INTRODUCTION TO DATA IX

2. Data Visualisation

Data visualisation is a graphical representation of data. There are different types of visualisation based on the type of attribute given.

Below are few famous visualisation methods for categorical and numerical features.

1. Bar Chart: It is only meant for "categorical attributes". The attribute value is represented by bars where, height of bar represents the count of categories taken by the attribute.

2. Histogram: It is only meant for "numerical attributes". Here, the attribute values are grouped into ranges where, ranges are decided by the user.

3. Line Graph: It is only meant for variables where there is a continuous change in value of attribute over time. More formally, it represents the trend.

4. Scatterplot: It is used to show relationship between two attributes.

- Univariate means one variable; bivariate means two variables.
- Bar chart, Histogram and Line Graph use univariate data while Scatterplot use bivariate data.

INTRODUCTION TO DATA X

Data Visualisation- Bar graph

Consider 20 people participated in rating a movie under categories namely, excellent, good, average and poor.

The data set was formed by the survey data as shown in Table 4.

Excellent	Good	Average	Poor
6	3	1	10

Table 4: Movie Survey data

Figure 5: Bar graph of movie review

- A **continuous variable** can take only numerical values.

INTRODUCTION TO DATA XI

Data Visualisation- Histogram

Consider age of 15 students to be analysed using Histograms.

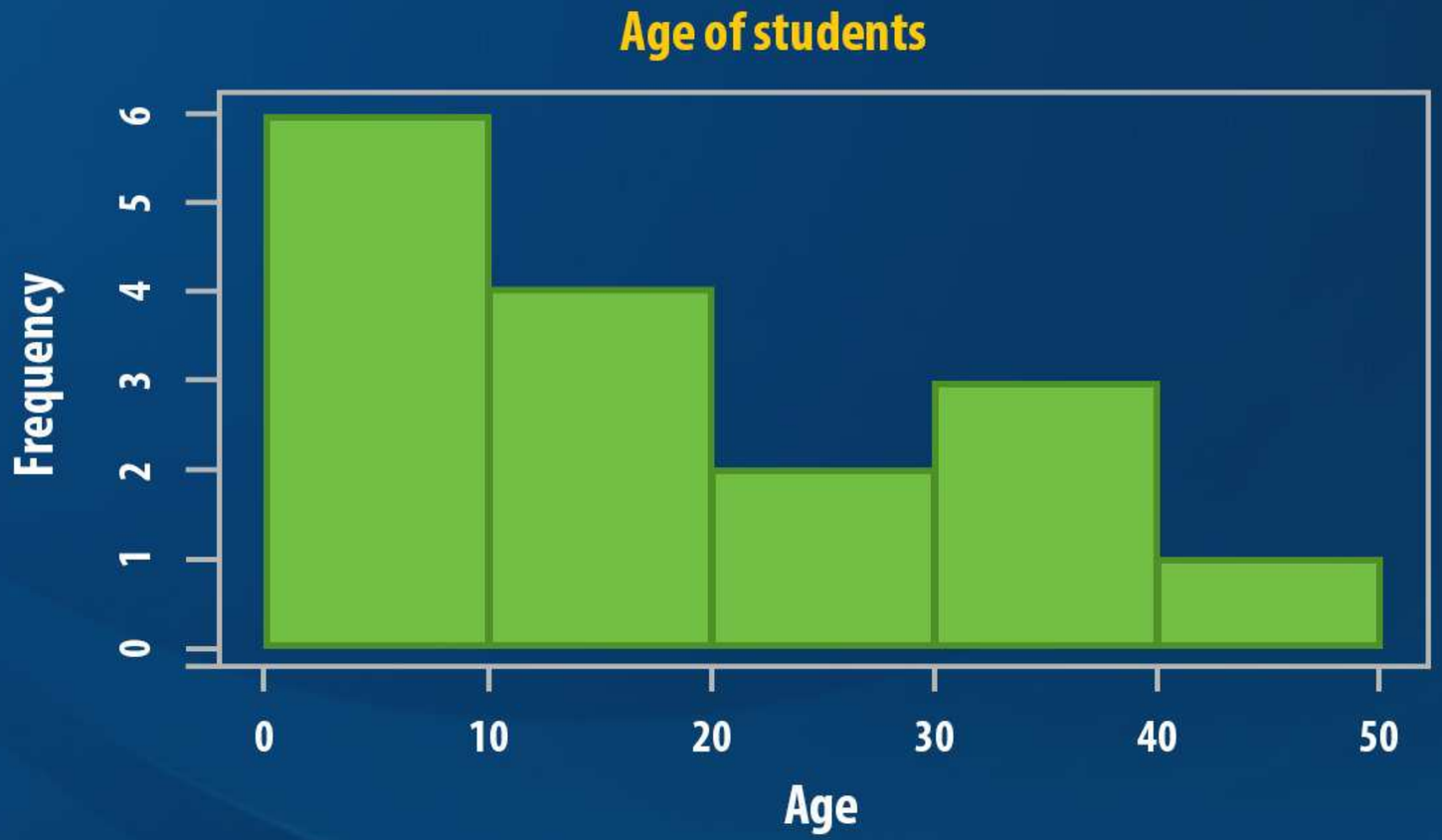

Figure 6: Histogram on age of 15 students

INTRODUCTION TO DATA XII

Data Visualisation- Scatterplot

It is very commonly used graph to understand relationship between two features.

Consider Figure 7 and Figure 8.

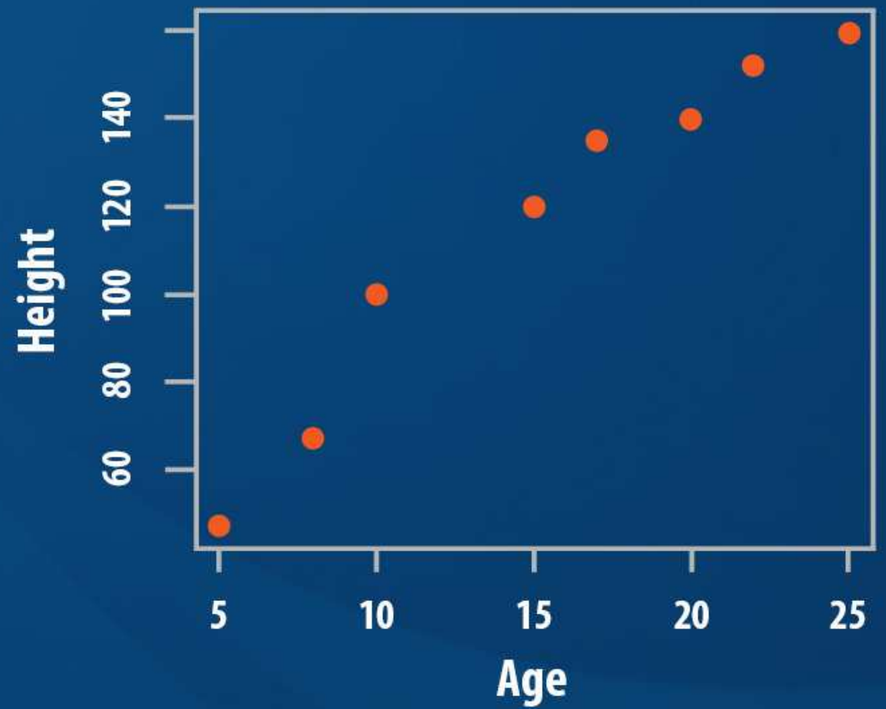

Figure 7: Example of Positive relations between Age and Height

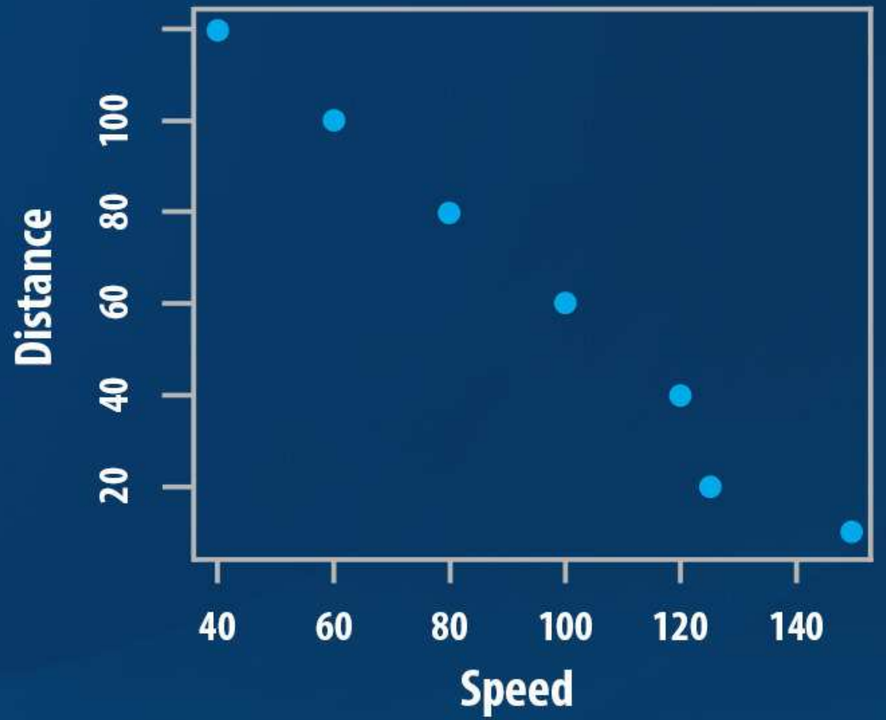

Figure 8: Example of negative relation between Speed and Distance

Figure 9: Revealing relationship between two features using Scatterplot

INTRODUCTION TO DATA XIII

Data Visualisation- Line Graph

Line graph shows the trend over time. Consider data set in Table 5 that stores sales of a product at each day of the week. Using line graph, we need to analyse whether sale is increasing or decreasing. The line graph on Table 5 is shown in Figure 10.

As shown in the figure, sale has gone down on Wednesday but grows after this day of week.

Day	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.	Sun.
Sales	16.8	23.89	7.8	17.89	19.0	20.2	22.2

Table 5: Sale of a product over the

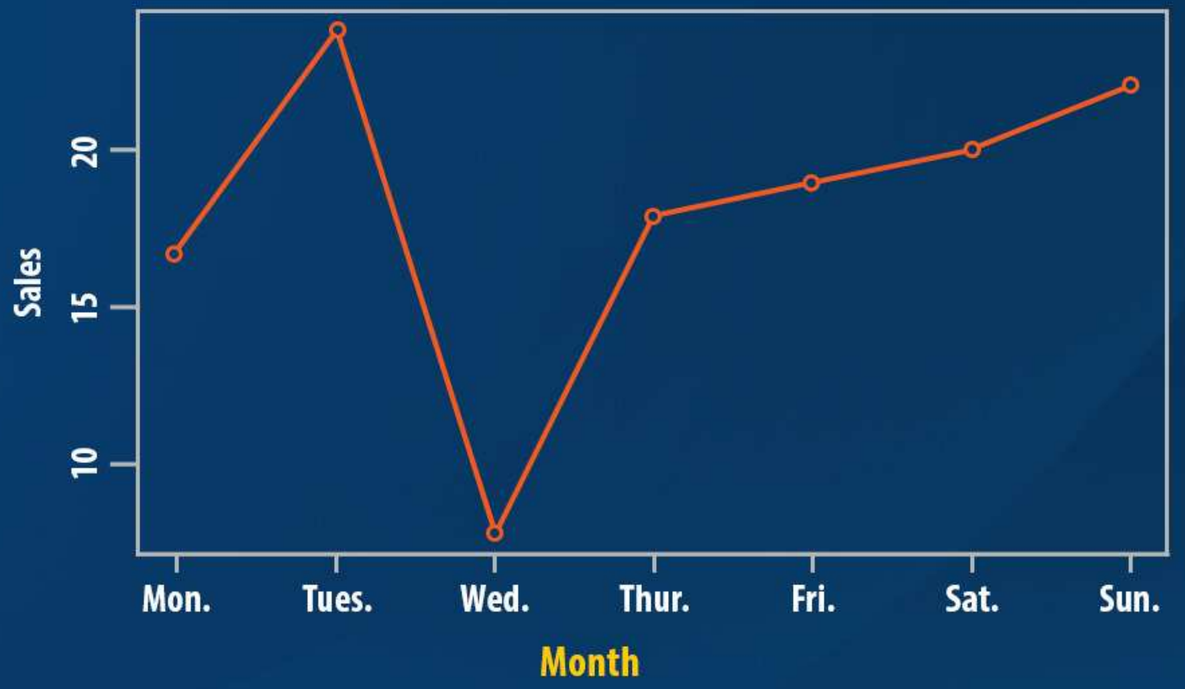

Figure 10: Sales of product over the week