



SNS COLLEGE OF ENGINEERING

Kurumbapalayam (Po), Coimbatore – 641 107

An Autonomous Institution

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
III YEAR /VI SEMESTER

Unit 2- GETTING INSIGHTS FROM DATA

Topic : Multivariate Statistics





Multivariate statistics



- ✓ Multivariate statistical methods are used to analyze the joint behavior of more than one random variable.
- ✓ Two types
- ✓ Location multivariate statistics
- ✓ Dispersion multivariate statistics



Location multivariate statistics



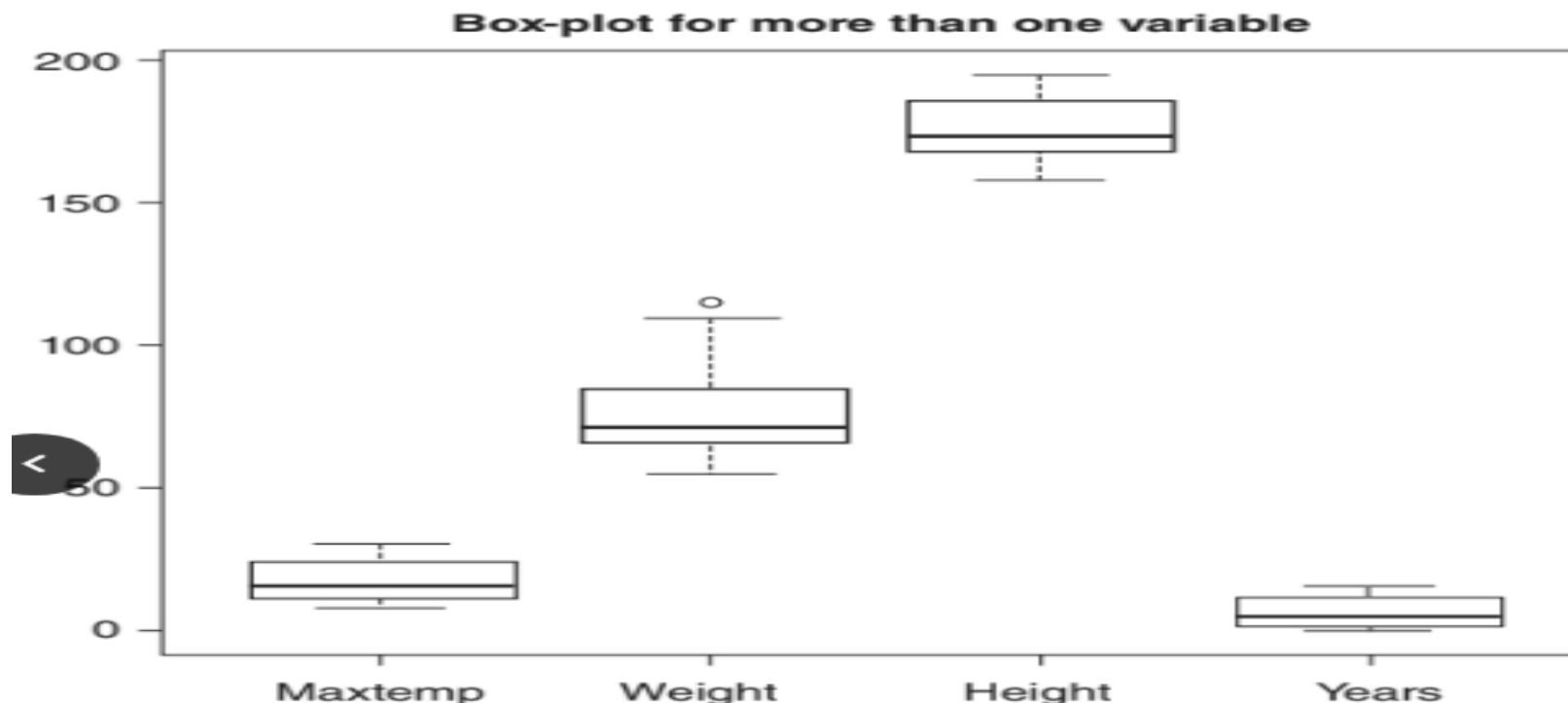
- ✓ To measure the location statistics when there are several attributes we just measure the location of each attribute.
- ✓ Thus, the multivariate location statistical values can be computed independently for each attribute.



Location statistics example



The main location statistics for the four attributes “maxtemp”, “height”, “weight” and “years” from the data set





Dispersion multivariate statistics



multivariate statistics, dispersion statistics, such as the amplitude, interquartile range, mean absolute deviation and standard deviation.

Dispersion statistics	Maxtemp	Weight	Height	Years
Amplitude	23.00	60.00	37.00	16.00
Interquartile range	11.75	17.50	14.75	9.50
\overline{MAD}	7.41	14.09	11.12	6.67
s	7.45	17.38	11.25	5.66



Relationship between two data



The relationship between two attributes is evaluated using the covariance or correlation,

	Maxtemp	Weight	Height	Years
Maxtemp	55.52	34.46	20.19	5.82
Weight	34.46	302.15	184.62	42.39
Height	20.19	184.62	126.53	14.03
Years	5.82	42.39	14.03	31.98



Relationship between two data



Pearson correlation is shown below

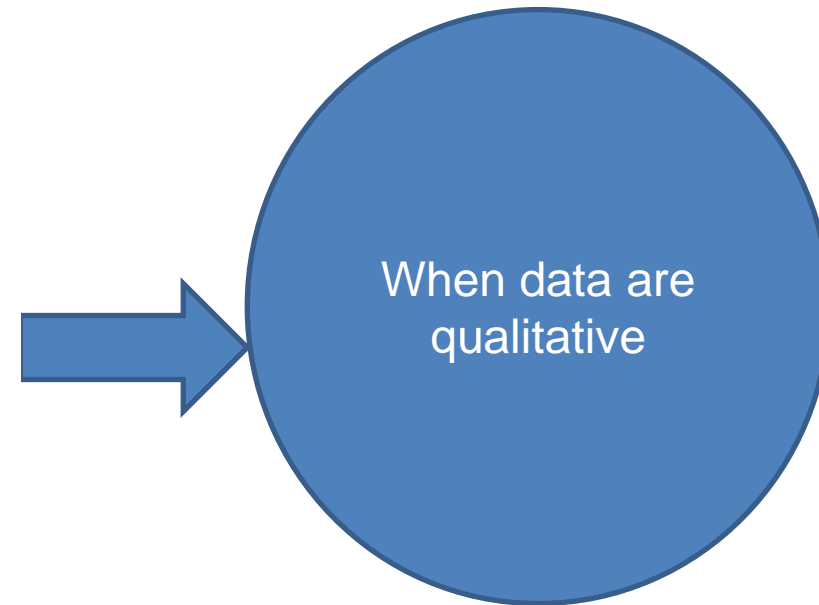
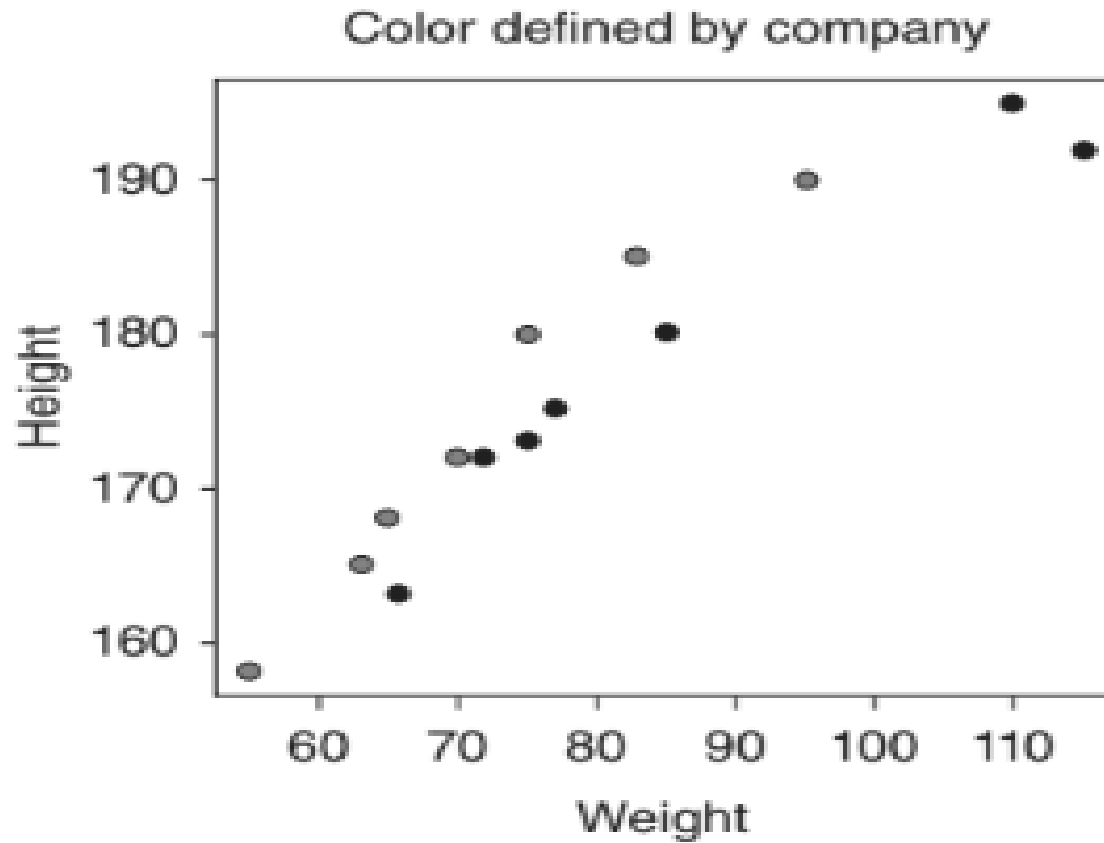
	Maxtemp	Weight	Height	Years
Maxtemp	1.00	0.27	0.24	0.14
Weight	0.27	1.00	0.94	0.43
Height	0.24	0.94	1.00	0.22
Years	0.14	0.43	0.22	1.00



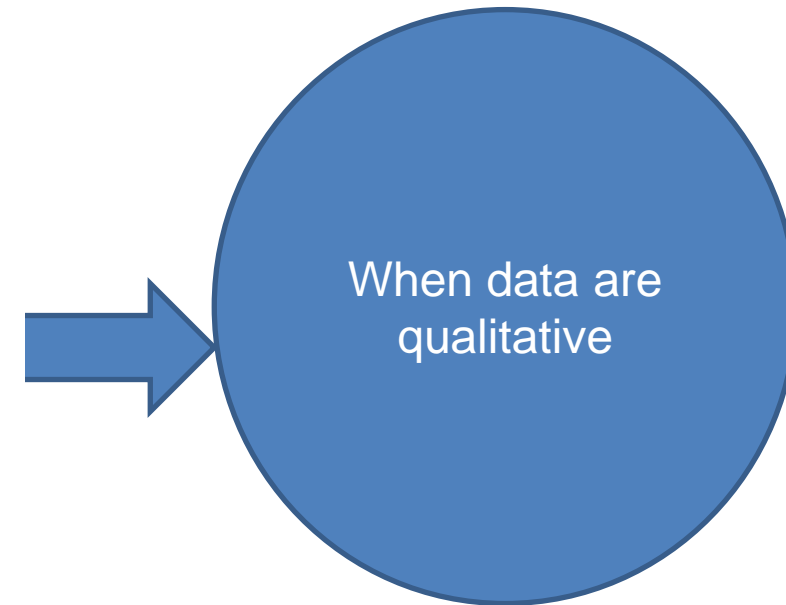
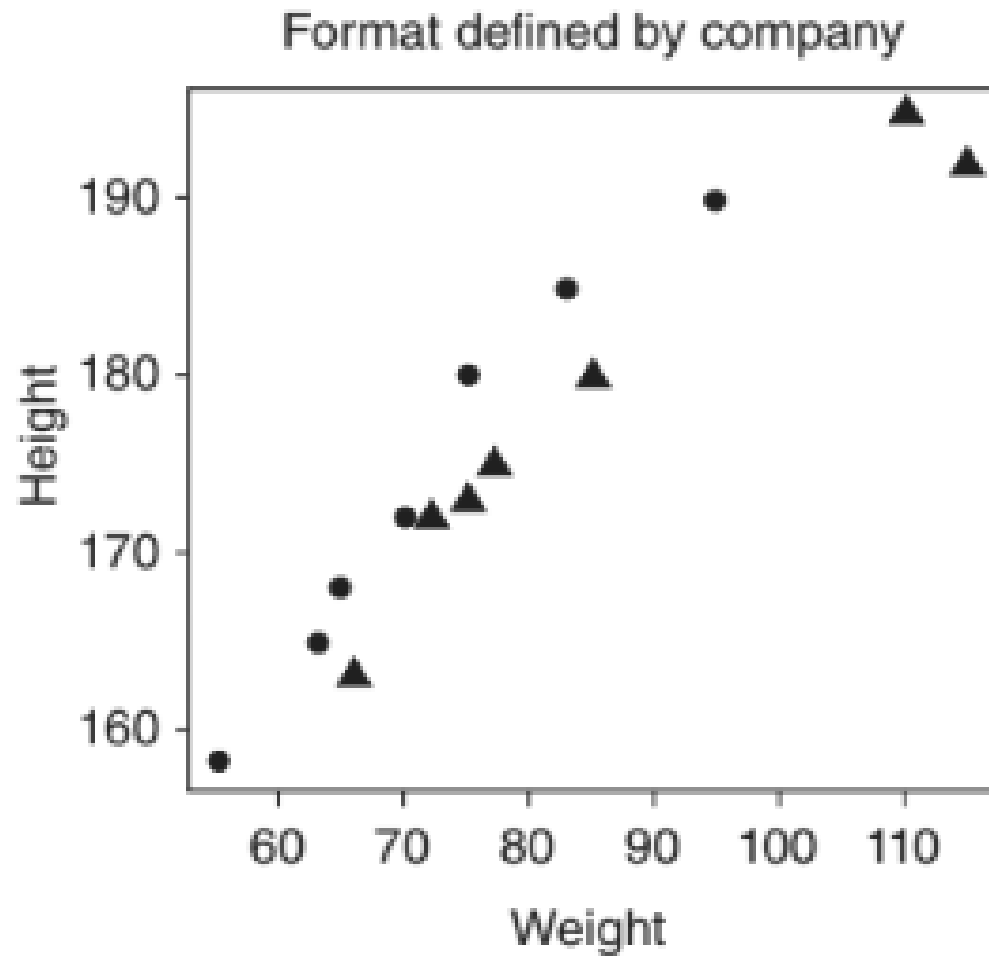
Scatter plot



Multivariate data visualization



Multivariate data visualization



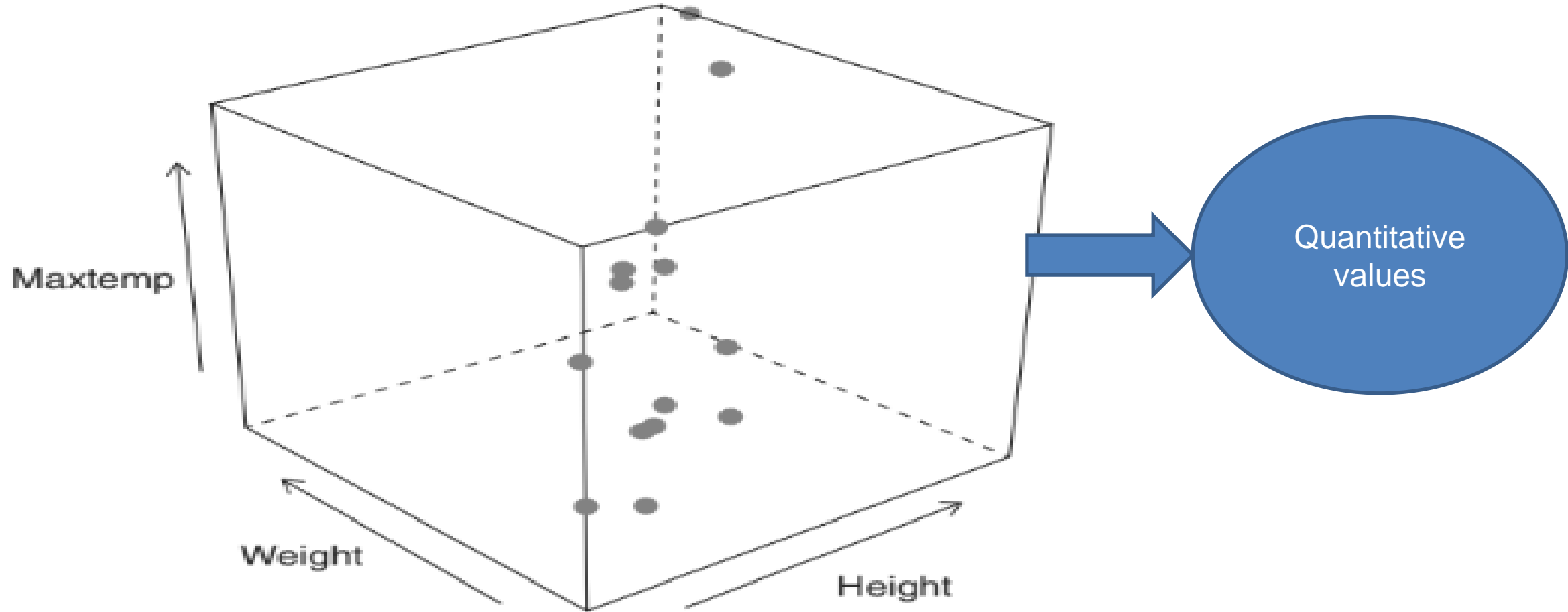


THREE DIMENSIONAL PLOT

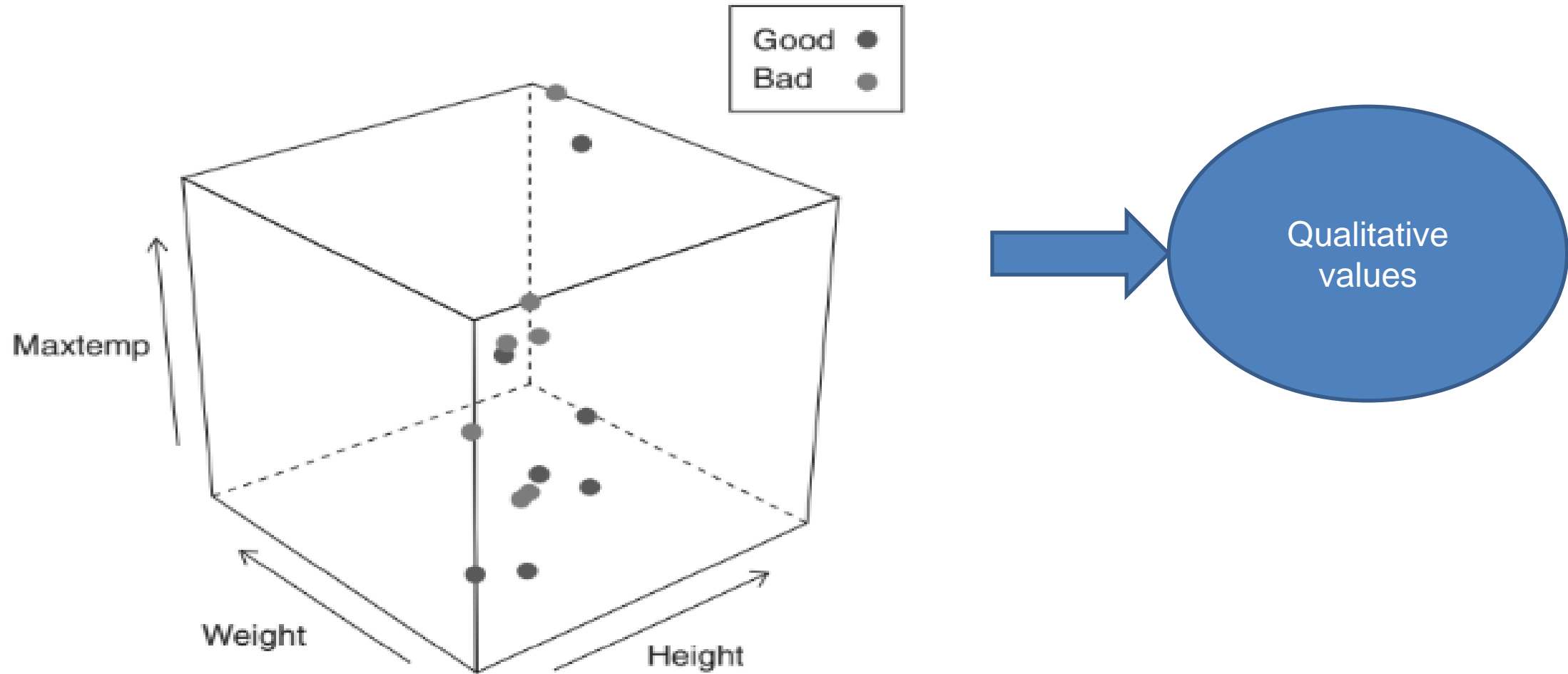


- ✓ Another approach to representing three attributes is to use a three dimensional plot, where each axis is associated with one of the attributes.
- ✓ It makes more sense to use this approach if the three attributes are quantitative,
- ✓ because the values of the corresponding attributes can be presented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

THREE DIMENSIONAL PLOT

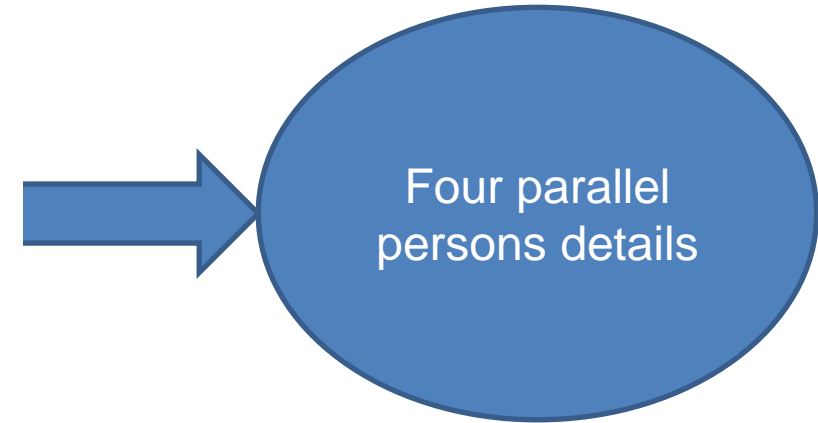
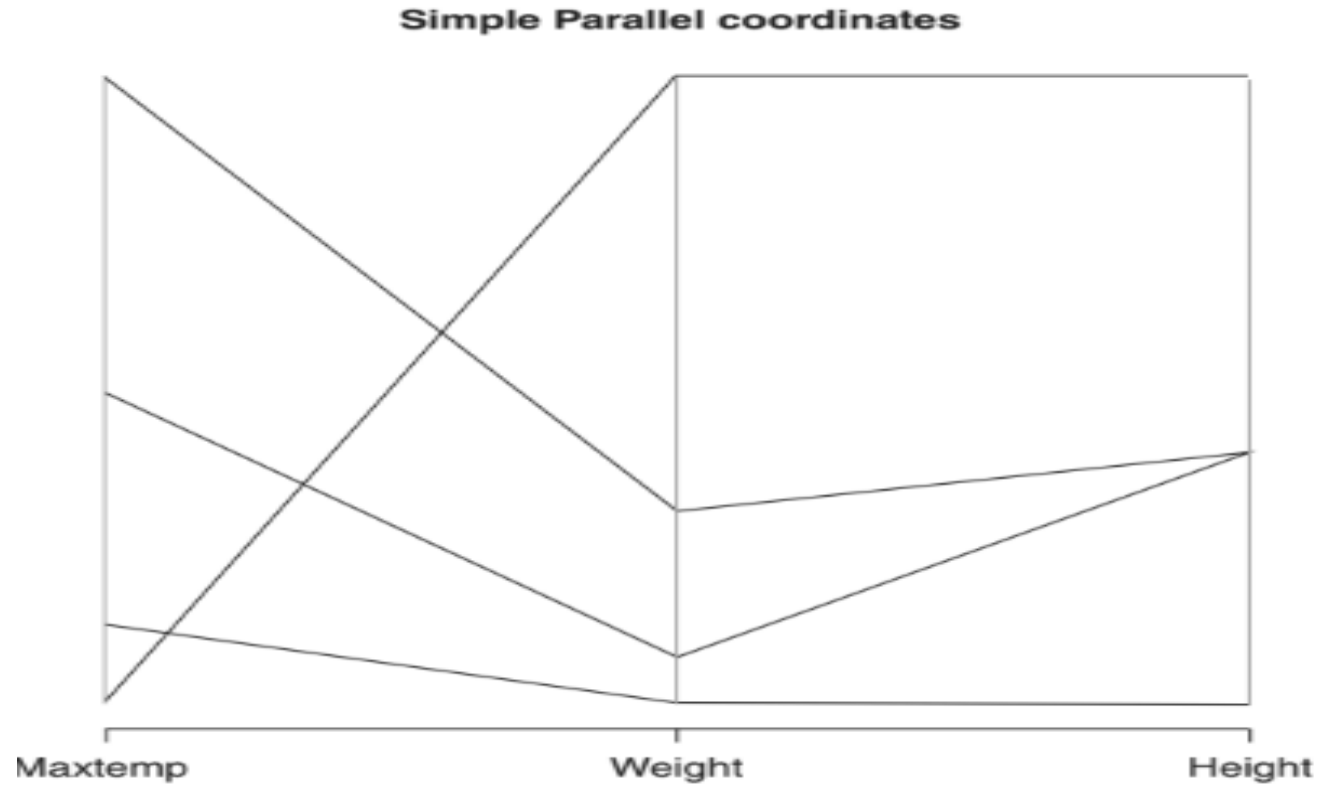


THREE DIMENSIONAL PLOT

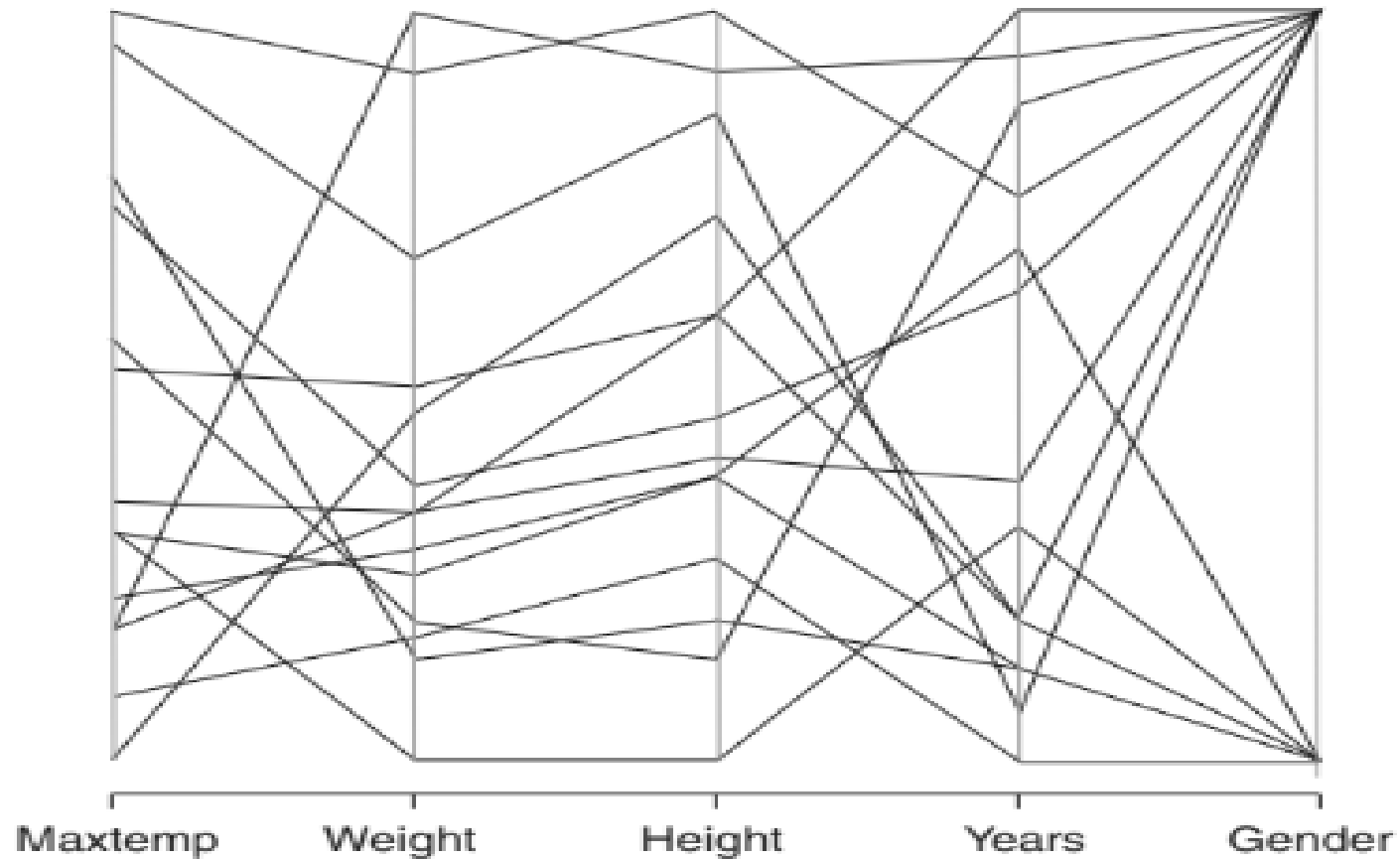




Parallel coordinate plot



Parallel coordinate plot

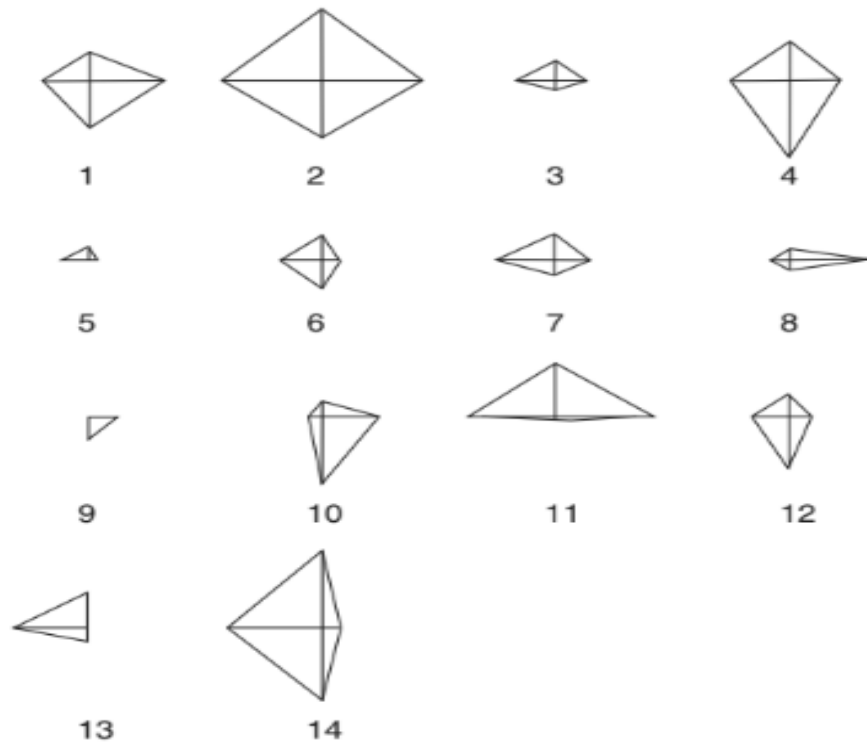


Four parallel persons details for 5 attributes

Also called spider plot

Star plot graph

Star plot



shows star plots for four quantitative attributes (max temp, 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 $[0.0, 1.0]$.



Chernoff face



- ✓ Chernoff faces, invented by applied mathematician, statistician and physicist Herman Chernoff in 1973, display multivariate data in the shape of a human face.
- ✓ The individual parts, such as eyes, ears, mouth and nose represent values of the variables by their shape, size, placement and orientation.
- ✓ It can be used to detect similarities between different items, but it is not the most efficient or the most accurate way to do so.



Chernoff face



- ✓ Chernoff faces are also useful for clustering,
- ✓ A trend in this area is the development and use of interactive visualization plots, where the user interacts with the plot to make the visual information
- ✓ more useful.

Chernoff face

Andrew



Bernhard



Carolina



Dennis



Eve



Fred



Gwyneth



Hayden



Irene



James



Kevin



Lea



Marcus



Nigel





Assessment 1



1. What are the most appropriate scales for the following examples?

- university students' exam marks
- level of urgency in the emergency room of a hospital
- classification of the animals in a zoo
- carbon dioxide levels in the atmosphere.

2. Present the absolute and relative frequencies and cumulative frequencies for the attribute "weight"





References



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

Thank You