Sai_Omkar_K_PS4

Load packages and data

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

data <- read.csv("ppha312x2021.csv") #working directory is same as data directory
```

Data cleaning

#Study raw data

```
str(data) #to show the structure of the data
```

```
## 'data.frame':    10149 obs. of  14 variables:
##  $ year    : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ statefip: chr  "Mississippi" "Mississippi" "Mississippi" "Mississippi" ...
##  $ met2013 : chr  "Memphis, TN-MS-AR" "Memphis, TN-MS-AR" "Memphis, TN-MS-AR" "Memphis, TN-MS-AR" ..
##  $ perwt   : num  93 96 81 29 92 83 112 66 67 25 ...
##  $ sex     : chr  "Male" "Male" "Male" "Female" ...
##  $ age     : chr  "43" "60" "40" "83" ...
##  $ race    : chr  "Black/African American/Negro" "White" "White" "White" ...
##  $ hispan  : chr  "Not Hispanic" "Not Hispanic" "Not Hispanic" "Not Hispanic" ...
##  $ bpl     : chr  "Tennessee" "Tennessee" "Tennessee" "Arkansas" ...
##  $ educd   : chr  "Grade 11" "1 or more years of college credit, no degree" "Grade 10" "1 or more yea
##  $ empstat : chr  "Not in labor force" "Not in labor force" "Not in labor force" "Not in labor force"
##  $ uhrswork: chr  "40" "N/A" "29" "N/A" ...
##  $ inctot  : int  23000 10800 14000 8400 10000 0 23000 4550 4100 13300 ...
##  $ incwage : int  23000 0 0 0 10000 0 23000 0 1700 0 ...
```

```
summary(data)
```

```
##       year        statefip            met2013              perwt
##  Min.   :2019   Length:10149       Length:10149       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:10149       Length:10149       Length:10149       Length:10149
##  Class :character   Class :character   Class :character   Class :character
```

1

```
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       bpl              educd            empstat           uhrswork
##   Length:10149      Length:10149      Length:10149      Length:10149
##   Class :character   Class :character   Class :character   Class :character
##   Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##       inctot          incwage
##   Min.   : -6900   Min.   :      0
##   1st Qu.:  11700   1st Qu.:      0
##   Median :  35000   Median : 25000
##   Mean   :1743385   Mean   :209524
##   3rd Qu.:  94000   3rd Qu.: 88000
##   Max.   :9999999   Max.   :999999
```

#Deleting the inctot values below 0 and keeping >= 0

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

##Question 1

New column: Hispanic

```
data %>% count(hispan)
```

```
##          hispan      n
## 1         Cuban     4
## 2       Mexican   212
## 3 Not Hispanic 9851
## 4         Other    52
## 5 Puerto Rican    21
```

```
data$isHispanic <- ifelse(data$hispan == 'Not Hispanic', 0, 1)
data <- data %>% mutate(isHispanic = as.factor(isHispanic))
summary(data$isHispanic)
```

```
##    0    1
## 9851  289
```

```
summary(data$hispan)
```

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

```
#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
```

##1b limit to (white & non-Hispanic) or (African/American & non-Hispanic) ##Using filter function to do the work

```r
data <- data %>% filter((race == 'White' & isHispanic == 0) |
                        (isAfricanAmerican == 1 & isHispanic == 0))
summary(data)
```

```
##       year         statefip            met2013              perwt
##  Min.   :2019   Length:9632        Length:9632        Min.   :    2.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:   57.0
##  Median :2019   Mode  :character   Mode  :character   Median :   84.0
##  Mean   :2019                                         Mean   :  114.2
##  3rd Qu.:2019                                         3rd Qu.:  131.0
##  Max.   :2019                                         Max.   : 1977.0
##      sex                age                race              hispan
##  Length:9632        Length:9632        Length:9632        Length:9632
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                educd             empstat             uhrswork
##  Length:9632        Length:9632        Length:9632        Length:9632
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot             incwage         isHispanic isAfricanAmerican
```

3

```
## Min.   :       0   Min.   :      0   0:9632      0:5738
## 1st Qu.:  11800   1st Qu.:      0   1:   0      1:3894
## Median :  35000   Median : 24000
## Mean   :1606598   Mean   :196023
## 3rd Qu.:  86000   3rd Qu.: 80000
## Max.   :9999999   Max.   :999999
```

##1C age between 25 to 59

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

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
data <- data %>% filter(age >= 25 & age <= 59)
summary(data)
```

```
##       year         statefip            met2013             perwt
##  Min.   :2019   Length:4175        Length:4175        Min.   :    4.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:   60.0
##  Median :2019   Mode  :character   Mode  :character   Median :   89.0
##  Mean   :2019                                         Mean   :  121.1
##  3rd Qu.:2019                                         3rd Qu.:  138.0
##  Max.   :2019                                         Max.   : 1977.0
##      sex                 age              race              hispan
##  Length:4175        Min.   :25.00   Length:4175        Length:4175
##  Class :character   1st Qu.:34.00   Class :character   Class :character
##  Mode  :character   Median :44.00   Mode  :character   Mode  :character
##                     Mean   :43.18
##                     3rd Qu.:53.00
##                     Max.   :59.00
##      bpl                educd             empstat            uhrswork
##  Length:4175        Length:4175        Length:4175        Length:4175
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot           incwage       isHispanic isAfricanAmerican
##  Min.   :     0   Min.   :     0   0:4175      0:2470
##  1st Qu.: 12000   1st Qu.:  2750   1:   0      1:1705
##  Median : 35000   Median : 31700
##  Mean   : 49841   Mean   : 44533
##  3rd Qu.: 62000   3rd Qu.: 60000
##  Max.   :756000   Max.   :476000
```

##1d New variable for female respondents

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

```
##    0    1
## 1973 2202
```

```
summary(data)
```

```
##       year          statefip           met2013              perwt
##  Min.   :2019    Length:4175       Length:4175         Min.   :    4.0
##  1st Qu.:2019    Class :character  Class :character    1st Qu.:   60.0
##  Median :2019    Mode  :character  Mode  :character    Median :   89.0
##  Mean   :2019                                          Mean   :  121.1
##  3rd Qu.:2019                                          3rd Qu.:  138.0
##  Max.   :2019                                          Max.   : 1977.0
##      sex               age            race              hispan
##  Length:4175       Min.   :25.00   Length:4175       Length:4175
##  Class :character  1st Qu.:34.00   Class :character  Class :character
##  Mode  :character  Median :44.00   Mode  :character  Mode  :character
##                    Mean   :43.18
##                    3rd Qu.:53.00
##                    Max.   :59.00
##      bpl               educd            empstat            uhrswork
##  Length:4175       Length:4175       Length:4175       Length:4175
##  Class :character  Class :character  Class :character  Class :character
##  Mode  :character  Mode  :character  Mode  :character  Mode  :character
##
##
##
##      inctot           incwage        isHispanic isAfricanAmerican isFemale
##  Min.   :     0   Min.   :     0   0:4175     0:2470            0:1973
##  1st Qu.: 12000   1st Qu.:  2750   1:   0     1:1705            1:2202
##  Median : 35000   Median : 31700
##  Mean   : 49841   Mean   : 44533
##  3rd Qu.: 62000   3rd Qu.: 60000
##  Max.   :756000   Max.   :476000
```

##1e New education variable with categories: ##1)Less than high school (including GED recipients), ##2)high school degree, ##3)some college (including associates degree), ##4)bachelor's degree, ##5)graduate degree

```
##1)Less than high school (including GED recipients),
data$edu_level[data$educd != "Master's degree" & data$educd != "Bachelor's degree" & data$educd != '1 o
##2)high school degree,
data$edu_level[data$educd == "Regular high school diploma"] <- 'high school degree'
##3)some college (including associates degree),
data$edu_level[data$educd == '1 or more years of college credit, no degree'] <- 'some college'
data$edu_level[data$educd == "Associate's degree, type not specified"] <- 'some college'
##4)bachelor's degree,
data$edu_level[data$educd == "Bachelor's degree"] <- "bachelors degree"
##5)graduate degree
data$edu_level[data$educd == "Master's degree"] <- "graduate degree"




summary(data$edu_level)
```

```
##    Length     Class      Mode
```

```
##       4175 character character
```

```r
data <- data %>% mutate(edu_level = as.factor(edu_level))

summary(data$edu_level)
```

```
##      bachelors degree        graduate degree     high school degree
##                   864                    388                    897
## less than high school           some college
##                  1005                   1021
```

##1f DUMMY variable for whether the respondent is employed.

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

```
##    0    1
## 1005 3170
```

```r
summary(data)
```

```
##       year         statefip            met2013              perwt
##  Min.   :2019   Length:4175        Length:4175        Min.   :    4.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:   60.0
##  Median :2019   Mode  :character   Mode  :character   Median :   89.0
##  Mean   :2019                                         Mean   :  121.1
##  3rd Qu.:2019                                         3rd Qu.:  138.0
##  Max.   :2019                                         Max.   : 1977.0
##      sex                 age            race              hispan
##  Length:4175        Min.   :25.00   Length:4175        Length:4175
##  Class :character   1st Qu.:34.00   Class :character   Class :character
##  Mode  :character   Median :44.00   Mode  :character   Mode  :character
##                     Mean   :43.18
##                     3rd Qu.:53.00
##                     Max.   :59.00
##      bpl                educd             empstat             uhrswork
##  Length:4175        Length:4175        Length:4175        Length:4175
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot           incwage        isHispanic isAfricanAmerican isFemale
##  Min.   :     0   Min.   :     0   0:4175     0:2470            0:1973
##  1st Qu.: 12000   1st Qu.:  2750   1:   0     1:1705            1:2202
##  Median : 35000   Median : 31700
##  Mean   : 49841   Mean   : 44533
##  3rd Qu.: 62000   3rd Qu.: 60000
##  Max.   :756000   Max.   :476000
##                  edu_level      isEmployed
##  bachelors degree     : 864   0:1005
```

6

```
##   graduate degree     : 388    1:3170
##   high school degree   : 897
##   less than high school:1005
##   some college         :1021
##
```

Question 2 Compare the educational attainment of African American to white respondents.

```
data_edu_attain_AAVsWhite <-
  data %>% filter(race == 'Black/African American/Negro' | race == 'White') %>%
  group_by(race) %>% count(race, edu_level) %>% group_by(edu_level) %>%
  mutate(perc_freq = round(n / sum(n), 3)*100)
data_edu_attain_AAVsWhite
```

```
## # A tibble: 10 x 4
## # Groups:   edu_level [5]
##     race                         edu_level            n perc_freq
##     <chr>                        <fct>            <int>     <dbl>
##  1 Black/African American/Negro bachelors degree     215      24.9
##  2 Black/African American/Negro graduate degree      114      29.4
##  3 Black/African American/Negro high school degree    516      57.5
##  4 Black/African American/Negro less than high school 426      42.4
##  5 Black/African American/Negro some college          434      42.5
##  6 White                        bachelors degree      649      75.1
##  7 White                        graduate degree       274      70.6
##  8 White                        high school degree     381      42.5
##  9 White                        less than high school  579      57.6
## 10 White                        some college           587      57.5
```

```
  data_edu_attain_AAVsWhite %>%
  ggplot(aes(x = edu_level, y = perc_freq, fill = race)) +
  geom_col(position='dodge') +
  geom_text(aes(label = perc_freq), vjust = -0.5, hjust = 0.5,  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  xlab('Education Level')+
  ylab('Educational attainment of African/American
        as compared to White')
```
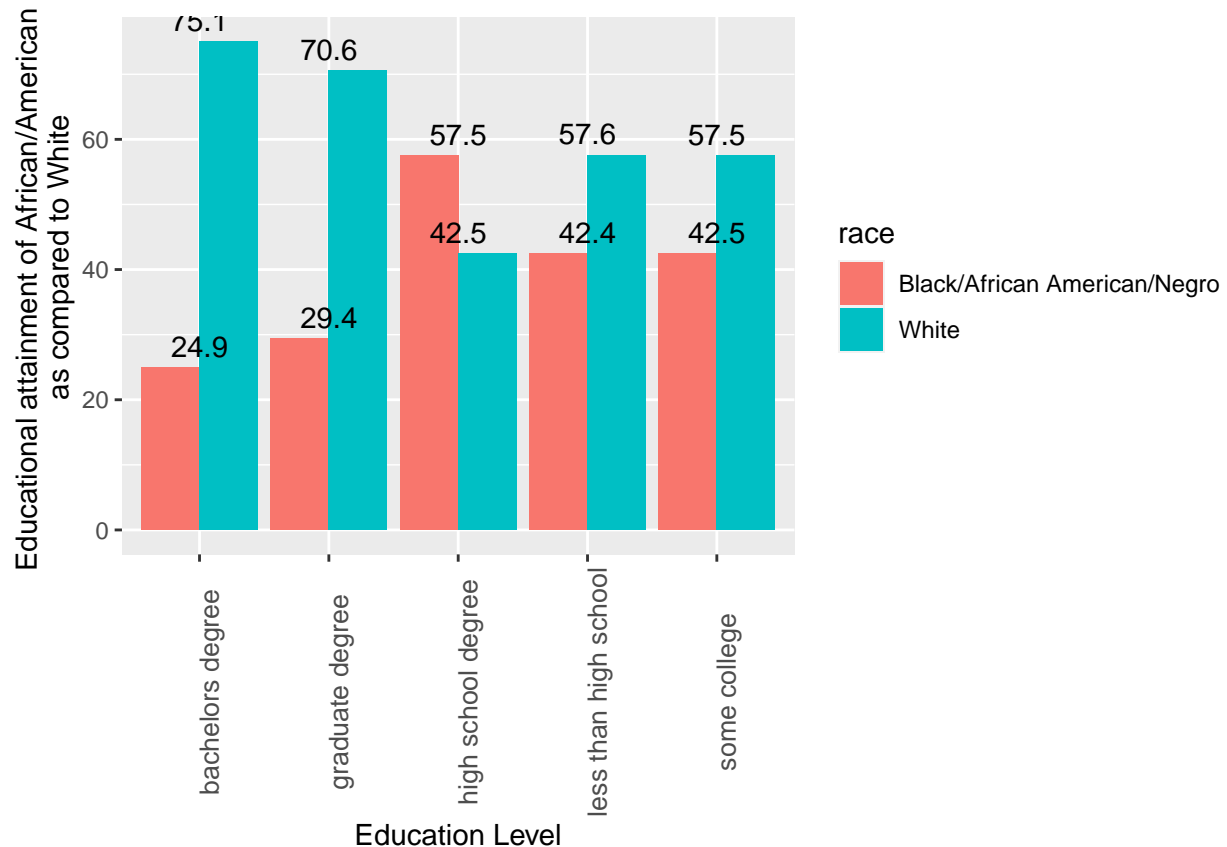
Question 3 Compare the employment rate by sex for African Americans and whites.

```
data_emp_by_sex_AAWhite <-
  data %>% filter(data$empstat != 'N/A' & data$empstat != 'Not in labor force' ) %>%
  filter(race == 'Black/African American/Negro' | race == 'White') %>%
  group_by(race, sex) %>% count(race, empstat) %>% group_by(race) %>%
  mutate(perc_freq = round(n / sum(n), 3)*100)

##checking if the filter happened as expected
summary(data_emp_by_sex_AAWhite)
```

```
##     race               sex               empstat                n
##  Length:8           Length:8           Length:8           Min.   :  23.00
##  Class :character   Class :character   Class :character   1st Qu.:  39.75
##  Mode  :character   Mode  :character   Mode  :character   Median :  266.50
##                                                           Mean   :  414.88
##                                                           3rd Qu.: 754.25
##                                                           Max.   :1061.00
##    perc_freq
##  Min.   : 1.100
```

8

```
## 1st Qu.: 2.875
## Median :21.000
## Mean   :25.000
## 3rd Qu.:47.075
## Max.   :54.700
```

```
head(data_emp_by_sex_AAWhite)
```

```
## # A tibble: 6 x 5
## # Groups:   race [2]
##   race                        sex    empstat        n perc_freq
##   <chr>                       <chr>  <chr>      <int>     <dbl>
## 1 Black/African American/Negro Female Employed     695      54.7
## 2 Black/African American/Negro Female Unemployed    51       4
## 3 Black/African American/Negro Male   Employed     482      38
## 4 Black/African American/Negro Male   Unemployed    42       3.3
## 5 White                        Female Employed     932      45.5
## 6 White                        Female Unemployed    33       1.6
```

```
##plotting the graph
data_emp_by_sex_AAWhite %>%
  ggplot(aes(x = empstat, y = perc_freq, fill = race)) +
  geom_col(position='dodge') +
  geom_text(aes(label = perc_freq),  colour = "black") +
  facet_grid(col = vars(sex)) +
  xlab('Employment status') +
  ylab('Employment rate by sex for African Americans and whites')
```
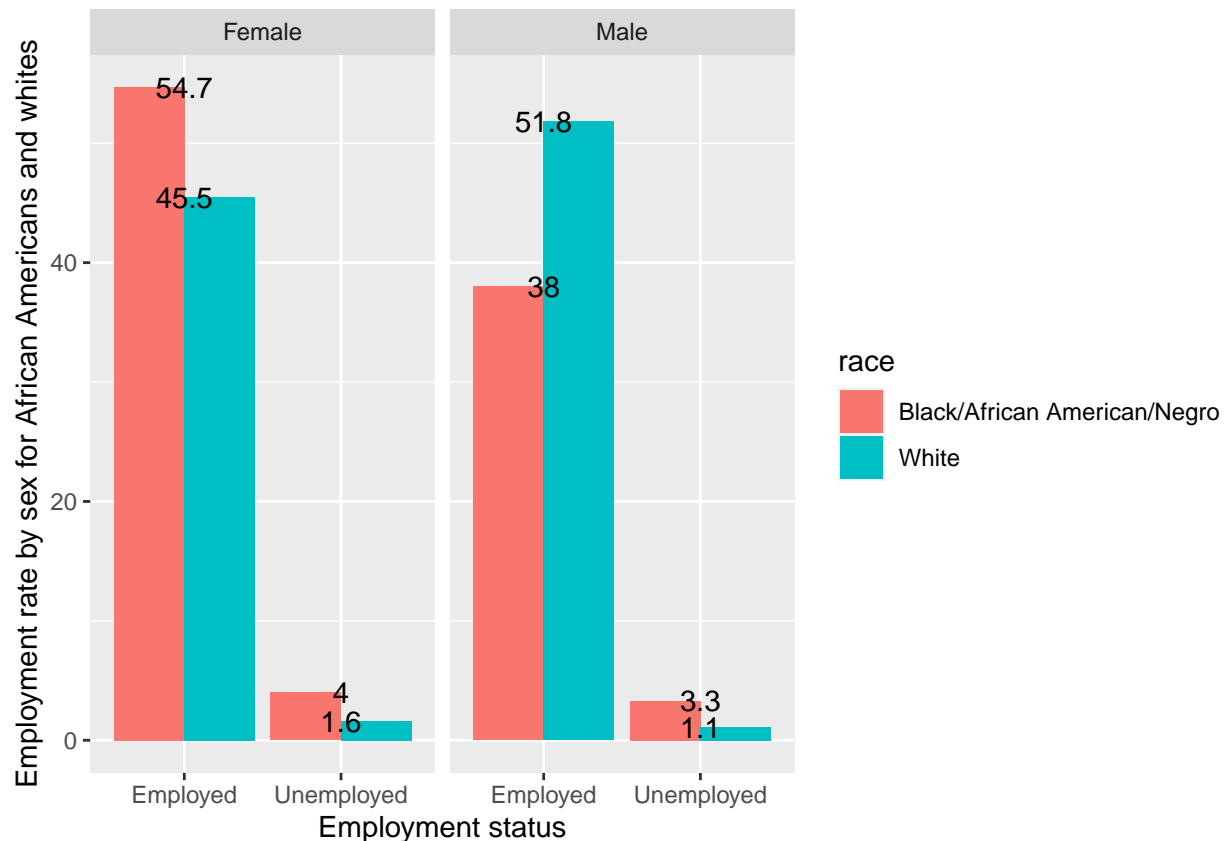
Employment rate by sex for African Americans and whites / Employment status

#Question4 Conditional on working, compare hours worked and its standard deviation by sex for African Americans and whites.

```
data_emp_hrs_sex_AAWhite <-
  data %>% filter(data$isEmployed == 1) %>%
  filter(race == 'Black/African American/Negro' | race == 'White') %>%
  group_by(race, sex) %>%
  summarize(sum_uhours = sum(as.numeric(uhrswork), na.rm = TRUE),
            mean_uhours = mean(as.numeric(uhrswork), na.rm = TRUE),
            sd_uhours = sd(as.numeric(uhrswork), na.rm = TRUE))
```

```
## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion
```

```
## Warning in is.data.frame(x): NAs introduced by coercion

## Warning in is.data.frame(x): NAs introduced by coercion

## Warning in is.data.frame(x): NAs introduced by coercion

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

```
summary(data_emp_hrs_sex_AAWhite)
```

```
##      race               sex               sum_uhours      mean_uhours
##  Length:4           Length:4           Min.   :20212    Min.   :39.42
##  Class :character   Class :character   1st Qu.:25641    1st Qu.:39.52
##  Mode  :character   Mode  :character   Median :32095    Median :40.79
##                                        Mean   :32783    Mean   :41.27
##                                        3rd Qu.:39237    3rd Qu.:42.54
##                                        Max.   :46730    Max.   :44.08
##     sd_uhours
##  Min.   : 9.662
##  1st Qu.: 9.881
##  Median :10.152
##  Mean   :10.109
##  3rd Qu.:10.379
##  Max.   :10.468
```

```
head(data_emp_hrs_sex_AAWhite)
```

```
## # A tibble: 4 x 5
## # Groups:   race [2]
##   race                        sex     sum_uhours mean_uhours sd_uhours
##   <chr>                       <chr>        <dbl>       <dbl>     <dbl>
## 1 Black/African American/Negro Female       27451        39.6      9.66
## 2 Black/African American/Negro Male         20212        42.0     10.5
## 3 White                       Female       36739        39.4     10.3
## 4 White                       Male         46730        44.1      9.95
```

```
data_emp_hrs_sex_AAWhite %>%
  ggplot(aes(x = sex, y = mean_uhours, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(mean_uhours)), colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  xlab('Sex')+
  ylab('Mean hours worked
        African Americans and whites grouped by sex')
```
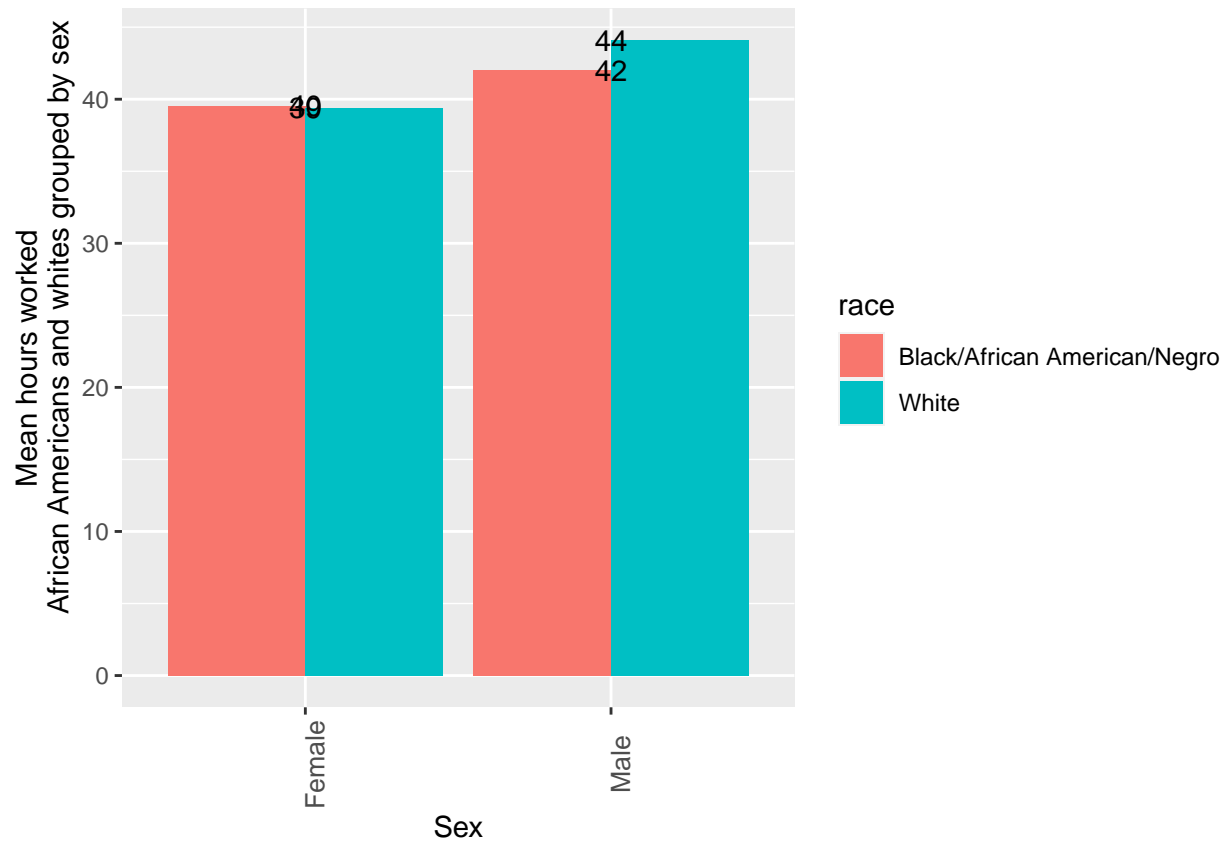
```
data_emp_hrs_sex_AAWhite %>%
  ggplot(aes(x = sex, y = sd_uhours, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sd_uhours)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  xlab('Sex')+
  ylab('Standard deviation of hours worked
        African Americans and whites grouped by sex')
```
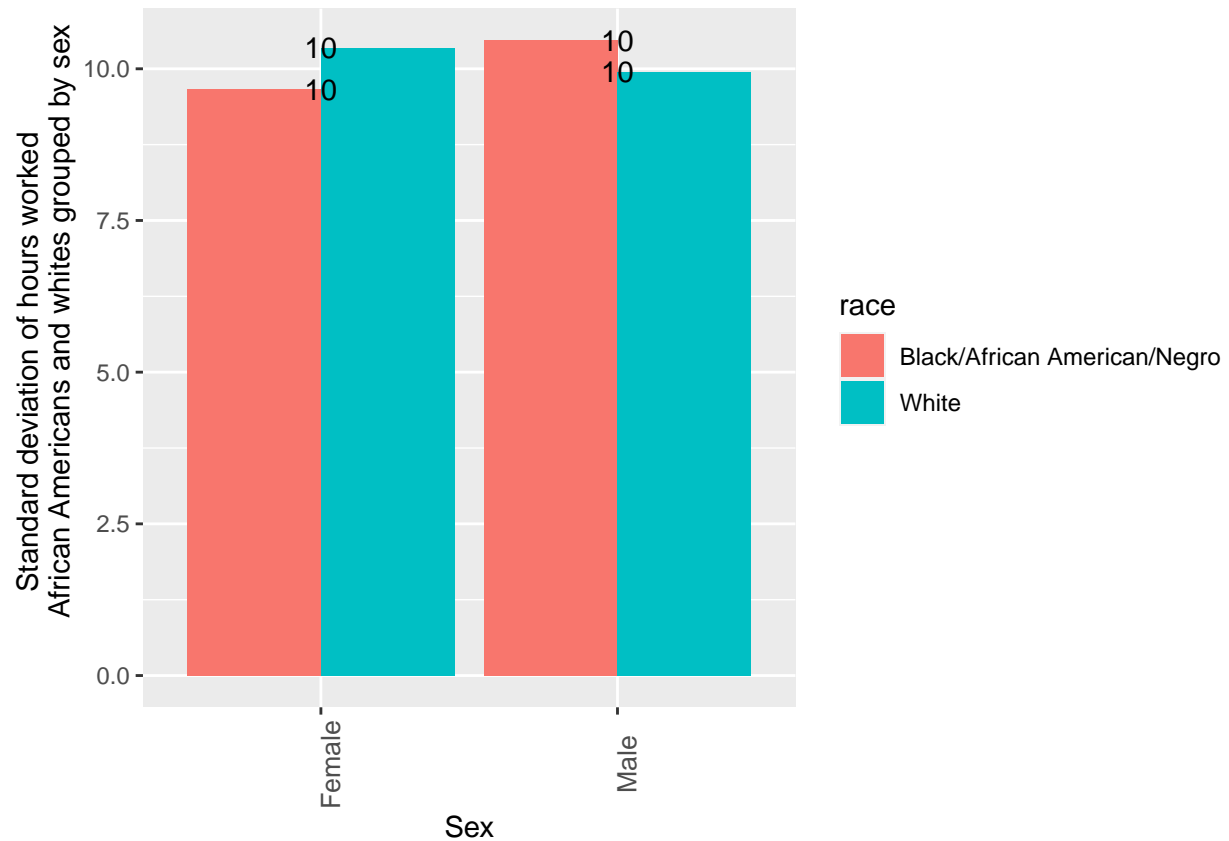
```
data_emp_hrs_sex_AAWhite %>%
  ggplot(aes(x = sex, y = sum_uhours, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sum_uhours)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  xlab('Sex')+
  ylab('Total hours worked
        African Americans and whites grouped by sex')
```
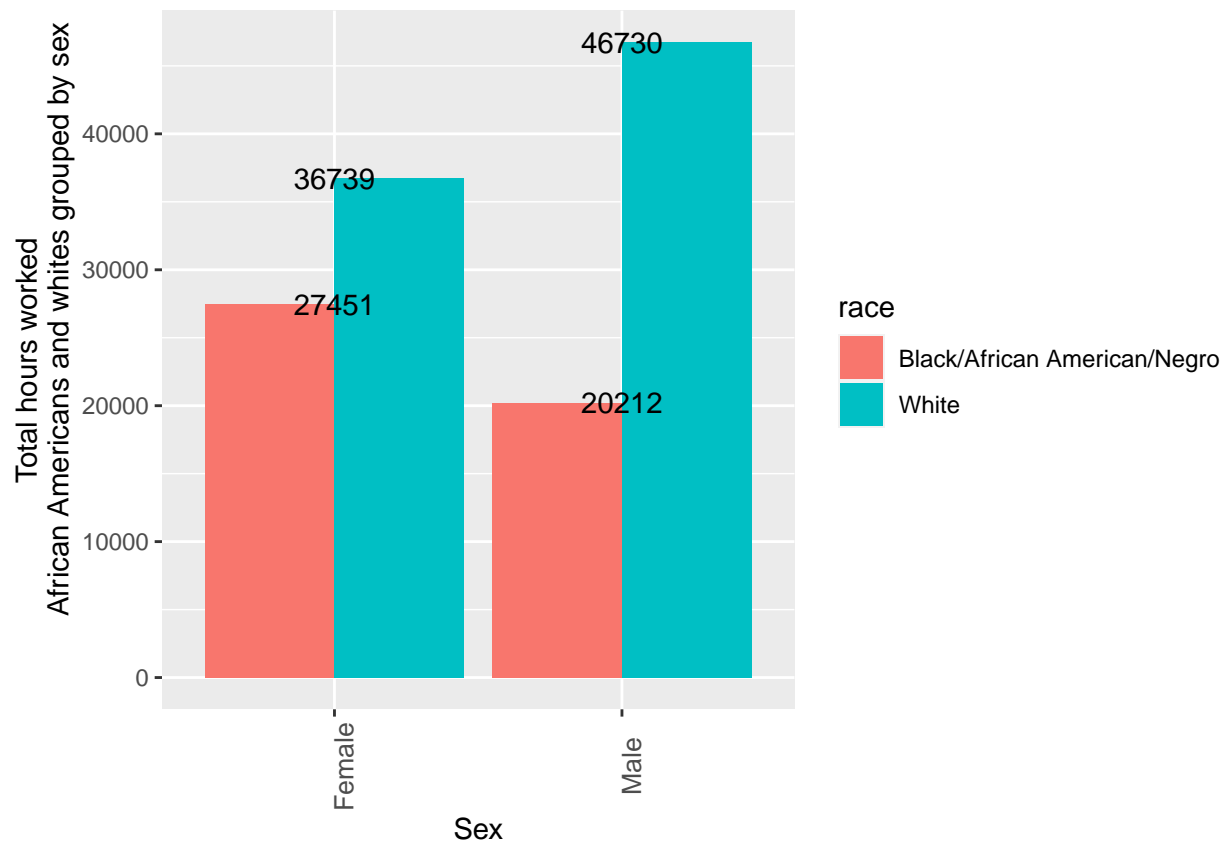
##Question5 Conditional on working, compare total income, its standard deviation, and its skewness by education and sex for African Americans and whites.

```
data_emp_income_sex_AAWhite <-
  data %>% filter(data$isEmployed == 1) %>%
  filter((race == 'Black/African American/Negro' | race == 'White') &
         (inctot != 9999999)) %>%
  group_by(race, edu_level, sex) %>%
  summarise(sum_income = sum(inctot, na.rm = TRUE),
            mean_income = mean(inctot, na.rm = TRUE),
            skew_income = skewness(inctot, na.rm = TRUE),
            sd_income = sd(as.numeric(inctot), na.rm = TRUE))
```

`summarise()` has grouped output by 'race', 'edu_level'. You can override using the `.groups` argumen

```
summary(data_emp_income_sex_AAWhite)
```

```
##      race                        edu_level       sex
##  Length:20          bachelors degree     :4   Length:20
##  Class :character   graduate degree      :4   Class :character
##  Mode  :character   high school degree   :4   Mode  :character
##                     less than high school:4
##                     some college         :4
##
##    sum_income         mean_income       skew_income        sd_income
```

14

```
##  Min.    : 1471500   Min.   : 26142   Min.   :0.6726   Min.    : 15531
##  1st Qu.: 4541395   1st Qu.: 41781   1st Qu.:2.1071   1st Qu.: 33007
##  Median : 6909502   Median : 56572   Median :2.5729   Median : 45004
##  Mean    : 9908165   Mean   : 60406   Mean   :2.8851   Mean    : 52410
##  3rd Qu.:11153788   3rd Qu.: 68219   3rd Qu.:3.6150   3rd Qu.: 57876
##  Max.    :31744790   Max.   :127067   Max.   :6.1633   Max.    :123310
```

```r
head(data_emp_income_sex_AAWhite)
```

```
## # A tibble: 6 x 7
## # Groups:   race, edu_level [3]
##   race          edu_level    sex    sum_income mean_income skew_income sd_income
##   <chr>         <fct>        <chr>       <int>       <dbl>       <dbl>     <dbl>
## 1 Black/African~ bachelors d~ Fema~     6046880      48765.        4.49    36337.
## 2 Black/African~ bachelors d~ Male      3675760      60258.        1.07    32705.
## 3 Black/African~ graduate de~ Fema~     5746100      67601.        3.08    47902.
## 4 Black/African~ graduate de~ Male      1471500      70071.        0.918   43764.
## 5 Black/African~ high school~ Fema~     4130510      26142.        0.673   15531.
## 6 Black/African~ high school~ Male      5917894      36757.        2.40    26456.
```

```r
tail(data_emp_income_sex_AAWhite)
```

```
## # A tibble: 6 x 7
## # Groups:   race, edu_level [3]
##   race  edu_level              sex    sum_income mean_income skew_income sd_income
##   <chr> <fct>                  <chr>       <int>       <dbl>       <dbl>     <dbl>
## 1 White high school degree     Female    4603700      35966.        2.01    22583.
## 2 White high school degree     Male      8302980      52885.        2.48    34837.
## 3 White less than high school  Female   10085540      63035.        3.63    71907.
## 4 White less than high school  Male     22708800      91200         2.73   123310.
## 5 White some college           Female    9358350      45429.        2.14    34261.
## 6 White some college           Male     17987640      66131.        2.99    50060.
```
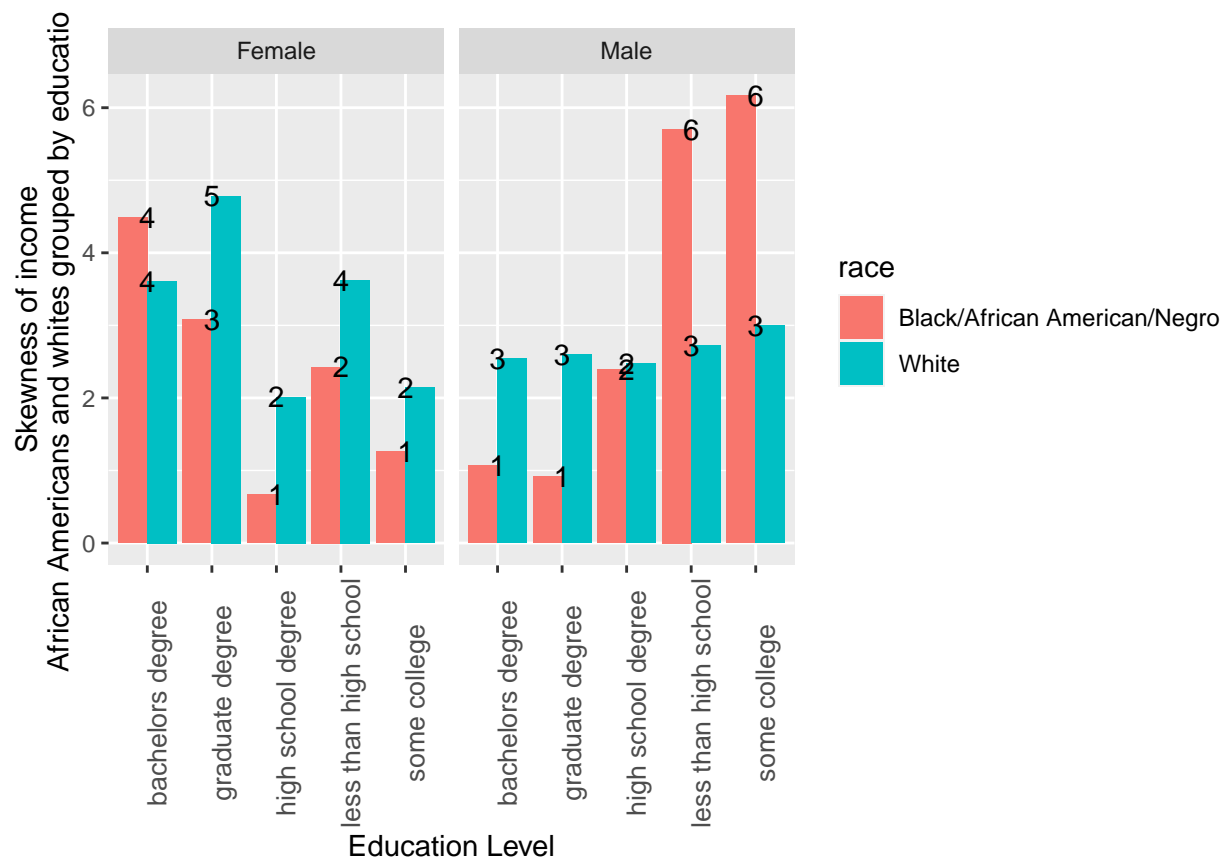
```r
data_emp_income_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = sum_income, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sum_income)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex)) +
  xlab('Education Level') +
  ylab('Total income of African Americans and whites
        grouped by education and sex ')
```
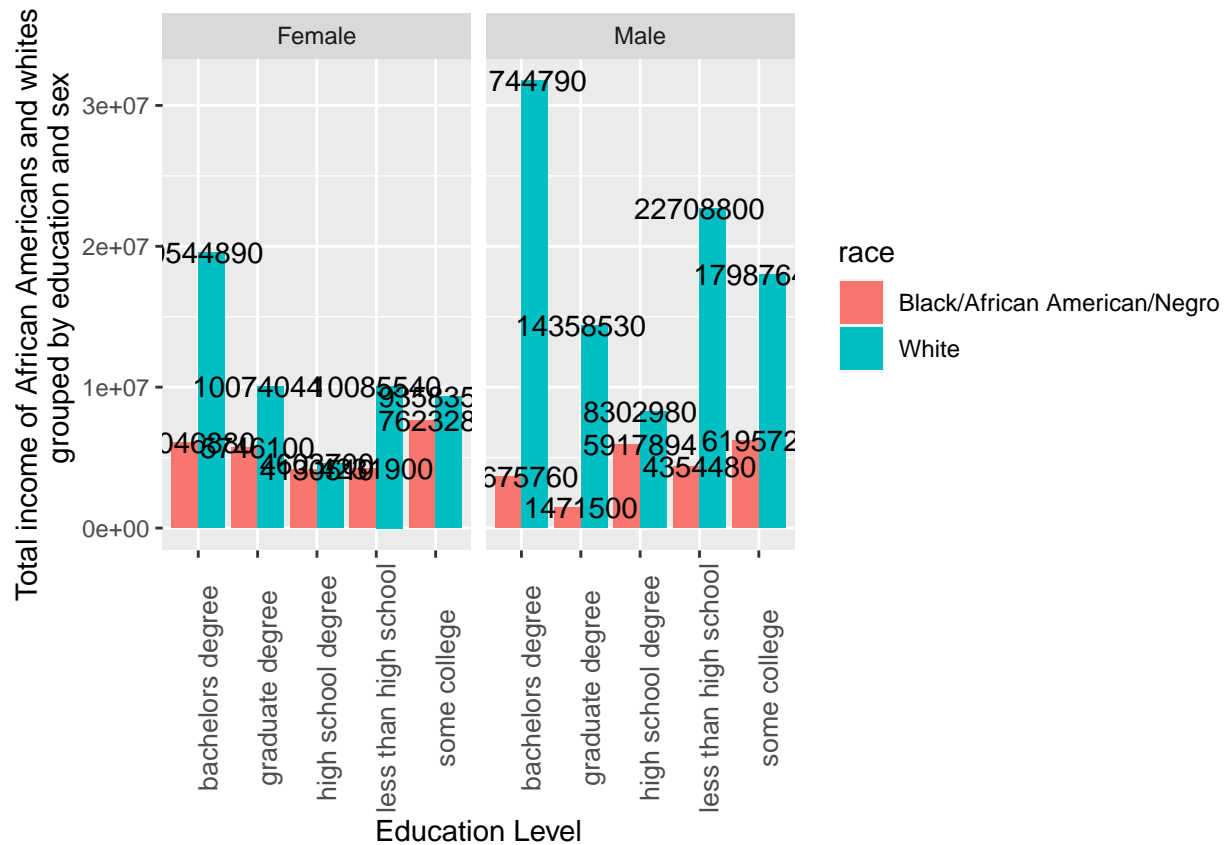
```
data_emp_income_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = mean_income, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(mean_income)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex)) +
  xlab('Education Level') +
  ylab('Mean income
        African Americans and whites grouped by education and sex')
```
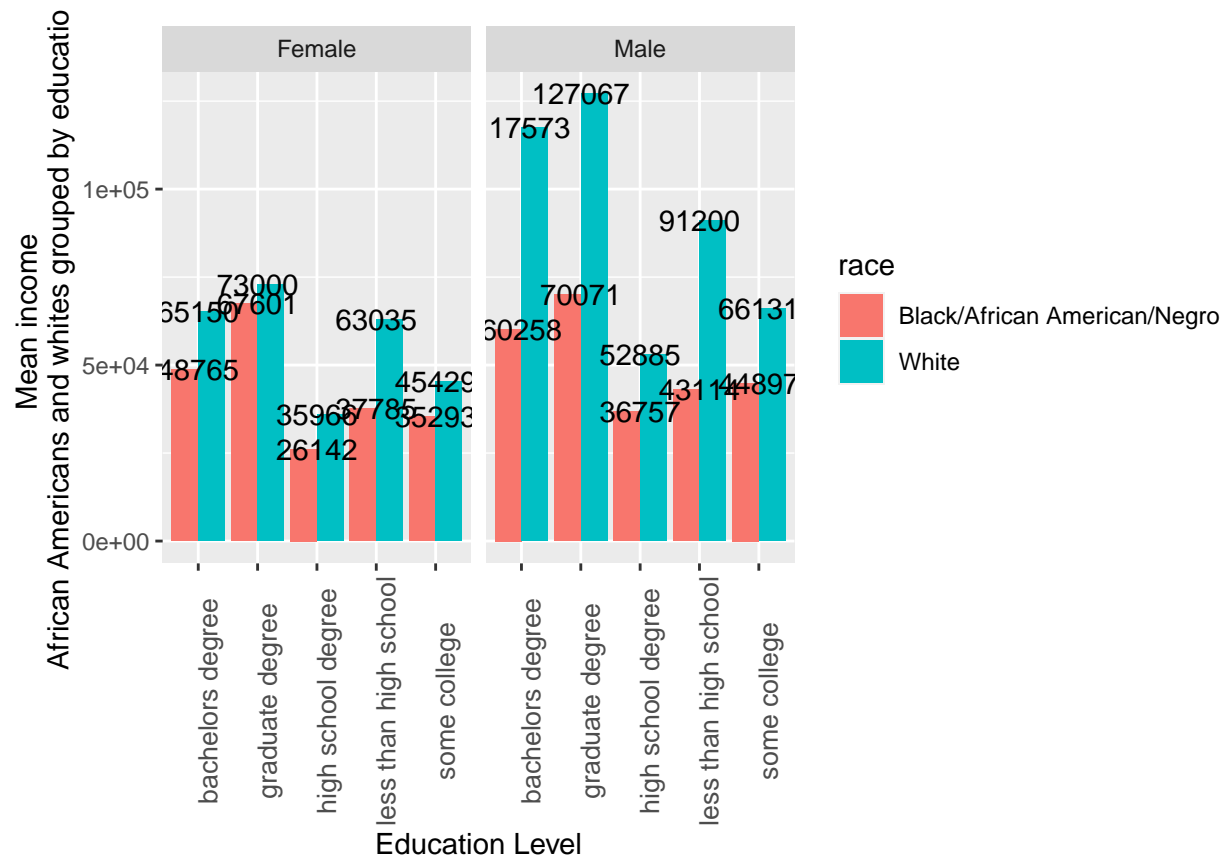
```
data_emp_income_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = skew_income, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(skew_income)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex)) +
  xlab('Education Level') +
  ylab('Skewness of income
        African Americans and whites grouped by education and sex')
```

```
data_emp_income_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = sd_income, fill = race )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sd_income)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex)) +
  xlab('Education Level') +
  ylab('Standard deviation of income
        African Americans and whites grouped by education and sex')
```
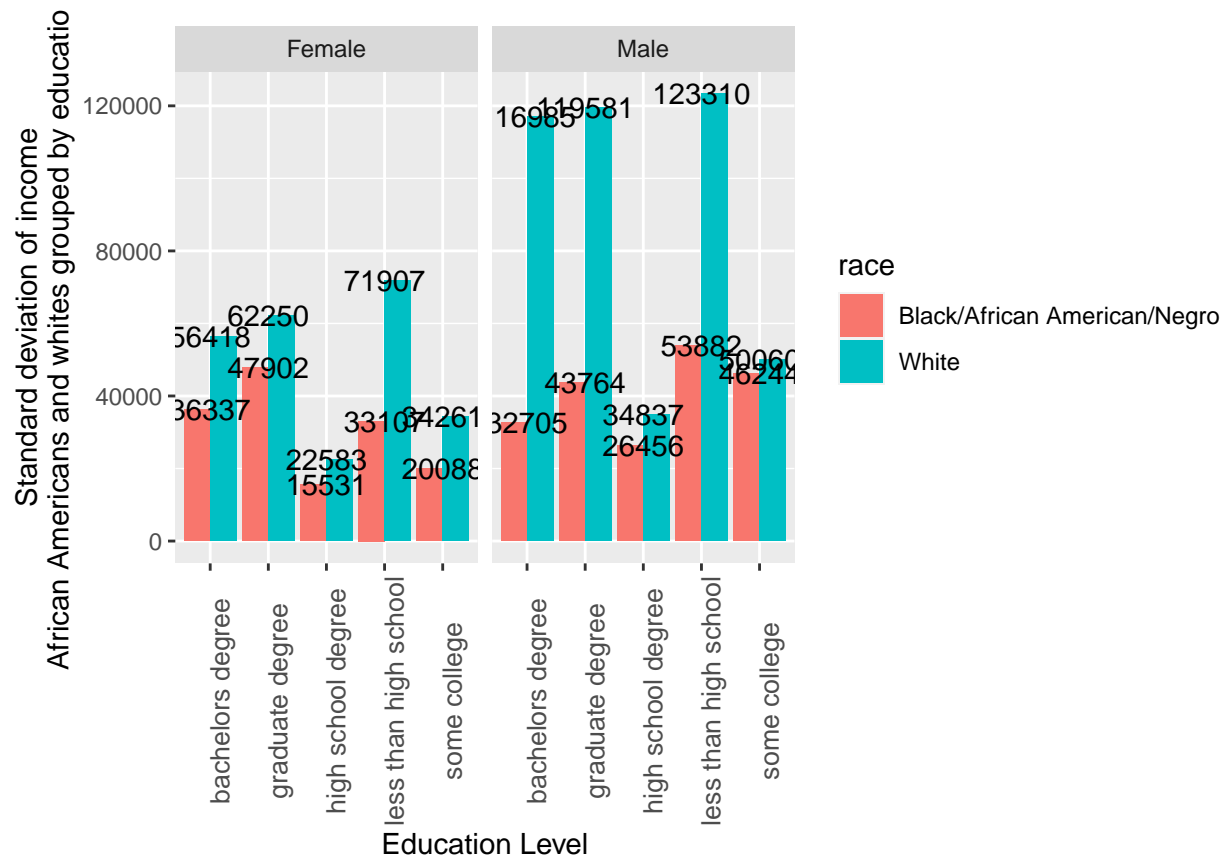
Question6 For those with positive wages, compare the wage income, its standard deviation, and its skewness by education and sex for African Americans and whites

```
##Removed the inconsistent values of incwage (99999) using filter before making a new variable and plot
data_wage_edu_sex_AA_White <-
  data %>% filter((race == 'Black/African American/Negro' | race == 'White')
                  & (incwage != 999999)) %>%
  group_by(race, edu_level, sex) %>%
  summarise(sum_incwage = sum(incwage, na.rm = TRUE),
            mean_incwage = mean(incwage, na.rm = TRUE),
            skew_incwage = skewness(incwage, na.rm = TRUE),
            sd_incwage = sd(incwage, na.rm = TRUE))
```

`summarise()` has grouped output by 'race', 'edu_level'. You can override using the `.groups` argumer

```
summary(data_wage_edu_sex_AA_White)
```

```
##      race                          edu_level      sex
##  Length:20         bachelors degree       :4    Length:20
##  Class :character  graduate degree        :4    Class :character
##  Mode  :character  high school degree     :4    Mode  :character
##                    less than high school:4
##                    some college           :4
##
##    sum_incwage        mean_incwage      skew_incwage      sd_incwage
```

```
## Min.    : 1361400    Min.    : 16427    Min.    :0.6696    Min.    : 17076
## 1st Qu.: 4376592    1st Qu.: 25623    1st Qu.:1.1125    1st Qu.: 29372
## Median : 6454375    Median : 40794    Median :1.9845    Median : 37481
## Mean    : 9296318    Mean    : 46550    Mean    :2.4566    Mean    : 47464
## 3rd Qu.:10460930    3rd Qu.: 57560    3rd Qu.:2.9663    3rd Qu.: 55843
## Max.    :28367550    Max.    :113322    Max.    :7.3605    Max.    :108285
```

```
head(data_wage_edu_sex_AA_White)
```

```
## # A tibble: 6 x 7
## # Groups:   race, edu_level [3]
##   race        edu_level    sex    sum_incwage mean_incwage skew_incwage sd_incwage
##   <chr>       <fct>        <chr>        <int>        <dbl>        <dbl>      <dbl>
## 1 Black/Afri~ bachelors ~  Fema~      5570100       38151.        0.724     27927.
## 2 Black/Afri~ bachelors ~  Male       3928600       56936.        0.819     35550.
## 3 Black/Afri~ graduate d~  Fema~      5188500       57016.        1.17      35553.
## 4 Black/Afri~ graduate d~  Male       1361400       59191.        0.900     47925.
## 5 Black/Afri~ high schoo~  Fema~      4106730       16427.        0.926     17076.
## 6 Black/Afri~ high schoo~  Male       5572900       20951.        1.44      24132.
```

```
tail(data_wage_edu_sex_AA_White)
```

```
## # A tibble: 6 x 7
## # Groups:   race, edu_level [3]
##   race  edu_level               sex    sum_incwage mean_incwage skew_incwage sd_incwage
##   <chr> <fct>                   <chr>        <int>        <dbl>        <dbl>      <dbl>
## 1 White high school degree      Female     4408800       22267.        1.67      24589.
## 2 White high school degree      Male       7948740       43436.        1.66      35396.
## 3 White less than high school   Female     9434400       37738.        3.89      60185.
## 4 White less than high school   Male      21869700       66473.        2.83     108285.
## 5 White some college            Female     9048180       31972.        2.00      34521.
## 6 White some college            Male      16321410       53689.        1.97      44002.
```
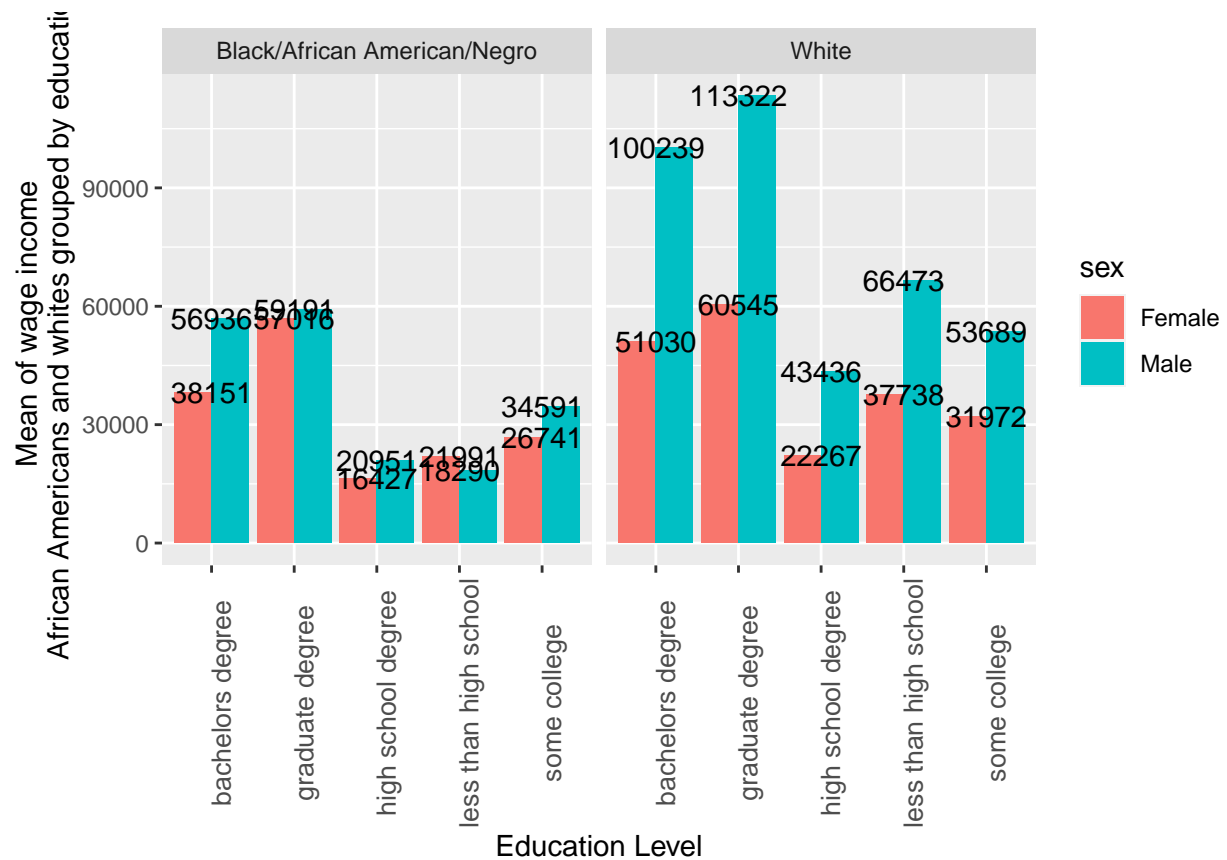
```
data_wage_edu_sex_AA_White %>%
  ggplot(aes(x = edu_level, y = sd_incwage, fill = sex )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sd_incwage)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(race)) +
  xlab('Education Level') +
  ylab('Standard deviation of wage income
        African Americans and Whites grouped by education & Sex')
```
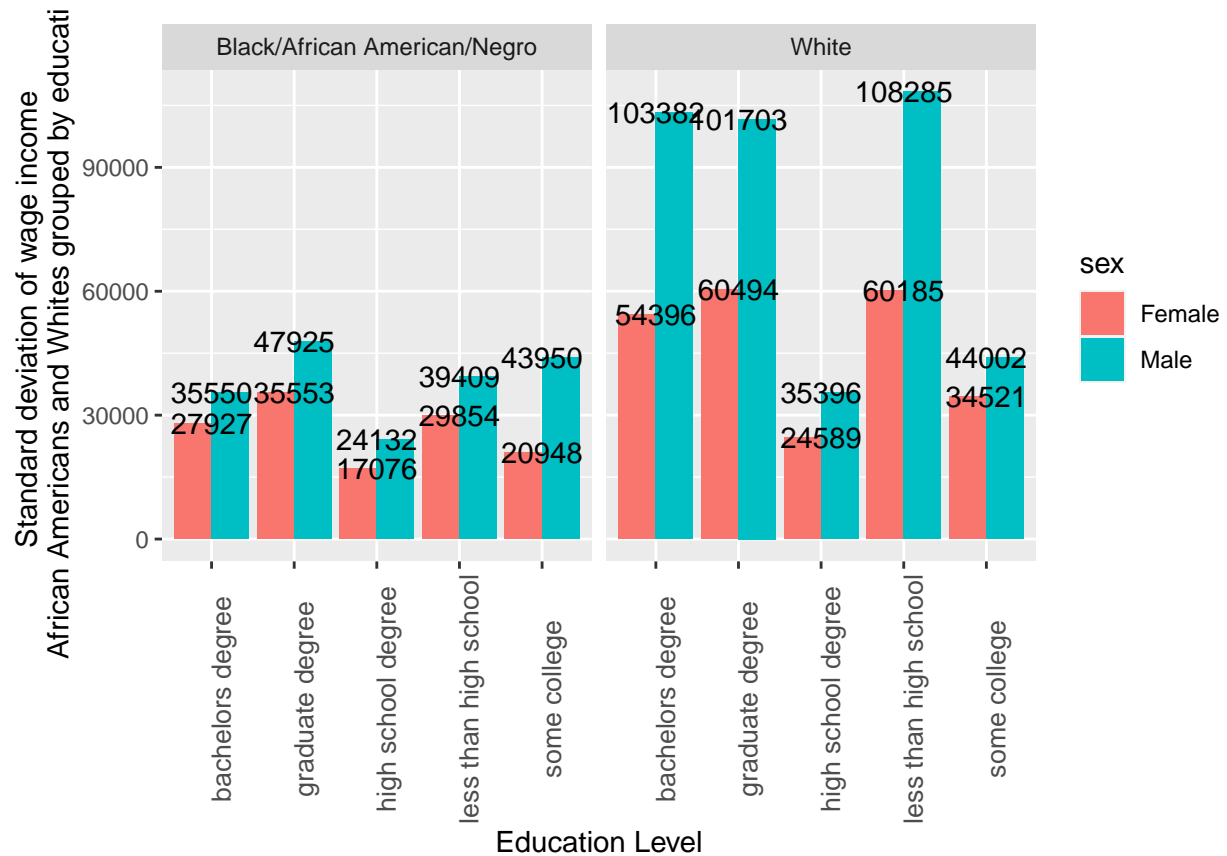
```
data_wage_edu_sex_AA_White %>%
  ggplot(aes(x = edu_level, y = skew_incwage, fill = sex )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(skew_incwage)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(race)) +
  xlab('Education Level') +
  ylab('Skewness of wage income
        African Americans and Whites grouped by education & Sex')
```
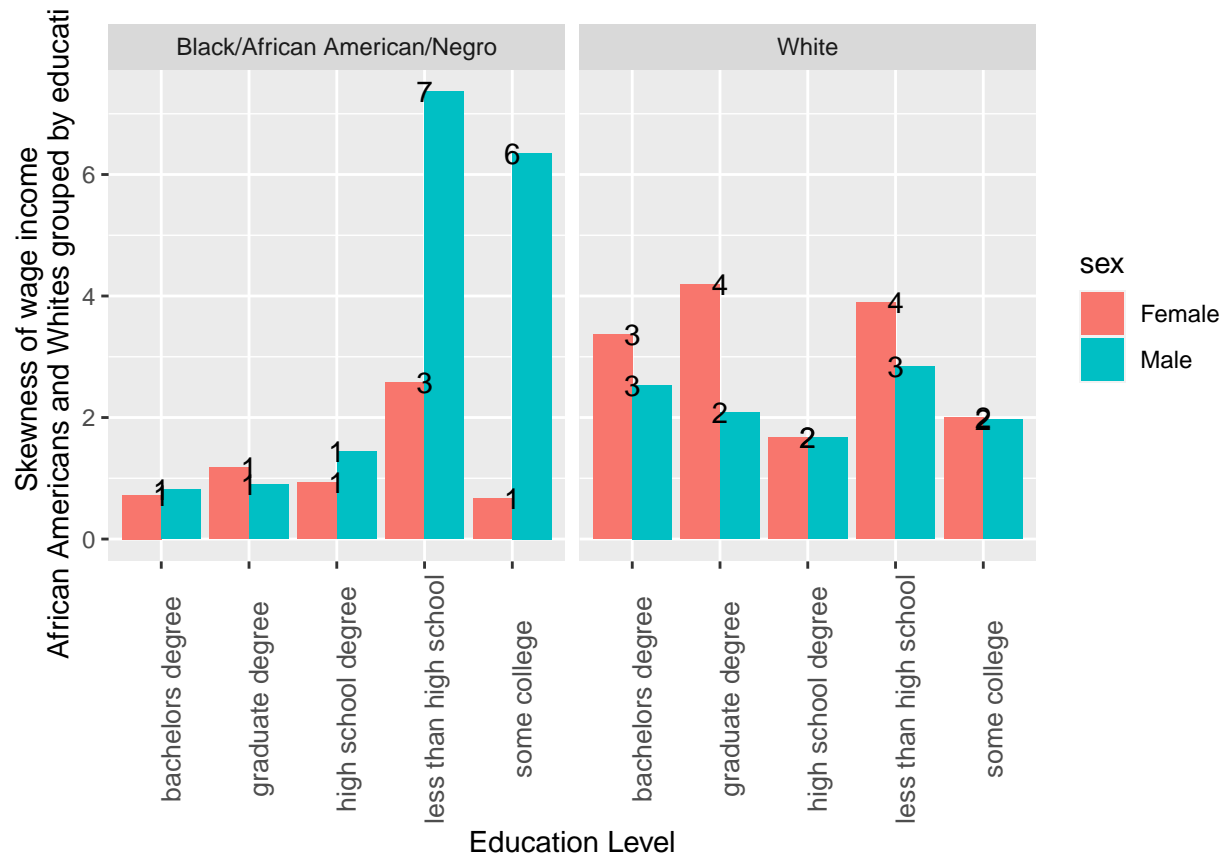
```
data_wage_edu_sex_AA_White %>%
  ggplot(aes(x = edu_level, y = mean_incwage, fill = sex )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(mean_incwage)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(race)) +
  xlab('Education Level') +
  ylab('Mean of wage income
        African Americans and whites grouped by education & Sex')
```

```
data_wage_edu_sex_AA_White %>%
  ggplot(aes(x = edu_level, y = sum_incwage, fill = sex )) +
  geom_col(position='dodge') +
  geom_text(aes(label = round(sum_incwage)),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(race)) +
  xlab('Education Level') +
  ylab('Total of wage income
        African Americans and whitess grouped by education & Sex')
```
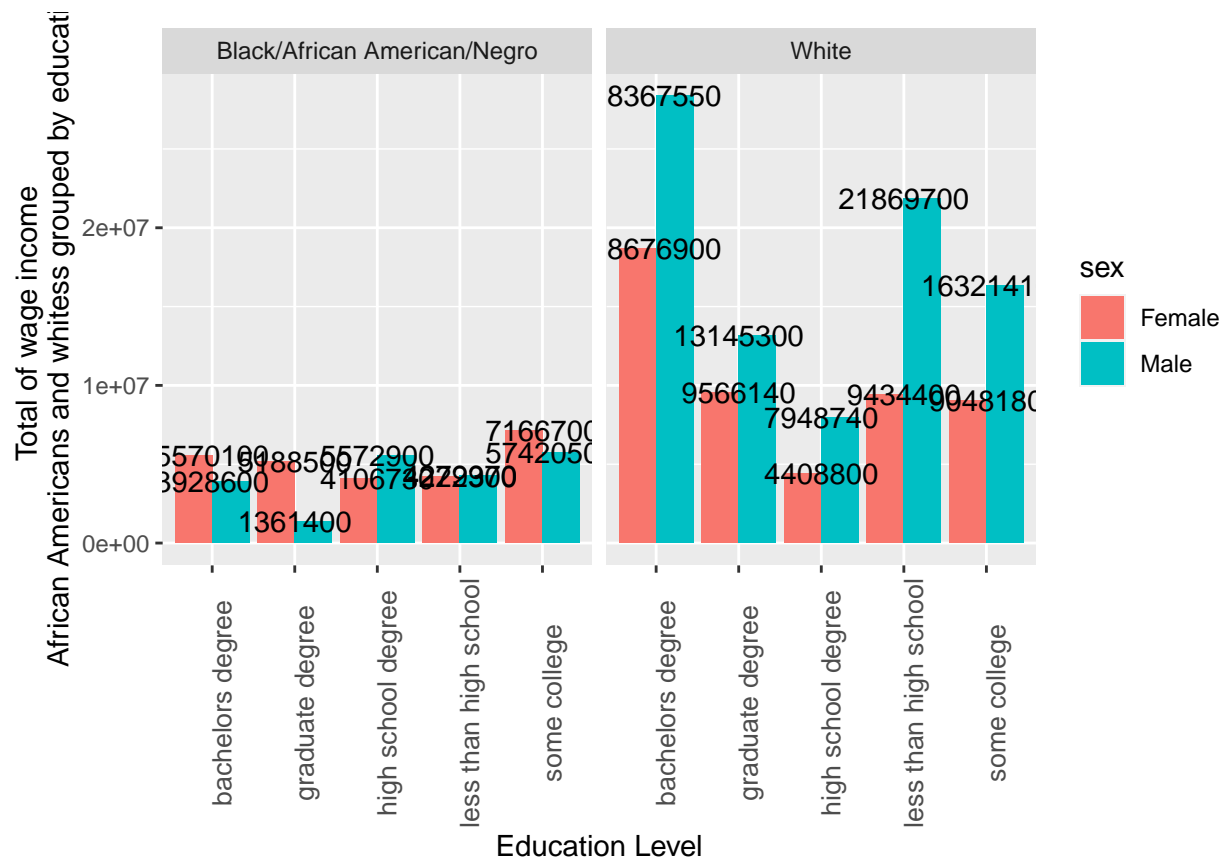
Question7 Calculate employment rates by age and sex for African Americans and whites.

```
data_empRate_age_sex_AAWhite <-
  data %>%
  filter(data$empstat != 'N/A' & data$empstat != 'Not in labor force' ) %>%
  filter(race == 'Black/African American/Negro' | race == 'White') %>%
  group_by(age, sex) %>% count(race, empstat) %>% group_by(age) %>%
  mutate(employment_rate = round(n / sum(n), 3)*100)

summary(data_empRate_age_sex_AAWhite)
```

```
##       age             sex                race              empstat
##  Min.   :25.00   Length:231         Length:231         Length:231
##  1st Qu.:32.00   Class :character   Class :character   Class :character
##  Median :41.00   Mode  :character   Mode  :character   Mode  :character
##  Mean   :41.32
##  3rd Qu.:50.00
##  Max.   :59.00
##        n           employment_rate
##  Min.   : 1.00   Min.   : 0.80
##  1st Qu.: 2.00   1st Qu.: 1.60
##  Median :14.00   Median :14.50
##  Mean   :14.37   Mean   :15.15
##  3rd Qu.:23.00   3rd Qu.:25.75
##  Max.   :55.00   Max.   :44.40
```

```
head(data_empRate_age_sex_AAWhite)
```

```
## # A tibble: 6 x 6
## # Groups:   age [1]
##     age sex    race                        empstat        n employment_rate
##   <dbl> <chr>  <chr>                       <chr>      <int>           <dbl>
## 1    25 Female Black/African American/Negro Employed       7            10.3
## 2    25 Female Black/African American/Negro Unemployed     3             4.4
## 3    25 Female White                        Employed      14            20.6
## 4    25 Male   Black/African American/Negro Employed      12            17.6
## 5    25 Male   Black/African American/Negro Unemployed     2             2.9
## 6    25 Male   White                        Employed      28            41.2
```

```
tail(data_empRate_age_sex_AAWhite)
```

```
## # A tibble: 6 x 6
## # Groups:   age [1]
##     age sex    race                        empstat        n employment_rate
##   <dbl> <chr>  <chr>                       <chr>      <int>           <dbl>
## 1    59 Female Black/African American/Negro Unemployed     4             3.1
## 2    59 Female White                        Employed      39            30.2
## 3    59 Female White                        Unemployed     1             0.8
## 4    59 Male   Black/African American/Negro Employed      17            13.2
## 5    59 Male   White                        Employed      33            25.6
## 6    59 Male   White                        Unemployed     1             0.8
```

```
data_empRate_age_sex_AAWhite %>%
  ggplot(aes(x = age, y = employment_rate, fill = race)) +
  geom_col(position='dodge', width = 1) +
  geom_text(aes(label = employment_rate),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex), space = 'free')+
  xlab('Age')+
  ylab('Employment rate
        African Americans and Whites grouped by Age & Sex')
```
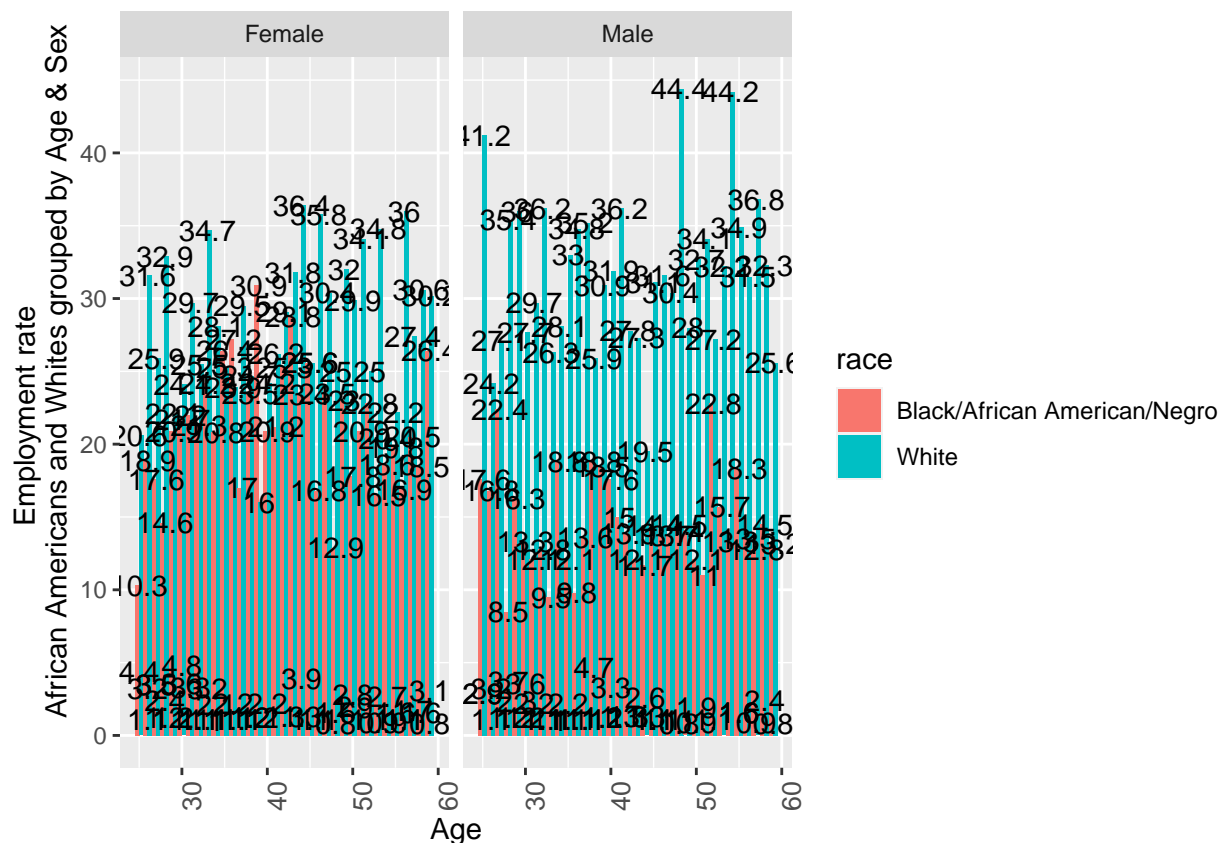
Question8 Conditional on working, compare the hours worked by education and sex for African Americans and whites.

```
data_emp_hrs_edu_sex_AAWhite <-
  data %>% filter(data$isEmployed == 1) %>%
  filter(race == 'Black/African American/Negro' | race == 'White') %>%
  group_by(race, edu_level, sex) %>%
  summarise(sum_uhours = sum(as.numeric(uhrswork), na.rm = TRUE),
            mean_uhours = mean(as.numeric(uhrswork), na.rm = TRUE),
            sd_uhours = sd(as.numeric(uhrswork), na.rm = TRUE))
```

```
## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mask$eval_all_summarise(quo): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion

## Warning in mean(as.numeric(uhrswork), na.rm = TRUE): NAs introduced by coercion

## Warning in is.data.frame(x): NAs introduced by coercion
```

```
## Warning in is.data.frame(x): NAs introduced by coercion

## Warning in is.data.frame(x): NAs introduced by coercion

## 'summarise()' has grouped output by 'race', 'edu_level'. You can override using the '.groups' argumen
```

```
summary(data_emp_hrs_edu_sex_AAWhite)
```

```
##      race                    edu_level     sex
##  Length:20         bachelors degree     :4   Length:20
##  Class :character  graduate degree      :4   Class :character
##  Mode  :character  high school degree   :4   Mode  :character
##                    less than high school:4
##                    some college         :4
##
##    sum_uhours      mean_uhours       sd_uhours
##  Min.   :  869   Min.   :38.24   Min.   : 6.809
##  1st Qu.: 4758   1st Qu.:39.22   1st Qu.: 9.032
##  Median : 5994   Median :41.45   Median :10.282
##  Mean   : 6557   Mean   :41.35   Mean   : 9.793
##  3rd Qu.: 8102   3rd Qu.:43.54   3rd Qu.:10.680
##  Max.   :12206   Max.   :44.88   Max.   :11.853
```

```
head(data_emp_hrs_edu_sex_AAWhite)
```

```
## # A tibble: 6 x 6
## # Groups:   race, edu_level [3]
##   race                        edu_level  sex    sum_uhours mean_uhours sd_uhours
##   <chr>                       <fct>      <chr>       <dbl>       <dbl>     <dbl>
## 1 Black/African American/Negro bachelors~ Fema~        4945        39.9      6.81
## 2 Black/African American/Negro bachelors~ Male         2679        43.9     10.5
## 3 Black/African American/Negro graduate ~ Fema~        3551        41.8      7.24
## 4 Black/African American/Negro graduate ~ Male          869        41.4      8.86
## 5 Black/African American/Negro high scho~ Fema~        6083        38.5     10.5
## 6 Black/African American/Negro high scho~ Male         6684        41.5      9.69
```

```
tail(data_emp_hrs_edu_sex_AAWhite)
```

```
## # A tibble: 6 x 6
## # Groups:   race, edu_level [3]
##   race  edu_level             sex     sum_uhours mean_uhours sd_uhours
##   <chr> <fct>                 <chr>        <dbl>       <dbl>     <dbl>
## 1 White high school degree    Female        4895        38.2      8.50
## 2 White high school degree    Male          6862        43.7      9.09
## 3 White less than high school Female        6283        39.3     10.0
## 4 White less than high school Male         10827        43.5     10.9
## 5 White some college          Female        7962        38.7     10.2
## 6 White some college          Male         12206        44.9     10.9
```

```
data_emp_hrs_edu_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = sum_uhours, fill = race)) +
  geom_col(position='dodge') +
  geom_text(aes(label = sum_uhours),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(col = vars(sex)) +
  xlab('Education Level')+
  ylab('Hours worked
          African Americans and whites grouped by Education & Sex')
```
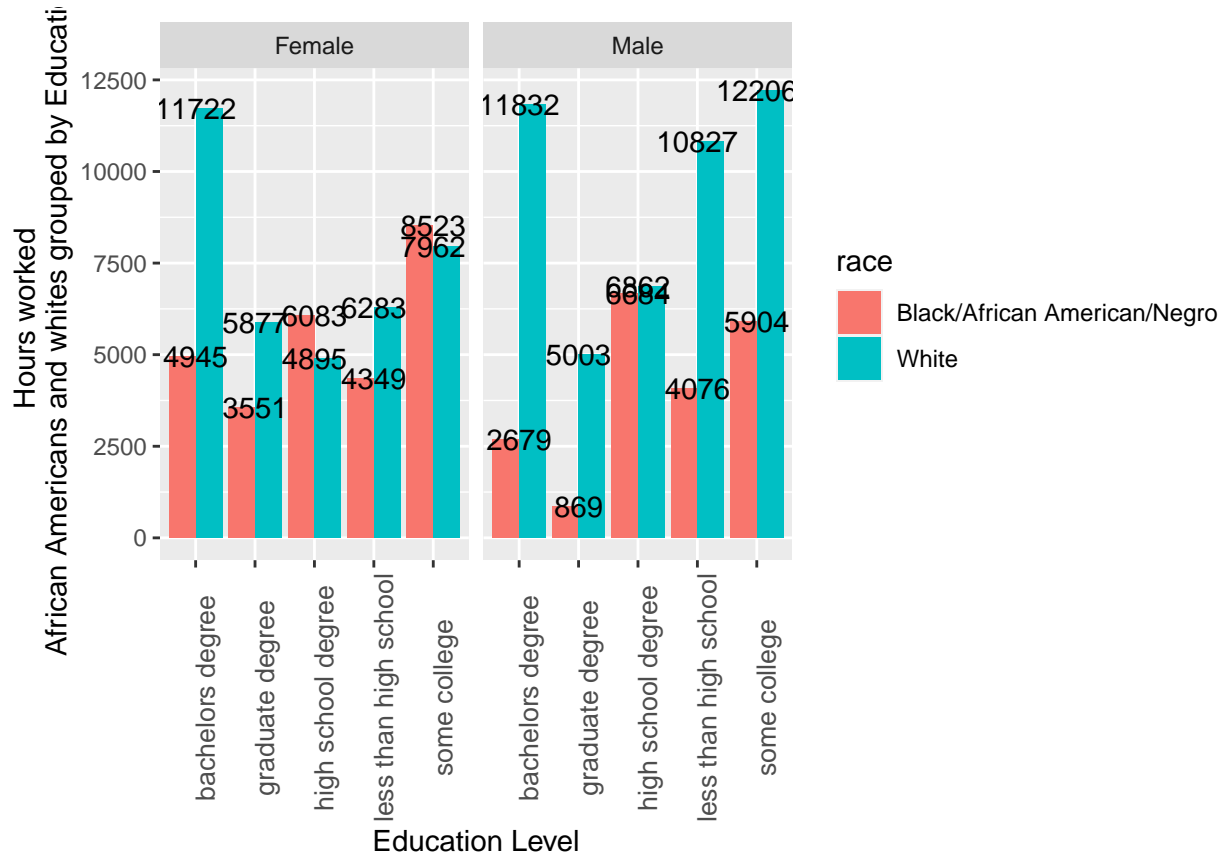
#Mean

```
data_emp_hrs_edu_sex_AAWhite %>%
  ggplot(aes(x = edu_level, y = mean_uhours, fill = race)) +
  geom_col(position='dodge') +
  geom_text(aes(label = mean_uhours),  colour = "black") +
  theme(axis.text.x=element_text(size=10, angle=90))+
  facet_grid(row = vars(race), col = vars(sex), space = 'free')+
  ylab('Mean Hours worked
        African Americans and whites grouped by Education and sex')
```

Mean Hours worked

African Americans and whites grouped by Education

edu_level

race

Black/African American/Negro

White