Name :- Badal Chauhan   Student Id :- 21711043

Sub :- Data Analysis with R

Univ. Roll No :- 2101056

3.
```r
library (dplyr)

setwd ("D:/rp/Pop/Data")

mydata <- read.csv ("D:/rp/Pop/pop.csv")

View (mydata)

str (mydata)

plot (mydata)

barplot (mydata $ GrowthRate, col = 'purple')

pie (mydata $ Year)

min (mydata $ Year)

max (mydata $ Year)

mean (mydata $ Year)

median (mydata $ Year)

sd (mydata $ GrowthRate)

var (mydata $ GrowthRate)

hist (mydata $ Year, col = 'blue')

boxplot (mydata $ year, col = 'red')
```

Badal
Chauhan

Name :- Badal Chauhan   Student Id :- 21711043

Sub :- Data Analysis with R

Univ. Roll No :- 2101050

```
dotchart (mydata $Year, color = 'black')

quantile (mydata $GrowthRate, 0.25)

quantile (mydata $GrowthRate, 0.75)

summary (mydata)
```

Badal
Chauhan

Name :- Badal Chauhan   Student Id :- 21711043

Sub :- Data Analysis with R

Univ. Roll No. :- 2101050

4. Discriptive Statistical Analysis :-

i). 75% of population have Growth - Rate more than 1.2025

ii). Average Growth - Rate is 1.129167

iii). Median of Growth - Rate is 1.11.

Inferential Statistical Analysis :-

Year	Population	Growth - Rate
Min. : 2010	Min : 1.234e +09	Min : 0.970
1st Qu : 2013	1st Qu : 1.277e +09	1st Qu : 1.035
Median : 2016	Median : 1.317e +09	Median : 1.116
Mean : 2016	Mean : 1.316e +09	Mean : 1.129
3rd Qu : 2018	3rd Qu : 1.356e +09	3rd Qu : 1.202
Max : 2021	Max : 1.393e +09	Max : 1.360

Badal
Chauhan

R Script:-

```
## loading the dplyr library
library(dplyr)


## setting working directory
setwd("D:/rp/Pop/Data")


## reading the dataset (.csv)
mydata<-read.csv("D:/rp/Pop/pop.csv")


## viewing dataset
View(mydata)


## displaying internal structure
str(mydata)


## drawing points (markers)
plot(mydata)


## representing data in rectangular bars with length of the bar proportional to the value of variable
barplot(mydata$GrowthRate, col = 'purple')


## representing data as slices of a circle with different colors
pie(mydata$Year)


## finding minimum element present in the dataset
min(mydata$Year)


## finding maximum element present in the dataset
max(mydata$Year)
```

```r
## calculating arithmetic mean of the dataset

mean(mydata$Year)


## calculating median (middle most value) in the dataset

median(mydata$Year)


## calculating standard diviation

sd(mydata$GrowthRate)


## calculating variance

var(mydata$GrowthRate)


## representing the frequencies of values of variables

hist(mydata$Year, col = 'blue')


## representing that how well distributed is the data in the dataset

boxplot(mydata$Year, col = 'red')


## representing specified data in the dot form

dotchart(mydata$Year, color = 'black')


## creating sample quantiles within a dataset with probability [0, 1]

quantile(mydata$GrowthRate, 0.25)


quantile(mydata$GrowthRate, 0.75)


## summary of the dataset

summary(mydata)
```
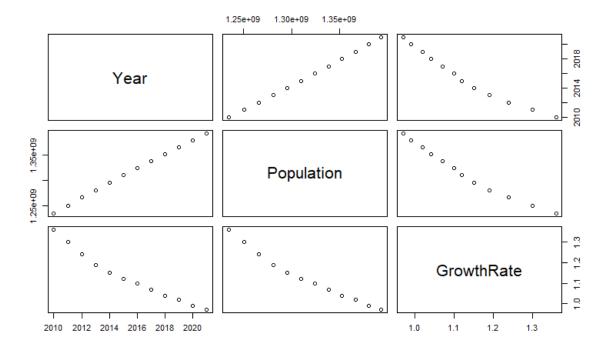
R Outputs:-

> ## loading the dplyr library

> 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

Warning message:

package 'dplyr' was built under R version 4.1.2

>

> ## setting working directory

> setwd("D:/rp/Pop/Data")

>

> ## reading the dataset (.csv)

> mydata<-read.csv("D:/rp/Pop/pop.csv")

>

> ## viewing dataset

> View(mydata)

>

> ## displaying internal structure

> str(mydata)

'data.frame':   12 obs. of 3 variables:

 $ Year   : int 2021 2020 2019 2018 2017 2016 2015 2014 2013 2012 ...
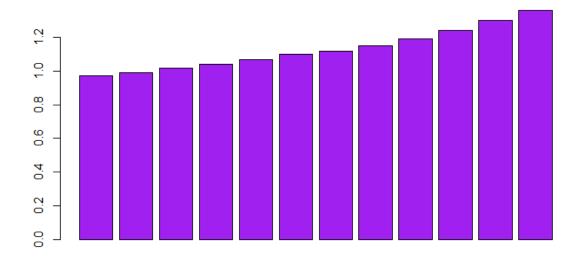
$ Population: int 1393409038 1380004385 1366417754 1352642280 1338676785 1324517249 1310152403 1295600772 1280842125 1265780247 ...

 $ GrowthRate: num 0.97 0.99 1.02 1.04 1.07 1.1 1.12 1.15 1.19 1.24 ...

>

> ## drawing points (markers)

> plot(mydata)

>

> ## representing data in rectangular bars with length of the bar proportional to the value of variable

> barplot(mydata$GrowthRate, col = 'purple')

>

> ## representing data as slices of a circle with different colors

> pie(mydata$Year)

>

> ## finding minimum element present in the dataset

> min(mydata$Year)

[1] 2010

>

> ## finding maximum element present in the dataset

> max(mydata$Year)

[1] 2021

>

> ## calculating arithmetic mean of the dataset

> mean(mydata$Year)

[1] 2015.5

>

> ## calculating median (middle most value) in the dataset

> median(mydata$Year)

[1] 2015.5

>

> ## calculating standard diviation
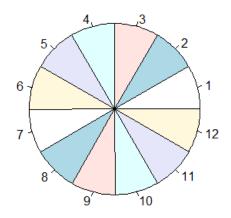
> sd(mydata$GrowthRate)

```
[1] 0.123543

>

> ## calculating variance

> var(mydata$GrowthRate)

[1] 0.01526288

>

> ## representing the frequencies of values of variables

> hist(mydata$Year, col = 'blue')

>

> ## representing that how well distributed is the data in the dataset

> boxplot(mydata$Year, col = 'red')

>

> ## representing specified data in the dot form

> dotchart(mydata$Year, color = 'black')

>

> ## creating sample quantiles within a dataset with probability [0, 1]

> quantile(mydata$GrowthRate, 0.25)

  25%

1.035

>

> quantile(mydata$GrowthRate, 0.75)

  75%

1.2025

>

> ## summary of the dataset

> summary(mydata)

     Year        Population         GrowthRate

 Min.   :2010   Min.   :1.234e+09   Min.   :0.970

 1st Qu.:2013   1st Qu.:1.277e+09   1st Qu.:1.035

 Median :2016   Median :1.317e+09   Median :1.110

 Mean   :2016   Mean   :1.316e+09   Mean   :1.129
```

3rd Qu.:2018   3rd Qu.:1.356e+09   3rd Qu.:1.202

Max.   :2021   Max.   :1.393e+09   Max.   :1.360

Histogram of mydata$Year