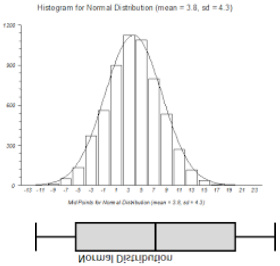
How to choose?

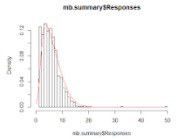
Descriptive statistics, graphs, tests

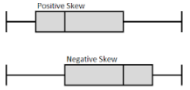
Descriptive statistics and Visuals

- Dependent on measurement type and shape of distribution
  - For continuous (Normally distributed) variables
    - Count, mean, standard deviation (minimum, maximum)
    - Histograms, dot plots, box plots, scatter plots

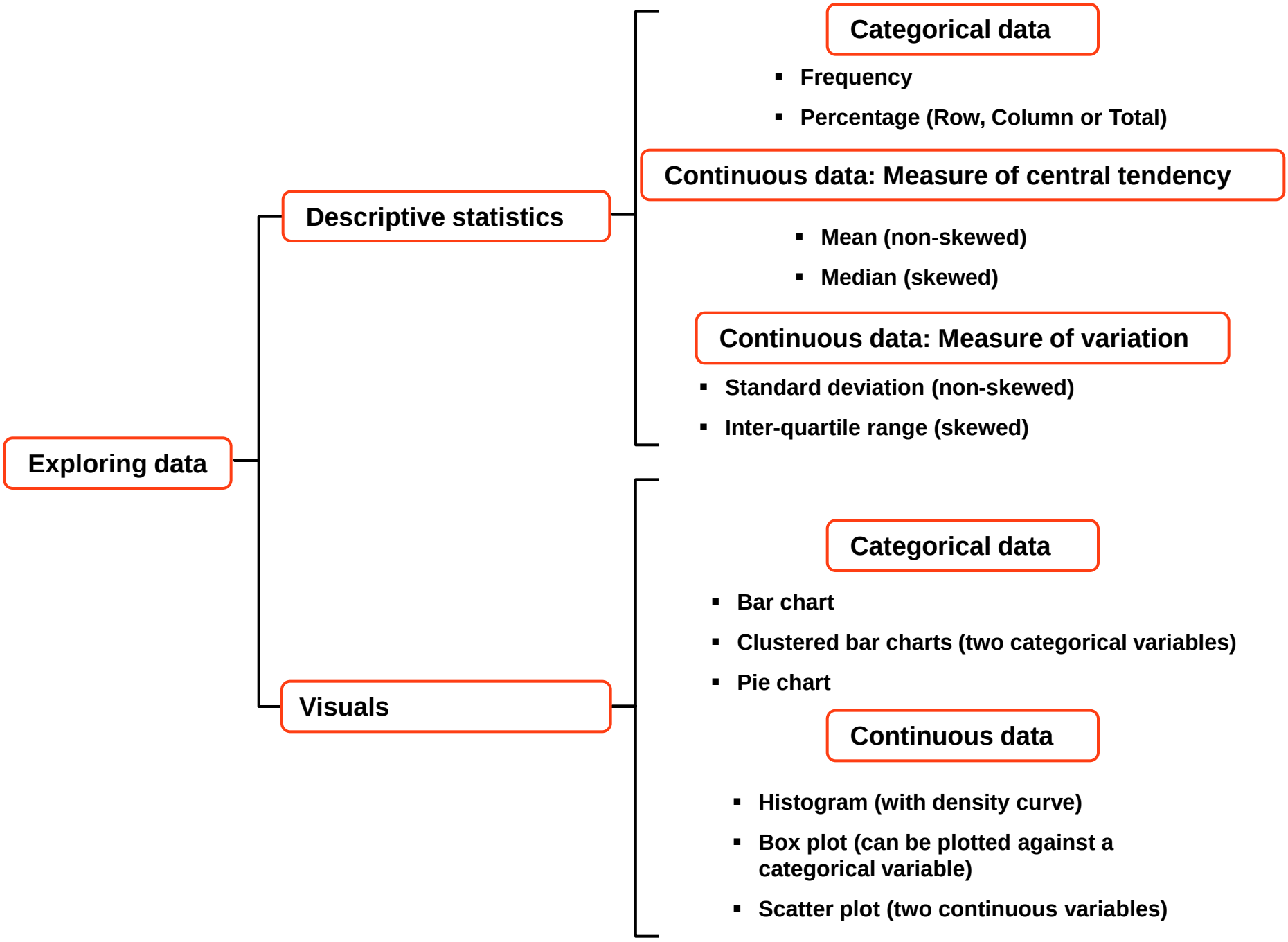


  - For continuous (skewed) variables
    - Count, median, inter-quartile range (minimum, maximum,)
    - Histograms, dot plots, box plots, scatter plots

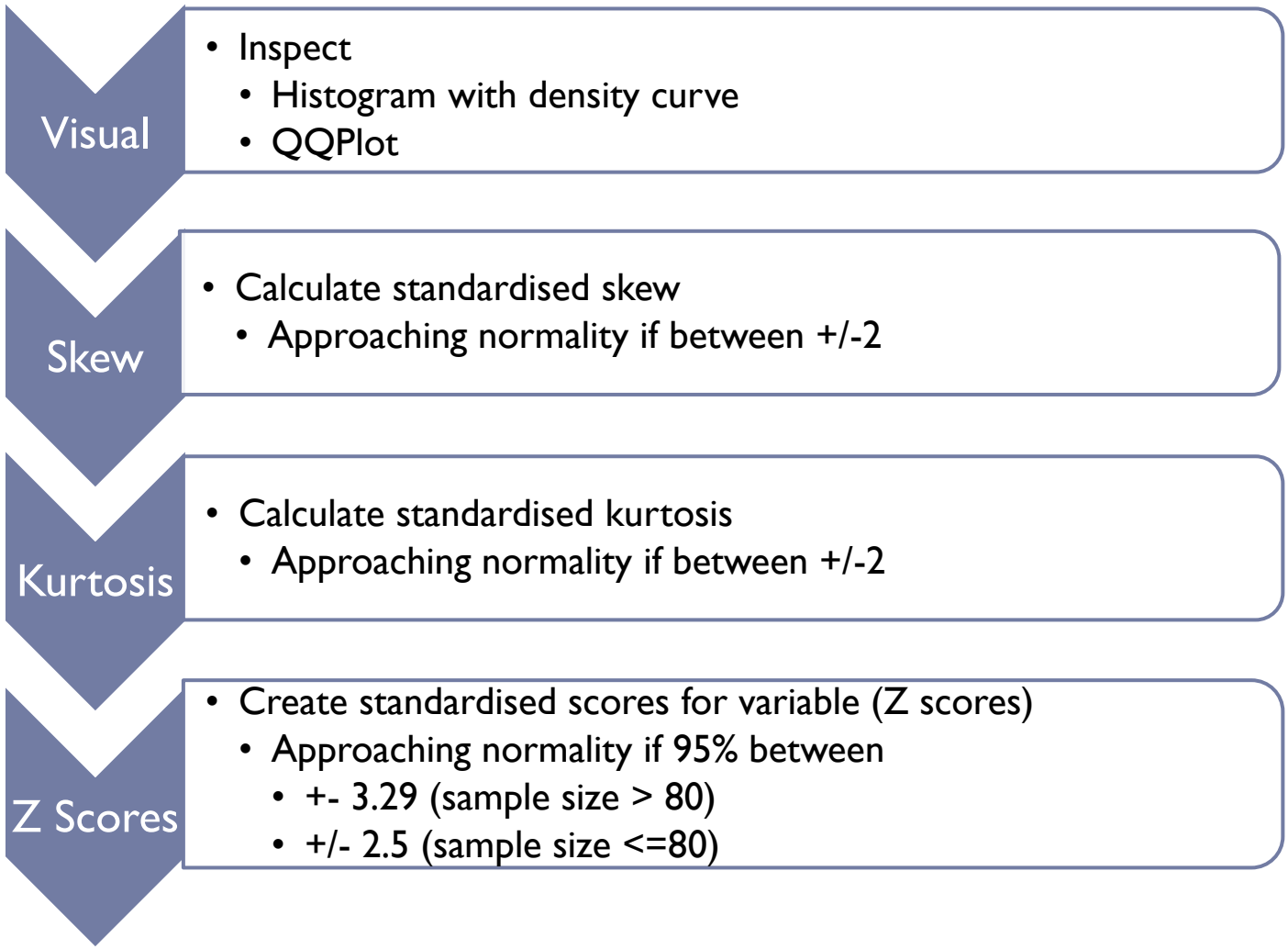




  - For categorical variables
    - Frequency counts, percentages
    - One-way tables, two-way tables
    - Bar charts, pie charts


Exploring data
Descriptive statistics
Categorical data
Frequency
Percentage (Row, Column or Total)
Continuous data: Measure of central tendency
Mean (non-skewed)
Median (skewed)
Continuous data: Measure of variation
Standard deviation (non-skewed)
Inter-quartile range (skewed)
Visuals
Categorical data
Bar chart
Clustered bar charts (two categorical variables)
Pie chart
Continuous data
Histogram (with density curve)
Box plot (can be plotted against a categorical variable)
Scatter plot (two continuous variables)

Deciding Normality



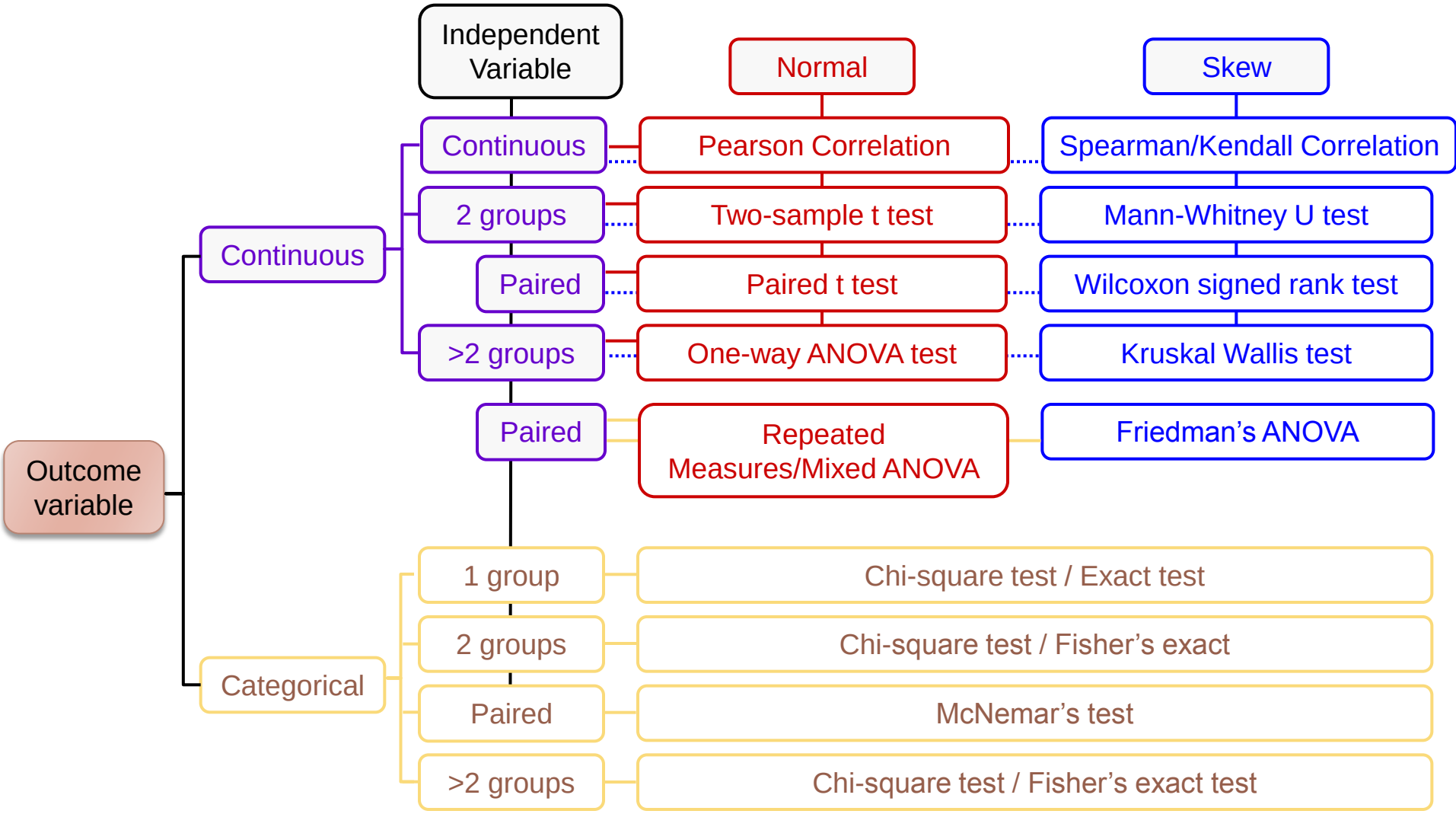
Deciding Normality

- There are statistical tests for normality
  - Shapiro-Wilks for small samples
  - Kolmogorov-Smirnov for large samples (>50)
- I have not covered these with you as they will usually return a result to indicate non-normality
  - THIS DOES NOT MEAN YOU CAN'T PROCEED WITH PARAMETRIC TESTS
- I expect you to complete a normality assessment in your CA as I have outlined during the module.
  - If you rely solely on a normality test then you will lose quite a few marks.

Which test?

- What is your measurement type?
  - For the outcome variable?
    - Continuous (Normal, Skew) or Categorical?
    - If more than one outcomes, are they paired or related?
  - For your independent variable?
    - Continuous or Categorical (1 group, 2 groups, more than 2 groups)
    - For 2 or more than 2 groups: Independent (Unrelated) / Paired (Related)

How to choose your test?



Parametric Difference Tests – Pre-Check

T-test

- Levene's test
  - Non-significant result variances homogenous in groups (var.equal=TRUE)
  - Significant result variances heterogeneous (var.equal=FALSE)

ANOVA

- Bartlett's test
  - Non-significant result variances homogeneous in groups (Tukey post-hoc)
  - Significant result variances heterogeneous (Games Howell post-hoc)

Independent variable/ Number of groups	Outcome(Dependent) variable		
	Continuous and Normally distributed (Parametric)	**Continuous and skewed / Ordinal (Non-parametric)**	**Binary (2 categories)**
Continuous	Pearson correlation `cor.test method='Pearson'`	Spearman/Kendall correlation `cor.test method='Spearman'/method='Kendall`	
2 independent groups	T test `t.test var.equal=TRUE/FALSE`	Mann-Whitney U test `coin::wilcox.test`	Chi-square test / Fisher's Exact `gmodels::Crosstable fisher=TRUE chisq=TRUE`
Paired (related) sample (2 time points)	Paired t test `t.test paired=TRUE/FALSE var.equal=TRUE/FALSE`	Wilcoxon signed rank test `coin::wilcox.test Paired=TRUE`	McNemar's test `mcnemar.test`
>2 independent groups	One-way ANOVA test `aov posthoc: Tukey/Games Howell posthoc Userfriendlyscience::onewaypost hoc=tukey/games-howell`	Kruskal-Wallis test `kruskal.test`	Chi-square test / Fisher's Exact Test `gmodels::Crosstable fisher=TRUE chisq=TRUE`
>2 related samples (>2 time points)	Repeated measures ANOVA `Aov`	Friedman's Test `Friedman.test`	