

Collecting data

Heike Hofmann, Hadley Wickham

Homework

- Now up on the website
- Two parts
- Due next Thursday
- Office hours MWF 1:30-2:30 and TTR 11-12 in 113 Pearson, or send email

Data voting results

Baseball	8
Crime	3
Movie rating	3
Google adsense	2
Healthy youth	2
Diamonds	1
STD	0
NASCAR	0

Objectives

- Try out a real-life data collection exercise
- Practice some of the skills you learned last time
- Discuss potential problems with collected data

The Challenge

- Seth Roberts has invented a new diet (the Shangri-La diet), and he wants to know how well it works
- How can he do it?

**What data
does he need?**

Data to collect

- Demographics
- Weight every week
- Compliance with diet ?

How could he collect it?

Hint: he has a website

Some ideas

- Get people to email him
- Comments on a blog
- Entries in a forum
- Specially design software

How is he collecting it?

www.sethroberts.net

Data is collected using forum software.
Each dieter has their own thread and each week updates an inline table with new data.
They then post a new message to that thread to flag that it has been updated.

**Are there problems
with this data?**

Is it still useful?

Some problems

- Biggest problem: self-reported
 - Might drop out if not going well
 - Might lie!
- No consistent format (easy to make transcription errors)
- How could you resolve these problems?

Your turn

Collect data for three or four people
Put it into a spreadsheet and format it nicely

How did it go?

How could you
make it easier?

**How can you avoid
repetition?**

Further challenges

- How reliable is your transcription? How could you figure that out? (ie. design an experiment)
- How could you automate the task? Start with a step-by-step description of exactly what you did
- If you could design a form specifically for this task, what would it look like?