Scatter Plots in R with `ggplot2`

2024-07-19

About

A scatter plot is a data visualization that shows the relationship between **two continuous variables**. It is useful for identifying patterns and outliers. We can also use an additional categorical or continuous variable to color the points and group them or show a third pattern in the data.

> **i** Assumed knowledge
>
> - You know how to install and load packages in R.
> - You know how to import data into R.
> - You recognise data frames and vectors.

> 💡 **Data structure**
>
> Your data should be **structured** in a way that makes it *easy* to plot. The ideal structure is **long**, i.e. one where each column represents a variable and each row an observation (Figure 1). You can either reshape your data in R or **move cells manually** in a spreadsheet program to achieve the desired structure. For boxplots comparing more than one group of data, a **categorical variable** representing the group should be present in the data.
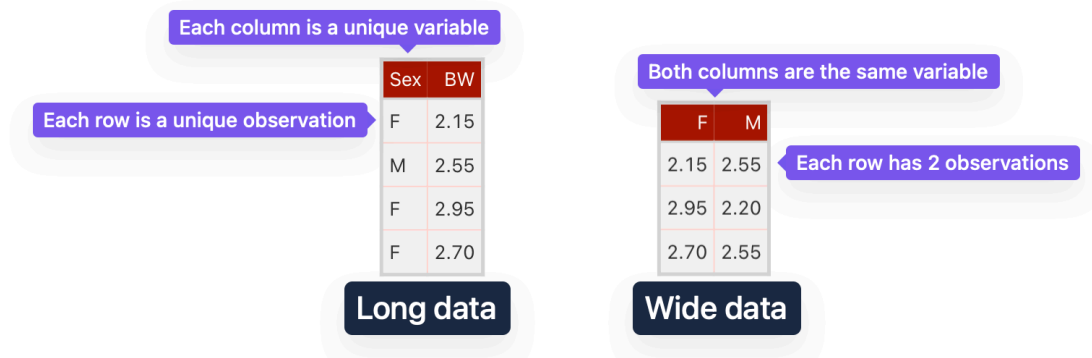>
>
>
> Figure 1: Long data (left) where each column is a different variable – e.g. Sex is categorical and BW is the measured, continuous response – is preferred over wide data (right), as it makes it easier to manipulate data when plotting.

Data

For this cheatsheet we will use part of the possums dataset used in BIOL2022 labs.

Import data

```
library(readxl) # load the readxl package
possums <- read_excel("possum_bw.xlsx") # read file, store as "possums" object
```