STAT GR5702 Exploratory Data Analysis and Visualization

Section 001 Spring 2017

TR 5:40pm-6:55pm; 501 Northwest Corner Building

Instructor: Prof. Joyce Robbins jtr13@columbia.edu

Office Hours: Wed, 12:30pm-2:00pm (or by appt.), 603 Watson Hall (612 W. 115th St.)

Course TAs: Ian Kinsella and Bridget Ratcliffe

Communication:

We will be using Piazza for online class discussion. You are strongly encouraged to post questions here first, where you can receive quick answers from your classmates as well as the TAs and instructor. You are also strongly encouraged to monitor the discussions and help out your classmates as much as possible. I expect the tone of the discussion to be civil and friendly and will not tolerate disrespect. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class page at: https://piazza.com/columbia/spring2017/statgr5702_001_2017_1/home (Links to an external site.)

Required Book:

Unwin, Antony. 2015. *Graphical Data Analysis with R (Links to an external site.)*. CRC Press. ISBN 978-1498715232

Recommended Books:

Viz Theory

Munzner, Tamara. 2014. *Visualization Analysis & Design (Links to an external site.)*. CRC Press. ISBN 978-1466508910
Robbins, Naomi. 2013. *Creating More Effective Graphs (Links to an external site.)*. Chart House. ISBN 978-0985911126

R Resources

Peng, Roger D. 2016. *R Programming for Data Science, 5th ed. (Links to an external site.)* Leanpub.
Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis, 2nd ed. (Links to an external site.)* Springer. ISBN: 978-3319242750
Wickham, Hadley and Garrett Grolemund. 2017. *R for Data Science (Links to an external site.)*. O'Reilly. Web version (Links to an external site.).

Grading:

Your grade will be based on the following:

35% Midterm (Tuesday, March 7) (all conceptual, no coding)

30% Homework Assignments (due Tues 1/31, Thurs 2/16, Tues 2/28, Tues 3/28, 11:5pm)

25% Final Project (due Thurs, 4/13, 11:59pm)

5% Peer Review of Final Projects (due Thurs, 4/20, 11:59pm)

5% Community Contribution (Write-up due Thurs, 4/27, 11:59pm) (see below)

Course Schedule:

* = homework due

Week	**D**ate	**Topics**	Textbook
1	1/17, 1/19	Data Visualization Theory and Practice	Ch. 1, 2
2	1/24, 1/26	Grammar of Graphics, Perception Studies	
3	1/31*, 2/2	Continuous Data	Ch. 3, 5
4	2/7, 2/9	Categorical Data, Multivariate Data	Ch. 4, 6
5	2/14, 2/16*	Multivariate Data (cont.)	Ch. 7, 8
6	2/21, 2/23	Temporal and Spatial Data	Ch. 11
7	2/28*, 3/2	Case Studies, Review	Ch. 10
8	3/7, 3/9	Midterm, Project Help	
		Spring Break	

(last homework due 3/28*)

Post Spring Break topics:

1. Hierarchical and Network Data
2. Animation
3. Interactivity
4. Color
5. Improving graphs for presentation
6. Data Journalism
7. Student suggested topics / presentations

Community Contribution

This fairly open-ended assignment provides an opportunity to receive credit for contributing to the collective learning of the class, and perhaps beyond. To complete the assignment you must submit a short description of your contribution. If appropriate, attach relevant files. (The due date is set at the end of the semester for allow for contributions related to the final project. However, you are encouraged to work on it and share as earlier as possible in order to be more helpful to the class.) **Use Piazza to ask for and offer help.**

There are many ways in which you can contribute:

- give a well-hearsed 5 minute lightning talk in class (live or video) on a datavis topic (theory or tool)
- create a cheatsheet or other resource
- be a Piazza super user
- write a tutorial for a tool that's not well documented
- translate a useful resource into another language
- build a viz product (ex. htmlwidget) for class use
- create a web site for sharing class resources publicly
- provide significant subject matter help to a classmate
- organize and a lead a help session on a topic you've mastered
- other...