

Handout #1, Data Analysis Project **Investigating the Effect of Race and Gender on Earnings in the US¹**

For the past couple weeks, we have been discussing inequality in America. Often, we think of power and inequality in terms of income. Who has more money, and thus, in general, more access to opportunities and resources? Who makes the most money, and who makes the least? Does income differ for men and women, and for whites and people of color? In this exercise, we will examine earnings data for all full-time workers in the US. The data come from the 2012-2016 Public Use Microdata Sample (PUMS). You will be able to examine data for the nation as a whole, for Kentucky, and for a state of your choosing.

LEARNING GOALS:

A. *Substantive: Students will be able to:*

1. Discuss and write about the influence of race and gender on earnings, by examining 2012-2016 Census data. You will be able to use data to explain whether men make more than women and whether whites earn more than people of color.
2. Use data to make national, regional and state comparisons in terms of earnings. You will be able to compare Kentucky to the nation as a whole, in terms of income inequality, as well as comparing Kentucky to another state of their choosing. The data shown in the tables for this worksheet come from <http://ssdan.net/datacounts/webchip/> which access the geo2016 data collection and then the data set Earn_States.

B. *Methodological/Quantitative Skills: Students will acquire the following skills, related to data analysis:*

1. Interpretation: Students will be able to:

- Read and report basic frequencies from a large data set
- Describe bivariate tables, both orally and in writing

2. Representation: Students will be able to:

- Take raw data in tabular form and create properly formatted and labeled tables

3. Application/Analysis: Students will be able to:

Manipulate variables in a large data set using a basic statistical package

- Identify independent and dependent variables
- Form hypotheses about the relationship between two variables
- Analyze relationship between two variables as presented in bivariate table

4. Communication: Students will be able to:

- Present hypotheses and findings for relationship between two variables in informal presentation and in a paper that tells a story about the effects of race and gender on income, using numbers as evidence

5. Confidence: Students will feel more comfortable

- reading and discussing data from a table
- writing about and using percentages to make an argument

VARIABLES: For this exercise, we will be examining the following variables: earnings, race, sex, and age. Although there are many ways in which each could be conceptualized, the following ways are those used by the U.S. Census Bureau and the Bureau of Labor Statistics:

¹ This exercise is based on a module originally developed by Tim Thornton, SUNY-Brockport.

EARNINGS: Money a person makes from working, as wages, salary, or a form of self-employment, expressed as an annual amount

SEX - individual's self-identification as either male or female.

AGE: - age divided into the following categories: 16-24, 25-34, 35-44, 45-54, 55-64, 65+

RACE (RaceLat) – individual's self-identification as:

- **Non-Hispanic White (NHwhite)** – all persons who indicated their race as white and not of Hispanic origin.
- **Black** – all persons who indicated their race as black.
- **Asian (or Pacific Islander)** – includes all persons who indicated their race or ethnicity as Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Cambodian, Hmong, Laotian, Thai, or other Asian as well as Hawaiian, Samoan, Guamanian or other Pacific Islander.
- **Hispanic**– persons of white or “other” races who identified themselves as Mexican, Puerto Rican, Cuban, or Other Spanish/Hispanic. This category can refer to ancestry, nationality group, lineage, or country of birth of the person's parents or ancestors before their arrival in the U.S.
- **American Indian (AmIndian)** – all persons who classified themselves as American Indian, Eskimo or Aleut.
- **Non-Hispanic Multiracial (NHMulti)** – any Non-Hispanic persons who identified as more than 1 race
- **Non-Hispanic Other (NHOther)** – any Non-Hispanic persons of a single race that was not white, black, American Indian/Alaskan Native (AIAN), Asian or Pacific Islander

FREQUENCIES: A frequency table gives an overall sense of the distribution of a particular variable or set of variables. Here are the frequencies for the variable RaceEth (race) for the PUMS sample of **full-time, year-round workers in 2016**.

RaceEth:

NHWhite	Black	AmIndian	Asian	NHMulti	NHOther	Hispanic
65.17%	10.95%	0.51%	5.67%	1.48%	0.35%	15.88%
69,251,560	11,632,096	545,951	6,020,063	1,570,512	369,151	16,873,896

Note that the first row of data gives percentages, and the second row gives the total number. Thus, in the U.S., 69,251,560 fulltime, year-round workers in 2016 were Non-Hispanic White, which represents 65.2% of fulltime workers. Generally for comparison purposes, we talk about percentages, rather than raw numbers. We see that about two-thirds of US full-time workers in 2016 were White, about 11% of all workers were Black, about 5.7% Asian, about 16% Hispanic, less than 1% were American Indian, and very few identified as Non-Hispanic Other or Non-Hispanic Multi-racial.

Now examine the frequencies for age, sex, and earnings below:

Sex:

Male	Female
57.32%	42.68%
60,913,292	45,349,937

Earnings

<25K	25-34K	35-49K	50-69K	70-99K	100K+
23.24%	15.74%	19.44%	17.24%	12.32%	12.03%
24,692,909	16,721,809	20,655,225	18,318,496	13,092,583	12,782,207

Age

16-24	25-34	35-44	45-54	55-64	65+
6.93%	23.3%	23.59%	24.87%	17.64%	3.67%
7,362,112	24,762,219	25,071,933	26,422,814	18,744,742	3,899,409

Answer the following:

- What percentage of all full-time workers are men? _____
- What percentage of all full-time workers make less than \$25,000? _____
Less than \$34,000: _____ \$50,000 or more? _____
- What percentage of all full-time workers are 16-24? _____ older than 65? _____

Now describe these frequencies in easily understood English. In other words, how would you describe all full-time workers in the year 2016?

Now you get to be the sociologists. You can see that about 23% of all full-time workers make less than \$25,000. Who are they? About 24% of all full-time workers make more than \$70,000. Who are they? Are they men or women? Are they Whites, African Americans, Hispanics, Asians? Which age group makes the most? The least? Begin by making hypotheses about what you expect to find:

HYPOTHESES:

- 1) SEX: _____ will have higher incomes than _____.
- 2) RACE/ETHNICITY: _____ will have the highest incomes, and _____ will have the lowest.
- 3) AGE: People in the age group _____ will earn the most money, and people in the _____ age group will earn the least.

In order to investigate your hypotheses, you will need to do cross-tabulations, also called bivariate tables. In a cross-tabulation, you are simply exploring the association between two variables. Before running these, think about what you really want to know. It makes sense to say that a person's sex affects his or her earnings. It does not make sense to say that earnings affect a person's sex.

The following two definitions are important:

INDEPENDENT VARIABLE (X) - the variable that influences or affects another variable

DEPENDENT VARIABLE (Y) - the variable that is influenced by, or depends upon, another variable

You can write the relationship between the two as $X \rightarrow Y$. In this case, we are interested in how sex influences earnings. Sex would be the independent variable (X), and earnings would be the dependent variable (Y). In other words, earnings to some extent, depend on sex:

Independent Variable \rightarrow Dependent variable
 $X \rightarrow Y$
Sex \rightarrow Earnings

Below is a bivariate table (cross-tabulation) of sex and earnings:

Table 1: 2016 Earnings by Sex for U.S. Full-time Year-Round Workers, ACS/PUMS

Earnings	Male	Female	Total
<25K	21.1%	26.2%	23.2%
25-34K	13.8%	18.3%	15.7%
35-49K	18.2%	21.1%	19.4%
50-69K	17.7%	16.6%	17.2%
70-99K	13.7%	10.4%	12.3%
100K+	15.5%	7.4%	12%
Total	100%	100%	100%
	60,913,292	45,349,937	106,263,229

source: Earn.dat: 2012-16, wgted PUMS, Frey/U-Mich for SSDAN

Do incomes differ for men and women in the US, and if so, how? Who actually makes more? Below, try to write a description of this table:

Here's my description: From Table 1, we see that female full-time workers typically make less than male full-time workers. For example, 44.5% of women make \$34,000 or less, compared to only 35% of men. By contrast, 29% of full-time male workers make \$70,000 or more, compared to only 18% of women.

Note how I have described the data. I started with a broad, generalized statement: "female full-time workers typically make less than male full-time workers." Then I used specific statistics from the table to

make my case. Note that since I **percentaged down the rows**, and I **compared across the columns**. In other words, I compared percentages **across men and women**.

Now let's examine the influence of race on earnings:

Table 2: 2016 Earnings by Race for U.S. Full-time Year-Round Workers, PUMS

Earnings	NHWhite	Black	AmIndian	Asian	NHMulti	NHOther	Hispanic	Total
<25K	19%	28.7%	31.4%	20.3%	23.9%	27.6%	37.6%	23.2%
25-34K	14.2%	20%	20%	11.7%	16.1%	18.3%	20.2%	15.7%
35-49K	19.7%	21.1%	20.7%	15%	19.6%	20.4%	18.5%	19.4%
50-69K	18.9%	15.6%	14.7%	15.9%	17.1%	15.4%	12.2%	17.2%
70-99K	13.9%	9.2%	8.2%	16.1%	12.3%	10.6%	6.9%	12.3%
100K+	14.3%	5.4%	5%	21%	11.1%	7.7%	4.6%	12%
Total	100%	100%	100%	100%	100%	100%	100%	100%
Total	69,251,560	11,632,096	545,951	6,020,063	1,570,512	369,151	16,873,896	106,263,22

Source: Source: Earn.dat: wgted PUMS 2012-16, Frey/U-Mich for SSDAN

When interpreting this table, you are going to want to compare **ACROSS** the categories (e.g. compare Whites to Blacks), not down the categories. To make comparisons easier, you might want to combine some categories. Which categories make the most sense to combine? How would you do this? Below, write a brief description of race differences in earnings. This will likely take you several drafts.