



Probability and Statistics
Spring 2021
Final Project

Group Members:

Eman Hassan (19I-2045)

Zain Ul Abideen Arshad (19I-1987)

Section: G

Submitted to: Sir Waqas Munir

Dated: 12/06/2021

- **Covid-19 Pakistan Data**

The data for Pakistan related to Covid-19 is as below, dated *10 June 2021*,

Data Of	Total Cases
Total/Confirmed Cases	937,434
Deaths	21,529
Recovered Cases	871,669
Total Tests	13,705,490
Active Cases	44,236
Critical Cases	2,967

The links to the data shown above are given below,

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

The above-mentioned website is the official website of Government of Pakistan which handles all the information related to COVID-19 and is updated on daily basis.

<https://github.com/owid/covid-19-data/tree/master/public/data>

And this link is of GitHub where statistics related to COVID-19 are updated for approximately all countries across the globe.

https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%22%20%5C1%20%22countries#countries

- **Summary of total cases province wise:**

The dataset used for this task is from:

https://en.wikipedia.org/wiki/Template:COVID-19_pandemic_data/Pakistan_medical_cases

Csv sheet used for this task is as follow,

1	date	punjab	sindh	kpk	balochista	gilgit	azad_kash	islamabad
2	2020-02-26	0	1					1
3	2020-02-27	0						
4	2020-02-28	0						
5	2020-02-29	0	2					2
6	2020-03-01	0						
7	2020-03-02	0						
8	2020-03-03	0				1		
9	2020-03-04	0						
10	2020-03-05	0	3					
11	2020-03-06	0						
12	2020-03-07	0						
13	2020-03-08	0	4					
14	2020-03-09	0	13					
15	2020-03-10	0	14		1	2		2
16	2020-03-11	0				3		
17	2020-03-12	0	15					
18	2020-03-13	0	16		6			3
19	2020-03-14	0	18					4
20	2020-03-15	1	34		11			
21	2020-03-16	0	150	15	14			

The dataset starts from 26 March,2020 and is till 22 May 2021. The data is then transformed into weekly dataset and then plotted. The following R code is used and from that the summary written below is obtained.

R-Code

```
data = read.csv(file.choose())
dates=data$date
dates=as.Date(data$date)
wd=cut(dates,"week")
pag=aggregate(punjab~wd,data,sum)
summary(pag)
sag=aggregate(sindh~wd,data,sum)
summary(sag)
kag=aggregate(kpk~wd,data,sum)
summary(kag)
bag=aggregate(balochistan~wd,data,sum)
```

```
summary(bag)
gag=aggregate(gilgit~wd,data,sum)
summary(gag)
ajkag=aggregate(azad_kashmir~wd,data,sum)
summary(ajkag)
summary(iag)
```

Summary

Provinces	Min	1 st Quartile	Median	Mean	3 rd Quartile	Max
Punjab	0	432831	702718	793076	1094217	2268001
Sindh	3	437845	976585	1013617	1712219	2073889
Balochistan	18	72095	109339	93087	131740	166206
KPK	105	194143	270876	327360	473195	883127
Islamabad	3	80110	122072	185059	289159	552007
Gilgit	1	9422	21294	19512	30251	37890
AJK	1	8833	24511	39840	63608	127056

- **Comparing number of deaths province wise:**

Punjab has 10,314 deaths till now, **Sindh** has 5,136 deaths recorded, **KPK** has 4,158 deaths recorded, **Islamabad** has the figure of 765, **Azad Kashmir** has 554, **Balochistan** has 289 deaths and **Gilgit** has 107 deaths until now. We would compare the data using bar graph.

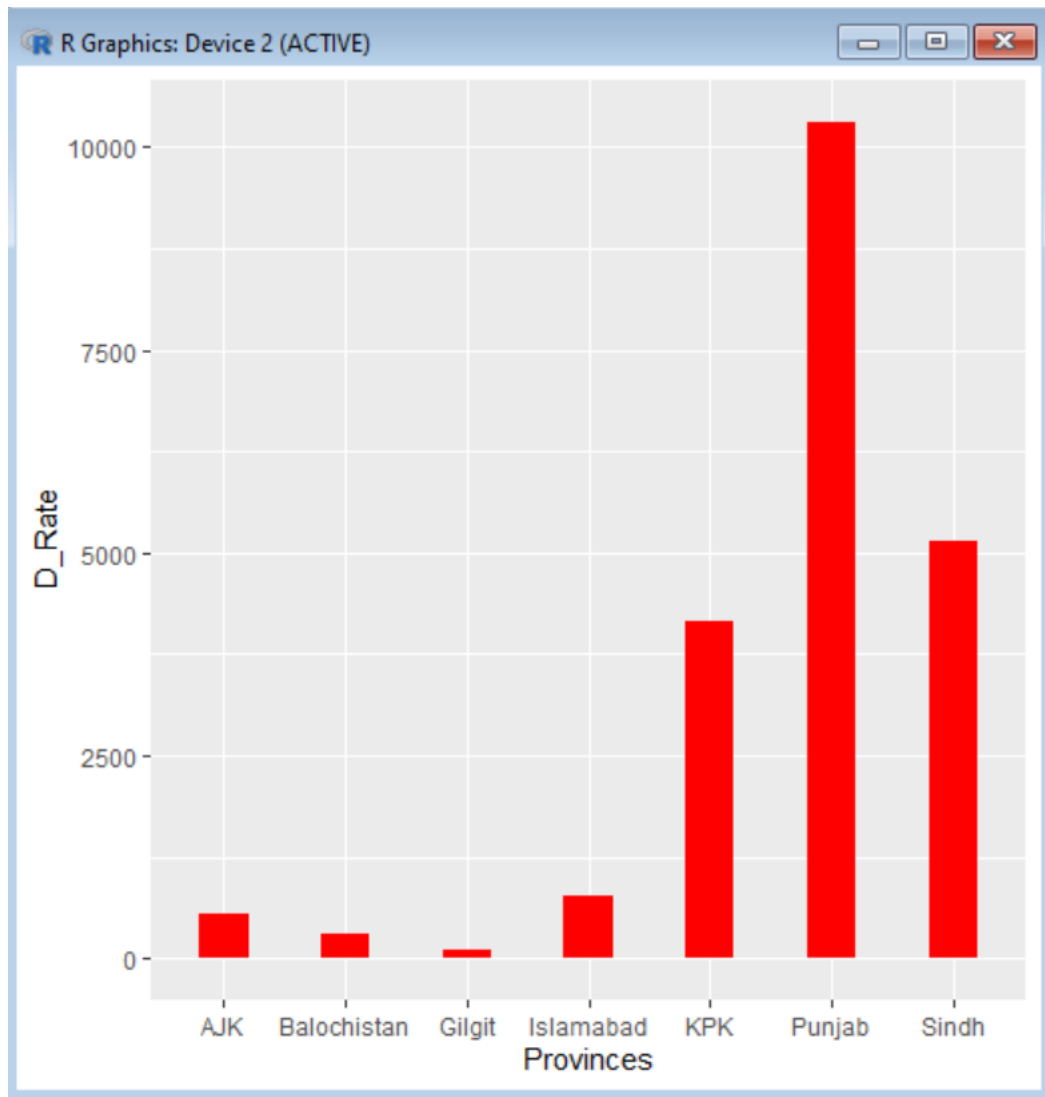
R-Code:

```
library(ggplot2)
D_Rate<-c(10314,5136,4158,765,554,289,107)
Provinces<-c("Punjab","Sindh","KPK","Islamabad","AJK","Balochistan","Gilgit")
```

```
data<-data.frame(D_Rate,Provinces)
```

```
ggplot(data,aes(x=Provinces, y=D_Rate)) + geom_bar(stat = "identity",  
width = 0.4, fill = "red")
```

Graph:

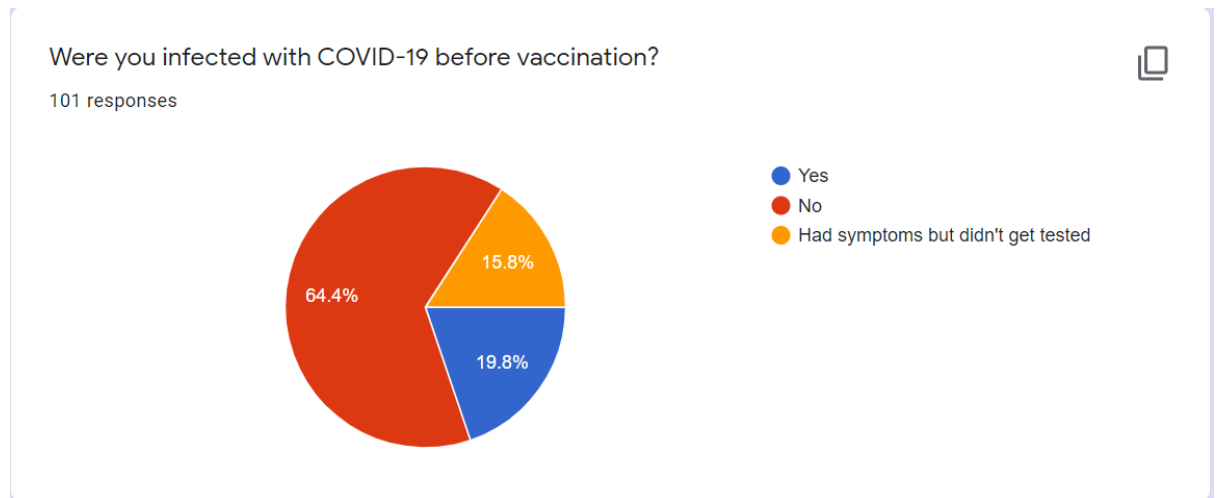


Barplot is the most suitable graphical form to present this data. The dataset is collected from: <https://covid.gov.pk/stats/pakistan>

- **Survey Form:**

A survey was conducted from a group of 100 people, where the people were asked about their concerns and thoughts about the vaccine and its safety. Here is the link to the form: <https://forms.gle/fbdmXN5CicBM2X2w5>

One of the questions that was asked in the survey was whether the respondents were infected with COVID or not. To which the following results are summarized,

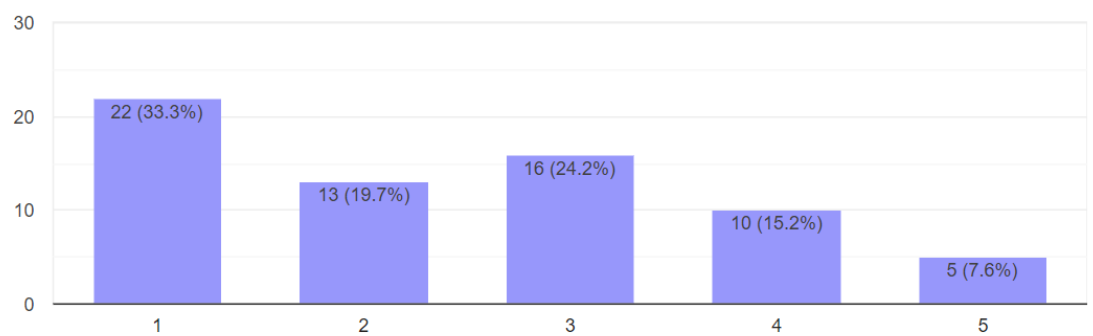


From the above chart, we can see that 64.4% of the respondents didn't contract corona virus, whereas 19.8% were confirmed patients and 15.8% of them had symptoms but they didn't get tested. So, majority of the respondents were not infected.

Another question asked was, if the respondents had contracted the virus what the was the severity of the symptoms to which we got the response,

What was the severity of the COVID symptoms on scale 1 to 5?

66 responses

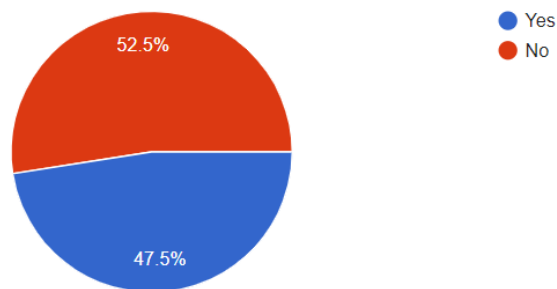


From the above bar graph, we can see that the 33.3% people had no to very low symptoms, and 7.6% people had high symptoms.

The next question asked was did the respondents get vaccinated or not and if yes, why vaccine did they opt for. To these questions following were the responses,

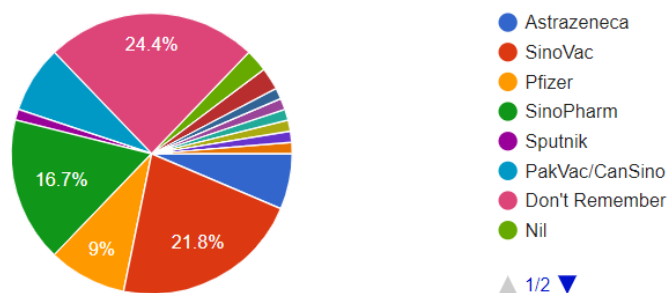
Did you get yourself vaccinated?

101 responses



If yes, which vaccine did you receive?

78 responses



The above charts represent that out of 101 people only 47.5% people got vaccinated and from the chart below we can see the vaccines which the respondents got.

Following this when the reason was asked that why the respondents didn't get vaccinated the following response was received,

If no, then why not have you gotten yourself vaccinated till yet?

46 responses

Because our vaccination has not been arrived in pakistan yet

Because the vaccine were only available for old age before and now I'll be doing my registration soon.

I have not registered myself till now

Because of insufficient vaccines in the area

I dont feel the need to get it

Dont have any specific reason

Wss not available for my age group

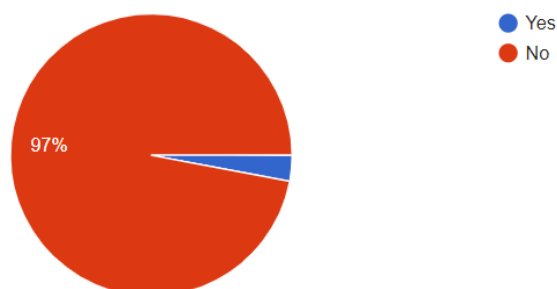
Will soon cuz just get registered.

Got no time for that.

Another question was whether the respondents contracted COVID after the vaccine to which 97% respondents selected NO.

Did you contract virus after vaccination?

101 responses



At last, the opinion of the respondents was asked whether they think the vaccine is safe or not. To which majority of the people responded that they think it is safe. Below are some of the responses which were recorded,

Do you think vaccine is safe? Why or why not. Give reason.

101 responses

Vaccine is definitely better than contracting the virus

Yes its totally safe. Me being a medical student is suggesting everyone in pakistan to get yourself vaccinated. Because corona is a severe disease and it must be fought away in some way. So kindly get yourself vaccinated

No because it has some side effects

The Vaccine is safe as all other vaccines are safe that we have taken since birth. Nothing to overthink about it so much.

I think yes it help you from severe covud attack.

Yes it's safe and beneficial for health

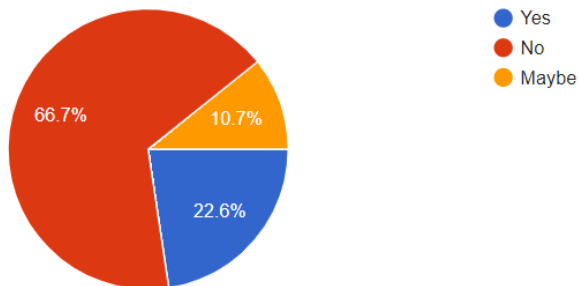
Vaccine is important for us to prevent this disease. And to protect ourselves and other members also.

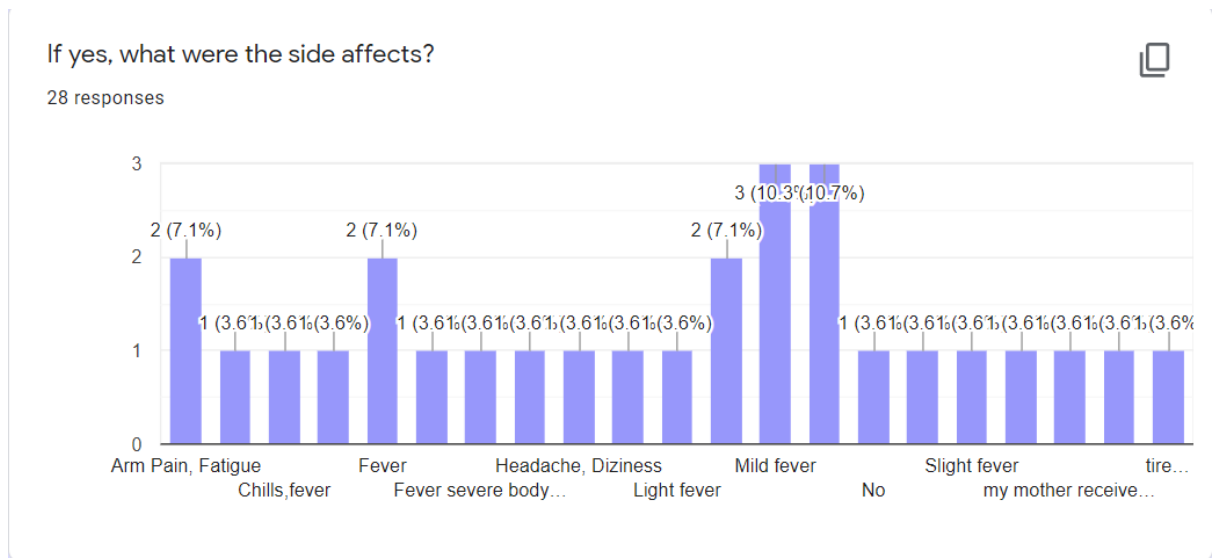
Yes it boosts our immunity

Some of the other questions that were asked are as below,

Did you feel any side affects after receiving the vaccine?

84 responses





So, from analysing the survey results, we can conclude that most people think that vaccine is safe. Few people have concerns, but they are not any major concerns.

Now, we would represent the group of vaccinated people – data obtained from above survey, through different graphical methods.

- **Representation of vaccinated people:**

	frequency	MidPoints	cumulative_frequency	Relative Freq.
[18,25)	18	21.5	18	0.37500000
[25,32)	15	28.5	33	0.31250000
[32,39)	1	35.5	34	0.02083333
[39,46)	7	42.5	41	0.14583333
[46,53)	2	49.5	43	0.04166667
[53,60)	2	56.5	45	0.04166667
[60,67)	3	63.5	48	0.06250000

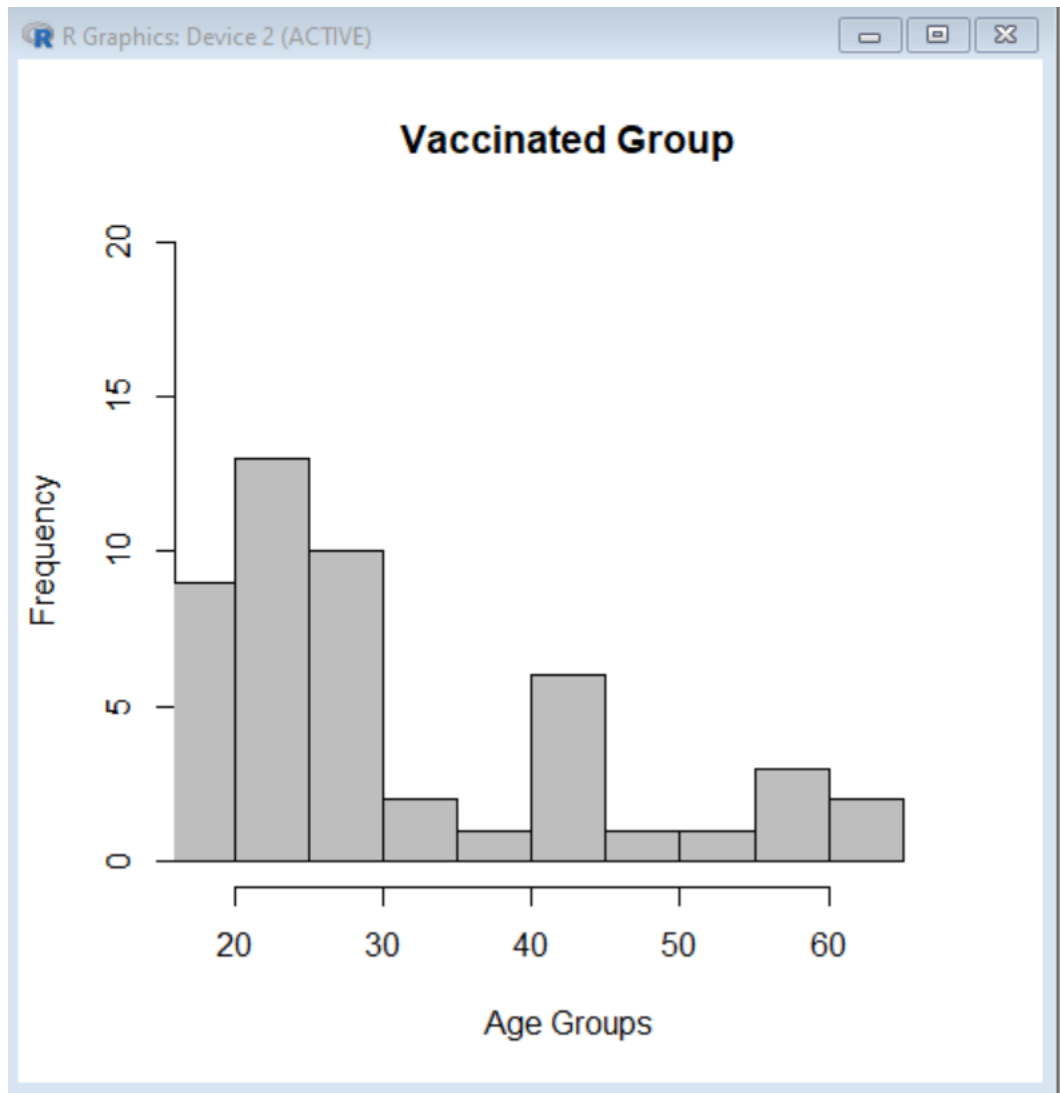
There were 48 people out of 101 people who were vaccinated. The above table is obtained from the following R code which would then be helpful for us to plot the asked graphs,

R-Code:

```
data=c(18,18,18,19,19,20,21,21,22,22,22,23,23,24,25,25,25,25,26,26,26,
27,27,27,27,28,28,19,20,19,21,29,31,35,40,41,49,45,43,44,44,42,52,60,5
8,56,65,61)
sort(data)
num<-length(data)
maximum<-max(data);
minimum<-min(data);
range<-(maximum-minimum)
noOfClasses=ceiling(1+3.322*log10(num))
height<-ceiling(range/noOfClasses)
breaks<-seq(minimum,maximum+height,height)
dataDistribution<-cut(data,breaks,right=F)
frequency = table(dataDistribution)
table = cbind(frequency)
MidPoints<-seq(minimum+height/2,(maximum+height)-height/2,height)
cumulative_frequency=cumsum(table)
RF = table/num
colnames(RF)<- "Rel f"
finalTable = cbind(frequency ,MidPoints,cumulative_frequency,RF)
finalTable
```

1. Histogram:

```
hist(finalTable,xlab="Age",ylab="Frequency")
h=hist(data,xlim=c(18,67),ylim=c(0,20),xlab = "Age Groups",col =
"grey",main="Vaccinated Group", breaks = 7)
```



2. Frequency Polygon:

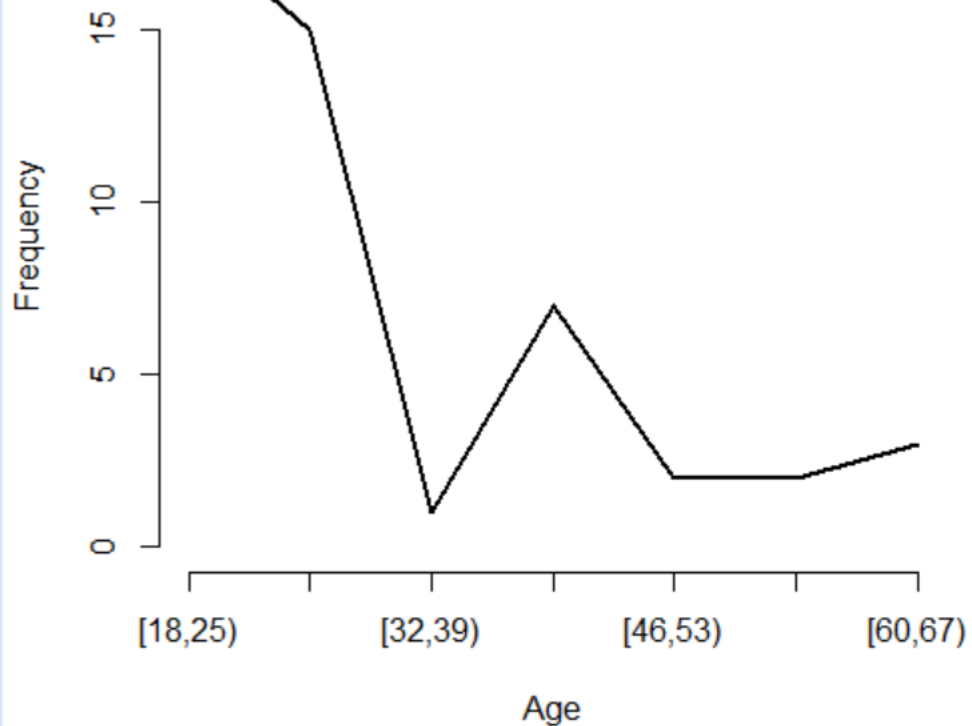
R-Code:

```
breaks=seq(18,67,7)
```

```
distribution=cut(data,breaks,right=FALSE)
```

```
values.freq=table(distribution)
```

```
plot(values.freq,xlab="Age",ylab="Frequency",type="l")
```



3. Cumulative Frequency Polygon:

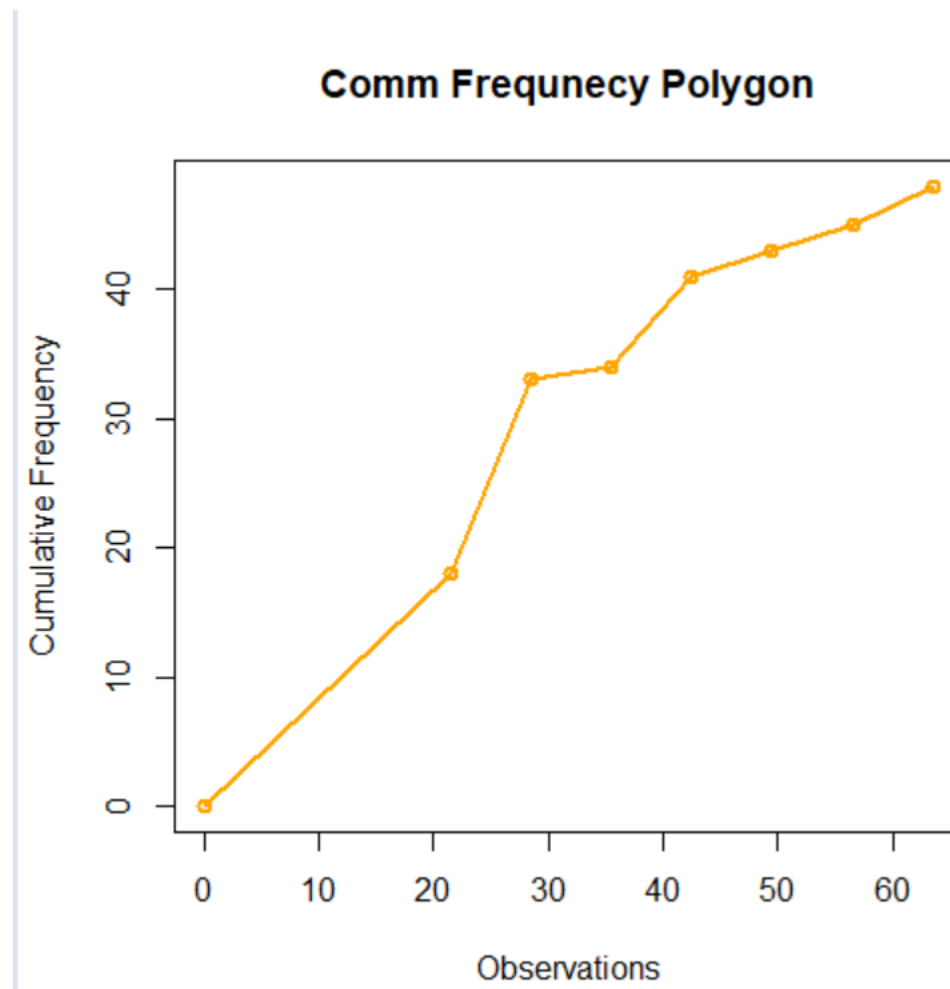
R-Code:

```
n<-length(data)
maxi1<-max(data);
maxi1
mini1<-min(data);
mini1
r1<-(maxi1-mini1)
r1
noOfClasses = ceiling(1+3.322*log10(length(data)))
noOfClasses
h1<-ceiling(r1/noOfClasses) # To find the width/size of class.
h1
MidPoints<-seq(mini1+h1/2,(maxi1+h1)-h1/2,h1)
MidPoints
breaks = seq(mini1,maxi1+h1,h1)
```

```

dataCut = cut(data, breaks, right=FALSE)
frequency = table(dataCut)
midpoint0 = c(0, MidPoints)
cumulativeFreq = c(0, cumsum(frequency))
plot(midpoint0, cumulativeFreq, main="Comm Frequency
Polygon", xlab="Observations", ylab="Cumulative Frequency")
lines(midpoint0, cumulativeFreq, col="orange", lwd=2, type="o")

```



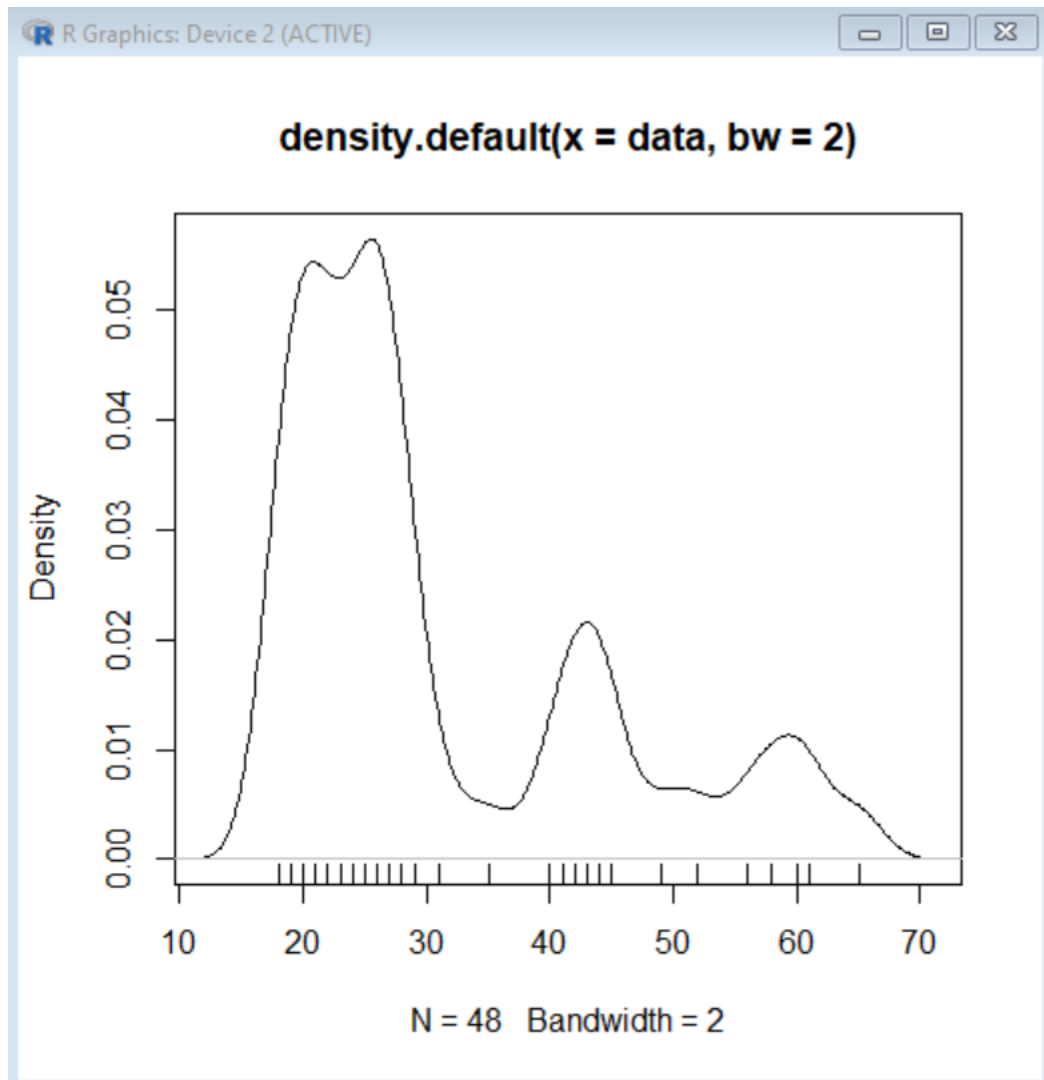
4. Rug Plot:

R-code:

```

plot(density(data,bw=2.0))
rug(data)

```



- **Comparison of Death Rate with different Countries:**

To compare the death rate of Pakistan with other countries, we have collected the data from the following website,

https://www.worldometers.info/coronavirus/?utm_campaign=homeAdvegas1?%22%20%5C%20%22countries#countries

The death rate can be calculated from the following formula,

Death_rate = (total deaths/total cases) x 100

1. Pakistan

Total cases: 939,931

Total deaths: 21,633

Death Rate: 2.30%

2. Turkey

Total cases: 5,319,359

Total deaths: 48,593

Death Rate: 0.9%

3. USA

Total cases: 34,306,446

Total deaths: 614,738

Death Rate: 1.79%

4. India

Total cases: 29,359,155

Total deaths: 367,097

Death Rate: 1.2%

5. Argentina

Total cases: 4,093,090

Total deaths: 84,628

Death Rate: 2.06%

R-Code:

```
rate <-c(2.3,0.9,1.79,1.2,2.06)  
labels<-c("Pakistan","Turkey","USA","India","Argentina")  
barplot(rate,names.arg=labels,xlab="Countries",ylab="Death Rate",col="pink",main="Death Rate Comparison")
```

Graph:

