[49] In general, we want students to interpret results more than we want them to produce results. If we ask a True/False question, we want the student to explain *why* a statement is true or is false, so that we can assess the thinking that lead to the answer chosen. However, sometimes the practicalities of teaching a large class mean that an appropriate exam question might be a multiple choice item that does not ask for explanation.

ITEM 2

Suppose that 20% of undergraduate students at a university own an iPad and 60% of graduate students at the university own an iPad. Is it reasonable to conclude that 40% (the average of 20% and 60%) of all students at the university (undergraduate and graduate students combined) own an iPad? Explain why or why not, as if to a college student who has not taken a statistics class.

ITEM 3

Suppose that you take a random sample of 100 houses currently for sale in California. Does the Central Limit Theorem suggest that a histogram of the house prices in the sample will display an approximately normal distribution? Explain briefly.

ITEM 4

Does everyone who scores below the median on this exam necessarily have a negative z-score for this exam? Explain.

ITEM 5

Describe a situation where a third variable could be masking the relationship between two variables.[50]

ITEM 6

Suppose that Nancy, who is statistically savvy, wants to compare the average costs of textbooks for students at her college between the fall and spring semesters of last year. Let μ_F and μ_S represent the two population means. You may assume that Nancy has taken several statistics courses and knows a lot about statistics, including how to interpret confidence intervals and hypothesis tests. You have random samples from each semester and are to analyze the data and write a report. You seek advice from four persons:

1. Rudd says, "Conduct an alpha=0.05 test of H_0: $\mu_F = \mu_S$ vs. H_A: $\mu_F \neq \mu_S$ and tell Nancy whether you reject H_0."

2. Linda says, "Report a 95% confidence interval for $\mu_F - \mu_S$."

3. Steve says, "Conduct a test of H_0: $\mu_F = \mu_S$ vs. H_A: $\mu_F \neq \mu_S$ and report to Nancy the p-value from the test."

4. Gloria says, "Compare $\overline{\mathbf{y}}_1$ to $\overline{\mathbf{y}}_2$. If $\overline{\mathbf{y}}_1 > \overline{\mathbf{y}}_2$, then test H_0: $\mu_F = \mu_S$ vs. H_A: $\mu_F > \mu_S$ using alpha =0.05 and tell Nancy whether you reject H_0. If $\overline{\mathbf{y}}_1 < \overline{\mathbf{y}}_2$, then test H_0: $\mu_F = \mu_S$ vs. H_A: $\mu_F < \mu_S$ using alpha = 0.05 and tell Nancy whether you reject H_0."

[50] Sample solution: For an observational study which assessed the association between coffee drinking and cancer, smoking status could mask (or ``confound") the relationship, since smoking could be associated with both coffee drinking and cancer (see also Appendix B, Multivariable Thinking).

Rank the four pieces of advice from worst to best and explain why you rank them as you do. That is, explain what makes one better than another.

Examples of Assessment Items Needing Improvement and Commentary

We next give some examples of assessment items with problems and commentary about the nature of the difficulty. We recommend that questions such as these should either be improved as discussed in the following section or dropped from use.

Assessment items to avoid using on tests: traditional True/False, pure computation without a context or interpretation, items with too much data to enter and compute or analyze, or items that only test memorization of definitions or formulas.

ITEM 7

A teacher taught two sections of elementary statistics last semester, each with 25 students, one at 8:00 a.m. and one at 4:00 p.m. The means and standard deviations for the final exams were 78 and 8 for the 8:00 a.m. class and 75 and 10 for the 4:00 p.m. class. In examining these numbers, it occurred to the teacher that the better students probably sign up for 8:00 a.m. class. So she decided to test whether the mean final exam scores were equal for her two groups of students. State the hypotheses and carry out the test.[51]

ITEM 8

An economist wants to compare the mean salaries for male and female CEOs. He gets a random sample of 10 of each and does a t-test. The resulting p-value is 0.045.[52]

1. State the null and alternative hypotheses.
2. Make a statistical conclusion.
3. State your conclusion in words that would be understood by someone with no training in statistics.

ITEM 9

Which of the following gives the definition of a p-value?[53]

A. It's the probability of rejecting the null hypothesis when the null hypothesis is true.
B. It's the probability of not rejecting the null hypothesis when the null hypothesis is true.
C. It's the probability of observing data as extreme as that observed.
D. It's the probability that the null hypothesis is true.

---

[51] Critique: The teacher has all the population data so there is no need to do statistical inference. In addition, the proposed design has serious flaws in terms of statistical practice.

[52] Critique: The question doesn't address the conditions necessary for a t-test, and with the small sample sizes, they are almost surely violated here. Salaries are almost surely skewed.

[53] Critique: None of these answers is quite correct. Answers B and D are clearly wrong; answer A is the level of significance; and answer C would be correct if it continued "…or more extreme, given that the null hypothesis is true."

Examples Showing Ways to Improve Assessment Items

ITEM 9 (REVISITED)
Which of the following gives the definition of a p-value?

CHANGED TO:
A randomized trial of the use of bed nets to prevent malaria in sub-Saharan Africa yielded a p-value of 0.001. Without resorting to jargon, interpret this result in the context of this study to someone without background knowledge of statistics.[54]

True/False items, even when well-written, do not provide much information about student knowledge because there is always a 50% chance of getting the item right without any knowledge of the topic. One approach is to change the items into forced-choice questions with three or more options.

ITEM 10
The value of the standard deviation of a data set depends on the center of the distribution. True or False

CHANGED TO:
Does the size of the standard deviation of a data set depend on the center of the distribution?
A. Yes, the higher the mean, the higher the standard deviation.
B. Yes, because you have to know the mean to calculate the standard deviation.
C. No, the size of the standard deviation is not affected by the location of the distribution.
D. No, because the standard deviation only measures how the values differ from each other, not how they differ from the mean.

ITEM 11
A correlation of +1 indicates a stronger association than a correlation of -1. True or False

REWRITTEN AS:

A recent article in an educational research journal reports a correlation of +0.8 between math achievement and overall math aptitude. It also reports a correlation of -0.8 between math achievement and a math anxiety test. Which of the following interpretations is the most correct?

A. The correlation of +0.8 indicates a stronger relationship than the correlation of -0.8.
B. The correlation of +0.8 is just as strong as the correlation of -0.8.
C. It is impossible to tell which correlation is stronger.

Context is important for helping students see and deal with statistical ideas in real-world situations.

---

[54] Sample solution: If bed nets were not associated with malaria prevalence then we'd only be likely to see a result this extreme or more extreme one time out of a thousand. Therefore we conclude that bed nets are very likely to prevent malaria.

ITEM 12

Once it is established that X and Y are highly correlated, what type of study needs to be done to establish that a change in X causes a change in Y?

A CONTEXT IS ADDED:

A researcher is studying the relationship between an experimental medicine and T4 lymphocyte cell levels in HIV/AIDS patients. The T4 lymphocytes, a part of the immune system, are found at reduced levels in patients with the HIV infection. Once it is established that the two variables – dosage of medicine, and T4 cell levels – are highly correlated, what type of study needs to be done to establish that a change in dosage causes a change in T4 cell levels?

A. correlational study
B. controlled experiment
C. prediction study
D. survey

Try to avoid repetitious/tedious calculations on exams that may become the focus of the problem for the students at the expense of concepts and interpretations.

ITEM 13

It was claimed that 1 out of 5 cardiologists takes an aspirin a day to prevent hardening of the arteries. Suppose the claim is true. If 1,500 cardiologists are selected at random, what is the probability that at least 275 of the 1,500 take an aspirin a day?[55]

ITEM 14

A first-year program course used a final exam that contained a 20-point essay question asking students to apply Darwinian principles to analyze the process of expansion in major league sports franchises. To check for consistency in grading among the four professors in the course, a random sample of six graded essays were selected from each instructor. The scores are summarized in the table below. Construct an ANOVA table to test for a difference in means among the four instructors.

Instructor			Scores			
Affinger	18	11	10	12	15	12
Beaulieu	14	14	11	14	11	14
Cleary	19	20	15	19	19	16
Dean	17	14	17	15	18	15

ITEM 14 (REVISITED)[56]

[55] Critique: This problem requires use of software to calculate the exact binomial or use of the normal approximation to the binomial. Computer output might be provided to augment this question and facilitate solution.

[56] Critique: The version of the question above requires a fair amount of pounding on the calculator to get the results and never even asks for an interpretation. The revision below still requires some calculation (which can be adjusted depending on the amount of computer output provided) but the calculations can be done relatively efficiently---especially by students who have a good sense of what the computer output is providing.

A first-year program course … (same intro as above) … The scores are summarized in the table below, along with some descriptive statistics for the entire sample and a portion of the one-way ANOVA output.

Descriptive Statistics

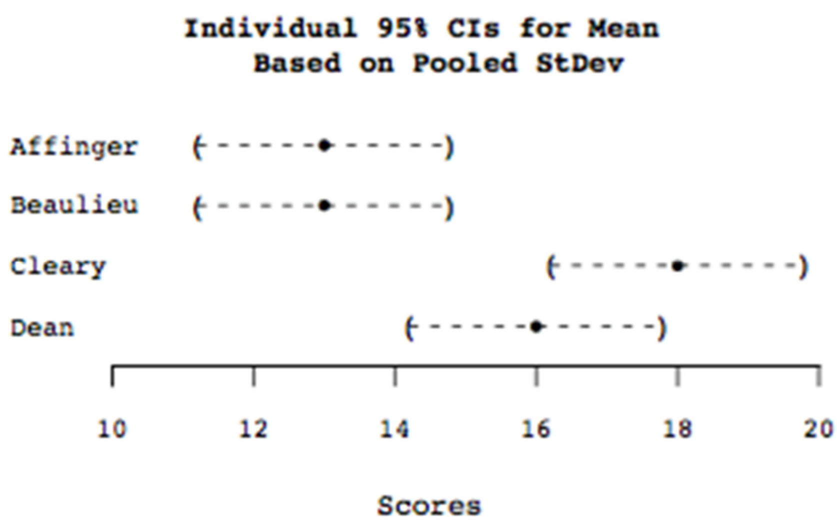
Variable	N	Mean	Median	TrMean	StDev	SEMean
Score	24.00	15.00	15.00	15.00	2.92	0.60

One-way Analysis of Variance

****ANOVA TABLE OMITTED****

Level	N	Mean	StDev
Affinger	6	13.00	2.97
Beaulieu	6	13.00	1.55
Cleary	6	18.00	2.00
Dean	6	16.00	1.55

Pooled StDev = 2.098



1. Unfortunately, we are missing the ANOVA table from the output. Use the information given above to construct the ANOVA table and conduct a test (5% level) for any significant differences among the average scores assigned by the four instructors. Be sure to include hypotheses and a conclusion. If you have trouble getting one part of the table that you need to complete the rest (or the next question), make a reasonable guess or ask for assistance (for a small point fee).

2. After completing the ANOVA table, construct a 95% confidence interval for the average score given by Dr. Affinger. *Note: Your answer should be consistent with the graphical display.*

Additional Examples of Good Assessment Items

ITEM 15

A study found that individuals who lived in houses with more than two bathrooms tended to have higher blood pressure than individuals who lived in houses with two or fewer bathrooms. Can a cause-and-effect conclusion be drawn from this? Why or why not?

ITEM 16

Researchers took random samples of subjects from two populations and applied a test to the data; the p-value for the test, using a non-directional (one-sided) alternative, was 0.06. For each of the following, say whether the statement is true or false and why.

1. There is a 6% chance that the two population distributions really are the same.

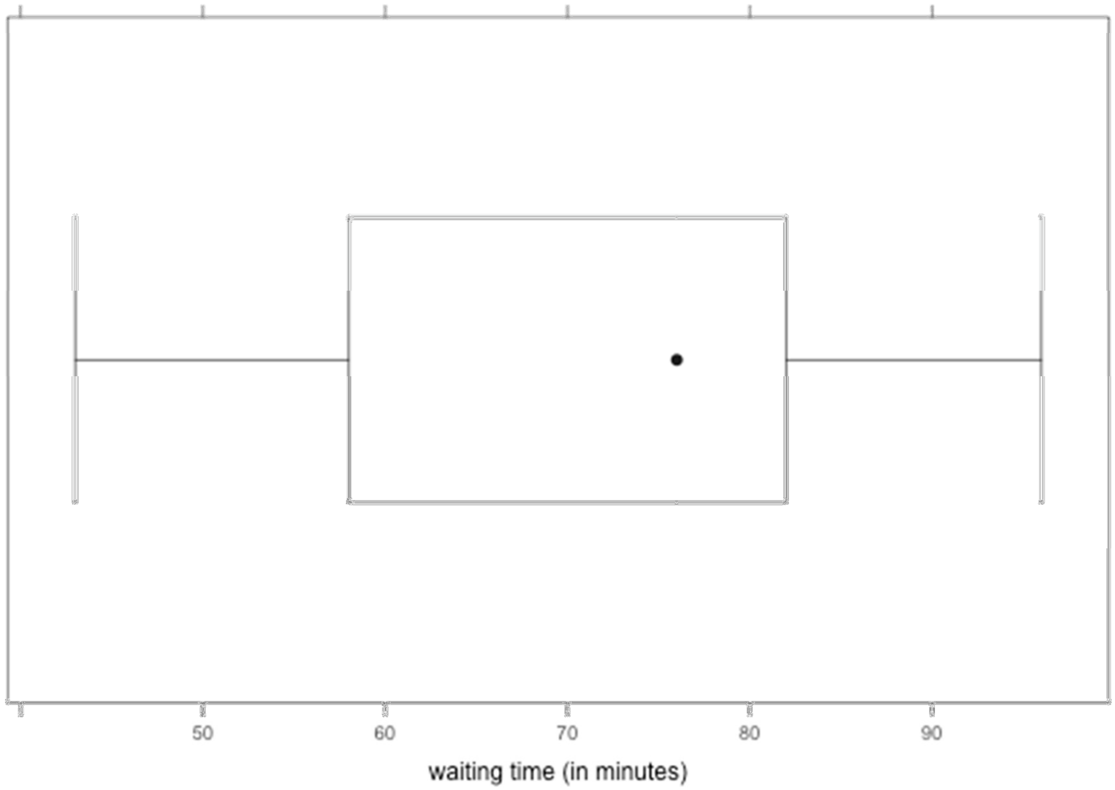
2. If the two population distributions really are the same, then a difference between the two samples as extreme as the difference that these researchers observed would only happen 6% of the time.

3. If a new study were done that compared the two populations, there is a 6% probability that H_0 would be rejected again.

4. If alpha = 0.05 and a directional alternative were used, and the data departed from H_0 in the direction specified by the alternative hypothesis, then H_0 would be rejected.

ITEM 17

As the name suggests, the Old Faithful geyser in Yellowstone National Park has eruptions that come at fairly predictable intervals, making it particularly attractive to tourists. Here is a boxplot of the times between eruptions recorded by an observer.



You are a busy tourist and have only 10 minutes to sit around and watch the geyser. But you can choose when to arrive. If the last eruption occurred at noon, what time should you arrive at the geyser to maximize your chances of seeing an eruption?

1. 12:50pm
2. 1:00pm
3. 1:05pm
4. 1:15pm
5. 1:25pm

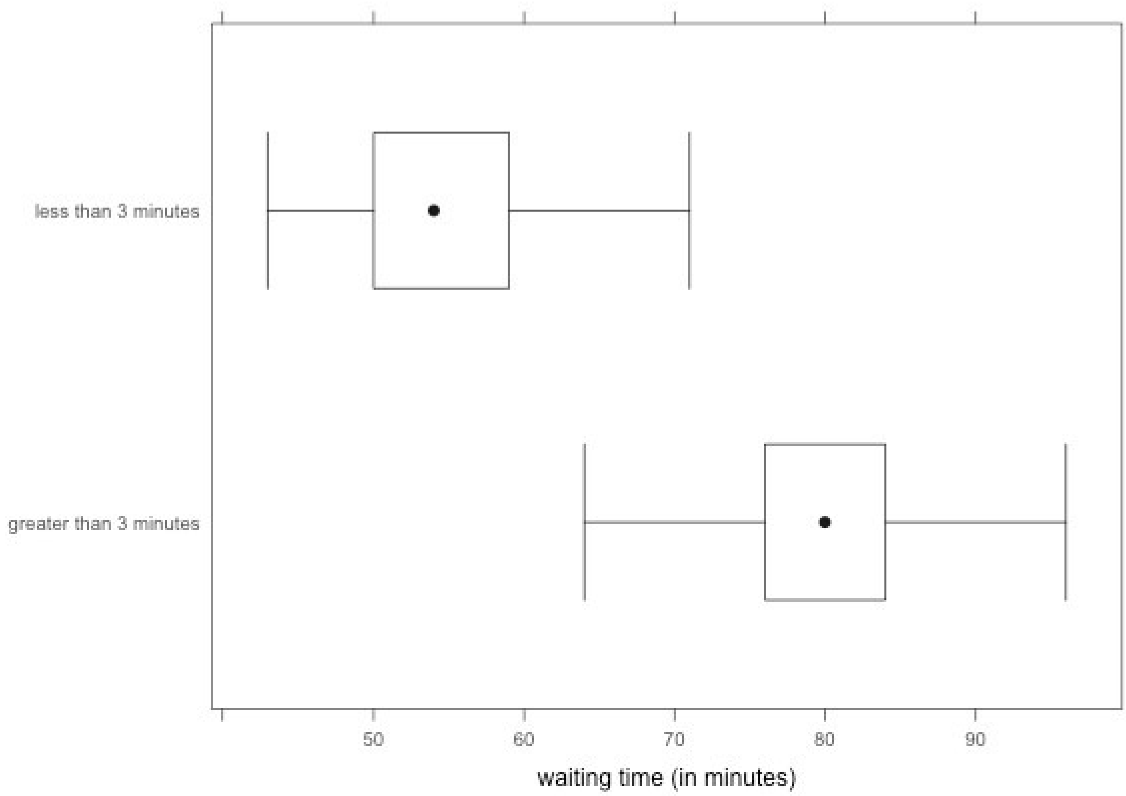
Roughly, what is the probability that in the best 10-minute interval, you will actually see the eruption?

1. 5%
2. 10%
3. 20%
4. 30%
5. 50%
6. 75%

A simple measure of how faithful is Old Faithful is the interquartile range. What is the interquartile range, according to the boxplot above?

1. 10 minutes
2. 15 minutes
3. 25 minutes
4. 35 minutes
5. 50 minutes
6. 75 minutes

Not only are you a busy tourist, you are a smart tourist. Having read about Old Faithful, you understand that the time between eruptions depends on how long the previous eruption lasted. Here's a box plot indicating the distribution of inter-eruption times when the previous eruption duration was less than three minutes.



You can easily ask the ranger what was the duration of the previous eruption. What is the best 10-minute interval to return (after a noon eruption) so that you will be most likely to see the next eruption, given that the previous eruption was less than three minutes in duration?

1. 12:30 to 12:40
2. 12:40 to 12:50
3. 12:50 to 1:00

4. 1:15 to 1:25
5. 1:25 to 1:35

How likely are you to see an eruption if you return for the most likely 10-minute interval?

1. 5%
2. 10%
3. 20%
4. 30%
5. 50%
6. 75%

ITEM 18

An article on the CNN web page begins with the sentence, "Family doctors overwhelmingly believe that religious faith can help patients heal, according to a survey released Monday." Later, the article states, "Medical researchers say the benefits of religion may be as simple as helping the immune system by reducing stress," and Dr. Harold Koenig is reported to say that "people who regularly attend church have half the rate of depression of infrequent churchgoers."

Use the language of statistics to critique the statement by Dr. Koenig and the claim, suggested by the article, that religious faith and practice help people fight depression. You will want to select some of the following words in your critique: observational study, experiment, blind, double-blind, precision, bias, sample, spurious, confounding, causation, association, random, valid, and reliable.

ITEM 19

A student weighed a sample of 100 industrial diamonds. She found that the sample average weight was 4.80 grams and the SD was 0.28 grams. *In the context of this setting,* explain what is meant by the sampling distribution of an average.

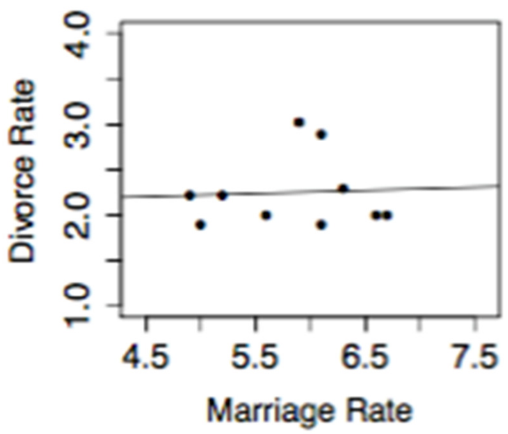
ITEM 20

A gardener wishes to compare the yields of three types of pea seeds---type A, type B, and type C. She randomly divides the type A seeds into three groups and plants some in the east part of her garden, some in the central part of the garden, and some in the west part of the garden. Then, she does the same with the type B seeds and type C seeds.

1. *What kind* of experimental design is the gardener using?
2. *Why* is this kind of design used in this situation? (Explain in the *context of the situation*.)

ITEM 21

The scatterplot shows how divorce rate and marriage rate (both as number per year per 1000 adults) are related for a collection of 10 countries. The regression line has been added to the plot.



1. The U.S. is not one of the 10 points in the original collection of countries. It happens that the U.S. has a higher marriage rate than any of the 10 countries. Moreover, the divorce rate for the U.S. is higher than one would expect, given the pattern of the other countries. How would adding the U.S. to the data set affect the regression line? Why?

2. Think about the scatterplot and regression line after the U.S. has been added to the data set. Provide a sketch of the residual plot. Label the axes and identify the U.S. on your plot with a triangle.

ITEM 22

Researchers wanted to compare two drugs, formoterol and salbutamol, in aerosol solution to a placebo for the treatment of patients who suffer from exercise-induced asthma. Patients were to take a drug or the placebo, do some exercise, and then have their "forced expiratory volume" measured. There were 30 subjects available.

1. Should this be an experiment or an observational study? Why?
2. Within the context of this setting, what is the placebo effect?
3. Briefly explain how to set up a randomized blocks design (RBD) here.
4. How would an RBD be helpful? That is, what is the main advantage of using an RBD in a setting like this?

ITEM 23

For each of the following three settings, state the type of analysis you would conduct (*e.g.*, one-sample t-test, regression, chi-square test of independence, chi-square goodness-of-fit test, etc.) if you had all the raw data and specify the explanatory and response variable on which you would perform the analysis, but *do not actually carry out the analysis.*

1. A student measured the effect of exercise on pulse for each of 13 students. She measured pulse before and after exercise (doing 30 jumping jacks) and found that the average change was 55.1 and the SD of the changes was 18.4. How would you analyze the data?

2. Three HIV treatments were tested for their effectiveness in preventing progression of HIV in children. Of 276 children given drug A, 259 lived and 17 died. Of 281 children given drug B, 274 lived and seven died. Of 274 children given drug C, 264 lived and 10 died. How would you analyze the data?

3. A researcher was interested in the relationship between blood pressure and physical activity. He measured the blood pressure and weekly total number of steps from a Fitbit for 125 women. How would you analyze these data?

ITEM 24

To compare a quantitative response variable across four groups, I selected random samples from each of the four groups and constructed parallel dotplots to compare the distributions across the four groups. I then conducted a test of H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ and rejected H_0 at the alpha = 0.05 level. I also tested H_0: $\mu_1 = \mu_2 = \mu_3$ and rejected H_0 at the alpha = 0.05 level. However, when I tested H_0: $\mu_2 = \mu_3$ using alpha = 0.05, I did *not* reject H_0. Likewise, when I tested H_0: $\mu_1 = \mu_4$ using alpha = 0.05, I did *not* reject H_0.

1. Sketch a graph of the parallel dotplots of the data. That is, based on what I told you about the tests, you should have an idea of how the data look. Use that idea to draw a graph. Indicate the sample means with triangles that you add to the dotplots.

2. It is possible to get data with the same sample means that you graphed in part 1, but for which the hypothesis H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ is not rejected at the alpha = 0.05 level. Provide a graph of this situation. That is, keep the same sample means (triangles) you had from part 1, but show how the data would have been different if H_0 were not to be rejected.

ITEM 25

Students collected data on a random sample of 12 breakfast cereals. They recorded x = fiber (in grams/ounce) and y = price (in cents/ounce). A scatterplot of the data shows a linear relationship. The fitted regression model is

$$\hat{y} = 17.42 + 0.62x$$

The sample correlation coefficient (r) is 0.23. The standard error of the sample slope is 0.81. Also, $s_{y|x} = 3.1$.

1. Find r^2 and interpret r^2 in the context of this problem.

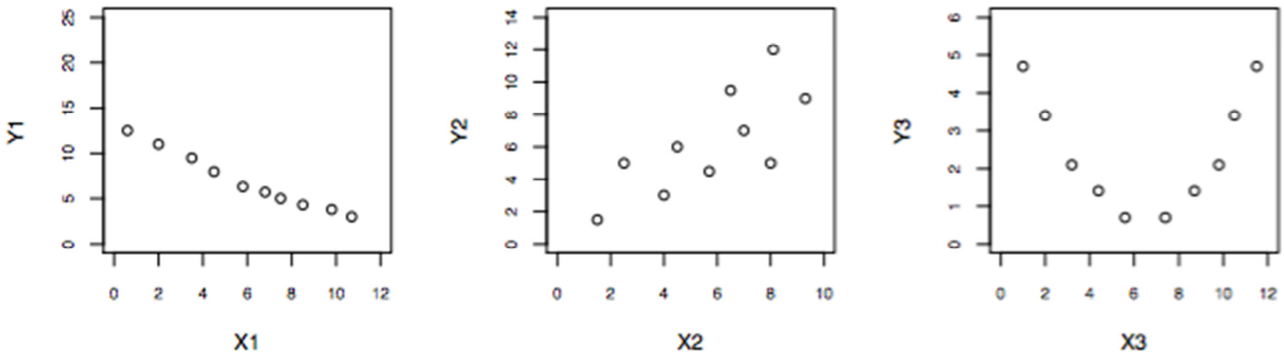
2. Suppose a cereal has 2.63 grams of fiber/ounce and costs 17.3 cents/ounce. What is the residual for this cereal?

3. Interpret the value of $s_{y|x}$ in the context of this problem. That is, what does it mean to say that $s_{y|x} = 3.1$?

4. In the context of this problem, explain what is meant by “the regression effect.”

ITEM 26

Give a rough estimate of the sample correlation for the data in each of the scatterplots below.



ITEM 27

Identify whether a scatterplot would or would not be an appropriate visual summary of the relationship between the variables. In each case, explain your reasoning.

1. Blood pressure and age
2. Region of country and opinion about stronger gun control laws
3. Verbal SAT and math SAT score
4. Handspan and gender (male or female)

ITEM 28

The paragraphs that follow each describe a situation that calls for some type of statistical analysis. For each, you should:

1. Give the name of an appropriate statistical procedure to apply (from the list below). You may use the same procedure more than once, and some questions might have more than one correct answer.

2. In some problems, you will also be given a p-value. Use it to reach a conclusion for that specific situation. Be sure to say something more than just Reject H_0 or Fail to Reject H_0. (Assume a 5% significance level.)

Some statistical procedures you might choose:

Confidence interval (for a mean, p, …)	Normal distribution
Determining sample size	Correlation
Test for a mean	Simple linear regression
Test for a proportion	Multiple regression
Difference in means (paired data)	Two-way table (chi-square test)
Difference in means (two independent samples)	ANOVA for difference in means
Difference in proportions	Two-way ANOVA for means

A. Researchers were commissioned by the Violence In Children's Television Investigative Monitors (VICTIM) to study the frequency of depictions of violent acts in Saturday morning TV fare. They selected a random sample of 40 shows that aired during this time period over a 12-week period. Suppose 28 of the 40 shows in the sample were judged to contain scenes depicting overtly violent acts. How should they use this information to make a statement about the population of all Saturday morning TV shows?

B. In one of his adventures, Sherlock Holmes found footprints made by the criminal at the scene of a crime and measured the distance between them. After sampling many people, measuring their height and length of stride, he confidently announced that he could predict the height of the suspect. How?

C. Anthropologists have found two burial mounds in the same region. They know several tribes lived in the region and that the tribes have been classified according to different lengths of skulls. They measure a random sample of skulls found in each burial mound and wish to determine if the two mounds were made by different tribes. (p-value = 0.0082)

D. The Career Planning Office is interested in seniors' plans and how they might relate to their majors. A large number of students are surveyed and classified according to their MAJOR (Natural Science, Social Science, Humanities) and FUTURE plans (Graduate School, Job, Undecided). Are the type of major and future plans related? (p-value = 0.047)

E. *Sophomore Magazine* asked a random sample of 15 year olds if they were sexually active (yes or no). They would like to see if there is a difference in the responses between boys and girls. (p-value = 0.029)

F. Every week during the Vietnam War, a body count (number of enemy killed) was reported by each army unit. The last digits of these numbers should be fairly random. However, suspicions arose that the counts might have been fabricated. To test this, a large random sample of body count figures was examined and the frequency with which the last digit was a 0 or a 5 was recorded. Psychologists have shown that people making up their own random numbers will use these digits less often than random chance would suggest (*i.e.*, 103 sounds like a more "real" count than 100). If the data were authentic counts, the proportion of numbers ending in 0 or 5 should be about 0.20. (p-value = 0.002)

G. The Hawaiian Planters Association is developing three new strains of pineapple (call them A, B, and C) to yield pulp with higher sugar content. Twenty plants of each variety (60 plants in all) are randomly distributed into a two-acre field. After harvesting, the resulting pineapples are measured for sugar content and the yields are recorded for each strain. Are there significant differences in average sugar content between the three strains? (p-value = 0.987)

ITEM 29

Some of the statistical inference techniques we have studied include:

A. One-sample z-procedures for a proportion
B. Two-sample z-procedures for comparing proportions
C. One-sample t-procedures for a mean
D. Two-sample t-procedures for comparing means
E. Paired-sample t-procedures
F. Chi-square procedures for two-way tables
G. ANOVA procedures
H. Linear regression procedures

For each of the following research questions, indicate (by letter) the appropriate statistical inference procedure for investigating the question.[57]

1. Economists compared starting salaries of new employees across three different groups: those with graduate degrees, those with only bachelor's degrees, and those with no higher education degrees.

2. A researcher investigated whether laughter increases blood flow by having subjects watch a humorous movie and a stressful movie, randomly deciding which movie the subject would see first, measuring the blood flow through the person's blood vessels while watching the movie.

3. Student researchers investigated whether balsa wood is less elastic after it has been immersed in water. They took 44 pieces of balsa wood and randomly assigned half to be immersed in water and the other half not to be. They measured the elasticity by seeing how far (in inches) the piece of wood would project a dime into the air.

4. Do more than two-thirds of students at a particular university have at least one class on Fridays during this term?

5. Are people more likely to fill in the missing letter in F A I _ with an L if they are given a red pen rather than a blue pen?

6. Is there an association between a college student's level of drinking alcohol (classified as none, some, or considerable) and her/his residence situation (classified as living on-campus, off-campus with parents, or off-campus but not with parents)?

7. A researcher used data from the American Time Use Survey (ATUS) to investigate whether high school math teachers tend to spend more time working per day than high school history teachers.

8. Biologists recorded the frequency of a cricket's chirps (in chirps per minute) and also the temperature (in degrees Fahrenheit) when the cricket measurement was recorded. They investigated whether chirp frequency is a significant predictor of temperature.

[57] The list of methods or examples can be shortened.

ITEM 30

How accurate are radon detectors sold to homeowners? To answer this question, university researchers placed 12 radon detectors in a chamber that exposed them to 105 picocuries per liter of radon. The detector readings found are below, along with some descriptive statistics.[58]

91.9 97.8 111.4 122.3 105.4 95.0
103.8 99.6 96.6 119.3 104.8 101.7

Variable	N	Mean	Median	TrMean	StDev	SEMean	Minimum	Q1	Q3
readings	12	104.13	102.75	103.51	9.40	2.71	122.30	96.90	109.90

1. Is there convincing evidence that the mean 20 readings of all detectors of this type differs from the true value of 105? Perform the appropriate hypothesis test with alpha = 0.05.
2. Explain what a Type I error associated with this situation would be.
3. Explain what a Type II error associated with this situation would be.
4. What is the probability of a Type II error if the reading of the detectors is too low by 5 picocuries (really 100 when it should read 105)?

ITEM 31

According to a U.S. Food and Drug Administration (FDA) study, a cup of coffee contains an average of 115 mg of caffeine, with the amount per cup ranging from 60 to 180 mg depending on the brewing method. Suppose you want to repeat the FDA study to obtain an estimate of the mean caffeine content to within 5 mg with 95% using your favorite brewing method. How many cups of coffee must you brew to be 95% confident? In problems such as this, we can estimate the standard deviation of the population to be 1/4 of the range.

ITEM 32

An internet company is planning to test which of two online ad campaigns is more effective in generating clicks on their site. Outline the design of an experiment you would use to achieve this goal. Assume you have money to place 500 ads for each of the two possible campaigns.

ITEM 33

A study of iron deficiency among infants compared samples of infants following different feeding regimens. One group contained breast-fed infants, while the children in another group were fed a standard baby formula without any iron supplements.

A graphical display indicated that the blood hemoglobin levels in children (both breast-fed and formula-fed) are approximately normally distributed in each group. Here are the summary results on blood hemoglobin levels at 12 months of age:

Group	Sample size	Sample mean	Sample SD
Breast-fed	230	13.3	1.7

[58] This item might be improved by providing more output (*e.g.*, 95% confidence interval) to allow students to tackle it without calculation or use of a table.

Formula	230	12.4	1.8

The two-sample t-test yielded a test statistic of 5.51 with 458 degrees of freedom. This is associated with a two-sided p-value that was less than 0.0001.

Interpret the results from the test statistic and p-value that are provided. Be sure to report the observed difference in groups in the context of the problem.

ITEM 34

A group of physicians subjected polygraph testing to the same careful testing given to medical diagnostic tests. They found that if 1,000 people were subjected to the polygraph and 500 told the truth and 500 lied, the polygraph would indicate that approximately 185 of the truth-tellers were liars and 120 of the liars were truth-tellers. In the application of the polygraph test, an individual is presumed to be a truth-teller until indicated that s/he is a liar. What is a Type I error in the context of this problem? What is the probability of a Type I error in the context of this problem? What is a Type II error in the context of this problem? What is the probability of a Type II error in the context of this problem?

ITEM 35

Audiologists recently developed a rehabilitation program for hearing-impaired patients in a Canadian program for senior citizens. A simple random sample of the 30 residents of a particular senior citizens home and the seniors were diagnosed for degree and type of sensorineural hearing loss which was coded as follows: 1 = hear within normal limits, 2 = high-frequency hearing loss, 3 = mild loss, 4 = mild-to-moderate loss, 5 = moderate loss, 6 = moderate-to-severe loss, and 7 = severe-to-profound loss. The data are as follows:

6	7	1	1	2	6	4	6	4	2	5	2	5	1	5
4	6	6	5	5	5	2	5	3	6	4	6	6	4	2

1. Create a boxplot of the data.
2. Write a brief description of the distribution of the data.
3. Find a 95% confidence interval for the mean hearing loss of senior citizens in this Canadian program. The mean and standard deviation of the above data are 4.2 and 1.808, respectively. Interpret the interval.

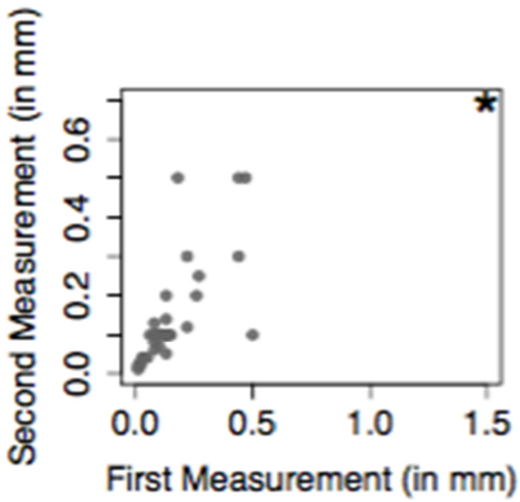
ITEM 36

A utility company was interested in knowing if agricultural customers would use less electricity during peak hours if their rates were different during those hours. Customers were randomly assigned to continue to get standard rates or to receive the time-of-day structure. Special meters were attached that recorded usage during peak and off-peak hours; the technician who read the meter did not know what rate structure each customer had.

1. Is this an observational study or experiment? Defend your answer.
2. What are the explanatory and response variables?
3. Identify a potential confounding variable in this work.
4. Is this a matched-pair design? Defend your answer.

ITEM 37

At the beginning of the semester, we measured the width of a page in our statistics book two times. Below is the scatterplot of the first measurement vs. the second measurement.



1. Describe the relationship between the variables.
2. What effect does the starred point have on the correlation coefficient? That is, if the starred point were removed, how would the correlation coefficient change, if at all?

ITEM 38

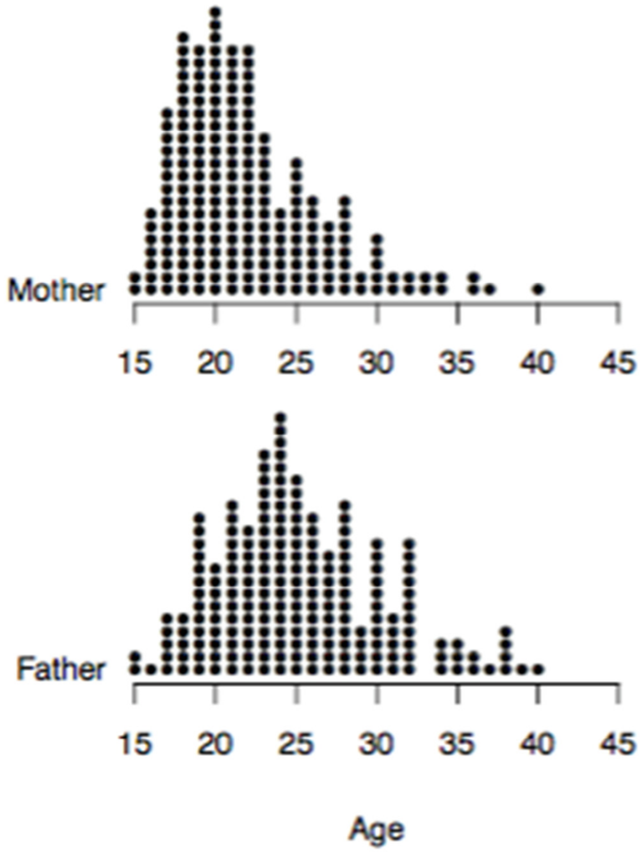
A study in the *Journal of Leisure Research* investigated the relationship between academic performance and leisure activities. Each in a sample of 159 high-school students was asked to state how many leisure activities they participated in weekly. From the list, activities that involved reading, writing, or arithmetic were labeled "academic leisure activities." Some of the results are in the table below:

	Mean	Standard Deviation
GPA	2.96	0.71
Number of leisure activities	12.38	5.07
Number of academic leisure activities	2.77	1.97

Based on these numbers (and knowing that the GPA is a value between 0 and 4 and the number of activities cannot be negative), discuss the potential skewness of each of the above variables.

ITEM 39

A random sample of 200 mothers and a separate random sample of 200 fathers were taken. The age of the mother when she had her first child and the age of the father when he had his first child were recorded.



1. Describe the data for the mothers' age.
2. Describe the data for the fathers' age.
3. Compare the distributions.
4. A suggestion is made to check the correlation between the ages if we wish to compare the two populations. Is this a good suggestion? Why or why not?

ITEM 40

When conducting a randomized experiment, the original randomization of units to treatment groups breaks the association between

1. the explanatory variable and the response variable.
2. the explanatory variable and confounding variables.
3. the response variable and confounding variables.

ITEM 41

When conducting a randomization test, the simulated re-randomization of units to treatment groups breaks the association between

1. the explanatory variable and the response variable.
2. the explanatory variable and confounding variables.
3. the response variable and confounding variables.

ITEM 42

For each of the following, circle your answer to indicate whether the quantity can NEVER be negative or can SOMETIMES be negative:

1. z-score	SOMETIMES	NEVER
2. Probability	SOMETIMES	NEVER
3. Test statistic	SOMETIMES	NEVER
4. Sample proportion	SOMETIMES	NEVER
5. Standard deviation	SOMETIMES	NEVER
6. Inter-quartile range	SOMETIMES	NEVER
7. Standard error	SOMETIMES	NEVER
8. p-value	SOMETIMES	NEVER
9. Slope coefficient	SOMETIMES	NEVER
10. Correlation coefficient	SOMETIMES	NEVER

ITEM 43

A high school statistics class wants to estimate the average number of chocolate chips in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a 95% confidence interval for the average number of chips per cookie (18.6 to 21.3). Indicate if each is VALID or INVALID.[59]

1. We are 95% confident that the confidence interval of 18.6 to 21.3 includes the true average number of chocolate chips per cookie.
   VALID INVALID
2. We are 95% confident that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.
   VALID INVALID
3. We expect 95% of the cookies to have between 18.6 and 21.3 chocolate chips.
   VALID INVALID

ITEM 44

Consider an observational study of the effects of second-hand smoke on health in which we want to compare non-smokers (i) who live with a smoker to (ii) those who do not live with a smoker. There are two ways in which independence is relevant in the sampling and data collection process. (a) Give an example in which one type of independence is met but the other is not; (b) give an example in which the other type of independence is met but the first is not.

ITEM 45

A terse report of a statistical test is given below:

> The P-value for a hypothesis test with hypotheses H_0: $\mu = 3$ versus H_A: $\mu \neq 3$ is 0.04.

Critique the following responses for clarity, completeness and correctness.

---

[59] Multiple True/False items of this sort can provide very useful information. If there is a single correct understanding for a statistical concept, but several known misunderstandings for the same concept, a multiple T/F item can provide information on whether or not a student correctly recognizes each of the misunderstandings as false or invalid.

1. This means that the probability of getting our test statistic is 0.04.

2. This means that the probability of getting a test statistic at least as extreme as ours is 0.04.

3. This means that if the null hypothesis is true, the probability of getting a test statistics at least as extreme as ours is 0.04

4. This means that if the null hypothesis is true, the probability of getting a test statistic less than or equal to the one we got is 0.04

5. This means that it is very unlikely that the result that was used to compute this P-value would have happened by pure chance alone, assuming that H_0 is true. Therefore we could conclude that the evidence is against the Null Hypothesis, and H_0 is probably not true.

6. The sentence means that assuming the population average is equal to three, the likelihood of getting an average as large as or larger than we got for our sample is about 4 percent.

7. The p-value is the probability that the data will be as extreme or more extreme as the alternate hypothesis suggests.

ITEM 46

Explain what the following sentence means:

> The interval (2.25, 2.75) is a 99% confidence interval for the mean GPA of UT students having between 45 and 60 credit hours.

Critique the following responses for clarity and correctness.

1. A 99% confidence interval is used to show that 99% of the time when you pick a sample from the population (students having between 45 and 60 credit hours) you will find a mean GPA in the interval (2.25, 2.75).

2. There is a 99% chance that $2.25 \leq \mu \leq 2.75$.

3. This means that if we took many, many simple random samples and constructed a confidence interval based on each sample, 99% of the resulting confidence intervals would contain the true mean.

ITEM 47

For each part, draw a scatterplot satisfying the conditions given, or else explain why the conditions are impossible:

1. Regression line has small positive slope and correlation is high and positive.
2. Regression line has large positive slope and correlation is high and positive.
3. Regression line has small positive slope and correlation is low and positive.

4. Regression line has large positive slope and correlation is low and positive.
5. Regression line has positive slope and correlation is negative.

ITEM 48

Rosiglitazone is the active ingredient in the controversial Type 2 diabetes medicine Avandia and has been linked to an increased risk of serious cardiovascular problems such as stroke, heart failure, and death. A common alternative treatment is pioglitazone, the active ingredient in a diabetes medicine called Actos. In a nationwide retrospective observational study of 227,571 Medicare beneficiaries aged 65 years or older, it was found that 2,593 of the 67,593 patients using rosiglitazone and 5,386 of the 159,978 using pioglitazone had serious cardiovascular problems. These data are summarized in the contingency table below.

		Cardiovascular problems		
		Yes	No	Total
Treatment	Rosiglitazone	2,593	65,000	67,593
	Pioglitazone	5,386	154,592	159,978
	Total	7,979	219,592	227,571

Determine if each of the following statements is true or false. If false, explain why. *Be careful:* The reasoning may be wrong even if the statement's conclusion is correct. In such cases, the statement should be considered false.

1. Since more patients on pioglitazone had cardiovascular problems (5,386 vs. 2,593), we can conclude that the rate of cardiovascular problems for those on a pioglitazone treatment is higher.

2. The data suggest that diabetic patients who are taking rosiglitazone are more likely to have cardiovascular problems since the rate of incidence was (2,593 / 67,593 = 0.038) 3.8% for patients on this treatment, while it was only (5,386 / 159,978 = 0.034) 3.4% for patients on pioglitazone.

3. The fact that the rate of incidence is higher for the rosiglitazone group proves that rosiglitazone causes serious cardiovascular problems.

4. Based on the information provided so far, we cannot tell if the difference between the rates of incidences is due to a relationship between the two variables or due to chance.

The next several items are based on simulation (resampling) methods.

ITEM 49

Rosiglitazone is the active ingredient in the controversial Type 2 diabetes medicine Avandia and has been linked to an increased risk of serious cardiovascular problems such as stroke, heart failure, and death. A common alternative treatment is pioglitazone, the active ingredient in a diabetes medicine called Actos.

A randomized study compared the rates of serious cardiovascular problems for diabetic patients on rosiglitazone and pioglitazone treatments. The table below summarizes the results of the study.

		Cardiovascular problems		
		Yes	No	Total
Treatment	Rosiglitazone	2,593	65,000	67,593
	Pioglitazone	5,386	154,592	159,978
	Total	7,979	219,592	227,571

1. What proportion of all patients had cardiovascular problems?

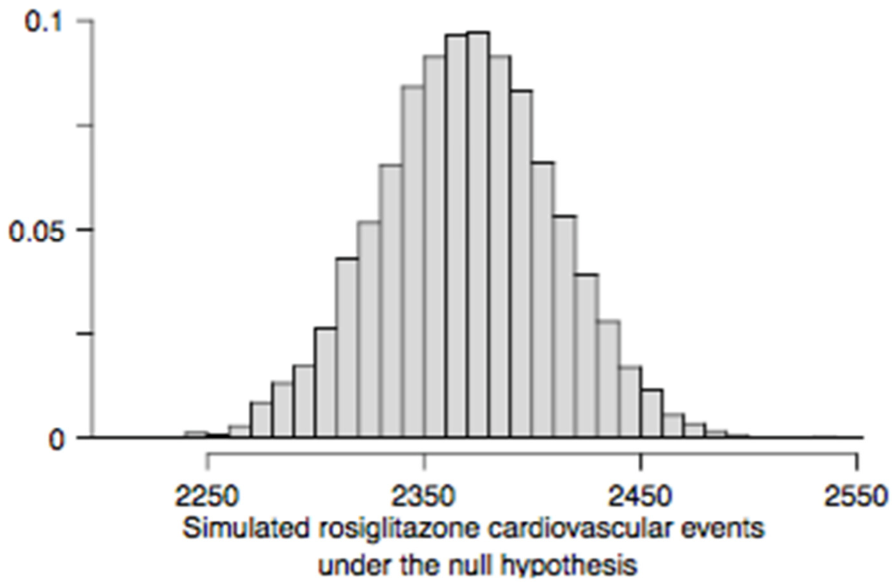
2. If the type of treatment and having cardiovascular problems were independent (null hypothesis), about how many patients in the rosiglitazone group would we expect to have had cardiovascular problems?

3. We can investigate the relationship between outcome and treatment in this study using a randomization technique. While in reality we would carry out the simulations required for randomization using statistical software, suppose we actually simulate using index cards. In order to simulate from the null hypothesis, which states that the outcomes were independent of the treatment, we write whether or not each patient had a cardiovascular problem on cards, shuffle all the cards together, and then deal them into two groups of size 67,593 and 159,978. We repeat this simulation 10,000 times and each time record the number of people in the rosiglitazone group who had cardiovascular problems. Below is a relative frequency histogram of these counts.

4. What are the claims being tested?

5. Compared to the number calculated in the second part, which would provide more support for the alternative hypothesis, *more* or *fewer* patients with cardiovascular problems in the rosiglitazone group?

6. What do the simulation results suggest about the relationship between taking rosiglitazone and having cardiovascular problems in diabetic patients?



ITEM 50

The Stanford Heart Transplant Study was a randomized trial of a new medical intervention. Of the 34 patients in the control group, 4 were alive at the end of the study. Of the 69 patients in the treatment group, 24 were alive. The contingency table below summarizes these results.

		Group		
		Control	Treatment	Total
Outcome	Alive	4	24	28
	Dead	30	45	75
	Total	34	69	103

1. What proportion of patients in the treatment group and what proportion of patients in the control group died?

2. One approach for investigating whether or not the treatment is effective is to use a randomization technique.

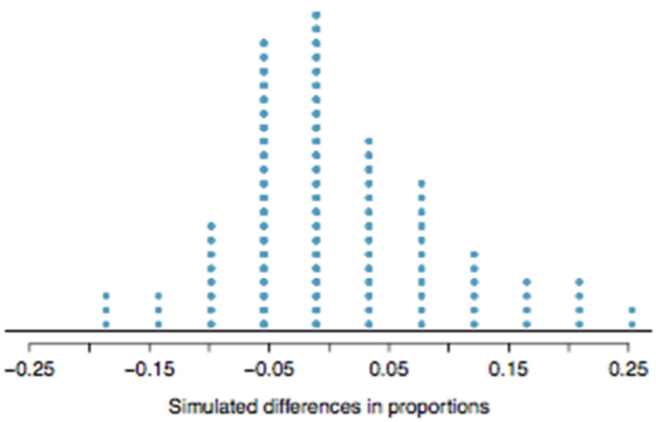
2.1 What are the claims being tested? Use correct null and alternative hypothesis notation

2.2 The steps below describes the set up for such approach, if we were to do it without using statistical software. Fill in the blanks with a number or phrase, whichever is appropriate.

- We write *alive* on _______ cards representing patients who were alive at the end of the study, and *dead* on ________ cards representing patients who were not.

- Then, we shuffle these cards and split them into two groups: one group of size _________ representing treatment, and another group of size __________ representing control.
- We calculate the difference between the proportion of *dead* cards in the treatment and control groups (treatment - control) and record this value. We repeat this many times to build a distribution centered at ____________.
- Lastly, we calculate the fraction of simulations where the simulated differences in proportions are ________.
- If this fraction is low, we conclude that it is unlikely to have observed such an outcome by chance and that the null hypothesis should be rejected in favor of the alternative.

2.3 What do the simulation results shown below suggest about the effectiveness of the transplant program?



ITEM 51

Researchers studying the effect of antibiotic treatment compared to symptomatic treatment for acute sinusitis randomly assigned 166 adults diagnosed with sinusitis into two groups. Participants in the antibiotic group received a 10-day course of an antibiotic, and the rest received symptomatic treatments as a placebo. These pills had the same taste and packaging as the antibiotic. At the end of the 10-day period patients were asked if they experienced improvement in symptoms since the beginning of the study. The distribution of responses is summarized below.

		Self-reported improvement in symptoms		
		Yes	No	Total
Treatment	Antibiotic	66	19	85
	Placebo	65	16	81
	Total	131	35	166

1. What type of a study is this?

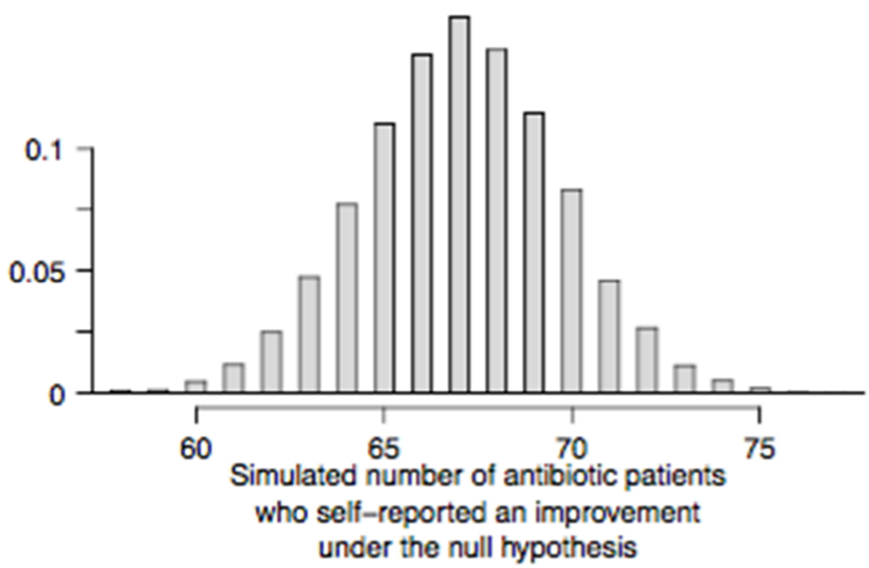
2. Does this study make use of blinding? Justify your answer.

3. Compute the difference in the proportions of patients who self-reported an improvement in symptoms in the two groups: $\hat{p}_{antibiotic} - \hat{p}_{placebo}$.

4. At first glance, does antibiotic or placebo appear to be more effective for the treatment of sinusitis? Explain your reasoning using appropriate statistics.

5. There are two competing claims that this study is used to compare: the null hypothesis that the antibiotic has no impact and the alternative hypothesis that it has an impact. Write out these competing claims in easy-to-understand language and in the context of the application.

6. Below is a histogram of simulation results computed under the null hypothesis. In each simulation, the summary value reported was the number of patients who received antibiotics and self-reported an improvement in symptoms. Write a conclusion for the hypothesis test in plain language. (Hint: Does the value observed in the study, 66, seem unusual in this distribution generated under the null hypothesis?)



Examples of Assessments for Presentations and Projects

Projects and presentations are an increasingly common component of introductory statistics courses.[60]

Projects provide an opportunity for students to learn statistics by doing statistics. They demonstrate that statistical practice includes formulating a statistical question, designing a plan

[60] Halvorsen's ICOTS 2010 paper (http://iase-web.org/documents/papers/icots8/ICOTS8_4G3_HALVORSEN.pdf) provides motivation for the use of projects as well as details of specific deliverables.

for collecting relevant data, using appropriate statistical methods for analyzing the data, and presenting results in a public setting such as a poster, oral presentation, or a paper (Halvorsen 2010).

Students have the opportunity to develop statistical questions that arise from broader research questions, to design data analysis plans, and to communicate results.

We provide a basic rubric for presentations and projects along with a sample numeric grading scheme.

Core Competency	**Needs Improvement**	**Basic**	**Surpassed**
Computation Perform computations	Computations contain errors and extraneous code	Computations are correct but contain extraneous/unnecessary computations	Computations are correct and properly identified and labeled
Analysis Choose and carry out analysis appropriate for data and context	Choice of analysis is overly simplistic, irrelevant, or missing key component	Analysis appropriate, but incomplete, or not important features and assumptions not made explicit	Analysis appropriate, complete, advanced, relevant, and informative
Synthesis Identify key features of the analysis, and interpret results (including context)	Conclusions are missing, incorrect, or not made based on results of analysis	Conclusions reasonable, but is partially correct or partially complete	Make relevant conclusions explicitly connected to analysis and to context
Visual presentation Communicate findings graphically clearly, precisely, and concisely	Inappropriate choice of plots; poorly labeled plots; plots missing	Plots convey information correctly but lack context for interpretation	Plots convey information correctly with adequate/appropriate reference information
Verbal Communicate findings in writing clearly, precisely, and concisely	Explanation is illogical, incorrect, or incoherent	Explanation is partially correct but incomplete or unconvincing	Explanation is correct, complete, and convincing

If needed, the competencies can be converted into a numeric score.

One might begin by giving a score of 85 for achieving basic competency in all 5 categories. Then we add to this score for competencies that surpass the basic level and subtract for those that need improvement. Three points might be added (subtracted) for each of the first three competencies that have surpassed the basic (need improvement), with four points added (subtracted) for the fourth competency that is surpassed (needs improvement) and five points for the fifth competency. In other words, it is increasingly challenging to surpass the basic competency, and it is increasingly problematic to not achieve basic competency. For example, if all five competencies are rated "surpassed," the score is 85 + 3*3 +4 + 5 = 100; if 4 competencies are rated "surpassed" and the fifth is "basic," then the score is 85 + 3*3 +4 = 95; and for 3 "surpassed," 1 "needs improvement," and 1 "basic," the score is 85 + 3*3 - 3 = 91. If a competency is missing, then 15 points are subtracted regardless of how many competencies are categorized as needing improvement.

APPENDIX F: Learning Environments

Instructors work in a variety of settings, and some readers may question whether they can adopt the GAISE recommendations. Different classroom situations have areas of greater and lesser challenges in the implementation of the recommendations. This appendix provides examples of ways to apply the recommendations in different environments. Five common instructional conditions explored are

- face-to-face, both large and small class sizes
- flipped (inverted) classes
- distance learning
- cooperative learning
- limited technology

The purpose of this appendix is to provide research references and a few inspiring examples for implementing GAISE teaching in courses where one or more of the recommendations for teaching appear difficult to employ.

Face-to-Face Courses

Whether instruction is in a classroom or an individual tutoring setting, instruction in statistics has primarily been in a face-to-face format. While times are changing and other approaches are now available, the majority of college and university teaching of statistics still occurs in a face-to-face environment. In these college settings, the class size ranges from small to medium to large and to what some would even call very (or extremely) large. In this appendix, we illustrate how to incorporate the GAISE recommendations in teaching situations made complex by class size. For example, some have questioned the feasibility of active learning in a large class setting. Others have found that with very small classes a simulation completed with manipulatives by the students in the class might not demonstrate the desired principle.

Small Classes:

Collecting data from students during class is suggested as a way to *foster active learning* and *integrate real data with a context and purpose*. Classes with low enrollment, however, cannot collect enough data to be used in the same ways that larger classes can.

<u>Example #1</u>: Physical Exploration

When an active, concrete illustration (*e.g.*, die rolling, card shuffling) is desirable prior to a computer simulation that demonstrates a concept, individual students can repeat the task more than once to help generate additional real data. Another alternative is to have the students complete a process and record the result just once in the classroom in order to understand the process, and then to use technology-based simulations, such as applets, to repeat the simulation many times quickly. Alternatively, the teacher could prime the pump with a simulated data set and then add class data to those starter data.

Example #2: Project Data from the Class

One way to overcome the issue of collecting a large enough sample for use in a class-focused project is to keep records of the data collected over several semesters. Another option is to collect and share data with colleagues across multiple sections of the course. Beginning with the data collected from the class members, a conversation of the limitations of the small sample size can motivate additional data collection from a larger sample of non-classmates.

To *teach statistical thinking*, *focus on conceptual understanding*, or *foster active learning,* peer-to-peer interactions are often an integral part of the educational experience. A small class necessarily means fewer peers to interact with, creating challenges for instructors.

Example #3: Cooperative Groups

Some faculty find that using cooperative groups is a great strategy for teaching statistics (see the last section of this appendix and Appendix C: *Activities, Projects, Data*). Small classes limit the size of the groups and/or the number of groups. Pairing of students after initial dyad discussions provides an opportunity to leverage collaboration. Although is tempting in a small class to let individual students work to their strengths, rotating the group member's roles ensures all students have opportunities to lead, record, present, etc. Distributing responsibility to individual students for presentation of some of the "light" topics in the course can nurture the sense of ownership for learning among the classmates.

Example #4: Student Presentation of the Results

Fewer students allow time for students to report results from their small group (or individual) work to the entire class. Peer review/evaluation of such presentations offers additional interaction, whether written or verbal, immediate or after class.

Large Classes:

While large classes provide a great opportunity for collecting large data sets, they produce their own set of challenges. For example, many excuses have been heard to not *foster active learning* through the use of groups in large classes: "the chairs do not move," "I won't be able to talk to all the groups," "it will be too loud," etc. Carbone (1998) indicates that active/cooperative learning can be effective in large classes as well as small ones and provides suggestions to *foster active learning* in large classes. For very large classes, an assistant, who might even be an advanced student, could be helpful for group supervision (Davidson 1990). In some cases there may be an opportunity to break a large statistics course into separate smaller lab or discussion sections in which group work and activities could be used.

Gelman and Nolan (2002) report that with careful selection, activities can be used successfully in large statistics classes and strongly encourage group work to promote student learning. In particular, they suggest that when selecting activities for large classes, choose those in which the majority of students remain seated and a limited number of students go to the board or make a presentation to the class.

Regardless of the class size, involving students in the course is important. Gelman and Nolan (2002) provide an example of "Active Homework": Throughout the semester, they suggest assigning pairs of students to go to the library to find data that is needed for class or brought up in a class discussion. For a small class, by the end of the semester, the entire class could have this experience. Another approach to this same activity is to have students find data from the web to bring to class.

Example #1: Working with Partners in Positive, Productive Ways

A modification of the think-pair-share method that has been recommended for large classes by Blumberg (2015) can be remembered with the acronym FSLC. These letters help the students remember to **Formulate** the answer on their own first, then **Share** it with a partner. The acronym specifically encourages important partner behaviors, and teachers are encouraged to not omit the last two steps of the process which are to ensure that students **Listen** carefully to the answer of their partner and then **Create** a new answer that uses both partner's information in a manner so that the new answer is better than each of the individual answers.

Example #2: Cooperative Groups

In large, tiered lecture halls with fixed chairs, students may find it logistically easier to work in pairs instead of groups. The instructor can use a think-pair-share structure and randomly call on a student to report the thinking for their group. Asking the question and then using a random number generator to determine the student to be selected can help keep students in a large class alert. Sampling with replacement ensures students know they could be called on again at any time.

Example #3: Data Collection Using Class Polling

Zullo and Cline's book *Teaching Mathematics with Classroom Voting: With and Without Clickers* includes three chapters on using clickers in introductory statistics courses. Examples include lesson plans for box plots, hypothesis testing, confidence intervals, and data collection. Furthermore, the text describes how to select lessons that are good for using classroom voting and how to use these approaches for developing conceptual understanding. Specific examples can be found in Appendix D: *Examples of Using Technology*.

Example #4: Using Online Surveys to Maximize Class Time

With the ever increasing number of free or low-cost online tools for developing surveys, professors can maximize class time by setting up online surveys to collect data either in (via a mobile device such as a phone) or out of class as an efficient means of data collection. The article by Taylor and Doehler (2014) includes activity ideas and implementation details for using survey software for data collection in introductory statistics. Specific examples can be found in Appendix D: *Examples of Using Technology*.

References/Resources:

Blumberg, P. (2008), *Developing Learner-centered Teaching: A Practical Guide for Faculty*. San Francisco: Jossey-Bass.

Blumberg, P. (2015), "Student Participation/Active Learning," University of the Sciences' Teaching and Learning Center. Available at http://www.usciences.edu/teaching/tips/spal.shtml#within

Carbone, E. (1998), *Teaching Large Classes: Tools and Strategies*, SAGE Publications.

Davidson, N. (1990), "Small-group Cooperative Learning in Mathematics," *Teaching and Learning Mathematics in the 1990s. The NCTM Yearbook,* 52-61.

Gelman, A., and Nolan, D. (2002), *Teaching Statistics: A Bag of Tricks*, Oxford: Oxford University Press.

Taylor, L., and Doehler, K. (2014), "Using Online Surveys to Promote and Assess Learning," *Teaching Statistics*, *36*, 34-40.

Zullo, H., and Cline, K. (2012), *Teaching Mathematics with Classroom Voting: With and Without Clickers*, MAA Notes [Washington, DC]: MAA, Mathematical Association of America.

Flipped (Inverted) Classes

With audio, video, and even graphical materials easier to develop and make available online, opportunities for efficiently sharing materials with students are expanding. Some faculty are utilizing those types of technology to restructure in-class and out of-class learning environments. Faculty use technology to provide online lectures for students to listen to and learn from outside of class. Now they have evolved into videos of the teaching providing lecture-type instruction or of slides with a voiceover which uses animations to help students visualize a concept. In some cases, students watch these videos prior to class. The students then come to the classroom ready to engage in *active learning* to solve more sophisticated problems than would be easy to solve at home or in isolation. This type of learning structure has been called the "flipped" or "inverted" classroom. The inverted classroom model offers multiple opportunities, both in and out of the classroom, for helping students develop *statistical thinking* and *conceptual understanding*.

Example 1: Out-of-class Videos and In-class Problem Solving Sections

Lape et al. (2014) found that students who watch videos outside of class and use class time in Engineering and Mathematics for problem solving sessions believe that the class time helped them learn the concepts more than students in the corresponding traditionally taught courses.

Example 2: Motivating Reading

Wilson (2013) provides incentives for reading the textbook outside of class by giving "reading quizzes," and while only around 60% of the students rated the readings as helpful, they did significantly increase the amount of reading they did for the course as compared to when the course was taught in the traditional lecture approach. Wilson used a model where the lecture material was moved outside of class and the homework, which was presented as application-type activities that could be done individually or in groups, was completed during the scheduled class time. She found for both student-reported learning and for student final exam grades that the flipped classroom teaching model was significantly better than the traditional model. Thus, in terms of the GAISE recommendations, instructors could focus reading and the corresponding quizzes on *conceptual understanding*. Furthermore, using the flipped classroom approach including reading quizzes on concepts could, without loss of content, make more time during the class period for engaging students in significant *active learning activities* (See also Appendix C: *Activities, Projects, Data*.)

Example 3: Informed Classroom Instruction

Strayer (2014) recommends teachers convey information to students outside of class to gain a response from the students prior to coming to the class. These responses should be viewed by teachers before class to better inform them how to teach the class during the face-to-face time with the students. It is particularly helpful if the teacher can construct the task to reveal students' conceptions as well as their misconceptions. Building on the knowledge gained from the out-of-class material, the teacher can be prepared to efficiently structure class discussions to extend the students' knowledge. While the task for Strayer's research was an algebra lesson for pre-service teachers, the lessons learned are transferable to flipped introductory statistics courses.

References/Resources:

Bishop, J., and Verleger, M. (2013), "The Flipped Classroom: A Survey of the Research," in *120th ASEE Annual Conference and Exposition*.

Flipped Learning Network. Available at http://fln.schoolwires.net/Page/1

Lape, N., Levy, R., Yong, D. H., Haushalter, K. A., Eddy, R., and Hankel, N. (2014), "Probing the Inverted Classroom: A Controlled Study of Teaching and Learning Outcomes in Undergraduate Engineering and Mathematics," Paper presented at 2014 ASEE Annual Conference, Indianapolis, Indiana.

Rayens, W. (2012), "Teaching Statistical Concepts in an Inverted Classroom," University of Kentucky. Available at http://www.causeweb.org/webinar/teaching/2012-01/.

Strayer, J. (2012), "How Learning in an Inverted Classroom Influences Cooperation, Innovation and Task Orientation," *Learning Environments Research*, *15*, 171-193.

Wilson, S. G. (2013), *Teaching of psychology: The flipped class: A method to address the challenges of an undergraduate statistics course.* Lawrence Erlbaum Associates, Inc.

Distance Learning

As technology continues to advance, the use of online instruction to teach statistics has also evolved. The instruction in the online course can be asynchronous, synchronous or partially synchronous. Some online courses are broadcast live to an audience that can both see and hear the professor and the teacher can see and hear all of the students in remote locations at the same time. Since an online class is sometimes a result of students having work schedules which make meeting face-to-face difficult, the online class can also take on an asynchronous format where students watch videos of the instructor or the textbook author on their own time. Another format for the online course is partially synchronous in which there is a combination of face-to-face meetings and online instruction. These courses may have different names such as hybrid, blended, or web-enhanced, and they may have different percentages of time spent in the online or the face-to-face environment. Today, there are Massive Open Online Courses (MOOCs) which provide opportunities for learning to tens of thousands of people. Moreover, the MOOCs can be led by an instructor with a set schedule for completing course materials or can be completely self-paced.

Complete definitions and best practices for each of these learning environments can be found at the Hidden Curriculum webpage (Abbott 2014). Common themes for best practices include maximizing the strengths of each approach to foster interaction between students for discussions and collaborations regarding learning. For the partially synchronous environment, similar to the flipped classroom environment, instructors should carefully consider how to use the in-class time to maximize learning based on the goals of the course.

As faculty design online courses based on the GAISE recommendations, the following examples about technology might be useful in making decisions about the learning environment in which different content is delivered.

<u>Example #1</u>: Data Collection

Teachers can *integrate real data with a context and a purpose* in distance learning environments by using online surveys for collecting data which can be shared with the entire class. This type of data collection about the students can create interest and foster interactions.

<u>Example #2:</u> Discussion Boards

Discussion boards can be used to have students describe how they would use statistics in their major. This can help students connect with other students in their major while they are learning more about applications of statistics in the real world. Critique of journalistic efforts to report scientific research can engender online conversation even asynchronously. The instructor can set up specific “question and answer” assignments where the students can use the discussion board to help each other better understand the material. Discussing the choice of analytical tool – by students for their coursework or by researchers whose reports are being critiqued – provides opportunity for *statistical thinking, focusing on concepts, using technology,* and *multivariable thinking.*

Example #3: Simulations

The teacher can create short videos demonstrating a simulation using an applet. Then the students can follow the example in the video to run their own simulations for similar problems, using *technology to explore concepts*. (See also Appendix D: *Examples of Using Technology*.) Students can be asked to post decision responses telling what they learned from a simulation to encourage *statistical thinking* and serve as an *assessment to improve and evaluate student learning*.

References/Resources:

Abbott, S. (Ed) (2014), "Hidden curriculum," *The glossary of education reform*. Available at http://edglossary.org/hidden-curriculum

Everson, M. G., and Garfield, J. (2008), "An Innovative Approach to Teaching Online Statistics Courses," *Technology Innovations in Statistics Education, 2*. Available at http://escholarship.org/uc/item/2v6124xr.

Mills, J. D., and Raju, D. (2011), "Teaching Statistics Online: A Decade's Review of the Literature about What Works," *Journal of Statistics Education, 19*. Available at http://www.amstat.org/publications/jse/v19n2/mills.pdf

Mocko, M. (2013), "Selecting Technology to Promote Learning in an Online Introductory Statistics Course," Technology Innovations in Statistics Education, 7. Available at http://escholarship.org/uc/item/596195sg.

Tudor, G. (2006), "Teaching Introductory Statistics Online – Satisfying the Students," *Journal of Statistics Education*, *14*. Available at http://www.amstat.org/publications/jse/v14n3/tudor.html

Ward, B. (2004), "The Best of Both Worlds: A Hybrid Statistics Course," *Journal of Statistics Education*, *12*. Available at http://www.amstat.org/publications/jse/v12n3/ward.html

Utts, J., Sommer, B., Acredolo, C., Maher, M., and Matthews H. (2003), "A Study Comparing Traditional and Hybrid Internet-Based Instruction in Introductory Statistics Classes," *Journal of Statistics Education*, *11*. Available at http://www.amstat.org/publications/jse/v11n3/utts.html

Cooperative Learning

A cluster of teaching/learning techniques (with a variety of names and purposes) that involve students working together can provide opportunities for implementing GAISE recommendations into statistics courses. Team-based (St. Clair and Chihara 2012), student-driven (Sovak 2010), cooperative (Garfield 1993) or collaborative (Roseth, Garfield, and Ben-Zvi 2008) learning, and guided investigations (Bailey, Spence, and Sinn 2013) have nuances as outlined in the given references, but all come down to opportunities to *foster active learning* in the classroom and *integrate real data with a context and a purpose*, often necessitating the use of *technology to analyze it*. The actual tasks assigned to small groups of students might incorporate the remaining recommendations by focusing on *statistical thinking* and *conceptual understanding*.

For institutions or instructors who design entire courses around these types of instruction, we provide a few more examples in addition to the larger collection in Appendix C: *Activities, Projects, Data*.

<u>Example #1</u> - Histogram Comparisons

Each student is assigned a pair of histograms (out of four such pairs) for which they must determine which has more variability. They then discuss their reasoning with a partner until consensus is gained on both pairs of histograms. New partnerships are made and each student must explain the reasoning to the new partner for both their own and their original partner's histograms. In the end, every member of the foursome has four well-reasoned examples for determining the relative size of variability. *Active learning* and a *conceptual understanding* of variability are inherent in this activity.

Inspired by Roseth, Garfield, and Ben-Zvi (2008).

<u>Example #2</u> – Coin Distribution

Students bring coins from home (specify pennies, nickels, etc.) and in their groups they sort them by minting dates, calculating the ages. Several descriptive graphs, tables, or measures might be made on the small collection before compiling the data from all the groups into the classroom sample and making further descriptives, perhaps using technology. The *use of real data* in practicing the construction of dot plots, histograms, and other graphs brings *active learning* to the classroom. This might be a review of earlier descriptive topics and/or serve as a launching point for whole class discussion of sample size, limitations due to sampling methods, outliers, and/or informal inference.

Inspired by National Council of Teachers of Mathematics (2014). *Principles to Actions: Ensuring Mathematical Success for All.* Reston, Virginia: National Council of Teachers of Mathematics.

<u>Example #3</u> – NFL Quarterback Salaries

In general, a jigsaw activity gives different information to different group members so that it requires cooperation and discussion to fit the pieces together before the final question(s) can be answered. Determining the best predictor (Pass Completion %, Touchdowns, or Yards per Game) of quarterback salary through r and r^2 can be just such a task. Providing each group member with one data set to compare to the salary allows each student the practice in calculating

r and r^2. In order to come to consensus on the *best* predictor; however, comparisons of numbers, graphs, and appropriate use of vocabulary is required. Follow-up questions can tap *statistical thinking* and *conceptual understanding* in addition to the *use of real data*, *active learning*, and *technology* that was used to analyze the data. There is also opportunity to address the reality of multi-variate predictors.

Inspired by STatistics Education Web, http://www.amstat.org/education/STEW/.

References/Resources:

Bailey, B., Spence, D. J., and Sinn, R. (2013), "Implementation of Discovery Projects in Statistics," *Journal of Statistics Education, 21*. Available at http://www.amstat.org/publications/jse/v21n3/bailey.pdf

[NSF-funded curriculum available at http://faculty.ung.edu/DJSpence/NSF/materials.html]

Garfield, J. (1993), "Teaching Statistics Using Small-Group Cooperative Learning," *Journal of Statistics Education*, *1*. Available at http://www.amstat.org/publications/jse/v1n1/garfield.html

Roseth, C. J., Garfield, J. B., and Ben-Zvi, D. (2008), "Collaboration in Learning and Teaching Statistics," *Journal of Statistics Education*, *16*. Available at http://www.amstat.org/publications/jse/v16n1/roseth.html

Sovak, M.M. (2010), *The Effect of Student-Driven Projects on the Development of Statistical Reasoning* (Doctoral Dissertation). Retrieved from http://iase-web.org/Publications.php?p=Dissertations.

St. Clair, K., and Chihara, L. (2012), "Team-Based Learning in a Statistical Literacy Class," *Journal of Statistics Education*, *20*. Available at http://www.amstat.org/publications/jse/v20n1/chihara.pdf

Limited Technology

Technology has had many forms and definitions over the years. Today, certain forms of technology are considered standard in many introductory statistics courses; however, that definition of "standard" varies by institution. Some classrooms for teaching statistics have statistical analysis software and software for visualizing statistical simulations on computers while other institutions might not have a computer lab large enough to seat an entire statistics class. Some courses meet in computer labs, while others might have a weekly lab session. At some schools, the only computer is at the instructor's station, while at others, students bring their own devices to class. And, unfortunately, there are schools where students don't have access to any technology at all. Even for instances when instructors feels that their classroom environments might be "technology deprived," there are still ways to provide instruction which support the GAISE recommendations.

Example #1: Teacher Demonstrations

If the course instructor has access to a computer projection system, the teacher can bring a laptop into the classroom to demonstrate using statistical software to analyze large data sets and provide an opportunity for students to see that computation is the least important task of a statistician. Instructors can often receive complementary or discounted statistical analysis software. Many of these tools include instructional videos to help the student learn how to use the software outside of teacher instructional time. Teachers can also demonstrate applets which allow students to quickly observe results of a simulation that would be too time-consuming using physical manipulatives. Online statistical analysis tools and applets can be found in the statistics education digital library, *CAUSEweb.org*, using the advanced search tool. (See Appendix D: *Examples of Using Technology* for additional examples.)

After demonstrating how to use statistical software, the teacher can provide the students with a handout of examples of statistical output for practice at interpreting analysis results. Students should be able to answer questions on exams interpreting the output of statistical software, such as the question below.

- The body temperatures of a random sample of 65 healthy male college students were taken. Researchers wanted to know if the body temperature of college-age males is different from the "normal" body temperature of 98.6 degrees Fahrenheit. Use the output from a statistical software package to test the researcher's hypothesis at the 5% significance level. Then state your conclusion in the context of the problem.

Hypothesis test results:
μ : Mean of variable
$H_0 : \mu = 98.6$
$H_A : \mu \neq 98.6$

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
male	98.104615	0.086669986	64	-5.7157574	<0.0001

Inspired by the ARTIST Assessment Builder, https://apps3.cehd.umn.edu/artist/user/login.asp.
Note: A free registration is required for using the Assessment Builder.

Example #2: Calculators for Statistical Analysis

When statistical software packages are not an option, graphing calculators with pre-programmed statistical functions can be used to minimize time spent by students on computation and maximize time spent on *conceptual understanding* and interpreting the statistical output in the context of the given problems. In conjunction with the demonstration of statistical software, students should understand that calculators are not the technology of practicing statisticians or researchers from other fields doing data analysis.

Example #3: Physical Manipulatives

Classrooms with limited technology should not be deprived of physical manipulatives or opportunities to actively engage students in learning statistics. For example, having students make a "living boxplot" based on student data is a great way to get the students to actually be the manipulatives and to visualize what it means to have 25% of the data in a given region. Scheaffer et al. (2004) provide instructions for that and other rich activities which might only require paper and a ruler to *teach statistical thinking, focus on conceptual understanding,* and *foster active learning*.

References/Resources:

ARTIST Assessment Builder. https://apps3.cehd.umn.edu/artist/user/login.asp. Assessment Builder FAQ https://apps3.cehd.umn.edu/artist/QA.html.

CAUSEweb.org. The Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) Digital Library for Statistics Education Resources, Research, and Professional Development.

Scheaffer, R., Watkins, A., Witmer, J., and Gnanadesikan, M. (2004), *Activity-based Statistics, Instructors Guide,* 2nd Ed., revised by Erickson, T. Key College Publishing.

Texas Instruments, TI 84 Activity Central. *Statistics: Find Activities that Support your Lessons*. Available at https://education.ti.com/en/84activitycentral/us/Statistics.