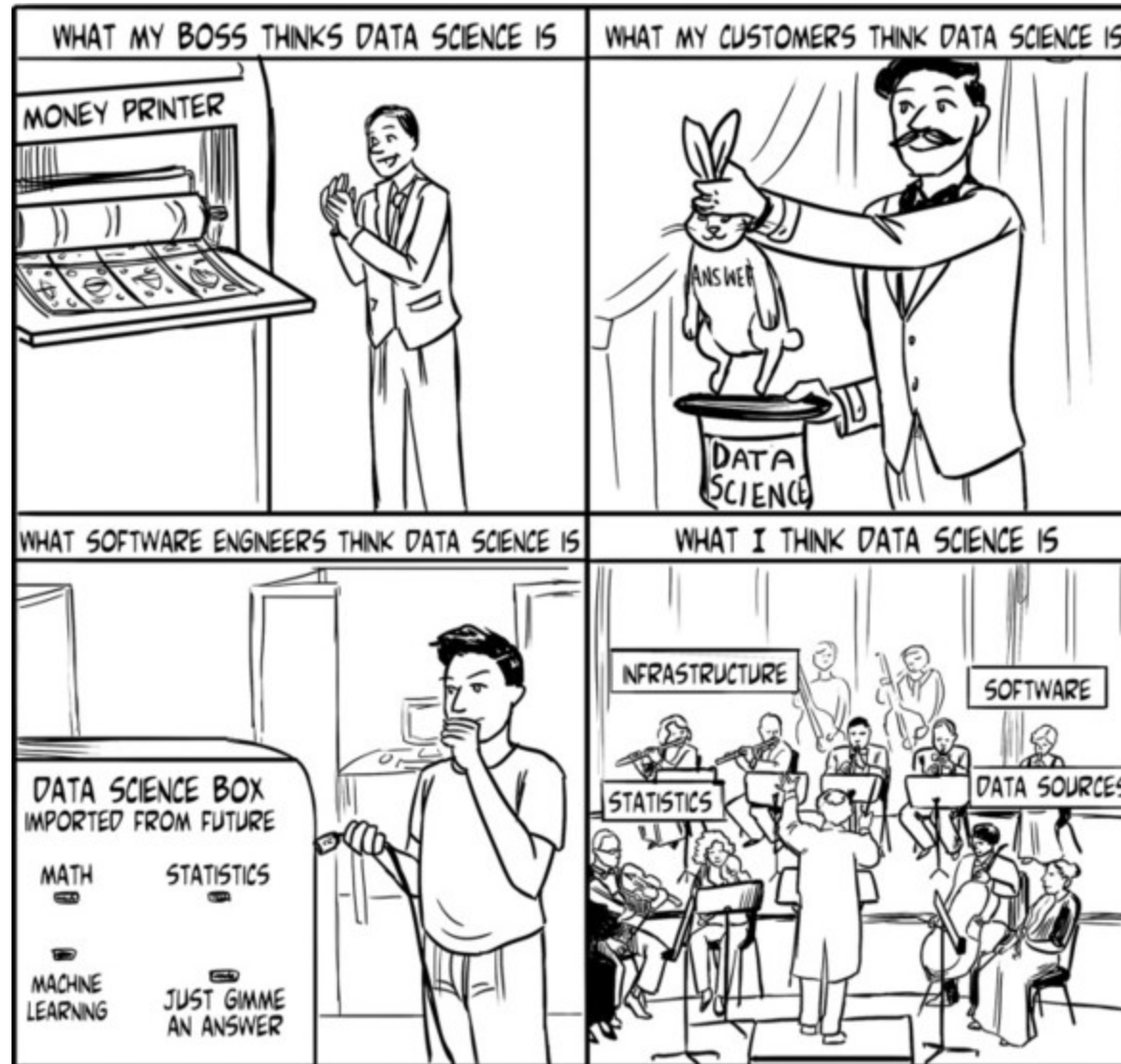


# **Data Science Capstone Project**

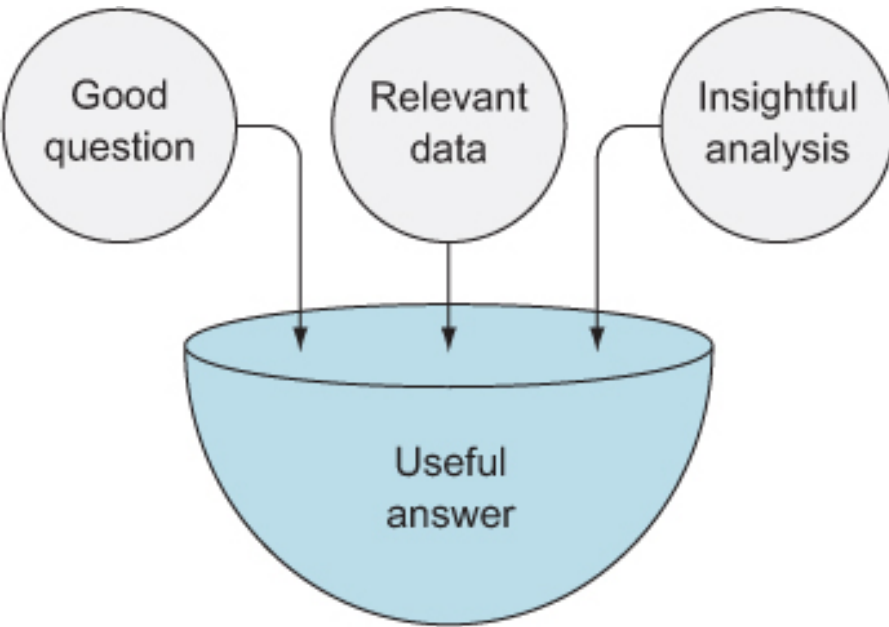
**PSTAT 197BC / CS190DE**

Alex Franks  
Sang-Yun Oh



[Image credit](#)

# Getting into Project mindset



- (and out of class/homework mindset)
- Classes often focus on what these components are
- For capstones, the sponsor, mentor, and team come up with these components

[Image credit](#)

# What you will work on ...

Data science projects as processes:

- **How to prepare** (save headaches later)
- **What to build** (set goals and execute your ideas)
- **Finish up project** (deliver output)

# How you will work ...

Data science project as team effort:

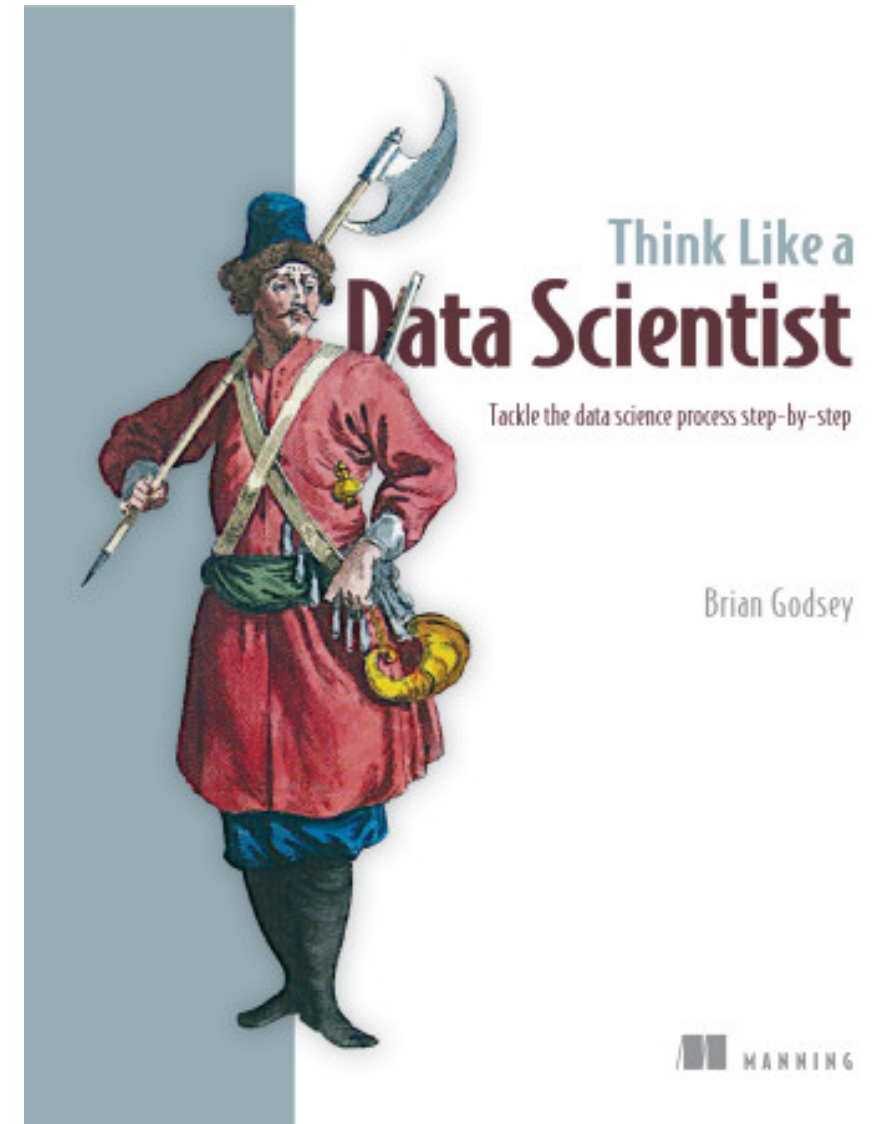
- Set goals and timeline
- Share responsibilities and teamwork
- Set and follow agreed-upon workflows and guidelines
- Establish common language and understanding
- Use same or compatible software tools

# Think like a Data Scientist











! Reading assignment:

Chapters 1 - 2

[Login](#) first then open [book link](#)



# Team work and reproducibility

- [Good enough practices](#)     
Data management, programming, collaborating, organizing projects, tracking work, and writing manuscripts
- [Reproducibility and collaboration](#)     
Case study in environmental science. Documentation, version control, team collaboration, and communication
- [Collaborative research teams](#)      
Diversity and performance of research teams. Importance of interpersonal skills and team functioning

# Topics in ethics and privacy

- [Ethics + Emerging Sciences](#)

Lectures

- [Big Data, Big Problems](#)

Size, aggregation (individual/communities), and informational harm

- [Data Ethics](#)

Institutional policies/procedures, regulations/laws, modification records, data origins, security plan, data access permissions, etc



# Version Control, Collaboration, Project Management

## Using Git and Github

- ★ [Git and Github for Poets](#), [more good videos](#)
- [Popular branching workflows](#)

## Collaborating with Github

- Project management board: [Overview](#), [Demo](#), [In-depth](#)
- [Tasks, Labels & Milestone](#)
- [Discussions](#)

# Grading 1

- **Meetings/Participation (individual): 20%**
  - Attendance in group meetings and lecture time sessions
- **Assigned Reading: 20%**
  - Must make at least one comment (question, answer, thought etc) on assigned readings

# Grading 2

- **Deliverables (group):**  
(More details to be added)
  - **Progress updates:** 30%  
3 progress updates in this Jupyter book
  - **End of Quarter Project Summary:** 20%  
Report and video presentation
- **Sponsor and mentor evaluations (individual):** 10%  
Based on your engagement and contribution to the project

# First Week - Immediately

- Send weekly meeting Doodle poll to [your team](#)
- Create [GitHub](#) account and [notify us via Google form](#)
- Join [Gitter Community](#)

## After private repository and Gitter room access:

- Introduce yourself to your team
- Create issue for Doodle poll, assign it, then close it when complete

# Before Next Week

**Reading Assignment** (before January 13, 2020):

Add at least one comment, question, or reply to [Google Drive files](#)

- *Think Like a Data Scientist*, Chapter 1
- *Good enough practices in scientific computing*