

EDA Project

Joshua Lemley

2024-09-23

Introduction

We are going to look at the data set "county". This dataset has 15 variables but we are going to focus on two. In We are interested in looking more into pverty and unemployment.

Univaritate Analysis

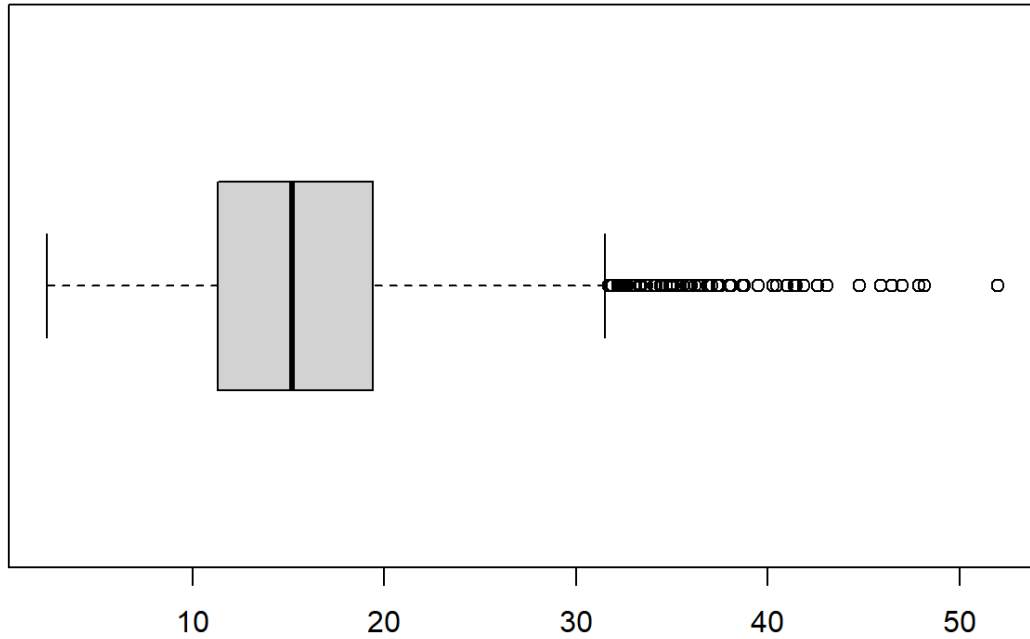
Poverty

```
summary(county$poverty)
```

```
##   Min. 1st Qu.  Median   Mean 3rd Qu.  Max.   NA's  
##  2.40  11.30  15.20  15.97  19.40  52.00    2
```

```
boxplot(county$poverty, horizontal = TRUE, main= "Distribution of Poverty in Counties")
```

Distribution of Poverty in Counties



This summary helps us see that the average percent of poverty in counties is around 16%. The lowest amount of poverty is 2.4% and highest 52%.

Income

```
table(county$unemployment_rate)
```

```
##
## 1.62 1.64 1.66 1.73 1.75 1.76 1.77 1.8 1.82 1.86 1.87 1.88 1.89
## 1 1 3 1 1 2 1 1 2 1 2 1 2
## 1.9 1.93 1.95 1.97 1.99 2 2.02 2.03 2.04 2.06 2.08 2.1 2.12
## 1 2 1 1 3 2 2 1 1 1 1 1 3
## 2.14 2.16 2.17 2.18 2.2 2.22 2.23 2.24 2.25 2.26 2.27 2.28 2.3
## 2 3 2 2 1 2 2 4 2 2 2 3 2
## 2.31 2.32 2.33 2.34 2.35 2.36 2.37 2.38 2.39 2.4 2.41 2.42 2.43
## 1 2 2 2 3 4 1 2 2 5 4 3 5
## 2.44 2.45 2.46 2.47 2.48 2.49 2.5 2.51 2.52 2.53 2.54 2.55 2.56
## 4 6 2 4 6 5 2 3 8 7 4 2 3
## 2.57 2.58 2.59 2.6 2.61 2.62 2.63 2.64 2.65 2.66 2.67 2.68 2.69
## 3 4 5 1 8 3 4 7 2 6 7 5 2
## 2.7 2.72 2.73 2.74 2.75 2.76 2.77 2.78 2.79 2.8 2.81 2.82 2.83
## 5 7 5 3 4 3 7 5 10 5 4 3 4
## 2.84 2.85 2.86 2.87 2.88 2.89 2.9 2.91 2.92 2.93 2.94 2.95 2.96
## 7 6 4 7 8 2 6 10 8 5 3 11 4
```

2.97 2.98 2.99 3 3.01 3.02 3.03 3.04 3.05 3.06 3.07 3.08 3.09
5 4 6 7 9 4 8 7 10 13 10 12 8
3.1 3.11 3.12 3.13 3.14 3.15 3.16 3.17 3.18 3.19 3.2 3.21 3.22
6 3 7 6 9 3 7 8 9 11 9 9 9
3.23 3.24 3.25 3.26 3.27 3.28 3.29 3.3 3.31 3.32 3.33 3.34 3.35
5 3 9 7 10 10 5 10 10 8 4 12 5
3.36 3.37 3.38 3.39 3.4 3.41 3.42 3.43 3.44 3.45 3.46 3.47 3.48
8 4 7 5 5 7 9 9 6 6 7 8 9
3.49 3.5 3.51 3.52 3.53 3.54 3.55 3.56 3.57 3.58 3.59 3.6 3.61
14 5 9 9 9 14 4 13 7 11 10 7 11
3.62 3.63 3.64 3.65 3.66 3.67 3.68 3.69 3.7 3.71 3.72 3.73 3.74
14 11 7 10 15 7 7 12 8 9 10 10 7
3.75 3.76 3.77 3.78 3.79 3.8 3.81 3.82 3.83 3.84 3.85 3.86 3.87
5 8 11 7 7 14 7 9 15 13 10 8 9
3.88 3.89 3.9 3.91 3.92 3.93 3.94 3.95 3.96 3.97 3.98 3.99 4
7 5 12 10 10 8 5 6 9 12 8 11 8
4.01 4.02 4.03 4.04 4.05 4.06 4.07 4.08 4.09 4.1 4.11 4.12 4.13
7 19 4 7 10 10 8 13 10 4 11 12 9
4.14 4.15 4.16 4.17 4.18 4.19 4.2 4.21 4.22 4.23 4.24 4.25 4.26
9 9 17 4 12 6 12 7 7 14 9 9 10
4.27 4.28 4.29 4.3 4.31 4.32 4.33 4.34 4.35 4.36 4.37 4.38 4.39
8 6 9 12 8 8 8 9 8 13 15 7 12
4.4 4.41 4.42 4.43 4.44 4.45 4.46 4.47 4.48 4.49 4.5 4.51 4.52
14 8 8 12 12 7 8 10 10 10 15 10 11
4.53 4.54 4.55 4.56 4.57 4.58 4.59 4.6 4.61 4.62 4.63 4.64 4.65
15 10 7 10 7 5 5 9 15 8 3 7 10
4.66 4.67 4.68 4.69 4.7 4.71 4.72 4.73 4.74 4.75 4.76 4.77 4.78
7 2 14 8 7 5 7 11 9 10 7 8 5
4.79 4.8 4.81 4.82 4.83 4.84 4.85 4.86 4.87 4.88 4.89 4.9 4.91
9 12 8 8 7 8 7 8 13 6 4 12 2
4.92 4.93 4.94 4.95 4.96 4.97 4.98 4.99 5 5.01 5.02 5.03 5.04
5 8 2 2 8 7 5 7 12 10 11 9 2
5.05 5.06 5.07 5.08 5.09 5.1 5.11 5.12 5.13 5.14 5.15 5.16 5.17
4 4 7 14 11 7 2 4 12 4 6 6 8
5.18 5.19 5.2 5.21 5.22 5.23 5.24 5.25 5.26 5.27 5.28 5.29 5.3
14 6 4 10 7 8 9 5 9 1 13 6 2
5.31 5.32 5.33 5.34 5.35 5.36 5.37 5.38 5.39 5.4 5.41 5.42 5.43
6 9 3 3 6 5 5 8 5 3 4 2 2
5.44 5.45 5.46 5.47 5.48 5.49 5.5 5.51 5.52 5.53 5.54 5.55 5.56
4 1 8 5 4 8 5 5 9 6 5 1 2
5.57 5.58 5.59 5.6 5.61 5.62 5.63 5.64 5.65 5.66 5.67 5.68 5.69
3 6 5 2 5 6 4 3 6 7 3 5 3
5.7 5.71 5.72 5.73 5.74 5.75 5.76 5.77 5.78 5.79 5.8 5.81 5.82
4 3 5 4 5 4 4 3 8 4 6 6 6
5.83 5.84 5.85 5.86 5.87 5.88 5.89 5.9 5.91 5.92 5.93 5.94 5.95
5 2 7 9 4 1 5 8 3 6 3 4 6
5.96 5.97 5.98 5.99 6 6.01 6.02 6.03 6.04 6.05 6.06 6.07 6.08
4 9 4 7 2 5 5 4 4 4 4 5 11
6.09 6.1 6.11 6.12 6.13 6.14 6.15 6.16 6.17 6.18 6.19 6.2 6.21

```

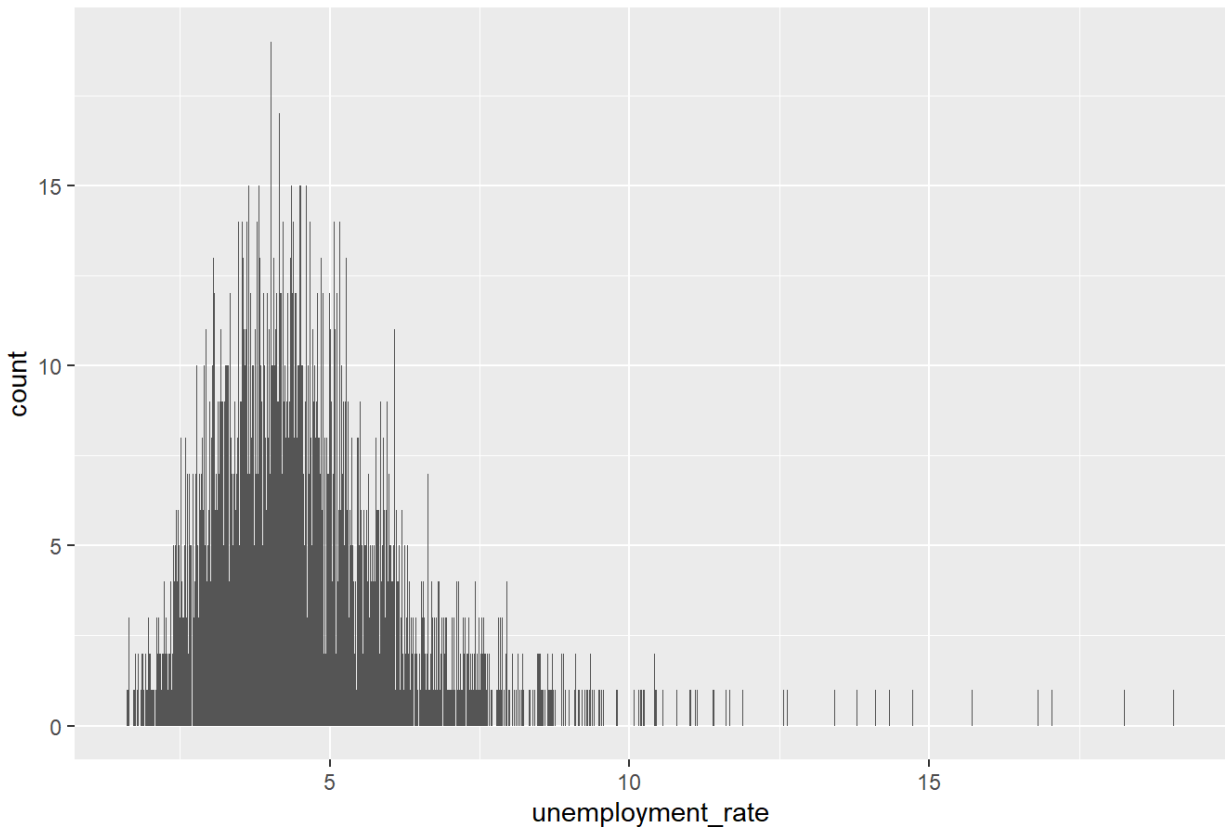
## 7 1 2 6 4 3 4 4 5 1 2 3 6
## 6.23 6.24 6.25 6.26 6.27 6.28 6.29 6.3 6.31 6.32 6.33 6.34 6.35
## 2 1 5 4 2 2 3 3 5 2 4 2 1
## 6.36 6.37 6.38 6.39 6.42 6.43 6.44 6.45 6.46 6.47 6.51 6.52 6.53
## 3 2 2 3 2 1 3 2 2 1 2 1 4
## 6.54 6.55 6.56 6.57 6.58 6.59 6.6 6.61 6.63 6.64 6.65 6.66 6.67
## 3 3 4 4 3 2 1 1 2 3 7 1 1
## 6.69 6.7 6.71 6.72 6.74 6.75 6.76 6.77 6.78 6.8 6.81 6.82 6.83
## 2 4 2 3 1 1 3 1 3 1 4 3 4
## 6.84 6.85 6.86 6.87 6.88 6.91 6.93 6.94 6.95 6.96 6.97 6.99 7
## 1 1 2 3 3 3 2 3 3 1 1 1 1
## 7.01 7.03 7.04 7.05 7.06 7.07 7.08 7.09 7.1 7.12 7.14 7.15 7.16
## 1 1 3 2 1 3 1 1 1 4 1 4 1
## 7.17 7.18 7.19 7.2 7.21 7.22 7.23 7.24 7.25 7.26 7.27 7.29 7.3
## 1 2 2 1 1 1 3 1 2 3 1 1 2
## 7.31 7.32 7.33 7.34 7.35 7.36 7.37 7.38 7.39 7.4 7.42 7.43 7.44
## 2 2 3 1 1 2 1 2 2 1 2 1 4
## 7.45 7.46 7.47 7.48 7.5 7.51 7.52 7.53 7.54 7.56 7.57 7.59 7.6
## 1 2 1 1 3 1 1 2 3 2 3 2 1
## 7.61 7.62 7.63 7.64 7.66 7.69 7.71 7.79 7.8 7.82 7.83 7.84 7.85
## 2 1 2 1 2 1 1 1 1 3 1 1 3
## 7.86 7.87 7.88 7.89 7.91 7.92 7.94 7.96 7.99 8 8.01 8.05 8.08
## 1 1 3 1 1 1 2 4 1 1 1 2 1
## 8.11 8.14 8.15 8.17 8.21 8.23 8.24 8.33 8.35 8.4 8.42 8.47 8.48
## 1 2 1 1 1 2 1 1 1 1 1 2 2
## 8.49 8.5 8.52 8.53 8.54 8.55 8.56 8.57 8.6 8.64 8.65 8.66 8.68
## 2 2 2 1 1 1 1 1 1 2 1 1 1
## 8.69 8.71 8.72 8.73 8.76 8.87 8.88 8.91 8.93 9 9.09 9.1 9.15
## 1 1 2 1 1 2 2 2 1 1 1 2 1
## 9.17 9.21 9.26 9.29 9.3 9.34 9.36 9.38 9.41 9.5 9.51 9.54 9.57
## 1 1 1 1 1 1 2 1 1 1 1 1 1
## 9.79 9.8 10.08 10.16 10.19 10.21 10.24 10.25 10.42 10.44 10.45 10.56 10.8
## 1 1 1 1 1 1 1 1 2 1 1 1 1
## 11.01 11.03 11.11 11.13 11.4 11.42 11.62 11.67 11.89 12.57 12.63 13.43 13.79
## 1 1 1 1 1 1 1 1 1 1 1 1 1
## 14.11 14.34 14.72 15.71 16.82 17.05 18.26 19.07
## 1 1 1 1 1 1 1 1

```

```
ggplot(county, aes(x=unemployment_rate)) +geom_bar() + ggtitle("Employment Rates of Counties")
```

```
## Warning: Removed 3 rows containing non-finite outside the scale range
## (`stat_count()`).
```

Employment Rates of Counties



This shows us that the unemployment rate is mainly around 5% in most counties.

Bivariate Analysis

```
summary(county$poverty, county$unemployment_rate)
```

```
##  Min. 1st Qu.  Median   Mean 3rd Qu.  Max.  NA's  
##  2.40  11.30  15.20  15.97  19.40  52.00   2
```

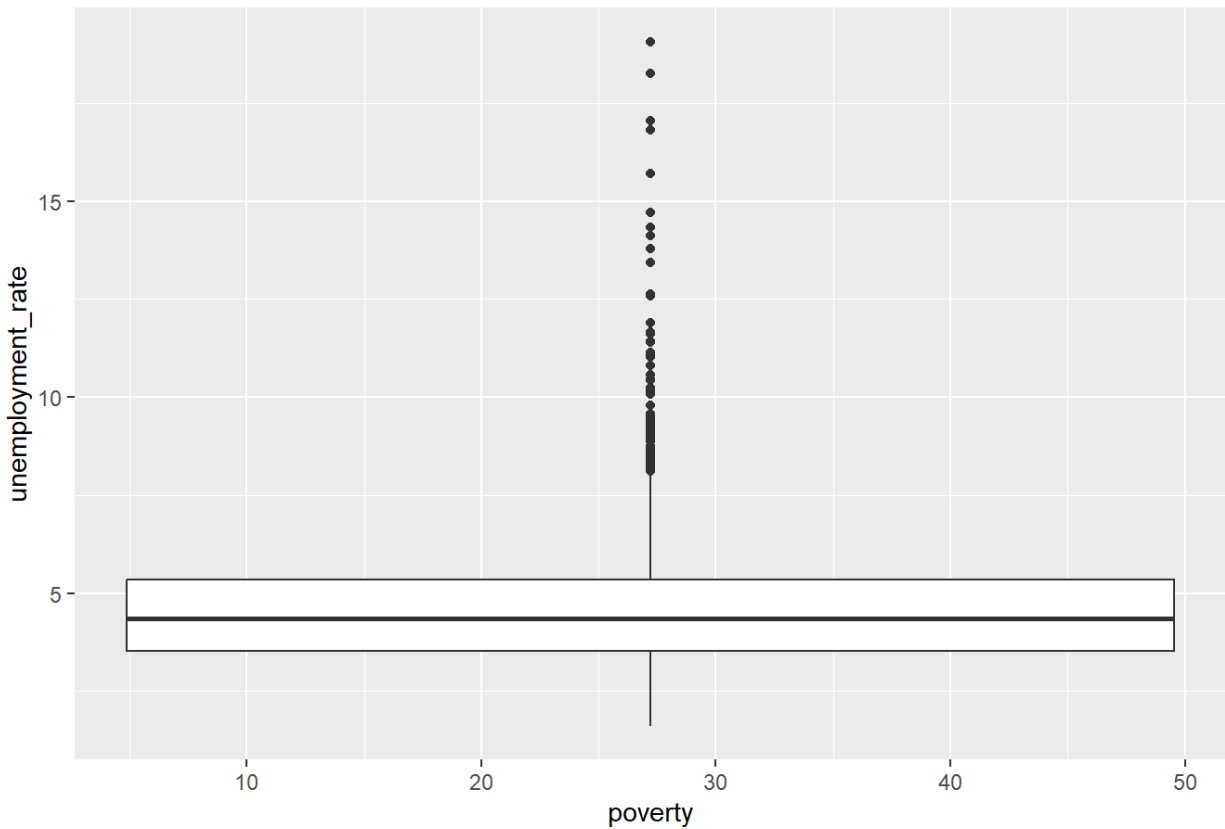
```
ggplot(county, aes(y=unemployment_rate, x=poverty)) +  
geom_boxplot() + ggtitle("Relationship Between Poverty and Unemployment Rate")
```

```
## Warning: Continuous x aesthetic  
## i did you forget `aes(group = ...)`?
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range  
## (`stat_boxplot()`).
```

```
## Warning: Removed 1 row containing non-finite outside the scale range  
## (`stat_boxplot()`).
```

Relationship Between Poverty and Unemployment Rate



The two graphics show us the relationship between poverty and the unemployment rate in counties across America.

Conclusion

The figures show us how the percent of poverty and unemployment rate are proportional across American Counties. We can see that the counties with higher poverty have higher unemployment rates compared to those who are lower. This seems to correspond with many counties in America.