

Name - Ayaz Goharani

Roll no - 2101009 (23)

Course - MCA - (B) 1st sem

Sub: scripting language / R
programming

(Que - 3) (Ans)

We're using the Titanic dataset to analyze:

load data:

```
2) titanic <- read.csv("C:/Users/Desktop/Titanic.csv",  
                        header = TRUE)
```

Peek your data:

`view(titanic)`

This helps us in familiarising with the dataset.

`head(titanic, 10)`

return first 10 rows.

`tail(titanic)`

return Bottom, 10 rows.

`summary(titanic)`

It is one of the most important function that helps in summarising each attribute in the dataset. It gives the descriptive statistics of the data.

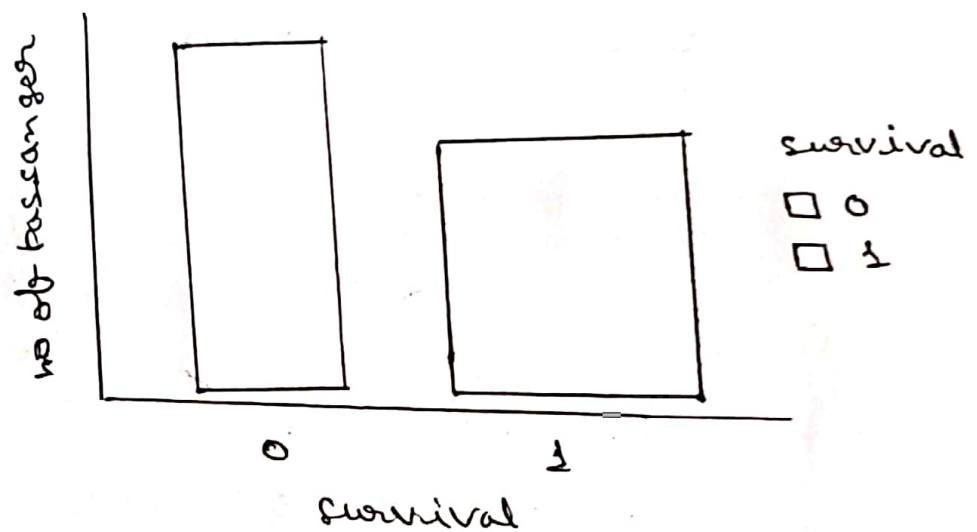
Name - Aalya Goswami

Roll no - 2101009

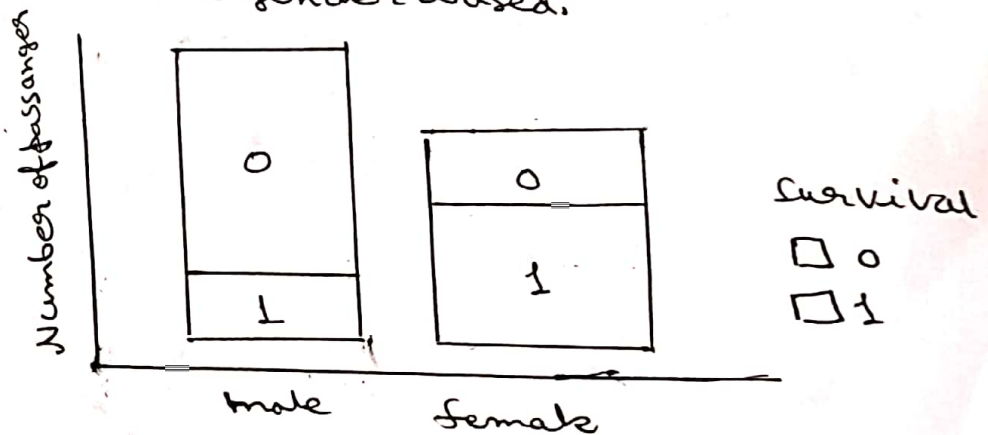
Analysis & Visualization:

Survival rate:

```
ggplot(titanic, aes(x = survived)) + geom_bar()
```



Survival rate gender based:



```
ggplot(titanic, aes(x = sex, fill = survival))  
+ theme_bw() + geom_bar() +  
labs(y = "Number of passengers",  
title = "Survival Rate by Gender").
```

Name - Ajay Chawani

Roll no - 2101009.

(Que 4) Ans)

Descriptive Statics:-

Summary:- Gives us the descriptive sets like

In case of Numerical data:

Gives mean, mode, median, Range.

Measures of Central Tendency.

1) mean (Titanic \$ fare) [on average person spent \$32 to board the ship]
32.20421

2) mode (Titanic \$ Age) [mode common Age on Titanic]
24

3) median (Titanic \$ fare)
14.52

Measure of spread

Range (Titanic \$ fare) [it shows lowest & highest value of fare]
0.000 512.3292 -

Name - Ajay Oswami

Roll no - 2101009

`var(titanic $fare)`

2469.437

`sqrt(var(titanic $fare))`

49.69343

Inferential statistics:

Hypothesis testing:

`new.data <- subset(titanic, $pclass == 1)`

\Rightarrow `test2 = function(a, b) {`

`sample-mean = mean(a)`

`pop-mean = mean(b)`

`c = nrow = (n)`

`varb = var(b)`

`data = (sample-mean, pop-mean) / sqrt(varb/c)`

`return(data)`

Call function :-

`2. test2(newdat $survived, titanic $survived,
 newdata)`

7.423828.