Daniel Chalhon

Relationship Between Income and Political Affiliation

The American political landscape is experiencing a sharp divide in party affiliation, with more people divided than ever before. According to one article from Pew Research, "across 10 political values Pew Research Center has tracked since 1994, there is now an average 36-percentage-point gap between Republicans and Republican-leaning independents and Democrats and Democratic leaners. In 1994, it was only 15 points" (Doherty 2017). One particularly interesting phenomenon that warrants exploring, is what shapes the views which cause individuals to lean or side with a certain political party. Often it is said that in America, Republicans are the party of the rich and Democrats are the party of the poor. "The caricature, common in the 1930s, is that the GOP is the party of the plutocrat in the Monopoly game, complete with top hat and tails. A Pew Research poll found 62 percent of Americans believe the Republicans favor the rich" (Barone 2017). This survey from Pew Research is further solidified by today's mainstream Republican policies which for example include tax cuts for the wealthy and fewer regulations for business, as recently seen. On the opposite side, it seems likely Americans would view Democrats as the party of the poor and middle class as their policies on higher taxes for the wealthy and income inequality would be more favorable to the aforementioned groups.

Acknowledging such commonplace public perceptions, this paper intends to explore the relationship between family income and one's political ideology. The choice of studying the variable of political ideology rather than party affiliation was primarily decided by the idea that party allegiance may very easily shift, and change based on a variety of factors (party leadership,

single issues such as immigration, etc.) while ones underlying ideology may be a more reliant indicator of their political views and long-term voting trends.

Considering the traditional values of modern conservatism from the Raegan era to present day one may associate such ideology with pro-business legislation, lower taxation, and a reduction in social spending. Conversely, one may associate the ideology of modern liberalism (Clinton era to present day) with increased social spending, increased business regulation, income inequality, and relatively higher taxation for wealthier Americans. Therefore, holding all other factors constant such as gender, race, life experience, geographical area and upbringing, the hypothesis is that based on the ideas each of these ideologies, as incomes rise individuals should become more conservative and as incomes lower individuals should become more liberal.

The database used was the Pew Research April 17 Political Survey. The survey was made up of adults age 18 and older living in the United States while being conducted by telephone. Of a nationally representative sample of 1,502 adults, 375 surveys were recorded by landline and 1,126 by cell phone. The interviews were administered in both English and Spanish. The sample population data was weighted (mathematically manipulated to more accurately represent the whole population) in two stages. The first stage involved compensation weights that adjusted for non-response bias (e.g. only those with extreme views would answer the phone or those with certain views would not want to answer) in addition to different probabilities of adult selection within each household (e.g. is the husband or wife answering the phone). The second stage of weighing involved balancing the sample population demographics to match population parameters, which is known as post-stratification weights. This type of weight is given to specific age-groups, genders, religions and other identifying categories that make the sample population much closer to the general population as a whole (April 17 Political Survey).

The specific variables measured can be found under the names “income” and “ideo”, representing Americans 2016 family income and current political ideology, respectively. Below is a table describing how both variables were coded.

** Responses of 10 and 9 respectively have been excluded from further analysis for accuracy purposes **

Income	**Ideo**
“Last year, that is in 2016, what was your total family income from all sources, before taxes? Just stop me when I get to the right category.”	*“In general, would you describe your political views as…”*
1 = Less than $10,000	*1 = Very Conservative*
2 = $10,000 to under $20,000	*2 = Conservative*
3 = $20,000 to under $30,000	*3 = Moderate*
4 = 30,000 to under $40,000	*4 = Liberal*
5 = 40,000 to under $50,000	*5 = Very Liberal*
6 = 50,000 to under $60,000	*9 = [Do Not Read] Don’t know / Refused*
7 = $75,000 to under $100,000	
8 = $100,000 to under $150,000	
9 = $150,000 or more	
10 = [Do Not Read] Don’t Know / Refused	

Analyzing the way the variables were coded reveals that ideo (political ideology) is a nominal variable while income (total 2016 family income) is an interval variable. This means that because the variable ideo has no specific way to order categories (can be logically arranged in any way) and has no measurable difference between categories, its variable type is considered to be nominal. The income variable on the other hand, is considered an interval variable as the categories or choices can be placed in a logical order and can be exactly measured within or between categories. Typically, income data would be considered to be ratio data (all properties of interval

data but has clear definitions of being exact), however, due to the nature of the way this variable was coded (1-9 instead of actual numbers), it should be considered as interval rather than ratio.

One way to begin drawing meaning from the data of variables is to inspect their summary statistics. Below is a table of summary statistics that describe each variable.

	Income	*Ideo*
Min.	*1*	*1*
1st Quartile Range	*3*	*2*
Median	*6*	*3*
3rd Quartile Range	*8*	*4*
Max	*9*	*5*
Standard Deviation	*2.510564*	*1.070763*
Mean	*5.518*	*2.891*
Number of Responses	*1395*	*1465*
Missing (NA)	*106*	*36*

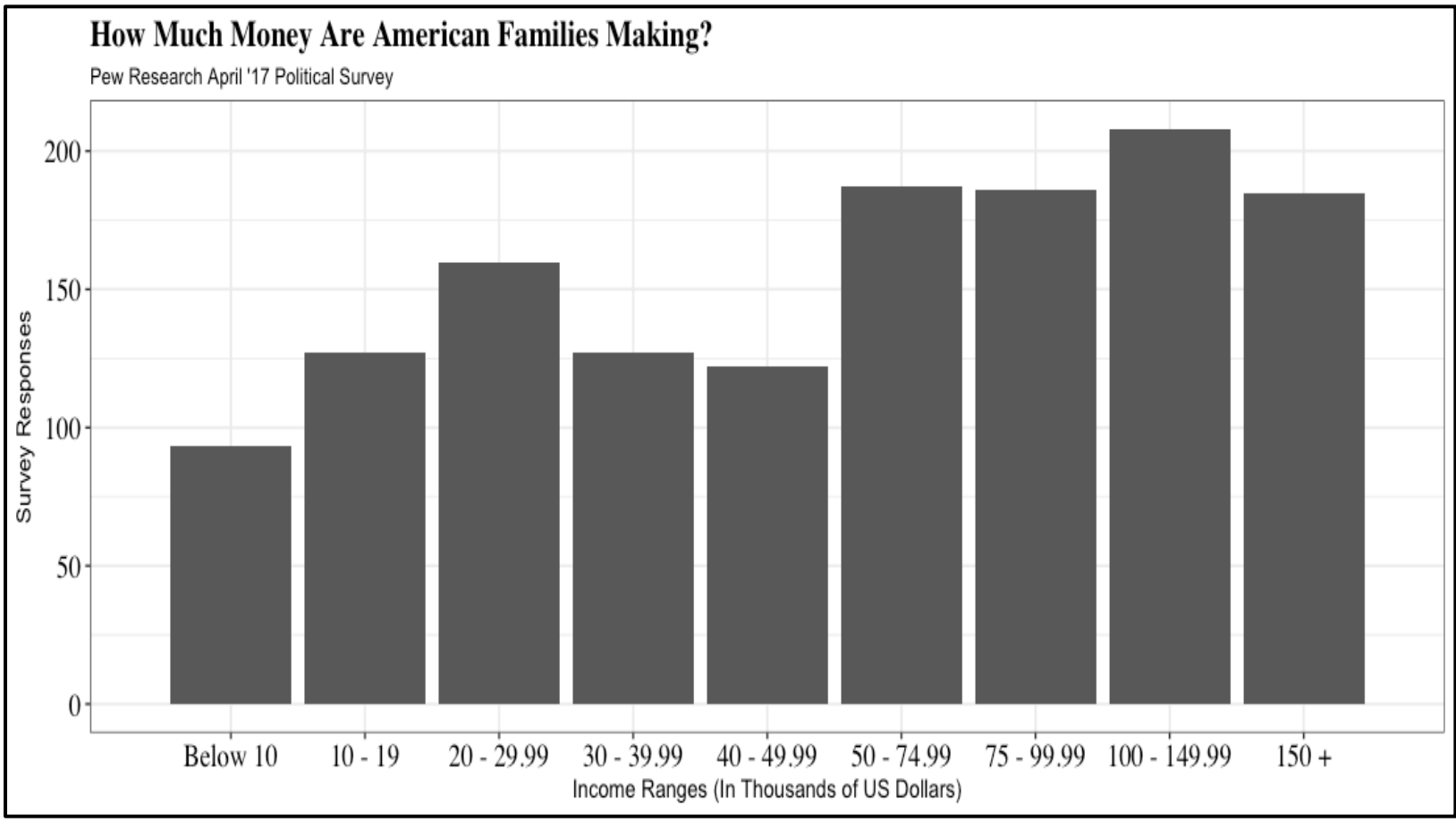
First looking at the variable income, we can see a median of 6 and a mean of 5.518. Knowing that the variable is coded from 1 to 9, a median of 6 would demonstrate that the data may be slightly skewed towards the right (meaning there is a greater representation of data on the right side of the graph), or in this case, higher incomes. In a perfectly normal distribution (data that forms a symmetrical bell-shaped curve), where the same variable was used, one would see a median of 5 which indicates there may be a skew. Loosely speaking, while a mean smaller than the median (5.518 vs 6) may suggest a potential leftward skew (more lower-income data), the current mean (5.518) is still greater than what would be the median in a perfectly normal distribution (5), which further supports the evidence of a rightward (higher) income skew.

The standard deviation, or how spread out the data is over the curve, is 2.510564. A low standard deviation signals that the data is more closely clustered around the mean while a high standard deviation may suggest that the data is dispersed farther away from the mean. This may denote that the data in the income variable is relatively distributed across options 1 to 9 with some potential clustering of data.

Lastly, quartile ranges can also help shed light on a variables data distribution. Quartile ranges split data into four categories based on how the numbers fall in relation to each other and are the last number in its range. For example, a 1st quartile range would consist of the last number in the first 25 percent of the data distribution, 2nd quartile range would be the median, 3rd quartile range would be the greatest or last number within 75 percent of data, and the 4th quartile range would be the max number within the data. Having quartile ranges of 3, 6, 8 and 9, one can see that the data may be skewed to the right based on a relatively high first quartile range and third quarter range. To further explain, if the last number in the first 25 percent of the data (first quartile range) is 3 with options from 1 to 9, it's possible that there is less data on the left (smaller numbers) which is boosting that quartile range. The same principle applies as well to the third quartile range.

Moving onto the ideo (ideology) variable, the mean is 2.891 with a median of 3. Since the mean is so close the median, this would indicate that the data has a slight left skew with a large concentration of data in the center. The standard deviation is 1.070763 signaling that the data is very closely clustered around the mean, and the quartile ranges are 2, 3, 4 and 5 which paints the picture of a relatively normal distribution. Ultimately, using descriptive statistics can only explain so much of the way data is distributed. Looking at a visual representation along with descriptive statistics are a much more effective way of analyzing data.

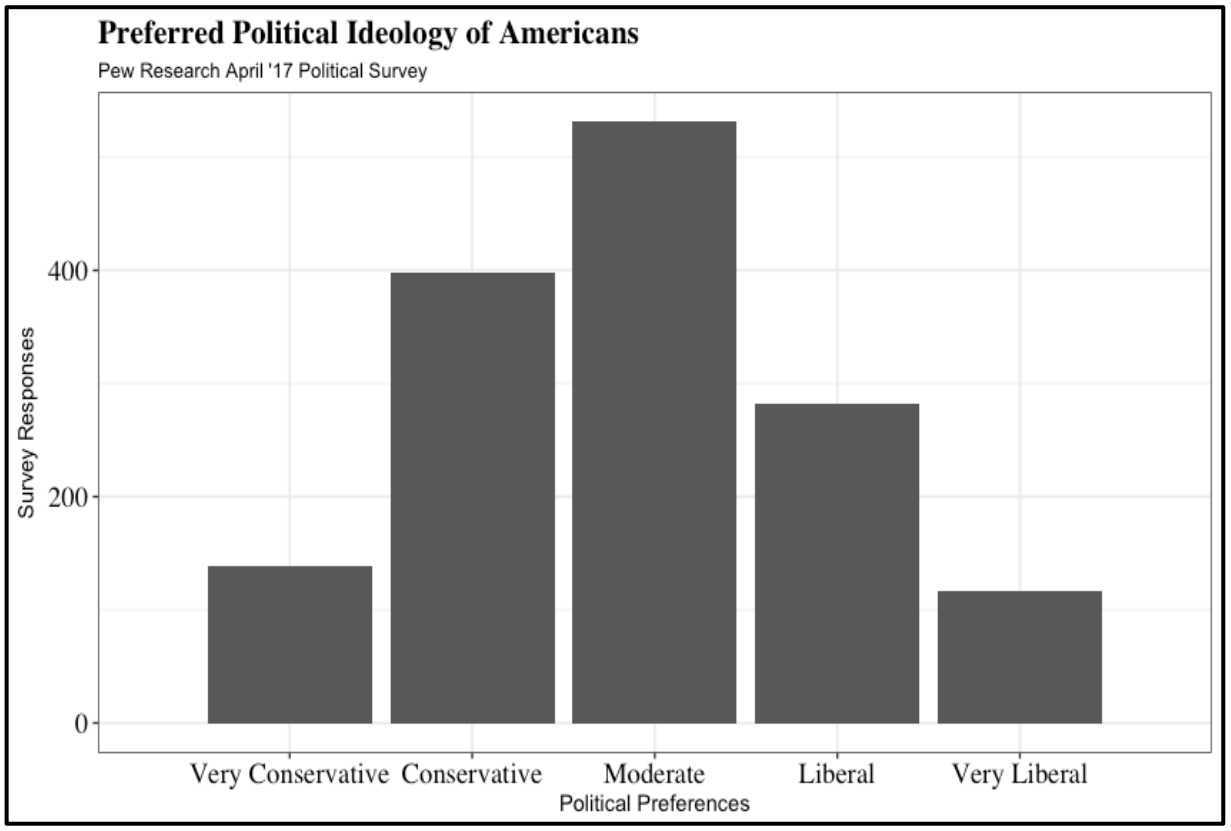
The bar graph below serves as a visual representation of the income variable.



Looking at the graph above one may notice the rightward skew of data, reaffirming the hypothesis drawn from the summary statistics. Furthermore, there seems to be a significant overrepresentation of higher income families as the actual median family income in the United States is $59,039 dollars (Davidson 2017). With a normal distribution, one would see the data much less conglomerated on the right and more even with the left side. The category with the most surveyed individuals has family incomes between $100,000 and $149,999 followed second by $50,000 – $74,999 (the median category). However, the next two highest categories revert back to higher income families with incomes of $75,000 – $99,999 and greater than $150,000. Based on this information, if the current hypothesis is current, a graph of political ideology should display higher levels of conservatism based on the higher representation of high family income individuals

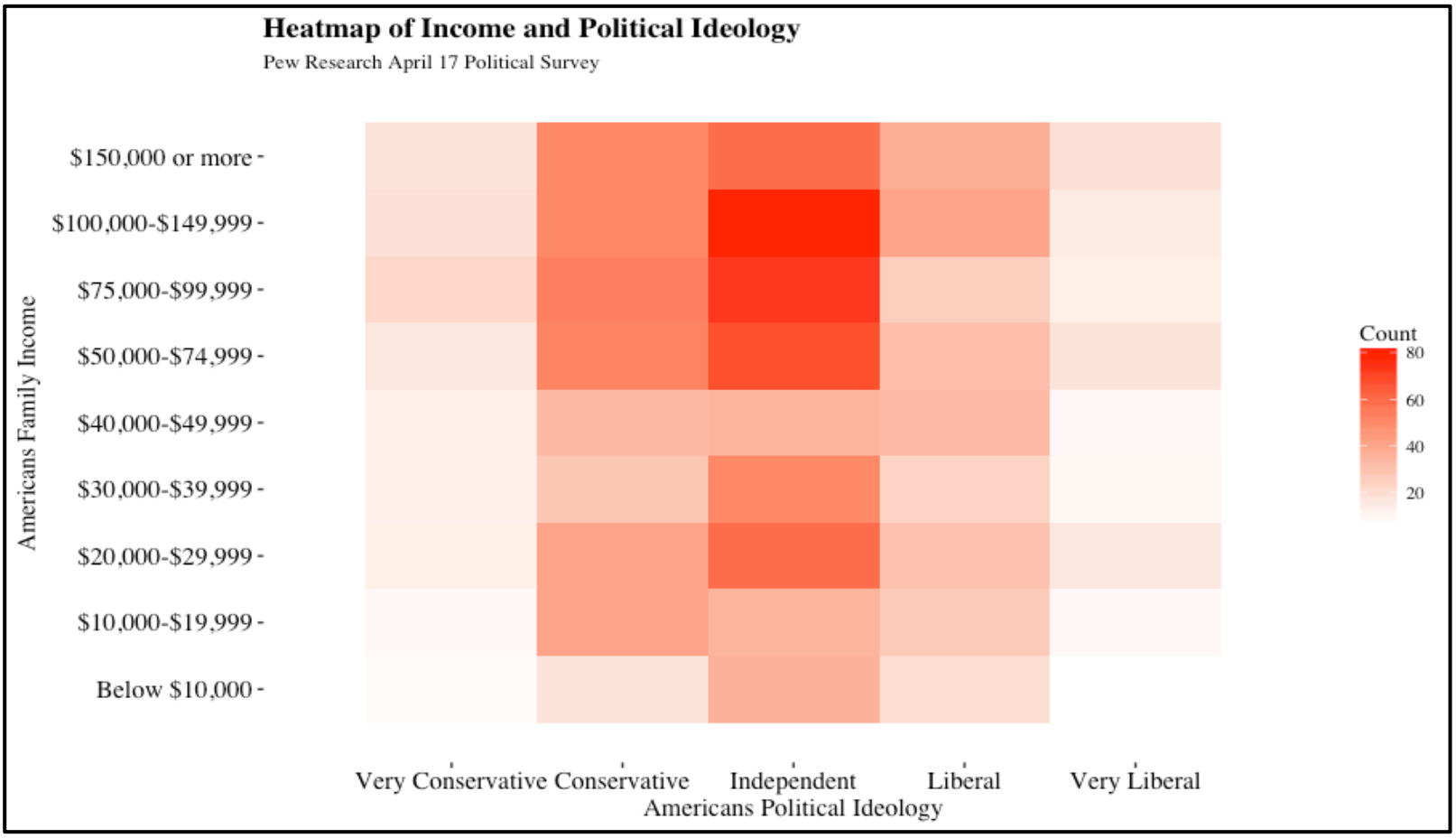
as well as lower levels of liberalism based on a smaller representation of lower family income individuals.

Towards the left is a graph describing the variable ideo (ideology).



As the summary statistics predicted, there is a large amount of data clustered around the mean and a leftward skew with a relatively normal distribution. In this graph, it seems most people identify as moderate with the next highest category being conservative, and then liberal. This may suggest some potential correlation between income and ideology as the survey overrepresented higher income families and seems to also have a higher number of conservatives. However, those who identified as very conservative tends to be relatively low in relation to the other categories. This could signal that there is a limit to how much wealth correlates with conservatism or that the link between the two variables may not be as clear as believed. A better representation of this relationship would be a graph that displays both variables.

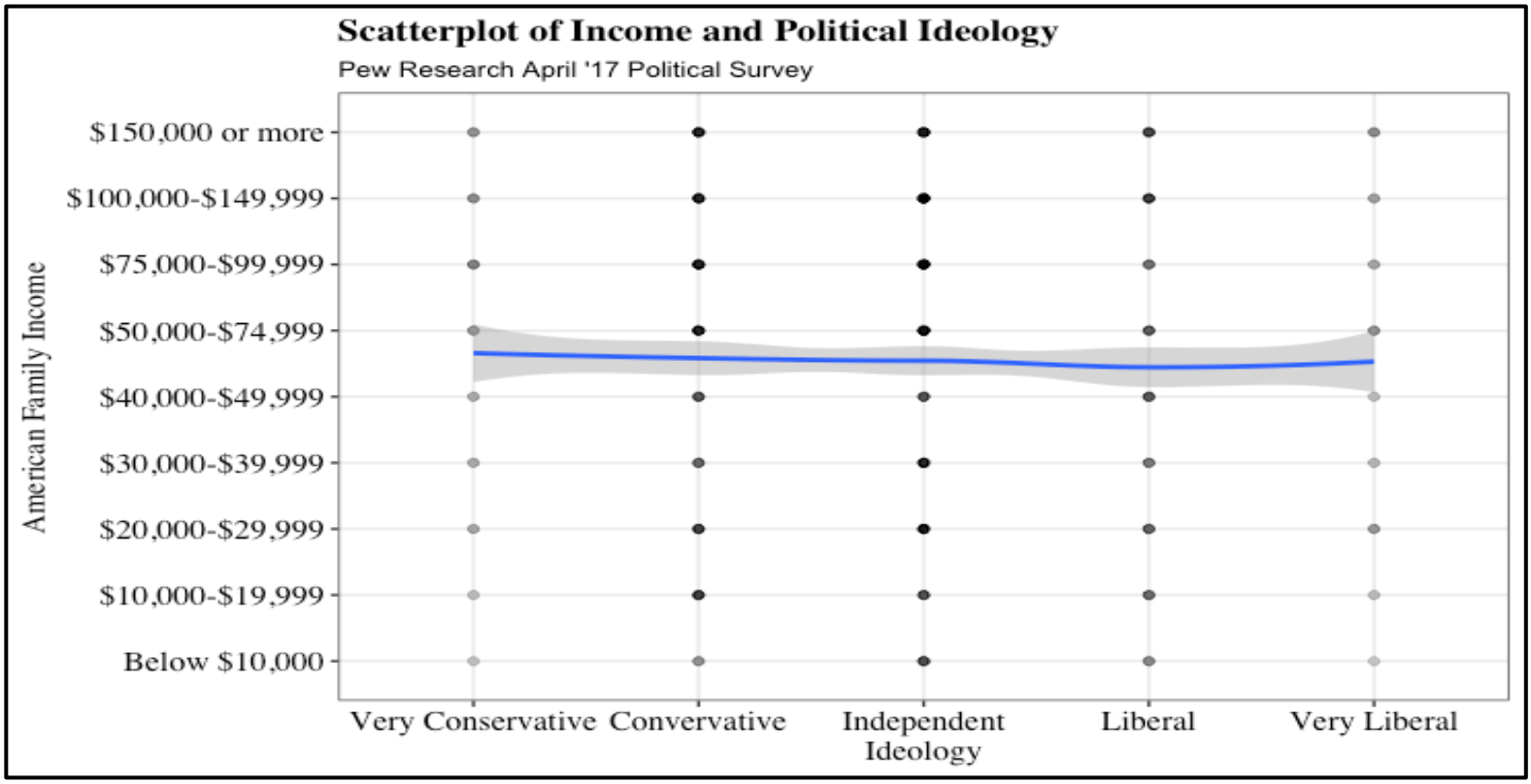
Below is a heat map showing the relationship between variables income and ideo.



Looking at the heat map, we can see reaffirmations of what was learned from previous graphs but also notice a few differences. As displayed, independent is the modal category and the predominate ideology for each income range besides $10,000 – 19,999. On both extremes of the spectrum, we rarely see individuals across all income ranges who identify as very liberal or very conservative. Furthermore, one may notice that more survey respondents identified as conservative than liberal. This effect seems to be more pronounced above the higher income ranges as the trend of a greater number of conservatives seems to appear at the $50,000 – 74,999 category, which is the median. Below this category, the difference between those who identify as liberals and conservatives seems to become smaller. This may indicate some sort of relationship whereas family income rises, individuals may lean slightly towards becoming more conservative in their

ideology. However, it is also possible that the survey has overrepresented conservative individuals and this result is just the product of sampling error. It is difficult to draw a definitive conclusion from the graphs analyzed so far.

Below is a scatterplot depicting variables income and ideo with a trend line.



Another useful way of identifying relationships between variables is to use a scatterplot with a trend line. The scatterplot enables one to spot certain patterns and trends occurring and can be more easily identified with a trend line. Unfortunately, it is difficult to visualize variables (ordinal and interval) that fall on a small scale such as these on a scatterplot. This type of graph tends to work best with continuous or ratio variables that can have large variations. The darkness of each point on the plot has been adjusted to reflect its weight in the plot (or lack thereof) meaning that more dense areas will be darker and less dense areas will be lighted. This, however, reveals similar conclusions to the previous heatmap. Fortunately, a trend line can help reveal further

information about both variables. Generally, the slope of a trend line will indicate the correlation between two variables. Therefore, a positive correlation would have an upward sloping line while a negative correlation would have a downward sloping line. Noticing that variables income and ideo have a horizontal trend line with little slope, this indicates that there is likely none or very little correlation between income range and political ideology.

Initially, this paper hypothesized that as income ranges rose, individuals would become more conservative. With the overrepresentation of high-income families within the survey, had that hypothesis been true, one should see a significant overrepresentation of conservatives. Looking at a heat map of both variables, there seemed that there may have been a slight correlation between the two, but the evidence was highly inconclusive and questionable. By analyzing a scatterplot with a trend line, this paper was able to conclude that there is no relationship between family income and political ideology.

Works Cited

Barone, Michael. "Which Party Is the Party of the 1 Percent?" Capital Research Center, Capital Research Center, 10 Jan. 2017, capitalresearch.org/article/party-one-percent/.

Davidson, Paul. "Median Household Income Hits $59,039, Rising for 2nd Straight Year." *USA Today*, Gannett Satellite Information Network, 13 Sept. 2017, www.usatoday.com/story/money/2017/09/12/household-incomes-rise-3-2-2016-2nd-straight-increase/655310001/.

Doherty, Carroll. "Key Takeaways on Americans' Growing Partisan Divide over Political Values." Pew Research Center, Pew Research Center, 5 Oct. 2017, www.pewresearch.org/fact-tank/2017/10/05/takeaways-on-americans-growing-partisan-divide-over-political-values/.

"April 17 Political Survey." Pew Research Center, Washington, D.C. (April 11, 2017) http://www.people-press.org/datasets/.