Ans 3) we are using here Titanic dataset to analyze:

Load data:

titanic ← read.csv ("C:/Users/Desktop/titanic.csv", header = TRUE, sto = ";")

Peek at your data:

→ view (titanic)

This help us for familiarizing with data set

→ head (titanic, 1)

return first 10 rows.

→ tail (titanic, 10)

return Bottom, 10 ~~years~~ rows.

→ names (titanic)

This help us in checking out all the variables in the data set

→ Summary (titanic)

It is one of the most important functions that help in summarizing each attribute in the dataset.

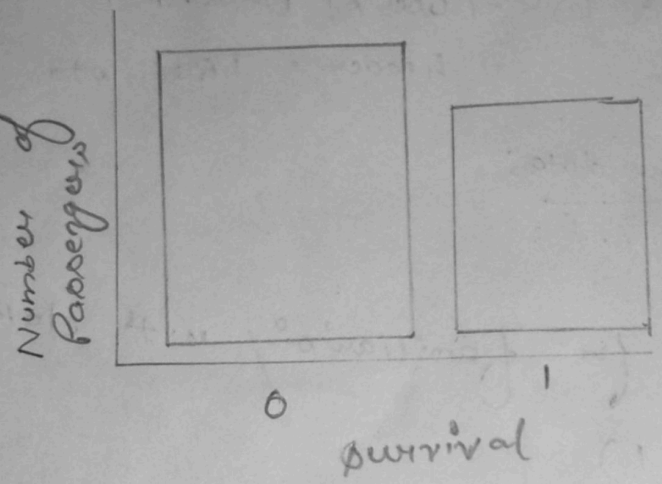
It gives the descriptive statistics of the data.

Rawat

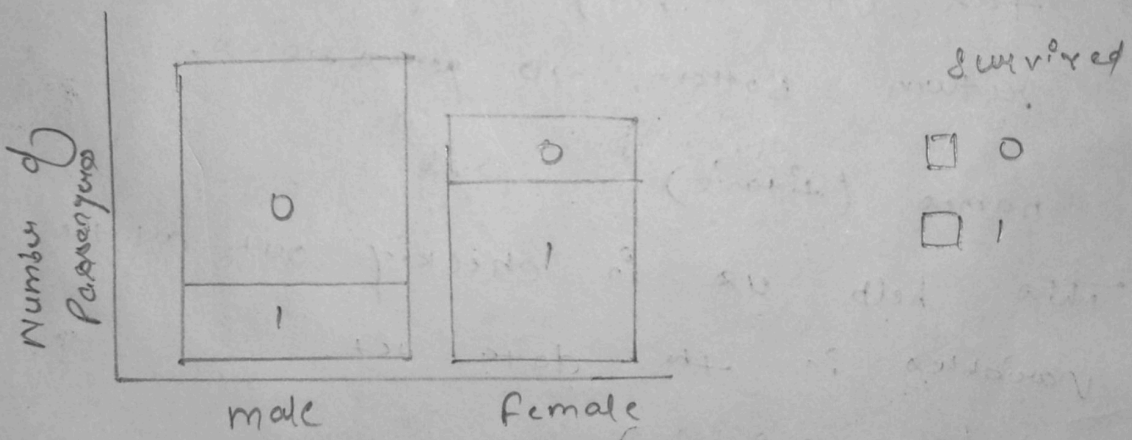
Analysis & Visualization

Survival Rate:

```
gg plot (titanic, aes(x = survived)) + geom.bar()
```



Survival Rate based gender:




```
gg plot ( titanic, aes ( x = sex, fill = survived )) +
        theme_bw () + geom_bar () +
        labs (y = "Number of passengers",
        title = "Survived Rate by Gender")
```

Rawat

Ans 4)

Descriptive Statics:

Summary: Gives us the descriptive acts like

In case of Numeral data:

Gives mean, mode, median, range

Measures of Central Tendency

=) mean (titanic $ fare) [on Average person spent $ 32 to board the titanic]

32.2042

=) mode (titanic $ Age) [mode common Age on titanic]

24

=) median (train $ fare)

14.542

Measure of spread:

range (titanic $ fare) [It shows lowest $ highest value of fare]

0.000 $ 12.3292

Gaut

=) var (titanic $ fare)

2469.437

=) sqrt (var (titanic $ fare)

49.69343

Inferential statistics:

Hypothesis Testing:

new_data - subset (titanic, $ pclass == 1)

=) test 2 = function (a, b, a){

sample.mean = mean(a)

pop.mean = mean (b)

c = nrow= (n)

varb = var (b)

data = (sample_mean, pop_mean)/sqrt (var.blc).

return data

Call function:

z.test 2 (newdata $ survived, titanic $ survived,

newdata)

7.423828

Ans