Ma 404 ∼ Probability & Statistics I
Learning to Efficiently Perform Statistical Computations

- Install R.

- Install RStudio.

- Open RStudio. (Start here if you are using a lab computer.)

- Create a manual/example file that illustrates every new command that you learn. Create an outline/-heading structure so that you can find examples of the command quickly. This will serve as your own quick reference guide, so you won't have to keep "relearning" how to run commands. Expect to need this information to complete your analysis project.

  This can be done efficiently in RStudio using RMarkdown files. They allow you to save and run code right in the file, but also allow you to organize it by section/type of analysis and add "comment" lines where you can include notes to yourself on what/how to do things.

- The following is a list of things you will find beneficial to understand as you learn how to use R to complete your project. Learn the skills in R no later than when we discuss them in class (some will never be discussed in class, but you'll need them to do what we do in class).

You want to acquire the following skills (either by just figuring it out, searching online, or using *R for Everyone* - you may borrow this book from me for short periods of time). Chapter references are from *R for Everyone* for your convenience.

- Chapter 2:

  - execute code in RStudio
  - create and save your work in an RStudio project ... I strongly recommend using an RMarkdown file, opened within a project

- Chapter 3:

  - understand the weakness of open source code/packages (things aren't always well defined and there are potential for errors)
  - be able to install/load packages

- Chapters 4-5:

  - saving to variables/names (case sensitive)
  - data formats (factors, numeric, dates)
  - data types (vectors, data frames, lists, selecting specific elements and columns from data)
  - adding comments to code files
  - dealing with empty/missing data
  - select just a subset of data (not in *R for Everyone*)
  - know how to deal with NAs in your data when running commands

- Chapter 6:

  - reading data into R

- Chapter 7:

  - create histograms, scatterplots, density plots, boxplot/violin plots, line graphs
  - use plotly for some beautiful data visualizations (not in *R for Everyone*)

- Chapters 11-12, 18:

  - aggregate, summarize, group by functions
  - summary stats (mean, sd, min, max, median, percentiles)

- Chapter 17:

  - generate a random variable simulated from a specific, known distribution
  - generate the density function value
  - generate the cdf value
  - generate the inverse of the cdf value

- Chapters 18-19, 21:

  - find confidence intervals
  - run T-Test, print results/summary (one-sample, two-sample, paired sample)
  - run ANOVA, print results/summary, find means for each group
  - run linear regression, plot the line and the data, print results/summary and model
  - determine if residuals are normally distributed, compare two or more models

- Chapters 27-29: optional material on typesetting/presenting/sharing work done in R