CPS Data Analysis

Kate Hyeon

2023-09-18

```r
library(tidyverse)
```

Load and inspect CPS data

```r
cps <- read.csv("cps_june_20-21.csv")

cps <- na.omit(cps)

summary(cps$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.00   30.00   40.00   40.77   52.00   64.00
```

```r
summary(cps$sex)
```

```
##    Length     Class      Mode
##     16876 character character
```

```r
summary(cps$race)
```

```
##    Length     Class      Mode
##     16876 character character
```

```r
summary(cps$college)
```

```
##    Length     Class      Mode
##     16876 character character
```

-The oldest person in our dataset is 64 years old.

-The sex column is a categorical binary variable, and there are more Male than Female.

-The race column is a categorical variable. The most common race in this data set is White-Hawaiian/Pacific Islander and the least common race in this data set American Indian-Asian.

-The college column is a categorical variable. More people had No college degree and the less people had College degree.

```r
mutate(cps, sex.fac = as.factor(sex))
```

```r
cps.temp1 <- cps %>%
  mutate(sex.fac = as.factor(sex),
         race.fac = as.factor(race),
         college.fac = as.factor(college)) %>%
  select(-serial, -ind)

head(cps.temp1, n = 5) %>%
  select(sex.fac, race.fac, college.fac, earnweek) %>%
  knitr::kable()
```

sex.fac	race.fac	college.fac	earnweek
Female	Black/Negro	No college degree	750.0
Female	Black/Negro	No college degree	1092.6
Male	Black/Negro	No college degree	760.0
Female	Black/Negro	No college degree	510.0
Male	Black/Negro	No college degree	800.0

```r
levels(cps.temp1$sex.fac)
```

```
## [1] "Female" "Male"
```

```r
levels(cps.temp1$race.fac)
```

```
##  [1] "American Indian-Asian"
##  [2] "American Indian-Hawaiian/Pacific Islander"
##  [3] "American Indian/Aleut/Eskimo"
##  [4] "Asian only"
##  [5] "Asian-Hawaiian/Pacific Islander"
##  [6] "Black-American Indian"
##  [7] "Black-American Indian-Asian"
##  [8] "Black-Asian"
##  [9] "Black/Negro"
## [10] "Four or five races, unspecified"
## [11] "Hawaiian/Pacific Islander only"
## [12] "White"
## [13] "White-American Indian"
## [14] "White-American Indian-Asian"
## [15] "White-Asian"
## [16] "White-Asian-Hawaiian/Pacific Islander"
## [17] "White-Black"
## [18] "White-Black--Hawaiian/Pacific Islander"
## [19] "White-Black-American Indian"
## [20] "White-Black-American Indian-Asian"
## [21] "White-Black-Asian"
## [22] "White-Hawaiian/Pacific Islander"
```

```
levels(cps.temp1$college.fac)
```

```
## [1] "College degree"     "No college degree"
```

-The as.factor() function allocated levels to the categorical variables, in the order that is ascending from left to right. From the level() function, we can see the categories in ascending order. In the sex category, there are more males than females. In the race category, there are generally more White people than Black, Asian, and American-Indian people. In the college degree category, there are more people with no college degree than there are those with a college degree.

```
cps_2020 <- cps.temp1 %>%
filter(year == 2020)

head(cps_2020, n = 5) %>%
  select(sex.fac, race.fac, college.fac, earnweek) %>%
  knitr::kable()
```

sex.fac	race.fac	college.fac	earnweek
Female	Black/Negro	No college degree	750.0
Female	Black/Negro	No college degree	1092.6
Male	Black/Negro	No college degree	760.0
Female	Black/Negro	No college degree	510.0
Male	Black/Negro	No college degree	800.0

```
rm(cps.temp1)
```

Descriptive Statistics

```
str(cps_2020)
```

```
## 'data.frame':    7970 obs. of  15 variables:
##  $ year      : int  2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...
##  $ month     : int  6 6 6 6 6 6 6 6 6 6 ...
##  $ statefip  : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ age       : int  44 47 45 29 28 59 25 24 56 42 ...
##  $ sex       : chr  "Female" "Female" "Male" "Female" ...
##  $ race      : chr  "Black/Negro" "Black/Negro" "Black/Negro" "Black/Negro" ...
##  $ college   : chr  "No college degree" "No college degree" "No college degree" "No college degree"
##  $ earnweek  : num  750 1093 760 510 800 ...
##  $ hrsworkt  : int  40 40 45 40 50 40 40 40 40 40 ...
##  $ hispanic  : chr  "Not Hispanic" "Not Hispanic" "Not Hispanic" "Not Hispanic" ...
##  $ hhid      : num  2.02e+13 2.02e+13 2.02e+13 2.02e+13 2.02e+13 ...
##  $ personid  : num  2.02e+13 2.02e+13 2.02e+13 2.02e+13 2.02e+13 ...
##  $ sex.fac   : Factor w/ 2 levels "Female","Male": 1 1 2 1 2 1 1 2 2 2 ...
##  $ race.fac  : Factor w/ 22 levels "American Indian-Asian",..: 9 9 9 9 9 12 12 12 12 9 ...
```

```
##  $ college.fac: Factor w/ 2 levels "College degree",..: 2 2 2 2 2 1 1 1 1 2 ...
##  - attr(*, "na.action")= 'omit' Named int [1:885] 51 64 68 71 91 148 149 152 160 161 ...
##   ..- attr(*, "names")= chr [1:885] "51" "64" "68" "71" ...
```

-The unit of observation is person.

```
summarise(cps_2020, n_distinct(personid))
```

```
summarise(cps_2020, n_distinct(hhid))
```

-The number of individuals observed is

.

-The

individuals are from

households.

```
cps_2020 %>%
  summarise(avg_age = mean(age),
            min_age = min(age),
            max_age = max(age)) %>%
  knitr::kable()
```

avg_age	min_age	max_age
41.1803	16	64

-The average age of the individuals in the sample is 41.2 years old.

-The youngest person is 16 years old.

-The oldest person is 64 years old.

Earnings per week for different groups in June 2020

```
max_earnings <- cps_2020 %>%
  summarise(max_earnings = max(earnweek),)

cps_2020 %>% arrange(desc(earnweek)) %>%
  head(n = 1) %>%
  select(earnweek) %>%
  knitr::kable()
```

earnweek
2884.5

-The top weekly earnings is 2884.5 dollars.

```
cps_max_earn <- cps_2020 %>%
  arrange(desc(earnweek))

cps_max_earn %>%
  select(age, sex, race) %>%
  head(n=1) %>%
  knitr::kable()
```

age	sex	race
39	Male	Asian only

-The age of the top weekly earner is 39, the sex is male, and the race is Asian only.

```
cps_2020 %>%
  arrange(desc(earnweek)) %>%
  select(age, sex, race, earnweek) %>%
  head(n=10) %>%
  knitr::kable()
```

age	sex	race	earnweek
39	Male	Asian only	2884.5
36	Male	White	2884.0
41	Male	White	2884.0
57	Female	White	2884.0
45	Male	White	2884.0
40	Male	White	2884.0
59	Male	White	2884.0
64	Male	White	2884.0
49	Male	Asian only	2884.0
36	Male	White	2884.0

```
nrow(cps_2020[cps_2020$earnweek > 2000, ])
```

```
## [1] 602
```

-602 individuals earned more than $2000 in weekly earnings.

Wage gaps between males and females

```
cps_2020_male <- cps_2020 %>%
  filter(sex.fac == "Male") %>%
  arrange(desc(earnweek))

cps_2020_female <- cps_2020 %>%
```

```
  filter(sex.fac == "Female") %>%
  arrange(desc(earnweek))

cps_2020_male %>%
  summarise(mean_earnings_male = mean(earnweek),
            min_earnings_male = min(earnweek),
            max_earnings_male = max(earnweek)) %>%
  knitr::kable()
```

mean_earnings_male	min_earnings_male	max_earnings_male
1101.879	0.01	2884.5

-For males, the average weekly earning is 1101.88, the minimum weekly earning is 0.01, and the maximum weekly earning is 2884.5.

-For females, the average weekly earning is 920.73, the minimum weekly earning is 0.23, and the maximum weekly earning is 2884.

-The gender gap in weekly earnings is 181.15.

Wage gaps between white males and Black females

```
cps_2020_wh_male <- cps_2020_male %>%
  filter(race.fac == "White")
cps_2020_bl_female <- cps_2020_female %>%
  filter(race.fac == "Black/Negro")
```

-The gender gap in weekly earnings is $263.03.

Wage gap between college educated white males and college educated Black females

```
cps_2020_wh_male_college <- cps_2020_male %>%
  filter(college.fac == "College degree" & race.fac == "White")
cps_2020_bl_female_college <- cps_2020_female %>%
  filter(college.fac == "College degree" & race.fac == "Black/Negro")
```

-The gender gap in weekly earnings is $345.83.