

Data Visualization & Storytelling

Fall 2016



Instructor: Michael Lopez

Office: Harder 217

Email: mlopez1@skidmore.edu

Twitter: @StatsbyLopez

Phone: 580-5297

Office Hours: Thursday 10-12 or by appointment

Class Meetings: WF 8:40-10:00 (Harder 108)

Website: <https://github.com/statsbylopez/DataViz>

Course Goals: Graphics are powerful tools for analyzing and communicating data and summary statistics of data. Proper visualization can yield improved understanding of both small and large data sets. Our course goals are to

- (1) Understand components of data visualization tools
- (2) Objectively critique and redesign graphics
- (3) Produce visualizations, and be able to eloquently describe them both orally and in writing

Course Skills:

- Computing skills using *R* & *R-studio*
- Univariate, multivariate, interactive, and animated plots using *ggplot2*
- Fundamental skills of reproducible data analysis using *RMarkdown*
- Basic tools for scraping data from the internet

Textbook: No textbook is required. Here is a list of recommended books

- ggplot2: Elegant graphics for data analysis*, by Hadley Wickham
- Visualizing Data* by William Cleveland
- Visual Display of Quantitative Information* by Edward Tufte
- Truthful Art* and *Functional Art* by Alberto Cairo

Computing: The use of the R statistical environment (downloadable from <http://www.r-project.org/>) with the RStudio interface (downloadable from www.rstudio.org) is thoroughly integrated.

Writing: Your ability to communicate results, which may be technical in nature, to your audience, which is likely to be non-technical, is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Grading:

Homework [20%] Homework is the most effective way to reinforce concepts learned in class. There will be weekly homework assignments.

RMarkdown will be used as to ensure reproducible work and of a standard format.

On the top of every homework, identify the names of *everyone* you collaborated with in doing problems. A basic principle of scholarship is that once gives credit to all who contributed to the findings. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. All students, staff and faculty are bound by the Skidmore College Honor Code.

Late homework is not accepted. I will drop the lowest homework score.

Homeworks are graded out of 5 points:

- 1-3 out of 5 points: Most questions attempted, minimal effort
- 4 of 5 points: All questions attempted, complete effort, graded questions incorrect
- 4.5 of 5 points: All questions attempted, complete effort, graded questions partially correct
- 5 of 5 points: All questions attempted, graded questions perfect

Projects & Presentation [25%] The major milestone in this class will be conducting (with a group) a data visualization on a question of interest to you. For each, you may collect primary data by hand or you can use data available on the Internet or from faculty research.

You will prepare a project proposal describing your study and obtain approval from me before you begin the investigation. During the last week of class, you (and your group) will give a 10 minute oral presentation of your study, including 2-3 high-caliber visualizations. We will spend time in class looking at what data is available on the web and about writing a project proposal.

Exams [45%] There will be three take-home exams, each comprising 15% of your grade. Exams will be posted on Github and due 48 hours later (generally, Monday-Wednesday).

Participation/Labs [10%] Active participation in class, engagement with the annotated readings notes, regular attendance, and completion of in-class labs will comprise the remainder of your grade.

Disability accommodations: Any student with special needs requiring accommodations should give me his/her memo of accommodations in a timely manner. It is the student's responsibility to follow up with me regarding all accommodations that require my participation. The student is advised to ensure full use of testing accommodations by coming to talk to me at least three days before any test.

Attendance: Your attendance in class is crucial, as is your punctuality. We are all going to learn this material together, so we need to have everyone present and working. Accommodations for an unavoidable absence can be made in advance via email; one necessary absence during the semester is not unusual; having more than two is uncommon.

Collaboration: Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete homework assignments, and prepare for exams. However, every word that you write must be your own. Copying and pasting sentences, paragraphs, or blocks of *R* code from another student is not acceptable and will receive no credit. No interaction with anyone but the instructor is allowed on any exams or quizzes. All students, staff and faculty are bound by the Skidmore College Honor Code.

Additional policies:

Cell phones: If your phone rings, I get to answer it! Respect the rest of the class and turn your phone on silent, airplane mode, or off

Computers: You may bring your own laptop to use for labs, or use the ones provided in Harder, I strongly recommend that computers be closed during note taking, and please be aware that your computer use is a distraction for your classmates. If you would like to use a computer during class, please sit in the back row.

Calendar (subject to change)

Date	Topics	Assignments Due
7-Sep	Introduction and History of Graphics	
9-Sep	<i>Lab 1: Intro to R/RMarkdown</i>	
14-Sep	Grammar of graphics	HW 1
16-Sep	<i>Lab 2: Data manipulation & dplyr</i>	
21-Sep	Continuous variables	HW 2
23-Sep	<i>Lab 3: ggplot2, continuous variables</i>	
28-Sep	Categorical variables	HW 3
30-Sep	<i>Lab 4: ggplot2, categorical variables</i>	
5-Oct	Exam 1	
7-Oct	<i>Lab 5: Multivariate graphs & dependencies</i>	
12-Oct	***No classes***	
14-Oct	Advanced Graphical Considerations	HW 4
19-Oct	Standard errors, trends, uncertainty	
21-Oct	<i>Lab 6: ggthemes and statistical uncertainty</i>	HW 5
26-Oct	Getting your own data	HW 6
28-Oct	<i>Lab 7: Data scraping</i>	
2-Nov	Maps & contours	HW 7
4-Nov	<i>Lab 8: Spatial tools in R**</i>	
9-Nov	Project Work	HW 8
11-Nov	<i>Lab 8: Election data</i>	
16-Nov	<i>Lab 9 – review and saving images</i>	
18-Nov	<i>Skype call with A. McCann, Exam 2</i>	Exam 2
30-Nov	Interactive plots & visualizations, color choices	
2-Dec	Project proposals	
7-Dec	Project work	HW 9
9-Dec	Project work	
14-Dec	Project presentations	
???		Project Papers

*Guest speaker: Allison McCann, Fivethirtyeight

**Guest speaker: Gregory Matthews, Loyola Chicago