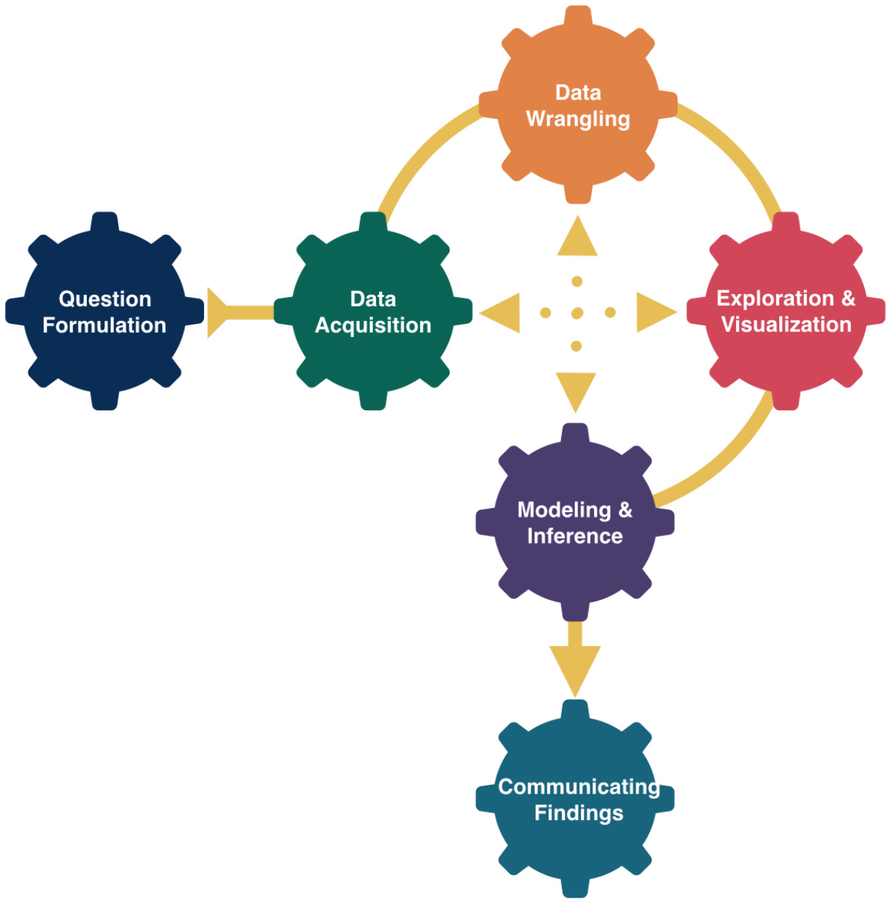


Computing in DATA 100

Kelly McConville

DATA 100

Week 2 | Spring 2026

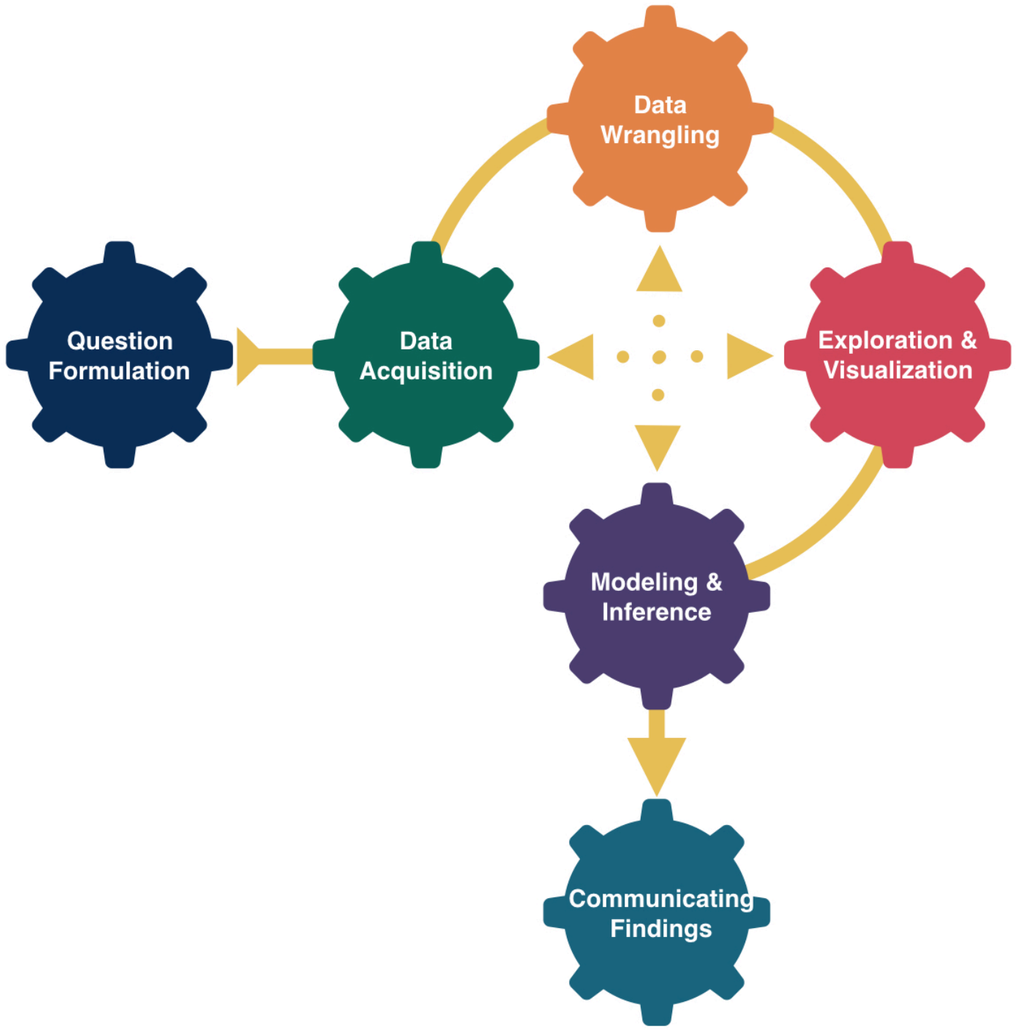
Announcements

- Class in full swing:
  - Office hours:
    - Mine: Tuesdays 2:30 - 4pm, Wednesdays 2-3pm, Taylor 212
    - Rebecca: Mondays and Wednesdays 4:30 - 6pm, Taylor 210
    - General Data Science Office Hours
  - The first problem set is due at noon on Thursday in Gradescope.

Day 2 Goals

- Discuss course structure & goals (lecture, lab, office hours, assessments…)
- Present important course policies (engagement, code of conduct, chatGPT, …)
- Get started in `RStudio` and with `Quarto` documents

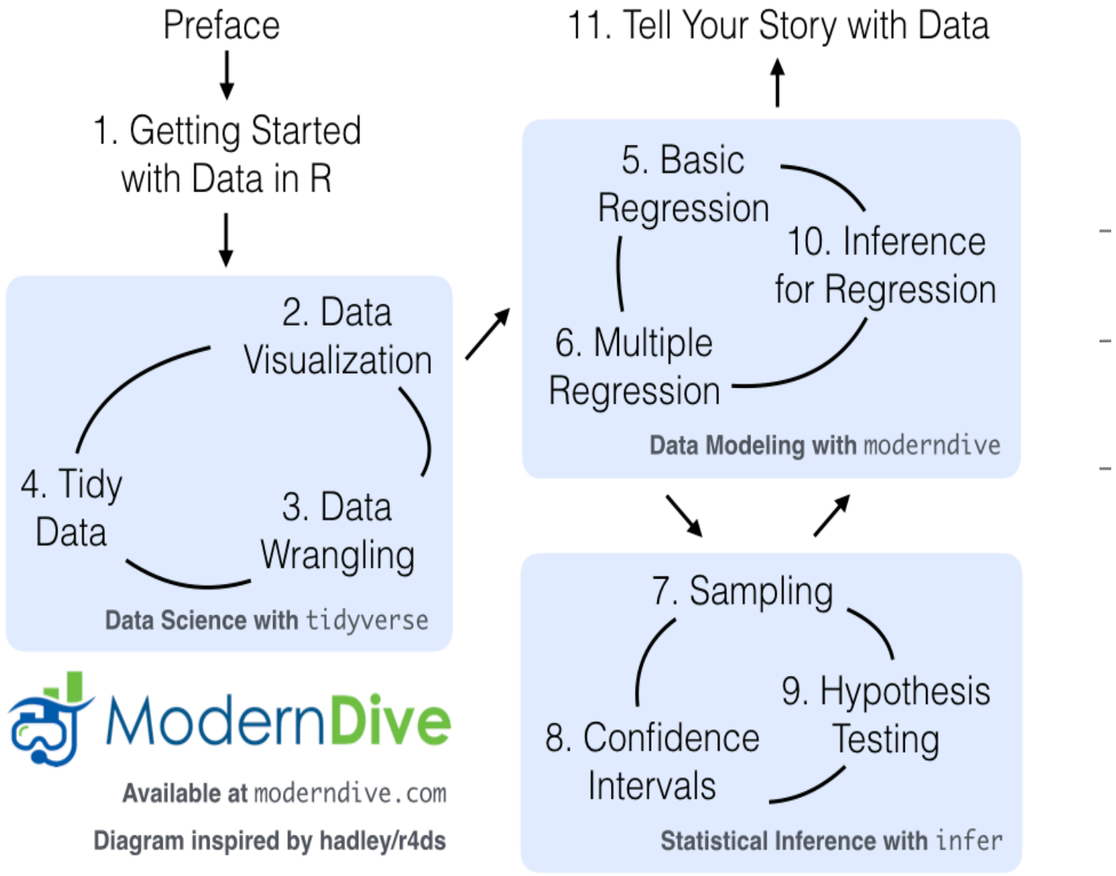
Overarching Goal: Learn how to extract knowledge from data.

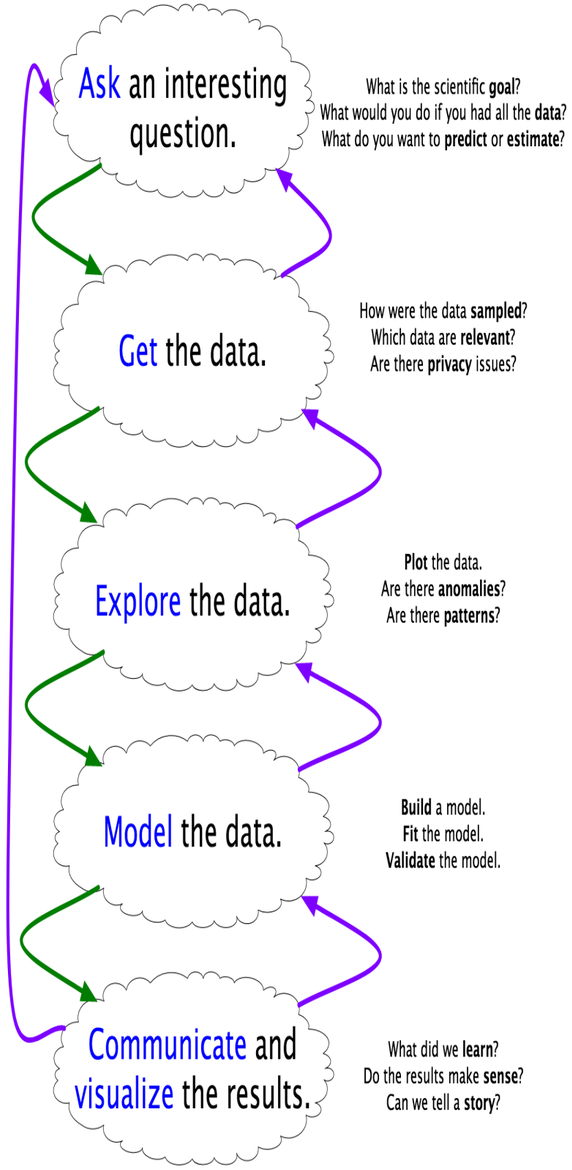


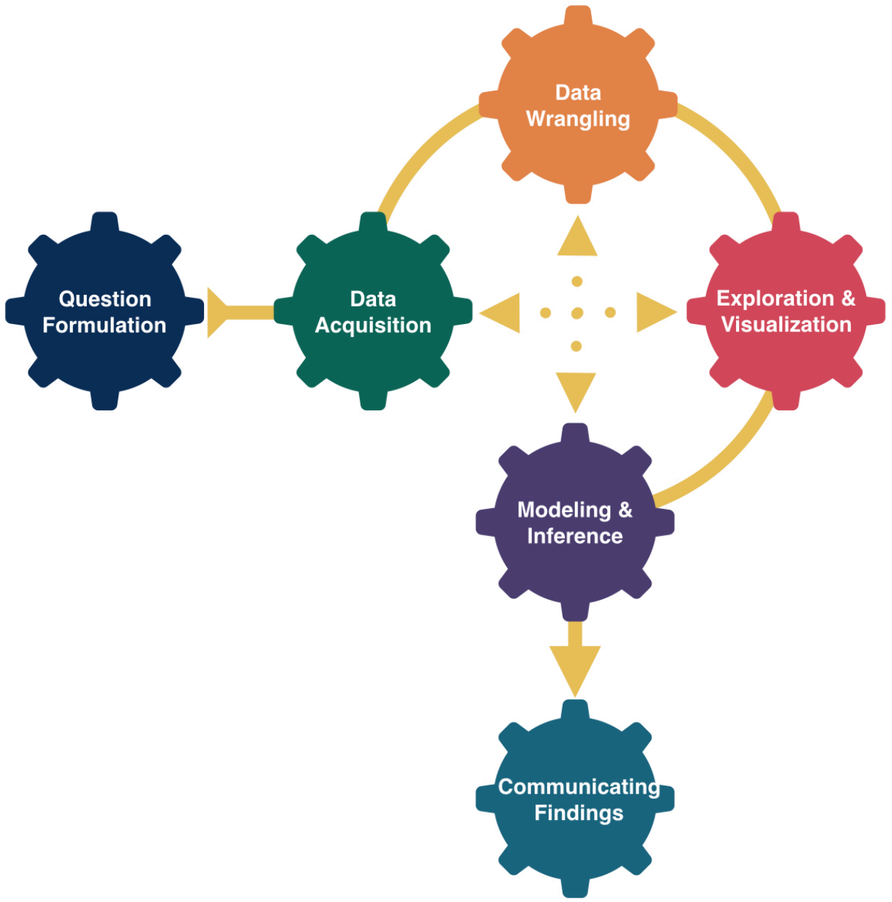
Course Learning Objectives

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

Data Analysis Workflow







Workflow

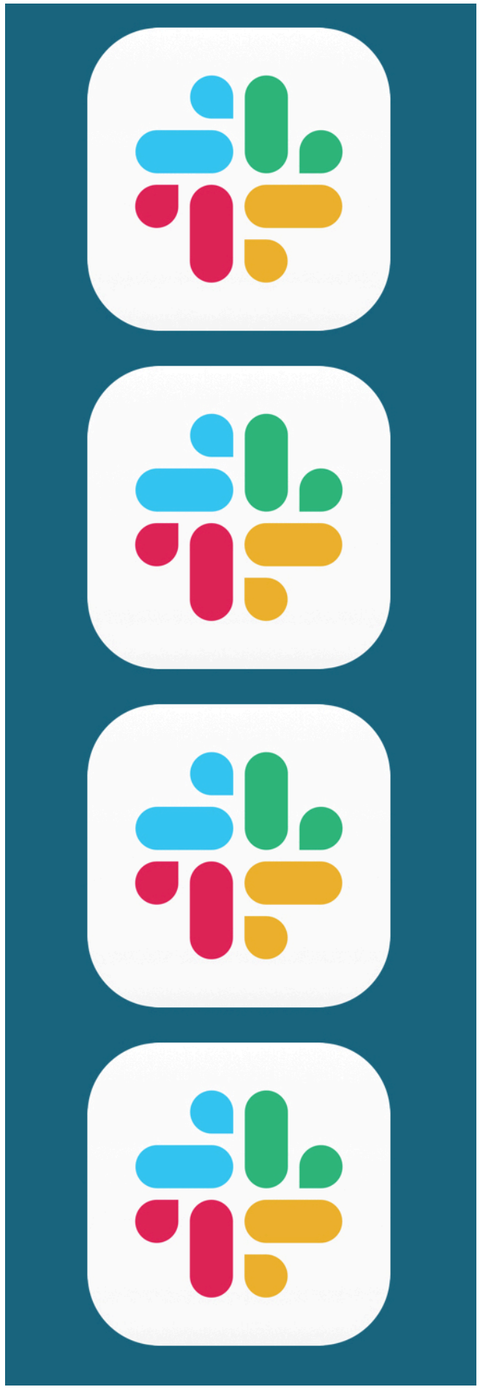
- Not a strict set of guidelines.
- General structure for extracting knowledge from data.
- Consider iteration, context, reproduciblity, and ethics throughout.

Forms of Assessment

- **Weekly quizzes:**
  - Address important concepts covered.
  - Find out what is still confusing.
  - Can drop one quiz grade.
- **Weekly problem sets:**
  - Practice concepts.
  - Time during lab will be devoted to starting the next p-set.
  - Can drop one p-set grade and get 4 extension days.
- **Exams:**
  - Format: In-class Exam & Oral Exam.
  - Will have two Mid-terms and Final.

Forms of Assessment

- **Participation/Engagement:**
  - In lecture and lab.
  - By March 6th, must:
    - Attend at least one office hour.
    - Post at least two messages on Slack.
  - By the end of the semester, must attend at least one data science event and reflect on the experience in this form



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

- Take the weekly quiz in lab.
- Turn in Problem Set 01 by noon on Thursday.
- Will receive Problem Set 02 in lab.

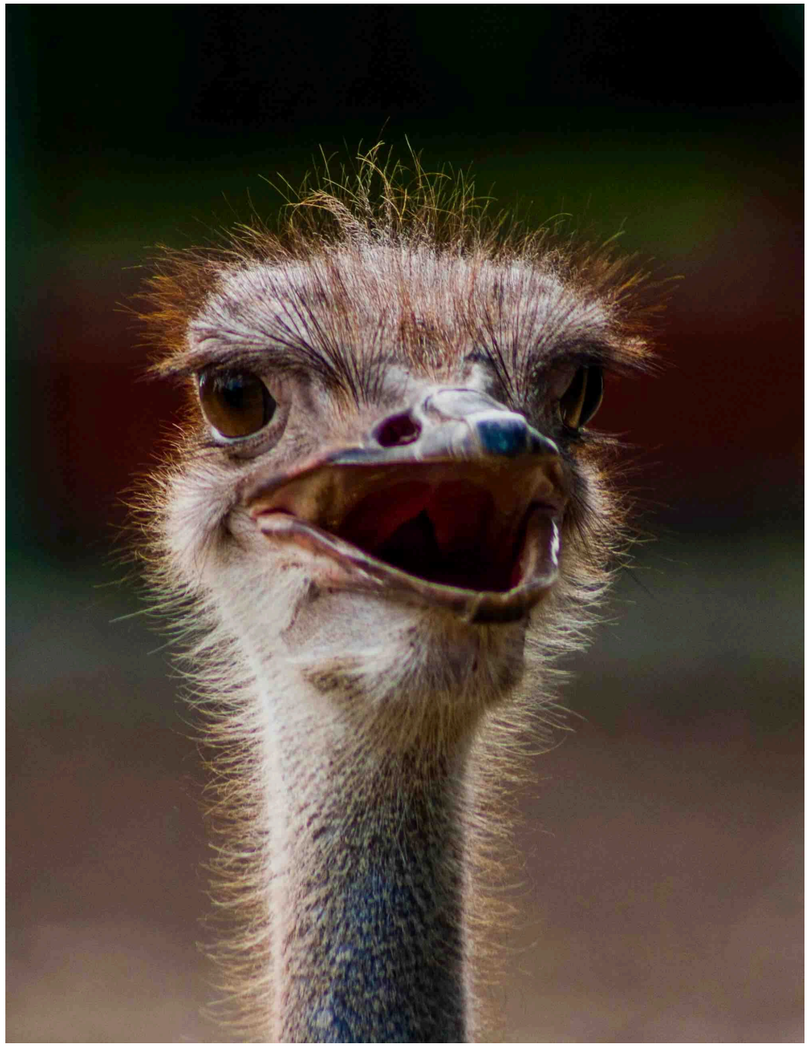
And, now let's talk about ChatGPT...

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

Professor McConville in Spring 2023

BBC Science Focus

Professor McConville after Spring 2023



Ricardo IV Tamayo

Generative AI Engagements

- Went to looooots of talks on generative AI
- Talked to ChatGPT, Gemini, and others
- Attended a workshop on data analysis with AI

ChatGPT Strengths

- Impressive but 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.

DATA 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 DATA 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 **near the end of** DATA 100 but will be detrimental to your learning **during most of** DATA 100.

AI Policy for DATA 100

We are going to take a two-pronged approach to AI in the course:

- (No AI) For the first 13 weeks of the course, you are not to use AI tools. Therefore, we expect that all work students submit during this period will be their own. This includes code, written work, and oral assessments. During this period, 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.
- (AI) Near the end of the course, we will address the question of how AI can help in data work. At that point in the semester, you are free to experiment with various AI tools on assignments.

In lecture and on individual assignments, we will make it clear which approach to AI students can take. If it is unclear, make sure to ask.

Engagement

- Being **actively present** is key.

- During lecture and lab, 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 in the second row **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.

Code of Conduct

> We expect all members of DATA 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.

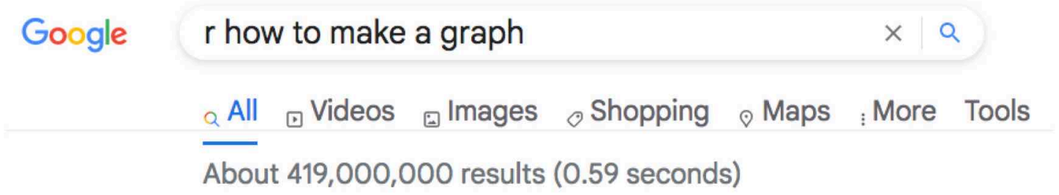
Ways to Get Support

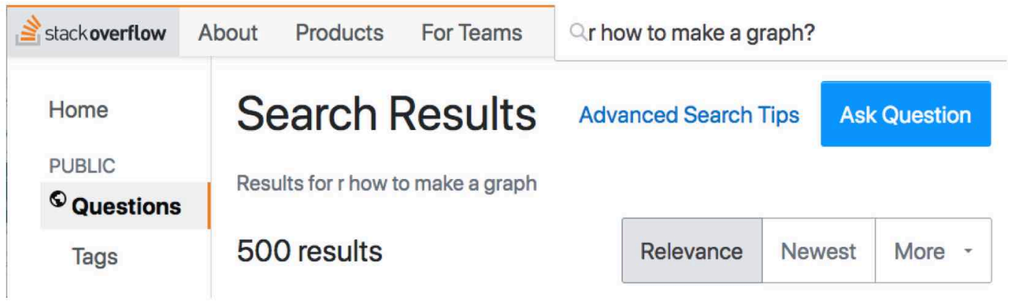
- **Attend, attend, attend** when you don’t feel ill
- **Participate** in the DATA 100 Slack Workspace
- **Come** to Office Hours
- **Create** study groups with your classmates
  - Don’t know anyone in the course?
    - Use Slack or lab to each to know other students in the course!

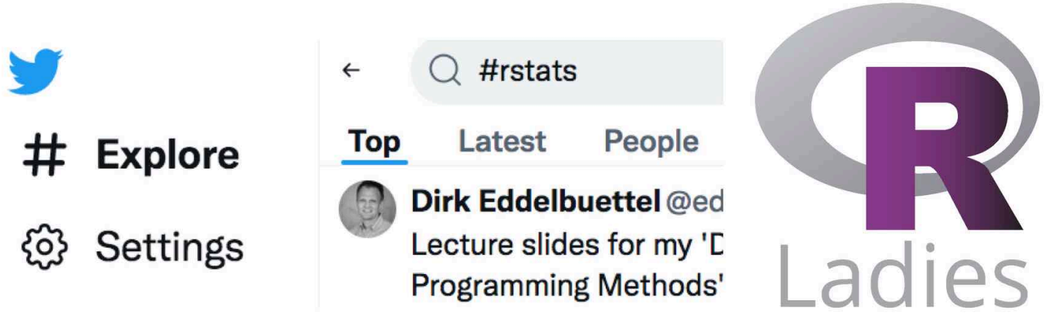
Computation in DATA 100


ggplot2
readr
tidyverse
reprex
xaringan
R
R Studio®
%>%
Leci n'est pas un pipe.
dplyr
quarto
forcats
stringr

Getting Help with R







- Novices asking the internet for R help = 😰
- **Get help from the DATA 100 teaching staff or classmates!**
  - Will start p-sets in lab 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?

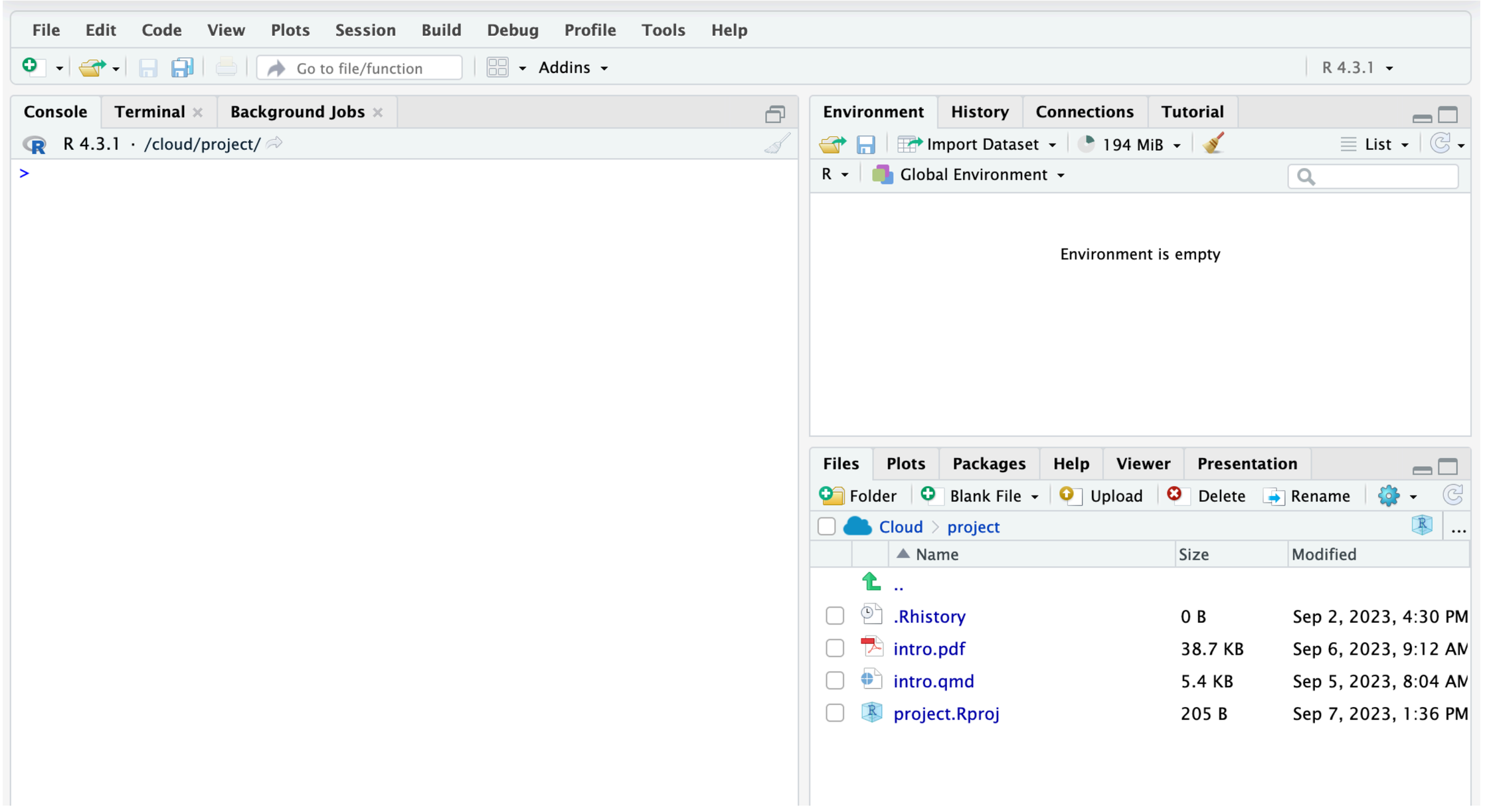
`R` is the name of the programming language.

`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


File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
R 4.3.1
Console
Terminal
Background Jobs
R 4.3.1 · /cloud/project/
>
Environment
History
Connections
Tutorial
Import Dataset
194 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Folder
Blank File
Upload
Delete
Rename
Cloud
project
Name
Size
Modified
..
.Rhistory 0 B Sep 2, 2023, 4:30 PM
intro.pdf 38.7 KB Sep 6, 2023, 9:12 AM
intro.qmd 5.4 KB Sep 5, 2023, 8:04 AM
project.Rproj 205 B Sep 7, 2023, 1:36 PM

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
```

Environment

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

Files et. al.

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

What questions do we have so far?

Let's get started on the introductory Quarto document.

Reminders

- Class in full swing:
  - Office hours:
    - Mine: Tuesdays 2:30 - 4pm, Wednesdays 2-3pm, Taylor 212
    - Rebecca: Mondays and Wednesdays 4:30 - 6pm, Taylor 210
    - General Data Science Office Hours
  - The first problem set is due at noon on Thursday in Gradescope.