THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Exploring Bivariate Data

Dr Michael Bulmer
School of Mathematics and Physics
The University of Queensland

Making Maths Matter 2026

Identifying and describing associations between two numerical variables

General Mathematics – Unit 3, Topic 1

Pearson Correlation

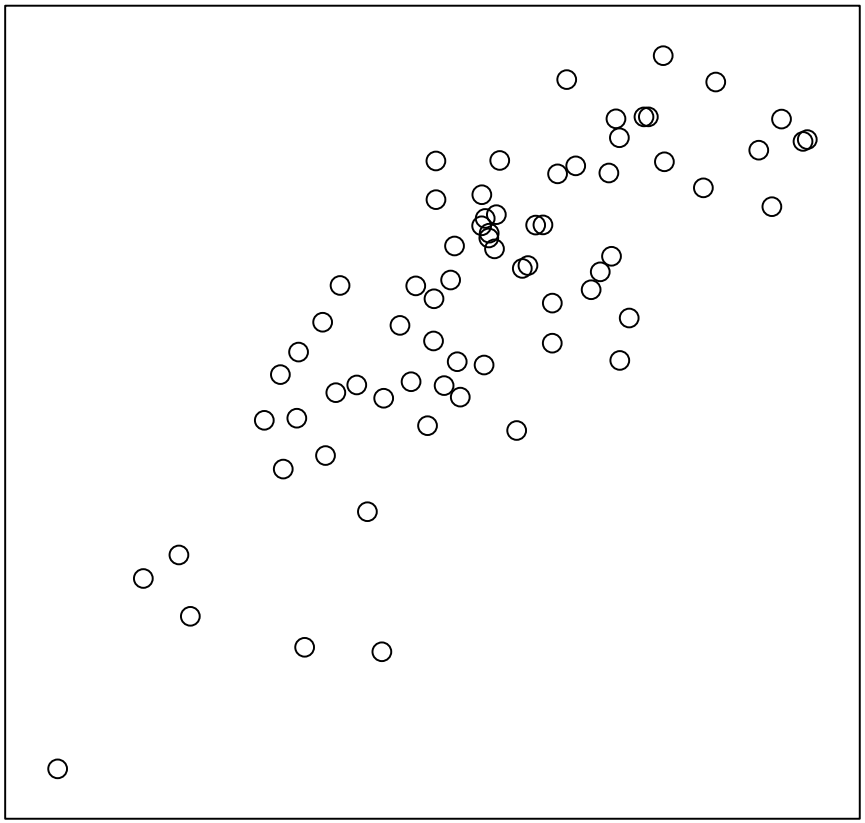
The *(Pearson) correlation coefficient*, r, measures the strength of a **linear** relationship.

If the points in our scatter plot are (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) then

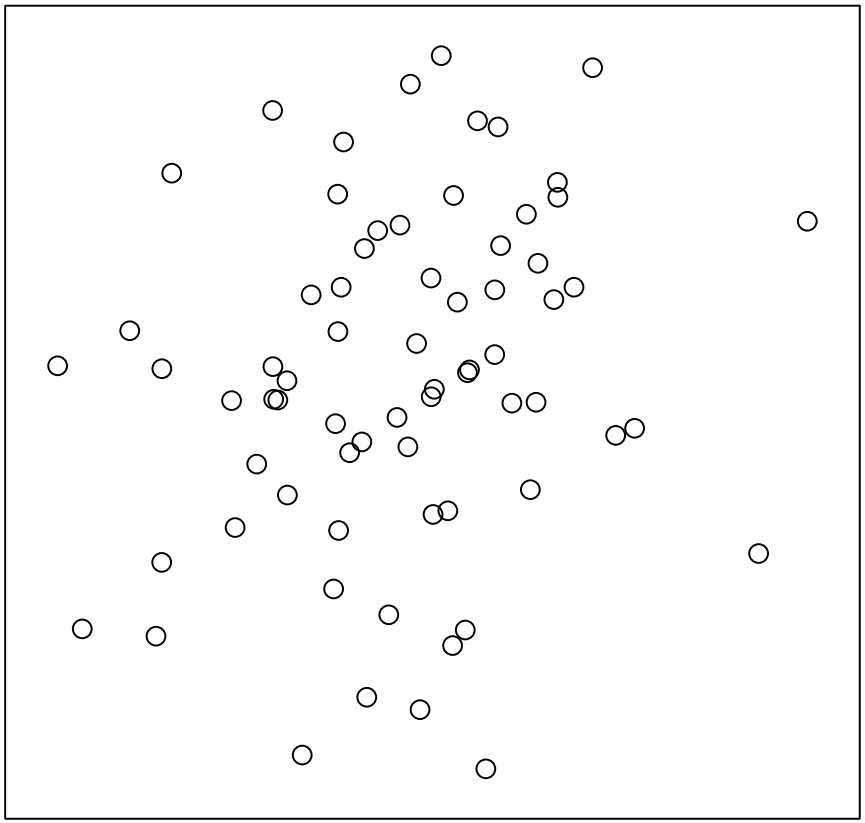
$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right),$$

where s_x and s_y are the sample standard deviations of the x and y values of the points.

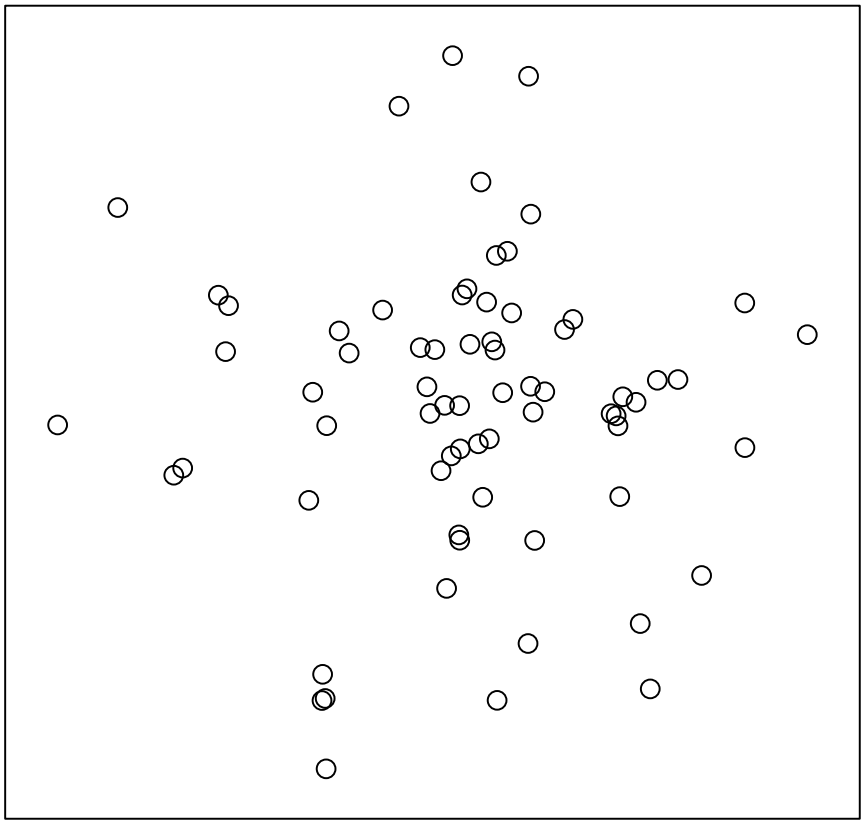
Pearson Correlation



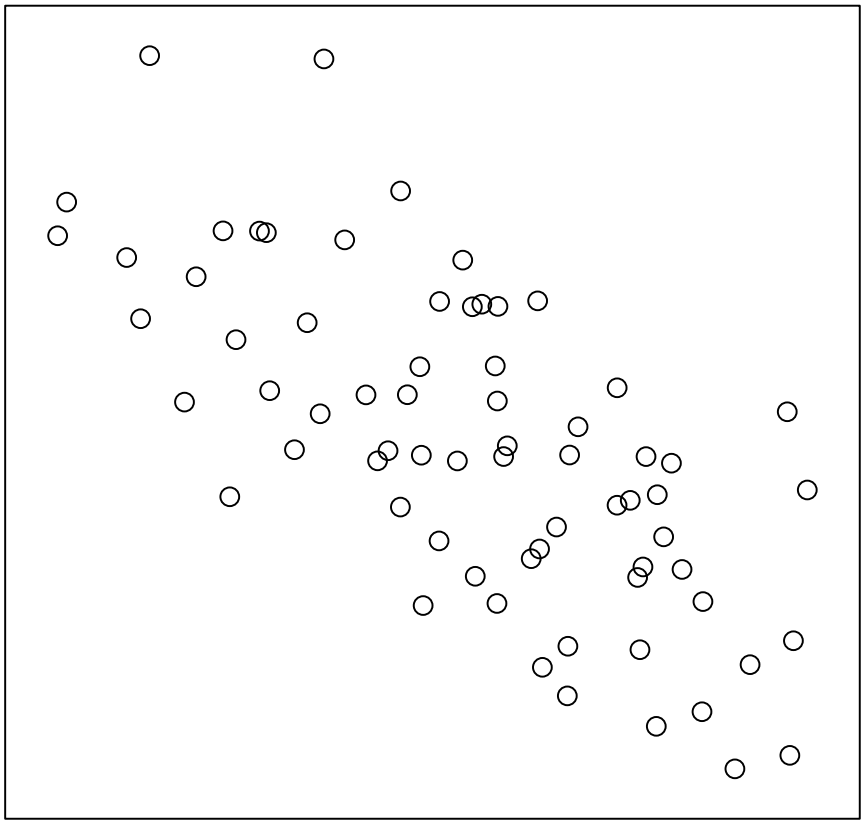
Pearson Correlation



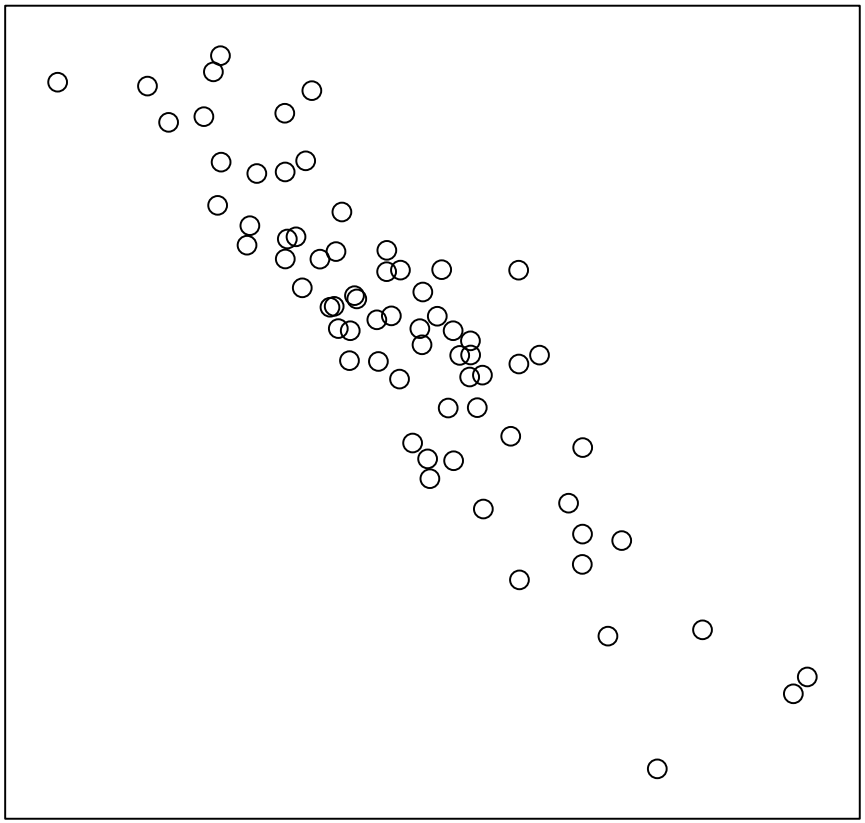
Pearson Correlation



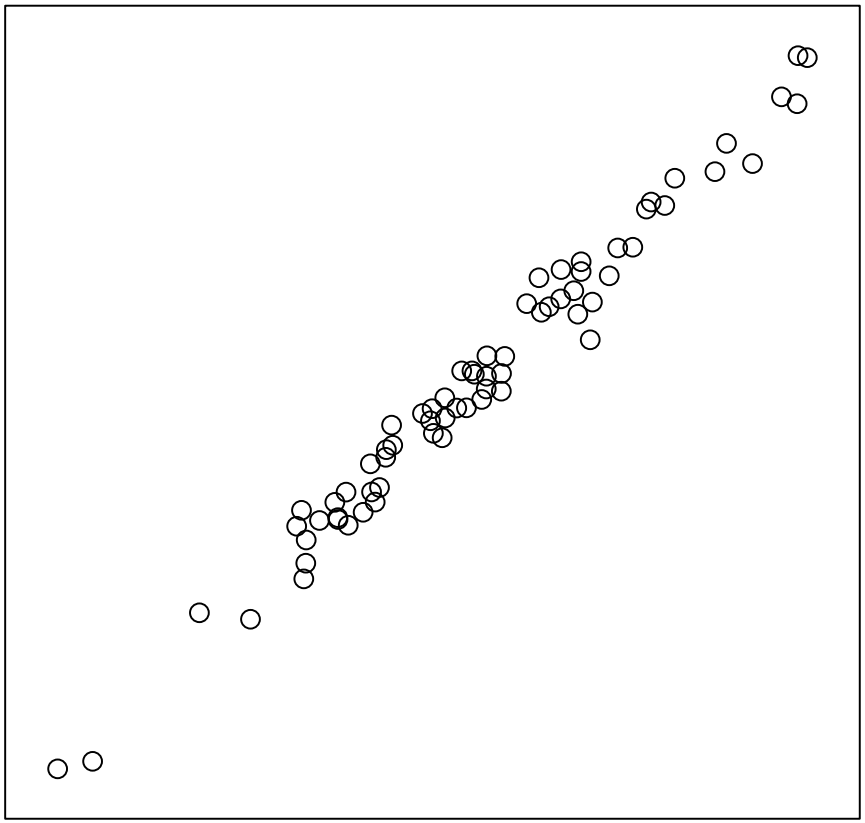
Pearson Correlation



Pearson Correlation



Pearson Correlation

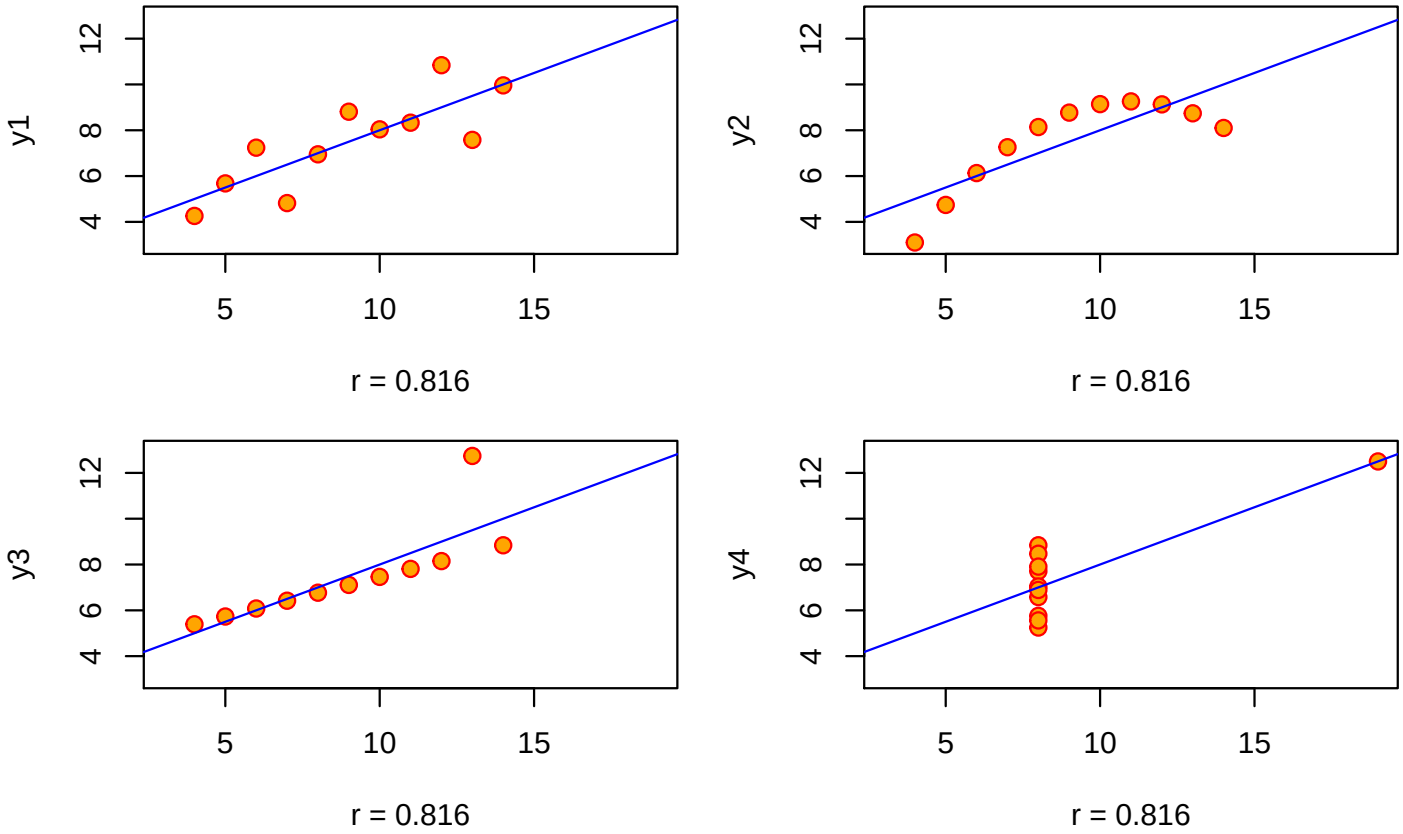


Anscombe Quartet

The correlation coefficient r is only appropriate for describing **linear relationships** between variables.

The famous 'Anscombe Quartet' gives a useful example of the importance of visualising your data.

Anscombe Quartet



R

- A “free software environment for statistical computing and graphics”
- Windows, Mac and Linux
- www.r-project.org

RStudio

- A development environment for working with R
- Projects, script editor, history and better tools for plots
- www.rstudio.com

The Islands

The Islands provides an online population of virtual human subjects for use in statistical investigations.

Schools are welcome to use this to support student projects.

islands.smp.uq.edu.au/register

Balance

A student project investigated a possible association between coordination and age on the Islands.

How do we measure the *total* variability in balance time?

```
balance = read.csv("balance.csv")
mean(balance$Time)
sd(balance$Time)
```

Balance and Age

Does knowing the age of a person help explain this variability?

```
plot(Time ~ Age, balance)
lm(Time ~ Age, balance)
abline(48.190,-0.308)
```

What does R^2 represent?

General Mathematics – Unit 3, Topic 1

$1.5 \times$ **IQR**

Consider a box plot of balance time.

Why is it '1.5' times the interquartile range?

General Mathematics – Unit 2, Topic 5

Australian Vehicle Prices

Let's explore an existing data set (from kaggle.com) which gives the prices of vehicles in Australia for 2023.

Designing Studies – Sample Size

Suppose you want to know if the correlation between two variables is at least ρ.

What sample size would you need to be reasonably confident of being able to detect the association?

Pearson Correlation