Notes 5/1/19

Review

* Recall the Data Analysis Cycle
  + Start with raw data 🡪data cleaning/processing/manipulation 🡪Core analysis🡪 slides🡪 paper
* Programming languages
  + Raw data 🡪 data objects: tables/vectors/lists/etc
* Notation in R for manipulation
  + Obj[…]
* File structure

Project/

README.md

Data/

Rawdata/

Cleandata/

(need a data dictionary in each with metadata)

Output/

Tables/

Summaries/

Code/

Functions/

Scripts/

Tests/

Report/

* Data
  + Tidy-data
  + Aggregations
  + Group\_by / select / filter / subsets / reshaping
  + Regex & strings (stringr)
* File formats
  + Excel data, csvs, rdata, plain text
  + Exporting that data
    - Sinking, write.table() etc
* Reproducibility
  + Relative paths
  + Files and subdirectories
  + Data.org principles
    - Karl Broman
    - Paul Murrel