



Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujse21

Teaching Statistics and Data Analysis with R

Mary C. Tucker, Stacy T. Shaw, Ji Y. Son & James W. Stigler

To cite this article: Mary C. Tucker, Stacy T. Shaw, Ji Y. Son & James W. Stigler (2022): Teaching Statistics and Data Analysis with R, Journal of Statistics and Data Science Education, DOI: 10.1080/26939169.2022.2089410

To link to this article: https://doi.org/10.1080/26939169.2022.2089410

б

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 19 Jul 2022.

ĺ	
1	ت

Submit your article to this journal 🖸

Article views: 4111



View related articles

Teaching Statistics and Data Analysis with R

Mary C. Tucker^{*}^a, Stacy T. Shaw^b, Ji Y. Son^c, and James W. Stigler^a

^aUniversity of California, Los Angeles, Los Angeles, CA; ^bWorcester Polytechnic Institutute, Worcester, MA; ^cCalifornia State University, Los Angeles, Los Angeles, CA

ABSTRACT

We developed an interactive online textbook that interleaves R programming activities with text as a way to facilitate students' understanding of statistical ideas while minimizing the cognitive and emotional burden of learning programming. In this exploratory study, we characterize the attitudes and experiences of 672 undergraduate students as they used our online textbook as part of a 10-week introductory course in statistics. Students expressed negative attitudes and concerns related to R at the beginning of the course, but most developed more positive attitudes after engaging with course materials, regardless of demographic characteristics or prior programming experience. Analysis of a subgroup of students revealed that change in attitudes toward R may be linked to students' patterns of engagement over time and students' perceptions of the learning environment.

1. Introduction

Statistics has long been taught in conjunction with data analysis. Yet, an enduring question is how best to integrate data analysis activities into the statistics curriculum. A dominant approach in U.S. universities is the lab-lecture model, in which students learn about a statistical concept such as the *t*-test during lecture and then apply that concept later in the computer lab. Lab activities often rely on prepackaged software tools, such as SPSS, and focus on producing statistical output, often with procedural instructions (ASA GAISE College working group 2016; McCulloch 2017). Although supplementing lecture with lab-based activities can provide students with meaningful learning opportunities, there are reasons to consider alternative models for integrating data analysis in the statistics curricula. One is that students who experience a separation between statistics concepts learned in lecture and the data analysis they perform in lab may end up thinking of statistics as a set of isolated skills and procedures rather than as a system of interconnected ideas (Reid and Petocz 2002). Traditional statistics instruction that focuses more on algebraic formulas and the procedures for calculating statistics and less on their interpretation perpetuates this problem (Moore 1992, 1997). Because the concepts, formulas, and data analysis procedures are not intertwined, the knowledge students are left with at the end of the course does not easily transfer to new situations (Fries et al. 2020; Son et al. 2021).

Lack of transfer is not the only reason to rethink the integration of statistical concepts and data analysis. Modern statistics has become an increasingly computational science (Nolan and Temple Lang 2010; Cobb 2015). Many of the most important concepts in modern statistics can only be understood through computational procedures for data analysis. Techniques such as simulation, bootstrapping, and permutation require students to imagine and then implement, a data generating process that can be used to generate reference distributions that can, in turn, be used to interpret actual distributions of data. These techniques have grown in popularity because now, for the first time in human history, the computational power to implement them is affordable.

Software for data analysis is also undergoing a transformation. The earliest packages—many of which are still in use today (such as SPSS)—made it easier and more convenient to do the calculations previously done by hand and then by electronic calculators. In lectures, students would learn the formulas, then in the lab they would apply the formulas to produce and interpret statistics. But software has changed as new techniques have emerged (ASA GAISE College working group 2016). It is not enough to get the answer to a question about a bootstrapped standard error. Students must be able to understand and reason about the computational process (not just the formula) that produced the standard error (Moore 1992; Cobb 2007a, 2007b; Nolan and Temple Lang 2010).

Thus, because the field of statistics has advanced in its practices, and because technology has changed, it is a good time to create, test, and improve new pedagogies for teaching statistics and data analysis. Many have taken up this challenge (e.g., Cobb et al. 1997; Finzer 2001; Konold and Miller 2004; Nolan and Temple Lang 2007; Rossman and Chance 2014, see also the *JSDSE* 2021 special issue on computing; Gould, Wong, and Ryan 2017; Pruim, Kaplan, and Horton 2017; Lock et al. 2021).

CONTACT Mary C. Tucker and *mctuck@gmail.com* Duriversity of California, Los Angeles, Los Angeles, CA.

© 2022 The Author(s). Published with license by Taylor and Francis Group, LLC.

KEYWORDS Data science education;

Higher education; R; Statistics education research



Check for updates

∂ OPEN ACCESS

Supplementary materials for this article are available online. Please go to www.tandfonline.com/ujse.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The moral rights of the named author(s) have been asserted.

Our interest is in how to introduce statistical programming languages (in our case using R) into the teaching of data analysis and in how to teach modern methods of data analysis as a means of helping students create a deep and transferable understanding of the core concepts that underlie statistics. We also want to connect our students to the fast-growing world of data science, and increase the numbers of students, especially those who are traditionally disadvantaged, who can envision the field of data science as a possible career pathway.

2. Teaching Data Analysis with R

R is a statistical computing environment for data analysis that has been widely adopted by researchers and industry professionals in STEM, the social sciences, and the humanities (R Core Team 2019). Though many data analysis tools are available, R offers several advantages. First, R is open-source and free, which reduces barriers to access. Second, R is flexible and powerful; it can be used for both simulation and data analysis. Unlike analyzing data using point-and click interfaces, analyzing data in R involves writing code, which can make statistical thinking and data analysis processes more visible and reproducible; thus, the use of R may offer students an additional representational tool to build their understanding. Third, R is used by a growing number of people across different fields and thus is more generalizable than other statistical software programs.

Although more and more educators are using R to teach data analysis (e.g., Baumer et al. 2014; Nolan and Temple Lang 2015; Mascaró, Sacristán, and Rufino 2016), many barriers remain. One barrier is the lack of integration between text-based instructional materials and R coding environments. For example, a growing number of textbooks supplement statistics content with R, but most require students to switch between the text and some other platform (such as RStudio installed on their own computer) to try out any R code (i.e., Dalgaard 2008; Field, Miles, and Field 2012; Nolan and Temple Lang 2015; Navarro 2020). For novice students, the initial difficulty of setting up R may be a barrier to the R component of the class (Çetinkaya-Rundel and Rundel 2018), and even though R is interspersed with the statistics concepts in the textbook, it might feel like something separate they have to do after they read the book.

In addition to challenges integrating R with course materials, many instructors report (both formally and informally) that students "hate programming" and fear students' initial feelings toward R will turn them off of statistics (Ward 2013; Rode and Ringel 2019). Because R is syntax-based (rather than having a point-and-click interface), there is a general perception that it may be harder to learn for some students—especially those who come in with no prior programming experience (Biehler 1997). For many instructors, just the idea of teaching R may be challenging, particularly when they themselves may have little or no experience with R. All of these factors lead to a persistent worry that the cognitive and emotional burden of learning R will make it more difficult for students to focus their attention on understanding statistical concepts (rOpenSci 2018).

We share some of these concerns. However, it is possible that these presumed negative effects of learning R might be mitigated by the way in which R is introduced and taught. If R is taught as a set of procedures to be learned separately from the main course content (e.g., concepts and algebraic formulas), then it will be hard for students to view it as a tool to support statistical thinking and understanding. Furthermore, students may not be able to see its relevance and value in their chosen field of study or to their lives more broadly. It is also possible that teachers who incorporate R attempt to teach too much programming, including concepts (e.g., for loops, pointers) that are not necessary for students taking an introductory statistics class nor directly aligned with the goal of teaching statistics for understanding.

2.1. Cognitive and Motivational Considerations

Like many things in education, we believe the effects of integrating R into the introductory statistics course will depend on how the integration is carried out. If students perceive R as something extra that must be learned, over and above the already challenging subject matter of statistics, they may feel that learning R is too costly and stressful. But if students see R as a tool for understanding and doing statistics, they may see R as valuable and relevant, and thus develop more positive attitudes toward R and toward statistics.

Our perspective draws on theory and research from cognitive psychology. According to Cognitive Load Theory (Sweller 1988, 1999, 2018), inherently complex learning materials (known as high intrinsic cognitive load) or materials that introduce too much information simultaneously (high extraneous cognitive load) can deplete students' cognitive resources and impede meaningful learning (Sweller 1988, 1999, 2018). Conversely, materials that stimulate mental activity in ways that are challenging, yet relevant (high germane cognitive load), can promote learning and transfer (Paas, Renkl, and Sweller 2004). Our perspective also draws on the expectancy-value-cost theory of achievement motivation (Barron and Hulleman 2015), which provides a useful framework for understanding how students' perceptions of the learning environment shape their choices, persistence, and motivation. When students perceive instructional activities to be valuable and relevant, their motivation and interest increases (Hulleman and Harackiewicz 2009; Hulleman et al. 2017). When students perceive instructional activities to be too costly (in time, resources, or stress), motivation and interest decreases (Barron and Hulleman 2015).

2.2. Setting the Stage

With these considerations in mind, in this project we aimed to engineer a new way of integrating the teaching of R into the teaching of statistics so as to promote students' statistical thinking and transfer. Our project builds directly on three developments: the MOSAIC project (making R syntax more transparent to novices), web 2.0 technology (making R programming environments more available to novices), and the emergence of the field of learning engineering (making student data from learning R more available to researchers).

The MOSAIC Project The first of these is the MOSAIC project (Pruim, Kaplan, and Horton 2017). Though R was not designed to support students' development of deep understanding of statistics concepts, the MOSAIC team has

written a number of R packages designed to do just that: to simplify the syntax of functions so that students can more easily understand the underlying structure. MOSAIC R functions not only enable students to explore and analyze data, but they also provide tools for representing and understanding complex statistical ideas, such as randomness.

Widespread Availability of Web2.0 Technology A second development is the ability to integrate R programming environments in cloud-hosted online instructional materials. Even with tools like MOSAIC, students still, until recently, had to either download and install an R integrated programming environment (IDE) such as R Studio on their own computers to run R or use computer labs with R already installed. This has necessitated the pedagogical separation between statistics and data analysis. Today, next generation web technologies make it possible to interleave explanatory text with R programming examples, right on the same web page. The barriers to full integration of statistics and data analysis having now been removed, paves the way for researchers to understand how to exploit the new platforms to benefit student learning.

Availability of Student Response Data Finally, the emergence of the interdisciplinary fields of learning engineering (Lieberman 2018; Thille 2018; Dede, Richards, and Saxberg 2019) has fostered new ways of capturing and analyzing students' interactions, responses, and learning on a moment-by-moment level as students work through course materials. Certainly, individual instructors have devised creative ways of integrating statistics and data analysis, but we have scant data available to assess the actual impact of these different pedagogical innovations on students' motivation and learning, especially in real time. This leads directly to our current project.

3. Project Overview

This study is part of a larger project in which we are building a technology platform and a set of working routines for developing, implementing, testing, and continuously improving online learning materials based on students' response data. The first prototype of this approach is a continuously improving online textbook for teaching introductory statistics, CourseKata Statistics and Data Science (available for preview at *www.coursekata.org*; see Stigler et al. 2020). The online book consists of 12 chapters and 144 pages, each of which interleave text, graphics, R coding exercises, and formative assessment questions (more than 1200 in all). We briefly describe the book in the next section.

3.1. Course Design

A major feature of the online book is the complete integration of data analysis using R into the introductory statistics curriculum. Students using our online textbook do not have to download, install or configure any software, or even switch from one window on a screen to another, in order to complete the pages of the book. We hypothesize that by interleaving R data analysis activities with the introduction of statistical concepts, and by building on the MOSAIC project (Pruim, Kaplan, and Horton 2017), we can make script-based programming accessible to all learners, regardless of background or prior programming experience, and

support deep learning of statistics without increasing cognitive load or eliciting negative emotions in students.

The pedagogical design of the book is based on the practicing connections hypothesis, which is grounded in theory and research from psychology and the learning sciences (Fries et al. 2020). The practicing connections hypothesis argues that in order to develop coherent and transferable knowledge in complex domains, like statistics, students need repeated opportunities to engage in deliberate practice making connections among key concepts, representations, and situations of the domain over time. Whereas most traditional statistics texts use algebraic notation to represent statistics concepts, our textbook uses R as a key representation, and integrates R according to the following principles:

- R exercises should help students to represent ideas and make connections between concepts, not just be used to compute answers. In our textbook, R exercises focus on using modeling, simulation, and visualization to understand abstract concepts like randomness and the data generating process. The R code should be used to help students learn the meaning of other important statistical representations (e.g., the algebraic notation of the General Linear Model, specific vocabulary).
- 2. R exercises should be interleaved throughout the text and embedded in the practice of data analysis. The textbook should interleave R exercises with other types of exercises (e.g., writing, multiple choice, categorization tasks) to provide opportunities for productive struggle, feedback, and deliberate practice. In other words, R exercises should be viewed as one part of an active opportunity for students to make connections.
- 3. Practice with R should start simple and become increasingly sophisticated and complex over time, providing students a gentle on-ramp to the adoption of R as a tool for doing and thinking about statistics and data analysis.
- 4. Expectations about what authentic users of R do should be taught in an authentic manner. Novices might expect to learn how to flawlessly generate code that works on the first try. Instead, a growth mindset toward R should be cultivated where students are taught to expect errors, forgetting, and frustration as part of the process for all data scientists. Using cheat sheets, asking for help, and searching online are not hacks just for beginners but something that even professional R users do.
- 5. Students should have opportunities to apply their developing skills to answer meaningful questions using real data so that they can experience, first-hand, the value of computation for understanding and doing statistics and the relevance of statistical thinking in their everyday lives.

In sum, the textbook should reduce the R learning curve and increase motivation wherever possible; focus on teaching R as a tool for understanding and doing statistics, rather than programming as an end in itself; draw on resources created by other statistics educators (e.g., MOSAIC) to reduce the number of functions students need to learn and to make those functions more understandable; scaffold R activities, and provide hints and feedback; cultivate a growth mindset orientation toward learning to program; and make it easy for students to get help, so they don't focus on memorization (e.g., R cheat sheet, glossary, help desk).

3.2. Study Aims

The goal of the current research was to present an initial "proof of concept." We examined three large classes of students' attitudes and experiences as they used the online textbook throughout a 10-week introductory course in statistics. Prior work has found that students are initially anxious viewing R output, but their anxiety abates over time (Rode and Ringel 2019). However, this finding has yet to be replicated in a larger sample size and within a course that requires students to write code, generate plots, and compare models in R (rather than look at R output). If our approach to integrating R as a tool for understanding statistics is effective, we expect that at the end of the course we will find that students not only showed gains in learning and transfer, but they also developed positive attitudes toward R (even for students who are initially hesitant about learning R programming).

We had three specific aims in this exploratory study. The primary aim was to determine if it is possible to use R in an introductory course with students who have limited mathematical and computer programming backgrounds, without negatively impacting their learning or motivation. Past research suggests that students hold negative beliefs and misconceptions about computer programming, which can influence their motivation and behavior during learning (Siek et al. 2006; Scott and Ghinea 2014; Cheryan, Master, and Meltzoff 2015; Google and Gallup 2015; Tek, Benli, and Deveci 2018). It is possible that even if R is a valuable skill for students to learn, that students would, on average, leave with a negative disposition to programming. But if the design of our course is successful, students may end the course feeling positive about R and perceiving the course to be valuable.

A second aim was to understand individual differences in how students experienced the course and in how differences in these experiences might produce different motivational and learning outcomes. In particular, we wanted to make sure that integrating R did not contribute to existing disparities by helping some groups of students and hindering others, such as female students, and students with race/ethnicities that are traditionally underrepresented in STEM fields (Riegle-Crumb and King 2010; Gallup 2016; Charles and Theibaud 2018).

A third aim was to identify potential challenges and opportunities for intervention to improve students' experiences learning statistics and data analysis with R. Even if most students have a positive experience using R in our course, it is likely that some students may struggle. Thus, our final aim is to identify students who do not develop positive attitudes toward R and identify how they may differ from students who do end up with positive attitudes. By better understanding the experiences of students who struggle, we hope to design improvements that might reduce their numbers in future iterations of our course.

4. Method

4.1. Study Context

The course, Introduction to Psychological Statistics, is an undergraduate course offered by the psychology department at a large public research university located in a major city in the western United States. The course provides a basic introduction to statistics and data analysis with an emphasis on its application to research in psychology. The goals are for students to understand basic concepts that underlie descriptive and inferential statistics and use them to make sense of new situations, to be prepared cognitively and emotionally to learn more advanced techniques in the future, and to be able to do basic data analysis using the R statistical programming language. Students majoring in psychology must complete the course with a grade of C- or better in order to remain in their degree program.

The three sections of the course described in this study are blended courses. In addition to standard lecture and office hours, each class incorporated online components. The online components included the interactive online textbook (CourseKata Statistics and Data Science), a question/answer forum, and links to online resources (e.g., a help desk and reference materials). Much of the course content was conveyed through the online interactive textbook, which included over 1200 embedded formative assessments, embedded R programming exercises, and practice quizzes. The first four chapters in the textbook provide students with a scaffolded initial introduction to R, including a description of the interactive programming environment as well as practice with basic programming concepts such as functions, variables, and data types. Students were not provided with supplemental instruction that introduced R outside of the content in the online textbook.

Lectures occurred twice weekly, were led by instructors, and focused on deepening understanding of concepts, connecting concepts, discussing examples, and answering questions. Students participated during lecture by answering questions posed by the instructor via an audience polling tool. Other face-to-face components included a separate, weekly, large-group discussion section which was used to administer quizzes and answer questions about course content. Quizzes were administered once every other week. On weeks when no quizzes were scheduled, discussion sections involved short review presentations and question and answer sessions led by graduate Teaching Associates. Students' final course grades comprised performance on quizzes (administered once every two weeks), completion of homework assignments (reading and completing all exercises in the assigned online textbook chapters), and performance on cumulative exam(s).

4.2. Participants

Data were collected from 672 undergraduate students who used our interactive online textbook as part of the course, Introduction to Psychological Statistics during the 2019–2020 academic year. Students were enrolled in one of three sections of the course at the University of California, Los Angeles. Each section was taught by a different instructor, but all used the same blended course format and interactive online textbook: CourseKata Statistics and Data Science. The data were collected as part of an ongoing project, which was approved by the Institutional Review Board at the University of California, Los Angeles (IRB No: 20-001033).

We initially obtained data from 789 students: 290 who took the course in Fall 2019 (Class A), 244 who took one section of the course in Winter 2020 (Class B), and 255 students who took

Table 1. Socio-demographic characteristics of participants in the sample.

Characteristic	Class A Fall 2019 Instructor A		Class B Winter 2020 Instructor B		Class C Winter 2020 Instructor C		Total Sample	
	n	%	п	%	n	%	п	%
Gender								
Female	183	75%	163	74%	146	70%	492	73%
Male	47	19%	54	25%	55	26	156	23%
Nonbinary/other	0	0%	1	<1%	5	1%	6	<1%
Did not disclose	14	8%	1	<1%	3	<1%	18	3%
Race/ethnicity								
White	64	26%	47	22%	73	35%	184	27%
Asian	71	29%	97	44%	83	40%	251	37%
Latino	52	21%	40	18%	33	16%	125	19%
Multi-racial/other	18	7%	11	5%	8	4%	37	6%
M East/N. African	15	6%	13	6%	3	1%	31	5%
Black/African Am.	10	4%	9	4%	6	3%	25	4%
Did not disclose	14	6%	2	<1%	3	1%	19	3%
Programming experience								
None	153	63%	98	45%	89	43%	340	51%
Some	91	37%	120	55%	120	57%	331	49%
GPA								
3.50-4.00	171	70%	123	56%	115	55%	409	61%
Parental education								
College educated mother	170	70%	148	68%	154	74%	472	70%
Total	244		219		209		672	

another section of the course in Winter 2020 (Class C). However, our analyses focus on only those students who (a) remained enrolled for the duration of the course, (b) completed at least one question on both the pre- and post- course surveys, and (c) agreed to share their data with the research team. Students who did not meet these criteria (n = 117) were excluded. Of these excluded students, 27 (23%) did not agree to share their data, 30 (26%) created an account but did not respond to any questions in the online textbook, and 60 (51%) did not complete either the pre- or post-course survey.

The resulting analytic sample (n = 672) was 73% female and self-identified as 37% East/Southeast Asian, 19%, Latino, 27% White, 5% Middle Eastern and North African, 4% Black/African American, and 5% more than one race/ethnicity/other. Table 1 contains sociodemographic information for students in the sample.

Due to small sample sizes, participants who did not disclose gender (n = 18) or selected nonbinary or other gender (n = 6) were not included in formal hypothesis tests comparing gender differences. Similarly, Black/African American students (n = 25), Middle Eastern and North African students (n = 31), and students who self-identified with more than one race/ethnicity/other (n = 37) were not included in formal hypothesis tests of racial/ethnic differences due to low frequencies. Instead, we present the raw data for these demographic subgroups and comment on how those data compare to patterns from larger demographic groups (e.g., male, female and Asian, Latino, White students).

4.3. Procedure

Students took a survey at the beginning and end of the course. The surveys included questions about students' experiences, expectations, attitudes, and motivational beliefs. A subset of the motivational belief questions was presented again on a survey embedded two-thirds of the way through the course content (chapter 8 in a 12-chapter book). During the course, students' interactions and performance on embedded formative assessments were collected within the online textbook. An overview of the study design is shown in Figure 1.

4.4. Measures

Measures were generated by students' responses to survey questions administered at three time points before and during the course and log-data from students' interactions within the online textbook. Questionnaires assessed students' attitudes and beliefs about learning and students' background and demographic characteristics. The measures are described below.

Attitudes toward Learning R Students rated their attitudes toward using R on a single item—"In this course, you will use/have used R (a programming language) to analyze data; how do you feel about this?" to which they responded using a 5point Likert scale from strongly negative to strongly positive. R attitude items were administered at the beginning and end of the course (t_1 and t_3). In addition to the attitude measure, students completed two additional R items at t_3 . Students rated their confidence in their ability to use R (e.g., "I am confident I could use R to analyze a new dataset") using a 5-point scale from not at all confident to extremely confident. They also rated how important R was for their learning using a 5-point scale from not at all important to important.

Concerns about the Course Following recommendations by Gal and Ginsburg (1994), we paired Likert items with an openended question in which students self-described their concerns at the beginning of the course. Students were asked to provide a short response to the following prompt, suggested by Gal and Ginsburg (1994): "When I think about this course, I'm Interactions within the online textbook and performance on end-ofchapter review activities recorded for chapters 1-12



3 Surveys **Pre** (t_1 , before chapter 1), **Mid** (t_2 , after chapter 8), **Post** (t_3 , after chapter 12),

Figure 1. Study design.

concerned that...(write "not at all" if you so feel)." Responses that mentioned "R," "coding," "computer programming," or "programming" were coded to create a dichotomous indicator of whether or not students mentioned computer programming as a concern (1 = mentioned computer programming as a concern, 0 = did not mention computer programming as a concern).

Expectancy, Value, and Cost We measured students' motivational beliefs using items from the expectancy-value-cost scale (Kosovich et al. 2015). Students rated their expectations for success in the course (e.g., "I know I can learn the material in this course," "I believe I can be successful in this course") using a 5-point scale from strongly disagree to strongly agree. Expectancies were measured at two timepoints: t_1 and t_2 . Perceived course value was measured using two items ("The content of this course is important for me," and "What I learn in this course will be useful in the future"). Students rated agreement using a 5-point scale from strongly disagree to strongly agree. Value items were administered at three timepoints: t_1 , t_2 , and t_3 . Perceived course cost was measured using three items—"I'm unable to put in the time needed to do well in this course," "I have to give up too much to do well in this course," and "This course is too stressful for me." Students rated their agreement using a 5-point scale from strongly disagree to strongly agree. Cost items were administered at a single timepoint, t_2 . We created aggregate measures of expectancy, value, and cost by averaging the scores for each construct at each time point.

Beliefs about Memorization On the pre- and post-course surveys (t_1 and t_3), students rated how much they felt the course required (or would require) memorization (e.g., "I expect that this course will require a lot of memorization" or "This course required a lot of memorization") using a 5-point scale from strongly disagree to strongly agree.

Engagement Engagement with course materials was measured using data from students' interactions within the online textbook. Measures were calculated on a chapter-by-chapter level for each student. We looked at the following engagement measures: *chapter review score*, calculated as the proportion of correct responses on the embedded review activity at the end of each chapter, *word count*, calculated as the number of words on a summary students were required to write for each chapter, *R performance*, calculated as the proportion of R activities that were correct on the first submission, *R attempts to correct*, for R activities that students answered incorrectly on the first attempt, the average number of attempts it took students to get the correct answer.

Sociodemographic Characteristics Several socio-demographic variables that could account for differences in students' attitudes and learning outcomes were included in the analyses. These variables included self-reported *gender* and *race/ethnicity*, *GPA*, and *parental education*. Students reported their cumulative grade point average by selecting from one of the following five categories: 3.50-4.00, 3.00-3.49, 2.50-2.99, 2.00-2.49, and less than 2.00. Parental education was measured by asking students to report the highest level of education attained by their mother. Responses were dummy coded to indicate the mother's education level (1 = college educated, 0 = not college educated).

Computer Programming Background Computer programming background was measured on the pre-survey (t_1) . Students selected one of the following options to indicate prior experience with computer programming: yes, I have taken a computer programming class; yes, I have used computer programming in a nonprogramming class; not a formal course, but I have tried programming on my own; no. The first three categories were combined to create a dichotomous indicator of previous programming experience (0 = no prior computer programming experience, 1 = some computer programming experience).

4.5. Statistical Analysis

Descriptive and inferential analyses were performed to characterize students' attitudes and beliefs at the beginning of the course (t_1), after chapter 8 (t_2), and at the end of the course (t_3) and to investigate changes in students' attitudes from beginning to end of the course ($t_3 - t_1$). Because R attitude, R confidence, R importance, and perceived memorization were each measured using a single, ordered item, they were treated as ordinal data and analyzed with nonparametric tests and ordinal regression using the ordinal package (Christensen 2018) in R (R Core Team 2019). Expectancy, value, and cost and engagement outcome variables were analyzed using one-way and repeated measures ANOVAs. We conducted all of our analysis in R version 3.6.2 (R Core Team 2019).

5. Results

5.1. Student Characteristics

About half (49.3%) of the students in our sample had been exposed to computer programming before taking our course. Of those students who had previously been exposed to computer programming (n = 331), 59.5% had taken a formal computer programming course, 32.3% had used programming in a non-computer programming course, and 8.2% had tried programming on their own. Consistent with underrepresentation of racial/ethnic minority groups in STEM in the United States, prior programming experience was associated with race/ethnicity, $X^2(3) = 22.82$, p < 0.001. Of students who self-identified as Latino/Hispanic, about 34% had previously been exposed to computer programming compared to 50% of students who self-identified as White and 61% of students who identified as Asian. A similar pattern of results was observed among students who self-identified as Black/African American (about 36% of whom had previous programming experience) and students who self-identified as Middle Eastern/North African (about 29% of whom reported prior programming experience). About half (49%) of students who self-identify as multi-racial or other race/ethnic background reported some computer programming experience.

Previous programming experience also varied as a function of parental education, $X^2(1) = 16.77$, p < 0.001. Students with college educated mothers (53.7%) were more likely to have prior programming experience than students with noncollege educated mothers (36.4%). Previous programming experience did not vary as a function of gender, $X^2(1) = 0.32$, p = 0.6.

5.2. Attitudes and Concerns at the Beginning of the Course

Prior to starting the course, the majority of students (73%) expressed negative or neutral attitudes toward learning R (e.g., "strongly negative," "negative," or "neither positive nor negative"). Initial R attitude ratings were distributed symmetrically, with a median of 3 (neither positive nor negative), and a range of 4 (see Figure 3). The most common response was "neither positive nor negative" (n = 239), which accounted for 36% of all student ratings. The remaining students showed a tendency toward more negative attitudes. For instance, 9% of students (n = 59) felt "very negative" while only 6% of students (n = 41) felt "very positive." Similarly, 30% of students felt "negative" (n = 195) while 21% (n = 138) of students felt "positive."

Table 2 shows the distribution of initial R attitude ratings for all students and broken down by programming experience, gender, and race/ethnicity. The table shows that "very negative" attitude ratings were more common among female students (10%), Latino students (15%), and students no previous programming experience (12%), whereas "very positive" attitude ratings were most common among White students (8%), male students (12%), and students with some previous programming experience (8%).

Mann-Whitney U tests and effect sizes, r, were conducted to detect any differences in initial R attitude ratings by programming experience and gender. Results showed an effect of programming experience (Z = 5.0, p < 0.0001, r = 0.12395% CI[0.049, 0.198]) and gender (Z = 3.7, p = 0.0002,r = 0.141, 95% CI[0.065, 0.217]), such that students with some programming experience (mean rank = 372.7) were more likely than students with no programming experience (mean rank = 300.3) to rate R positively. Similarly, male students (mean rank = 384.6) were more likely than female students (mean rank = 322.2) to rate R positively. A Kruskal-Wallis test further revealed that initial R attitude ratings varied according to students' race/ethnicity, H(2) = 9.5, p = 0.009. A post hoc Dunn test with Bonferroni adjustment revealed statistically discernible differences between White students and Latino students (Z = 2.9, p = 0.011) such that White students (mean rank = 307.8) were more likely than Latino students (mean rank = 255.3) to rank R positively. There were no other statistically discernible differences in initial R attitudes by race/ethnicity at the beginning of the course.

When asked to describe their concerns about the course (before starting) on an open-ended question, approximately one-third of students (32%) mentioned R, coding, or computer programming. Chi-square tests of independence revealed that females were more likely than males (40% of females vs. 22% of males) to mention R as a concern, $X^2(1) = 10.9$, p < 0.001, but there were no detectable differences in the proportion of students mentioning R as a concern by race/ethnicity, $X^2(2) = 0.8$, p = 0.7. The proportions of students who mentioned R

Table 2. Distribution of initial R attitude ratings overall and broken down by programming experience, gender, and race/ethnicity.

	Pre-course R Attitude Ratings (t ₁)					
	1	2	3	4	5	
	Strongly negative	Somewhat negative	Neither positive nor negative	Somewhat positive	Strongly positive	
All students	59(9%)	195	239	138	41(6%)	672
Programming experience						
None	4(12%)	113	120	53	13(4%)	340
Some	18(5%)	82	119	84	28(9%)	331
Gender						
Female	48(10%)	149	181	93	21(4%)	492
Male	9(6%)	38	49	41	19(12%)	156
Nonbinary/other	0(0%)	4	2	0	0(0%)	6
Did not disclose	2(11%)	4	7	4	1(1%)	18
Race/ethnicity						
Asian	22(9%)	79	88	48	14(6%)	251
Latino	19(15%)	32	50	19	5(4%)	125
White	11(6%)	51	55	52	15(8%)	184
Multi-racial/other	1(3%)	13	16	2	3(9%)	35
M. East/N. African	2(7%)	8	12	7	2(7%)	31
Black/African Am.	2(8%)	7	9	6	1(4%)	25
Did not disclose	2(11%)	5	7	4	1(5%)	19

as a concern among Black/African American students (25%), Middle Eastern and North African students (32%), students who identified as multi-racial or other race/ethnicity (39%), and students who did not disclose their race/ethnic identity (26%) were similar to the proportions of students who mentioned R as a concern among White (34%), Latino (30%), and Asian (33%) students. Similarly, there was no discernable difference between the proportion of students with programming experience who mentioned R as a concern (33%) and the proportion of students without programming experience who mentioned R as a concern (32%) to mention R as a concern, $X^2(1) = 0.10$, p = 0.8.

Though students with and without prior programming experience expressed concern about learning R at similar rates, the content of their responses differed subtly. For example, students with some programming experience seemed concerned because of their previous experience. One student wrote:

"I worry that I'll enter and leave the course with absolutely no overarching understanding of the applicability of coding, because this has been the case in every class I've taken that required coding."

Another simply reported: "When I took a computer programming course at community college, I struggled a lot."

On the other hand, students with no programming experience tended to express a generalized fear of coding and their ability to learn it: "I won't enjoy computer coding because I've never done it before," one student wrote. Other students simply reported that, "Programming is hard," and "I will be bad at programming."

5.3. Changes in Attitudes toward R

Figure 2 shows the distribution of students' R attitude ratings at the beginning and end of the course. Overall, students rated R more positively at the end of the course than at the beginning. Whereas 74% of students rated their attitude as either negative or neutral at the beginning of the course, this percentage was reduced to only 34% by the end of the course. The percent of students who felt "strongly positive," about using R, likewise, went from 6% to 19%; more than triple the number of students



Figure 2. Distribution of R attitude ratings at the beginning and end of the course. *Note:* Vertical lines represent the median rating score. R attitude was measured using a 5-point scale from 1 = very negative to 5 = very positive.

from the beginning to end of the course. A paired signed-rank test detected a difference between the median R attitude rating at the end of the course (Median = 4) and the median R attitude rating at the beginning of the course (Median = 3), Z = 13.9, p < 0.0001.

Table 3 shows the breakdown of R attitude ratings at the end of the course by R attitude ratings at the beginning of the course. About 80% (n = 202) of the 254 students who initially held negative attitudes toward learning R developed more positive attitudes by the end of the course. Further, over half (57%) of those students who felt "strongly negative" or "somewhat negative" at the beginning of the course felt "somewhat positive" or "strongly positive" about R at the end of the course.

 Table 3. Frequency of R attitude ratings at the beginning and at the end of the course.

Pre-course R attitude ratings (t_1)	Post-course R attitude ratings (t_3)					n
	1	2	3	4	5	
1 = Strongly negative	7	9 ^a	14 ^a	25 ^b	4 ^b	59
2 = Somewhat negative	14	31	34 ^a	88 ^b	28 ^b	195
3 = Neither positive nor negative	9	25	53	120	32	239
4 = Somewhat positive	4	4	16	72	42	138
5 = Strongly positive	1	3	3	15	19	41

Note: ^{*a*}Students who started out negative and increased their attitudes toward R. ^{*b*}Students who started out negative and ended the course feeling positive toward R.

To investigate differences in R attitudes across the two timepoints $(t_3 - t_1)$ and to see if change in R attitude ratings varied based on individual student characteristics, we used repeated ordinal regression analysis. We fit separate models to estimate the effect of time on R attitude ratings and to estimate the main and interactive effects of time and prior programming experience, gender, race/ethnicity, and concern about learning R at the beginning of the course. In each ordinal regression model, R attitude rating was an ordinal outcome variable with time (t_1 and t_3) and the student characteristic (group) as predictors, with students as the random factor.

There was a discernible difference between the likelihood of rating R positively at the beginning of the course compared to the end of the course (likelihood ratio $X^2(1) = 224, p < 224$ 0.0001). Across all students, the odds of rating R more positively were higher at the end of the course (t_3) than at the beginning of the course (t_1) . Though, overall, students changed their attitudes toward R to be more positive, some groups of students increased their attitudes more than others. Specifically, there was a statistically discernible difference in the pattern of R ratings for female compared to male students (likelihood ratio $X^{2}(1) = 4.44, p = 0.0001$ such that the odds of female students increasing their R attitude ratings from beginning to end of the course were higher than the odds of males increasing their R attitude ratings from beginning to the end of the course. There was also a discernible difference in the pattern of R ratings for students who did and those who did not mention R as a concern (likelihood ratio $X^2(1) = 10.0, p = 0.002$) such that students who mentioned R as a concern were more likely to increase their R attitude ratings to be positive from beginning to end of the course than students who did not mention R as a concern.

5.4. Confidence and Importance Ratings

In addition to R attitude ratings, we looked at students' ratings of how confident they felt using R to analyze a new dataset and their rating of how important they felt R was for their learning of course material. The frequencies for R confidence ratings and R importance ratings are shown in Tables 4 and 5, respectively.

As Table 4 shows, students reported feeling confident in their ability to use R at the end of the course. However, some groups of students reported higher confidence than others. Mann-Whitney U tests revealed that R confidence ratings differed by prior programming experience (Z = 4.6, p < 0.0001) and gender (Z = 2.8, p = 0.005). Students

with some prior programming experience (mean rank = 367.3) rated their confidence higher than students without any prior programming experience (mean rank = 302.3) and male students (mean rank = 359.1) rated their confidence higher than female students (mean rank = 359.1). A Kruskal-Wallis test further revealed a detectable difference in the average end-of-course R attitude ratings by students' race/ethnicity, H(2) = 14.2, p = 0.0008. A post hoc Dunn test with Bonferroni adjustment revealed that White students reported higher confidence than Asian students (Z = 2.5, p = 0.04) and Latinx students (Z = 3.7, p = 0.0007). Although Black/African American students (n = 25) and Middle Eastern/North African students (n = 31) were not included in this analysis due to small sample sizes, their responses are similar to the responses of Latino and Asian students.

In terms of R importance, overall, students felt that R was important for their learning in this course. The most common responses were "very important" (37%) and "extremely important" (36%), and this pattern was similar within each subgroup of students. Mann-Whitney U tests showed that importance ratings did not differ by previous programming experience (Z = 1.5, p = 0.1) or gender (Z = 1.4, p = 0.2). Similarly, a Kruskal-Wallis test revealed that R importance did not vary by race/ethnicity for Asian, Latino, and White students (H(2) = 4.2, p = 0.1). As with R confidence ratings, Black/African American students (36% of whom rated R as extremely important) and Middle Eastern/North African students (42% of whom rated R as extremely important) ratings of R importance were similar to the responses of Latino (39% of whom rated R as extremely important), Asian (40% of whom rated R as extremely important), and White students (30% of whom rated R as extremely important).

5.5. Students Who Expressed Negative Attitudes on the Presurvey

Though the majority (80%) of students who started off feeling negatively toward R shifted their feelings to be more positive at the end of the course, there were students whose attitudes did not improve. To better understand the students' experiences that might have contributed to these differences, we identified two subgroups of students:

- The *negative-to-positive* students were 297 students who rated their attitude toward learning R as "very negative," "somewhat negative," or "neither positive nor negative" at the beginning of the course (t₁) and at the end of the course (t₃) rated their attitude as "somewhat positive" or "very positive."
- The *negative-to-negative* students were 196 students who also rated their attitudes as "very negative," "somewhat negative," or "neither positive nor negative" at the beginning of the course (t₁) and again rated their attitudes as "very negative," "somewhat negative," or "neither positive nor negative" at the end of the course (t₃).

We compared these two subgroups of students, which we will refer to as the *neg-pos* and *neg-neg*, respectively, in an effort to identify how to improve future students' experiences with learning R. We compared socio-demographic characteristics as well as two categories of students' experiences: learning-related

Table 4. Distribution of R confidence ratings at the end of the course overall and broken down by programming experience, gender, and race/ethnicity.

	Post-course R confidence ratings (t_3)					
	1	2	3	4	5	
	Not at all confident	Somewhat confident	Moderately confident	Very confident	Extremely confident	
All students	23(3%)	84	220	277	65(10%)	669
Programming Experience						
None	16(5%)	54	124	116	27(8%)	337
Some	7(2%)	30	96	160	38(11%)	331
Gender						
Female	17(4%)	66	165	202	40(8%)	490
Male	6(4%)	11	45	70	23(15%)	155
Nonbinary/other	0(0%)	1	2	1	2(33%)	6
Did not disclose	0(0%)	6	8	4	0(0%)	18
Race/ethnicity						
Asian	7(3%)	30	86	104	23(9%)	250
Latino	7(6%)	23	38	46	9(7%)	123
White	4(2%)	14	54	83	29(16%)	184
Multi-racial/other	2(6%)	3	11	19	0(0%)	35
M. East/ N. African	1(3%)	6	10	12	2(7%)	31
Black/African Am.	2(8%)	2	11	8	2(8%)	25
Did not disclose	0(0%)	6	9	4	0(0%)	19

Table 5. Distribution of R importance ratings at the end of the course overall and broken down by programming experience, gender, and race/ethnicity.

	Post-course R Importance Ratings (t_3)					
	1 Not at all important	2 Somewhat important	3 Moderately important	4 Very important	5 Extremely important	
All students	14(2%)	42	121	249	241(36%)	667
Programming Experience						
None	7(2%)	25	71	114	119(35%)	336
Some	7(2%)	17	50	134	122(37%)	330
Gender						
Female	11(2%)	28	83	179	186(38%)	487
Male	1(< 1%)	11	33	62	49(31%)	156
Nonbinary/other	0(0%)	1	0	2	3(5%)	6
Did not disclose	2(11%)	2	5	6	3(17%)	18
Race/ethnicity						
Asian	3(1%)	11	44	92	98(40%)	248
Latino	2(2%)	9	22	43	48(39%)	124
White	4(2%)	12	35	78	54(30%)	183
Multi-racial/other	1(3%)	3	6	9	16(46%)	35
M. East/ N. African	1(3%)	4	4	9	13(42%)	31
Black/African American	1(4%)	1	5	9	9(36%)	25
Did not disclose	2(11%)	2	5	7	3(16%)	19

beliefs (e.g., success expectancies, perceptions of value and cost, and beliefs about memorization) and engagement with course materials.

Characteristics of the Two Subgroups There were no discernible differences in the two subgroups of students at the start of the course. Students in the two subgroups showed similar distributions of prior programming experience, $X^2(1) = 0.88$, p = 0.3, gender $X^2(1) = 0.35$, p = 0.6, race/ethnicity, $X^2(2) = 4.1$, p = 0.1, and grade point average $X^2(4) = 8.9$, p = 0.06. The groups also did not differ in terms of the proportion of students who mentioned R as a concern at the beginning of the course $X^2(1) = 0.69$, p = 0.4. Raw data for the socio-demographic characteristics of students in the neg-neg and neg-pos groups are presented in Table 6.

Belief Trajectories We used repeated measures ANOVA to compare changes in expectancy and value beliefs over time for the two subgroups of students (neg-neg vs. neg-pos). As Figure 3 shows, there was an interaction between subgroup and time for perceived course value, F(2,948) = 31.5, p < 0.0001.

Though both groups valued the course content at the beginning of the course, the neg-neg group came to value the course less over time more than the neg-pos group. Paired sample *t*-tests and Cohen's *d* calculations revealed that, although both groups decreased their perceived value from timepoint 1 to timepoint 2, students in the neg-neg group, t(180) = 9.0, p < 0.001, d = 0.71, decreased their perceived value more than students in the neg-pos group, t(277) = 3.4, p = 0.0009, d = 0.23. Students in the neg-neg group also showed a decrease in mean value from time point 2 to timepoint 3, t(180) = 2.3, p = 0.02, d = 0.16, whereas students in the neg-pos group did not, t(277) = 0.28, p = 0.8, d = 0.02.

A repeated measures ANOVA revealed an interaction between subgroup and time for expectancy, F(1, 465) = 20.9, p < 0.0001. As Figure 4 shows, both groups of students developed lower expectations for success from beginning to end of the course. However, the neg-neg group showed a greater decrease in their expectations from timepoint 1 to timepoint 2, t(179) = 7.8, p < 0.0001, d = 0.72, than did the neg-pos group, t(286) = 2.13, p = 0.03, d = 0.16. A one-way ANOVA

 Table 6.
 Socio-demographic characteristics of students in neg-neg and neg-pos subgroups.

	Subgroup		
	Neg-pos	Neg-neg	
Total students Programming Experience	297	196	
None	160	114	
Some	137 (42%)	82 (46%)	
Gender			
Female	233 (79%)	145 (74%)	
Male	56	40	
Nonbinary/other	4	2	
Did not disclose	4	9	
Race/ethnicity			
Asian	122 (41%)	67 (34%)	
Latino	54 (18%)	47 (24%)	
White	76 (26%)	41 (21%)	
Multi-racial/other	21 (6%)	11 (6%)	
Middle Eastern/North African	10 (3%)	12 (6%)	
Black/African American	9 (3%)	9 (5%)	
Did not disclose	5 (2%)	9 (5%)	



Figure 3. Perceived course value at the beginning of the course (t1), after chapter 8 (t2), and at the end of the course (t3) for students in the neg-neg and neg-pos subgroups.

comparing perceived course cost for the two subgroups at time 2 further revealed that students in the neg-neg group (M = 3.42, SD = 0.95) felt the course was more costly (i.e., the course was time consuming and stressful) than students in the neg-pos group (M = 2.61, SD = 0.97), F(1, 476) = 80.1, p < 0.0001.

We used repeated ordinal regression to compare changes in conceptions that the course required memorization from timepoint 1 to timepoint 3 across the two subgroups. Holding time constant, the odds of students in the neg-neg group agreeing that the course required memorization were more than twice as high as the odds of a student in the neg-pos group agreeing that the course required memorization $(\exp(\beta) = 2.43, \text{likelihood} \text{ ratio } X^2(1) = 29.1, p < 0.0001)$. There was also an interaction between time and subgroup, such that the odds that students in the neg-neg group would increase their agreement that the course required memorization were more than twice as high as the odds that students in the neg-neg group would increase their agreement that the course required memorization were more than twice as high as the odds that than students in the neg-pos group would increase their agreement that the course required memorization were more than twice as high as the odds that than students in the neg-pos group would increase their agreement that the course required memorization were more than twice as high as the odds that than students in the neg-pos group would increase their agreement that the course required memorization were more than twice as high as the odds that than students in the neg-pos group would increase their agreement that the course required memorization were more than twice as high as the odds that than students in the neg-pos group would increase their agreement that the course required memorization over



Figure 4. Expectations for success at the beginning of the course (t1) and after chapter 8 (t2) for students in the neg-neg and neg-pos subgroups.

time $(\exp(\beta) = 2.15)$, likelihood ratio $X^2(1) = 9.10$, p = 0.003). Taken together, these results suggest that students who maintained negative or neutral attitudes toward R throughout the course may have developed different patterns of beliefs than students who developed more positive attitudes toward R.

Engagement with Course Materials over Time One reason why a student might develop more negative beliefs is if they experience failure or perceive their performance to be low. Alternatively, experiencing success might lead students to develop more positive beliefs. To explore this hypothesis, we visually inspected students' performance on the R exercises, measured as the proportion of exercises answered correctly on the first attempt (standardized for each chapter, as the difficulty of R exercises varied) and on the review activities at the end of each chapter (also standardized by chapter). As Figures 5 and 6 show, the neg-pos students performed better on end of chapter review questions (Figure 5) and the embedded R activities (Figure 6) than neg-neg students.

Students who ended the course feeling negative perceived the course to be more costly. One reason they might perceive the course as costly is if they are spending a disproportionate amount of time or effort on the course compared to students who end the course feeling positive. We looked at two proxy measures of time and effort to explore this idea. First, we looked at the number of attempts it took students to arrive at the correct answer for R exercises they got wrong on their first attempt. Next, we looked at the word count of students' chapter summaries, reasoning that the length of their written response would positively correlate with student effort. As Figure 7 shows, the neg-pos and neg-neg groups did not show statistically discernible differences in the number of attempts on R exercises that were initially incorrect, but students who ended positive tended to make more attempts than those who ended negative. Similarly, neg-pos students showed a trend of writing more words on average for the end of chapter summaries (Figure 8).



Figure 5. Performance on end-of-chapter review questions for students in the negneg and neg-pos groups.

Note: Performance was measured as the proportion of end-of-chapter review questions answered correctly. A *z*-score was calculated for each student as a measure of average performance on the review questions in each chapter relative to other students. The graph shows estimates and 95% confidence intervals for each group by chapter.



Figure 6. Proportion of R exercises answered correctly on the first attempt for each chapter for students in the neg-neg and neg-pos subgroup.

Note. Performance was measured as the proportion of R activities that were submitted correctly on the first attempt. A *z*-score was calculated for each student as a measure of average performance on all R activities in each chapter relative to other students. The graph shows estimates and 95% confidence intervals for each group by chapter.

6. Discussion

The goal of this study was to explore whether integration of R coding within an interactive online textbook can reduce the perceived cost of learning programming for students while increasing its value and relevance to statistical understanding and practice. Specifically, we were interested in whether integrating R programming in a way that facilitates student understanding of statistics would lead psychology students taking an introductory statistics course to develop more positive attitudes toward R and improve their motivation to learn statistics.



Figure 7. Attempts to arrive at the correct answer on incorrect R activities for each chapter for students in the neg-neg and neg-pos subgroups.

Note: Average attempts to arrive at the correct answer for each chapter for each student were calculated by dividing the total number of attempts on R exercises in each chapter by the total number of R exercises in that chapter the student answered incorrectly on the first attempt. A *z*-score was calculated for each student as a measure of average attempts to correct on R activities in each chapter relative to other students. The graph shows estimates and 95% confidence intervals for each group by chapter.



Figure 8. Word count for end-of-chapter summaries for students in the neg-neg and neg-pos subgroups.

Note: Word count was calculated as the total number of words students used in their end-of-chapter summaries A *z*-score was calculated for each student as a measure of summary length for each chapter relative to other students. The graph shows estimates and 95% confidence intervals for each group by chapter.

In line with previous research (i.e., Anderson et al. 2008; Baser 2013) our results show that, in the beginning, some students tend to hold negative attitudes toward programming. Yet, we found that most students (84%) ended the course either positively disposed or neutral toward R. This corroborates past evidence that suggests although students may initially be more anxious about computing, their anxiety greatly abates with practice (Du, Wimmer, and Rada 2016; Rode and Ringel 2019).

Even more promising, we found that students, in general, developed more positive attitudes toward R over time and

this pattern appeared similar for students of different genders, race/ethnic backgrounds and different levels of prior experience. In fact, students who were the most concerned about learning programming at the beginning of the course showed the greatest increase in their attitudes toward programming after engaging with course materials. These findings suggest that, in line with our hypothesis and findings from other successful projects like MOSAIC, that when students are introduced to programming languages like R in a way that supports understanding, they can develop more positive attitudes toward programming. Similar results were found by Charters et al. (2014) in a population of adults learning computer programming for the first time. Our study extends these findings to a new population: undergraduate students learning computer programming in the context of an introductory course in statistics.

Although the disparities in R attitudes between students from different backgrounds became less pronounced over time, there were still detectable differences in students' R confidence and perceptions of the importance of R for their learning. For example, male students and White students felt more confident in their ability to use R to analyze a new dataset than women and Asian and Latino students. Thus, although providing students with opportunities to work with R to promote understanding may help to narrow discrepancies between demographic groups, it does not completely erase them. This finding is similar to other findings that show women tend to have lower confidence in STEM domains compared to male students, despite equal levels of preparation and previous performance (Ellis, Fosdick, and Rasmussen 2016).

While most students who entered the course with negative attitudes toward R developed positive attitudes toward R by the end of the course, some students maintained their negative attitudes throughout the course. Students' difficulties learning programming are well documented in the literature (i.e., Qian and Lehman 2017), but little is known about what differentiates students who come to enjoy programming from those who do not. Most studies have focused on the relationship between initial attitudes and learning. For example, Tai (2003) found that students who held positive attitudes toward computerassisted learning environments demonstrated greater learning of computer programming. However, there is considerable value in studying the experiences of students who do not go on to forming positive identities, on a more fine-grained level over time.

Our findings provide some preliminary insights into students' experiences and belief trajectories over time. Students who began the class with negative attitudes toward R but later reported positive attitudes (neg-pos) and students who never adopted positive attitudes (neg-neg) did not differ in terms of demographics or initial course perceptions and expectations. However, partway through the course, students in the neg-neg group valued the course less, had lower expectations for success, and believed that the course was more costly. What might drive these differences?

Some clues as to why students had such different experiences can be found in their approaches to learning: the neg-neg group believed (both at the beginning and end of the course) that learning would require a lot of memorization. They generally performed worse on the R coding exercises and end of chapter review activities. Past research has shown that students who conceive of learning as memorization hold an interconnected pattern of beliefs that can be maladaptive to learning (Säljö 1979; Van Rossum, Deijkers, and Hamer 1985). For example, students who view learning as memorization often tend to view their ability to learn as fixed, stable, and unchanging (Chan 2008) and are more likely to believe that learning is about testing, calculation, and practice than understanding and connecting concepts (Liang and Tsai 2010). Students who conceive of learning as memorization may also adopt a "surface level" approach to learning, be less interested in the course material, and show lower self-efficacy (Tsai et al. 2011). Our results show partial support for a relation between beliefs and performance, as students in the neg-neg group performed lower on the end of chapter review activities as well as the R programming activities, suggesting that they may have a different approach to the course material that leads to lower performance in general.

Further, we found that although the neg-neg students perceived the course to be more time-consuming than the negpos students, their patterns of engagement with course materials suggest they spent equivalent amounts of time on the course, and they did not spend more time on R exercise attempts, nor did they write more on open-response questions. In fact, students who adhered to their negative attitudes tended to make fewer attempts and used fewer words than students who developed more positive attitudes. This suggests that, though objectively, both groups of students put roughly equal effort into the course with pos-pos students sometimes showing greater effort, students who had negative attitudes toward R at the end of the course subjectively felt the course was more time-consuming.

These results, while exploratory, raise questions about the role of students' experiences and behavior and the role of these factors in attitude formation. Past studies have found experience factors to be only moderately correlated with students' attitudes (i.e., Garland and Noyes 2004). Our results indicate that experiences and behaviors during learning—specifically students' performance and perceptions of how valuable and costly the course is—may differentiate students who develop more positive attitudes from those who maintain initial negative attitudes.

Strengths, Limitations, and Areas for Future Research This exploratory study contributes to the growing body of knowledge of how computational tools can be integrated to improve statistics teaching and learning. A strength of this study is that we were able to longitudinally track attitudes and link these beliefs to demographic and performance data. As a proof of concept, our results were promising. Students reported positive experiences using R during our course, and this pattern was similar for students across varying levels of prior experience. One reason for the success may be the way we integrated R in the course. For example, in addition to designing our integration to reduce the cognitive load of learning R and making the experience of using R more valuable to students, we implemented additional supports such as a help desk that was embedded directly into the online textbook, a glossary of R functions, and a cheat sheet that students could download and use as a reference. The effect of these approaches warrants more exploration in the future.

This study was limited in a number of ways. First, we measured students' attitudes toward R using a single rating item at the beginning and end of the course. However, students' attitudes toward R are likely more complex and may vary substantially at different points in the course. As our results show, students' attitudes toward learning R may also differ from their confidence and from how important they perceived R to be for their learning. Future studies should investigate the different dimensions of students' R attitudes and how they change over time throughout the course in relation to contextual factors such as overall course performance or the difficulty of the material. It may also be helpful to include more qualitative measures of students' R attitudes to develop a richer and more complete understanding of how students are feeling about learning R.

Second, our investigations of the relationship between change in students' attitudes toward R and their motivational beliefs and experiences throughout the course were mostly exploratory. More research is needed to understand the relationship and direction between these variables. Future studies could investigate whether interventions that target students' motivational beliefs or early experiences using R might lead those students who maintained negative attitudes toward R to develop more positive attitudes by the end of the course. For example, past research has shown that feedback is important for learning programming. Perhaps changing the nature or content of the feedback students receive when they answer a question incorrectly may buffer against negative experiences when students make mistakes learning R. Similarly, adding additional scaffolds to make learning R less costly or identifying and reaching out to students who struggle with R early in the course may improve their subsequent learning experiences.

Other important limitations of the present study include the small sample sizes in some socio-demographic subgroups (e.g., African American/Black, Middle Eastern/North African, Multiracial/other race/ethnic identity, and nonbinary people) and the treatment of demographic subgroups in our survey. To better capture the diversity of students' experiences, we are collecting data from a more diverse sample of students at the university, community college, and high school levels. Future studies will include expanded formal statistical analyses that include important, yet, traditionally excluded groups whose experiences we were not able to fully capture in this study. Additionally, we have updated our demographic survey items to better capture the diversity of student backgrounds. For example, we included additional questions to take into account the vastly different backgrounds of Asian American students and to differentiate between Asian American students and international students from Asia.

Finally, an important unanswered question concerns how learning R might benefit students' learning of statistics concepts. Though we did not address the relationship between learning R and students' statistics understanding in this study, we encourage future research in this area. For example, it would be interesting to know what types of experiences using R are useful for developing understanding and transfer of statistics concepts, and whether, once they have learned how to use R to explore data, students apply those strategies to understand and explore datasets they encounter in the future.

Funding

This project has been made possible in part by a grant from the Chan Zuckerberg Foundation, and a Learning Lab from the Governor's Office of Planning and Research, California (OPR18115).

References

- Anderson, N., Lankshear, C., Timms, C., and Courtney, L. (2008), "Because it's Boring, Irrelevant and I Don't Like Computers': Why High School Girls Avoid Professionally-Oriented ICT Subjects," *Computers & Education*, 50, 1304–1318. https://doi.org/10.1016/j.compedu.2006.12.003.
- ASA GAISE College working group, Carver, R., Everson, M., Gabrosek, J., Rowell, G. H., Horton, N. J., Lock, R., Mocko, M., Rossman, A., Velleman, P., Witmer, J., and Wood, B. (2016), "Guidelines for Assessment and Instruction in Statistics Education: College report, 2016," Available at https://www.amstat.org/asa/education/Guidelines-for-Assessmentand-Instruction-in-Statistics-Education-Reports.aspx.
- Barron, K. E., and Hulleman, C. S. (2015), "Expectancy-Value-Cost Model of Motivation," in *International Encyclopedia of the Social & Behavioral Sciences*, J. D. Wright, pp. 503–509, Amsterdam: Elsevier. https://doi.org/10.1016/b978-0-08-097086-8.26099-6.
- Baser, M. (2013), "Attitude, Gender, and Achievement in Computer Programming," *Middle East Journal of Scientific Research*, 14, 248–255. https://doi.org/10.5829/idosi.mejsr.2013.14.2.2007.
- Baumer, B., Çetinkaya-Rundel, M., Bray, A., Loi, L., and Horton, N. J. (2014), "R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics," *Technology Innovations in Statistics Education*, 8, 1–30.
- Biehler, R. (1997), "Software for Learning and for Doing Statistics," *International Statistical Review*, 65, 167–189. https://doi.org/10.1111/j.1751 5823.1997.tb00399.x.
- Çetinkaya-Rundel, M., and Rundel, C. (2018), "Infrastructure and Tools for Teaching Computing Throughout the Statistical Curriculum," *The American Statistician*, 72, 58–65. https://doi.org/10.1080/ 00031305.2017.1397549.
- Chan, K. (2008), "Hong Kong Teacher Education Students' Epistemological Beliefs and their Relations with Conceptions of Learning and Learning Strategies," *The Asia Pacific Education Researcher*, 16, 36–50. https://doi.org/10.3860/taper.v16i2.265.
- Charles, M., and Theibaud, S. (2018), Gender and Stem: Understanding Segregation in Science, Technology, Engineering, and Mathematics. Basel: Mdpi.
- Charters, P., Lee, M. J., Ko, A. J., and Loksa, D. (2014), "Challenging Stereotypes and Changing Attitudes." in SIGCSE Proceedings. https://doi.org/10.1145/2538862.2538938.
- Cheryan, S., Master, A., and Meltzoff, A. N. (2015), "Cultural Stereotypes as Gatekeepers: Increasing Girls' Interest in Computer Science and Engineering by Diversifying Stereotypes," *Frontiers in Psychology*, 6, 49– 49. https://doi.org/10.3389/fpsyg.2015.00049.
- Christensen, R. (2018), "ordinal Regression Models for Ordinal Data": R package version 2018.8-25. Available at *http://www.cran.r-project.org/package=ordinal/.*
- Cobb, G. (2007a), "One Possible Frame for Thinking about Experiential Learning," *International Statistical Review*, 75, 336–347. https://doi. org/:10.1111/j.1751-5823.2007.00034.x.
- (2007b), "The Introductory Statistics Course: A Ptolemaic Curriculum?," *Technology Innovations in Statistics Education*, 1, 1–17. https://escholarship.org/content/qt6hb3k0nz/qt6hb3k0nz.pdf.
- (2015), "Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum from the Ground Up," *The American Statistician*, 69, 266–282. https://doi.org/10.1080/ 00031305.2015.1093029.
- Cobb, P., Gravemeijer, K. P., Bowers, J., and Doorman, M. (1997), Statistical Minitoolss [Applets and Applications], Nashville and Utrecht: Vanderbilt University and Freudenthal Institute – Utrecht University.
- Dalgaard, P. (2008), Introductory Statistics with R. New York: Springer.
- Dede, C., Richards, J., and Saxberg, B. (2019), *Learning Engineering for Online Education: Theoretical Contexts and Design-Based Examples*, New York: Routledge.
- Du, J., Wimmer, H., and Rada, R. (2016), "Hour of Code": Can it Change Students' Attitudes Toward Programming?" *Journal of Information Technology Education: Innovations in Practice*, 15, 053–073. http://doi.org/10.28945/3421.
- Ellis, J., Fosdick, B. K., and Rasmussen, C. (2016), "Women 1.5 Times More Likely to Leave STEM Pipeline after Calculus than Men: Lack of Mathematical Confidence a Potential Culprit," *PloS One*, 11, e0157447. https://doi.org/10.1371/journal.pone.0157447.

- Field, A., Miles, J., and Field, Z. (2012), *Discovering Statistics Using R*, London: Sage Publications Ltd.
- Finzer, W. (2001), Fathom Dynamic Statistics [Computer Software], Emeryville, CA: Key Curriculum Press.
- Fries, L., Son, J. Y., Givvin, K. B., and Stigler, J. W. (2020), "Practicing Connections: A Framework Guide Instructional Design for Developing Understanding in Complex Domains," *Educational Psychology Review*, 33, 739–762. https://doi.org/10.1007/s10648-020-09561-x.
- Gal, L., and Ginsburg, L. (1994), "The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework," *Journal of Statistics Education*, 2, 1–16. https://doi.org/10.1080/10691898.1994.11910471.
- Gallup (2016), Trends in the State of Computer Science in U.S. K-12 Schools. Gallup.
- Garland, K. J., and Noyes, J. M. (2004), "Computer Experience: A Poor Predictor of Computer Attitudes," *Computers in Human Behavior*, 20, 823–840. https://doi.org/10.1016/j.chb.2003.11.010.
- Google and Gallup. (2015), Images of Computer Science: Perceptions Among Students, Parents, and Educators in the US.
- Gould, R., Wong, R., and Ryan, C. N. (2017), *Essential Statistics: Exploring the World through Data*. Boston, MA: Pearson.
- Hulleman, C. S., and Harackiewicz, J. M. (2009), "Promoting Interest and Performance in High School Science Classes," *Science*, 326, 1410–1412. https://doi.org/10.1126/science.1177067.
- Hulleman, C. S., Kosovich, J. J., Barron, K. E., and Daniel, D. B. (2017), "Making Connections: Replicating and Extending the Utility Value Intervention in the Classroom," *Journal of Educational Psychology*, 109, 387–404. https://doi.org/10.1037/edu0000146.
- Konold, C., and Miller, C. (2004), *Tinkerplots [Computer Software]*, Amherst, MA: University of Massachussetts Amherst.
- Kosovich, J. J., Hulleman, C. S., Barron, K. E., and Getty, S. (2015), "A Practical Measure of Student Motivation: Establishing Validity Evidence for the Expectancy-Value-Cost Scale in Middle School," *The Journal of Early Adolescence*, 35, 790–816. https://doi.org/10.1177/0272431614556890.
- Liang, J., and Tsai, C. (2010), "Relational Analysis of College Science-Major Students' Epistemological Beliefs toward Science and Conceptions of Learning Science," *International Journal of Science Education*, 32, 2273– 2289. https://doi.org/10.1080/09500690903397796.
- Lieberman, M. (2018), "Learning Engineers Inch Toward the Spotlight." Inside Higher Ed. Available at https://www.insidehighered.com/digitallearning/article/2018/09/26/learning-engineers-pose-challenges-andopportunities-improving.
- Lock, R. H., Lock, P. H., Morgan, K. L., Lock, E. F., and Lock, D. F. (2021), Statistics: Unlocking the Power of Data, Hoboken, NJ: Wiley.
- Mascaró, M., Sacristán, A. I., and Rufino, M. M. (2016), "For the Love of Statistics: Appreciating and Learning to Apply Experimental Analysis and Statistics through Computer Programming Activities," *Teaching Mathematics and Its Applications*, 35, 74–87. https://doi. org/10.1093/teamat/hrw006.
- McCulloch, R. S. (2017), "Learning Outcomes in a Laboratory Environment vs. Classroom for Statistics Instruction: An Alternative Approach Using Statistical Software," *International Journal of Higher Education*, 6, 131– 142. https://doi.org/10.5430/ijhe.v6n5p131.
- Moore, D. S. (1992), "Teaching Statistics as a Respectable Subject." In *Statistics for the Twenty-First Century*, eds. F. Gordon and S. Gordon, pp. 14–25, Washington, DC: Mathematical Association of America.
 (1997), "New Pedagogy and New Content: The Case of Statistics," *International Statistical Review*, 65, 123–165.
- Navarro, D. (2020), "Learning Statistics with R." Available at https://learningstatisticswithr.com.
- Nolan, D., and Temple Lang, D. (2007), "Dynamic, Interactive Documents for Teaching Statistical Practice," *International Statistical Review*, 75, 295–321.
- (2010), "Computing in the Statistics Curricula," *The American Statistician*, 64, 97–107.
- (2015), *Data Science in R*, Boca Raton, FL: Chapman and Hall/CRC. Paas, F., Renkl, A., and Sweller, J. (2004), "Cognitive Load Theory: Instruc-
- Paas, F., Kenki, A., and Swenet, J. (2004), Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture," *Instructional Science*, 32, 1–8. http://www. jstor.org/stable/41953634.
- Pruim, R., Kaplan, D. T., and Horton, N. J. (2017), "The Mosaic Package: Helping Students to 'Think with Data' Using R," *The R Journal*, 9, 77– 102.

- Qian, Y., and Lehman, J. (2017), "Students' Misconceptions and Other Difficulties in Introductory Programming: A Literature Review," ACM Transactions on Computing Education, 18, 1–24. https://doi.org/10. 1145/3077618.
- R Core Team (2019), "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. Available at *http://www.R-project.org/*.
- Reid, A., and Petocz, P. (2002), "Students' Conceptions of Statistics: A Phenomenographic Study," *Journal of Statistics Education*, 10, 1–18.
- Riegle-Crumb, C., and King, B. (2010), "Questioning a White Male Advantage in STEM: Examining Disparities in College Major by Gender and Race/Ethnicity," *Educational Researcher*, 39, 656–664. https://doi. org/10.3102/0013189x10391657.
- Rode, J. B., and Ringel, M. M. (2019), "Statistical Software Output in the Classroom: A Comparison of R and SPSS," *Teaching of Psychology*, 46, 319–327. https://doi.org/10.1177/0098628319872605.
- rOpenSci (2018), "rOpenSci Educators Collaborative: What are the challenges when teaching science with R?" R Bloggers. Available at https://www.r-bloggers.com/2018/07/ropensci-educators-collaborative-what-are-the-challenges-when-teaching-science-with-r/.
- Rossman, A., and Chance, B. (2014), "Using Simulation-based Inference for Learning Introductory Statistics," Wiley Interdisciplinary Reviews: Computational Statistics, 6, 211–221.
- Säljö, R. (1979), "Learning about Learning," *Higher Education*, 8, 443–451. https://doi.org/10.1007/bf01680533.
- Scott, M. J., and Ghinea, G. (2014), "On the Domain-Specificity of Mindsets: The Relationship between Aptitude Beliefs and Programming Practice," *IEEE Transactions on Education*, 57, 169–174. https://doi. org/10.1109/te.2013.2288700.
- Siek, K. A., Connelly, K., Stephano, A., Menzel, S., Bauer, J., and Plale, B. (2006), "Breaking the Geek Myth: Addressing Young Women's Misperceptions about Technology Careers," *Learning & Leading with Technol*ogy, 33, 19–22.
- Son, J. Y., Blake, A. B., Fries, L., and Stigler, J. W. (2021), "Modeling First: Applying Learning Science to the Teaching of Introductory Statistics," *Journal of Statistics and Data Science Education*, 29, 4–18. https://doi.org/10.1080/10691898.2020.1844106.
- Stigler, J. S., Son, J. Y., Givvin, K. B., Blake, A. B., Fries, L., Shaw, S. T., and Tucker, M. C. (2020), "The Better Book Approach for Education Research and Development," *Teachers College Record: The Voice of Schol*arship in Education, 122, 1–32.
- Sweller, J. (1988), "Cognitive Load during Problem Solving," Cognitive Science, 12, 257–285.
- ——— (1999), Instructional Design in Technical Areas, Melbourne: ACER Press.
- (2018), "Instructional Design," in *Encyclopedia of Evolutionary Psychological Science*, eds. T. K. Shackelford and V. A. Weekes-Shackelford, pp. 1–5, Cham: Springer. https://doi.org/10.1007/978-3-319-16999-6_2438-1.
- Tai, D. Y. (2003), "A Study on the Effects of Spatial Ability in Promoting the Logical Thinking Abilities of Students with Regard to Programming Language," World Transactions on Engineering and Technology Education, 2, 251–254.
- Tek, F. B., Benli, K. S., and Deveci, E. (2018), "Implicit Theories and Self-Efficacy in an Introductory Programming Course," *IEEE Transactions on Education*, 61, 218–225. https://doi.org/10.1109/te.2017.2789183.
- Thille, C. (2018), Bridging Learning Research and Teaching Practice for the Public Good: The Learning Engineer, New York: TIAA Institute.
- Tsai, C., Jessie Ho, H. N., Liang, J., and Lin, H. (2011), "Scientific Epistemic Beliefs, Conceptions of Learning Science and Self-efficacy of Learning Science Among High School Students," *Learning and Instruction*, 21, 757–769. https://doi.org/10.1016/j.learninstruc.2011.05.002.
- Van Rossum, E. J., Deijkers, R., and Hamer, R. (1985), "Students' Learning Conceptions and their Interpretation of Significant Educational Concepts," *Higher Education*, 14, 617–641. https://doi.org/10.1007/ bf00136501.
- Ward, B. W. (2013), "What's Better—R, SAS*, SPSS*, or Stata*? Thoughts for Instructors of Statistics and Research Methods Courses," *Journal* of Applied Social Science, 7, 115–120. https://doi.org/10.1177/193672 441245057.