BIOL 351: Data Visualization and Communication for Biologists

Dr. Kristine Grayson and Dr. Angie Hilliker

Spring 2021

Course Description:

Data sets are increasingly becoming larger and more complex, whether used to direct advertising to your social media feed, track human impact on climate change, or understand the genetic underpinnings of cancer. The need to communicate complex data in clear, accurate, and engaging ways to both novice and expert audiences remains essential.

In this course, we will focus on the presentation of datasets in biology and learn visualization approaches that accurately represents data in compelling ways. You will learn about the common pitfalls, mistakes, and misleading practices in data presentation. We will use tools such as Tableau and Program R to manage and process large data sets and visualize trends in data. You will also explore more creative and visually compelling ways to present data to better communicate your results and support your conclusions. We will explore cutting-edge topics in biology and the connections with current events through data. The skills we will build are important across many careers and for guiding decision-making in other aspects of your life.

Pre-requisites: BIOL 202; Course credit: Upper level BIOL Elective

*Students should not be able to receive credit for this class and PSYC 359: Data Visualization and Analysis (currently taught by Andrew Bell) due to content overlap. PSCY 359 is an upper level special topics course that counts as an elective for the Psychology major. It is also a required course for Psychology Majors completing the Neuroscience Concentration.

Course Segments:

- 1. Data management and organization
 - a. Emphasis: tidy data principles, advanced spreadsheeting, introduction to coding
 - b. Software: Excel and Program R
- 2. Data types and graphical approaches for visualize data
 - a. Emphasis: Methods and aesthetics in producing accurate and informative figures
 - b. Software: Tableau and Program R
- 3. Inference from data
 - a. Emphasis: Causality, statistical significance, misrepresentation Note: this course will not cover the learning outcomes from statistics courses that relate to selecting appropriate statistical tests, understanding their underlying assumptions, and executing analyses
- 4. Accessing and cleaning data
 - a. Emphasis: Navigation of public data sources and tools for big data
- 5. Data workflow and establishing reproducible data
 - a. Emphasis: Ethics and scientific misconduct

Sample Readings:

- Data Visualization Principles:
 - <u>Fundamentals of Data Visualization</u>: A Primer on Making Informative and Compelling Figures by Claus O. Wilke.
 - <u>Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm</u> (2015) Tracey Weissgerber et al. PLOS Biology 13(4): e1002128.
 - <u>Reveal, Don't Conceal: Transforming Data Visualization to Improve Transparency</u> (2019) Tracey Weissgerber et al. Circulation 140:1506–1518
- Working in Program R:
 - <u>R for Data Science: Import, Tidy, Transform, Visualize, and Model Data</u> by Hadley Wickham and Garrett Grolemund.
- Exploring the Aesthetics of Data Visualization:
 - o <u>Observe, Collect, Draw: A Visual Journal</u> by Giorgia Lupi and Stefanie Posevec

Sample Activities and Assessments:

- Literature discussions of primary research papers to examine data formats and interpretation. Accessing public data associated with papers and exploring figure types and formatting.
- Case studies on flawed data presentation.
- Journaling and visualization practice exercises using data collected from your own daily activities and from assigned data sets.
- Learning Tableau and Program R via online activities and tutorials.
- Tidy Tuesday style challenges (archived data sets on <u>GitHub</u>; weekly challenge on Twitter <u>#tidytuesday</u>).
- Literature discussions of papers focused on data reproducibility, presentation, and storage.
- Capstone project where students propose and develop a graphical exploration of a biological topic using authentic datasets.