

# Project++

Mai Luu

## Data Explanation

### The general context of the data

The data contains a record of Employment in 4 States: Alabama, Alaska, Arizona, and Arkansas surrounding in 2018.

### What the observation represent? And the information measured for each observation.

Each record details rural or urban code. If the continuum code is from 1-3, which means it is in Rural. Otherwise, the area is in the Urban area. The data has 185 observations which are in order from State AL to AZ. The data also provides the number of Employment, Unemployment, and Median Household Income so the audience can easily compare the employment rate of each state throughout the year. The employed\_2018 columns represent the number of people who are employed. And the same to Unemployed\_2018 column. The median\_Household is the median income of one household in the year 2018. It is measured by a thousand dollars. The rural\_urban\_continuum\_code is identified from 1-3. 1 has the most population, 2 has the population from 250,000 to 1 million, and 3 has the least population in the rural area. In Urban Area: 5,7,9 not adjacent to the metro area, in descending order from the most population to the least 4,6,8 adjacent to a metro area, in the descending area from the most population to the least.

## Data Characteristic

The type of data is a sample. The full dataset includes 50 states in the US and it provides unemployment and employment from the past 10 years. The data is collected from the

United States Department of Agriculture page. The selection method is downloading the full dataset to Excel Sheet and sort the data to only the first 4 State I want to compare with. I chose the data which is the most recent (2018) and eliminate the dataset before that year.

## Numerical summary

### The average of unemployment and employment by rural-urban area

```
library(ggplot2)
library(tidyverse)
```

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

```
## v tibble  2.1.3      v dplyr    0.8.3
## v tidyr   1.0.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0
## v purrr   0.3.3
```

```
## -- Conflicts ----- tidyverse_conflict
```

```
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
Unemployment<-read.csv("/Users/mailuu/Desktop/STAT3080/Unemployment.csv", header=TRUE,
stringsAsFactors =FALSE,
fileEncoding = "latin1")
```

```
TheAverage<- group_by(Unemployment, Urban_Rural)
summarize(TheAverage, avg.unemployment=mean(Unemployed_2018),
          avg.employment=mean(Employed_2018))
```

```
## # A tibble: 2 x 3
##   Urban_Rural avg.unemployment avg.employment
##   <chr>         <dbl>         <dbl>
## 1 Rural         2928.         63190.
## 2 Urban         466.          8737.
```

The average of unemployment and employment in rural area higher than urban area.

Determine the average number of occurrences of employment and unemployment over the states with the household income that is less than \$50,000

```
lessthan50000 <- Unemployment$Median_Household_Income_2018 < 50000
lessthan50000list<-Unemployment[lessthan50000,]
mean(lessthan50000list$Employed_2018)
```

```
## [1] 17542
```

```
mean(lessthan50000list$Unemployed_2018)
```

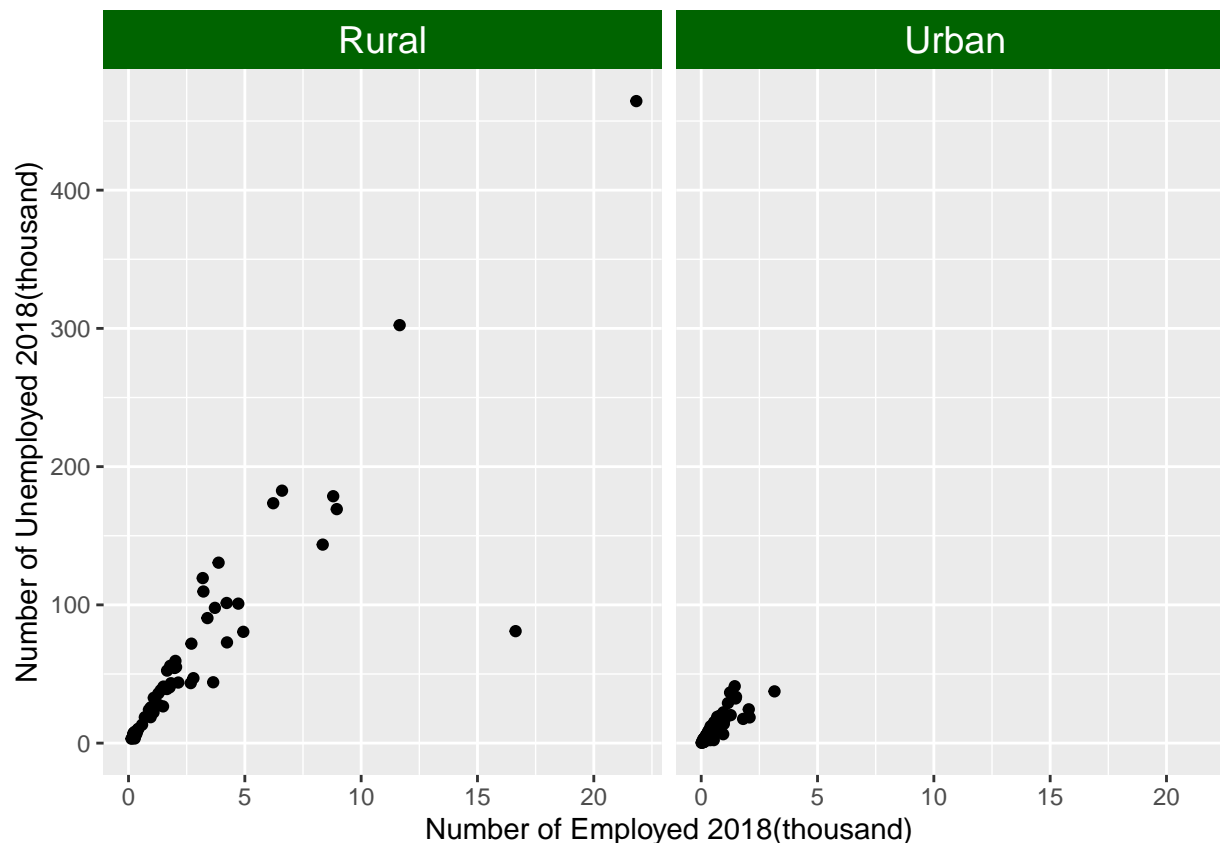
```
## [1] 924.1087
```

The average number of occurrences of employment over the states with the household income less than \$50,000 is 17542 The average number of occurrences of unemployment over the states with the household income less than \$50,000 is 924.1087

## Graphical Summary

The median of Employed 2018 of each State separated by urban or rural

```
Rural_Urban <- ggplot(Unemployment,aes(x=Unemployed_2018/1000, y=Employed_2018/1000)) +
  geom_point() +
  facet_grid(.~Urban_Rural)
Rural_Urban + theme(strip.background=element_rect(fill="darkgreen"),
  strip.text.x=element_text(size=14, color="white")) +
  labs(x="Number of Employed 2018(thousand)",
  y="Number of Unemployed 2018(thousand)")
```



```
Median_Income<-round(Unemployment$Median_Household_Income_2018/1000,digits=0)
Medi_Unemp <-ggplot(Unemployment, aes(x=Median_Income,
  y=Unemployed_2018, color=Unemployment$i..States)) +
  geom_point(shape=5) +
  scale_color_manual(values=c(
    "AL"="orange","AK"="navy","AZ"="dark gray","AR"="blue")) +
  scale_fill_manual(values=c(
    "AL"="orange","AK"="navy","AZ"="dark gray","AR"="blue"))
Medi_Unemp + labs ( title="Median Household income with Unemployed rate 2018",
  x="Income of Household (thousand)", y="Number of Unemployment") +
  scale_color_manual(values=c("AL"="orange","AK"="navy","AZ"="dark gray","AR"="blue"),
  labels=c("Alabama","Alaska","Arizona","Arkansas"), name="States") +
  theme (legend.position=c(1,1), legend.justification=c(1,1))
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```

Median Household income with Unemployed rate 2018



## Conclusion

The average unemployment and employment in rural areas higher than the urban areas. It indicates that urban area is more competitive and has a limited job available. The last graph shows that having a low income is not much related to the high unemployment in 4 States According to the graph, a lower Income does not have a higher rate of unemployment. The unemployment rate is still quite low with the low household income The income of the household is mostly around \$40,000 to \$50,000. The State that has the highest employment and unemployment rate is Arizona Arkansas has the highest household income and unemployment. The number of unemployment and employment in a rural area is more varied and spread out more in the urban area.