Assignment 2

Sara Arman

```
library(tidycensus)
library(tidyverse)
library(ggplot2)
library(ggthemes)
library(viridis)
```

For my first step, I am loading my data set from Assignment 1.

This data set includes the following variables: 1. Sex, “SEX”, categorical 2. Wages or salary income past 12 months, “WAGP”, continuous 3. Public assistance income past 12 months, “PAP”, continuous 4. Educational attainment, “SCHL”, categorical 5. Citizenship status, CIT, categorical 6. Age, AGEP, continuous

I have filtered my data to only include wages and public assistance income greater than -1. This has narrowed my number of observations down from 70,131 to 59,952.

```
person_data <-read_csv('SaraArman-DataSet1_Final.csv') %>%
  filter(WAGP> -1 & PAP> -1)
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

Continous Variables

My first continous variable is Wages over the past 12 months (WAGP) so I wil find the:

1. mean
2. standard deviation
3. 95% CI for Population Mean
4. Inter Quartile Range
5. And then I will create a histogram.

```
WAGP_summary <- summary(person_data$WAGP)
WAGP_summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0   14300   40284   59000  620000
```

Mean

```
WAGP_summary["Mean"]
```

```
##     Mean
## 40284.03
```

Standard Deviation

```
sd(person_data$WAGP)
```

```
## [1] 70259.9
```

95% CI for Population mean

```
conf_int<- t.test(person_data$WAGP)
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$WAGP
## t = 140.4, df = 59958, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  39721.64 40846.42
## sample estimates:
## mean of x
##  40284.03
```

IQR (IQR = Q3-Q1 = 47000)

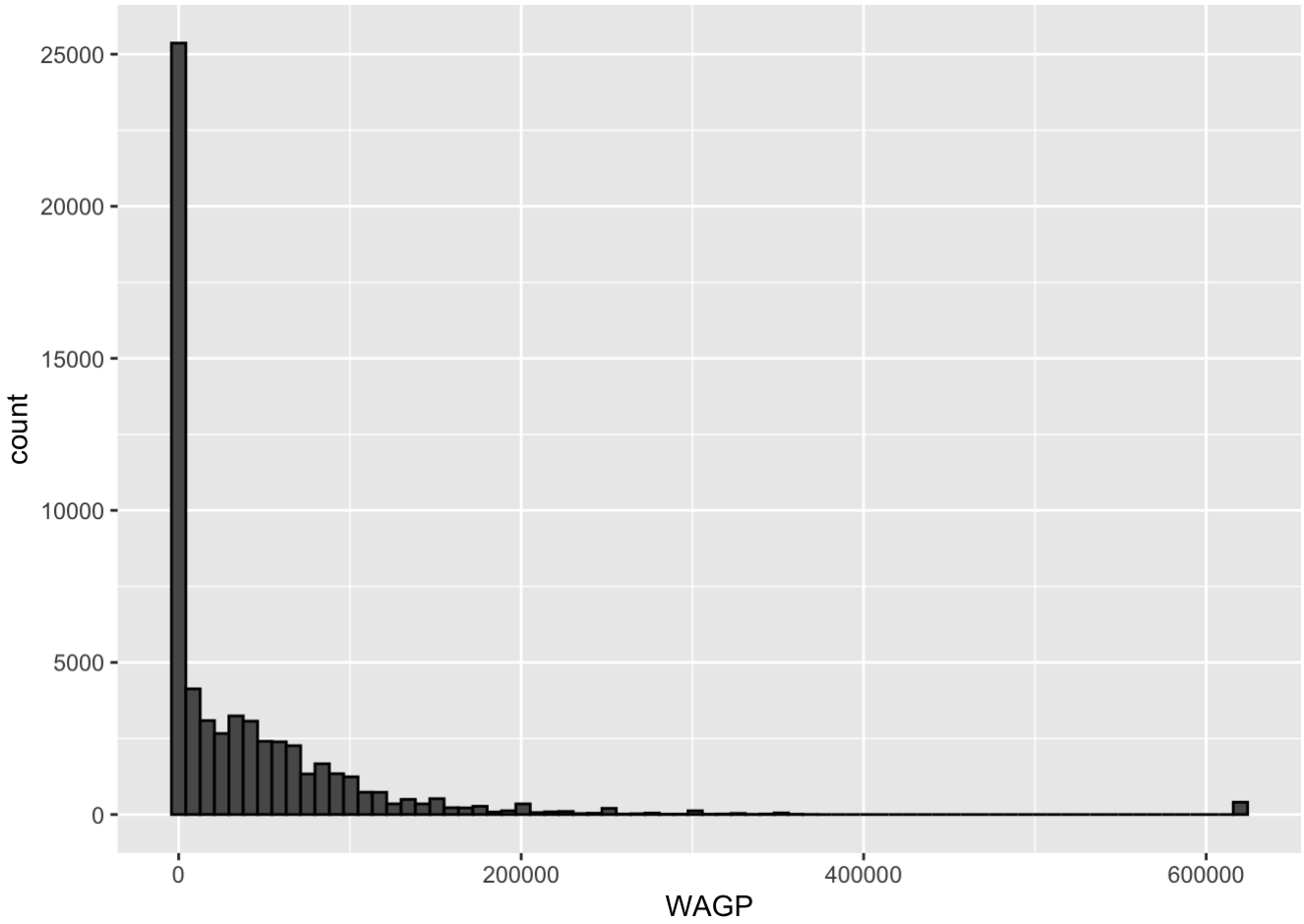
```
quantile(person_data$WAGP)
```

```
##     0%    25%    50%    75%   100%
##      0      0  14300  59000 620000
```

Histogram

This histogram was not surprising, as it shows that most people make less than 100,000 each year, with 25,000 people making 0 a year.

```
options(scipen = 999)
ggplot(person_data, aes(x = WAGP)) +
  geom_histogram(color = "black" , bins = 75)
```




```
scale_x_continuous(name = "Wages and Income Over the Past 12 Months")
```

```
## <ScaleContinuousPosition>
##  Range:
##  Limits:    0 --    1
```

```
                    breaks = breaks <- seq(0, 10, by = 1)
```

My second continous variable is Public Assistance Income Recieved over past 12 months, PAP

```
PAP_summary <- summary(person_data$PAP)
PAP_summary
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##     0.00     0.00     0.00    39.56     0.00 16000.00
```

Mean

```
PAP_summary["Mean"]
```

```
##     Mean
## 39.55703
```

Standard Deviation

```
sd(person_data$PAP)
```

```
## [1] 520.3467
```

95% CI

```
conf_int <- t.test(person_data$PAP)
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$PAP
## t = 18.615, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  35.39196 43.72210
## sample estimates:
## mean of x
##  39.55703
```

IQR (IQR = Q3-Q1)

```
quantile(person_data$PAP)
```

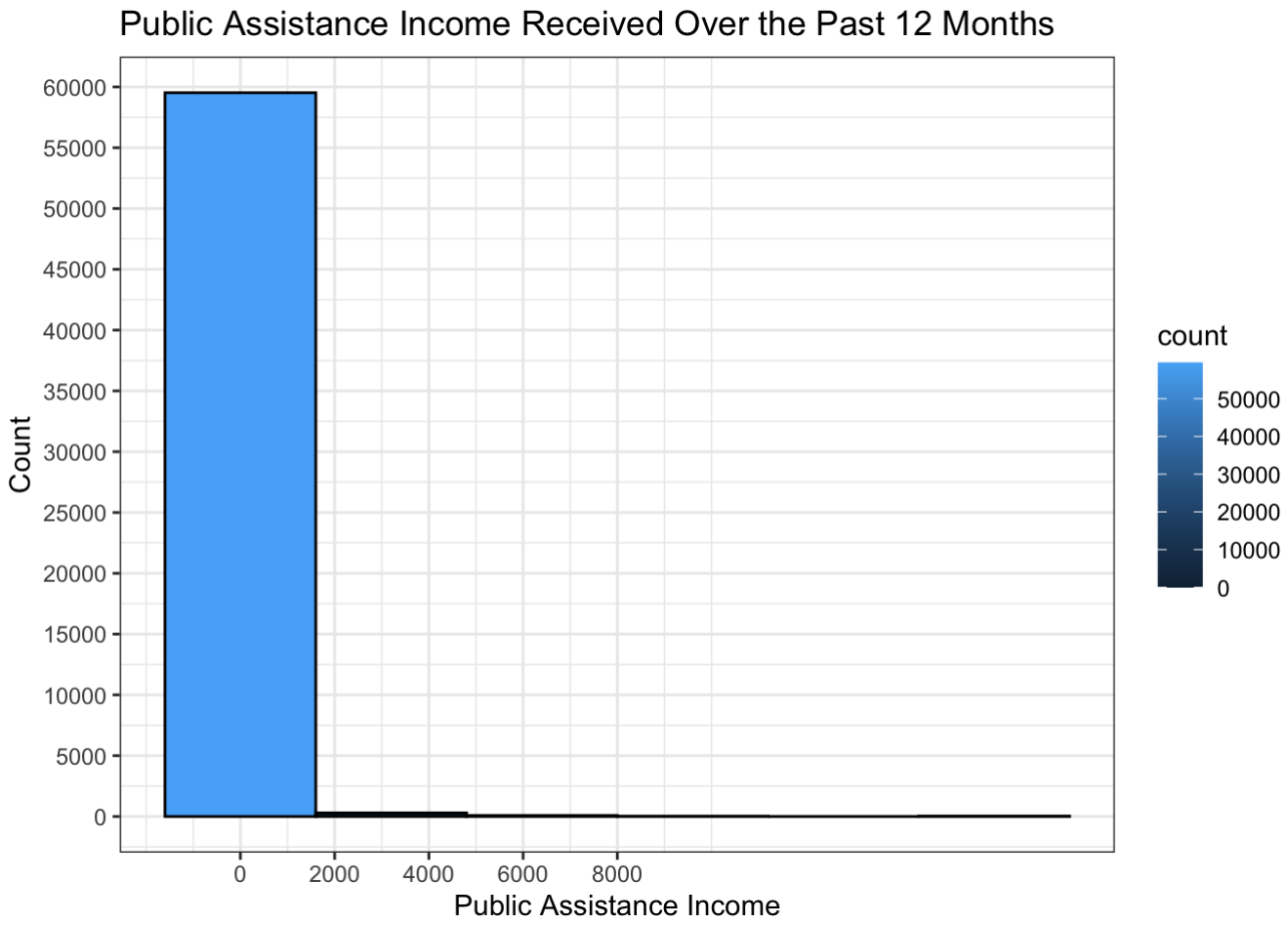
```
##    0%   25%   50%   75%  100%
##     0     0     0     0 16000
```

Histogram

This histogram shows that most people recevied $0 in pubic assistance income, while a very small subset of the population recieved between $2000 and $4000.

```
ggplot(person_data,
       aes(x = PAP)) +
geom_histogram(aes(fill = ..count..), bins = 6, color = "black", weight = 3) +
scale_x_continuous(name = "Public Assistance Income",
                   breaks = seq(0, 8000, by = 2000)) +
scale_y_continuous(name = "Count",
                  breaks = seq(0, 60000, by = 5000)) +
labs(title = 'Public Assistance Income Received Over the Past 12 Months')+
 theme_bw()
```

```
## Warning: Ignoring unknown parameters: weight
```




```
coord_flip()
```

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

My third continous variable is age

```
AGEP_summary <- summary(person_data$AGEP)
AGEP_summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   15.00   30.00   48.00   47.78   63.00   95.00
```

Mean

```
AGEP_summary["Mean"]
```

```
##    Mean
## 47.7846
```

Standard Deviation

```
sd(person_data$AGEP)
```

```
## [1] 19.99209
```

95% CI

```
conf_int <- t.test(person_data$AGEP)
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$AGEP
## t = 585.27, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  47.62458 47.94463
## sample estimates:
## mean of x
##   47.7846
```

```
quantile(person_data$AGEP)
```

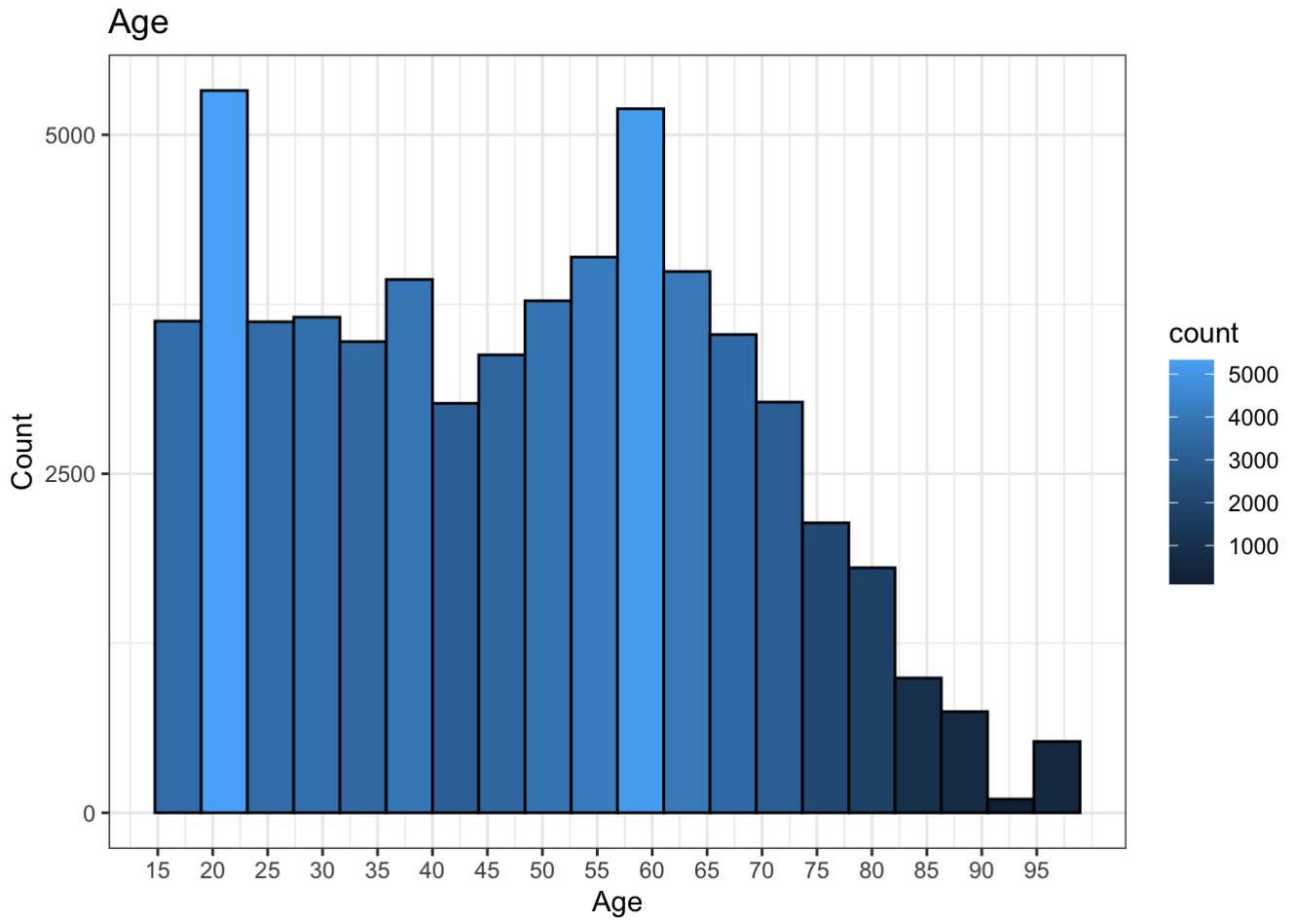
```
##   0%  25%  50%  75% 100%
##   15   30   48   63   95
```

Histogram

This histogram shows that there are many people who live in Massachusetts who are between the ages of 20-25, and 55-65.

```
ggplot(person_data,
       aes(x = AGEP)) +
geom_histogram(aes(fill = ..count..), bins = 20, color = "black", weight = 5) +
scale_x_continuous(name = "Age",
                   breaks = seq(0, 95, by = 5)) +
scale_y_continuous(name = "Count",
                  breaks = seq(0, 30000, by = 2500)) +
labs(title = 'Age')+
 theme_bw()
```

```
## Warning: Ignoring unknown parameters: weight
```




```
coord_flip()
```

Table I created of my continuous variables.

Characteristics of the 2018 American Community Survey in Massachusetts

Statistic	Wages/Income	Public Assistance Income	Age
Sample Mean	40284.03	39.55703	47.7
Std. Deviation	70259.9	520.3	19.9
95% CI	t = 140.4 95% CI: 39721.64 40846.42	t = 18.6 95%CI: 35.39196 43.72210	t = 585.2 95% CI: 47.62458 47.94463
IQR	59000	0	33

```
## <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
##     aspect: function
##     backtransform_range: function
##     clip: on
##     default: FALSE
##     distance: function
##     expand: TRUE
##     is_free: function
##     is_linear: function
##     labels: function
##     limits: list
##     modify_scales: function
##     range: function
##     render_axis_h: function
##     render_axis_v: function
##     render_bg: function
##     render_fg: function
##     setup_data: function
##     setup_layout: function
##     setup_panel_guides: function
##     setup_panel_params: function
##     setup_params: function
##     train_panel_guides: function
##     transform: function
##     super:  <ggproto object: Class CoordFlip, CoordCartesian, Coord, gg>
```

Categorical Variables

For each categorical variable in my data set (SEX, SCHL, CIT), I will calculate the 95-percent confidence interval for the proportion of the population in each category.

My first categorical variable is Sex

```
unique(person_data$SEX_label)
```

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

```
table(person_data$SEX_label)
```

```
##
## Female   Male
##  31262  28697
```

```
SEX_summary <- summary(person_data$SEX)
SEX_summary
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   2.000   1.521   2.000   2.000
```

```
table(person_data$SEX_label) / sum(table(person_data$SEX_label))
```

```
##
##    Female      Male
## 0.5213896 0.4786104
```

```
mean(person_data$SEX_label == "Female")
```

```
## [1] 0.5213896
```

```
mean(person_data$SEX_label == "Male")
```

```
## [1] 0.4786104
```

```
conf_int <- t.test(person_data$SEX_label == "Female")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SEX_label == "Female"
## t = 255.57, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.5173910 0.5253882
## sample estimates:
## mean of x
## 0.5213896
```

```
conf_int <- t.test(person_data$SEX_label == "Male")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SEX_label == "Male"
## t = 234.6, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.4746118 0.4826090
## sample estimates:
## mean of x
## 0.4786104
```

My second categorical variable is Educational Attainment (SCHL)

I am finding the confidence interval for each category within educational attainment.

```
unique(person_data$SCHL_label)
```

```
##  [1] "1 or more years of college credit, no degree"
##  [2] "Some college, but less than 1 year"
##  [3] "Regular high school diploma"
##  [4] "Master's degree"
##  [5] "Associate's degree"
##  [6] "Bachelor's degree"
##  [7] "No schooling completed"
##  [8] "Grade 4"
##  [9] "Grade 11"
## [10] "GED or alternative credential"
## [11] "Grade 8"
## [12] "Grade 9"
## [13] "Grade 10"
## [14] "Professional degree beyond a bachelor's degree"
## [15] "12th grade - no diploma"
## [16] "Grade 5"
## [17] "Grade 7"
## [18] "Grade 6"
## [19] "Doctorate degree"
## [20] "Grade 3"
## [21] "Grade 1"
## [22] "Grade 2"
## [23] "Nursery school, preschool"
## [24] "Kindergarten"
```

```
SCHL_summary <- summary(person_data$SCHL_label)
SCHL_summary
```

```
##    Length     Class      Mode
##     59959 character character
```

```
table(person_data$SCHL_label)
```

```
##
##   1 or more years of college credit, no degree
##                                           7359
##                         12th grade - no diploma
##                                            888
##                              Associate's degree
##                                           4035
##                               Bachelor's degree
##                                          13684
##                                Doctorate degree
##                                           1643
##                  GED or alternative credential
##                                           1479
##                                        Grade 1
##                                             20
##                                       Grade 10
##                                           1348
##                                       Grade 11
##                                           1515
##                                        Grade 2
##                                             43
##                                        Grade 3
##                                             73
##                                        Grade 4
##                                            130
##                                        Grade 5
##                                            115
##                                        Grade 6
##                                            206
##                                        Grade 7
##                                            165
##                                        Grade 8
##                                            688
##                                        Grade 9
##                                           1208
##                                    Kindergarten
##                                             11
##                                 Master's degree
##                                           7745
##                          No schooling completed
##                                            845
##                        Nursery school, preschool
##                                             10
## Professional degree beyond a bachelor's degree
##                                           1746
##                       Regular high school diploma
##                                          11564
##             Some college, but less than 1 year
##                                           3439
```

```
conf_int <- t.test (person_data$SCHL_label == "1 or more years of college credit, no deg
ree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "1 or more years of college credit, no degree"
## t = 91.588, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1201073 0.1253604
## sample estimates:
## mean of x
## 0.1227339
```

```
conf_int <- t.test (person_data$SCHL_label == "12th grade - no diploma")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "12th grade - no diploma"
## t = 30.022, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01384324 0.01577700
## sample estimates:
##  mean of x
## 0.01481012
```

```
conf_int <- t.test (person_data$SCHL_label == "Associate's degree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Associate's degree"
## t = 65.773, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.06529059 0.06930138
## sample estimates:
##  mean of x
## 0.06729599
```

```
conf_int <- t.test (person_data$SCHL_label == "Bachelor's degree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Bachelor's degree"
## t = 133.15, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.2248632 0.2315820
## sample estimates:
## mean of x
## 0.2282226
```

```
conf_int <- t.test (person_data$SCHL_label == "Doctorate degree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Doctorate degree"
## t = 41.101, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02609531 0.02870881
## sample estimates:
##  mean of x
## 0.02740206
```

```
conf_int <- t.test (person_data$SCHL_label == "GED or alternative credential")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "GED or alternative credential"
## t = 38.941, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02342530 0.02590841
## sample estimates:
##  mean of x
## 0.02466686
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 1")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 1"
## t = 4.4728, df = 59958, p-value = 0.000007733
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.0001873945 0.0004797280
## sample estimates:
##    mean of x
## 0.0003335613
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 10")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 10"
## t = 37.135, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02129540 0.02366865
## sample estimates:
##  mean of x
## 0.02248203
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 11")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 11"
## t = 39.424, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02401108 0.02652346
## sample estimates:
##  mean of x
## 0.02526727
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 2")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 2"
## t = 6.5597, df = 59958, p-value = 0.00000000005434
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.0005028753 0.0009314382
## sample estimates:
##    mean of x
## 0.0007171567
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 3")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 3"
## t = 8.5491, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.0009383709 0.0014966264
## sample estimates:
##   mean of x
## 0.001217499
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 4")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 4"
## t = 11.414, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.001795837 0.002540460
## sample estimates:
##   mean of x
## 0.002168148
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 5")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 5"
## t = 10.734, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.001567760 0.002268195
## sample estimates:
##   mean of x
## 0.001917977
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 6")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 6"
## t = 14.377, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.002967307 0.003904055
## sample estimates:
##   mean of x
## 0.003435681
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 7")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 7"
## t = 12.863, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.002332556 0.003171204
## sample estimates:
##  mean of x
## 0.00275188
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 8")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 8"
## t = 26.381, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01062201 0.01232701
## sample estimates:
##  mean of x
## 0.01147451
```

```
conf_int <- t.test (person_data$SCHL_label == "Grade 9")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Grade 9"
## t = 35.112, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01902244 0.02127176
## sample estimates:
## mean of x
## 0.0201471
```

```
conf_int <- t.test (person_data$SCHL_label == "Kindergarten")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Kindergarten"
## t = 3.3169, df = 59958, p-value = 0.0009108
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.00007505038 0.00029186701
## sample estimates:
##    mean of x
## 0.0001834587
```

```
conf_int <- t.test (person_data$SCHL_label == "Master's degree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Master's degree"
## t = 94.306, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.1264870 0.1318562
## sample estimates:
## mean of x
## 0.1291716
```

```
conf_int <- t.test (person_data$SCHL_label == "No schooling completed")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "No schooling completed"
## t = 29.276, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01314944 0.01503649
## sample estimates:
##  mean of x
## 0.01409296
```

```
conf_int <- t.test (person_data$SCHL_label == "Nursery school, preschool")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Nursery school, preschool"
## t = 3.1625, df = 59958, p-value = 0.001565
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.0000634165 0.0002701448
## sample estimates:
##    mean of x
## 0.0001667806
```

```
conf_int <- t.test (person_data$SCHL_label == "Professional degree beyond a bachelor's d
egree")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Professional degree beyond a bachelor's degree"
## t = 42.407, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.02777400 0.03046579
## sample estimates:
## mean of x
## 0.0291199
```

```
conf_int <- t.test (person_data$SCHL_label == " Regular high school diploma")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == " Regular high school diploma"
## t = NaN, df = 59958, p-value = NA
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  NaN NaN
## sample estimates:
## mean of x
##         0
```

```
conf_int <- t.test (person_data$SCHL_label == "Some college, but less than 1 year")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$SCHL_label == "Some college, but less than 1 year"
## t = 60.4, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.05549465 0.05921707
## sample estimates:
##  mean of x
## 0.05735586
```

My third categorical variable is Citizenship status , CIT

```
unique(person_data$CIT_label)
```

```
## [1] "Born in the U.S."
## [2] "Not a citizen of the U.S."
## [3] "U.S. citizen by naturalization"
## [4] "Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas"
## [5] "Born abroad of American parent(s)"
```

```
CIT_summary <- summary(person_data$CIT_label)
CIT_summary
```

```
##    Length     Class      Mode
##     59959 character character
```

```
table(person_data$CIT_label)
```

```
##
##                                          Born abroad of American parent(s)
##                                                                        650
## Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas
##                                                                        920
##                                                             Born in the U.S.
##                                                                      48379
##                                                 Not a citizen of the U.S.
##                                                                       4118
##                                            U.S. citizen by naturalization
##                                                                       5892
```

```
table(person_data$CIT_label) / sum(table(person_data$CIT_label))
```

```
##
##                                          Born abroad of American parent(s)
##                                                                 0.01084074
## Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas
##                                                                 0.01534382
##                                                             Born in the U.S.
##                                                                 0.80686803
##                                                 Not a citizen of the U.S.
##                                                                 0.06868026
##                                            U.S. citizen by naturalization
##                                                                 0.09826715
```

```
conf_int <- t.test (person_data$CIT_label == "Born abroad of American parent(s)")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$CIT_label == "Born abroad of American parent(s)"
## t = 25.634, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.01001185 0.01166963
## sample estimates:
##  mean of x
## 0.01084074
```

```
conf_int <- t.test (person_data$CIT_label == "Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas ")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$CIT_label == "Born in Puerto Rico, Guam, the U.S. Virgin Islands, or the Northern Marianas "
## t = NaN, df = 59958, p-value = NA
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  NaN NaN
## sample estimates:
## mean of x
##         0
```

```
conf_int <- t.test (person_data$CIT_label == "Born in the U.S.")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$CIT_label == "Born in the U.S."
## t = 500.49, df = 59958, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.8037082 0.8100278
## sample estimates:
## mean of x
##  0.806868
```

```
conf_int <- t.test (person_data$CIT_label == "Not a citizen of the U.S. ")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$CIT_label == "Not a citizen of the U.S. "
## t = NaN, df = 59958, p-value = NA
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  NaN NaN
## sample estimates:
## mean of x
##         0
```

```
conf_int <- t.test (person_data$CIT_label == "U.S. citizen by naturalization ")
conf_int
```

```
##
##  One Sample t-test
##
## data:  person_data$CIT_label == "U.S. citizen by naturalization "
## t = NaN, df = 59958, p-value = NA
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  NaN NaN
## sample estimates:
## mean of x
##         0
```