Data extraction, transformation, and loading in R

Ernest Guevarra

2022-02-07

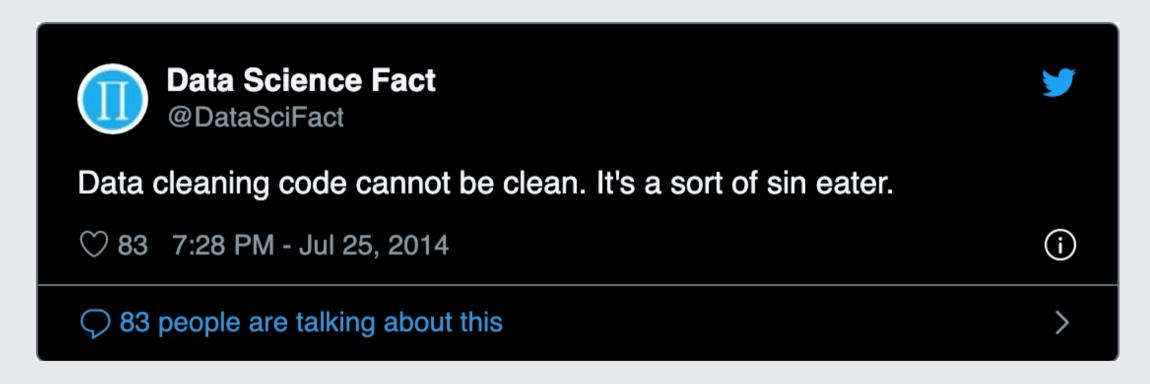
Outline

- Concepts and considerations about data
- Practical session

Concepts and considerations about data in tweets and quotes

Consideration #1:

Data cleaning/processing/checking is an important, but very challenging, step when working with data



@StatFact tweet



@AllisonReichel tweet

and I'm still pretty sure some of the data is missing, but it could still be here, in this ONE HUNDRED SHEET excel file

@RallidaeRule tweet



@GhazalGulati, via @AmeliaMN tweet

Classroom data are like teddy bears; real data are like a grizzly with salmon blood dripping out its mouth.

@JennyBryan, via @sgrifter tweet



@naupakaz, via @kwbroman tweet

Consideration #2

Big data is what everyone is talking about



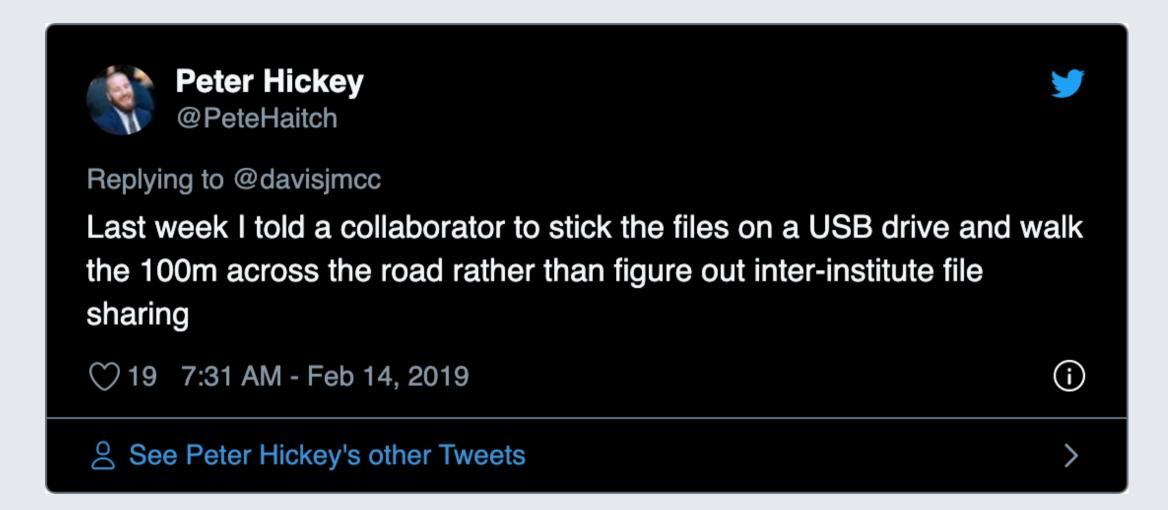
@mlipsitch, via @callin_bull tweet

Let's start the "titanic data" movement. Data too big to fail.

@neilfws with assist from @aaronquinlan, tweet

Consideration #3

Data sharing requires a well-thought out process that everyone can follow



@PeteHaitch tweet

Consideration #4 Working with data in R is great!



@tggleeson tweet

Practical session

Practical session

- We will all go through Exercise #1 in the Practical R for Epidemiologists book
- This is also one of your next assignment for this lecture series

Questions?

Thank you!

Slides can be viewed at https://OxfordIHTM.github.io/open-reproducible-science/session3.html

PDF version of slides can be downloaded at https://OxfordIHTM.github.io/open-reproducible-science/pdf/session3-working-with-data-in-r.pdf

R scripts for slides available here