STA 235H - Introduction

Fall 2022

McCombs School of Business, UT Austin

Welcome to STA 235H
Data Science for Business Applications

Introductions

About the instruction team

Prof: Magdalena Bennett, Ph.D.

- Assistant Professor in the Stats Group (IROM department)
- Ph.D. in Economics of Education, Columbia University
- Research: Causal Inference (+ ML) applied to social policies (e.g. education).

T.A.: Shentao Yang (Ph.D. student)

T.A.: Katie Aufricht (3rd-year Honors)

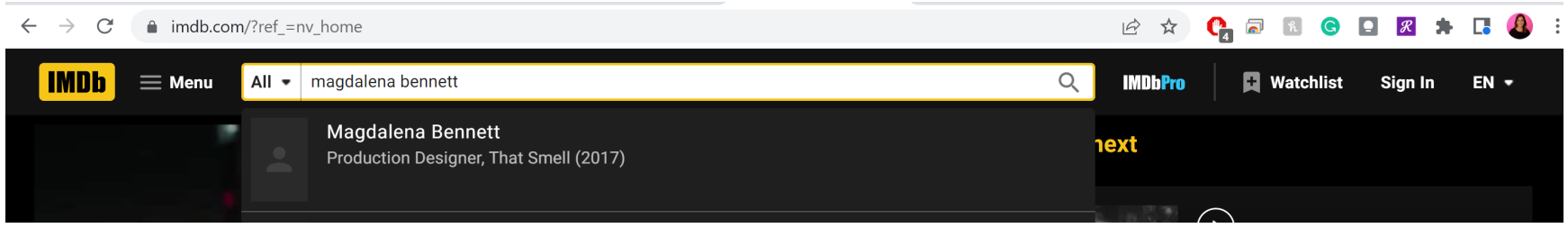
T.A.: Isabella Hsu (3rd-year Honors)

Introduce yourself!

Interesting (or uninteresting) fact about yourself

Interesting fact about me?

… I have a credit on IMDB.



Introduce yourself!

Interesting (or uninteresting) fact about yourself

Let's review the syllabus

Please, read the syllabus!

- Task before our first class: $\approx$ 60% of students completed it.
- How did you do in the survey?

About this course

- Objective:

"[G]ain the tools you need to tackle real-world problems from a quantitative perspective."

You don't need to be a data scientist for this class to be useful!

About this course

1) Multiple Regression

2) Causal Inference

3) Prediction

How, when, and where?

- In-person (Fall 2022): 2 hrs/week at UTC 1.102
- Drop-in Office Hours:

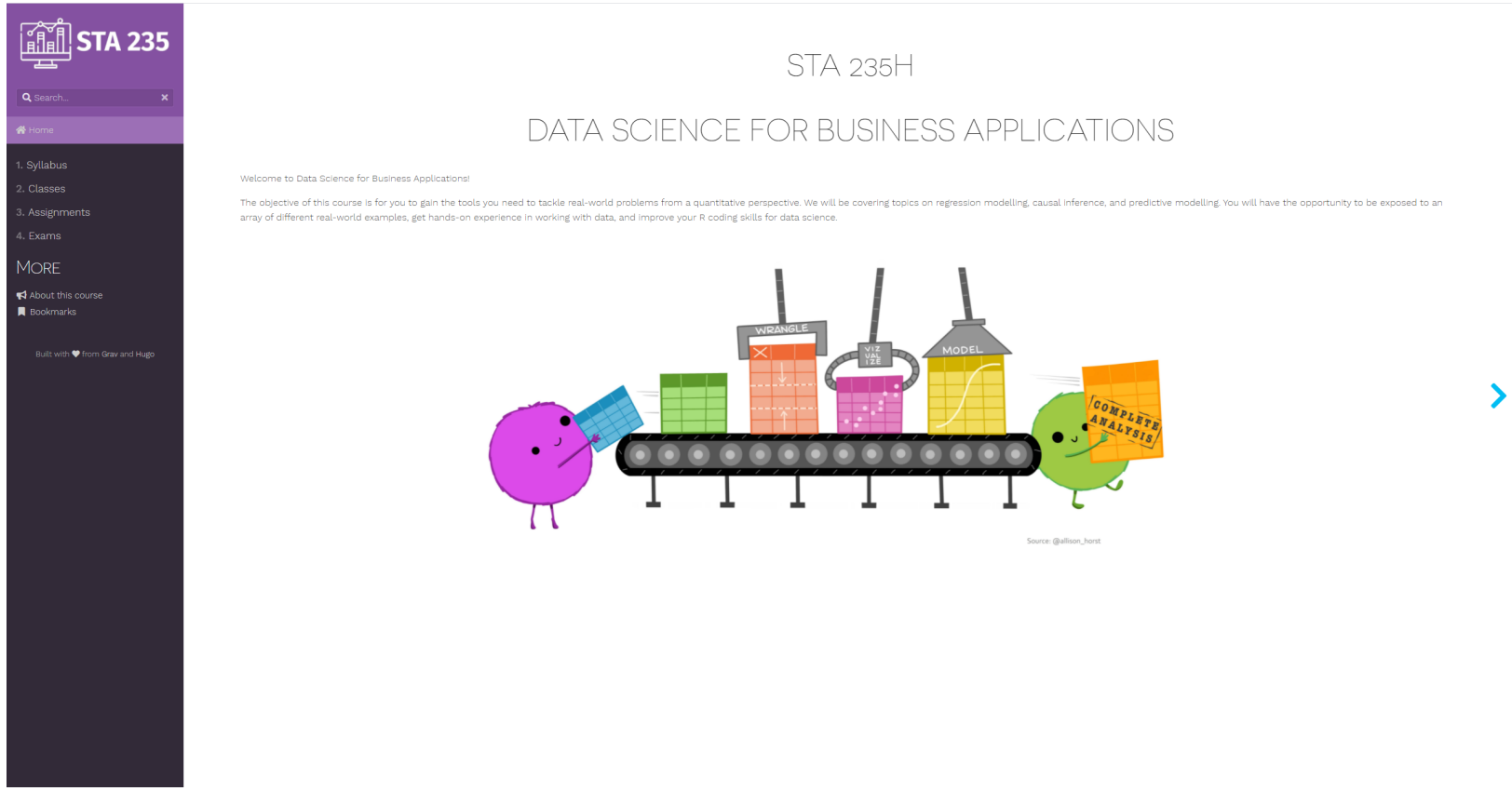
Prof. Bennett:
Tue 3:30 - 5:30 PM
Thu 3:30 - 5:30 PM

T.A.s:
TBD

- Other times available upon request

How, when, and where? (Cont.)

http://sta235.netlify.app



Classroom Norms

- Please, be on time.
- Participate and ask questions! (cold-calling can be used to loosen the atmosphere)
- Bring your laptop: We will be doing in-class coding (let me know if you have any issues with this point).

What will you need?

- A laptop to bring to class.
- R & R Studio
- Required Books:
    - Angrist, J. & J. Pischke. (2015). "Mastering Metrics". Princeton University Press. (*Buy used or new*)
    - James, G et. al. (2021). "An Introduction to Statistical Learning with Applications in R". Springer. (*Available online*)

How to succeed in this course?

- Attend class
- Slides are uploaded before class (*not self-contained*). Take notes but focus on understanding.
- Ask questions during class!
- Complete all readings and assignments by the assigned date
- Get an early start on assignments and follow the submission guidelines

Assignments, Exams, and Grading

- CAVEAT ABOUT TIME TO DO THE ASSIGNMENTS.

Assignments, Exams, and Grading

- Just in Time Teaching (JITT) assignments (10%):
  - Short online questionnaires about readings and/or material.
  - Submit by 11:59 PM on Sunday (for Tue class) or Tuesday (for Thu class) before that week's class.
  - Graded for completion (new material) and correctness (for material already seen). You can re-take it as many times as you want!
- 7 homework assignments (30%):
  - Assignments include both written questions and code.
  - You can drop one hw assignment (only 6 will count).
  - Assignments are individual. No collaboration, copying, or plagiarism will be accepted.

Read submission guidelines

Assignments, Exams, and Grading (Cont.)

- **Midterm (15%) and Final Exam (20%):**
  - Take-home exam (timed). Final exam is cumulative.
- **Final project (20%):**
  - Group project about data analysis.
- **Participation (5%):**
  - Attendance will be taken on 5 random classes. You can be absent in one of them without justification.
  - If you miss more than one (1) of those classes, you can make up with participation.

Assignments, Exams, and Grading (Cont.)

- There are no extensions for assignments, but you can drop drop one (1) HW and drop one (1) JITT.
    - Assignments are posted with enough time to account for frequent issues (e.g. sickness, interviews, etc.)

Please reach out to the instruction team if you have any issues

- Cutoffs for final letter grade:

Grade	A	A-	B+	B	B-	C+	C	C-	D	F
Cutoff	94%	90%	87%	84%	80%	77%	70%	65%	60%	<60%

There can be curving in individual assignments, but NOT for the final grade.

There will not be extra credit

Communicating with the instruction team

- Email address: m.bennett@austin.utexas.edu
  - Use the subject STA 235H - Your subject.
  - Email me directly for questions related to course administration.
  - Usually respond in 1 business day.
  - Please, do not send messages through Canvas.
  - *General questions should be posted on Canvas (Chatter)*
- Chatter:
  - Forum style discussion board.
  - Quickest way to get an answer about class material.
  - You can post with your name or anonymously (see instructions on Canvas)

Collaborations and Academic Integrity

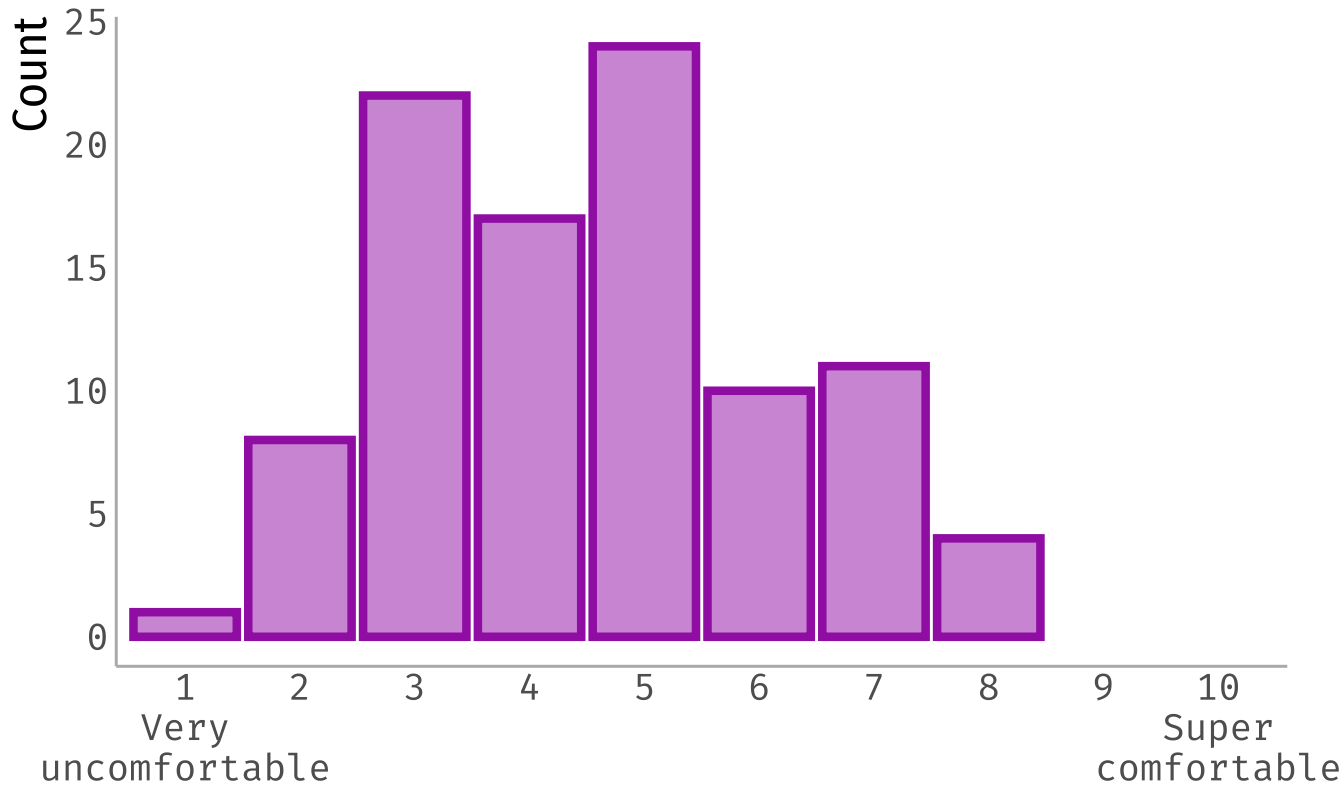
- You are encouraged to form study groups!
    - Studying or discussing class material with others does not mean you can copy or directly collaborate on assignments.
    - Students are responsible for their own work. All of it.
- Do not share your files with other students
    - If we find any evidence of copying or plagiarism, all students involved will be subject to disciplinary measures.
- Remember to give credit where credit is due!
    - Use citations and references when you use someone else's work.

What questions do you have?

Your expectations

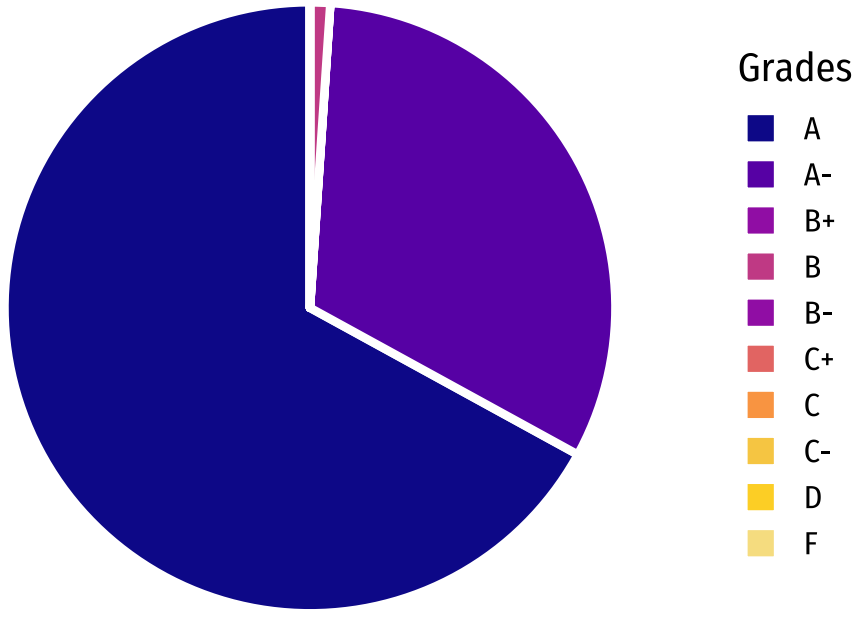
What do you expect to learn from this course?

How comfortable are you with R?



What grade do you expect to get?

- Confidence is great (but also hard work)



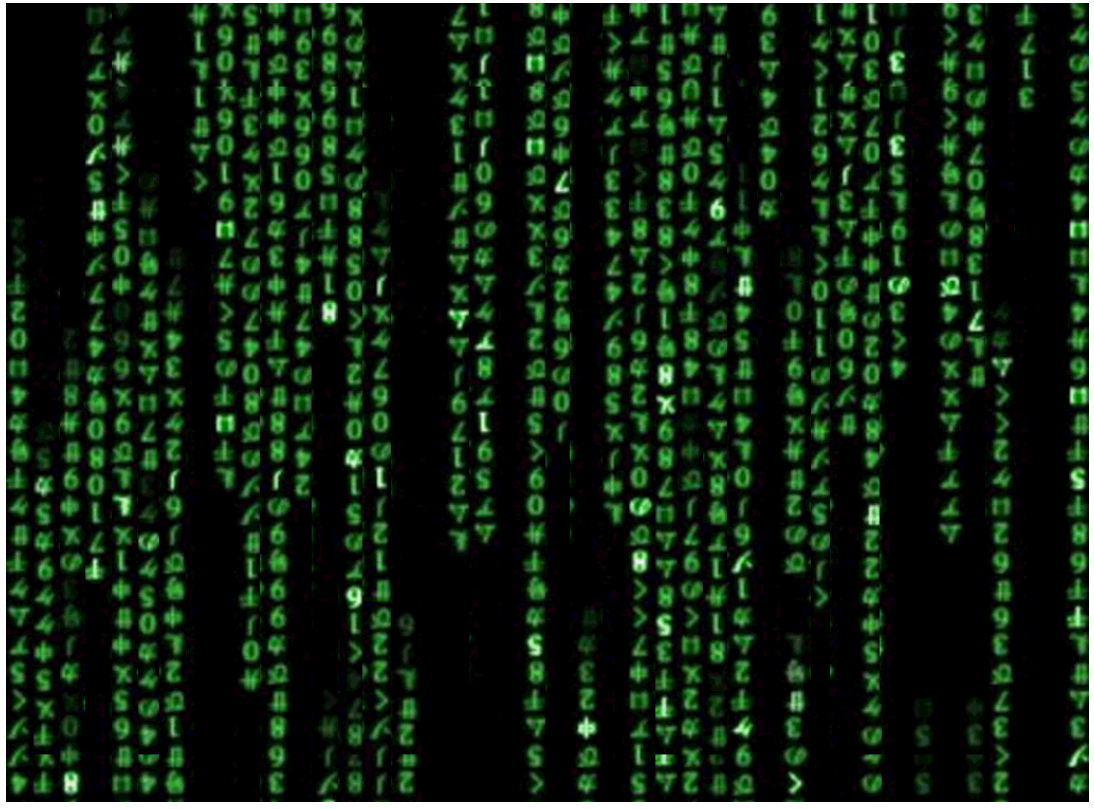
A brief motivation

What is Data Science?

What are we going to see in this course?

What should I expect to learn by the end of the semester?

What is Data Science?



Data Science tasks

By Hernán, Hsu, and Healy:

Description

Prediction

Causal Inference

Data Science tasks

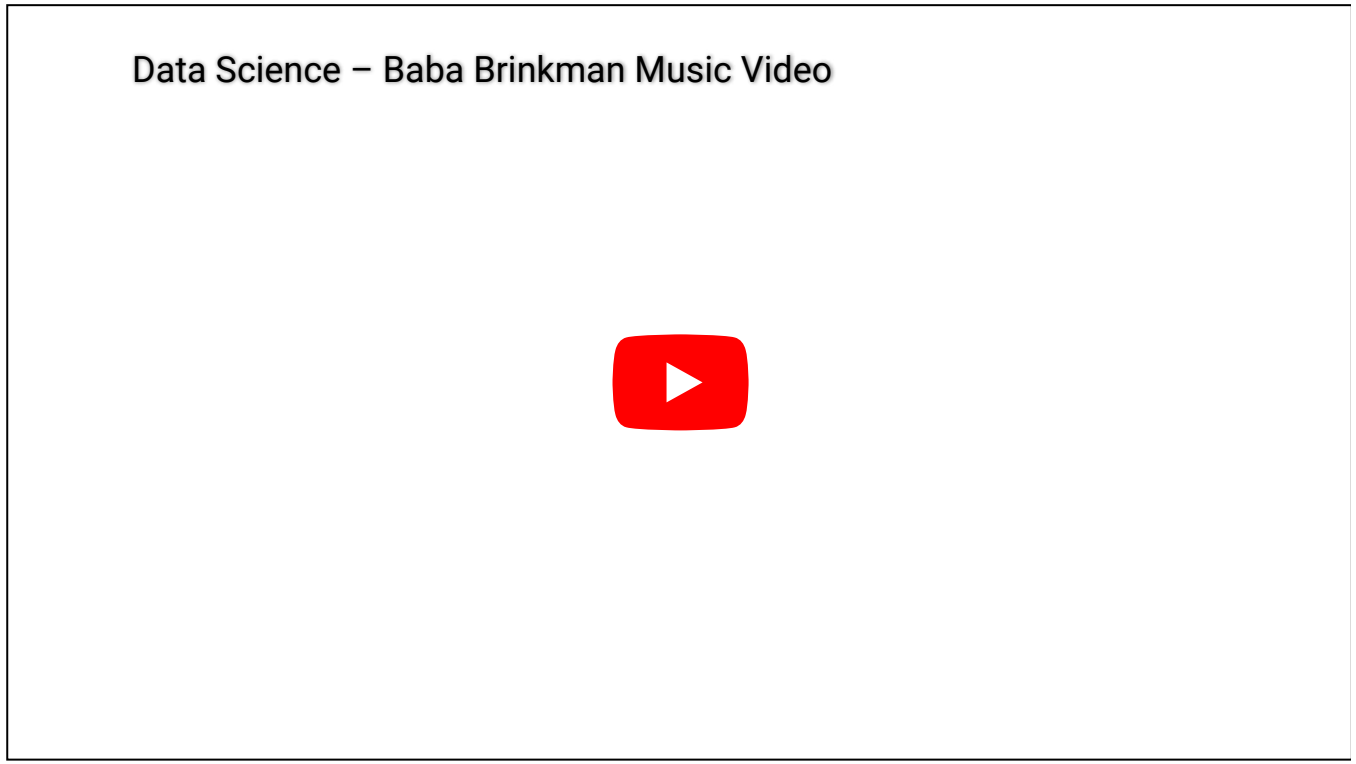
Can we classify our customers into different segments?

What is the probability of a shopper coming back to our website?

What is the effect of increasing our advertising budget on our total revenue?

We'll review all of these in this class!

Data Science vs. Statistics?



> "But it's a shallow journey if ONLY the machine's learning"

Lübke, K. (2020). "Causality in Data Science Education". *Causal Data Science Meeting*

After this course...

1) Bridge the gap between the "what" and the "how"

2) Be critical consumers of "Data Science"

Some notes before the break

- We will be using a seating chart for contact-tracing purposes:
    - Choose your preferred seat during the break and then write it down on the sitting chart I'll pass around.
- "Services for Students with Disabilities (SSD) is seeking the assistance of students to serve as volunteer notetakers."
    - Volunteers will be eligible to receive volunteer hours in appreciation for their time.
    - If you are a good notetaker and interested in helping other students, please contact me after class.