De-Normalising data

Heike Hofmann, Hadley Wickham

Homework

- This week's is due, next week's is on the website

- Your job: ask and answer a question of the data

Outline

- Denormalise Data

- Practice

Less duplication

More consistent

Normalisation

Small pieces joined together

Harder to edit

Harder to view

Denormalisation

- So, now no redundancy, but unfortunately few statistical algorithms work with normalised data

- So have to **de**normalise it

- We'll use excel now, and R later

Reminder: Freeze panes

- Very useful for large spreadsheets

- Ensures you can see the id variables at all times

- Panes are frozen at top-left of current selection

- Try it out now

vlookup

- Looks up data in another "table" using a key

- Do exercises in Juice Analytics spreadsheet

Why normalise?

- Reduce redundancy to detect possible errors

- Useful way of thinking about objects under study

- Connections with hierarchical linear models

Shangri-La diet data

- Download it.

- Compare it with your results (look up who is still in the program now)

- How can we calculate weight loss and bmi change?

What do we need?

- Calculate actual date

- Convert height to single column

- Compute first weight, and date (use vlookup)

- Reference from date-level data

- Calculate bmi, and bmi change