Crime

Katie Raymond

2025-06-24

R Markdown

I am looking at the Crime Data for my final assignment. There are only 11 variables listed and I will be focusing on the two variables "crime" and "region". Looking at crime by region helps us see how crime rates vary depending on where people live. It can highlight areas that might need more support or safety resources. By comparing the regions, we can start to understand how local factors such as the economy or community might play a role in crime. This kind of insight can help with making decions about crime prevention and safety based on location.

```
library(readxl)
Crime_Data <-read_excel("C:/Users/kayra/OneDrive/Desktop/Math 130/Data/Crime_Data.xlsx")
```

```
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)
```

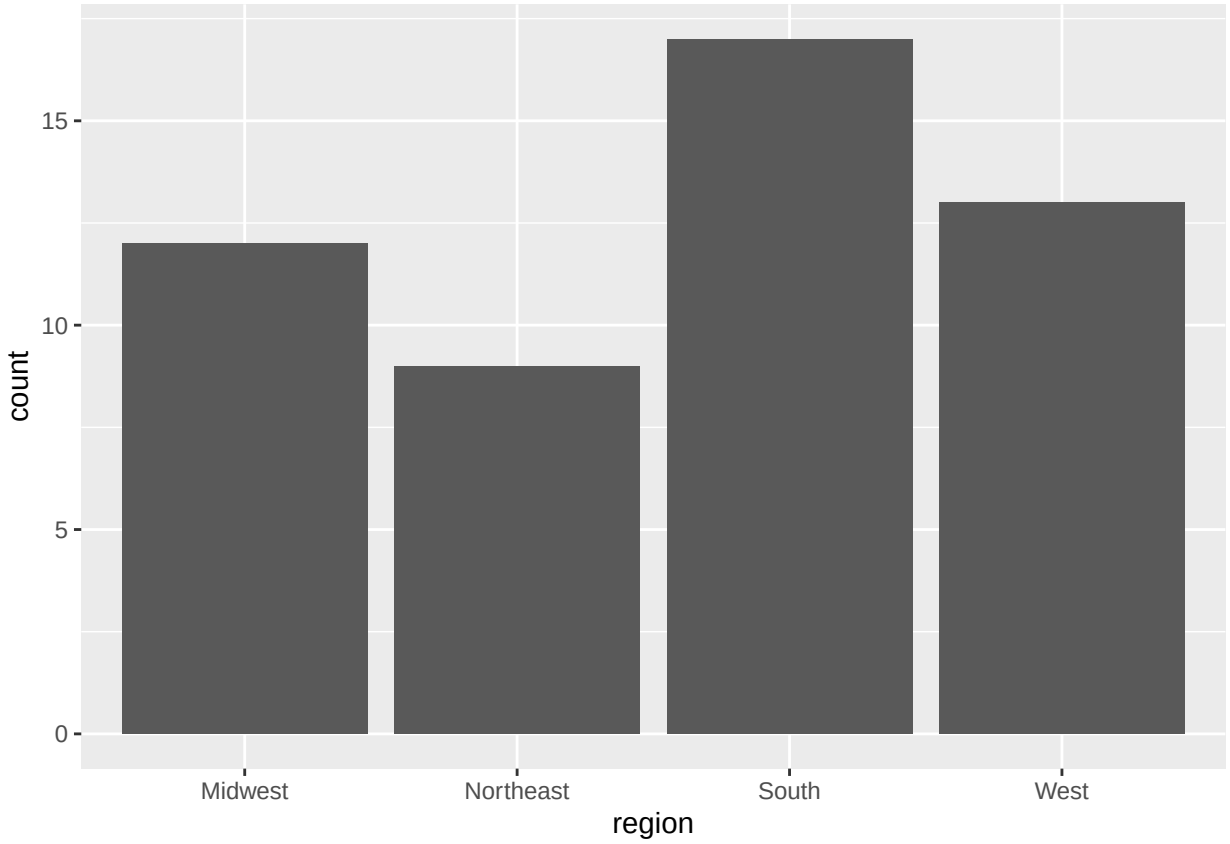
```
head(Crime_Data)
```

```
## # A tibble: 6 x 11
##     sid state region division       crime murder pctmetro pctwhite pcths poverty
##   <dbl> <chr> <chr>  <chr>          <dbl>  <dbl>    <dbl>    <dbl> <dbl>   <dbl>
## 1     1 ak    West   Pacific          761   9        41.8     75.2  86.6    9.10
## 2     2 al    South  East South Ce~   780  11.6      67.4     73.5  66.9   17.4
## 3     3 ar    South  West South Ce~   593  10.2      44.7     82.9  66.3   20
## 4     4 az    West   Mountain         715   8.60     84.7     88.6  78.7   15.4
## 5     5 ca    West   Pacific         1078  13.1      96.7     79.3  76.2   18.2
## 6     6 co    West   Mountain         567   5.80     81.8     92.5  84.4    9.90
## # i 1 more variable: single <dbl>
```

Univariate Analysis

Variable 1: Region

```
ggplot(Crime_Data, aes(region))+geom_bar()
```




```
table(Crime_Data$region)
```

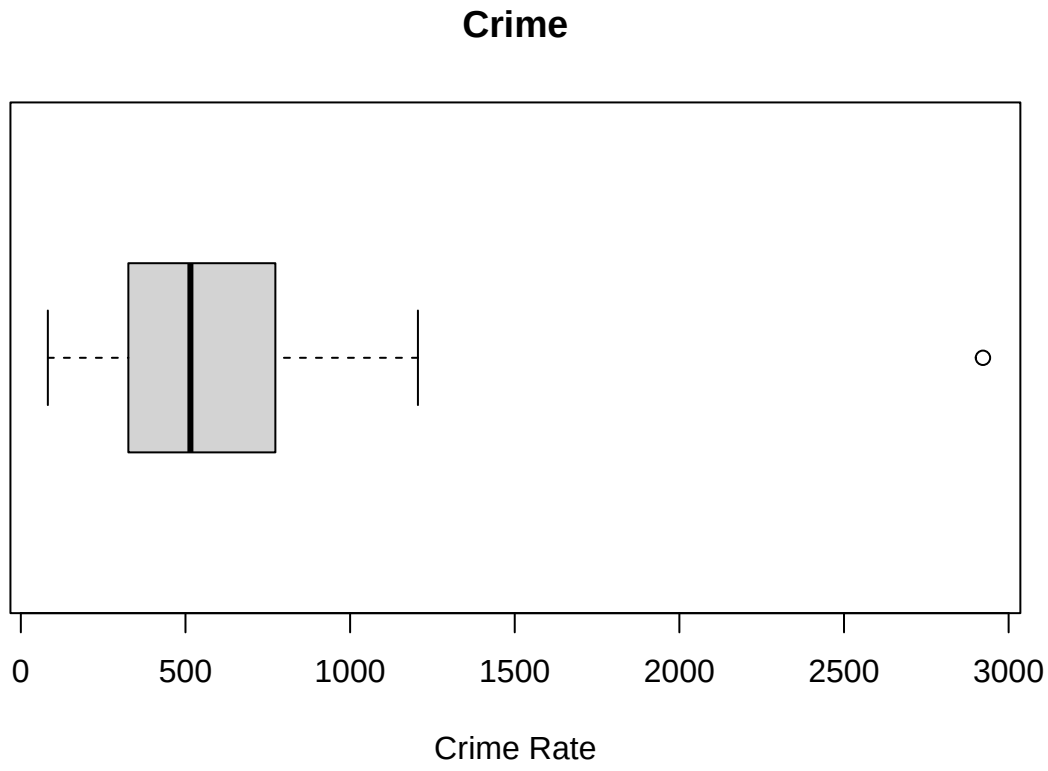
```
##
##   Midwest Northeast     South      West
##        12         9        17        13
```

Variable #2: Crime

```
summary(Crime_Data$crime)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    82.0   326.5   515.0   612.8   773.0  2922.0
```

```
boxplot(Crime_Data$crime, horizontal = TRUE, main="Crime", xlab="Crime Rate")
```



Bivariate Analysis:

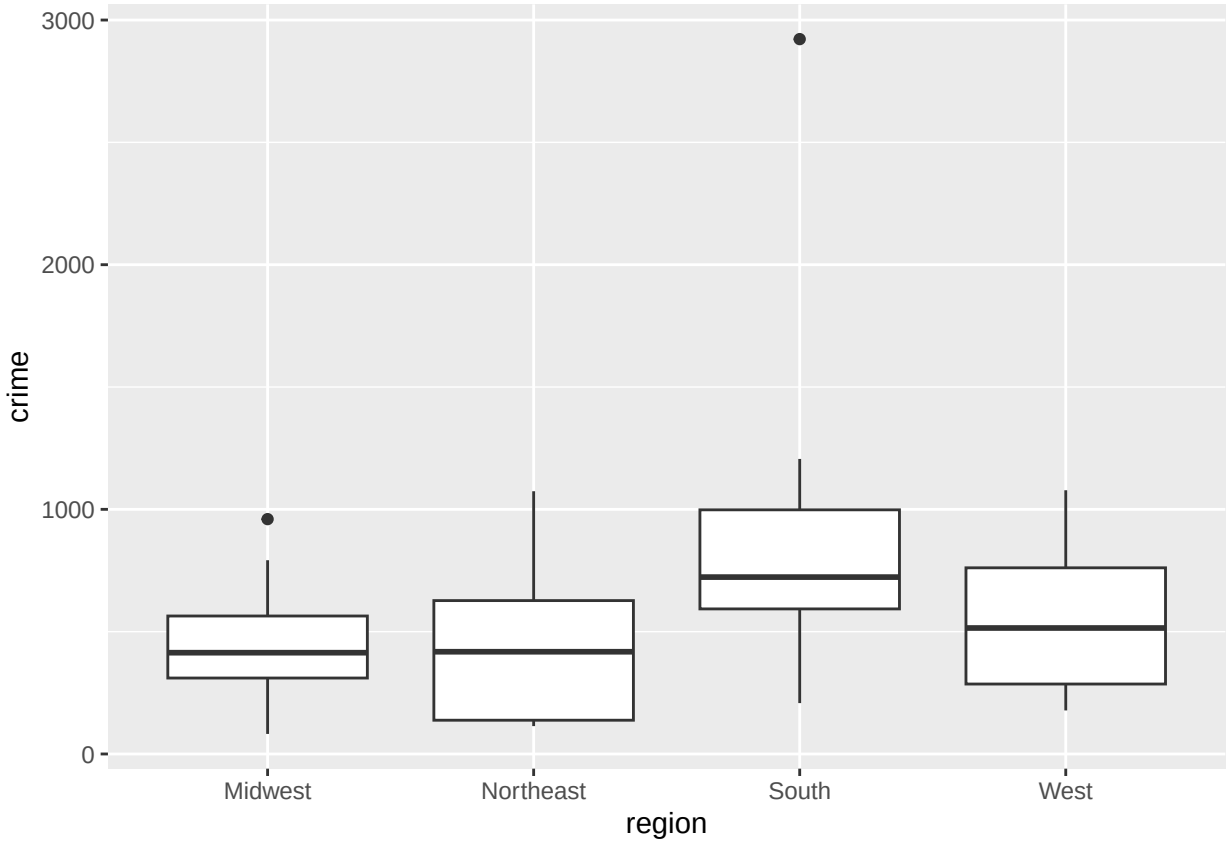
How much Crime by Region?

```
library(dplyr)

Crime_Data %>%
  group_by(region) %>%
  summarise(
    count = n(),
    mean = mean(crime, na.rm = TRUE),
    sd = sd(crime, na.rm = TRUE),
    min = min(crime, na.rm = TRUE),
    q1 = quantile(crime, 0.25, na.rm = TRUE),
    median = median(crime, na.rm = TRUE),
    q3 = quantile(crime, 0.75, na.rm = TRUE),
    max = max(crime, na.rm = TRUE)
  )
```

```
## # A tibble: 4 x 9
##   region    count  mean    sd   min    q1 median    q3   max
##   <chr>     <int> <dbl> <dbl> <dbl> <dbl>  <dbl> <dbl> <dbl>
## 1 Midwest      12  461.  259.    82  310.    414   564   960
## 2 Northeast     9  462.  329.   114  138     418   627  1074
## 3 South        17  842.  596.   208  593     723   998  2922
## 4 West         13  558.  293.   178  286     515   761  1078
```

```
ggplot(Crime_Data, aes(x=region, y=crime)) + geom_boxplot()
```



The South reports the highest average crime rate of 841.88 and median of 723, indicating a significantly greater crime to other regions. It also has the widest spread of crime rates, with a high standard deviation of 595.99 and a max value of 2922 which is a rather exteme outlier. This outlier is likely skewing the mean to be higher than any other region and affecting the spread greatly. If the outlier were to be removed, the northeast would have the largest spread. The midwest and northeast have much lower crime averages (460.92 and 462.22 respectively). The midwest also presents the smallest spread. The west has a bit of a larger average than the midwest and northeast with a mean of 557.85

Conclusion: It does appear that there is a connection between where a region is located and how high its crime rates tend to be. The South stands out with noticeably higher crime rates compared to other regions, which suggests that location could play a role. While we'd need more analysis to be sure, the patterns we see make it likely that region and crime rate are related.