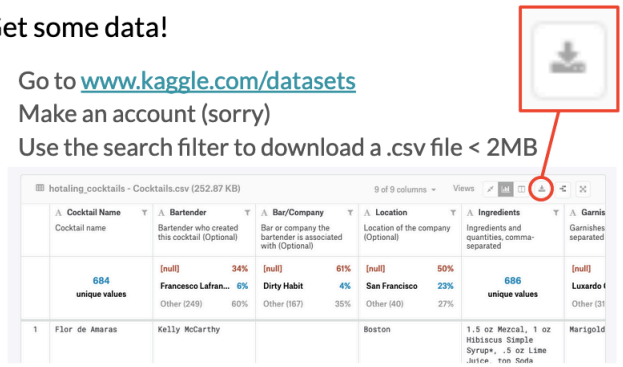
Get some data!



1. Go to www.kaggle.com/datasets
2. Make an account (sorry)
3. Use the search filter to download a .csv file < 2MB

Make a folder structure and project

1. First make a project folder with the name of your project
2. Inside your project folder include folders for:
   a. Raw data, Clean data
   b. Cleaning, Analysis
   c. Graphs, Logs
   d. Archive
3. Create a new Project in Stata in your project folder
4. Drag your folder structure into the project manager!

Convert you raw data to .dta

1. Move your .csv data to your Raw Data folder
2. Create a new .do file in your Cleaning folder
3. Add starter code and comment block
4. `import delimited "Raw Data/YOUR_CSV.csv", varnames(1)`
5. `save "Clean Data/NICE_DATA.dta", replace`
6. Drag your nice new data from the finder into your project manager folder tree

Analyze!

1. Generate summary statistics for a couple variables
2. Make 2 histograms for two interesting continuous variables
   a. `histogram varname`
3. Output each figure to your Plots folder
   a. `graph save "Graphs/filename", replace`
4. Make a scatter plot of the two variables and save to your graphs folder.
5. Play with the advanced stata graphics in the Day 2 folder!

1. Draw a picture that describes each of these mathematical expressions.

$\forall X \in E, X \in K$

$\{X : X \in E, X \in K\}$

$\{X : X \notin E\}$

$\{X : X \in E \text{ or } X \in K\}$