Intro to Data Manipulation

Week 4 DS4IR

Gina Reynolds, April 2020

> Wrangle. "From Middle English wranglen, from Low German wrangeln ("to wrangle"), frequentative form of Low German wrangen ("to struggle, make an uproar") and German rangeln ("to wrestle").
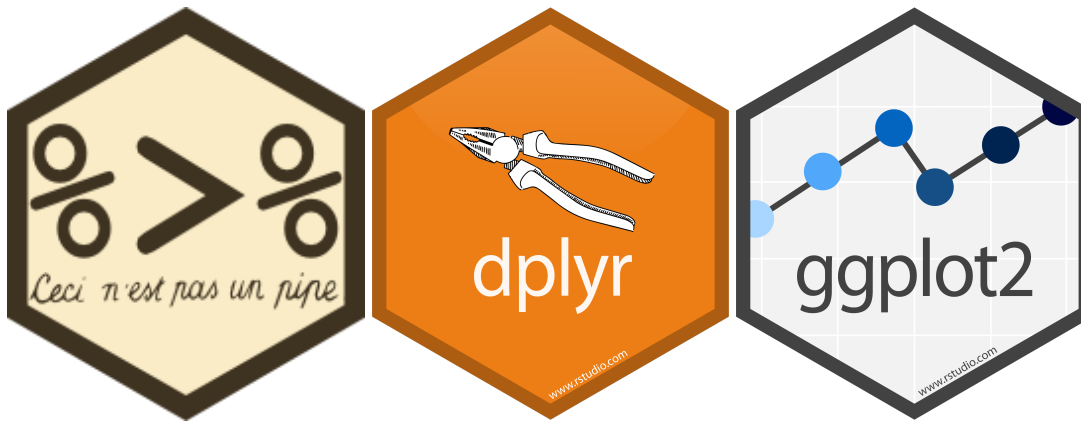

Tidy data

> Happy families are all alike; every unhappy family is unhappy in its own way. — Leo Tolstoy Tidy datasets are all alike, but every messy dataset is messy in its own way. — Hadley Wickham

---

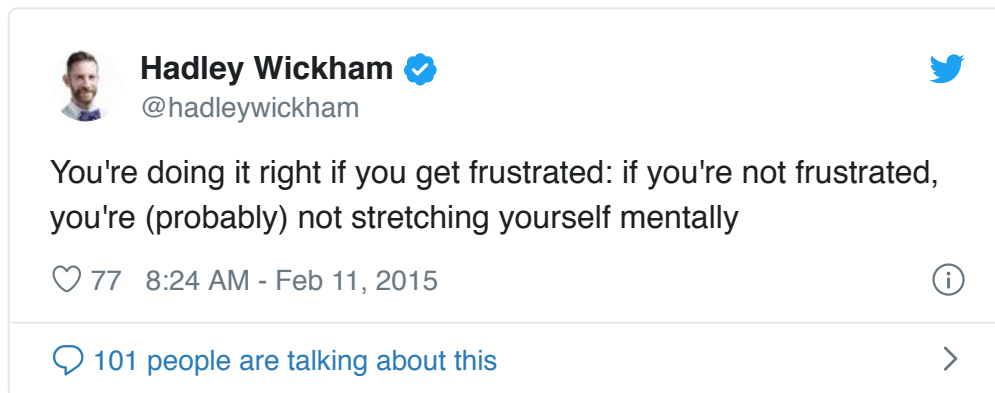- a little review - sequential workflow w/ pipe operator, booleans, dplyr

The Treachery of Images, *La Trahison des images*, painter René Magritte

dplyr (https://dplyr.tidyverse.org/)

- Combining booleans
- Summarizing
- Special topics - regex, dates, country codes
- homework/practice

- Remember - final project proposal due tomorrow…

Hadley Wickham ✔
@hadleywickham

You're doing it right if you get frustrated: if you're not frustrated, you're (probably) not stretching yourself mentally

♡ 77   8:24 AM - Feb 11, 2015

ⓘ

💬 101 people are talking about this   ›

"80% of my time is spent on cleaning the data"

–

"18% of my time is running summary statistics"