SNS COLLEGE OF TECHNOLOGY

Coimbatore-35.
An Autonomous Institution

COURSE NAME : DATA ANALYTICS

II YEAR/ IV SEMESTER

UNIT – II Getting Insights from Data

Topic: *Bivariate Analysis*

Dr.K.Sangeetha
HoD
Department of Computer Science and Engineering

Descriptive Statistics (cont..)

Descriptive Bivariate Analysis :

- Pairs of attributes, and their relative behavior

- It is organized according to the scale types of the attributes: quantitative, nominal and ordinal.

- When one of the attributes of the pair is qualitative – that is, nominal or ordinal – and the other is quantitative

Descriptive Statistics (cont..)

Descriptive Bivariate Analysis :

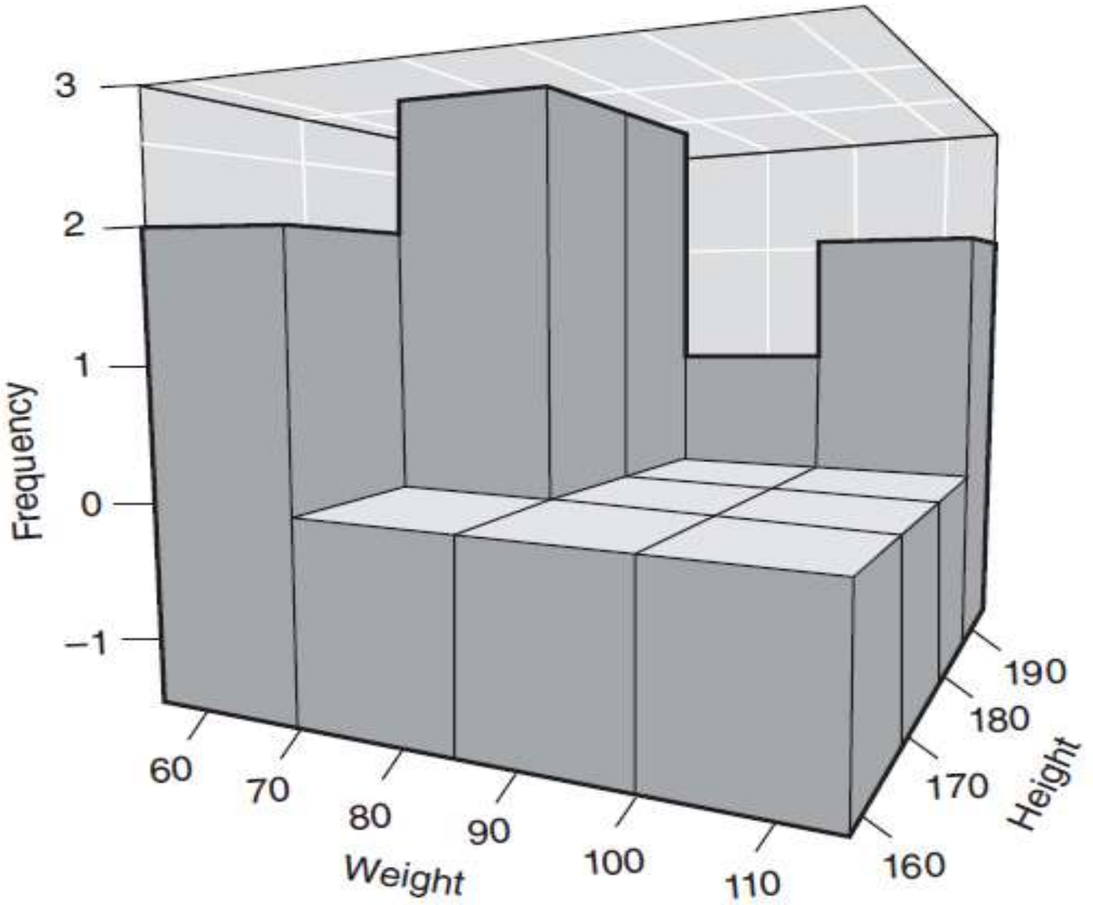
1. Two Quantitative Attributes:

- Several visualization techniques that can visually show the distribution of points with two quantitative attributes.

- One of these techniques is an extension of the histogram called a three-dimensional histogram.

Descriptive Statistics (cont..)

1. Two Quantitative Attributes:

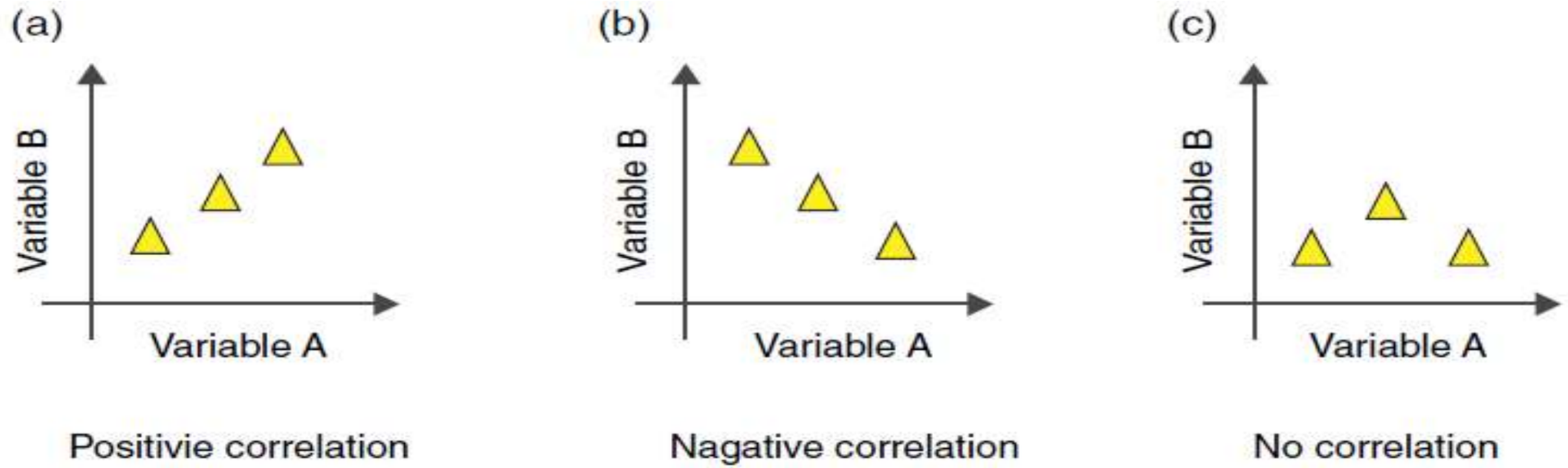


3D histogram for attributes “weight” and “height”.

Descriptive Statistics (cont..)

1. Two Quantitative Attributes:

Examples of **correlations between two attributes**, A and B: a positive correlation, a negative correlation and a lack of correlation.



Descriptive Statistics (cont..)

2. Two Qualitative Attributes, at Least one of them Nominal :

- When the attributes are both qualitative with at least one nominal, contingency tables are used.
- Contingency tables present the joint frequencies, facilitating the identification of interactions between the two attributes.

- They have a matrix-like format, with cells in a square and labels at the left and top.
- On the right most column are the totals per row while in the bottom most row are the totals per column.
- The bottom right-hand corner has the total number of values.

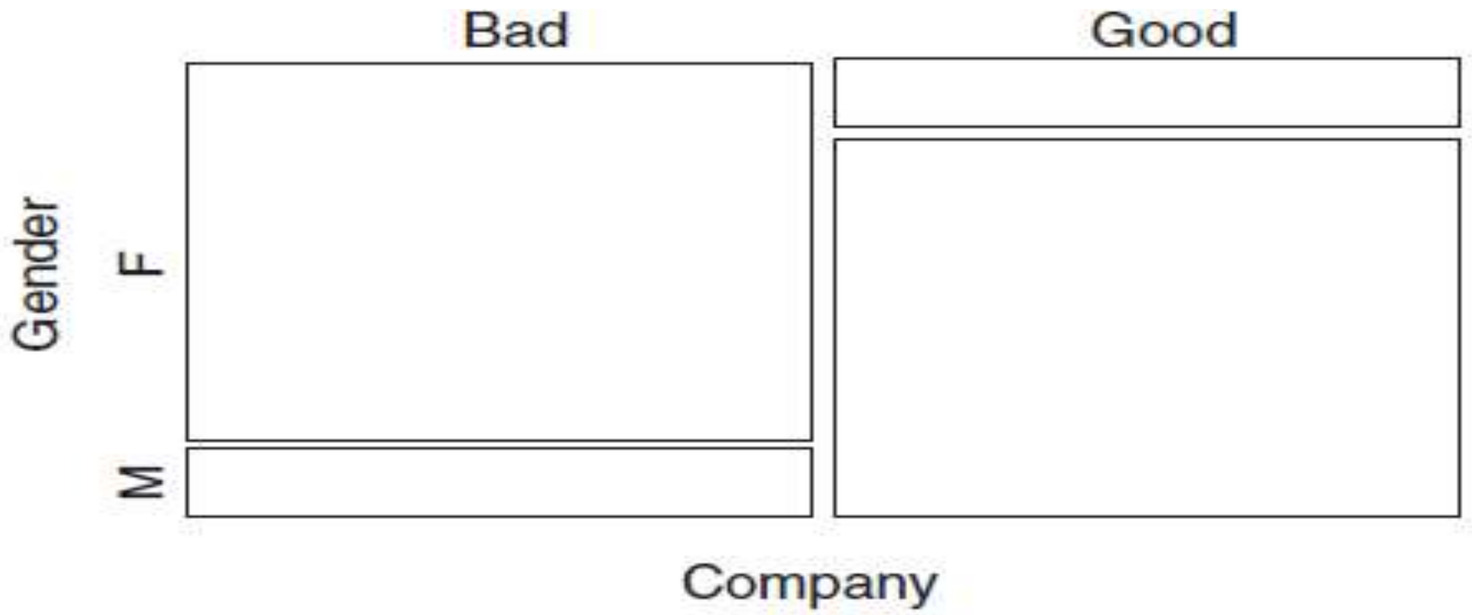
Descriptive Statistics (cont..)

Contingency table :

		Company		
		Good	Bad	
Gender	Male	6	2	8
	Female	1	5	6
		7	7	14

Contingency table with absolute joint frequencies for “company” and “gender”.

Descriptive Statistics (cont..)



Mosaic plot for "company" and "gender".

Descriptive Statistics (cont..)

3. Two Ordinal Attributes :

Scatter plots with ordinal attributes usually have the problem that there are many values falling at the same point, making it impossible to evaluate the number of values per point.

In order to avoid this problem, some software packages use a jitter effect which add a random deviation to the values, making it possible to evaluate how large the cloud is.

Contingency tables can be used and mosaic plots too.The values should be in increasing order.

Descriptive Multivariate Analysis

- **Multivariate analysis -** analysis of a data set which explores more than two attributes.

- **Four types are :**

1. Multivariate Frequencies
2. Multivariate Data Visualization
3. Multivariate Statistics
4. Infographics and Word Clouds

Descriptive Multivariate Analysis

1. Multivariate Frequencies

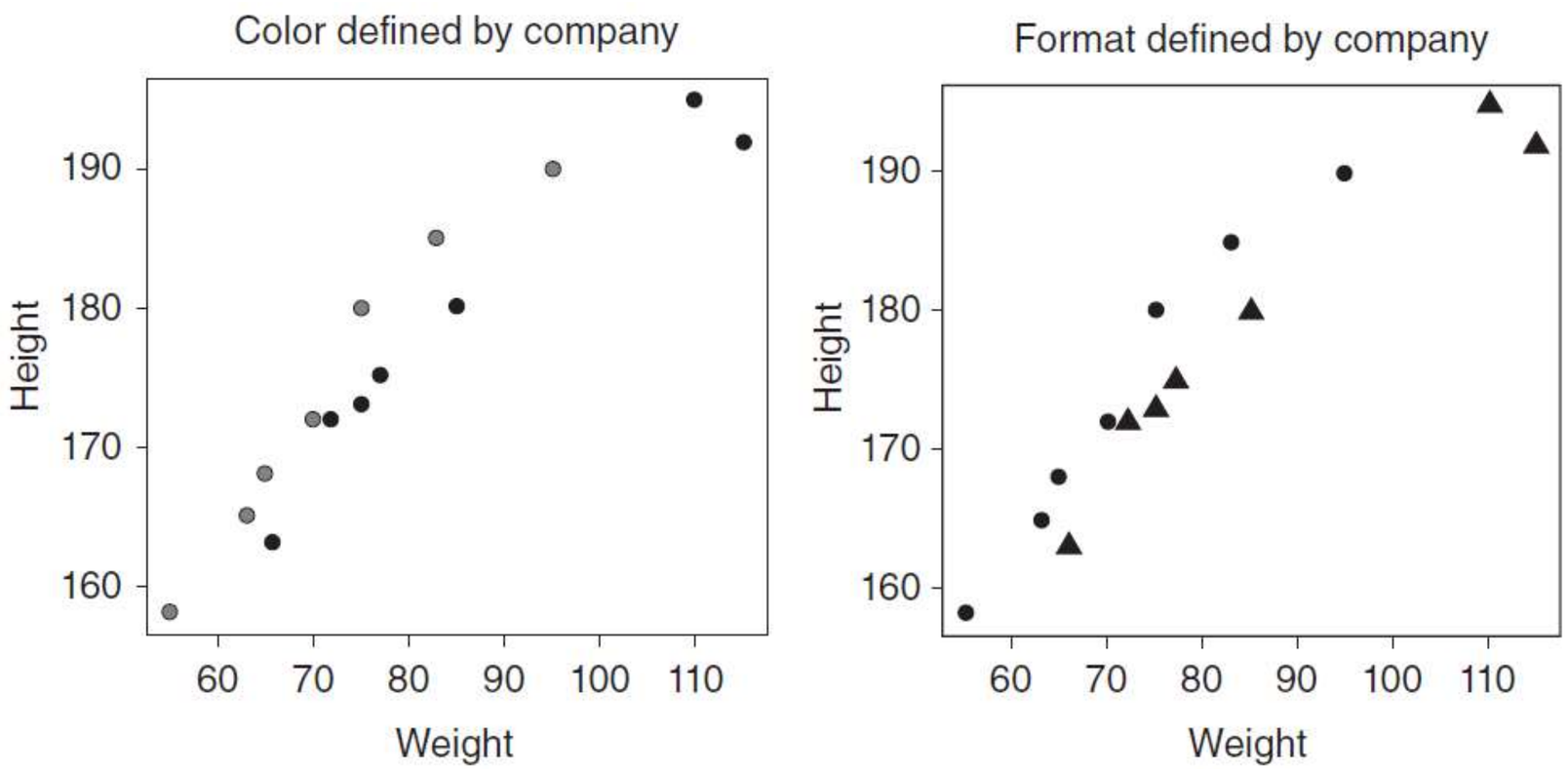
❖ Computed independently for each attribute

❖ For each attribute, the following frequency measures can be taken:

- absolute frequency
- relative frequency
- absolute cumulative frequency
- relative cumulative frequency

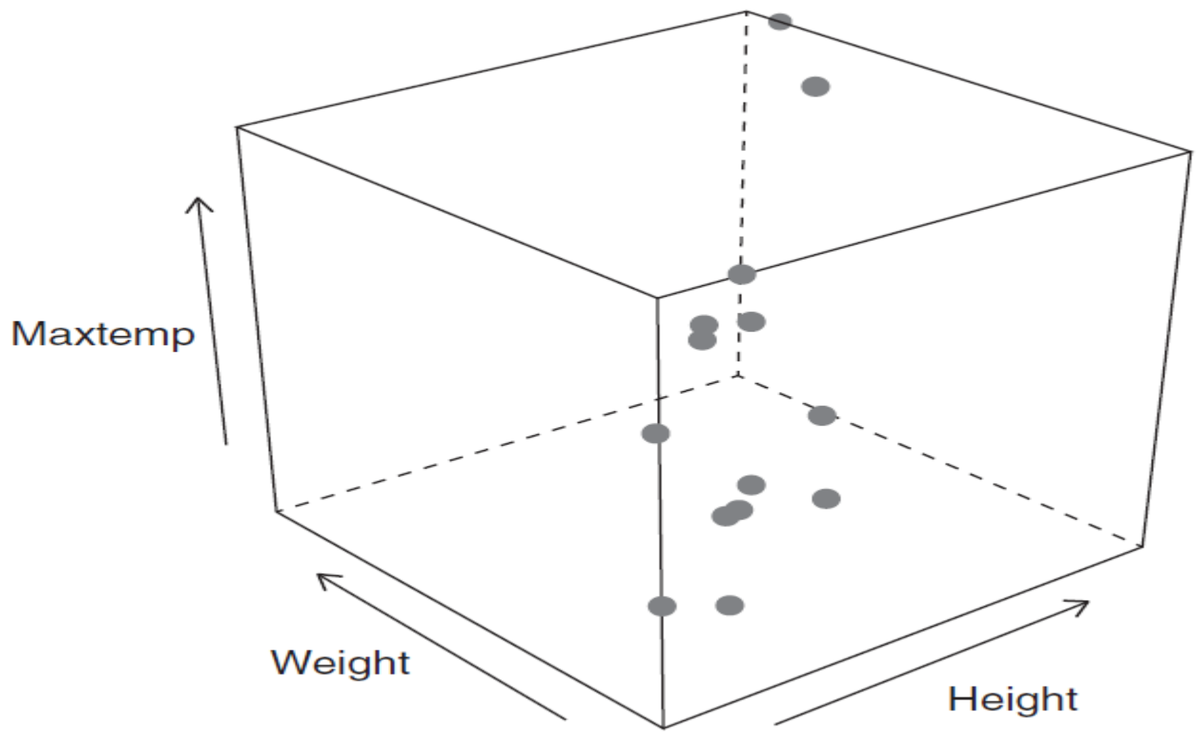
Descriptive Multivariate Analysis

2. Multivariate Data Visualization



Descriptive Multivariate Analysis

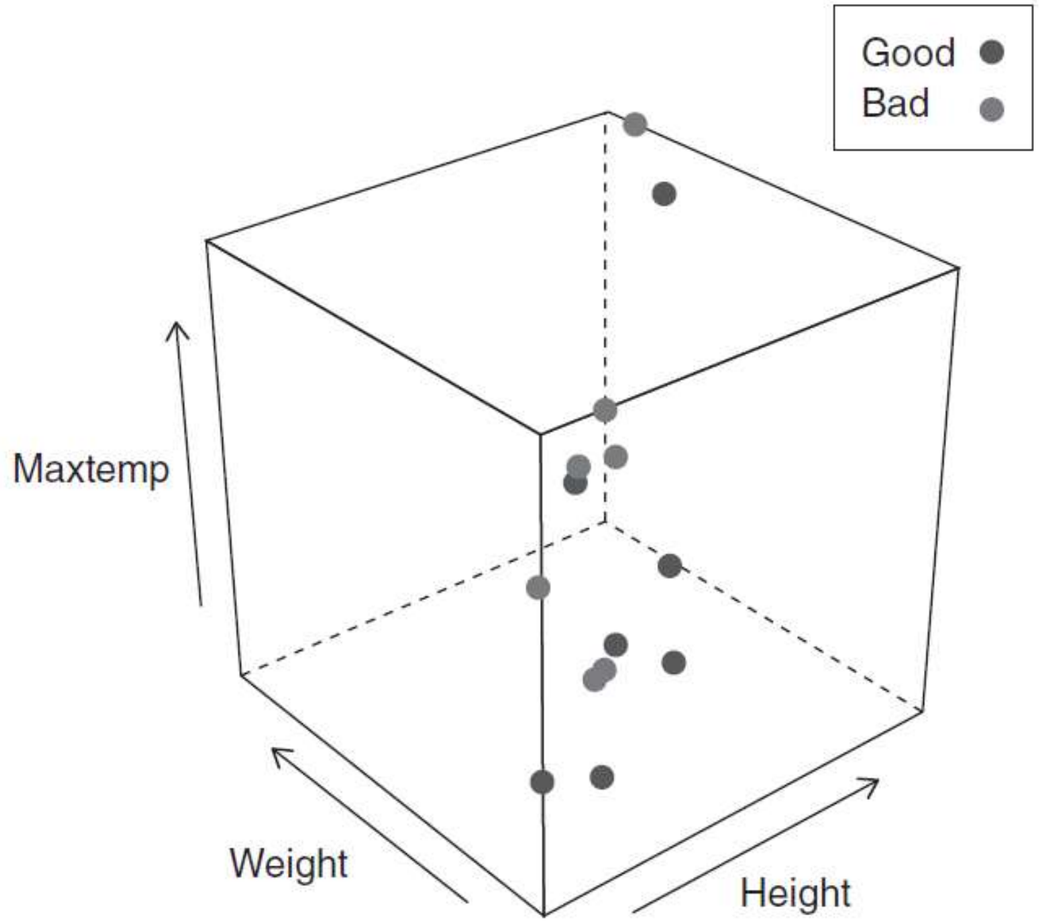
2. Multivariate Data Visualization



Descriptive Multivariate Analysis

2. Multivariate Data Visualization

Figure Plot for four attributes of the friends data set using color for the forth attribute.



Descriptive Multivariate Analysis

2. Multivariate Data Visualization

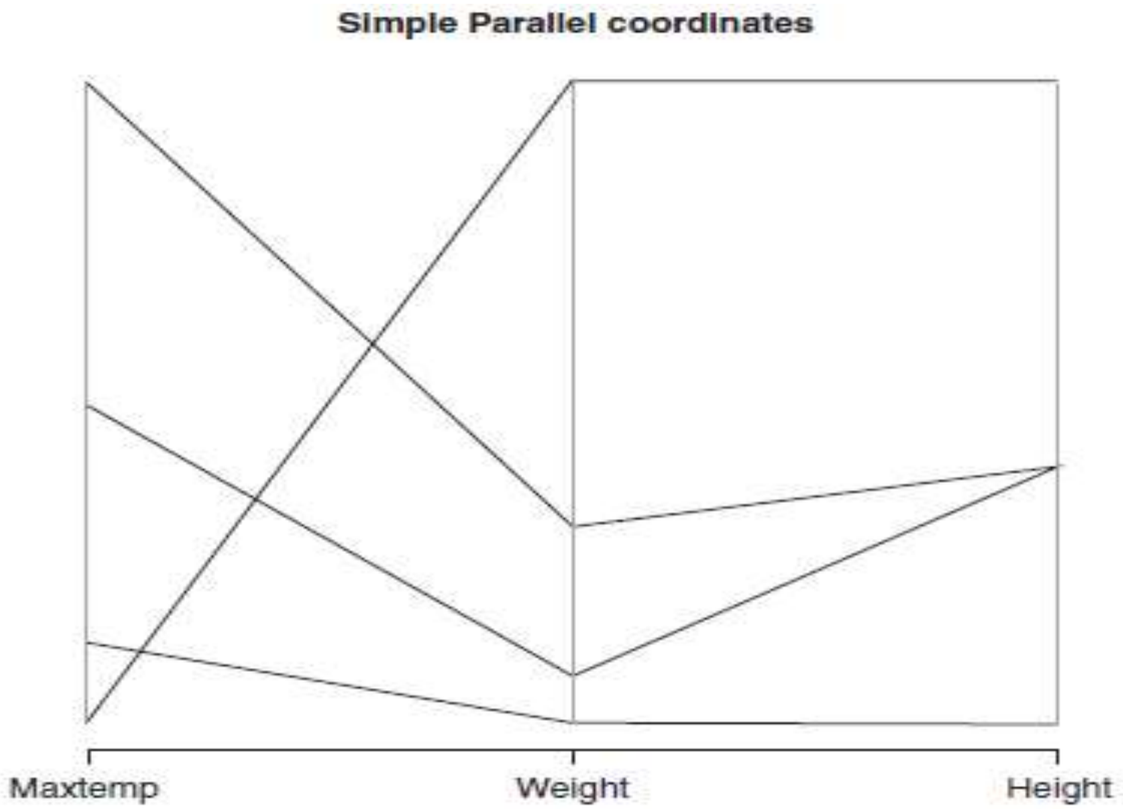


Figure 3.5 Parallel coordinate plot for three attributes.

Descriptive Multivariate Analysis

2. Multivariate Data Visualization

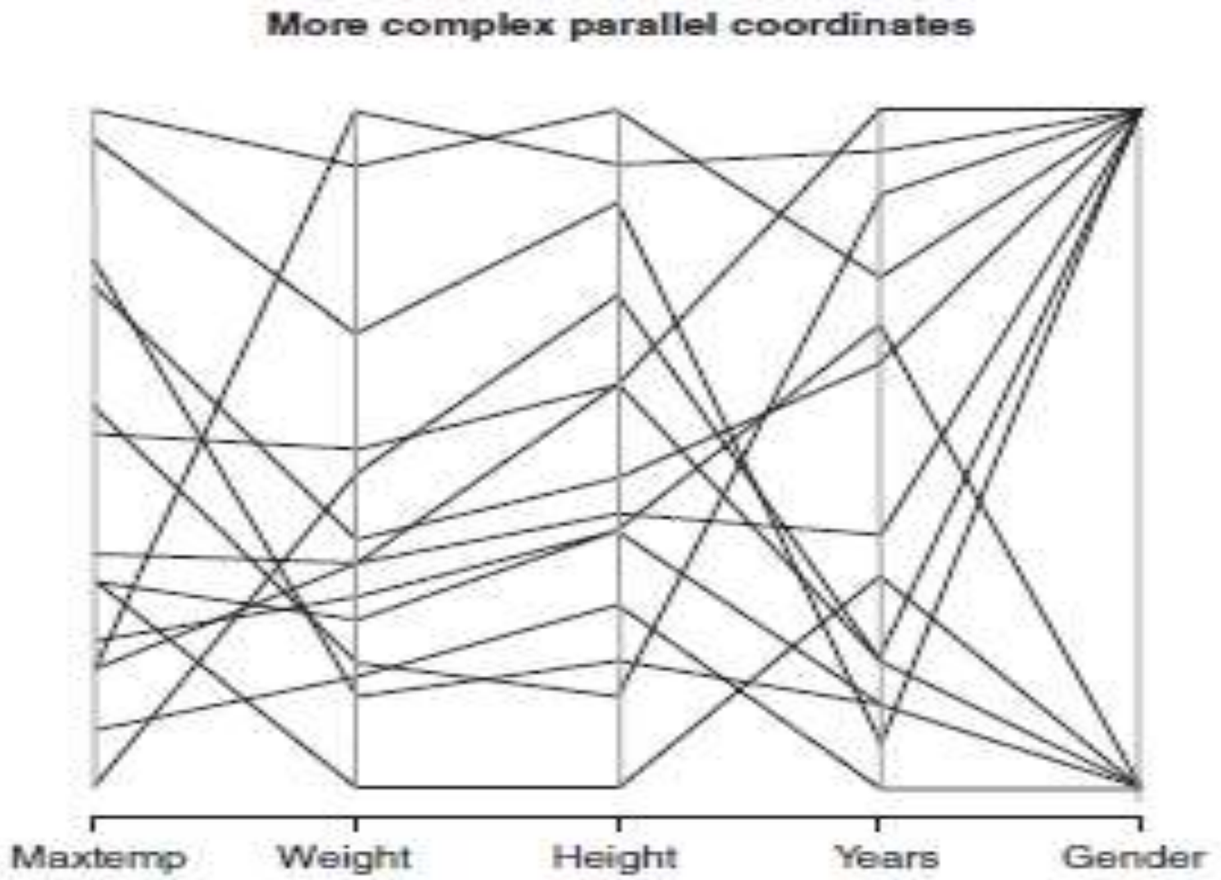
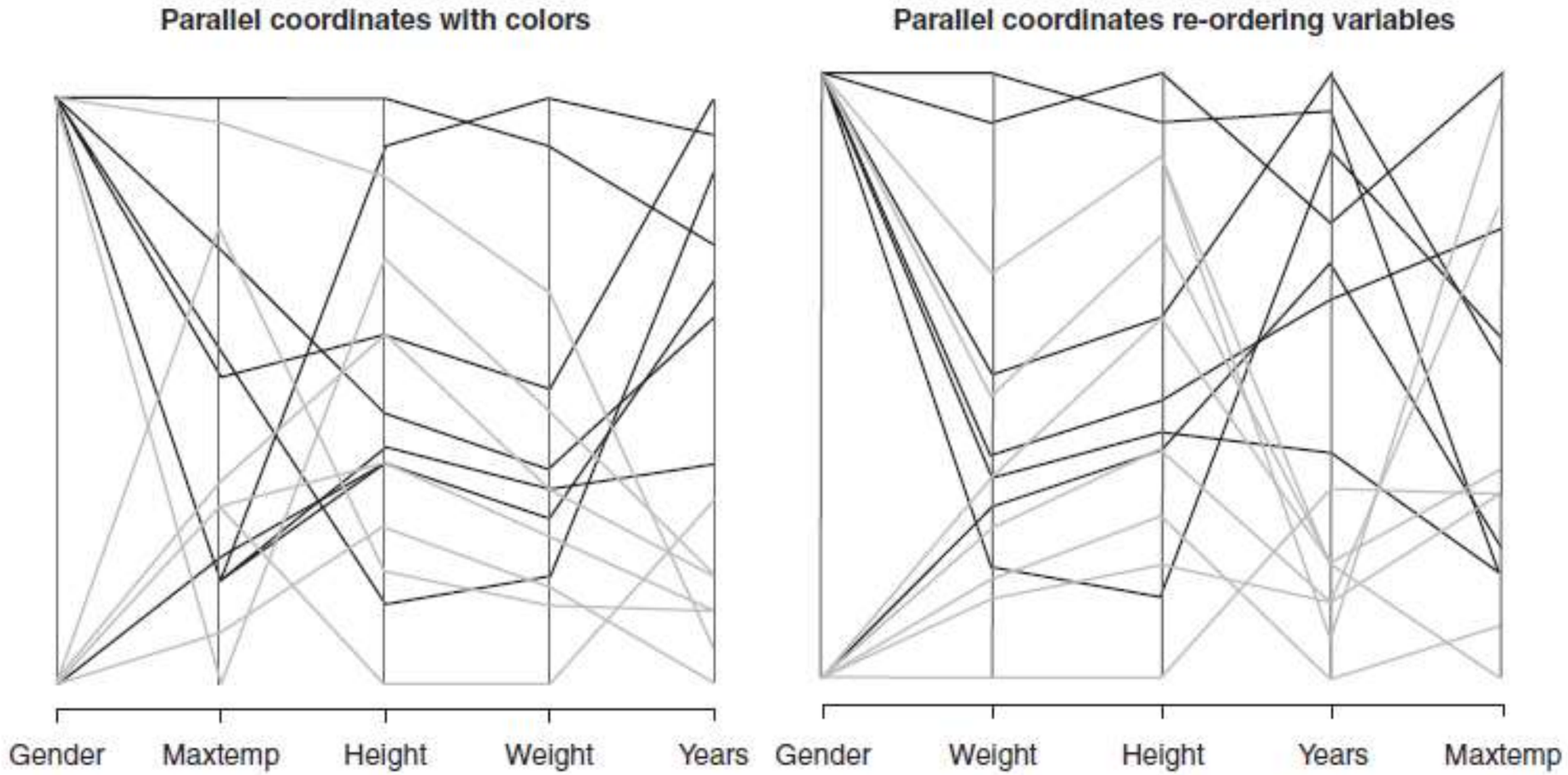


Figure 3.6 Parallel coordinate plot for five attributes.

Descriptive Multivariate Analysis

2. Multivariate Data Visualization



Descriptive Multivariate Analysis

3. Multivariate Statistics

❖ Multivariate statistics are a simple extension of the univariate statistics

Assessment 1

1. What is **Univariate Data visualization**?

Ans : _____

2. What is **Descriptive Bivariate Analysis** ?

Ans : _____

References

TEXT BOOKS

1. Joao Moreira, Andre Carvalho, Tomás Horvath – “A General Introduction to Data Analytics” – Wiley -2018

REFERENCES

1 Dean J, —Big Data, Data Mining and Machine learning, Wiley publications, 2014.

2 Provost F and Fawcett T, —Data Science for Business, O‘Reilly Media Inc, 2013.

3 Janert PK, —Data Analysis with Open Source Tools, O‘Reilly Media Inc, 2011. .

4 Weiss SM, Indurkhya N and Zhang T, —Fundamentals of Predictive Text Mining, Springer-Verlag London Limited, 2010.

5. Runkler T A, - Data Analytics: Models and Algorithms for Intelligent data analysis,Springer, 2012