\* STATA Workshop

\* Jan. 25, 2017

\* Empirical Reasoning Center

\* World Values Survey Wave 6 (2010-2013)

\* http://www.worldvaluessurvey.us/WVSOnline.jsp

\* The following line clears current data

\*

\*

\* Note: for help with a function use the command "help"

\* Setting a working directory

\* File>Change working directory>

\* In the pop up window you want the find the folder that contains the data for this workshop

\* Clicking through these menus is the same as running the following line

\* 1

\*

\* Importing data using the file menu

\* File>Import>Text Data (.csv)

\* Then click Browse to find the wvs\_subset.csv in the workshop folder

\* The following line will import the data

\* 2

\*

\* Saving a STATA file using the file menu

\* File>Save as>

\* Keep this STATA dataset in the same workshop folder

\* The following line will save the data

\* 3

\*

\* Opening a STATA file using the file menu

\* File>Open

\* Only ".dta" STATA datasets can be opened this way. Note that this is different from importing an Excel, CSV, or any other dataset type.

\* The following line will open a STATA dataset

\* 4

\*

\* Viewing a dataset

\* The following line displays the dataset like a spreadsheet

\* 5

\*

\* Tabulating a variable

\* A frequency table provides all of the different values that variable takes on and the corresponding number of observations

\* The following line will create a frequency table of the v2 variable

\* 6

\*

\* Renaming a variable

\* The following lines change the name of the v2 variable to Country, tabulate v2 and Country, and relabel the Country variable

\* 7

\*

\* Labeling data values

\* The following lines create a set of labels that translate 156 to China and 840 to US, apply the Country labels to the Country variable, and tabulate the Country variable

\* 8

\*

\* Tabulating two variables

\* A cross tabulation looks at the number of observations along two dimensions

\* The following lines rename the v240 variable Gender, and tabulate the Country and Gender variables

\* 9

\*

\* Note: Gender = 1 is male and = 2 is female

\* The following line adds the row and column percentages to the cross tab

\* 10

\*

\* Working with string data types

\* The following lines rename the v160 variable thirtyyrold and tabulate it

\* 11

\*

\* The thirtyyrold question is: 'Please tell me how acceptable or unacceptable you think most people in [country] would find it if a suitably qualified 30 year old was appointed as their boss?

\* 1 = completely unacceptable & 10 = completely acceptable

\* Summarizing a variable

\* A variable summary provides the basic statistics

\* The following line summarizes the thirtyyrold variable

\* 12

\*

\* The following line provides more information about a variable, like data type

\* 13

\*

\* The following line replaces all observations that have NA as the value for thirtyyrold to missing

\* 14

\*

\* Note: missing for a string variable is "" & missing for a numeric variable is .

\* The following line changes thirtyyrold from a string variable to a numeric variable and summarizes thirtyyrold again

\* 15

\*

\* Charting variables

\* The following line creates a histogram of the thirtyyrold variable

\* 16

\*

\* The following lines rename the v164 variable seventyyrold, replace missing values, and destring it

\* 17

\*

\* Note: the seventyyrold question is: 'Please tell me how acceptable or unacceptable you think most people in [country] would find it if a suitably qualified 70\* year old was appointed as their boss?

\* The following line produces a specified summary of the seventyyrold variable

\* 18

\*

\* Using if statements

\* If statements allow you to analyze subsets of data

\* The following line tabulates the seventyyrold variable only if the individual is from China

\* 19

\*

\* Using the by command

\* A by command allows you to run analyses subsetted by every value of another variable

\* The following line tabulates the seventyyrold variable by Country

\* 20

\*

\* The following line uses tabstat to summarize both the thirtyyrold and seventyyrold variables by country

\* 21

\*

\* Generating a new variable

\* The following lines create a blank variable called US, and assign the value of 1 if the individual is from the US and 0 if the individual is from China

\* 22

\*

\* Recoding a variable

\* The following lines rename variable v242 age, tabulate it, create a categorical age variable, and tabulate it

\* 23

\*

\* Creating correlation tables

\* The following line produces a 3 by 3 correlation table for thirtyyrold, seventyyrold, and age

\* 24

\*

\* Running a regression

\* Does one's age affect how comfortable they are working for a 30-year old?

\* The following line runs a linear regression with thirtyyrold as the dependent variable and age as the independent variable

\* 25

\*

\* In the US does one's age affect how comfortable they are working for a 30-year old?

\* The following line runs a linear regression with thirtyyrold as the dependent variable and age as the independent variable just in the US.

\* 26

\*

\* Does one's age and country (US vs. China) affect how comfortable they are working for a 30-year old?

\* The following line runs a linear regression with thirtyyrold as the dependent variable and age & US as independent variables

\* 27

\*

\* Does one's age and country (US vs. China) affect how comfortable they are working for a 70-year old?

\* The following line runs a linear regression with seventyyrold as the dependent variable and age & US as independent variables

\* 28

\*

\* Creating scatter charts

\* Graphics>Twoway graph>Plots>Create

\* Clicking through these menus is the same as running the following line

\* 29

\*

\* Now, save the STATA dataset again.