SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A' Grade
Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

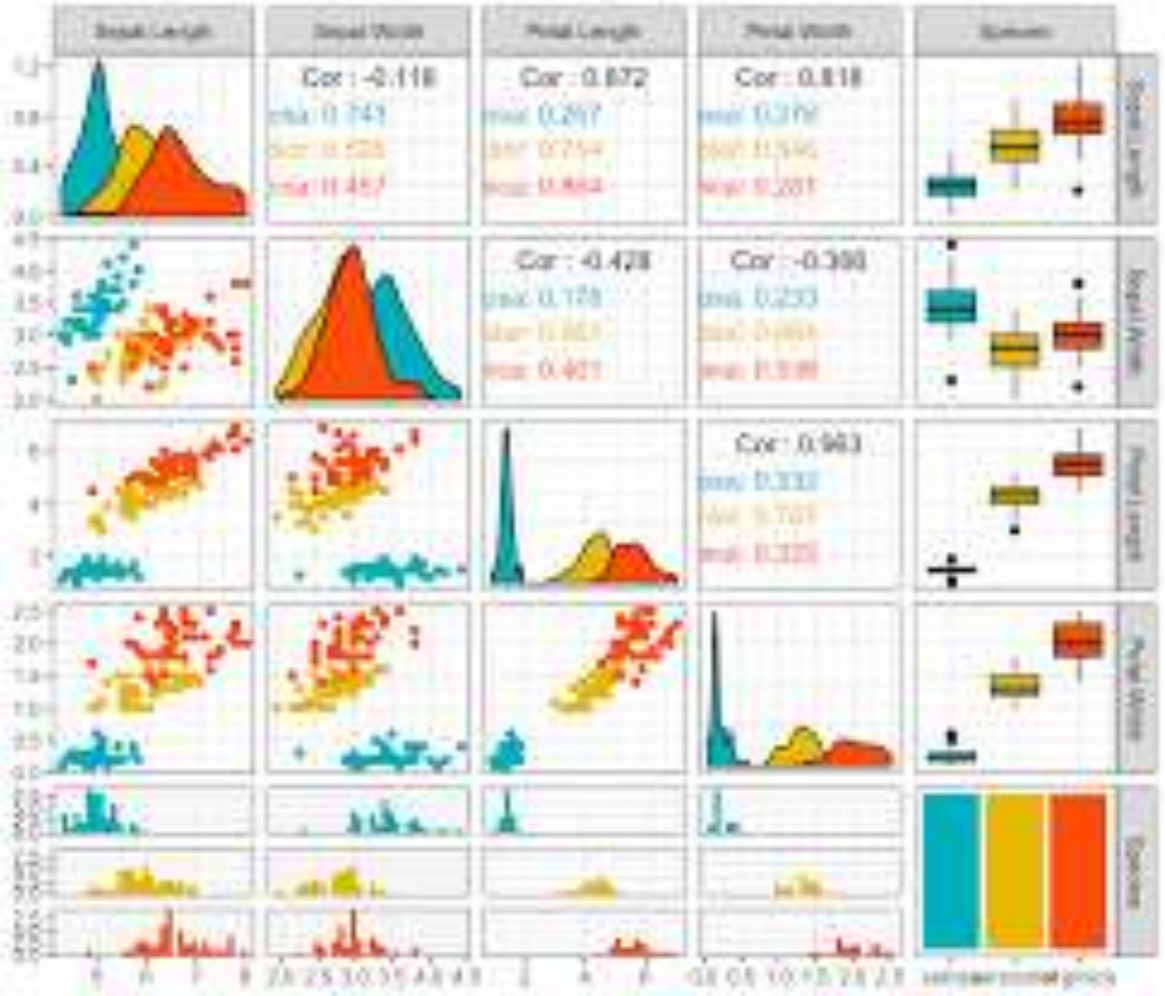
DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY

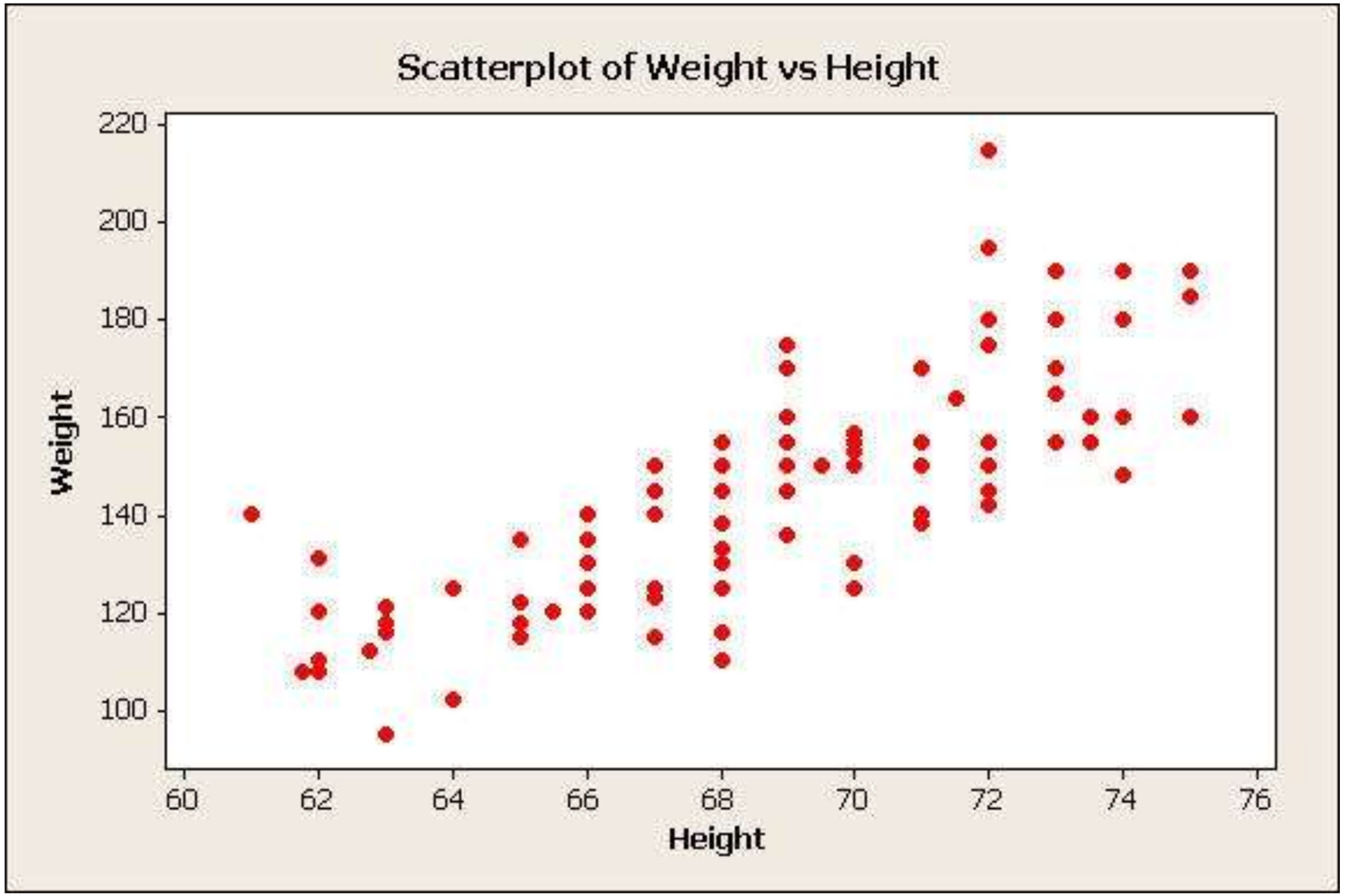
COURSE NAME :19CS407 DATA ANALYTICS WITH R

II YEAR /IV SEMESTER

Unit 2- GETTING INSIGHTS FROM DATA, DATA QUALITY AND PREPROCESSING

Topic : Multivariate Data Visualization


Cor : -0.118
Cor : 0.872
Cor : 0.818
Cor : -0.428
Cor : -0.366
Cor : 0.963


Scatterplot of Weight vs Height
Weight
Height
220
200
180
160
140
120
100
60
62
64
66
68
70
72
74
76

Multivariate Data Visualization

- ✓ the previous plots can be extended to represent a small number of extra attributes.
- ✓ Additionally, new visualization approaches and techniques are continuously being created to deal with new types of data, new approaches to results interpretation and new data analysis tasks.

Multivariate Data Visualization

- ✓ Depending on the number of attributes, and the need to represent spatial and/or temporal aspects of the data, different plots can be used.

- ✓ This how multivariate data can be visually represented in different ways and the main benefits of each of these alternatives

Multivariate Data Visualization

- When the multivariate data has three attributes, or one can only analyze three attributes from a multivariate data set, the data can still be visualized in a bivariate plot, associating the values of the third attribute with how each data object is represented in the plot.

- If the third attribute is quantitative, the value can be represented by the size of the object representation in the plot

Multivariate Data Visualization

- The size of each object in the plot is proportional to the value of the third attribute for this object.
- If, on the other hand, the third attribute is qualitative, its value can be represented in the plot by either the color or by the shape of the object.
- The number of colors or shapes will be the number of values the attribute can assume.
- In classification tasks, color and shape are usually employed to represent the class labels

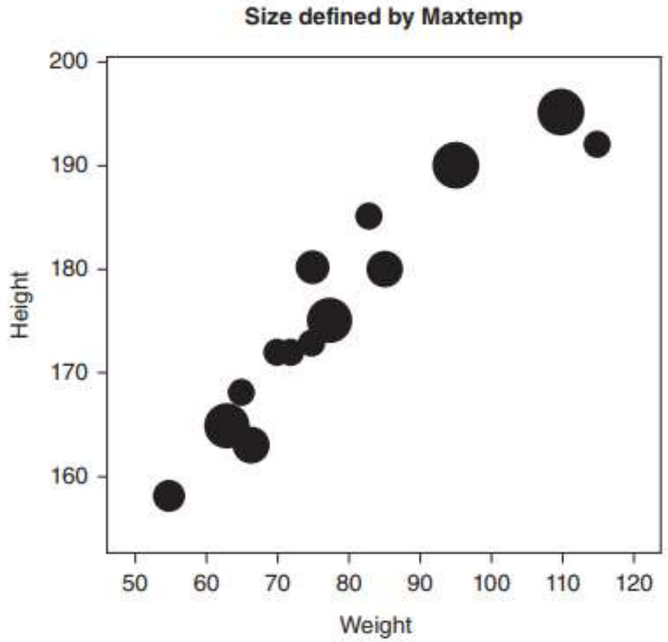


Figure 3.1 Plot of objects with three attributes.

Multivariate Data Visualization

- As an example, Figure shows two plots where the third attribute is qualitative.
- On the right-hand size it represents each qualitative value as a different shape. On the left-hand size, it represents each qualitative value by a different color.

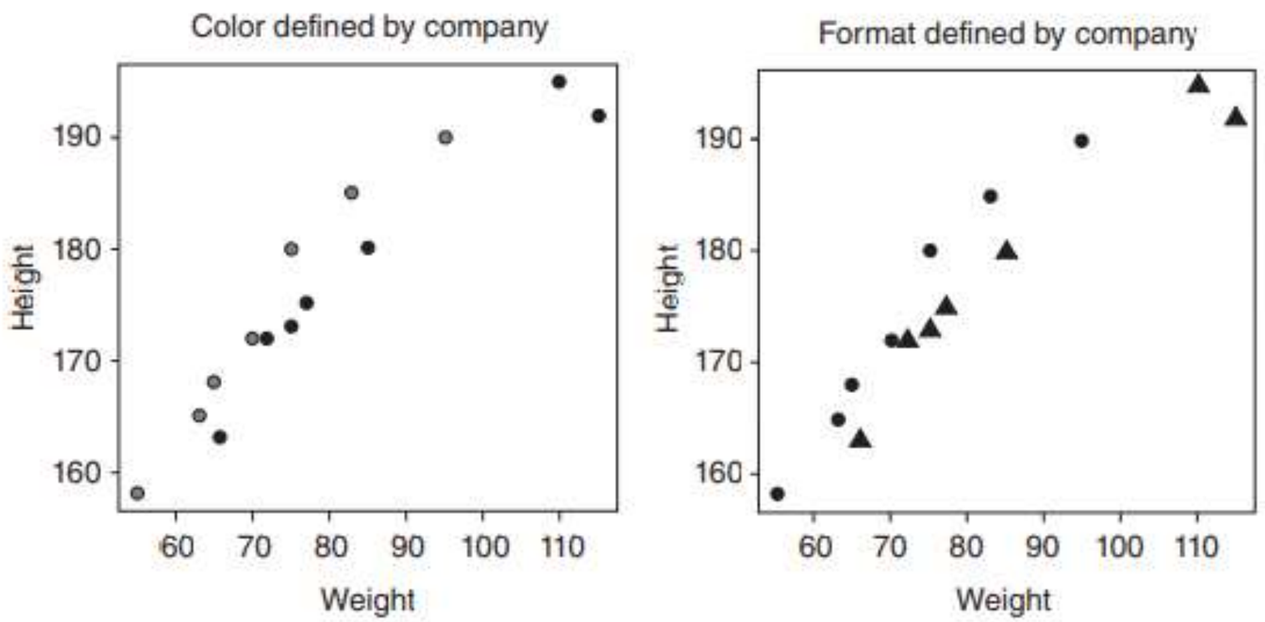


Figure 3.2 Two alternatives for a plot of three attributes, where the the third attribute is qualitative.

Multivariate Data Visualization

- Another approach to representing three attributes is to use a three-dimensional plot, where each axis is associated with one of the attributes.
- It makes more sense to use this approach if the three attributes are quantitative, because the values of the corresponding attributes can be presented on each axis assuming an order among them

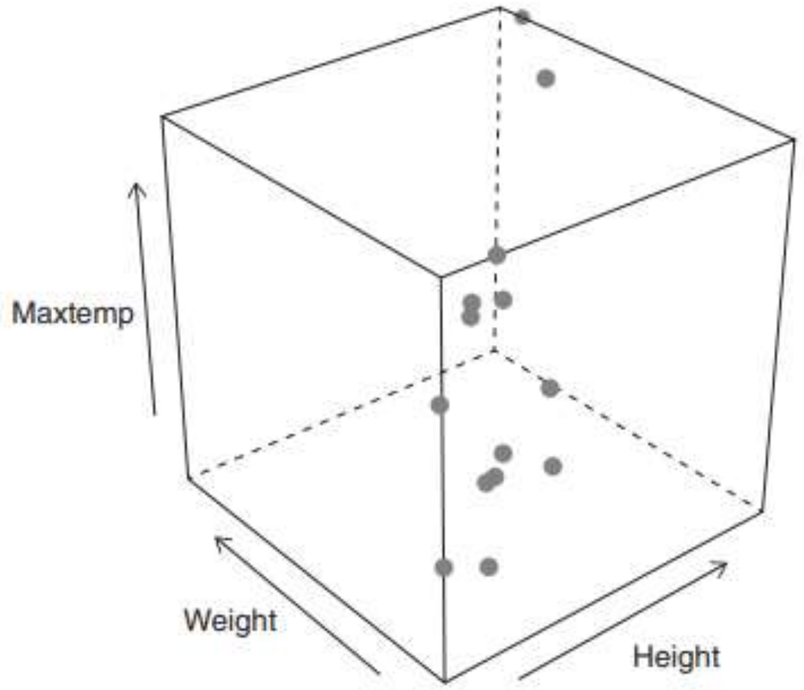


Figure 3.3 Plot for three attributes from the contacts data set.

Multivariate Data Visualization

- example using a three-dimensional plot for three quantitative attributes from the contacts data set.

- One may ask how can we represent the relationships between more than three attributes.

- A straightforward approach would be to modify the three-dimensional graph from Figure 3.3, representing the fourth attribute through the size, color or shape of the plotted object. This is shown in in Figure 3.4, using different colors for the attribute

Multivariate Data Visualization

- Although we can also do some tricks to represent more than two predictive attributes in the previous plots by using 3D versions and different formats and colors, not all plots will allow this, or, when they do, the resulting plot can be very confusing.

- For instance, depending on the plot chosen, color and shape will not preserve the original order and magnitude of the quantitative values, only the different values. Additionally, some qualitative values will not be naturally represented by different object sizes

Multivariate Data Visualization

- A different plot should be used. There are also some plots specifically for more than two quantitative attributes. These usually describe the quantitative attributes of a data set.

- One of the most popular is the parallel coordinates plot, also known as a profile plot. Each object in a data set is represented by a sequence of lines crossing several equally spaced parallel vertical axes, one for each attribute

parallel coordinate plot

- parallel coordinate plot for four of our contacts, using three quantitative predictive attributes. The quantitative attributes occupy positions on the vertical axes related to their values.
- Each object is represented by a sequence of lines crossing the vertical axes at heights that represent the attribute values.

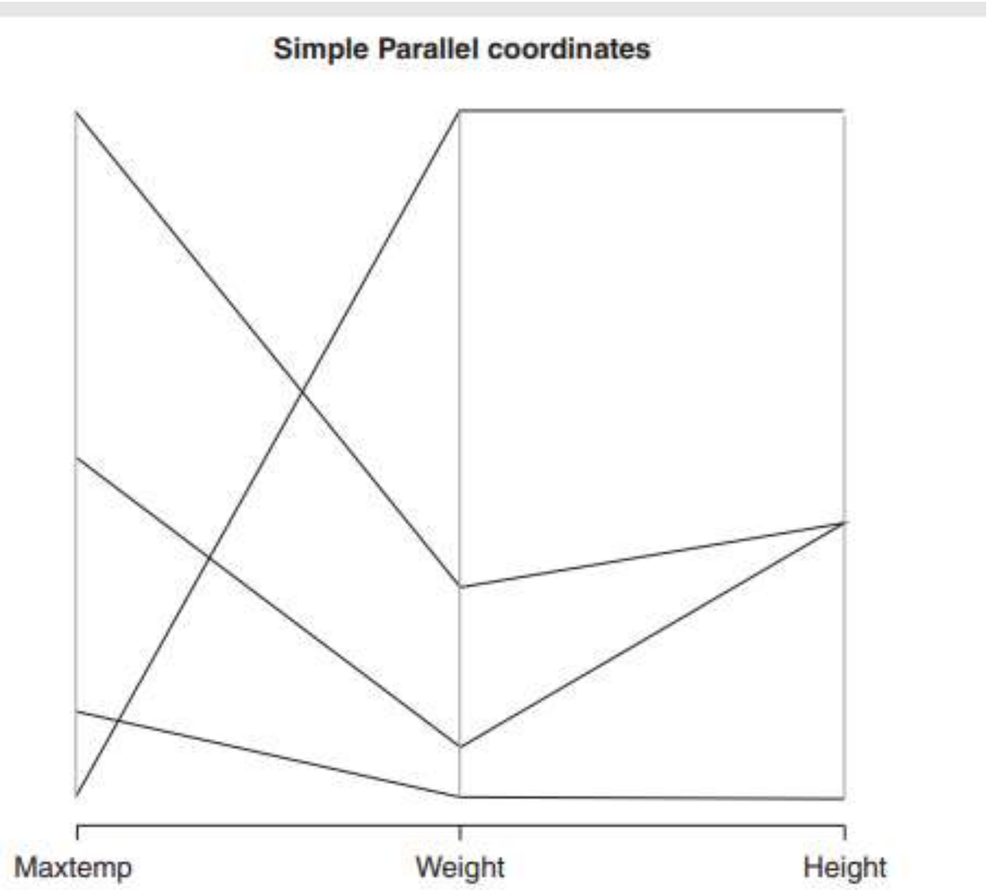


Figure 3.5 Parallel coordinate plot for three attributes.

parallel coordinate plot

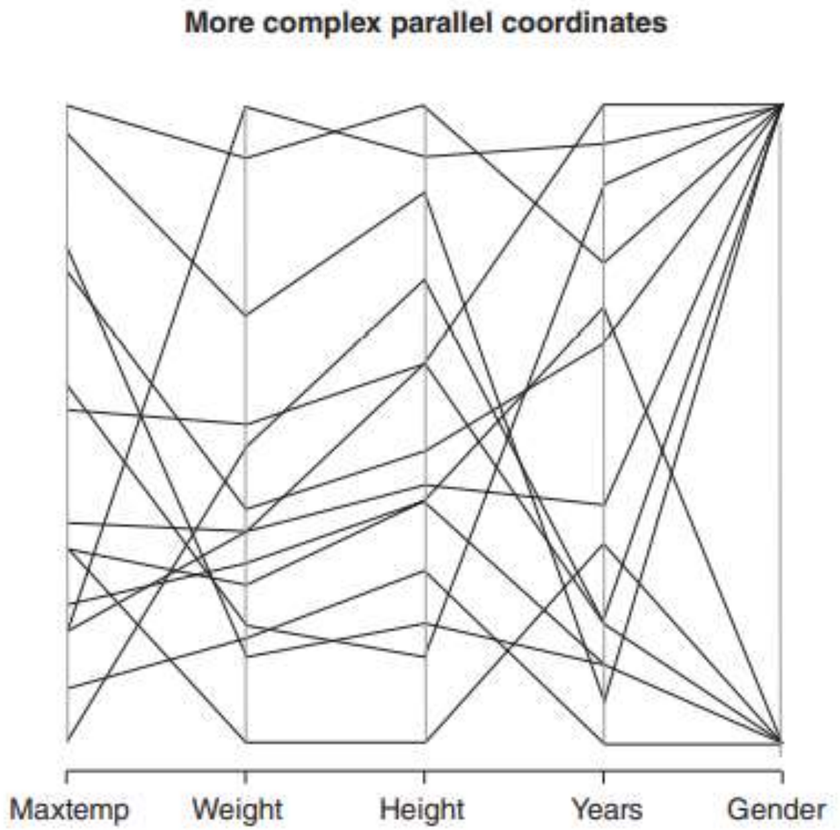


Figure 3.6 Parallel coordinate plot for five attributes.

parallel coordinate plot

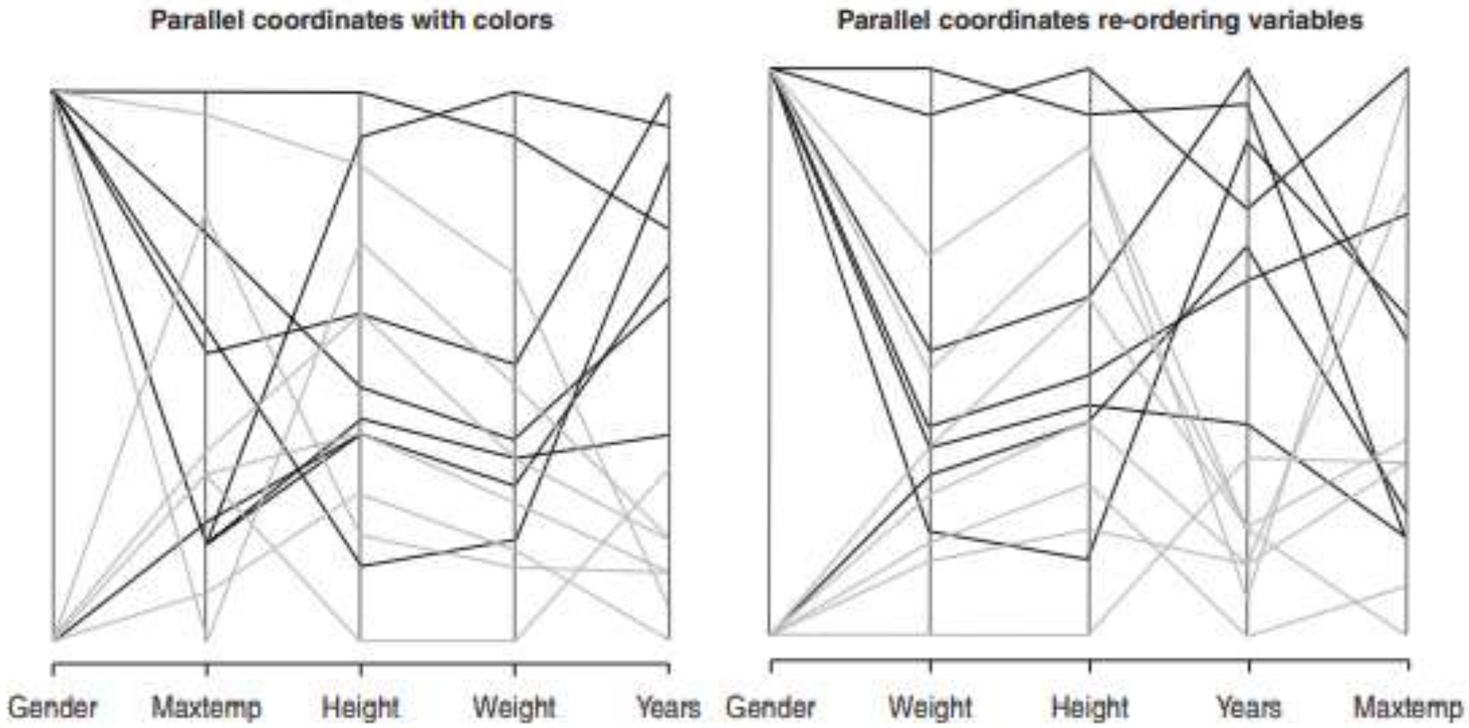


Figure 3.7 Parallel coordinate plots for multiple attributes: left, using a different style of line for contacts who are good and bad company; right, with the order of the attributes changed as well.

Star Plot

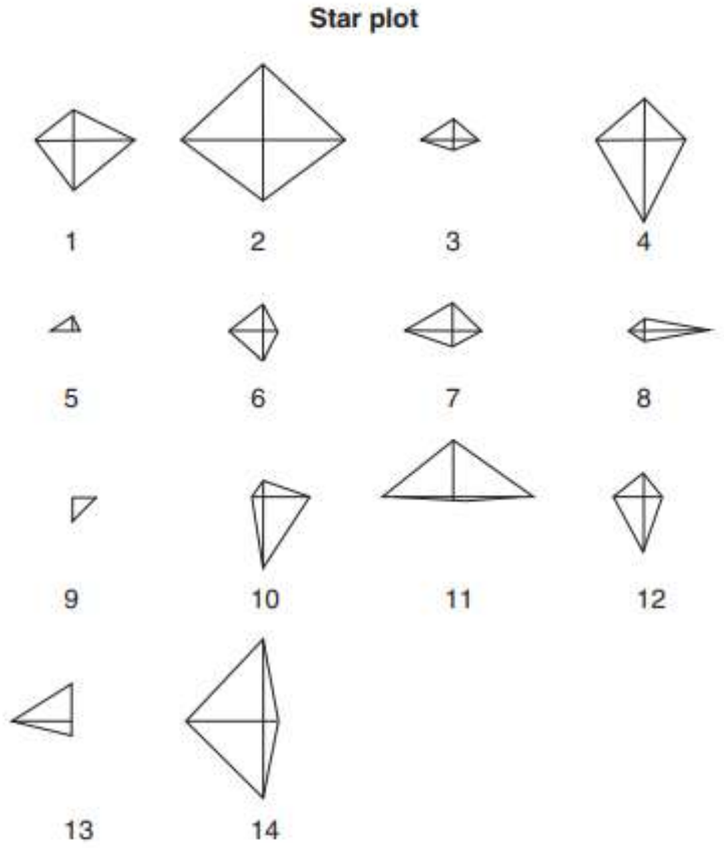


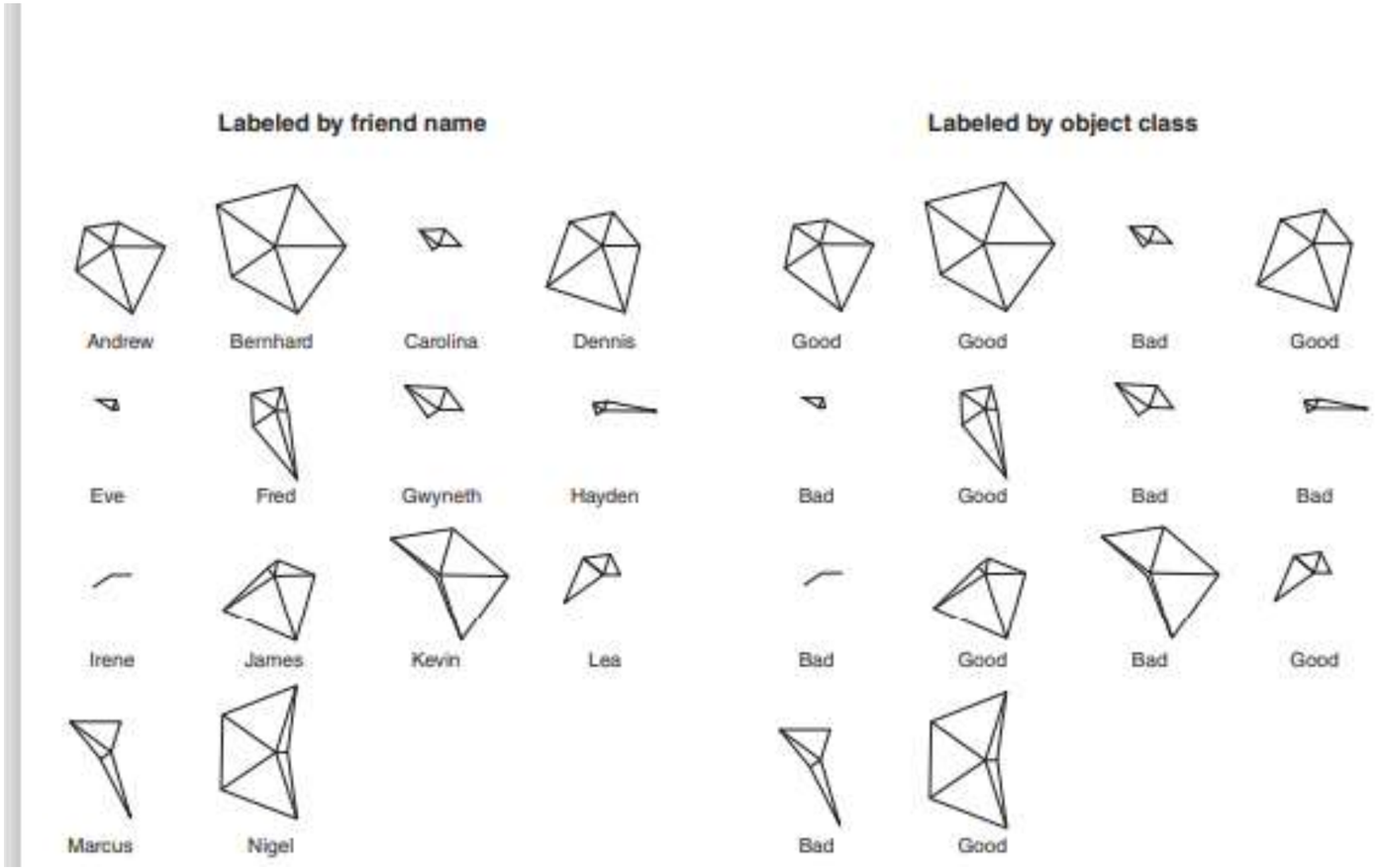
Figure 3.8 Star plot with the value of each attribute for each object in our contacts data set.

Star Plot

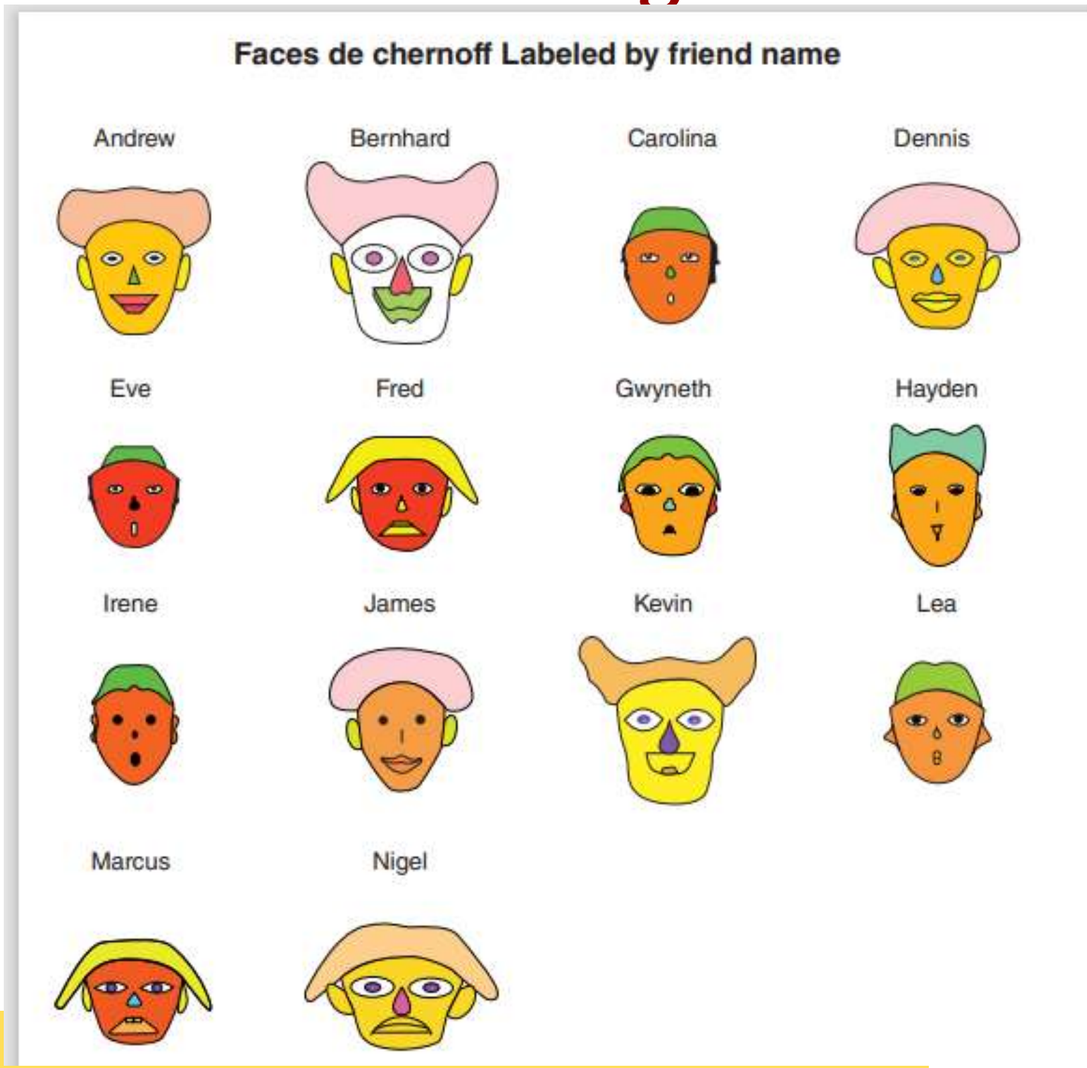
- star plots for four quantitative attributes (maxterm, height, weight and years). To avoid the predominance of attributes with larger values in the plot, all attributes have their values normalized to the interval [0.0, 1.0].
- When the value of an attribute is close to 0.0, its corresponding star will be too close to the center to be see n. This is clear in the star labeled “Irene”

Contact	Maxtemp	Weight	Height	Years	Gender	Company
Andrew	25	77	175	10	M	Good
Bernhard	31	110	195	12	M	Good
Carolina	15	70	172	2	F	Bad
Dennis	20	85	180	16	M	Good
Eve	10	65	168	0	F	Bad
Fred	12	75	173	6	M	Good
Gwyneth	16	75	180	3	F	Bad
Hayden	26	63	165	2	F	Bad
Irene	15	55	158	5	F	Bad
James	21	66	163	14	M	Good
Kevin	30	95	190	1	M	Bad
Lea	13	72	172	11	F	Good
Marcus	8	83	185	3	F	Bad
Nigel	12	115	192	15	M	Good

Star plot with the value of each attribute for each object in contacts data set



Visualization of the objects in our contacts data set using Chernoff faces.



Assessment 1

References

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

Thank You