**Stata Cheat Sheet** (replace highlighted sections to match your data):

**Setting up your do-file**

Set working directory: cd "C:\Users\favero\Desktop"

Start log file (if you get an error when you run this, try ending log file first):

log using "project.txt", replace text

End log file (include at the end of your file): log close

**Reading in data**

Open a Stata file: use "covid.dta"

Open a CSV file: insheet using "hw3.csv"

**Data management**

Clear dataset so you can read in another dataset: clear

Delete extra data: keep if year == 2004

Bring variables (columns) to the front of the dataset: order variable(s)

Create a new variable: gen new\_variable\_name = [expression]

* Example: gen age = 2019 - birth

Recode a variable: recode old\_variable ( old\_value(s) = new\_value ) ( old\_value(s) = new\_value ), gen( new\_variable\_name )

* Example: recode race (2/5 = 0) (1 = 1), gen(white)
* See details on syntax here: <http://wlm.userweb.mwn.de/Stata/wstatrec.htm>

**Summarizing 1 variable at a time**

Histogram: hist variable

Boxplot: graph box variable

Stripplot: stripplot variable, jitter(5)

* The first time you run this, you’ll need to install the package: ssc install stripplot

Frequency table: tab variable

* If you want to also show missing values: tab variable, missing

Summary statistics for continuous variables (including percentiles): sum variable(s), detail

**Associations between 2 variables**

1 quant. & 1 qual. variable: bys qual\_variable: sum quant\_variable

* To test for significance: oneway quant\_variable qual\_variable, tab
  + Alternative if qual. var. is binary: ttest usps, by(healthcare)
* Graph: graph box quant\_variable, over( qual\_variable )
  + Alternative: stripplot quant\_variable, over( qual\_variable )
    - Again, you’ll need to install the package the first time: ssc install stripplot

2 quant. variables: corr variable\_1 variable\_2

* To test for significance: pwcorr variable\_1 variable\_2, sig
* For a confidence interval: corrci variable\_1 variable\_2
  + The first time you run this, you’ll need to install the package: findit pr0041\_2
* Graph: twoway scatter variable\_1 variable\_2

2 qual. variables: tab variable\_1 variable\_2 , row

* To get percentages by columns instead: tab variable\_1 variable\_2 , col
* To test for significance: tab variable\_1 variable\_2, row chi2
* Graph: graph bar, over( variable\_1 ) by( variable\_2 )

**Regression**

reg dependent\_variableindependent\_variable(s)

* Example (1 independent variable): reg edinstruct democrat
* Example (2 independent variables): reg edinstruct democrat poptotal