



**PES University, Bangalore**

(Established under Karnataka Act No. 16 of 2013)

**UE19CS203 – STATISTICS FOR DATA SCIENCE**

**Unit-1 - Introduction to Data Science**

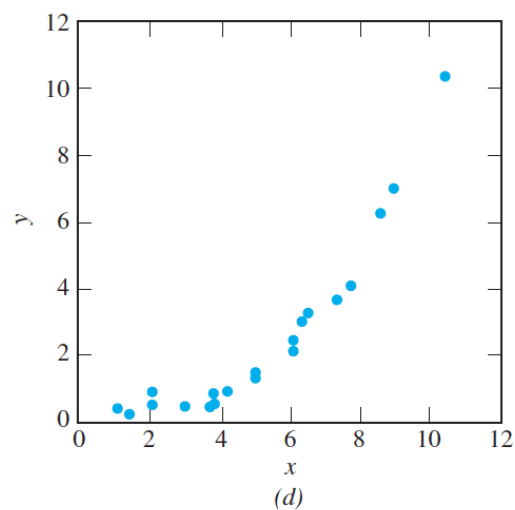
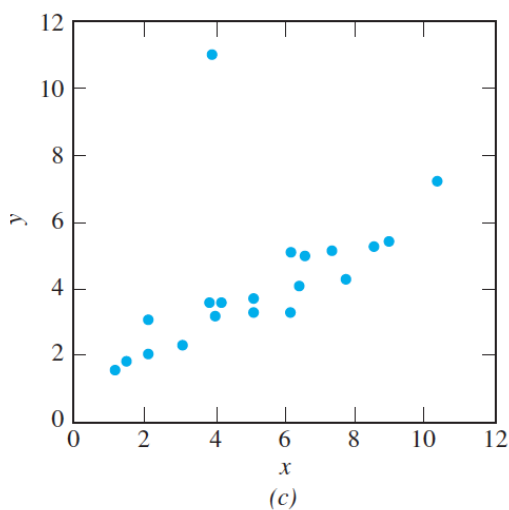
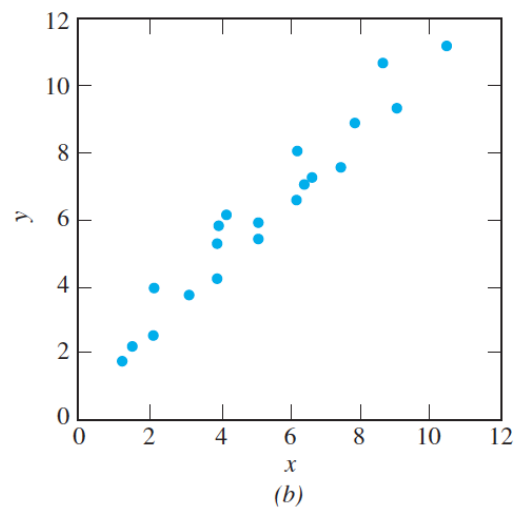
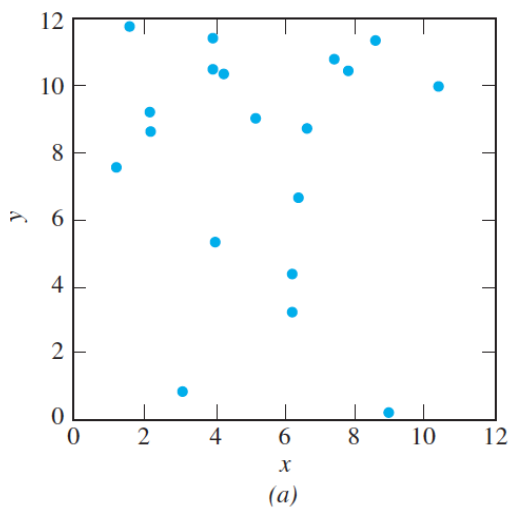
**QUESTION BANK**

**Data Visualization Techniques – Scatter Plot, Bar Chart, Heat Map**

**Exercises for Section 1.3**

**[Text Book Exercise – Section 1.3 – Q. No.[17 – 18] – Pg. No. [42 – 43]]**

1. Match each scatterplot to the statement that best describes it.

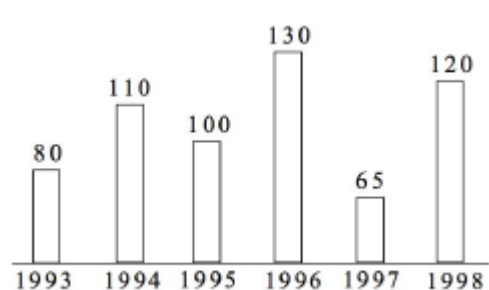


- a) The relationship between x and y is approximately linear.
- b) The relationship between x and y is nonlinear.
- c) There isn't much of any relationship between x and y.
- d) The relationship between x and y is approximately linear, except for an outlier.

2. For the following data:

|   |     |     |     |     |     |      |      |      |      |      |
|---|-----|-----|-----|-----|-----|------|------|------|------|------|
| X | 1.4 | 2.4 | 4.0 | 4.9 | 5.7 | 6.3  | 7.8  | 9.0  | 9.3  | 11.0 |
| Y | 2.3 | 3.7 | 5.7 | 9.9 | 6.9 | 15.8 | 15.4 | 36.9 | 34.6 | 53.2 |

- a) Make a scatterplot of y versus x. Is the relationship between x and y approximately linear, or is it nonlinear?
  - b) Compute the natural logarithm of each y value. This is known as making a log transformation of y. Make a scatterplot of  $\ln y$  versus x. Is the relationship between x and  $\ln y$  approximately linear, or is it nonlinear?
  - c) In general, it is easier to work with quantities that have an approximate linear relationship than with quantities that have a nonlinear relationship. For these data, do you think it would be easier to work with x and y or with x and  $\ln y$ ? Explain.
3. The bar graph provided below gives the data of the production of food grains over the years which is (1000 tonnes). Study the bar chart and answer the questions.



- a) In the case of how many years was the production below the average production of the given years?
  - b) What was the percentage drop in production from 1996 to 1997?
  - c) In which year was the production 50% of the total production in 1993 and 1998 together?
  - d) If the production in 1999 will be above the average production of the given years, which of the following could be the minimum production for 1999?
  - e) What was the approximate percentage increase in production from 1993 to 1994?
4. What are Heat Maps? And How does Heat Map data reports?
5. What is the purpose of using Heat Maps?