1. What is statistics
2. Types of statistics
3. Inferential statistics
4. Descriptive statistics
5. What is population and what is sample
6. What is statistic and parameter
7. Types of data
8. Numerical vs categorical
9. Continuous and discrete
10. Levels of data

* Nominal
* Ordinal
* Interval
* Ratio

Data analysis :

1. Numerical data
2. Categorical data

In a class there 40 members girls are there and 50 members boys are there

The above statement you need to represent a tabular format

| Gender | No of students |
| --- | --- |
| Girl | 40 |
| Boys | 50 |

Frequency table:

| Class | Frequency |
| --- | --- |
| Girls | 40 |
| Boys | 50 |

In India the domestic cars are 50 , foreign cars are 30

| class | frequency |
| --- | --- |
| Domestic cars | 50 |
| Foreign | 30 |

For any categorical information the tabular form of data is : Frequency table

In a xyz hospital number of diabetic patients are 100, non diabetic are 200

Y

Yes

No

Yes

No

Yes

No

Yes

No

Q1) how many classes are there

* 2

Q2) what are those classes : yes and no

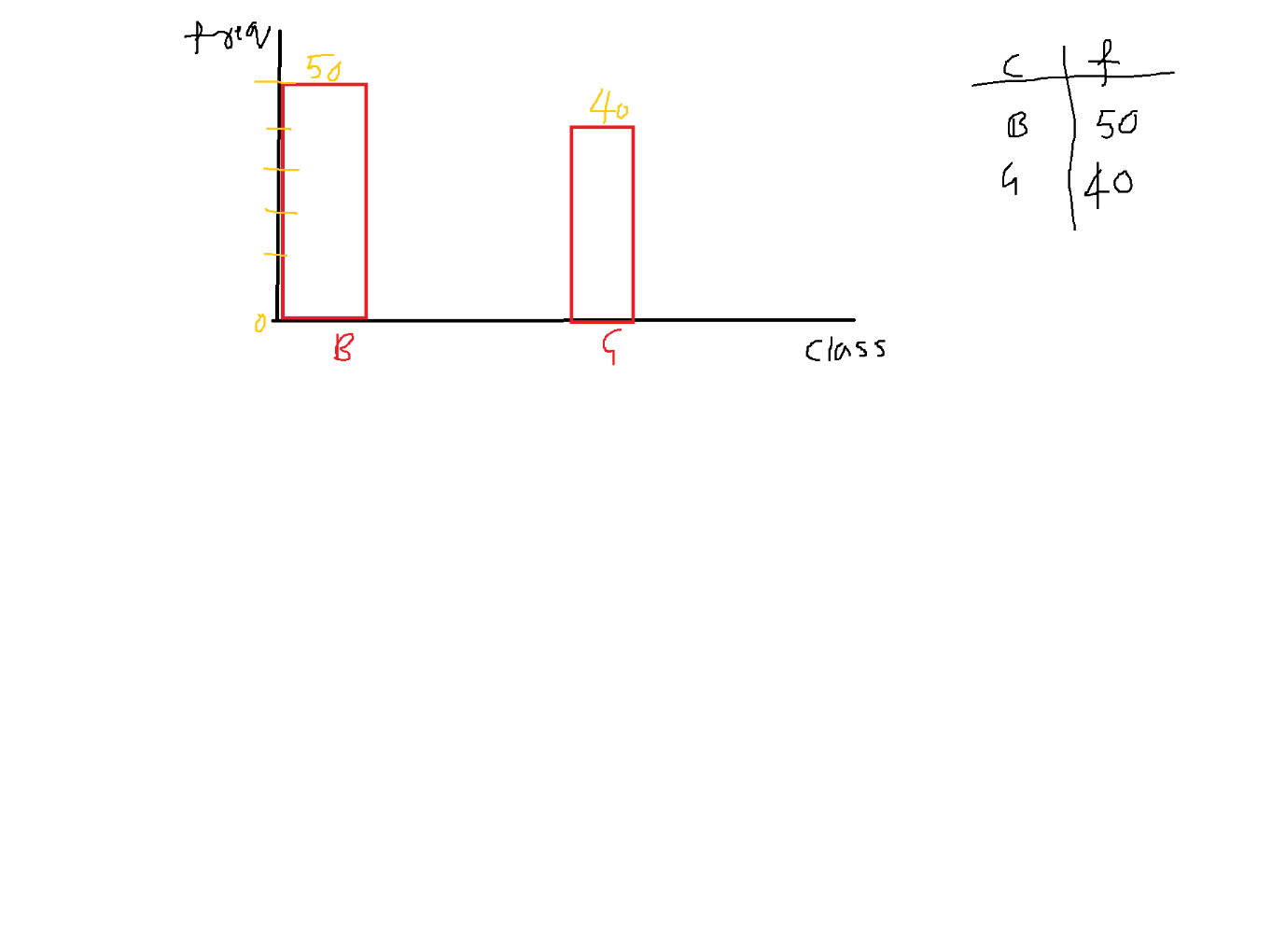
| Class | Frequency |
| --- | --- |
| Yes | 4 |
| No | 4 |

For categorical data types the tabular representation is : Frequency table

Graphical representation:

Boys=========== 50

Girls ========== 40



The above plot is bar chart/bar plot/bar graph

x- axis: class =============== categorical

y-axis: frequency ============ number

if you want to draw a bar plot one column: categorical

another column: Numerical

=========================================

In your class 10 members are there

5 7 8 12 15

3 20 23 30 25 ============ > raw data

Out of these 10 members : 5 Girls 5 boys ========= > frequency table / bar graph / cat vs numer

5 7 8 12 15

3 20 23 30 25

What is the suitable interpretation of above raw data

We can divide data into some intervals

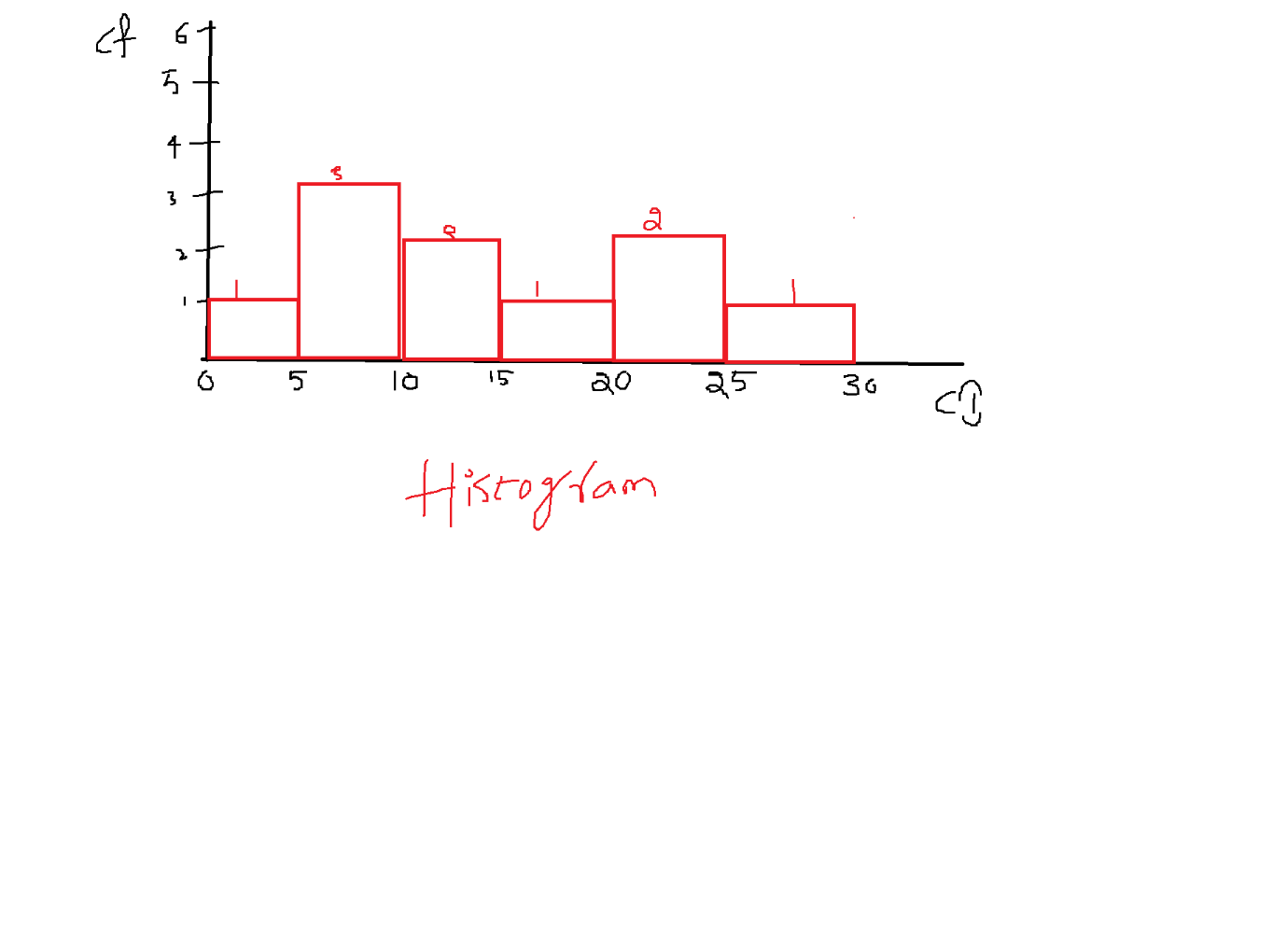
And we can count the number of students fall in those interval

Frequency distribution table

| Class intervals | Class frequency |
| --- | --- |
| 0-5 | 1 |
| 5-10 | 3 |
| 10-15 | 2 |
| 15-20 | 1 |
| 20-25 | 2 |
| 25-30 | 1 |

For any numerical column======= > frequency distribution table

Graphical representation : Histogram



A histogram is an approximate representation of the distribution of numerical data.

Numerical columns represent using Histogram

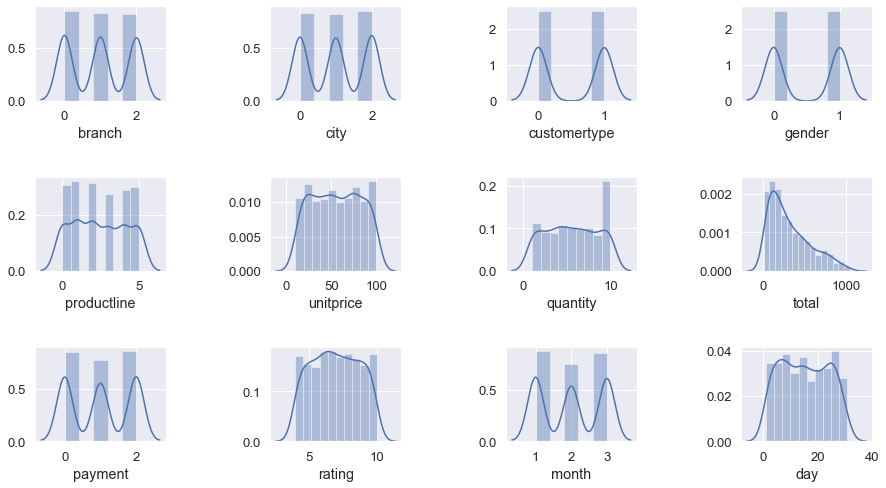
It is a rectangular boxes

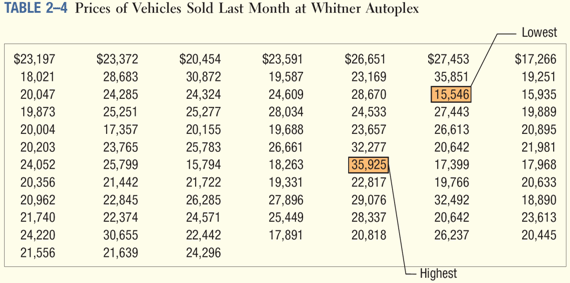
x-axis: Numerical

Y-axis: Numerical

| Categorical | Numerical |
| --- | --- |
| Tabular representation: Frequency table how many members are present in each class | Frequency distribution table , in each class interval how many members are fall |
| x-axis: categorical : class  y-axis: numerical : frequency | x-axis: Numerical : class interval  y-axis: Numerical : class frequency |
| Graphical representation : Bar graph | Histogram |

Data distribution





Data distribution is generated by histogram

Histogram will generate by class interval vs class frequency

Class interval vs class frequency is generated from raw data

Insight ============== observation

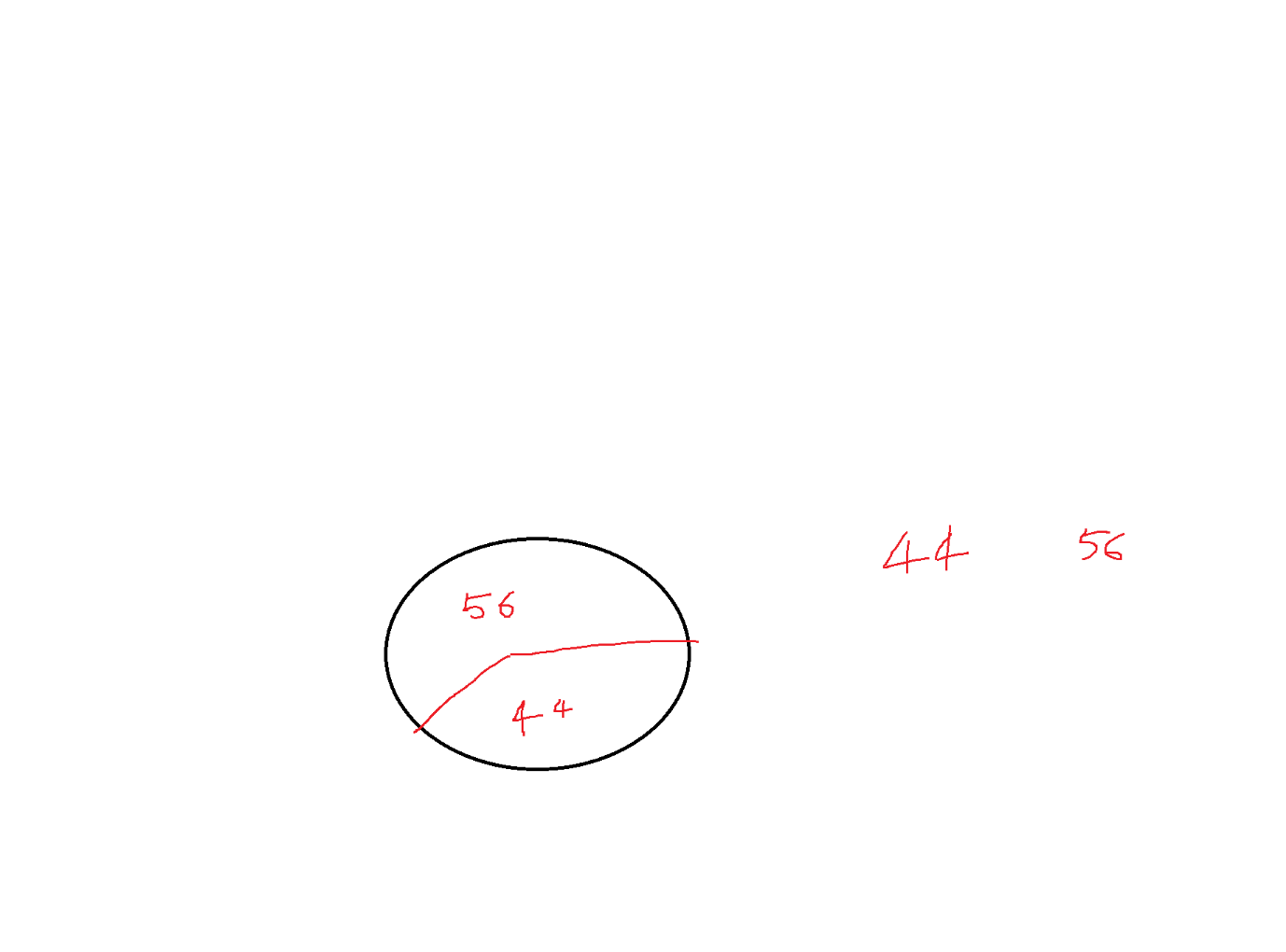
Lets go back to frequency table

| Gender | No of students | Relative frequency |
| --- | --- | --- |
| Girl | 40 | 44% |
| Boys | 50 | 56% |

| class | frequency | Relative frequency |
| --- | --- | --- |
| Girl | 40 ==== out of 100 | 44% ==== wrt 100 |
| Boys | 50 | 56% |

Pie-chart:

To make relative frequency



Mutually exclusive

If you toss a coin ======== head or tail

Whenever head occurs ========= tail will not occur (mutually exclusive)

There is only one job is there ======== two applicants

Steps to implement frequency distribution table:

How many class intervals you need to consider

How much class width

Step-1: Find the min and max value

Min=15k Max=35k

Step-2: Total how many observations : 80

Step-3:

So number of classes are 7

Step-4: width of the class

(Max-min)/k

(35-15)/7= 3

* Min : 15k
* Max: 36k
* Classes= 7
* Width=3

| Class interval | Class Frequency1 |
| --- | --- |
| 15-18 |  |
| 18-21 |  |
| 21-24 |  |
| 24-27 |  |
| 27-30 |  |
| 30-33 |  |
| 33-36 |  |

