

Data wrangling 2



Chat

Name one nice thing that happened to you this week.

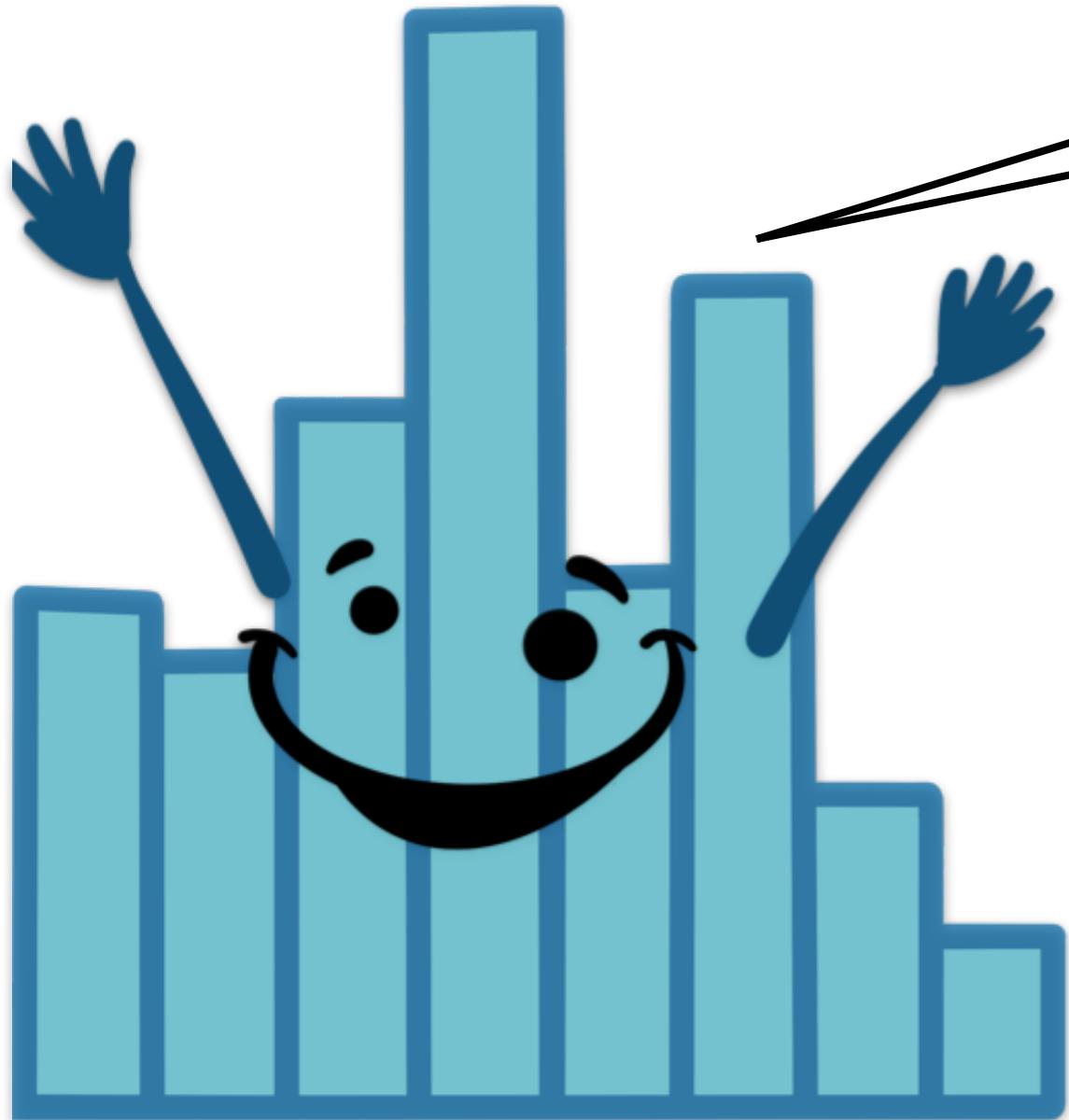
To: Everyone ▾ More ▾

Type message here...



01/22/2021

Remember to
record the
lecture!



Data wrangling 2



Chat

Name one nice thing that happened to you this week.

To: Everyone ▾ More ▾

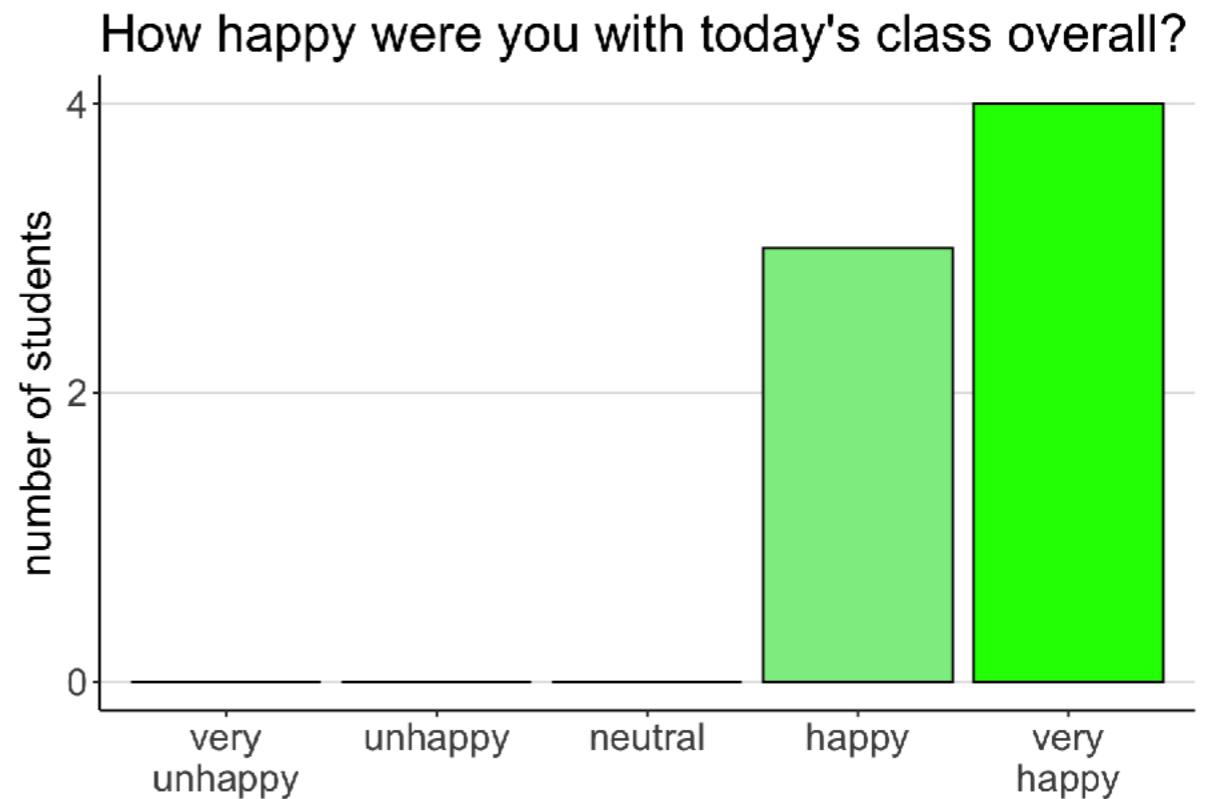
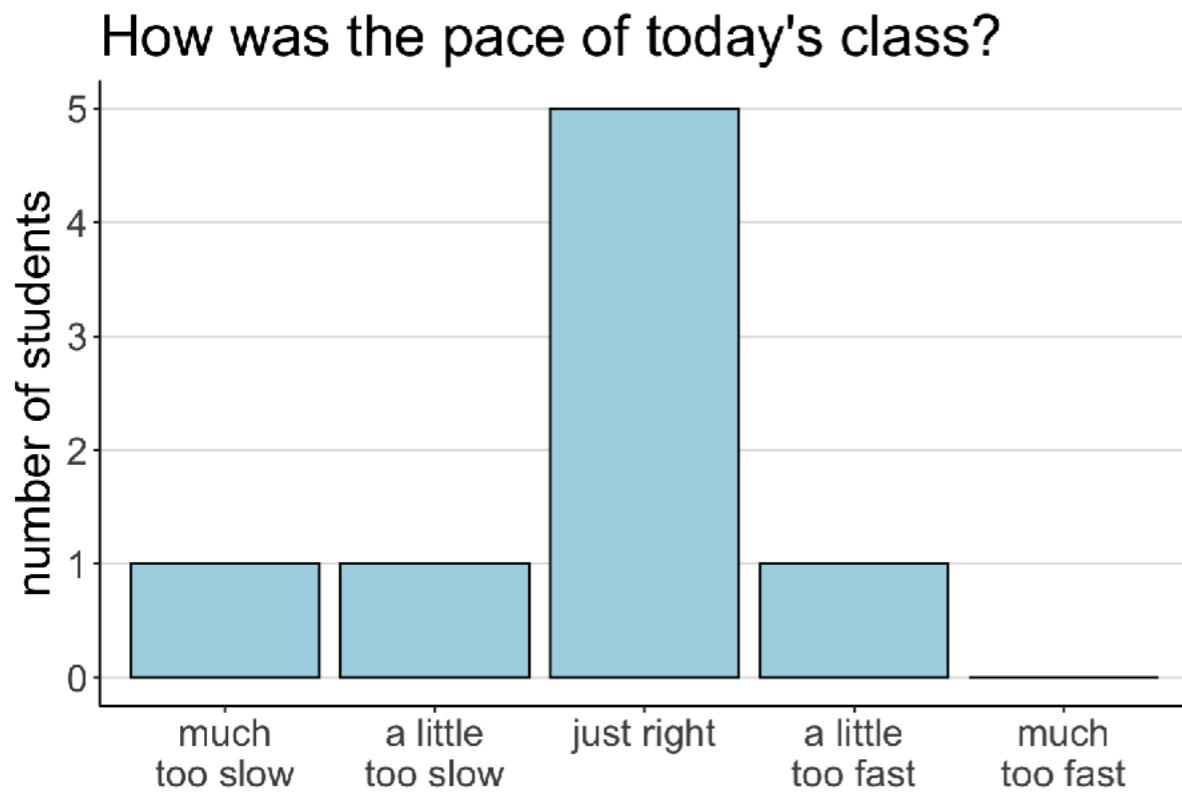
Type message here...



01/22/2021

Your feedback

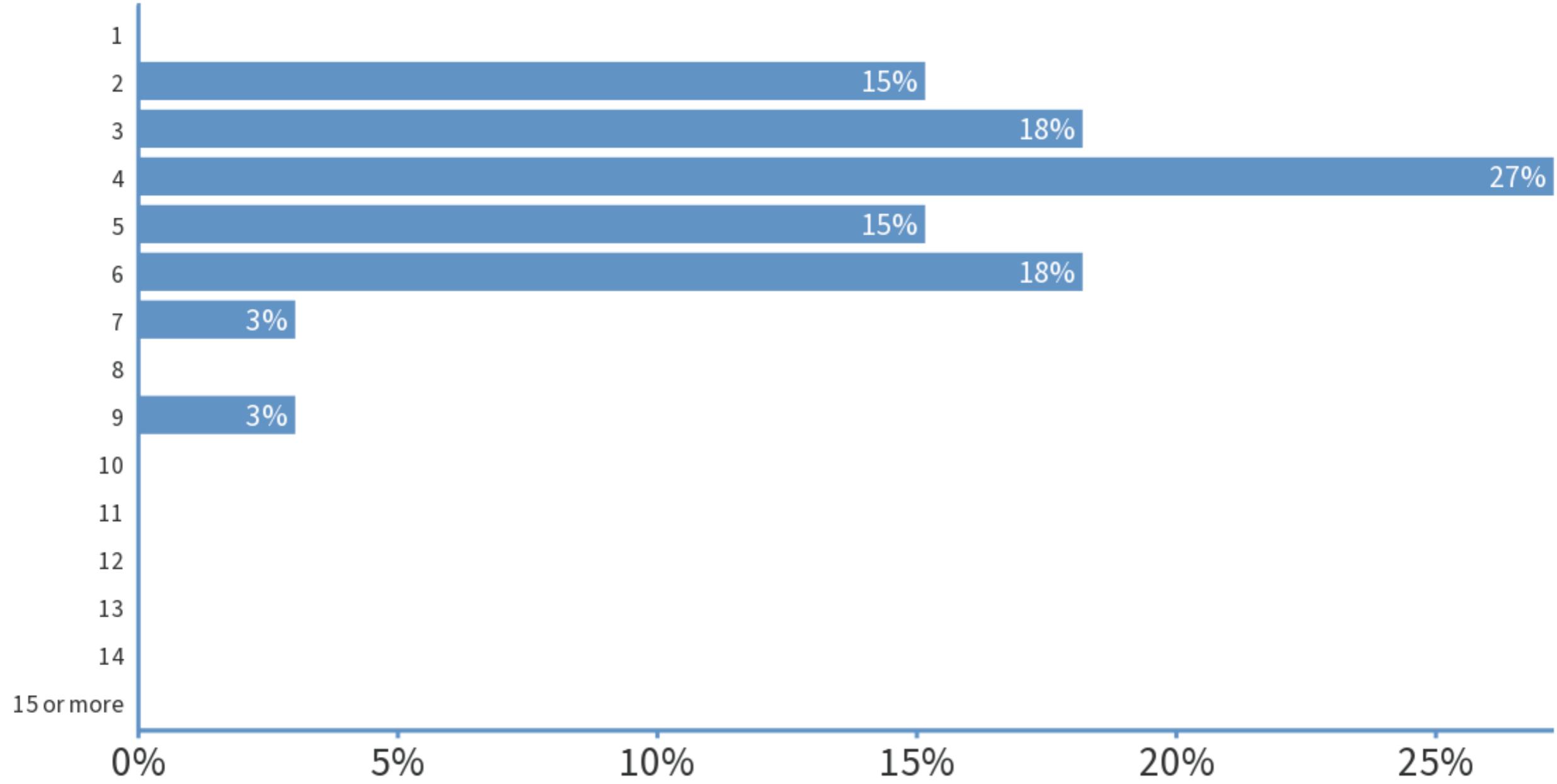
Your feedback



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

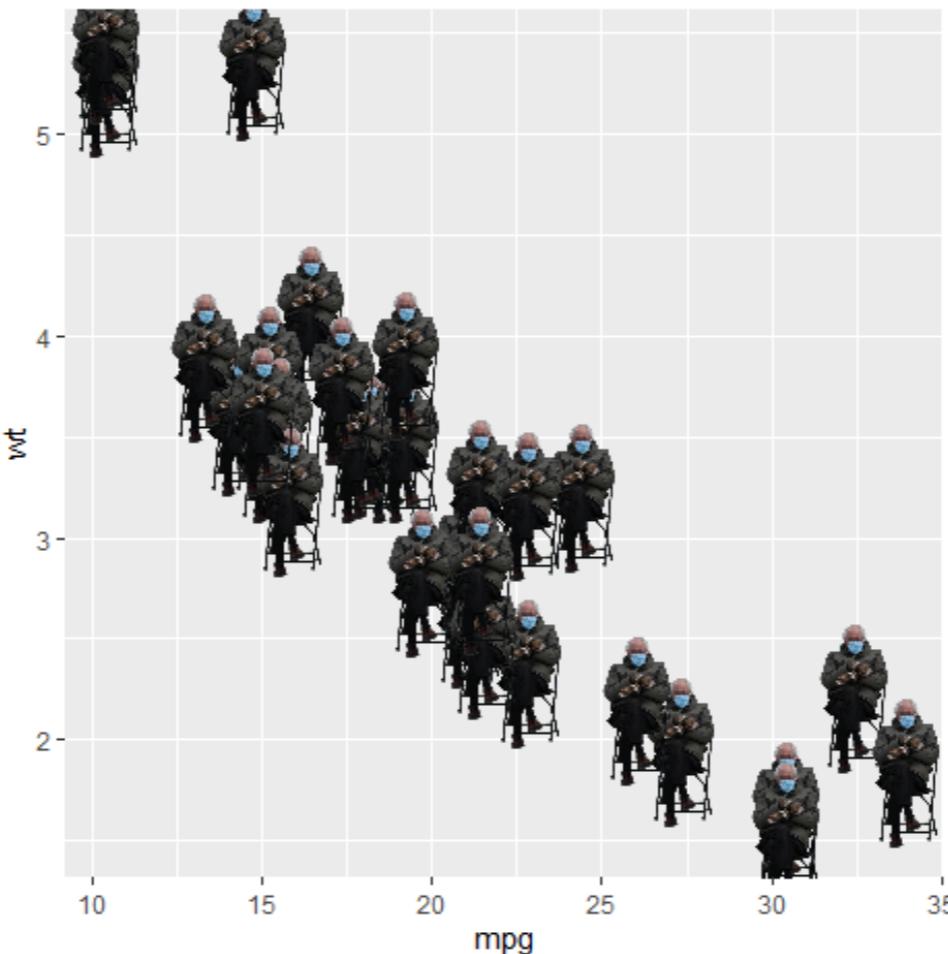
ggbernie

A ggplot2 geom for adding Bernie Sanders

This is a package inspired by a [tweet](#) by [@samuelmehr](#)

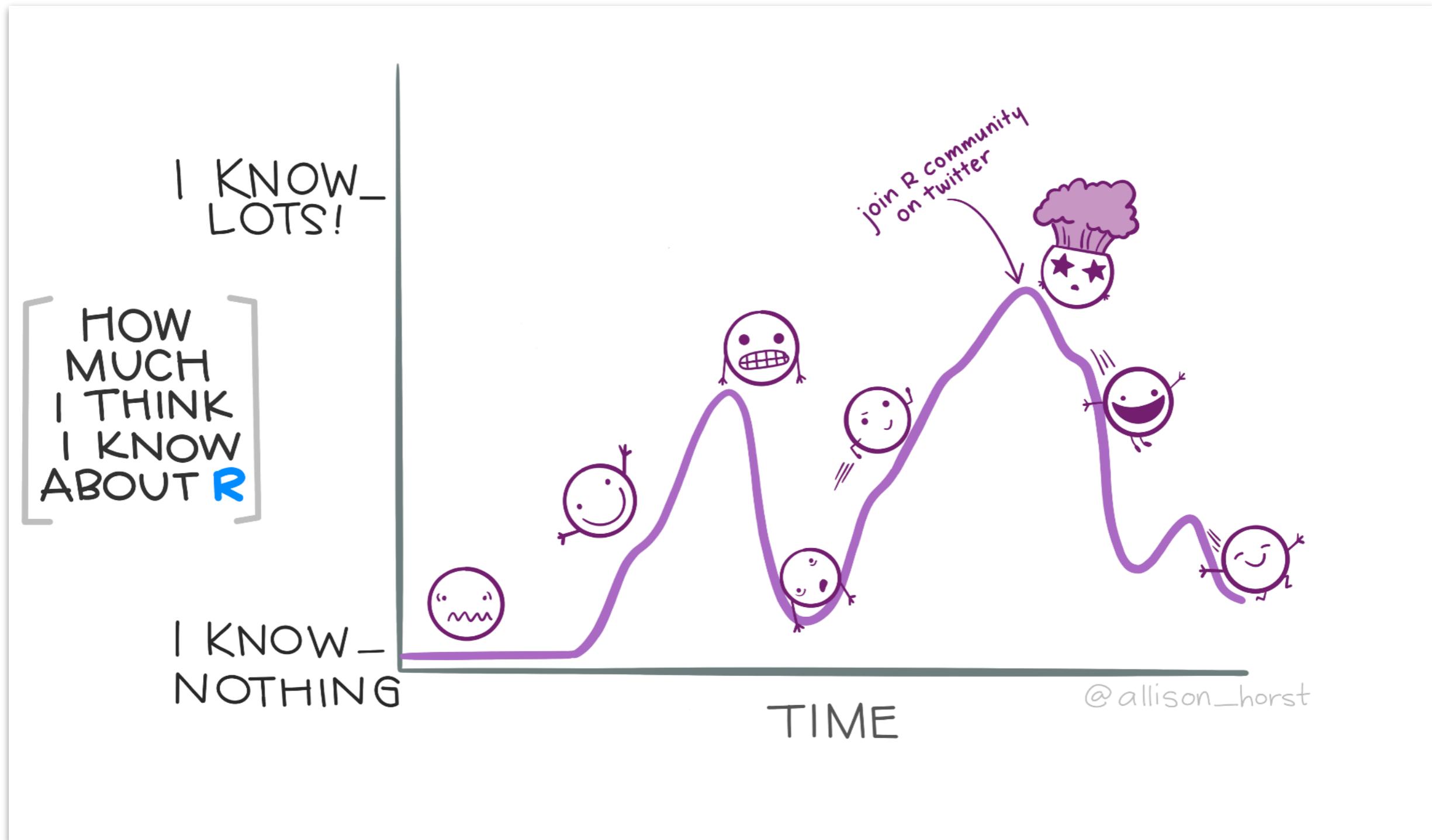
Bernie sitting

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



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

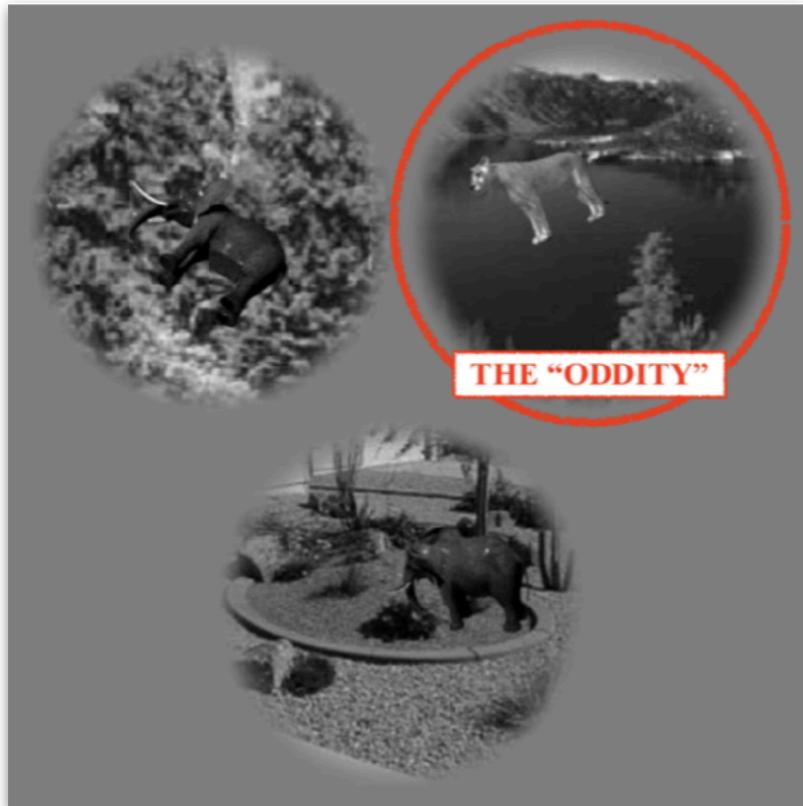
The R learning/knowledge curve



Logistics

Homework 2

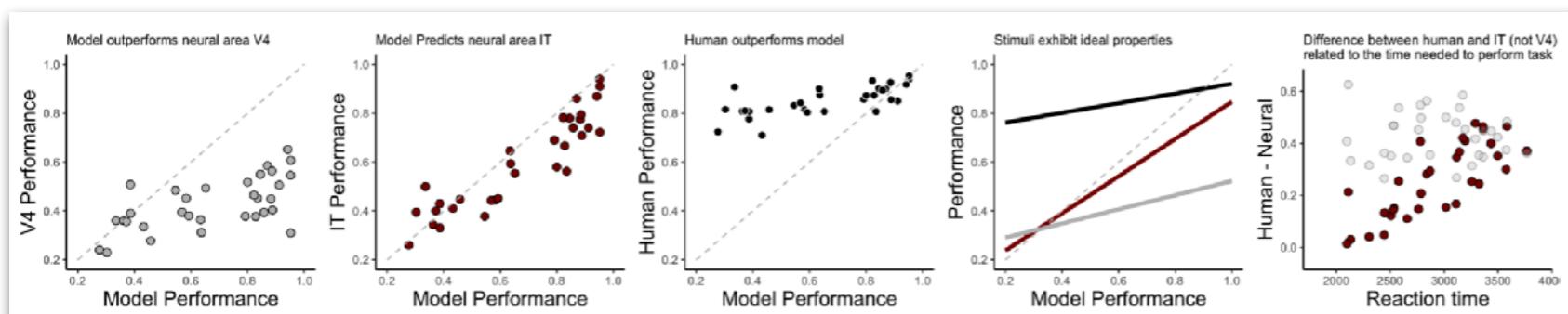
due Thursday 28th, at 8pm



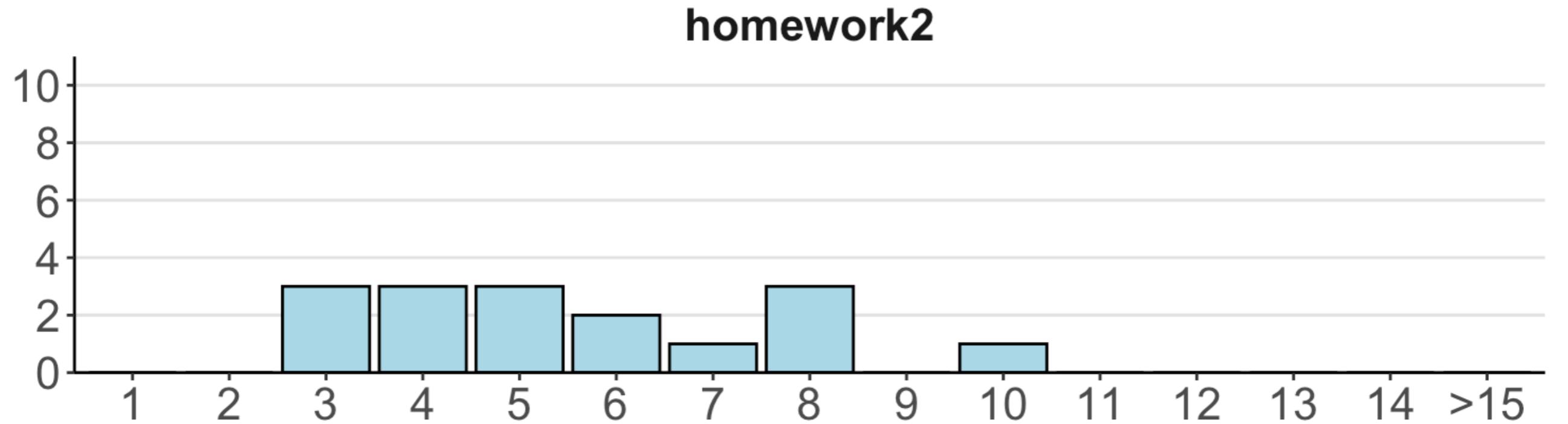
cool experiment

NAME	RELATION Relationship to subject of child to adult	SOME DATA		FURTHER DESCRIPTION		EXPLANATION
		Age	Sex	Food	Color	
Turner Lee	Fried	0	Male	10	Red	Lee
Middleton	Male	1	Male	10	Yellow	Yellow
George B. Johnson	Male	1	Male	10	Green	Green
Lee, son of John	Male	1	Male	10	Blue	Blue
Archie	Male	1	Male	10	Red	Red
Captain, Sir Archibald Macleod	Male	1	Male	10	Black	Black
Gordon, Captain E.	Fried	0	Male	10	Red	Red
Matthew	Male	1	Male	10	Yellow	Yellow
George	Male	1	Male	10	Green	Green
Hawking, Carl F.	Male	1	Male	10	Blue	Blue
Loring, B.	Male	1	Male	10	Red	Red
Tom	Male	1	Male	10	Yellow	Yellow
Frank G.	Male	1	Male	10	Green	Green
Franklin	Male	1	Male	10	Blue	Blue
Braxton, Marion	Fried	0	Female	10	Red	Red
Lucile	Male	1	Male	10	Yellow	Yellow
Merle	Male	1	Male	10	Green	Green
Royal	Male	1	Male	10	Blue	Blue
Wright	Male	1	Male	10	Red	Red
Conner, Ruth	Male	1	Male	10	Yellow	Yellow
Hannaway, William J.	Fried	0	Male	10	Green	Green
John	Male	1	Male	10	Blue	Blue
Edward, George W.	Male	1	Male	10	Red	Red
Eliza	Male	1	Male	10	Yellow	Yellow
Agnes	Male	1	Male	10	Green	Green
Hannaway, John S.	Male	1	Male	10	Blue	Blue

messy dataset



Homework 2



how long it took people last year

Data wrangling time ...

dplyr : go wrangling



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

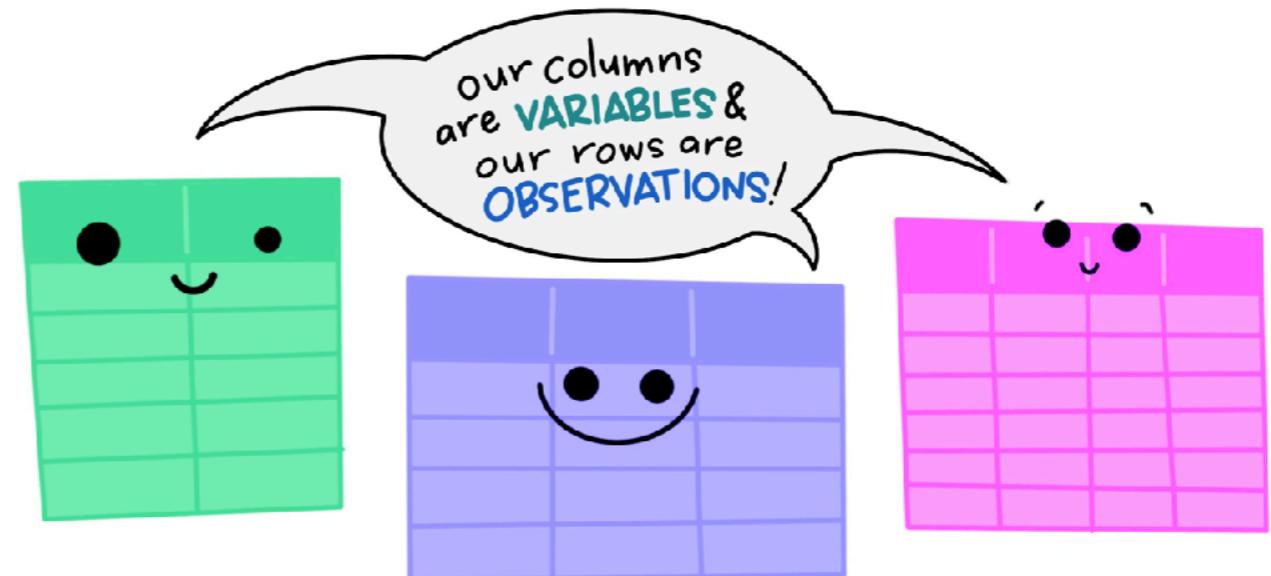
each row an observation

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

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...”

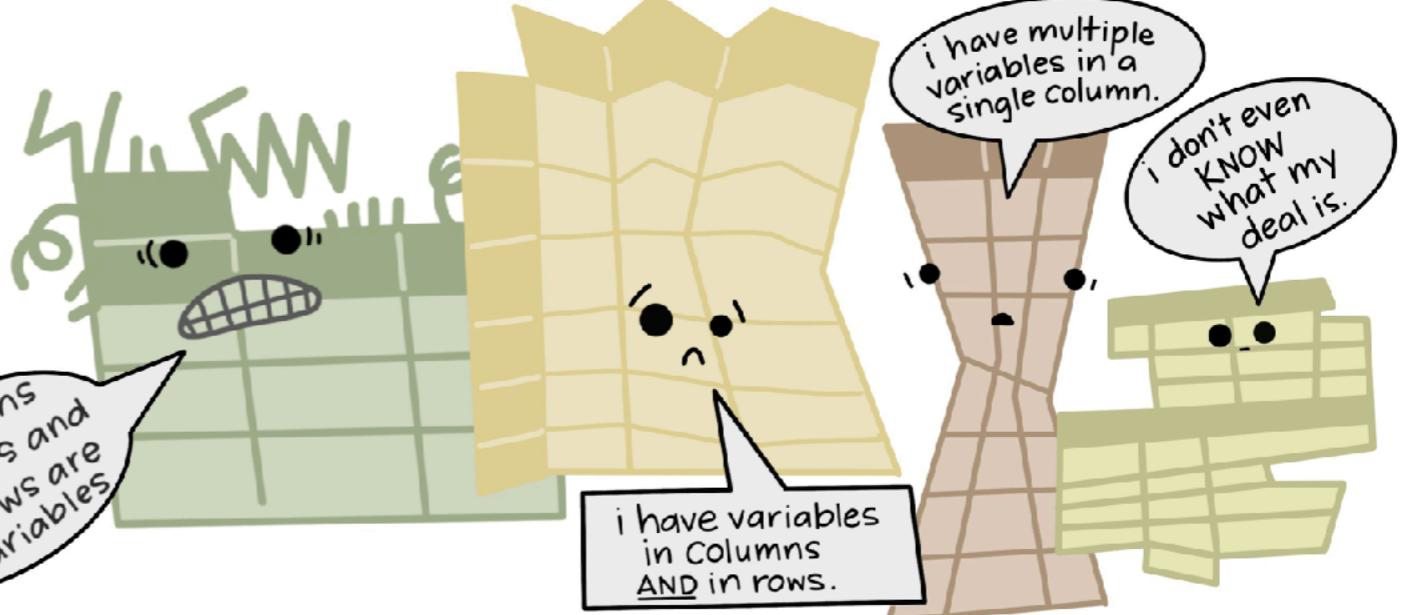


“...but every messy dataset is
messy in its own way.”

-HADLEY WICKHAM

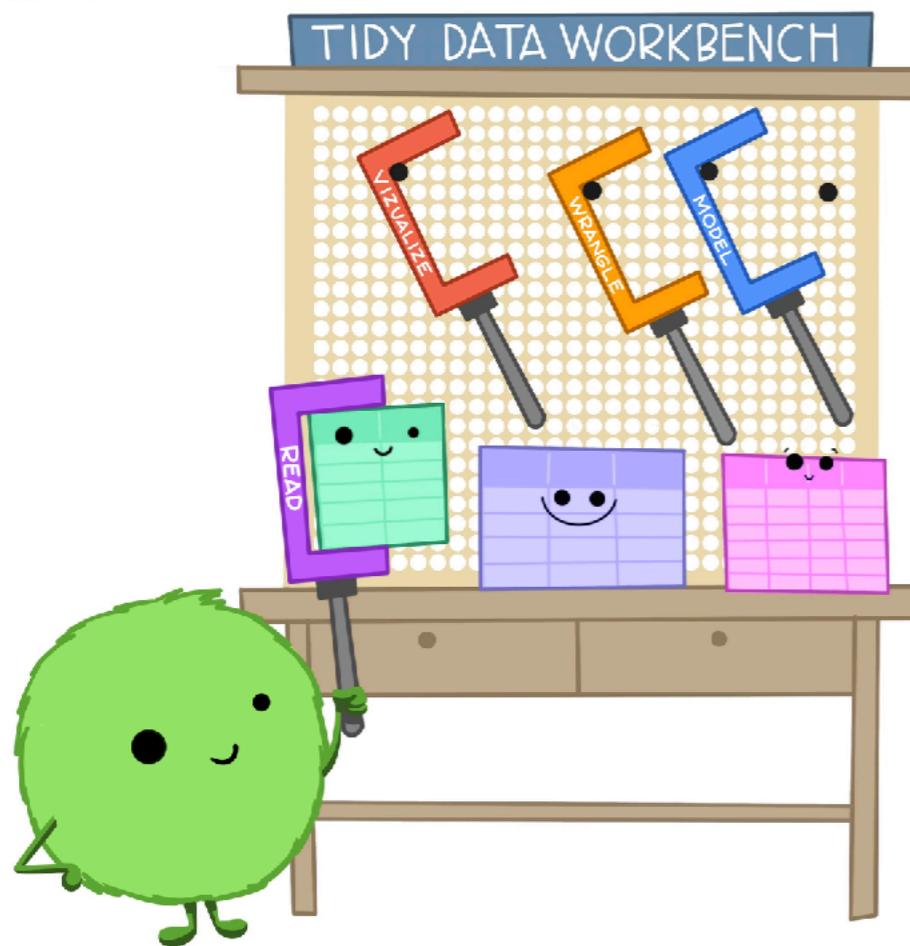


my columns
are values and
my rows are
variables

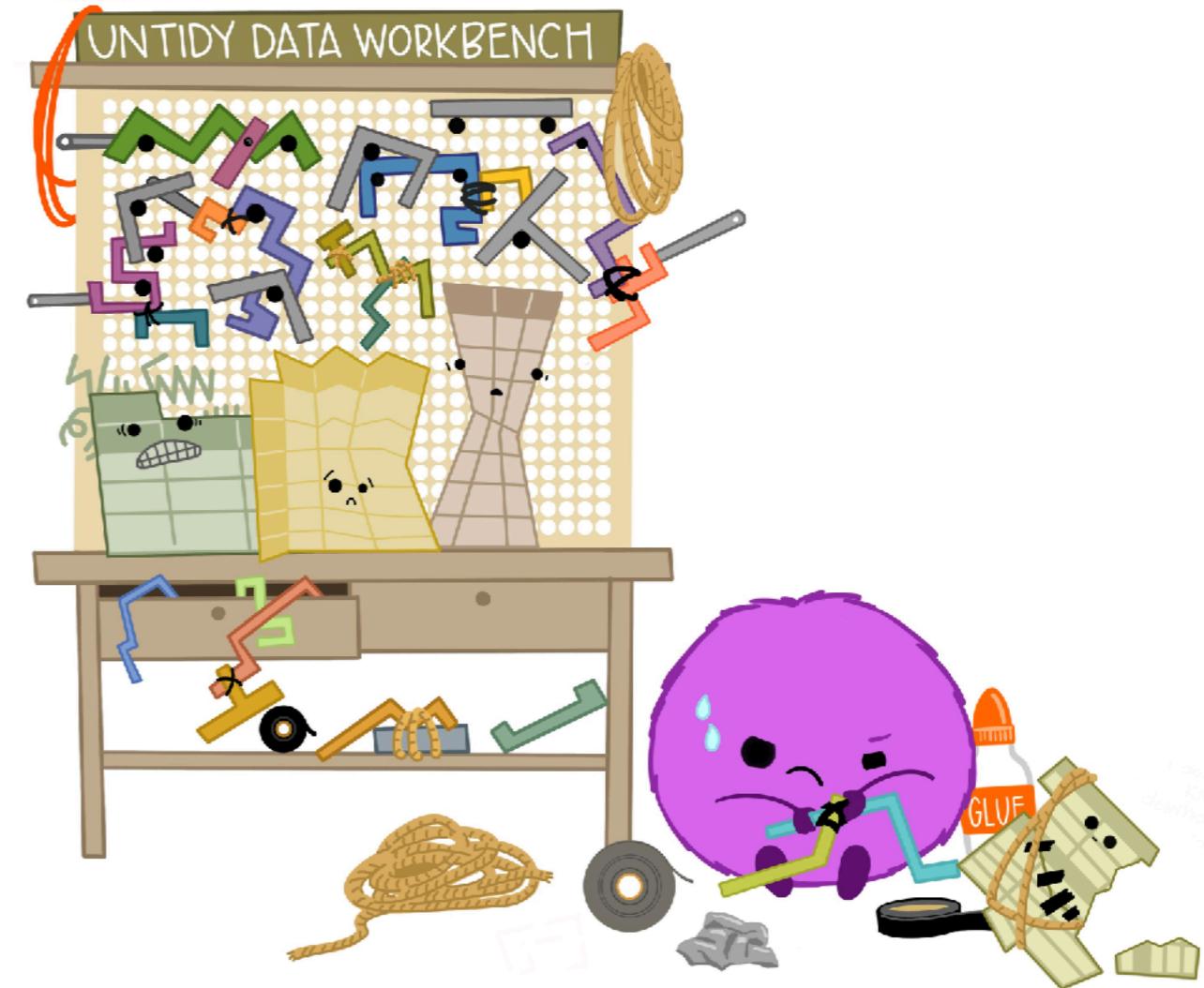


Tidy data

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(.)

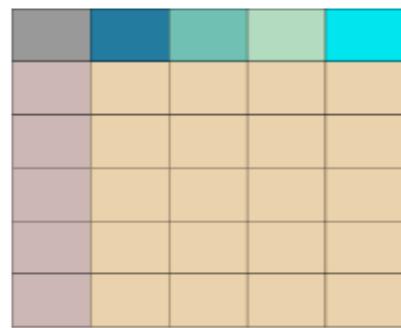
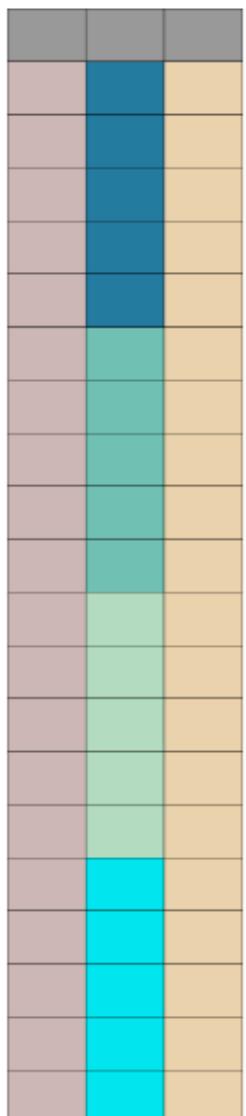
select all
variables that
are not numeric

pivot_longer()



a

b



wide

- Group 1
- Group 2
- Group 3
- Group 4
- Data
- Header
- ID

long



pivot_wider()

`left_join()`

`left_join(x, y)`

1	x1
2	x2
3	x3

1	y1
2	y2
4	y4

`left_join()`

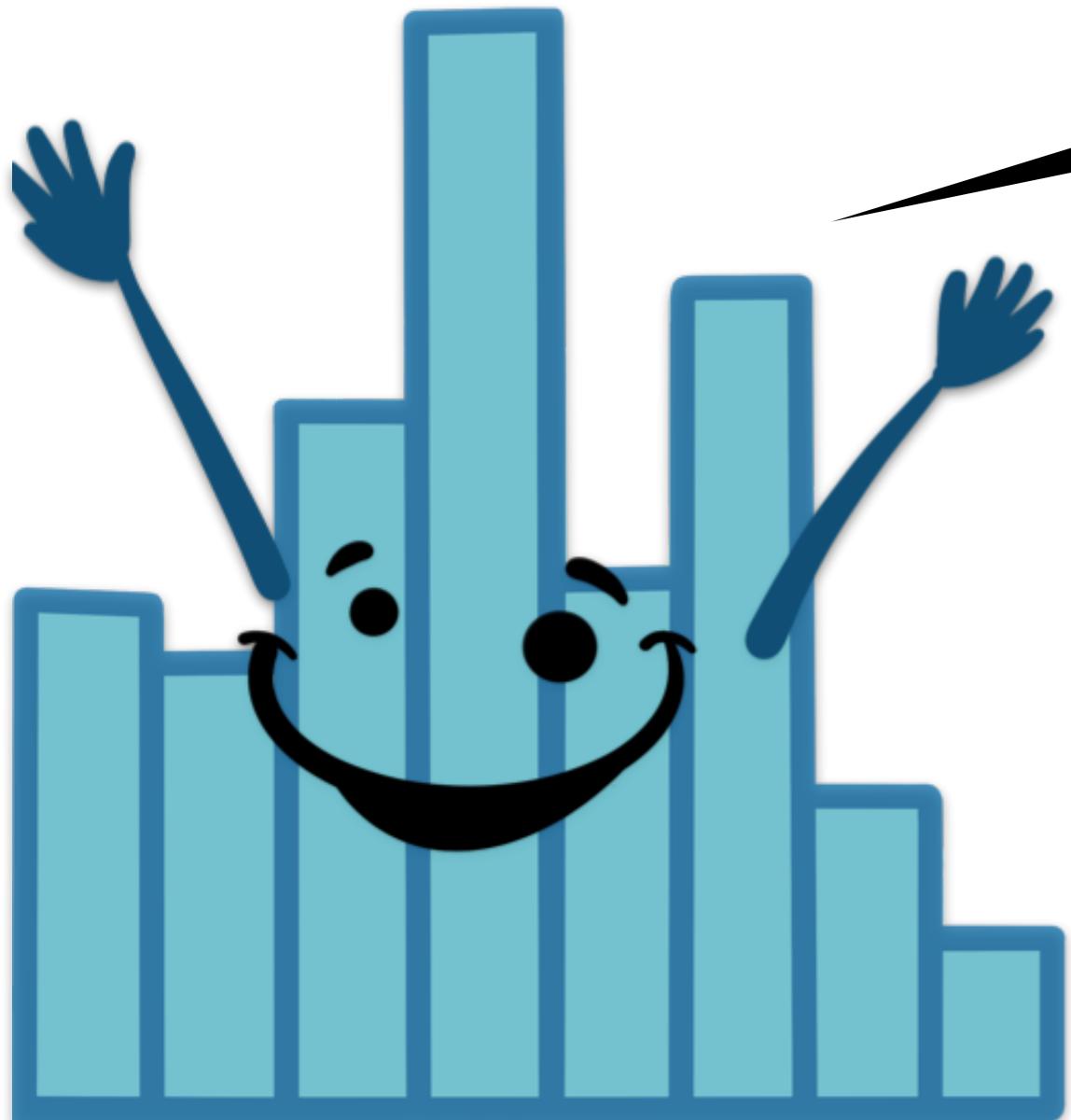
`left_join(x, y)`

1	x1
2	x2
3	x3

1	y1
2	y2
4	y4
2	y5

01:00

stretch break!



Feedback

How was the pace of today's class?

much a little just a little much
too too right too too
slow slow

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!