Intro survey (google form)
Mrs. Mutner liked to go over a few of her rules on the first day of school.
About Stat 133
Stats Major

Prereqs
- Calculus
- Calculus II
- Multivariable Calculus
- Linear Algebra

Core
- Stat 133 Computing
- Stat 134 Probability
- Stat 135 Statistics

Elective
- Stat 150 Stochastic Processes
- Stat 151A Linear Modeling
- Stat 152 Sampling Surveys
- Stat 153 Times Series
- Stat 154 Predictive Modeling
- Stat 155 Game Theory
- Stat 158 Design of Experiments
- Stat 159 Reproducible Research
Roles for Stat 133

- Statistics
- Applied Math
- Data Science
- Other Majors

Core Course
Option for Data100
Data Science Cluster
misc
My Philosophy
DATA: BY THE NUMBERS

NUMBER OF YEARS TO GET DATA: 3
YES! FINALLY!

NUMBER OF YEARS TO INTERPRET DATA: 2
what does it all mean??

NUMBER OF YEARS TO WRITE ABOUT DATA: 1.5
blah blah blah blah...

NUMBER OF SLIDES TO PRESENT DATA: 1
RESULTS that’s it?

Data Preparation

- Acquisition
- Storage
- Cleaning
- Processing
- Tidying
- Reshaping
- Wrangling
Analysis

- Exploration
- Description
- Visualization
- Hypothesis Tests
- Inference
- Simulation
- Model Fitting
Reports

- Document(s)
- Article(s)
- Book(s)
- Poster(s)
- Blog post(s)
- Dissertation
- News
Communication

- Oral
- Print
- Web
- Audio
- Video
- Multimedia
- Other
Traditionally, this is where most teaching focuses on...
However ...

(Almost) no one teaches this!

In practice these are where we spend most of our time
Course Content
Course cornerstones

- Data Manipulation
- Data Visualization
- Reporting Tools
- Programming Concepts
- Data Technologies
- R & other tools
Data Tables

1. Data Tables
2. Selecting and Filtering
3. Reshaping
4. Aggregation & Group by operations
5. Joins and Merges
Taking Care of Data

1. Storing Tables (files & formats)
2. Data Dictionary (metadata)
3. Data Organization
4. Cleaning
5. Data Tidying
Data Visualization

1. Visualization basics
2. Colors
3. Design and Aesthetics considerations
4. Efficient displays
5. Good and bad practices
Programming Concepts

1. Emphasis on **data analysis**
2. Data types and data structures
3. Control flow structures
4. Functions
5. Regular Expressions
Reporting Tools

1. Markdown syntax
2. LaTeX (mostly equations)
3. Dynamic Documents
4. Shiny Apps
5. Writing reports
R and other tools

1. R
2. RStudio
3. Command Line (Bash)
4. Unix filters & utilities
Instruction

In-person instruction

Lecture: more conceptual/theory

Lab: practice
Website & bCourses

**Units:** weekly topics

- Slides, readings, cheatsheets, files
- Lab materials
- Assignments
- Submissions
Grading Structure

8% Lab work (weekly; drop 2 lowest)

35% HW (6 assignments; drop lowest)

27% Apps (3 shiny apps; no drops)

8% Midterm

22% Final exam
Enrollment

Waitlist

Concurrent-enrollment
Some Comments
Remarks

Very hands-on course
Expect to do A LOT OF WORK outside class
Deceptively simple
Very easy to fall behind
Course Format

**Lecture**: conceptual stuff, demos, case studies, examples, review some code

**Lab**: practical work using R, command line, git

**Homework**: follow the work of labs, plus some challenges
My Expectations

Don’t expect that you’ll become a data scientist (that takes years of hard work)

Instead: give you solid foundations about data analysis

Expose you to different “data technologies”
Ultimate Goals

Understand different types of data (e.g. files, forms, formats)

Know how to access information stored in different formats

Know how to do data manipulation and processing in R

Be better prepared to crunch data
Becoming a data scientist is a \textit{(yearslong)} \textbf{marathon} \ldots not a \textit{(one semester)} \textbf{sprint}
Intro survey (google form)