

Name - Mansi Tiwari

Course - MCA

Section - A

Student id - 21712002

Roll no - 2161114

Q. ③ Analyze any csv dataset using R

library(dplyr)

library(ggplot2)

setwd("M:/along 18p")

getwd()

data <- read.csv("mud1.csv")

view(data)

head(data)

tail(data)

tail(data, 10)

str(data)

summary(data)

data\$state.Length

sum(is.na(data))

ggplot(data, aes(y = state, x = murders)) + geom\_bar(stat = "identity")

ggplot(data, aes(y = state, x = gunmurders)) + geom\_bar(stat = "identity")

ggplot(data, aes(y = Population Density, x = murders, group = Population Density, color = murders)) + geom\_line() + geom\_point()

ggplot(data, aes(x = murders, y = region)) + geom\_boxplot()

data\_size <- factor(data)

str(data\_size)

summary(data\_size)

level(data\_size)

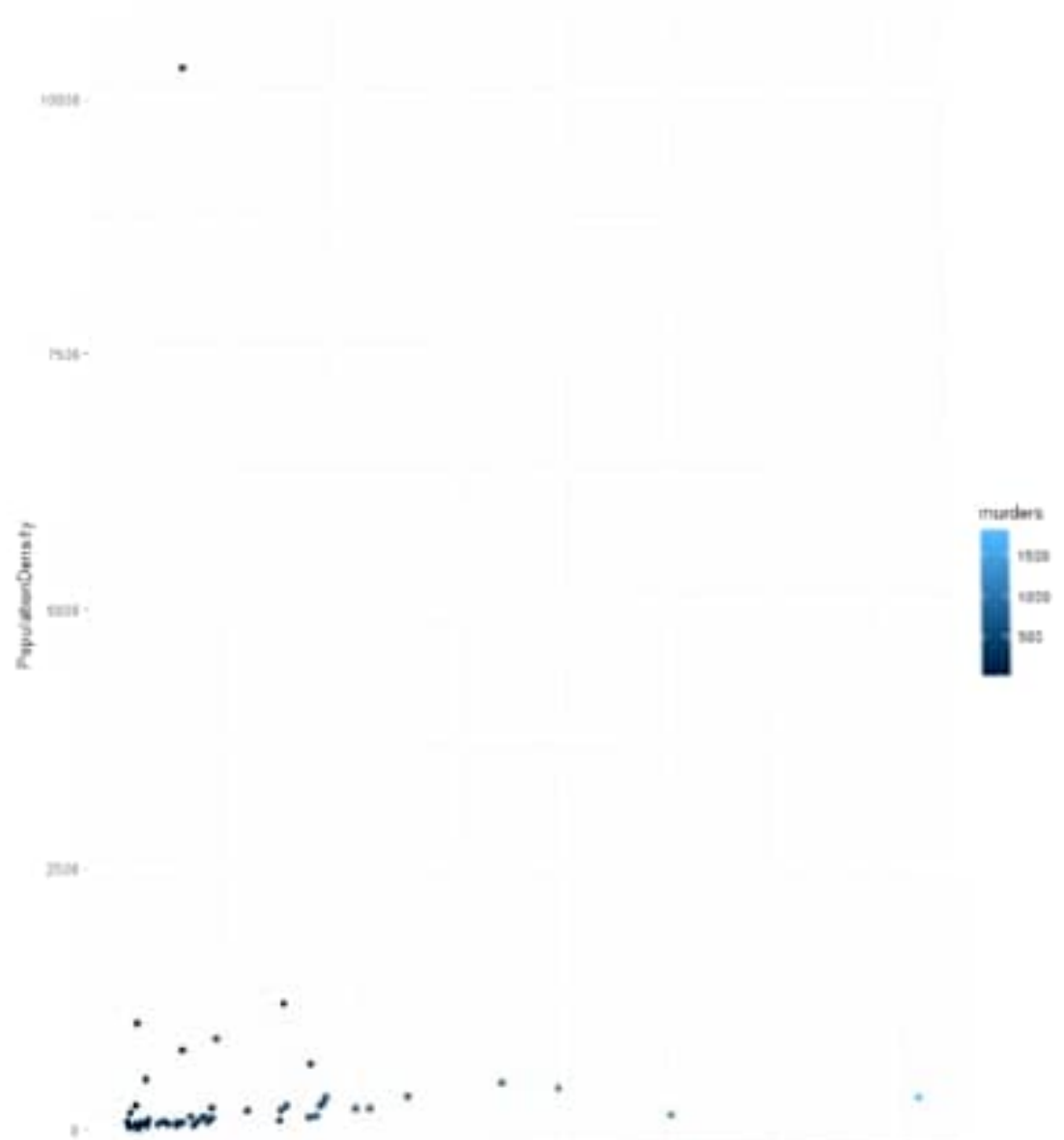
data\_table <- (data\_size)

pie(data\_table)

ggplot(data, aes(y = murders, fill = region, x = region)) + geom\_bar(width = 1, stat = "identity") + coord\_polar("x", start = 0)

```
#scatterplot
```

```
ggplot(data, aes(y = PopulationDensity, x  
=murders, group=PopulationDensity,  
colour=murders)) +geom_line() + geom_point()
```

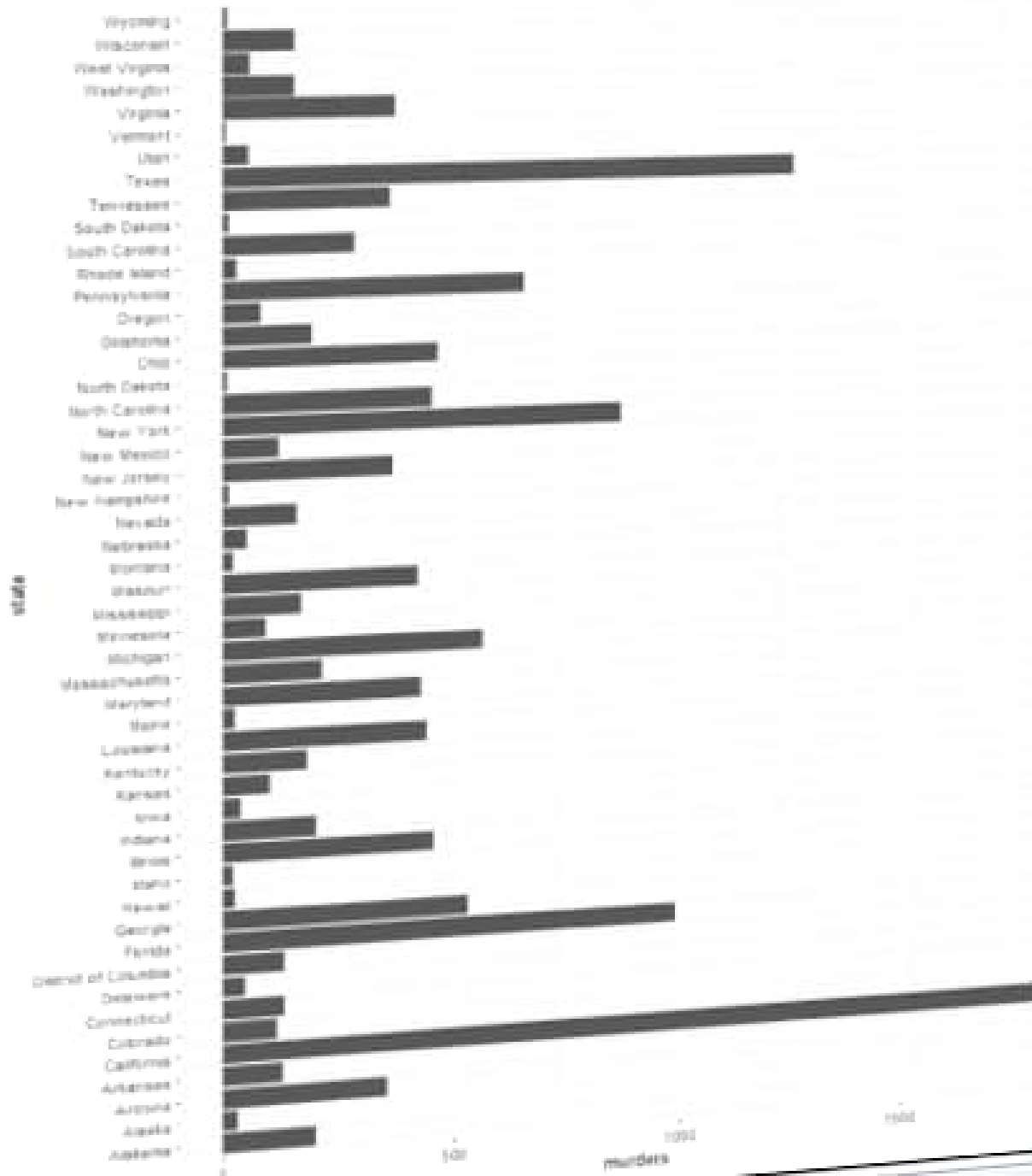


## #piechart

```
ggplot(data,aes(y=murders,fill=region,x=region)) +  
geom_bar(width=1,stat="identity")+coord_polar("x",s  
tart=0)
```



```
ggplot(data, aes(y=state, x=murders))+  
geom_bar(stat = "identity")
```





Q43 Discuss Descriptive and inferential statistics of above dataset.

Descriptive statistics -

Descriptive statistics are a part of statistics that can be used to describe data. It is used to summarize the attributes of a sample in such a way that a pattern can be drawn from the group. It enables researchers to present data in a more meaningful way such that easy interpretation can be made.

Population -

Min  $\rightarrow$  563626

Mean  $\rightarrow$  4339367.6075769

Median  $\rightarrow$  4339367

Population density -

Min  $\rightarrow$  1.264

Mean  $\rightarrow$  394.549

Median  $\rightarrow$  102.600

Max.  $\rightarrow$  10298.000

1<sup>st</sup> Qu  $\rightarrow$  46.185

3<sup>rd</sup> Qu  $\rightarrow$  224.350

Murders -

Min  $\rightarrow$  7.0

1<sup>st</sup> Qu  $\rightarrow$  49.0

Median  $\rightarrow$  151.0

Mean  $\rightarrow$  273.2

3<sup>rd</sup> Qu  $\rightarrow$  394.0

Max.  $\rightarrow$  1811.0

Gunmurders  $\rightarrow$

Min  $\rightarrow$  2.0

1<sup>st</sup> Qu  $\rightarrow$  94.5

Median  $\rightarrow$  97.0

Mean  $\rightarrow$  184.4

3<sup>rd</sup> Qu  $\rightarrow$  268.0

Max.  $\rightarrow$  1257.0



gunownership -

Min - 0.0360

1<sup>st</sup> Qu - 0.3055

Median - 0.3580

Mean - 0.3695

3<sup>rd</sup> Qu - 0.4400

Max - 0.5970

### Inferential Statistics

Inferential statistics is a branch of statistics that is used to make inferences about the population by analyzing a sample. When the population data is very large it becomes difficult to use it. In such cases, certain samples are taken that are representative of the entire population. Inferential statistics draws conclusions regarding the population using these samples. Sampling strategies such as simple random sampling, cluster sampling, stratified sampling, and systematic sampling need to be used in order to choose correct samples from the population. Some methodologies used in order to choose correct samples from the population. Some methodologies used in inferential statistics are as follows:-

(I) Hypothesis Testing (z test, t test)

(II) Regression Analysis. (Check relationship b/w dependent & independent variable)