



StatPlayground: A Sandbox for Learning Practical Statistics

Krishna Subramanian^(✉), Jeanine Bonot, Radu A. Coanda, and Jan Borchers

RWTH Aachen University, 52056 Aachen, Germany
{krishna,borchers}@cs.rwth-aachen.de,
{jeanine.bonot,radu.coanda}@rwth-aachen.de

Abstract. Inferential statistics is a frequent task in several fields such as HCI, Psychology, and Medicine. Research shows that inferential statistics is often used incorrectly because the underlying statistical concepts are misunderstood. From interviews with students in an HCI lab, we find that, in addition to theoretical knowledge of statistics, novice analysts require statistical know-how, i.e., practical knowledge of how various data characteristics are inter-related and how they influence significance test selection and statistics, to analyze data. However, current learning resources such as books and online searches are not adequate to help learn statistical know-how. As a possible solution, we present StatPlayground, an interactive web app that supports exploratory learning of statistical know-how. StatPlayground does this by allowing users to modify data via direct-manipulation of visualizations, to see how those changes affect other data characteristics such as the shape of the distribution and variance of the data, as well as the resulting significance test and statistics such as effect size and p -value. StatPlayground can be combined with traditional teaching methods and can help prepare students for real-world analysis. Our evaluation of StatPlayground with graduate students shows the potential of StatPlayground to help learn statistical know-how and design implications for simulation tools for learning statistics.

Keywords: Statistical analysis · Exploratory learning · Data visualization

1 Introduction

Inferential statistics is used to validate experimental findings in several scientific fields such as HCI [6], Medicine [2], and Psychology [1]. Over the past two decades, however, research has shown that inferential statistics is often used incorrectly [6, 16, 19]. A potential reason is the lack of adequate understanding of statistical concepts [6, 14, 28].

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-29384-0_9) contains supplementary material, which is available to authorized users.

To understand how students prepare to do statistical analysis, we interviewed twelve students (undergraduate, graduate, and doctoral) from the HCI research lab at our university who analyze real-world data. We found that, while students had formal training in statistics, they had to learn practical statistics and statistical tool usage in order to perform analysis. We refer to this practical knowledge of statistics as *statistical know-how*. Examples include understanding the influence of outliers on the shape of a distribution, knowledge of which test to perform when data has a uniform distribution, and understanding how data transformations could affect test selection and statistics. Resources that students use in order to prepare to do analysis either do not help gather statistical know-how or are often inaccurate and untrustworthy. Furthermore, with these resources, applying the knowledge to the context of their own analysis was reported to be difficult.

As a potential solution to help students learn statistical know-how, we introduce StatPlayground. StatPlayground is a web app that visualizes *hypothetical* data, computes the inferential statistics automatically by checking the assumptions of the data, and visualizes the results. Users manipulate the data to see how this affects other data characteristics and the resulting statistics in real time. E.g., adding an outlier to a distribution that is normally distributed, might result in the distribution being no longer normal, and the selection of a non-parametric test. StatPlayground facilitates exploratory learning to help users construct their own statistical know-how at their own pace. Unlike current statistical simulations that focus on few concepts, StatPlayground can help users understand how several statistical concepts are interrelated. Although statistical significance testing is still the most common analysis method in HCI, it does suffer from methodological issues [10, 19]. For this reason, StatPlayground supports frequently used significance tests, but encourages the use of methods considered more meaningful such as *effect-size* and *95% confidence intervals* [9].

This paper contributes the following: (1) Results of a formative study that informs us about current statistical practice in HCI, (2) StatPlayground, a web application for exploratory learning of statistical know-how, and (3) findings from a preliminary user study to validate StatPlayground and the resulting design implications for teaching statistics using simulations.

2 Motivation and Related Work

We first describe the formative study that motivated StatPlayground and then discuss prior research in statistical education, simulations, and direct manipulation techniques.

2.1 Motivational Study: Statistical Practice in HCI Research

While previous research has investigated how students learn statistics in schools, not much is known about how students in universities prepare for a real-world analysis task. To investigate this, we interviewed 12 students (3 female) from the

local HCI lab. Our participants include two undergraduate, eight graduate, and two doctoral students. All participants but one had performed at least one analysis in the past two months (median = 10 days). Three participants had ongoing analysis during the time period of our study, which we observed. On average, interviews lasted 45 min and observations lasted 1 hour. We analyzed the transcripts and observation videos by applying the *contextual design* methodology [4]. While our sampling accounts for variations in participants' educational background and experience, it is limited to HCI domain. The major findings from our study are:

Formal Education is Not Adequate for Real-World Analysis: We found that although students had taken a statistical course earlier, most (11/12) had to learn statistics again before working on a real-world analysis task. This might be because statistical courses tend to teach predominantly theoretical concepts. To perform real-world analysis, students gather two types of knowledge: statistical know-how and statistical tool usage. Statistical know-how, as we discussed earlier, involves an understanding of how various statistical concepts are inter-related. “Which test should I apply here?”, “How do outliers affect my distribution?”, and “What happens if my data is left skewed?” are some examples of statistical know-how students reported having to learn before conducting analysis.

Current Resources Do Not Help Learn Statistical Know-How: Students reported resorting to *just-in-time learning* of statistics to gather statistical know-how. As a result, traditional learning methods like books, lecture slides, and online courses, while considered useful, were reported to be infeasible to help prepare to do statistics. Students reported that, while using these resources, they had to spend too much time to identify information that is *relevant* for their analysis. Tutorials like ps4hci¹ and hcistats² helped learn tool usage, but not statistical know-how. On the other hand, online searches and web articles helped students answer specific questions about statistical know-how, but often had inaccurate or unreliable information. Even when students found reliable sources, they had to apply the knowledge in those sources to the context of their analysis, which is not straightforward.

2.2 Statistical Literacy and Principles in Statistical Education

Prior research in statistical education has shown that *active learning* tools can benefit students. E.g., Garfield suggests that instructors should allow students to construct knowledge of statistical concepts on their own based on prior knowledge [14]. Learning tools that facilitate this should allow students' guesses and predictions to be confronted with actual results through real-time feedback. In a similar vein, Biehler also recommends the use of a ‘*sandbox*’ tool that allows students to explore statistical concepts by themselves [5]. Furthermore, Stat-Playground has the potential to help students gain the ability to make sense of

¹ <https://depts.washington.edu/madlab/proj/ps4hci/>.

² <http://yatani.jp/teaching/doku.php?id=hcistats:start>.

statistical information [11], the ability to identify the relationships between statistical concepts [14], have better data awareness [23], and understand statistical procedures better [13], all of which are considered to be key skills a statistician should have.

2.3 Simulations

The use of simulations for learning is not novel. E.g., simulations have been used to teach electrical circuits [7] and mathematics [22]. In statistical education, Rice Virtual Lab in Statistics³, Seeing Theory⁴, and Wise Interface for Statistical Education⁵ are some prominent simulations. While research on the effectiveness of simulations on statistical education is sparse, there are some promising results. Simulations lead to better learning outcomes in comparison to traditional teaching methods [27]. But there is a caveat—not all simulations are effective; their *design* and how they are used play a critical role. E.g., students learn better from simulations when they are given appropriate questions to answer [17].

StatPlayground extends current statistical simulation tools in two ways. First, while current simulations focus on teaching a few statistical concepts, StatPlayground can help users understand how several statistical concepts are inter-related. This helps users gain a more *holistic* understanding of statistical concepts. Second, current simulations are *rigid*—they depict one example under certain fixed conditions. As a result, the knowledge transfer from simulations to the context of user’s own analysis could be tedious and lead to further complications. StatPlayground allows for more flexibility by allowing the user to shape the data in several ways: Change the distribution type, change measures of central tendency and spread, add or remove individual data points, etc.

2.4 Direct Manipulation of Visualizations

In StatPlayground, users directly manipulate visualizations to modify data characteristics. Direct manipulation [25] is an established interaction technique that offers several benefits like real-time feedback and more user engagement [20]. We used prior research works that utilized direct manipulation successfully to manipulate data [18, 20, 24] to design StatPlayground. E.g., *foreshadowing*, a technique we use in StatPlayground to provide *feedforward* to the user to preview resulting statistics (see Fig. 3), is inspired by *OctoPocus* [3] and *Design by Dragging* [8].

3 System Design

We used an iterative design process to develop StatPlayground⁶. Initially, a proof-of-concept prototype was evaluated with five experts, all of whom have at

³ <http://onlinestatbook.com/rvls.html>.

⁴ <http://seeingtheory.io>.

⁵ <http://wise.cgu.edu>.

⁶ See supplements.

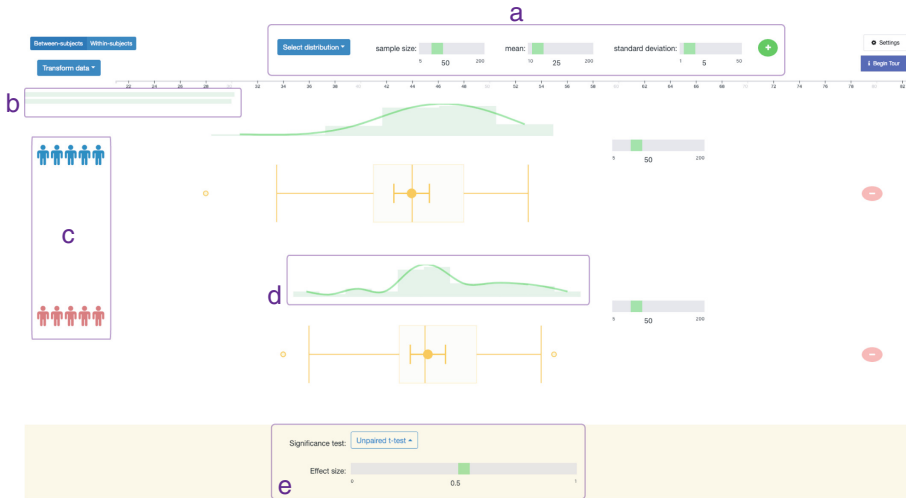


Fig. 1. StatPlayground contains three main panels: distribution generator (a), which can be used to generate the common data distributions by specifying parameters; results (e), which display the results of the current data configuration; plot, which uses box plot to visualize data distributions. Additionally, StatPlayground visualizes other data characteristics: shape of the distribution (d), homogeneity of the variances (b), and experimental design (c).

least two years of experience performing analysis; two experts teach statistics to undergraduate students. As a result of this evaluation, we identified features to be added to and removed from StatPlayground as well as some usability improvements.

Layout: StatPlayground has three panels: *data generation* (Fig. 1a), *plot*, and *results* (Fig. 1e). To begin with, the user generates common distributions like normal, uniform, and binomial distributions by setting the relevant parameters in the data generation panel. The generated data is then visualized as a box plot that shows the mean, median, inter-quartile range, and outliers of the distribution, as well as the confidence interval of the mean. Other data characteristics that influence inferential statistics are also visualized: shape of the distribution as a histogram curve (Fig. 1d), variance of each distribution as an animated bar chart at the top-left corner (Fig. 1b), and experimental design of the factor (Fig. 1c).

Interaction Design: Users interact with the plots in StatPlayground through *direct manipulation*. Click-and-dragging interface elements *modifies* the underlying data e.g., click-dragging a mean modifies the underlying mean of a distribution by modifying individual data points. Users can also create or delete data by clicking on either side of the whiskers and by clicking on an existing data point respectively. E.g., Fig. 3 (left) shows the interface after an outlier has been added; the histogram curve now indicates that the *test for normality* has failed.

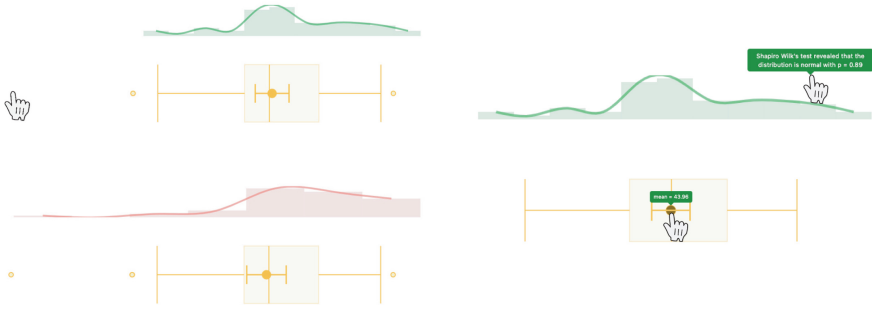


Fig. 2. Left: Users can create, edit, or delete data by directly interacting with the boxplot. Here, the user clicks to create an outlier. StatPlayground visualizes the change in data characteristics: the measures of central tendency shifts and the distribution is no longer normal! **Right:** StatPlayground uses progressive disclosure to reveal detailed information. User hovers over interface elements to view tooltips.

To avoid cluttering the interface, StatPlayground uses *progressive disclosure* to provide information on demand. E.g., hovering over any data point shows the characteristic and value (Fig. 3 bottom-right) and hovering over the histogram curve reveals statistics (Fig. 3 top-right).

Hovering over interface elements for 2s shows *fine-level controls* that can be used to lock/unlock and set the range of values for data characteristics. This can be used to ask more fine-grained, what-if questions like “*Would the effect size of this comparison of two distributions increase even if the means of both distributions remain the same i.e., locked?*” Additionally, during direct manipulation, StatPlayground visualizes possible results the user can expect to reach along the path of her manipulation via *foreshadowing*. E.g., in Fig. 3, the user is click-dragging the mean of a distribution further away from the other distribution and foreshadowing helps see that moving them closer results in *smaller* effect sizes.

Statistical Computations in StatPlayground: Computing descriptive and inferential statistics from data is done mostly natively via JavaScript libraries⁷. However, inverse-computation of potential datasets from given inferential statistics (p -value, effect size), used in *foreshadowing*, is an interesting non-deterministic problem that has not been solved earlier. Prior research [21] is limited to inverse computation of datasets from *descriptive statistics*, not inferential statistics. Therefore, we built a Python package, *Cheno*⁸ to perform inverse-computations of data from inferential statistics.

⁷ <https://github.com/pieterprovoost/jerzy>, <https://jstat.github.io>.

⁸ Cheno is an anagram of Cohen, a famous statistician who pioneered statistical power and effect size. See supplements.

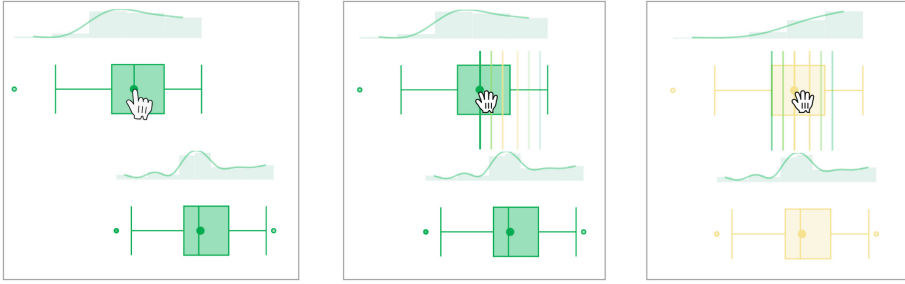


Fig. 3. StatPlayground uses *feedforward* to help users preemptively view statistical results. **Left:** Two distributions are being compared; user changes the mean of the distribution above by click-dragging it towards the mean of the distribution below. **Middle and right:** Feedforward elements indicate various effect sizes that would result upon further dragging.

4 Evaluation

We evaluated StatPlayground holistically to gauge its potential to help learn practical statistics.

Participants: 13 participants (4 female) volunteered to take part in the study. All participants are university students and have taken an introductory statistics course and/or performed an analysis earlier.

Procedure: Participants were presented with a hypothetical dataset and a set of questions to answer using StatPlayground. Questions were *story-driven* e.g., “assume that of all participants in a text-entry experiment, one participant has practiced touch-typing; how does she (i.e., the outlier) affect your analysis?” Follow-up questions were used to test if the participant was able to understand a concept that had come up earlier. We used a question-driven approach instead of open-ended exploration to motivate our participants. Participants were encouraged to think-aloud and we captured the audio and screen of the session.

Findings and Discussion: By analysing the collected data, we identified instances where StatPlayground helped participants gather statistical know-how e.g., how the measures of central tendency (mean, median) and spread (variance) influence the shape of the distribution, how effect sizes work in comparison to p -values, and how confidence intervals work. Participants found StatPlayground to be *easy to use* and *fun*. P3 compared StatPlayground to traditional learning methods, commenting “this is a more intuitive way of grasping the concepts and understanding how they relate to each other.” P13 mentioned that if she were to continue using this tool, she could even *predict* the statistics from box plots. Nevertheless, some participants required more help than what was provided in the tutorial to understand certain concepts. Some participants also had difficulty using and interpreting fine-level control and homogeneity of variances graph respectively. We have since modified the app based on their feedback.

5 Implications and Future Work

Interaction Modes: Participants had varied motivation levels. Some participants were quite adventurous and explored StatPlayground without many prompts, while others were cautious (“*I could trial and error, but I am not sure whether what I observe will be interesting.*” -P7). StatPlayground could cater to both of these target user groups. Cautious novices could use StatPlayground in a *guided mode*, where the interaction is driven by our prompts e.g., as in-place embellishments in visualizations [26]. More adventurous users and experts, e.g., course instructors, could use StatPlayground in an *unguided mode* to learn and teach statistical concepts. Before using StatPlayground in a classroom, however, instructors should use prior research on statistical education [12,28] to design the prompts.

Help Compare Statistical Analysis Paradigms: During our study, we observed that participants compared the benefit of effect size and p -values, particularly noting how effect size tends to be less binary than p -values. StatPlayground could thus be used as a platform to compare paradigms of statistical methods e.g., Bayesian vs. frequentist inference.

Possible Misuse: Although StatPlayground is intended to be used as a learning tool, few participants wondered whether StatPlayground could be used for analyzing their own data e.g., as a quick-and-dirty way to *pilot test* analysis. However, since StatPlayground allows users to modify their data characteristics, using it for real-world analysis could lead the user to wander down the “*garden of forking paths*” [15]; the user would not be able to differentiate between exploratory and confirmatory analyses. To avoid this, StatPlayground currently does not allow users to work with their data.

6 Conclusion

We presented StatPlayground, an exploratory learning interface that allows students to construct their own understanding of statistical know-how by directly interacting with the visualization interface. While the user study indicates benefits of such an approach, future work can evaluate its effect on learning outcomes and statistical practice. Such exploratory learning interfaces provide an immersive, self-paced medium to bridge the gap between knowing the concepts in theory and applying them to a real-world problem. We are working on making StatPlayground available online at hci.rwth-aachen.de/statplayground.

Acknowledgements. We would like to thank Nur Hamdan and Chat Wacharamanatham for their valuable feedback during the course of this project.

References

1. Aron, A., Coups, E., Aron, E.: Statistics for Psychology. Always Learning, 6th edn. Pearson, Upper Saddle River (2013). <https://books.google.de/books?id=JtQPwAACAAJ>

2. Banerjee, A., Jadhav, S., Bhawalkar, J.: Probability, clinical decision making and hypothesis testing. *Ind. Psychiatry J.* **18**(1), 64–69 (2009). <https://doi.org/10.4103/0972-6748.57864>
3. Bau, O., Mackay, W.E.: OctoPocus: a dynamic guide for learning gesture-based command sets. In: *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology, UIST 2008*, pp. 37–46. ACM, New York (2008). <https://doi.org/10.1145/1449715.1449724>
4. Beyer, H., Holtzblatt, K.: *Contextual Design: Defining Customer-Centered Systems*. Interactive Technologies. Elsevier Science (1998). <https://books.google.de/books?id=sVKuMvaFzjQC>
5. Biehler, R.: Software for learning and for doing statistics. *Int. Stat. Rev.* **65**(2), 167–189 (1997). <https://doi.org/10.1111/j.1751-5823.1997.tb00399.x>
6. Cairns, P.: HCI... not as it should be: inferential statistics in HCI research. In: *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...But Not As We Know It - Volume 1, BCS-HCI 2007*, pp. 195–201. British Computer Society, Swinton (2007). <http://dl.acm.org/citation.cfm?id=1531294.1531321>
7. Carlsen, D.D., Andre, T.: Use of a microcomputer simulation and conceptual change text to overcome student preconceptions about electric circuits. *J. Comput. Based Instruction* **19**(4), 105–109 (1992). <https://eric.ed.gov/?id=EJ457935>
8. Coffey, D., Lin, C., Erdman, A.G., Keefe, D.F.: Design by dragging: an interface for creative forward and inverse design with simulation ensembles. *IEEE Trans. Visual Comput. Graphics* **19**(12), 2783–2791 (2013). <https://doi.org/10.1109/TVCG.2013.147>
9. Cumming, G.: The new statistics: why and how. *Psychol. Sci.* **25**(1), 7–29 (2014). <https://doi.org/10.1177/0956797613504966>. PMID: 24220629
10. Dragicevic, P.: Fair statistical communication in HCI. In: Robertson, J., Kaptein, M. (eds.) *Modern Statistical Methods for HCI*. HIS, pp. 291–330. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26633-6_13
11. Gal, I.: Adults’ statistical literacy: meanings, components, responsibilities. *Int. Stat. Rev.* **70**(1), 1–25. <https://doi.org/10.1111/j.1751-5823.2002.tb00336.x>
12. Garfield, J.: How students learn statistics. *International Statistical Review/Revue Internationale de Statistique* **63**(1), 25–34 (1995). <http://www.jstor.org/stable/1403775>
13. Garfield, J.: The challenge of developing statistical reasoning. *J. Stat. Educ.* **10**(3), null (2002). <https://doi.org/10.1080/10691898.2002.11910676>
14. Garfield, J., Ben-Zvi, D.: How students learn statistics revisited: a current review of research on teaching and learning statistics. *Int. Stat. Rev.* **75**(3), 372–396. <https://doi.org/10.1111/j.1751-5823.2007.00029.x>
15. Gelman, A., Loken, E.: The garden of forking paths: why multiple comparisons can be a problem, even when there is no “Fishing Expedition” or “p-Hacking” and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University (2013)
16. Gray, W.D., Salzman, M.C.: Damaged merchandise? A review of experiments that compare usability evaluation methods. *Hum. Comput. Inter.* **13**(3), 203–261 (1998). <https://doi.org/10.1207/s15327051hci13032>
17. de Jong, T., Härtel, H., Swaak, J., van Joolingen, W.: Support for simulation-based learning: the effect of assignments in learning about transmission lines. In: Díaz de Ilarraza Sánchez, A., Fernández de Castro, I. (eds.) *CALISCE 1996*. LNCS, vol. 1108, pp. 9–26. Springer, Heidelberg (1996). <https://doi.org/10.1007/BFb0022586>

18. Kandel, S., Paepcke, A., Hellerstein, J., Heer, J.: Wrangler: interactive visual specification of data transformation scripts. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2011, pp. 3363–3372. ACM, New York (2011). <https://doi.org/10.1145/1978942.1979444>
19. Kaptein, M., Robertson, J.: Rethinking statistical analysis methods for CHI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2012, pp. 1105–1114. ACM, New York (2012). <https://doi.org/10.1145/2207676.2208557>
20. Kondo, B., Collins, C.: DimpVis: exploring time-varying information visualizations by direct manipulation. *IEEE Trans. Visual Comput. Graphics* **20**(12), 2003–2012 (2014). <https://doi.org/10.1109/TVCG.2014.2346250>
21. Matejka, J., Fitzmaurice, G.: Same stats, different graphs: generating datasets with varied appearance and identical statistics through simulated annealing. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI 2017, pp. 1290–1294. ACM, New York (2017). <https://doi.org/10.1145/3025453.3025912>
22. Reed, S.K.: Effect of computer graphics on improving estimates to algebra word problems. *J. Educ. Psychol.* **77**(3), 285 (1985). <https://doi.org/10.1037/0022-0663.77.3.285>
23. Rumsey, D.J.: Statistical literacy as a goal for introductory statistics courses. *J. Stat. Educ.* **10**(3), null (2002). <https://doi.org/10.1080/10691898.2002.11910678>
24. Rzeszutowski, J.M., Kittur, A.: Kinetica: naturalistic multi-touch data visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2014, pp. 897–906. ACM, New York (2014). <https://doi.org/10.1145/2556288.2557231>
25. Shneiderman, B.: Direct manipulation for comprehensible, predictable and controllable user interfaces. In: Proceedings of the 2nd International Conference on Intelligent User Interfaces, IUI 1997, pp. 33–39. ACM, New York (1997). <https://doi.org/10.1145/238218.238281>
26. Srinivasan, A., Drucker, S.M., Endert, A., Stasko, J.: Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Trans. Visual Comput. Graphics* **25**(1), 672–681 (2019). <https://doi.org/10.1109/TVCG.2018.2865145>
27. Wackerly, D., Lang, J.: ExplorStat—active demonstration of statistical concepts. In: Joint Statistics Meeting, Chicago (1996)
28. Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., Chang, B.: What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *J. Stat. Educ.* **16**(2) (2008). <https://doi.org/10.1080/10691898.2008.11889566>