# Lecture 2.3 – Reading and Writing Files

## Specific Learning Objectives:

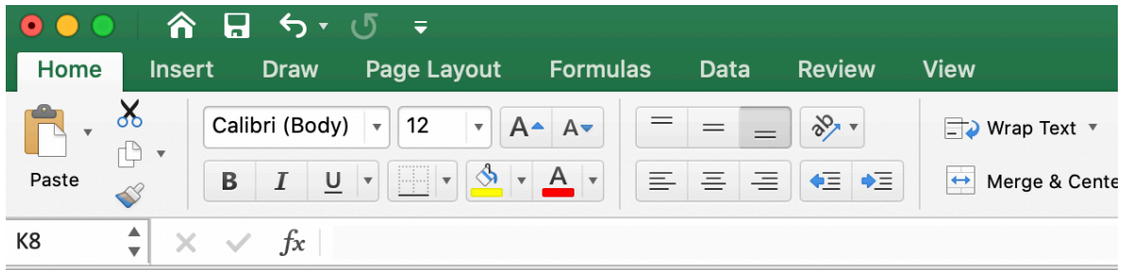
### 3.1 – Load, clean, and organize data in R.

# Using Spreadsheets Safely

- I probably can't convince you to not use spreadsheets, so you may as well use them SAFELY (so you can't upload the data to R later).

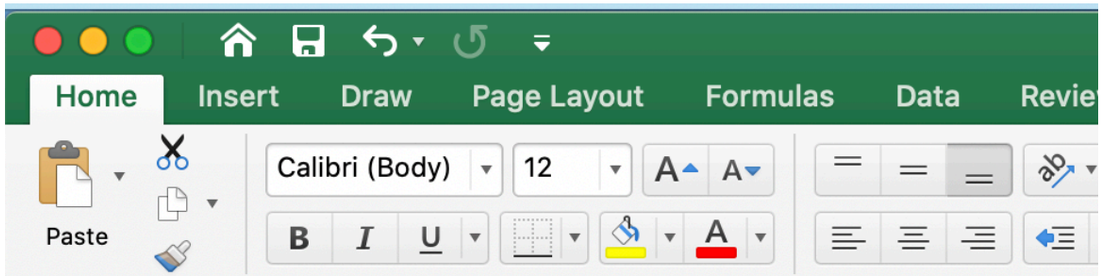
  - Align your data in the upper left, leaving no extra columns on either side.

## DON'T 🙅‍♀️




| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | | | | | | | | | |
| 4 | | | | | | | | | |
| 5 | | | | | | | | | |
| 6 | | | Blue Crab Data | | | | | | |
| 7 | | | | | | | | | |
| 8 | | | Re | Rep | Orient | current | Model | Speed | |
| 9 | | | 1.59 | 1 | B | 0 | C | 0.43839515 | |
| 10 | | | 1.59 | 2 | B | 0 | C | 0.51984394 | |
| 11 | | | 1.59 | 3 | B | 0 | C | 0.4109238 | |
| 12 | | | 1.83 | 1 | B | 4 | C | 0.5698963 | |
| 13 | | | 1.83 | 2 | B | 4 | C | 0.72409106 | |
| 14 | | | 1.83 | 3 | B | 4 | C | 0.57350835 | |
| 15 | | | 2.82 | 1 | B | 15 | C | 1.10059978 | |
| 16 | | | 2.82 | 2 | B | 15 | C | 1.53748625 | |
| 17 | | | 2.82 | 3 | B | 15 | C | 1.27945825 | |
| 18 | | | 0.57 | 1 | B | 0 | C | 0.08905971 | |
| 19 | | | 0.57 | 2 | B | 0 | C | 0.12984493 | |
| 20 | | | 0.57 | 3 | B | 0 | C | 0.07397596 | |
| 21 | | | 0.23 | 1 | B | 4 | C | 0.04501641 | |
| 22 | | | 0.23 | 2 | B | 4 | C | 0.03796072 | |
| 23 | | | 0.23 | 3 | B | 4 | C | 0.04804528 | |
| 24 | | | 0.94 | 1 | B | 15 | C | 0.16302115 | |
| 25 | | | | | | | | | |
| 26 | | | | | | | | | |
| 27 | | | | | | | | | |

## DO 👍🏻



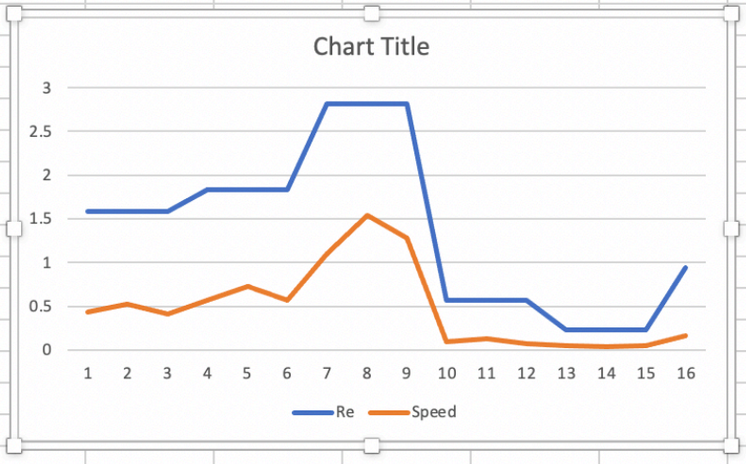

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Re | Rep | Orient | current | Model | Speed | Refresh |
| 2 | 1.59 | 1 | B | 0 | C | 0.43839515 | |
| 3 | 1.59 | 2 | B | 0 | C | 0.51984394 | |
| 4 | 1.59 | 3 | B | 0 | C | 0.4109238 | |
| 5 | 1.83 | 1 | B | 4 | C | 0.5698963 | |
| 6 | 1.83 | 2 | B | 4 | C | 0.72409106 | |
| 7 | 1.83 | 3 | B | 4 | C | 0.57350835 | |
| 8 | 2.82 | 1 | B | 15 | C | 1.10059978 | |
| 9 | 2.82 | 2 | B | 15 | C | 1.53748625 | |
| 10 | 2.82 | 3 | B | 15 | C | 1.27945825 | |
| 11 | 0.57 | 1 | B | 0 | C | 0.08905971 | |
| 12 | 0.57 | 2 | B | 0 | C | 0.12984493 | |
| 13 | 0.57 | 3 | B | 0 | C | 0.07397596 | |
| 14 | 0.23 | 1 | B | 4 | C | 0.04501641 | |
| 15 | 0.23 | 2 | B | 4 | C | 0.03796072 | |
| 16 | 0.23 | 3 | B | 4 | C | 0.04804528 | |
| 17 | 0.94 | 1 | B | 15 | C | 0.16302115 | |
| 18 | 0.94 | 2 | B | 15 | C | 0.20110097 | |
| 19 | 0.94 | 3 | B | 15 | C | 0.15687138 | |

# Using Spreadsheets Safely

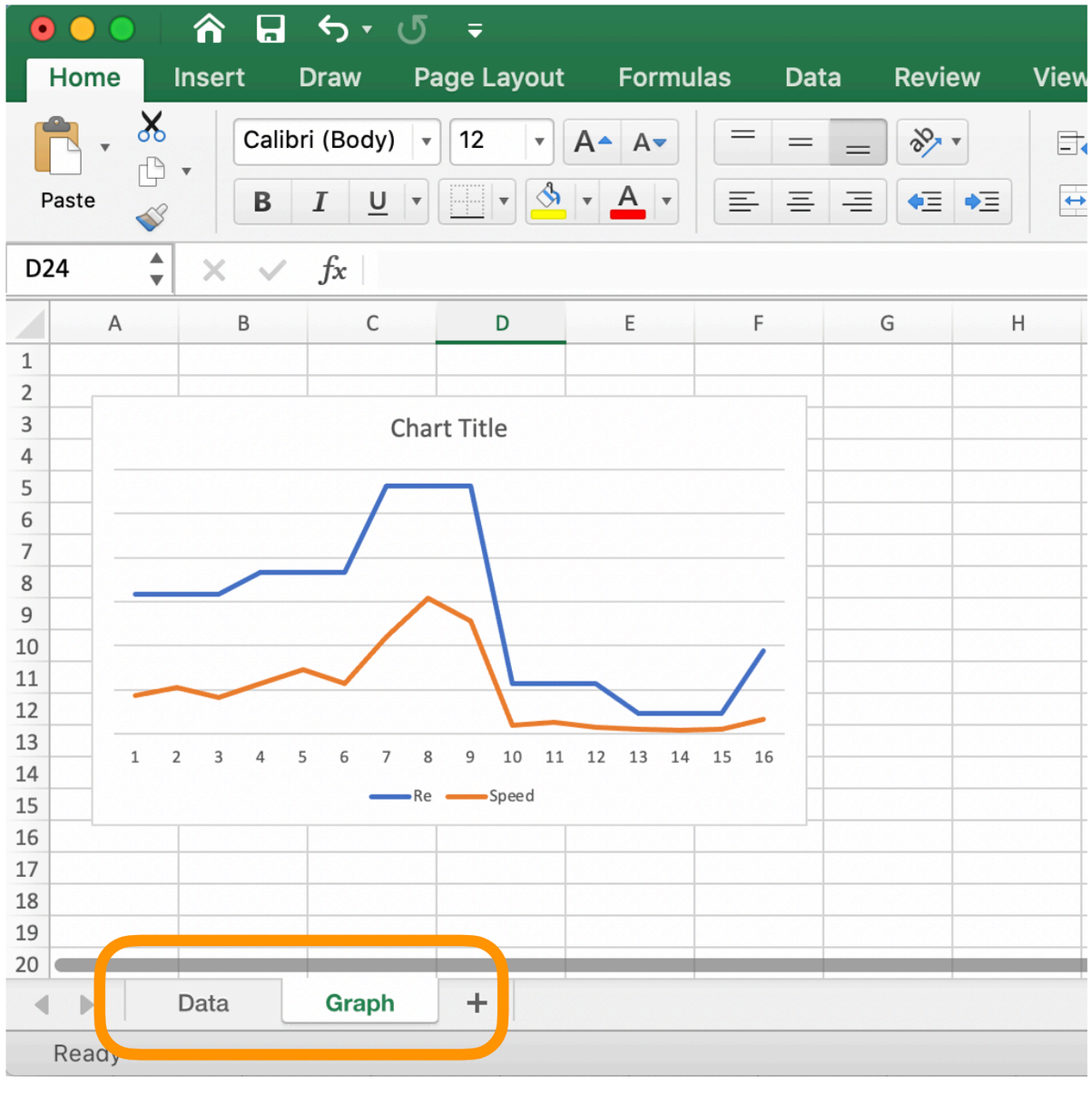
- Put objects like charts and graphs on a separate sheet from the data.

## DON'T 🤦🏼‍♀️

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 6 | | | Blue Crab Data | | | | | | |
| 7 | | | | | | | | | |
| 8 | | | Re | Rep | Orient | current | Model | Speed | |
| 9 | | | 1.59 | 1 | B | 0 | C | 0.43839515 | |
| 10 | | | 1.59 | 2 | B | 0 | C | 0.51984394 | |
| 11 | | | 1.59 | 3 | B | 0 | C | 0.4109238 | |
| 12 | | | 1.83 | 1 | B | 4 | C | 0.5698963 | |
| 13 | | | 1.83 | 2 | B | 4 | C | 0.72409106 | |
| 14 | | | 1.83 | 3 | B | 4 | C | 0.57350835 | |
| 15 | | | 2.82 | 1 | B | 15 | C | 1.10059978 | |
| 16 | | | 2.82 | 2 | B | 15 | C | 1.53748625 | |
| 17 | | | 2.82 | 3 | B | 15 | C | 1.27945825 | |
| 18 | | | 0.57 | 1 | B | 0 | C | 0.08905971 | |
| 19 | | | 0.57 | 2 | B | 0 | C | 0.12984493 | |
| 20 | | | 0.57 | 3 | B | 0 | C | 0.07397596 | |
| 21 | | | 0.23 | 1 | B | 4 | C | 0.04501641 | |
| 22 | | | 0.23 | 2 | B | 4 | C | 0.03796072 | |
| 23 | | | 0.23 | 3 | B | 4 | C | 0.04804528 | |
| 24 | | | 0.94 | 1 | B | 15 | C | 0.16302115 | |




## DO 👍🏻



# Using Spreadsheets Safely

- Only put one data set per sheet.

## DON'T 🙅

J6 | Baby crab data

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | | | Blue Crab Data | | | | | | | Baby crab data | | | | | | | |
| 8 | | | Re | Rep | Orient | current | Model | Speed | | number | size (mm) | log size | sex | Hair diameter | Hair length | Cuticle thickness | |
| 9 | | | 1.59 | 1 | B | 0 | C | 0.43839515 | | 1 | 27 | 3.29583687 | m | 8.10E-06 | 648 | 7.68E-07 | |
| 10 | | | 1.59 | 2 | B | 0 | C | 0.51984394 | | 2 | 8.5 | 2.14006616 | f | 6.73E-06 | 465 | 7.19E-07 | |
| 11 | | | 1.59 | 3 | B | 0 | C | 0.4109238 | | 3 | 11 | 2.39789527 | f | | | | |
| 12 | | | 1.83 | 1 | B | 4 | C | 0.5698963 | | 4 | 5 | 1.60943791 | f | 5.68E-06 | 374 | | |
| 13 | | | 1.83 | 2 | B | 4 | C | 0.72409106 | | 5 | 11.5 | 2.44234704 | m | 7.44E-06 | 484 | 8.74E-07 | |
| 14 | | | 1.83 | 3 | B | 4 | C | 0.57350835 | | 6 | 28 | 3.33220451 | m | 8.49E-06 | 660 | 1.01E-06 | |
| 15 | | | 2.82 | 1 | B | 15 | C | 1.10059978 | | 7 | 25 | 3.21887582 | f | 8.01E-06 | 543 | 1.01E-06 | |
| 16 | | | 2.82 | 2 | B | 15 | C | 1.53748625 | | 8 | 15 | 2.7080502 | m | 7.00E-06 | 505 | 7.04E-07 | |
| 17 | | | 2.82 | 3 | B | 15 | C | 1.27945825 | | 9 | 22 | 3.09104245 | f | 8.81E-06 | 552 | | |
| 18 | | | 0.57 | 1 | B | 0 | C | 0.08905971 | | 10 | 4.9 | 1.58923521 | m | 5.69E-06 | 347 | 6.39E-07 | |
| 19 | | | 0.57 | 2 | B | 0 | C | 0.12984493 | | 11 | 5.85 | 1.76644166 | m | 5.16E-06 | 356 | | |
| 20 | | | 0.57 | 3 | B | 0 | C | 0.07397596 | | 12 | 7.12 | 1.96290773 | m | 5.53E-06 | 400 | 6.31E-07 | |
| 21 | | | 0.23 | 1 | B | 4 | C | 0.04501641 | | 13 | 16 | 2.77258872 | m | 8.38E-06 | 435 | 7.91E-07 | |
| 22 | | | 0.23 | 2 | B | 4 | C | 0.03796072 | | 14 | 13 | 2.56494936 | f | 7.75E-06 | 468 | 5.93E-07 | |
| 23 | | | 0.23 | 3 | B | 4 | C | 0.04804528 | | | | | | | | | |
| 24 | | | 0.94 | 1 | B | 15 | C | 0.16302115 | | | | | | | | | |

## DO 👍

D19

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | number | size (mm) | log size | sex | Hair diameter | Hair length | Cuticle thickness | | |
| 2 | 1 | 27 | 3.29583687 | m | 8.10E-06 | 648 | 7.68E-07 | | |
| 3 | 2 | 8.5 | 2.14006616 | f | 6.73E-06 | 465 | 7.19E-07 | | |
| 4 | 3 | 11 | 2.39789527 | f | | | | | |
| 5 | 4 | 5 | 1.60943791 | f | 5.68E-06 | 374 | | | |
| 6 | 5 | 11.5 | 2.44234704 | m | 7.44E-06 | 484 | 8.74E-07 | | |
| 7 | 6 | 28 | 3.33220451 | m | 8.49E-06 | 660 | 1.01E-06 | | |
| 8 | 7 | 25 | 3.21887582 | f | 8.01E-06 | 543 | 1.01E-06 | | |
| 9 | 8 | 15 | 2.7080502 | m | 7.00E-06 | 505 | 7.04E-07 | | |
| 10 | 9 | 22 | 3.09104245 | f | 8.81E-06 | 552 | | | |
| 11 | 10 | 4.9 | 1.58923521 | m | 5.69E-06 | 347 | 6.39E-07 | | |
| 12 | 11 | 5.85 | 1.76644166 | m | 5.16E-06 | 356 | | | |
| 13 | 12 | 7.12 | 1.96290773 | m | 5.53E-06 | 400 | 6.31E-07 | | |
| 14 | 13 | 16 | 2.77258872 | m | 8.38E-06 | 435 | 7.91E-07 | | |
| 15 | 14 | 13 | 2.56494936 | f | 7.75E-06 | 468 | 5.93E-07 | | |
| 16 | | | | | | | | | |
| 17 | | | | | | | | | |
| 18 | | | | | | | | | |

BlueCrabData | BabyCrabData | +

Ready

# Using Spreadsheets Safely

- Use plain text and avoid special characters and inserting things like pictures, clipart, and video.


- If you need to write notes, make a “comments” column and restrict text to that column. Don’t write comments in random cells!

- If a column should be numeric, make sure ALL data is numeric in that column.

- Any calculations should (ideally) go in another sheet.

# Check Your Understanding

**Import the data set in `tomatohaul2021.csv` as `tomato.data`. The first column should be row names, the dates should be characters, and there is a header.**

# Check Your Understanding

**Use the pipe operator to create `babycrab.clean` object that takes `babycrab.data` and cleans it with each of the cleaning steps at once but in this order:**

1. **cleans column names**
2. **removes empty rows and columns**
3. **removes all rows with NA values**

# Check Your Understanding

**Pick a data frame from `data()` and save it as a csv file in the `data/` folder.**

# In-class Exercises

1. **Work on Assignment 2.2 using R Markdown.**

# Action Items

1. **Complete Assignment 2.2 using R Markdown.**

2. **Read Davies Chapter 8 and Chang Chapters 1-2 for next time.**