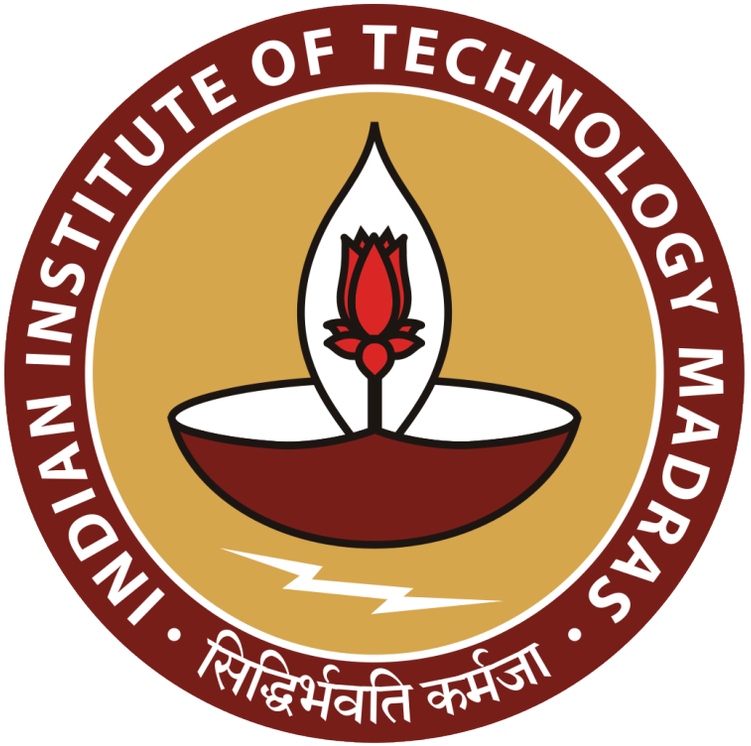


IIT Madras

ONLINE DEGREE

Statistics for Data Science -1

Lecture 4.4: Association between two numerical variables-Scatterplot

Usha Mohan

Indian Institute of Technology Madras

Introduction

- To understand the association between two numerical variables.
- Learn how to construct scatter plots and interpret association in scatter plots.
- Summarize association with a line.
- Correlation matrix.

Scatter plot

We use a scatterplot to look for association between numerical variables.

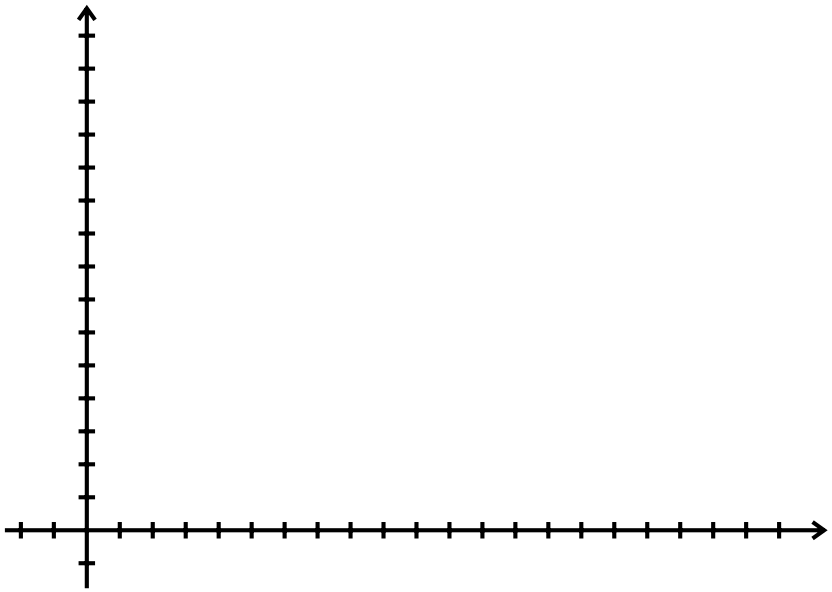
Definition

A scatter plot is a graph that displays pairs of values as points on a two-dimensional plane.

- To decide which variable to put on the x-axis and which to put on the y-axis, display the variable you would like to explain along the y-axis (referred as response variable) and the variable which explains on x-axis (referred as explanatory variable).

Example 1

Age (years)	Height (cms)
1	75
2	85
3	94
4	101
5	108



Example 2: Prices of homes

A real estate agent collected the prices of different sizes of homes. He wanted to see what was the relationship between the price of a home and size of a home. In particular, he wanted to know if the prices of homes increased linearly with the size or in any other way? To answer the question, he collected data on 15 homes. The data he recorded was

1. Size of a home measured in 1000 of square feet.
2. Price of a home measured in lakh of rupees.

Housing data

	Size (1000 Square feet)	Price (INR Lakhs)
1	0.8	68
2	1	81
3	1.1	72
4	1.3	91
5	1.6	87
6	1.8	56
7	2.3	83
8	2.3	112
9	2.5	93
10	2.5	98
11	2.7	136
12	3.1	109
13	3.1	122
14	3.2	159
15	3.4	170

Scatter plot using google sheets

Step 1: Highlight data you want to plot

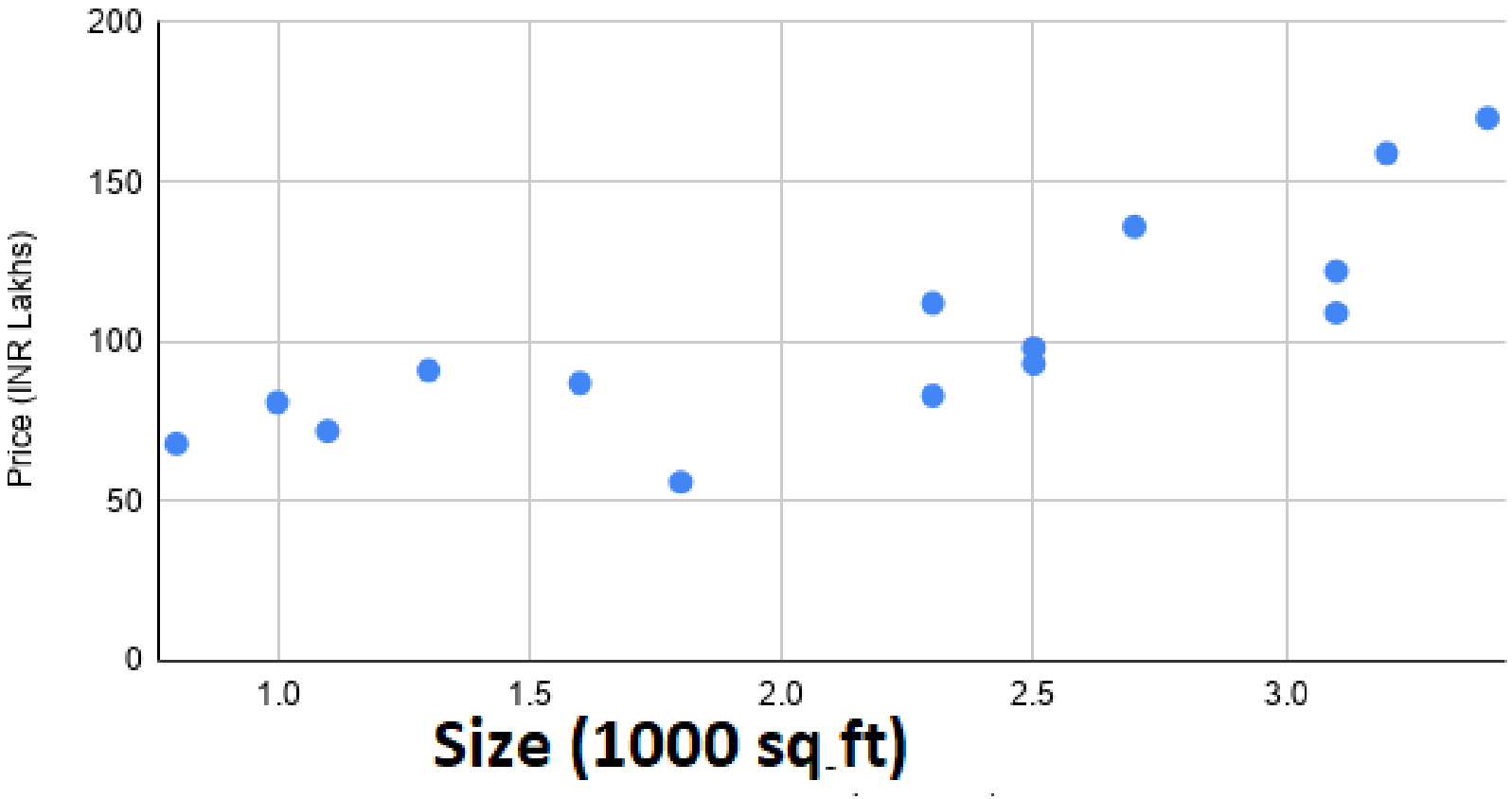
Step 2: Insert - chart- choose scatter chart

Step 3: Under $X-$axis tab, choose your explanatory variable.

Step 4: Under series tab, the response variable.

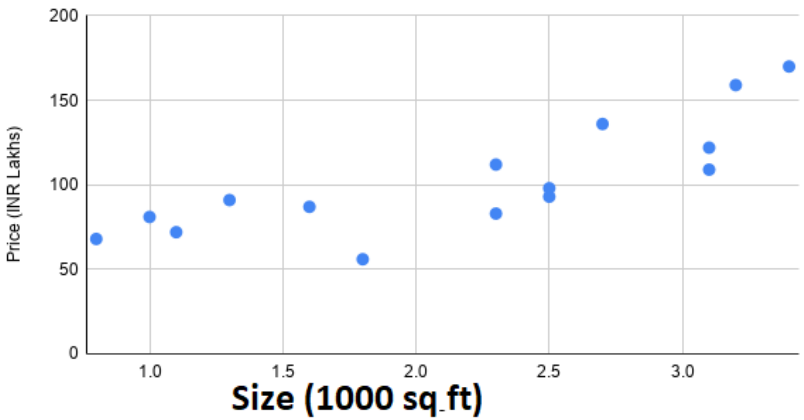
Step 5: Label the title of the chart, axes appropriately.

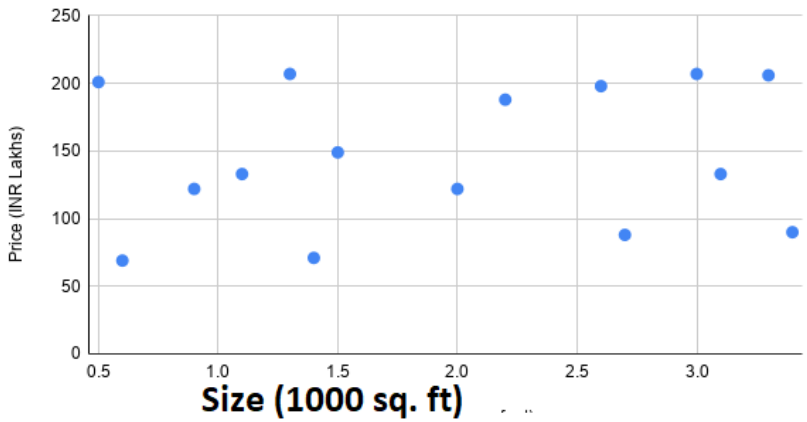
Scatter plot



Visual test for association

- Do we see a pattern in the scatter plot?
  - In other words, if I know about the x-value, can I use it to say something about the y-value or guess y-value?





Section summary

1. Draw a scatter plot
2. Notion of explanatory variable and response variable.
3. Visual test for association