DS 413/613 HOMEWORK WEBSCRAPING CENSUS BUREAU DATA

**1** a} Show and use a census API key that gives you access to the **ACS** data. Do not use my API key, use and show your own key.

b) Using the link provided in your notes, secure a Census Bureau API key. Run the census code that requires usage of the API key and then use R coding to produce a table that shows the totals for Asian Males for ages 67 to 69 by state for the year 2000. The identifier code is **P012D021**

c) Show and use R code to find the mean, median, ,max, min, Q1, and Q3 for the median ages.

d) Show and use R code (tidyverse/dplyr) coding to find the top ten states with highest populations of Asian Males whose ages are between 67 and 69.

**2** a) Using the link provided in your notes, use and show R coding to produce a table that shows the median ages for Hispanic or Latino women for the year 2010 (Hint: the 8 character variable code starts with characters P013. Search in your table to get the other four characters. (Hint: Ctrl F speeds up the search process)

b) Show and use R code to find the mean, median, ,max, min, Q1, and Q3 for the median ages.

c) Use ggplot coding to produce a Histogram of vertical orientation for the median ages for the table that you produced for 2a.

d) Produce a coding chunk using dplyr functions to generate a table that gives results for values that are greater than or equal to a median age of 25.

e) Use and show R code to merge your table produced in 1b with a similar table for the year 2010. Your table should have the exact columns, column names and table content shown below. The first five rows of the tibble that you are to produce is given below.

# A tibble: 52 x 5

GEOID NAME value00 value10 Differences

<chr> <chr> <dbl> <dbl> <dbl>

1 01 Alabama 118 327 209

2 02 Alaska 118 258 140

3 04 Arizona 547 1187 640

4 05 Arkansas 98 184 86

5 06 Californ~ 28524 44587 16063

**3**a) Using **ACS** census data from 2015, show and use R code to produce a tibble that shows the population total estimates and the margin of errors for white males ages 35 - 44 in the counties of California. The required variable code starts with the characters BO1OO1. Use the table to find the other characters.

The first five rows of your data table are provided below:

Text

Description automatically generated

b Use dplyr functions to change your table of part a so that it reflects estimates that are greater than $30,000 dollars and list the estimates in descending order.

c) Use and show ggplot R coding to produce a scatter plot that features x = natural log of **moe** plotted against y = natural log of **estimate.** Does your plot suggest a linear relationship between the varibles ? If so, what general trend can be inferred? (Use the full data table that you generated for part b)

d) Use and show R code that will produce the following graph for the data generated in part c

Chart, scatter chart

Description automatically generated with medium confidence