STA 235H - Introduction

Fall 2021

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
- Research: Causal Inference (+ ML) applied to social policies (e.g. education).

T.A.: Pedro Santos (Ph.D. student)

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

Introduce yourself!

Interesting (or uninteresting) fact about yourself

Interesting fact about me?

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.
- There was also an Easter Egg in the syllabus

12% fo students found it!

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

- Structure:

1) Multiple Regression

2) Causal Inference

3) Prediction

How, when, and where?

- In-person (Fall 2021): 2 hrs/week at UTC 1.130
- Online Office Hours:

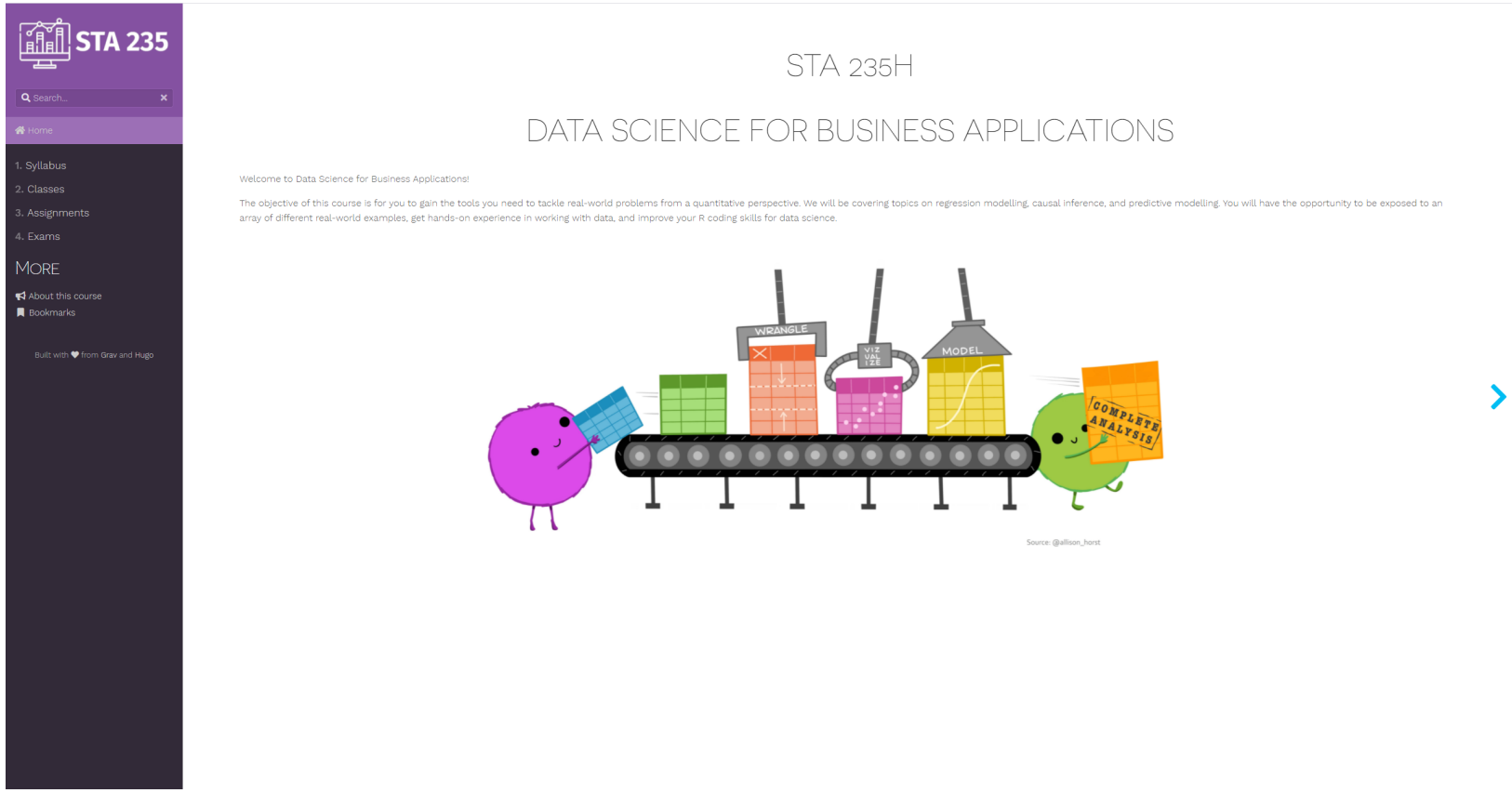
Prof. Bennett:
Tue 3:30 - 5:00 PM
Thu 3:30 - 5:00 PM

T.A.s:
TBD
(R intro session)

- Appointments by calendly
- 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).

Let's talk a little bit about COVID-19 measures...

Keep yourself and other SAFE

Masks are not mandatory, but are highly encouraged

If you're not feeling well, please stay home

Check out more COVID-19 resources in the Syllabus

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. Take notes but focus on understanding
- Ask questions during class
- Complete all readings and assignments by the suggested (or assigned) date
- Get an early start on assignments and follow the submission guidelines

Caveat: We are on a pandemic, so reach out to the instruction team if you are having trouble

Assignments, Exams, and Grading

- Just in Time Teaching (JITT) assignments (10%):
  - Short online questionnaires about readings 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.
- 6 group homework assignments (30%):
  - Assignments include both written questions and code.
  - Groups (3-4) will be randomly assigned by the instruction team
  - No copying or plagiarism will be accepted.

Read submission guidelines

Assignments, Exams, and Grading (Cont.)

- Midterm and Final Exam (20% each):
  - Take-home exam. Final exam is cumulative.
- Final project (20%):
  - Group project about prediction.

Assignments, Exams, and Grading (Cont.)

- You get one (1) 24-hour extension in a homework assignment, and you can drop one (1) JITT.

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%

Assume there is no grade curving (if I do, it will always be in your favor).

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.
    - *General questions should be posted on Canvas*
- Canvas discussion board:
    - Quickest way to get an answer about class material.
    - Do not send messages through Canvas.

Collaborations and Academic Integrity

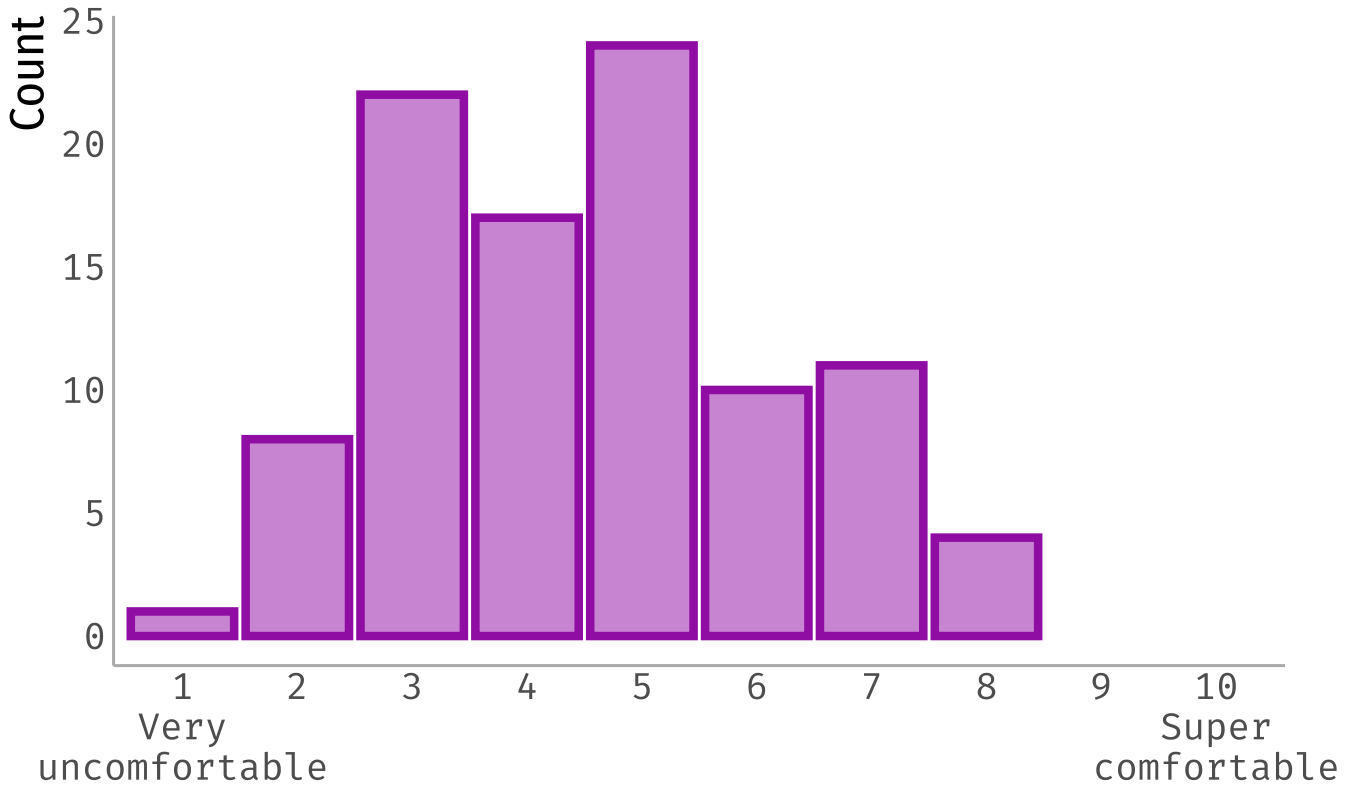
- You are encouraged to form study groups!
  - Studying or discussing assignments with others does not mean "divide and conquer".
  - Students are responsible for their own work. All of it.
- Do not share your files with other students
  - If we find any evidence of copying or plagiarism, all students 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)

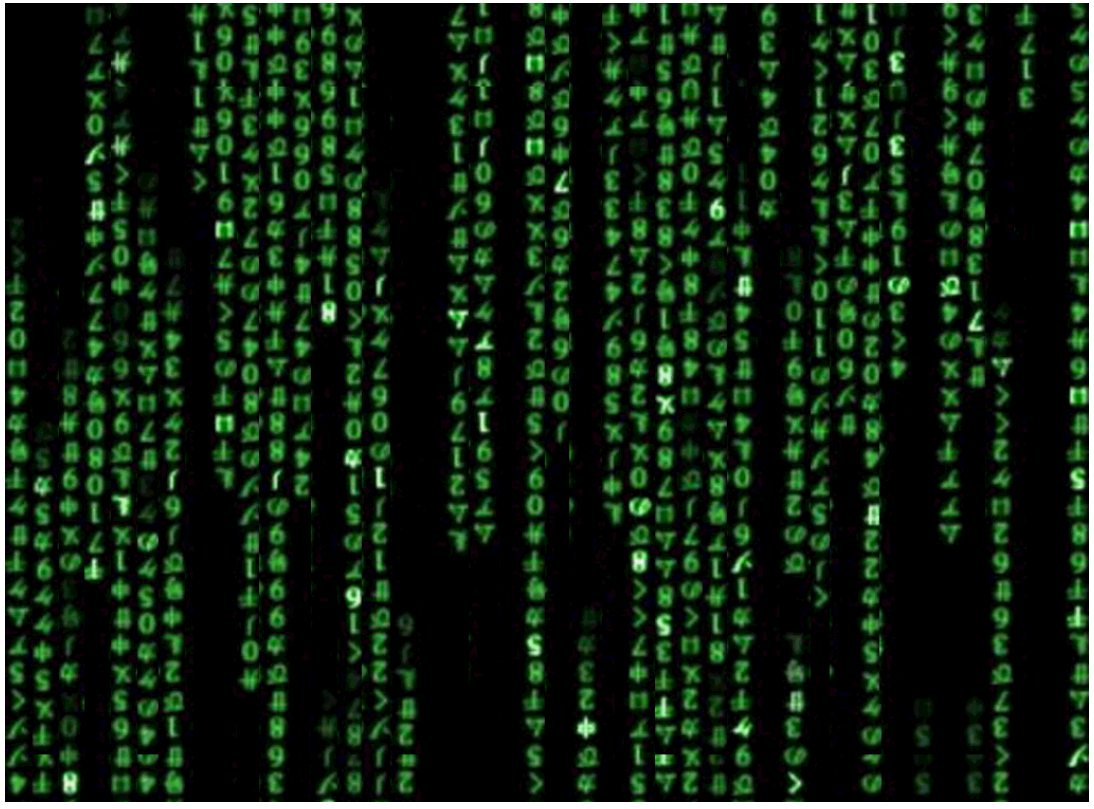
A brief motivation

What is Data Science?

What are we going to see in this course?

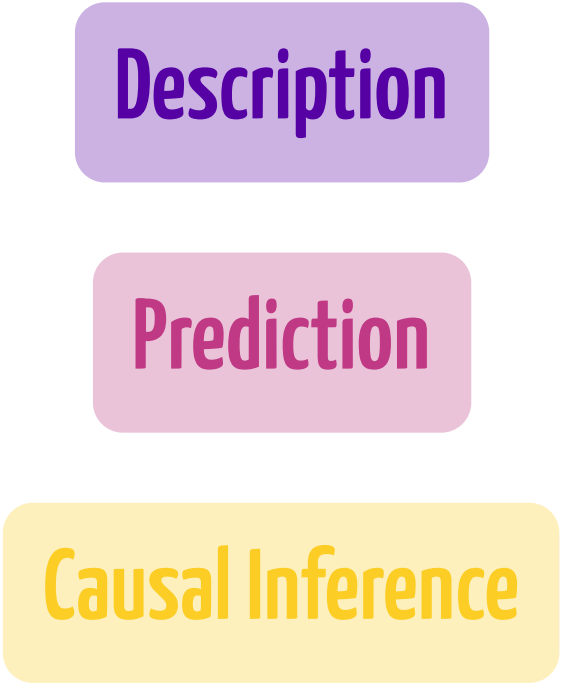
What should I expect to learn by the end of the semester?

What is Data Science?



Data Science tasks

By Hernán, Hsu, and Healy:



Data Science tasks

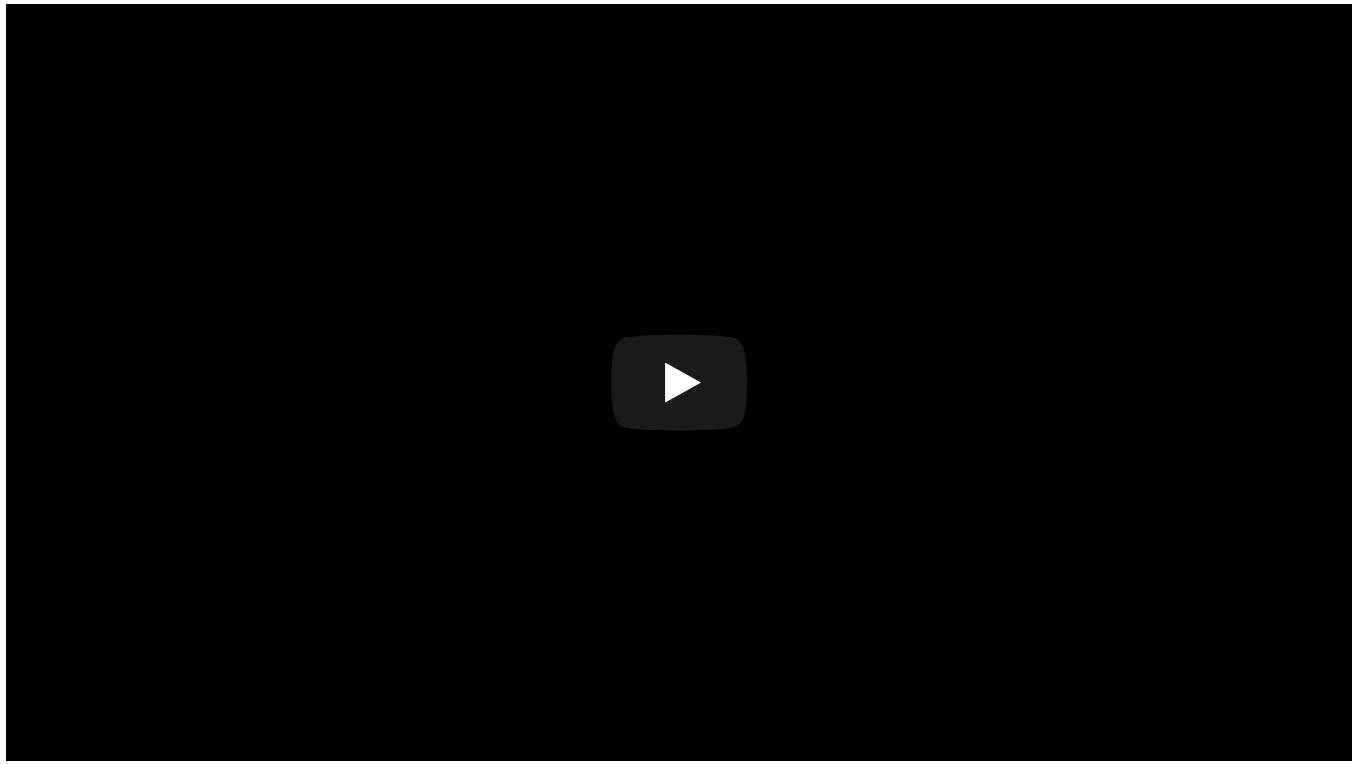
Can we classify our customers into different segments?

What is the probability of a shopper to come 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"