

# STA 235 - Introduction

*Spring 2021*

McCombs School of Business, UT Austin

# Welcome to STA 235

# Data Science for Business Applications

# Introductions

# About the instruction team

## Prof: Magdalena Bennett, Ph.D.

- Assistant Professor in the Stats Group.
- Research: Causal Inference + Econ of Education

## TA: Ricardo Lara

- Ph.D. candidate in Petroleum Engineering



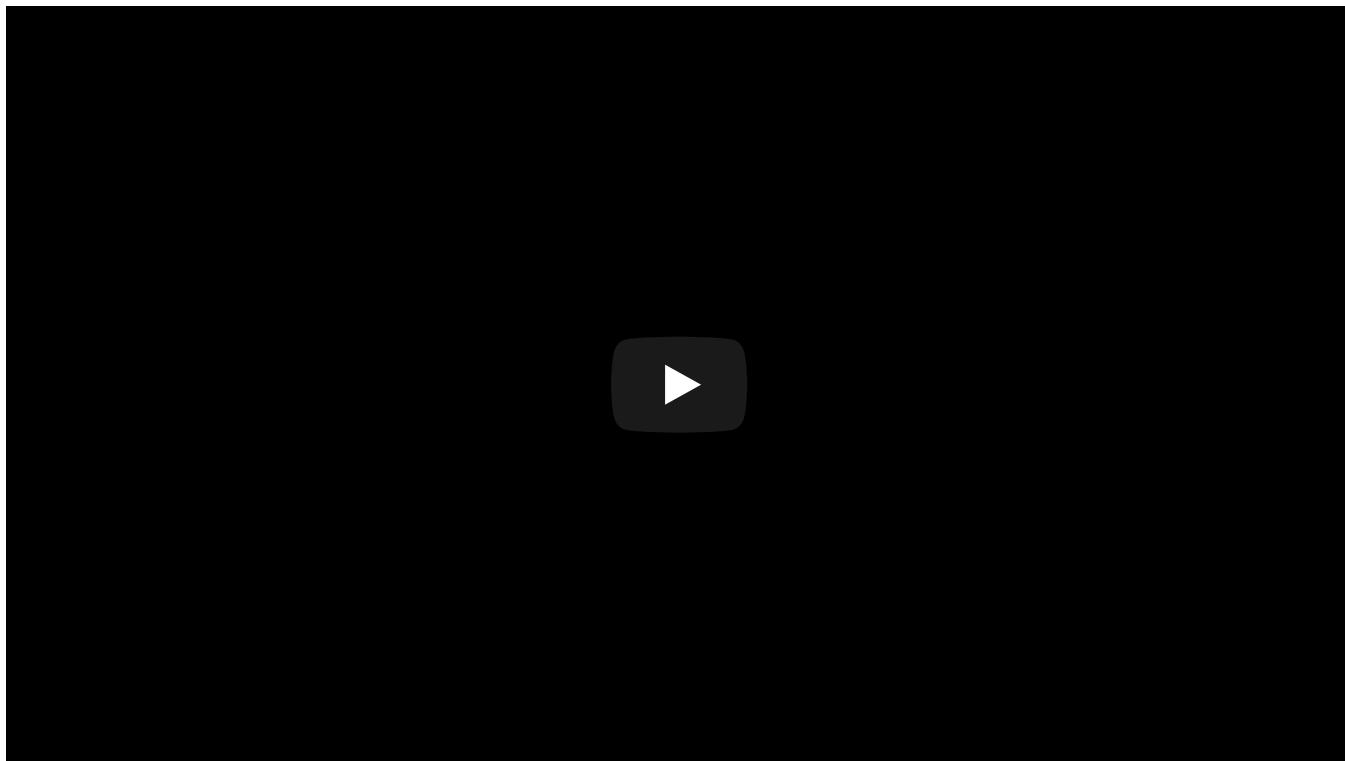
# Introduce yourself!

Interesting (or uninteresting) fact about yourself

or Walk-up song?

Let's review the syllabus

# Please, read the syllabus!



# About this course

- **Objective:**

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

- **Structure:**

1) Multiple Regression

2) Causal Inference

3) Prediction

# How, when, and where?

- **Online session for Spring 2021:** Mon at 12:30 - 2:15 PM (Sec. 1) and 2:30 - 4:15 PM (Sec. 2)
  - Be punctual!
- **Online Office Hours:**

**Prof. Bennett:**  
**Mon 5:00 - 6:00 PM**  
**Thur 4:00 - 5:00 PM**

**T.A.:**  
***TBD (before exams & assignments)***

- Appointments by **calendly**
- Other times available upon request

# How, when, and where? (Cont.)

<http://sta235.netlify.app>

The screenshot shows a web browser displaying the STA 235 website at <http://sta235.netlify.app>. The left side features a sidebar with a purple header containing the STA 235 logo and a search bar. Below the header, the sidebar lists navigation links: Home, 1. Syllabus, 2. Classes, 3. Discussion, and a MORE section with links for About this course, Bookmarks, and GitHub repo. At the bottom of the sidebar, it says "Built with ❤ from Grav and Hugo". The main content area has a white background with large text: "STA 235" and "DATA SCIENCE FOR BUSINESS APPLICATIONS". Below this, a welcome message reads: "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 modelling. 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." To the right of the text is a blue right-pointing arrow. At the bottom center is a cartoon illustration of two characters, one pink and one green, working on a conveyor belt with various data-related components like a bar chart, a scatter plot, and a model icon.

# "Classroom" Norms

- Keep your **video on** if connection allows it.
- Keep your **mic muted** unless speaking.
- **Raise your hand** to speak.
- Be **mindful** when using the **chat** - would you make this comment online to your neighbor?



# What will you need?

- A **computer** with internet connection.
  - Microphone and webcam (especially for office hours)
- **R & R Studio**
- **Required Books:**
  - Angrist, J. & J. Pischke. (2015). "Mastering Metrics". Princeton University Press.
  - James, G et. al. (2017). "An Introduction to Statistical Learning with Applications in R". Springer. (Available online)



# How to succeed in this course?

- **Attend class**
- Classes are recorded, so focus on **understanding**
- **Ask questions** during class
- Complete all **readings** and **assignments** by the suggested (or assigned) date
  - Caveat: We are on a pandemic, so reach out to the instruction team
- Get an **early** start on assignments and **follow the submission guidelines**

**Reach out if you need help!**

# Assignments, Exams, and Grading

- **3 homework assignments (20%) + 1 project (15%):**

- Assignments include both written questions and code.
- Prediction project: similar to a long homework.
- No copying or plagiarism will be accepted.

**Read submission guidelines**

- **Just in Time Teaching (JITT) assignments (10%):**

- Short online questionnaires about readings or material.
- Submit by midnight on Sunday before class.
- Graded for completion.

# Assignments, Exams, and Grading (Cont.)

- **Midterm exam (25%):**
  - In-class exam (online)
- **Final exam (30%):**
  - Common time for both sections (May 13th 2:00 - 5:00 PM)
  - Let me know as soon as possible if you have another work or non-academic commitment
- 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).

# Communicating with the instructor team

- Email address: **m.bennett@austin.utexas.edu**
  - Use the subject **STA 235 - Your subject**.
  - Email me directly for questions related to course administration.
  - Usually respond in 1 business day.
  - General questions should be posted on Piazza
- **Piazza 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.

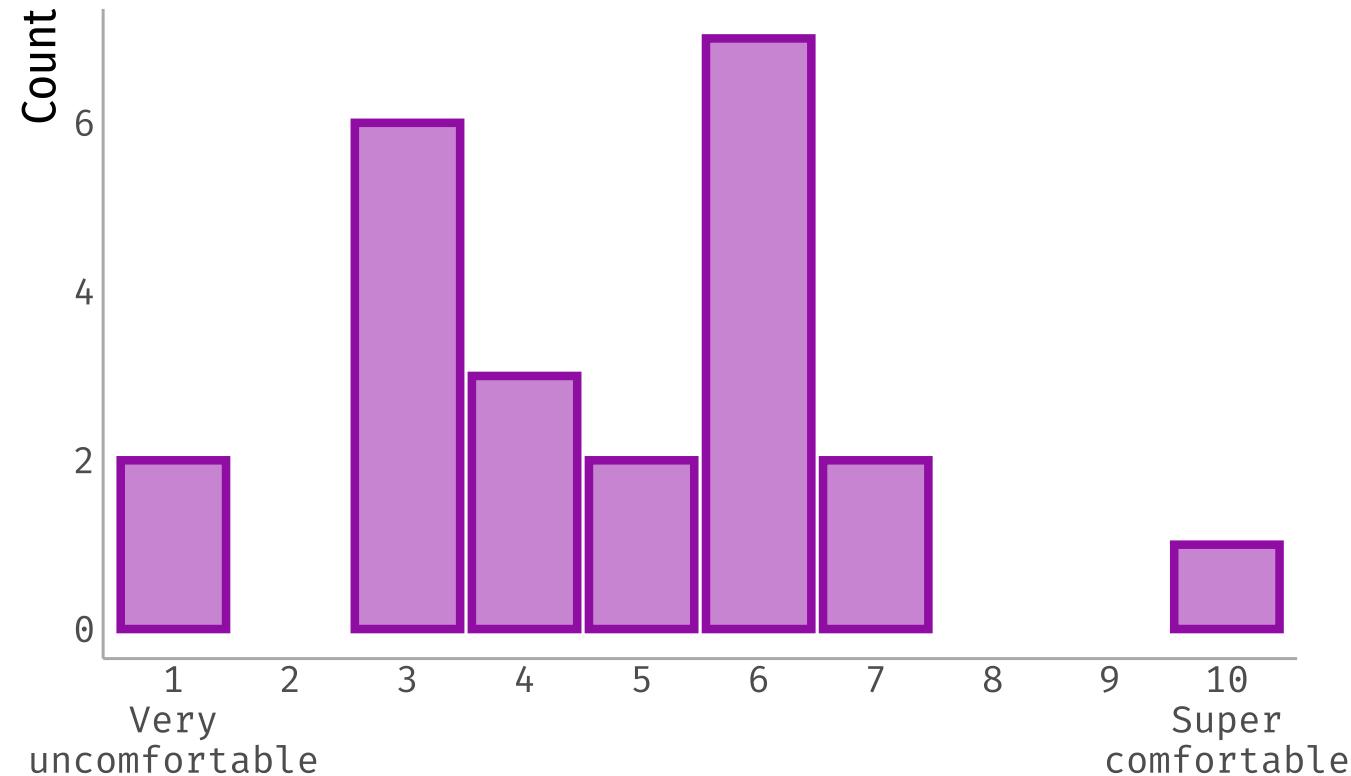
What questions do you have?

# Your expectations

# What do you expect to learn from this course?

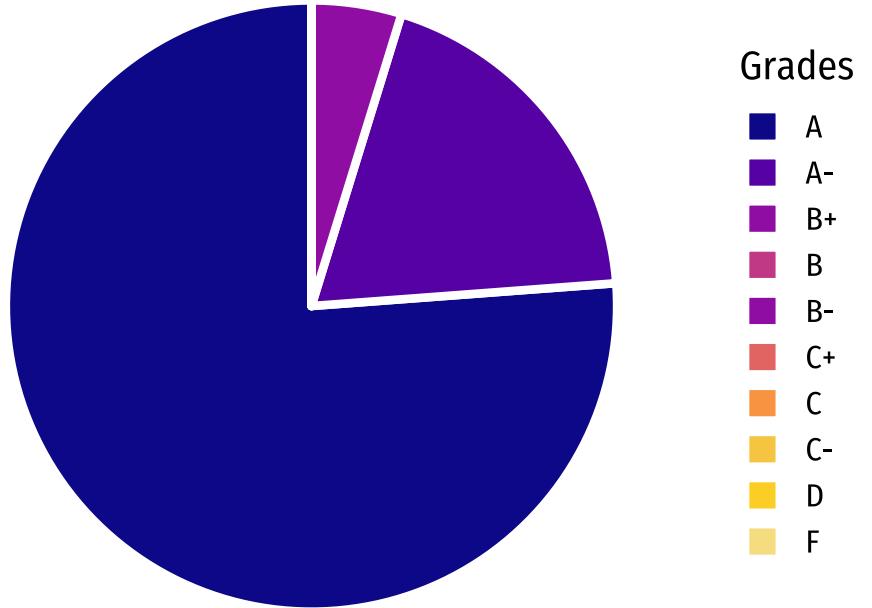
A word cloud centered around the term "regression". Other prominent words include "data", "business", "analytics", "learning", "models", "science", "apply", "hope", "skills", "solve", "quantitative", and "problems". The words are colored in various shades of blue, green, orange, and red, representing different semantic clusters or parts of speech.

# 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

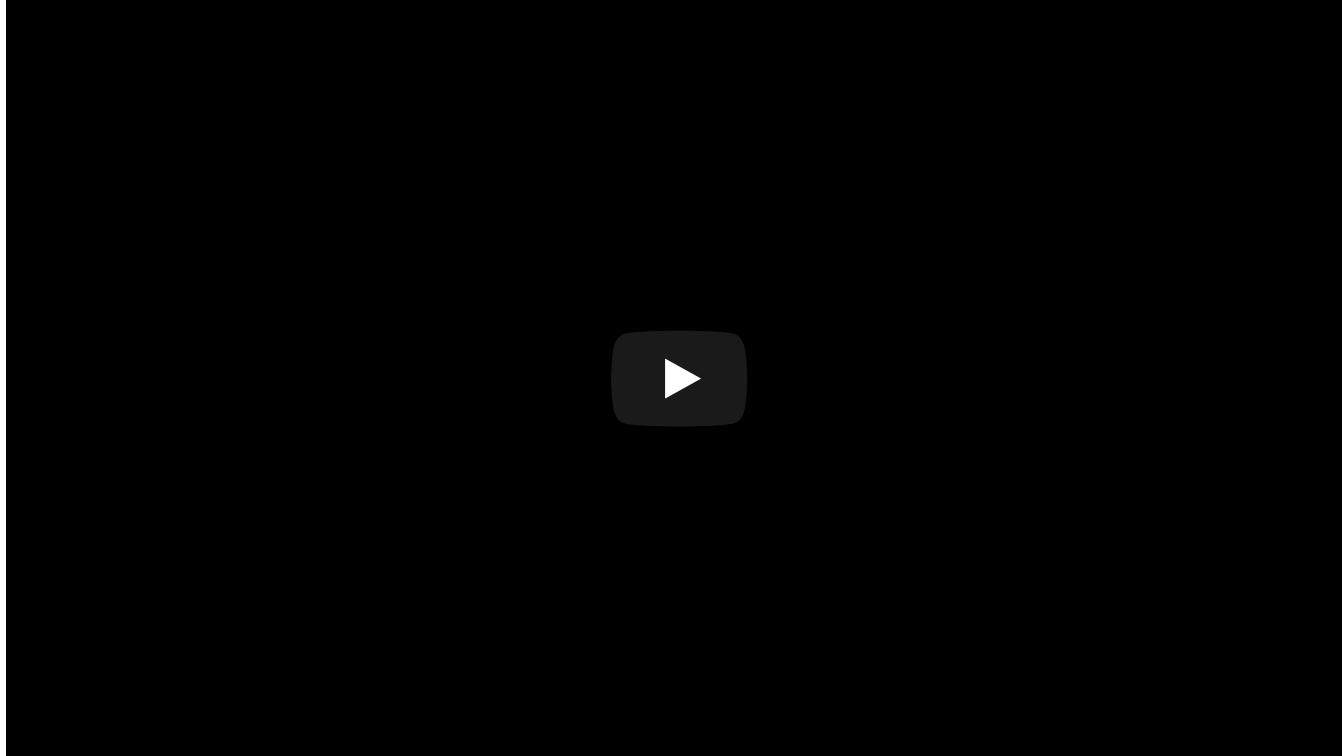
Can we classify our customers into different segments?

What is the probability of a shopper 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"