# Welcome to Intro to Data Science

Dr. Maria Tackett

08.27.19

# Welcome!

# What is Data Science?

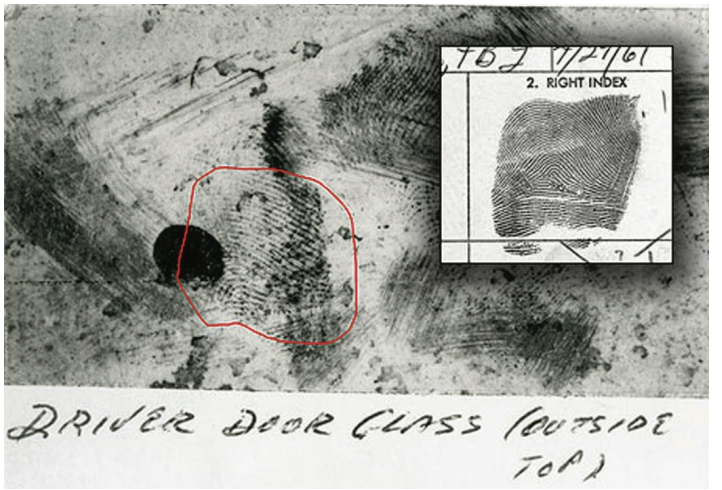
*"Data science is a concept to unify statistics, data analysis, machine learning and their related methods in order to understand and analyze actual phenomena with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science."*

-Wikipedia

# Instructor

[Prof. Maria Tackett]

maria.tackett@duke.edu
Old Chem 118B
Th 1p - 2:30p

# Teaching Assistants

Max Bartlett

- maxwell.bartlett@duke.edu
- Old Chem 203B
- Mon 6p - 8p

Meredith Brown

- meredith.brown@duke.edu
- Old Chem 203B
- Mon 12:30p - 2:30p

Steven Herrera

- rosvid.herrera.tenorio@duke.edu
- Old Chem 203B
- Thu 5:30p - 7:30p

Jessie Ou

- yanchen.ou@duke.edu
- Old Chem 203B
- Sun 12p - 2p

# Teaching Assistants


Malavi Ravindran

malavi.ravindran@duke.edu
Old Chem 203B
Mon 3p - 5p

Becky Tang

becky.tang@duke.edu
Old Chem 203B
Wed 3p - 5p


# Where to find information

- Course website: http://bit.ly/sta199-fa19
- GitHub (assignments): https://github.com/sta199-fa19

# Course Objectives

- Learn to explore, visualize, and analyze data in a reproducible and shareable manner
- Gain experience in data wrangling and munging, exploratory data analysis, predictive modeling, and data visualization
- Work on problems and case studies inspired by and based on real-world questions and data
- Learn to effectively communicate results through written assignments and final project presentation

# Examples of Data Science

# Billboard Hot 100 Analytics

Analysis: https://towardsdatascience.com/billboard-hot-100-analytics-using-data-to-understand-the-shift-in-popular-music-in-the-last-60-ac3919d39b49

GitHub repository with data and R scripts

# Atlas of Redistricting

https://projects.fivethirtyeight.com/redistricting-maps/north-carolina/

# A Year as Told by Fitbit

https://livefreeordichotomize.com/2017/12/27/a-year-as-told-by-fitbit/

# Your Turn!

# Create a GitHub account

Go to https://github.com/, and create an account (unless you already have one). After you create your account, click here and enter your GitHub username.

Tips for creating a username from Happy Git with R.

- Incorporate your actual name!
- Reuse your username from other contexts if you can, e.g., Twitter or Slack.
- Pick a username you will be comfortable revealing to your future boss.
- Shorter is better than longer.
- Be as unique as possible in as few characters as possible.
- Make it timeless. Don't highlight your current university, employer, or place of residence.
- Avoid words laden with special meaning in programming, like `NA`.

Raise your hand if you have any questions.

# Join RStudio.cloud

- Go to http://bit.ly/sta199-fa19-rstudio, and log in with your GitHub credentials.
- You should see a project called *UN Votes*. Click "Copy"; this will create your copy of the project and launch it.

Raise your hand if you have any questions.

# UN Votes Analysis

1. In the Files pane in the bottom right corner, spot the file called `unvotes.Rmd`. Open it, and then click on the "Knit" button.

2. Go back to the file and put your name in the author field at the top of the file (in the `yaml` -- we will discuss what this is at a later date). Knit again.

3. Change the country names in parts 1 and 2 to countries that interest you. The spelling and capitalization must match what's in the data, so you can use the Appendix to see the correct spelling and capitalization. Knit again.

You have made your first data visualization!

Raise your hand if you have any questions.

# Discussion

Discuss the following with a partner.

1. Start by introducing yourself! Name, year, major/ academic interest, favorite hobby.

2. Consider the plot in Part 1. Describe how the voting behaviors of the four countries have changed over time.

3. Consider the plot in Part 2.

    - On which issues have the two countries voted most similarly in recent years?
    - On which issues have they voted most differently in recent years?
    - Has this changed over time?

# Course Policies

# Class Meetings

## Lecture

- Focus on concepts behind data analysis
- Interactive lecture that includes examples and hands-on exercises
- Bring fully-charged laptop to every lecture
    - Please let me know as soon as possible if you do not have access to a laptop

## Lab

- Focus on computing using R `tidyverse` syntax
- Apply concepts from lecture to case study scenarios
- Work on labs in teams of 3 - 4
- Bring fully-charged laptop to every lab

# Textbooks

- OpenIntro Statistics, 4th Edition
  - Free PDF available online. Hard copy available for purchase.
  - Assigned readings about statistical content
- R for Data Science
  - Free online version. Hard copy available for purchase.
  - Assigned readings and resource for R coding using `tidyverse` syntax.

# Activities & Assessments

- Homework: Individual assignments combining conceptual and computational skills. *Lowest score dropped.*
- Labs: Team assignments focusing on computational skills. Start labs on Friday and due Tuesday. *Lowest score will be dropped.*
- Exams: Two take-home exams.
- Final Project: Team project presented during the final exam period, **Saturday, December 14, 7p - 10p**. You must complete the project and present in class to pass the course.

# Activities & Assessments

- Application Exercises: Exercises usually started in class and completed in teams by the next class. Check/no check grade.
- Writing Exercises: Short writing prompts completed individually along with peer reviews.
- Teamwork: Teams of 3-4 based on survey and pretest results. Consistent throughout the semester. Periodic peer evaluations.

# Grade Calculation

| Component | Weight |
|---|---|
| Homework | 20% |
| Labs | 15% |
| Exam 1 | 17.5% |
| Exam 2 | 17.5% |
| Final Project | 15% |
| Participation & Application Exercises | 5% |
| Writing Exercises | 5% |
| Teamwork | 5% |

- If you have a cumulative numerical average of 90 - 100, you are guaranteed at least an A-, 80 - 89 at least a B-, and 70 - 79 at least a C-. The exact ranges for letter grades will be determined after Exam 2.
- You are expected to attend lectures and labs. Excessive absences or tardiness can impact your final course grade.

# Excused Absences

- Students who miss a class due to a scheduled varsity trip, religious holiday, or short-term illness should fill out the respective form.
  - These excused absences do not excuse you from assigned work.
- If you have a personal or family emergency or chronic health condition that affects your ability to participate in class, please contact your academic dean's office.
- Exam dates cannot be changed and no make-up exams will be given.

# Late Work & Regrade Requests

- Homework assignments:
  - Late but within 24 hours of deadline: 20% penalty
  - Not accepted if submitted any later
- Late work will not be accepted for the take-home exams or final project.
- Regrade requests must be submitted within one week of when the assignment is returned using the link posted in the course syllabus

# Academic Honesty

All work for this class should be done in accordance with the Duke Community Standard.

> To uphold the Duke Community Standard:
>
> - I will not lie, cheat, or steal in my academic endeavors;
> - I will conduct myself honorably in all my endeavors; and
> - I will act if the Standard is compromised.

Any violations will automatically result in a grade of 0 on the assignment and will be reported to Office of Student Conduct for further action.

# Reusing Code

- Unless explicitly stated otherwise, you may make use of online resources (e.g. StackOverflow) for coding examples on assignments. If you directly use code from an outside source (or use it as inspiration), you must or explicitly cite where you obtained the code. Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism.
- On individual assignments, you may discuss the assignment with one another; however, you may not directly share code or write up with other students.
- On team assignments, you may not directly share code or write up with another team. Unauthorized sharing of the code or write up will be considered a violation for all students involved.

# Where to find help

- **If you have a question during lecture or lab, feel free to ask it!** There are likely other students with the same question, so by asking you will create a learning opportunity for everyone.
- **Office Hours**: A lot of questions are most effectively answered in-person, so office hours are a valuable resource. Please use them!
- **Piazza**: Outside of class and office hours, any general questions about course content or assignments should be posted on Piazza since there are likely other students with the same questions.

# Academic Resource Center

Sometimes you may need help with the class that is beyond what can be provided by the teaching team. In that instance, I encourage you to visit the Academic Resource Center.

The Academic Resource Center (ARC) offers free services to all students during their undergraduate careers at Duke. Services include Learning Consultations, Peer Tutoring and Study Groups, ADHD/LD Coaching, Outreach Workshops, and more. Because learning is a process unique to every individual, they work with each student to discover and develop their own academic strategy for success at Duke. Contact the ARC to schedule an appointment. Undergraduates in any year, studying any discipline can benefit! Contact ARC@duke.edu, 919-684-5917, 211 Academic Advising Center Building, East Campus – behind Marketplace.

# Technology

- You should bring a laptop to every lecture and lab session. Outlets are limited, so make sure it is fully-charged.
- Ensure the volume on all devices is set to mute.
- Refrain from engaging in activities not related to the class discussion. Browsing the web and social media, excessive messaging, playing games, etc. is not only a distraction for you but is also a distraction for everyone around you.

# Accessibility

Please contact the Student Disability Access Office (SDAO) if there is an element of the course that is not accessible to you. There you can engage in a confidential conversation about the process for requesting reasonable accommodations.

Please note that accommodations are not provided retroactively, so please contact them as soon as possible. More information can be found online at access.duke.edu.

# Inclusion

In this course, we will strive to create a learning environment that is welcoming to all students and that is in alignment with Duke's Commitment to Diversity and Inclusion. If there is any aspect of the class that is not welcoming or accessible to you, please let me know immediately.

Additionally, if you are experiencing something outside of class that is affecting your performance in the course, please feel free to talk with me and/or your academic dean.

# Questions?

# Announcements

- Fill out the **Getting To Know You Survey on Sakai** - due 9/3 at 11:59p
- My office hours start this week: Thu, 1p - 2:30p and by appointment
- TA office hours start next week
- **Please see me if you are on the waiting list**