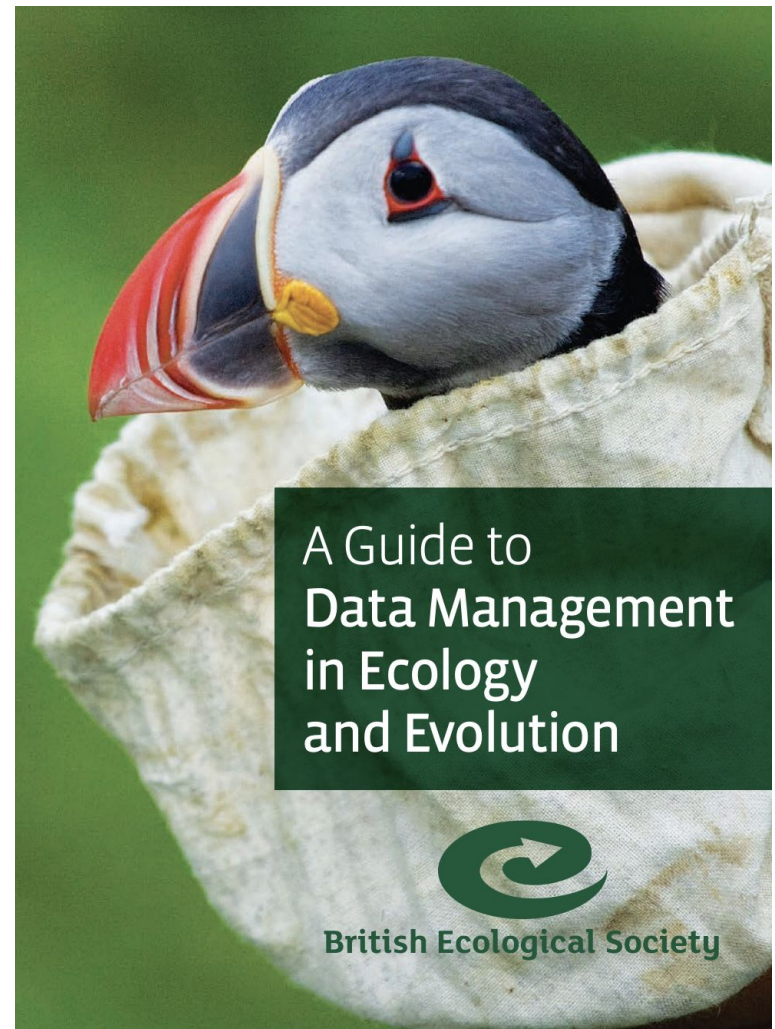
The background features a honeycomb pattern of hexagons. Some hexagons contain logos for R packages: 'dplyr' (top left, with colorful fish), 'readr' (bottom left, with a grid icon), 'stringr' (bottom middle-left, with a string icon), 'purrr' (bottom middle-right, with a cat icon), and a colorful geometric logo (bottom right).

Intro to Tidyverse

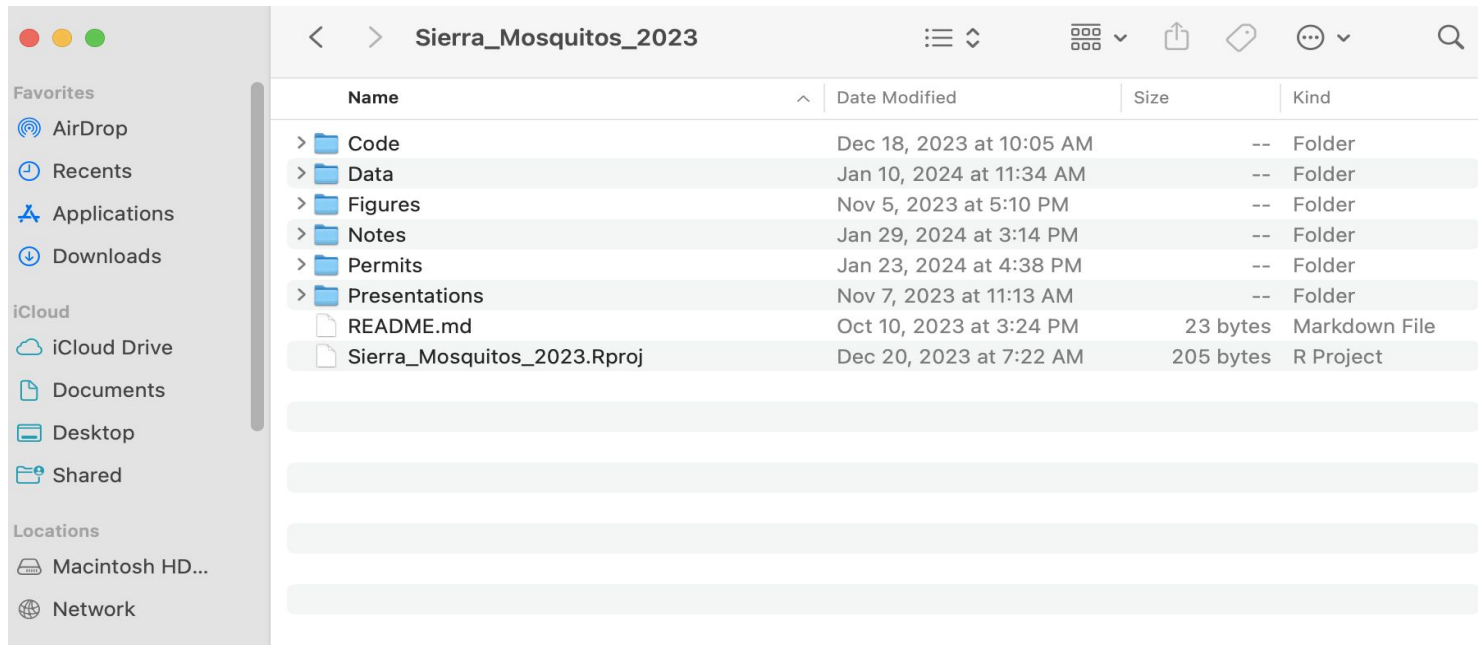
And data management!!!

Data Management

1. Create a git repository online and an RStudio project.
2. Set up an analysis-friendly folder structure



Folder Structure



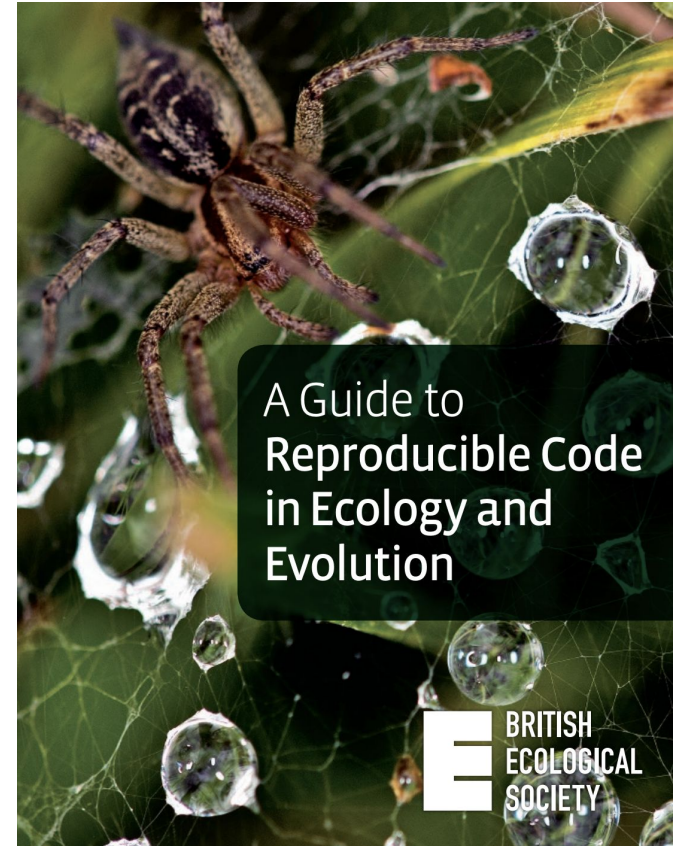
The screenshot shows a macOS Finder window titled 'Sierra_Mosquitos_2023'. The left sidebar contains 'Favorites' (AirDrop, Recents, Applications, Downloads), 'iCloud' (iCloud Drive, Documents, Desktop, Shared), and 'Locations' (Macintosh HD..., Network). The main pane displays a table of the folder's contents.

Name	Date Modified	Size	Kind
> Code	Dec 18, 2023 at 10:05 AM	--	Folder
> Data	Jan 10, 2024 at 11:34 AM	--	Folder
> Figures	Nov 5, 2023 at 5:10 PM	--	Folder
> Notes	Jan 29, 2024 at 3:14 PM	--	Folder
> Permits	Jan 23, 2024 at 4:38 PM	--	Folder
> Presentations	Nov 7, 2023 at 11:13 AM	--	Folder
README.md	Oct 10, 2023 at 3:24 PM	23 bytes	Markdown File
Sierra_Mosquitos_2023.Rproj	Dec 20, 2023 at 7:22 AM	205 bytes	R Project

[A professional GitHub repository](#)

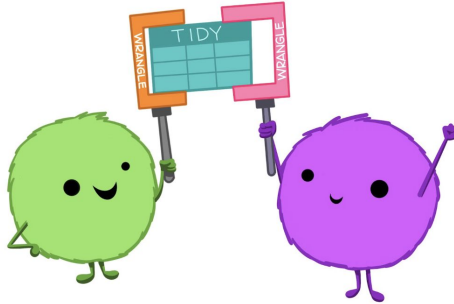
Reproducible Code

A data analyses are **reproducible** if all the information (data, files, etc.) required are available for someone else to re-do your entire analysis.



Benefits of Reproducible Code

- You can figure out what you did six months from now.
- You can quickly make adjustments to code or data.
- You can pass along or share a project with others.
- You can give useful code examples to people who want to extend your research.



Requirements of Reproducible Code

- All your raw data should be saved in the project directory.
- Scripts should be included for all code - data cleaning, analyses, figures, etc.
- You should include details on the versions of any software used.
- If possible, there should be no “by hand” steps.

clean data

“**TIDY DATA** is a standard way of mapping the meaning of a dataset to its structure.”
—HADLEY WICKHAM

In tidy data:

- each variable forms a column
- each observation forms a row
- each cell is a single measurement

each column a variable

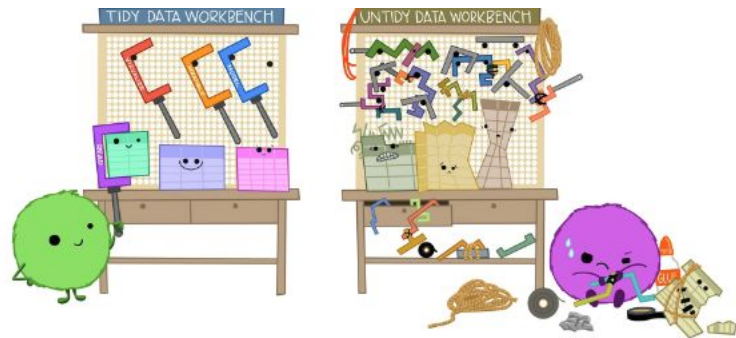
id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

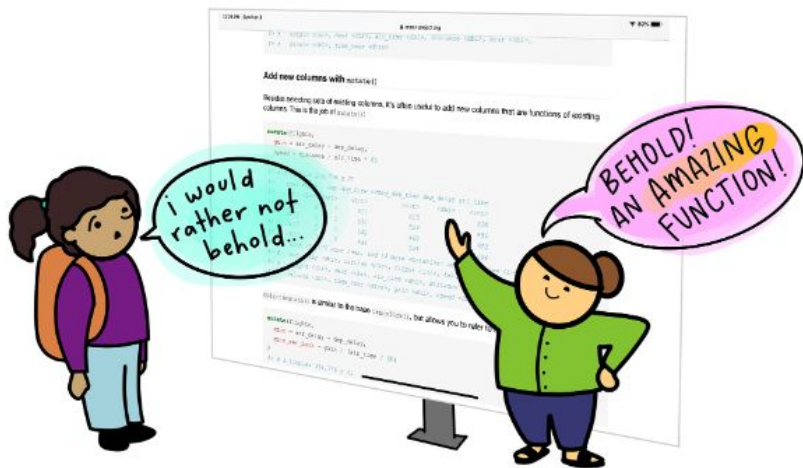
Wickham, H. (2014). Tidy Data. Journal of Statistical Software 59 (10). DOI: 10.18637/jss.v059.i10

You can start this while doing manual entry!

	A	B	C	D	E	F	G	H	I	J
1	sampler	Lake	Date	Time_start	Time_end	Total_time	snow_melt	black_mosq	culiseta	culex_tarsali
2	SC	Heart	14-Jul	18:01	19:57	116	158	0	6	0
3	SL_MR	Heart	19-Jul	15:03	16:03	60	16	1	0	0
4	SC	Heart	19-Jul	15:03	16:03	60	18	0	0	0
5	SC	Heart	25-Jul	18:00	19:00	60	233	2	1	0
6	SL	Heart	25-Jul	18:00	19:00	60	93	5	0	0
7	SC	Heart	25-Jul	19:00	20:00	60	398	6	0	0
8	MR	Heart	25-Jul	19:00	20:00	60	170	1	0	0
9	SC	Heart	3-Aug	15:20	16:20	60	44	0	1	0
10	SL	Heart	3-Aug	15:15	16:15	60	22	0	0	0
11	SC	Heart	3-Aug	16:20	17:20	60	110	0	1	0
12	SL	Heart	3-Aug	16:15	17:15	60	50	0	1	0
13	CC	Heart	15-Aug	16:00	17:00	60	29	0	0	0
14	SC	Heart	15-Aug	16:00	16:30	30	41	0	0	0
15	SC	Heart	15-Aug	16:30	17:00	30	50	0	0	0
16	SC	Heart	22-Aug	17:04	18:04	60	193	0	0	0
17	KK	Heart	22-Aug	17:04	18:04	60	108	0	0	0
18	SL_MR	Hidden	14-Jul	18:01	19:57	116	93	1	1	0
19	SL_MR	Hidden	19-Jul	17:03	18:03	60	21	1	1	0
20	SC	Hidden	19-Jul	17:03	18:03	60	37	1	0	0
21	MR	Hidden	25-Jul	15:00	16:00	60	19	0	0	0
22	SC	Hidden	25-Jul	15:00	16:00	60	6	0	0	0
23	SC	Hidden	25-Jul	16:00	17:00	60	25	0	1	0
24	SL	Hidden	25-Jul	16:00	17:00	60	63	0	0	0

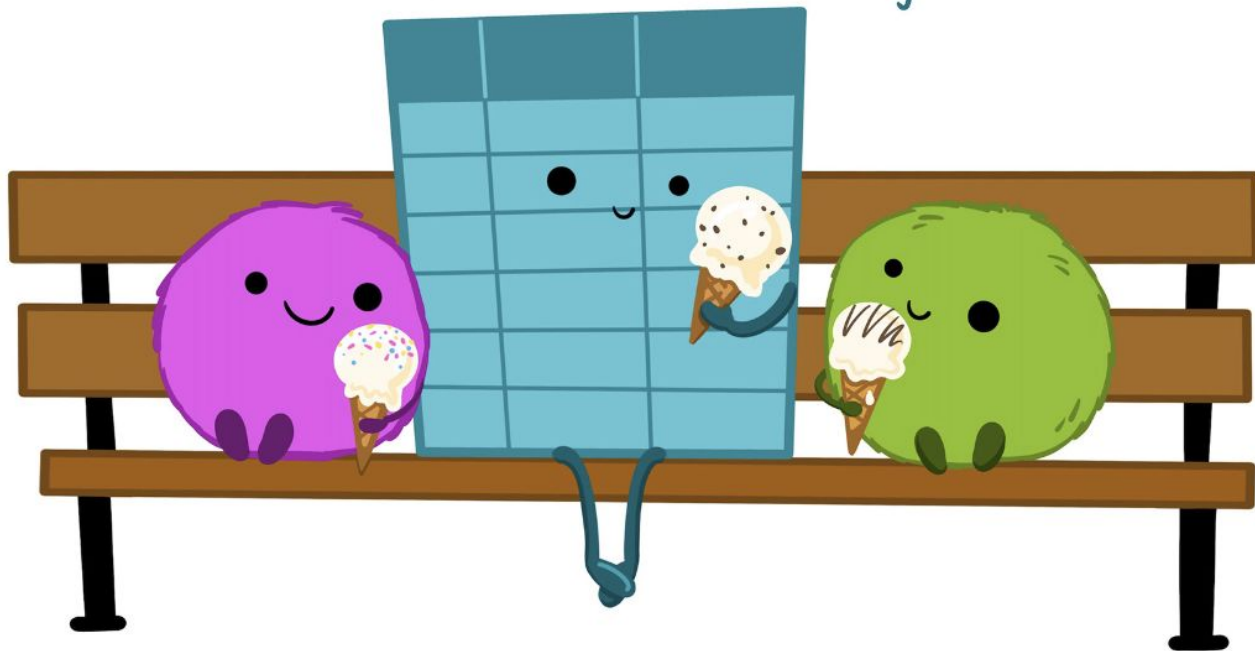


Let's take a closer look at some packages!



Let's talk about Tidyverse:

make friends with tidy data.



Overview of Tidyverse

Today we will be focusing on the packages we find most helpful:

1. **Dplyr** - within dataframe changes
2. **Lubridate** - time series
3. **Tidyr** - reshaping and tidying data (between wide and long formats)

There are other packages that are included in Tidyverse!

1. **ggplot**- graphical presentation of data
2. **purrr**- combining dataframes
3. **Tibble**- making dataframes in R (great for example data)
4. **readr**- reading and writing data in various formats
5. **across**- used to apply a function or transformation to multiple columns of a data frame simultaneously

Quick Intro to Nemo_BO Dataset

- Collected at Blue Oak Ranch (BO)
- Data includes life-history, morphological, and physiological traits for *Nemophila menziesii* (Baby Blue Eyes, Boraginaceae)
- We can use the data to estimate phenotypic selection
 - (the process by which the phenotype of an organism influences its reproductive success in a given environment)



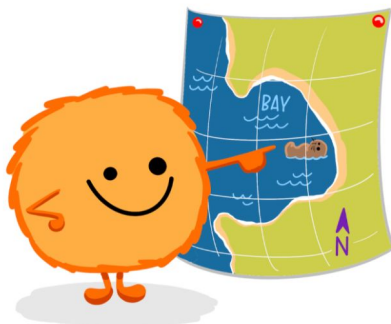
Dplyr: filter

- used to subset data frames based on specific conditions or criteria
- It allows you to extract rows from a dataframe that meet certain logical conditions.

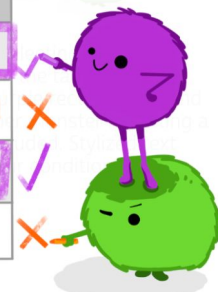
`dplyr::filter()` KEEP ROWS THAT satisfy your CONDITIONS

keep rows from... this data... ONLY IF... type is "otter" AND site is "bay"

```
filter(df, type == "otter" & site == "bay")
```



type	food	site
otter	urchin	bay
shark	seal	channel
otter	abalone	bay
otter	crab	wharf



@alison_horst

Dplyr: Mutate

- Create new columns or modify existing columns in a dataframe based on transformations or calculations
- It allows you to add calculated columns to your data frame without modifying original data



Dplyr: case_when

- It allows you to define multiple conditions and corresponding values to be assigned based on those conditions
- creating new variables based on complex conditional logic in your data analysis or data wrangling tasks

Time series ACF series

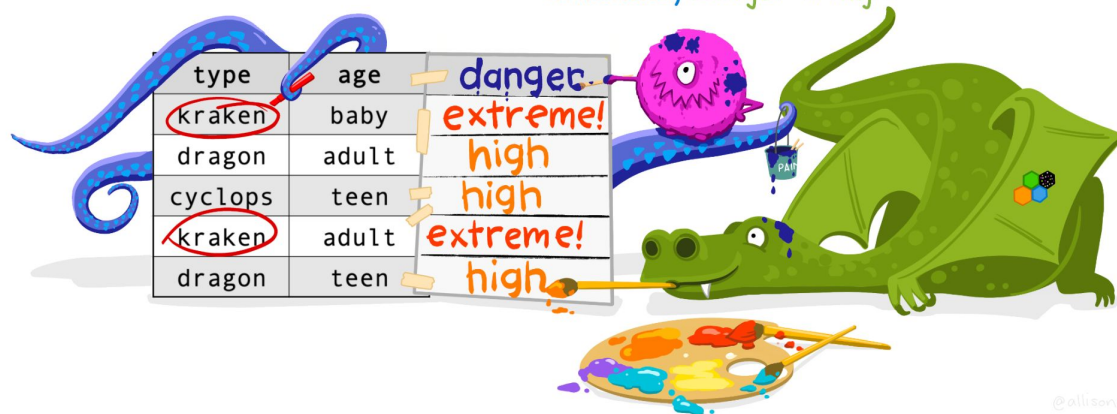
Artistic license

Not on reading "linear", branching across the top of a data table
assumes the function to each column. Student test results.

dplyr::case_when() IF ELSE... (but you love it?)

df %>% ADD COLUMN 'danger'
mutate(danger = case_when(type == "kraken" ~ "extreme!",
TRUE ~ "high"))

IF type is kraken THEN danger is extreme!
TRUE ~ "high"
OTHERWISE, danger is high.



Lubridate

- designed to make working with dates and times easier and more intuitive
- It provides functions to parse, manipulate, and work with date-time objects in a user-friendly manner



here("data", "my_data.csv")

- generates a path to a file named "my_data.csv" located in the "data" directory relative to the project's root directory
- This path remains consistent regardless of the current working directory, simplifying code maintenance and sharing.

