USArrests

Chriss Jordan Oboa

2022-11-01

Loading Library

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

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
```

```
## v tibble  3.1.8     v purrr   0.3.5
## v tidyr   1.2.1     v stringr 1.4.1
## v readr   2.1.3     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggdist)
library(ggridges)
```

```
##
## Attaching package: 'ggridges'
##
## The following objects are masked from 'package:ggdist':
##
##     scale_point_color_continuous, scale_point_color_discrete,
##     scale_point_colour_continuous, scale_point_colour_discrete,
##     scale_point_fill_continuous, scale_point_fill_discrete,
##     scale_point_size_continuous
```

```
library(car)
```

```
## Loading required package: carData
##
## Attaching package: 'car'
##
## The following object is masked from 'package:purrr':
```

```
##
##     some
##
## The following object is masked from 'package:dplyr':
##
##     recode
```

Loading the dataset

```
data("USArrests")
head(USArrests)
```

```
##            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
```

1) Provide a short narrative description of the dataset. The USArreest dataset is a 1973 national United States dataset which contains statistic on arrests per 100,000 residents for as Murder, Assault, Urban Pop and Rape.

2) Find the mean, median, mode, range (min, max) for each crime.

```
summary(USArrests)
```

```
##      Murder          Assault         UrbanPop          Rape
##  Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
##  1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
##  Median : 7.250   Median :159.0   Median :66.00   Median :20.10
##  Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
##  3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
##  Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
```

Mode of Muder crimes:

```
mode = function(){
    return(sort(-table(USArrests$Murder))[1])
}
mode()
```

```
## 2.1
##  -2
```

Mode of Assault crimes:

```
mode = function(){
    return(sort(-table(USArrests$Assault))[1])
}
mode()
```

```
## 120
##  -3
```

Mode of Urban Pop crimes:

```
mode = function(){
    return(sort(-table(USArrests$UrbanPop))[1])
```

```
}
mode()
```

```
## 66
## -4
```

Mode of Rape crimes:

```
mode = function(){
    return(sort(-table(USArrests$Rape))[1])
}
mode()
```

```
## 14.9
##   -2
```

3) Answer the following questions by performing different computations. Useheadings or sentences to describe what you are doing:

i) Which state is the safest for each one of the crime categories?

a) the safest based on Murder

```
#the safest based on Murder
USArrests [which(USArrests$Murder == min(USArrests$Murder)),]
```

```
##              Murder Assault UrbanPop Rape
## North Dakota    0.8      45       44  7.3
```

b) the safest based on Assault

```
#the safest based on Assault
USArrests [which(USArrests$Assault == min(USArrests$Assault)),]
```

```
##              Murder Assault UrbanPop Rape
## North Dakota    0.8      45       44  7.3
```

c) the safest based on UrbanPop

```
#the safest based on UrbanPop
USArrests [which(USArrests$UrbanPop == min(USArrests$UrbanPop)),]
```

```
##         Murder Assault UrbanPop Rape
## Vermont    2.2      48       32 11.2
```

d) the safest based on Rape

```
#the safest based on Rape
USArrests [which(USArrests$Rape == min(USArrests$Rape)),]
```

```
##              Murder Assault UrbanPop Rape
## North Dakota    0.8      45       44  7.3
```

ii) Which state is the least safe for each one of the crime categories?

a) the least safe based on Murder

```
#the least safe based on Murder
USArrests [which(USArrests$Murder == max(USArrests$Murder)),]
```

```
##         Murder Assault UrbanPop Rape
## Georgia   17.4     211       60 25.8
```

b) the least safe based on Asault

```
#the least safe based on Assault
USArrests [which(USArrests$Assault == max(USArrests$Assault)),]
```

```
##                Murder Assault UrbanPop Rape
## North Carolina     13     337       45 16.1
```

c) the least safe based on UrbanPop

```
#the least safe based on UrbanPop
USArrests [which(USArrests$UrbanPop == max(USArrests$UrbanPop)),]
```

```
##            Murder Assault UrbanPop Rape
## California      9     276       91 40.6
```

d) the least safe based on Rape

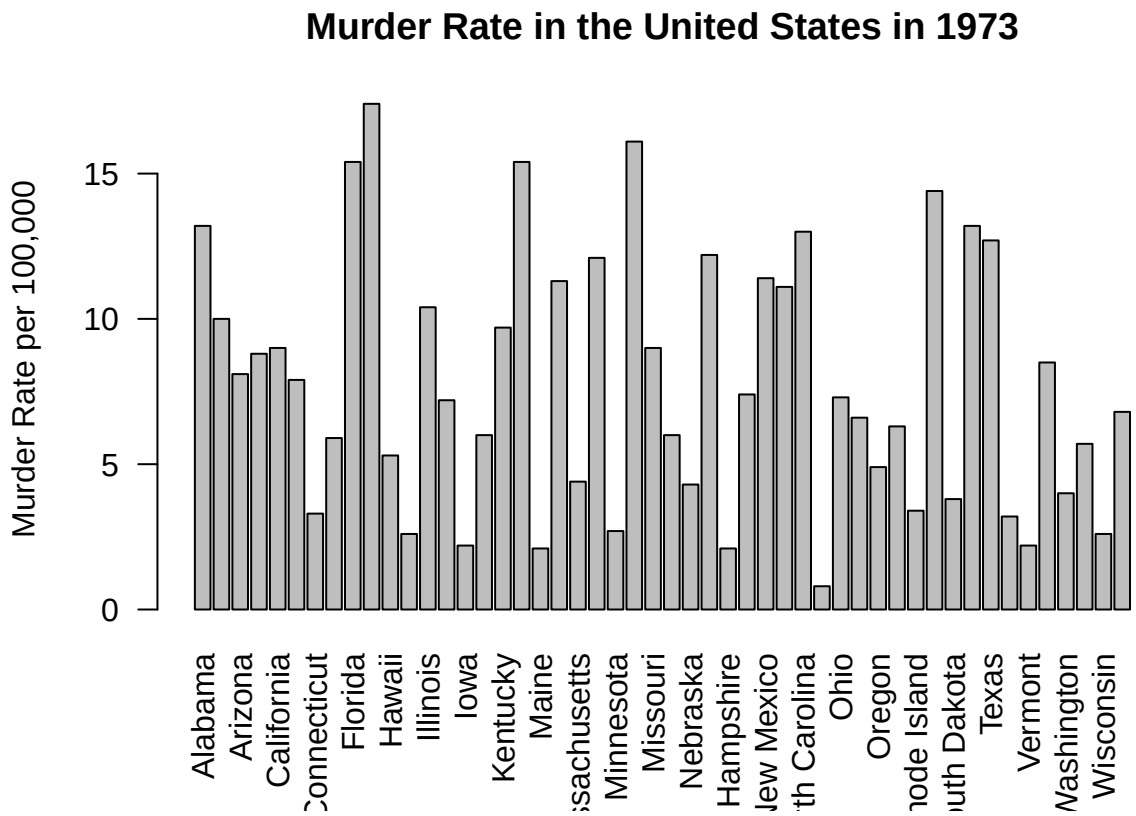
```
#the least safe based on Rape
USArrests [which(USArrests$Rape == max(USArrests$Rape)),]
```

```
##        Murder Assault UrbanPop Rape
## Nevada   12.2     252       81   46
```

iii) Provide one bar chart that provides crime information for DC, MD and VA.

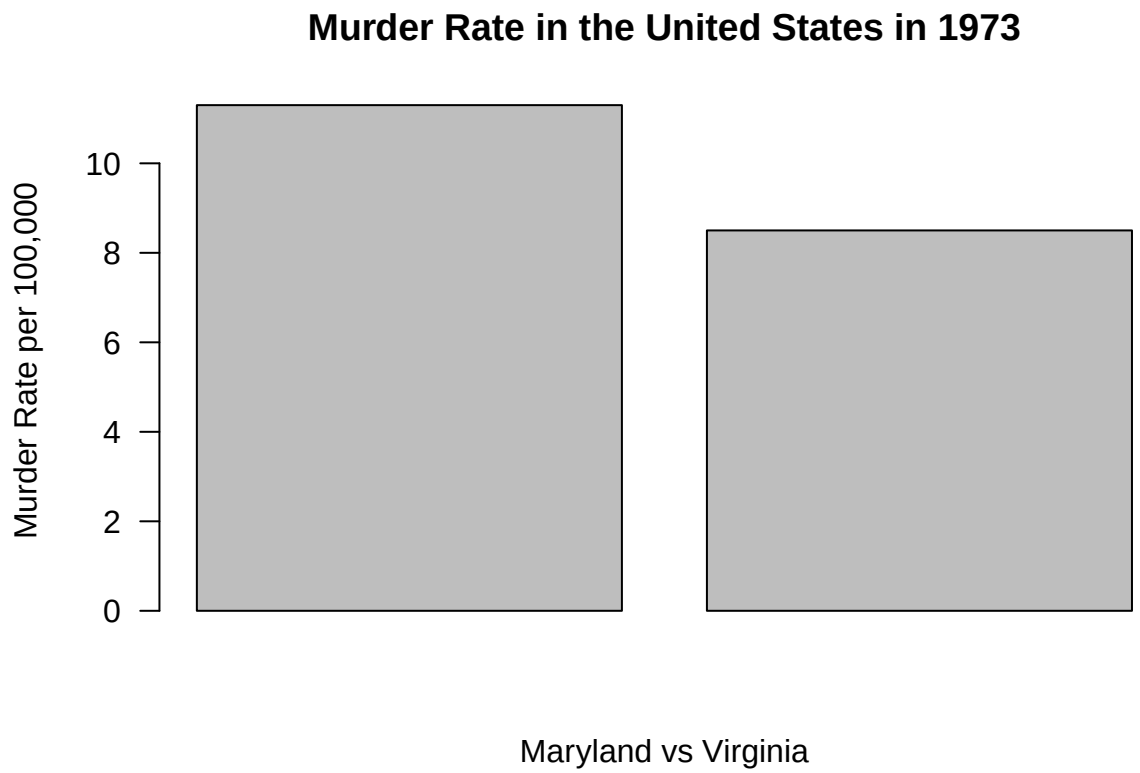
Describe what the chart is presenting. It describes MD and VA Murder rate per 100,000 in the United States in 1973. However, DC data are not found in the dataset.

```
state.names = row.names(USArrests)
barplot(USArrests$Murder, names.arg = state.names, las = 2, ylab = "Murder Rate per 100,000",
    main = "Murder Rate in the United States in 1973")
```




```
#DC, MD and VA Murder rate per 100,000
state.names = row.names(USArrests[c("Maryland","Virginia"),1])
```

```
barplot(USArrests[c("Maryland","Virginia"),1], names.arg = state.names, las = 2, ylab = "Murder Rate pe
    main = "Murder Rate in the United States in 1973")
```



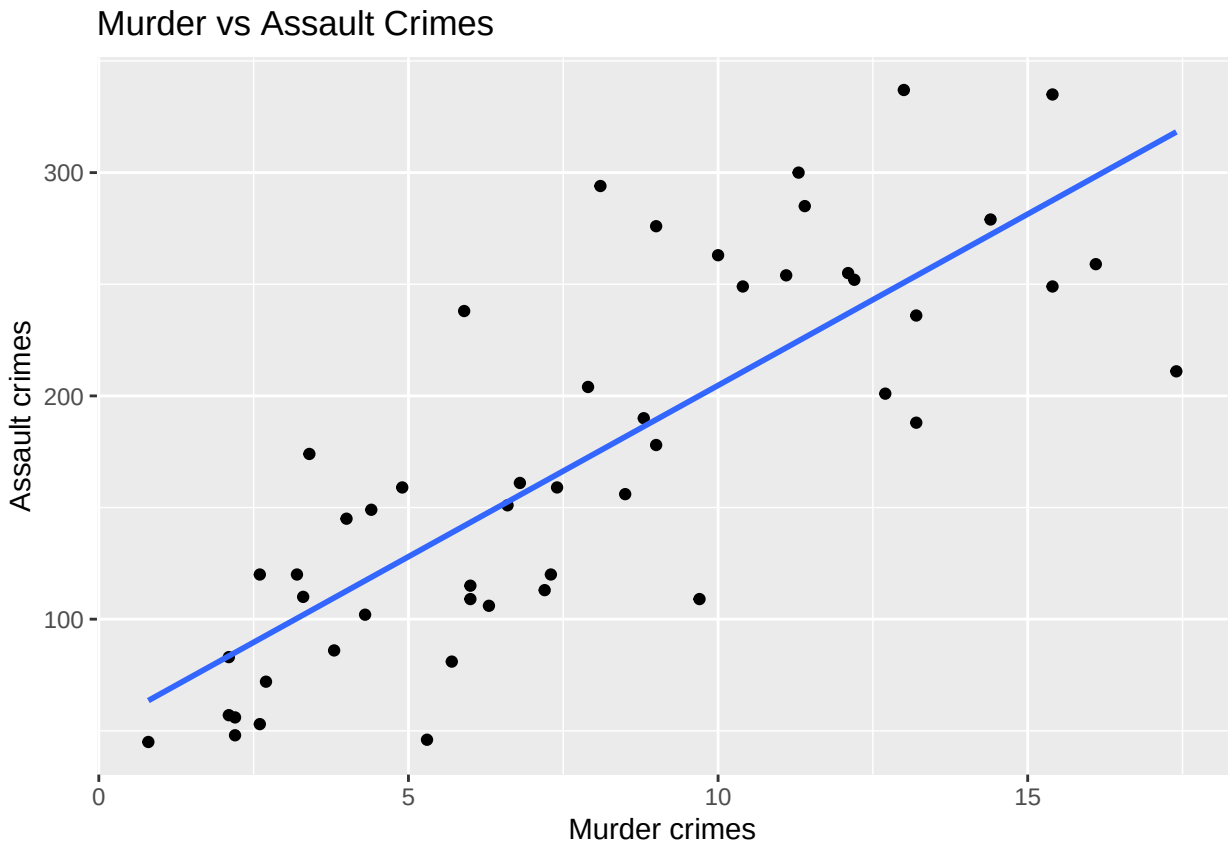
iv) Ask a question that may be answered by plotting the data. Is there any positive relationship or corelation between Murder and Assault?

Provide a chart of your choice (but not a bar chart) that would help you answer the question and what the answer is.

```
ggplot(data = USArrests, mapping = aes(x = Murder, y= Assault))+
  geom_point (binwidth = 0.01)+ geom_smooth(method = "lm", se=FALSE) +
  labs(x = "Murder crimes", y= "Assault crimes",
    title = "Murder vs Assault Crimes")
```

```
## Warning: Ignoring unknown parameters: binwidth
```

```
## `geom_smooth()` using formula 'y ~ x'
```


Murder vs Assault Crimes
Assault crimes
300
200
100
0
5
10
15
Murder crimes

End