

Homework 2

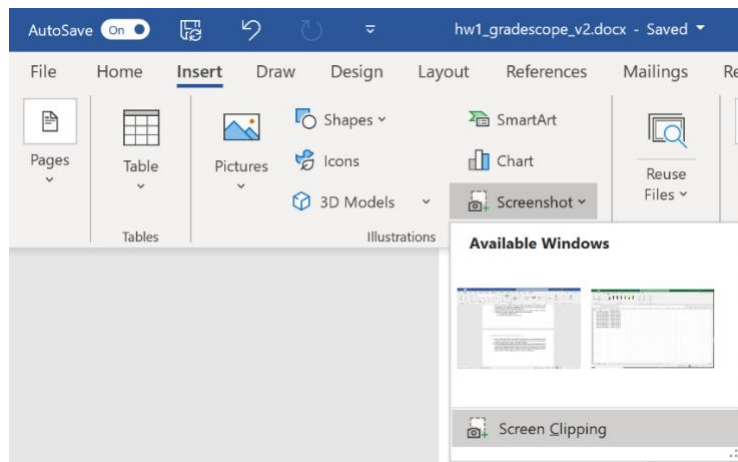
Homework 1 is due on **February 10 (Wednesday) at 11:59pm** through the Gradescope tab on Canvas. All submissions must be in PDF format. To save a Word document as a PDF, click “File > Save as Adobe PDF.”

Download the dataset `anes_timeseries_2016.dta` and the codebook, `anes_timeseries_2016_userguidecodebook.pdf` either through Canvas or the [ANES Website](#). You will be analyzing data from the American National Election Study 2016 Time Series.

Please follow the instructions closely and format your answers as requested.

To insert a screenshot into your Word document:

1. Open Stata, then go back to your Word document.
2. In your Word document, click “Insert,” then click on the “Screenshot” drop-down menu and click on “Screen Clipping.”
3. Select only the part of the Stata output you are being asked to insert, and it will show up on your Word document.



Please write down the names of the people you have worked with (if any) in the box below:

Part I: Exploring and Cleaning the Dataset [30 points]

1. Load the dataset with the `use` command in your do file. **Paste** the code from your do file below:

3 pts

2. Generate a table of frequencies for `v161158x` (Party ID). This is an ordinal 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.

2 pts

3. Create a new variable called `partyid`. In doing this, you should go through the following steps below. Paste the code for each step on the column on the right.

a) Generate a new variable, <code>partyid</code> , and set it equal to <code>v161158x</code> .		2 pts
b) Recode <code>partyid</code> 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. <i>Hint: if you wanted to recode variable “var1” so that values between 1 and 5 are coded as missing values, it would look like <code>recode var1 1/5 = .</code></i>		2 pts
c) Create a set of value labels so that 1 is labeled as “1. Democrat”, 2 is labeled as “2. Independent”, and 3 is labeled as “3. Republican”.		2 pts
d) Assign the set of values labels generated in question 4c to the variable <code>partyid</code> .		2 pts
e) After having labeled the values, it is time to label the variable. Label		

<code>partyid</code> as “Party Identification.”	
---	--

2 pts

4. Generate a table of frequencies for `V161087` (Feeling Thermometer: Republican Presidential candidate). This variable describes how respondents rate Donald Trump on a scale from 0 to 100. Paste the code.

4 pts

--

Rename `V161087` to `trumpft`.

--

5. Generate a table of frequencies for `V162110` (Feeling Thermometer: Police). This variable describes how respondents rate the Police on a scale from 0 to 100. Paste the code.

4 pts

--

Rename `V162110` to `policeft`.

--

At the bottom of the table, you will see that 998 denotes “Don’t know” and 999 denotes “Don’t recognize.” Recode `policeft` so that 998 and 999 are coded as missing values.

--

6. Generate a table of frequencies for `V161267` (Respondent age). Paste the code.

4 pts

Rename `V161267` to `age`.

7. Notice how in `trumpft`, `policeft`, and `age`, missing values are coded as negative numbers. Recode the negative numbers in these variables to missing values. Bonus point: do this using a single line of code. Paste your code below.

3 pts

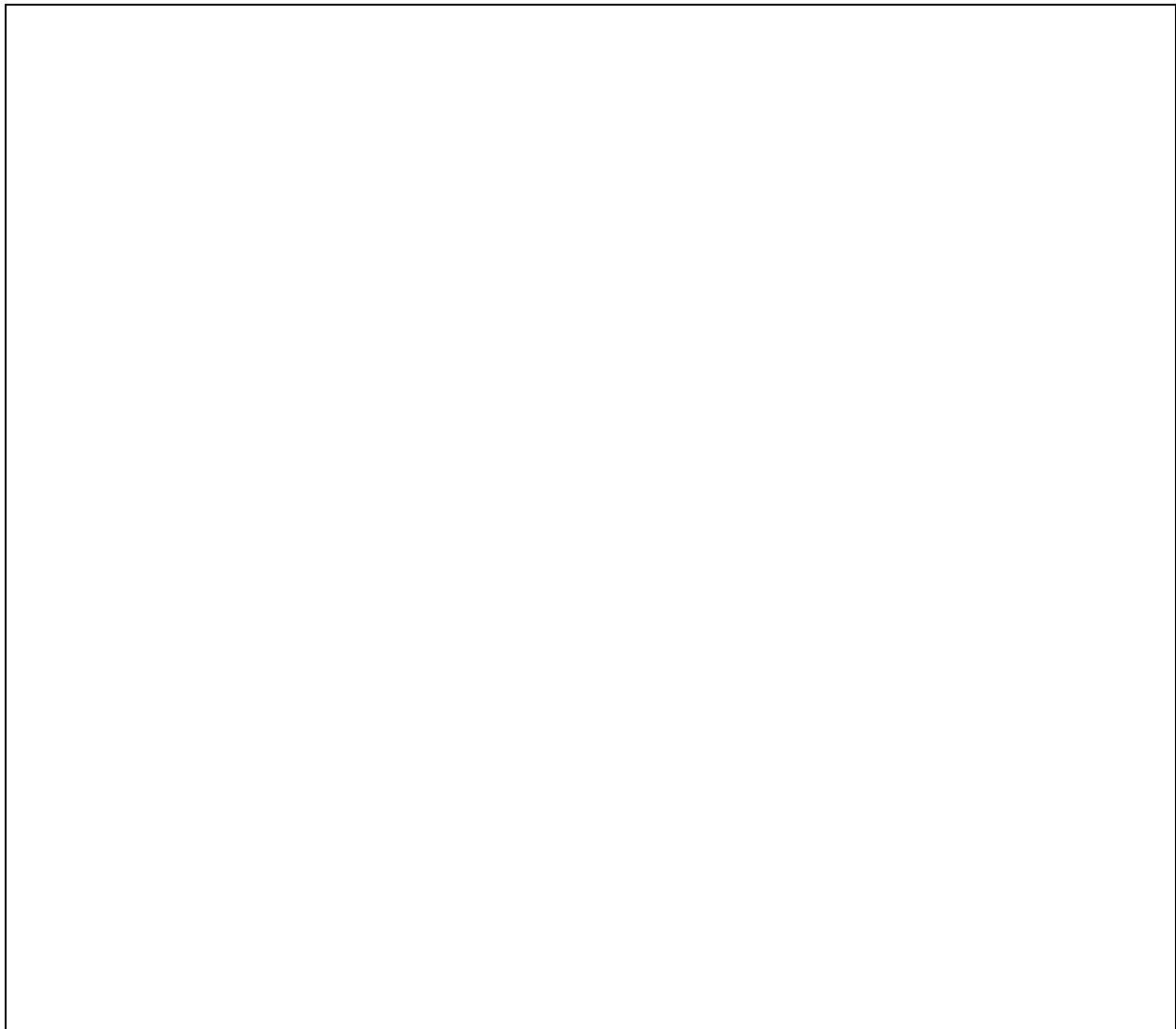
Hint: to recode multiple variables at once, try `recode var1 var2 var3 (...)`

Part II: Descriptive Statistics [35 points]

8. Generate histograms with the frequencies (not proportions) for
- a. trumpft,
 - b. policeft, and
 - c. age.

10 pts

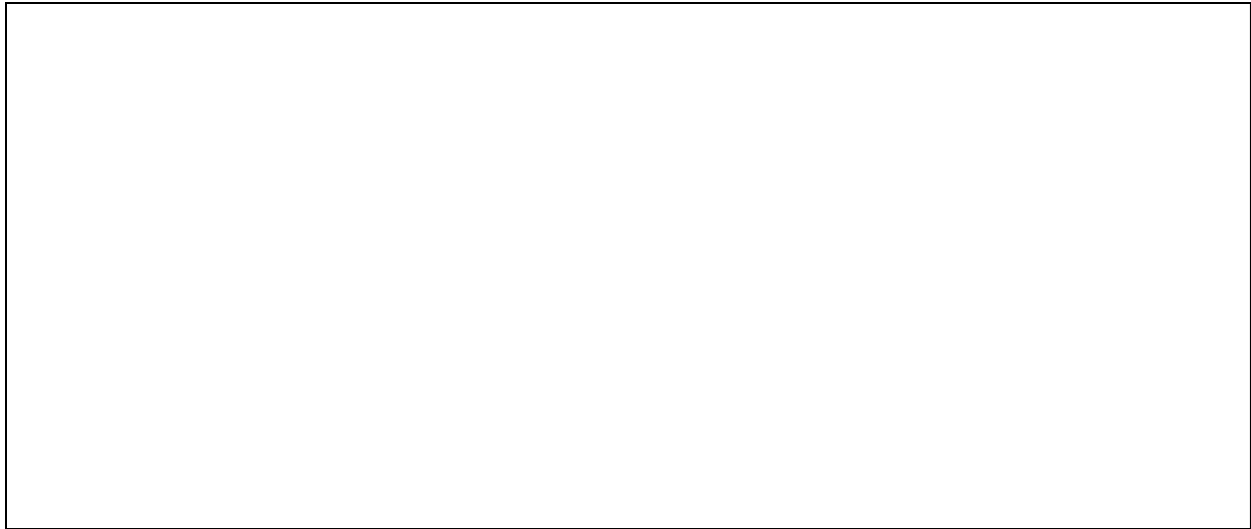
In each histogram, set the width of the bars to 10, and label the x axis with tick marks from 0 to 100, with one tick mark at every multiple of 10. Add titles to your histograms. Insert **screenshots** of your histograms below. Feel free to resize your histograms so that all of them fit in the box.



ECON/POLI 5D: Data Analytics for the Social Sciences

9. Generate a bar plot of the frequencies (not proportions) for `partyid`. Be sure to add a title to your plot. Insert a **screenshot** of your bar plot below.

8 pts



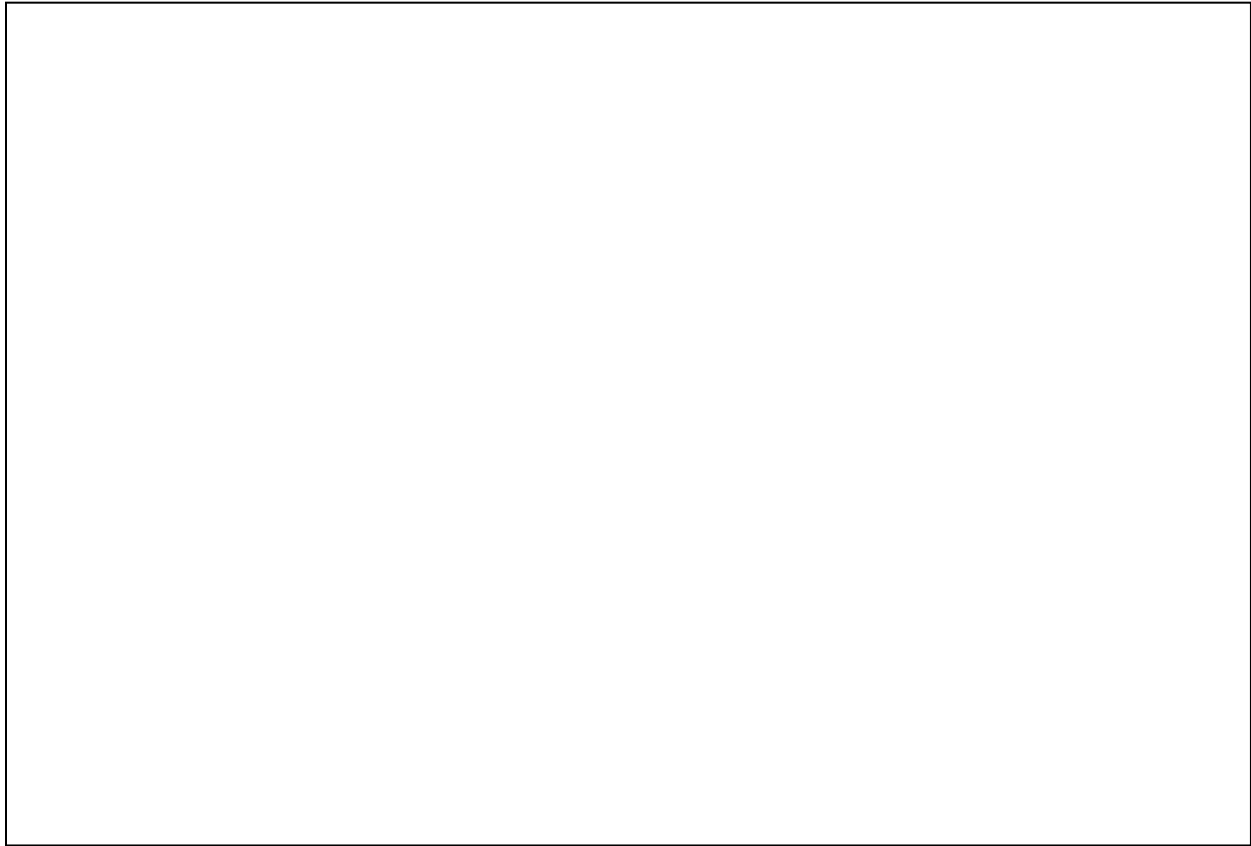
10. Now, let us compare how people with different party IDs rate Trump. Generate a bivariate bar plot with `partyid` on the x axis and the means of `trumpft` by group on the y axis. Add a title to your plot. Insert a **screenshot** below.

8 pts



11. Generate a scatter plot with **age** on the x axis and **policeft** on the y axis. Add a regression line (line of best fit) and a title. Insert a **screenshot** below.

9 pts



Judging from the scatter plot and the regression line, do older people tend to view the police more favorably or less favorably than younger people? Type your answer below.



Part III: Regression Analysis [35 points]

Because the ANES is a survey, we have to tell Stata how it was designed. This includes letting Stata know the weight of each observation. Declare the survey design using the following command:

```
svyset [pweight=v160101], strata(v160201) psu(v160202)
```

12. Test the hypothesis


“Older people are more likely to view the Police favorably than younger people.”

using regression analysis. Before running the regression:

a) What is the dependent variable?		4 pts
b) What is the independent variable?		4 pts
c) What is the null hypothesis?		4 pts

13. Run a regression to test the hypothesis that older people are more likely to view the Police favorably than younger people. Hint: because it is a survey put “svy:” before the regression command.

a. Insert a screenshot of the regression results below:



5 pts

- b. According to the regression results, what is the predicted Police Feeling Thermometer of a respondent who is 20 years old? What is the predicted Police Feeling Thermometer of a 40 year old?

5 pts

- c. Judging from the 95% confidence intervals for the slope estimate, can we reject the null hypothesis at a 95% confidence level?

4 pts

- d. What can we conclude about the relationship between age and views of the Police in the American public? Does the evidence from the regression results support the hypothesis being tested?

5 pts

- e. What percentage of the variation in the Police Feeling Thermometer is explained by age?

4 pts

Part IV: Final Project [Ungraded]

This last part is meant to prompt you to start thinking about your final project. **Although these tasks and questions will not be graded, we expect you to do your best at answering them.** Before answering:

- Visit the Google Drive folder [ECON/POLI 5D – Datasets](#), and read carefully through the [Datasets Summary](#) spreadsheet.
- Based on your reading of the [Datasets Summary](#) spreadsheet, select one dataset that seems interesting to you. (Will not have to stick to this dataset for your final project).
- Skim the codebook of the dataset of your choice and think about the variables included in the dataset (if you choose the NBA dataset, open the spreadsheet on Excel and skim through the variable names).

14. Based on what you have seen, what would be an interesting research question that could be answered using this dataset? (You will not have to stick to this question for your final project).

15. Download the dataset of your choice and load it on Stata. Choose at least two variables to describe using `summarize`, `tabulate`, or both—whatever makes the most sense to you. Insert screenshots of the results below (if you use `tabulate` and end up with a long table, your screenshot does not have to include the whole thing).