Wage Difference Between Local Government Workers and Private Sector Workers

Eric Liu

2023-03-06

Assignment 1: Download the Data

```
# prefer to use the raw STATA data set here
library(haven)
morg96 = read_dta("morg96.dta")
```

Assignment 2: Clean the Data and Perform Basic Analysis

Clean and Filter the Data

```
# workers ages 25 to 55
age = which(morg96$age %in% c(25:55))
morg96_1 = morg96[c(age), ]
# working at least 35 hours during his last week
hourslw = which(morg96_1$hourslw %in% c(35:99))
morg96_2 = morg96_1[c(hourslw), ]
# exclude workers who report being self employed
self = which(morg96_2$class94 %in% c(6,7))
morg96_new = morg96_2[-self, ]
# summary(morg96_new)
# str(morg96_new)
```

Select Four Seperate Sectors

```
# private sector workers
a1 = which(morg96_new$class94 %in% c(4,5))
private = morg96_new[c(a1), ]
# local government workers
a2 =  which(morg96_new$class94 %in% c(3))
local = morg96_new[c(a2), ]
# state government workers
a3 = which(morg96_new$class94 %in% c(2))
state = morg96_new[c(a3), ]
# federal government workers
a4 = which(morg96_new$class94 %in% c(1))
federal = morg96_new[c(a4), ]
```

Summary of Private Sector Workers

```r
# install.packages(psych)
library(psych)
# get the min, max, mean, and SD of private sector workers
describe(private$earnwke)
```

```
##    vars     n   mean     sd median trimmed    mad min  max range skew kurtosis
## X1    1 71709 605.71 373.42    500  549.49 284.66   0 1923  1923  1.5     2.39
##      se
## X1 1.39
```

```r
describe(private$age)
```

```
##    vars     n  mean   sd median trimmed   mad min max range skew kurtosis   se
## X1    1 71709 38.47 8.32     38   38.25 10.38  25  55    30 0.18    -1.01 0.03
```

```r
# rename database
private_sex = as.data.frame(table(private$sex))
names(private_sex) = c("Sex","Count")
private_sex[ ,1] = c("Male", "Female")
private_schooling = as.data.frame(table(private$grade92))
names(private_schooling) = c("Education","Count")
private_schooling[ ,1] = c("Less than 1st Grade","1st - 4th Grade","5th or 6th Grade",
                           "7th or 8th Grade","9th Grade","10th Grade","11th Grade",
                           "12th Grade NO DIPLOMA","High School Graduate, Diploma or GED",
                           "Some College but No Degree","Associate Degree -- Occupational/Vocational",
                           "Associate Degree -- Academic Program","Bachelor's Degree (e.g. BA,AB,BS)",
                           "Master's Degree (e.g. MA,MS,MEng,Med,MSW,MBA)",
                           "Professional School Deg (e.g. MD,DDS,DVM,LLB,JD)",
                           "Doctorate Degree (e.g. PhD, EdD)")
# perform bar plots of sex and years of schooling completed
# install.packages(ggplot)
library(ggplot2)
ggplot(private_sex, aes(x = Sex, y = Count, fill = Sex)) +
    ggtitle("Private Sector Workers by Sex") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Sex") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal()
```

Private Sector Workers by Sex
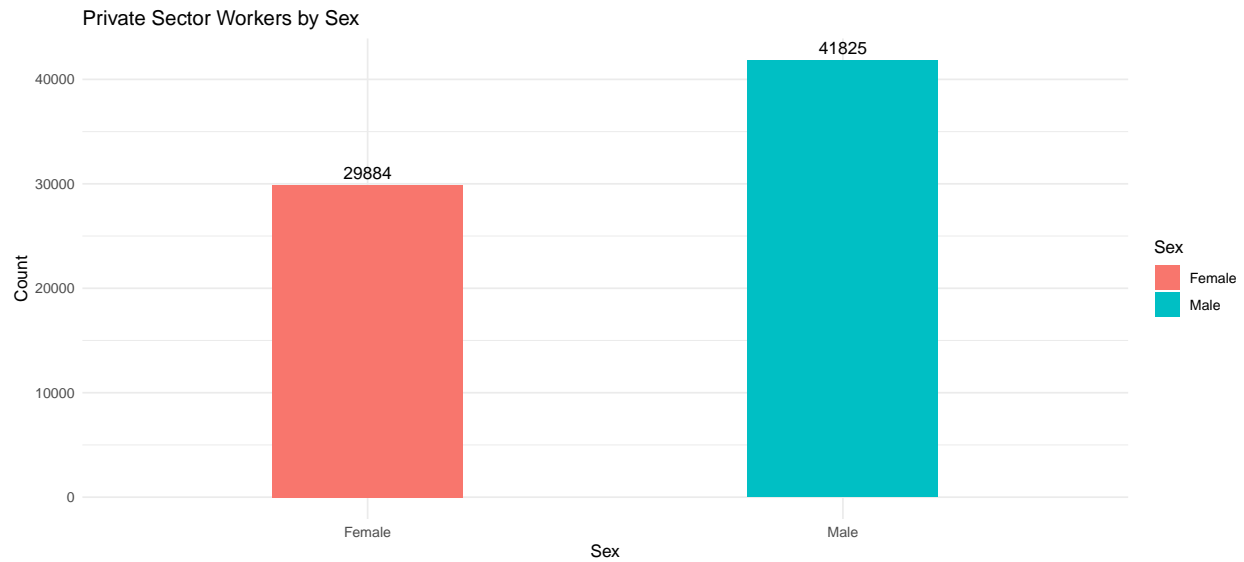

```
ggplot(private_schooling, aes(x = Education, y = Count, fill = Education)) +
    ggtitle("Private Sector Workers by Years of Schooling Completed") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Education") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal() +
    theme(axis.text.x = element_blank())
```

Private Sector Workers by Years of Schooling Completed
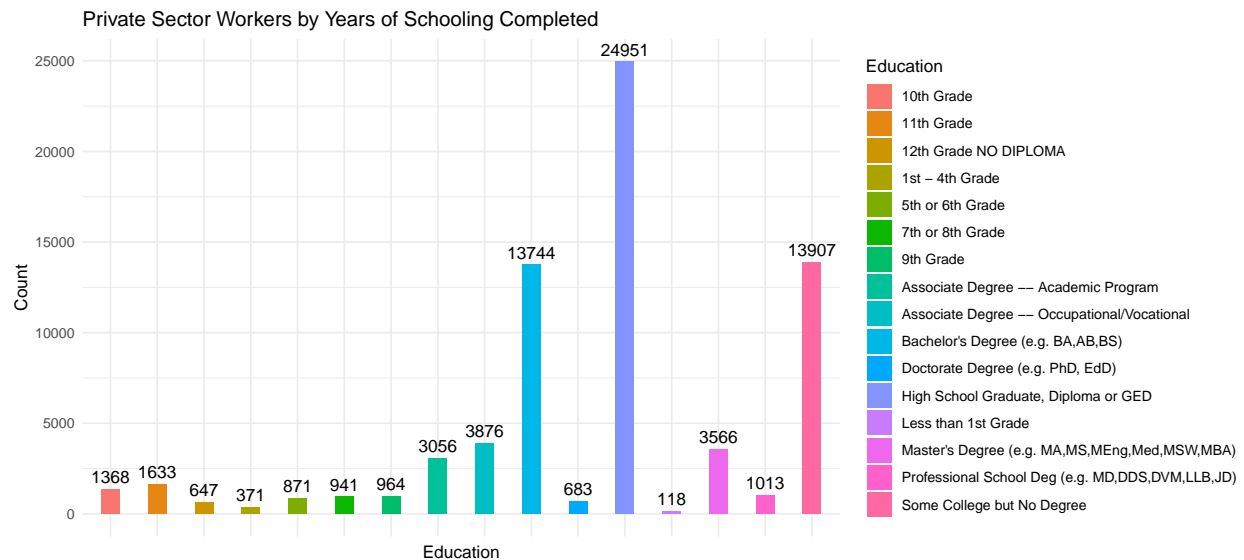
Private Sector Workers

Parameters	Min	Max	Mean	Standard Deviation
Private Sector Workers' Usual Weekly Earnings	0	1923	605.71	373.42
Private Sector Workers' Age	25	55	38.47	8.32

Summary of Local Government Workers

```
describe(local$earnwke)
```

```
##    vars    n   mean     sd median trimmed    mad min  max range skew kurtosis
## X1    1 7697 633.32 330.24    576  598.21 286.14   0 1923  1923 1.13     1.62
##      se
## X1 3.76
```

```
describe(local$age)
```

```
##    vars    n  mean  sd median trimmed   mad min max range  skew kurtosis   se
## X1    1 7697 40.85 8.3     42   41.05 10.38  25  55    30 -0.19    -1.03 0.09
```

```r
local_sex = as.data.frame(table(local$sex))
names(local_sex) = c("Sex","Count")
local_sex[ ,1] = c("Male", "Female")
local_schooling = as.data.frame(table(local$grade92))
names(local_schooling) = c("Education","Count")
local_schooling[ ,1] = c("Less than 1st Grade","1st - 4th Grade","5th or 6th Grade",
                         "7th or 8th Grade","9th Grade","10th Grade","11th Grade",
                         "12th Grade NO DIPLOMA","High School Graduate, Diploma or GED",
                         "Some College but No Degree","Associate Degree -- Occupational/Vocational",
                         "Associate Degree -- Academic Program","Bachelor's Degree (e.g. BA,AB,BS)",
                         "Master's Degree (e.g. MA,MS,MEng,Med,MSW,MBA)",
                         "Professional School Deg (e.g. MD,DDS,DVM,LLB,JD)",
                         "Doctorate Degree (e.g. PhD, EdD)")
```

```r
ggplot(local_sex, aes(x = Sex, y = Count, fill = Sex)) +
    ggtitle("Local Government Workers by Sex") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Sex") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal()
```
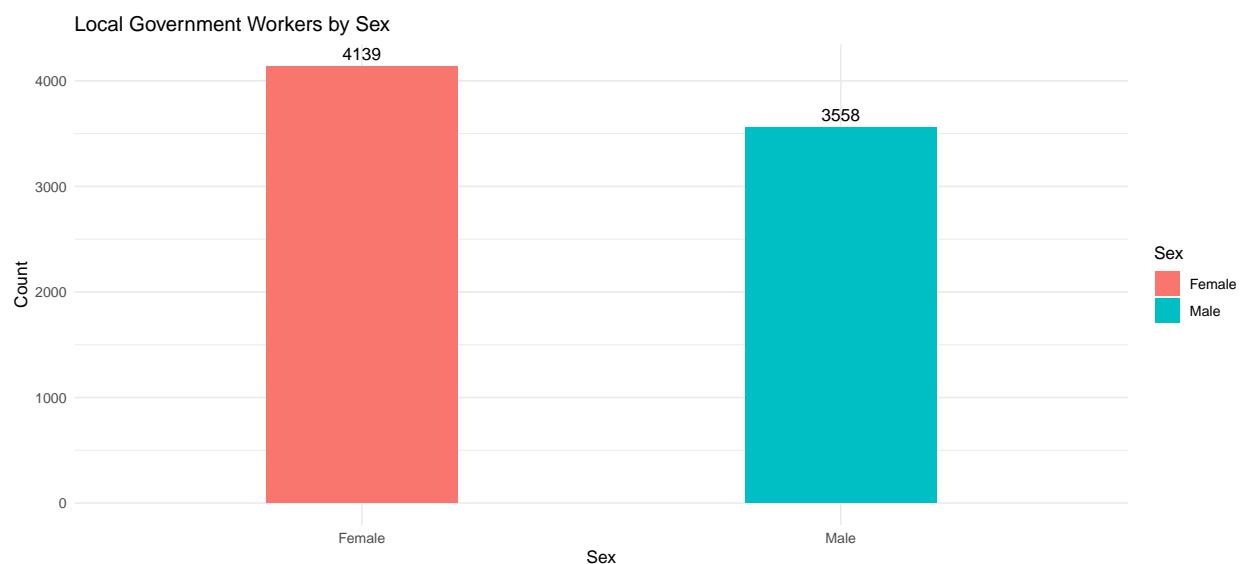
```
ggplot(local_schooling, aes(x = Education, y = Count, fill = Education)) +
    ggtitle("Local Government Workers by Years of Schooling Completed") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Education") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal() +
    theme(axis.text.x = element_blank())
```
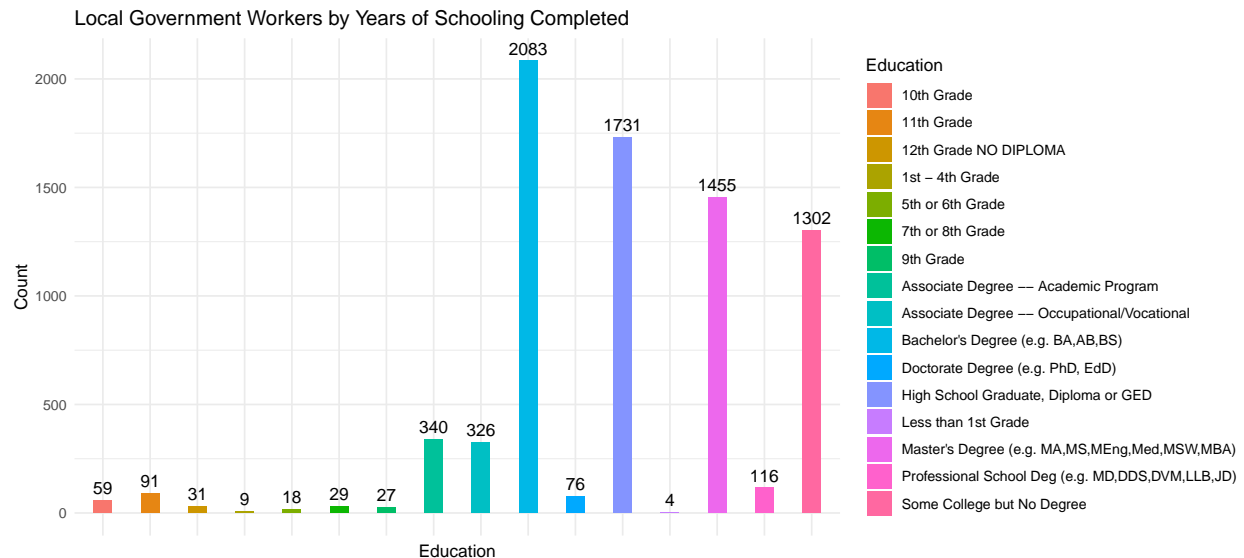

Local Government Workers

Parameters	Min	Max	Mean	Standard Deviation
Local Government Workers' Usual Weekly Earnings	0	1923	633.32	330.24
Local Government Workers' Age	25	55	40.85	8.3

Summary of State Government Workers

```
describe(state$earnwke)
```

```
##      vars    n   mean     sd median trimmed    mad min  max range skew kurtosis
## X1     1 4357 619.84 321.86    553   580.2 256.49   0 1923  1923 1.45     2.91
##        se
## X1 4.88
```

```
describe(state$age)
```

```
##      vars    n  mean   sd median trimmed   mad min max range  skew kurtosis   se
## X1     1 4357 40.79 8.16     41   40.94 10.38  25  55    30 -0.15    -0.99 0.12
```

```
state_sex = as.data.frame(table(state$sex))
names(state_sex) = c("Sex","Count")
state_sex[ ,1] = c("Male", "Female")
state_schooling = as.data.frame(table(state$grade92))
```

```
names(state_schooling) = c("Education","Count")
state_schooling[ ,1] = c( "1st - 4th Grade","5th or 6th Grade",
                          "7th or 8th Grade","9th Grade","10th Grade","11th Grade",
                          "12th Grade NO DIPLOMA","High School Graduate, Diploma or GED",
                          "Some College but No Degree","Associate Degree -- Occupational/Vocational",
                          "Associate Degree -- Academic Program","Bachelor's Degree (e.g. BA,AB,BS)",
                          "Master's Degree (e.g. MA,MS,MEng,Med,MSW,MBA)",
                          "Professional School Deg (e.g. MD,DDS,DVM,LLB,JD)",
                          "Doctorate Degree (e.g. PhD, EdD)")

ggplot(state_sex, aes(x = Sex, y = Count, fill = Sex)) +
    ggtitle("State Government Workers by Sex") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Sex") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal()
```
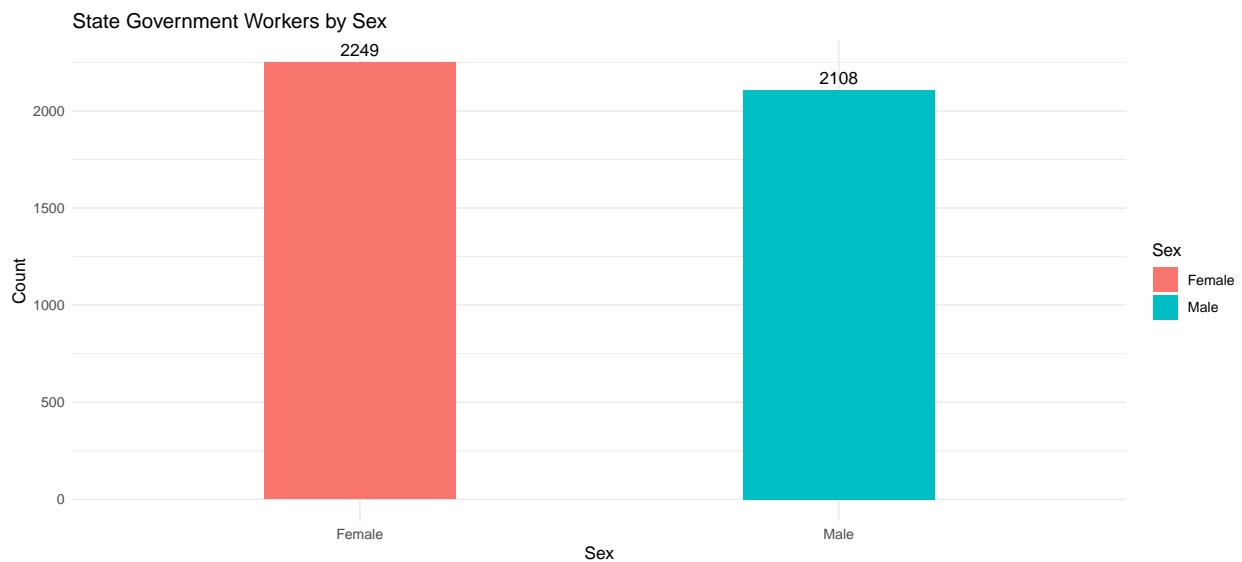


```
ggplot(state_schooling, aes(x = Education, y = Count, fill = Education)) +
    ggtitle("State Government Workers by Years of Schooling Completed") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Education") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal() +
    theme(axis.text.x = element_blank())
```

6

State Government Workers by Years of Schooling Completed
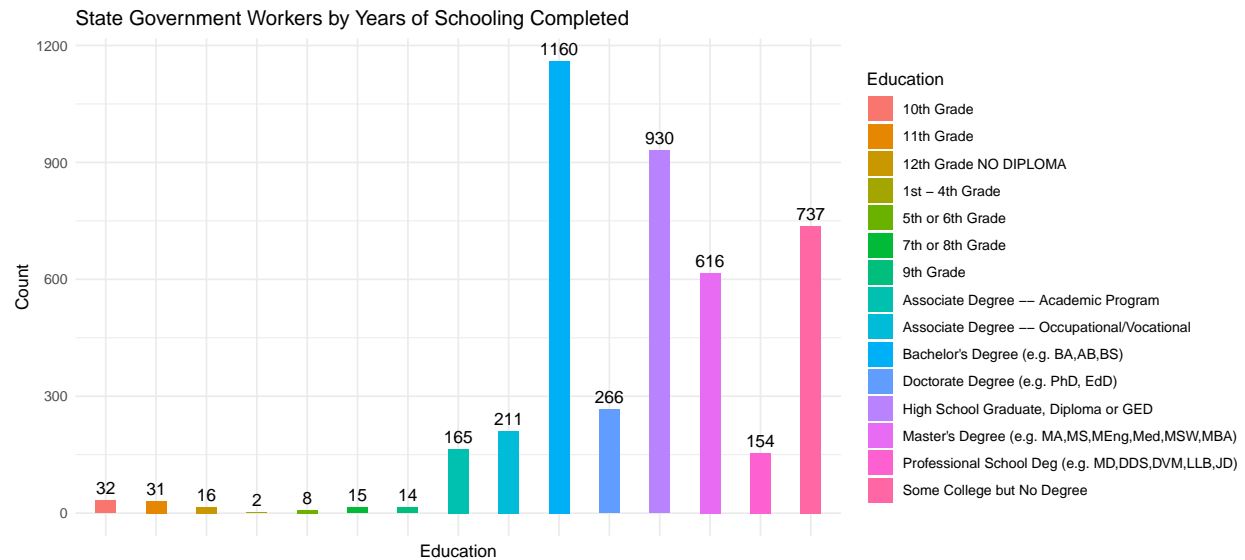
State Government Workers

Parameters	Min	Max	Mean	Standard Deviation
State Government Workers' Usual Weekly Earnings	0	1923	619.84	321.86
State Government Workers' Age	25	55	40.79	8.16

Summary of Federal Government Workers

```
describe(federal$earnwke)
```

```
##      vars    n   mean     sd median trimmed    mad min  max range skew kurtosis
## X1     1 3168 725.59 349.45    673   689.3 286.14   0 1923  1923 1.07     1.33
##        se
## X1 6.21
```

```
describe(federal$age)
```

```
##      vars    n  mean  sd median trimmed mad min max range  skew kurtosis   se
## X1     1 3168 41.55 7.8     42   41.81 8.9  25  55    30 -0.25    -0.89 0.14
```

```
federal_sex = as.data.frame(table(federal$sex))
names(federal_sex) = c("Sex","Count")
federal_sex[ ,1] = c("Male", "Female")
federal_schooling = as.data.frame(table(federal$grade92))
names(federal_schooling) = c("Education","Count")
federal_schooling[ ,1] = c("1st - 4th Grade","5th or 6th Grade",
                "7th or 8th Grade","9th Grade","10th Grade","11th Grade",
                "12th Grade NO DIPLOMA","High School Graduate, Diploma or GED",
                "Some College but No Degree","Associate Degree -- Occupational/Vocational",
                "Associate Degree -- Academic Program","Bachelor's Degree (e.g. BA,AB,BS)",
                "Master's Degree (e.g. MA,MS,MEng,Med,MSW,MBA)",
                "Professional School Deg (e.g. MD,DDS,DVM,LLB,JD)",
```

```
                          "Doctorate Degree (e.g. PhD, EdD)")

ggplot(federal_sex, aes(x = Sex, y = Count, fill = Sex)) +
    ggtitle("Federal Government Workers by Sex") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Sex") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal()
```


Federal Government Workers by Sex
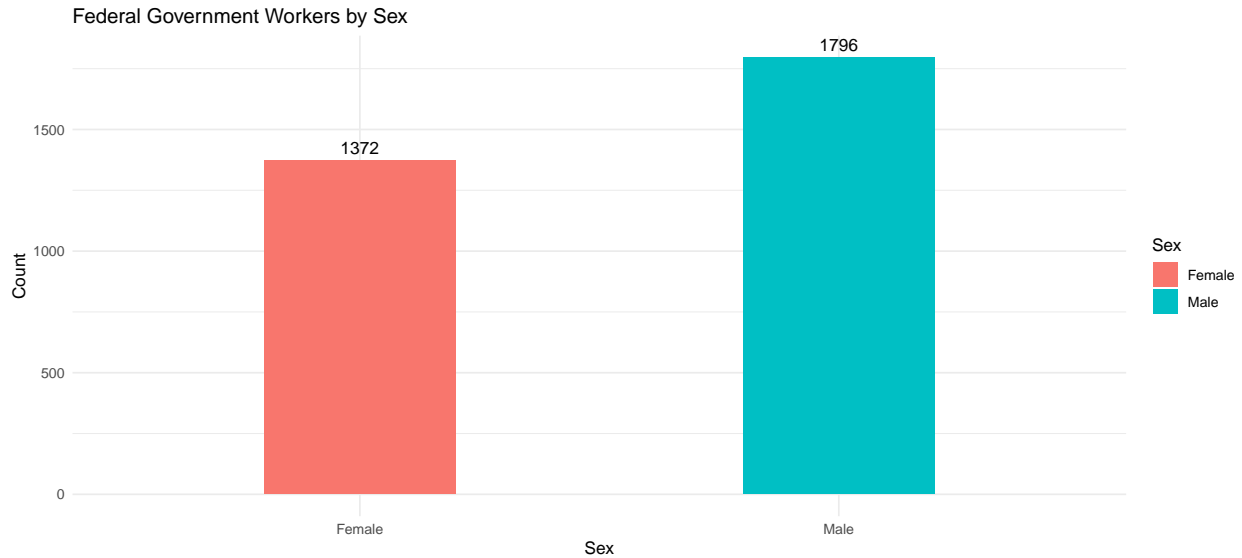
```
ggplot(federal_schooling, aes(x = Education, y = Count, fill = Education)) +
    ggtitle("Federal Government Workers by Years of Schooling Completed") +
    geom_bar(stat = "identity", width = 0.4) + xlab("Education") + ylab("Count") +
    geom_text(aes(label = Count), vjust = -0.5, size = 4) + theme_minimal() +
    theme(axis.text.x = element_blank())
```


Federal Government Workers by Years of Schooling Completed
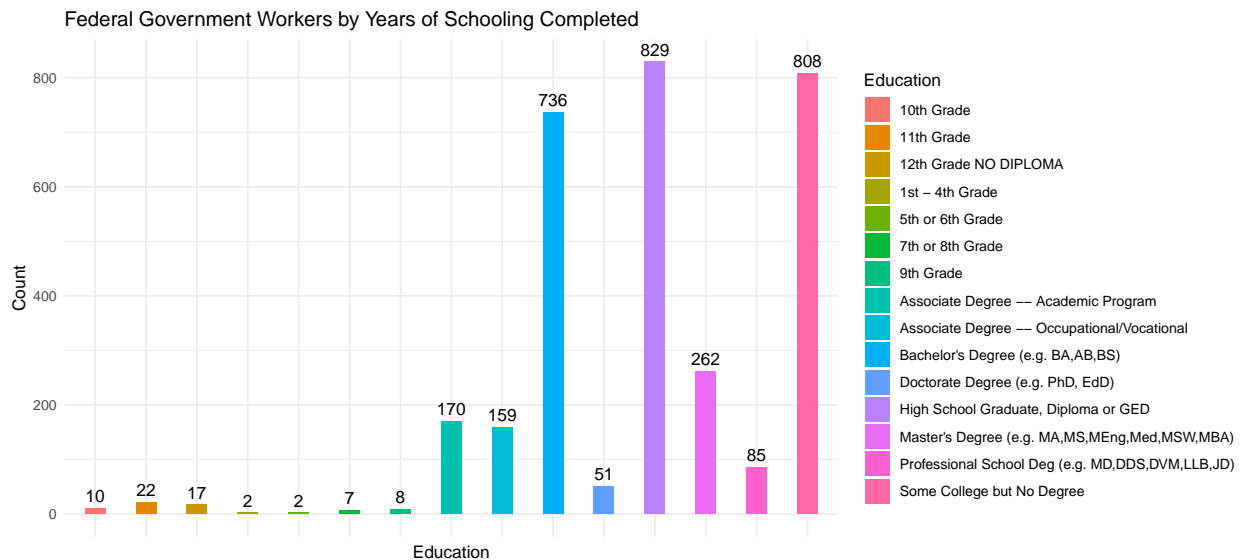
Federal Government Workers

Parameters	Min	Max	Mean	Standard Deviation
Federal Government Workers' Usual Weekly Earnings	0	1923	725.59	349.45
Federal Government Workers' Age	25	55	41.55	7.8

Interpretation: From these four brief summaries, there is an increasing trend among the mean of age of Private sector workers (38.47), state government workers (40.79), local government workers (40.85), and federal government (41.55). In addition, for usual weekly earnings, federal government workers have the highest mean value (725.59), and private sector workers have the lowest mean value (605.71). The minimal age, weekly earnings, and maximum age, weekly earnings for each sector are the same. Furthermore, 58.3% of private sector workers are males, and 56.7% of federal government workers are males. On th other hand, 53.8% of local government workers are females, and 51.6% of state government workers are female. High school graduate is the major level for private sector workers (34.8%), and federal government workers (26.2%). Meanwhile, bachelor's degree is the major level for local government workers (27.1%), and state government workers (26.6%).

Assignment 3: Merge Data on the Amount of Land within an MSA Unavailable for Real-estatedevelopment

```
library(haven)
housing_supply = read_dta("housing_supply.dta")
# install.packages(dplyr)
library(dplyr)
morg96_new = morg96_new %>% left_join(housing_supply %>% select(msafips, unaval), by = "msafips")
```

Assignment 4: Who Seems to Earn More? The Public (Local Government Workers) or Private Sector Workers?

```
# exclude federal government workers and state government workers
morg96_local = morg96_new[-c(a3,a4), ]
# nature log of weekly earnings for each worker
ln_earnwk = log(morg96_local$earnwke)
morg96_local$ln_earnwk = ln_earnwk
# 0/1 binary variable of whether the worker is employed by the local government or not
morg96_local$local_condition = NA
a2_1 =  which(morg96_local$class94 %in% c(3))
morg96_local[c(a2_1), 79] = 1
morg96_local[-c(a2_1), 79] = 0
# since nature log of 0 is -inf, we need to get rid of that
a5 = which(morg96_local$ln_earnwk %in% c(NA))
morg96_local = morg96_local[-c(a5), ]
a6 = which(morg96_local$ln_earnwk %in% c(-Inf))
morg96_local = morg96_local[-c(a6), ]
# 0/1 binary variable is a categorical variable
morg96_local$local_condition = as.factor(morg96_local$local_condition)

# linear regression model on nature log of weekly earnings and workers condition
lm_1 = lm(ln_earnwk ~ local_condition, data = morg96_local)
summary(lm_1)
```

```
## 
## Call:
## lm(formula = ln_earnwk ~ local_condition, data = morg96_local)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3136 -0.3926 -0.0224  0.4030  1.3226
## 
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.238998   0.002201 2834.78   <2e-16 ***
## local_condition1 0.074564   0.007067   10.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 0.5889 on 79273 degrees of freedom
## Multiple R-squared:  0.001402,   Adjusted R-squared:  0.00139
## F-statistic: 111.3 on 1 and 79273 DF,  p-value: < 2.2e-16
```

Interpretation: From the summary, we can tell that after switching from not local government workers to local government workers, the nature log of weekly earnings increase by 0.075 which means that the public sector workers (local government workers) earn more. However, the adjusted R-squared value (0.0014) for this model is very small which means that this model is not satisfied.

Assignment 5: Add More Controls for Workers' Demographics to the Regressions

```r
morg96_local$sex = as.factor(morg96_local$sex)
morg96_local$race = as.factor(morg96_local$race)
morg96_local$ethnic = as.factor(morg96_local$ethnic)
# square of the age of the workers
morg96_local$age_sq = (morg96_local$age)^2
morg96_local$grade92 = as.factor(morg96_local$grade92)
lm_2 = lm(ln_earnwk ~ local_condition + sex + race + ethnic + age +
                    age_sq + grade92, data = morg96_local)
summary(lm_2)
```

```
## 
## Call:
## lm(formula = ln_earnwk ~ local_condition + sex + race + ethnic +
##     age + age_sq + grade92, data = morg96_local)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.7788 -0.2946  0.0156  0.3207  2.2219
## 
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.321e+00  6.089e-02  70.960  < 2e-16 ***
## local_condition1 -4.340e-02  6.090e-03  -7.126 1.04e-12 ***
## sex2             -3.079e-01  3.570e-03 -86.241  < 2e-16 ***
## race2            -1.814e-01  6.066e-03 -29.900  < 2e-16 ***
```

```
## race3             -1.066e-01  1.829e-02  -5.826 5.70e-09 ***
## race4             -9.655e-02  9.274e-03 -10.410  < 2e-16 ***
## ethnic2           -7.173e-02  5.793e-02  -1.238 0.215656
## ethnic3           -1.010e-01  1.697e-02  -5.952 2.65e-09 ***
## ethnic4            5.960e-02  2.335e-02   2.552 0.010712 *
## ethnic5           -3.129e-02  3.136e-02  -0.998 0.318392
## ethnic6           -7.023e-02  1.916e-02  -3.665 0.000247 ***
## ethnic7           -1.318e-02  2.465e-02  -0.535 0.592968
## ethnic8            9.957e-02  1.299e-02   7.663 1.84e-14 ***
## ethnic9           -5.450e-02  5.493e-02  -0.992 0.321160
## age                6.110e-02  2.050e-03  29.812  < 2e-16 ***
## age_sq            -6.518e-04  2.593e-05 -25.134  < 2e-16 ***
## grade9232          1.001e-01  5.132e-02   1.950 0.051201 .
## grade9233          1.352e-01  4.759e-02   2.841 0.004492 **
## grade9234          2.182e-01  4.748e-02   4.596 4.32e-06 ***
## grade9235          1.957e-01  4.743e-02   4.126 3.70e-05 ***
## grade9236          2.824e-01  4.680e-02   6.035 1.60e-09 ***
## grade9237          2.858e-01  4.649e-02   6.147 7.95e-10 ***
## grade9238          3.329e-01  4.869e-02   6.837 8.15e-12 ***
## grade9239          4.892e-01  4.511e-02  10.845  < 2e-16 ***
## grade9240          6.096e-01  4.519e-02  13.488  < 2e-16 ***
## grade9241          6.580e-01  4.567e-02  14.408  < 2e-16 ***
## grade9242          7.189e-01  4.581e-02  15.690  < 2e-16 ***
## grade9243          9.296e-01  4.520e-02  20.565  < 2e-16 ***
## grade9244          1.113e+00  4.559e-02  24.421  < 2e-16 ***
## grade9245          1.279e+00  4.735e-02  27.021  < 2e-16 ***
## grade9246          1.220e+00  4.843e-02  25.195  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4925 on 79012 degrees of freedom
##   (232 observations deleted due to missingness)
## Multiple R-squared:  0.3016, Adjusted R-squared:  0.3013
## F-statistic:  1137 on 30 and 79012 DF,  p-value: < 2.2e-16
```

Interpretation: From the summary, we can tell that after switching from not local government workers to local government workers, the nature log of weekly earnings decrease by 0.043 which means that the public sector workers (local government workers) earn less after adding controls for workers' demographics to the regressions. The adjusted R-squared value is 0.3013 which is bigger than the previous model (0.0014). As the result, by adding more variables, this model is better than the first one.

Assignment 6: How the Local Government Worker-Private Sector Wage Gap Differs Across Metropolitan Statistical Areas

```
# general linear regression model
morg96_local$msafips = as.factor(morg96_local$msafips)
lm_3 = lm(ln_earnwk ~ local_condition + sex + race + ethnic + age +
                      age_sq + grade92 + msafips + unaval, data = morg96_local)
# summary(lm_3)

# install.packages(sandwich)
# install.packages(lmtest)
```

```
# install.packages(tidyverse)
# install.packages(broom)
library(sandwich)
library(lmtest)
library(tidyverse)
library(broom)
# get degree of freedom
# cluster the standard errors by metropolitan statistical area
length(unique(morg96_local$msafips))
```

```
## [1] 243
```

```
# there are 243 unique metropolitan statistical area, so 243-1 = 242
lm_3_clustered = coeftest(lm_3, vcov = vcovCL, type = "HC1",
                          df = 242,
                          cluster =~ msafips)
tidy(lm_3_clustered, conf.int = TRUE) %>% filter(term == "local_condition1")
```

```
## # A tibble: 1 x 7
##   term             estimate std.error statistic p.value conf.low conf.high
##   <chr>               <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 local_condition1 -0.00253    0.0108    -0.234   0.815  -0.0238    0.0187
```

Interpretation: From the summary, we can tell that after switching from not local government workers to local government workers, the nature log of weekly earnings decrease by 0.003 which means that the public sector workers (local government workers) earn less after adding cluster metropolitan statistical areas to the model. However, the p-value of variable local_condition (0.815) is bigger than 0.05, which means that this variable is not statistically significant to the nature log of weekly earnings.

Assignment 7: Workers With at Least a 4 Year College Degree vs Those With Less Education

```
a7 = which(morg96_local$grade92 %in% c(31:42))
four_year_local = morg96_local[-c(a7), ]
non_four_year_local = morg96_local[c(a7), ]

# workers with at least a 4 year college degree
length(unique(four_year_local$msafips))
```

```
## [1] 243
```

```
lm_4 = lm(ln_earnwk ~ local_condition + sex + race + ethnic + age +
                   age_sq + grade92 + msafips + unaval, data = four_year_local)
lm_4_clustered = coeftest(lm_4, vcov = vcovCL, type = "HC1",
                          df = 242,
                          cluster =~ msafips)
tidy(lm_4_clustered, conf.int = TRUE) %>% filter(term == "local_condition1")
```

```
## # A tibble: 1 x 7
##   term          estimate std.error statistic p.value conf.low conf.high
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 local_condition1  -0.0289    0.0137     -2.11  0.0361  -0.0559   -0.00188
```

```r
# workers with less than a 4 year college degree
length(unique(non_four_year_local$msafips))
```

```
## [1] 243
```

```r
lm_5 = lm(ln_earnwk ~ local_condition + sex + race + ethnic + age +
                    age_sq + grade92 + msafips + unaval, data = non_four_year_local)
lm_5_clustered = coeftest(lm_5, vcov = vcovCL, type = "HC1",
                        df = 242,
                        cluster =~ msafips)
tidy(lm_5_clustered, conf.int = TRUE) %>% filter(term == "local_condition1")
```

```
## # A tibble: 1 x 7
##   term          estimate std.error statistic p.value conf.low conf.high
##   <chr>            <dbl>     <dbl>     <dbl>   <dbl>    <dbl>     <dbl>
## 1 local_condition1  0.00779   0.0142     0.550   0.583  -0.0201    0.0357
```

Interpretation: For workers with at least a 4 year college degree, we can tell that after switching from not local government workers to local government workers, the nature log of weekly earnings decrease by 0.029 which means that the public sector workers (local government workers) earn less after adding cluster metropolitan statistical areas to the model. Furthermore, the p-value of variable local_condition (0.036) is smaller than 0.05, which means that this variable is statistically significant to the nature log of weekly earnings. On the other hand, for workers whose education level are less than 4 year college degree, we find out that after switching from not local government workers to local government workers, the nature log of weekly earnings increase by 0.008 which means that the public sector workers (local government workers) earn more after adding cluster metropolitan statistical areas to the model. However, the p-value of variable local_condition (0.583) is bigger than 0.05, which means that this variable is not statistically significant to the nature log of weekly earnings.

Questions

1. The second problem needs mean, standard deviation, min value, and max value of: workers' usual weekly earnings, workers' years of schooling completed, workers' gender, and worker's age. However, I do not know how to get these five value for workers' year of schooling completed and gender. For workers' years of schooling completed, variable "grade92", there is not a clear year list for the workers, such as 1st to 4th, 7th or 8th. As the result, for workers' year of schooling completed and gender, I choose to use the bar plot to show the number and pattern of these two variables.

2. There is a class of worker (code:7, without pay) which I do not know if it belongs to private sector workers or not. In this report, I consider workers without pay as part of private sector workers since there are other classes, such as Self-employed, and local, state, and federal government.

3. For federal, and state government workers, the variable grade92, years of schooling completed, shows that they do not have code 31: Less than 1st grade.