Pre-Doc Task 1: Labor Force Participation

Meredith Gavin

6/20/2023

PREP WORK

Libraries

```
library(haven)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6          v purrr   0.3.4
## v tibble  3.1.8          v stringr 1.4.0.9000
## v tidyr   1.2.0.9000     v forcats 0.5.1
## v readr   2.1.2

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(readr)
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

Set-Up

```
#set working directory as pathname
task1_wd <- "/Users/meredithgavin/Desktop/pre-doc-data-tasks/pre_doc_task_1"
```

QUESTION 1 - Labor Market Analysis

Goal: We'd like you to use these data to produce your best answer to the following question: how have hourly wages ("wage") and labor force participation ("lfp") evolved for skilled and unskilled workers since 1976? (a) Please summarize the key trends for wages and labor force participation. (b) Among men older than age 25, which groups of people have had the biggest changes in labor force participation? (c) What factors do you think are driving these patterns? What evidence might you want to assemble to test these hypotheses if you were to investigate them further?

Load dataset

```
cps_wages_lfp <- read_csv("cps_wages_LFP.csv")
```

```
## Rows: 6883923 Columns: 23
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (12): statefip, month, age, sex, race, hispan, educ, empstat, wkswork2, ...
## dbl (11): year, wtsupp, occ, wkswork1, uhrsworkly, inctot, incwage, white, s...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

explore data

```
str(cps_wages_lfp)
```

```
## spc_tbl_ [6,883,923 x 23] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ year      : num [1:6883923] 1977 1977 1977 1977 1977 ...
##  $ statefip  : chr [1:6883923] "alabama" "alabama" "alabama" "alabama" ...
##  $ month     : chr [1:6883923] "march" "march" "march" "march" ...
##  $ wtsupp    : num [1:6883923] 1444 1593 1230 1473 1503 ...
##  $ age       : chr [1:6883923] "15" "4" "43" "34" ...
##  $ sex       : chr [1:6883923] "male" "female" "male" "female" ...
##  $ race      : chr [1:6883923] "white" "white" "white" "white" ...
##  $ hispan    : chr [1:6883923] "not hispanic" "not hispanic" "not hispanic" "not hispanic" ...
##  $ educ      : chr [1:6883923] "grade 8" NA "grade 4" "grade 8" ...
##  $ empstat   : chr [1:6883923] "nilf, school" NA "at work" "at work" ...
##  $ occ       : num [1:6883923] NA NA 473 283 NA 753 623 NA 703 NA ...
##  $ wkswork1  : num [1:6883923] NA NA 52 52 NA 39 13 NA 36 NA ...
##  $ wkswork2  : chr [1:6883923] NA NA "50-52 weeks" "50-52 weeks" ...
##  $ uhrsworkly: num [1:6883923] NA NA 58 57 NA 50 55 NA 20 NA ...
##  $ inctot    : num [1:6883923] 0 NA 11200 0 3043 ...
##  $ incwage   : num [1:6883923] 0 NA 10400 0 0 3300 600 NA 1170 0 ...
##  $ age_group : chr [1:6883923] "age < 25" "age < 25" "25 <= age < 45" "25 <= age < 45" ...
##  $ white     : num [1:6883923] 1 1 1 1 1 1 1 0 1 1 ...
##  $ skilled   : num [1:6883923] 0 NA 0 0 0 0 0 NA 0 0 ...
##  $ hours     : num [1:6883923] NA NA 3016 2964 NA ...
##  $ wage      : num [1:6883923] NA NA 3.45 0 NA ...
##  $ lfp       : chr [1:6883923] "Not in labor force" NA "In labor force" "In labor force" ...
##  $ empstatid : chr [1:6883923] "Not In Labor Force" NA "Employed" "Employed" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   year = col_double(),
##   ..   statefip = col_character(),
##   ..   month = col_character(),
##   ..   wtsupp = col_double(),
##   ..   age = col_character(),
##   ..   sex = col_character(),
##   ..   race = col_character(),
##   ..   hispan = col_character(),
##   ..   educ = col_character(),
##   ..   empstat = col_character(),
##   ..   occ = col_double(),
##   ..   wkswork1 = col_double(),
##   ..   wkswork2 = col_character(),
##   ..   uhrsworkly = col_double(),
##   ..   inctot = col_double(),
##   ..   incwage = col_double(),
##   ..   age_group = col_character(),
##   ..   white = col_double(),
##   ..   skilled = col_double(),
##   ..   hours = col_double(),
##   ..   wage = col_double(),
##   ..   lfp = col_character(),
##   ..   empstatid = col_character()
```

```
##   .. )
##  - attr(*, "problems")=<externalptr>
```

(a) Please summarize the key trends for wages and labor force participation.

create month-year date column

```
cps_wages_lfp <-
  cps_wages_lfp %>%
  mutate(date = with(., sprintf("%d-%s", year, month)))
```

plot time series for labor force participation

```
options(scipen = 999) #remove scientific notation from axes

lfp_timeseries <- cps_wages_lfp %>%
  group_by(year, lfp) %>%
  summarise(n = n()) %>%
  drop_na(lfp) %>%
  ggplot(mapping = aes(x = year, y = n, color = lfp)) +
  geom_line() +
  scale_color_manual(values = c("dark green", "red")) +
  labs(title = "U.S. Labor Force Participation Over Time",
       subtitle = "1976 - 2015",
         y = "Observations",
         x = "Year",
       color = "",
       caption = "Data Source: U.S. Current Population Survey (CPS)") +
  theme_minimal()
```

```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
png("lfp_timeseries.png")
```

plot time series for wages

```
options(scipen = 999) #remove scientific notation from axes


wages_timeseries <- cps_wages_lfp %>%
  group_by(year) %>%
  summarise(mean = mean(wage, na.rm = TRUE),
            median = median(wage, na.rm = TRUE)) %>%
  pivot_longer(
    cols = c("mean", "median"),
    names_to = "type",
```

```
    values_to = "value"
  ) %>%
  ggplot(mapping = aes(x = year, y = value, color = type)) +
  geom_line() +
  scale_color_manual(values = c("dark blue", "light blue")) +
  labs(
    title = "U.S. Wages Over Time",
    subtitle = "1976 - 2015",
    caption = "Data Source: U.S. Current Population Survey (CPS)",
    color = "",
    y = "Wage",
    x = "Year"
  ) +
  theme_minimal()

png("wages_timeseries.png")
```

(b) Among men older than age 25, which groups of people have had the biggest changes in labor force participation?

Subset the data to include only men over 25

```
cps_men_over25 <- cps_wages_lfp %>%
  filter(sex == "male" & age > 25) %>%
  mutate(lfp_binary = recode(lfp,
                             "Not in labor force" = "0",
                             "In labor force" = "1"))
```

plot total labor force participation by year

```
lfp_over25_timeseries <- cps_men_over25 %>%
  filter(lfp != "NA") %>%
  group_by(year, lfp) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  ggplot(mapping = aes(x = year, y = n, color = lfp)) +
  geom_line() +
  labs(title = "Aggregate Labor Force Participation",
       subtitle = "1976 to 2015",
       x = "Year",
       y = "Total Participants",
       color = "",
       caption = "Source: U.S. Current Population Survey") +
  theme_minimal()
```

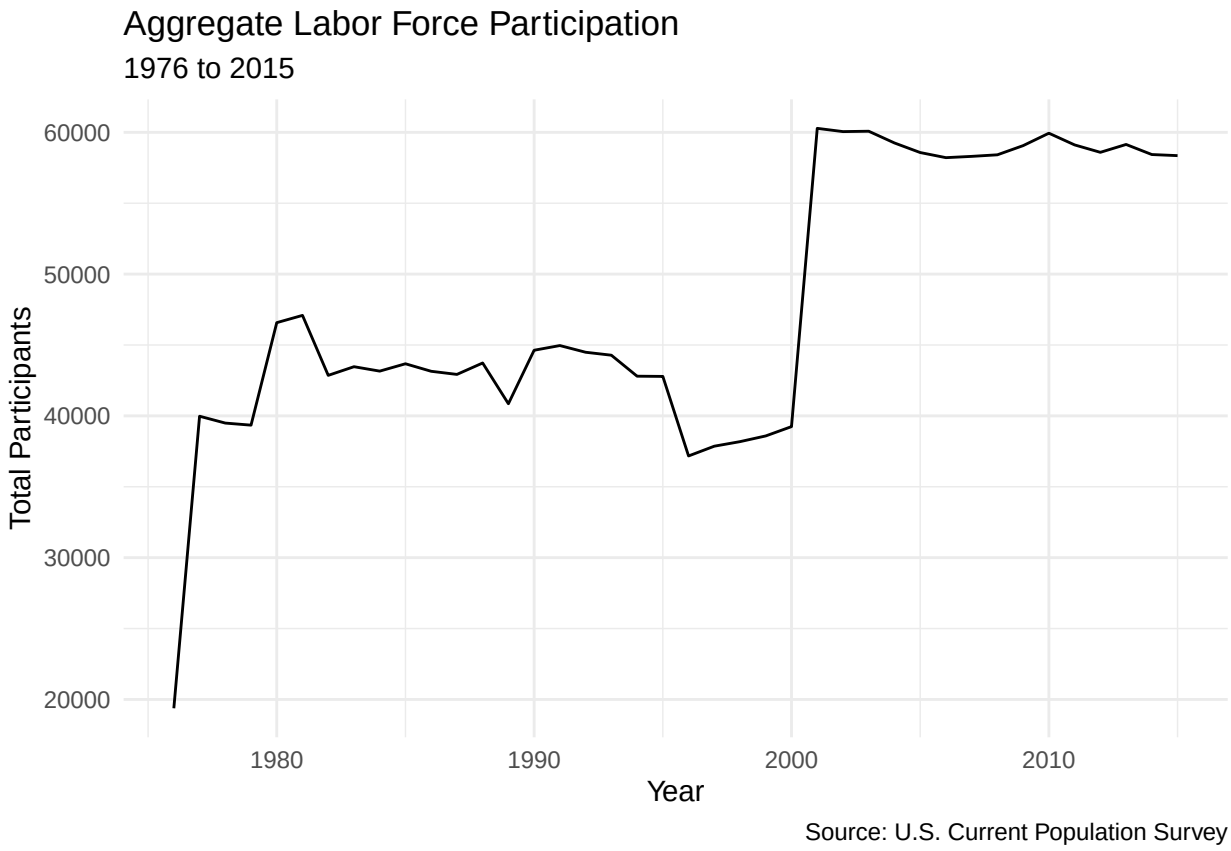
```
## `summarise()` has grouped output by 'year'. You can override using the
## `.groups` argument.
```

```
ggsave("lfp_over25_timeseries.png")
```

```
## Saving 6.5 x 4.5 in image
```

plot percent labor force participation by year

```
cps_men_over25 %>%
  filter(lfp != "NA") %>%
  group_by(year) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  ggplot(mapping = aes(x = year, y = n)) +
  geom_line() +
  labs(title = "Aggregate Labor Force Participation",
       subtitle = "1976 to 2015",
       x = "Year",
       y = "Total Participants",
       caption = "Source: U.S. Current Population Survey") +
  theme_minimal()
```



find percent change in labor force participation year to year and average

```
cps_over25_grouped <- cps_men_over25 %>%
  group_by(statefip, lfp, year) %>%
  summarise(n = n()) %>%
  filter(lfp != "NA") %>%
  mutate(pct_change = ((n - lag(n)) / lag(n)*100)) %>%
  filter(pct_change != "NA") %>%
  summarise_at(vars(pct_change), list(mean = mean))
```

```
## `summarise()` has grouped output by 'statefip', 'lfp'. You can override using
## the `.groups` argument.
```

Average percent change in lfp - states with greatest and smallest changes

```
not_lf_state <- cps_over25_grouped %>%
  filter(lfp == "Not in labor force") %>%
  arrange(mean) %>%
  filter(mean >= 5.0979609 | mean <=    1.0583456) %>%
  ggplot(mapping = aes(x = mean, y = reorder(statefip, -mean))) +
  geom_col(fill = "dark green") +
  labs(x = "Mean Percentage Change",
       y = "State",
       title = "Mean Percent Change of People Not in Labor Force",
       subtitle = "1976 to 2015",
       caption = "Source: U.S. Current Population Survey") +
  theme_minimal()

ggsave("not_lf_state.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
lf_state <- cps_over25_grouped %>%
  filter(lfp == "In labor force") %>%
  arrange(mean) %>%
  filter(mean >= 4.5171540 | mean <= 0.2296526) %>%
    ggplot(mapping = aes(x = mean, y = reorder(statefip, -mean))) +
  geom_col(fill = "dark green") +
  labs(x = "Mean Percentage Change",
       y = "State",
       title = "Mean Percent Change of People in Labor Force",
       subtitle = "1976 to 2015",
       caption = "Source: U.S. Current Population Survey") +
  theme_minimal()

ggsave("lf_state.png")
```

```
## Saving 6.5 x 4.5 in image
```

Plot labor force participation by age group

```
lfp_age_group <- cps_men_over25 %>%
  filter(lfp != "NA") %>%
  group_by(year, lfp, age_group) %>%
  summarise(n = n()) %>%
  arrange(desc(n)) %>%
  ggplot(mapping = aes(x = year, y = n, color = age_group)) +
  geom_line() +
  labs(
    title = "Labor Force Participation by Age Group",
    subtitle = "1976 to 2015",
    x = "",
    y = "",
    color = "",
    caption = "Source: U.S. Current Population Survey"
  ) +
  theme_minimal() +
  facet_wrap(~ lfp)
```

```
## `summarise()` has grouped output by 'year', 'lfp'. You can override using the
## `.groups` argument.
```

```
ggsave("lfp_age_group.png")
```

```
## Saving 6.5 x 4.5 in image
```