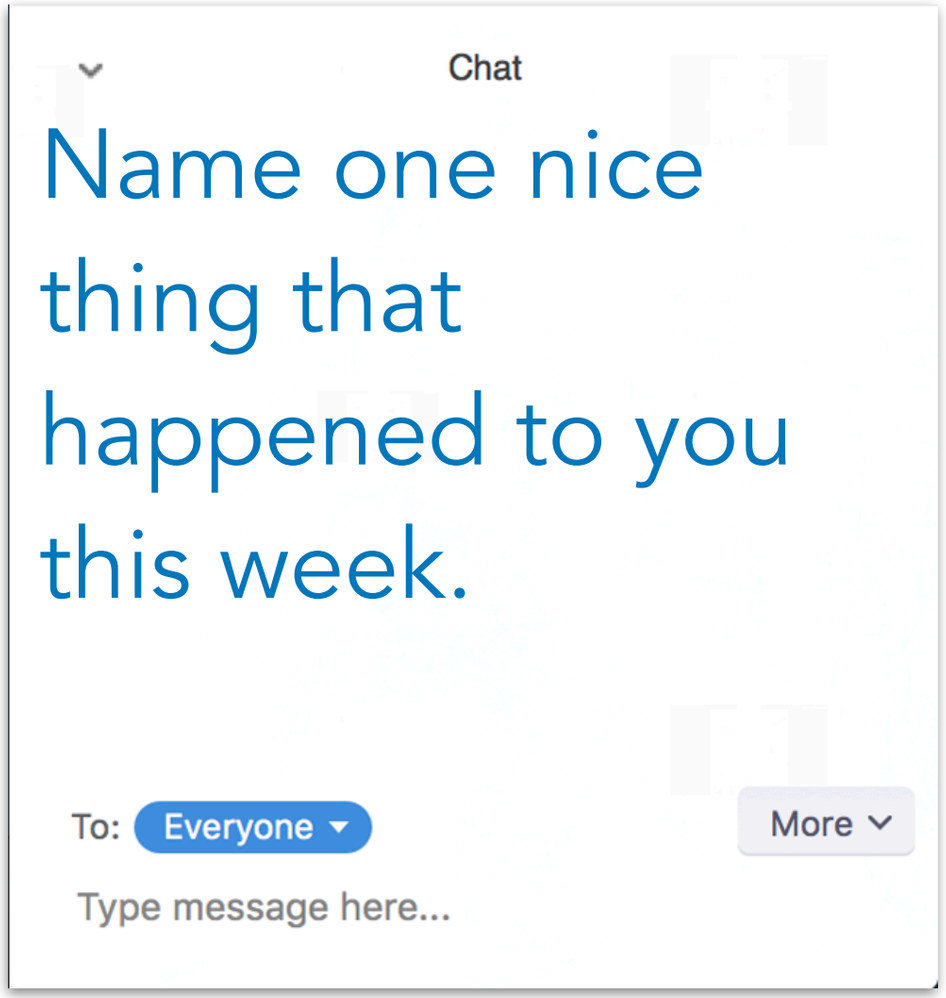
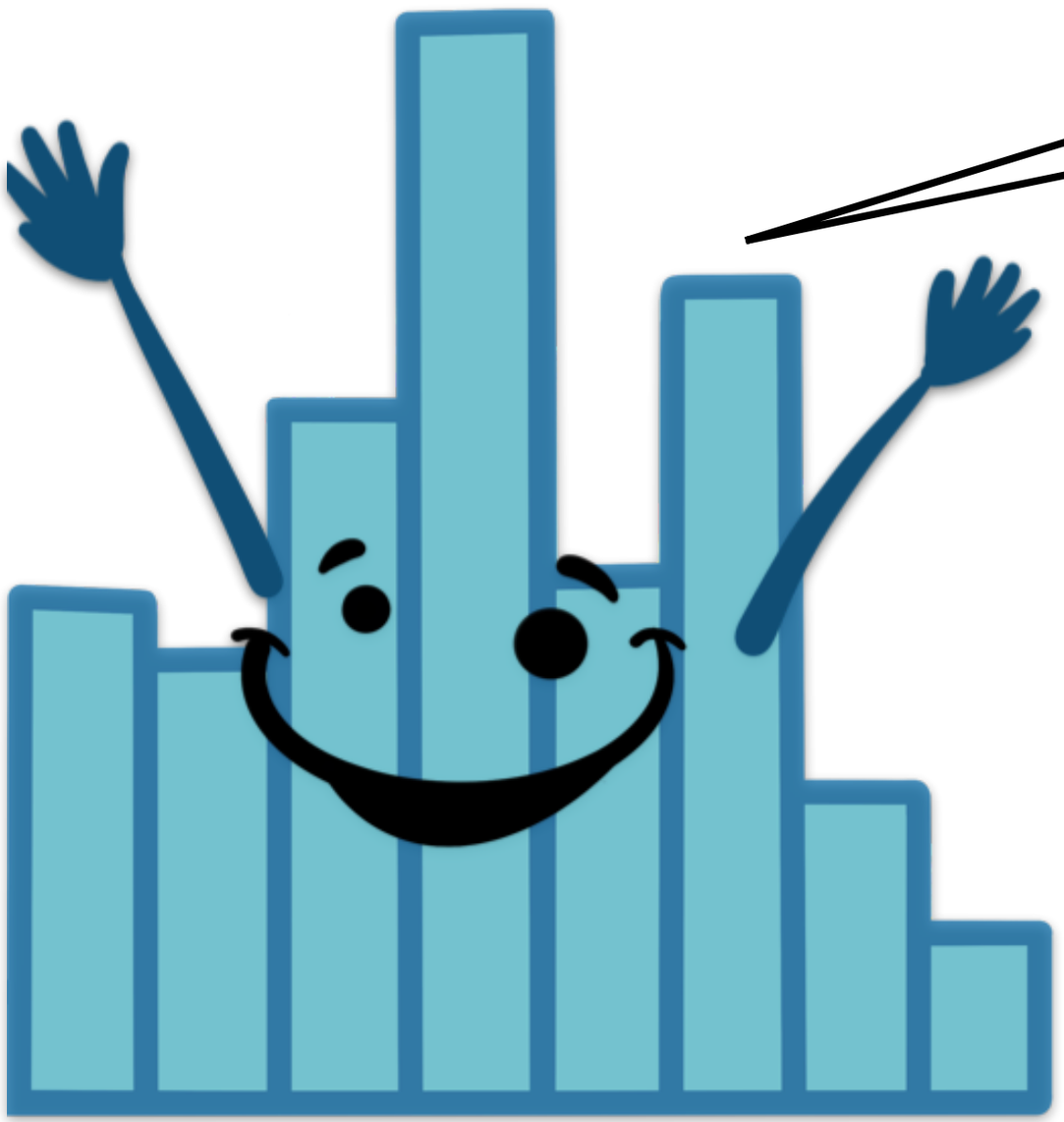
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

Data wrangling 2

"Data don't make any sense,
we will have to resort to statistics."



01/22/2021

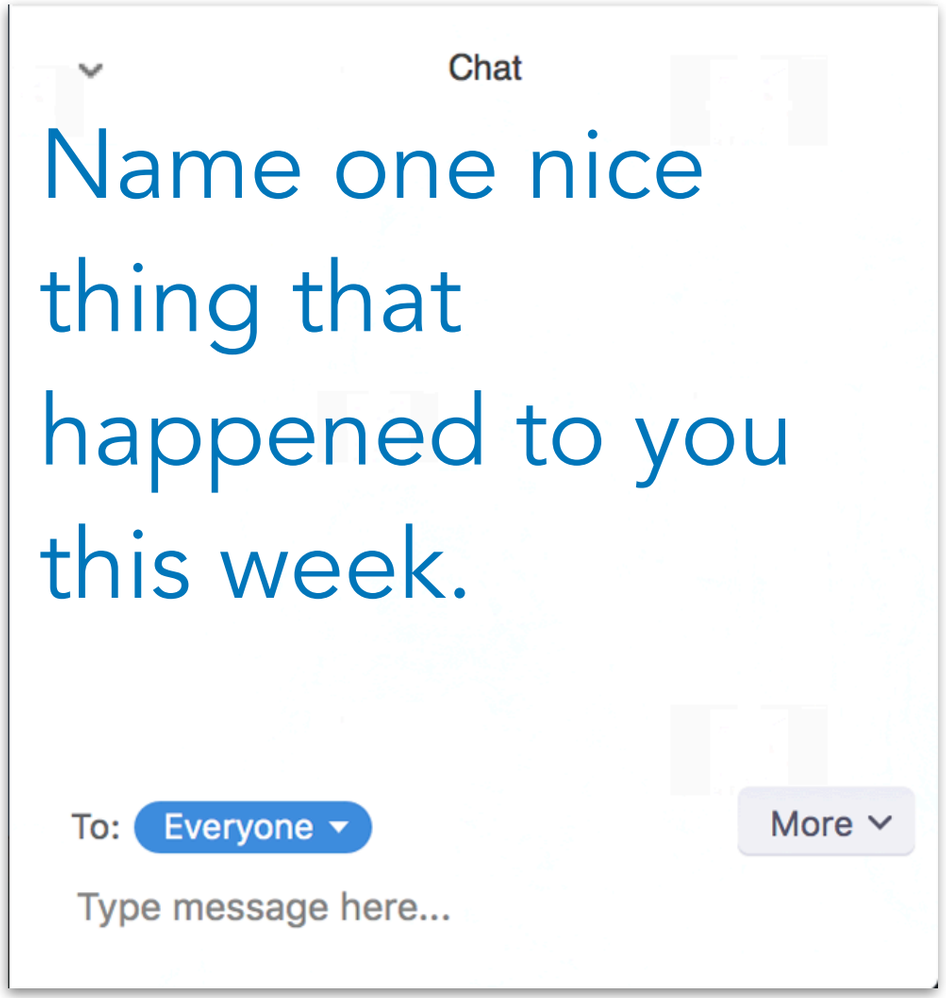


Remember to record the lecture!

PSYCH 252: Statistical Methods for Behavioral and Social Sciences

Data wrangling 2

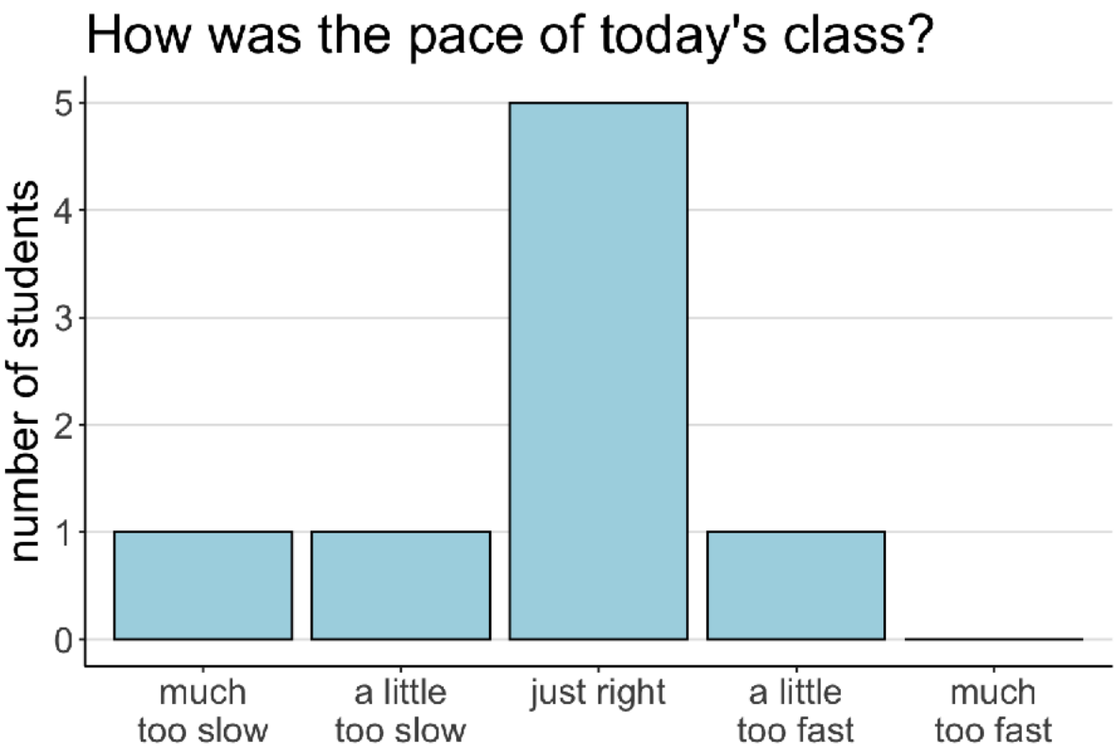
"Data don't make any sense,
we will have to resort to statistics."

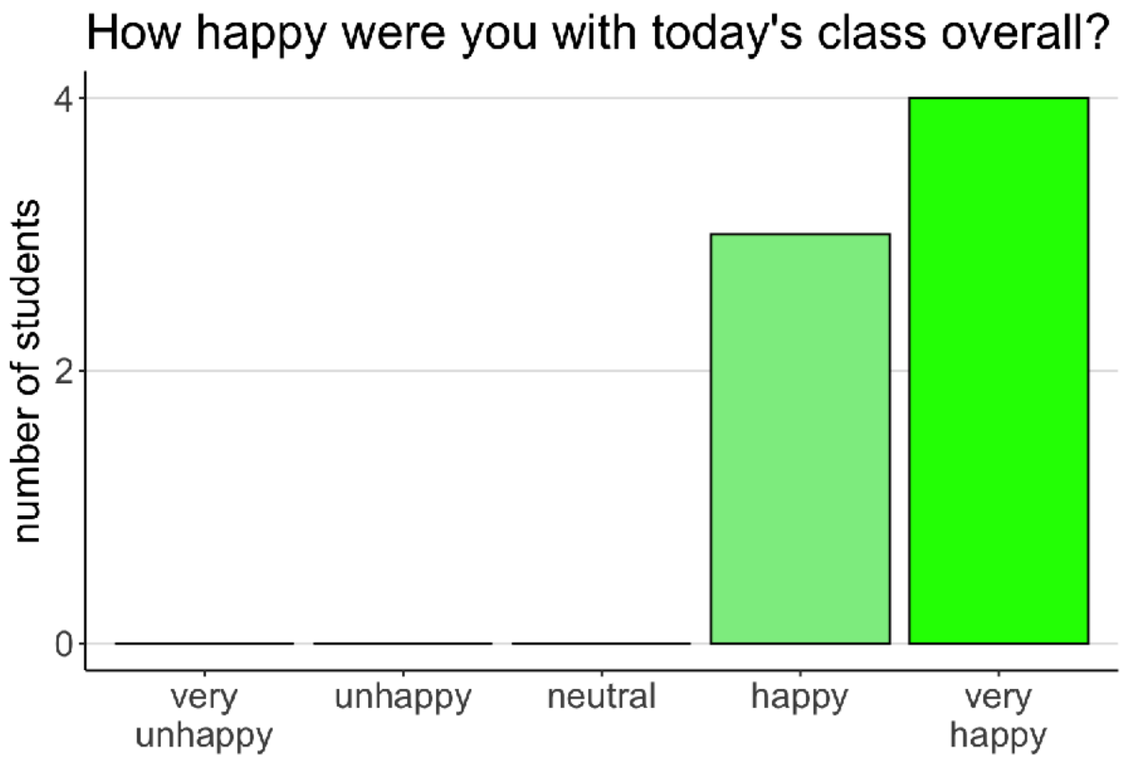


01/22/2021

Your feedback

Your feedback

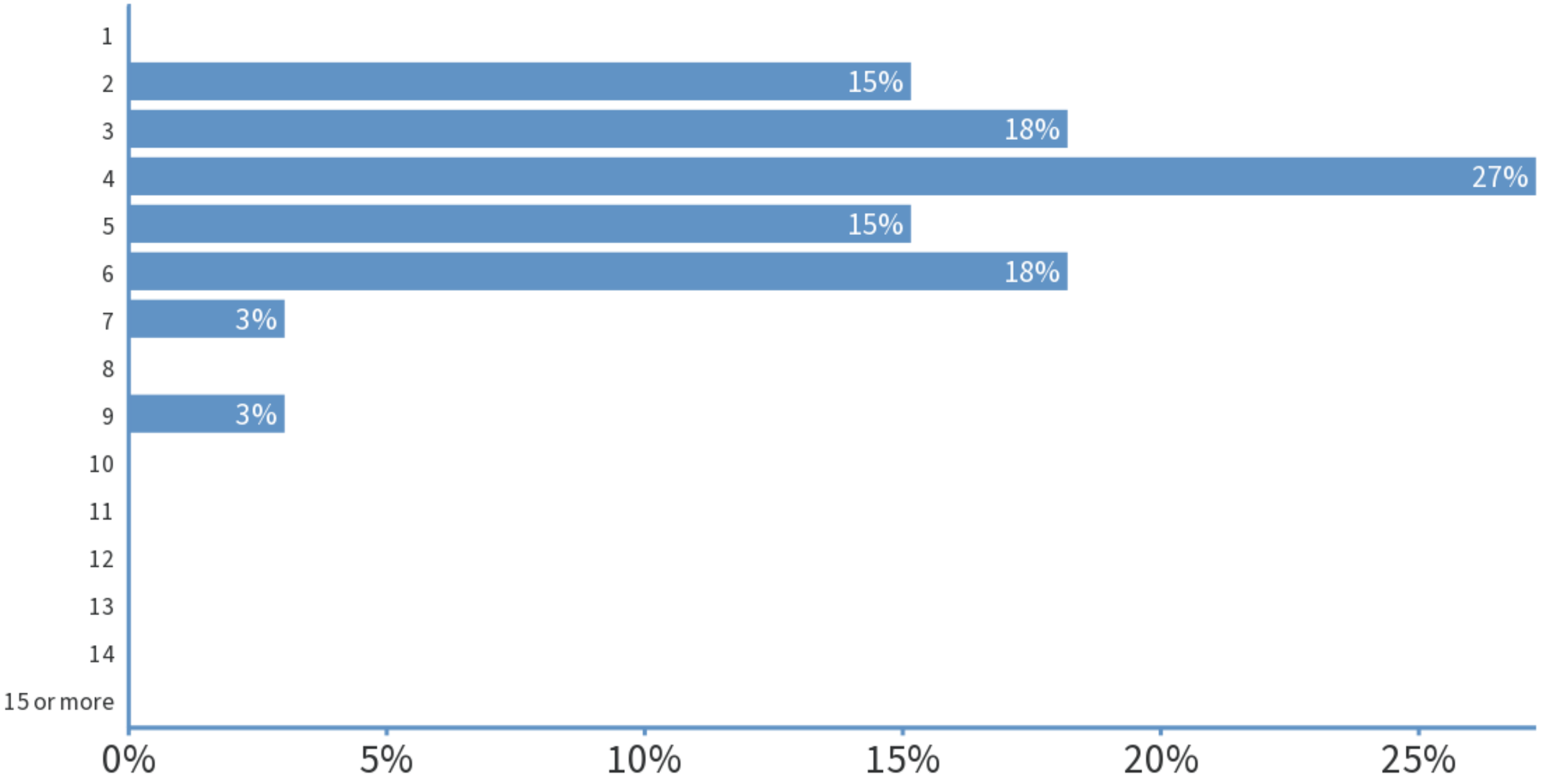




I thought the pacing was better than vis 2 and appreciated the pauses for us to get to code chunks! Thanks for taking these into consideration :)

I think this class might be helpful for someone with no prior programming experience, but I had felt fairly comfortable already and had taken 251, so this class felt slow.

How many hours did it take you to complete Homework 1?



Things that came up

Powerful visualizations with ggplot2

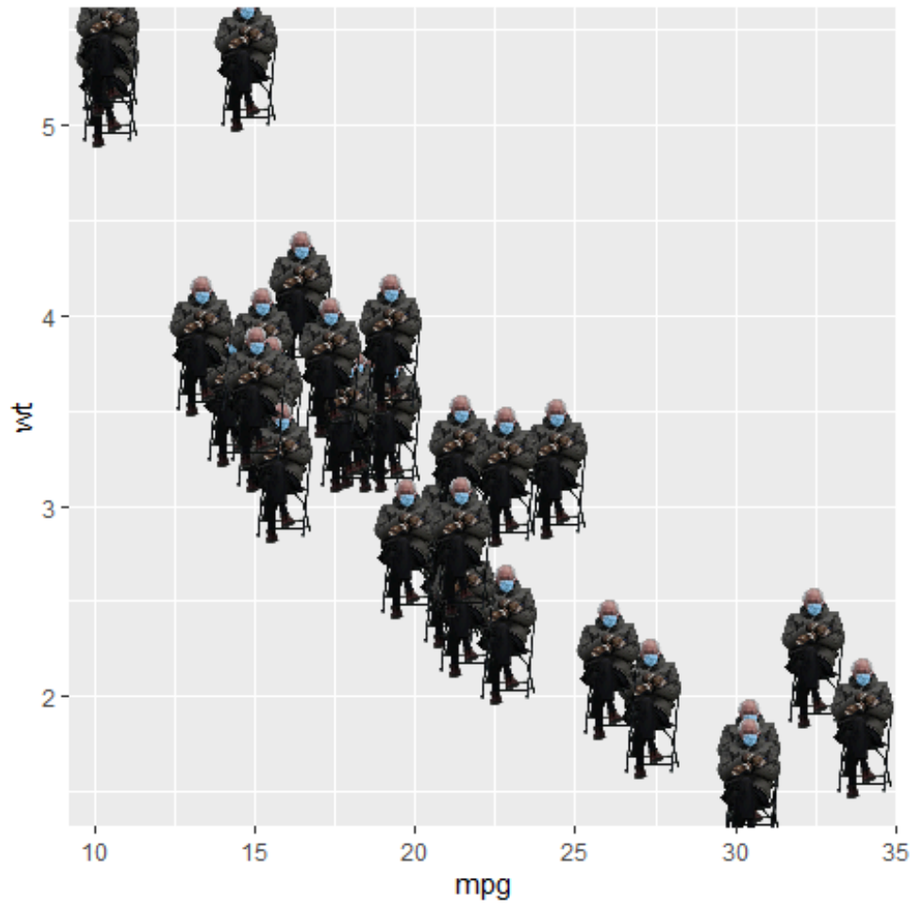
ggbernie

A ggplot2 geom for adding Bernie Sandners

This is a package inspired by a tweet by @samuelmehr

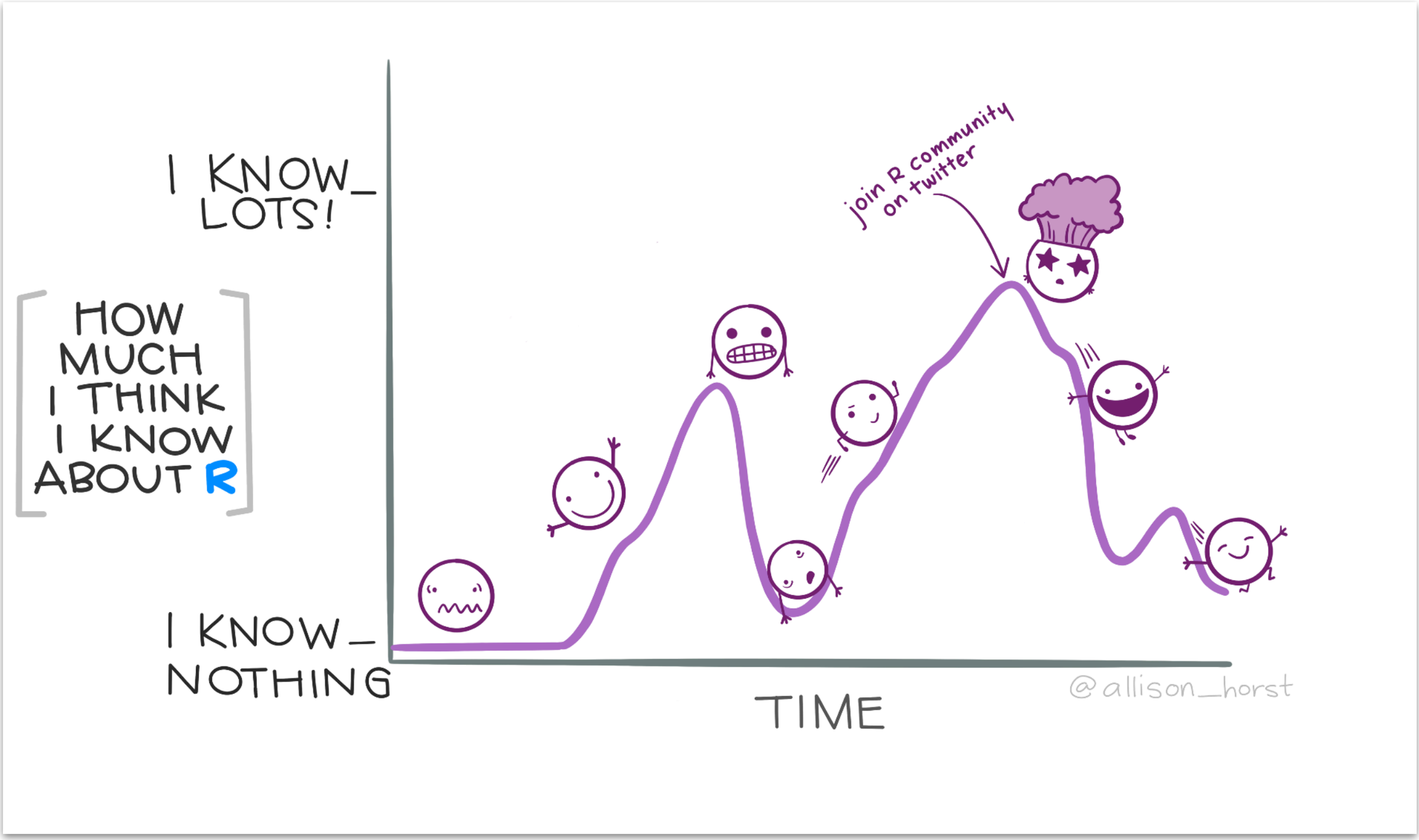
Bernie sitting

```
ggplot(mtcars) +
  geom_bernie(aes(mpg, wt), bernie = "sitting")
```



https://github.com/R-CoderDotCom/ggbernie/

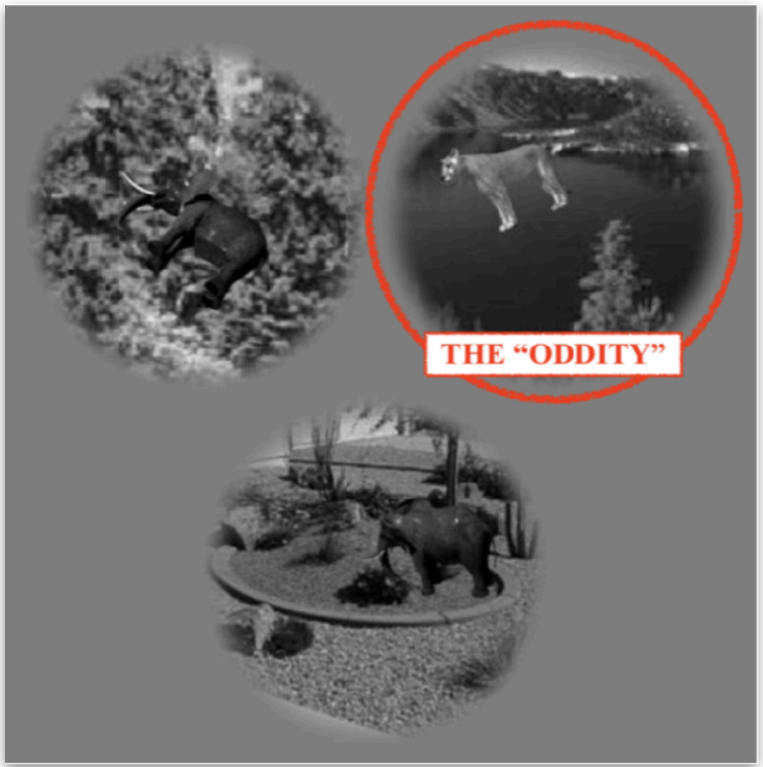
The R learning/knowledge curve



Logistics

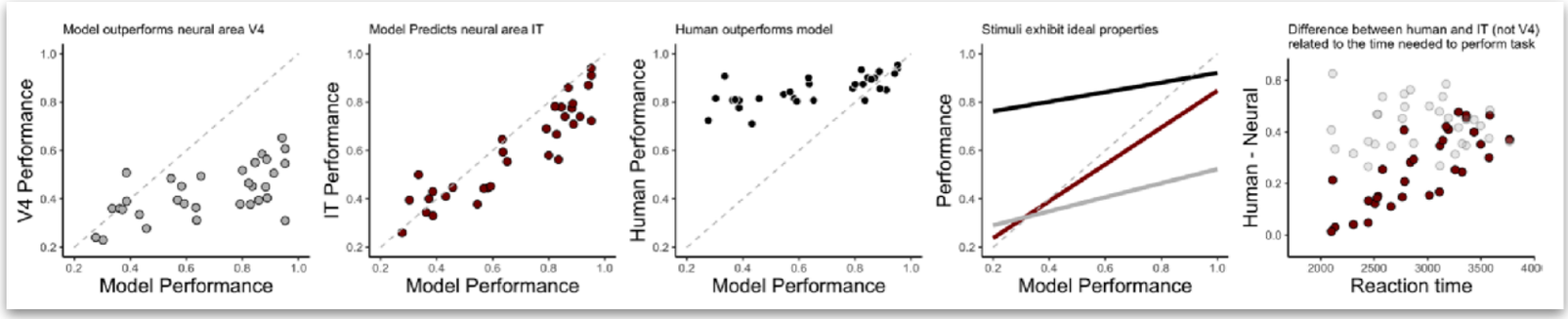
Homework 2

due Thursday 28th, at 8pm



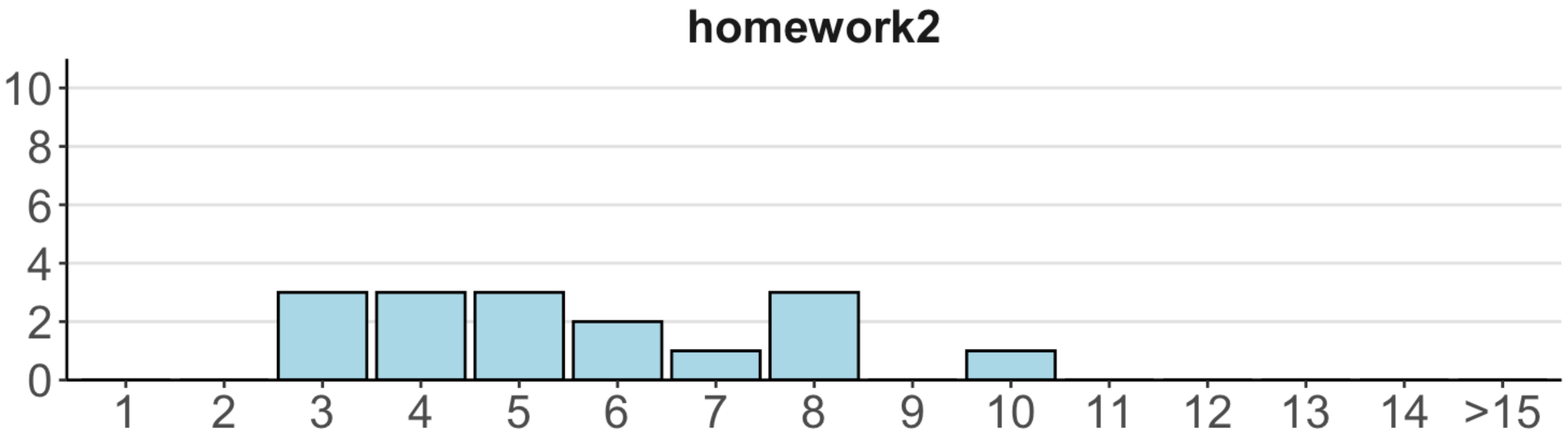
cool experiment

messy dataset



neat plot!

Homework 2



how long it took people last year

Data wrangling time ...

Tidy 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

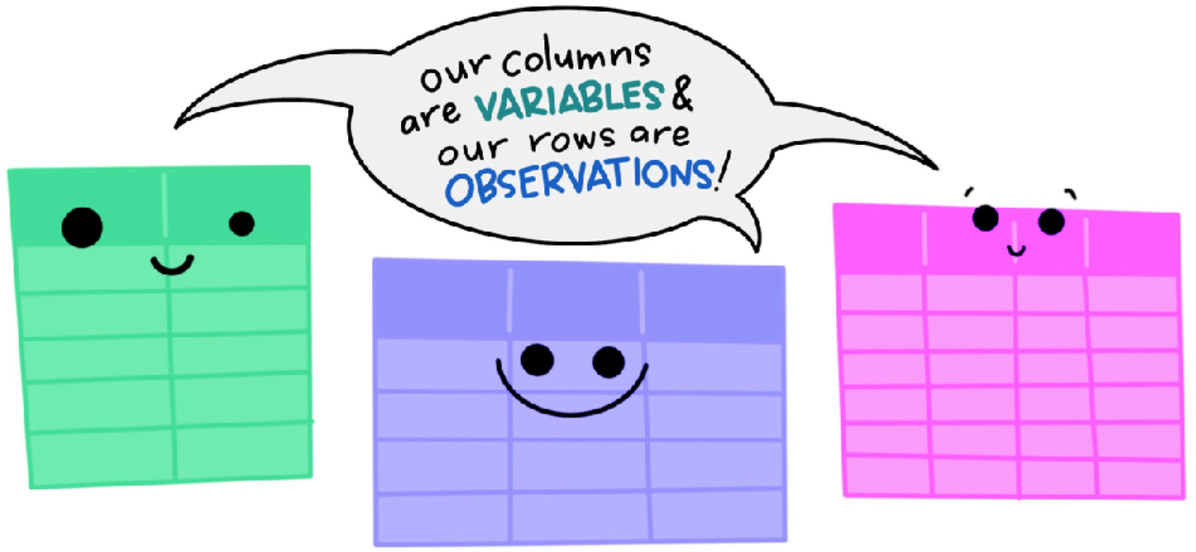
id	name	color
1	floof	gray
2	max	black
3	cat	orange
4	donut	gray
5	merlin	black
6	panda	calico

each row an observation

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

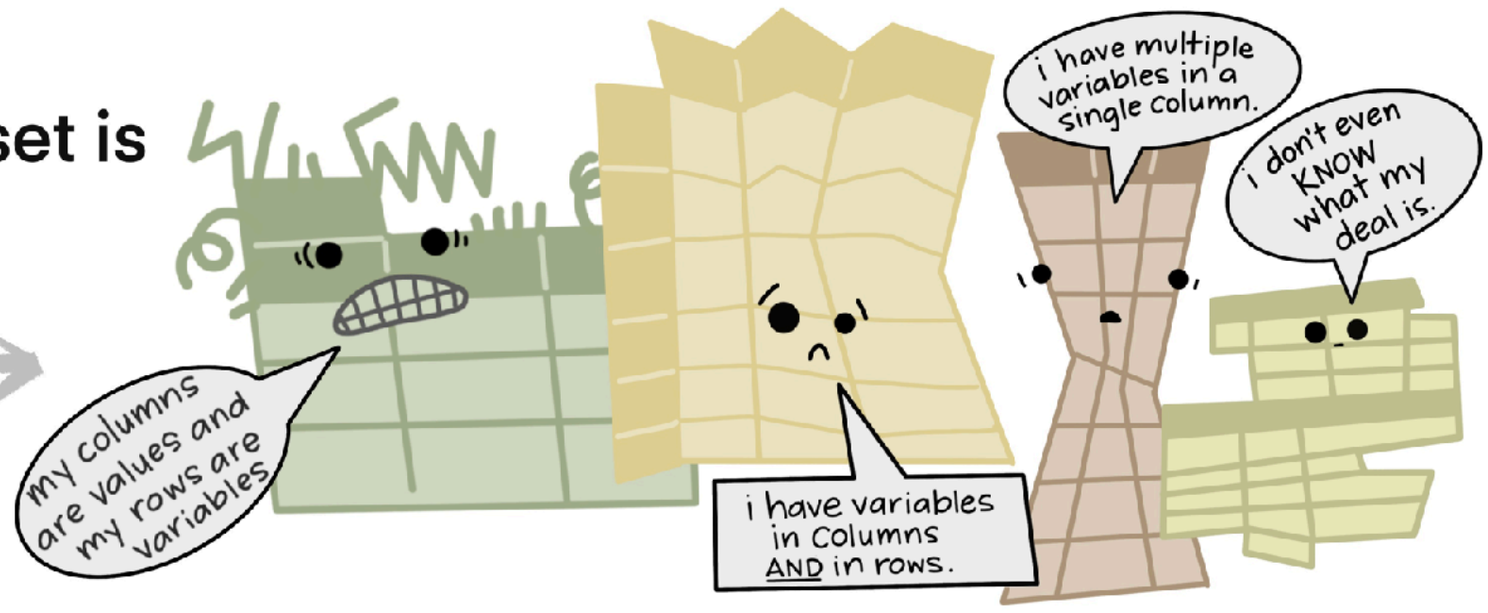
Tidy data

The standard structure of tidy data means that "tidy datasets are all alike..."



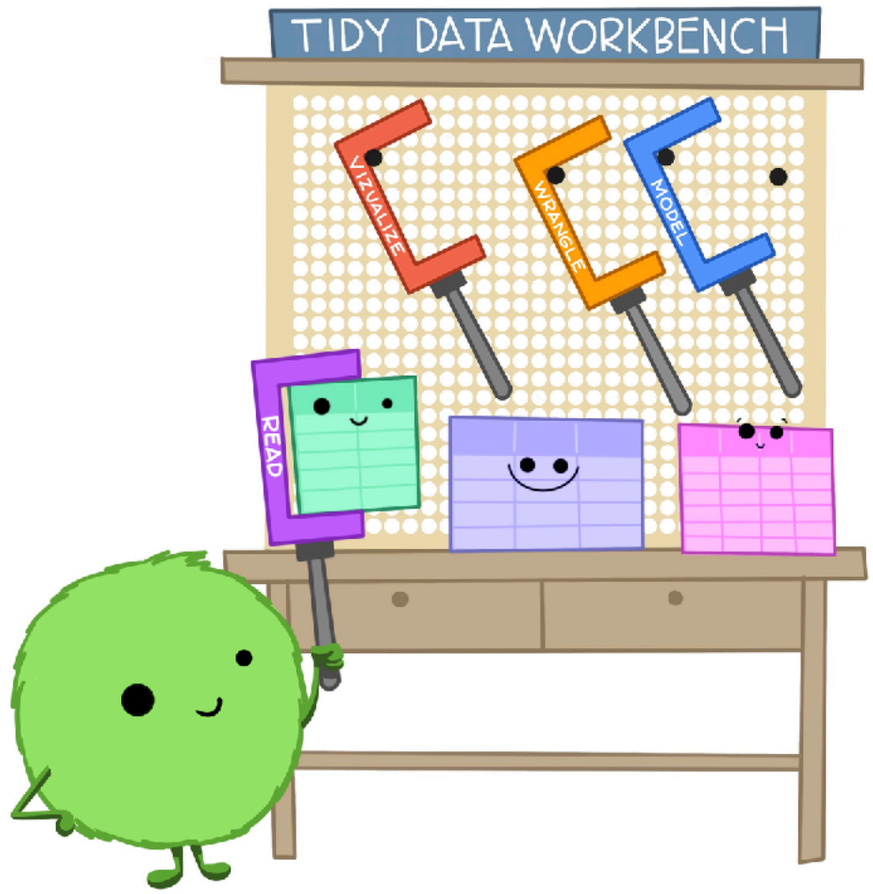
"...but every messy dataset is messy in its own way."

–HADLEY WICKHAM

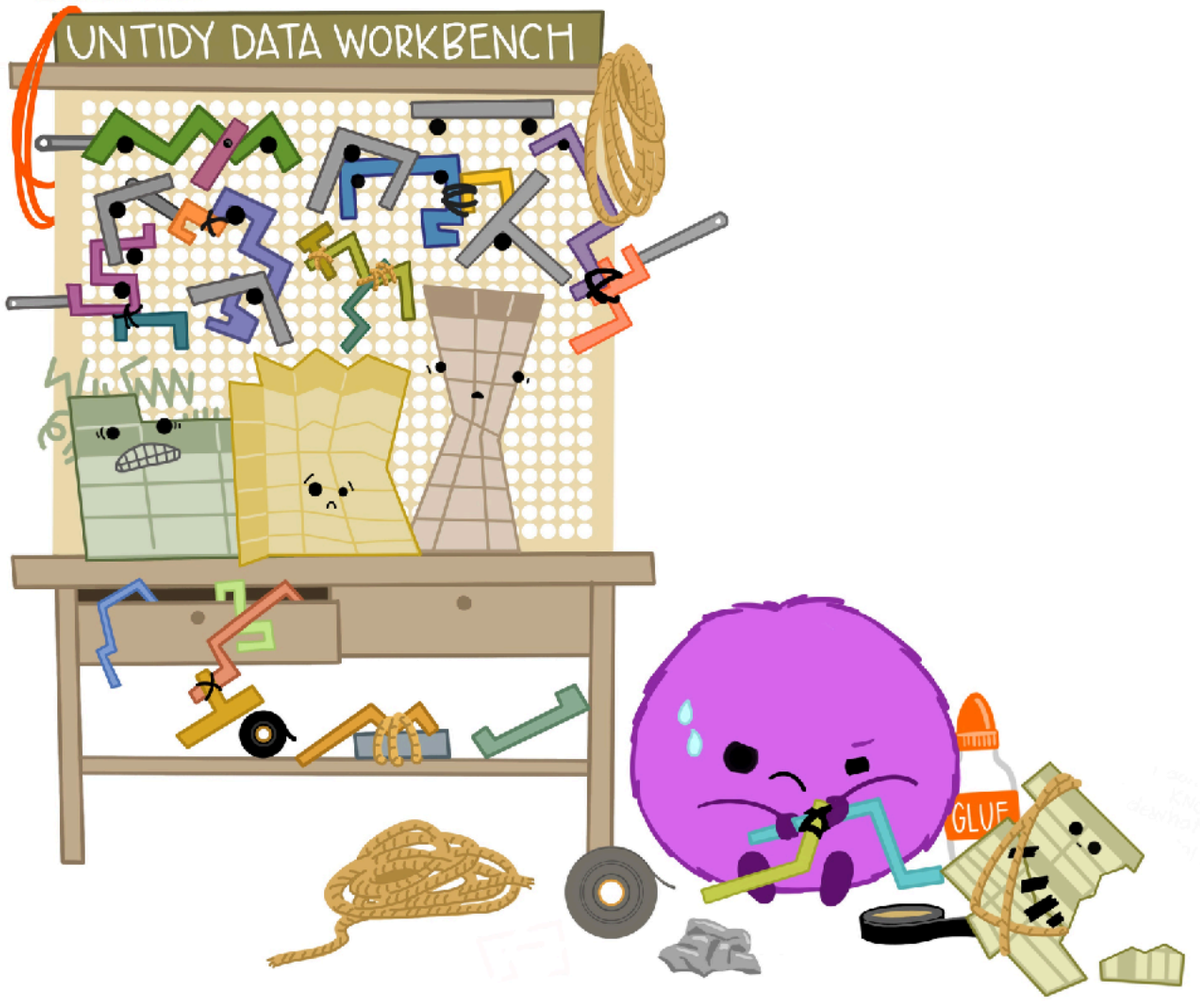


Tidy data

When working with tidy data, we can use the same tools in similar ways for different datasets...



...but working with untidy data often means reinventing the wheel with one-time approaches that are hard to iterate or reuse.



Using functions with `tidyverse` verbs

```
1 df.starwars %>%
2   select(where(fn = is.numeric))
```

my recommendation

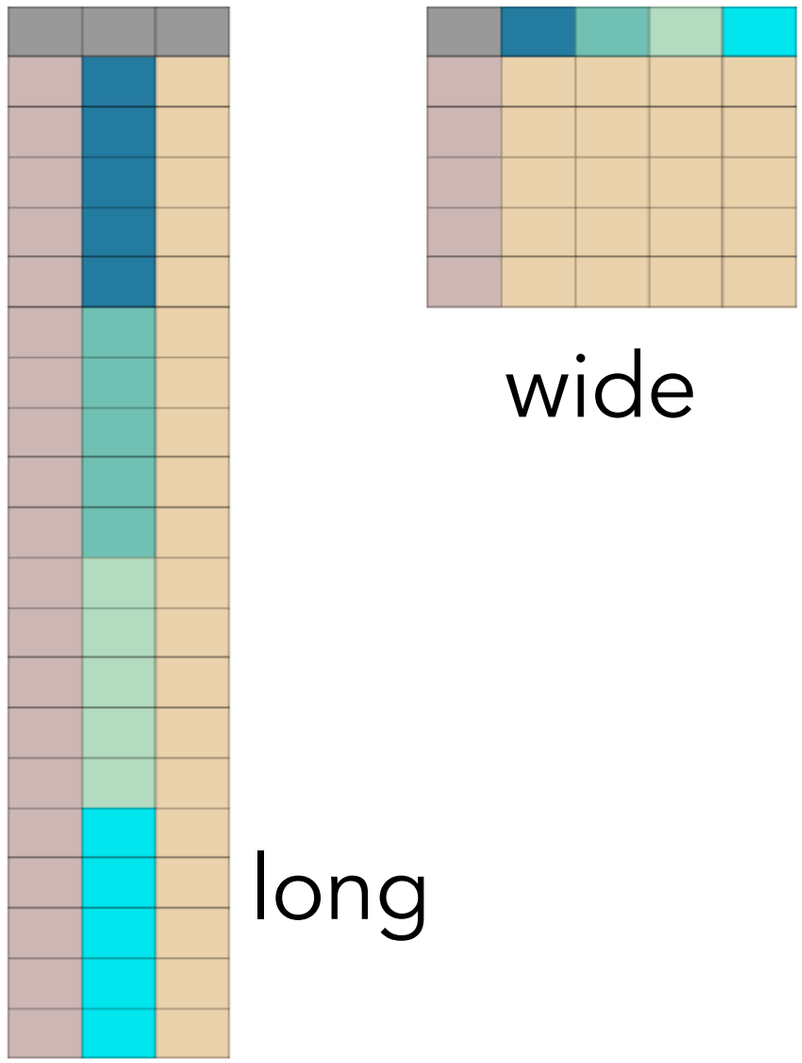
- `fn = is.numeric`
- `fn = "is.numeric"`
- `fn = function(x){is.numeric(x)}`
- `fn = ~ is.numeric(.)`

flexible, short, works well with other verbs we'll learn about later

- `fn = ~ !is.numeric(.)`

select all variables that are not numeric

pivot_longer()

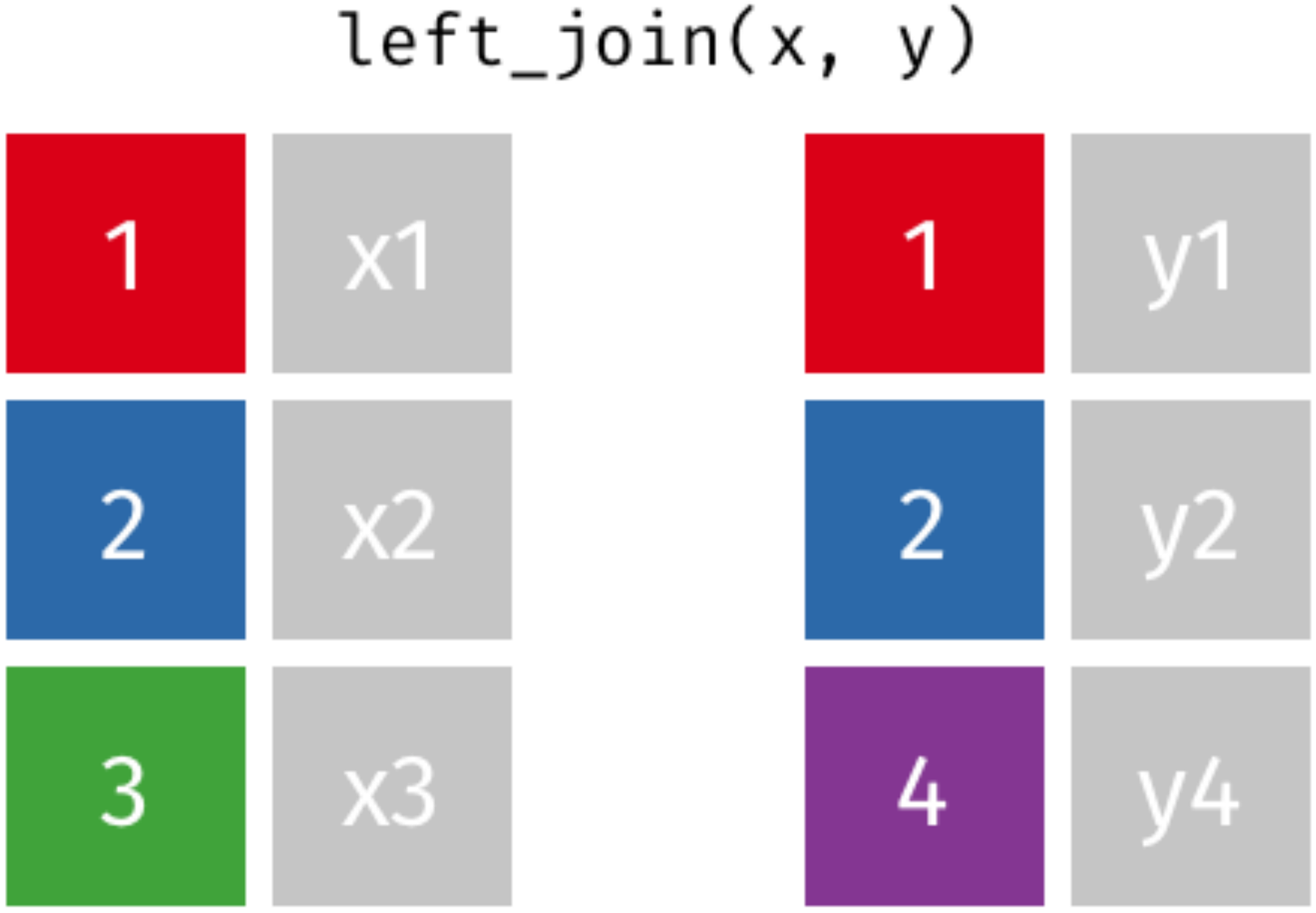


long

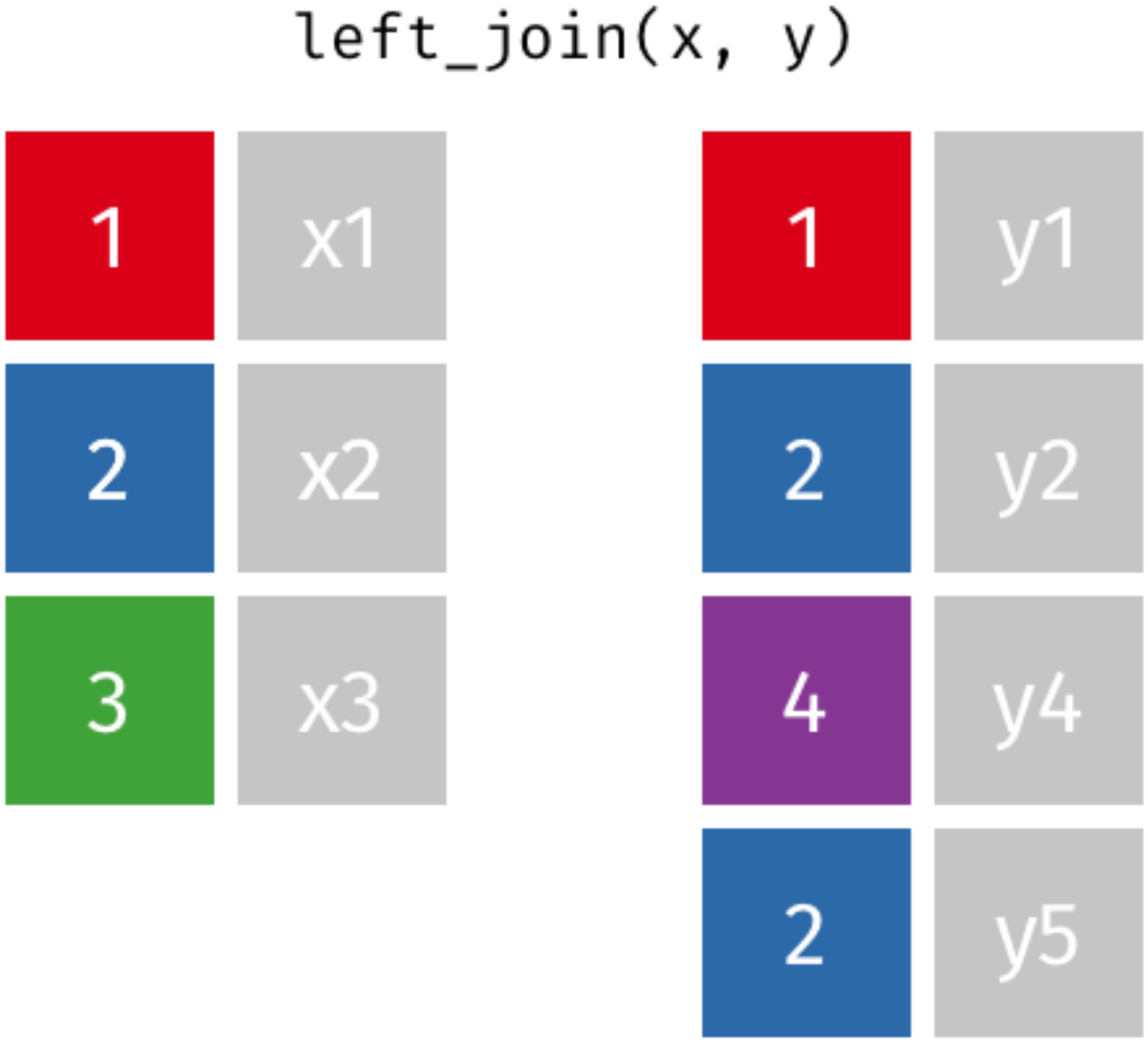
wide

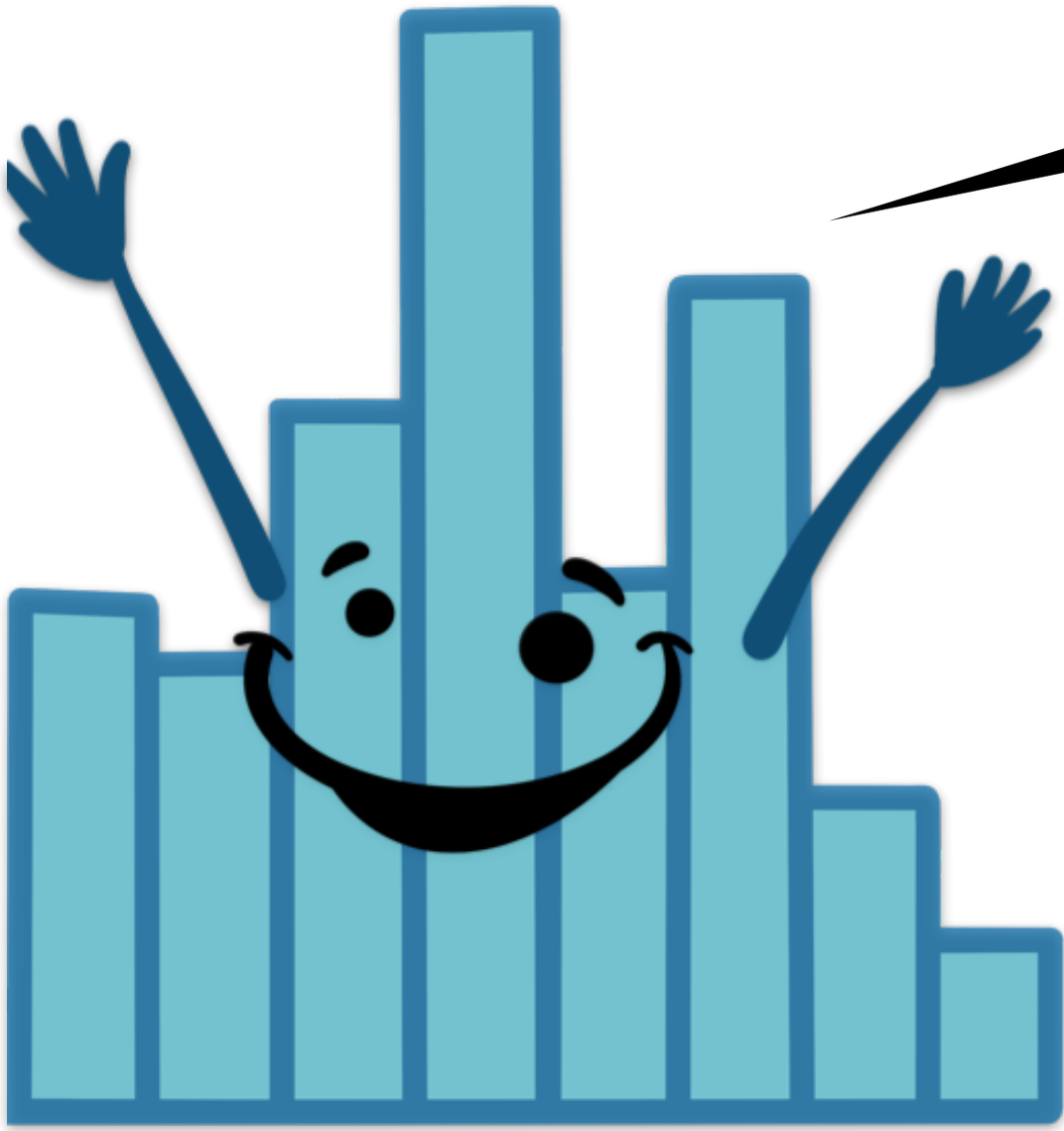
pivot_wider()

left_join()



left_join()



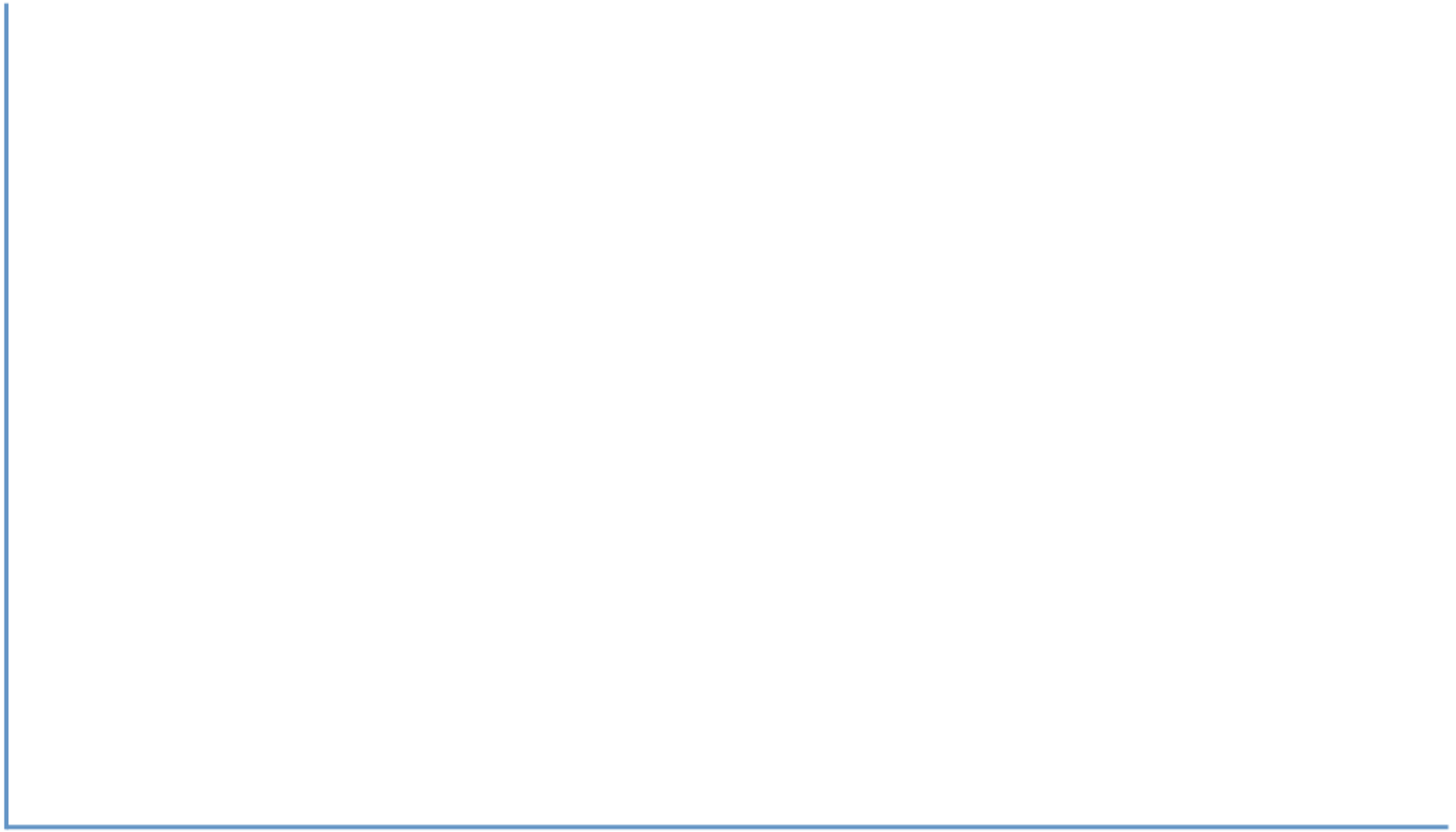


01:00

stretch break!

Feedback

How was the pace of today's class?



How happy were you with today's class overall?

What did you like about today's class? What could be improved next time?

Thank you!