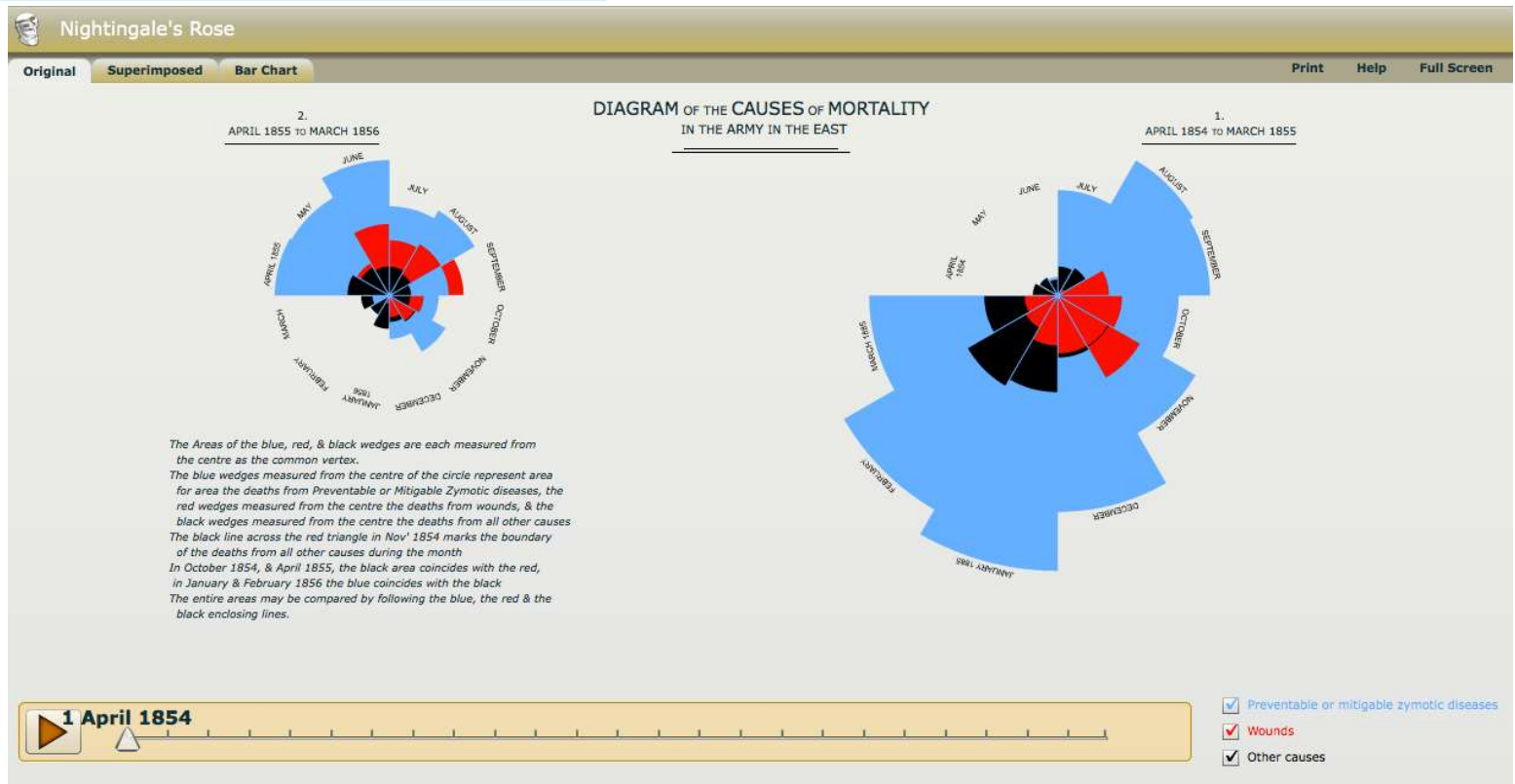THE UNIVERSITY OF
WESTERN AUSTRALIA

MGMT5504

SEMESTER 2 2022– LECTURE 2

Recap

What Does Statistics Mean?

↘ *Descriptive statistics*

Collecting, presenting and describing data

- E.g., number of people, trends in employment
- Numerical and graphical tools

Recap

➢ *Inferential statistics*

Make an inference (draw conclusions and/or make decisions) about a population from a sample

- Population Parameter vs Sample Statistics
- Estimation
- Hypothesis testing

Sample statistics (known)

Population parameters (unknown but can be estimated from sample statistics)

Sample

Population
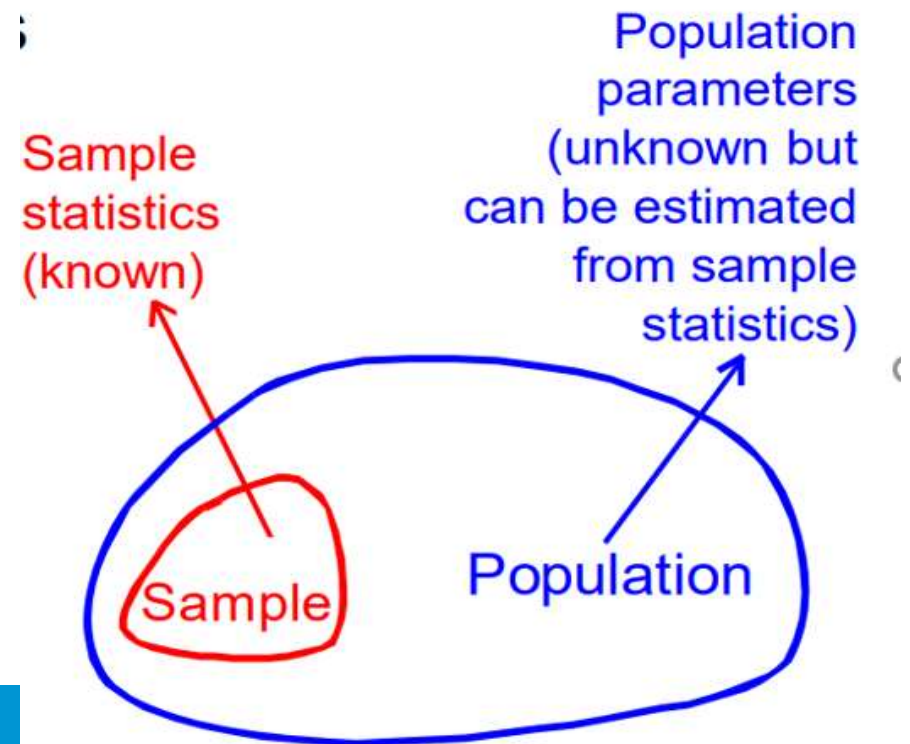
Recap

Sampling Method

Probability Sampling

Simple Random sampling
Stratified Random sampling
Cluster sampling

Non-probability Sampling

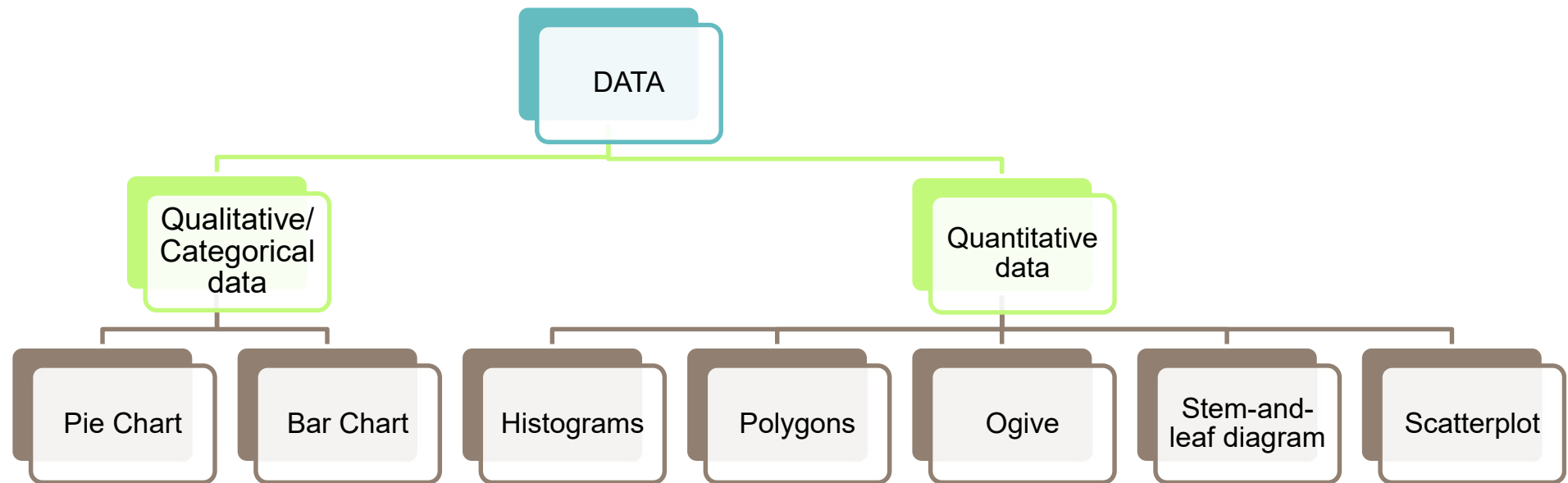| Snowball sampling |
| --- |
| Convenience sample |

Recap

Types of Data/Variables

↘ Cross-sectional and time-series/longitudinal

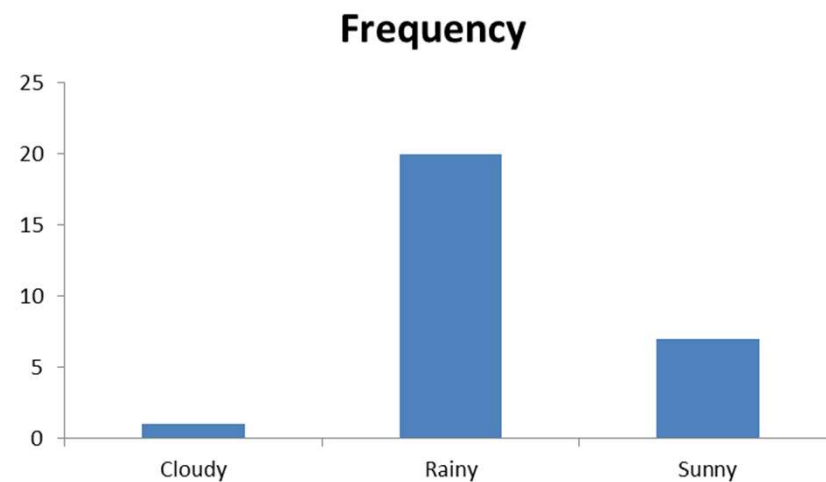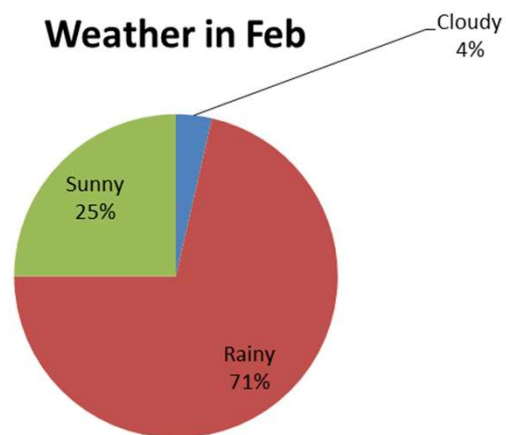↘ Categorical – nominal, ordinal

↘ Continuous/Numerical – interval, ratio

Descriptive Analysis-graphic analysis

Example 1: Categorical variable

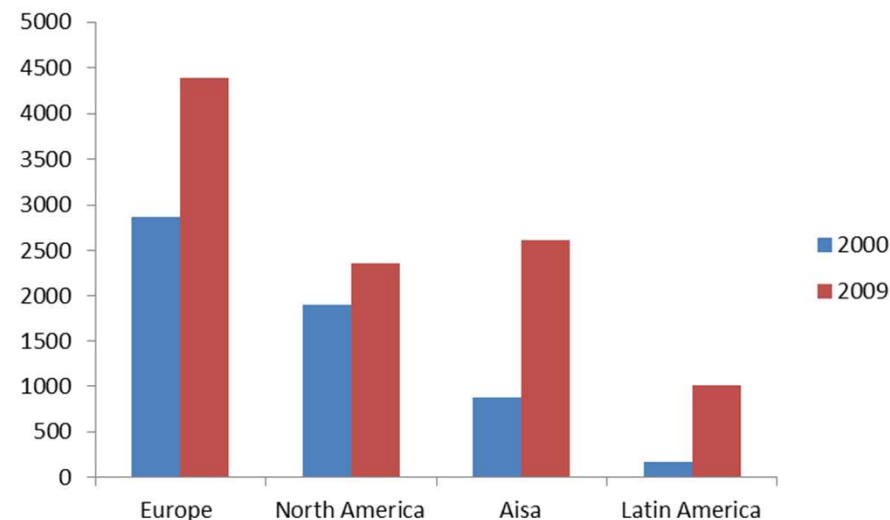| Weather	Days in Feb
Cloudy	1
Rainy	20
Sunny	7

Weather in Feb

- Cloudy 4%
- Sunny 25%
- Rainy 71%

Frequency

Example 2: Categorical variable

Adidas's net sales by region (millions of euros)

Region	2000	2009
Europe	2860	4384
North America	1906	2360
Aisa	875	2614
Latin America	171	1006
Total	5812	10364

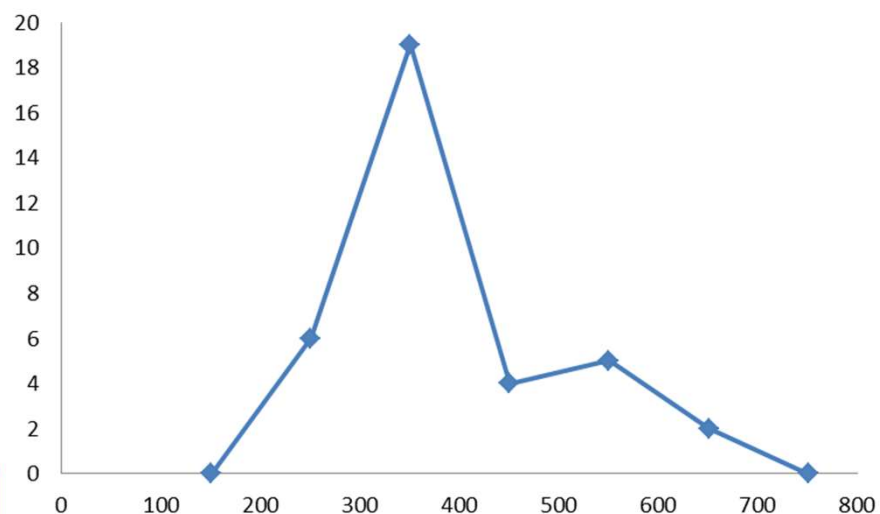**Adidas's net sale by region, 2000**
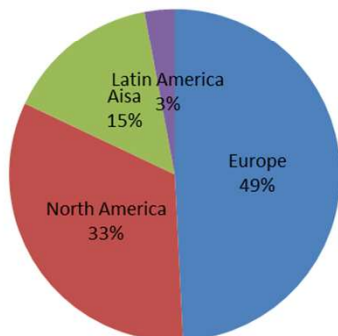
Adidas's net sale by region, 2009
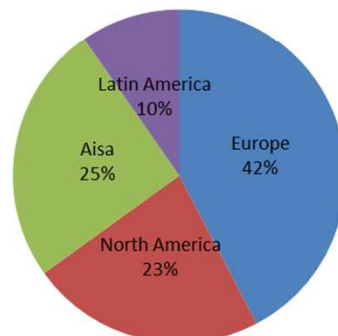
Example: Quantitate variable
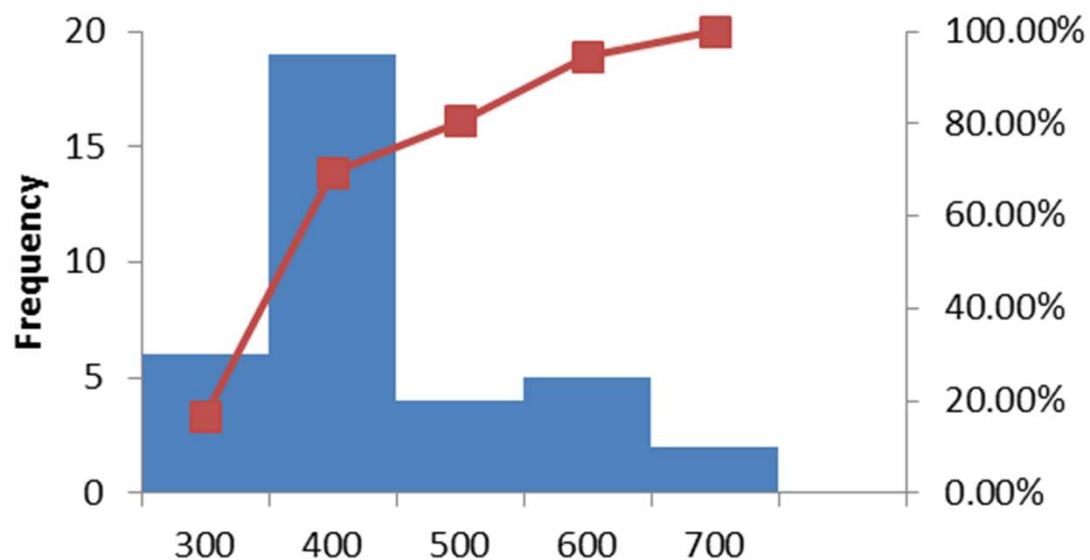
House sales prices in Warrnambool, 2014

349	435	525	315
349	399.9	229	335
299	331.1	331.1	480
330	299.9	339	239
289.9	375	399.9	330
629	315	695	209
595	355	355	519
339.9	385	329	520
595	449.9	499.9	399.9

Difference between Categorical and quantitative data
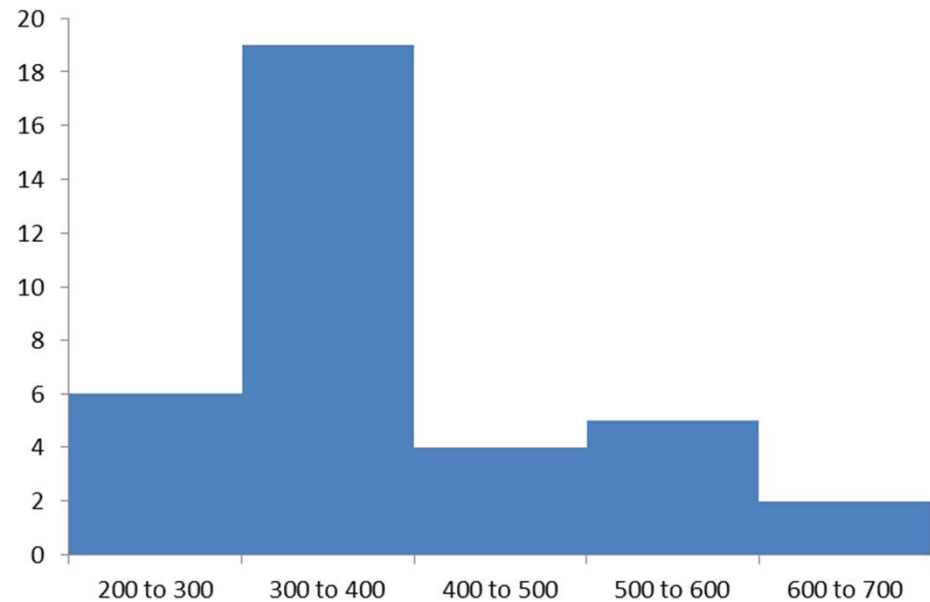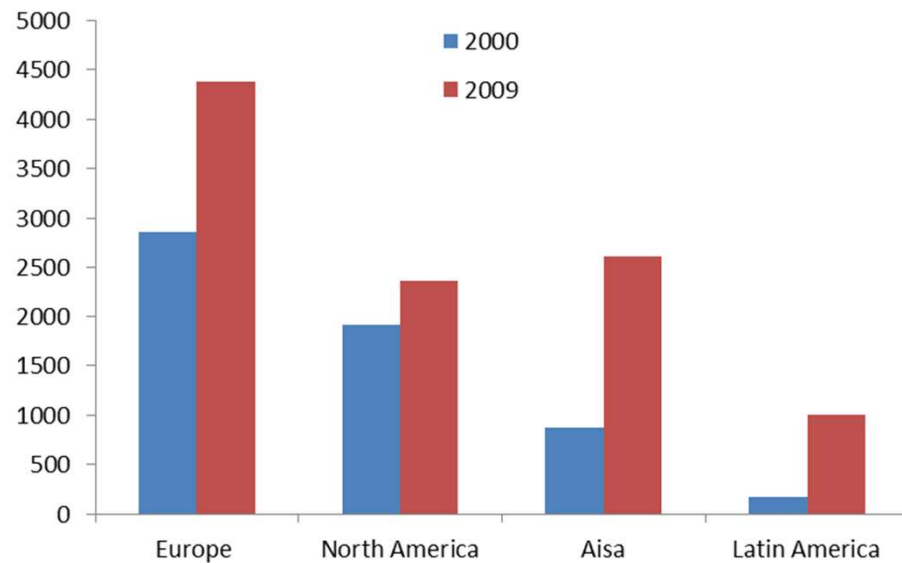
Stem and Leaf Diagram

↘ A simple way to see distribution details for quantitative data

↘ Separate the sorted data series into leading digits (stem) and trailing digits (leaves)

↘ List all stems from low to high

↘ For each stem, list all associated leaves

↘ Example: 12, 13, 17, 21, 24, 26, 27, 28, 30, 32, 35, 37, 38, 41, 43, 44, 46, 53, 58
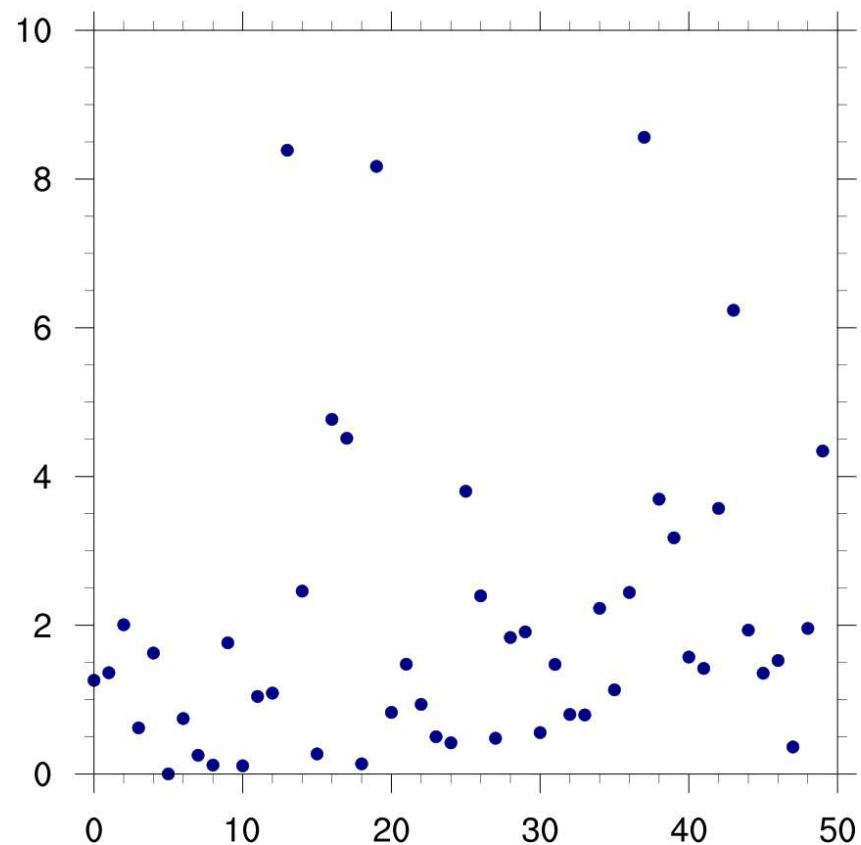
Stem	Leaves
1	2 3 7
2	1 4 4 6 7 8
3	0 2 5 7 8
4	1 3 4 6
5	3 8

Scatter Plot

SCATTER PLOT
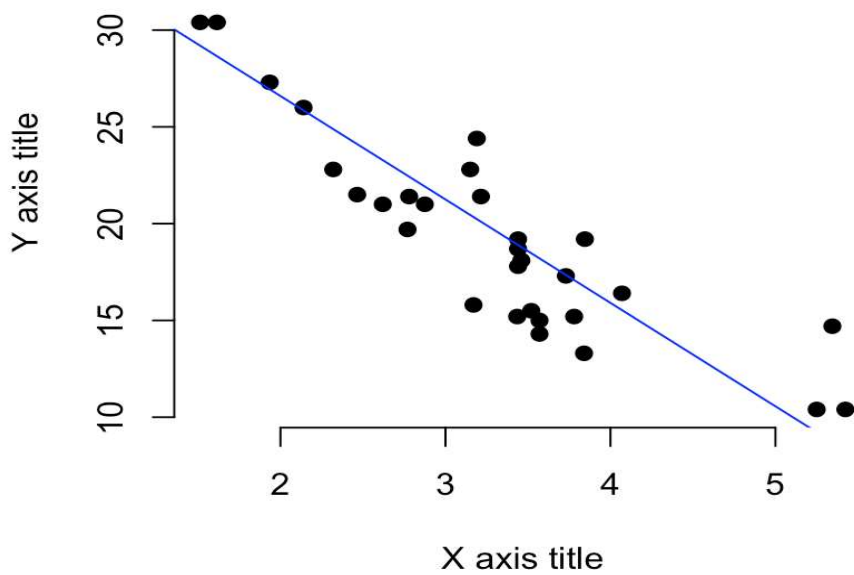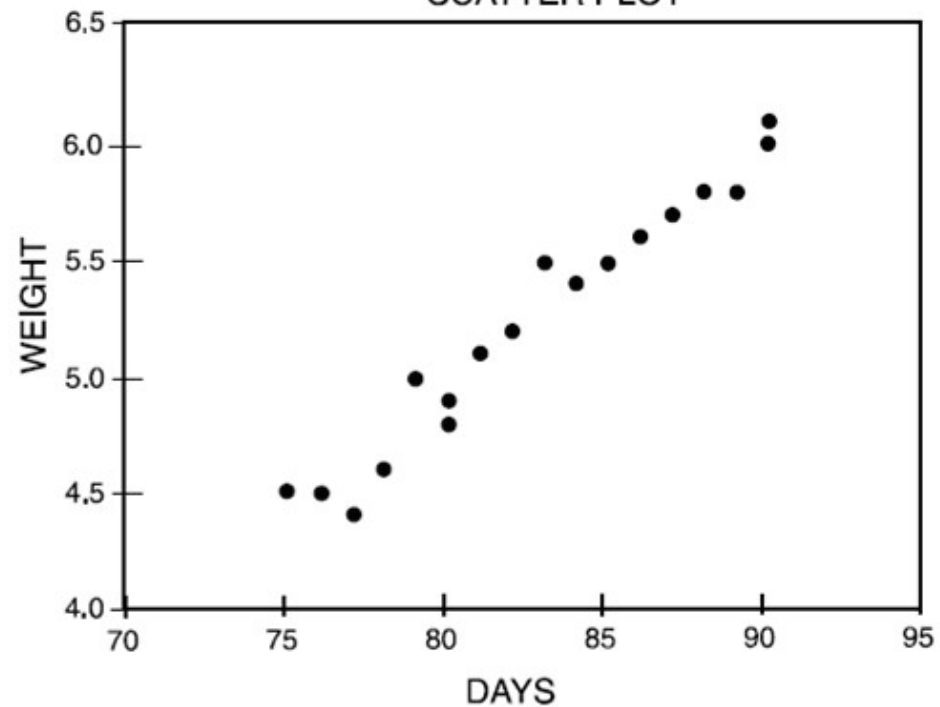
Numbers, numbers...

↘ Go to www.menti.com and enter the code 5122 0377

Descriptive Statistics

↘ Central tendency
- Mean
- Median
- Mode
- Percentile
- Quartile

↘ Measures of dispersion
- Range
- Variance and standard deviation
- Coefficient of variation

$$= \frac{standard\ deviation}{mean}$$

Calculation for central location

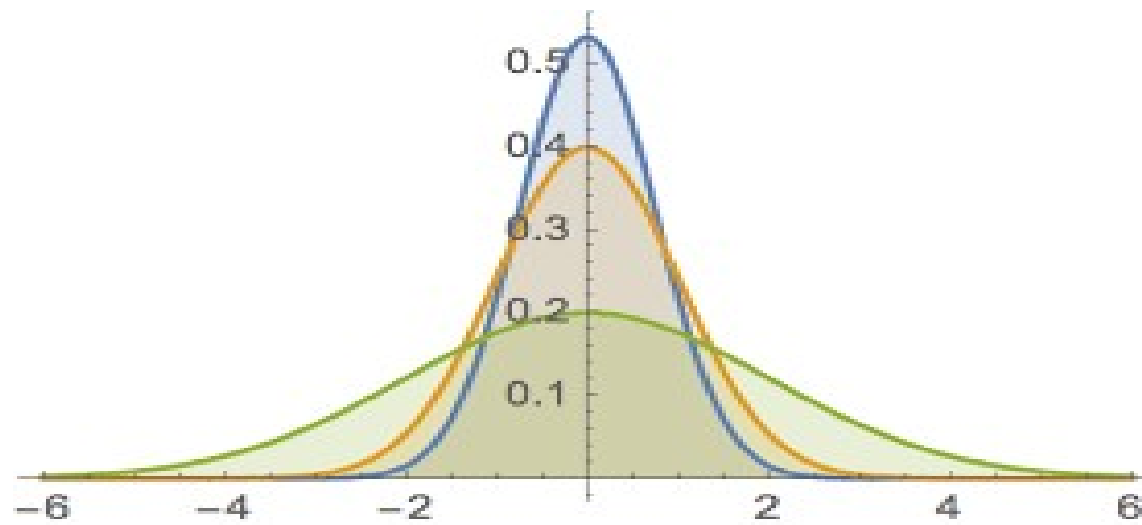| Title	Salary ($)
Administrative assistant	40,000
Research assistant	40,000
Computer programmer	65,000
Senior research associate	90,000
Senior sales associate	145,000
Chief financial officer	150,000
President (and owner)	550,000

↘ Mean=$\frac{\Sigma x_i}{N}$

$=\frac{40000+40000+65000+90000+145000+150000+550000}{7}$=$154,286

↘ Mode = 40,000

↘ Percentile (p^{th} percentile)

L_p=(n+1)$\frac{p}{100}$

↘ Median

L_{50}=(7+1)$\frac{50}{100}$=4 , thus Median=$90,000

↘ Quartile (25th percentile, 50th percentile, 75th percentile)

L_{25}=(7+1)$\frac{25}{100}$=2 , thus 1st Quartile=$40,000

L_{50}=(7+1)$\frac{50}{100}$=4 , thus 2nd Quartile=$90,000

L_{75}=(7+1)$\frac{75}{100}$=6 , thus 3rd Quartile=$150,000

Question: Calculate and interpret the 60th percentile salary of employees. $L_{60} = 4.8$, thus 60% of employees' salary is below $134,000

Box-Whisker plot

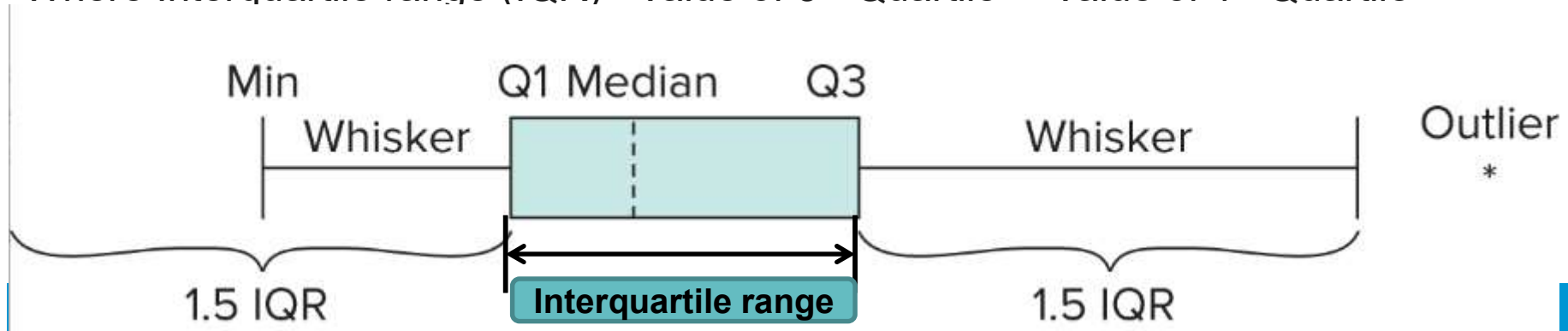↘ Minimum

↘ 1st Quartile

↘ Median (2nd Quartile)

↘ 3rd Quartile

↘ Maximum

To determine the outlier

Smaller than 1.5×Interquartile Range below the 1st Quartile, OR

Larger than 1.5×Interquartile Range above the 3rd Quartile

Where Interquartile range (IQR)= Value of 3rd Quartile − Value of 1st Quartile

Calculation for measures of dispersion

Title	Salary ($)
Administrative assistant	40,000
Research assistant	40,000
Computer programmer	65,000
Senior research associate	90,000
Senior sales associate	145,000
Chief financial officer	150,000
President (and owner)	550,000

➤ Range=max value-min value

$\quad$ =\$550,000-\$40,000=\$510,000

➤ Variance and standard deviation

Variance for sample: $s^2 = \dfrac{\sum(X_i-\bar{X})^2}{n-1}$

Variance for population: $\sigma^2 = \dfrac{\sum(X_i-\mu)^2}{N}$

$$\sigma^2 = \frac{(40000-154286)^2 + (40000-154286)^2 + (65000-154286)^2 + \cdots + (550000-154286)^2}{7}$$

$\quad$ =27845918367

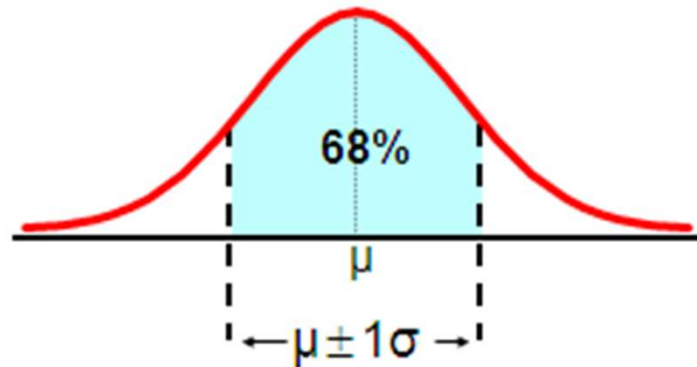Standard deviation: $\sigma = \sqrt{\sigma^2}$=166870

Empirical Rule

Symmetric and bell-shaped distribution:

68% of observations fall in the interval $\mu \pm \sigma$

95% of observations fall in the interval $\mu \pm 2\sigma$

99.7% of observations fall in the interval $\mu \pm 3\sigma$

Example: 280 students in one lecture class with a mean sore 74 and a standard deviation of 8 (assuming distribution is symmetric and bell-shaped).

1. Approximately how many students scored between 58 and 90?

2. Approximately how many students scored more than 90?

↘ To summarise qualitative data the suitable graphical tools to use are
_____ .

↘ For a relatively symmetric and bell-shaped distribution, approximately
_____ of all observations fall in the interval including 2 standard deviations
from the mean.

↘ Range = difference between _____ and
_____ .

↘ To display a relationship between two variables one can use
_____ .

↘ _____ is a relative measure of dispersion.

Recap

↘ You are given the following parameters for two populations:

↘ $\mu_1 = 700$ $\mu_2 = 29,000$

↘ $\sigma_1 = 150$ $\sigma_2 = 150$

Which population has data values more variable relative to the size of population mean?

Apply the empirical rule to find out:

68%   R: [700-150,700+150]

95%   R: [700-2*150, 700-2*150]

99.7%   R: [250,1150]

Break