

DEVELOPING A FIRST-YEAR SEMINAR COURSE IN STATISTICS AND DATA SCIENCE

Aimee Schwab-McCoy

Xavier University

Cincinnati, OH, United States of America

schwaba1@xavier.edu

Statistical literacy is an increasingly important skill for today's students. Undergraduate enrollments in statistics at both the introductory and advanced level have skyrocketed, and statistics education researchers have done much to modernize the curricula and increase student engagement. The course describes in this paper explores an alternative approach to statistical literacy and data science: a discussion-oriented, first-year seminar course. The seminar course emphasizes real data problems, student-led discussions and critiques, and the use of statistics in media and policy decisions. This paper will discuss the structure and justification, the content, and challenges instructors may face when adapting the course to their own institution.

INTRODUCTION

There have been calls in the statistics education literature for “socially-minded” introductory courses in statistics that promote the growth of statistical literacy framed in relevant, real-world problems (Tishkovskaya & Lancaster, 2010). The growing need for statistically-literate individuals in the workforce means shifting the traditional perceptions of what an introductory statistics course should look like. One possible goal for a course emphasizing statistical literacy is the development of “statistical citizens” (Rumsey, 2002). These are students who will go on to be statistical consumers, who need to think critically about data and have an awareness of the importance, misuse, and serious impacts of data on their daily lives. Since then, “Stat 101” courses have been developed with an eye toward modern statistical methods such as randomization tests and the bootstrap (Hesterberg, 2015; Lock, Lock, Lock, Lock, & Lock, 2012; Tintle et al., 2016). These courses often trade traditional emphases on calculation and methodology for intuitive understanding and technology-driven results. Statistics education has exploded as a discipline, with an entire issue of *The American Statistician* recently devoted to advances in teaching statistics at the university level. However, the emphasis on “Stat 101” has been at the expense of other courses in the curriculum, which are not nearly as well-represented in the statistics education literature (Horton & Hardin, 2015). Additionally, the explosion of “big data” has revolutionized the way statisticians work, and brings big implications for the way we teach statistics (Ridgway, 2015). As instructors, the responsibility falls to us to design courses that appeal to both groups: the statistically sophisticated and the statistical novices, and introduces students to the wealth of possibilities available in the world of big data.

This paper will explore the development of a first-year seminar course in statistics and data science at a private institution in the Midwestern United States. Course structure, statistical content, reading assignments and assessments, and anticipated pedagogical challenges will be discussed. A sample lesson plan will be presented to differentiate the seminar approach from a traditional statistics classroom, along with some preliminary student feedback. This paper will ideally serve as a model for developing similar courses at other institutions, or provide the inspiration for incorporating seminar-style lessons in existing courses.

SEMINAR FORMAT COURSE FOR FIRST YEAR STUDENTS

Statistics and data science are just one growth area in higher education. A trend that crosses disciplinary boundaries is the first-year seminar course: courses that are explicitly designed for freshmen

students to introduce them to the university environment and model deep study of a topic or discipline. The National Resource Center for First-Year Experience and Students in Transition in the United States has identified over 150 colleges and universities that offer or require some form of a first-year seminar course (National Resource Center, 2016). First-year experiences, such as these seminar-style courses, have measureable positive effects on student GPAs and retention (Jamelske, 2009). Seminar-style first-year courses also have the potential to serve as “recruiting pipelines” for science, technology, engineering, and mathematics (STEM) disciplines. Student interviews suggest that seminar courses in STEM may help empower students to pursue further study and career paths they may not have considered otherwise (Sweeder & Strong, 2012).

Seminar courses are a major departure from the traditional courses offered in both statistics, and STEM in general. These courses are characterized by readings of challenging texts (essays, book chapters, journal articles), followed by in-depth discussion of the texts during class. In some ways, this is similar to the popular “flipped classroom” approach (Kuiper, Carver, Posner, & Everson, 2015; McGee, Stokes, & Nadolsky, 2016; Winquist & Carlson, 2014), since students are reading course material before class. Where the seminar course differs is in the physical classroom experience. Instead of completing activities or group work, as would be common in a statistics “flipped classroom”, students are engaging in active discussion about the assigned readings, and delving deeper into the conceptual issues presented by the texts. The instructor’s role in seminar courses is to facilitate and guide the conversation, and provide an academic perspective on the topic at hand. Formative assessments such as low-stakes reaction papers and quizzes, discussion participation and leadership, and higher-stakes essays are used instead of traditional summative assessments like exams (Keup et al., 2011).

In many cases, first-year seminar courses are an institutional effort. Xavier University, a private, Jesuit Catholic institution in the Midwestern United States, recently instituted a first-year seminar program that all students are required to complete. The courses, designated Core 100, are designed to introduce students to interdisciplinary thought, college life, and the “greater good”. The first-year seminar program launched in Fall 2015, after the revision of the new core curriculum. These courses should encourage critical and creative thinking, employ effective research skills, and teach students to construct an argument and use evidence to support it. Faculty apply to teach the course by writing a topic proposal, but may choose any topic. Seminar offerings during the first year of the course included courses from areas across the humanities, social sciences, and arts, however, few courses were proposed from STEM disciplines. *Core 100: Life in the Data Deluge* will be one of the first.

CORE 100: LIFE IN THE DATA DELUGE

Core 100: Life in the Data Deluge is all about statistics and data science in students’ daily lives. Students will learn statistical concepts motivated by engaging, data-driven examples. The statistical topics for the Core 100 course will be surprisingly similar to those on a traditional “Stat 101” syllabus. During their assigned readings for the course, students will be exposed to:

- Experimental and observational studies
- Graphical and numerical summaries of data
- Sampling variability
- Concepts of probability
- Hypothesis testing, specifically the logic and conclusions
- Interval estimation and margin of error
- Regression models
- Classification models
- Machine learning algorithms and concepts

- Data collection (random samples, “ $n = all$ ”)

Despite the similarity in statistical content, this is emphatically not a “Stat 101” course. For example, sampling variability and confidence intervals will be motivated by a unit discussing political polls and the American Presidential election: why do poll results vary, and what changes from poll to poll? What is the margin of error, and why is it necessary? Is the estimated poll number actually the true population approval rating? Sports analytics and sabermetrics lead naturally into statistical modeling: how do we make predictions about a team’s performance? What information is necessary to make an accurate prediction, and what information is extraneous? Graphical summaries of data can go beyond producing histograms or dotplots by examining infographics in the news media. A discussion of what infographics do well, and what they perhaps do not, introduces the same ideas as a traditional textbook treatment of data summaries. Topics from data science, such as machine learning, web scraping, and algorithmic thought are another important component of the course – and typically missing from the introductory statistics syllabus.

A hallmark of seminar-style courses is “challenging readings”: students are expected to read new texts outside their comfort zone and synthesize those texts into deeper class discussions. Class readings will be selected from popular statistics websites and blogs, and books with substantial content in statistical literacy and quantitative literacy. For example, GAPMINDER may be used to introduce students to data visualization in a fun and interactive way, without requiring specialized knowledge in statistical software (Le, 2013). The popular blog FiveThirtyEight regularly publishes pieces featuring statistical analyses of current events in politics, sports, and science (FiveThirtyEight, 2016). Readings will also be selected from three recently released books, all of which use concepts from statistics and data science in interesting applications – such as online dating, virtual poker, and earthquake forecasting.

- *The Signal and the Noise* (Silver, 2015)
- *Big Data* (Mayer-Schonberger & Cukier, 2014)
- *Dataclysm* (Rudder, 2014)

These particular resources were selected for their presentation of advanced statistical content such as multivariate and time series analyses in relevant and easy to understand ways. Using recognizable and interesting examples for first-year college students invites students to participate in discussions in more meaningful ways and encourages them to ask questions about the statistical methods underlying the “cool result”. Additionally, students will practice writing skills through short reflection papers and conduct individual data collection and analysis projects.

One advantage of the seminar course is its dynamic nature. Student projects are encouraged in the traditional classroom to strengthen connections to course content (Carver et al., 2016; Tishkovskaya & Lancaster, 2012; Zieffler et al., 2008); in seminar courses connections are made not only through projects, but through student-led discussion. During the semester, students will have the chance to lead discussion in teams on a data-related topic of their choosing. The students will choose the readings (from a set of possible topics), prepare thought-provoking questions for their peers to consider before class, and finally lead the in-class discussion. Students are encouraged to choose topics they are passionate about and required to meet individually with the instructor to discuss their plans before class. Seminar courses at this particular institution are capped at 15 students, so it is possible to give students opportunities to lead discussion for an entire class period without sacrificing weeks or months on a syllabus. Larger courses may need to assign leaders to work in groups of three or four.

EXAMPLE: PERSONAL DATA AND TARGETED ADVERTISING

Many statistics educators may not have experience teaching a seminar-style course. Core 100: Life in the Data Deluge is currently in its first iteration, having officially launched as a course in Fall 2016. An example lesson plan related to student personal data on social networking sites and targeted advertising follows, with preliminary student reactions, to hopefully illustrate the differences between a seminar-style course and a traditional statistics course.

At the start of the semester, many first-year students are unaware of the role data plays in their lives. They may be aware of personal data as an abstract concept – they post pictures, statuses, and updates on social media all the time without an understanding of the depth of information that is actually available. To introduce students to these ideas of personal privacy, large-scale data collection, and targeted advertising, a class period was devoted to learning about what information is actually available to advertisers through social networking sites such as Facebook.

This lesson started with three recent articles from *Slate* and *The Washington Post* about targeted ads and Facebook (Dewey, 2016; Wagner, 2016; Weingarten, 2016). In August 2016 Facebook made some of their data collection methodology public, including a list of 98 data points that they collect on all users to sell to advertisers. As a starting point, students were asked to read these three articles and conduct a short survey. Previous class discussion had suggested that students felt there was a generational gap in regards to “big data”. Some students felt that since they had grown up in the age of big data, they were more comfortable with things like targeted ads than their parents, grandparents, and even older siblings. For their survey, students asked three of their peers and two people they considered “older” the following question,

“How do you feel about your online data being recorded by companies and used to target advertising to you?”

Students were instructed to take careful notes on what their respondents said, and bring them to class the next day. During class discussion, students were split into smaller groups and asked to look for common themes in the responses. Generally speaking, their “peers” were more likely to be comfortable with their online data being recorded. However, the students were surprised that a large number of their “older” respondents were comfortable, or even supportive, of targeting advertising.

The next part of the discussion involved a critique of the survey itself. Students were asked to think about how they would modify their survey to collect more representative data. This was early in the semester, yet the students had already read research descriptions in previous articles. This meant that they had had some exposure to research descriptions and intuitively were developing a “good idea” for what made a survey informative or representative. Students in this class recognized that random sampling was not a part of their survey, and that their survey was much smaller than ones they had read about in the course. They also recognized that their roommates, dormitory neighbors, parents, and cafeteria workers probably did not constitute a representative sample of the American adult population. Random, representative, sufficient sampling is an important concept in any introductory statistics course, and one that students were able to intuit themselves in an informal discussion, without a traditional lecture or lesson.

Following our discussion of their survey results, students were asked to reflect on their personal reactions to the list of data points Facebook provides to advertisers. The objective was for students to think about these data points individually, and digest the wealth of information available to retailers. To help their reflection, they completed an activity called “Your consumer profile, visualized”. Students were provided with a box of crayons and a table of data points available to advertisers, categorized into groups like “basic information”, “life events”, “employment history”, and “financial history”. Students were instructed to color in each cell in the table according to how comfortable they felt with that particular

piece of information being available. This served two purposes: students were able to create a data visualization by hand and (hopefully) recognize the usefulness of visualizations. Second, and maybe more importantly, students had to think about each data point individually and make a decision. Instructions and a complete list of variables can be found in the Appendix.

After about 15-20 minutes, the students finished their reaction heat maps and posted them in the classroom. Once all heat maps were on the board, they were asked to look for common themes and trends. Students thought they were comfortable with advertisers having access to their personal data before the activity, but after doing the activity some areas of concern were clearly present. Students noticed:

- The “basic information” such as name, gender, location was almost unanimously acceptable. Of the demographic variables listed, “ethnic affinity” was the one students were most uncomfortable with.
- “Financial information” like the number of credit lines, spending habits, and which banks or credit cards the customer uses were very uncomfortable for most students. Students were also uncomfortable with information like their income, household composition, and travel habits being unavailable to online retailers.
- Some students were generally more comfortable with their data being available to retailers, and some were not. Many students tended to assign middle rankings to these variables – which indicated that they were find with the data being available, but found it “creepy”.

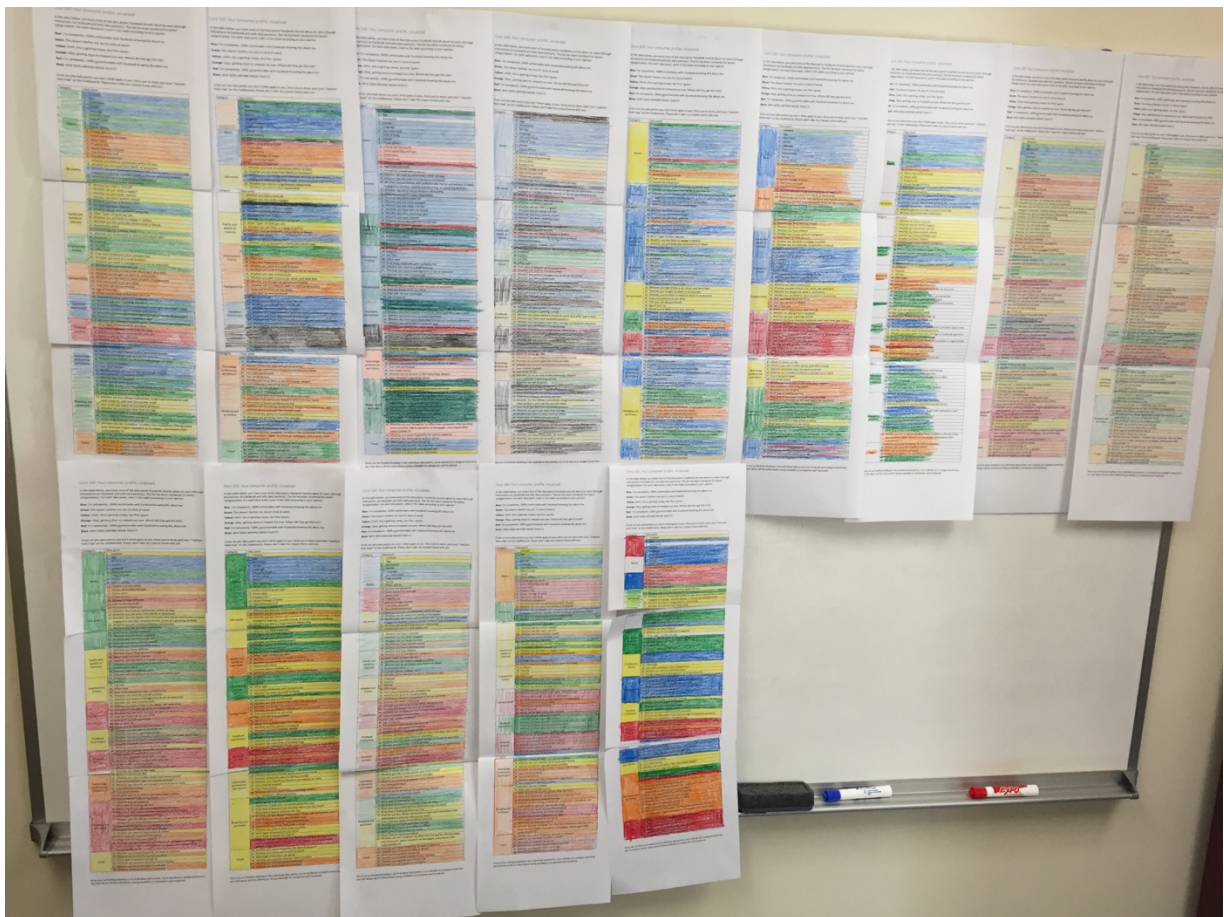


Figure 1. Student "reaction heat maps" to 88 personal data points collected by Facebook to sell to advertisers. Students shaded the cells in the table based on how comfortable they were with a particular data point being available. Blue represented the

highest level of comfort, followed by green, yellow, orange, and finally red, which indicated the highest level of discomfort. Students also shaded cells in black if they didn't want that data available to advertisers under any circumstances.

After displaying the reaction heat maps (shown in *Figure 1*), the instructor asked students if completing this activity changed their opinions on the data that was collected. One student stated,

“Doing this activity made it more tangible the data that is being collected. It was one thing to read the article, but it didn't really sink in how much data they collect until I had to think about each one.”

This learning activity challenged students' perceptions of personal data collection, and raised their awareness of the issue. Students taking this course are all new to the college experience, so they may not have thought about or been exposed to privacy issues before taking this course. Statistical issues covered in this lesson included sampling design, representative samples, and data visualization. Students also learned about the value of data collection, and were hopefully motivated to learn more about data science algorithms and details later in the semester. Students were active at all points during the class discussion, and expected to participate in the course at a deeper level than in a typical introductory statistics course. Even though this course is still in the pilot stages and this activity occurred early in the semester, this is an encouraging example for the promise of a data-oriented freshman seminar.

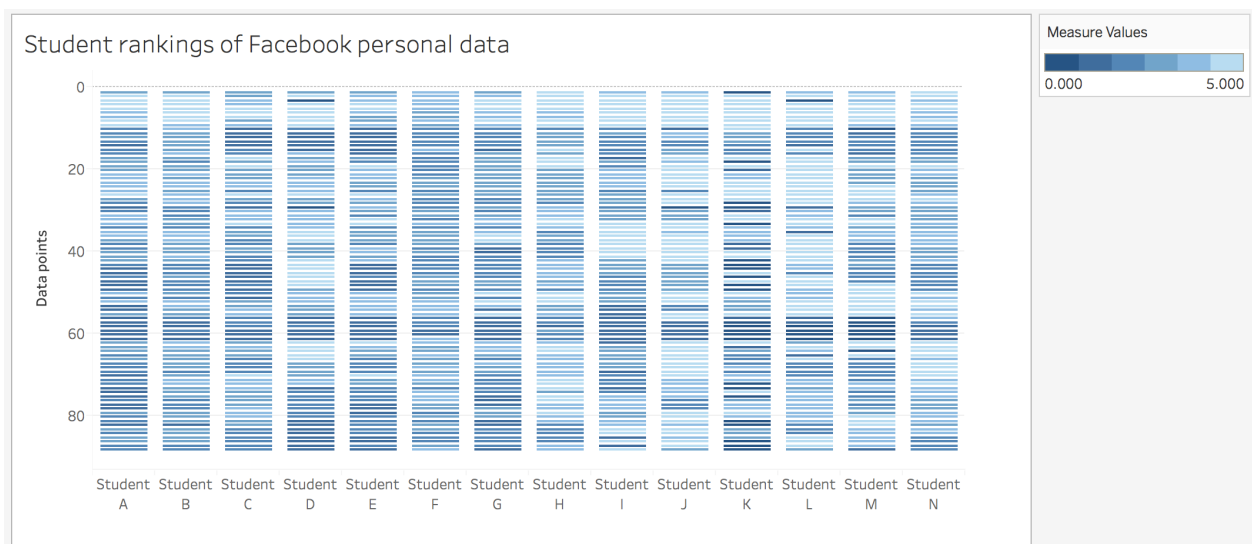


Figure 2. Student responses to the 88 data points, reproduced in Tableau. Lighter values indicate higher levels of comfort, and darker values indicate lower levels. A complete list of the variables presented, as well as the student handout, are available in the Appendix.

CORE 100 AND THE GAISE GUIDELINES

The American Statistical Association published the original Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report in 2005 (Aliaga et al., 2005). The six GAISE College Report recommendations were revised in 2015, to reflect the improvements in technology and overall changes in pedagogy (Carver et al., 2016). The updated GAISE recommendations are as follows:

1. Teach statistical thinking.
2. Focus on conceptual understanding.
3. Integrate real data with a context and purpose.
4. Foster active learning.

5. Use technology to explore concepts and analyze data.
6. Use assessments to improve and evaluate student learning.

Core 100: Life in the Data Deluge falls very much in line with these recommendations. Statistical literacy is the overall goal of the course, with students being taught to critically read, discuss, and write about statistics and data in the media. Since Core 100 focuses on challenging readings taken from all walks of society, artificial data does not appear anywhere in the course. All example data analyses presented to students are real, and within their original context. One potential downside of the course is the lack of emphasis on procedural knowledge. However, despite the fact that students may not work with statistical software hands-on or conduct statistical analyses themselves, they are very involved in the statistical communication and conceptual understanding on these analyses. Students will learn to interpret results from technology that are presented in the media as statistical consumers instead of “statistical producers” (Rumsey, 2002). Students will also develop a critical sense for data manipulation, collection, and presentation – many of the same statistical literacy skills we strive for in the traditional “Stat 101” curriculum.

CHALLENGES AND OPPORTUNITIES

Each new course offers challenges for the instructor, Core 100 is certainly no exception. The biggest potential challenge for instructors teaching a first-year seminar course is the format: a discussion-based course would likely be a first for statistics instructors. In fact, many instructors coming to statistics from a STEM background may have never taken a seminar-style course, let alone taught one. Seminar courses can (and should) be student-led as much as possible, but this takes some control over class content out of the instructor’s hands. A possible strategy could be to assign discussion questions along with the planned reading to help frame the students’ thoughts before class. Many guides have been prepared to help instructors prepare for teaching seminar courses for the first time, such as *The First-Year Seminar: Designing, Implementing, and Assessing Courses to Support Student Learning & Success* (Keup et al., 2011).

Assessment is another challenge. Small writing components are now common in many introductory statistics courses, however Core 100 emphasizes writing as one of the primary assessment methods in the course. The emphasis on writing has multiple pedagogical benefits. Seminar-style courses increase student exposure to written statistical communication through reading statistical reports and journaling about their experiences in the classroom. Students also participate in daily guided discussions and critical questioning of their readings in the course, which increases student engagement and understanding (Theoret & Luna, 2009). Instructors may also find it useful to reach out to colleagues in the humanities who teach seminar courses and assess writing more regularly for advice on how to best structure and prepare for the course. Reaching beyond our traditional “bag of tricks” for teaching introductory statistics can be a wonderful opportunity for professional development as quantitative instructors.

There are some important learning opportunities for students taking the course. Core 100 is based completely on texts about statistics and social issues, rather than any statistics textbook. This fosters a deep conceptual understanding of statistics *in context*. Students immediately see the relevance of statistics and the importance of thinking about variability, randomness, and statistical assumptions because the entire course content is packaged within interesting concepts. This comes however with a tradeoff, a decrease in the coverage of statistical techniques. At the end of Core 100, students should not be expected to be able to find a p-value, or find a regression model. Student success in this course is instead measured by their understanding – for example, whether they can interpret a p-value and explain controversies related to significance testing such as “p-hacking”.

The course also provides new content opportunities. Relatively advanced topics in statistics can be covered at a purely conceptual level. One example is nonlinear or multiple regression, which is rarely covered in introductory statistics. Once students understand the idea of a single explanatory variable being used to predict or describe a single response variable based on a straight line, it is not much of a conceptual leap to picture a curve being used instead, or more than one explanatory variable. Classification methods are another example; classification algorithms rely on complex calculations most first-year students are not mathematically prepared for, but the basic idea behind them is intuitive. Advanced data visualization methods such as infographics, spatial maps, or social network graphs are used widely in popular media sources such as FiveThirtyEight and simple to explain visually to students.

Instructors interested in developing seminar-style courses without an existing institutional framework are encouraged to start small. Reading news articles and blog posts about data can easily be incorporated into existing statistical courses, both at the introductory and advanced level. Leading students in discussions of statistics and data in the media can increase their engagement, as long as the topic is one they feel a connection to or the relevance of. Students can also gain a deeper understanding of the need for statistical literacy and practice statistical communication in an informal setting.

CONCLUSIONS

Core 100: Life in the Data Deluge presents an alternative model for teaching concepts from statistics and data science in a seminar-style course targeted toward first-year students. The seminar style course offers a flexibility in content and delivery that is unusual in many introductory courses, which provides challenges and opportunities for new learning experiences. The course is necessarily fluid, the world of big data is changing constantly and students are always interested in the “hot new thing”.

Lessons from the seminar-style course can also be used to enrich an existing curriculum. Instructors can incorporate articles or blog posts from popular media sources that communicate a statistical topic, and lead a short discussion over the article. In a more traditional course, seminar activities may need to be focused on a specific topic from the syllabus (such as sampling techniques or margin of error), and students may need guided questions to prompt discussion.

The course launched in Fall 2016, however preliminary feedback from the first few weeks of the course have suggested that students are beginning to appreciate the promise of big data, and the potential problems it poses. Future iterations of the course will be subject to more rigorous data collection and evaluation, however I am hopeful that such a course will attract new students to statistics and STEM disciplines as a whole, develop their statistical literacy skills, and foster a lifelong interest in data and statistics.

ACKNOWLEDGEMENTS

The author would like to thank the reviewers, presenters, and discussants for their thoughtful comments and questions at IASE Berlin. Many of the discussions at that conference inspired new subject applications for this course, and I am grateful to my colleagues for their feedback.

APPENDIX: Your consumer profile, visualized

The table below lists 88 of the data points Facebook records about its users (through interactions on Facebook and with data partners). The list has been reordered from what you read in *The Washington Post* for better categorization. For each data point, color in the table according to your opinion.

Blue: I'm completely, 100% comfortable with Facebook knowing this about me.

Green: This doesn't bother me, but it's kind of weird.

Yellow: Uhhh, this is getting creepy, but fine I guess.

Orange: Okay, getting close to creeped out now. Where did they get this info?

Red: I'm completely, 100% uncomfortable with Facebook knowing this about me.

Black: WHY DOES ANYONE KNOW THIS?!?!

Cross out any data points you don't think apply to you. Once you're done, post your "reaction heat map" on the chalkboards. Please don't take my crayons home with you.

Category	Data point
Basics	1. Location
	2. Age
	3. Generation
	4. Gender
	5. Language
	6. Education level
	7. Field of study
	8. School
	9. Ethnic affinity
	10. Income and net worth
	11. Home ownership and type
	12. Home value
	13. Property size
	14. Square footage of home
	15. Year home was built
	16. Household comparison
Life events	17. Whether you have an anniversary within 30 days
	18. Whether you are away from family or hometown
	19. Whether you are friends with someone who has an anniversary, is newly engaged or married, recently moved, or has an upcoming birthday
	20. Whether you are in a long-distance relationship
	21. Whether you are in a new relationship
	22. Whether you have a new job
	23. Whether you are newly engaged
	24. Whether you are newly married
	25. Whether you have recently moved
	26. Whether you have a birthday soon
Family and beliefs or interests	27. Whether you have children
	28. Whether you or your partner is pregnant
	29. What "type" of mom you are
	30. Whether you are likely to engage in politics
	31. Whether you are politically conservative or liberal
	32. Relationship status
	33. Interests (sports, hobbies, etc.)
	34. Whether you've donated to charity (and which type)
Employment history	35. Employer
	36. Industry
	37. Job title
	38. Office type
	39. How many employees your company has

Category	Data point
	40. Whether you work at a small business
	41. Whether you work in management or are an executive
Transportation	42. Whether you own a motorcycle
	43. Whether you plan to buy a car, when, and what kind
	44. Whether you bought car parts or accessories
	45. Whether you are likely to need car parts or accessories
	46. Style and brand of car you drive
	47. Year your car was purchased
	48. Age of your car
	49. Where you're likely to buy your next car
Facebook interactions	50. Whether you play games in Facebook
	51. Whether you own a gaming console
	52. Whether you have created a Facebook event (and what type it was)
	53. Whether you have used Facebook payments
	54. Whether you have spent more than average on Facebook payments
	55. Whether you administer a Facebook page
Financial services	56. Whether you belong to a credit union, national bank, or regional bank
	57. What types of investments you have
	58. How many available credit lines you have
	59. Whether you are an active credit card user
	60. Which credit card(s) you have
	61. Whether you carry a balance on your credit card
Technology preferences and history	62. Whether you listen to the radio
	63. Which TV shows you like
	64. Whether you use a mobile device (and which brand)
	65. Internet connection type and service provider
	66. Whether you recently purchased a smartphone or tablet
	67. Your internet browser
	68. Your email service
	69. Whether you're an early or late technology adapter
	70. Your computer's operating system
Shopping and purchases	71. Whether you use coupons
	72. What type of clothing you buy
	73. What time of year your household shops the most
	74. What kind of groceries you buy (and where you buy them)
	75. What kind of beauty products you buy
	76. Whether you buy allergy medications, cough/cold medications, pain relief products, and over-the-counter meds
	77. How much you spend on household products
	78. Whether you purchase more than average
	79. Whether you tend to shop online or offline
	80. What types of restaurants you eat at
	81. What kinds of stores you shop at
	82. Whether you are "receptive" to offers from companies offering online auto insurance, higher education or mortgages, and prepaid debit cards/satellite TV
Travel	83. Whether you travel frequently, for work or pleasure
	84. Whether you commute to work or school
	85. What type of vacations you go on
	86. Whether you recently returned from a vacation
	87. Whether you recently used a travel app
	88. Whether you participate in a timeshare

Once you've finished shading in the individual data points, try to decide at a category level how you feel about all this information being available to companies and Facebook.

REFERENCES

- Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., ... Witmer, J. (2005). GAISE College Report. *American Statistical Association*.
- Carver, R. H., Everson, M. G., Gabrosek, J., Holmes Rowell, G., Horton, N. J., Lock, R., ... Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education: College Report*. American Statistical Association.
- Dewey, C. (2016, August 19). 98 personal data points that Facebook uses to target ads to you. *The Washington Post*.
- FiveThirtyEight. (2016). <http://fivethirtyeight.com/>.
- Hesterberg, T. C. (2015). What Teachers Should Know about the Bootstrap: Resampling in the Undergraduate Statistics Curriculum. *The American Statistician*.
- Horton, N. J., & Hardin, J. S. (2015). Teaching the Next Generation of Statistics Students to “Think With Data”: Special Issue on Statistics and the Undergraduate Curriculum. *The American Statistician*, 69(4), 259–265.
- <http://www.sc.edu/fye>. (2016). National Resource Center for the First-Year Experience and Students in Transition.
- Jamelske, E. (2009). Measuring the impact of a university first-year experience program on student GPA and retention. *Higher Education*, 57(3), 373–391.
- Keup, J. R., Hunter, M. S., Groccia, J. E., Garner, B., Latino, J. A., Ashcraft, M., ... Petschauer, J. W. (2011). *The First-Year Seminar: Designing, Implementing, and Assessing Courses to Support Student Learning & Success*. Columbia, SC: National Resource Center for the First-Year Experience & Students in Transition (University of South Carolina).
- Kuiper, S. R., Carver, R. H., Posner, M. A., & Everson, M. G. (2015). Four Perspectives on Flipping the Statistics Classroom: Changing Pedagogy to Enhance Student-Centered Learning. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 25(8), 655–682.
- Le, D. T. (2013). Bringing data to life into an introductory statistics course with GAPMINDER. *Teaching Statistics*, 35(3), 114–122.
- Lock, R., Lock, P. F., Lock, K., Lock, E. F., & Lock, D. F. (2012). *Statistics: Unlocking the Power of Data*. Wiley.
- Mayer-Schonberger, V., & Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*.
- McGee, M., Stokes, L., & Nadolsky, P. (2016). Just-in-Time Teaching in Statistics Classrooms. *Journal of Statistics Education*, 24(1).
- Ridgway, J. (2015). Implications of the Data Revolution for Statistics Education. *International Statistical Review*, 1–22.
- Rudder, C. (2014). *Dataclism: Who We Are (When We Think No One’s Looking)*. Broadway Books.
- Rumsey, D. J. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses. *Journal of Statistics Education*, 10(3).
- Silver, N. (2015). *The Signal and the Noise: Why So Many Predictions Fail -- but Some Don’t*. Penguin Books.
- Sweeder, R. D., & Strong, P. E. (2012). Impact of a Sophomore Seminar on the Desire of STEM Majors to Pursue a Science Career. *Journal of STEM Education*, 13(3), 52–61.
- Theoret, J. M., & Luna, A. (2009). Thinking Statistically in Writing: Journals and Discussion Boards in an Introductory Statistics Course. *International Journal of Teaching and Learning in Higher Education*, 21(1), 57–65.
- Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2016). *Introduction to Statistical Investigations*. Wiley.
- Tishkovskaya, S., & Lancaster, G. (2012). Statistical Education in the 21st Century: a Review of Challenges, Teaching Innovations and Strategies for Reform. *Journal of Statistics Education*, 20(2).
- Tishkovskaya, S., & Lancaster, G. A. (2010). Identified Problems in Teaching and Learning Statistics.

- International Conference on Teaching Statistics Proceedings*, 8, 2–5.
- Wagner, L. (2016, August 22). Facebook Pulled Back the Curtain on Targeted Advertising. Yikes. *Slate*.
- Weingarten, E. (2016, August 8). There’s No Such Thing as Innocuous Personal Data. *Slate*.
- Winqvist, J. R., & Carlson, K. A. (2014). Flipped Statistics Class Results: Better Performance Than Lecture Over One Year Later. *Journal of Statistics Education*, 22(3).
- Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What Does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature. *Journal of Statistics Education*, 16(2), 1–25.