Name- Arpit Gautam Sec- C Std-Id- 21711205

3.) Here, I am taking data set of Titanic ship that contains historical records of all passengers who on-boarded the titanic

Data set contains following variables ÷

- Passenger Id : Serial no.
- Survived : 0 - dead & 1 - survived.
- Pclass - Ticket class - 1st, 2nd or 3rd
- Name - Name of passenger.
- Sex - M / F
- Ticket - Serial No.
- Cabin - Cabin No.

Analysis ÷

1.) View(titanic) ÷ This helps us in familiarsing with data set.

2.) head(titanic, n) / tail(titanic, n) ÷ It arranges first n data either from starting or end. By default value of n is 5.

3.) names(titanic) ÷ This helps us in checking out all the variables in data set.

4.) Str (titanic) ÷ This helps in understanding the structure of data set, data type of each attribute & no. of rows & column present in data.

5.) ~~Arrange~~ Filter → my newdata <- filter (titanic, age > 20)

④ Descriptive Statistics my new data ÷

In order to find descriptive statistics, we use Summary (titanic) command ÷

Passenger Id	Survived	Sex
Min. : 1.0	Min. : 0.0	female : 314
1st Qu. : 223.5	1st Qu. : 0.0	male : 577
Median : 446.0	Median : 0.0	
Mean : 446.0	Mean : 0.38	
3rd Qu. : 668.5	3rd Qu : 1.0	
Max : 891.0	Max : 1.0	

Age	Fare
Min. : 10	Min. : 0.0
1st Qu. : 20.12	1st Qu : 7.91
Median : 28.0	Median : 14.51
Mean : 29.7	Median : 32.56
3rd Qu. : 38.0	3rd Qu : 31.0
Max : 80.0	Max : 512.33

Inferential Statistics

1) Bar graph :- gg plot (titanic , aes (x= survived)) + geom-bar()

*) From Bar graph, it can be infered that only 38.3% of the passengers did survived.

2) Gender Based survivel :- gg plot (titanic , aes (x= sex , fill = survived)) + theme_bw() + geom_bar() + labs (y= " Passengers No.", title = " Survival Rate by Gender")

*) From graph , the survival rate amongst the women was significantly higher when compared to men.

3) Histogram (Survival Rates basis age) :- ggplot (titanic , aes (x= Age, fill = Survived)) + theme_bw() + geom- histogram (binwidth = 5) + labs (y= "No. of Passengers", x= "Age")

*) From graph , we infered that for age < 10 section in graph, the survival rate is high. & the survival rate is low for age beyond 45.