Course Syllabus

Jump to Today

STAT GR5702 Exploratory Data Analysis and Visualization (EDAV), Spring 2018

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

Instructor: Prof. Joyce Robbins jtr13@columbia.edu

(Please write "GR5702" or "EDAV" somewhere in your email so I know what class you're in.)

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

Head TA: Jing Wu jw3233@columbia.edu

(Please address all correspondence about homework to Jing; do not attempt to contact the graders whose names may appear on the homework assignments.)

Communication:

We will be using Piazza for online class discussion. You are encouraged to post questions here first, where you can receive quick answers from your classmates. You are also 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.

Required Book:

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

Recommended Book:

Murray, Scott. 2017. *Interactive Data Visualization for the Web (Links to an external site.)Links to an external site., 2nd edition*.

ISBN-13: 978-1491921289

(If you buy this book, make sure you get the 2nd edition.)

Grading:

Your grade will be based on the following:

20%	Test #1 **(Feb 27)**
20%	Test #2 **(Apr 17)**
25%	Homework Assignments **(Jan 30, Feb 13, Mar 6, Mar 27, Apr 10)**
5%	Community Contribution **(Apr 3)**
25%	Final Project **(Apr 24)**
5%	Peer Review of Final Projects **(May 1)**

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-rehearsed 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
- ..

You may draw on and expand existing resources. When doing so, it is critical that you cite your sources.

Topics

I. Understanding Data

1. Introduction to exploratory data analysis and data visualization
2. Perception
3. Continuous variables
4. Discrete variables
5. Dependency relationships
6. Multivariate categorical variables
7. Temporal data
8. Spatial data

II. Context: Data Science Pipeline

1. Collect
2. Import
3. Clean
4. Transform
5. **Visualize**
6. Model
7. Communicate

III. Tools

1. static: R (base graphics / ggplot2)
2. interactive: Plotly, htmlwidgets, Shiny, D3 + ...
3. version control: Git / GitHub
4. communication: Rmarkdown + ...