

we are seeing here Titanic dataset to analyse
Load data:

```
titanic <- read.csv('c:/user/Desktop/titanic.csv',  

  header = TRUE, as.is = TRUE)
```

Peek at your data

→ view(titanic)

This helps us to familiarise 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.

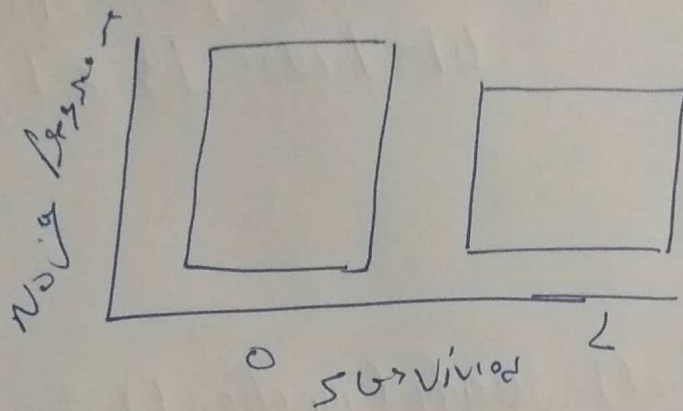
→ summary(titanic)

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.

Analysis & Visualization

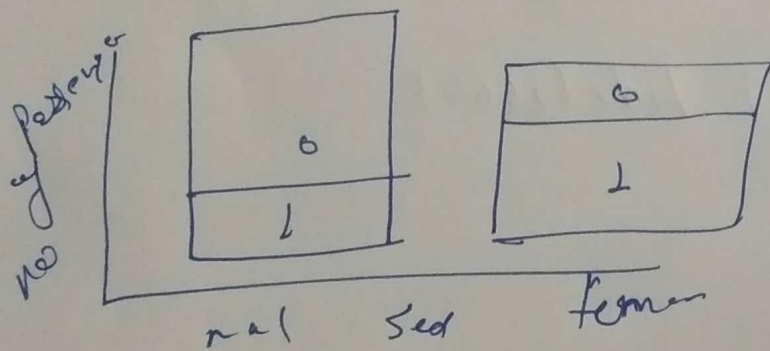
Survival Rate

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



Survived
 0
 1

Survival Rate by Gender

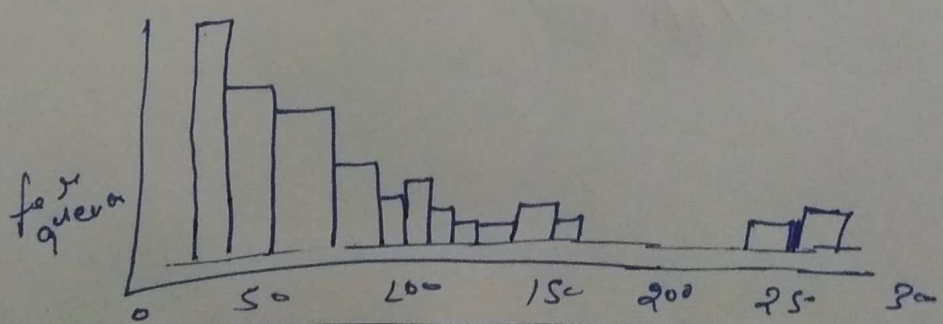


Survived
 0
 1

```
ggplot(titanic, aes(x=Gender, fill=Survived)) +  
  theme_bw() + geom_bar() +  
  labs(y="Number of Passengers",  
        title="Survival Rate by Gender")
```

Distribution of Fare

```
hist(titanic$Fare, main="Fare Distribution", xlab="Fare",  
      col="grey", breaks=40, xlim=c(0, 300))
```



Descriptive Statistics : Descriptive statistics are used to describe the characteristics or features of a dataset. The

term 'descriptive statistics' can be used to describe both individual quantitative observations as well as the overall process of obtaining insight from these data. We can use descriptive statistics to describe both an entire population or an individual sample.

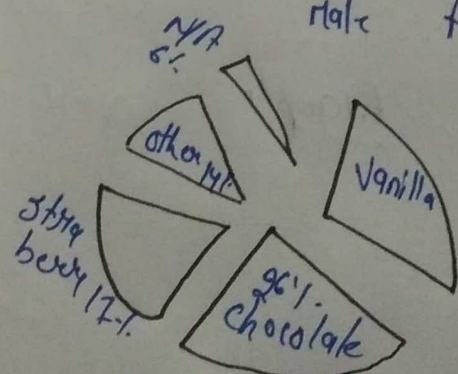
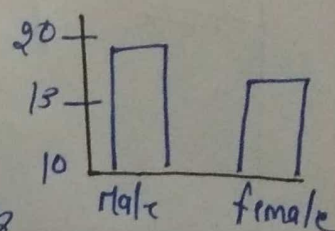
Data can be summarized and represented in an accurate way using charts, tables and graphs.

For example we have marks of 1000 students and we may be interested in the overall performance of those student and the distribution as well as the spread of marks. Descriptive statistics provide us the tools to define our data in a most understandable and appropriate way.

Respondent	Age	Gender	favorite ice cream
1	36	M	Vanilla
2	22	F	chocolate
3	61	M	other
4	88	F	vanilla
5	31	M	vanilla
6	83	F	chocolate
7	14	M	vanilla

Raw data

Age
Mean 42.6
Standard Dev. 21.9



Descriptive statistics

Inferential Statistics : It is about using data from sample and then making inferences about the larger Population from which the sample is drawn. The goal of the inferential statistics is to draw conclusions from a sample and generalize them to the Population. It determines the Probability of the characteristics of the sample using Probability theory. The most common methodologies used are hypothesis tests etc.

for example

Suppose we are interested in the exam marks of all the students in India. But it is not feasible to measure the exam marks of all the student in India. So now we will measure the marks of a small sample of students, for example 2000 students. This sample will now represent the large Population of Indian students. We would consider this sample for our statistical study for studying the Population from which it is deduced.