

MOHAMMAD SAIM ASLAM:19I-0461

HUSNAIN ZAHID : 19I-2193

SEMESTER PROJECT

SUBMITTED TO : SIR WAQAS MUNIR

SUBMITTED ON :12/06/2021

Questionnaire:

<https://forms.gle/FjBjFeFMF6sowiUm6>

part1:

i will attach the file from which it will read the data and I have read the the data from point1 file

```
data=read.csv(file.choose())
data
```

Confirmed. Cases	deaths	recovered. Cases	active. Case	total. Tests
939,931	21,633	875,581	42,717	13,781,668
Critical. Situation				
3,557				

Links:

https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data/Pakistan_medical_cases
<https://covid.gov.pk/>

PART 2:

##punjab data and I have read the data from Punjab_data file

```
data=read.csv(file.choose())
ex_date=as.Date(data$date)
```

```
week_date=cut(ex_date,"week")
new_wd= gsub(","," ", data$cases)
ew=as.numeric(new_wd)
d=aggregate (ew~week_date, data, sum)
summary (d)
```

```
Min. : 1
1st Qu.: 534640
Median : 710014
Mean : 812952
3rd Qu.:1109277
Max. :2275925
```

##sindh data and I have read the data from sindh_data file

```
data=read.csv(file.choose())
ex_date=as.Date(data$date)
```

```
week_date=cut(ex_date,"week")
```

```
new_wd= gsub(", ", "", data$cases)
ew=as.numeric(new_wd)
d=aggregate(ew~week_date, data, sum)
summary(d)
```

```
Min. : 23
1st Qu.: 530207
Median : 986074
Mean :1029454
3rd Qu.:1693998
Max. :2087357
```

KPK data and I have read the data from kpk_data file

```
data=read.csv(file.choose())
ex_date=as.Date(data$date)

week_date=cut(ex_date,"week")
new_wd= gsub(", ", "", data$cases)
ew=as.numeric(new_wd)
d=aggregate(ew~week_date, data, sum)
summary(d)
```

```
Min. : 105
1st Qu.:194143
Median :270876
Mean :327360
3rd Qu.:473195
Max. :883127
```

Balochistan data and I have read the data from balochistan_data file

```
data=read.csv(file.choose())
ex_date=as.Date(data$date)

week_date=cut(ex_date,"week")
new_wd= gsub(", ", "", data$cases)
ew=as.numeric(new_wd)
d=aggregate(ew~week_date, data, sum)
summary(d)
```

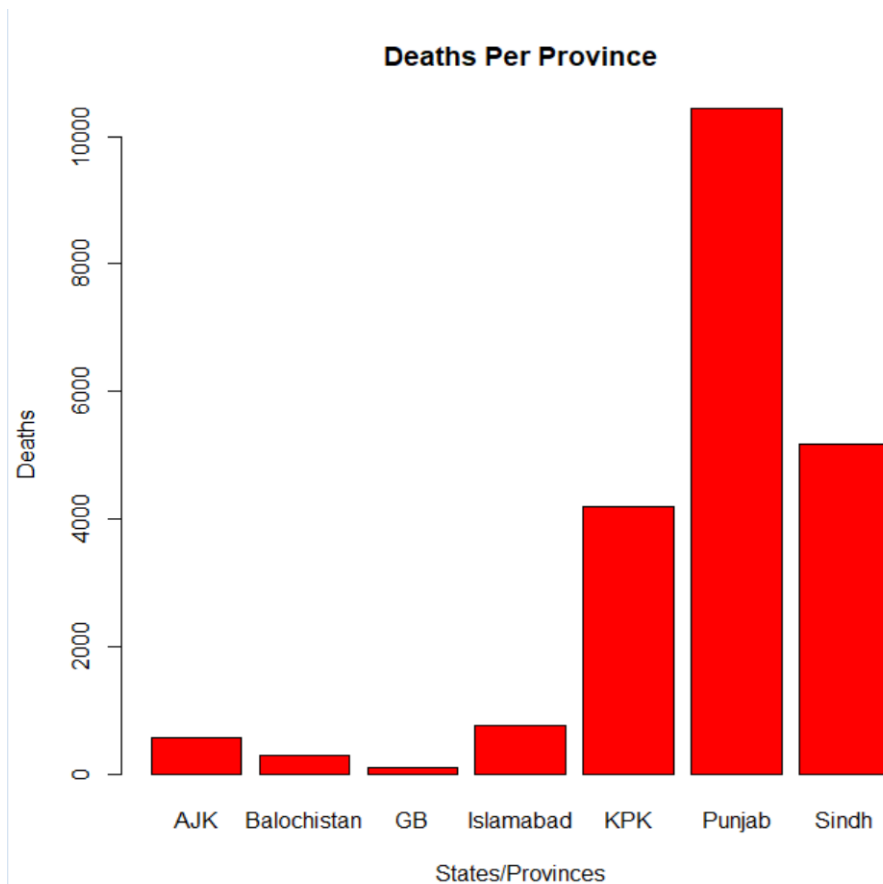
```
Min. : 47
1st Qu.: 73364
Median :108601
Mean : 92878
```

3rd Qu.:131393
Max. :167012

Part-3

```
> H<-c(559,291,107,768,4185,10436,5183)
> M<-c("AJK","Balochistan","GB","Islamabad","KPK","Punjab","Sindh")
> barplot(H,names.arg=M,xlab="States/Provinces",ylab="Deaths",col="red",main="Deaths Per Province",border="black")
```

Graph:



Data:

Deaths by Provinces:

AJK: 559

Balochistan: 291

GB: 107

Islamabad: 768

KPK: 4,185

Punjab: 10,436

Sindh: 5,183

Dated: June 10,2021

Source: <https://covid.gov.pk/stats/pakistan>

Explanation:

The values mention above are directly plotted using **barplot()** function. As, day-by-day date for province wise deaths wasn't available.

Part 4:

<https://forms.gle/FjBjFeFMF6sowiUm6>

Part 5:

i have read the data from frq file

data=read.csv(file.choose())

classes=seq(20,75,5)

extract=data\$Age

cutter=cut(extract,classes,right="FALSE")

table(cutter)

frq_dis=cbind(table(cutter))

frq_dis

com_frq=cumsum(frq_dis)

com_frq

frequency distribution:

[20,25) 33

[25,30) 8

[30,35) 2

[35,40) 3

[40,45) 3

[45,50) 0

[50,55) 6

[55,60) 4

[60,65) 4

[65,70) 6

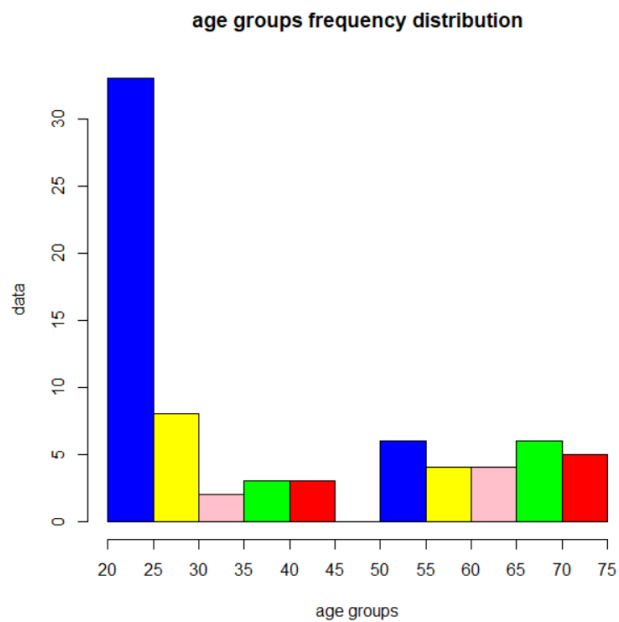
[70,75) 5

Cumulative frequency:

33 41 43 46 49 49 55 59 63 69 74

##histogram

```
with(data,hist(gh[Age>=20 & Age<=75],xlab="age groups",ylab="data",main="age groups frequency  
distribution",col=c("blue","yellow","pink","green","red","grey"),breaks=seq(20,75,by=5),xaxp=c(20,75,1  
1)))
```

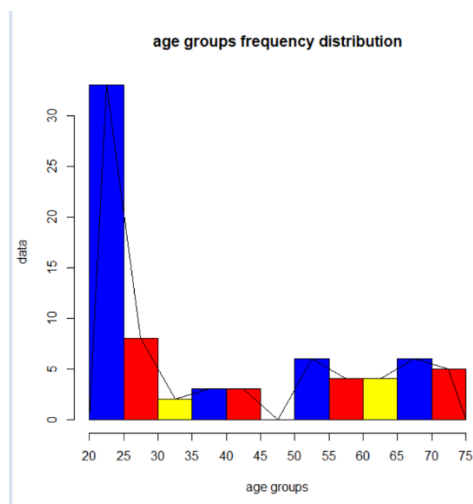


Frequency polygon:

```
##frequencypolygon
```

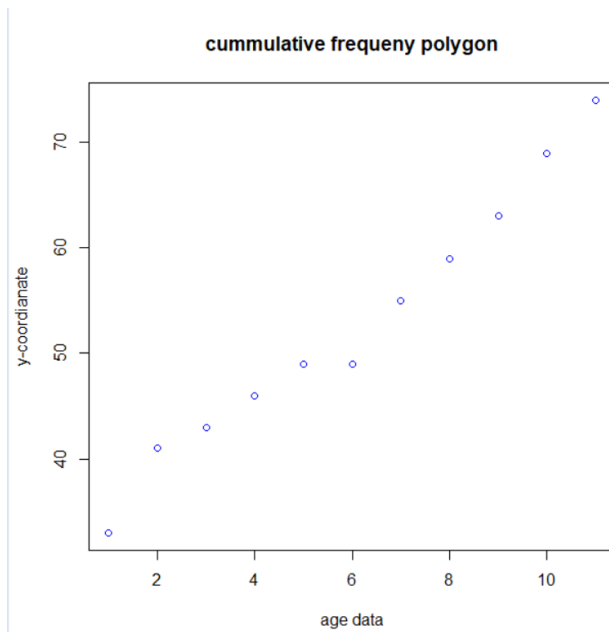
```
library(UsingR)
```

```
with(data,simple.freqpoly(gh[Age>=20 & Age<=75],xlab="age groups",ylab="data",main="age groups
frequency distribution",col=c("blue","red","yellow"),breaks=seq(20,75,by=5),xaxp=c(20,75,11)))
```



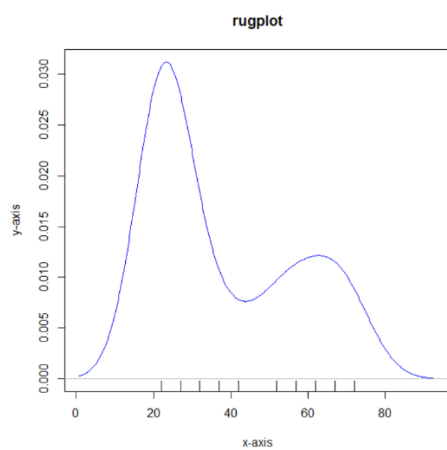
##cumulative frequency polygon:

```
plot(com_frq,main="cumulative frequeny polygon",xlab="age data",ylab="y-coordianate",col="blue")
```



##rugplot

```
with({  
  plot(density(extract),xlab="x-axis",ylab="y-axis",main="rugplot",col="blue")  
  rug(extract,ticks=0.03,side=1,lwd=0.5)  
})
```

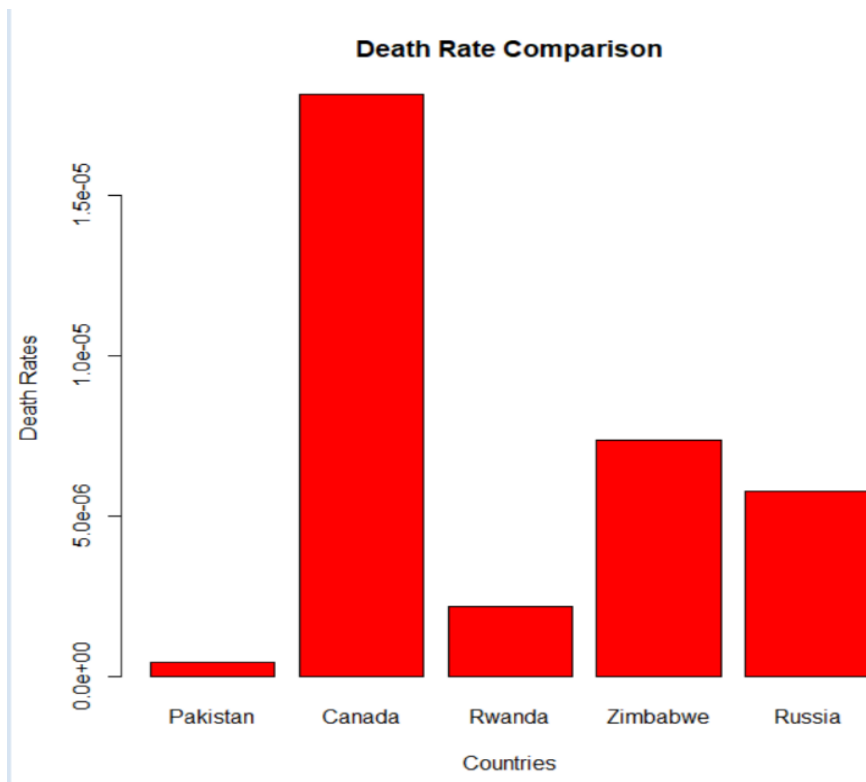


Part-6

The files Canada data, Pakistan data, Russia data, Rwanda data are used in 6th point

```
> data=read.csv(file.choose())
> pakistan=tail(data$pakistan, n=1) / tail(data$pop0, n=1)
> data=read.csv(file.choose())
> canada=tail(data$canada, n=1) / tail(data$pop1, n=1)
> data=read.csv(file.choose())
> rwanda=tail(data$rwanda, n=1) / tail(data$pop2, n=1)
> data=read.csv(file.choose())
> zimbabwe=tail(data$zimbabwe, n=1) / tail(data$pop3, n=1)
> data=read.csv(file.choose())
> russia=tail(data$russia, n=1) / tail(data$pop4, n=1)
> H<-c(pakistan,canada,rwanda,zimbabwe,russia)
> M<-c("Pakistan","Canada","Rwanda","Zimbabwe","Russia")
> barplot(H,names.arg=M,xlab="Countries",ylab="Death Rates",col="red",main="Death Rate
Comparison",border="black")
```

Graph:



Explanation:

In the first statement, CSV file is being loaded up. Then in **pakistan** variable, the result of the division of the total deaths and population is being stored, which is the death rate. I have used **tail()** command in the calculations which returns the last row of the CSV of a particular table. In my opinion it was more efficient and faster than other methods. Same procedure has been used for other countries as well. Then all death rates are stored in **H** vector. Names of the countries are stored in **M** vector. Then all values are being plotted by **barplot()** function.
