ECON 490: *Current Population Survey (CPS) Class Activity*

For this activity, we will explore the CPS capstone data set posted to Dropbox. You will also get some practice working with data in R in a less structured context than what we've seen previously in the R coding activities.

To get started:

- Download the CPS data set as well as the PDF documentation from Dropbox – there is a link on the "Capstone Data Resources" page on Canvas.
- Open the data in R using either 1) `File > Open File...` or 2) `setwd() + readRDS()`
- Create a new R script file and include "ECON 490 CPS Activity: Your Name" at the top as a comment.

As you work through the questions below, write out your answers (with comments to indicate each question) in your code file. Everyone will answer the first set of questions below; you'll then answer the questions corresponding to your group number in the second section. Everyone should submit their own code file.

Questions for Everyone

The following questions are the kinds of basic questions you should ask whenever opening a new data set. Check each of the questions below in your data set (use a comment to separate each question, e.g., `# Q1`, etc.).

Q1) What years are covered in this data set?

Q2) How many observations are included in your data set? From the documentation, what is each row of the data?

Q3) What states do we observe in the data? What Census regions do we observe?

Group Questions

Everyone should have a group number – you only have to complete the questions below corresponding to your specific group number!

GROUP 1: *Migration/Mobility*

We want to explore what economic and demographic factors are correlated with mobility. In other words, what characteristics are associated with people deciding to move/stay put?

- Start by reading the documentation to learn about the migration status variable.
- Then, create a 0/1 binary variable set equal to 1 if someone moved last year and 0 otherwise.
- Next, identify two other variables and explore the correlation of these variables with your new 0/1 binary mobility variable.
    - Make sure to pay attention to how these other variables are recorded! In particular, make sure that you are handling missing values and factor variables appropriately.
- Finally, you should generate the following output:
    - At least one conditional summary statistic where you calculate average mobility rates (i.e., the average of your binary mobility variable) across the distribution of one of the other variables you've selected.
    - At least one regression where you use your 0/1 migration indicator as an outcome variable and the 2 other variables you've selected as explanatory variables.

- In comments with your code, interpret the coefficients from your regression output. What do they mean in practical terms? Are the results what you were expecting?

GROUP 2: *Labor Force Participation*

We want to explore what factors are correlated with whether or not individuals are in the labor force. In other words, what characteristics are associated with people deciding to work/participate in the labor market?

- Start by reading the documentation to learn about the labor force status variable.
    - Note that sometimes, employment (whether or not you currently have a job) is defined conditional on being in the labor force – don't get these two variables mixed up! Here, we want to focus on whether or not you're in the labor force at all, regardless of your employment status.
- Next, create a 0/1 binary variable set equal to 1 if someone is in the labor force and 0 otherwise.
    - Technically, the labor force variable is already binary, but we only want to include binary variables in OLS regressions when we've recoded them as 0/1 variables.
- Then, identify two other variables and explore the correlation of these variables with your new 0/1 binary labor force participation variable.
    - Make sure to pay attention to how these other variables are recorded! In particular, make sure that you are handling missing values and factor variables appropriately.
- Finally, you should have the following output:
    - At least one conditional summary statistic where you calculate average labor force participation rates (i.e., the average of your binary labor force variable) across the distribution of one of the other variables you've selected.
    - At least one regression where you use your 0/1 labor force indicator as an outcome variable and the 2 other variables you've selected as explanatory variables.
- In comments with your code, interpret the coefficients from your regression output. What do they mean in practical terms? Are the results what you were expecting?

GROUP 3: *Earnings from Employment*

We want to explore what factors are correlated with employment earnings. In other words, what economic and demographic characteristics are associated with higher or lower incomes?

- Start by reading the documentation to learn about the `inc.wage` variable.
    - Pay careful attention to who reports wage income in this data set. You'll want to use the `filter()` to include an appropriate sample of individuals. The easiest option is to filter your sample to only people with non-missing `inc.wage` values.
- Next, identify two other variables and explore the correlation of these variables with wage earnings.
    - Make sure to pay attention to how these other variables are recorded! In particular, make sure that you are handling missing values and factor variables appropriately.
- Finally, you should generate the following output:
    - At least one conditional summary statistic where you calculate average wages across the distribution of one of the other variables you've selected.
    - At least one regression where you use wage earnings as an outcome variable and the 2 other variables you've selected as explanatory variables.
- In comments with your code, interpret the coefficients from your regression output. What do they mean in practical terms? Are the results what you were expecting?