

# Welcome to Stat 480

Heike Hofmann, Hadley Wickham

# Outline

- Introduction
- Objectives, software, modules
- Assessment
- What do you already know?
- What are you interested in?
- Getting started with excel

**had.co.nz/stat480**

heike@iastate.edu

stat480@had.co.nz

# Objectives

- Learn how to deal with complex, messy, real data
- Gain familiarity with basic data collection, storage and manipulation
- Use graphics to explore and understand data
- Fluently reshape data into the most convenient form for analysis or reporting
- Automate cleaning and analysis in Excel, R and SAS.

# Software

- Excel
  - R
  - SAS
- 
- All on lab computers. You probably already have excel. R is free. SAS is expensive

# Modules

- Data collection and organization with Excel
- Automating analyses with R
- Graphics
- Structuring and restructuring data
- SAS, the industry standard

# Assessment

- A short homework every week + 3 bigger projects
- No mid term, no final.
- 40% homework, 60% projects
- I will publish the best answers on the class website anonymously, if that's not ok, please note that on your assignment

# Homework

- One every week
- 2-3 hours
- Revise what we covered, plus synthesise some new information

# Projects

- 3 bigger projects
- Will be open ended, and will involve a substantial write up (10+ pages)
- Work in groups (same for all projects)
- I still haven't decided what they will be ...

# Disability and sickness

- Make sure to let me know (in advance)
- Keep on top of the weekly homework or you will get swamped!

# Lectures

- Electronic copy of the slides will be available on the website
- But you'll need to take your own notes!
  - If you really want complete notes, organise a roster with others in the class
- Don't goof off on the computers!
- If you're bored, complain!

# What do you know already?

- Excel? Pivot tables? Macros? Autofill? Autofilter?
- A programming language? CS course? R? SAS?
- Database theory? Do you know what third normal form is?
- Have you used a text editor?

- Baseball salaries and performance
- Crime data (incl. type, time, place, demographics etc.)
- Google adsense earnings
- Healthy youth survey
- Movie rating data from the IMDB
- Election Data
- Pricing diamonds
  
- Any thing else you can think of?