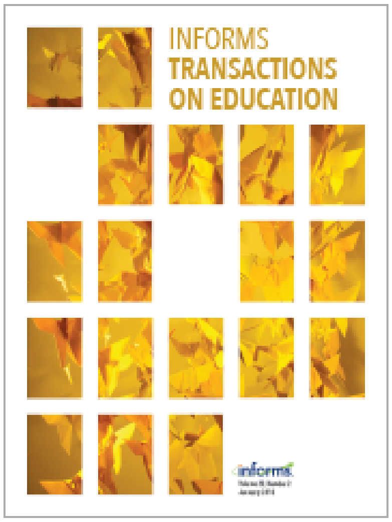


INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dynamic Interactive Visualizations: Implications of Seeing, Doing, and Playing for Quantitative Analysis Pedagogy

Shailesh S. Kulkarni, Bin Mai, S. Yasaman Amirkiaee, Hakan Tarakci

informs
http://pubsonline.informs.org/journal/ited/

INFORMS TRANSACTIONS ON EDUCATION
Vol. 19, No. 3, May 2019, pp. 121–142
ISSN 1532-0545 (online)

Dynamic Interactive Visualizations: Implications of Seeing, Doing, and Playing for Quantitative Analysis Pedagogy

Shailesh S. Kulkarni,[a,*] Bin Mai,[b] S. Yasaman Amirkiaee,[a] Hakan Tarakci[a]

[a] Information Technology and Decision Sciences Department, College of Business, University of North Texas, Denton, Texas 76201;
[b] Educational Administration and Human Resource Development, Texas A&M University, College Station, Texas 77843

*Deceased

Contact: binmai@tamu.edu (BM); seyedeyasaman.amirkiaee@unt.edu (SYA); hakan.tarakci@unt.edu, http://orcid.org/0000-0001-9509-5409 (HT)

Received: September 13, 2016
Accepted: October 11, 2018
Published Online in Articles in Advance: May 16, 2019

https://doi.org/10.1287/ited.2018.0203

Abstract. Teaching relatively complex quantitative topics in statistics, operations management, and management science to undergraduate as well as graduate students can pose numerous pedagogical challenges. However, several topics in these disciplines are amenable to exposition by means of dynamic interactive visualizations. We present a sample of such visualizations and discuss their implications for pedagogy. We also discuss how optimization modeling and a single data set can be leveraged to expose students to multiple variants of covering models and, more importantly, how visualizations can be used to quickly demonstrate the differences in these models. We present empirical evidence that by using dynamic interactive visualizations we are able to enhance the student learning experience.

Keywords: visualization • interactive • quantitative analysis • pedagogy

1. Introduction and Motivation

This paper introduces the use of dynamic interactive visualizations as effective teaching aids for quantitative analysis. Our goal is to show how such visualizations can be leveraged quite effectively to demonstrate relatively complex topics in business statistics (Stat), operations management (OM), and operations research/management science (OR/MS). Furthermore, we show how optimization modeling and data can be tied intimately to and brought alive with visualizations. We present evidence that the use of dynamic interactive visualizations indeed improves the student learning experience.

Compared with other literature in visualization as a pedagogical tool, in this paper, we emphasize "dynamic interactivity" as an important aspect of all our visualizations. Usually, software visualization tools are defined as "a collection of techniques which allows [users] to see the innards of program execution clearly" (Eisenstadt et al. 1992, p. 335), and in pedagogical practice, "visualization of the activities...might be an important aspect for improving...education. Computer-based visualizations like animations and simulations are effective teaching-learning resources." (Banerjee et al. 2013, p. 176). However, according to Naps et al. (2002), such a technology, no matter how well it is designed, is of little educational value unless it engages learners in an active learning activity. Researchers have attempted to use a wide variety of visualization tools in pedagogical practice to investigate their impact on student learning (e.g., Kumar 2015, Shoufan et al. 2015, and Garay et al. 2019). In this paper, we present the application of visualization tools enhanced by dynamic interactivity, which has received little attention in the existing literature, and show that such an approach indeed enhances students' learning engagement and improves their learning experience.

We use the dynamic interactivity moniker somewhat broadly. It could mean the ability to, during the visualization presentation, change parameters and see the impact on graphs, charts, and outcomes instantly. It could also signify the ability to see visual "pop-outs" (Keim et al. 2006, 2008, 2015) as a user plays around with a visualization. It could also mean clicking and selecting graphical objects and visualizing their impact. We provide such examples in this paper. Finally, such visualizations could be used for data analysis, estimation and hypothesis testing, optimization, or simply to illustrate complex concepts that derive from multiple methodological streams.

Furthermore, we focus on the dynamic interactivity capability provided specifically by Mathematica. Mathematica is a modern technical computing system with some unique native capabilities that are arguably not easily replicable in other computing systems. Researchers have long recognized the values of potent technical computing systems such as Mathematica. Ramsden (1997, p. 15) stated that the new (at the time) technology tool Mathematica "opens up new, exciting avenues for educators" and explained, "the revolutionary potential of new technology lies, however, in our finding new things to do with it." Since its first release in 1998, Mathematica has been widely adopted in education subjects such as math (e.g., Liu and Deng 2011), science and technology (e.g., Sato et al. 2003), and engineering (e.g., Saunders-Smits et al. 2005). However, the more recent advanced features, especially the interactive visualization functionalities, provided by Mathematica have just started to be integrated in teaching practices, and the evaluation of the impacts of such features on pedagogy has been sparse.

Compared with other computing systems such as Microsoft (MS) Excel and Visual Basic for Applications (VBA) that provide some visualization capabilities, Mathematica has some specific advantages in terms of education delivery. Excel falls short when it comes to reasonably serious, large-scale models and computation. For example, the sales representative (rep) example that we discuss in Section 3.3 cannot be solved with the student version of Excel. Creating interactive manipulations needs a single command, which is provided by the eponymous "Manipulate" in Mathematica and acts as a generalized wrapper to all other computations held within it. This is an advantage of the functional programming paradigm that Mathematica is built on. VBA, on the other hand, is procedural and thereby comes with some of the common shortcomings of procedural languages. More importantly, the sheer computing capability and vast array of inbuilt functions that one can leverage with Mathematica have made it a complete technical computing system. To that end, if one were to go beyond what we present in this paper (and the authors have, in several other advanced classes), then we would be remiss not to mention that Mathematica holds a unique spot in the vast space of tools used to create dynamic interactive visualizations.

Stat, OR/MS, and OM courses have typically been taught in traditional classroom settings. More recently, several universities have been moving toward online models of pedagogy, as is evident from the ubiquitous use of Blackboard and such other online pedagogical delivery portals. Online teaching is perhaps unparalleled in its ability to allow students to individually customize the time they need to understand and master a given topic. Besides this, discussion boards, wikis, and customized chat rooms for topics can enhance discussion and the free exchange of ideas among students who would otherwise either work alone or form their own little cliques. We believe, however, that both the online and especially the traditional classroom experience can be enhanced with dynamic interactive visualizations. To this end, we elucidate our experience in the classroom when using several such visualizations and also summarize the perceived student experience.

The authors of this study teach large sections of Stat and OR/MS courses at the undergraduate level. They also teach decision modeling at the graduate level. The typical student profile at the undergraduate level comprises students from all majors in the Stat course and predominantly operations supply chain, logistics, and decision sciences majors in the OR/MS course. The graduate-level decision modeling course appeals to a broader student body comprising majors as diverse as music and finance to journalism. At the undergraduate level, we deliver online quizzes and homework assignments using customized software because it allows for self-paced multimedia-interfaced learning. In all courses, we allow the use of laptop computers during class sessions and during examinations. We have tutoring labs and extensively use classroom technology such as dual screens, Apple TV, visualizers, and student response systems ("clickers") that allow us to simultaneously use multiple sources of information and analysis. Despite harnessing these multiple modes of delivering pedagogy, we find that students struggle with quantitative subject matter and find it difficult to master, which, in turn, leads to their failing to appreciate the conceptual issues in our discipline. Some of this is the result of student maturity as well as teaching these courses in large sections, but we believe that we can further assist our students in integrating the conceptual framework in our body of knowledge by including a set of dynamic interactive visualizations that allows for students to "see," "do," and "play" in real time. An inadvertent positive externality of this approach is that it also breaks the monotony of traditional lecture delivery. On the basis of our interactions with students over the years and from comments that we have received in our formal course evaluations, we believe that a dynamic visual component added to pedagogical delivery leads to furthering of the conceptual understanding of seemingly difficult subject matter and makes it more accessible to students.

The remainder of this paper is organized as follows. In Section 2, we provide a brief review of the literature that specifically addresses the need for effectiveness of visualizations for pedagogy. In Section 3 (and in three subsections therein), we discuss specific visualizations that we have used for Stat-, OM-, and MS-related topics. In Section 4, we discuss a classroom exercise that students do with "covering" models. This exercise begins with a hands-on (paper and pencil) exercise using an actual map and extends to our dynamic interactive visualizations. We use a single data set to demonstrate several types of covering models and provide visualizations for each model, which makes their individual exposition and highlighting differences between them much easier than traditional algebraic analysis. We also provide the empirical evidence that the dynamic interactive visualization techniques we adopted in our teaching do improve the student learning outcomes significantly. In Section 5, we present the empirical results of the impacts of our visualization approach on the student learning experience. Finally, in Section 6, we conclude with the limitations of using visualizations in the classroom and ideas for future research.

2. Visualizations for Quantitative Analysis Pedagogy: A Brief Literature Review and a Conceptual Model

2.1. A Brief Literature Review

Among the studies that discuss the instructional role of visualizations, several of them directly concentrate on the management field. However, they focus on introducing or developing a specific visualization tool or software within a specific area for solving a specific problem. One of the earliest attempts in this field has been summarized by Jones (1996), who discusses visualization application in optimization with applications to network routing, facility location, and so on. In some newer works, Baloukas et al. (2005, 2009) demonstrate the role of visualization in pedagogy by introducing new software that provides an animated demonstration of the uncapacitated primal network simplex algorithm. Markham and Palocsay (2006) focus on the "scenarios" visualization with Excel's built-in Scenario tool and provide some examples in breakeven analysis, queuing theory, and decision analysis. Lazaridis et al. (2007) introduce new software on linear programming to improve pedagogy. Akçay et al. (2012) use information visualization and data-mining techniques to examine and interpret the results of data envelopment analysis (DEA) models by SmartDEA software. Evans (2015) emphasizes the prescriptive analytics aspect of using visualization and interprets the optimization results of standard Excel Solver output reports in a simple way to show the visualization effectiveness for facilitating data communication. Additionally, Andrienko et al. (2007) call for concerted cross-disciplinary efforts for incorporating geovisual analytics in support of space-related decision making.

In the next section, we review studies on brain mechanisms during the learning process and how visualization can facilitate this process. We also propose our model for the visualization role and position in pedagogy based on cognitive load theory and literature reviews.

2.2. Developing a Conceptual Model

To demonstrate how we use visualization as an instructional tool, we first describe the human brain mechanism when the learning process occurs. In fact, precise recognition of human perception and cognition systems is an initial step in designing appropriate instructional tools. Therefore, we start with cognitive load theory, originally developed in the early 1980s. Its main purpose is to help design instructional tools, and it has considered information structures, knowledge of human cognition, and their interactions (VanMerrienboer and Sweller 2005). *Cognitive load* is defined as the amount of cognitive resources required for accomplishing a function within the brain. Cognitive load can also be considered as *memory demand* (Wickens et al. 2015), which is controllable by manipulating instructional formats (Huang et al. 2009).

As Huang et al. (2009, p. 141) explain, cognitive load theory emphasizes the limitations of working memory and considers cognitive load associated with the learning of complex cognitive tasks as the major factor determining the success of instructional methods. To provide more space of memory for knowledge acquisition, the instructional aid should be designed in alignment with cognitive processes.

The main factors that affect working memory load can be placed into two categories: intrinsic cognitive load and extraneous cognitive load. The first category includes the intrinsic nature of the learning materials or tasks themselves, and the second one includes the manner in which the materials or tasks are presented. Instructional interventions can alter the extraneous cognitive load (Van Merrienboer and Sweller 2005). Sweller et al. (1998) show that instructors can use visualization as an instructional tool to generate two effects to decrease extraneous cognitive load: split attention effect and redundancy effect. Split attention effect reduces extraneous cognitive load by replacing multiple sources of information (usually pictures and the accompanying text) with a single, integrated source of information so that there is no need to mentally integrate them. Redundancy effect emphasizes replacing multiple sources of information that is self-contained with one source of information to avoid unnecessary processing of redundant information (Van Merrienboer and Sweller 2005).

Huang et al. (2009) show and measure the relationship between information that is provided by visualization and human cognitive loads. They point out that proper visualization can serve as external memory to reduce demand on human memory and cognitive processes that are needed for understanding and communicating information from the data. They provide a conceptual construct of cognitive load in the context of graph visualization. They propose that cognitive load is influenced by six casual factors: domain complexity, data complexity, visual complexity, task complexity, demographic complexity, and time complexity. Further, by considering mental effort, response time, and response accuracy, they measure visualization efficiency.

Huang et al. (2009) show that an optimal performance may not always be expected by keeping the cognitive load low because that might cause boredom and concentration loss. Additionally, experiencing low cognitive load in the presence of visualization of complex data means that the visual tool is not working efficiently (i.e., little information is processing within the brain). In this point of view, within the capacity of human memory, a well-designed visualization tool will increase the cognitive load sufficiently beyond the point of concentration loss by reducing visual complexity as much as possible and still manage to maintain the memory space for processing complex data and tasks.

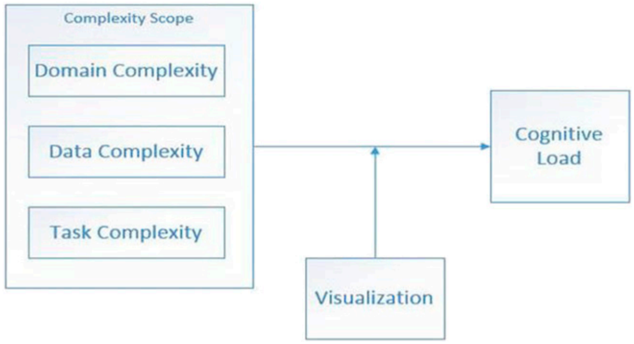
Based on the cognitive load theory and Huang et al. (2009), we developed the conceptual model for our pedagogy facilitated by visualization (Figure 1).

Here are the definitions of the model constructs:

- Domain Complexity refers to the extent of intricateness in the nature of problems that are proposed in the field.
- Data Complexity includes the volume of the data and information that is tied to the problem, as well as the number of objects and their attributes and relations in the data.
- Task Complexity depends on the number of actions that should be done and the volume of information required for processing these actions (Wood 1986).
- Visualization (interactive) involves visual representations of abstract data to magnify cognition and empower the viewer to acquire knowledge about the hidden structures of and causal relationships within the data.
- Cognitive Load is the amount of cognitive resources required for accomplishing a function within the brain.

On the basis of our model in this study, we describe how we use visualization to facilitate learning processes in courses with sophisticated concepts, models, and theories. We strategically select courses from three areas that contain problems and elements that are well aligned with three common casual factors as identified in our model: domain complexity, data complexity, and task complexity. These three courses, as we mentioned earlier, are Stat, OM, and OR/MS. The following sections provide details on using visualization as an effective instructional tool in these (arguably) complex areas.

Figure 1. The Visualization Position in Pedagogy



3. Interactive Visualizations for Statistics, Operations Management, and Management Science

In this section, we introduce and discuss specific visualizations that we have developed spanning the three areas mentioned above. Of course, there is sometimes considerable overlap between topics. For example, an economic order quantity (EOQ) model may be taught in an OM course but may also be covered as a basic example of nonlinear optimization in an OR/MS course. Nevertheless, our focus is not on the distinction of these visualizations per functional area but rather on their functionality in easing "quantitative analysis" pedagogy as it pertains to any functional area. We are interested in the domain complexity, data complexity, and task complexity of the three pedagogy areas and focus on how Mathematica's interactive visualization functionalities can influence the cognitive loads and facilitate the students' decision making in solving problems in these areas.

We have extensively used Mathematica, one of the world's foremost technical computing systems, in creating the visualizations. There are three reasons for this. First, the functional programming paradigm used by Mathematica allows for coding relatively complex visualizations in a fraction of the time and effort that it would take to code using other procedural languages. Second, students can download the free Wolfram CDF Player R (available at http://www.wolfram.com) and embed it within any standard internet browser, thereby obviating the need for special software. Finally, programming in Mathematica affords access to a vast array of sophisticated functions and routines specifically catered to Stats and MS, and their free portal (http://demonstrations.wolfram.com) allows download and reuse of code that has already been created for thousands of visualizations. Please see Appendix A for useful guides and tutorials on Wolfram's website for Mathematica.

3.1. Business Statistics

We have created several interactive visualizations for statistical estimation and hypothesis testing. The topics that we cover in our core undergraduate course (the second course in a two-course sequence) include, among others, hypothesis testing for a single mean, hypothesis testing for differences in two means (large and small populations, equal and unequal variances), hypothesis testing for a single proportion, hypothesis testing for differences in two proportions (large samples), hypothesis testing for differences in variances, and singe-factor analysis of variance (ANOVA).

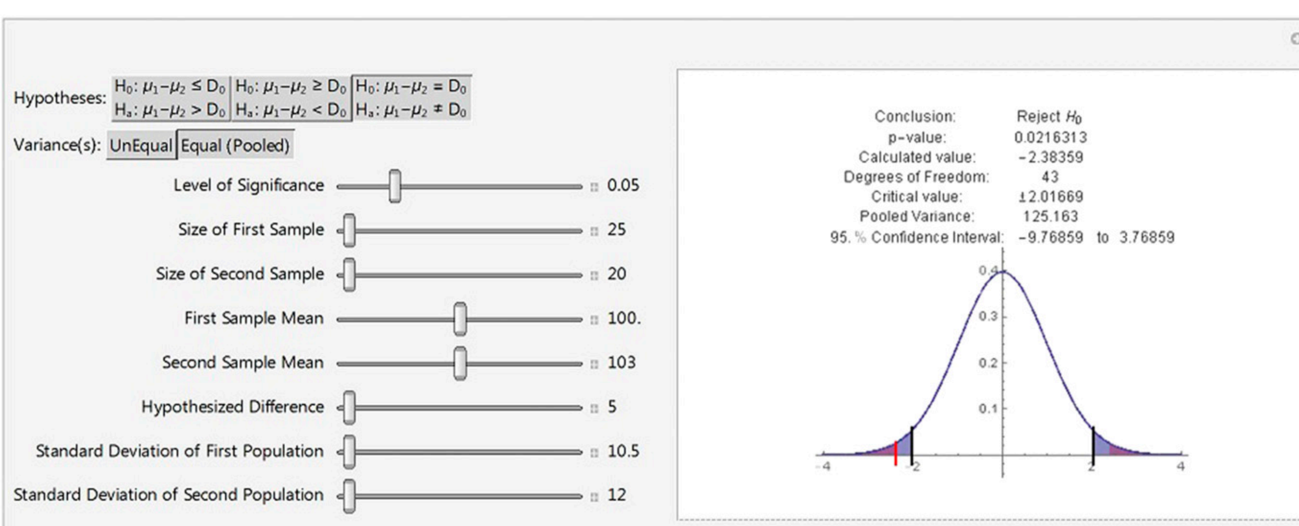
Published in INFORMS Transactions on Education on May 16, 2019 as DOI: 10.1287/ited.2018.0203.
This article has not been copyedited or formatted. The final version may differ from this version.

In business statistics, such concepts are fairly sophisticated, as reflected by students' frequent feedback and comments. Although the data sets the students are handling are usually straightforward and unambiguous, it has been observed that it has been a quite complicated process for students to clearly understand the constructions of the hypotheses, the relationships between the data and the hypotheses, the most appropriate test to conduct on the data, and, importantly, the proper interpretations of the test results with regard to the hypotheses. Each step of this process is mentally demanding. Applying our conceptual model presented in Section 2.2, we observe that in this case both the domain complexity and task complexity are very high, whereas the data complexity is relatively low. Therefore, the visualization tools we develop should focus on illustrating the underlying relationships among the data available, the hypotheses constructed, and the various tests that can be performed on the data. In this way, we would be able to limit the overall cognitive load of the students and facilitate their development of deep understanding of the essence of the problems they face and their ability to make the best decision.

We have developed interactive visualizations for all these topics besides creating interactive Gaussian, Student t, F, and χ^2 charts. Our experience over the years has been that a majority of students have great difficulty in understanding and grasping concepts such as, for example, the p-value. In fact, the p-value and its explanation have been much maligned in the popular press, too (Aschwanden 2015). In Figure 2, we depict a screen shot of visualization for hypothesis testing for differences in two means (large and small populations, equal or unequal variances). This is a particularly challenging topic for students for various reasons, including deciding on which (equal or unequal variances) test to perform. The particular style that is shown in Figure 2 is common across almost all the visualizations that we have developed for business statistics. Students can enter sample statistics in the fields either by using slider bars or by opening up the fields. They can select any level of significance (and the range is not limited to the end point of the slider bars) by entering it into the requisite field. They can select the direction of the hypothesis test and also the equal or unequal variance option. The visualization dynamically responds to these choices. An interesting and useful feature from the point of view of pedagogy is the explanation of the level of significance, how it relates to the area under the probability distribution, and more importantly, how it relates to the p-value and, in turn, how all this relates to the statistical decision and conclusion of the test of hypothesis. This can be achieved by gradually changing the data or the level of significance such that the p-value and level of significance coincide and then one exceeds the other. Students almost immediately see the connection when it is presented to them in such a dynamic visual manner. This becomes the proverbial "Aha!" moment.

Students in our core course have used these demonstrations over several years for solving homework problems and also during examinations and in general have found them very useful for understanding fundamental concepts. This is formally and now routinely reflected in our course evaluation comments. We have also created interactive demonstrations for the Gaussian (normal), Student t, χ^2, and F distributions. These come in handy in lieu of paper charts. All these interactive charts allow for reverse look-up of p-values, too. An example of the F

Figure 2. Demonstration for Estimation and Hypothesis Testing



distribution (which students find especially cumbersome to reverse look-up for p-values when given as a paper chart) is shown in Figure 3.

3.2. Operations Management

Below we discuss a sample out of several visualizations that we have developed to illustrate fundamental topics in OM. Although the potential for applications is vast, we restrict ourselves to the economic order quantity model, inventory management for short-life-cycle products (the "newsvendor model"), and finally, a basic model in location analysis. We have also developed visualizations for our PhD seminars that cover advanced topics such as risk pooling via inventory aggregation and contracts and coordination in supply chains, but we forego their discussion in this paper. The key point to note in these visualizations is that the idiom design relies to some extent on "preattentive" features (Treisman and Gelade 1980, Treisman 1985) that quickly capture the reader's attention. Preattentive features are a limited set of visual features that are detected very rapidly by low-level, fast-acting visual processes. Example of such preattentive features could be a colored "ball" sliding up and down a curve or changes in shaded areas, object size, etc.

The EOQ model is typically the first model/paradigm that students see when they study inventory management. Figure 4(a) shows an EOQ visualization with levers that allow the various costs, interest rate, and demand rate to be adjusted. As can be expected, the curve changes dynamically. What is interesting in an EOQ model is the sensitivity of the total cost to an order quantity that deviates from the optimal. The total cost is much less sensitive as the order quantity exceeds the EOQ compared with when it falls below the EOQ. With the "Heuristic Order Quantity" lever,

Figure 3. Demonstration for F Distribution

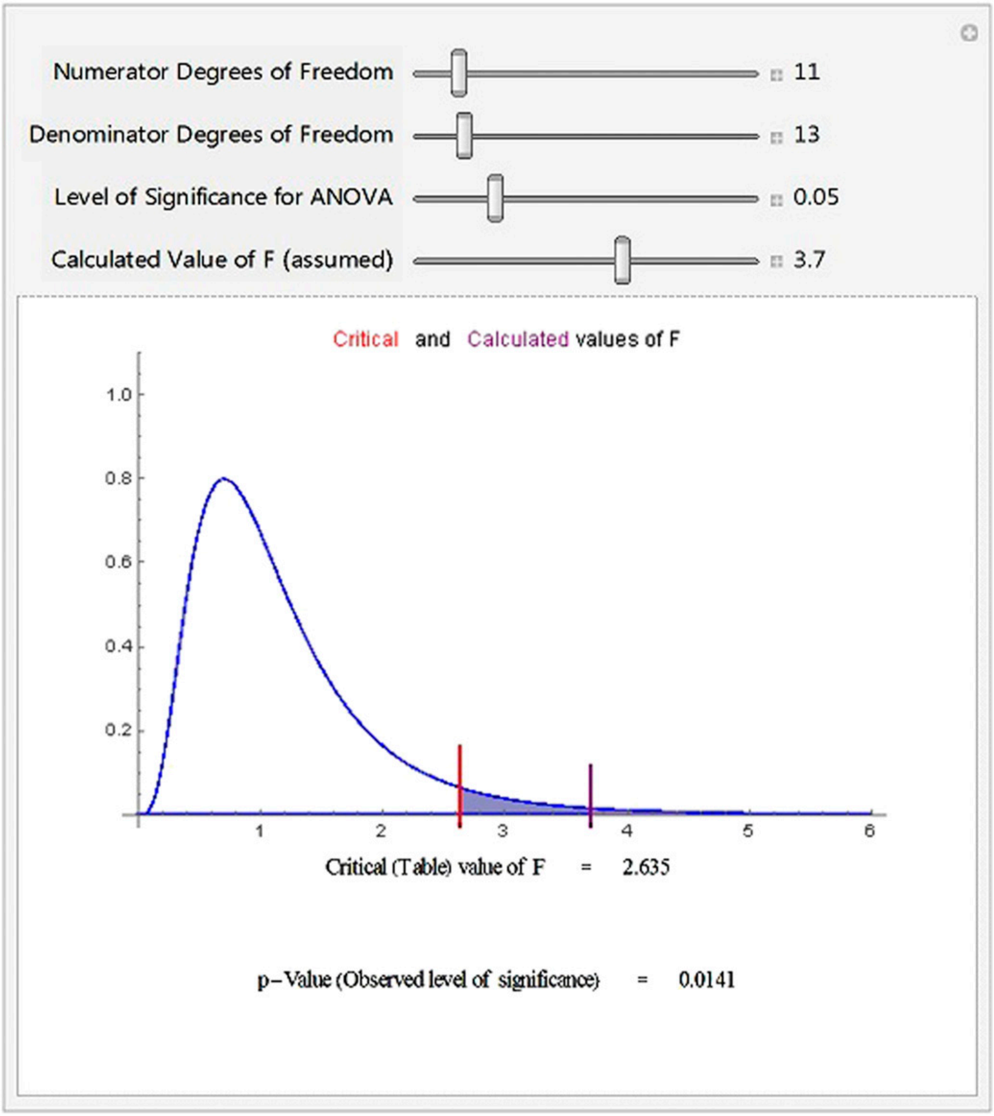
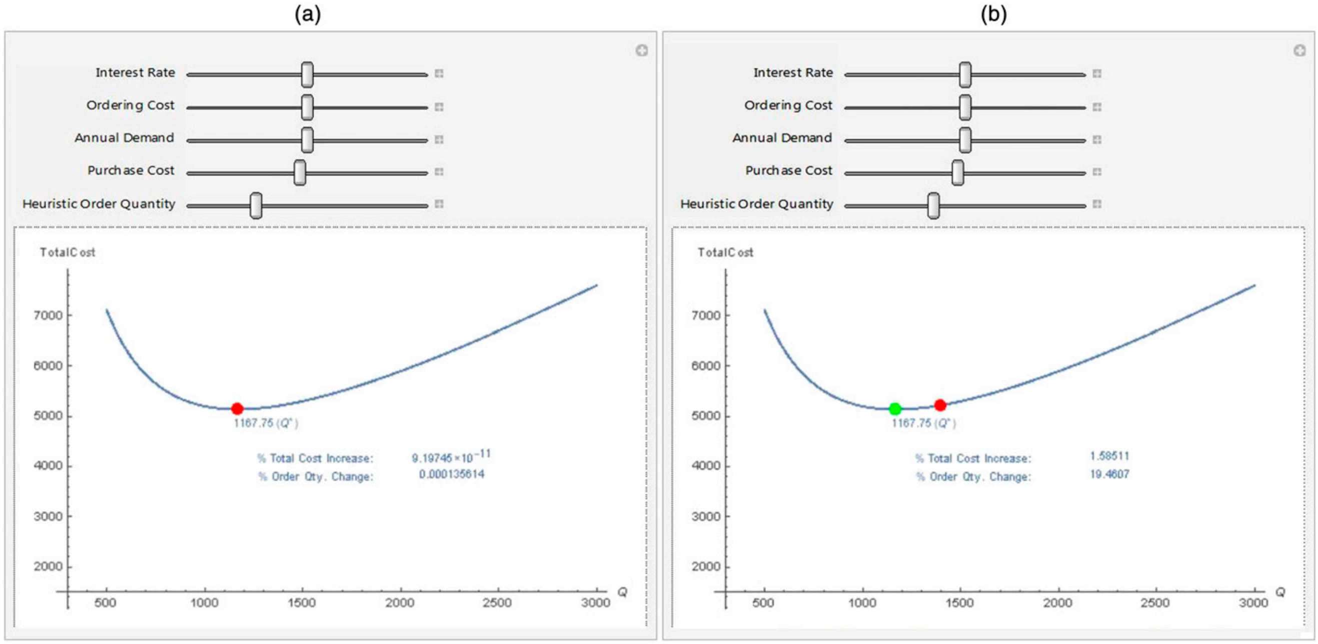


Figure 4. (a) The EOQ Model and (b) Parametric Sensitivity of the EOQ Model



this is easy to demonstrate. In Figure 4(b), we see a red ball that rolls away from the (now) green ball that is static (the EOQ). The percentage order quantity change and the percentage total cost increase change dynamically in the pane. One can conduct the same exercise by making the heuristic order quantity less than the EOQ and quickly see that the percentage cost increase is relatively more pronounced. Although it would not be difficult to show the ordering cost and holding cost curves too as part of the visualization, we have chosen not to do that to avoid “chart clutter” (Tufte 2001) and to focus on the concept of sensitivity.

The newsvendor problem is a fundamental model in stochastic inventory management and is also a basic application of stochastic programming with recourse. However, introducing the newsvendor model with the classic critical fractile calculations and spreadsheet implementation often leaves a majority of students confused. The visualization presented in Figure 5, (a) and (b) (the case and data are derived from Chopra and Meindl 2015), allows for the manipulation of various model primitives that help decide the critical fractile and thereby the optimal order quantity. As evidenced in the EOQ visualization, we allow for a heuristic (aka naive) order quantity and see its impact on the profit. This is given as “Percent Profit Decline due to (Naive) Q” in the pane [note that Figure 5(a) shows both the optimal and heuristic order quantity coinciding, whereas Figure 5(b) shows the heuristic order quantity at mean demand—the red bar had slid away from the black bar—and thereby the impact on profit].

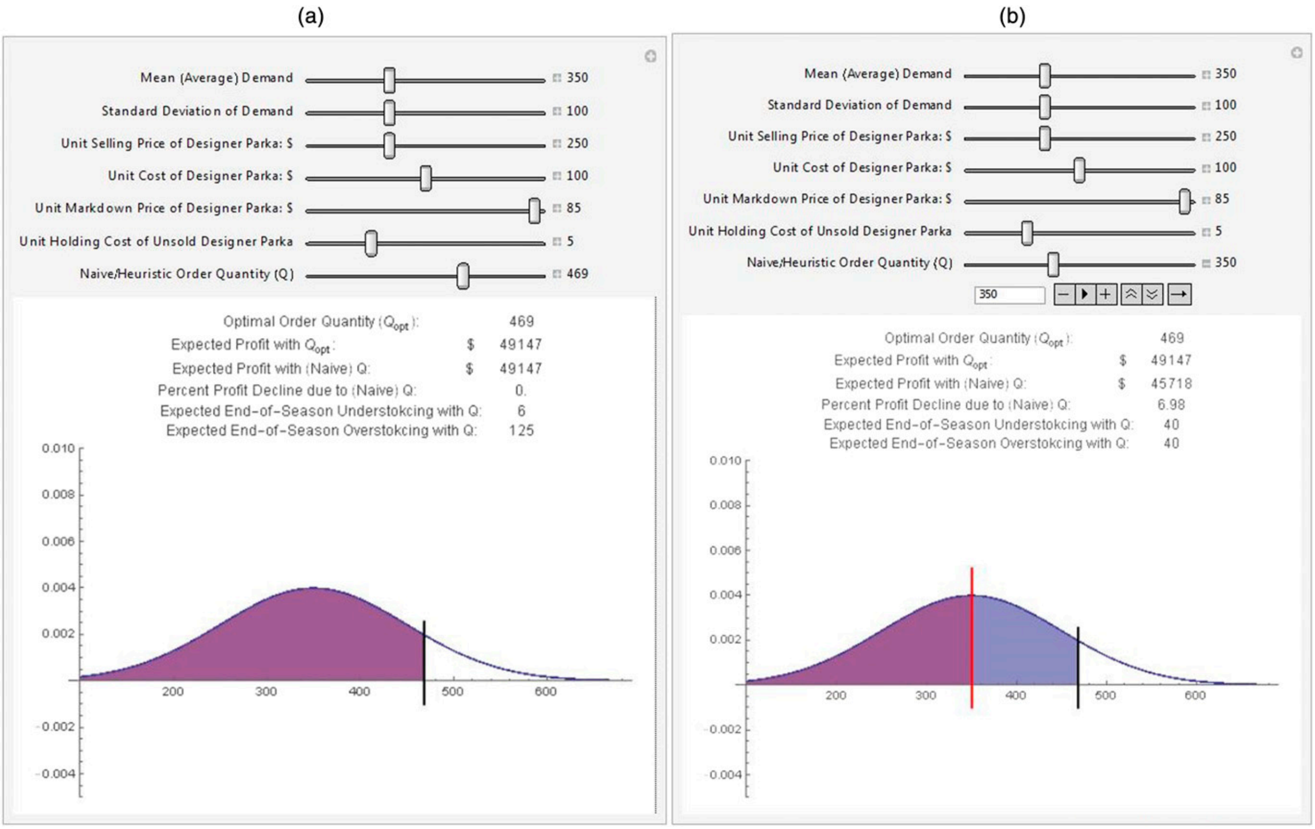
From the descriptions of the problem, we can observe that according to our conceptual model, the domain, data, and task complexities in this case are all high. Therefore, we need to develop a visualization tool that can handle all the complexities required by the problem and maintain a reasonable level of overall cognitive load.

This allows for easy experimentation. We typically pose questions related to the optimal order quantity and the total cost before sliding each lever. Some effects are apparent—such as what happens to “Q_{opt}” and expected profit as the mean demand increases or when the selling price increases, etc.—and students typically anticipate these effects. However, the impact of standard deviation makes students pause a bit ex post, which, experimenting with the slider, allows for students to see the change in the shape of the distribution and the dynamic movement of Q_{opt}. Moreover, the visualization shows the expected end-of-season understocking and overstocking quantities. These provide some points of discussion, especially in relation to the other model parameters. More complex questions can be posed, such as sensitivity of the percent profit decline when the heuristic order quantity is above/below Q_{opt} when Q_{opt} itself is above/below the mean demand, and answered immediately using the visualization.

Another fundamental model that is shown in introductory OM classes is the classic “facility location model.” In this model, one has to locate a facility, such as a distribution hub, given a set of geographically separated existing facilities with some type of

Published in INFORMS Transactions on Education on May 16, 2019 as DOI: 10.1287/ited.2018.0203.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 5. (a) The Newsvendor Model and (b) Parametric Sensitivity of the Newsvendor Model



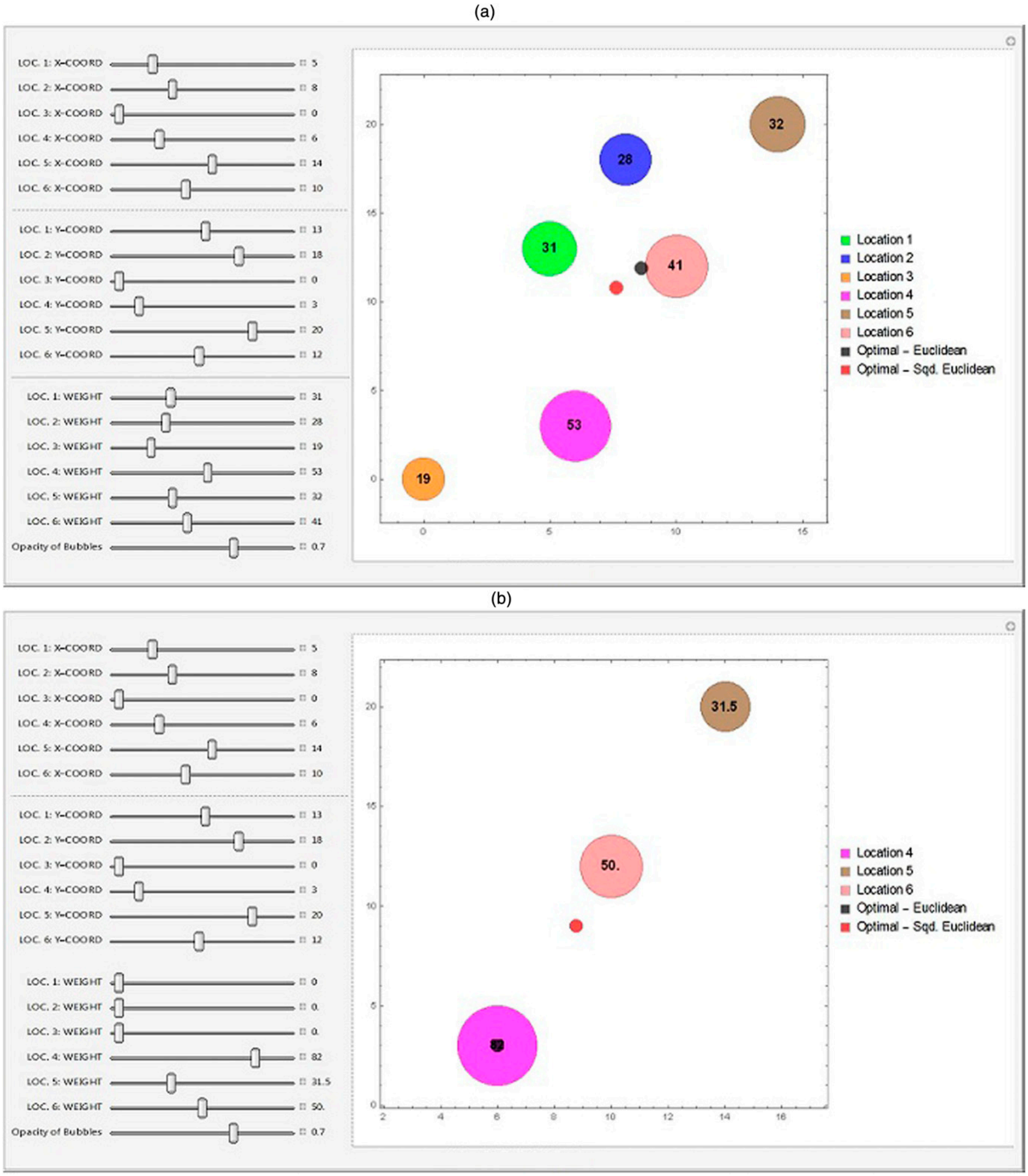
demand requirement. Typically, the objective is to minimize the demand/volume-weighted Euclidean or squared Euclidean distance and finding the optimal location. There has been some discussion in the literature on the merits of using the Euclidean versus the squared Euclidean (see Kuo and White 2004, Kulkarni 2011). The visualization in Figure 6(a) shows both optimal locations derived with both criteria simultaneously, thereby fostering some interesting class discussion. A typical question that we pose is, "Which optimal location seems to be more sensitive to changes in demand/volume weights and/or locations of the existing facilities?" Students experiment with the sliders and quickly come to their own conclusion. We also drive weights of a subset of locations to zero and assess the impact on optimal location using both metrics when there are three or fewer locations [Figure 6(b)]. This shows the relative sensitivity of both metrics more clearly and leads to interesting class discussion about which metric to use for optimal location and policy implications for the same.

3.3. Management Science

We use several visualizations to illustrate fundamental topics in OR/MS. Although we have customized these demonstrations to suit our course-specific needs (such as using problem-specific data for the shortest-path problem, the binary knapsack problem, etc.), we have at times used these demonstrations directly without any changes. For example, a traveling salesman "game," freely available at http://demonstrations.wolfram.com/TravelingSalesmanGame/, is shown in Figure 7. For a problem such as traveling salesman, although the question is usually presented in an easy-to-understand format, the underlying domain knowledge is very complicated. Although the data presented for the problem are intuitive and unambiguous, the task of solving the problem is highly sophisticated. Again, the visualization tools deployed must be able to limit the overall cognitive load. Use of such visualizations, which depart from pure data visualizations but nevertheless serve as powerful tools for visualizing abstract modeling-related concepts, make class proceedings fun, interactive, and seemingly less monotonous.

An important visualization on the same line as above that we have developed and deployed in our OR/MS courses is the graphical linear programming solver. Figure 8 provides a snapshot of the solver. The solver proves highly useful in pedagogy and also in illustrating important concepts in sensitivity analysis. Students can dynamically observe the change in the optimal pivot (illustrated by the red dot), as well as observe changes in the feasible region. Furthermore,

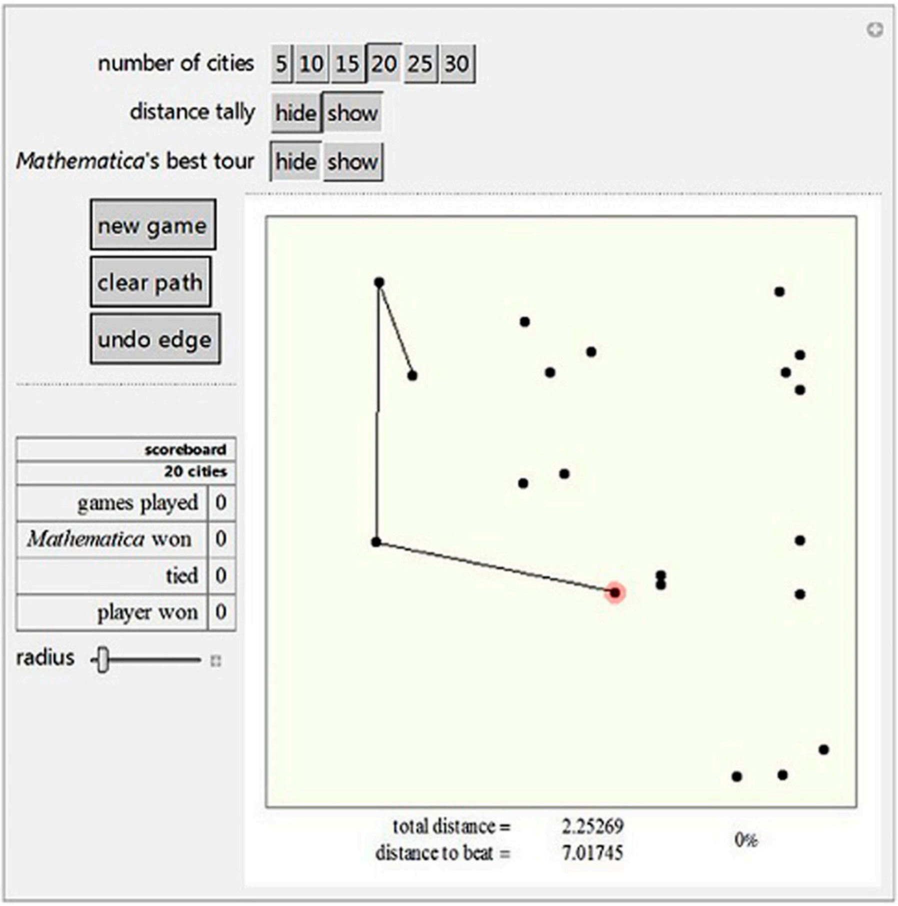
Figure 6. (a) An Optimal Facility Location—Euclidean Versus Squared Euclidean Metrics and (b) Optimal Facility Location with Three Locations



students have access to this textbook-free, platform-free solver that they can use gratis and for as long as they wish.

To reinforce fundamental concepts, we also do a number of hands-on exercises in class. One popular exercise to illustrate the development of a basic mathematical programming model and to further illustrate shadow prices is the "tile puzzle." In this class exercise, students build predesigned objects with some beginning inventory. The goal is to maximize

Figure 7. The Traveling Salesman Game



profit. We refer the reader to Simpson and Hancock (2010) for a detailed description of this exercise. In regard to the context of this paper, however, we have developed an interactive visualization that allows students to dynamically explore the same concepts that would take repeated solving in MS Excel or illustration via a static chart if using the new Analytic Solver Platform's parametric sensitivity analysis capability. In summary, students do the hands-on exercise, see/develop the algebraic model and thereafter the MS Excel representation of the model, and finally, experiment in much more detail with the interactive visualization, having already developed an understanding of the basic model on account of the hands-on exercise and MS Excel work. The tile puzzle interactive visualization is shown in Figure 9.

Finally, at times, we also supplement solutions to homework problems using interactive visualization. An example follows:

"The province of PulauPulau has decided to secede from the world-famous island nation of PulauPulau-Bompa. The original country will now be called Bompa, and the breakaway province (now a new country) retains its name as PulauPulau. While the separation has been otherwise amicable, there is however one contentious issue. There is a circular patch of land of radius 1 mile that is purported to contain a vast source of energy. The problem is that no one knows exactly what type of energy it is, how big the energy field is, or where it is located. Engineers at PulauPulau and Bompa have agreed to divide this circular patch of land between them. They have decided to trace an arc using a segment of length 'L' such that PulauPulau gets the part that is above the arc and Bompa get the part below it, and each country gets an equal share of this coveted land."

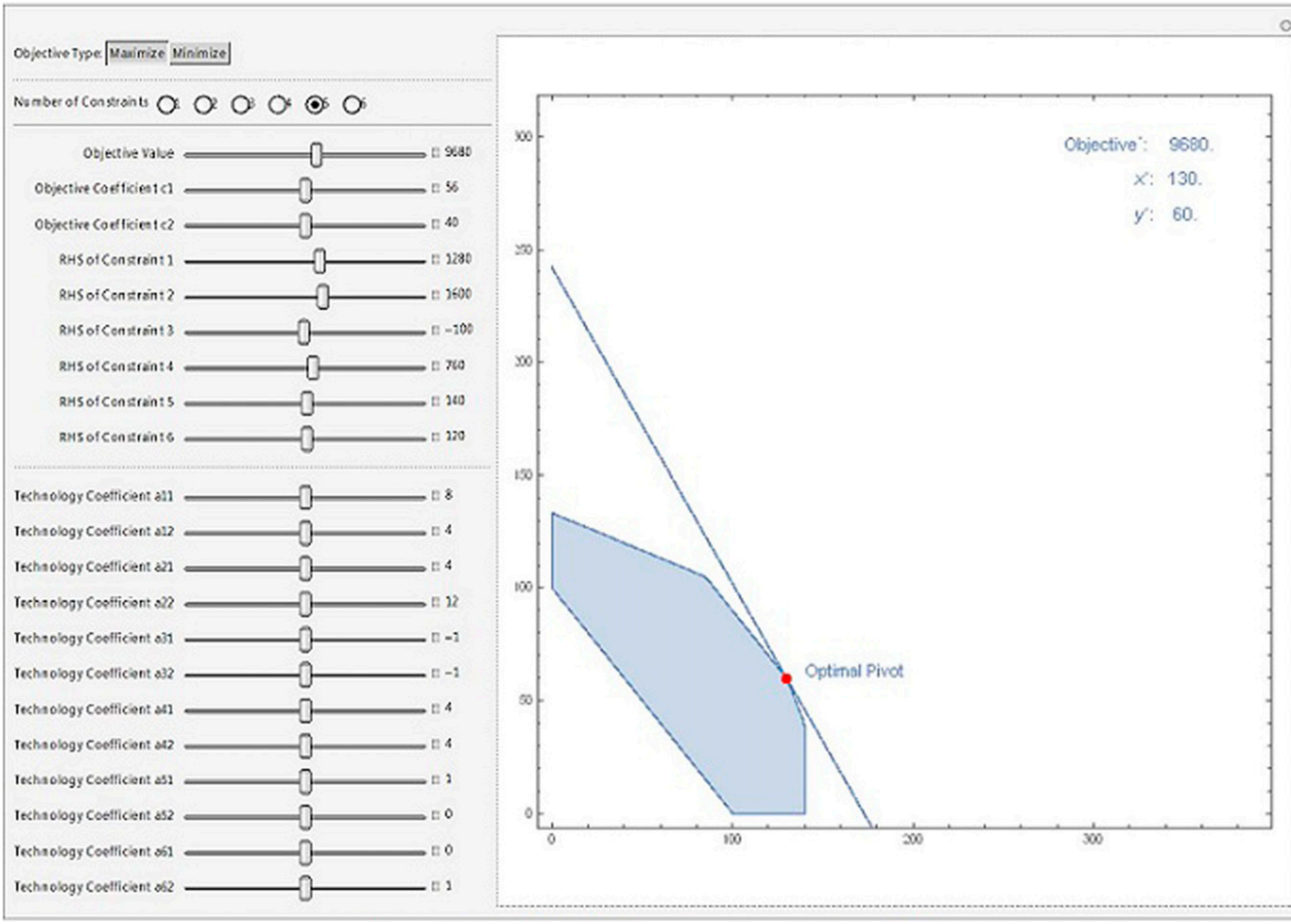
Although the solution to the above homework problem involves some basic geometry and is used to illustrate the use of the "Data Table" and "Goal Seek" macros within MS Excel, we also provide an interactive visualization (Figure 10) so that students can see how the solution actually works and get a better feel for the visual representation of the problem.

4. Modeling and Visualization of Covering Models: A Detailed Classroom Exercise

In this section, we demonstrate how a single data set can be leveraged to inspire discussion about variants

Published in INFORMS Transactions on Education on May 16, 2019 as DOI: 10.1287/ited.2018.0203.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 8. Graphical Linear Programming Solver



of a real-world problem and how visualizations can supplement understanding of the similarities and differences between the problem variants.

A doctoral course in integer programming introduced us to the basic "set cover" problem discussed below. Although the focus of that course was on algorithm development, we have found the problem setting to be useful enough to be used across undergraduate and graduate classes. The CHB, Inc., banking location problem (see Camm et al. 2017) discusses taking advantage of a change in banking laws in the state of Ohio. The idea is to "cover" the state of Ohio in its entirety with "principal places of business" (PPBs) such that each PPB covers the county in which it is located and any contiguous counties, as evidenced by those with which it shares a border. Given the "0–1" adjacency matrix that denotes county adjacencies (1 for adjacent, 0 for not), the problem of optimally covering the state of Ohio with the minimal number of PPBs is easily modeled as a classic "set cover" problem and is solved efficiently in MS Excel.

We, however, begin class proceedings by conducting a "map coloring" exercise after explaining the bank location problem. Students are formed into teams and are given a paper map of the state of Ohio and are asked to select crayons of two distinct colors (say blue and yellow). A county painted in blue is to be marked as a PPB. Students are instructed to minimize the number of blue units by using whichever strategy they deem appropriate and are told that once a PPB is selected, its contiguous counties have to be colored in yellow because they too are covered by virtue of being adjacent to a blue county (PPB). The state of Ohio has 88 counties, and students quickly realize that the exercise ends once all counties are colored (either blue or yellow), and the task at hand is to minimize the blue-colored counties (number of PPBs). At the end of the exercise, students typically have 16–20 PPBs (the optimal solution is 15). Most students realize that a greedy approach (i.e., selecting the first PPB as the one with the maximum number of neighbors) results in being forced to select a couple of extra PPBs at a later stage because of a few noncontiguous counties being left uncovered. The integer programming formulation is then shared with students, and MS Excel is used to solve the problem.

Figure 9. The "Tile Puzzle"

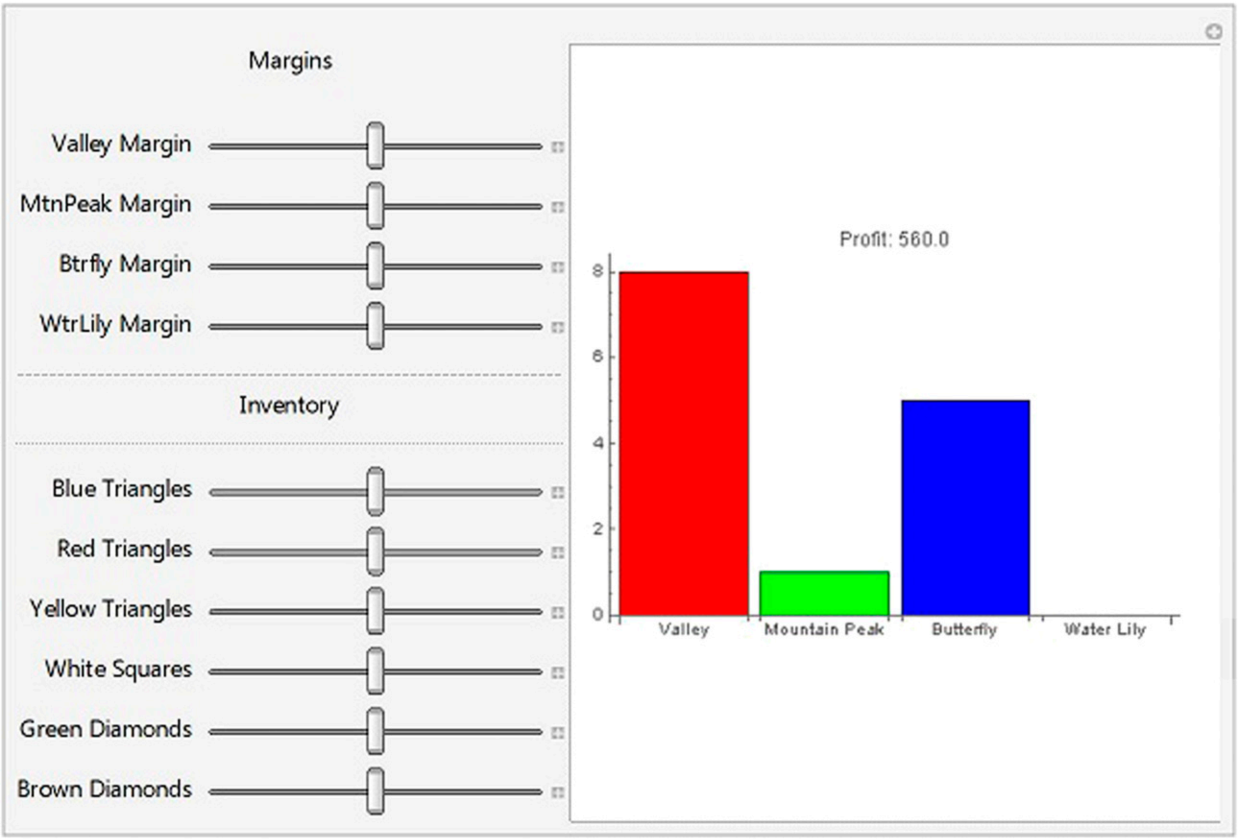
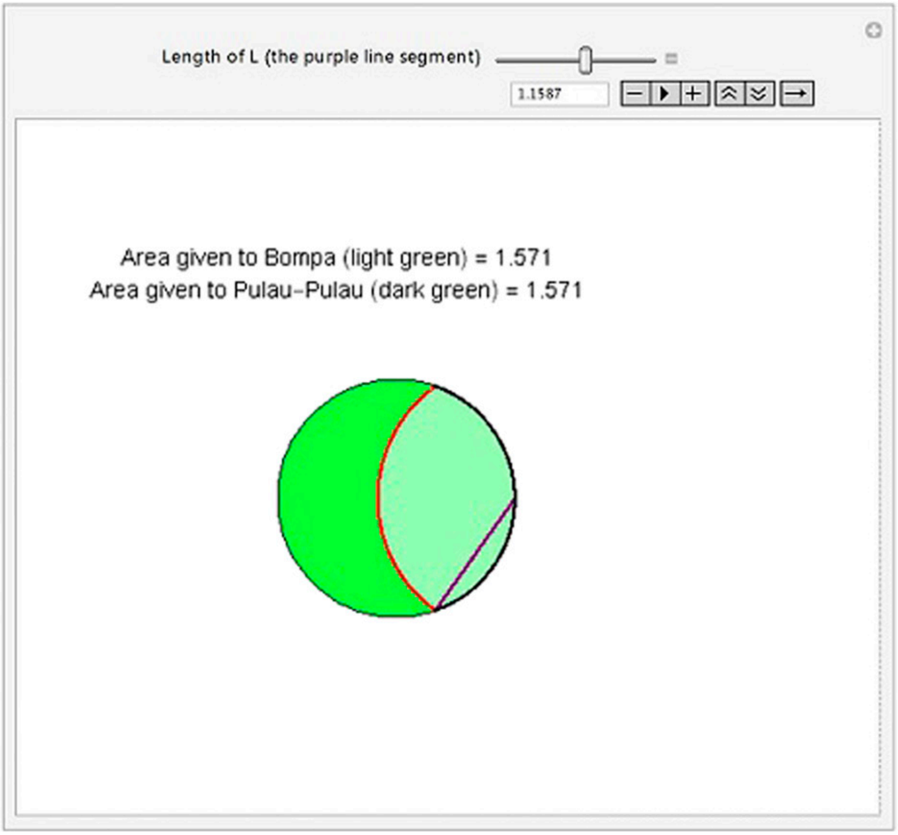


Figure 10. Interactive Solution to PulauPulau-Bompa



A variant of the problem is then introduced. Students are given another map of the state of Ohio, but now with county populations given on the map and inscribed within each county. The counties are shaded in advance according to their populations, with highly populated counties being conspicuously darker. The problem posed is to optimally locate a given number (we typically ask students to assume five) of PPBs such that the population covered is maximized. Student teams mark these on the map after working through their individual (and typically greedy) heuristics. The integer programming formulation ("max cover") is then shared with students, and MS Excel is used to solve the problem. Finally, another variant of the problem is introduced, and students are provided another set of shaded maps with the county populations embedded as in the "max cover" exercise. We call this the "sales rep" problem in that PPBs are now akin to sales reps. Although maximizing population coverage with a given number of sales reps is still the objective, the catch is that each sales rep can now cover the county in which she is located and only one more contiguous county. Students quickly go about their heuristics and marking the maps. The most concise integer programming formulation is then shared with students (with some interesting intermediate quirks that are outside the scope of this paper), and it is demonstrated that it is far too large to be solved with the student version of the Solver macro that comes bundled with MS Excel. We solve the problem using Xpress-MP or the professional version of the Solver macro.

Visualizations related to the "set cover," "max cover," and "sales rep" problems discussed above are shown in Figures 11, 12, and 13, respectively. Figure 11 serves as a good demonstration of how one can transform a map into a node-edge representation, and it also uses the color "channel" to demonstrate PPBs and coverage. In short, the visualization (which dynamically solves the "set cover" based on embedded data and renders the output shown) visually demonstrates coverage and shows where the PPBs have to be optimally located. For this particular problem, the optimal number of PPBs to cover all 88 counties is 15, and the red dots in Figure 11 represent the counties where a PPB should be established. The yellow counties, then, represent all the other counties that can be serviced through these PPBs because they are adjacent to at least one PPB. More interesting, when students see the interactive visualization shown in Figure 12 for the "max cover," they begin to see patterns emerge as they select five, six, seven, etc. PPBs. Concomitant with county populations, they try to make sense of how the optimization (which is again dynamic and embedded in the visualization) is choosing the optimal locations.

Figure 11. Interactive "Set Cover" Visualization for PPB Location in Ohio

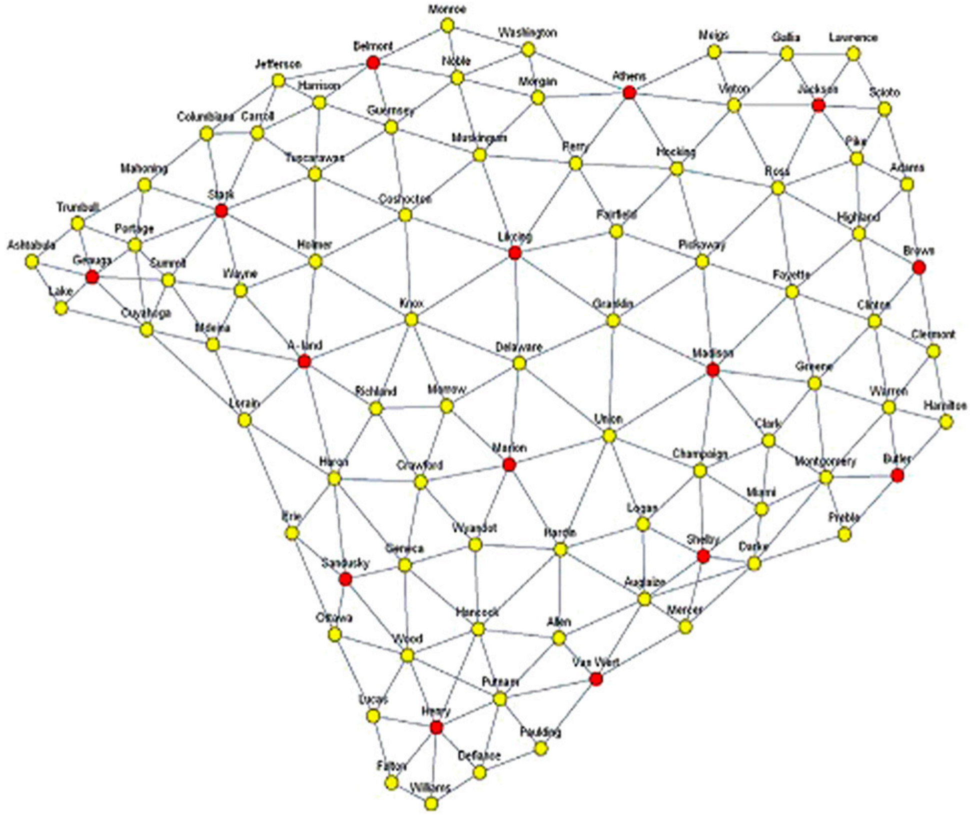
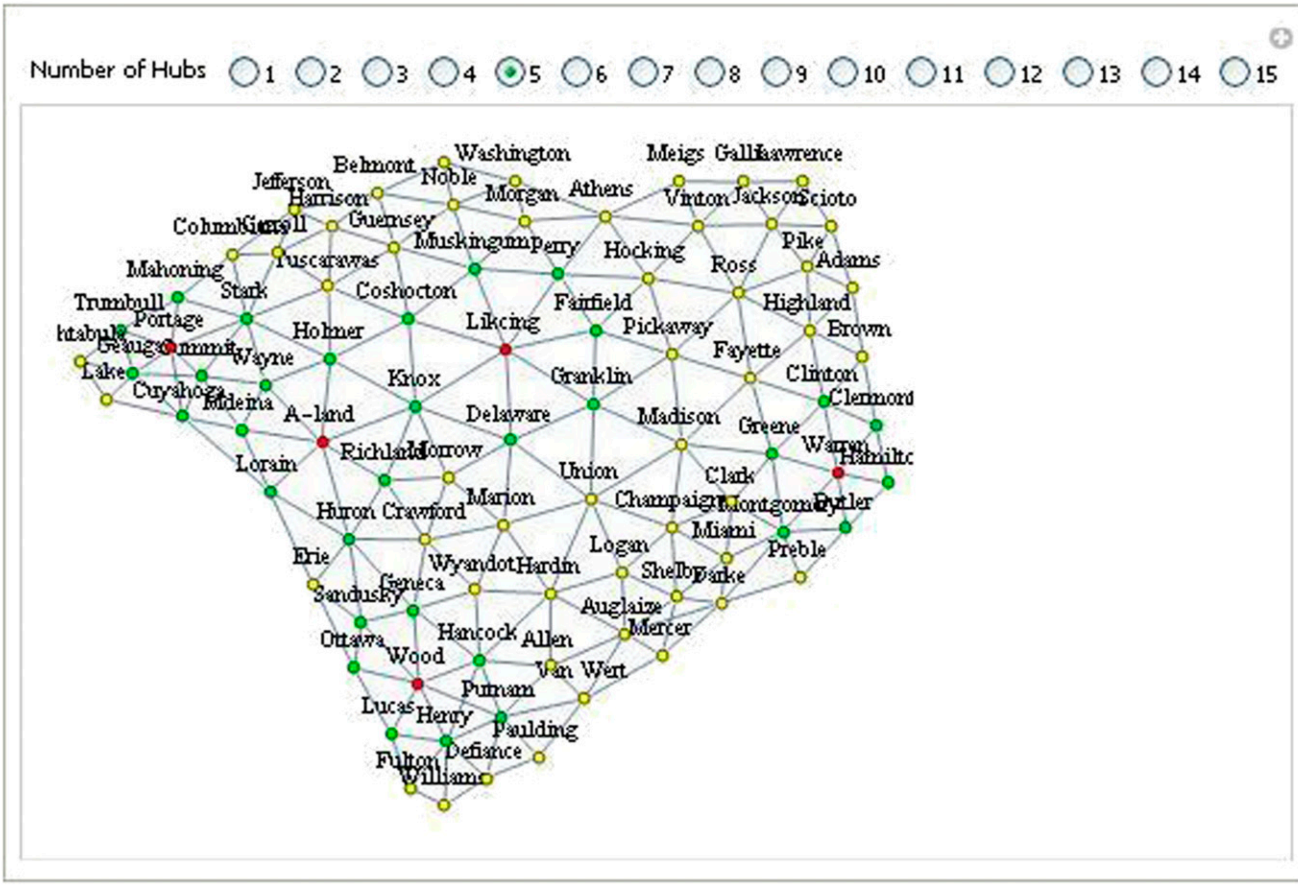


Figure 12. Interactive "Max Cover" Visualization for PPB Location in Ohio

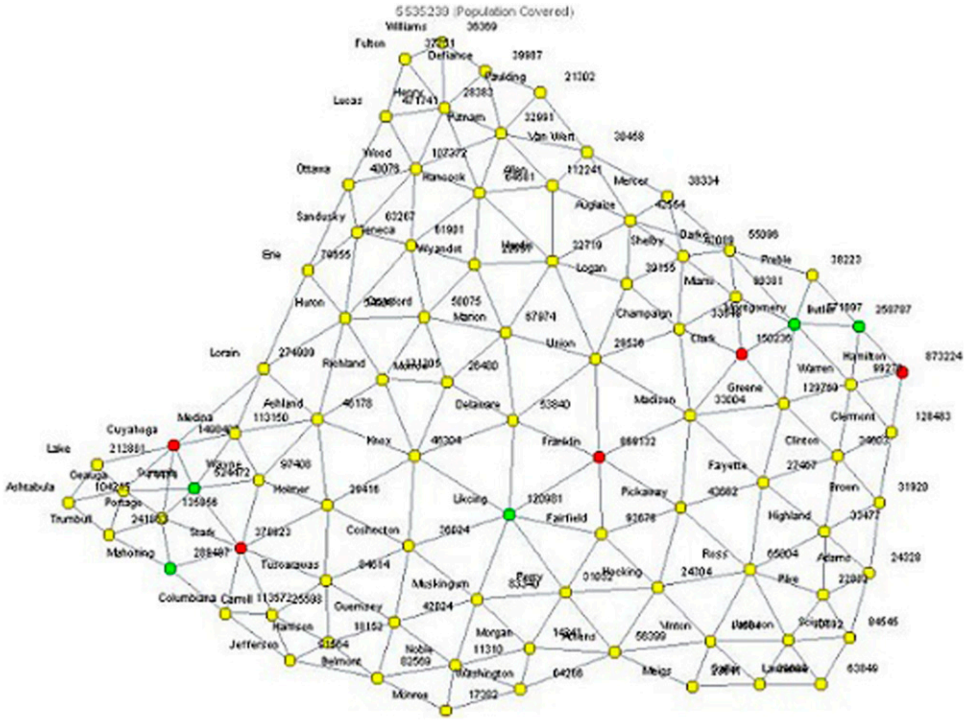


At this point, some students also realize that the "set cover" serves as the key upper bound for the "max cover" in that 15 counties is the maximum number of PPBs in the radio button bar for the "max cover" visualization, albeit optimization is still needed to know where to place those PPBs. For the particular problem with five PPBs, the optimal solution to the "max cover" problem can be seen in Figure 12. Here the red nodes represent the counties where a PPB should be established, the green nodes are the covered counties, and the yellow nodes are the uncovered/unserviced counties. Figure 12 shows that the five optimal counties for PPBs are Ashland, Licking, Summitt, Warren, and Wood. Finally, the "sales rep" visualization shown in Figure 13, when seen in tandem with the earlier two visualizations, now makes the connection between these series of problems much more evident.

The color channel is again put to use, with main counties in red (where the sales rep is located), secondary county in green (the one she selects among the neighbors), and uncovered counties in yellow. For this particular problem with five sales reps, we find that the optimal counties to place them are Clark, Cuyahoga, Franklin, Hamilton, and Stark. Furthermore, students click on the various radio buttons denoting the number of sales reps and dynamically see which "pairs" of counties are selected, thereby fostering discussion. Although students do not need a visualization to realize why "44" is the final radio button (considering that Ohio has 88 counties), what is more interesting is the dynamic rendering of the optimization as one goes from, say, 35 to 40 to 44 sales reps. Students see the "shift" in optimal locations to adjust for additional reps, which again leads to interesting class discussion as to why the optimization behaved the way it did. An extension of the "sales rep" model is when reps are allowed to select two or more contiguous counties as neighbors. We discuss the modeling extension (which is simple yet quite clever), and although we do not have a visualization for it, we ask students how such an extension could be rendered. We invariably have a few students realize that all it takes in terms of visualization design is to have another radio button bar that denotes the number of neighbors that can be selected. We conclude the class exercise by urging students to take our code for the existing "sales rep" visualization and

Published in INFORMS Transactions on Education on May 16, 2019 as DOI: 10.1287/ited.2018.0203.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 13. Interactive "Sales Rep" Visualization for Ohio



render the visualization for the extension. There are typically a few enthusiastic takers.

5. Impact of Visualization on Student Learning Experience

In this section, we describe the experiment we conducted to investigate the impacts of visualization on student learning experience. We present the empirical results of our analyses on the experiment outcome and demonstrate the impact of visualization.

We presented a group of three data analysis exercises to a group of 93 students. The three exercises include two Mathematica exercises and one Excel assignment. The first Mathematica exercise requires the students to use Mathematica to experiment with size of the sample and the significance level to determine the effect of each of the results of the ANOVA test provided from a formal survey in one of the authors' classes and answer a series of question regarding the data analysis. The second Mathematica exercise requires the students to collect specified data and use Mathematica to experiment with the collected data and the significance level to determine the effect of each of the results of the ANOVA test from the collected data. The Excel assignment requires the students to use Excel to conduct statistical analyses on a given set of sample data of the sales prices of gasoline ($/gallon, at pump) at different retailing pumps/locations, to conduct ANOVA (using KPK Excel macros), test a given set of hypotheses, and answer several questions. Although the two Mathematica exercises are designed such that they rely heavily on the visualization tools to help the students to understand the concepts and practices of ANOVA analysis, the macros provide the students with ANOVA results in a table only. Detailed descriptions of these three exercises and assignment are available in Appendices B, C, and D.

The two Mathematica exercises are optional, but the Excel assignment is mandatory for the students, and if they choose to complete the Mathematica exercises, they must complete them before working on the Excel assignment. So some students completed both Mathematica exercises, some completed only the first Mathematica exercise, some completed only the second Mathematica exercises, and some did not complete any Mathematica exercises. We recorded every student's score on the Excel assignment.

We grouped the student scores into four categories: the scores of those who completed both Mathematica exercises, the scores of those who completed only the first Mathematica exercise, the scores of those who completed only the second Mathematica exercises, and finally, the scores of those who did not complete any Mathematica exercises.

Thus we construct our null hypothesis:

Hypothesis 1. *There is no significant difference among the mean scores for the four groups.*

Table 1. ANOVA Results for All Four Categories

Group	Count	Sum	Average	Variance		
Took both MQ1 and MQ2	20	82	4.1	0.831579		
Took MQ1 but not MQ2	24	84	3.5	1.391304		
Took MQ2 but not MQ1	10	36	3.6	1.155556		
Took neither MQ1 nor MQ2	39	122	3.128205	1.588394		
Source of variation	SS	df	MS	*F*	*p*-Value*	*F* crit
Between groups	12.66683	3	4.222277	3.169584	**0.02819**	2.706999
Within groups	118.559	89	1.332123			
Total	131.2258	92				

Notes. MQ1, Mathematica Exercise 1; MQ2, Mathematica Exercise 2; MS, Mean Square; SS, Sum of Squares.
*The relevant p-value for Hypothesis 1 (0.02819) is boldfaced.

Further, to specifically study the positive impact of visualization on the student learning outcomes, we formulate the following null hypotheses:

Hypothesis 2a. *The mean score for the students who complete both Mathematica exercises is less than or equal to the scores of students who complete only the first Mathematica exercise (one-tailed).*

Hypothesis 2b. *The mean score for the students who complete both Mathematica exercises is less than or equal to the scores of students who complete only the second Mathematica exercise (one-tailed).*

Hypothesis 2c. *The mean score for the students who complete both Mathematica exercises is less than or equal to the scores of students who complete no Mathematica exercises (one-tailed).*

Table 2. *t*-Test Results for the Pair of Two Categories

	Took both MQ1 and MQ2	Took MQ1 but not MQ2
Mean	4.1	3.5
Variance	0.831578947	1.391304348
Observations	20	24
Hypothesized mean difference	0	
df	42	
t Stat	1.901650498	
P(*T*<=*t*) one-tail*	**0.032045522**	
t Critical one-tail	1.681952357	
P(*T*<=*t*) two-tail*	0.064091044	
t Critical two-tail	2.018081703	
	Took both MQ1 and MQ2	Took MQ2 but not MQ1
Mean	4.1	3.6
Variance	0.831578947	1.155555556
Observations	20	10
Hypothesized mean difference	0	
df	16	
t Stat	1.261345977	
P(*T*<=*t*) one-tail*	**0.112632592**	
t Critical one-tail	1.745883676	
P(*T*<=*t*) two-tail*	0.225265183	
t Critical two-tail	2.119905299	
	Took both MQ1 and MQ2	Took none of MQ1, MQ2
Mean	4.1	3.128205128
Variance	0.831578947	1.588394062
Observations	20	39
Hypothesized mean difference	0	
df	50	
t Stat	3.38731992	
P(*T*<=*t*) one-tail*	**0.000691429**	
t Critical one-tail	1.675905025	
P(*T*<=*t*) two-tail*	0.001382858	
t Critical two-tail	2.008559112	

*The relevant *p*-values for Hypotheses 2a (0.032045522), 2b (0.112632592), and 2c (0.000691429) are boldfaced in the table above.

We conducted an ANOVA to test Hypothesis 1. The results are presented in Table 1. The results clearly reject the null Hypothesis 1 at a 0.05 significance level. There is a significant difference among the mean scores for the four categories.

We then conducted a *t*-test for testing Hypothesis 2. The results are presented in Table 2. These results show that at the 0.05 significance level, the students who complete both Mathematica exercises scored significantly higher in their Excel assignments compared with those who did not complete any Mathematica exercises, as well as with those who completed only the first Mathematica exercise. The null Hypotheses 2b and 2c are rejected. In other words, our experiment results indicate that using visualization tools does have a positive impact on the student learning outcomes.

6. Limitations, Conclusion, and Future Work

In this paper, we have discussed the use of dynamic interactive visualizations in teaching quantitatively oriented material, be it operations management, statistics, or management science. The goal of this paper was to introduce myriad ways in which one can leverage such visualization in tandem with hands-on classroom exercises. We also provide empirical evidence that using visualization tools does enhance students' learning experience and improve their learning outcomes.

Of course, what we presented here was rather stylized and short of a formal, generalized framework for using visualization for quantitative analysis pedagogy. Such a goal definitely merits attention but was not the focus of our paper. Our visualization has been productively used by our students, and they have gone on to develop visualizations of their own at their places of work. We have also shared our work freely with faculty from other universities in the United States and around the world. In future research, we hope to formally capture the before (not using visualization) versus after (using visualization) cognitive load reduction in a laboratory setting and understand in more detail what idiosyncrasies of topical content make certain areas more versus less amenable to usefulness of visualization as catalysts for improved comprehension.

Acknowledgments

As the other three authors, we would like to dedicate this paper to our late colleague, Shailesh S. Kulkarni, who passed away suddenly and unexpectedly on July 6, 2018. He was a great husband, father, friend, colleague, teacher, scholar, and human being. He will be sorely missed.

Appendix A

In this appendix, we provide links to some useful guides and tutorials on Wolfram's website for Mathematica. These include documentation as well as videos.

1. The Center for Mathematica Resources, with links to all types of guides: https://www.wolfram.com/mathematica/resources/
2. *A Beginner's Guide to Mathematica*, a paperback book: http://www.wolfram.com/books/profile.cgi?id=6010
3. *An Elementary Introduction to the Wolfram Language*, an online book: https://www.wolfram.com/language/elementary-introduction/2nd-ed/
4. A quick how-to list for commonly used functions in Mathematica: https://reference.wolfram.com/language/guide/HowToTopics.html
5. The homepage for Wolfram University, a repository of video tutorials and interactive courses: http://www.wolfram.com/wolfram-u/
6. An online interactive course for newcomers to Wolfram and Mathematica: http://www.wolfram.com/wolfram-u/an-elementary-introduction-to-the-wolfram-language/
7. A video course for basic programming: http://www.wolfram.com/wolfram-u/basic-programming/
8. The homepage for Wolfram Demonstrations: http://demonstrations.wolfram.com/
9. The guide page for contributing to Wolfram Demonstrations: http://demonstrations.wolfram.com/participate.php
10. An interactive demonstration about two-population *t*-tests on the Wolfram Demonstrations website written by two of the authors: https://demonstrations.wolfram.com/ComparingTwoMeansUsingIndependentSamplesOfUnknownVariance/
11. An interactive demonstration about the newsvendor model on the Wolfram Demonstrations website written by one of the authors: https://demonstrations.wolfram.com/CapacityPlanningForShortLifeCycleProductsTheNewsvendorModel/

Appendix B. DSCI 3710 Mathematica Laboratory: Exercise 1

All students have access to Mathematica in our labs, but you can do this laboratory at home if you download Mathematica's Player from Wolfram's website. Access will be discussed further in class, but you can go to Wolfram's online site to download their free player. http://www.wolfram.com/products/player/.

These instructions are an outline of the steps you need to take in order to complete your laboratory assignment. They should serve as a guide and are not intended to provide each step and detail of what you should do because the laboratory is intended to provide you with a vehicle for investigating concepts.

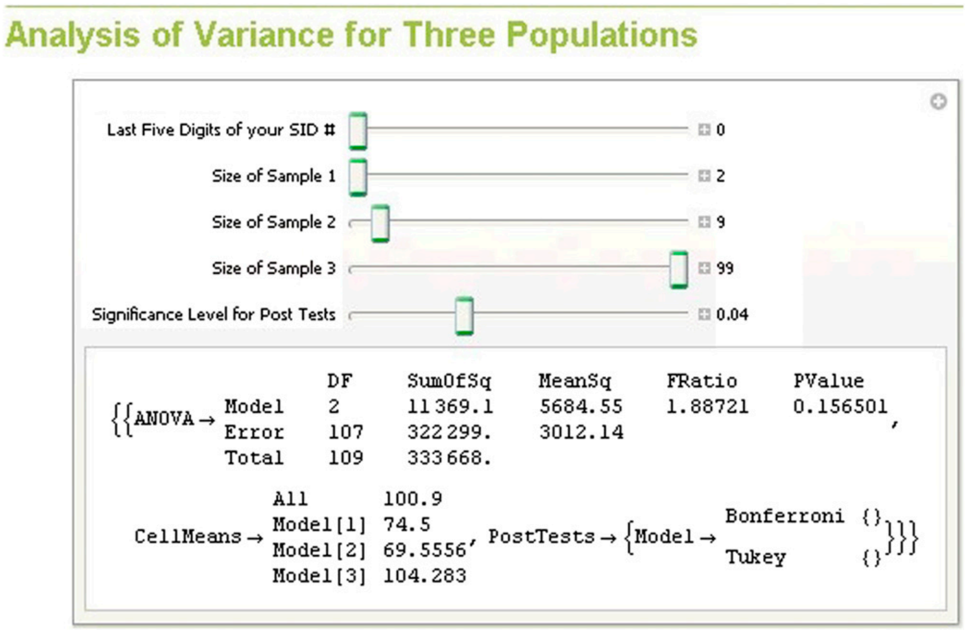
Please read the instructions listed below before you begin the laboratory exercise. By doing so, you will have a clearer understanding of the purpose of the laboratory exercise and what you can do to investigate the concepts involved.

Description

The intent of this laboratory exercise is to have you use Mathematica (or the Player) to experiment with size of the sample and the significance level to determine the effect of each on the results of the ANOVA test. The steps that follow are intended to guide your experimentation. Figure B.1

Published in INFORMS Transactions on Education on May 16, 2019 as DOI: 10.1287/ited.2018.0203.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure B.1. Mathematica Exercise #1 Sample Screenshot



shows a screenshot of the laboratory layout in Matematica for this ANOVA experiment.

1. Use your mouse to move the student ID (SID) slider so that the display shows the last five digits of your SID.
2. Experiment with changing the sample size for each of the three samples by moving the slider for each sample from the left to the right.
   a. Take some notes about how the calculated *F*-statistic (Fratio) changes as the sample sizes are increased.
   b. Take some notes about how the *p*-value changes as the sample sizes are increased.
   c. Take notes some notes about how the PostHoc tests change as the sample sizes are increased.
3. Experiment with the statistical significance by moving the slider from the left to the right.
   a. Take some notes about how the calculated *F*-statistic (Fratio) changes as the sample sizes are increased.
   b. Take some notes about how the *p*-value changes as the sample sizes are increased.
   c. Take notes some notes about how the PostHoc tests change as the sample sizes are increased.
4. Print the screen for one of our experiments with your SID by pressing “Print screen.” Then print your name on that page and bring the printed page to class.
5. Once you are done, see if you can answer the questions that follow.

The questions below are much like the quiz you will have as WEBTEST for this assignment. You should attempt to answer these before coming to the experiential class session on this laboratory. In class you and other students who completed this laboratory can discuss the assignment, your results, and the answers to these questions. If you understand how to answer these questions, you should have no difficulty with those that will be asked. However, **these questions are not the exact questions that you will have to answer.**

1. What is the result of increasing the sample size on the calculated value of the test statistics?
   a. The calculated value of the *F*-statistic increases.
   b. The calculated value of the *F*-statistic decreases.
   c. The calculated *p*-value increases.
   d. The calculated *p*-value decreases.
   e. The value remains unchanged.
2. What is the result of increasing the sample size on the table value of the test statistics?
   a. The calculated value of the *F*-statistic increases.
   b. The calculated value of the *F*-statistic decreases.
   c. The calculated *p*-value increases.
   d. The calculated *p*-value decreases.
   e. The value remains unchanged.
3. What is the result of increasing the sample size on the *p*-value of the test statistics?
   a. The calculated value of the *F*-statistic increases.
   b. The calculated value of the *F*-statistic decreases.
   c. The calculated *p*-value increases.
   d. The calculated *p*-value decreases.
   e. The value remains unchanged.
4. What is the result of increasing the significance level on the *p*-value of the test statistics?
   a. The calculated value of the *F*-statistic increases.
   b. The calculated value of the *F*-statistic decreases.
   c. The calculated *p*-value increases.
   d. The calculated *p*-value decreases.
   e. The value remains unchanged.
5. What is the result of increasing the significance level on the calculated value of the test statistics?
   a. The calculated value of the *F*-statistic increases.
   b. The calculated value of the *F*-statistic decreases.
   c. The calculated *p*-value increases.
   d. The calculated *p*-value decreases.
   e. The value remains unchanged.

Appendix C. DSCI 3710 Mathematica Group Laboratory: Exercise 2

All students have access to Mathematica in the COBA computer labs. Access will be discussed further in your class.

These instructions outline the steps you need to take in order to complete your laboratory assignment. They should serve as a guide and are not intended to provide each step and detail of what you should do because the laboratory is intended to provide you with a vehicle for investigating concepts.

Please read the instructions listed below before you begin the laboratory exercise. By doing so, you will have a clearer understanding of the purpose of the laboratory exercise and what you can do to investigate the concepts involved.

Description

The intent of this laboratory exercise is to have you work in a group of six to collect data and experiment using Mathematica with the sample data and the significance level to determine the effect of each on the results of the ANOVA test. The steps that follow are intended to guide your experimentation.

Getting Data

For this group exercise you will need to work as a team to collect data that you can enter into the sample fields in Mathematica.

Think of a problem that interests you for which you can get data for 3 to 5 samples. For example:

a. Your team could collect the price of 5 values meals at 3 more different fast food restaurants.

b. Your team could ask 5 undergraduate male students, 5 undergraduate female students, 5 graduate male students, and 5 graduate female students how much they spend on gas for their autos each week.

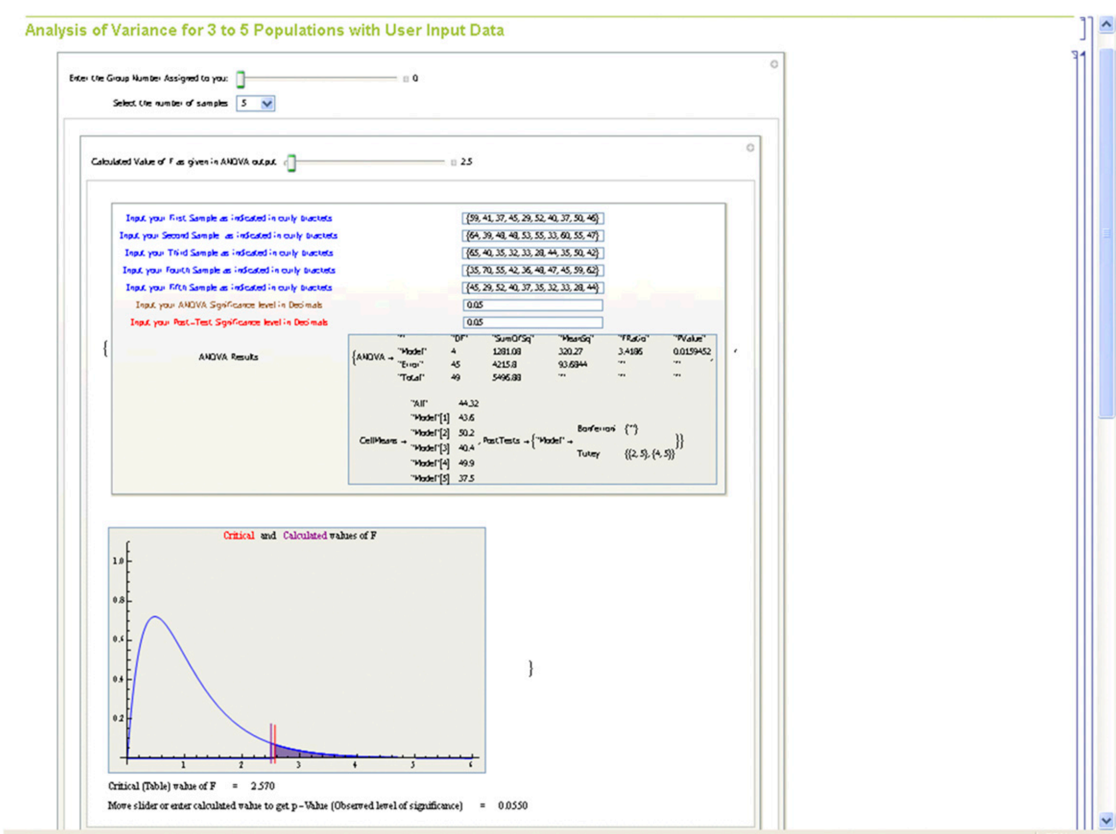
Figure C.1 shows a screenshot of the laboratory layout in Mathematica for this ANOVA group experiment.

1. Use your mouse to move the group number slider so that the display shows the group number that was assigned to your team.

2. Experiment with changing the numerical values in at least one of the samples. For example, try adding a constant to each value, subtracting a constant from each value, multiply each value by 2, and finally divide each value by 2.

a. Take some notes about how the calculated F-statistic changes.

Figure C.1. Mathematica Exercise #2 Sample Screenshot



b. Take some notes about how the p-value changes.

c. Take notes some notes about how the PostHoc tests change.

3. Experiment with the statistical significance.

a. Take some notes about how the calculated F-statistic changes.

b. Take some notes about how the p-value changes.

c. Take notes some notes about how the PostHoc tests change.

4. Print the screen for one of our experiments with your group number by pressing print screen. Then print each of your names on that page and bring the printed page to class.

5. Once you are done see if you can answer the questions that follow.

The questions on the next page are much like the quiz you will have as WEBTEST for this assignment. You should attempt to answer these before coming to the experiential class session on this laboratory. In class your group can discuss the assignment, your results, and the answers to these questions. If you understand how to answer these questions, you should have no difficulty with those that will be asked. However, **these questions are not the exact questions that you will have to answer.**

6. What is the result of increasing the numerical values in one of the samples?

a. The calculated value of the F-statistic increases.
b. The calculated value of the F-statistic decreases.
c. The calculated p-value increases.
d. The calculated p-value decreases.
e. The value remains unchanged.

7. What is the result of decreasing the numerical values in one of the samples?

a. The calculated value of the F-statistic increases.
b. The calculated value of the F-statistic decreases.
c. The calculated p-value increases.
d. The calculated p-value decreases.
e. The value remains unchanged.

8. What is the result of increasing the significance level on the p-value of the test statistics?

a. The calculated value of the F-statistic increases.
b. The calculated value of the F-statistic decreases.
c. The calculated p-value increases.
d. The calculated p-value decreases.
e. The value remains unchanged.

9. What is the result of increasing the significance level on the calculated value of the test statistics?

a. The calculated value of the F-statistic increases.
b. The calculated value of the F-statistic decreases.
c. The calculated p-value increases.
d. The calculated p-value decreases.
e. The value remains unchanged.

10. What is the result of decreasing the significance level on the table value of the test statistics?

a. The calculated value of the F-statistic increases.
b. The calculated value of the F-statistic decreases.
c. The calculated p-value increases.
d. The calculated p-value decreases.
e. The value remains unchanged.

Appendix D. DSCI 3710 Excel Assignment 2

We assume that all students have had an introduction to spreadsheets. The instructions given below are an outline of the steps you need to take in order to complete your assignment. **These instructions build on the ones given for the first assignment** and therefore should be regarded as a guide and not as a detailed map of the complete procedure. Further help on maneuvers in Excel is available by clicking on the Excel Help menu.

Please be sure to read the case description for each problem before you begin the computer exercise. By so doing, you will have a clearer understanding of the purpose of the exercise and how you will conduct the analysis. This could help reduce the amount of time you spend in the computer laboratory on this assignment. **Answer the questions listed in this handout.**

Currentprices.com keeps a record of the sales prices of gasoline ($/gallon, at pump) at different retailing pumps/locations. The data on regular unleaded gasoline are recorded at 37 different pumps at 4 different locations, viz., Allen, Blaze, Corlis, and Dustin. The data are presented in the spreadsheet entitled Assgt#2.xls.

You have to: (1) analyze the data for the existence of any difference between the true mean prices at the four different locations using the ANOVA procedure; (2) conduct Tukey's multiple comparison procedure on the data; (3) construct a 95% confidence interval for the mean price of regular gasoline in Allen; and (4) construct a 95% confidence interval for the difference between the mean prices in Dustin and Blaze. The general procedure is outlined below.

1. Open the file Assgt#2.xls.

2. Insert a row (under Edit menu) at the top of the spreadsheet and then label the columns A, B, C, and D appropriately (Allen, Blaze, Corlis, and Dustin, for example).

3. To conduct the ANOVA, select the Add-Ins tab (in 2003 use the KPK pull-down menu) and then choose KPK data analysis/One Factor ANOVA.

4. Highlight the input range for the data in the four columns (A1–D38). Click on the output radio button, and choose cell F1 for the output range. Accept the default alpha of 0.05 for the ANOVA procedure.

5. Select (by placing checkmarks in the field) Tukey Test for Difference between means, at the default alpha of 5%. You should then use Excel or hand calculate both (a) and (b).

a. To calculate the confidence interval for difference between two means, select the appropriate group numbers (viz., 4 and 2) and do the following. For the confidence interval for the difference in two means, you need to use the pooled standard deviation from the ANOVA output (square root of MSE) and build your own formula for the difference in two means or hand calculate and enter the values.

b. To calculate the confidence interval for single group mean, select the appropriate group (viz., 1) and either build the formula in Excel using the pooled standard deviation or get an approximation of this interval selecting the confidence interval as the method for the t-test in the One Population Inference Option from the KPK menu.

6. To reveal any hidden output in your ANOVA table, drag the column line to expand the width of the appropriate column. An example of what your output should look like is shown below. Note: Some numbers have been replaced by x's (Table D.1).

Table D.1. Assignment #2 ANOVA Table

ANOVA: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Allen	37	xx	1.734	xx		
Blaze	37	xx	1.643	xx		
Corliz	37	xx	1.681	xx		
Dustin	37	xx	1.989	xx		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	Xx	3	xx	52.135	xx	2.667
Within Groups	Xx	xx	xx			
Total	Xx	147				
TUKEY MULTIPLE COMPARISON TEST						
Critical Q	Distance	Alpha				
3.68	Xx	0.05				
Means joined by a double line are not significantly different.						
Blaze	Corliz	Allen	Dustin			
xx	Xx	xx	xx			
Means joined by a double line are not significantly different.						
Blaze	Corliz	Allen	Dustin			
xx	Xx	xx	xx			
95% CI FOR MEAN OF Allen						
Mean of Allen	Critical t	Standard error	Lower limit	Upper limit		
xx	Xx	xx	xx	xx		
95% CI FOR Dustin – Blaze						
Difference	Critical t	Standard error	Lower limit	Upper limit		
0.346486	Xx	0.030587	xx	xx		

Experiential Exercise

As you work this assignment, think about the following questions. Then form a team of 3 to 5 and discuss each of the following. You can engage in this discussion by meeting, or your group can use a wiki to engage in an online discussion. Instructions for setting up a wiki are provided on the Excel assignment page of our course website.

1. What wording tells you the alternative and null hypothesis?
2. What wording tells you the type of statistical test to perform? For example, is a *z*- or *t*-statistic appropriate?
3. What wording tells you that this is a one- or two-tail hypothesis test?
4. What were the steps you used to obtain the calculated value of the test statistic?
5. What were the steps you used to obtain the critical value of the test statistic?
6. How do you use the calculated and critical value to make a statistical decision about this test?
7. How do you obtain the *p*-value for the test statistic?
8. How do you use the *p*-value and the level of significance to make a statistical decision about this test?
9. How does the result of your test relate to the statistical significance of your findings?
10. What managerial implications can you conclude from the results of your test?

To be ready for the "Excel Quiz 2" HLS Web Test, you should prepare the following:

1. The correct output for the ANOVA and comparison of means.
2. The five-step hypothesis test for the difference between the mean gasoline prices.
3. Which location can be singled out as having a mean price that is different from others (by the Tukey's multiple comparison procedure)?

The Excel assignments are each graded via a short Excel Quiz in **HLS Web Test** that is open for about 48 hours as listed in the syllabus and in your HLS progress report. You are expected to use your output and written answers to complete the quiz. You are not required to turn in the output.

The questions below are much like the quiz you will have in WEBTEST. If you can answer these, you should have no difficulty with those that will be asked. However, **these questions are not the exact questions that you will have to answer.**

The correct answers to the sample questions are marked with an asterisk (*).

1. Which location is most likely to have the lowest price per gallon?
   A. Allen B. Blaze* C. Corlis D. Dustin
2. What is the *p*-value (observed level of significance) for the ANOVA *F*-statistic that tests the equality of gasoline prices in the four locations?
   A. 0.902 B. 0.017 C. 0.0000* D. 2.67 E. 0.0796
3. The decision and conclusion based on the ANOVA test of equality of mean prices is:
   A. Reject Ho, conclude there is evidence of a difference.*
   B. F.T.R. Ho, conclude there is insufficient evidence of a difference.
   C. Reject Ho, conclude there is insufficient evidence of a difference.
   D. F.T.R. Ho, conclude there is evidence of a difference.
   E. Neither reject nor F.T.R. Ho, test is inconclusive.
4. What is the Tukey's critical difference between means?
   A. 0.902 B. 0.017 C. 0.0000 D. 2.67 E. 0.0796*

5. How much variation (in terms of sum of squares) is explained by location?

A. 2.7* B. 2.5 C. 5.20 D. 0.90 E. 0.017

Thanks

References

Akçay AE, Ertek G, Büyüközkan G (2012) Analyzing the solutions of DEA through information visualization and data mining techniques: SmartDEA framework. *Expert Systems Appl.* 39(9): 7763–7775.

Andrienko G, Andrienko N, Jankowski P, Keim D, Kraak MJ, MacEachren A, Wrobel S (2007) Geovisual analytics for spatial decision support: Setting the research agenda. *Internat. J. Geographical Inform. Sci.* 21(8):839–857.

Aschwanden C (2015) Not even scientists can easily explain P-values. Accessed October 10, 2018, http://fivethirtyeight.com/features/not-even-scientists-can-easily-explain-p-values/.

Baloukas T, Paparrizos K, Sifaleras A (2005) Promoting operations research education using a new web-accessible didactic tool. *Proc. 7th Balkan Conf. Oper. Res.* (University of Bucharest, Bucharest, Romania), 393–398.

Baloukas T, Paparrizos K, Sifaleras A (2009) Teaching note—An animated demonstration of the uncapacitated Network Simplex algorithm. *INFORMS Trans. Ed.* 10(1):34–40.

Banerjee G, Patwardhan M, Mavinkurve M (2013) Teaching with visualizations in classroom setting: Mapping instructional strategies to instructional objectives. *Internat. Conf. Tech. Ed.* (Conference Publishing Services, Los Alamitos, CA), 176–183.

Camm J, Cochran J, Fry M, Ohlmann J, Anderson D (2017) *Essentials of Business Analytics*, 2nd ed. (Cengage, Boston), 626–627.

Chopra S, Meindl P (2015) *Supply Chain Management: Strategy, Planning, and Operation*, 6th ed. (Pearson Publishing, London).

Eisenstadt M, Price BA, Domingue J (1992) Software visualization as a pedagogical tool. *Instructional Sci.* 21(5):335–364.

Evans JR (2015) Teaching note—Analytic visualization of solver output. *INFORMS Trans. Ed.* 15(3):232–239.

Garay GR, Tchernykh A, Drozdov AY, Garichev SN, Nesmachnow S, Torres-Martinez M (2019) Visualization of VHDL-based simulations as a pedagogical tool for supporting computer science education. *J. Comput. Sci.* Forthcoming.

Huang W, Eades P, Hong SH (2009) Measuring effectiveness of graph visualizations: A cognitive load perspective. *Inform. Visualization* 8(3):139–152.

Jones CV (1996) *Visualization and Optimization* (Springer Science & Business Media, New York).

Keim D, Andrienko G, Fekete JD, Görg C, Kohlhammer J, Melançon G (2008) Visual analytics: Definition, process, and challenges. Kerren A, Stasko JT, Fekete J-D, North C, eds. *Information Visualization* (Springer, Berlin), 154–175.

Keim DA, Mansmann F, Schneidewind J, Ziegler H (2006) Challenges in visual data analysis. *Proc. 10th IEEE Internat. Conf. Inform. Visualisation (IV'06)* (IEEE Computer Society, Los Alamitos, CA), 9–16.

Keim DA, Munzner T, Rossi F, Verleysen M (2015) Bridging information visualization with machine learning (Dagstuhl Seminar 15101). *Dagstuhl Rep.* 5(3):1–27.

Kulkarni SS (2011) Analytical clarifications and an easy Excel-based implementation of a basic location model in operations management. *Internat. J. Inform. Oper. Management Ed.* 4(1):22–34.

Kumar AN (2015) The effectiveness of visualization for learning expression evaluation. *Proc. 46th ACM Tech. Sympos. Comput. Sci. Ed.* (ACM, New York), 362–367.

Kuo C, White RE (2004) A note on the treatment of the center-of-gravity method in operations management textbooks. *Decision Sci. J. Innovative Ed.* 2(2):219–226.

Lazaridis V, Paparrizos K, Samaras N, Sifaleras A (2007) Visual LinProg: A web-based educational software for linear programming. *Comput. Appl. Engrg. Ed.* 15(1):1–14.

Liu LH, Deng FA (2011) Reformation of mathematical experiment course system and construction of practical platform in college. *Modern Comput.* 6:9.

Markham IS, Palocsay SW (2006) Scenario analysis in spreadsheets with Excel's Scenario tool. *INFORMS Trans. Ed.* 6(2): 23–31.

Naps TL, Rößling G, Almstrum V, Dann W, Fleischer R, Hundhausen C, Velázquez-Iturbide JÁ (2002) Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bull.* 35(2):131–152.

Ramsden P (1997) Mathematica in education: Old wine in new bottles or a whole new vineyard? *Innovation in Mathematics*, vol. 15 (WIT Transactions on Engineering Sciences, Southampton, UK), 419–426.

Sato H, Muro K, Hasegawa A, Zhao D (2003) Mathematica as a powerful and practical tool for displaying animated three-dimensional structures of the Earth's interior. *Geochemistry Geophys. Geosystems* 4(4):1–8.

Saunders-Smits GN, Gürdal Z, Hol J (2005) Innovative teaching of aircraft structural analysis and design courses—Mathematica in an engineering education environment *Proc. 2005 Amer. Soc. Engrg. Ed. Annual Conf. Exposition* (American Society for Engineering Education, Washington, DC), 8143–8153.

Shoufan A, Lu Z, Huss SA (2015) A web-based visualization and animation platform for digital logic design. *IEEE Trans. Learn. Tech.* 8(2):225–239.

Simpson NC, Hancock PG (2010) The tile puzzle: An experiential exercise in introductory Management Science. *BRC Acad. J. Ed.* 1(1):33–46.

Sweller J, van Merriënboer JJG, Paas F (1998) Cognitive architecture and instructional design. *Ed. Psych. Rev.* 10:251–296.

Treisman A (1985) Preattentive processing in vision. *Comput. Vision Graphics Image Processing* 31:156–177.

Treisman A, Gelade G (1980) A feature-integration theory of attention. *Cognitive Psych.* 12:97–136.

Tufte E (2001) *The Visual Display of Quantitative Information* (Graphics Press, Cheshire, CT).

Van Merrienboer JJ, Sweller J (2005) Cognitive load theory and complex learning: Recent developments and future directions. *Ed. Psych. Rev.* 17(2):147–177.

Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) *Engineering Psychology and Human Performance* (Routledge Taylor & Francis Group, London).

Wood RE (1986) Task complexity: Definition of the construct. *Organ. Behav. Human Decision Processes* 37(1):60–82.