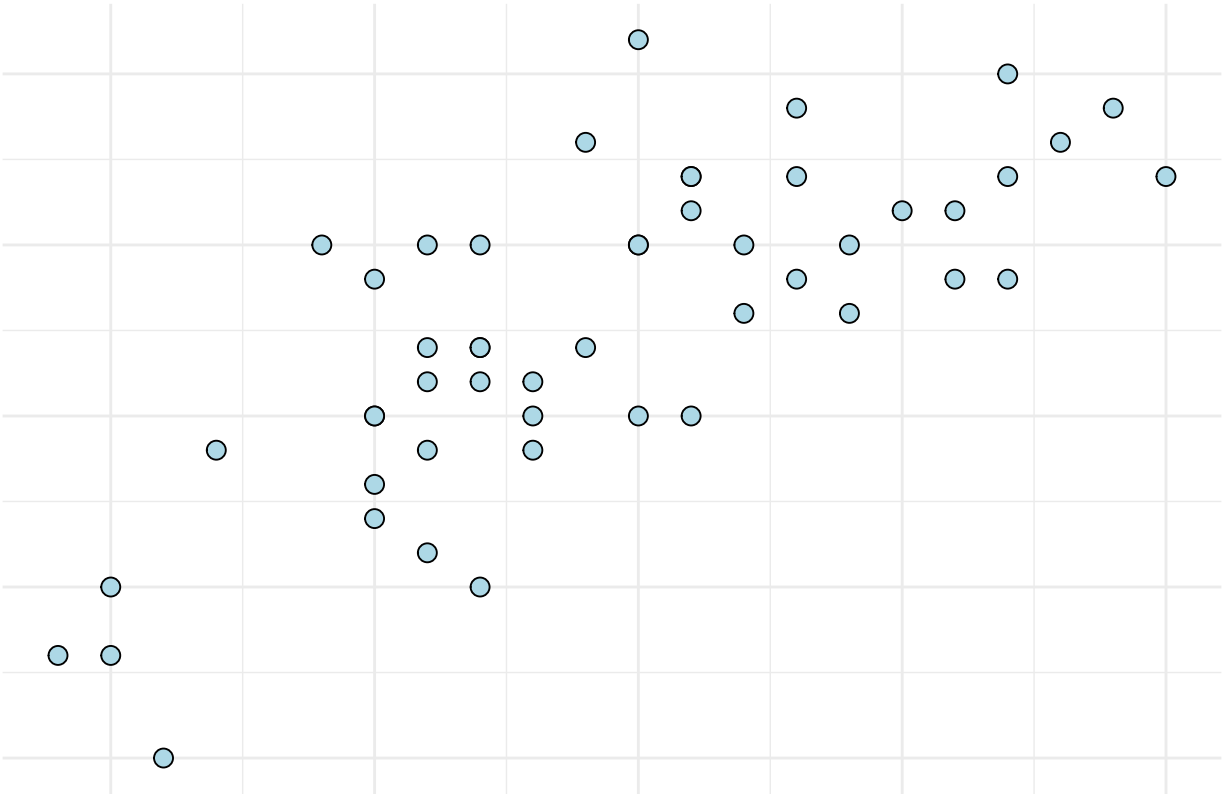
Correlation

What is a correlation?

Correlation analysis is a type of statistical analysis used to determine both the strength and direction of a relationship between two variables. The correlation coefficient ranges from -1 (negative correlation) to +1 (positive correlation). Correlation is very useful are identifying patterns in data but does not imply that one variable causes another.



Are these variables related?

When to use a correlation analysis?

When you are in a situation where you have two **continuous** variables.

Examples could include; comparing hours spent revising and exam performance, jaw size and bite force relationship and dietary sugar content and tooth demineralisation levels.

Assumptions of a T-test

The assumptions of a T-test that must be met are as follows:

1) Quantitative variables
2) Normally distributed data (Close to a bell shaped curve)
3) Linearity
4) Independence of observations (One value must not directly influence another)
5) Variation should be consistent as values increase (Homoscedastic)

Different types of correlation

Type of test: Pearson

Example: The relationship between daily flossing time (minutes) and interproximal probe depth (mm)

Pearson is used when data is normally distributed

Type of test: Spearman

Example: The correlation between a plaque index rank and a gum health rank.

Spearman used when data is not normally distributed

How to implement in RStudio

```
# running a correlation analysis
cor.test(x = variable1, y = variable2, method = "spearman") # or pearson depending on the test
```

Interpreting results

After you run your correlation analysis, you should be considering a few different things in the output: the p-value, the effect size (correlation coefficient), and the direction of the relationship.

The p-value

The p-value tells you how likely it is to observe your results (a correlation of the magnitude you found) if there is actually no real relationship between the variables in the wider population. So a low p-value means that it is very unlikely to see your data when there is no real difference. A threshold of 0.05 is commonly used in research and this means there is a 5% chance the results you observe are due to chance and not another reason.

Effect size (Correlation coefficient)

The effect size in a correlation analysis is directly represented by the correlation coefficient itself (Pearson's r, Spearman's rho). It tells you how strong or meaningful the relationship between your variables actually is. Generally speaking; 0.2 = small effect, 0.5 = medium effect, 0.8+ = large effect.