CS5821 Project

Kahlil Cole, Ayomide Awojobi

April 13, 2021

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse
1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   1.4.0
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'purrr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

```
## -- Conflicts ------------------------------------------
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(tidyr)
library(ggplot2)



# Reading dataset into R
dataset <- read.csv("Occupational_Employment_and_Wage_Estimates.csv",
stringsAsFactors=T)

# number of observations in dataset
dim(dataset) # there are 6563 observations in the dataset
```

```
## [1] 6563   12
```

```r
# Remove missing values from dataset
newdata <- na.omit(dataset)

# number of observations in newdata
dim(newdata) # there are now 5473 observations in the newdata. 1090
observations
```

```
## [1] 5473   12
```

```r
# of missing values were removed

# Show column names of newdata
names(newdata) #ID is the number of observations, year is 2020.
```

```
##  [1] "ID"                "Year"              "Area.code"
##  [4] "Area.name"         "SOC.code"          "Occupational.title"
##  [7] "Employment"        "Average.wage"      "X25th.Percentile"
## [10] "X50th.Percentile"  "X75th.Percentile"  "Annual.wage"
```

```r
# Occupaiton title shows the type of occupation in the area, employment shows
the number
# of people employed to the particular occupation in the year 2020. Average
wage shows the
# average pay per hour for the occupation in the year 2020. Annual wage shows
to yearly pay
# for the occupation in that year.

# Area names

levels(newdata$Area.name) # there are 16 different area names in the newdata
all in
```

```
##  [1] "Bellingham, WA"
##  [2] "Bremerton-Silverdale, WA"
##  [3] "Eastern Washington nonmetropolitan area"
##  [4] "Kennewick-Richland, WA"
##  [5] "Lewiston, ID-WA"
```

```
##  [6] "Longview, WA"
##  [7] "Mount Vernon-Anacortes, WA"
##  [8] "Olympia-Tumwater, WA"
##  [9] "Portland-Vancouver-Hillsboro, OR-WA"
## [10] "Seattle-Tacoma-Bellevue, WA"
## [11] "Spokane-Spokane Valley, WA"
## [12] "Walla Walla, WA"
## [13] "Washington"
## [14] "Wenatchee, WA"
## [15] "Western Washington nonmetropolitan area"
## [16] "Yakima, WA"

# the state of Washington

# Occupation titles
head(levels(newdata$Occupational.title),20) # there are 756 occupation titles
ranging from

##  [1] "Accountants and Auditors"
##  [2] "Actors"
##  [3] "Actuaries"
##  [4] "Acupuncturists and Healthcare Diagnosing or Treating Practitioners,
All Other"
##  [5] "Adhesive Bonding Machine Operators and Tenders"
##  [6] "Administrative Law Judges, Adjudicators, and Hearing Officers"
##  [7] "Administrative Services and Facilities Managers"
##  [8] "Adult Literacy, Remedial Ed, GED Teachers/Instructors"
##  [9] "Advertising and Promotions Managers"
## [10] "Advertising Sales Agents"
## [11] "Aerospace Engineering and Operations Technologists and Technicians"
## [12] "Aerospace Engineers"
## [13] "Agents and Business Managers of Artists, Performers, and Athletes"
## [14] "Agricultural and Food Science Technicians"
## [15] "Agricultural Equipment Operators"
## [16] "Agricultural Inspectors"
## [17] "Agricultural Sciences Teachers, Postsecondary"
## [18] "Agricultural Workers, All Other"
## [19] "Air Traffic Controllers"
## [20] "Aircraft Cargo Handling Supervisors"

# Accountant and Auditors to Zoologists and Wildlife Biologists. Only showing
the first 20 to save space.

# Employment numbers
max.employment <- newdata %>% filter(Employment == max(Employment))
max.employment$Occupational.title

## [1] Retail Salespersons
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife
Biologists
```

```r
# maximum employement number is 106,073 and the occupation title for this is
Retail Salesperson.
# avergae wage for this occupation that has the highest employement number is
$17.79 an hour
# with an annual wage of $36,988

min.employment <- newdata %>% filter(Employment == min(Employment))
min.employment$Occupational.title
```

```
##  [1] Postal Service Mail Sorters, Processors, and Processing Machine
Operators
##  [2] Emergency Management Directors
##  [3] Judges, Magistrate Judges, and Magistrates
##  [4] Stationary Engineers and Boiler Operators
##  [5] Judicial Law Clerks
##  [6] Farm and Home Management Educators
##  [7] Photographers
##  [8] Financial & Investment Analysts, Financial Risk Specialists, All
Other
##  [9] Broadcast Technicians
## [10] Purchasing Managers
## [11] Nuclear Medicine Technologists
## [12] Health and Safety Engineers, Except Mining Safety Engineers and
Inspectors
## [13] Credit Authorizers, Checkers, and Clerks
## [14] Editors
## [15] Occupational Therapy Assistants
## [16] Credit Counselors
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife
Biologists
```

```r
# minimum employment number is 10 with 16 different occupational titles which
are; postal service
# mail sorters, emergency management directors, Judges and Magistrates,
Stationary Engineers,
# judicial law clerks, Farm and Home management educators, Photographers,
Financial and Investment
# analysts, Financial Risk specialists, Broadcast Technicians, Purchasing
managers, nuclear medicine technologists
# Health and safety engineers, credit authorizers, editors, occupational
therapy assistant, credit counselors.

# Average wage
min.avgwage <- newdata %>% filter(Average.wage == min(Average.wage))

min.avgwage$Average.wage # the minimum average wage is $10.6 an hour
```

```
## [1] 10.6
```

```r
min.avgwage$Occupational.title # the occupation that has the smallest average
was is bartender
```

```
## [1] Bartenders
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife
Biologists
```

```
min.avgwage$Employment # the number of people employed that have the smallest
average wage is 89 and are bartenders
```

```
## [1] 89
```

```
max.avgwage <- newdata %>% filter(Average.wage == max(Average.wage))
```

```
max.avgwage$Average.wage # the maximum average wage is $94.5 an hour
```

```
## [1] 94.5
```

```
max.avgwage$Occupational.title # the occupation with the highest average wage
is Nurse Anesthetists
```

```
## [1] Nurse Anesthetists
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife
Biologists
```

```
max.avgwage$Employment # the number of people employed that have the highest
average wage is 624
```

```
## [1] 624
```

```
# Annual wage
max.annualwage <- newdata %>% filter(Annual.wage == max(Annual.wage))
```

```
max.annualwage$Annual.wage # the maximum annual wage is $196,568
```

```
## [1] 196568
```

```
max.annualwage$Occupational.title # the occupation that has the highest
annual wage is Nurse Anesthetists
```

```
## [1] Nurse Anesthetists
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife
Biologists
```

```
max.annualwage$Employment # the number of people employed with the highest
annual wage is 624
```

```
## [1] 624
```

```
min.annualwage <- newdata %>% filter(Annual.wage == min(Annual.wage))
```

```
min.annualwage$Annual.wage # the lowest annual wage is $22,043
```

```
## [1] 22043
```

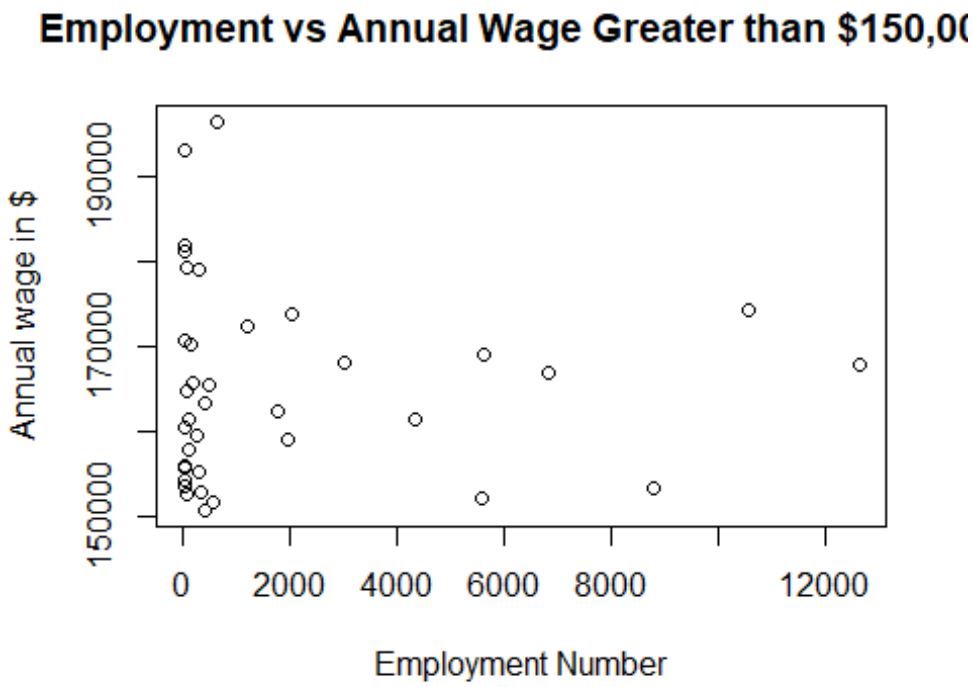
```r
min.annualwage$Occupational.title # the occupation that has the lowest annual wage is Bartenders
```

```
## [1] Bartenders
## 756 Levels: Accountants and Auditors Actors ... Zoologists and Wildlife Biologists
```

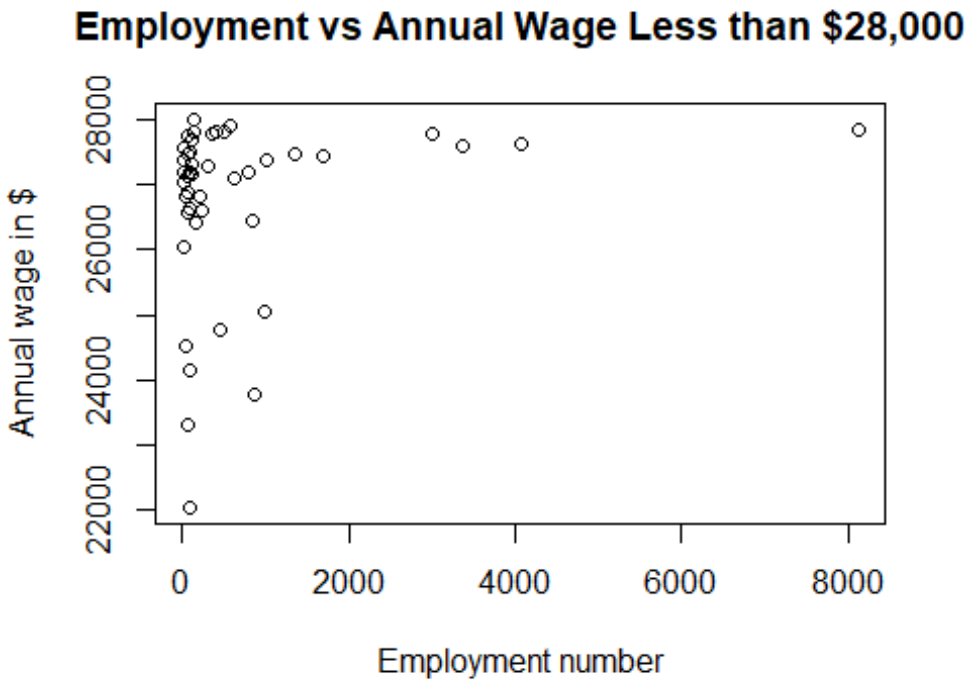
```r
min.annualwage$Employment # the number of people employed with the lowest annual wage is 89
```

```
## [1] 89
```

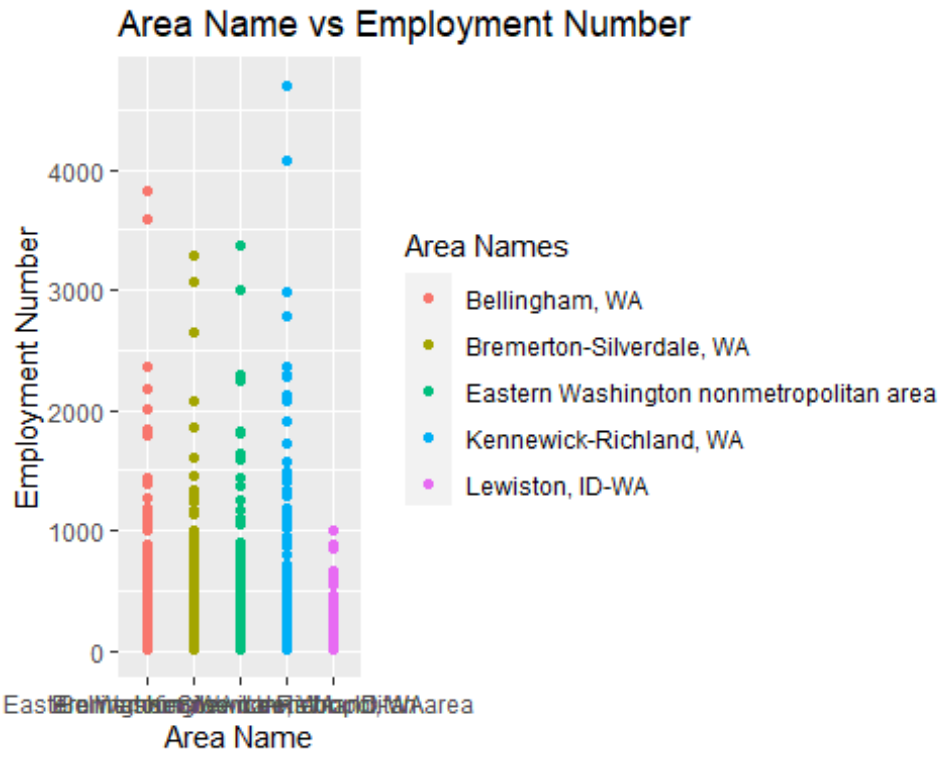
```r
# Comparing employment and annual wage
high.annual <- newdata %>% filter(Annual.wage >= 150000) #Annual wage is higher than $150,000
plot(high.annual$Employment, high.annual$Annual.wage, xlab = "Employment Number", ylab = "Annual wage in $", main = "Employment vs Annual Wage Greater than $150,000")
```



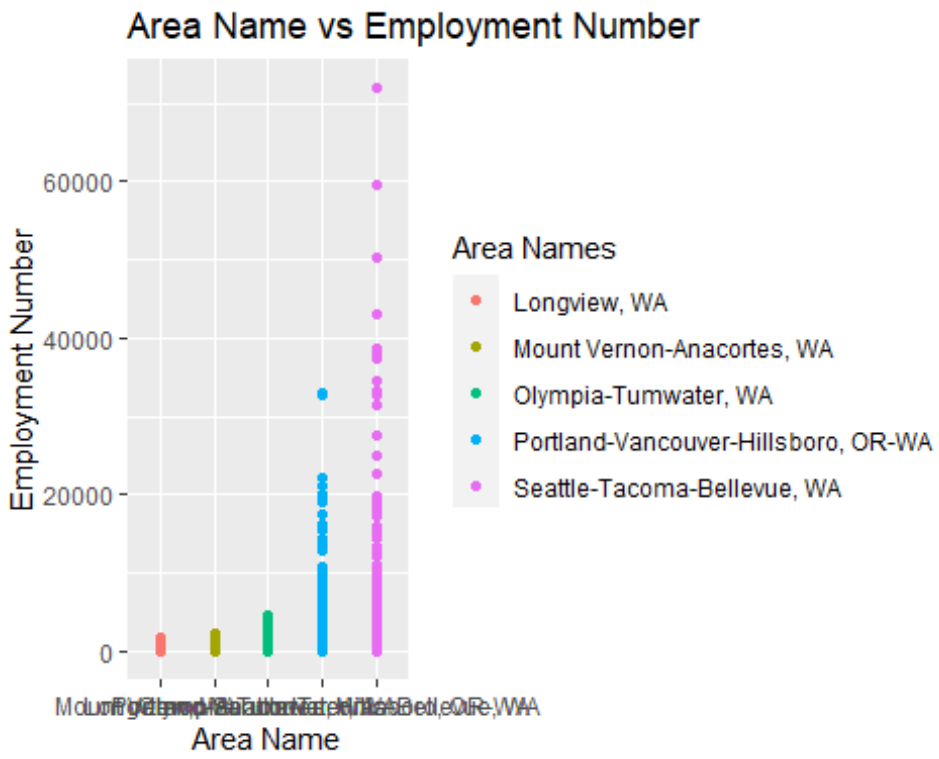

```r
# talk about distribution between wage and employment number when annual wage is greater than 150,000

low.annual <- newdata %>% filter(Annual.wage <= 28000) # Annual wage is lower than $28,000
plot(low.annual$Employment, low.annual$Annual.wage, xlab = "Employment number", ylab = "Annual wage in $", main = "Employment vs Annual Wage Less than $28,000")
```



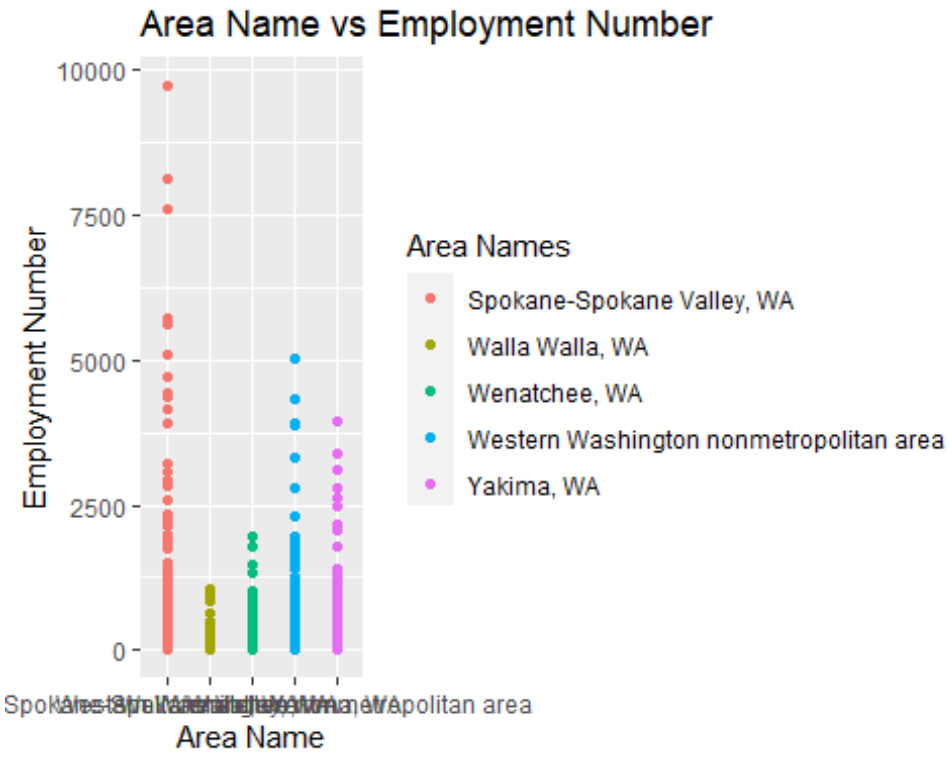

```
# talk about distribution between wage and employment number when annual wage
is less than 28,000

# Comparing Area and Employment numbers with plots
area1 <- newdata %>% filter(Area.name %in% c("Bellingham, WA", "Bremerton-
Silverdale, WA", "Eastern Washington nonmetropolitan area", "Kennewick-
Richland, WA", "Lewiston, ID-WA"))
ggplot(data = area1, aes(x = area1$Area.name, y = area1$Employment, col =
area1$Area.name)) +
  geom_point() + labs(x = "Area Name", y = "Employment Number", title = "Area
Name vs Employment Number", col = "Area Names")
```




```
area2 <- newdata %>% filter(Area.name %in% c("Longview, WA", "Mount Vernon-
Anacortes, WA", "Olympia-Tumwater, WA", "Portland-Vancouver-Hillsboro, OR-
WA", "Seattle-Tacoma-Bellevue, WA"))
ggplot(data = area2, aes(x = area2$Area.name, y = area2$Employment, col =
area2$Area.name)) +
  geom_point() + labs(x = "Area Name", y = "Employment Number", title = "Area
Name vs Employment Number", col = "Area Names")
```




```
area3 <- newdata %>% filter(Area.name %in% c("Spokane-Spokane Valley, WA",
"Walla Walla, WA","Wenatchee, WA", "Western Washington nonmetropolitan area",
"Yakima, WA"))
ggplot(data = area3, aes(x = area3$Area.name, y = area3$Employment, col =
area3$Area.name)) +
  geom_point() + labs(x = "Area Name", y = "Employment Number", title = "Area
Name vs Employment Number", col = "Area Names")
```




```
# talk about distribution between areas and number of people employed

# Most reliable place to gain employemt
# highest employment numbers from the graphs are from the areas "Spokane-
Spokane Valley, WA" and Seattle-Tacoma-Bellevue, WA

area.SSV <- newdata %>% filter(Area.name == "Spokane-Spokane Valley, WA")
area.STB <- newdata %>% filter(Area.name == "Seattle-Tacoma-Bellevue, WA")

sum(area.SSV$Employment) # total number of people employed in the Spokane-
Spokane Valley, WA area is 219698

## [1] 219698

sum(area.STB$Employment) # total number of people employed in the Seattle-
Tacoma-Bellevue area is 1,896,626

## [1] 1896626

mean(area.SSV$Average.wage) # the mean of the average pay per hour in the
Spokane-Spokane Valley, WA area is 28.22703

## [1] 28.22703

mean(area.STB$Average.wage) # the mean of the average pay per hour in the
Seattle-Tacoma-Bellevue area is 32.9287

## [1] 32.9287
```

```
# Conclusion is that there is a higher chance of gaining employment in the
Seattle-Tacoma-Bellevue, WA area as it has the most number of
# employed individuals and a good average wage across all employment titles.
```