

Final Project

Kailey Mullins

02/25/22

```
knitr::opts_chunk$set(fig.width=6, fig.height=4) # This sets all figure sizes in the document unless ot
library(dplyr)
```

```
##
## 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(ggplot2)

library(readxl)
CD <-read_excel("/Users/KaileyMullins/Desktop/MATH/Crime_Data.xlsx", sheet=1, col_names = TRUE)
```

FIRST QUESTION

1. Introduction: The excel sheet I am using in the Crime Data set. The set includes 51 observations and 11 variables. The data set is about the crime rates of murder and crime in all 51 states. It also gives some other information on the what percentages of races, poverty, and the region. My research question is what relgion has the highest murder rate in the United States?Im interesting in finding out about this because I want to know what region in the United States as the most murders, which coincides with how much crime is in that region.

SECOND QUESTION

2. Univariate Exploration: The two variables that is under consideration is murder rates and the region. The mean of murder is 8.27 and there is 12 states in the midwest, 9 states in the northeast, 17 states in south, and 13 states in the west. The minimum rate of murders in a state is 1.6 in Maine and the highest rate of murders is in Lousiana.

```
table (CD$region , useNA = "always" )
```

```
##
##   Midwest Northeast   South   West   <NA>
##      12         9      17     13      0
```

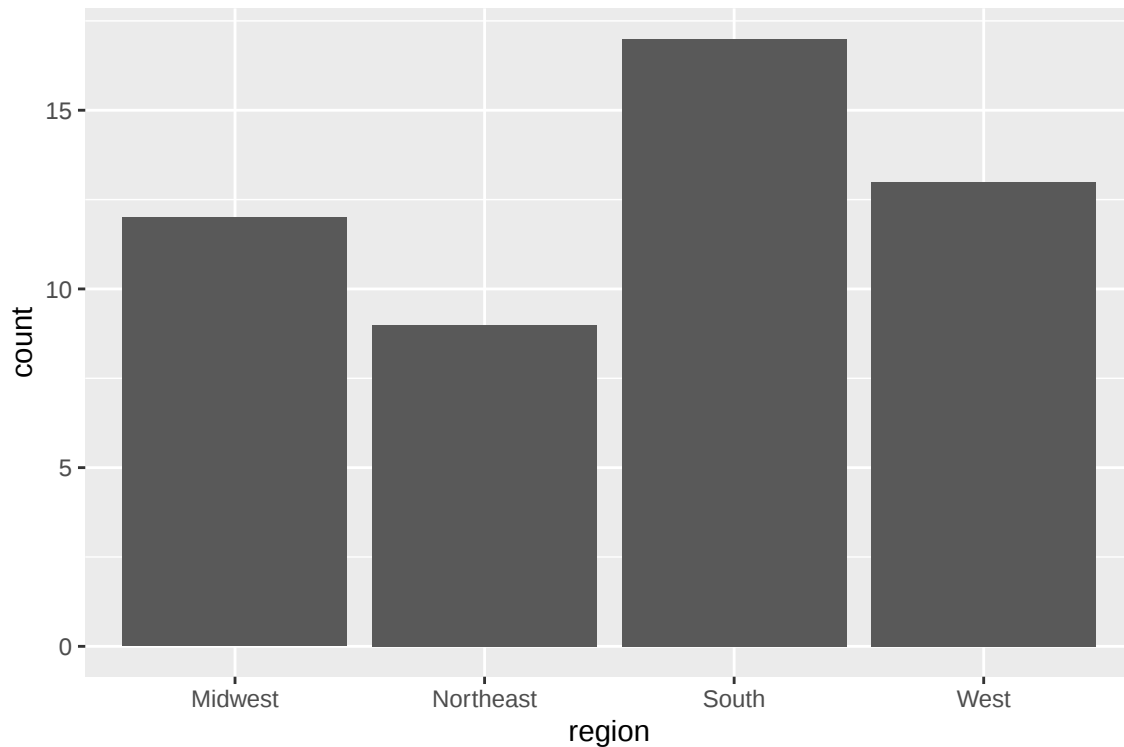
```
summary(CD$murder)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.600   3.900   6.800   8.727  10.350   78.500
```

```
CD$murder %>% mean()
```

```
## [1] 8.727451
```

```
ggplot(CD, aes(x=region, FILL= murder)) + geom_bar(position = "dodge")
```



```
round(CD$murder, digits= 3)
```

```
## [1]  9.0 11.6 10.2  8.6 13.1  5.8  6.3  5.0  8.9 11.4  3.8  2.3  2.9 11.4  7.5
## [16]  6.4  6.6 20.3  3.9 12.7  1.6  9.8  3.4 11.3 13.5  3.0 11.3  1.7  3.9  2.0
## [31]  5.3  8.0 10.4 13.3  6.0  8.4  4.6  6.8  3.9 10.3  3.4 10.2 11.9  3.1  8.3
## [46]  3.6  5.2  4.4  6.9  3.4 78.5
```

THIRD QUESTION Bivariate Exploration

```
table(CD$region, CD$murder)
```

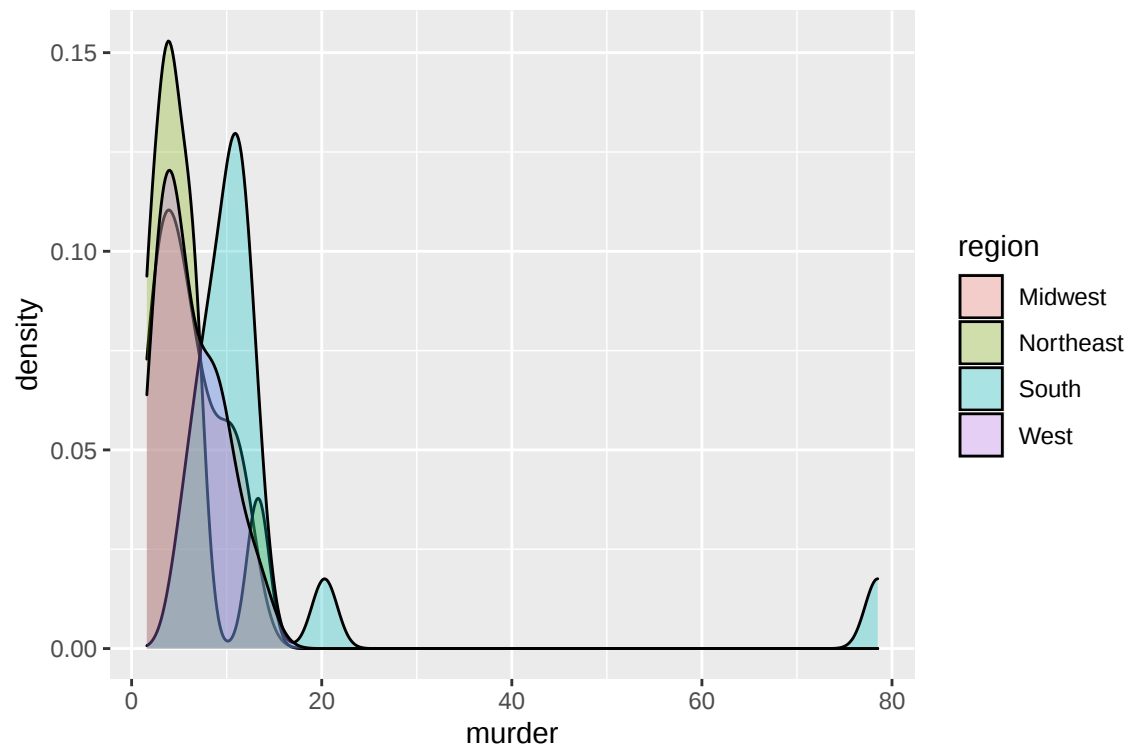
```
##
##      1.600000024 1.700000048 2 2.299999952 2.900000095 3 3.099999905
## Midwest      0      1 0      1      0 0      0
```

```

## Northeast      1      0 1      0      0 0      0
## South          0      0 0      0      0 0      0
## West           0      0 0      0      1 1      1
##
##      3.400000095 3.599999905 3.799999952 3.900000095 4.400000095
## Midwest        2      0      0      1      1
## Northeast      0      1      0      2      0
## South          0      0      0      0      0
## West           1      0      1      0      0
##
##      4.599999905 5 5.199999809 5.300000191 5.800000191 6 6.300000191
## Midwest        0 0      0      0      0 1      0
## Northeast      0 0      0      1      0 0      1
## South          0 1      0      0      0 0      0
## West           1 0      1      0      1 0      0
##
##      6.400000095 6.599999905 6.800000191 6.900000095 7.5 8 8.300000191
## Midwest        1      0      0      0 1 0      0
## Northeast      0      0      1      0 0 0      0
## South          0      1      0      1 0 0      1
## West           0      0      0      0 0 1      0
##
##      8.399999619 8.600000381 8.899999619 9 9.800000191 10.19999981
## Midwest        0      0      0 0      1      0
## Northeast      0      0      0 0      0      0
## South          1      0      1 0      0      2
## West           0      1      0 1      0      0
##
##      10.30000019 10.39999962 11.30000019 11.39999962 11.60000038
## Midwest        0      0      1      1      0
## Northeast      0      0      0      0      0
## South          1      0      1      1      1
## West           0      1      0      0      0
##
##      11.89999962 12.69999981 13.10000038 13.30000019 13.5 20.29999924
## Midwest        0      0      0      0 0      0
## Northeast      0      0      0      1 0      0
## South          1      1      0      0 1      1
## West           0      0      1      0 0      0
##
##      78.5
## Midwest        0
## Northeast      0
## South          1
## West           0

```

```
ggplot(CD, aes(x=murder, fill=region)) + geom_density(alpha=.3)
```



```
prop.table(table(CD$region, CD$murder))
```

```
##
##      1.600000024 1.700000048      2 2.299999952 2.900000095
## Midwest      0.00000000 0.01960784 0.00000000 0.01960784 0.00000000
## Northeast 0.01960784 0.00000000 0.01960784 0.00000000 0.00000000
## South      0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## West       0.00000000 0.00000000 0.00000000 0.00000000 0.01960784
##
##      3 3.099999905 3.400000095 3.599999905 3.799999952
## Midwest      0.00000000 0.00000000 0.03921569 0.00000000 0.00000000
## Northeast 0.00000000 0.00000000 0.00000000 0.01960784 0.00000000
## South      0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## West       0.01960784 0.01960784 0.01960784 0.00000000 0.01960784
##
##      3.900000095 4.400000095 4.599999905      5 5.199999809
## Midwest      0.01960784 0.01960784 0.00000000 0.00000000 0.00000000
## Northeast 0.03921569 0.00000000 0.00000000 0.00000000 0.00000000
## South      0.00000000 0.00000000 0.00000000 0.01960784 0.00000000
## West       0.00000000 0.00000000 0.01960784 0.00000000 0.01960784
##
##      5.300000191 5.800000191      6 6.300000191 6.400000095
## Midwest      0.00000000 0.00000000 0.01960784 0.00000000 0.01960784
## Northeast 0.01960784 0.00000000 0.00000000 0.01960784 0.00000000
## South      0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## West       0.00000000 0.01960784 0.00000000 0.00000000 0.00000000
##
##      6.599999905 6.800000191 6.900000095      7.5      8
```

```
## Midwest 0.00000000 0.00000000 0.00000000 0.01960784 0.00000000
## Northeast 0.00000000 0.01960784 0.00000000 0.00000000 0.00000000
## South 0.01960784 0.00000000 0.01960784 0.00000000 0.00000000
## West 0.00000000 0.00000000 0.00000000 0.00000000 0.01960784
##
## 8.300000191 8.399999619 8.600000381 8.899999619 9
## Midwest 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## Northeast 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## South 0.01960784 0.01960784 0.00000000 0.01960784 0.00000000
## West 0.00000000 0.00000000 0.01960784 0.00000000 0.01960784
##
## 9.800000191 10.19999981 10.30000019 10.39999962 11.30000019
## Midwest 0.01960784 0.00000000 0.00000000 0.00000000 0.01960784
## Northeast 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## South 0.00000000 0.03921569 0.01960784 0.00000000 0.01960784
## West 0.00000000 0.00000000 0.00000000 0.01960784 0.00000000
##
## 11.39999962 11.60000038 11.89999962 12.69999981 13.10000038
## Midwest 0.01960784 0.00000000 0.00000000 0.00000000 0.00000000
## Northeast 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
## South 0.01960784 0.01960784 0.01960784 0.01960784 0.00000000
## West 0.00000000 0.00000000 0.00000000 0.00000000 0.01960784
##
## 13.30000019 13.5 20.29999924 78.5
## Midwest 0.00000000 0.00000000 0.00000000 0.00000000
## Northeast 0.01960784 0.00000000 0.00000000 0.00000000
## South 0.00000000 0.01960784 0.01960784 0.01960784
## West 0.00000000 0.00000000 0.00000000 0.00000000
```

```
by_region <- group_by(CD, region, murder)
summarize(by_region)
```

```
## 'summarise()' has grouped output by 'region'. You can override using the
## '.groups' argument.
```

```
## # A tibble: 48 x 2
## # Groups:   region [4]
##   region murder
##   <chr>   <dbl>
## 1 Midwest 1.70
## 2 Midwest 2.30
## 3 Midwest 3.40
## 4 Midwest 3.90
## 5 Midwest 4.40
## 6 Midwest 6
## 7 Midwest 6.40
## 8 Midwest 7.5
## 9 Midwest 9.80
## 10 Midwest 11.3
## # ... with 38 more rows
```

```
summarize(by_region, count=n())
```

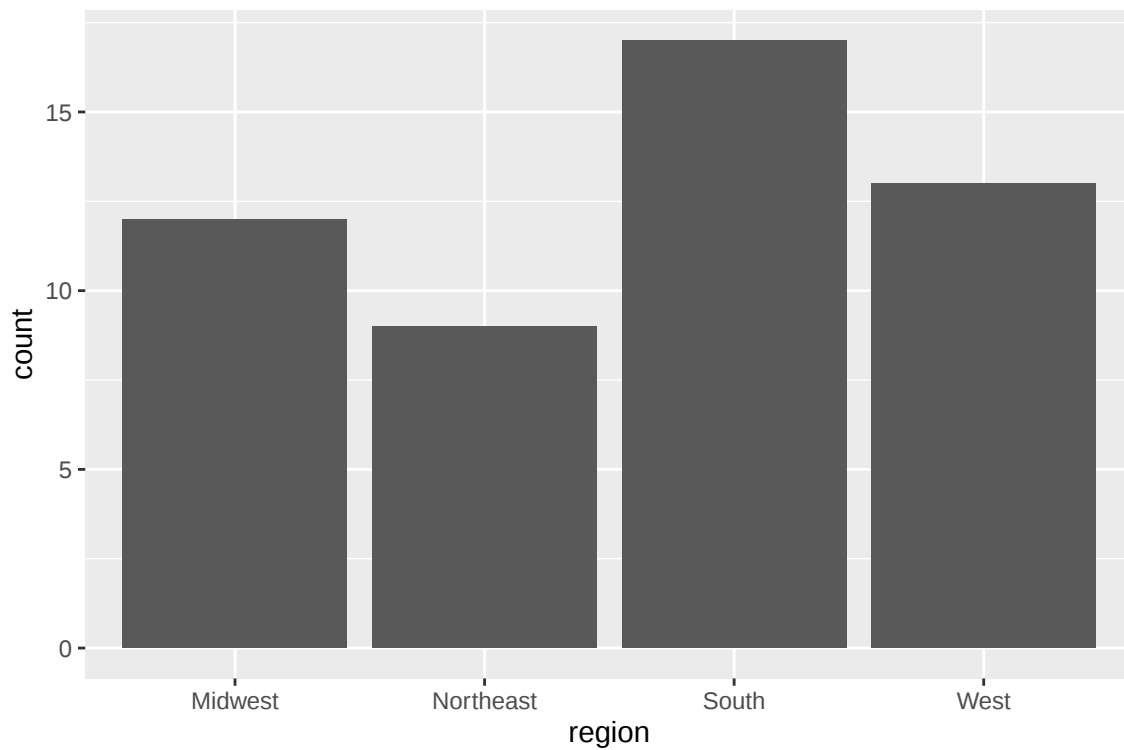
```
## 'summarise()' has grouped output by 'region'. You can override using the  
## '.groups' argument.
```

```
## # A tibble: 48 x 3  
## # Groups:   region [4]  
##   region murder count  
##   <chr>   <dbl> <int>  
## 1 Midwest  1.70     1  
## 2 Midwest  2.30     1  
## 3 Midwest  3.40     2  
## 4 Midwest  3.90     1  
## 5 Midwest  4.40     1  
## 6 Midwest  6       1  
## 7 Midwest  6.40     1  
## 8 Midwest  7.5     1  
## 9 Midwest  9.80     1  
## 10 Midwest 11.3     1  
## # ... with 38 more rows
```

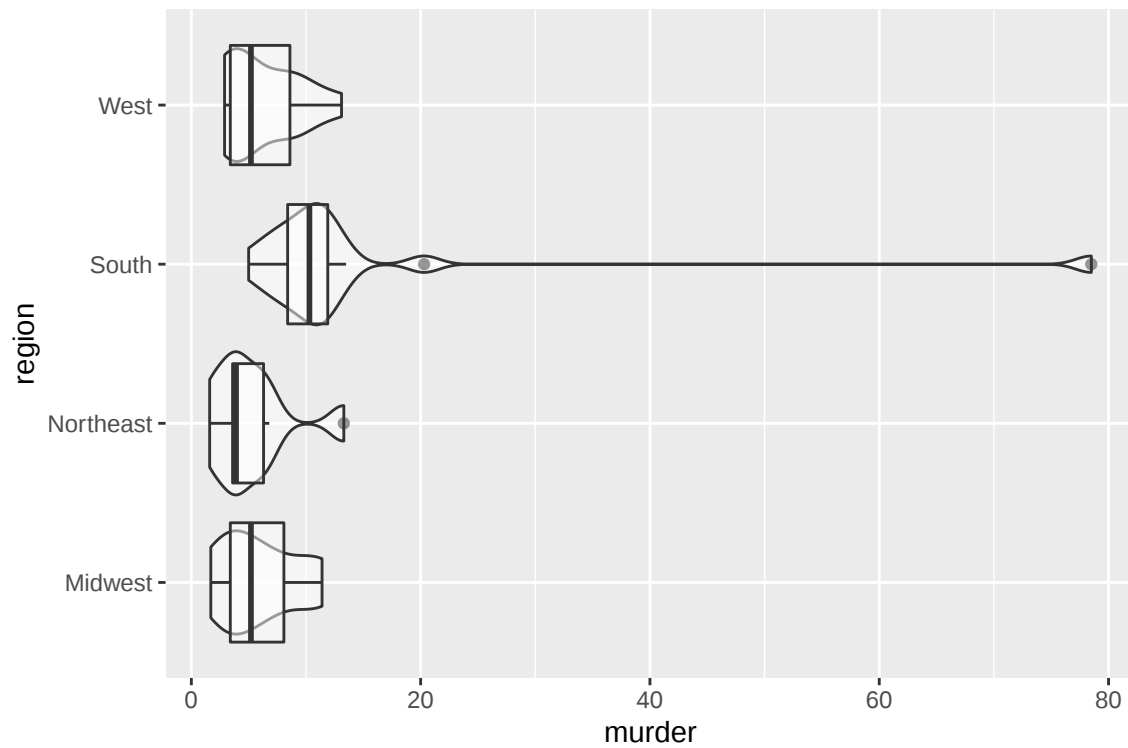
The table shows regions for each murder rate.

The density plot shows that the majority of the highest is in northeast.

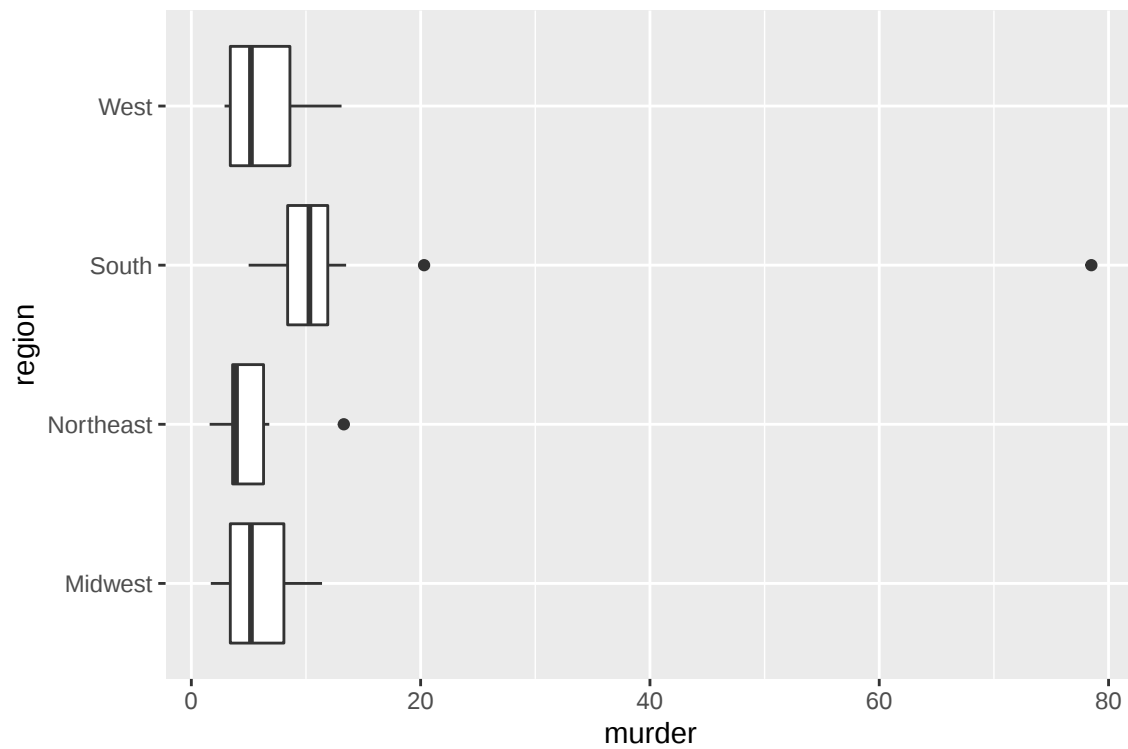
```
ggplot(CD, aes(x=region, FILL= murder)) + geom_bar(position = "dodge")
```



```
ggplot(CD, aes(x=murder, y=region, fill=murder))+geom_violin(alpha=0.5)+geom_boxplot(alpha=0.5)
```



```
ggplot(CD, aes(x=murder, y=region, fill=murder))+geom_boxplot()
```



First

graph- Shows that the South has the most states in the region. Second graph- Shows the south has the most murders and the state with the highest percentage. Third graph- Shows the same as the second graph but without violins. FOURTH QUESTION: Conclusion:

Through all the graphs I found that the Southern states have the highest rate of murders. I did not have a hypothesis, however I would have thought the West would have the highest crime rate because it is the most populated. The data did not support my thinking that the West would have the highest rate of murders.