Q1: What type of data do we have?

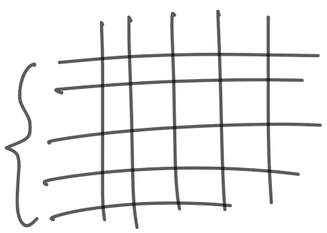
IDA = Initial Data Analysis
- main qualities of data
- suggests population
- background
- structure
- wrangling
- summaries

Research Questions

(i) Size (tidy data)

p variables = attributes of subjects

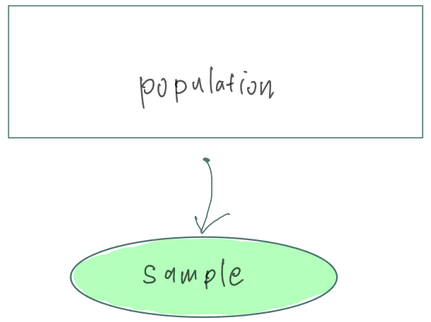
n subjects



p = dimension ☺

- univariate (1)
- bivariate (2)
- multivariate (2+)

Module 1: Exploring Data



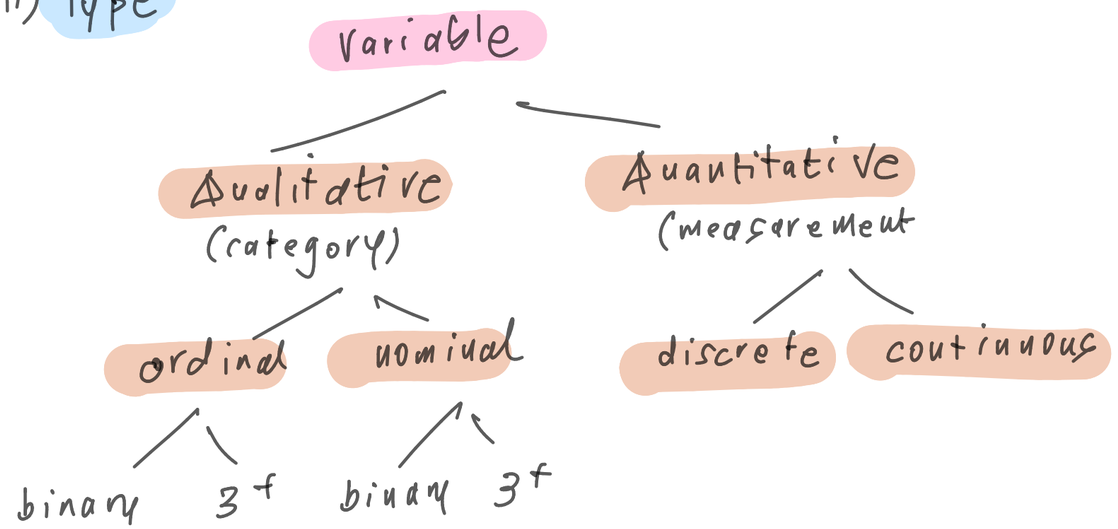
T1 : Design of Experiments

T2 : Data & Graphical Summaries

T3 : Numerical Summaries

LO3: Produce, interpret & compare graphical & numerical summaries using base R & ggplot.

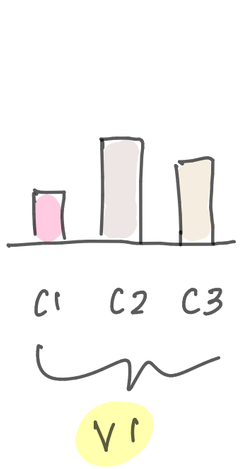
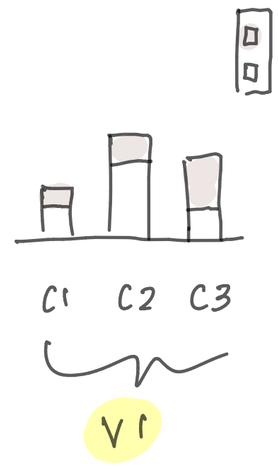
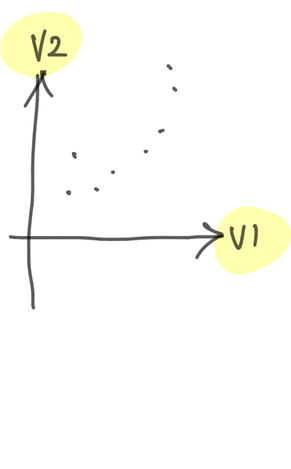
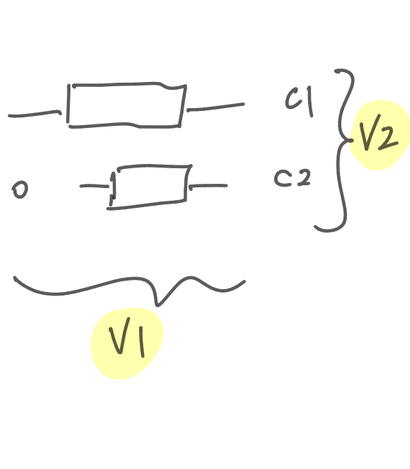
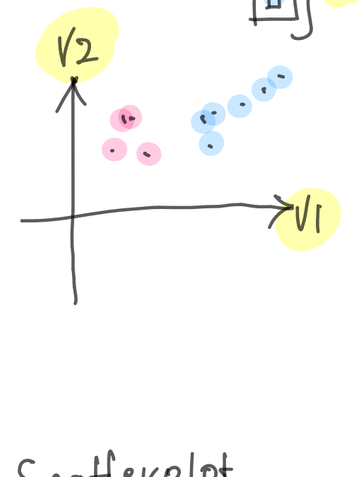
(ii) Type



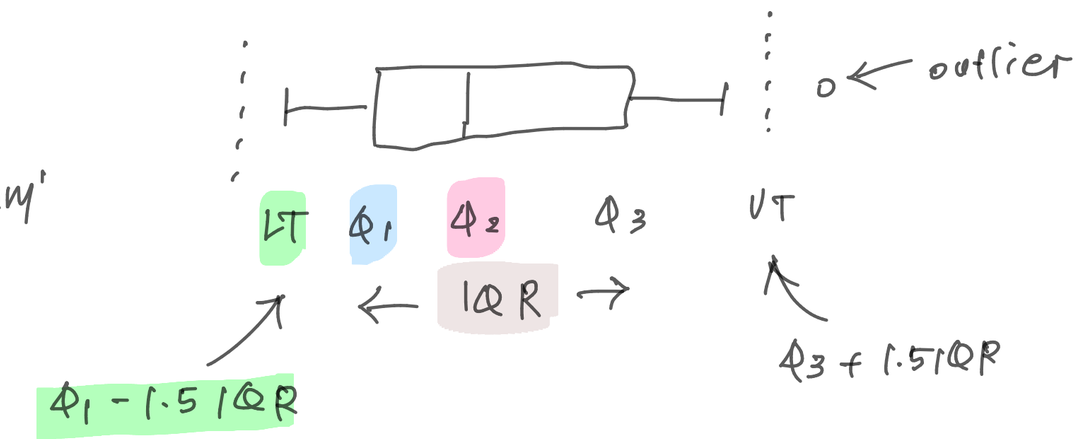
Note:
- ID ≠ variable DV4/21
- variables can be Qual or Quant DV4/23

Q2: How can we visualise data?

Data viz
- good. = tells an interesting story in a visually appealing way.
- poor = "chart junk" = visually boring, or distracting
- bad = misleading story

Variables	1 Qual	2 Qual	1 Quant	2 Quant	1 Quant 1 Qual	2 Quant 1 Qual
						
	Barplot	Double Barplot	Histogram Boxplot	Scatterplot	Side-by-side Boxplot	Scatterplot with colour ↖ or size/ shape ...

Note:
- Histograms of Barplots DV5/12-15
- Mistakes with histograms DV5/21-22
- Boxplot is a visual 'numerical summary' DV5/24, DV8



LT = lower threshold
Q1 = 1st Quartile (25% data below)
Q2 = 2nd Quartile (50% data below) = Median
IQR = Interquartile Range (50% data between)

How to produce data viz in RStudio?

The iris data is already in RStudio!

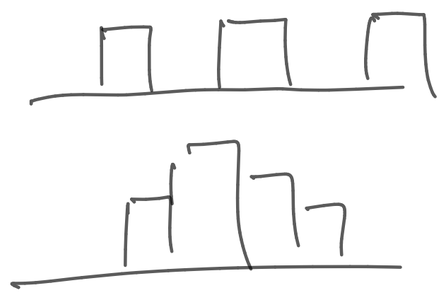
iris

Quant				Qual
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species

150X5

Base R

```
barplot ( table ( iris $ Species ))
hist ( iris $ Sepal.Length )
```



ggplot

```
library ( tidyverse)
ggplot ( iris, aes ( x = Species ))
    + geom_bar ( )
ggplot ( iris, aes ( x = Sepal.Length ))
    + geom_histogram ( )
```