Scatterplots and Correlation

Summary

1. Scatterplots are data displays that is simply plotting points.
2. Correlation means that the data points tend upwards or downwards, or possibly neither.

Scatterplots

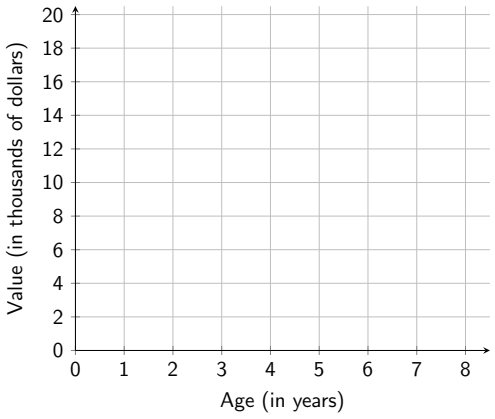
Scatterplot

A **scatterplot** is a visual display which can be used to examine an association between two variables, usually x and y.

- The independent variable, x, is called the **explanatory variable**.
- The dependent variable, y, is called the **response variable**.
- Scatterplots allow us to see if there is a relationship between the two variables.

Example 1. The table below shows the age of a certain model of car (in years) with the cars current value (in thousands of dollars). Create a scatterplot for the data.

Age	Value
2	15
3	12
3	13
2	14
4	13
5	10
6	10.5
1	16.5
0	18
4	14
7	11



Correlation

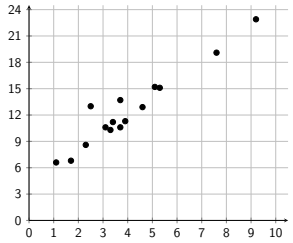
Often times, the data in a scatterplot has some pattern to it.

Correlation

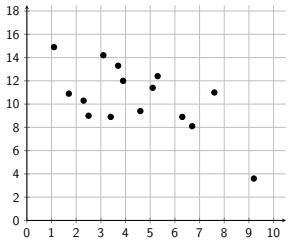
A **correlation** between two variables examines how the response variable's (y) values change as the explanatory variable's (x) values change.

There are 3 correlation types: positive, negative, and none (a.k.a. no correlation)

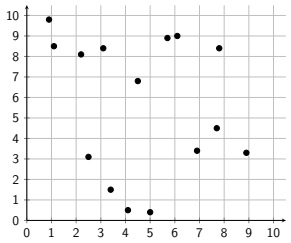
Positive Correlation
As x increases, so does y



Negative Correlation
As x increases, y decreases

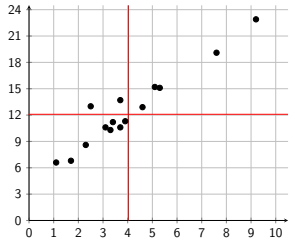


No Correlation
No visible patter between x and y

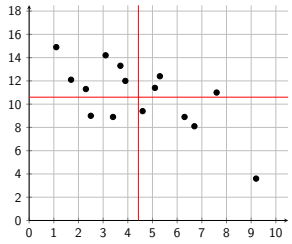


Means of x- and y-coordinates in red; along with count of points.

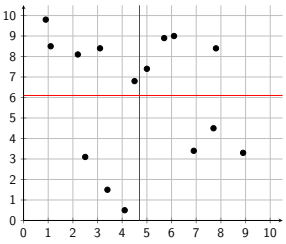
Positive Correlation
More points in Quads 1 and 3



Negative Correlation
More points in Quads 2 and 4



No Correlation
Just about same number of points in each



Correlation vs. Causation

***** VERY IMPORTANT *****

Just because there may be a strong correlation (an **association**) between two variables **DOES NOT MEAN THAT ONE CAUSES THE OTHER TO HAPPEN.**

For instance, dogs with larger paws tend to have larger weights, but we can not conclude that large paws cause a large weight.

If there is a strong correlation, there may be lurking variable(2) and/or confounding at play.

Lurking Variable

A **lurking variable** is an explanatory variable that has an influence in the outcome of a study or experiment but is not considered in the study or experiment.

Confounding

Confounding occurs when we can not distinguish the effect(s) one (or many) explanatory has (have) on a response variable.