Infusing Data Visualization into Intro Stat Using Tableau

Stacey Hancock
Department of Mathematical Sciences, MSU

International Conference on Teaching Statistics
Kyoto, Japan
July 13, 2018

Outline

• Why data visualization in the intro course?
• Why Tableau in the intro course?
• Tableau activity
• Assessment and results
• Improvements and next steps
GAISE College Report

1. Teach statistical thinking.
   - Teach statistics as an investigative process of problem-solving and decision-making.
   - Give students experience with multivariable 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.

A state legislator in a guest column writes:

“A scatter-plot graph of tuition versus graduation rates in the 50 states reveals a strong correlation—low tuition is associated with low graduation rates, high with high...

The clear expectation is that taxpayers should fork over escalating amounts to keep tuition low... [but] low tuition spurs a high dropout rate.” (emphasis added)
Graduation Rate vs In-State Tuition

Data source:
U.S. Department of Education National Center for Education Statistics
https://nces.ed.gov
Why Data Visualization in the Intro Course?

- easily investigate causal claims in the news!
- multivariable thinking in the first week
- explore “real data with a context and purpose”
- component of data literacy
Why Tableau?

- powerful and authentic
- free for academic use
- provides a platform for sharing work (Tableau Public)
- drag-and-drop interface that facilitates data exploration
- produces quality graphics without requiring programming skills
- requires minimal training in how to use the software

Intro Stat at Montana State University

- Flipped format
- Students working in groups of three, tables of nine
- Simulation-based curriculum (Tintle et al.)
- About 20-25 sections of 40 students each semester

http://www.montana.edu/pdc/projects/2013/teal.htm
Tableau Activity (Fall 2017)

One 75-minute class period:
1. Classroom discussion of data visualization in the news (10 min)
2. Instructor demo of Tableau (10 min)
3. Activity (60 min):
   • Current Population Survey data with step-by-step instructions
   • Roller Coaster data
4. Instructor wrap-up (5 min)

1. What are the observational units in these data?
2. What are the variables or summary measures displayed in the plot and what “graphical perception tasks” (Cleveland, 1994) are used for each variable, e.g., position, area, angle? What are the variable types (quantitative or categorical)?
3. What would the data set look like in a spreadsheet?
4. What is the overall message the author is attempting to communicate through the visualization?
5. What comparisons are easy to make using this data visualization? Hard to make?
6. How could we improve this visualization?

http://nyti.ms/2pC2HZy
Current Population Survey Data

- Mix of categorical and quantitative variables, e.g., hourly wage (US dollars per hour), number of years of education, sex, age (years), sector of the economy (sales, clerical, etc.)
- Wang, Rush & Horton (2017) describe data visualization activity using these data with R and R Studio
- Use this data set to give step-by-step instructors for different types of plots in Tableau

Roller Coaster Data

- Sample of 157 roller coasters in the US that opened between 1915 and 2016
- 15 Variables: Coaster, Park, City, State, Type, Design, Year Opened, Top Speed, Max Height, Drop, Length, Duration, Inversions?, # of Inversions, Age Group
- Some activity questions from ESTEEM: Enhancing Statistics Teacher Education with E-Modules (http://hirise.fi.ncsu.edu/project/esteem/)
- Activity worksheet posted at: http://www.math.montana.edu/courses/s216/TableauActivityS18.pdf
Roller Coaster Data

1. How fast is the fastest roller coaster? What is the name of this roller coaster?
2. Do newer roller coasters seem to have higher top speed than older ones? Create a data visualization that answers this question.
3. Are newer roller coasters more likely to be made of steel than older ones? What about roller coasters with a high top speed?
4. Is there a relationship between top speed and maximum height? Describe this relationship.
5. Investigate some other factors that you think might have a relationship with top speed. Create a Dashboard comprised of data visualizations which you think summarize the relationship between top speed and at least two other variables that seem to be related to top speed. Add text to the Dashboard and write a few sentences describing what you find.
Assessment

Student learning:
• Tableau questions on most homework assignments
• Tableau visualization given on exams
• Course project

Tableau activity:
• Fall 2017: Online student survey in week 14 (one 1 section)
• Spring 2018: Pre- and post-activity online student survey (all sections)
Qualitative Results

• Student feedback in fall 2017 mixed
  • only 4 of 34 agreed that Tableau had enhanced their understanding of statistics
  • several commented that Tableau helped make data easier to interpret and gave them a better understanding of how to visualize data
  • need more time in-class, guidance, and practice with the software
• Spring 2018
  • enjoyed working in a team on activity
  • valued instructor guidance throughout activity
  • viewed activity as a good introduction to visualizing data
Improvements in Spring 2018

• Two class days for activity:
  Day 1: Basic descriptive statistics and plots, instructor demo of Tableau;
  Day 2: Tableau activity with CPS and roller coaster data
• Revisit Tableau often (in-class activities, homework assignments, class project)

Thank you!

stacey.hancock@montana.edu
References


