Bootstrapping..

Load packages and data

```
suppressMessages(library(dplyr))
library(ggplot2)
library(moments)
library(haven)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v forcats   1.0.0      v stringr   1.5.1
## v lubridate 1.9.3      v tibble    3.2.1
## v purrr     1.0.2      v tidyr     1.3.1
## v readr     2.1.5
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(boot)

data <- read.csv("ppha312x2021.csv") #working directory is same as data directory
# set seed to ensure same results
set.seed(110821)
```

```
data <- data %>% filter(inctot >= 0)
```

```
summary(data$age)
```

```
##    Length     Class      Mode
##     10140 character character
```

```
#desc(data$age)
```

```
#max(data$age)
```

```
summary(data)
```

```
##      year         statefip           met2013              perwt
##  Min.   :2019   Length:10140      Length:10140       Min.   :   2.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:  57.0
##  Median :2019   Mode  :character   Mode  :character   Median :  85.0
##  Mean   :2019                                         Mean   : 115.6
```

```
##   3rd Qu.:2019                                    3rd Qu.: 133.0
##   Max.   :2019                                    Max.   :1977.0
##       sex                age                race              hispan
##   Length:10140       Length:10140       Length:10140       Length:10140
##   Class :character   Class :character   Class :character   Class :character
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       bpl               educd             empstat            uhrswork
##   Length:10140       Length:10140       Length:10140       Length:10140
##   Class :character   Class :character   Class :character   Class :character
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot            incwage
##   Min.   :      0    Min.   :     0
##   1st Qu.:  11775    1st Qu.:     0
##   Median :  35000    Median : 25000
##   Mean   :1744935    Mean   :209709
##   3rd Qu.:  94000    3rd Qu.: 88250
##   Max.   :9999999    Max.   :999999
```

```r
#New Column : African American
data %>% count(race)
```

```
##                              race    n
## 1 American Indian or Alaska Native    9
## 2      Black/African American/Negro 3899
## 3                           Chinese    9
## 4                          Japanese    3
## 5  Other Asian or Pacific Islander   61
## 6                  Other race, nec   93
## 7         Three or more major races   15
## 8                  Two major races  136
## 9                            White 5915
```

```r
data$isAfricanAmerican <- ifelse(data$race == 'Black/African American/Negro', 1, 0)
data <- data %>% mutate(isAfricanAmerican = as.factor(isAfricanAmerican))
summary(data$isAfricanAmerican)
```

```
##    0    1
## 6241 3899
```

```r
head(data$isAfricanAmerican)
```

```
## [1] 1 0 0 0 1 1
## Levels: 0 1
```

```r
data$isFemale <- ifelse(data$sex == 'Female', 1, 0)
data <- data %>% mutate(isFemale = as.factor(isFemale))
summary(data$isFemale)
```

```
##    0    1
## 4823 5317
```

```r
summary(data)
```

```
##       year        statefip            met2013              perwt
##  Min.   :2019    Length:10140       Length:10140       Min.   :   2.0
##  1st Qu.:2019    Class :character   Class :character   1st Qu.:  57.0
##  Median :2019    Mode  :character   Mode  :character   Median :  85.0
##  Mean   :2019                                          Mean   : 115.6
##  3rd Qu.:2019                                          3rd Qu.: 133.0
##  Max.   :2019                                          Max.   :1977.0
##      sex                age                race              hispan
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                educd              empstat             uhrswork
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot            incwage        isAfricanAmerican isFemale
##  Min.   :       0   Min.   :     0   0:6241            0:4823
##  1st Qu.:   11775   1st Qu.:     0   1:3899            1:5317
##  Median :   35000   Median : 25000
##  Mean   : 1744935   Mean   :209709
##  3rd Qu.:   94000   3rd Qu.: 88250
##  Max.   : 9999999   Max.   :999999
```

```r
data <- data %>% mutate(isEmployed = ifelse(data$empstat == 'Employed', 1, 0))
data <- data %>% mutate(isEmployed = as.factor(isEmployed))
summary(data$isEmployed)
```

```
##    0    1
## 5478 4662
```

```r
summary(data)
```

```
##       year        statefip            met2013              perwt
##  Min.   :2019    Length:10140       Length:10140       Min.   :   2.0
##  1st Qu.:2019    Class :character   Class :character   1st Qu.:  57.0
##  Median :2019    Mode  :character   Mode  :character   Median :  85.0
```

```
##  Mean    :2019                                    Mean    : 115.6
##  3rd Qu.:2019                                    3rd Qu.: 133.0
##  Max.    :2019                                    Max.    :1977.0
##      sex                age                race              hispan
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                educd              empstat            uhrswork
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot             incwage          isAfricanAmerican isFemale isEmployed
##  Min.   :      0    Min.   :      0    0:6241            0:4823   0:5478
##  1st Qu.:  11775    1st Qu.:      0    1:3899            1:5317   1:4662
##  Median :  35000    Median :  25000
##  Mean   :1744935    Mean   :209709
##  3rd Qu.:  94000    3rd Qu.: 88250
##  Max.   :9999999    Max.   :999999
```

```r
#formatting age to be numeric using as.numeric
data <- data %>% mutate(age = as.numeric(age))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `age = as.numeric(age)`.
## Caused by warning:
## ! NAs introduced by coercion
```

```r
data_1 <- data %>%
  filter(isFemale == '1' & isAfricanAmerican == '1' & isEmployed == '1') #%>%
 # select(incwage,age)

data_1 <- data_1[, c("incwage", "age")]

summary(data_1)
```

```
##     incwage            age
##  Min.   :     0    Min.   :16.00
##  1st Qu.: 15000    1st Qu.:33.00
##  Median : 29000    Median :44.00
##  Mean   : 33869    Mean   :44.21
##  3rd Qu.: 46000    3rd Qu.:56.00
##  Max.   :476000    Max.   :86.00
```

```r
print(colnames(data_1))
```

```
## [1] "incwage" "age"
```

```r
corr <- function(x, indices){
  data <- x[indices, ]
  corr <- cor(data[,1], data[,2])
  return( corr )
}

coefficient <- boot(data_1[],
                     corr,
                     10000)


coefficient
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data_1[], statistic = corr, R = 10000)
##
##
## Bootstrap Statistics :
##      original        bias     std. error
## t1* 0.1497379 0.0008852872  0.02755512
```

#The standard error of the correlation coefficient between labor wages, worker ages is ~ 0.0278

#Filter white, men, employed

```r
data_2 <- data %>% filter(data$race == "White" & data$isEmployed == '1' & data$isFemale == '0')

data_2 <- data_2[, c("incwage", "age")]

coefficient <- boot(data_2[],
                     corr,
                     10000)


coefficient
```

```
##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = data_2[], statistic = corr, R = 10000)
##
##
## Bootstrap Statistics :
##      original        bias     std. error
## t1* 0.1748645 0.0004523001  0.02119825
```

#The standard error of the correlation coefficient between labor wages, worker ages is ~ 0.021