

Summarising data

Heike Hofmann, Hadley Wickham

Outline

- Missing data
- Autofilter
- Simplifying the data: Pivot tables

Missing data

- Excel is inconsistent - sometimes zeroes, sometimes missing, sometimes error
- What happens when a blank is in a “simple” mathematical calculation (one cell involved)?
- What happens when a blank is in a formula (several cells involved)?

Get the data

- Use yours from last time
- Or download shangri-la-completed.xls from the website

Autofilter

- Easily subset and sort a spreadsheet
- Easy to use and powerful
- Data | Filter | Autofilter

Select

- All women
- All people with unknown age
- 6' women
- Youngest 5 people
- Sort by name

Shangri-La diet summary

- Reduce weight by date measurements to first and last
- Summarise weight loss for each person by
 - weight change
 - percent weight change
 - BMI change

Questions about the data

Questions about the data

- What people took part in it?
- Are males and females different?
- What effect does starting weight have?
Target weight?
- Do we draw different conclusions by looking
at bmi, or weight loss, or % weight loss?
- Are there seasonal differences?

Pivot tables

- Dynamically aggregate and summarise the data
- Drag and drop variables
 - Measured variables in the middle, id variables in rows and columns (and pages)
- Data | Pivot table report

**Create a pivot table
and try and answer
your questions**

Grouping

- Useful to simplify variables with many values
- This (in excel) is called grouping
- Group ages in decades, height into something reasonable, dates into months

Field settings

- Change summary type (average, sum, standard deviation)
- Formatting
- Special calculations

**Work on your
homework**