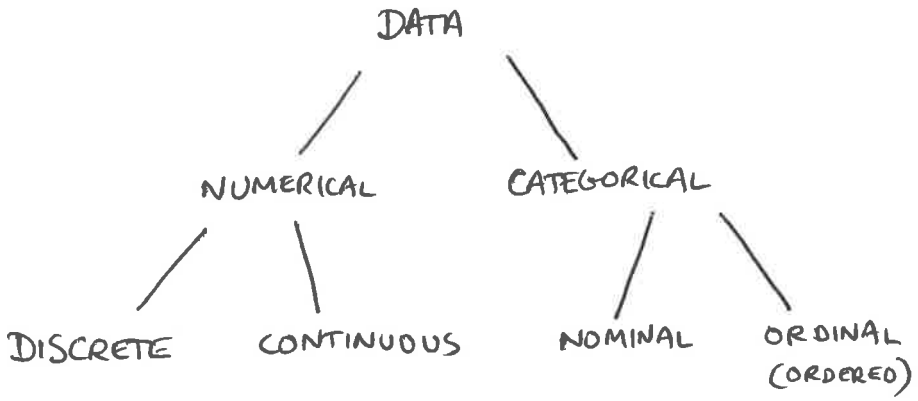
STATISTICS

3 PARTS
- COLLECTING DATA [GATHERING INFORMATION]
- ANALYSING DATA [AVERAGE, SPREAD]
- DISPLAYING / PRESENTING DATA [GRAPHS, CHARTS, ETC]

TYPES OF DATA : [LOTS OF WORDS TO LEARN, (SORRY)]



- PRIMARY DATA
- SECONDARY DATA
- UNI-VARIATE
- BI-VARIATE

SAMPLING

- RANDOM SAMPLE
- BIASED SAMPLE
- RELIABILITY OF SAMPLE
- POPULATION / SAMPLE

SEE DEFINITIONS SHEET FOR DEFINITION OF EACH TERM, AND EXAMPLES OF EACH

COLLECTING DATA /(SURVEYS)

- FACE-TO-FACE
- TELEPHONE
- POSTAL QUESTIONNAIRE
- ONLINE QUESTIONNAIRE
- OBSERVATION

KNOW PROS / CONS OF EACH :
- COST
- BIASED SAMPLE
- CLEAR / EASY TO UNDERSTAND

Statistics Definitions

Type of Data	Definition	Sample Question/Example of Data
Numerical Data	Data which is recorded as numbers	How many brothers/sisters do you have?
Discrete (Numerical) Data	Can only have a fixed number of values/answers	How many bedrooms are in your house? What is your shoe size? (note: can't be 11.345)
Continuous (Numerical) Data	Can have an infinite number of possible answers, is usually measured on a scale	What is your height?
Categorical Data	Data which is not recorded as numbers	How do you get to school?
Ordinal Data	Data which can be ordered in some way	Junior Cert Grades (A, B, C, D, etc) Month of Birth
Nominal Data	Categorical data which can't be ordered	What mobile phone network do you use? What is your favourite film?

Other definitions:

Data can be **Primary/Secondary**

- **Primary data** is collected by or for the person who is going to use it.
- **Secondary data** is data which is taken from another source

Data can be **Uni-variate** or **Bi-variate**

- **Uni-variate** means that you're just interested in one thing at a time, for example, the height of students in a school
- **Bi-variate** data is "linked"/ "paired" data – so you might be interested in the hours spent studying and the marks in an exam of students in the school, to see if there is a link between the two...

Samples

- The **population** is the entire group that is being studied
- A **sample** is a group that is taken/selected from the population

A **Simple Random Sample** is a sample in which each person in the population has an equal chance of being selected

A **Biased Sample** is a sample which does not fairly represent the population. For example, if I was trying to find out what the most popular sport in Dublin was, and I decide to ask 1,000 people coming out of the All-Ireland Hurling Final, this might be a biased sample.

Miscellaneous

A **Leading Question** is one which suggests a possible answer. For example: "Taxes are too high: Should they be reduced?"

DESIGNING QUESTIONNAIRES

NEEDS TO (BE):

- CLEAR / EASY TO UNDERSTAND
- USEFUL / RELEVANT.
- ALLOW ALL POSSIBLE ANSWERS
- HAVE NO LEADING QUESTIONS ← QUESTIONS WHICH SUGGEST A POSSIBLE ANSWER
- ASK ONLY ONE QUESTION AT A TIME.

PRESENTING DATA

WE SHOULD BE FAMILIAR WITH THESE.

- BAR CHARTS / LINE PLOTS
- PIE CHARTS
- STEM & LEAF DIAGRAMS
- HISTOGRAMS [LIKE A BAR CHART]
- SCATTERPLOTS.

STEM-AND-LEAF DIAGRAMS

- SEPARATE EACH VALUE INTO 2 PARTS ← STEM / LEAF

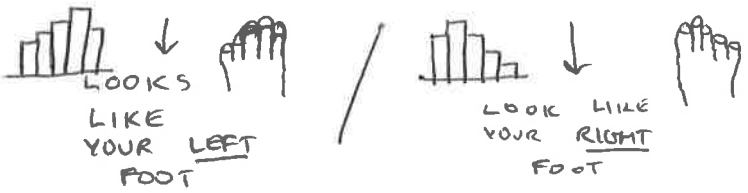
  eg 27 BECOMES 2|7

- ALL LEAVES WHICH HAVE THE SAME STEM ARE PLOTTED TOGETHER
  - LIKE A BAR CHART ON ITS SIDE
  - ARRANGE LEAVES IN ORDER OF SIZE.
- MUST HAVE A "KEY"
- EVERY PERSON / ITEM IS REPRESENTED BY A LEAF

BACK TO BACK STEM AND LEAF DIAGRAMS CAN BE USED TO COMPARE TWO DIFFERENT GROUPS.

HISTOGRAM [LIKE A BAR CHART - JOINED UP - NO GAPS]

- USED FOR "CONTINUOUS" DATA. eg HEIGHT
- ALLOWS YOU TO SEE THE SHAPE OF THE DATA / DISTRIBUTION
- EG SYMMETRIC / (LEFT) NEGATIVELY SKEWED / (RIGHT) POSITIVELY SKEWED



- SHOWS QUICKLY IF THERE ARE MORE BIG NUMBERS OR SMALL NUMBERS IN THE DISTRIBUTION

SCATTER GRAPHS / CORRELATION [DON'T GET ⊕/⊖ CORRELATION MIXED UP WITH ⊕/⊖ SKEWED]

- SCATTER GRAPHS SHOW LINKS BETWEEN 2 "VARIABLES"

  eg WE MIGHT TAKE PEOPLES HEIGHT + WEIGHT MEASUREMENTS AND SEE IF THERE'S A LINK.

  i.e. ARE TALLER PEOPLE USUALLY HEAVIER THAN SHORTER PEOPLE ?

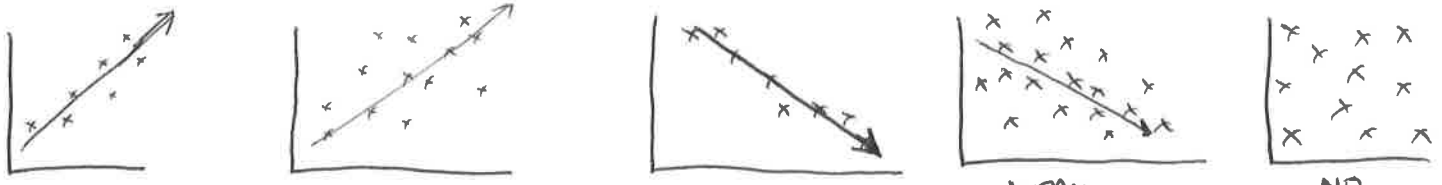
- 1st TRY TO DESCRIBE IN WORDS WHAT THE SCATTER GRAPH MEANS.
- USES "PAIRED" / BIVARIATE DATA
- CORRELATION COEFFICIENT r

> IF THE CORRELATION IS STRONG r WILL BE CLOSE TO 1 OR -1

- CORRELATION - ARE THEY LINKED ?

  MENTION
  - STRONG / WEAK
  - POSITIVE / NEGATIVE. (GOING UP / DOWN)

eg



STRONG POSITIVE ($r = 0.9$) | WEAK POSITIVE ($r = 0.6$) | STRONG NEGATIVE ($r = -0.95$) | WEAK NEGATIVE ($r = -0.5$) | NO CORRELATION ($r = 0$)

STEM AND LEAF CHART

Stem	Leaf
0	7, 9
1	3, 4
2	9
3	5
4	3, 4, 4, 7, (9)
5	2, 2, 3, 5, 7, 7, 8
6	1, 2, 5, 8, 9
7	3, 4, 5, 9
8	4, 7
9	1

Key: 3|5 = 35

- EACH ITEM OF DATA HAS ITS OWN LEAF.
- NUMBERS GO IN ORDER OF SIZE
- THIS 9 DOESN'T MEAN 9. IT MEANS 49 BECAUSE ITS STEM IS 4
- DON'T FORGET THE KEY.

BACK TO BACK STEM AND LEAF

Leaf (Girls)	Stem	Leaf (Boys)
6	0	9
9, 7, 2	1	3, 4, 5
	2	9
6, 6	3	5, 7
8, 6, 6, 6, 4, 2, 2	4	3, 4, 4, 9
9, 8, 6, 4	5	2, 2, 3, 5, 7, 7
8, 2	6	1, 2, 5, 8, 8, 9
9, 6, 5, 4	7	3, 4, 5, 9
5, 2, 0	8	4, 7
9, 8, 4, 3	9	1

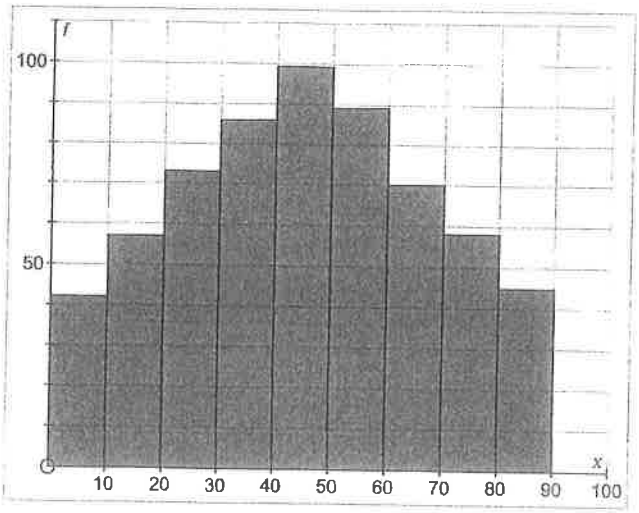
Key: 2|6| = 62

Key: |3|5 = 35

NOTE: KEY IS BACKWARDS

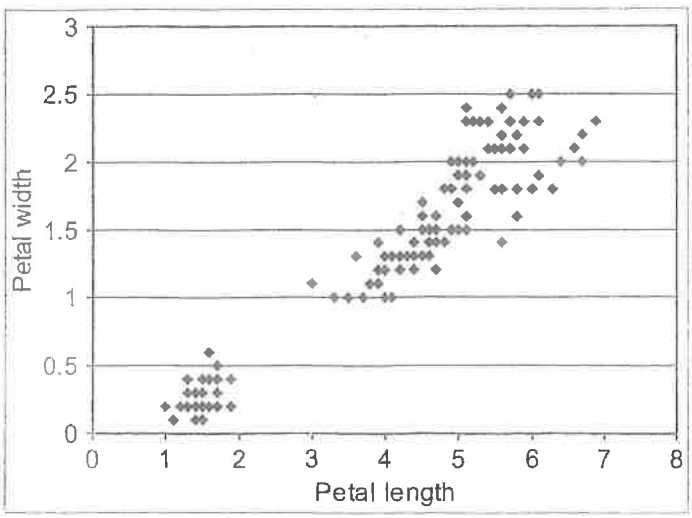
LEAF STEM

HISTOGRAM



- SYMMETRIC DISTRIBUTION
- NOTE: BARS JOINED TOGETHER - NO GAPS

SCATTER GRAPH



- EACH DOT REPRESENTS ONE "PETAL" - (HEIGHT + WIDTH)
- TRY TO DESCRIBE IN WORDS

eg "THERE IS A STRONG POSITIVE LINK / CORRELATION BETWEEN PETAL WIDTH/LENGTH. i.e. IF A PETAL IS LONG, IT IS USUALLY WIDE AS WELL."