Name Ishika Samant   Sign - Ishika

Univ. Roll№ 2101005
Gur.   PV- 21010005
Date   15/03/22
Code   PMC 103
Cours.   MCA   Sec D

Q3

We are using Titanic dataset to analyze

<u>Load data</u>

```
titanic <- read.csv("C:/users/Desktop./titanic.csv", header =
                                      True, sep = ",")
```

<u>Peek at your data</u>

= View (titanic)

This help us to familize with the data set.

= head (titanic, 10)

return first 10 rows

- tail (titanic, 10)

return Bottom , 10, rows

= names (titanic)

This helps us in checking out all the variables in the data set.
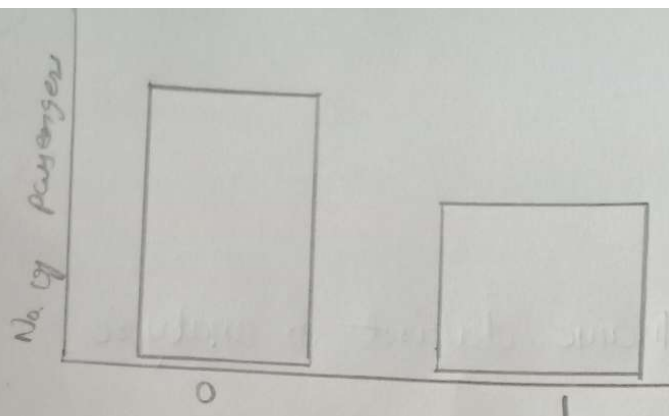
Summary (titer)

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

It gives the descriptive statistics of the data
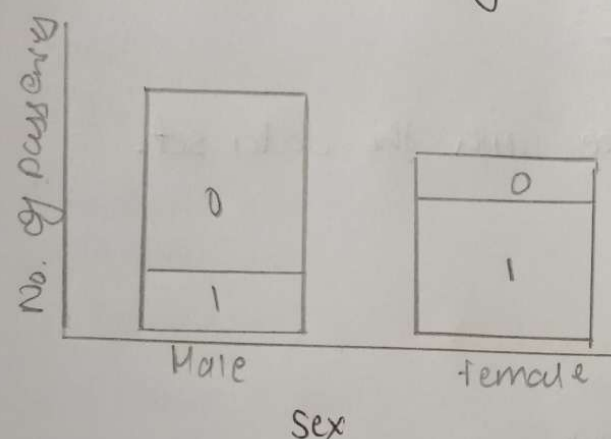
<u>Analysis   of   Visulizations</u>

• <u>Survival rate</u>

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

(y-axis: No. of Passengers)

Survival
☐ 0
☐ 1

0   1

Survived

Survival rate based genders –

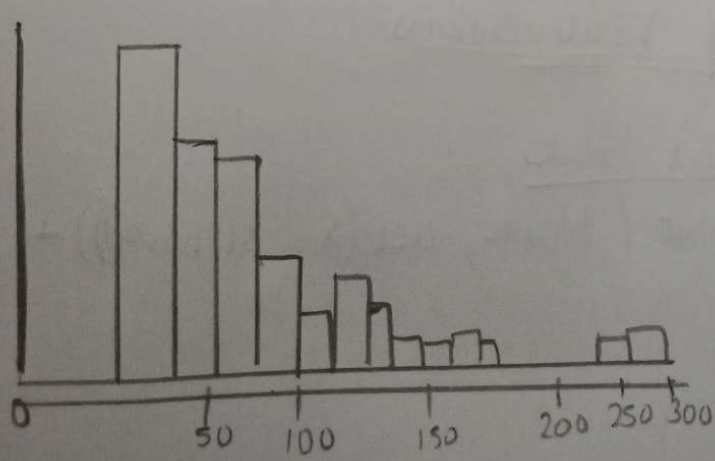(y-axis: No. of Passengers)

Male   female

Sex

Survival
☐ 0
☐ 1

ggplot(titanic, aes(x = sex, fill = survival)) +
    theme_bw() + geombar() + labs(y = 'Number of Passengers',
    title = "Survival Rate by gender")

Distributing of fare rate
hist(titanic $ fare, main = "fare Per person", x lab = "fare",
    col = 'grey', breaks = 40, xlim = c(0, 300))

Fare Per Person

50   100   150   200 250 300

Answer

1) Answer

<u>Descriptive Statistics</u>

<u>Summary</u> — Gives us the descriptive stats like

In case of Numerical data :—
Gives Mean, Mode Median, Range

<u>Measure of Central Trending</u> :—

= mean (titanic $ fare)   [an average person spent $ 32
32.20421   to board the titanic]

= mode (titanic $ Age)   [most common age on
= 24   titanic]

= median (train $ fare)
14.542

<u>Measure of Spread</u> :—

→ Arrange (titanic $ fare)   [It shows lowest and
0.000   512.3292   highest value of fare]

⇒ van (titanic $ fare)
2469.437

= sqrt van (titanic $ fare))
49.69343

<u>Inferential Statistics</u> :—

— Hypothesis Testing :—

new_data <— subset (titanic, titanic $ pclass == 1).

= .test 2 — function (a, b, n) {

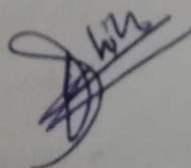samplemean = mean(a)
pop_mean = mean (b)
c = nrow(n)
van_b = van(b)
zeta = (sample-mean-pop-mean) / sqrt(van_b/c))
return zeta.

call function
2. test2 (new_data $ survived, titanic $ survivel, new_data)
    7.423828