Problem Set 1

Haylee Ham

1. DATA

The United States Current Population Survey (CPS) can be accessed and downloaded on the website of the National Bureau of Economic Research (NBER). Although NBER stores this data, the data collection is sponsored by and carried out by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics (BLS). The data from recent months can be accessed on the U.S. Census Bureaus site (https://www.census.gov/programs-surveys/cps.html). The NBER website (http://nber.org/data/cps_basic.html) has curated all available CPS data from January 1976 to the present.

The CPS data has been used in a variety of fields in the past several decades. One much cited paper in the field of econometrics, titled How much should we trust differences-in-differences estimates?, used the state-level data on female wages from CPS data in order to generate placebo laws and estimate difference-in-difference estimates, a method which, according to their findings, understates the standard deviations of the estimators (Bertrand, et. al., 2004). A paper in the area of public health used the CPS data along with other cross-sectional survey data sources in 15 countries in order to make a survey of tobacco use (Giovino, et. al., 2012). The study found that there are high proportions of smoking among men, early onset of smoking among women, and low quit rates. CPS data has often been used to study wage inequality. One key paper in this area used CPS data from 1963 to 1989 to find that real average weekly wages for low-skilled workers fell by 5% while wages for high-skilled workers during this time rose by 40%, mostly due to rising demand for skill rather than years of experience (Juhn, et. al., 1993). Another oft-cited paper in the field of wage inequality used the CPS data from 1973-1992 in order to study the mechanism behind the rise in wage inequality during this time (Dinardo, et. al., 1996). This study found that de-unionization, shocks in supply and demand, and the minimum wages loss of real value explained much of the rise in inequality in the last quarter of the 20th century. Interestingly, the CPS data was also used in a study that rebutted the findings of the two previously discussed studies (Autor, et. al., 2008). This study, which used CPS data from 1963 to 2005, found little supporting evidence for the previously found motivation driving wage inequality of the falling real value of the minimum wage. Rather, the authors of Trends in U.S. wage inequality: Revising the revisionists found wage inequality patterns show an increase in demand for high-skill and low-skill workers and a sharp decline for middle-skill workers due to technology advances that aided highly educated workers, substituted for jobs previously requiring workers with median years of education, and increased the need for less educated workers.

The CPS data is a survey of 60,000 households across all states in the United States as well as the District of Columbia. Samples from each of the areas surveyed in the CPS are constructed in order to produce representative estimates that reflect the demographic makeup of each area. In order to be eligible, a respondent must not be a member of the armed forces, a resident of a medical institution, and must at least 15 years old. When selected, a household is surveyed for four consecutive months, not surveyed for the following eight months, and then surveyed again for four more consecutive months. These surveys are conducted over the phone or in person and surveyors generally only speak to one representative of the household, usually the household member who rents or owns the house. In order to reflect the population of the entire nation, each person surveyed is weighted by the inverse of the probability of said person being in the sample. Weighting is also used to correct for household non-response. More weight is added to demographically-similar households within the same sample area that did respond to the survey. The questions included in the survey are generally concerning labor force participation, but questions concerning veteran status, student status, job tenure, and income, and other topics frequently appear. Supplements to the standard CPS survey collect more varied data across a range of fields.

1.1 References

Autor, D. H., L. F. Katz, and M. S. Kearney. "Trends in U.S. Wage Inequality: Revising the Revisionists." Review of Economics and Statistics, vol. 90, no. 2, 2008, pp. 300-323, SCOPUS, www.scopus.com, doi:10.1162/rest.90.2.300.

Bertrand, M., E. Duflo, and S. Mullainathan. "How Much should we Trust Differences-in-Differences Estimates?" Quarterly Journal of Economics, vol. 119, no. 1, 2004, pp. 249-275, SCOPUS, www.scopus.com, doi:10.1162/003355304772839588.

Dinardo, J., N. M. Fortin, and T. Lemieux. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." Econometrica, vol. 64, no. 5, 1996, pp. 1001-1044, SCOPUS, www.scopus.com.

Giovino, G. A., et al. "Tobacco use in 3 Billion Individuals from 16 Countries: An Analysis of Nationally Representative Cross-Sectional Household Surveys." The Lancet, vol. 380, no. 9842, 2012, pp. 668-679, SCOPUS, www.scopus.com, doi:10.1016/S0140-6736(12)61085-X.

Juhn, C., K. M. Murphy, and B. Pierce. "Wage Inequality and the Rise in Returns to Skill." Journal of Political Economy, vol. 101, no. 3, 1993, pp. 410-442, SCOPUS, www.scopus.com, doi:10.1086/261881.

2. FIGURES AND TABLES

Table 1. Descriptive Statistics for Key Continuous Variables

Variable	**Mean**	**Median**	**Min**	**Max**	**Std. Dev.**
Total number of persons in household	2.9105	3.0	0	16	1.8945
Person's age	33.6522	33.0	-1	85	25.4789
Hours per week worked at main job	13.6795	-1.0	-4	99	20.3843
Hours per week worked at other jobs	-0.7344	-1.0	-4	81	2.4246
Sum of all hours worked per week	13.8365	-1.0	-4	160	20.7649

Table 2. Descriptive Statistics for Key Categorical Variables

Variable	**No. of Unique Values**	**Mode**
Family income	16	100,000 - 149,999 USD
Metropolitan area size	6	Not identified or nonmetropolitan
Highest level of education	16	High school graduate
Geographic region	4	South
State	51	California
Marital status	6	Married, spouse present
Sex	3	Female
Race	26	White only

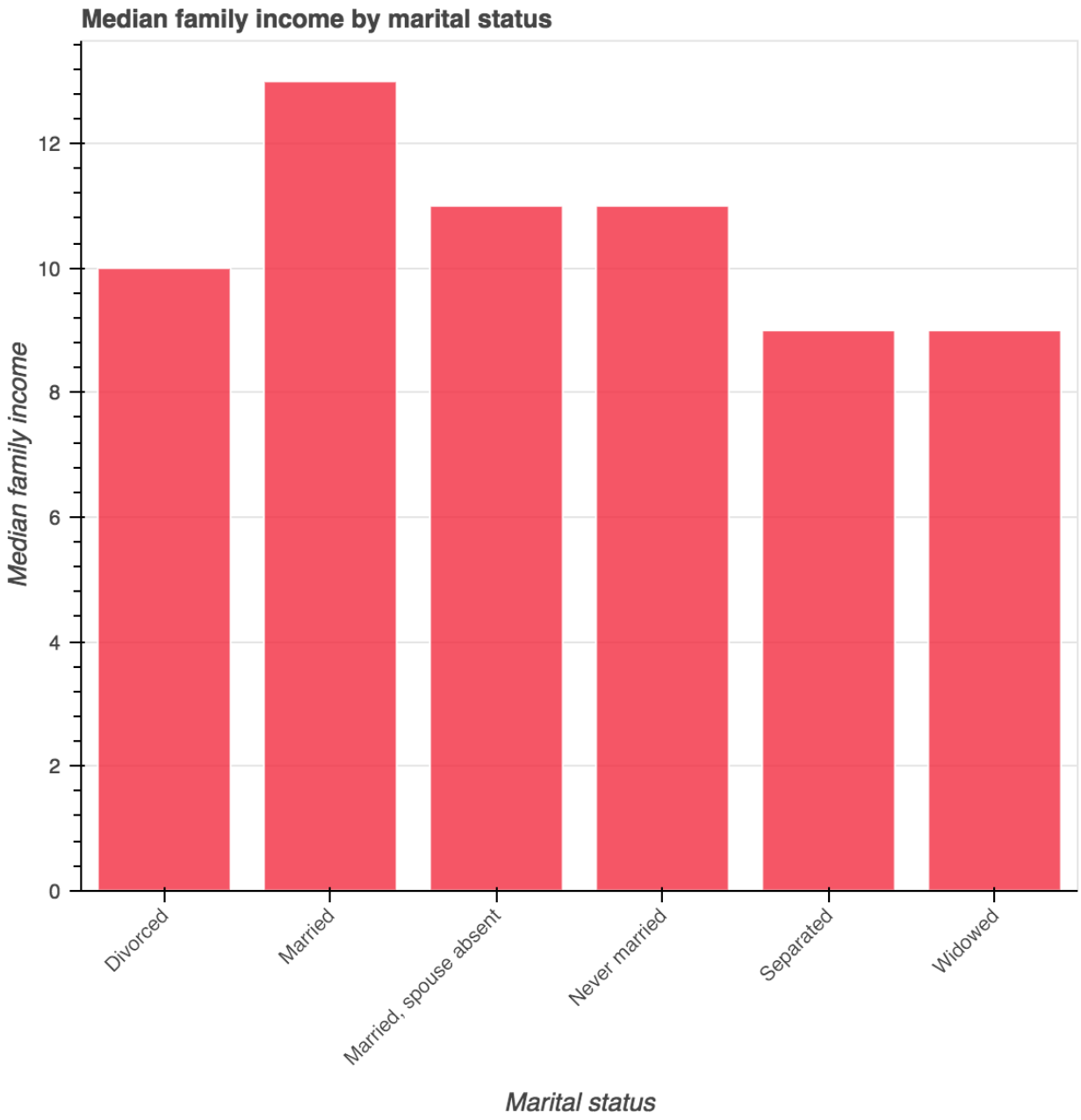

Figure 1. This bar graphs shows the median family income in each category of martial status. From this graph, it can be seen that married respondents have the highest median family income. Interestingly, respondents who are married with an absent spouse have a median family income two brackets higher than those who are separated.

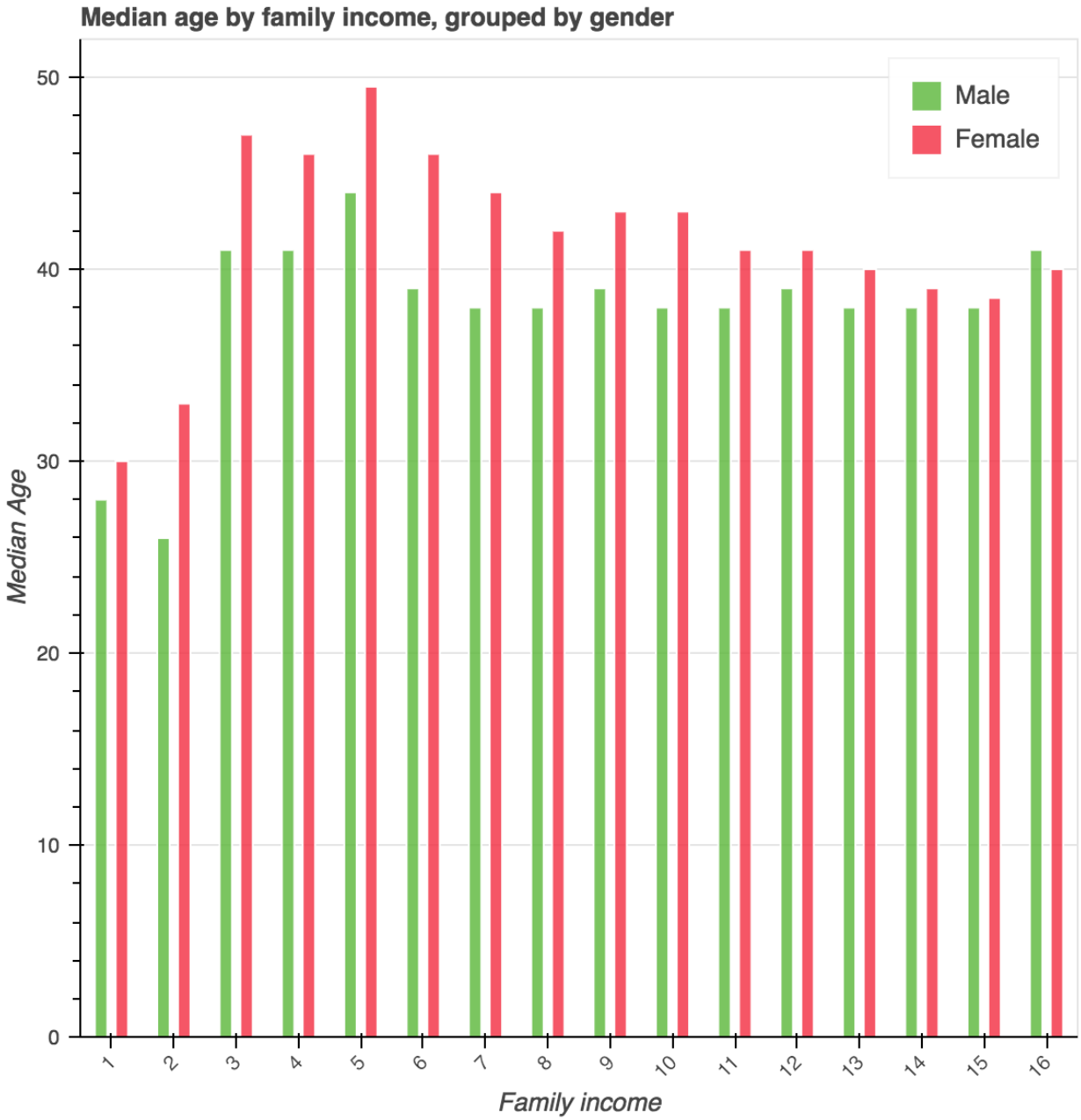

Figure 2. This bar graph shows the median age in each family income bracket, divided by gender. The median age of the lowest two brackets, which is less than 7,500 USD, is significantly lower than those in higher income brackets. It can also be seen that for all but the highest income bracket, the median age of females is higher than the median age of males. This may simply be an reflection of the median age of all men being lower than the median age of all women. This is in fact the case; the median age of men is 38 and the median age of women is 40. However, it is interesting to note that the median age of women generally declines above the household income of 15,000 USD, whereas the median age of men remains roughly the same above that threshold.

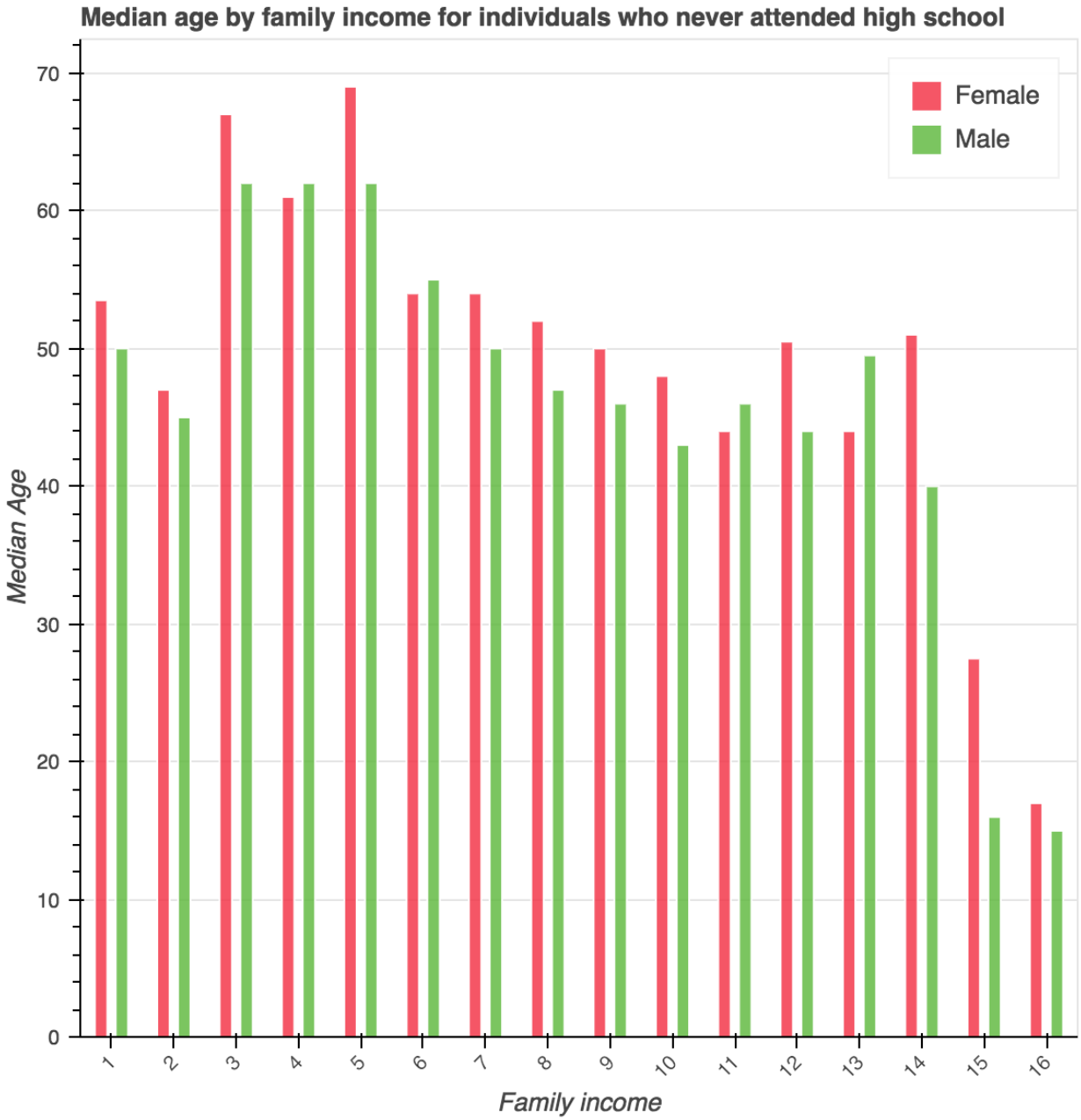

Figure 3. The graph showing median age by family income is revisited, however this time showing only individuals who never attended high school. These individuals may have attended some schooling, even up to 8th grade, however they never enrolled in high school. In this context, it can again be seen that the median age of females is generally higher than that of males in each category. However, a sharp decline is seen in the median age of both genders in the highest categories of family income. This is presumably due to the fact that the members of the highest categories of family income are children in a home where a parent or guardian is earning the stated family income and the children are not yet old enough to be enrolled in high school.

Table 3. Descriptive Statistics for Key Variables for Males

Variable	**Mode**	**Mean**	**Median**	**Std. Dev.**
Family income	100,000 - 149,999 USD	NA	NA	NA
Metropolitan area size	Not identified or nonmetropolitan	NA	NA	NA
Highest level of education	High school graduate	NA	NA	NA
Geographic region	South	NA	NA	NA
State	California	NA	NA	NA
Marital status	Married, spouse present	NA	NA	NA
Person's age	NA	38.1560	38.0	22.8317
Hours per week worked at main job	NA	18.1002	-1.0	22.1885
Hours per week worked at other jobs	NA	-0.6857	-1.0	2.6938
Sum of all hours worked per week	NA	-0.9107	-1.0	0.4412

Table 4. Descriptive Statistics for Key Variables for Females

Variable	**Mode**	**Mean**	**Median**	**Std. Dev.**
Family income	100,000 - 149,999 USD	NA	NA	NA
Metropolitan area size	Not identified or nonmetropolitan	NA	NA	NA
Highest level of education	High school graduate	NA	NA	NA
Geographic region	South	NA	NA	NA
State	California	NA	NA	NA
Marital status	Married, spouse present	NA	NA	NA
Person's age	NA	40.1606	40.0	23.1699
Hours per week worked at main job	NA	14.0842	-1.0	19.6740
Hours per week worked at other jobs	NA	-0.6977	-1.0	2.5265
Sum of all hours worked per week	NA	-0.9263	-1.0	0.4219