Developing an R Shiny Interactive Module for Teaching Statistics to High School Students

Preprint · July 2024

DOI: 10.13140/RG.2.2.28736.96005

1 author:

Jeffery Painter
Georgia Institute of Technology

Developing an R Shiny Interactive Module for Teaching Statistics to High School Students

Jeffery Painter[1*]

[1*]Georgia Institute of Technology, Raleigh, NC, USA.

Corresponding author(s). E-mail(s): jpainter32@gatech.edu;

Abstract

The following research investigates building an R Shiny interactive module to support instructors who teach statistics online to high school students. The framework allows for building interactive simulations and data explorations while facilitating classroom management through assessments and visual dashboards to gauge student performance.

Keywords: Pedagogy, R, Shiny, Teaching Statistics

1 Introduction

The demand for enhanced statistical literacy among students has reached unprecedented levels in today's data-centric world. As data becomes omnipresent, the capacity to comprehend, manipulate, and explore data stands as a crucial skill for students across diverse disciplines. Traditionally, the learning of statistical concepts and data manipulations has relied on lectures and rote memorization of formulas. Although there has been ongoing development in more advanced tools, such as web-based applets, to teach statistics, the focus has predominantly been on college undergraduates and graduate-level studies. However, there is a critical gap in addressing the specific needs of high school students and providing supportive tools for their instructors, an area that has been relatively overlooked until now.

1.1 The Need for Statistical Literacy

In order to discuss the need for statistical literacy, one must first give a general definition to better contextualize an argument for why it is necessary. While not exactly the same, one cannot have statistical literacy without also considering data literacy [1], and for simplicity, this report will generally group the concepts together.

One such definition is provided by Callingham, 2006 and purports that statistical literacy is:

> ...the ability to understand and critically evaluate statistical results that permeate our daily lives – coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions. [2]

Statistical (and data) literacy is of upmost importance in today's data-driven world for several reasons. First, statistical literacy is required to help us make more informed decisions about the world around us. Data is more abundant than ever and having the skills to analyze and interpret data allows people to make better choices in various aspects of their lives, from personal finances to healthcare. There is a marked difference in how one is able to think and solve problems given even minor statistical training, allowing individuals to better understand problems and issues relating to data interpretation [3].

As noted by Wolff et al, "as daily interactions with data become evermore commonplace and individuals more frequently make judgments from data and make decisions regarding the use of their own personal data" statistics and data literacy are becoming essential life skills [1].

Furthermore, statistical literacy fosters critical thinking and problem-solving abilities. It enables individuals to ask the right questions, evaluate the reliability of data sources, and discern meaningful insights from vast datasets while avoiding information overload and fatigue [4]. These skills are not only valuable in academic and professional settings, but also in everyday situations.

Statistical literacy goes beyond just mere data analysis as data literacy is essential for responsible citizenship. In an age where data privacy and security are paramount, understanding how data is collected, used, and shared empowers individuals to protect their digital identities and advocate for privacy rights.

Lastly, statistical and data literacy is a highly sought-after skill in the job market. Employers across various industries value candidates who can analyze data to drive informed decision-making and business strategies. Therefore, being data-literate enhances one's career prospects and employability. "Nearly 70 percent of the workforce would be expected to use data heavily in their work by 2025." [5]

In summary, statistical literacy is crucial because it equips individuals with the skills needed to make informed decisions, enhances critical thinking, promotes responsible citizenship, and improves career opportunities in an increasingly data-centric world.

1.2 The Current State of Statistical Education in High Schools

Statistical education in high schools across the United States and abroad is a topic of increasing importance, given the growing significance of data analysis in today's world. However, the current state of statistical education in these institutions is marked by considerable variability. In 2016, it was reported that less than 25% of high school students had taken a stand-alone statistics course while nearly 75% of high schools in

America offer a statistics course [6]. While the number of schools offering the statistics courses seems relatively high, it is still a far cry from being universally available. And even at those schools where it is offered, the number of students who voluntarily choose to take it is quite low. In communities with less resources and higher rates of ethnic minorities, these rates can be even far less [7].

While some schools offer robust and modernized statistics curricula, many others rely on traditional teaching methods that emphasize theory and formulaic calculations over practical application and data analysis. Statistical investigation and use of technology hardly occur in the current approach. The emphasis on calculating statistical measures contributes insufficiently to interpreting, critically evaluating and reasoning with data [8].

One common issue in high school statistics education is the lack of interactivity and engagement. Traditional teaching approaches often revolve around lectures, textbooks, and formula memorization, leaving students disconnected from real-world statistical applications [9]. This approach fails to captivate students' interest and hampers their ability to grasp complex statistical concepts, such as statistical power and sample size computation, which are increasingly important in fields beyond mathematics, including social sciences, health sciences, and business.

Furthermore, there is a significant gap in the availability of interactive and technology-driven tools for high school statistics education. While technology has revolutionized the field of statistics, its integration into high school classrooms remains limited [10] [11]. The absence of user-friendly, interactive applications tailored to high school students hinders their ability to explore data, visualize statistical concepts, and engage actively in the learning process. As a result, many high school graduates lack the statistical literacy required to navigate the data-driven world they encounter in college and beyond. Gelman and Glickman (2000) advocate for class-participation demonstrations as a more effective way to introduce students to data and sampling distributions compared to traditional lectures [12].

In summary, the current state of statistical education in high schools is characterized by a lack of uniformity and innovation. While some schools offer opportunities for students to engage in the data investigation process [13], many others rely on outdated teaching methods that do not adequately prepare students for the data-centric challenges of the modern world. Bridging this gap by introducing interactive, technology-driven tools and modern teaching approaches could significantly enhance statistical education and equip high school students with the skills they need to excel in a data-driven society.

2 Related Work

2.0.1 R Shiny as a Viable Teaching Platform

Recognizing the transformative potential of interactive learning in enhancing students' grasp of statistical concepts, the need for an effective teaching platform becomes apparent. R Shiny, a versatile and user-friendly platform for crafting interactive web applications, has emerged as a promising solution to bridge the gap between traditional teaching methods and modern, engaging learning experiences. Related works are explored below which demonstrate how R Shiny can serve as a valuable teaching platform, empowering both instructors and students in the realm of statistical education.

Several studies highlight the evolving use of R Shiny as a platform for teaching statistical concepts, dating back to 2008. However, the predominant focus has been on students at the undergraduate and graduate school levels. The imperative to develop more effective tools tailored to younger high school students remains a pressing need.

Ragasa's (2008) investigation into computer-assisted instruction versus traditional methods in a basic statistics class for college students underscores the positive impact of technology, such as R Shiny applications, on student achievement [14]. Miranda Freire's (2019) exploration of probability density functions, while centered on university students, aligns with the goal of enhancing high school students' statistical comprehension through R Shiny applications [15].

Kasprzak et al. (2020) provide a historical perspective on the growth of Shiny applications from 2012 to 2018. They emphasize R Shiny's ability to create specialized and user-friendly applications that can be easily shared and incorporated into workflows. While their focus is on research and collaboration, the insights into Shiny's progress and its utility in sharing statistical research visually and reproducibly are valuable. This background information helps contextualize the potential of R Shiny applications as educational tools [16].

Von Borries and de Castro Quadros (2022) introduce the ROC App, showcasing Shiny's capacity to aid students in understanding ROC curves, offering insights into creating educational applications for high school students [17]. Lazarski's (2021) comprehensive high school lesson plan using "Shiny Dice" illustrates the successful integration of an applet-based learning tool into the high school curriculum, providing valuable guidance for educators who wish to incorporate R Shiny applications into their lesson plans [13].

3 Implementation

This project focused on building an infrastructure to help construct and manage R Shiny web applications for use in a high school classroom. In addition, a simple lesson plan was developed to exemplify the use of this framework which focused on exploring the concepts of power and Type I and Type II error rates through data simulation tools built in R Shiny.

The implementation of the framework to support multiple teachers running multiple course sections required first developing a data model which supports user management, an access control layer and an authentication system. None of which are standard components when one typically builds an R shiny application today. The security and access layer concepts are borrowed from traditional web application development and help support the functionality required to assist educators with deliverying content to a large number of students while collecting assessment data that supports both data driven and visual analytics through an educator dashboard.

The dashboard view allows the educator to quickly assess where students are having difficulty and where they may need to spend more time to help students better understand statistical concepts and improve student performance.

All development was conducted within the RStudio IDE (integrated development environment). Some external libraries were utilized such as *shinyauthr*[1] and *shinyalert*[2] to help with user authentication and UI alerts.

For database storage, the SQLite database was employed and access is provided through the *RSQLite*[3] package.

4 Methodology

Development took place over many weeks. The initial lesson plan was developed in conjunction with two high school AP Statistics teachers who work at a private high school in Cary, NC. The instructors discussed the current limitations they experience when using R Shiny applications in the classroom, and their input was used to help guide the design and development of the current implementation. Furthermore, the instructors tested and evaluated the framework providing valuable feedback and recommended improvements along the way.

Generally, an R Shiny application is divided into two primary components, the UI (user interface) and server. When developing an R Shiny application, most development takes place in a single file typically called *app.R*. To help develop a more generalized framework for supporting educators who want to develop their own R shiny application, this framework has modularized the app.R into several sub-components which isolate the application by user role accessing the application.

Upon further inspection of our app.R, you will find that the UI includes multiple files (one each for the admin, teacher and student roles), as well as four individual source files for the server side of the application. There is a single *common_server.R* that includes the components required for user authentication, as well as individual server files corresponding again to each of the three access roles.

By separating the roles in this way, it provides a framework which is more easily maintained and also reduces the effort required for an educator to start taking advantage of the framework that has been created. This allows for easy re-use while allowing future users of this platform to create their own lesson plans and make full use of the assessment and dashboard frameworks that have been created in this project.

[1] https://cran.r-project.org/web/packages/shinyauthr/index.html
[2] https://cran.r-project.org/web/packages/shinyalert/index.html
[3] https://cran.r-project.org/web/packages/RSQLite/index.html

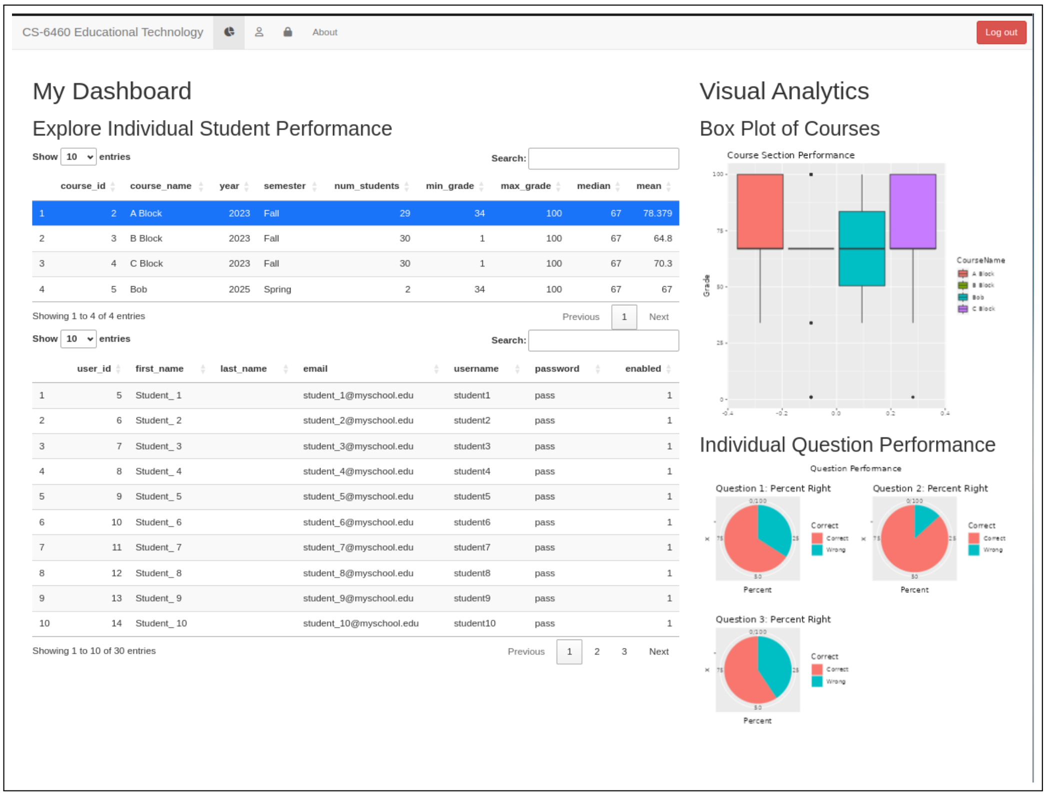

Fig. 1 Educator Dashboard View

5 Results

The results of this project are the creation of a framework which is capable of supporting multiple educators, running multiple courses and managing the assessment process of students taking advantage of simulation and data exploration tools built in an R Shiny environment.

The framework provides a unified way for educators to collect and analyze the performance of their students not only in a single class, but across multiple course sections. The framework has a dashboard with visual analytics demonstrating the ability for an instructor to assess performance across individual questions in the assessments they create, as well as the ability to deep dive into a single student's performance and gain an understanding of how many attempts it took the student to successfully complete their data investigation.

As shown in Figure 1, the instructor has a high level dashboard that provides both a data driven exploration of student assessments as well as data visualizations with broader overview of all the course sections they may be teaching. When a student logs into the R Shiny framework, they are presented with the data explorer and simulation tools as shown in Figure 2. Here, the student can read through the analysis task that

Fig. 2 Student Simulation Exploration

was developed to help them explore the concepts of statistical power, alpha and effect size, as well as how each of these factors impact Type I and Type II error rates. The data exploration task that was developed for this project has the students compare potential alternative chicken feeds from the one that a farmer is currently using to try and understand what effects various feeds might have on the growth of the farmer's chickens.

Once the student has completed exploring the data, they are able to test their understanding by answering a series of assessment questions. The system grades their responses in real time, and if they answer incorrectly, prompts the student to try again. Once the student has successfully completed answering all of the questions, they are alerted to their success.

6 Known Limitations

6.1 Issues Facing Statistical Instruction in High Schools Today

Several issues plague statistical instruction in high schools today, including limited use of technology, curriculum variability, and restricted access to data.

Research indicates that much of high school statistics education revolves around preparing students for the Advanced Placement (AP) exam, which offers college credit opportunities. While this can be advantageous for students, it often means they prioritize learning test-specific content over gaining a deep understanding of statistical and data literacy. High-stakes standardized tests sometimes influence curriculum decisions, promoting test-focused teaching instead of comprehensive statistical comprehension, discouraging innovative approaches [18].

High school statistics education often lags behind in integrating technology and interactive tools. Traditional teaching methods, such as lectures and textbooks, prevail, limiting students' exposure to modern statistical software and data visualization tools. Further, the high school statistics curriculum is highly *calculator* focused, and the use of alternative technologies (programming in R or Python) are not typically considered as they would be in a college course [19].

Second, variability in the curriculum is lacking, as many teachers themselves lack confidence in their own statistical abilities. Therefore, teaching this course usually goes back to more traditional pedagogical methods of assigning text book readings, lecture and less focus on developing deep statistical theory and knowledge. This is exactly the issue that [20] was attempting to address in his commentary on issues facing undergraduate instruction at the time of his writing. The content and depth of statistics courses can vary widely between high schools.

This may relate in part due to the data literacy of the teachers themselves. As reported, "teachers therefore must know how to use data. Becoming data literate is a skill set that educators must acquire and nurture throughout their careers." [21]

While some schools offer comprehensive statistics curricula, others provide only basic coverage or do not include statistics at all. Schools which have greater resources have been able to create quite remarkably sophisticated data investigation activities [13], but this is not the norm and expecting teachers to have these types of skills and capabilities is a high bar to achieve.

Teaching statistics faces several challenges today. Access to meaningful real-world data for students can be difficult. As shown in [22], there is a real advantage in building educational lessons that that emphasizes the value of engaging datasets connected to students' daily lives.

Additionally, statistics education is underrepresented in high schools nationally, with many not requiring it and students lacking opportunities unless they choose elective courses, potentially deepening educational inequalities [6] [7]. Furthermore, a shortage of qualified statistics instructors exists, as reported by [23]. Lee and Harrison (2021) note that while student enrollment in AP Statistics is rising, the number of capable teachers has not kept pace. Addressing these challenges necessitates a shift toward modern, interactive, and tech-driven statistics education, coupled with improved teacher training and standardized curricular guidelines to ensure consistent, high-quality instruction across high schools.

7 Conclusion

7.1 Preparing Teachers to be Successful in Teaching Statistics in K-12 Curriculum

To facilitate effective teaching of statistics in K-12 curriculum, teachers require essential resources and support. Research emphasizes the importance of nurturing mathematical and statistical mindsets to engage students and enhance their proficiency in critical statistical methods, contributing to a statistically literate society [24] [25].

Teachers need comprehensive support mechanisms, including professional development, specialized training, and access to educational technologies, enabling data exploration and fostering statistical and data literacy among students [26] [27].

Furthermore, teachers must acknowledge potential biases stemming from their expertise, which could hinder effective knowledge transfer to non-expert students [28]. Overcoming these biases is crucial for successful statistics instruction.

Additionally, educators require access to suitable tools and technologies for delivering course content effectively. Keengwe and Kidd (2010) provide historical context, tracing the evolution of online and distance education, highlighting the need for specialized K-12 statistical education tools [11]. While existing statistical materials are predominantly designed for graduate and undergraduate levels [29] [30] [31].

By providing teachers with these critical forms of support and resources, educational institutions can empower them to deliver high-quality statistics education in K-12, enhancing students' statistical literacy and preparing them for future academic and professional pursuits.

8 Future Work

While this project made great progress towards developing an interactive platform for enabling educators to take greater advantage of building R Shiny web applications, there is still much that could be done to enhance its capabilities. Future work should focus on enabling instructors to more easily create new assessment questions, and have those automatically populated in the student view. Currently, the assessment questions are hard coded. Furthermore, the dashboards and data visualizations that exist have been created with high level goals of demonstrating their capabilities. Once the application is actually being used, a survey of educators to better understand the types of reports that would be most beneficial to them to understand student performance could be developed. One example that could be investigated is capturing the time a student spends with the simulation tool and using that information to better understand which students are using the system more, either in actual time spent or how frequently they login to the system.

There are many opportunities to expand the capabilities of this framework to accommodate the needs of instructors teaching statistics at the high school level. The integration of additional modules and lesson plans is another area that deserves more work. Furthermore, integrating multiple modules into a single app should be possible, and putting controls into place that could allow instructors to release new modules as students progress over time could also benefit this application framework.

9 Declarations

Data availability There are no data associated with this paper. However, all source code is freely available from https://github.com/jlpainter/CS6460.

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

Conflict of interests This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue.

Funding No funding sources are attributable to the creation of this manuscript.

References

[1] Wolff, A., Gooch, D., Montaner, J.J.C., Rashid, U., Kortuem, G.: Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics **12**(3) (2016)

[2] Callingham, R.: Assessing statistical literacy: A question of interpretation. In: International Conference on Teaching Statistics (2006)

[3] Fong, G.T., Krantz, D.H., Nisbett, R.E.: The effects of statistical training on thinking about everyday problems. Cognitive psychology **18**(3), 253–292 (1986)

[4] Stanley, O.E.: Information overload: Causes, Symptoms, Consequences and Solutions. Asian Journal of Information Science and Technology **11**(2), 1–6 (2021)

[5] Pothier, W., Condon, P.: Cultivating a Data Literate Workforce: Considerations for Librarians. portal: Libraries and the Academy **23**(4), 629–636 (2023)

[6] Loewus, L.: Just 1 in 4 High School Seniors Have Taken Statistics. https://www.edweek.org/leadership/just-1-in-4-high-school-seniors-have-taken-statistics/2016/09 Accessed 2023-09-29

[7] Gagnon, D.J., Mattingly, M.J.: Limited access to AP courses for students in smaller and more isolated rural school districts (2015)

[8] Dijke-Droogers, M., Drijvers, P., Tolboom, J.: Enhancing statistical literacy. In: CERME 10 (2017)

[9] Boaler, J., Levitt, S.D.: Opinion: Modern high school math should be about data science—not Algebra 2. The Los Angeles Times (2019)

[10] Dinov, I.D., Sanchez, J., Christou, N.: Pedagogical utilization and assessment of the statistic online computational resource in introductory probability and statistics courses. Computers & Education **50**(1), 284–300 (2008)

[11] Keengwe, J., Kidd, T.T., *et al.*: Towards best practices in online learning and teaching in higher education. MERLOT Journal of Online Learning and Teaching **6**(2), 533–541 (2010)

[12] Gelman, A., Glickman, M.E.: Some class-participation demonstrations for introductory probability and statistics. Journal of Educational and Behavioral Statistics **25**(1), 84–100 (2000)

[13] Lazarski, C.: Developing the Theory of Hypothesis Testing: An Exploration. https://www.statisticsteacher.org/2021/03/29/developing-the-theory-of-hypothesis-testing-an-exploration/ Accessed 2023-09-10

[14] Ragasa, C.Y.: A comparison of computer-assisted instruction and the traditional method of teaching basic statistics. Journal of Statistics Education **16**(1) (2008)

[15] Miranda Freire, S.: Using shiny to illustrate the probability density function concept. Teaching Statistics **41**(1), 30–35 (2019)

[16] Kasprzak, P., Mitchell, L., Kravchuk, O., Timmins, A.: Six years of shiny in research-collaborative development of web tools in r. R Journal **12**(2) (2020)

[17] Von Borries, G.F., Castro Quadros, A.V.: Roc app: An application to understand roc curves. Brazilian Journal of Biometrics **40**(2) (2022)

[18] Chu, J.M.: Preparing for the AP exam: The dangers of teaching to the test. The History Teacher **33**(4), 511–520 (2000)

[19] Chance, B., Ben-Zvi, D., Garfield, J., Medina, E.: The role of technology in improving student learning of statistics. Technology Innovations in Statistics Education **1**(1) (2007)

[20] Moore, D.S.: New pedagogy and new content: The case of statistics. International statistical review **65**(2), 123–137 (1997)

[21] Mandinach, E.B., Gummer, E.S.: Every teacher should succeed with data literacy. Phi Delta Kappan **97**(8), 43–46 (2016)

[22] Vinué, G.: A web application for interactive visualization of european basketball data. big data **8**(1), 70–86 (2020)

[23] Lee, H.S., Harrison, T.: Trends in teaching advanced placement statistics: Results from a national survey. Journal of Statistics and Data Science Education **29**(3), 317–327 (2021)

[24] Bailey, N.G., McCulloch, A.W.: Describing critical statistical literacy habits of mind. The Journal of Mathematical Behavior **70**, 101063 (2023)

[25] LaMar, T., Boaler, J.: The importance and emergence of K-12 data science. Phi Delta Kappan **103**(1), 49–53 (2021)

[26] Gal, I., Ginsburg, L., Schau, C.: Monitoring attitudes and beliefs in statistics education. The assessment challenge in statistics education **12**, 37–51 (1997)

[27] MacGillivray, H.: Why Teaching Statistics is so important and challenging. Wiley Online Library (2018)

[28] Shatz, I.: The curse of knowledge when teaching statistics. Teaching Statistics **45**(1), 22–26 (2023)

[29] Fawcett, L.: Using interactive shiny applications to facilitate research-informed learning and teaching. Journal of Statistics Education **26**(1), 2–16 (2018)

[30] Stratton, C., Green, J.L., Hoegh, A.: Not just normal: Exploring power with shiny apps. Technology Innovations in Statistics Education **13**(1) (2021)

[31] Arnholt, A.T.: Using a shiny app to teach the concept of power. Teaching Statistics **41**(3), 79–84 (2019)