COVID-19 VISUALIZATION AND PREDICTION

DATASET USED:

"C:\Users\nivet\Downloads\corona_tested_individuals_ver_006.english.csv"

Import the dataset:

```
>library(readr)
>my_data1<-read.csv("C:/Users/nivet/Downloads/corona_tested_individuals_ver_006.english.csv")
```

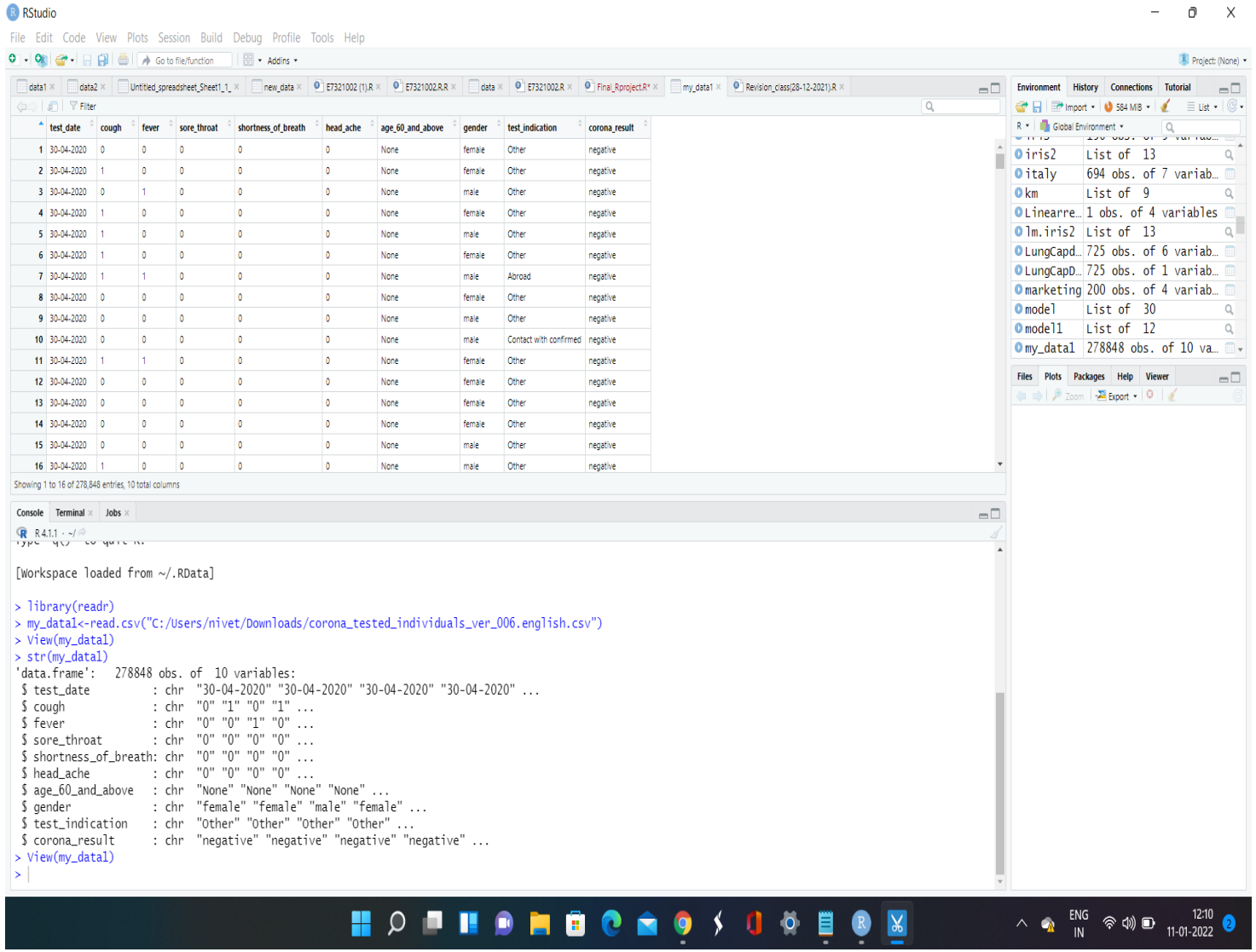
Now,

Exploratory data analysis:

```
>str(my_data1)
```

This gives the no. of rows and no. of columns in the dataset and the column names .

```
'data.frame': 278848 obs. of  10 variables:
 $ test_date         : chr  "30-04-2020" "30-04-2020" "30-04-2020" "30-04-2020" ...
 $ cough             : chr  "0" "1" "0" "1" ...
 $ fever             : chr  "0" "0" "1" "0" ...
 $ sore_throat       : chr  "0" "0" "0" "0" ...
 $ shortness_of_breath: chr  "0" "0" "0" "0" ...
 $ head_ache         : chr  "0" "0" "0" "0" ...
 $ age_60_and_above  : chr  "None" "None" "None" "None" ...
 $ gender            : chr  "female" "female" "male" "female" ...
 $ test_indication   : chr  "Other" "Other" "Other" "Other" ...
 $ corona_result     : chr  "negative" "negative" "negative" "negative" …
```



- We can see that our dataset contains 278848 rows of 10 columns.
- Now,

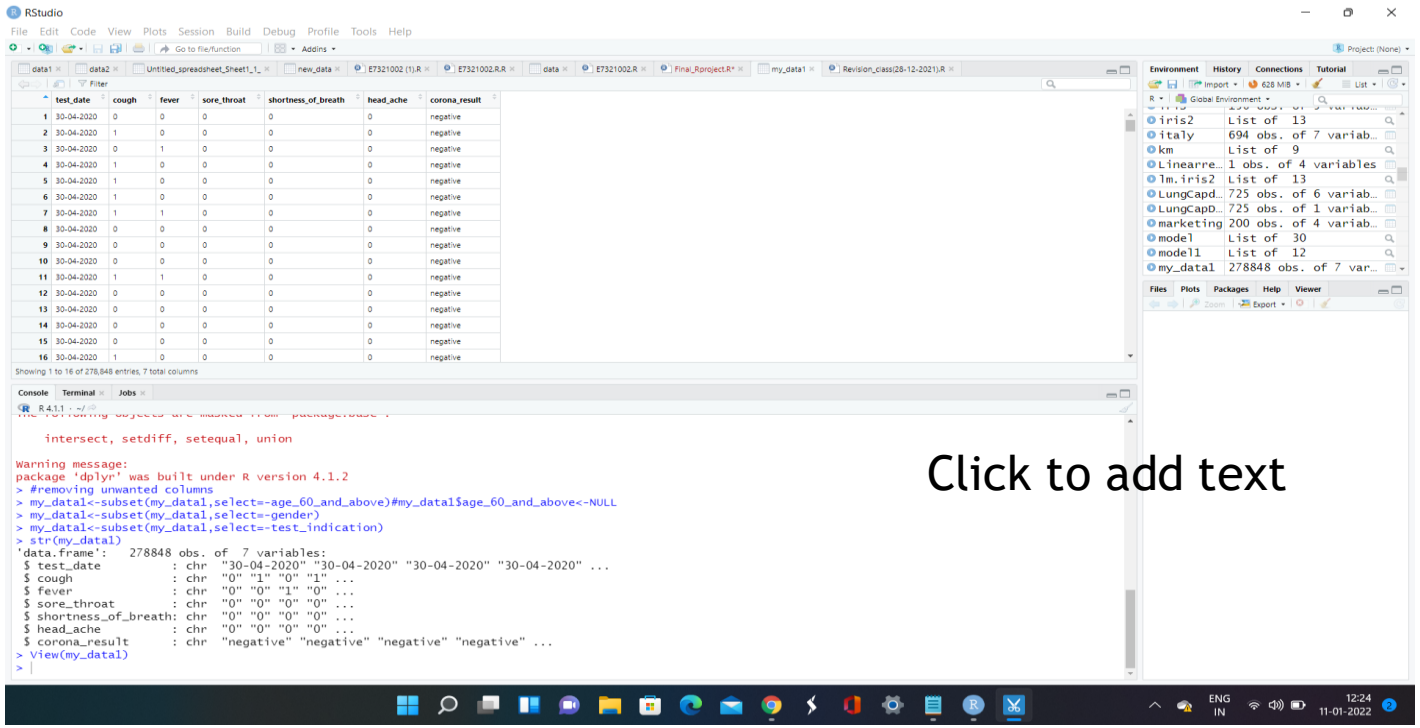
>View(my_data1)

- REMOVING UNWANTED COLUMNS:
- my_data1<-subset(my_data1,select=-age_60_and_above)#my_data1$age_60_and_above<-NULL
- my_data1<-subset(my_data1,select=-gender)
- my_data1<-subset(my_data1,select=-test_indication)
- Now,

```
>str(my_data1)
'data.frame': 278848 obs. of  7 variables:
 $ test_date          : chr  "30-04-2020" "30-04-2020" "30-04-2020" "30-04-2020" ...
 $ cough              : chr  "0" "1" "0" "1" ...
 $ fever              : chr  "0" "0" "1" "0" ...
 $ sore_throat        : chr  "0" "0" "0" "0" ...
 $ shortness_of_breath: chr  "0" "0" "0" "0" ...
 $ head_ache          : chr  "0" "0" "0" "0" ...
 $ corona_result      : chr  "negative" "negative" "negative" "negative" …
```

We can see that,the columns "age_60_and_above","gender", and "test_indication" are removed.

- We can see this:
- >View(my_data1)



Click to add text

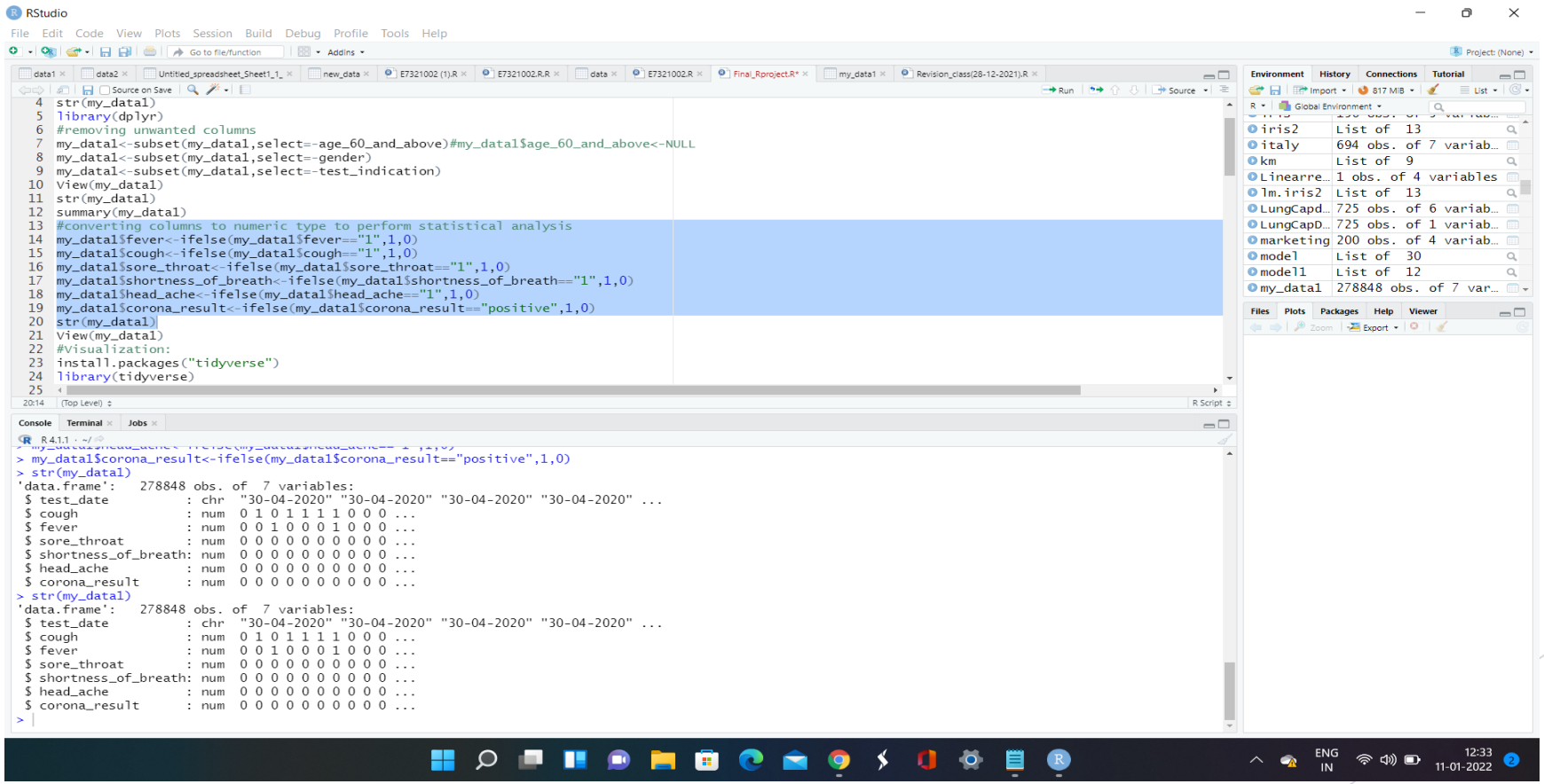
```
>summary(my_data1)#This gives the column names and their length and their type.
```

- test_date cough fever sore_throat shortness_of_breath head_ache corona_result
- Length:278848 Length:278848 Length:278848 Length:278848 Length:278848 Length:278848 Length:278848
- Class :character Class :character Class :character Class :character Class :character Class :character Class :character
- Mode :character Mode :character Mode :character Mode :character Mode :character Mode :character Mode :character
- We can see that all columns are of character type. So,let us convert this into numeric type to perform statistical analysis using ifelse function.

```
>my_data1$fever<-ifelse(my_data1$fever=="1",1,0)

>my_data1$cough<-ifelse(my_data1$cough=="1",1,0)

>my_data1$sore_throat<-ifelse(my_data1$sore_throat=="1",1,0)

>my_data1$shortness_of_breath<-ifelse(my_data1$shortness_of_breath=="1",1,0)

>my_data1$head_ache<-ifelse(my_data1$head_ache=="1",1,0)

>my_data1$corona_result<-ifelse(my_data1$corona_result=="positive",1,0)
```

By doing this, the null values also get replaced by 0.

- Now,
- >str(my_data1)
- We can see that except "test_date" column,all other columns have been converted to numeric type.

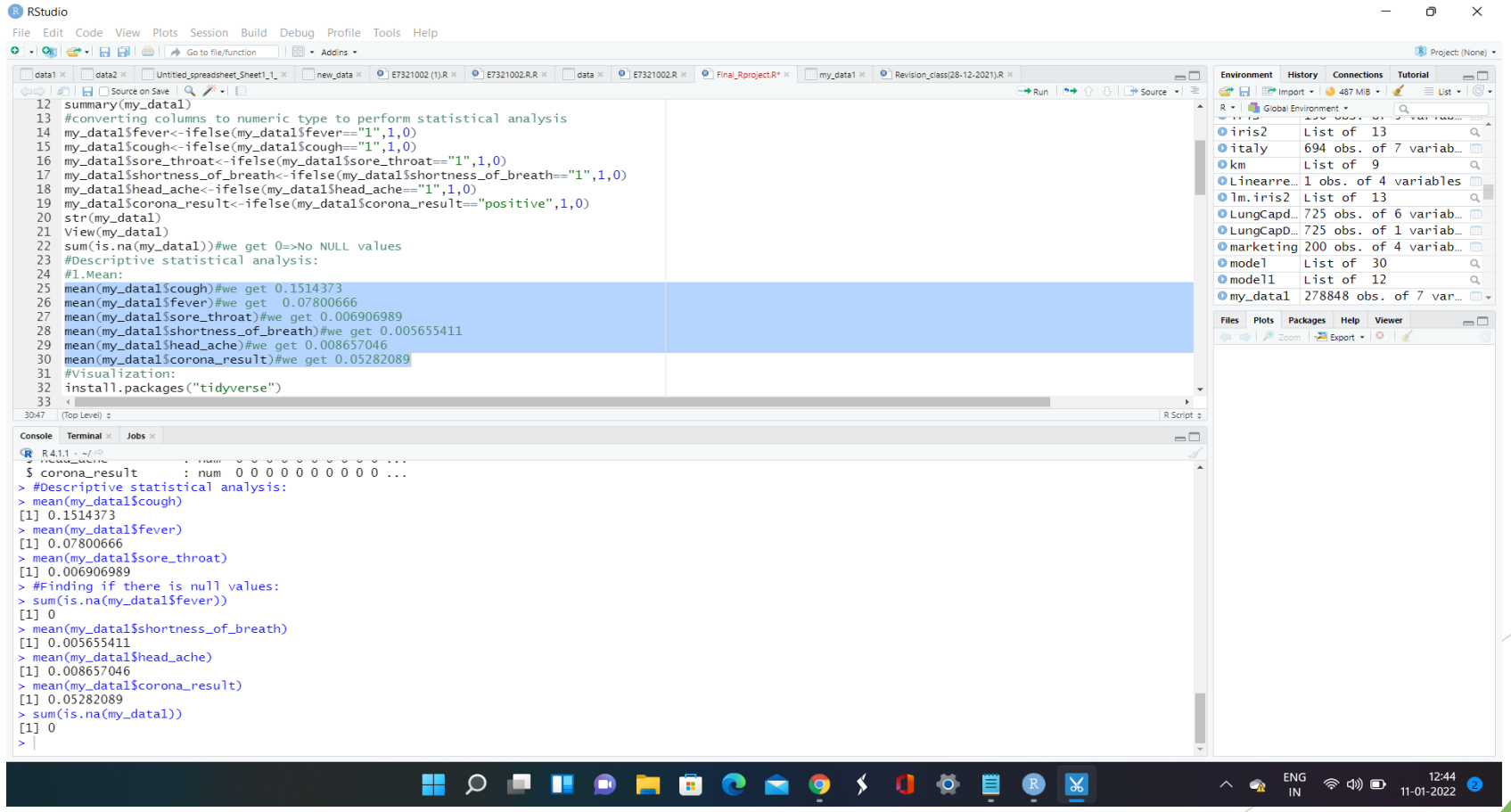


Checking if there is Null values:

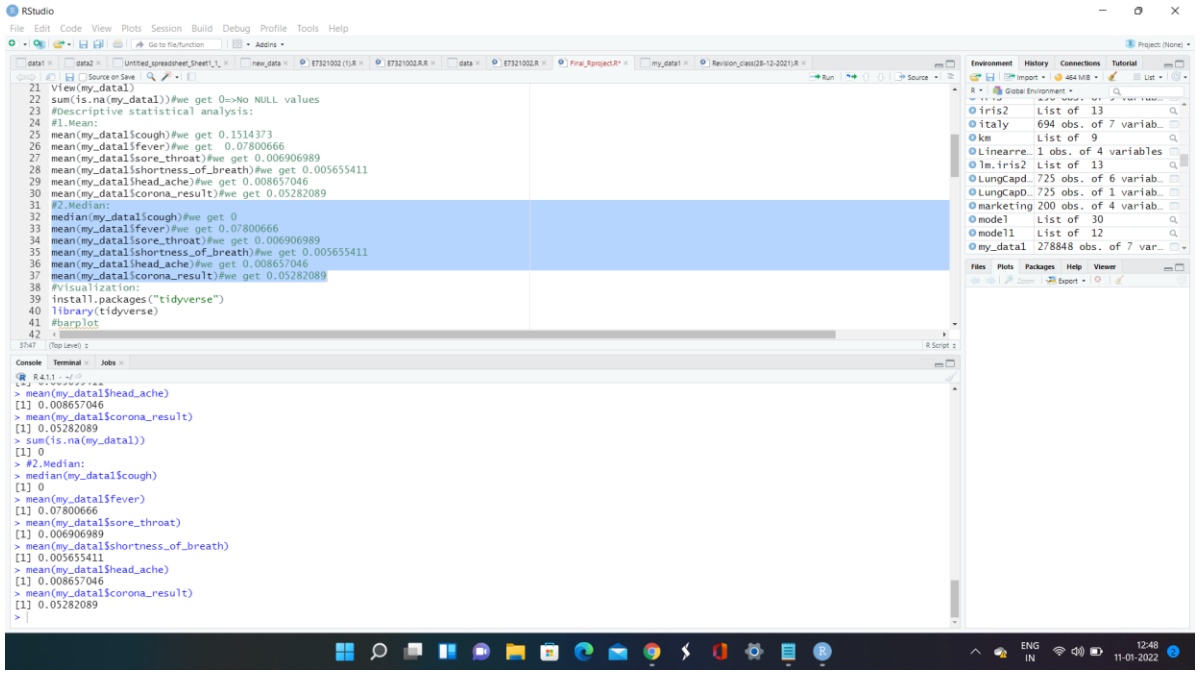
>sum(is.na(my_data1))#we get 0=>No NULL values

DESCRIPTIVE STATISTICAL ANALYSIS:

1.Mean:



- 2.MEDIAN:



3. MODE:

>table(as.vector(my_data1$cough))

0 1

236620 42228

Interpretation:

We can see that 0 occurs most times=>0 is the mode for this column.

Similarly, we can find the mode for other columns.

4.VARIANCE:

```
>var(my_data1$sore_throat)
  0.006859307
```

Similarly, we can find the varaince of other columns

5.STANDARD DEVIATION:

```
>sqrt(var(my_data1$corona_result))
 0.2236762
```

Similarly,we can find the standard deviation for other columns

6. RANGE:

range() gives min and max value of that column

```
>range(my_data1$shortness_of_breath)#we get min=0 and max=1
>range(my_data1$fever)#we get min=0 and max=1
```

Similarly we can find the range of other columns

TO USE skewness() and kurtosis() method, install moments package:

```
>install.packages("moments")
>library(moments)
```

7. SKEWNESS:

```
> skewness(my_data1$head_ache)
```

10.60762

We get 10.60762=>positively skewed

8. KURTOSIS:

```
>kurtosis(my_data1$cough)
```

4.781854

We get 4.781854=>Leptokurtic

DATA VISUALIZATION:

Install the packages "tidyverse" and "ggfortify"

```
>install.packages("tidyverse")
>library(tidyverse)
>install.packages("ggfortify")
>library(ggplot2)
```

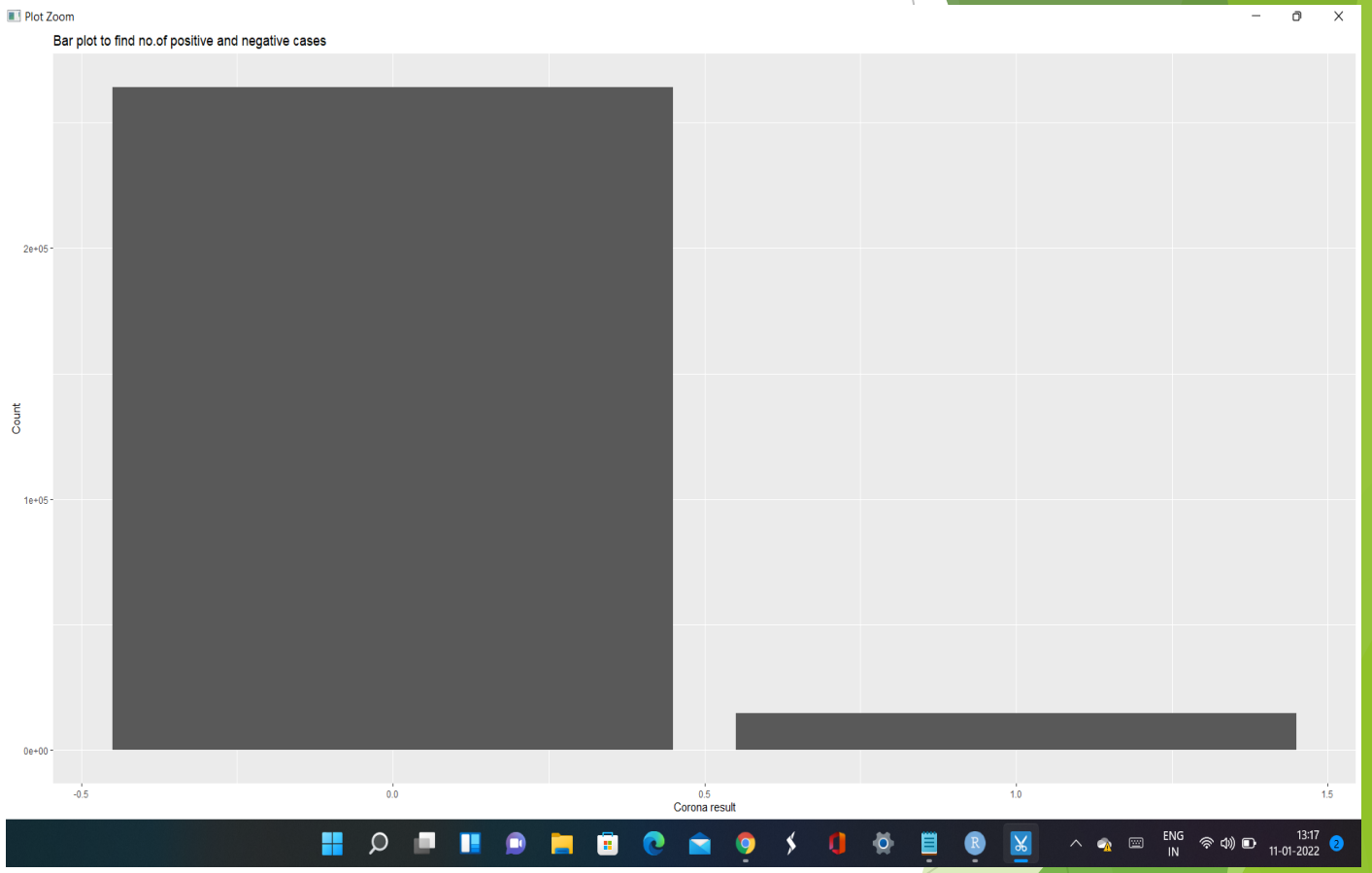
BAR PLOT:

Bar plot w.r.t "corona_result" column

```
ggplot(my_data1,aes(x=corona_result))+
  geom_bar()+
  xlab("Corona result")+
   ylab("Count")+
   ggtitle("Bar plot to find no.of positive and negative cases")
```

INTERPRETATION:

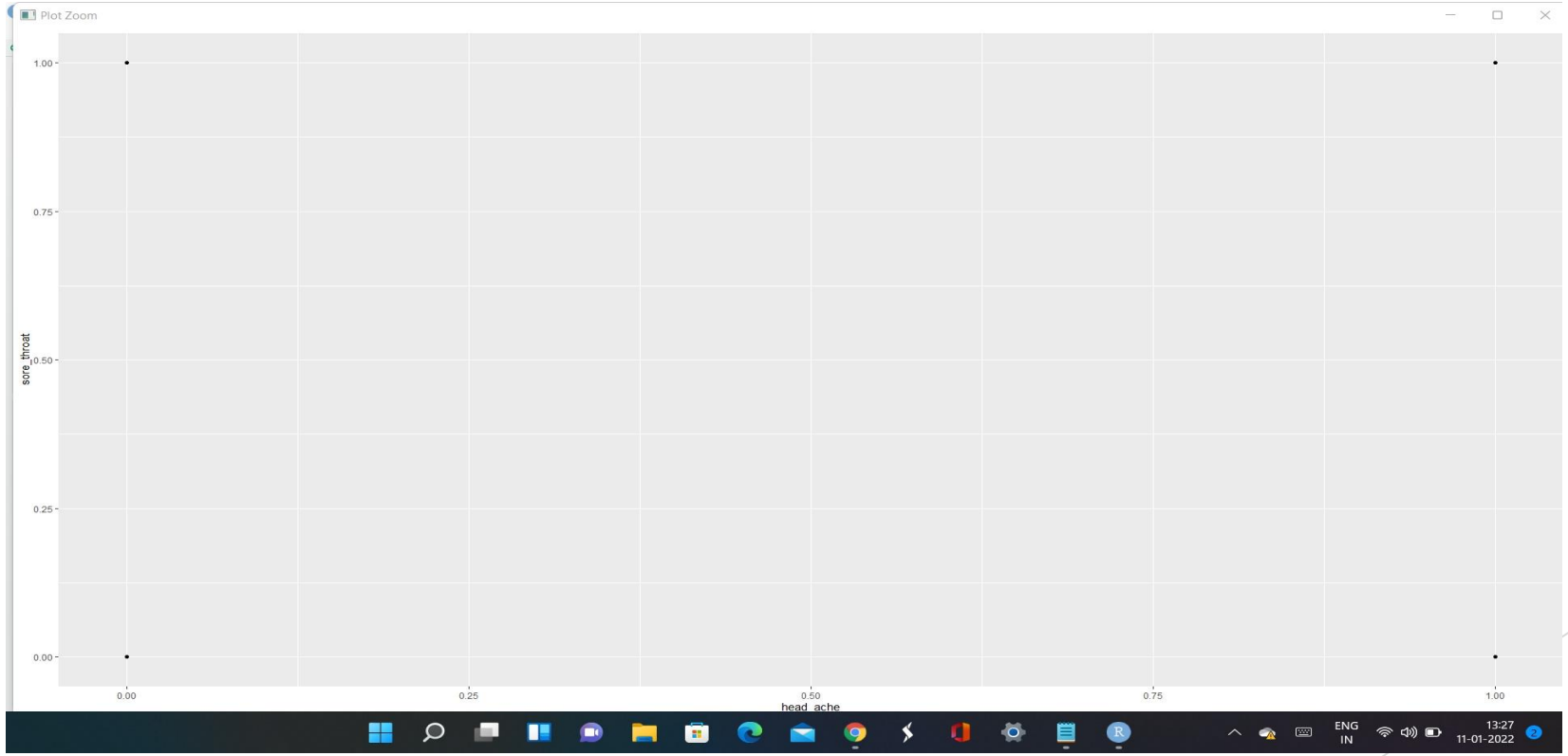
From the graph, we can see that, in the dataset, only few have been tested positive



SCATTER PLOT:

```
>ggplot(data=my_data1,aes(x=head_ache,y=sore_throat))+
  geom_point()
```

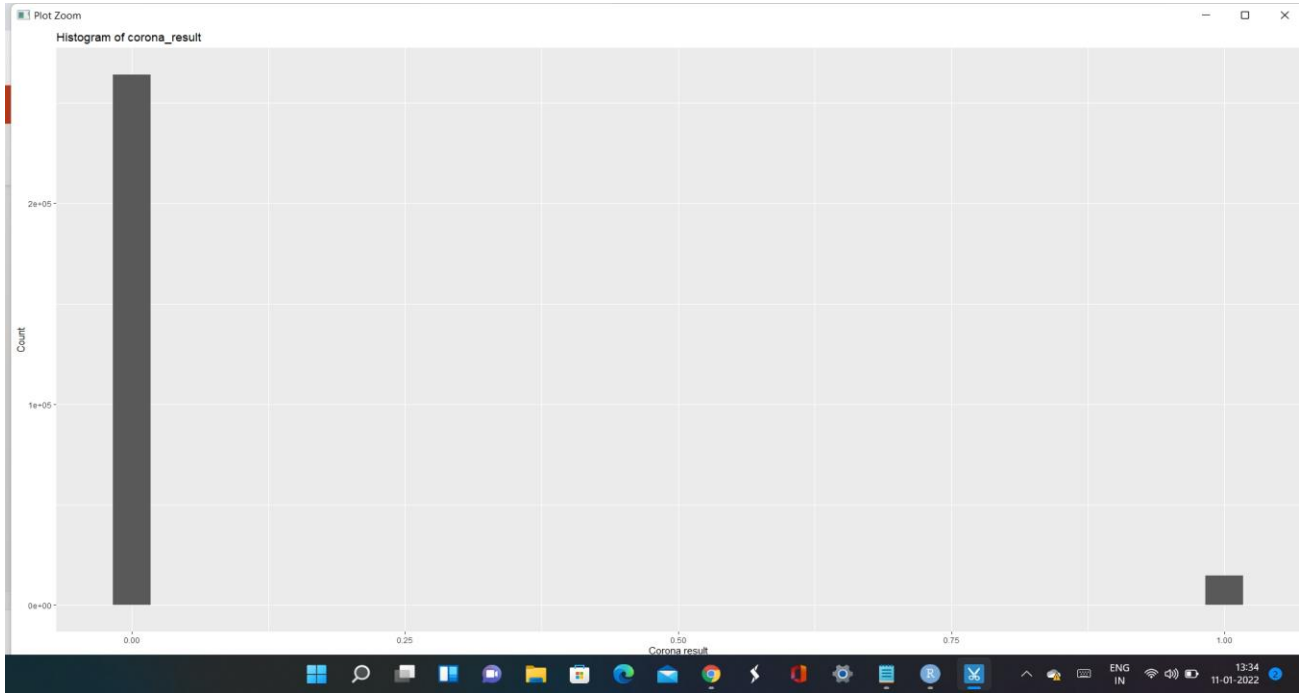
Since,all our values are either 0 or 1 in our dataset,scatter plots are not useful.



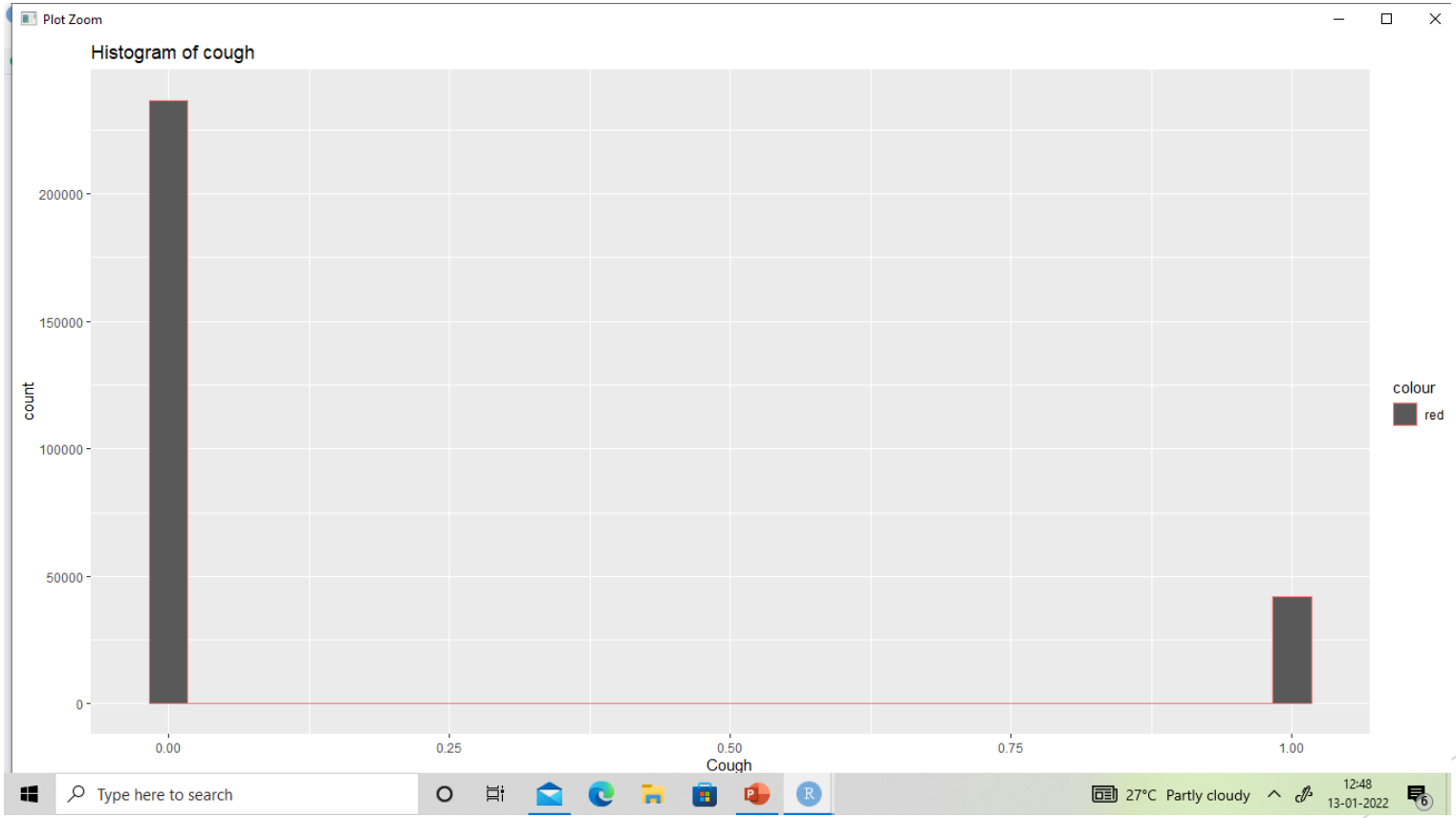
- HISTOGRAM:
- (For variable corona result)

```
>ggplot(my_data1,aes(x=corona_result))+
  geom_histogram(aes(fill=head_ache))+
  xlab("Corona result")+
  ylab("Count")+
   ggtitle("Histogram of corona_result")
```

We can see that ,most of the people's corona_Result is 0=>negative.



- HISTOGRAM : (For variable cough)
- ggplot(data=my_data1,aes(x=cough))+
- geom_histogram(aes(col="red"))+
- xlab("Cough")+
- ylab("count")+
- ggtitle("Histogram of cough")



STATISTICAL ANALYSIS :

For hypothesis testing , we need these packages:(“tidyverse”,”ggpubr”,”rstatix”)

Installing the packages:

```
install.packages("tidyverse")

library(tidyverse)

install.packages("ggpubr")

ibrary(ggpubr)

install.packages("rstatix")

library(rstatix)
```

- 1.T-TEST
- Single-t-test
- i) H0:mean of cough column is 0
- t.test(my_data1$cough,mu=0,alt="two.sided",var.equal = F,conf.level = 0.95)
- INTERPRETATION:
- From the analysis ,we get p< 2/2e-16 <0.05 =>reject H0=>mean of cough column is not 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> t.test(my_data1$cough,mu=0,alt="two.sided",var.equal = F,conf.level = 0.95)

        One Sample t-test

data:  my_data1$cough
t = 223.08, df = 278847, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.1501068 0.1527679
sample estimates:
mean of x
0.1514373

>
```

Type here to search

- ii) H0:mean of fever column is 0
- t.test(my_data1$fever,mu=0,alt="two.sided",var.equal = F,conf.level = 0.95)
- INTERPRETATION:
- From the analysis , we get p<2.2e-16 < 0.05 => H0 is rejected => mean of fever column is not 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> t.test(my_data1$fever,mu=0,alt="two.sided",var.equal = F,conf.level = 0.95)

        One Sample t-test

data:  my_data1$fever
t = 153.6, df = 278847, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.07701126 0.07900205
sample estimates:
 mean of x
0.07800666

>
```

Type here to search

- iii) H0:mean of sore throat column is 0
- t.test(my_data1$sore_throat,mu=0,alt="two.sided",var.equal = F , conf.level = 0.99)
- INTERPRETATION:
- From the analysis ,we get p<2.2e-16 < 0.05 => H0 is rejected => mean of sore throat column is not 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> t.test(my_data1$sore_throat,mu=0,alt="two.sided",var.equal = F,conf.level = 0.99)

        One Sample t-test

data:  my_data1$sore_throat
t = 44.038, df = 278847, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 0.006502993 0.007310984
sample estimates:
  mean of x
0.006906989

>
```

Type here to search

- iv)H0:mean of shortness of breath column is 0
- t.test(my_data1$shortness_of_breath,mu=0,alt="two.sided",var.equal = F,conf.level = 0.95)
- INTERPRETATION:
- From the analysis ,we get p<2.2e-16 < 0.05 => H0 is rejected => mean of shortness of breath column is not 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> #iv)H0:mean of shortness of breath column is 0
> t.test(my_data1$shortness_of_breath,mu=0,alt="two.sided",var.equal = F,conf.level =
 0.95)

        One Sample t-test

data:  my_data1$shortness_of_breath
t = 39.824, df = 278847, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.005377076 0.005933746
sample estimates:
  mean of x
0.005655411

>
```

Type here to search

- PAIRED t-test
- i)H0:Difference in mean of cough and fever is 0
- t.test(my_data1$cough,my_data1$fever,mu=0,alt="two.sided",var.equal = F,paired=T,conf.level=0.99)
- INTERPRETATION:
- From the analysis ,we get t = 115.32, df = 278847, p-value < 2.2e-16 which is less than 0.05.
- So H0 rejected ,alternative hypothesis: true difference in means is not equal to 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> #i)H0:Difference in mean of cough and fever is 0
> t.test(my_data1$cough,my_data1$fever,mu=0,alt="two.sided",var.equal = F,paired=T,con
f.level=0.99)

        Paired t-test

data:  my_data1$cough and my_data1$fever
t = 115.32, df = 278847, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 0.07179051 0.07507086
sample estimates:
mean of the differences
             0.07343069

>
```

Type here to search

- ii)H0: Difference in Mean of cough and shortness of breath is equal to 0
- t.test(my_data1$cough,my_data1$shortness_of_breath,mu=0,alt="two.sided",var.equal = F,paired=T,conf.level=0.99)
- INTERPRETATION:
- From the analysis ,we get = 214.84, df = 278847, p-value < 2.2e-16 which is less than 0.05.
- So H0 rejected ,alternative hypothesis: true difference in means is not equal to 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> #ii)H0: Difference in Mean of cough and shortness of breath is equal to 0
> t.test(my_data1$cough,my_data1$shortness_of_breath,mu=0,alt="two.sided",var.equal =
 F,paired=T,conf.level=0.99)

        Paired t-test

data:  my_data1$cough and my_data1$shortness_of_breath
t = 214.84, df = 278847, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 0.1440341 0.1475298
sample estimates:
mean of the differences
               0.1457819

>
```

Type here to search

- iii)H0: Difference in Mean of cough and corona result is equal to 0
- t.test(my_data1$cough,my_data1$corona_result,mu=0,alt="two.sided",var.equal = F,paired=T,conf.level=0.99)
- INTERPRETATION:
- From the analysis ,we get t = 135.68, df = 278847, p-value < 2.2e-16 which is less than 0.05.
- So H0 rejected ,alternative hypothesis: true difference in means is not equal to 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> #iii)H0: Difference in Mean of cough and corona result is equal to 0
> t.test(my_data1$cough,my_data1$corona_result,mu=0,alt="two.sided",var.equal = F,paire
d=T,conf.level=0.99)

        Paired t-test

data:  my_data1$cough and my_data1$corona_result
t = 135.68, df = 278847, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 0.09674425 0.10048866
sample estimates:
mean of the differences
             0.09861645

>
```

Type here to search

- iv)H0: Difference in Mean of cough and headache is equal to 0
- t.test(my_data1$cough,my_data1$head_ache,mu=0,alt="two.sided",var.equal = F,paired=T,conf.level=0.99)
- INTERPRETATION:
- From the analysis ,we get t = 209.63, df = 278847, p-value < 2.2e-16 which is less than 0.05.
- So H0 rejected ,alternative hypothesis: true difference in means is not equal to 0

Console Terminal Jobs

R 4.1.2 · ~/

```
> #iv)H0: Difference in Mean of cough and headache is equal to 0
> t.test(my_data1$cough,my_data1$head_ache,mu=0,alt="two.sided",var.equal = F,paired=T,
conf.level=0.99)

        Paired t-test

data:  my_data1$cough and my_data1$head_ache
t = 209.63, df = 278847, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
99 percent confidence interval:
 0.1410259 0.1445347
sample estimates:
mean of the differences
              0.1427803

>
```

Type here to search

- 2.CORRELATION:
- For performing correlation we need these packages:(“devtools”,”ggpubr”)
- Installing the packages:
- install.packages("devtools")
- library(devtools)
- install.packages("ggpubr")
- library(ggpubr)
- 1.Testing correlation for fever and cough using pearson Method
- H0:there is no correlation btw fever and cough
- cor(my_data1$fever,my_data1$cough,method="pearson")
- INTERPRETATION:
- we get rho=0.454386 which is not equal to 0.so H0 rejected=>there is correlation bw fever and cough
- 2.Testing correlation for fever and sore throat using pearson Method
- H0:there is no correlation bw fever and sore throat
- cor(my_data1$fever,my_data1$sore_throat,method="pearson")
- INTERPRETATION:
- From the analysis ,we get rho=0.1228317 which is not equal to 0.so H0 rejected=>there is correlation btw fever and sore_throat

```
 0.1410259 0.1445347
sample estimates:
mean of the differences
              0.1427803

> #H0:there is no correlation bw fever and cough
> cor(my_data1$fever,my_data1$cough,method="pearson")
[1] 0.454386
> #we get rho=0.454386 which is not equal to 0.so H0 rejected=>there is correlation bw
 fever and cough
> #H0:there is no correlation bw fever and sore throat
> cor(my_data1$fever,my_data1$sore_throat,method="pearson")
[1] 0.1228317
>
```

Performing correl Method Using agricolae package:

```
install.packages("agricolae")

library(agricolae)
```

Correlation using correl Method:

H0:there is no correlation btw fever and headache

- correl(my_data1$fever,my_data1$head_ache,method="pearson")
- INTERPRETATION:
- From the analysis ,we get rho=0.1688408 which is not equal to 0.so H0 rejected=>there is correlation btw fever and headache
- Prforming cor.test Using stats package:
- install.packages("stats")
- library(stats)
- H0:there is no correlation bw fever and corona result
- cor.test(my_data1$fever,my_data1$corona_result,method="pearson",conf.level = 0.95)
- INTERPRETATION:
- From the analysis ,we get cor=0.2636491 which is not equal to 0.so H0 rejected=>There is correlation between fever and corona_result

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Final_Rproject.R | Final_Rproject[329].R* | c | cor | my_data1 | corona_tested_datset

```r
#Using agricolae package:
install.packages("agricolae")
library(agricolae)
#H0:there is no correlation bw fever and headache
correl(my_data1$fever,my_data1$head_ache,method="pearson")
#we get rho=0.1688408 which is not equal to 0.so H0 rejected=>there is correlatic
#Using stats package:
install.packages("stats")
```

147:1 (Top Level) R Script

Console | Terminal | Jobs

R 4.1.2 · ~/

```
> correl(my_data1$fever,my_data1$head_ache,method="pearson")
$stat
[1] 90.45651

$rho
[1] 0.1688408

$pvalue
[1] 0

> #H0:there is no correlation bw fever and corona result
> cor.test(my_data1$fever,my_data1$corona_result,method="pearson",conf.level = 0.95)

        Pearson's product-moment correlation

data:  my_data1$fever and my_data1$corona_result
t = 144.33, df = 278846, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2601921 0.2670994
sample estimates:
      cor
0.2636491

>
```

Type here to search

- 3.ANOVA
- one-way anova
- i)H0:head_ache has no impact on corona_result
- aov1<-aov(my_data1$corona_result~my_data1$head_ache)
- summary(aov1)
- INTERPRETATION:
- From the analysis ,we get p<2e-16<0.05=>H0 rejected=>head_ache has impact on corona_result
- ii)H0:cough has no impact on corona_Result
- aov2<-aov(my_data1$corona_result~my_data1$cough)
- summary(aov2)
- INTERPRETATION:
- From the analysis , we get p<2e-16<0.05=>H0 rejected=>cough has impact on corona_result
- iii)H0:fever has no impact on corona_result
- aov3<-aov(my_data1$corona_result~my_data1$fever)
- summary(aov3)
- INTERPRETATION:
- From the analysis , we get p<2e-16<0.05=>H0 rejected=>fever has impact on corona_result

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Final_Rproject.R | Final_Rproject[329].R* | C | cor | my_data1 | corona_tested_datset

```r
summary(aov1)
#we get p<2e-16<0.05=>H0 rejected=>head_ache has impact on corona_result
#ii)H0:cough has no impact on corona_Result
aov2<-aov(my_data1$corona_result~my_data1$cough)
```

162:1 (Top Level) R Script

Console | Terminal | Jobs

R 4.1.2 · ~/

```
> #3.anova
> #one-way anova
> #i)H0:head_ache has no impact on corona_result
> aov1<-aov(my_data1$corona_result~my_data1$head_ache)
> summary(aov1)
                     Df Sum Sq Mean Sq F value Pr(>F)
my_data1$head_ache    1   1856    1856   42789 <2e-16 ***
Residuals        278846  12095       0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #we get p<2e-16<0.05=>H0 rejected=>head_ache has impact on corona_result
> #ii)H0:cough has no impact on corona_Result
> aov2<-aov(my_data1$corona_result~my_data1$cough)
> summary(aov2)
                 Df Sum Sq Mean Sq F value Pr(>F)
my_data1$cough    1    529   528.9   10988 <2e-16 ***
Residuals    278846  13422     0.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #we get p<2e-16<0.05=>H0 rejected=>cough has impact on corona_result
> #iii)H0:fever has no impact on corona_result
> aov3<-aov(my_data1$corona_result~my_data1$fever)
> summary(aov3)
                 Df Sum Sq Mean Sq F value Pr(>F)
my_data1$fever    1    970   969.7   20831 <2e-16 ***
Residuals    278846  12981     0.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Environment | History | Connections | Tutorial

Import Dataset · 299 MiB · List

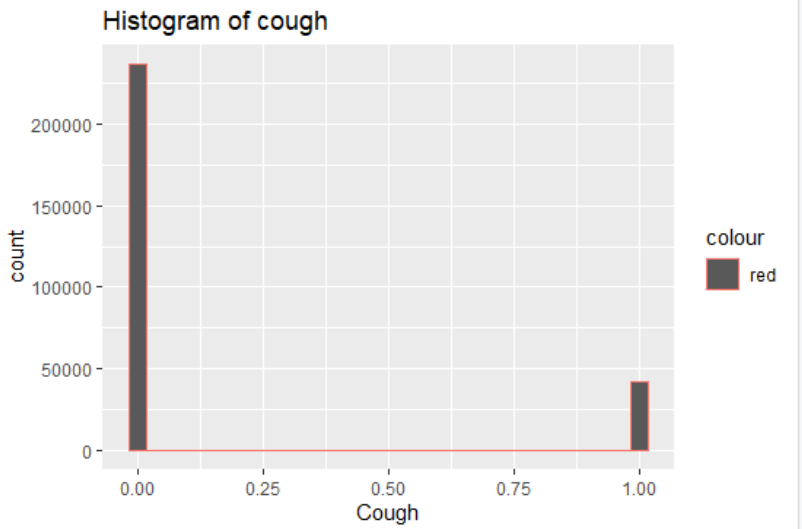
R · Global Environment

Data

aov1	Large aov (12 elements, 71.4 MB)
aov2	Large aov (12 elements, 71.4 MB)
aov3	Large aov (12 elements, 71.4 MB)
avo1	List of 14
custData	num [1:620, 1:3] 1.269 1.269 1.183 0.924 0...
data	num [1:620, 1:3] 1.269 1.269 1.183 0.924 0...
decomposeAP	List of 6

Files | Plots | Packages | Help | Viewer

Zoom · Export · Publish



Type here to search

27°C Partly cloudy 14:52 13-01-2022

- 2 way anova
- i)H0:fever and sore_throat has no impact on cough
- str(my_data1)
- aov4<-aov(cough~fever+sore_throat,data=my_data1)
- aov4
- summary(aov4)
- INTERPRETATION:
- From the analysis , we get p<2e-16<0.05,so H0 rejected=>fever and sore_throat has impact on cough
- ii)H0:fever,sore_throat,head_Ache,shortness_of_breath and cough together have no impact on corona_result
- aov5<-aov(corona_result~fever+sore_throat+head_ache+shortness_of_breath+cough,data=my_data1)
- summary(aov5)
- INTERPRETATION:
- From the analysis , we get p<2e-16<0.05,so H0 rejected=>fever, sore_throat, head_Ache, shortness_of_breath and cough together have impact on corona_result

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Final_Rproject.R | Final_Rproject[329].R* | C | cor | my_data1 | corona_tested_datset

```r
# 2 way anova
#i)H0:fever and sore_throat has no impact on cough
str(my_data1)
aov4<-aov(cough~fever+sore_throat,data=my_data1)
aov4
summary(aov4)
#we get p<2e-16<0.05,so H0 rejected=>fever and sore_throat has impact on cough
#ii)H0:fever,sore_throat,head_Ache,shortness_of_breath and cough together have no impact
aov5<-aov(corona_result~fever+sore_throat+head_ache+shortness_of_breath+cough,data=my_da
summary(aov5)
```

174:1 (Top Level) R Script

Console | Terminal | Jobs

R 4.1.2 · ~/

```
> aov4<-aov(cough~fever+sore_throat,data=my_data1)
> summary(aov4)
               Df Sum Sq Mean Sq F value Pr(>F)
fever           1   7398    7398   72885 <2e-16 ***
sore_throat     1    130     130    1283 <2e-16 ***
Residuals  278845  28305       0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> #we get p<2e-16<0.05,so H0 rejected=>fever and sore_throat has impact on cough
> #ii)H0:fever,sore_throat,head_Ache,shortness_of_breath and cough together have no impact on c
orona_result
> aov5<-aov(corona_result~fever+sore_throat+head_ache+shortness_of_breath+cough,data=my_data1)
> summary(aov5)
                      Df Sum Sq Mean Sq F value Pr(>F)
fever                  1    970   969.7   24731 <2e-16 ***
sore_throat            1    839   838.7   21389 <2e-16 ***
head_ache              1    957   957.1   24407 <2e-16 ***
shortness_of_breath    1    208   207.6    5295 <2e-16 ***
cough                  1     44    43.8    1116 <2e-16 ***
Residuals         278842  10934     0.0
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Environment | History | Connections | Tutorial

Import Dataset · 383 MiB · List

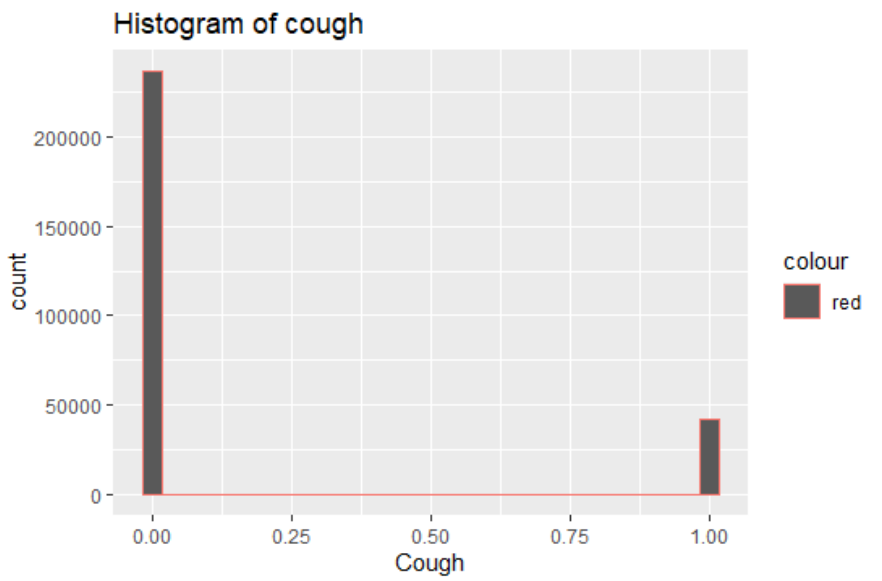
R · Global Environment

Data

aov1	Large aov (12 elements, 71.4 MB)
aov2	Large aov (12 elements, 71.4 MB)
aov3	Large aov (12 elements, 71.4 MB)
aov4	Large aov (12 elements, 75.9 MB)
aov5	Large aov (12 elements, 89.3 MB)
avo1	List of 14

Files | Plots | Packages | Help | Viewer

Zoom · Export · Publish



Type here to search

27°C Partly cloudy 15:00 13-01-2022

Since our dataset contains class label, we perform classification.

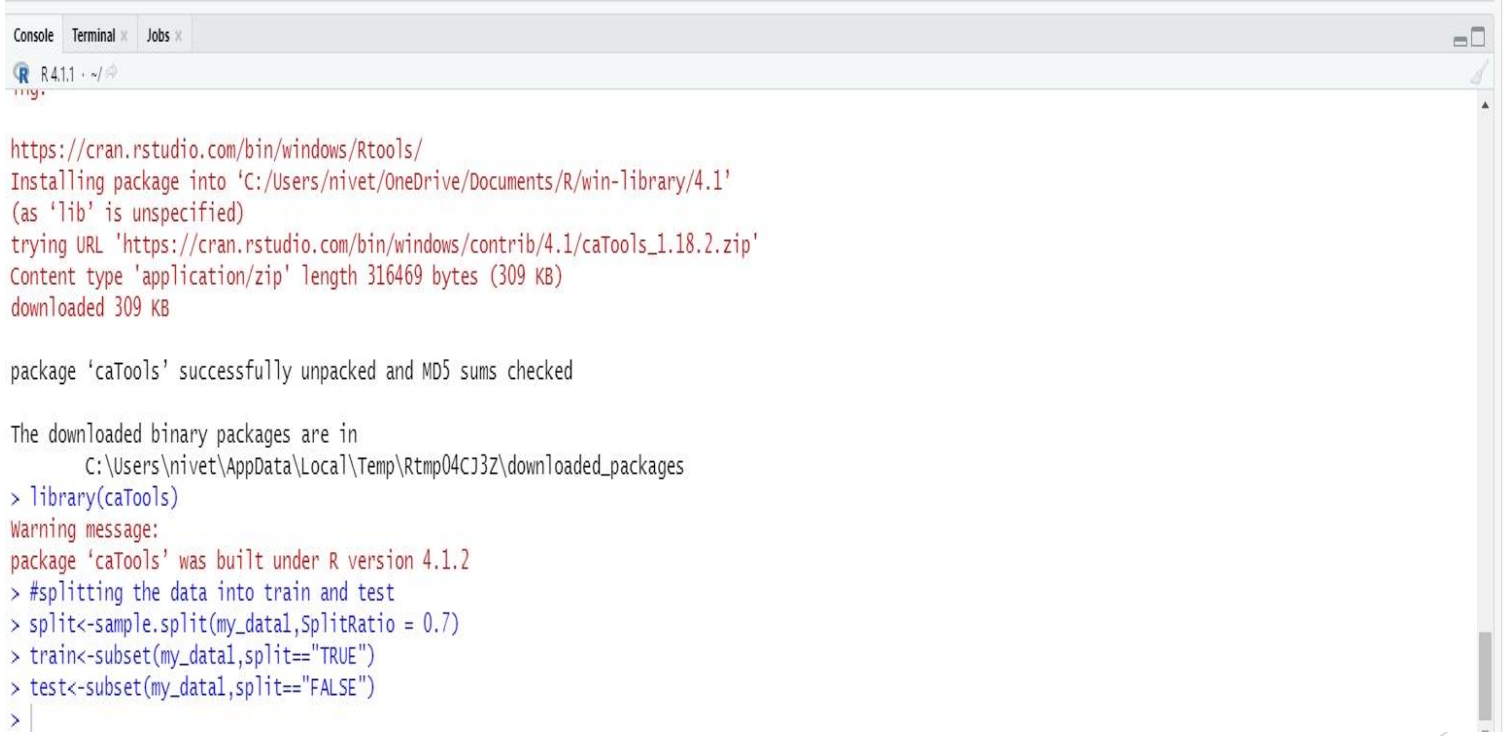
LOGISTIC REGRESSION:

Install caTools package:

```
>install.packages("caTools")
>library(caTools)
```

Split the dataset into train and test:

```
>split<-sample.split(my_data1,SplitRatio = 0.7)
>train<-subset(my_data1,split=="TRUE")
>test<-subset(my_data1,split=="FALSE")
```



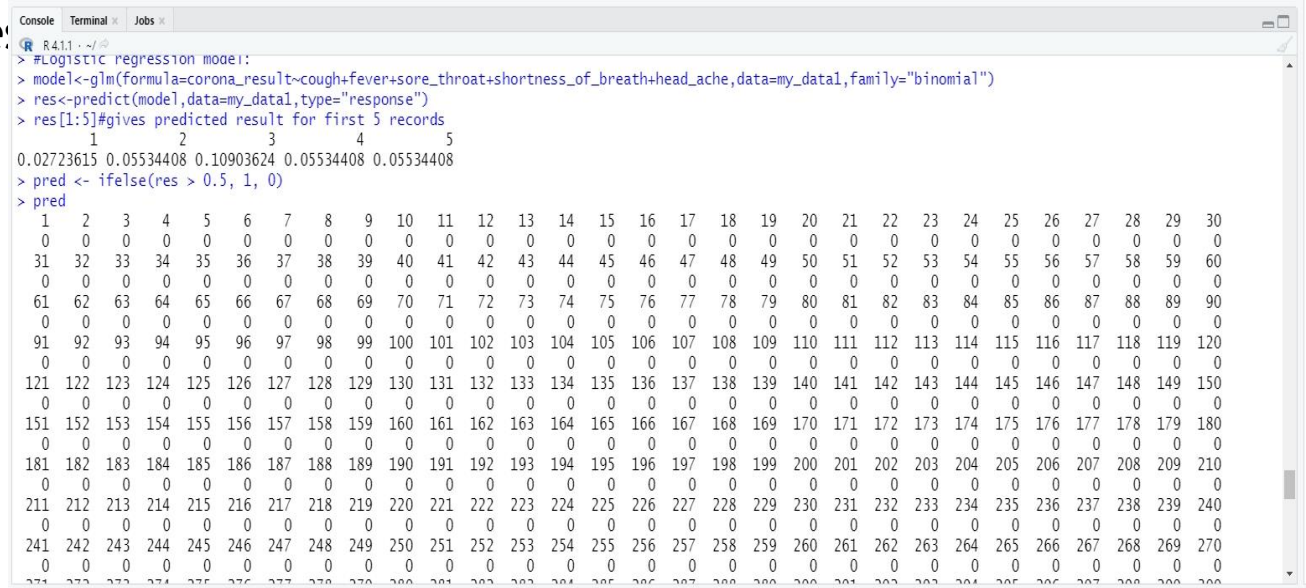
We build the logistic regression model w.r.t the dependent variable corona_result and independent variables-cough+fever+sore_throat+shortness_of_breath+head_ache.
i.e corona_result will depend upon cough,fever,sore_throat,shortness_of_breath and head_Ache.

Use data to be entire data and give family attribute to be "binomial"=>gives 1 out of 2 results only.

```
>model<-
glm(formula=corona_result~cough+fever+sore_throat+shortness_of_breath+head_ache,data=my_d
ata1,family="binomial")
```

Use predict() method to predict the value based on the model.

```
>res<-predict(model,data=my_data1,type="response")
>res[1:5]#gives predicted result for first 5 records
>pred <- ifelse(res > 0.5, 1, 0)#gives 1 if res>0.5 else gives 0
>pred #give
```



```
Console   Terminal   Jobs
R 4.1.1 · ~/
> #Logistic regression model:
> model<-glm(formula=corona_result~cough+fever+sore_throat+shortness_of_breath+head_ache,data=my_data1,family="binomial")
> res<-predict(model,data=my_data1,type="response")
> res[1:5]#gives predicted result for first 5 records
        1          2          3          4          5
0.02723615 0.05534408 0.10903624 0.05534408 0.05534408
> pred <- ifelse(res > 0.5, 1, 0)
> pred
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
 31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
 61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
 91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
```

FINDING CONFUSION MATRIX:

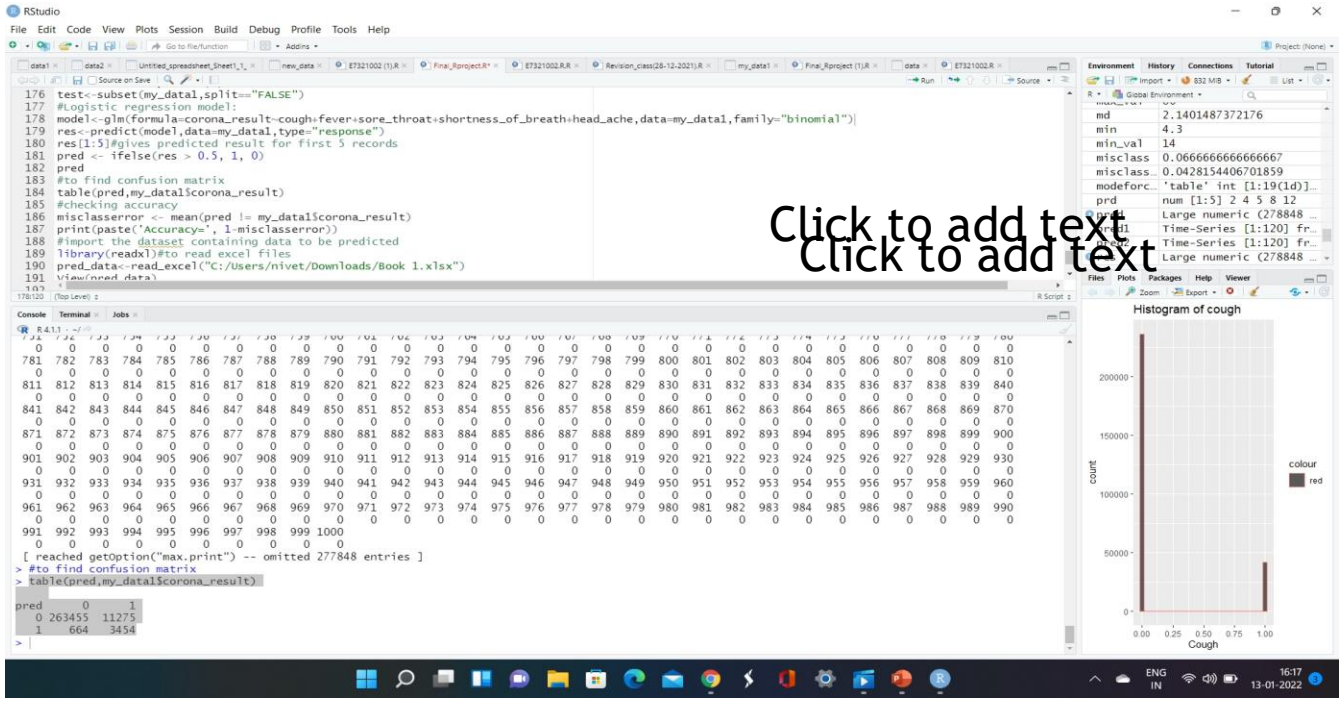
Confusion matrix tells you how much values are predicted correctly.

Method used:

table(Predicted_Value,Actual_Value)

Code:

```
>table(pred,my_data1$corona_result)
```



INTERPRETATION:

We can see that 263455 values are correctly predicted as 0 to be 0 and 3454 values are predicted 1 to be 1 correctly.

CHECKING ACCURACY PF PREDICTION:

First find misclasserror,which is got by taking mean of predicted values which are not equal to the actual value.

```
>misclasserror <- mean(pred != my_data1$corona_result)
```

Accuracy is got by subtracting misclasserror from 1.

```
>print(paste('Accuracy=', 1-misclasserror))
```

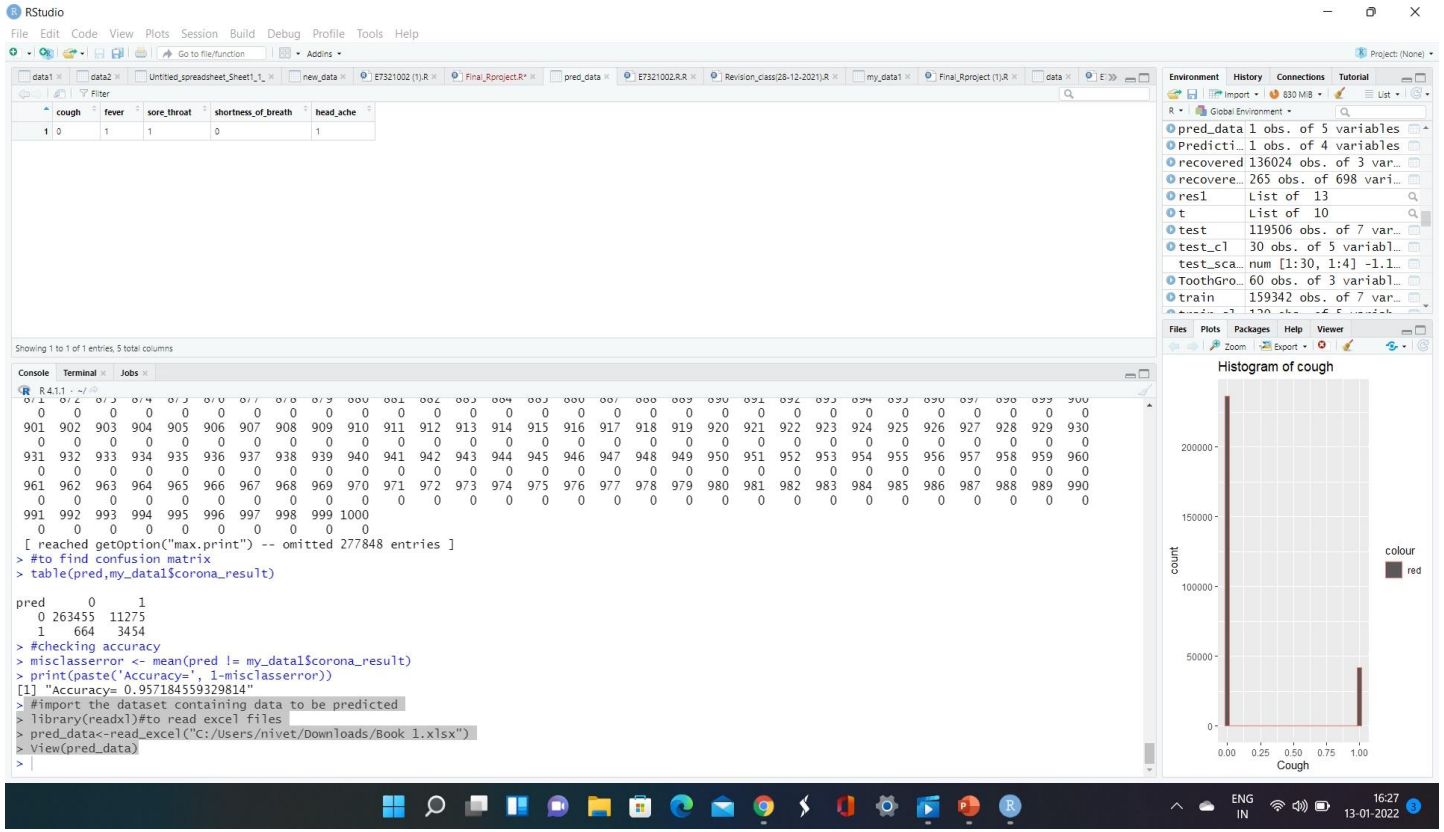
```
Console   Terminal ×   Jobs ×
R 4.1.1 · ~/
811  812  813  814  815  816  817  818  819  820  821  822  823  824  825  826  827  828  829  830  831  832  833  834  835  836  837  838  839  840
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
841  842  843  844  845  846  847  848  849  850  851  852  853  854  855  856  857  858  859  860  861  862  863  864  865  866  867  868  869  870
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
871  872  873  874  875  876  877  878  879  880  881  882  883  884  885  886  887  888  889  890  891  892  893  894  895  896  897  898  899  900
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
901  902  903  904  905  906  907  908  909  910  911  912  913  914  915  916  917  918  919  920  921  922  923  924  925  926  927  928  929  930
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
931  932  933  934  935  936  937  938  939  940  941  942  943  944  945  946  947  948  949  950  951  952  953  954  955  956  957  958  959  960
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
961  962  963  964  965  966  967  968  969  970  971  972  973  974  975  976  977  978  979  980  981  982  983  984  985  986  987  988  989  990
  0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0    0
991  992  993  994  995  996  997  998  999 1000
  0    0    0    0    0    0    0    0    0    0
 [ reached getOption("max.print") -- omitted 277848 entries ]
> #to find confusion matrix
> table(pred,my_data1$corona_result)

pred      0      1
   0 263455  11275
   1    664   3454
> #checking accuracy
> misclasserror <- mean(pred != my_data1$corona_result)
> print(paste('Accuracy=', 1-misclasserror))
[1] "Accuracy= 0.957184559329814"
>
```

We get accuracy to be 0.957184559329814=>95.7% data are predicted correctly=>Model works very well.

PREDICTING FOR NEW DATA:

Import the dataset containing the new data.The new dataset should not have the dependent variable column.

Based on the independent variables values,the value for the dependent variable will be predicted.



USE predict() method to predict for new data:

Inside predict() method, assign the name of new dataset to "newdata" attribute.

CODE:

```
>predict(model,newdata=pred_data,type="response")
```

OUTPUT:

0.9965705

INTERPRETATION:

We get 0.9965705 which is approximately equal to 1=>The person has been tested covid positive.

```
Console   Terminal ×   Jobs ×
R 4.1.1 · ~/
[ reached getOption("max.print") -- omitted 277846 entries ]
> #to find confusion matrix
> table(pred,my_data1$corona_result)

pred      0      1
   0 263455  11275
   1    664   3454
> #checking accuracy
> misclasserror <- mean(pred != my_data1$corona_result)
> print(paste('Accuracy=', 1-misclasserror))
[1] "Accuracy= 0.957184559329814"
> #import the dataset containing data to be predicted
> library(readxl)#to read excel files
> pred_data<-read_excel("C:/Users/nivet/Downloads/Book 1.xlsx")
> View(pred_data)
> str(pred_data)
tibble [1 x 5] (S3: tbl_df/tbl/data.frame)
 $ cough              : num 0
 $ fever              : num 1
 $ sore_throat        : num 1
 $ shortness_of_breath: num 0
 $ head_ache          : num 1
> predict(model,newdata=pred_data,type="response")#gives predicted value=0.9965705 for new data
        1
0.9965705
>
```

TEAM MEMBERS:

NIVETHA C- E7121015

NIVETHA K- E7321002

MAHA LAKSHMI P- E7321003

BRINDA G- E7321004

KUSHITHA G- E7321007

- THANK YOU