Week 5: Functions

Data Analysis for Psychology in R 1

Patrick Sturt

Department of Psychology
The University of Edinburgh

Weeks Learning Objectives

1. Understand the basic principles of functions.
2. Understand concept of data transformations.
3. Understand the calculation of z-scores.

Topics for today

- What is a function?
- Linear and non-linear functions
- How do we use functions in statistics?
- An example of z-scores

What is a function?

- A function takes an **input**, **does something**, and provides an **output.**
- **Input**

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- **Doing something**

$$f(x) = x - 2$$

- **An output**

$$y = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Functions and relations

- It is important to think of the function as showing the *relationship* between input and output.
- We can link this to the idea of relationships from week 4.
- The function links an input (predictors, x), to an output (outcome, y)
- So we can write

$$y = f(x) = x - 2$$

Visualising Functions

- An important tool in understanding functions is to plot them.
- So let's look at the following:

$$y = f(x) = 10 + 2x$$

Visualising Functions

- Our input x is a vector of numbers:

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{bmatrix}$$

Visualising Simple Functions

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

```
func_x <- tibble(
  x = c(1,2,3,4,5,6,7,8),
  y = 10 + (2*x)
)
```

- `tibble` is used to create a data set
- `x` is our original data entered as a vector of numbers using `c()`
- `y` is the output of the function `f(x) = 10+(2*x)`

Visualising Simple Functions

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

$$y = f(x) = 10 + 2x$$

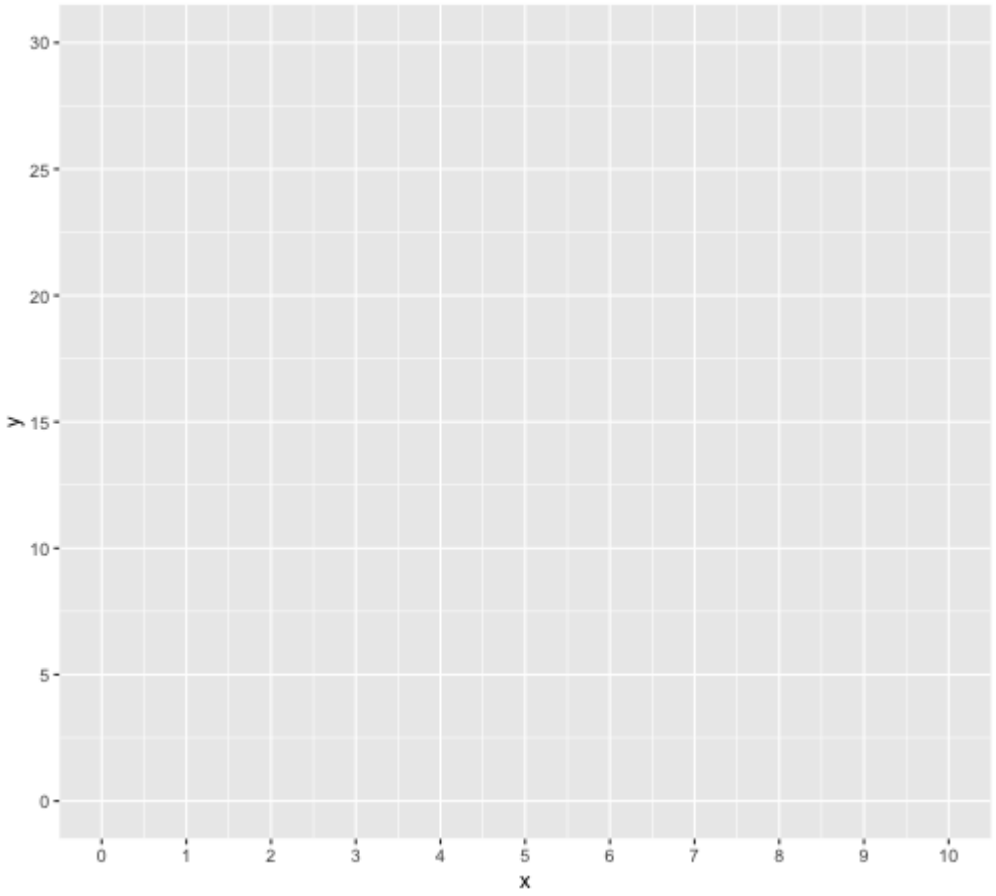
- Example row 1:

$$10 + (2 * 1) = 12$$

- Example row 5:

$$10 + (2 * 5) = 20$$

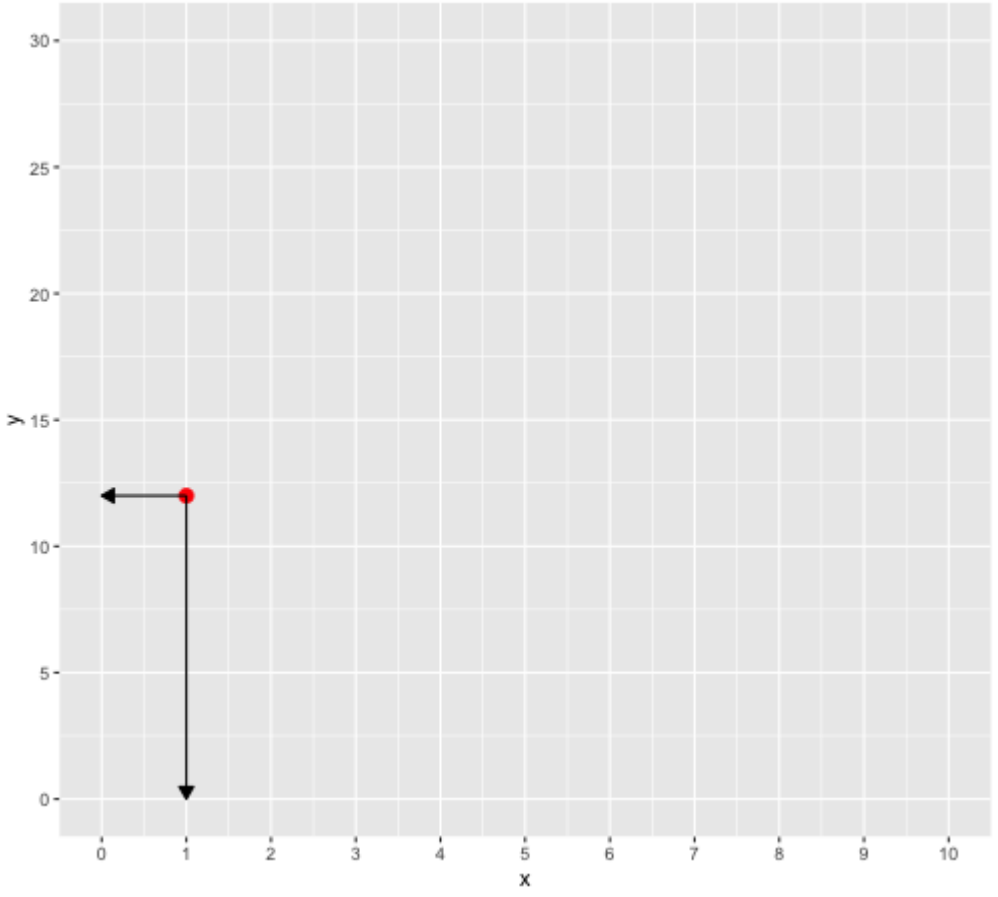
Visualising Functions



Our Data

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

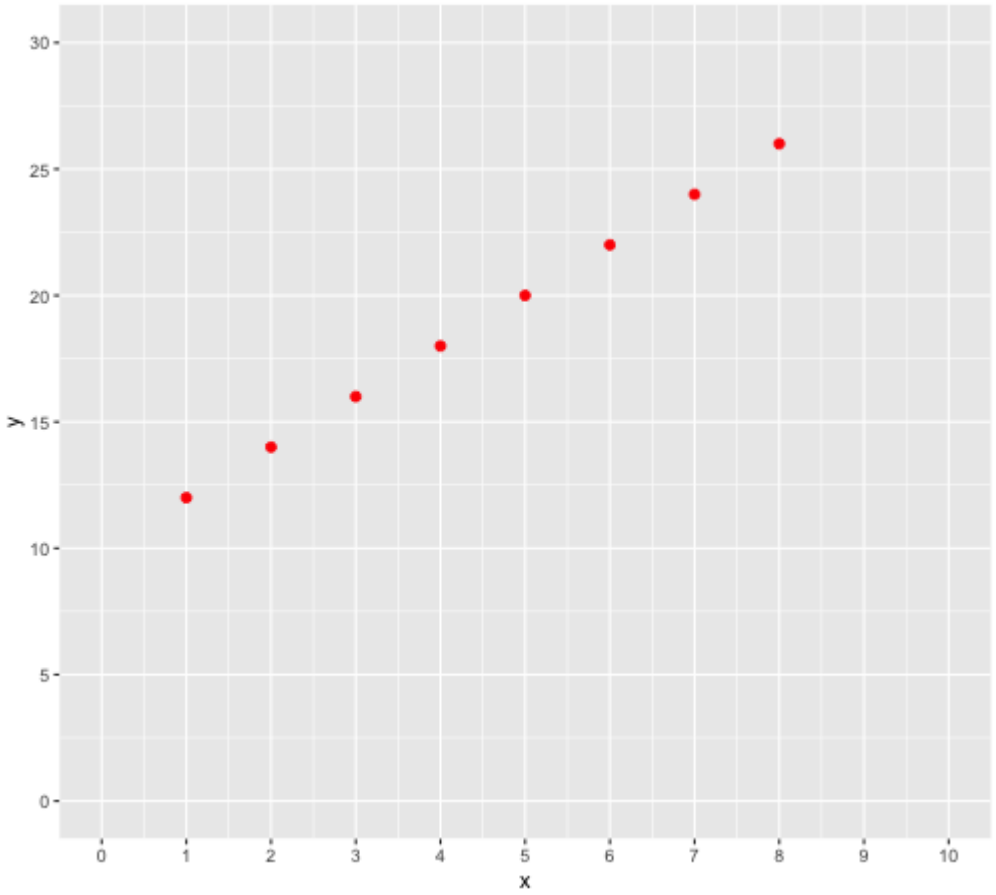
Visualising Functions



Our Data

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

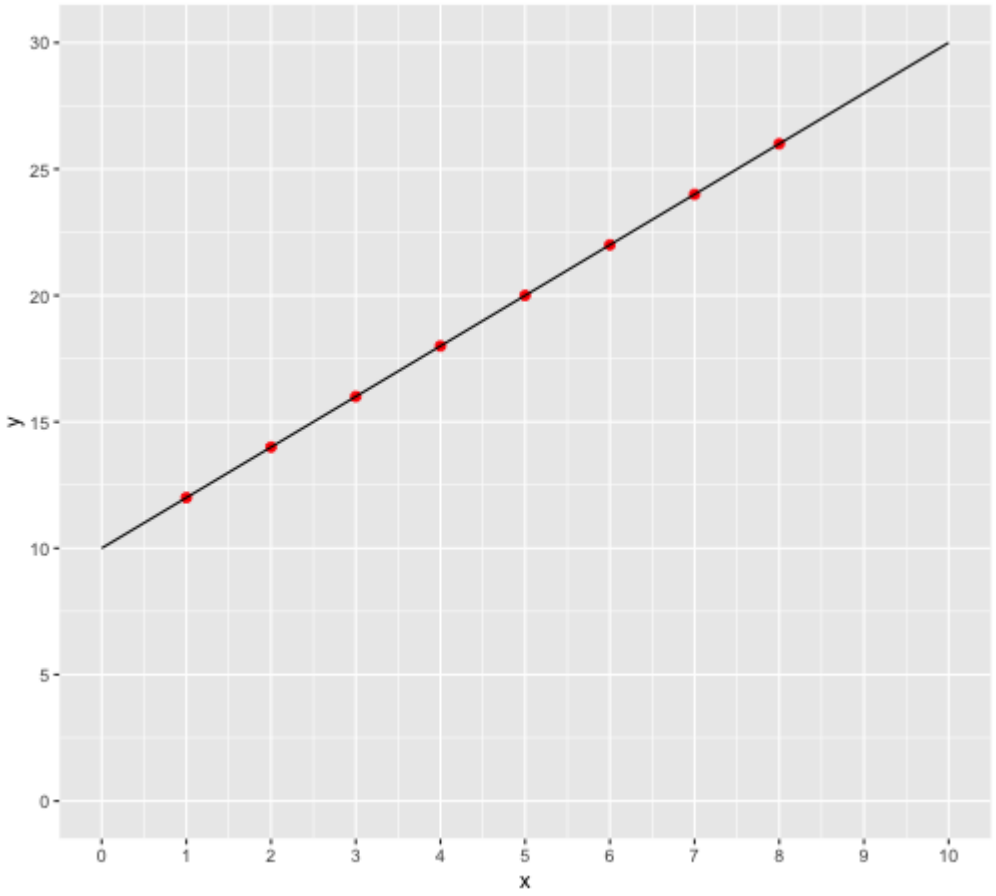
Visualising Functions



Our Data

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

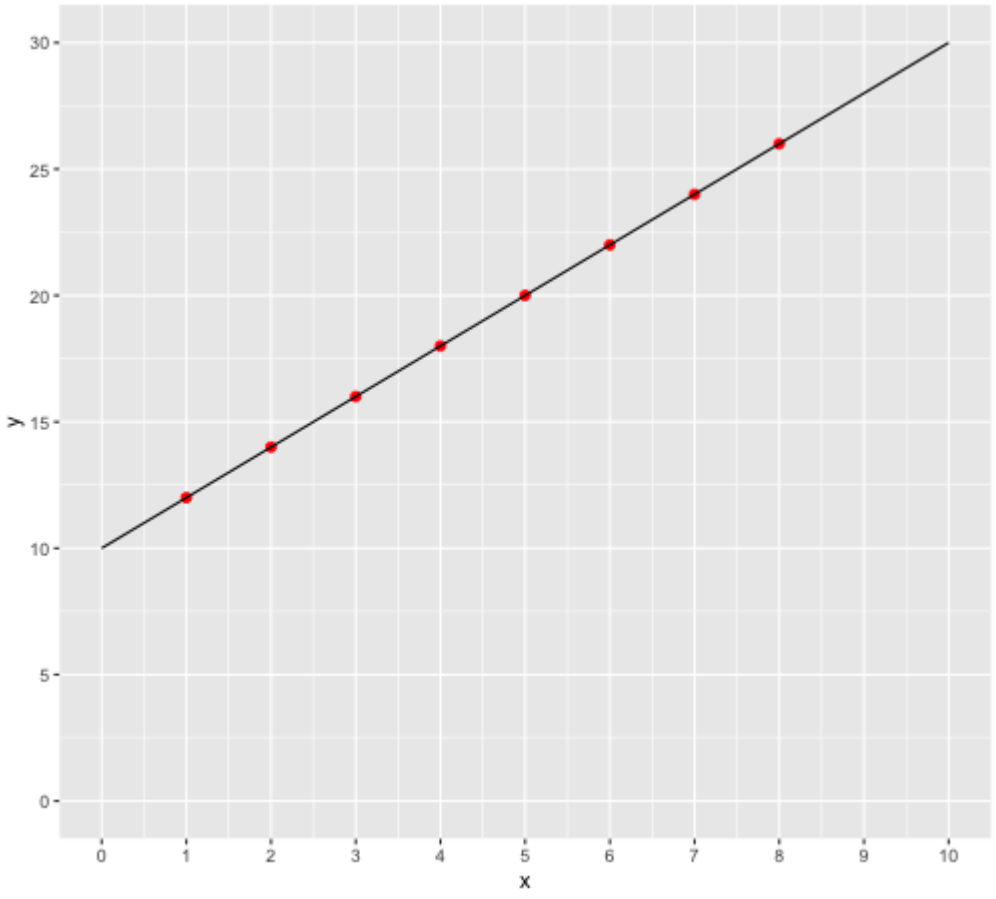
Visualising Functions



Our Data

x	y
1	12
2	14
3	16
4	18
5	20
6	22
7	24
8	26

Visualising Functions (R-code)



R-code

```
fun1 <- function(x) 10+(2*x)

ggplot(func_x, aes(x, y)) +
  geom_point(colour = "red", size = 2) +
  stat_function(fun = fun1) +
  scale_x_continuous(name = "x",
                     breaks = c(0:10),
                     labels = c(0:10),
                     limits = c(0,10)) +
  scale_y_continuous(name = "y",
                     breaks = c(seq(0,30,5)),
                     labels = c(seq(0,30,5)),
                     limits = c(0,30))
```

Multiple arguments

- Functions can take multiple arguments. Consider:

$$y = f(x, z) = 10 + (x * z)$$

- Where:

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$z = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Multiple arguments

x	z
1	1
1	2
1	3
2	1
2	2
2	3
3	1
3	2
3	3

- Notice that when we have multiple inputs, our rows correspond to pairs of inputs.
- So x = 1, pairs with:
  - z = 1
  - z = 2
  - z = 3
- and so on for all values of x

Multiple arguments

x	z	f(x,z)
1	1	11
1	2	12
1	3	13
2	1	12
2	2	14
2	3	16
3	1	13
3	2	16
3	3	19

$$y = f(x, z) = 10 + (x * z)$$

- Example 1, row 2

$$10 + (1 * 2) = 12$$

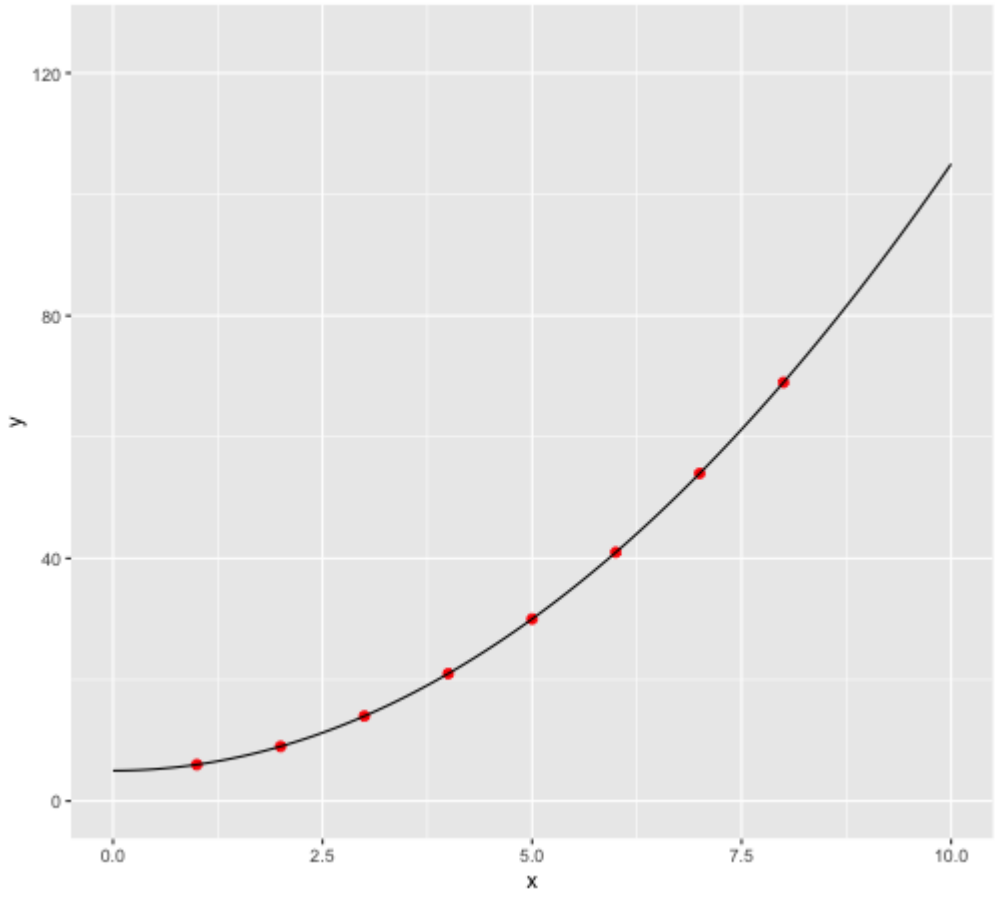
- Example, row 8

$$10 + (3 * 2) = 16$$

Linear vs non-linear functions

- Each of the examples so far have been linear functions.
    - If we plot them, we get a straight line (or flat surface)
- Can also have non-linear functions:
    - A non-linear function would contain powers or roots

Non-linear functions



Example of non-linear function

$$y = f(x) = 5 + x^2$$

Why are functions important?

- There are going to be lots of examples of functions in action.
- Two primary examples are:
    - **Data transformations**
    - **Describing formal models**
- We will start with transformations, and come back to models at the end of the course.

z-scores

- One of the most common transformations in data analysis is standardizing variables.
- What is standardizing?
    - It is putting all variables onto the same scale so they can be compared.
- We refer to standardized variables as z-scores (the reason we will explain later)
- z-score:

$$z = \frac{x - \mu}{\sigma}$$

Z-score for measured variable

- z-score for x:

$$z_{x_i} = \frac{x_i - \bar{x}}{s_x}$$

- Where
  - x_i = individual score on x
  - $\bar{x}$ = mean of x
  - s_x = standard deviation of x

z-scores

- A z-score will have a mean = 0, and a SD = 1.
- What this means is there is a standard way to interpret z-scores.
    - z-score = 1.5, means a respondent is 1.5 SD above the mean.
    - z-score = -2, means a respondent is 2 SD below the mean.

Summary of today

- Functions take input, do something, and produce an output.
- Functions can have multiple arguments, be linear or non-linear
- Typically we will visualize functions
- We use functions frequently in statistics.
- In fact almost everything we are going to see involves functions.

Next tasks

- This week:
  - Complete your lab
  - Come to office hours
- Quiz on continuous data
  - Open Monday 09:00
  - Closes Sunday 17:00
  - Feedback available after due date
- Week 6
  - No lectures on Monday and Tuesday
  - Labs will still take place
  - In the labs
  - check the feedback given for your formative report
  - study the example solutions and ask questions if anything is unclear