

# Computing in Stat 100

Kelly McConville

Stat 100

Week 2 | Fall 2023

# Announcements

- Class in full swing:
  - **Sections**: Can find your assigned section in my.harvard but need to go to the linked spreadsheet to find the room!
  - **Office hours**
  - Wrap-ups on Th 3-4pm and Fri 10:30 - 11:30am in SC 309
  - Lecture quiz will be released in **Gradescope** after class on Wednesday.

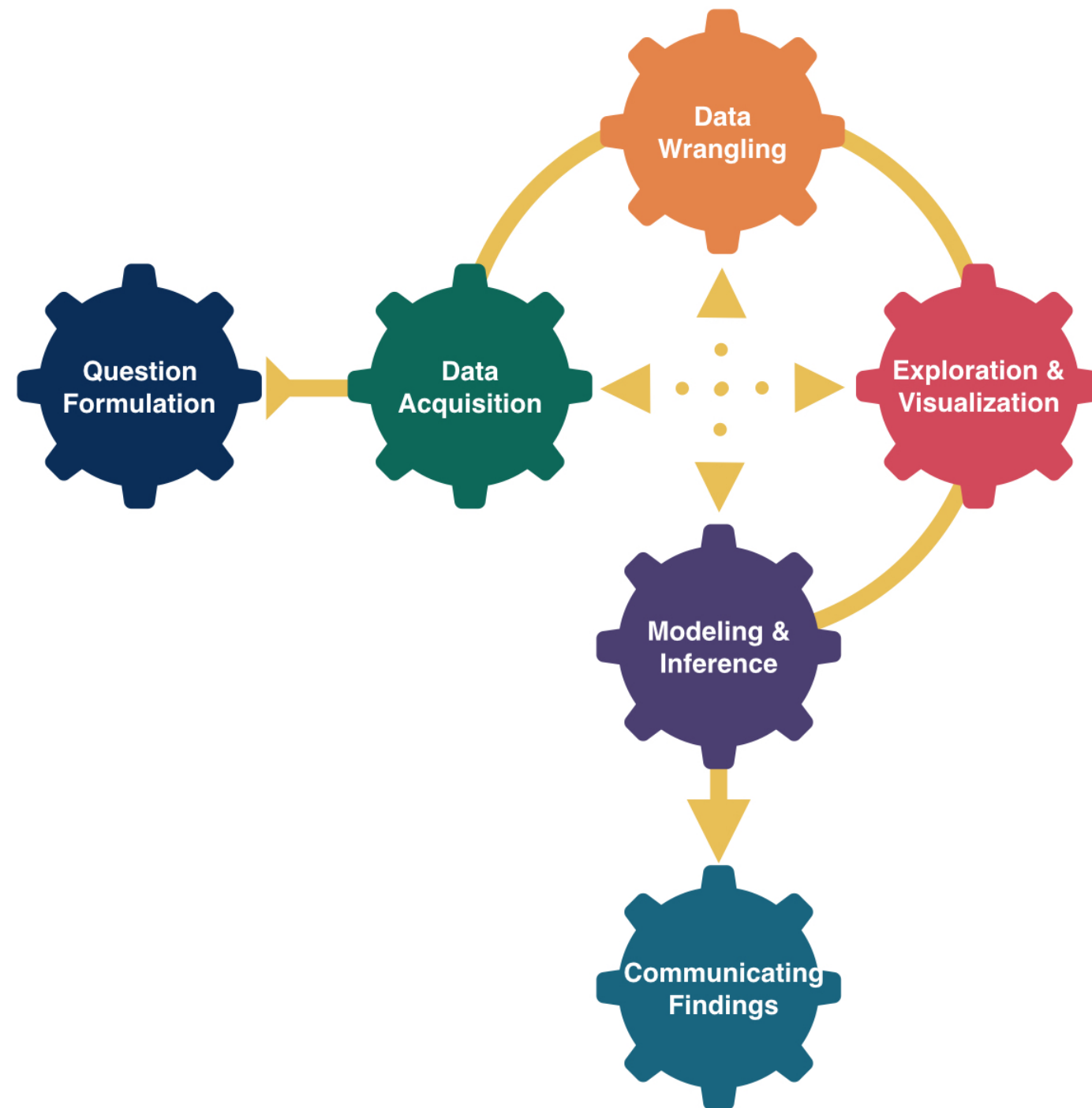


# Day 2 Goals

- Discuss course structure & goals (lecture, section, wrap-ups, office hours, assessments...)
- Present important course policies (engagement, code of conduct, chatGPT, ...)
- Get started in **RStudio** and with **Quarto** documents

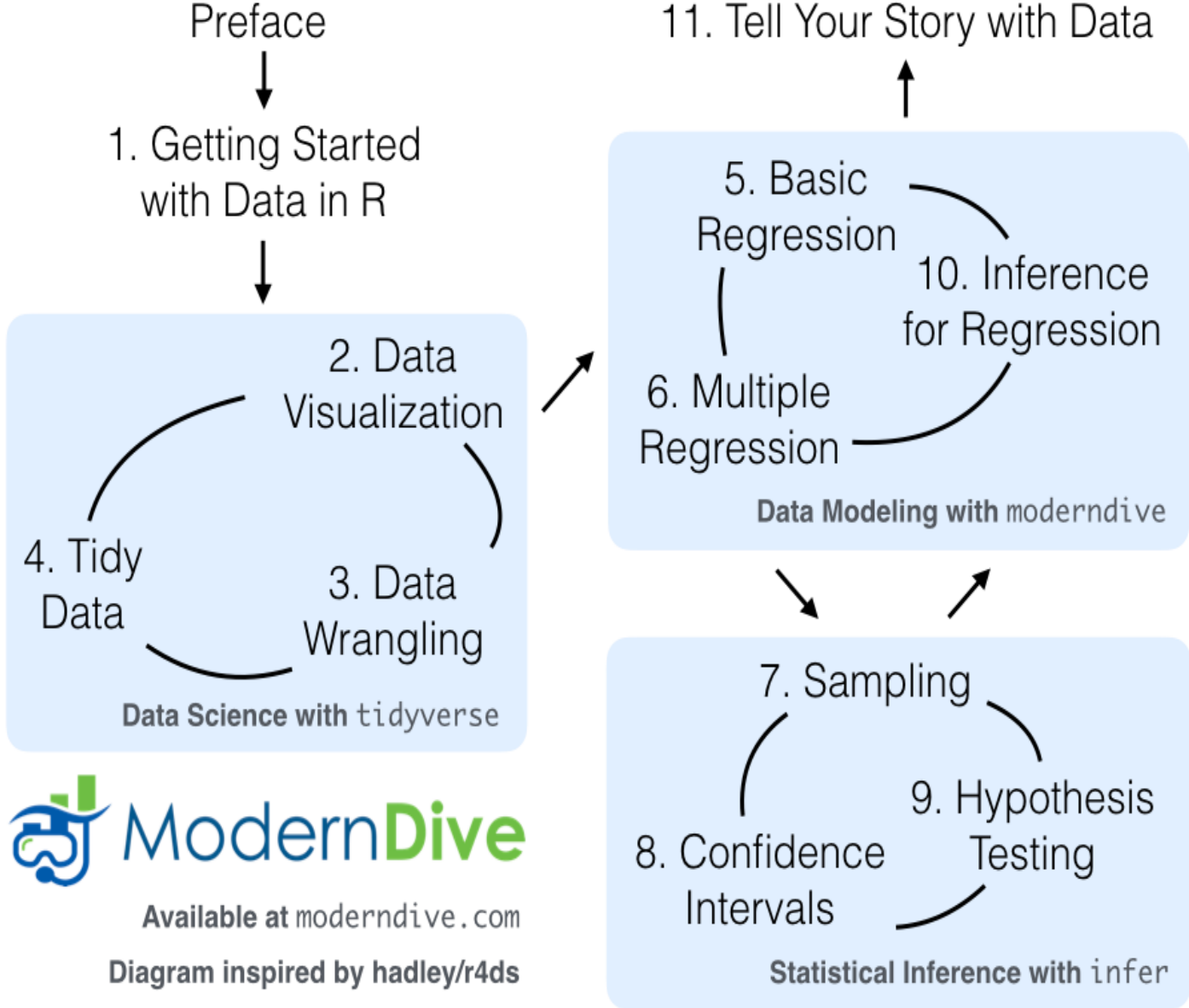
**Overarching Goal: Learn how to  
extract knowledge from data.**


# Course Learning Objectives

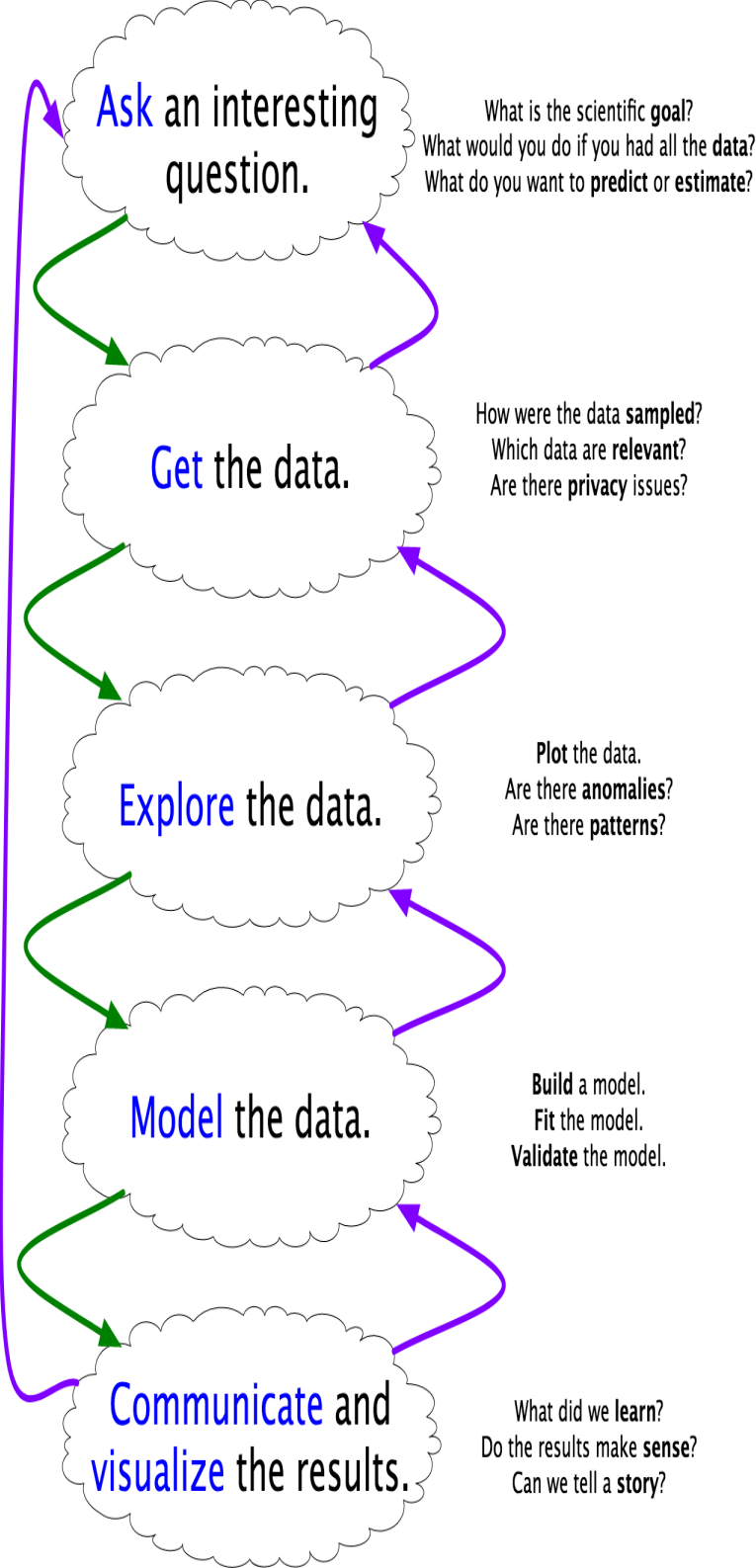


- Learn how to **analyze** data.
- Acquire good data **habits**.
- Develop **statistical thinking** and **problem-solving** skills.

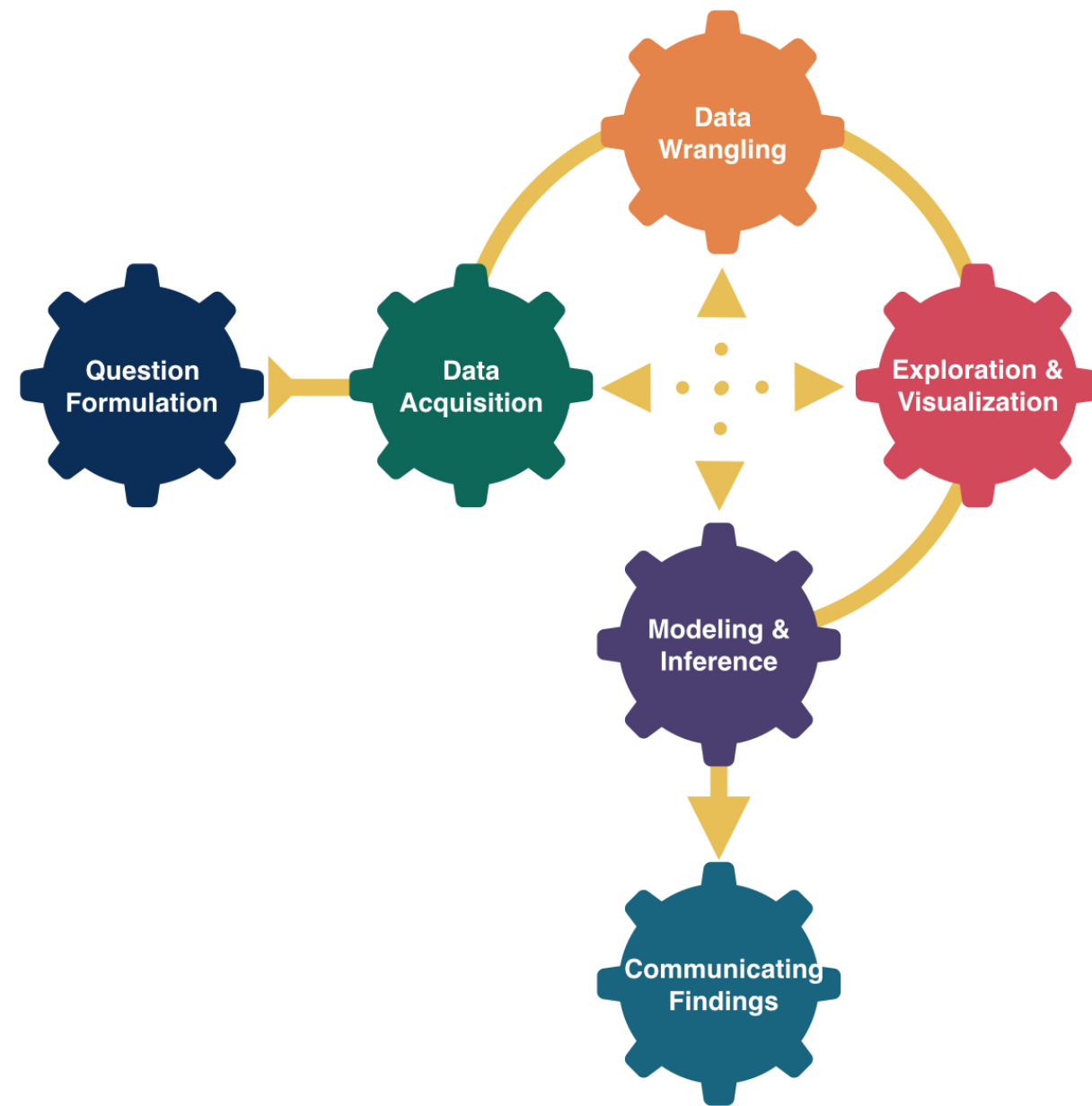
# Data Analysis Workflow



 **ModernDive**  
 Available at [moderndive.com](https://moderndive.com)  
 Diagram inspired by hadley/r4ds



# Workflow



- Not a strict set of guidelines.
- General structure for **extracting knowledge from data**.
- Consider **iteration, context, reproducibility, and ethics** throughout.

# Stat 100 Weekly Flow

Mon	Tues	Wed	Thurs	Fri
<b>10:30</b> —		<b>10:30</b> —		<b>10:30</b> —
<b>11:45</b> <b>Lecture</b>		<b>11:45</b> <b>Lecture</b>		<b>11:30</b> <b>Wrap-Up</b>
		<b>11:45</b> <b>P-Set and Quiz</b> <b>Released</b>		<b>11:45</b> <b>Quiz Due</b>
		<b>— 60 Minute Section —</b>		
			<b>3:00</b> —	
			<b>4:00</b> <b>Wrap-Up</b>	
	<b>5:00</b> <b>P-Set Due</b>			
	<b>----- Many Many Office Hours -----</b>			

# Forms of Assessment

- **Weekly lecture quizzes:**
  - Address important concepts covered.
  - Find out what is still confusing.
  - Can drop one quiz grade.
- **Exams:**
  - Format: In-class Exam & Oral Exam.
  - Will have both a Mid-term and Final.
- **Weekly problem sets:**
  - Practice concepts.
  - Time during section will be devoted to starting the next p-set.
  - Can drop one p-set grade and get 4 extension days.
- **Participation/Engagement:**
  - In lecture and section.
  - By Oct 9th, must:
    - Attend at least one **office hour**.
    - Post at least two messages on Slack.



# What Slack posts count?

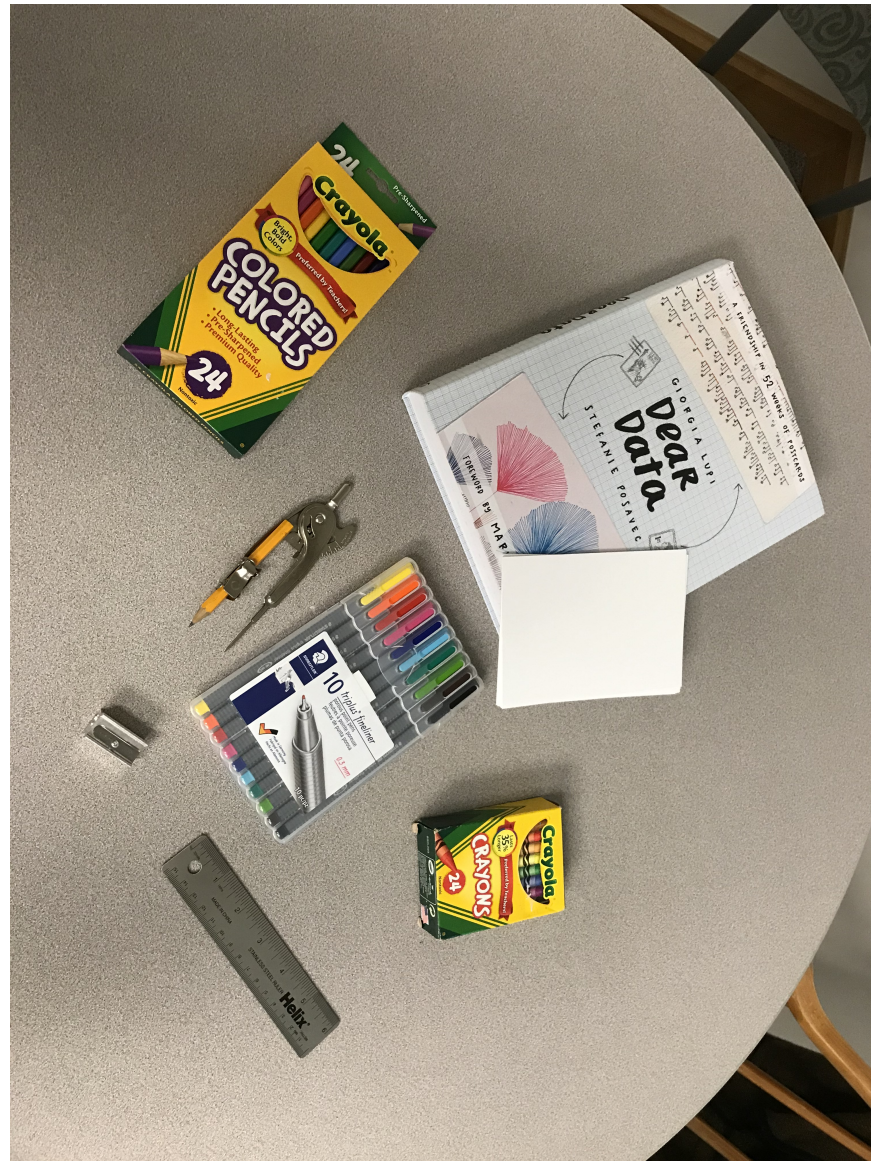
- Asking a question about course content
- Answering someone else's question
- Posting a useful resource and why you found it helpful
- Creating an example that illustrates a recent concept
- Sharing an article and talking about the stats discussed in the article
- Answering our start-of-term ice breaker!
- If you have never used Slack before, don't worry. We are here to help!



# This Week's Assessments

- Lecture quiz releases on Wed at 11:45am.
- **P-Set 1** will be posted to Posit Cloud on Wednesday.
  - Due the following Tuesday, September 19th at 5pm on Gradescope.
  - In section this week, you will get started on P-Set 1 and will learn how to submit to Gradescope.

# Problem 1 of Problem Set 1: Create your own Dear Data postcard!



- **Step 1:** Collect data on some aspect of your life.
- **Step 2:** Find a story in your data and determine your postcard recipient.
- **Step 3:** Figure out how you want to visualize the story.
- **Step 4+:** Visualize your data on a blank postcard.
  - Will receive in Section.

# And, now let's talk about ChatGPT...

I will refer to ChatGPT but I generally mean the various generative AI tools.



# Professor McConville in Spring 2023





# Professor McConville in Summer 2023



Ricardo IV Tamayo

## Generative AI Engagements

- Participated in Stats Dept reading group on generative AI
- Went to conference talks on generative AI
- Talked to ChatGPT

# ChatGPT Strengths

- Very dangerous **automated** data analysis.
  - Hard to determine if its conclusions are based on the data at hand or the data the large language model was built on.
- A **personalized** tutor (especially for coding) that tells you the answers.
- Instantly generating written blurbs.

# Stat 100 Goals

- Learn that data analysis is a **humanistic** endeavor where context drives the data analysis process.
  - Still need to be careful about biases and data quality.
- For you to **learn** to code and to **think** statistically.
  - The Stat 100 Teaching Team will help support that learning.
- For you to learn how to communicate about your data work and data ethics.

# My View on the (Current) Role of ChatGPT in Data Work

- To generate code to realize **fully conceived** ideas that aren't novel or advanced.
- **Seasoned data scientist** can/should:
  - Verify the correctness of the code.
  - Identify any assumptions or defaults ChatGPT is employing.
  - Run the code in **R** and draw conclusions based on the data to reduce risk of perpetuating biases that exist in ChatGPT's training data.
- So, ChatGPT might be useful to you **after** Stat 100 but will be detrimental to your learning **during** Stat 100.



# AI Policy for Stat 100

While A.I. tools, such as ChatGPT, are being used to generate code and analyze data, goals of this course are to develop your own ability to write code and to thoughtfully extract knowledge from data. Therefore, we expect that all work students submit for this course will be their own. This includes code, written work, and oral assessments. We specifically forbid the use of ChatGPT or any other generative artificial intelligence (AI) tools at all stages of the work process, including preliminary ones, unless the assignment specifically states that it is allowed. Violations of this policy will be considered academic misconduct. We draw your attention to the fact that different classes at Harvard could implement different AI policies, and it is the student's responsibility to conform to expectations for each course.

The goal of this policy is not to lessen your access to support but to ensure we are achieving the learning objectives of the course. Please see the [Avenues for Help](#) section of the syllabus for ways to get support in Stat 100.

# Engagement



- Being **actively present** is key.

- During lecture and section, remove distractions.
  - When we are on our computers, close email, social media, news, etc.
  - If you will be using a computer/tablet for taking notes, please sit outside the technology-free zone (first 6 rows of the middle section) **starting next lecture** so as not to distract classmates.
  - Hide your phone.
- I have high expectations but know that all of you (regardless of your stats or computing background) have the ability to meet them.

# Identity and the Classroom

- Acknowledge that my perspectives and experiences have shaped how I teach this course and how I approach my data work
- Some of my identities place me in dominant groups while others in marginalized groups
- Strive to bring examples and scholarly contributions that value knowledge from folks with a wide variety of identities
- Strive to be a open listener and recognize your thoughts as a generous offer and a vote of confidence in my ability to hear and be transformed by you
- Ask that you reflect on your own identities, privileges, and power and how they impact your engagement with Stat 100



# Code of Conduct

We expect all members of STAT 100 to make participation a harassment-free experience for everyone, regardless of age, body size, visible or invisible disability, ethnicity, sex characteristics, gender identity and expression, level of experience, education, socio-economic status, nationality, personal appearance, race, religion, or sexual identity and orientation.

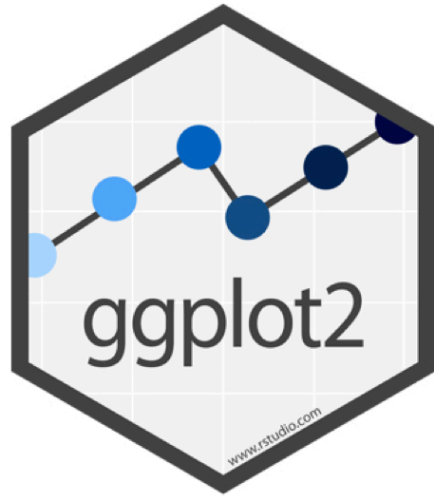
We expect everyone to act and interact in ways that contribute to an open, welcoming diverse, inclusive, and healthy community of learners. You can contribute to a positive learning environment by demonstrating empathy and kindness, being respectful of differing viewpoints and experiences, and giving and gracefully accepting constructive feedback.

This Code of Conduct is adapted from the [Contributor Covenant](#), version 2.0.

# Ways to Get Support

- **Attend, attend, attend** when you don't feel ill
- **Participate** in the Stat 100 Slack Workspace
- **Come** to Office Hours and Wrap-up Sessions
- **Create** study groups with your classmates
  - Don't know anyone in the course?
    - Use Slack or Section to each to know other students in the course!
    - Fill out this [Stat 100 study group matching form](#)

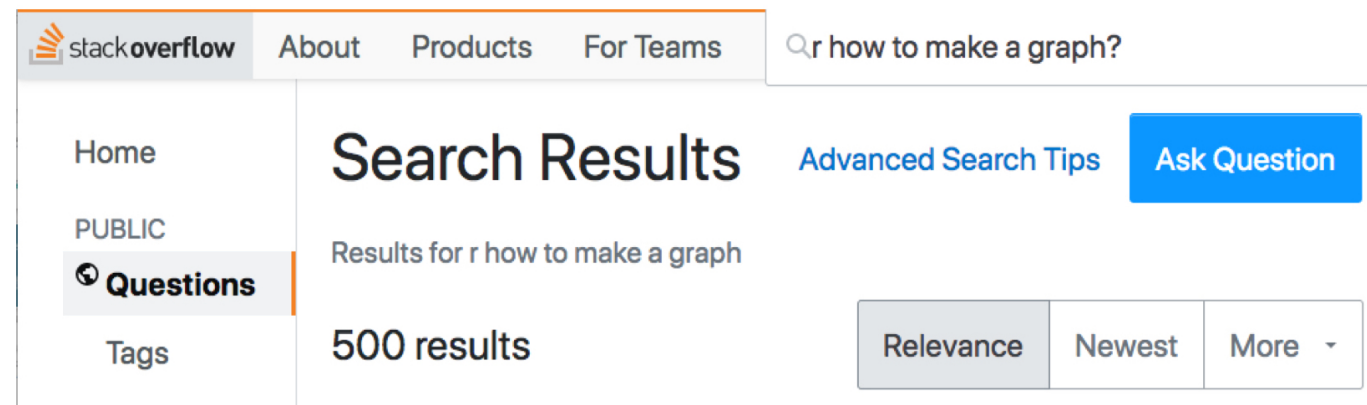
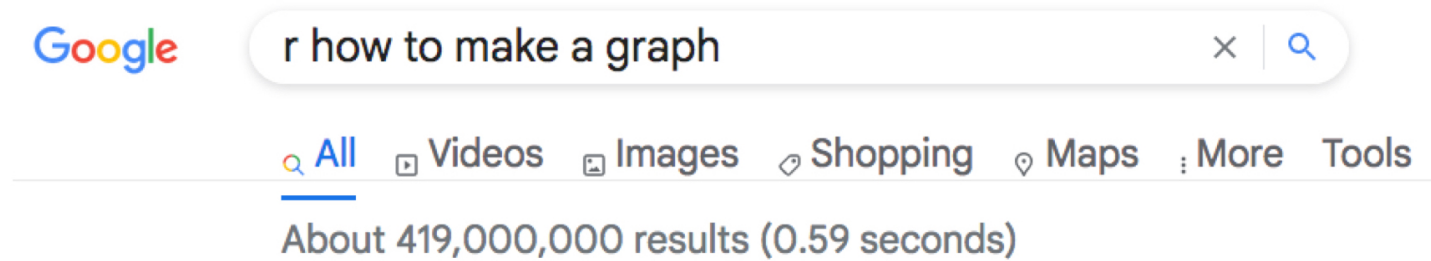
# Computation in Stat 100



Studio<sup>®</sup>



# Getting Help with R



# Explore

Settings



#rstats

Top

Latest

People



**Dirk Eddelbuettel** @ed  
Lecture slides for my 'C  
Programming Methods'



- Novices asking the internet for R help = 😓
- **Get help from the Stat 100 teaching staff or classmates!**
  - Will start p-sets in section each week.
  - Use the Slack **#q-and-a** channel.
- Get help **early** before 😡 sets in!
  - Be prepared for missing commas and quotes, capitalization issues, etc...
- Later in the semester, will learn tricks for effectively getting R help online.



# Language: **R**-This, **R**-That, **Q**-What?



**R** is the name of the programming language.



**RStudio** is the pretty interface and is hosted on a Posit Cloud Server.



**Quarto** is the type of file where we will record all of our work (code, output, narrative).

**Let's Get Started with R! Hop into  
our Posit Cloud site now.**

# Main Components of RStudio Lay-Out



# Main Components of RStudio Lay-Out

## Console

- Sideways caret called **prompt**.
- Where you run code.
- Let's try it:

```
1 6 * 2 / 3  
[1] 4
```

## Files et. al.

- **Files**: Accesses files in that project.
- **Plots**: Contains graphs we create.
- **Packages**: Installs and loads packages.
- **Help**: Displays help files.

## Environment

- Lists items stored in your session.
- Will add some items soon!

**What questions do we have so far?**

**Let's get started on the  
introductory Quarto document.**

# Reminders

- Class in full swing:
  - **Sections**: Can find your assigned section in my.harvard but need to go to the linked spreadsheet to find the room!
  - **Office hours**
  - Wrap-ups on Th 3-4pm and Fri 10:30 - 11:30am in SC 309
  - Lecture quiz will be released in **Gradescope** after class on Wednesday.