



STAT 545A, Exploratory Data Analysis 2016/17 Term 1 Instructor: Jenny Bryan

Time and Place: T/R, 9:30-11:00 AM, Sep 06, 2016 to Oct 20, 2016, ESB 2012

Course webpage: <http://stat545.com/>

Objectives: Students will learn

- to explore, groom, visualize and analyze data
- make all of that reproducible, reusable, and shareable
- using R

Prerequisites: There are no official pre-requisites for Stat 545A but most students will have had at least one prior statistics course. Please consult the instructor if in doubt.

Textbook/course materials: All course material is presented in the class or online

Assessment: Course mark = 75% homework, 15% peer review, 10% engagement (i.e. attendance, participation in discussion in-class and/or online)

Students submit homework once a week. The end result may still resemble a “final project”, but one that has been built up gradually over the term.

Topics: Selected topics may include (but are not limited to):

- Introduction to R and the RStudio IDE; scripts, the workspace, RStudio Projects
- Generate reports from R scripts and R Markdown
- Care and feeding of data in R
- Data aggregation; “apply” functions, plyr, dplyr
- Data visualization with ggplot2
- Graphs and descriptive stats for quantitative and categorical variables
- Writing R functions
- Coding style and project organization
- Version control with Git; collaboration via GitHub
- Character data; regular expressions
- Interactive pages, apps, and graphics with Shiny and ggvis
- Get data off the web and expose data, code, results on the web
- Distribute data and code via an R package
- Automate an analytical pipeline, e.g. via Make

Consult the course webpage for more information on specific topics to be covered.