Assignment3 Memo

Dr. Niladri Chakraborty

2025-04-22

1. Description of the dataset

This data set contains statistics, in arrests per 100,000 residents for assault, murder, and rape in each of the 50 US states in 1973. Also given is the percent of the population living in urban areas.

2. Preview of the USArrests dataset

```
library(tidyverse)

data = USArrests
head(data)

##            Murder Assault UrbanPop Rape
## Alabama      13.2     236       58 21.2
## Alaska       10.0     263       48 44.5
## Arizona       8.1     294       80 31.0
## Arkansas      8.8     190       50 19.5
## California    9.0     276       91 40.6
## Colorado      7.9     204       78 38.7

str(data)

## 'data.frame':    50 obs. of  4 variables:
##  $ Murder  : num  13.2 10 8.1 8.8 9 7.9 3.3 5.9 15.4 17.4 ...
##  $ Assault : int  236 263 294 190 276 204 110 238 335 211 ...
##  $ UrbanPop: int  58 48 80 50 91 78 77 72 80 60 ...
##  $ Rape    : num  21.2 44.5 31 19.5 40.6 38.7 11.1 15.8 31.9 25.8 ...
```
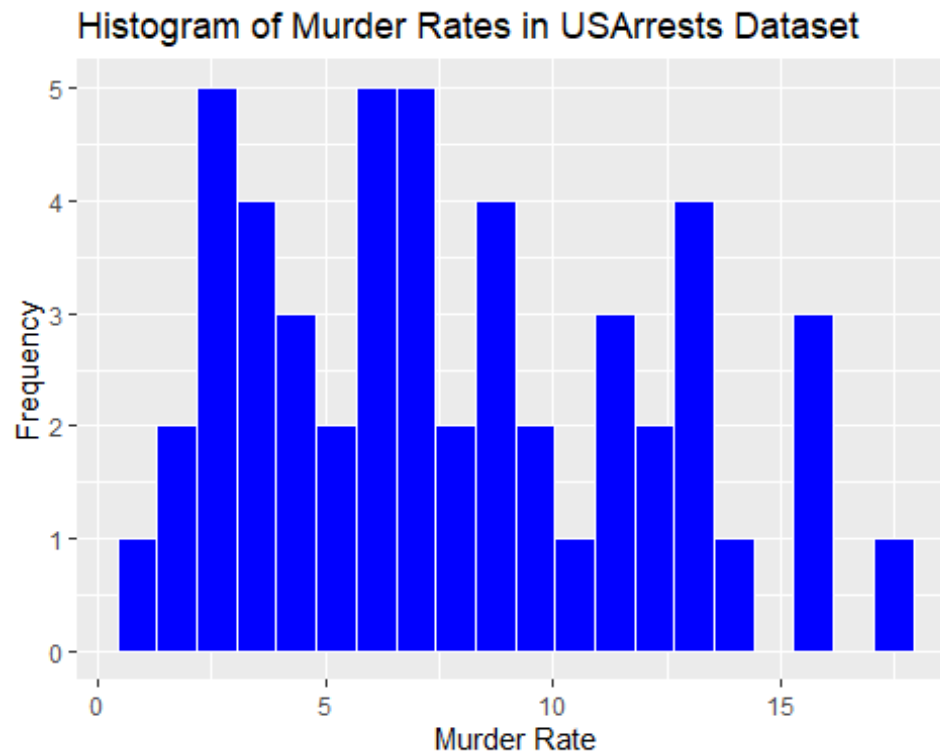
3. A Histogram of the Murder rate in USArrests dataset

```
ggplot(data,aes(x = data$Murder))+geom_histogram(fill = "blue",
color="white", bins = 20)+labs(x="Murder Rate", y =
"Frequency")+ggtitle("Histogram of Murder Rates in USArrests Dataset")
```

## Histogram of Murder Rates in USArrests Dataset

4. A scatterplot of Murder vs. Assault

```
ggplot(data, aes(x = data$Murder, y = data$Assault))+geom_point()+
  labs(x="Murder Rate", y = "Assault Rate")+
  ggtitle("Scatterplot of Murder and Assault Rates in USArrests Dataset")
```

Scatterplot of Murder and Assault Rates in USArrests
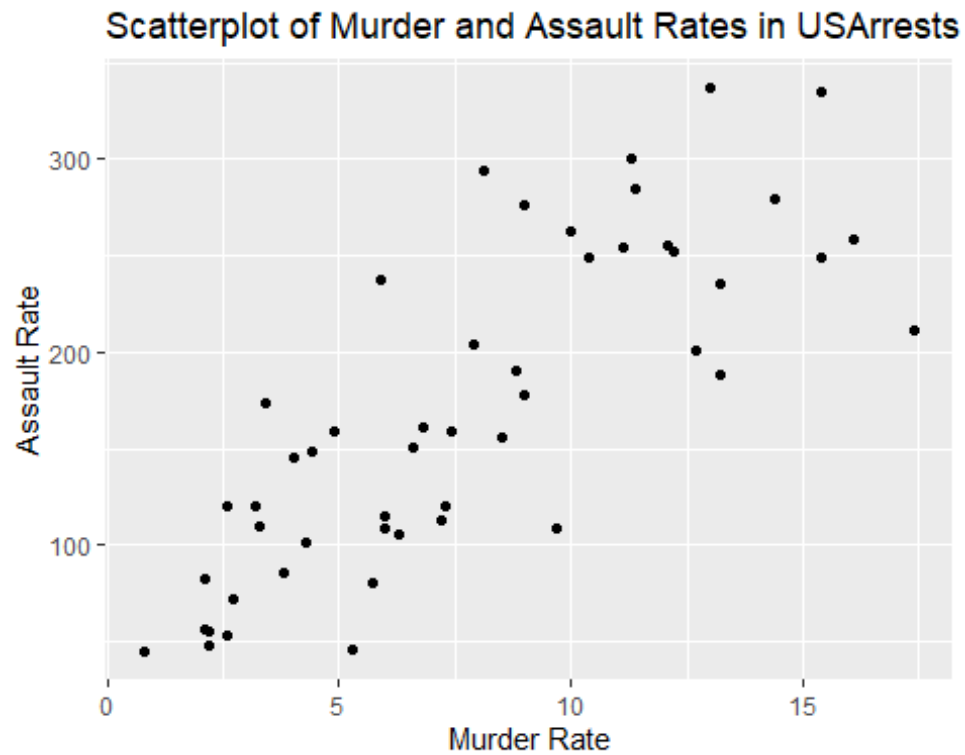
5. A line chart of Rape over Time

```
ggplot(data, aes(x = c(1:nrow(data)),y = data$Rape))+geom_line(color =
"blue")
```
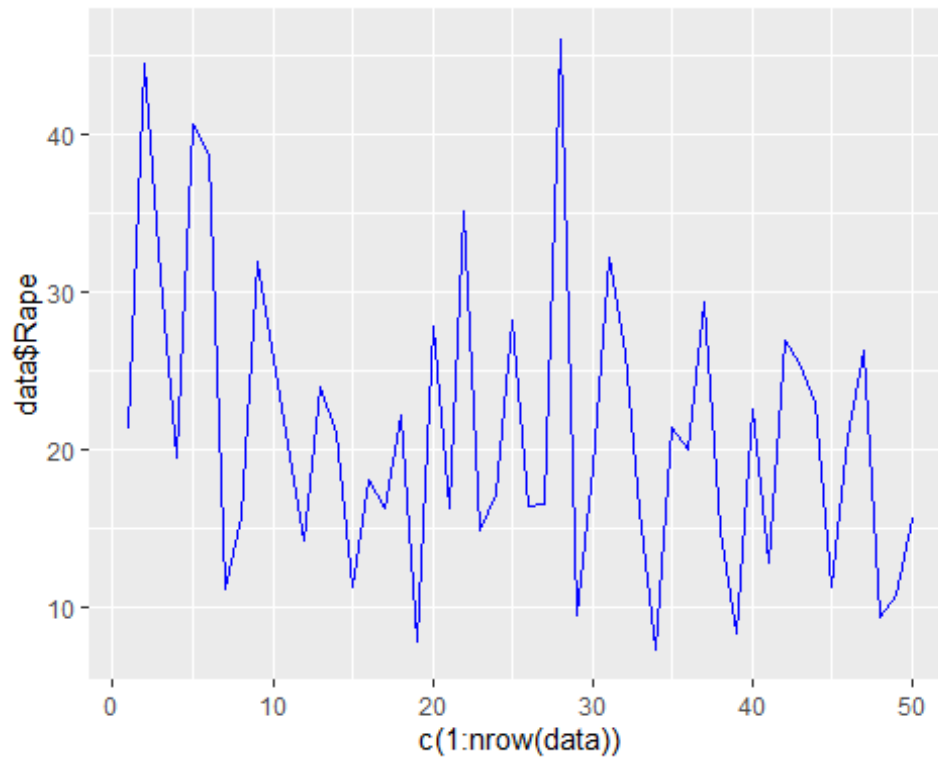
6. A correlation between urban population and the rate of rape

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.3.1
```
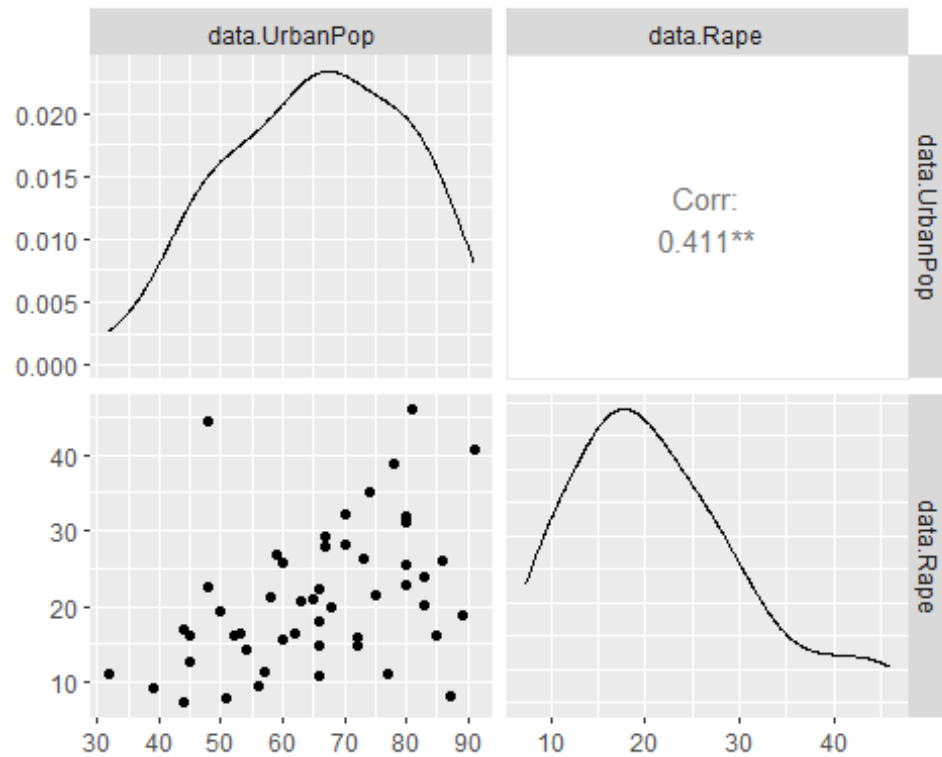
```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```r
data1 = data.frame(data$UrbanPop, data$Rape)
ggpairs(data1)
```

The correlation plot clearly shows a positive correlation between the urban population size and the rate of rape. The correlation coefficient is 0.411 and it is significant. The density plot shows that the rate of rape has positively skewed distribution. This indicates that the rate of rape is mostly concentrated towards the lower values. On the other hand, urban population has negatively skewed distribution. This shows that older population of above 50 is more widespread in all the cities in the dataset.