

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 very cute pup.



Introduce yourself!

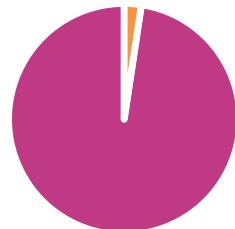
Interesting (or uninteresting) fact about yourself

Let's review the syllabus

Please, read the syllabus!

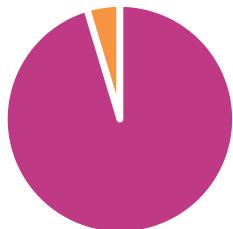
How did you do in the survey?

Where is Prof. Bennett's office?



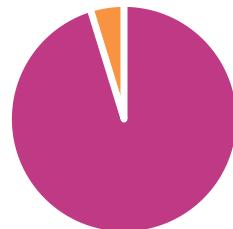
■ CBA Building, 6th Floor
■ GBS Building, 6th Floor

Does attendance count towards your grade?



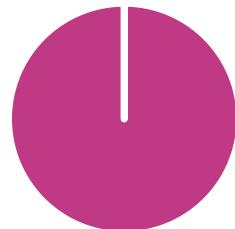
■ No
■ Yes

JITTs are due every week



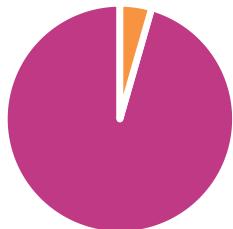
■ FALSE
■ TRUE

How many homework assignments can you drop?



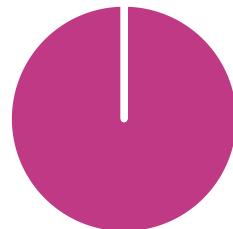
■ 1

It's ok to submit late assignments because there are extensions



■ FALSE
■ TRUE

Do you have to be present for the final project presentations?



■ Yes

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:

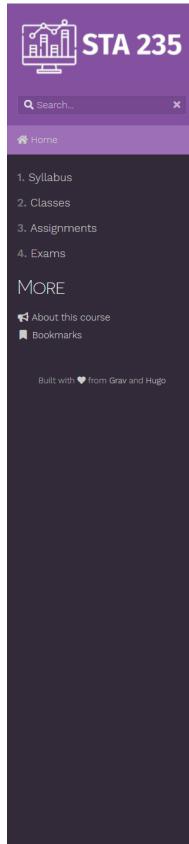
Prof. Bennett:
Tue 3:30 - 5:30 PM
Thu 3:30 - 5:30 PM

T.A.s:
Weekly + HW weeks (TBD)

- Other times available upon request

How, when, and where? (Cont.)

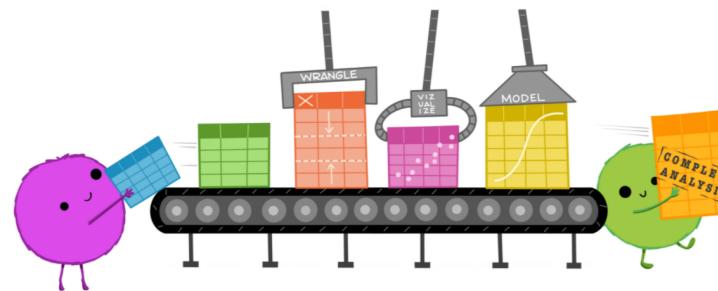
<http://sta235.netlify.app>



STA 235H

DATA SCIENCE FOR BUSINESS APPLICATIONS

Welcome to Data Science for Business Applications!
The objective of this course is for you to gain the tools you need to tackle real-world problems from a quantitative perspective. We will be covering topics on regression modelling, causal inference, and predictive modeling. You will have the opportunity to be exposed to an array of different real-world examples, get hands-on experience in working with data, and improve your R coding skills for data science.



Source: @allison_horst

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

- All assignments and exams are **take-home** and **open-book**.
- Provide **structured flexibility** in submission of assignments to avoid academic inequality:
 - E.g. HW assignments ~ 2-3hrs → 1 week: Accounts for (typical) foreseeable and unforeseeable events
 - Drop an assignment
 - No extensions
- In the case of other circumstances that do not fit in the previous description, **please reach out**.

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%):**
 - All in Canvas.
 - 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 (restricted time). 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 penalty.
 - If you miss more than one (1) of those classes, you can make up with participation.

Assignments, Exams, and Grading (Cont.)

- Assignments and exams are **usually curved**
 - Final grade **will not be curved**.
- 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%

- Cutoff scores are strict (no rounding)

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 (Chatter)
- **Please, do not send messages through Canvas.**
- Note: This is Ms. Bennett



Communicating with the instruction team

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)

The screenshot shows the Chatter interface for a course titled "Fa22 - DATA SCI FOR BUSN APPS: HON (05340)". The top navigation bar includes search and filter options. The main area displays a single post from Magdalena Bennett, dated 8/8/2022 at 2:32 PM, with the message "Welcome!". Below the post, there is a welcome message and some posting rules. At the bottom of the screen, there are buttons for "Unpin", "Lock", and "Comment".

CHATTER

Search posts...

Fa22 - DATA SCI FOR BUSN APPS: HON (05340)

All Unread My Posts

Magdalena Bennett 8/8/2022 2:32 PM

Welcome!

...

Welcome!

Magdalena Bennett 8/8/2022 2:32 PM (Edited)

Hello everyone!

This forum is designed to have a more fluid discussion about the contents of the course. You can post questions you might have on assignments, something that was not very clear in class, random thoughts... whatever you want! (as long as you keep it polite).

Please refer to the following rules when posting here:

- **Use a relevant post title:** Make it easy for people to answer (or read the answers to) your post. Include a concise description of what your question refers to! Doesn't need to be super complete, but don't make it too generic (e.g. "Question about ridge regression", "What is up with potential outcomes?")
- **Include relevant tags:** I have created tags for assignments, general questions, announcements, etc. Please include the appropriate tag on your question so it's easy to find!
- **Be polite:** I usually recommend using your name when posting, because no one should ever feel self-conscious about asking a question (and it's good that you get used to it!), but in this forum you can post anonymously if you want to. I do ask that you keep a polite tone towards everyone here, though; do not write something that you wouldn't sign your name on.



Unpin Lock Comment

+ Write a Post

Collaborations and Academic Integrity

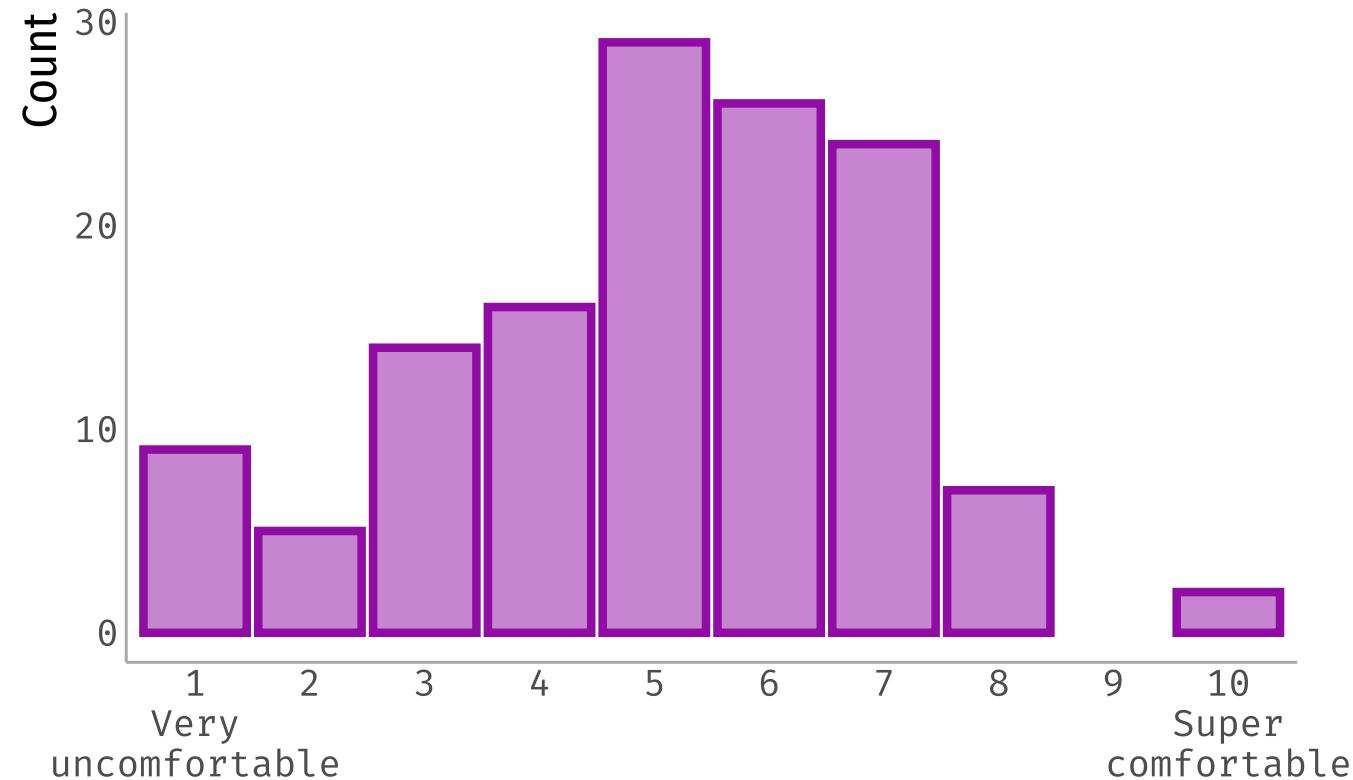
- You are encouraged to form study groups!
 - Studying or discussing class material with others does **not** mean you can copy or directly collaborate on assignments.
 - Students are responsible for their own work. All of it.
- Do not share your files with other students
 - If we find any evidence of 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

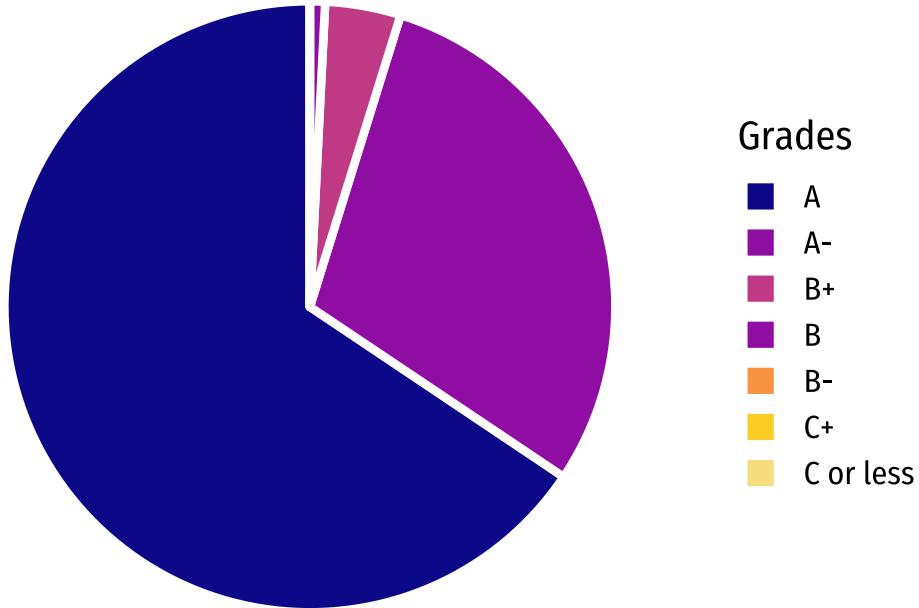
What do you expect to learn from this course?

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?

Data Science – Baba Brinkman Music Video



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

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.
- "Disability & Access (D&A) 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.