

# U6614: Assignment 2: Assessing gender wage gaps using the Current Population Survey

Your Name (your-uni)

2021-09-17

*Please submit your knitted .pdf file along with the corresponding R markdown (.rmd) via Courseworks by 11:59pm on Monday, February 1st.*

*Before knitting your rmd file as a pdf, you will need to install TinyTex for Latex distribution by running the following code:*

```
tinytex::install_tinytex()
```

*Please visit [this](#) link for more information on TinyTex installation.*

## 1 Load and inspect CPS data:

### a) Inspect the data frame and data types for each column

- make sure to inspect the age, sex, race, college columns

### b) Use the mutate function to create new column for sex

- `sex.fac = as.factor(sex),`
- check if it worked by calling the `str()` function

### c) Include sex.fac in a new data frame called cps.temp1

- also create factors for race and college education,
- use a pipe to exclude the columns for serial, ind
- after creating cps.temp1, print the first 5 observations

### d) Inspect race.fac, sex.fac, and college.fac using the levels() function

- what package is the levels() function located in?

### e) Use filter() to only include rows only for June 2020

- store as a new object cps\_2020,
- print the first 5 observations,
- confirm your data only includes observations for 2020

### f) Remove the cps.temp1 object from memory using the rm() function

## 2 Describe the cps\_2020 data frame

- a) What is the unit of observation?
- b) How many individuals are observed? from how many households?
- c) What is the average age of individuals in the sample? Youngest and oldest person?

## 3 Earnings per week for different groups in June 2020

- a) Find the observation for the top weekly earnings using the summarise() function
  - assign this to a new object called med\_earnings
- b) Find max weekly earnings using the arrange function instead of summarise
- c) Use the filter function to subset for the observation with max weekly earnings
  - don't hardcode the max earnings to filter on, refer to the max\_earnings object from a),
  - store in new data frame cps\_max\_earn,
  - confirm it worked
- d) What is the age, sex, and race of the top weekly earner in the sample?
- e) List the age, sex, and race of the top 10 weekly earners in the sample
- f) How many individuals earned more than \$2000 in weekly earnings?

## 4 Wage gaps between males and females:

- a) Use the filter function to subset observations for males
  - assign to new data frame, cps\_2020\_male,
  - sort in descending order of weekly earnings
  - check if it worked
- b) Repeat part a for females and create a new data frame, cps\_2020\_female
- c) Use summarise to find mean, min & max for males and females, separately
  - name each statistic appropriately (i.e. name each column in the 1-row table of stats)
  - what is the gender gap in mean weekly earnings?
- d) What is the wage gap in weekly earnings between white males and Black females?
- e) What is the wage gap between college educated white males and college educated Black females?

*NOTE: the exercises above are done using weekly earnings, but can easily be converted to hourly wages*