

## Tidy data and data files

## Background mathematics

- ▶ Addition
- ▶ Subtraction
- ▶ Multiplication
- ▶ Division
- ▶ Exponents
- ▶ Roots

# Addition

- ▶ Natural numbers
- ▶ 1, 2, 3, 4, ...

### **Addition**

- ▶ Natural numbers
- ▶ 1, 2, 3, 4, ...

### **Subtraction**

- ▶ Integers
- ▶ ..., -2, -1, 0, 1, 2, ...

## Inequalities

**Greater than ( $>$ )**

$$6 > 5$$

## Inequalities

**Greater than ( $>$ )**

$$6 > 5$$

**Less than ( $<$ )**

$$-4 < 2$$

## Inequalities

**Greater than or equal to ( $\geq$ )**

$$P \geq 0.1$$

## Inequalities

**Greater than or equal to ( $\geq$ )**

$$P \geq 0.1$$

**Less than or equal to ( $\leq$ )**

$$P \leq 0.05$$

## Multiplication and Division

### **Multiplication**

▶ Add a number multiple times

▶  $4 + 4 + 4 = 12$

▶  $4 \times 3 = 12$

▶  $4 * 3 = 12$  or  $4(3) = 12$

## Multiplication and Division

### Multiplication

- ▶ Add a number multiple times
- ▶  $4 + 4 + 4 = 12$
- ▶  $4 \times 3 = 12$
- ▶  $4 * 3 = 12$  or  $4(3) = 12$

### Division

- ▶  $12 \div 3 = 4$
- ▶  $12/3 = 4$
- ▶ *Rational* numbers

### Exponent

- ▶ Multiply a number multiple times
- ▶  $4 \times 4 \times 4 = 64$
- ▶  $4^3 = 64$

### Exponent

- ▶ Multiply a number multiple times
- ▶  $4 \times 4 \times 4 = 64$
- ▶  $4^3 = 64$

### Roots

- ▶  $\sqrt[3]{64} = 4$
- ▶  $64^{\frac{1}{3}} = 4$
- ▶ *Real* numbers ( $\sqrt{2} \approx 1.414$ )

## Negative exponents

$$4^{-1} = \frac{1}{4}$$

$$4^{-2} = \frac{1}{16}$$

**Applies to units**

$$6 \text{ g } l^{-1} = 6 \text{ g}/l$$

## Logarithms

Exponent to which a number needs to be raised to get another number,

$$10^3 = 1000.$$

10 raised to the power of 3 equals 1000,

$$\log_{10}(1000) = 3.$$

## Logarithms in figures

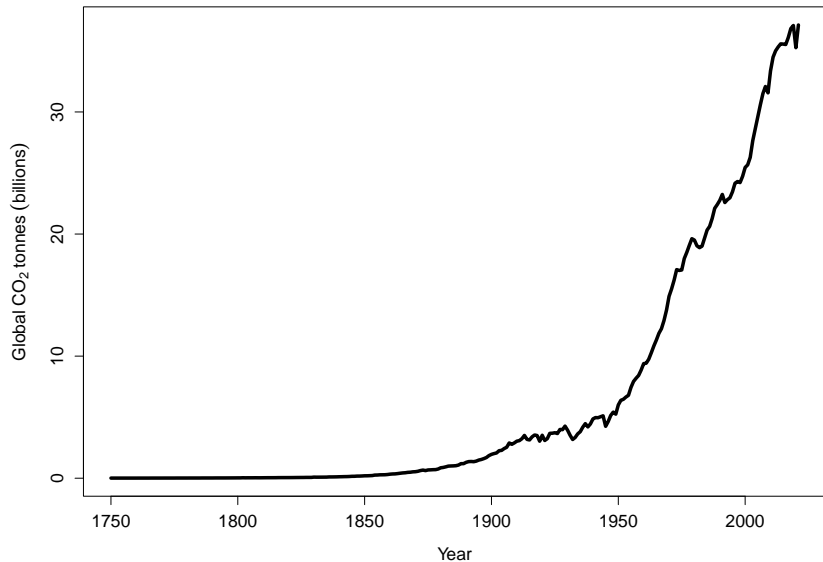


Figure 1: Global carbon dioxide emissions 1750-2021.

## Logarithms in figures

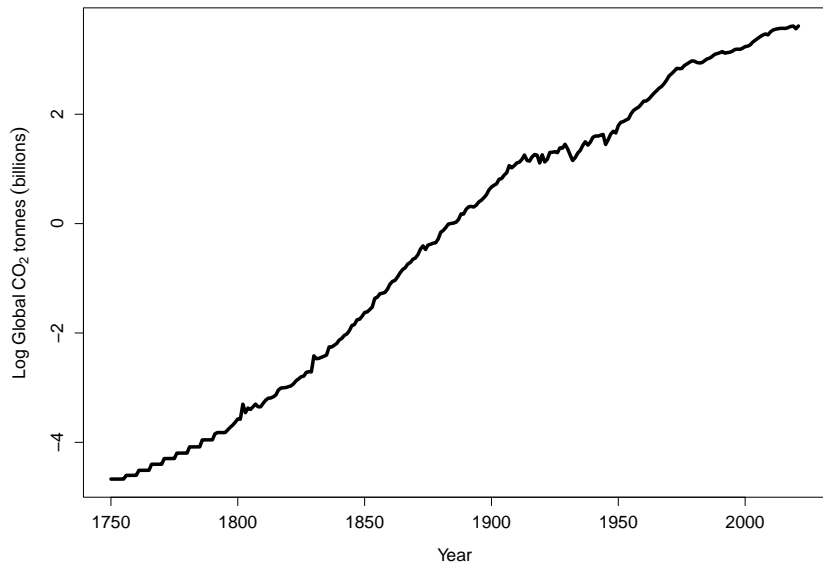


Figure 2: Natural logarithm of global carbon dioxide emissions 1750-2021.

## Natural logarithms

Exponent to which a number needs to be raised to get another number,

$$e^1 \approx 2.718282.$$

Sometimes log denoted 'ln',

$$\ln(2.718282) = 1.$$

## Order of operations

$$x = 3^2 + 2(1 + 3)^2 - 6 \times 0$$

## Order of operations

1. Brackets (parentheses)
2. Exponents & radicals
3. Multiplication & division
4. Addition & subtraction

## Order of operations

Division bars imply parentheses,

$$x = \frac{2^2 + 1}{3^2 + 2}.$$

Above is  $(2^2 + 1)/(3^2 + 2)$ ,

$$x = \frac{(2^2 + 1)}{(3^2 + 2)} = \frac{(4 + 1)}{(9 + 2)} = \frac{5}{11}.$$

## Order of operations

Radicals imply parentheses,

$$x = \sqrt{3 + 4^2} = \sqrt{(3 + 4^2)} \approx 3.59.$$

## Tidy data: Getting data in the right format



Figure 3: Dr Becky Boulton collects data from nest boxes in the field (A), then processes nest material in the lab (B).

## Data collection is messy



Figure 4: Sonoran Desert Rock Fig in the desert of Baja, Mexico.

## Data collection is messy

DATE (m)	SPECIES	SITE NO.	TREE NO.	FRUIT NO.	FRT LENG	FRT WID	FRT
5/4/10	F- <i>pet</i>	70	70	1	15	18	14
5/10/10	F- <i>pet</i>	70	70	2	17	19	15
5/10/10	F- <i>pet</i>	70	70	3	21	21	16
5/11/10	F- <i>pet</i>	70	70	4			
5/11/10	F- <i>pet</i>	70	70	5	15	16	14
5/11/10	F- <i>pet</i>	70	70	6	16	16	15

Figure 5: A portion of a lab notebook used to record measurements of fig fruits from different trees in 2010.

## Observations and variables

- ▶ **Observation:** Units of sample
- ▶ **Variable:** Unit measurement

## Observations and variables

- ▶ **Observation:** Fig trees
- ▶ **Variable:** Tree height,  
location

## Three characteristics of tidy data

Summarised by Wickham (2014)<sup>1</sup>:

1. Each **variable** gets its own column.
2. Each **observation** gets its own row.
3. Different units of observation require different data files.

---

<sup>1</sup>Wickham, H. 2014. *J. Stat. Softw.* 59:1-23.

## Three characteristics of tidy data

Tree	Species	Height (m)	Leaf length (cm)
1	Oak	20.3	8.1
2	Oak	25.4	9.4
3	Maple	18.2	12.5
4	Maple	16.7	11.3

## File extensions and types

- ▶ **File extensions** (e.g., PDF, JPG, DOCX)
- ▶ **Text files** (e.g., TXT, CSV, HTML)
- ▶ **Binary files** (e.g., PNG, MP3, PDF)

## File extensions and types

- ▶ Text files can be opened by a plain text editor
- ▶ Binary files require a specific program
- ▶ Text files (e.g., CSV) are more secure long-term

## Files stored in nested folders

- ▶ Important to organise files in a project
- ▶ One folder for a project, nested files
- ▶ Recommend a folder for SCIU4T4

# Files stored in nested folders

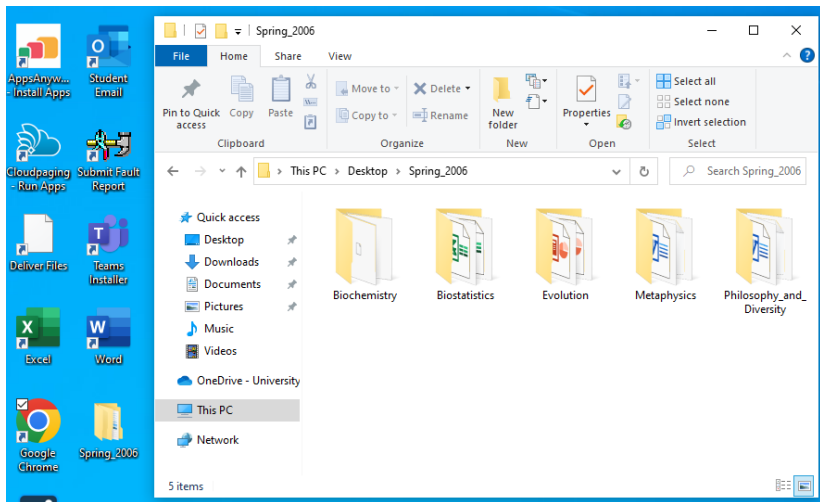


Figure 6: Windows file directory showing the file organisation of modules taken during spring 2006. In this case, the 'Spring\_2006' folder is located on the desktop; the path to the folder is visible in the toolbar above the folders.