

Welcome to Stat 579

Heike Hofmann

About me ...

Professor of Statistics

my research area is
exploratory data analysis
and statistical computing,
heavily involved in CSAFE



I have been teaching Stat 579 and other courses in
statistical computing since 2006

About me ...

Professor of Statistics

my research area is
exploratory data analysis
and statistical computing,
heavily involved in CSAFE



I have been teaching Stat 579 and other courses in
statistical computing since 2006

easiest to reach at
hofmann@iastate.edu

... what about you?

Before we get started ...

- course has two sections
 - 1 (Thursday, Statistics graduate students)
 - 2 (Tuesday, graduate students from outside)
- time: 12:10 - 2:00 - we will meet on Zoom
- You are asked to go over the slides and do suggested your turns

Outline

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

“Tall-order” Objective

- become computationally proficient to do statistical data analysis

Objectives

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

Additional Readings

- “Data Technologies” by Paul Murrell, available as pdf under Creative Commons License at <http://www.stat.auckland.ac.nz/~paul/ItDT/>
- “R For Data Science” by Hadley Wickham and Garret Grolemund at <http://r4ds.had.co.nz/>
- Recommended: “ggplot2” by Hadley Wickham (email me for the pdf or use Springer link for

Setup of Slides

Learn about **Concepts** of Statistical Computing
Practice **Skills**

about 2:1 split of time in reading vs practice,
set an alarm to see whether you can solve your
turns in that amount of time

Software: R

Assessments

- One homework assignment every week + quizzes to test understanding of basics
- One midterm exam
- final (team) project & presentation
- 20% homework, 15% quizzes, 35% midterm, 30% final project
- extra credit is listed as 0% ... so it is extra

Homework

- One every week
- 2-3 hours
- Revise what we covered, plus synthesize some new information
- I will publish the best answers on the class website anonymously, if that's not ok, please note that on your assignment

Midterm

- we need to figure out whether we can do this proctored during our regular meeting time ...
- Open note, open book, open internet ... but no collaboration with anybody
- tentatively scheduled for week of Oct 26

Final Project

- A bigger project (multiple due dates)
- Will be open ended, and will involve a substantial write up (~10 pages)
- Work in a team (3-4 members)
- Talk to each other! Find a group!

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, organize a roster with others in the class
- Don't goof off on the computers!
- If you're bored, complain!

What do you know already?

- Have you used R before?
- ... for more than a year?
- the tidyverse?
- RMarkdown?
- git?

What is this class about?

- Very Data Centric
- I want to know about topics you're interested in
- Go to Wufoo Site to let me know your favorites and make suggestions:
<https://heike.wufoo.com/forms/topics-of-interest/>

- Sports
 - e.g. Baseball salaries and performance
- Crime data (incl. type, time, place, demographics etc.)
- Health
 - e.g. fitness statistics, or disease rates, or health care
- Movies
 - e.g. ratings/box office revenues from IMDB
- Climate/Weather Data
- Travel data, e.g. US flights
- National or Global Issues (world poverty, carbon footprint, ...)
- Anything else you can think of?