Homework 2

Homework 2 is due April 30th at 11:59 PM. All submissions should be in PDF format and be submitted through Gradescope. Note: You need to tag the individuals problems in Gradescope according to the instructions. Gradescope will allow you to highlight, for example, the response from Question 1 and then mark it as question 1.

Download anes_2016.dta which is available in Canvas. Create a folder somewhere convenient on your computer named "hw2". In that folder, put the anes_2016.dta.

Download the do-file hw2_blank.do. This do-file has a few hints to get you started. You can continue to fill in this do file to complete your homework.

The first thing to do is to change your working directory to the folder named "hw2". If your data is in the correct place, and you have correctly changed your working directory, you should now be able to type "use anes_2016.dta, clear" to load the data.

Most questions ask you to copy and paste code from your do-file and sometimes to also report the result of your code (for example, the average of a variable). A few questions will ask you to insert a screenshot into your Word document (for example, a figure). There are a few different ways to do this. One option is to simply take a screenshot and drag into the word document.

Part I: Practice in Stata

1. Generate a table of frequencies for `V161158x` (Party ID). This is a categorical variable that takes on seven values, from "Strong Democrat" to "Strong Republican." Notice that missing values are coded -9 and -8. **Paste** the code from your do-file below. (5 points)

   ```
   tab V161158x
   ```

2. Create a new variable called `partyid`. In doing this, you should go through the steps listed below:

a. Generate a new variable, `partyid`, and set it to be equal to `V161158x`. **Paste** the code from your do-file below (2 points)

> gen partyid = V161158x

b. Recode `partyid` so Democrats are coded 1, Independents are coded 2, Republicans are coded 3, and missing values are coded accordingly. Classify independents leaning towards a party as a member of that party. (***Hint: type help recode to look up the "recode" command, which is a command we did not cover in lecture. If you wanted to recode variable "var1" so that values between 1 and 5 are coded as missing values, it would look like*** `recode var1 1/5 = .)` **Paste** the code from your do-file below (5 points)

> recode partyid (1/3 = 1) (4 = 2) (5/7 = 3) (8 9 = .)

c. Label the variable "partyid" "Party Identification". **Paste** the code from your do-file below (3 points)

> label var partyid "Party Identification"

3. The variable `V161087` (Feeling Thermometer: Republican Presidential candidate). This variable describes how respondents rate Donald Trump on a scale from 0 to 100. A rating of 0 means "Very cold or unfavorable feeling". A rating of 50 indicates "no feeling at all" and a score of 100 means "very warm or favorable." Participants can respond with any number between 0 to 100, with higher numbers being more favorable to Trump.

a. Rename `V161087 to trumpft.` **Paste** the code from your do-file below (2 points)

> rename V161087 trumpft

b. Replace the value of this variable to missing (remember missing is a period in Stata) if the value of the variable is currently -99 or -88. These two codes both indicate missing information. **Paste** the code from your do-file below (5 points)

```
replace trumpft = . if trumpft == -99 | trumpft == -88
```

c. Compute the **average** value of trumpft. **Paste** the code from your do-file below **and** report the average value of trumpft (5 points)

```
mean(trumpft)
36.95744
```

4. Summary Statistics for trumpft.

a. Compute the average value of trumpft for Democrats. Paste both the code below and report the average value. (5 points)

```
mean(trumpft) if partyid == 1
14.08906
```

b. Compute the average value of trumpft for Independents. Paste both the code below and report the average value. (5 points)

```
mean(trumpft) if partyid == 2
35.58772
```

c. Compute the average value of trumpft for Republicans. Paste both the code below and report the average value. (5 points)

```
mean(trumpft) if partyid == 3
62.88495
```

d. Compute the 25th and 75th percentile of trumpft for all individuals in the dataset. Paste the code below and report the 25th and 75th percentile (5 points)

> summarize trumpft, detail
>
> 25th percentile = 0; 75th percentile = 70

5. Summary statistics for `V161267` (Respondent age).

   a. Rename `V161267` to `age`. **Paste** the code from your do-file below (2 points)

   > rename V161267 age

   b. Replace the value of age to missing if the value of age is less than zero. **Paste** the code from your do-file below (3 points)

   > replace age = . if age < 0

   c. Compute the **median** value of age. **Paste** the code from your do-file below and report the median age. (5 points)
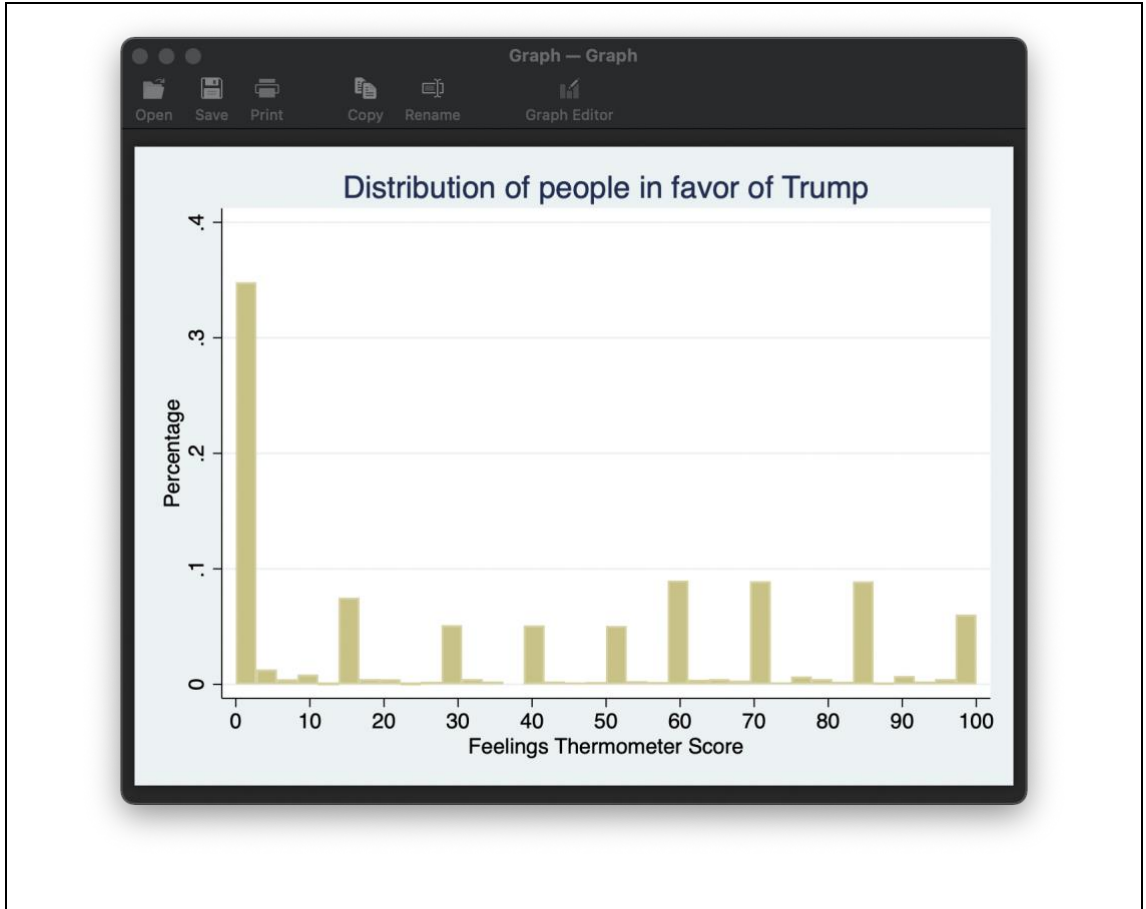
   > sum(age), detail
   > median = 50

6.

   a. Generate a histogram with the **fractions of sample** for trumpft. In the histogram, set the width of the bars to 10, and label the x axis with tick marks from 0 to 100, with one tick mark at every multiple of 10. Add titles to your histogram.

   Paste a Screenshot of the histogram in a. `trumpft` below (5 points)

b. Briefly interpret (1-3 sentences) the histogram that you created. (5 points)

The graph depicts that there is an overwhelming plurality of people surveyed who had negative feelings towards the Republican President. As we can see, approximately 35% of the people surveyed reported a score of 1-5 for the Feelings Thermometer variable.

7.

a. Generate a scatter plot with `age` on the x axis and `trumpft` on the y axis. Add titles. Paste a screenshot of your scatter plot below. (5 points)
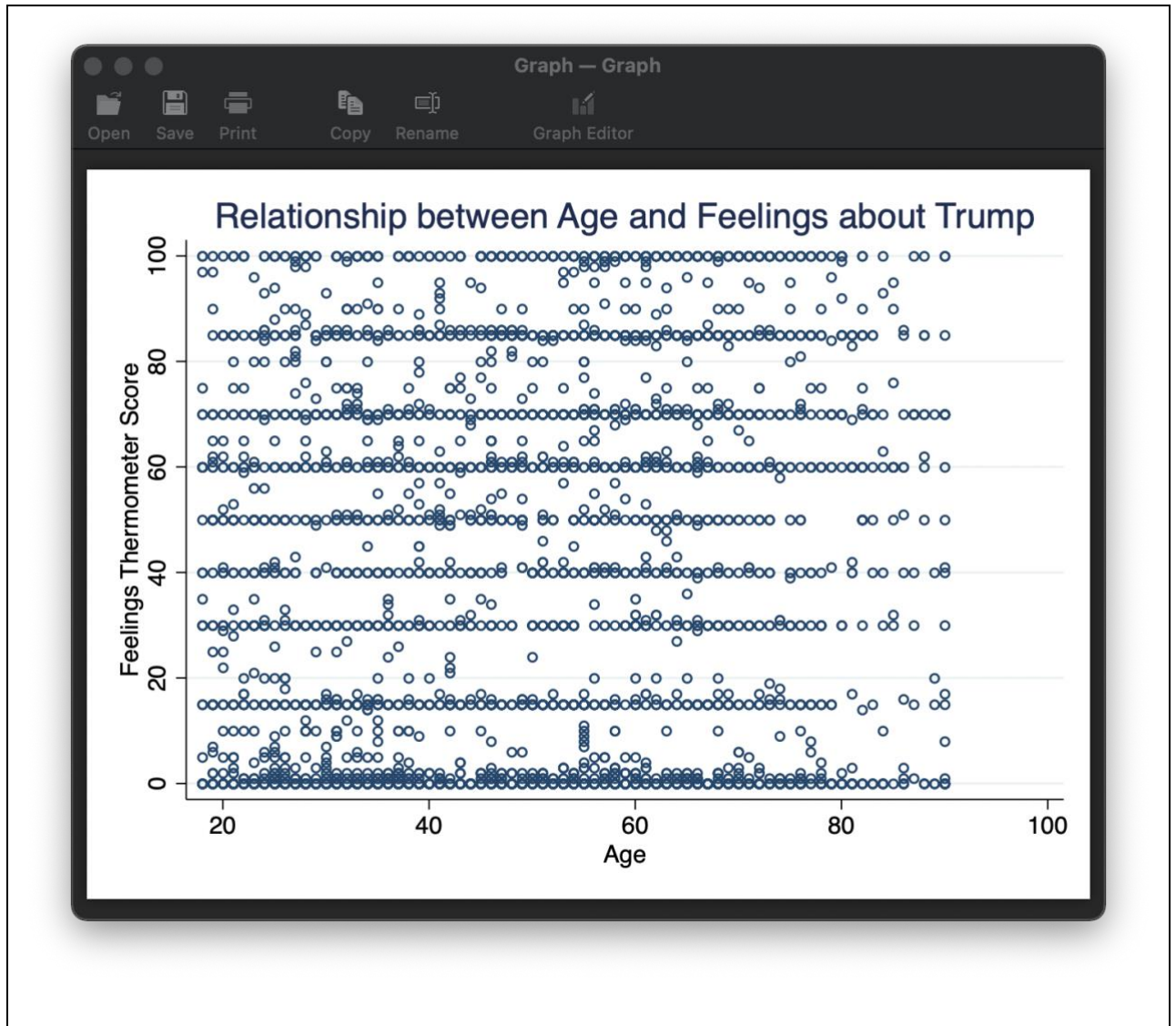
b. Briefly interpret your scatterplot (1-2 sentences). If you find it difficult to interpret, explain why. (5 points)

Given that this graph is difficult to read, there appears to be no trend in the attitudes towards President Trump based on age. This could make sense since Political Preference is not directly influenced by age, many young & old people lean Republican.

8. Reshaping the data

    a. In this dataset there are PRE (before election) and POST (after election) variables for each respondent. The variable V161086 is the feeling thermometer for the Democrat candidate before the election and V162078 is the feeling thermometer for the Democrat candidate after the election. Rename V161086 demft0 and rename the variable V162078 to demft1. Paste your code below. (2 points)

    ```
    rename V161086 demft0

    rename V162078 demft1
    ```

    b. Respondent_id is a variable that includes the unique value of a respondent. Given the data structure, is this a wide or long dataset? (4 points)

    This dataset is wide because instead of dedicating a variable to whether the observation is pre/post, it instead dedicates pre/post versions of each variable. To make this dataset long, one would create a variable that states pre or post, and this would double the number of rows in the dataset.

    c. Keep only the variables, "respondent_id", "demft0", and "demft1". Paste the code below that accomplishes this. (4 points)

    ```
    keep respondent_id demft0 demft1
    ```

    d. Reshape the data (i.e. if you think the data is currently wide, reshape it from wide to long. If you think it is currently long, reshape from long to wide. Only one of these options will actually work in this case and it is your job to figure out which.) Paste the code that reshapes the data below. (5 points)

    ```
    reshape long demft, i(respondent_id) j(ba_election)
    ```

    Note: ba_election is a dummy variable; 0 = before, 1 = after

Part II: Final Project

This last part is to prompt you to start thinking about your final project and understanding what goes into a good final project.

In the Modules tab there is now a Module named Final Exam Instructions and Materials. First, read through the Final Exam Description.

Next, select one of the three examples. You can choose based on which you think you will find the most interesting. The next questions will relate to the example that you decided to read.

1. What is the question of interest? (1 points)

    > The question of interest is, "do areas with higher rent have better health outcomes?", suggesting a positive correlation (higher rent = healthier).

2. What is the data that is used in the example you chose to read? (1 points)

    > 2016 Country Health Rankings & Roadmaps dataset to measure health outcomes
    >
    > 2016 Zillow Rent Index to measure price of rent per square foot

3. Take a screenshot of one figure in the example that you decided to read. Write a few sentences explaining what the figure shows and how it relates to the question of interest. (3 points)
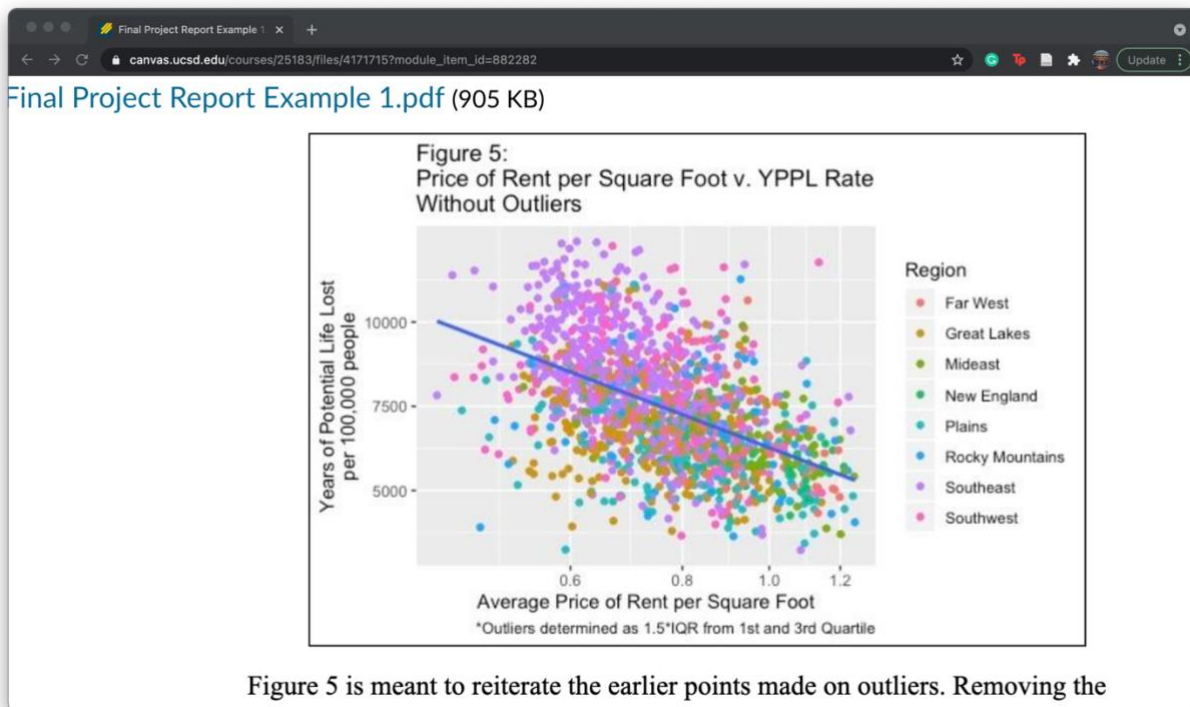
Figure 5 is meant to reiterate the earlier points made on outliers. Removing the

The graph demonstrates a negative relationship between Years of Potential Life Lost and Price of Rent per square ft. This suggests that people living in higher income areas have lower years of potential life lost. This seems in line with the author's hypothesis.

4. What is a question that you might want to explore in your final project? You will not need to commit to any question yet. This is just to get you started thinking about it and will be graded on completion. (3 points)

What is the relationship between rates of violence and level of income of areas? I'm interested in this because initially one assumes higher income areas have less violence; however, rural areas have lower incomes than urban areas and tend to experience less violence.