

Data Representation

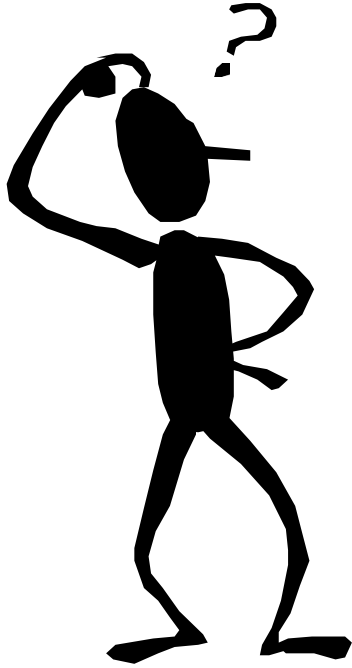
- KPIs visualised more clearly
- Unify data
- Apply AI driven analytics to reveal relationships between KPIs

Overview

- illustrative task of selecting one object from many on the basis of the numerical values of their attributes
- Example:
- What are the Attributes considered for buying a car?

- Selecting item based on numerical values of their attributes
- Attributes(variables) for buying a car?
 - Price
 - passenger capacity
 - petrol consumption

Visualisation Designer Challenge



- Presenting the data
- Allowing it to be rearranged

Conclusion: Decision should be facilitated

Dimensionality

- High volume of data in Commerce and Industry
- Attributes increases complexity increases- Dimensionality of Problem

Designer:

how many features can be incorporated?

- Techniques available for handling high dimensionality?

Napolean march map example

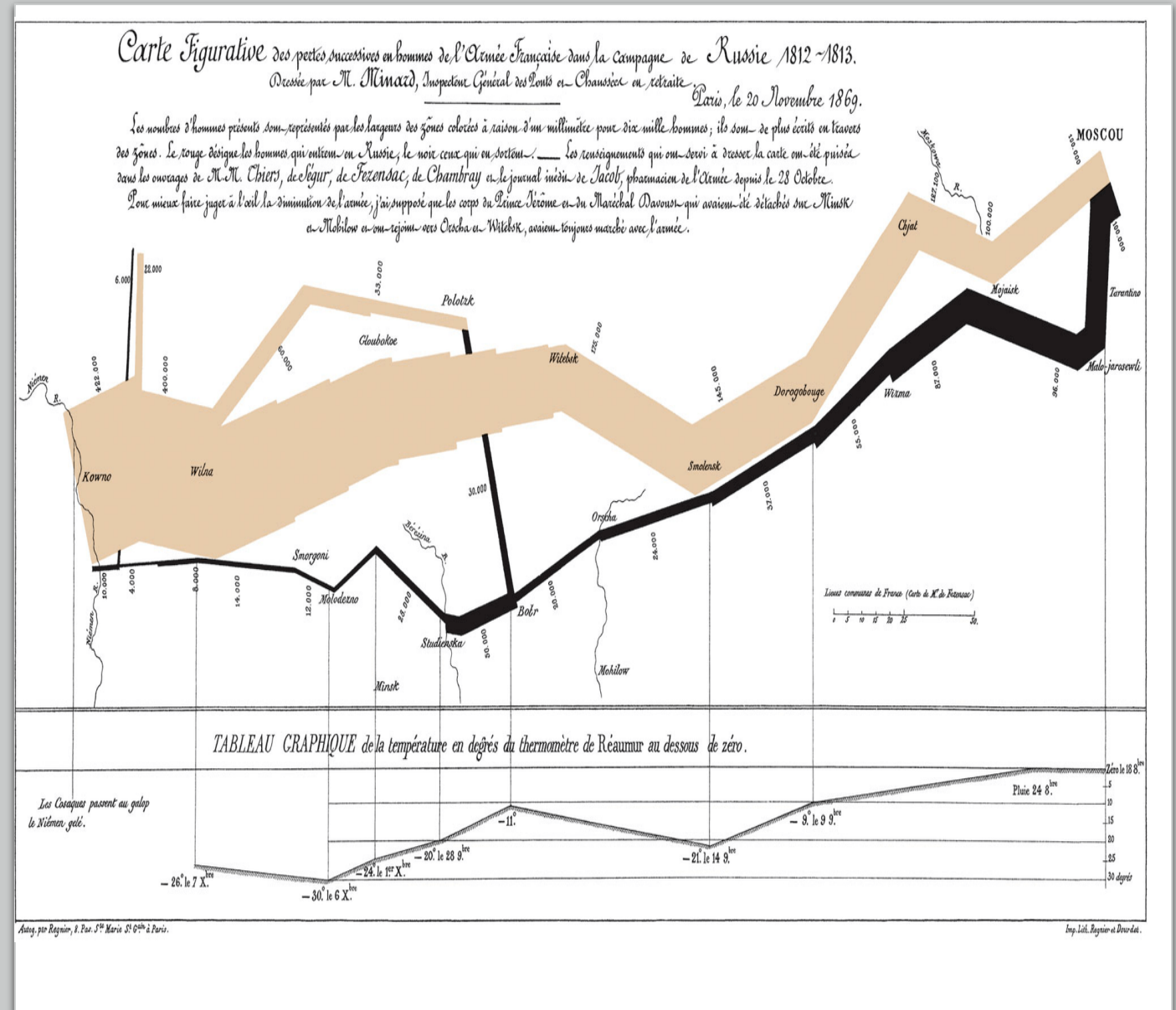
location,

temperature,

size of army,

direction,

etc.



Dimensionality

- Dimensions: Number of variables or attributes
- Univariate data - 1 variable
Car: cost
- Bivariate data - 2 variables
Car: cost, model
- Trivariate data - 3 variables
Car: cost, model, year
- Hypervariate or multivariate data - more than 3 variables
Car: cost, model, year, make, miles for gallon, no. of cylinders, weight, ...

Univariate Data

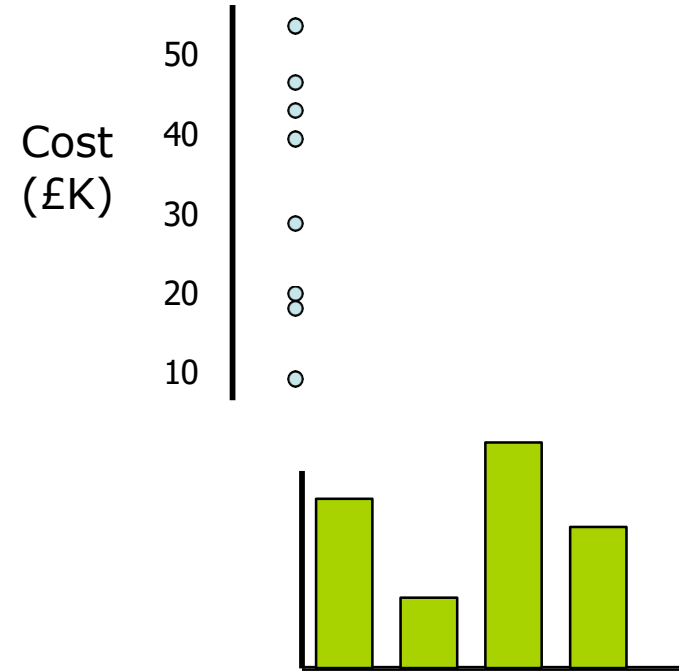
Car	Price (£)
BMW	51,395
Mercedes	50,850
Mercedes	41,000
Saab	39,085
BMW	38,000
Ford	36,950
BMW	34,550
Morgan	32,000
Rover	31,300
Ford	29,250

Univariate Data

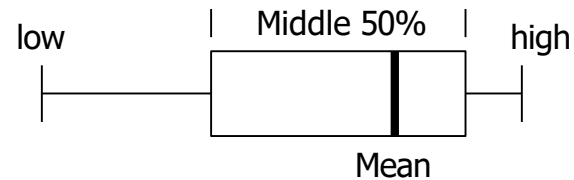
Saab	28,750
Mercedes	28,000
Ford	27,600
Ford	25,950
Land Rover	24,000
Jeep	23,200
Nissan	20,000
Vauxhall	18,500
Nissan	17,400
Ford	17,000
Ford	16,500
Nissan	15,500
Ford	14,950
Ford	12,000

Univariate Data

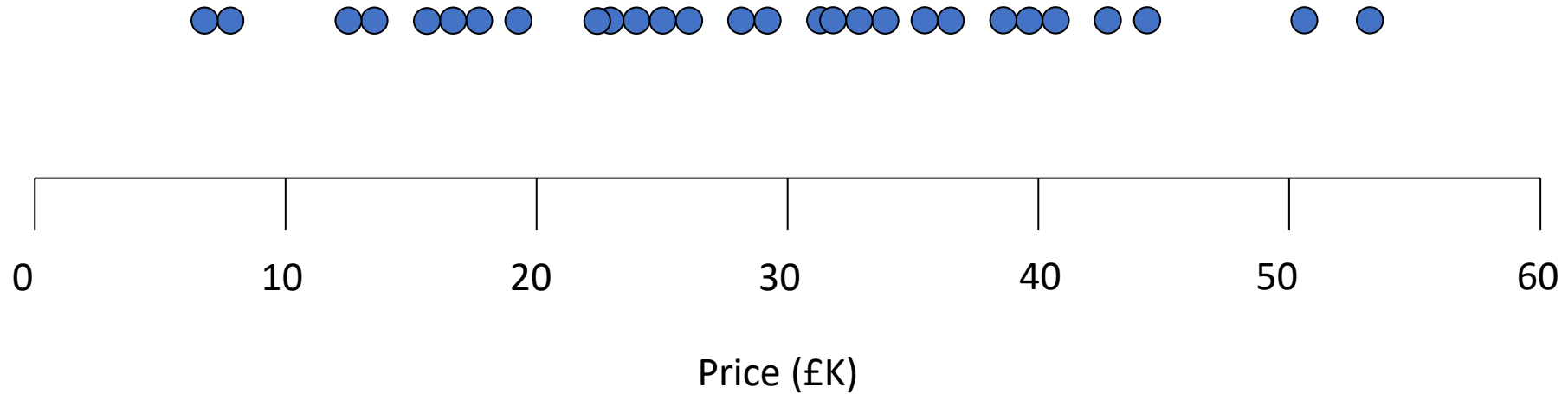
- Different representations
- In form of points against some scale
(points can be labeled)
- In forms of aggregation:
Histogram



box and whisker
plot

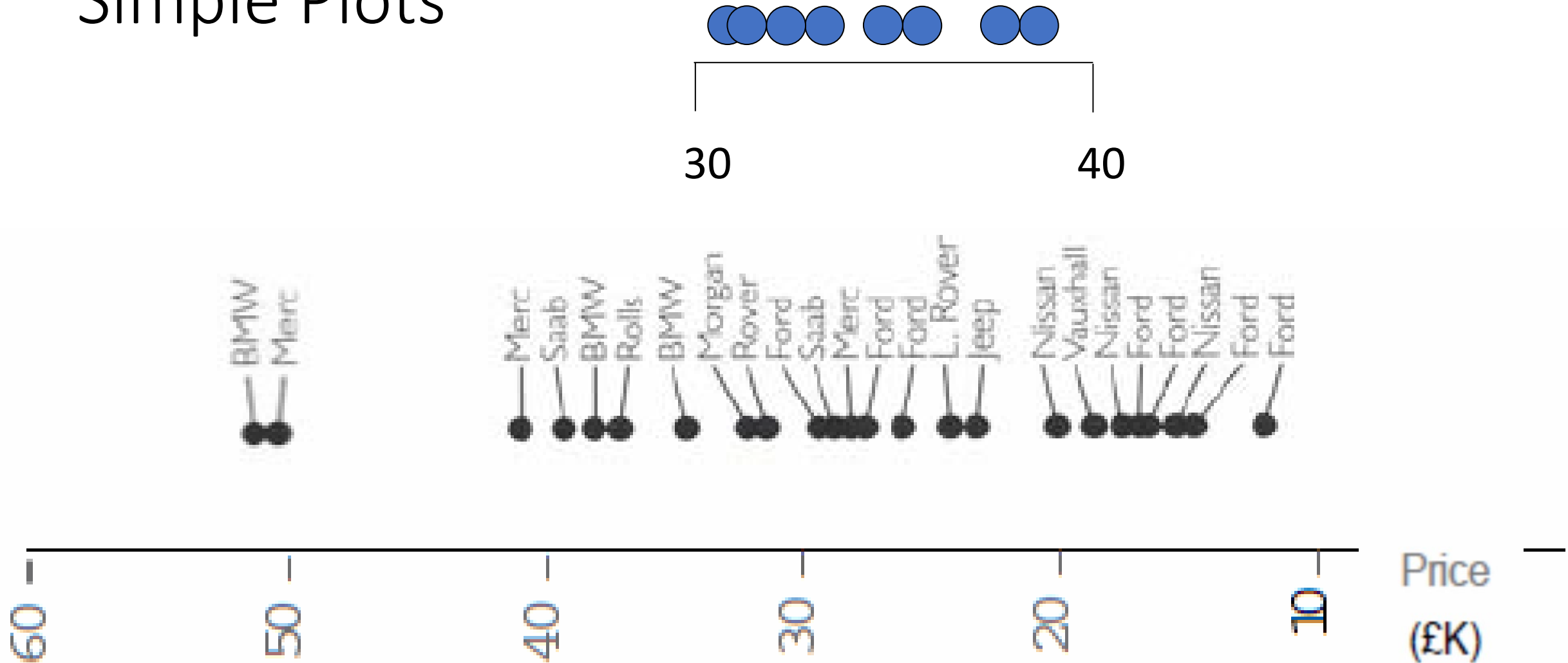


Simple Plot

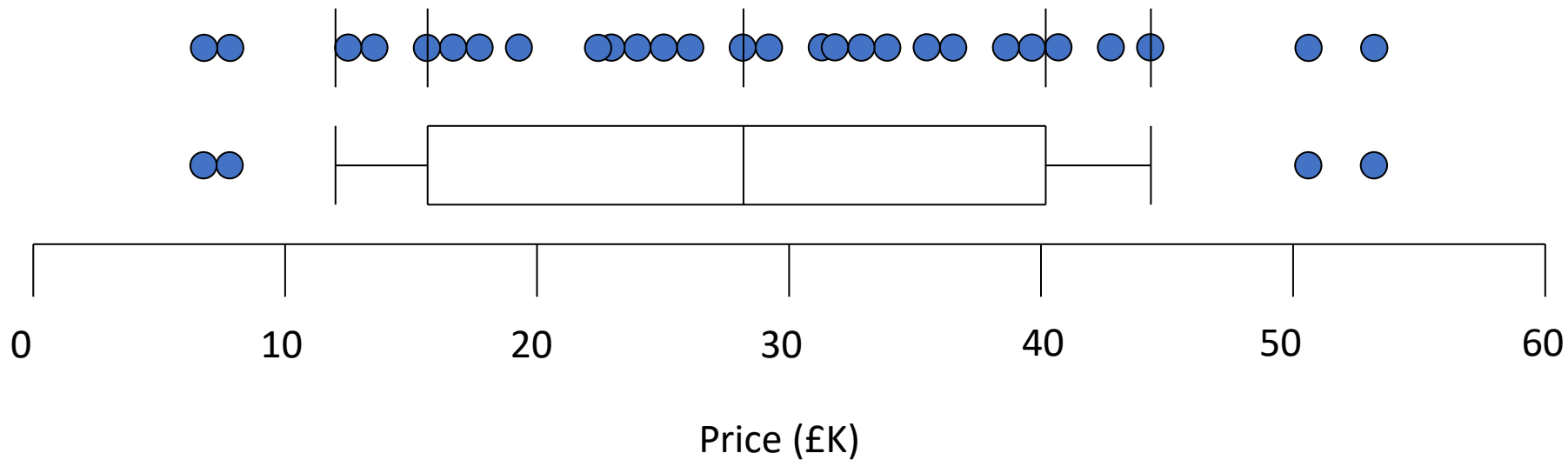


- Details identified:
- Lowest and highest values
- General distribution of points
- Bunching of points

Simple Plots

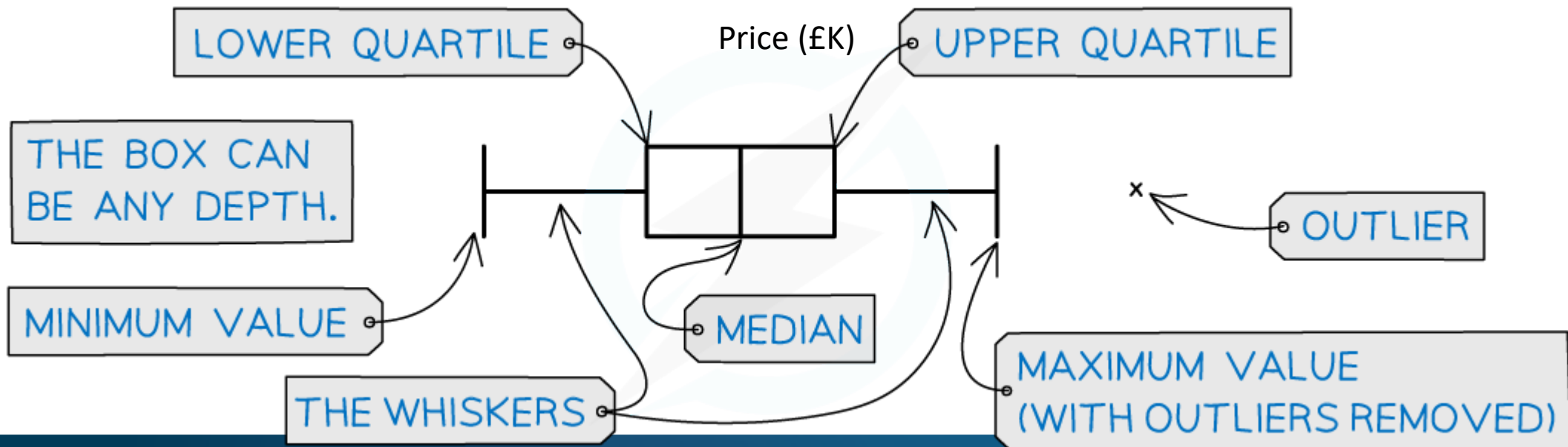
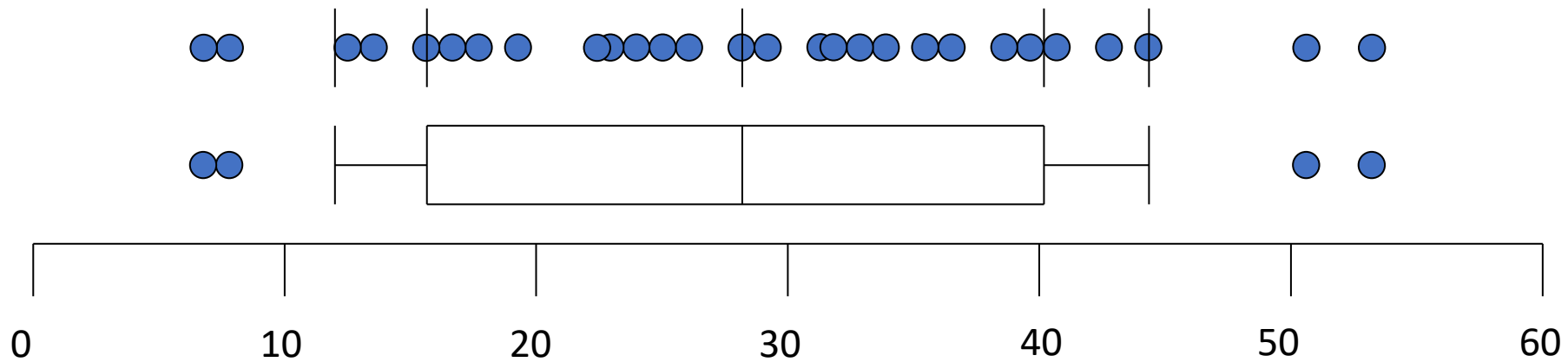


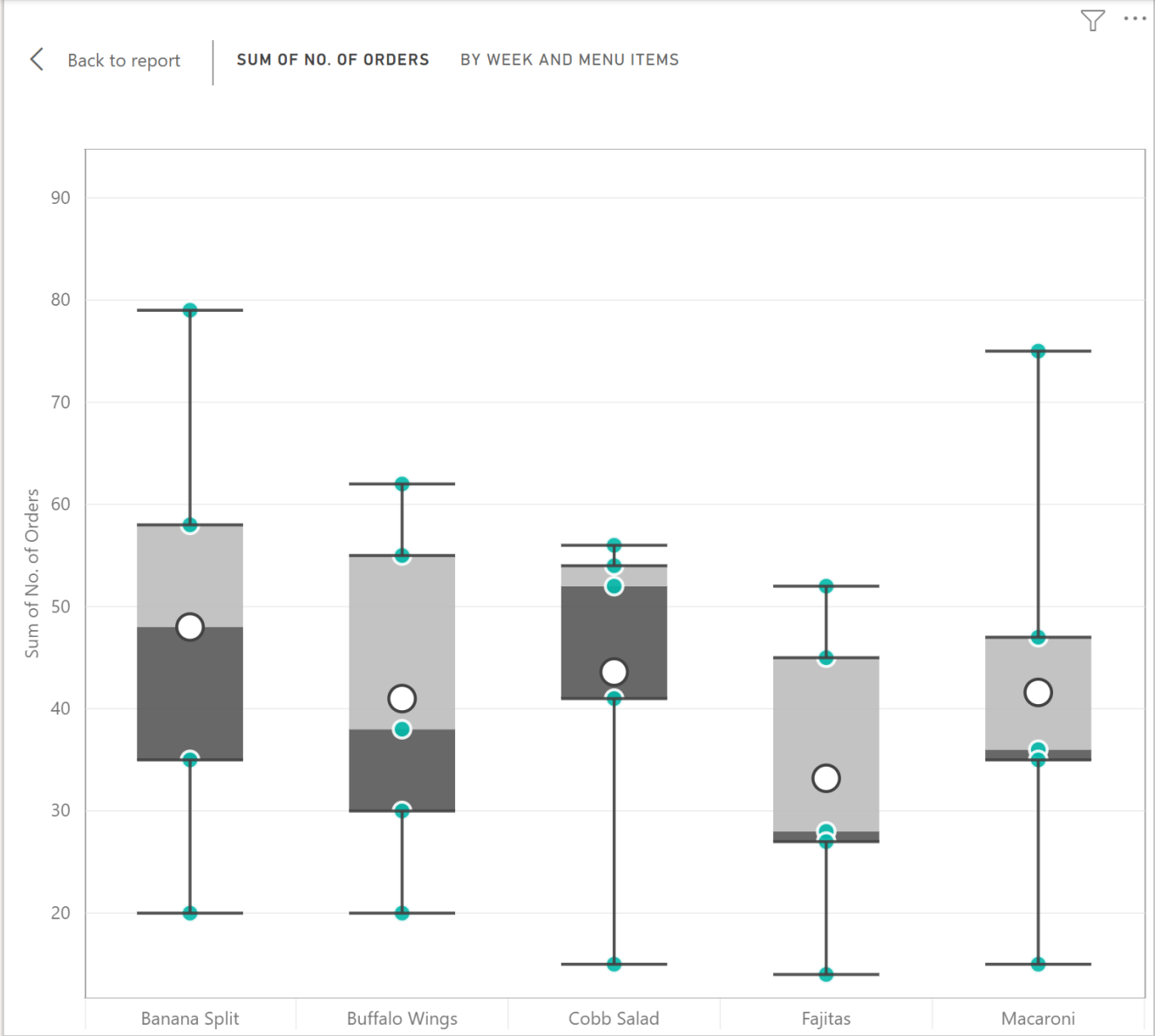
Tukey Plots/Box and whisker plot



- Aggregation introduced by Tukey Box Plot
- This shows
 - The 25th, 50th and 75th percentiles as a box
 - The 10th and 90th percentiles as lines
 - Anything beyond these as 'outliers'

Tukey Plots





Filters

Search

Filters on this visual

Week is (All)

Sum of No. of Orders is (All)

Menu Items is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

Add data fields here

Visualizations

Build visual

Visual icons

Fields

Search

financials

Orders

Sheet1

Menu Items

No. of Orders

Week

Week

Axis category I

Menu Items

Axis category II

Add data fields here

Value

Sum of No. of Orders

Dots size

Add data fields here

Legend

Add data fields here

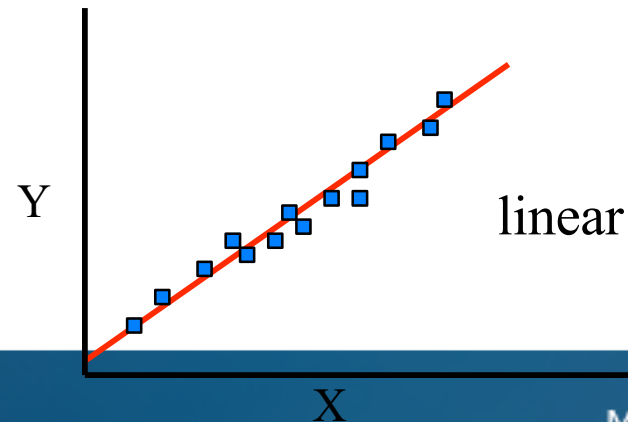
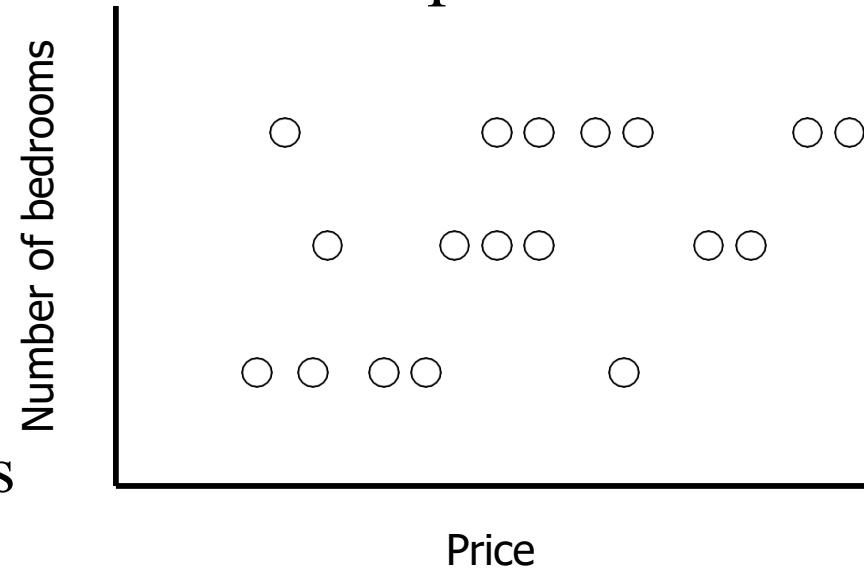
Exercise

- Open the file class exercise for univariate data.xlsx in moodle and complete the task listed

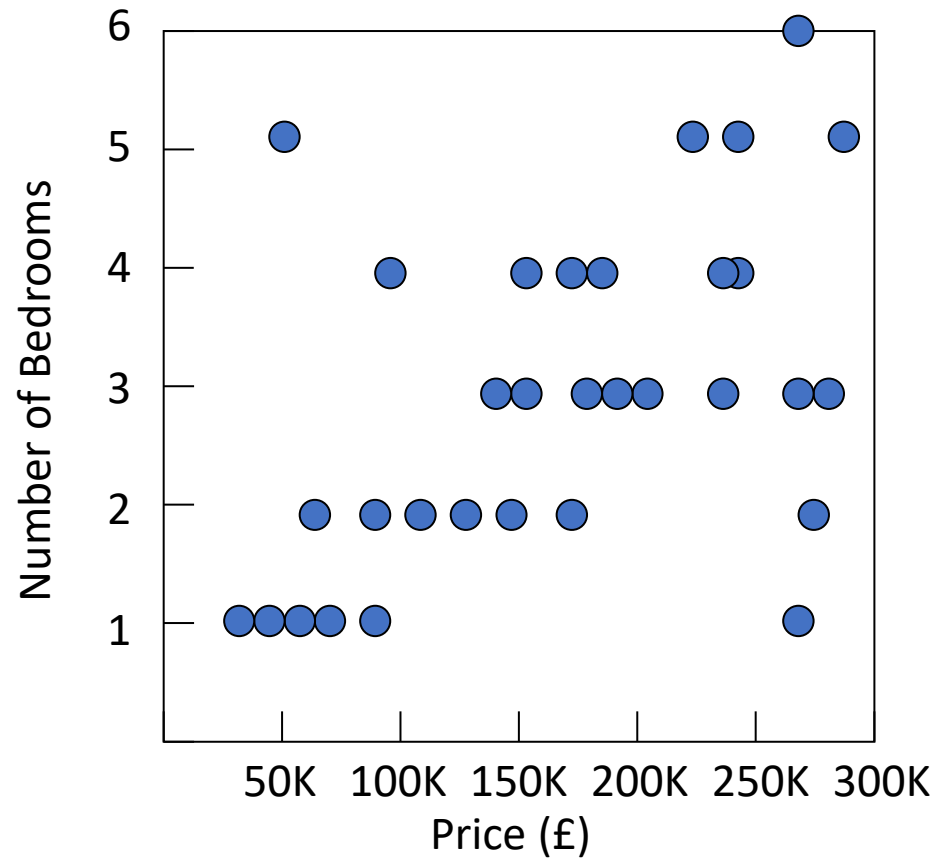
Bivariate Data

Example: house-hunting

- Scatter plot of one variable against other
 - In forms of aggregations or groups
- Two histograms
- Two box plots

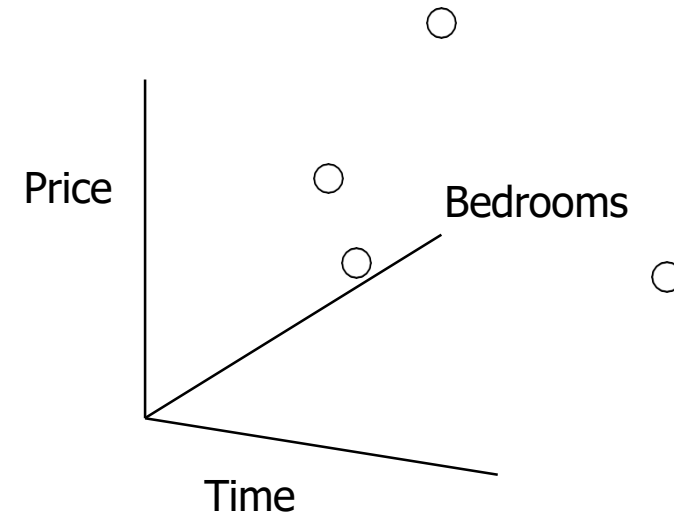


Boxplot-Bivariate Data



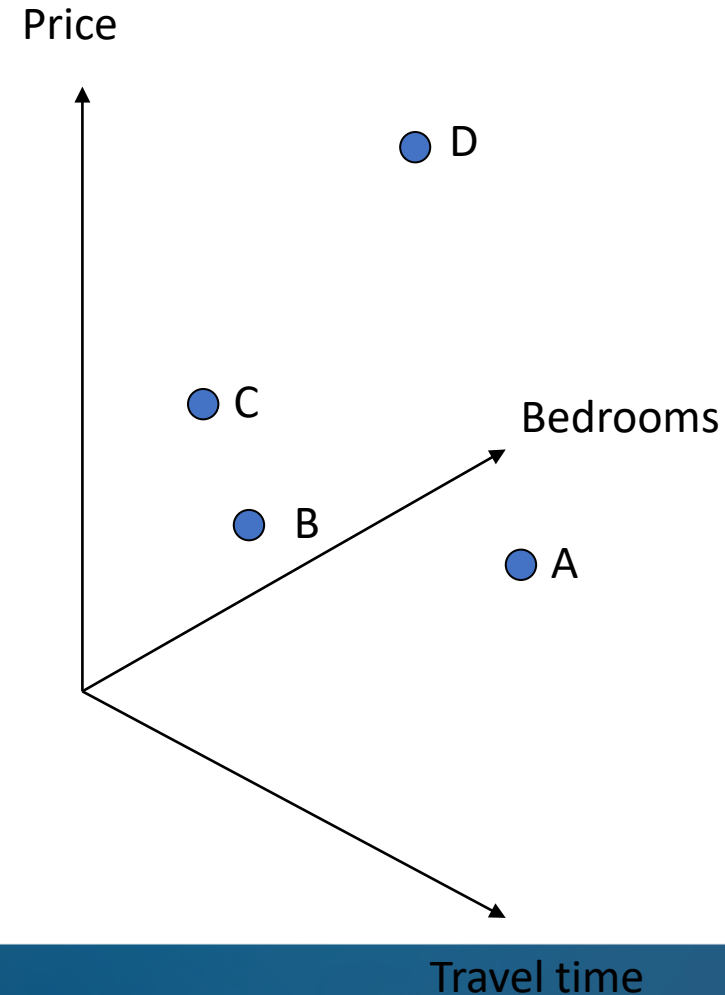
Trivariate Data

- 3D world in 2D graphic representation
- Scatter plot showing three axes

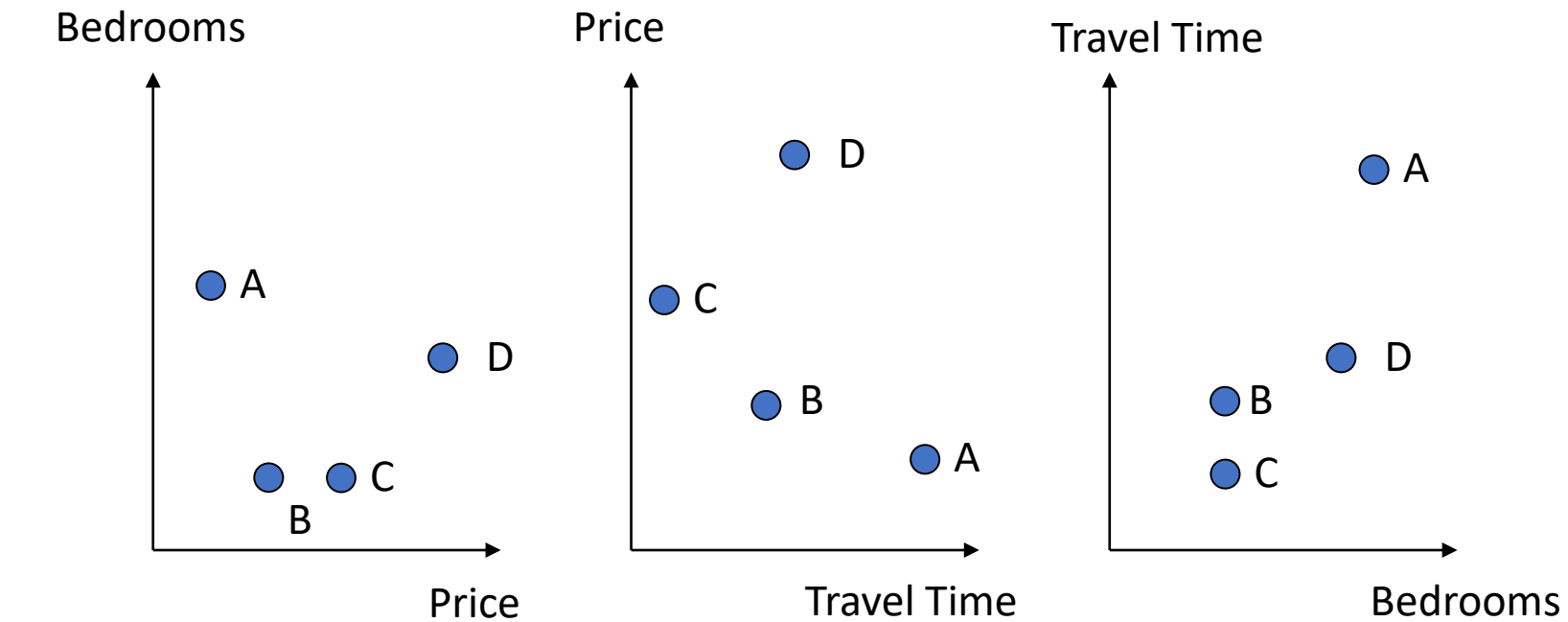


Trivariate Data

- Can we decide if A has a greater value of Price than B?
- No
- We could aid this by projecting each point onto each axis



Trivariate Data



Projection of the points onto all pairs of axes

Scatterplot Matrix



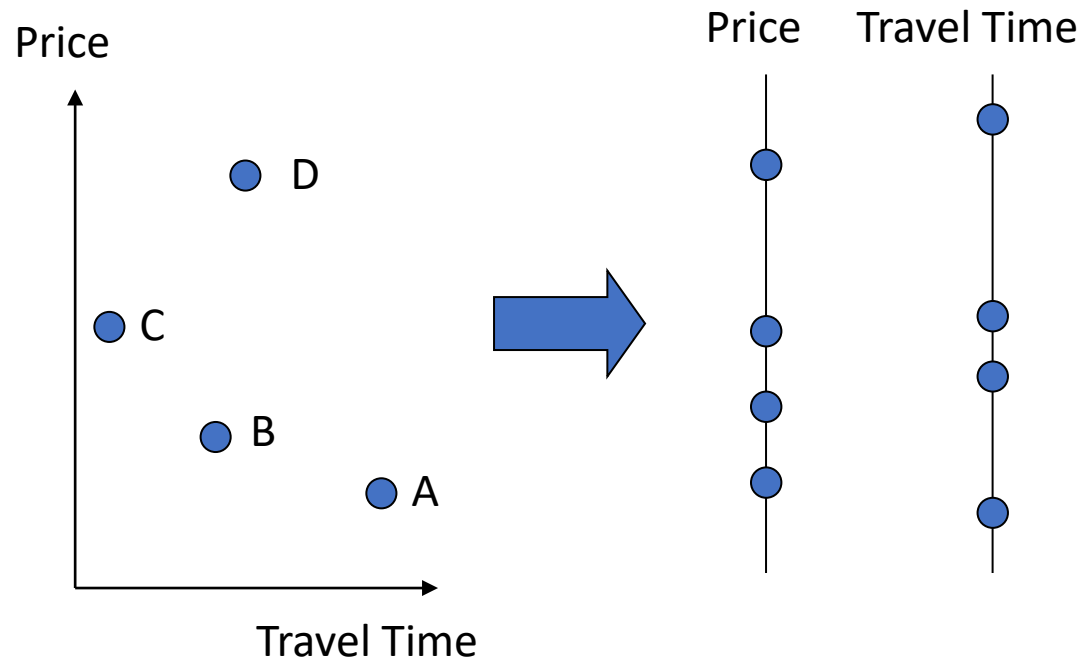
Trivariate Data

- The scatterplot matrix contains as many scatterplots as there are pairs of parameters
- For more than five parameters this becomes unworkable
- If we have N objects with M parameters we end up with $N \times M$ points in total
- Labelling this many points is a problem

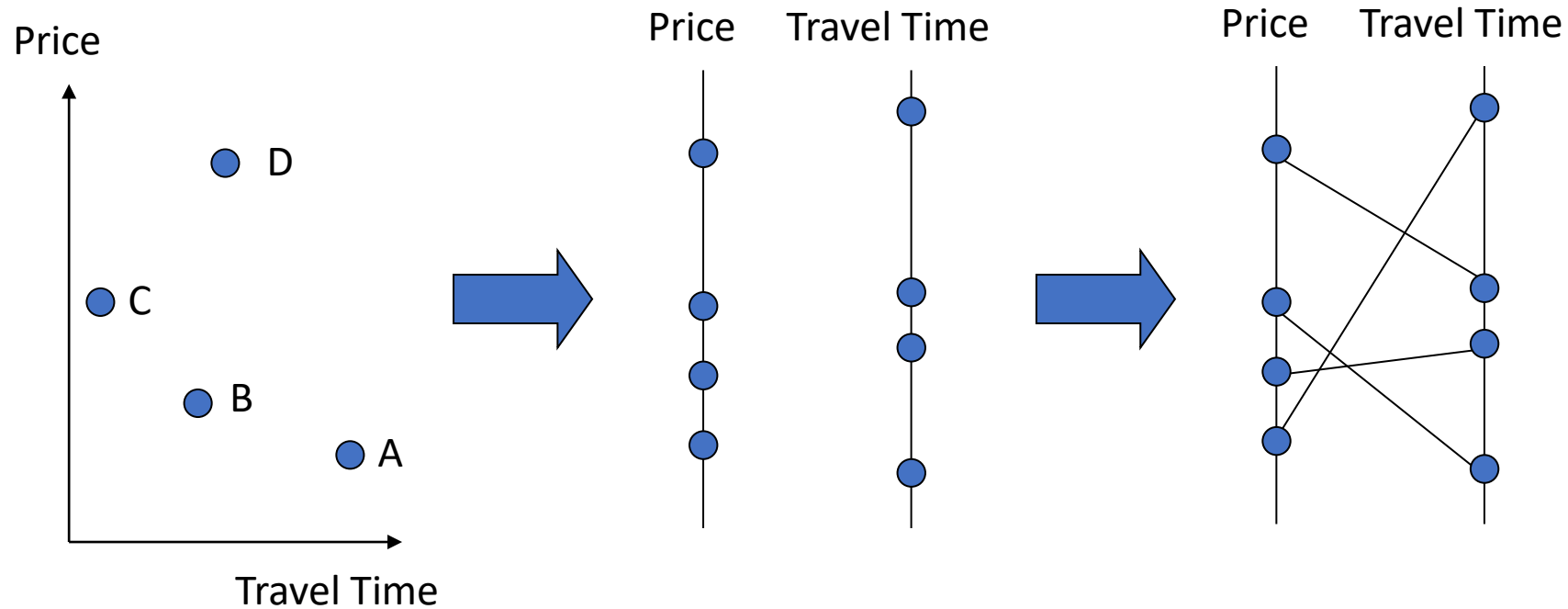
Hypervariate Data (Multivariate data)

- 3 or more variables
- helps in the reduction and simplification of data without losing any important details.
- One solution is to use *parallel coordinate plots*
- Handles large number of variables.
- All axes of multidimensional space arranged in parallel to each other
- Each data point appears once on each axis

Parallel Coordinate Plot

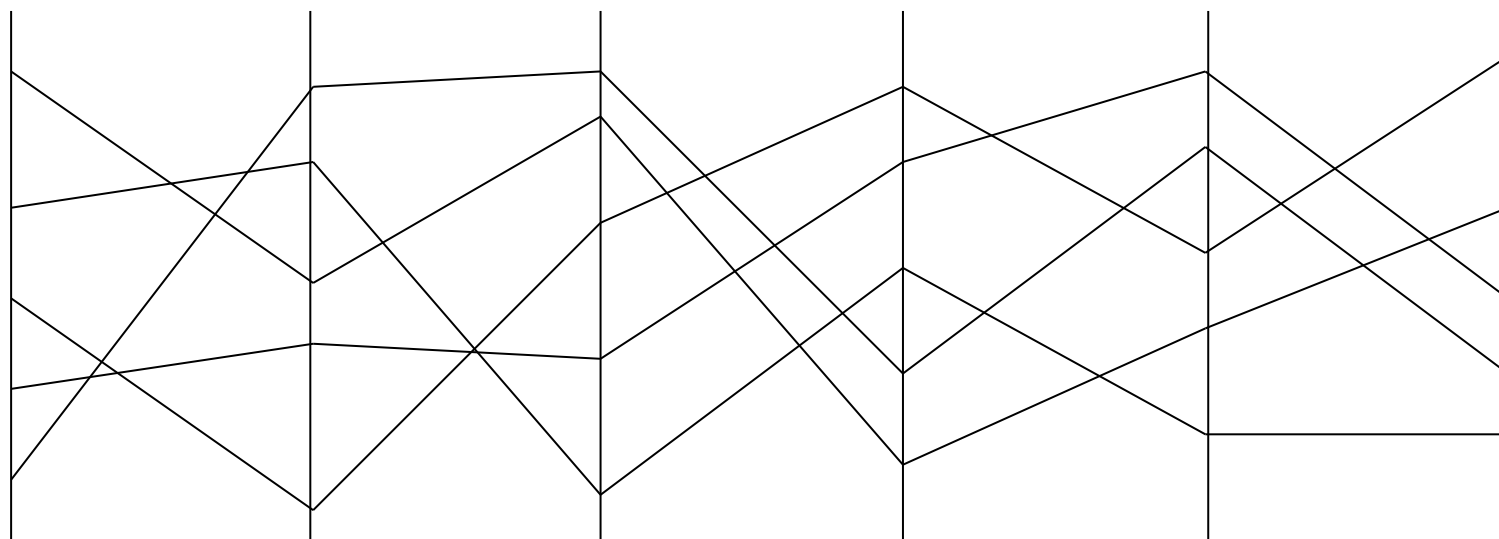


Parallel Coordinate Plot



Multivariate Data

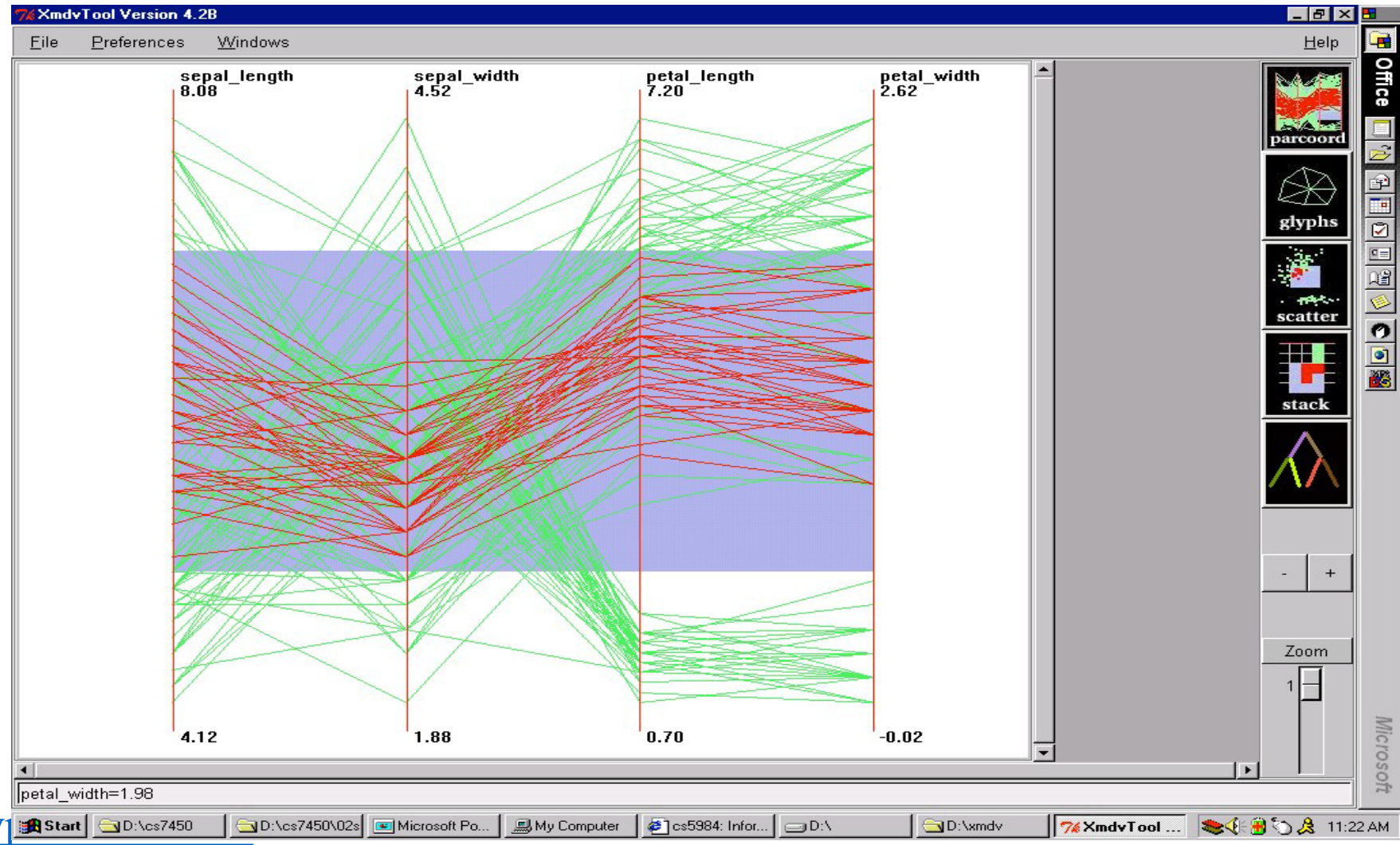
Parallel Coordinate plots for six variables



XmdvTool

XmdvTool is a public domain software for interactive visual exploration of multivariate datasets

Includes parallel coordinates

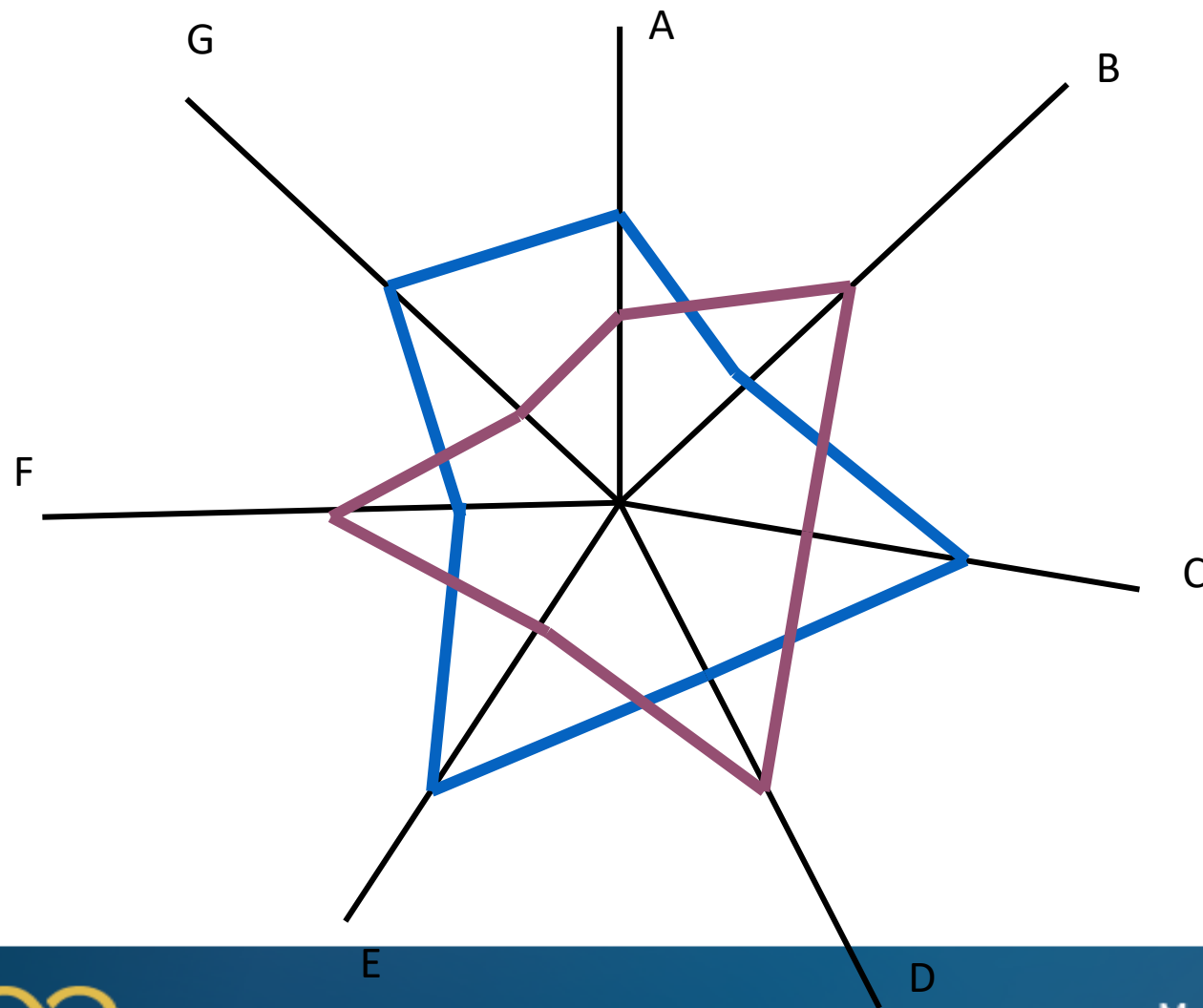


<http://davis.wi>

Hypervariate Data

- A variation on this approach is to use a *starplot*
- In this case the axes radiate from a common origin
- Multiple objects can be compared on the basis of their shapes

Starplot



Exercise

	Area (km ²)	Coastline (km)	Land boundaries (km)	Population (July 2005)	Median age (years)	Life expectancy (years)
Iceland	103,000	4,970	0	296,737	34	80.19
Finland	338,145	1,250	2,681	5,223,442	40.97	78.35
Sweden	449,964	3,218	2,233	9,001,774	40.6	80.4
Norway	324,220	25,148	2,542	4,593,041	38.17	79.4
Denmark	43,094	7,314	68	5,432,335	39.47	77.62

Produce the following diagrams from the data

1. A scatterplot of area v. population
2. A scatterplot matrix of area v. coastline v. land boundaries.
3. A parallel plot of all six attributes
4. A starplot of all six attributes

In each case, 'brush' Iceland in blue and Sweden in red.

Are there any interesting relationships within the data?

Summary

- Univariate
 - 1 dimension, Tukey Box Plots
- Bivariate
 - 2 dimensions, scatterplots, histograms
- Trivariate
 - 3 dimensions, scatterplot matrix, brushing
- Hypervariate
 - >3 dimensions, parallel coordinate plots, starplots