“Visualization can surprise you, but your models should never surprise you.”

* Hadley Wickham

**GOALS**

Goal 1: install R, learn to be familiar with the R environment, learn basic commands

Goal 2: learn to read in files, manipulate data, merge across spreadsheets, and clean data

Goal 3: run basic descriptive statistics and visualize/QC your data

Goal 4: run basic linear regression and other statistical tests, and understand the output

Goal 5: visualize your statistical results

We will use Tidyverse! https://s3.amazonaws.com/assets.datacamp.com/blog\_assets/Tidyverse+Cheat+Sheet.pdf

**Week 1 6/27/19:** Goals 1 and 2

Basics:

* Installing Packages: Installing packages.Rmd
  + Resources: <https://www.statmethods.net/r-tutorial/index.html>
* Importing data
  + Resources: <https://www.datacamp.com/community/tutorials/r-data-import-tutorial>
  + <https://www.datacamp.com/courses/dplyr-data-manipulation-r-tutorial/>

Data Types:

* Learn R data types of variables

Data Cleaning:

* Changing variable names
* Wide to long format; long to wide format
* Changing levels within a factor
* How to replace values (e.g., “999” to NA)
* Creating variables (e.g., creating a dummy variable or a “total” variable signifying sum of a scale)
  + We will be piping throughout

**NO SESSION 7/4/19**

**Week 2 7/11/19:** Goal 2

Data Cleaning

* Continuing cleaning

Summarizing Data:

* Determine the type of each variable
* in terms of R variable types -> class() and scales of measurement (nominal, interval, ordinal, ratio)
* What is the accurate measure of central tendency for each?
* How can you describe the variance?
* Think about whether you want to transform any variables
* In order apply appropriate statistics: some tests are meant for categorical variables, and some are meant for continuous variables

**Week 3 7/18/19:** Goal 3

Visualization Goals:

* Why it is always a good idea to visualize your data
  + Histograms
* Show the data without distortion
* Need to transform any variables?
* Use color, shape, and location to encourage comparisons
* Minimize visual clutter (maximize your information to ink ratio)
* The two questions that I ask myself before I create any data visualization are: 1) What type of variable(s) am I plotting? 2) What comparison do I want to make salient for the viewer (possibly myself)?
* How to interpret some common data visualizations
* How to identify misleading statistical graphics
* Use of ggplot: <http://r-statistics.co/Complete-Ggplot2-Tutorial-Part1-With-R-Code.html>  
  <https://rstudio-pubs-static.s3.amazonaws.com/265910_5c99663d94f34549ba6389c595cf7282.html>

**Week 4 7/25/19:** Goal 4

Statistical Tests

* lm: <https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R> <https://www.google.com/amp/s/biologyforfun.wordpress.com/2014/04/08/interpreting-interaction-coefficient-in-r-part1-lm/amp/>
* Logistic regression
* T-tests (and t-distribution)
* Paired Sample t-test: compares means from same groups at different times (e.g. pre-post intervention)
  + t.test(x, y, paired = TRUE, data = df)
* One Sample t-test: tests the mean of a single group against a known mean
  + t.test(x, mu = 75, data = df)
* Correlations
  + Pearson’s
  + ICC
  + Spearman’s rank-order
  + cor.test(x, y, alternative = c("two.sided", "less", "greater"), method = c("pearson", "kendall", "spearman"), exact = NULL, conf.level = 0.95, continuity = FALSE, ...)
* Chi-squares
* Modeling categorical data
  + chisq.test(df)
* Effect sizes
  + CohensD = meanDifference/pooledSD
  + *r*
  + standardized beta
  + *R2*
  + OR

**Week 5 8/1/19:** Goal 5

Applying visualization to statistical model

* geom\_bar
* geom\_histogram
* geom\_point
* facet\_wrap
* geom\_errorbar
* plotting residuals

**Week 6 8/8/19:**

* Interpretation
* Working on own projects
* Any outstanding issues

**More Resources**

Hadley Wickham’s R book: <http://r4ds.had.co.nz/> → review all of this once you’re done and/or as you go through the topics in the tutorials above

R Cookbook: <http://www.cookbook-r.com/Graphs/> → I recommend buying a physical copy of this book if you want to get deep into ggplot

Discovering Statistics in R: <https://www.amazon.com/Discovering-Statistics-Using-Andy-Field/dp/1446200469/ref=sr_1_1?s=books&ie=UTF8&qid=1522796679&sr=1-1&keywords=discovering+statistics+using+r>

**Last Assignment: obtain public dataset and clean it!**