Problem Set #1

MACS 30200
Lerong Wang

1 Part 1

Problem 1

The U.S. current population survey has been conducted by the Census Bureau for over 60 years for the Bureau of Labor Statistics as one of the major source of data on labor force. The monthly survey addresses basic issues such as labor force status, employment, job search, occupation with supplemental questions on a variety of other subjects such as education, migration, poverty, etc. The data can be accessed from the website of the United States Census Bureau or the National Bureau of Economic Research.

Problem 2

The data has been cited or used as the key data source in many research papers. For Example:

- Hirsch, Barry T., and David A. Macpherson. Union Membership and Coverage Database from the Current Population Survey: Note. SSRN Electronic Journal, 2003, doi:10.2139/ssrn.367781
- Shopland, D. R., et al. Cigarette Smoking Among U.S. Adults by State and Region: Estimates From the Current Population Survey. JNCI Journal of the National Cancer Institute, vol. 88, no. 23, Apr. 1996, pp. 17481758., doi:10.1093/jnci/88.23.1748

Problem 3

The current population survey is conducted monthly with a sample size of about 60000 U.S. households. A housing unit in the CPS is interviewed for four months and then dropped out of the sample for the next eight months and is brought back in the following four month.

Problem 4

A table that gives descriptive statistics for at least 8 key variables

	huprscnt	hrnumhou	hrmis	hrmonth	pulineno	puhroff2	hwhhwtln	hubusl1
count	148054.000000	148054.000000	148054.000000	148054.0	148054.000000	148054.000000	148054.000000	148054.000000
mean	0.633857	2.858498	4.523343	6.0	1.692997	-0.543113	1.035352	-0.760831
std	0.946577	1.873351	2.287926	0.0	1.685767	3.232578	0.985392	0.745734
min	0.000000	0.000000	1.000000	6.0	-1.000000	-3.000000	-1.000000	-3.000000
25%	0.000000	2.000000	3.000000	6.0	1.000000	-1.000000	1.000000	-1.000000
50%	0.000000	3.000000	5.000000	6.0	2.000000	-1.000000	1.000000	-1.000000
75%	1.000000	4.000000	7.000000	6.0	3.000000	-1.000000	2.000000	-1.000000
max	9.000000	15.000000	8.000000	6.0	16.000000	80.000000	9.000000	7.000000

Problem 5

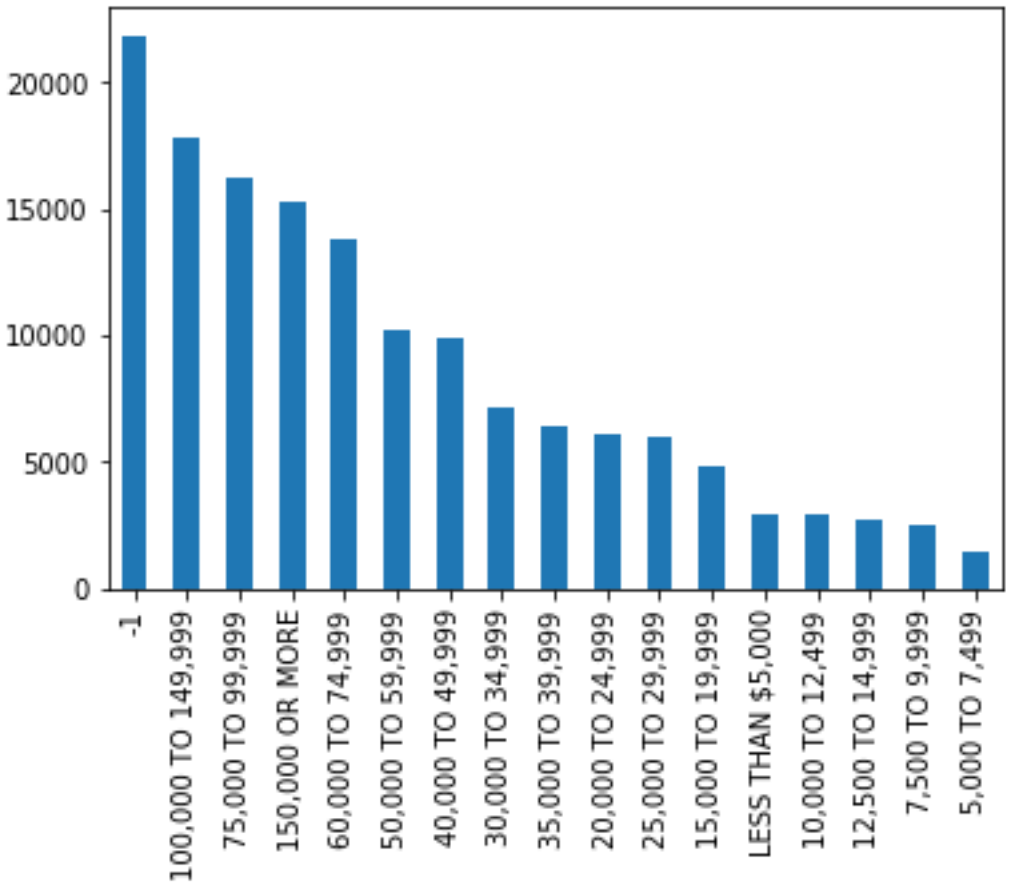

Figure 1: family income bar plot

This is a plot about variable "hefaminc", which is the family income. As we can see from the result, if we eliminate -1, which means no responses, 100000 to 149999 range has most respondents, and 5000 to 7499 range has the least.

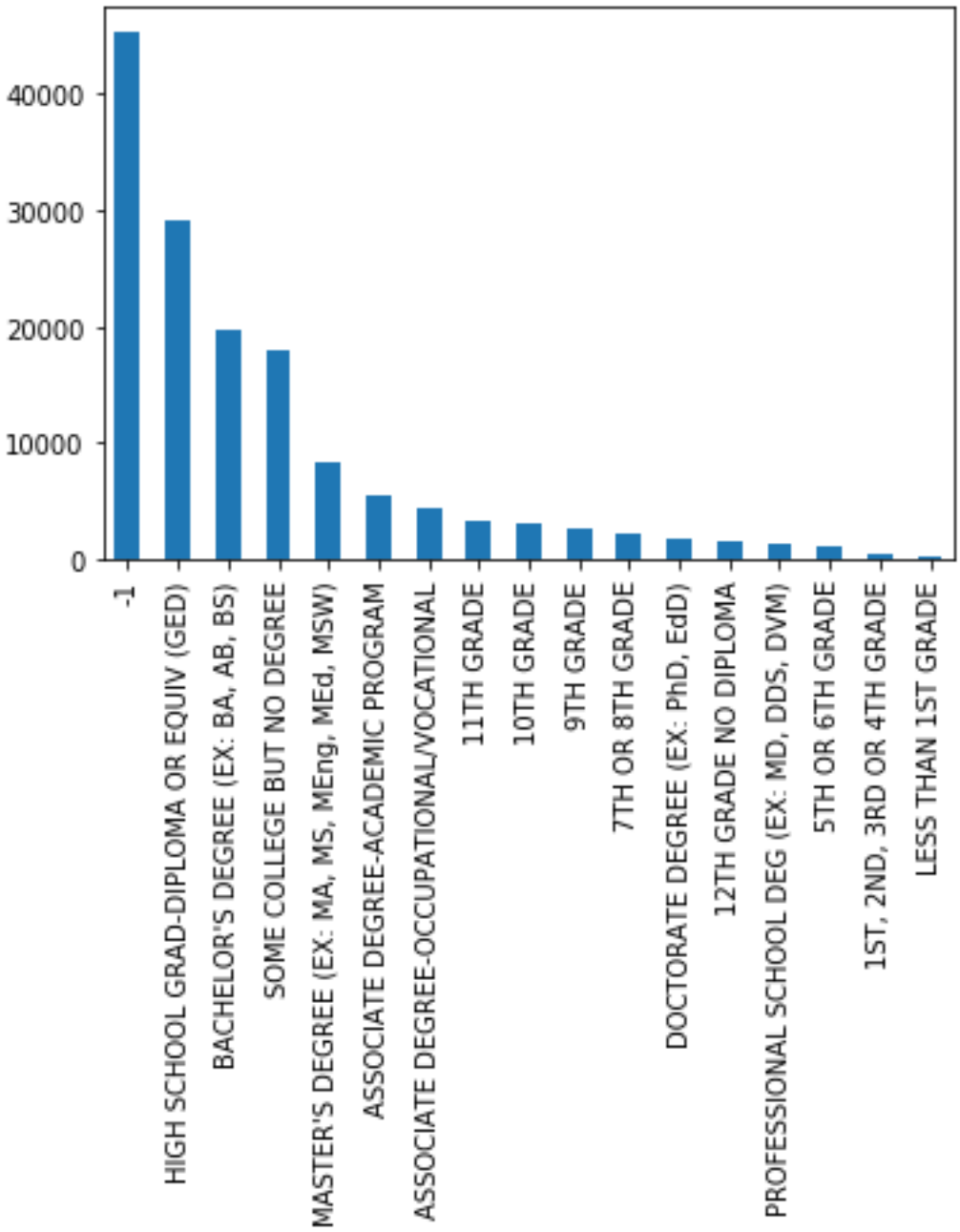

Figure 2: education level bar plot

This is a plot about variable "peeduca", which is the highest level of school completed or degree received. Not too surprisingly, high school grad-diploma or equiv has most people, and then bachelor's degree. Less than 1st grade has the least respondents.

Problem 6

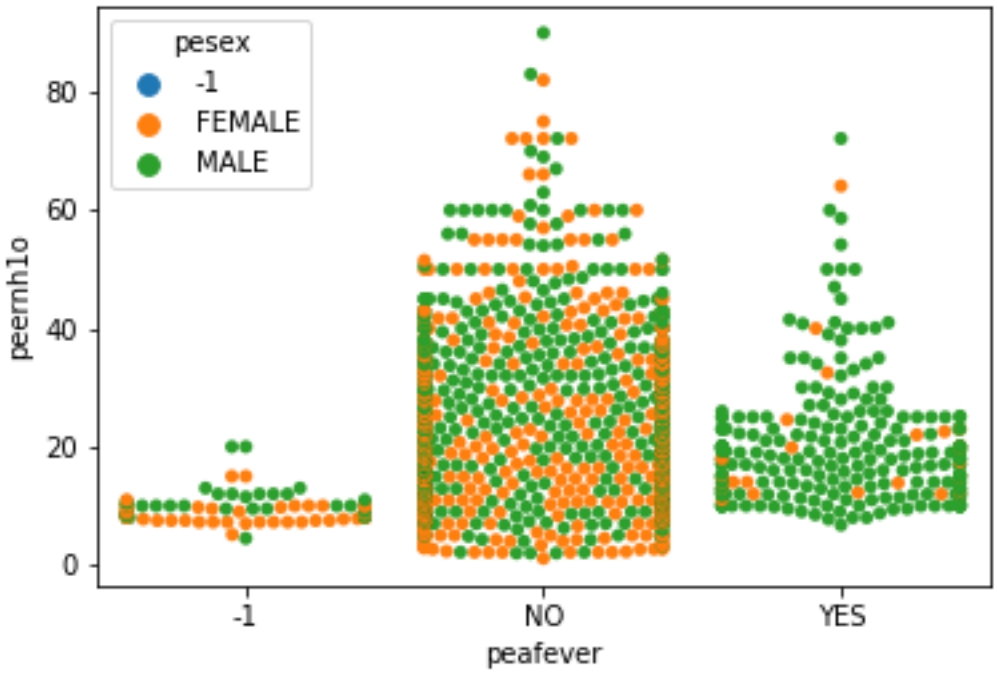

Figure 3

This is a plot showing the relationship of peernh10, which is the out variable for hourly rate of pay, and peafever, which means whether the respondents have ever served on active duty in the U.S. armed forces. A hue parameter on sex was added to differentiate between female and male. As we can see most respondents have not served in the U.S. armed forces before and among them, more female seem to have hourly rate of pay below 20 than male. For those who have served in the U.S. armed forces, male dominated.

2 Part 2

Assigned paper: Boxell, L., Gentzkow, M., Shapiro, J. M. (2017). Is the internet causing political polarization? Evidence from demographics (No. w23258). National Bureau of Economic Research.

Problem 1
This paper studies how trends in political polarization relate to respondents propensities to obtain news or information online or from social media. More specifically, the paper studies whether the demographic differences in the measures of political polarization that have already existed are consistent with the important role for the internet and social media.

Problem 2
The primary sources of data are the survey data from the American National Election Studies (ANES) 1948-2012 Time Series Cumulative, 2008 Time Series Study, and 2012 Time Series Study data sets. Since the ANES does not include information about social media usage, the study supplemented it with microdata from the Pew Research Center to plot trends in social media use from 2005 to 2012.

Problem 3
The measure of political polarization index was grounded by the theories from past literature. Political polarization was calculated by nine measures of polarization which were referenced from prior work. The nine measures are partisan affect polarization, ideological affect polarization, partisan sorting, straight-ticket, issue consistency, issue divergence, partisan-ideology polarization, perceived partisan-ideology polarization and religious polarization.

Problem 4
The paper uses both descriptive study and identification exercise. There are figures and tables which give us descriptive information about the information that was measuring in the research, and highlight interesting slices. For example, figure 1 shows trends in internet access and social media use by age group and figure 2 shows the trends in political polarization. From figure 2, we can see a clear trend that all of the nine measures of political polarization increase over time individually, and so does the overall index of political polarization.

However, the author also identified relationship in this paper, which is a characteristic of identification paper. The paper provided analysis and visualization to show that how trends in political polarization may be consistent with ones demographic group. It identified relationship of the change of political polarization index and age group.

Problem 5
The overall index of polarization is defined as:

$$Index_t = \frac{1}{|M|} \sum_{m \in M} m_t / m_{1996}$$

where M is the set of all nine polarization measures. This index was computed for different group of respondents, and the author normalized the group level values base on the overall 1996 value m_{1996}

In order to get the trends in polarization by age groups, the author did some quantitative detail and standard errors for inference. The results show that for every measure, except religious polarization, the oldest age group had larger changes in polarization than youngest age group.
The author also analyzed the trends in polarization according to predicted internet access and by actual internet access. By predicted internet access, the results show

that the group with greater possibility of having access to internet has the trend of growing more slowly in polarization between 1996 and 2012. By actual internet access, the results show that the respondents with internet access have greater polarization in 2012 than those without internet access, but the trends are parallel between the two groups between 1996 and 2012.

Problem 6

First, the author did not provide an explanation on why the nine measures of polarization referenced from prior work are sufficient enough to construct the overall polarization index. I personally think it would be better to provide a clear explanation and analysis.

Second, though visualization is good, but it is probably not sufficient enough to determine the relationship, so I would suggest to use more rigorous statistical analysis to analyze the relationship between political polarization and social media use among different age groups.

Reference

Boxell, L., Gentzkow, M., Shapiro, J. M. (2017). Is the internet causing political polarization? Evidence from demographics (No. w23258). National Bureau of Economic Research.

Current Population Survey (CPS) Basic Monthly Data at the NBER, nber.org/data/cpsbasic.html.