

Definition: *Defense against the Dark Arts* is the science of Data.

How should it be collected?

How should it be summarized?

How should it be displayed?

How should it be interpreted?

Where are the pitfalls?

Jargon

Usual structure of data set.

Individuals, subjects, cases, experimental units are all words used for the people or animals or plants or things on which measurements are made.

Variables: the things measured.

Example: *case by variable* presentation.

Data on sea urchins:

Urchin ID	Age	Size
3997	6.91	57.5
991	0.91	9.5
2163	2.41	29.5
15	0.49	0.5
2202	2.41	30.5
2862	3.42	44.5
1575	1.41	24.5
293	0.49	2.5
358	0.49	3.5
:	:	:

Comment: 9 cases (of 250) shown, 3 variables.

Comment: Notice poor scientific form – no units listed for Age or Size in on-line source.

Example: Weather on Christmas Day, 2011,
YVR

Subset of variables available

Time	Temp	Hum	Weather
:	:	:	:
08:00	4.5	90	Cloudy
09:00	6.2	82	Cloudy
10:00	7.5	76	Rain
11:00	7.7	75	Rain
12:00	8.4	72	Rain
13:00	8.0	80	Mostly Cloudy
14:00	8.2	71	Mostly Cloudy
15:00	9.1	70	Mostly Cloudy
16:00	7.8	73	Mainly Clear
:	:	:	:

Jargon: variable types.

Nominal: categories with no particular order.

Examples: Variable Sex has 2 “levels”: Male and Female. Variable Eye colour has levels like “blue”, “hazel”, “brown”.

Ordinal: categories with an order.

Examples: 5 point scales: “Stephen Harper is doing a good job: Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree”

Weather: Rain, Cloudy, Mostly Cloudy, Mainly Clear are ordered.

(But other unordered categories possible: snow, fog, hail.)

Categorical: either **Nominal** or **Ordinal**.

Also called **Qualitative**.

Quantitative: numerical variable like value in \$, age, height, weight.

Interval: quantitative variable for which distance from 1 to 2 is same as from 3 to 4.

Ratio: Like **Interval** but with a natural value for 0.

Discrete: Used for both Categorical variables and for variable with only integer values.

Continuous: values between integers (in principle as finely measured as desired)

Examples: Mass is ratio, temperature in degrees Celsius is interval, number of murders in a week in Vancouver is quantitative but discrete, temperature is continuous.

Note: 5 point scales (“Likert” scales in Psychology) often assigned numbers say 1-5 or 0-4. But is difference between “Strongly agree” and “agree” same as between “agree” and “neutral”?

Why the jargon?

Sometimes helps identify suitable methods of data presentation, summarization and analysis.

WARNING: many different forms of statistical jargon in use in different disciplines.

Social Sciences: use **nominal**, **ordinal**, **interval** and **ratio**.

Math Stat: use **categorical**, **ordered categorical**, **quantitative**, **discrete**, **continuous**.

WARNING: all labels are sometimes open to debate. Is money “discrete”? (Integer number of pennies but huge number of possible values.)

Data Collection Exercise: VOLUNTARY

On blank sheet of paper please provide:

- 1) Height
- 2) Weight
- 3) Sex
- 4) Value of Coins in pocket / purse
- 5) SFU credits completed: this semester excluded.

PLEASE DO NOT PUT YOUR NAME ON THIS.

Give to me at end of class or put in box outside Stat Workshop.

PURPOSE: provide data set to display and summarize.