Name: ____ Math 227 / Fall 2019 / Prof. Soto

Module 2. In Class Activity 2. Sections 2.4, 2.5, 2.6, 2.7, 2.8

- Open Data Camp Sandbox.

Question 1.

Display the first six cases of the **ACS** data frame.

Run the instruction **ACS$Sex <-factor(ACS$Sex, levels =c (0,1), labels = c (‘Female’, ‘Male’)).**

Display the first six cases of the **ACS** data frame again.

What changes do you notice?

Question 2. Change the 0’s and 1’s from the USCitizen variable in the **ACS** data frame to read Non Citizen and Citizen respectively. What is the instruction that you must type?

Question 3. What is the instruction that you must type in order to display the first 10 records of the **ACS** data frame but only including the Age and Income variables?

Question 4. Let us practice creating a histogram for the Income variable of our ACS data frame. Use the following instruction: **gf_histogram(~Income, data=ACS, color="red", fill="yellow")**

Now create a histogram for the Age variable of the ACS data frame (use the example above but include **binwidth=10** as an attribute inside the formula. Which age subgroup has the largest frequency? How many people are in this subgroup?

Question 5. Suppose we want to clean the ACS data to avoid having the records that list "NA" under Income. To do this we write: **ACS.CleanIncome <-filter(ACS, Income!="NA").** Notice that this instruction saves the clean data into a new data frame with name **ACS.CleanIncome.**

Now filter the **ACS.CleanIncome** data frame to save only the married people into another new frame **ACS.CleanIncome_Married**. Create a histogram for variable Age of this last data frame. What is the group with the highest frequency? About how many people we have in this group?

Question 6. We are going to use the USStates data frame (you can read a description in the last page).

Use the following two instructions to help you answer the question below:

USStatesByHS <- arrange(USStates,HighSchool)

select(USStatesByHS, State, HighSchool)

What is the state with the highest graduation rate in high school?

Question 7. Create a histogram of the **HouseholdIncome** variable from the **USStates** data frame. Do not include the binwidth attribute; instead write "**bins=10**" to have only ten bins in the histogram. Be as accurate as possible when answering the following: What is the mean household income range with the highest frequency? How many states do we have in this range?

Question 8. Create a histogram of the **Smokers** variable from the **USStates** data frame. Let your histogram have 7 bins. Be as accurate as possible when answering the following: What is the range for the percent of smokers with the highest frequency? How many states do we have in this range?

USStates

A data frame with 50 observations on the following 17 variables.

- `State` Name of state
- `HouseholdIncome` Mean household income (in dollars)
- `IQ` Mean IQ score of residents
- `McCainVote` Percentage of votes for John McCain in 2008 Presidential election
- `Region` Area of the country: `MW`=Midwest, `NE`=Northeast, `S`=South, or `W`=West
- `ObamaMcCain` Which 2008 Presidential candidate won state? `M`=McCain or `O`=Obama
- `Pres2008` Which 2008 Presidential candidate won state? `M`=McCain or `O`=Obama
- `Population` Number of residents (in millions)
- `EighthGradeMath` a numeric vector
- `HighSchool` Percentage of high school graduates
- `GSP` Gross State Product (dollars per capita)
- `FiveVegetables` Percentage of residents who eat at least five servings of fruits/vegetables per day
- `Smokers` Percentage of residents who smoke
- `PhysicalActivity` Percentage of residents who have competed in a physical activity in past month
- `Obese` Percentage of residents classified as obese
- `College` Percentage of residents with college degrees
- `NonWhite` Percentage of residents who are not white
- `HeavyDrinkers` Percentage of residents who drink heavily

Source

Various online sources, mostly at www.census.gov