Write a Short Description of the Data you Chose and Why:

When I initially read the description for this assignment, I got excited about picking the data points to analyze because most of my previous projects have included pre-chosen datasets. My mind immediately went to political data because politics interests me and I know traditionally there are a lot of surveys and studies done with political views and election results. Therefore, I filtered the data by politics. When scrolling through the potential variables I could analyze, I wanted to get a somewhat wide range of topics. In fact, my data has no numeric variables. But, I believe that my topics all paint a large cohesive picture of the political nature of the country. The topics that I decided to pick for this lab were political party affiliation (partyid), does the respondent follow public affairs (civic), how much influence the respondent feels they have on local government (locinflu), how interested the respondent is in politics (intpol), whether or not the respondent has given money to a candidate or political cause (gavepol), and whether or not the respondent believes the public has little control over politics (poleff1).

Now, moving into analyzing the data and potential answers respondents gave; all of the variables that I decided to analyze had multiple answer options. First, looking at political party affiliation, the specific question posed was "Generally speaking, do you usually think of yourself as a Republican, Democrat, Independent, or what?". The answers provided ranged from no answer, do not know/can not choose, strong democrat, not very strong democrat, independent close to democrat, independent, independent close to republican, not very strong republican, strong republican, to other party. From this variable, there are 71,905 valid cases and 485 missing cases. Next, when looking at public affairs, the question posed was "Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?" Along with the answer options provided in the question, some people answered with "no answer", and "do not know/can not choose". From this variable, there are 1,839 valid cases and 67,007 missing cases. Moving onto how much respondents feel like citizens have influence over local government decisions; the question posed was "How much influence do you think people like you can have over local government decisions -- a lot, a moderate amount, a little, or none at all?" Additionally, respondents could answer with "no answer" or "do not know/can not choose". There were 1,779 valid cases and 67,067 missing cases for this variable. I then also wanted to analyze how much people are interested in politics to compare the previous answers to. The question posed here was "How interested are you in politics and national affairs? Are you very interested, somewhat interested, only slightly interested, or not at all interested?" As is consistent with the other questions, respondents could answer with "no answer" and "do not know/can not choose" and from this dataset there are 1,805 valid cases and 67,041 missing cases. Next, I also wanted to see who donated money to a political candidate or cause. Those respondents were asked "In the past three or four years, have you contributed money to a

political party or candidate or to any other political cause?" This was a simple Yes or No question with some respondents also claiming "no answer". This variable has 1,807 valid cases and 67,039 missing cases. The last variables I picked analyzed whether respondents believed the public had control over politics and they were asked "Please indicate whether you agree or disagree with [the following] statement. The public has little control over what politicians do in office." Respondents answered with agree, disagree, no answer, and do not know/can not choose. This final variable had 636 valid cases and 68,210 missing cases.

Describe your Findings in 1-2 pages:

To begin, I am going to describe the cleaning processes I decided to use for EDA. First, I checked for missing variables and dropped rows that did not have data for all of the columns. Ultimately, no rows were dropped. Next, the data for inapplicable has a ".i" in front of the word, so I removed that and turned the variable into Nan with numpy. I then renamed labels within the variable columns to make the data easier to manage and analyze, along with renaming the columns so that viewers could more easily understand what each column's data was about.

Then, I moved onto EDA visualizations and numeric data. I decided to create histograms for each of my variables in order to visualize the data in a neat manner. It is important to note that for Follows Politics, Local Influence, Interest In Politics, Donated To Politics, and Public Control Over Politics, there were a lot of NaN variables as a result of the political party variable collecting data decades prior to and after the other variables. This does make the graphs for those variables look a tad ugly. I tried to dropna to make it look cleaner, but for some reason that never quite worked, so I decided to include value counts at the bottom for a numeric visualization of the distribution of data. Furthermore, I was able to create a pie chart for political parties, with the help of the included AI feature which was cool because I haven't made a pie chart yet on collab. I believe that this cart simply added another level of visualization of the data and really shows the different distributions of the political party people see themselves in. I also created a kernel density plot with this data to see how political party classification changed throughout the years that I had data for and I think it is fascinating. Looking first at the democrat line, in 1970, democrats had the highest number of followers, dipped down some but then went back up in 1990, all while continuing to remain the party with the most followers. Then, in 1990, the republican line came up and surpassed the democrat line, and the two essentially flipped. Around 2005, there was a sharp decline in the number of democrats and republicans reported, while the number of independents did not decrease by as much, so it then became the party with the most respondents until the end of the data. This surprises me because in this day and age, everyone seems very polarized and I believed that more people were picking a party rather than choosing independent, but this data contrasts that. The other variable responses mostly lined up with my predictions for them, with people mostly following politics, believing that they have some sort of degree of influence over local politics, people being mostly interested in politics, a larger number of people who have not donated to political organizations vs. those that have, have more people agreeing that there is public control over politics.

Towards the bottom of the file, I decided to try and cross tabulate the data. Specifically, I wanted to see what political party people were associated with when answering the other questions related to the variables. These numbers portray some of the general actions of members of the democratic party, republican party, and independent party while also noting that everyone does act differently in the political world even if they believe the same things. Not all Republicans

believe that they have a little bit of influence over local politics while some news outlets today claim that republicans think they can heavily affect politics. Furthermore, it is important to note that of those who answered, a greater percentage of republicans donated money to political organizations than democrats or independents which is also noteworthy. In this last presidential election, the democratic candidate received more money in donations than the republican candidate. Overall, there were a lot of No Answers and NaN variables so there are no definite conclusions that can be drawn from this data, but it is interesting to look at and visualize to find trends in the data.