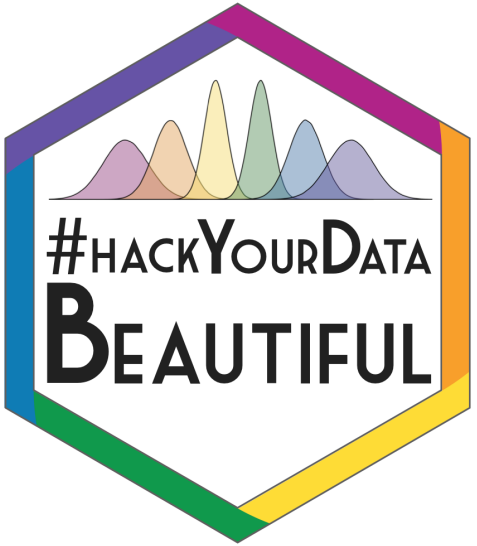


Introduction to Tidy Data

Shannon McNee

@ShannonMcNee2

Published Paper
Data Analysis

Questions we might ask ourselves

- How can I organise this data?
- How can I make this data meaningful?
- How can I make this data tidy?

And maybe even more importantly…

- Why should I even bother with tidying my data?

dplyr
ggplot2
tibble
readr
tidyr
purrr

The Wickham Six

90% of data analysis can be reduced to these six English verbs:

- select
- filter
- mutate
- arrange
- group_by
- summarise

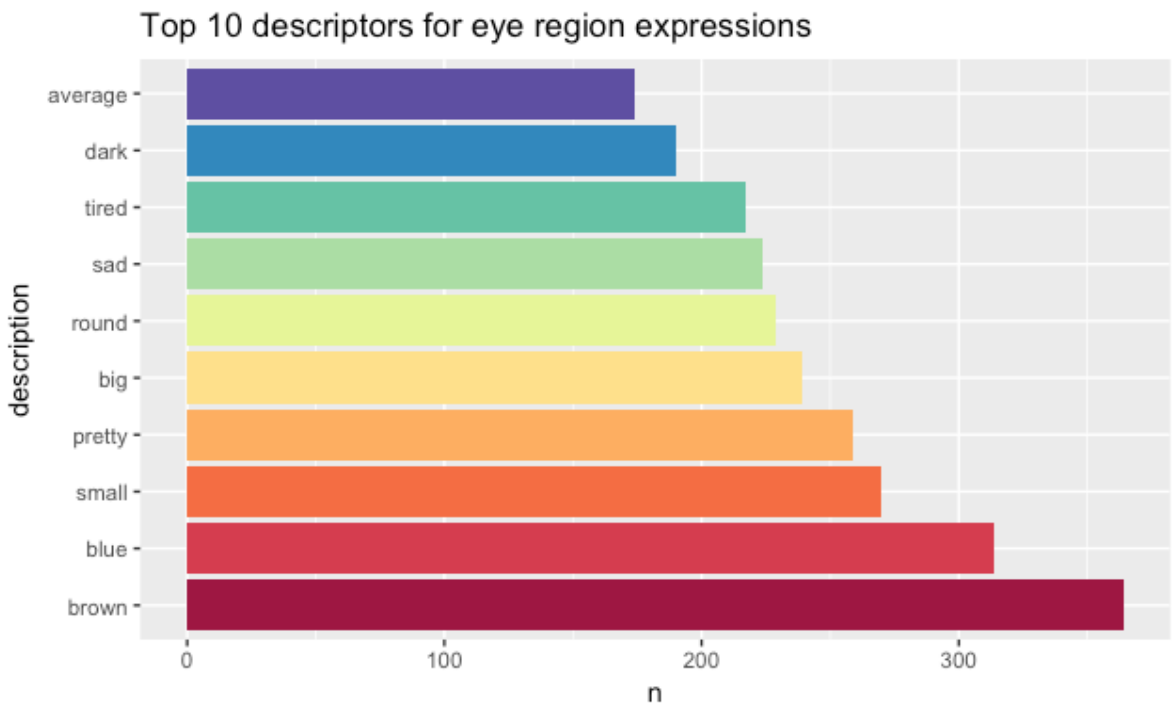
In addition:

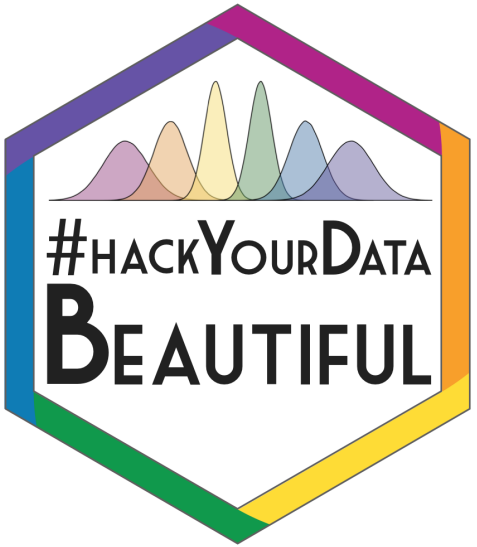
- gather
- spread
- separate

Telling a story through code

= “and then”

user_id	sex	age	t1	t2	t3	t4	t5	t6	t7
508844	male	19.0	empty	light	cheery	sad	focused	creative	lost
508966	female	20.4	bored	cafinated	nice	off kilter	bitchy	average	slightly dejected
508976	female	24.8	Dark high on drugs	Brown bug eyed	Greenish brown bangs	Dark look diff directions	Mexican eyes alret	Gray sketch	Too far brown
509196	female	14.6	soft brown	green, brown	green	hurt	Rich Brown	grey blue	soft
509286	female	16.7	brown						brown,droopy
509400	male	*NA*	wide						shy
509503	female	36.2	brown tired						hazel downturned
509665	female	20.7	far apart						nice shape
509787	female	5.0	tired						scared
509848	male	18	tired						wide
510008	female		ored						concerned eyes
510154			sy tired						normal
510		4.5	unimpressed						surprised
		18.4	droopy hooded eyes						average
	emale	14.2	wide						wide
517	male	15.4	ired						green
510715	female	18							average normal whatever
									Distant
									Average





Thank you for listening!