Welcome to STA 199

Lecture 0

Chris Hallstrom

2024-01-11

Hello world!

Meet the prof

Dr. Mine Çetinkaya-Rundel

Professor of the Practice

Old Chem 213

Meet the course team

- Mary Knox (Course coordinator)
- Betsy Bersson (Head TA)
- Chris Oswald
- Jasmine Wang
- Mert Bildirici
- Will Tirone
- Krish Bansal
- Caitrin Murphy
- Avery Hodges
- Jon Campbell

- Foxx Hart
- Netra MittalAlexa Fahrer
- Devin Johnson
- Konnie Huang
- Li Fan
- Kelly Huang
- Leah Johnson
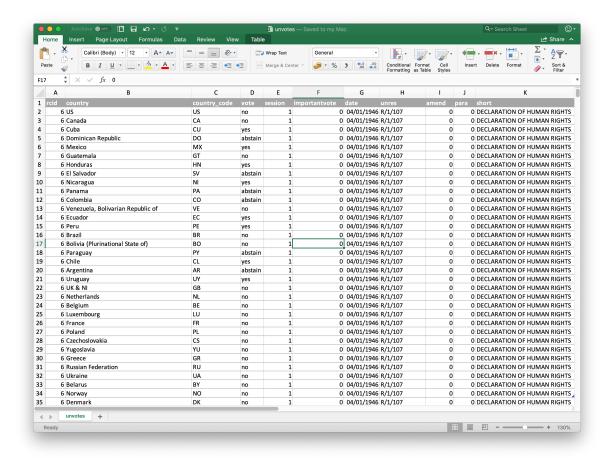- Lisa Zhang
- Miles King
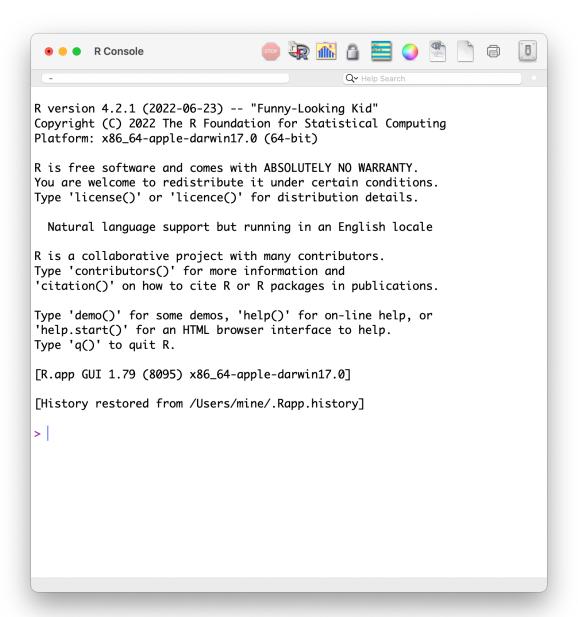- Noah Obuya

Meet each other!

Meet data science

- Data science is an exciting discipline that allows you to turn raw data into understanding, insight, and knowledge.

- We're going to learn to do this in a `tidy` way – more on that later!

- This is a course on introduction to data science, with an emphasis on statistical thinking.
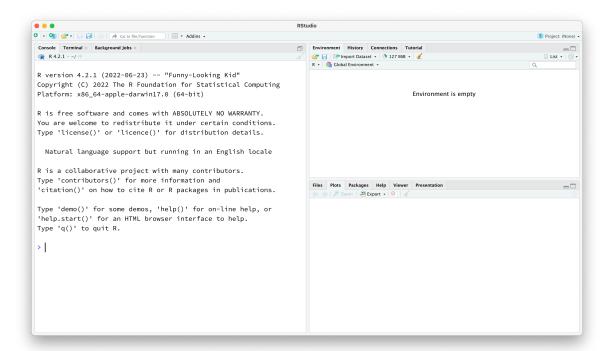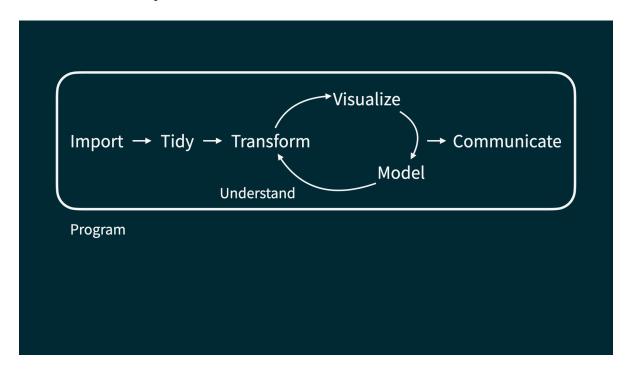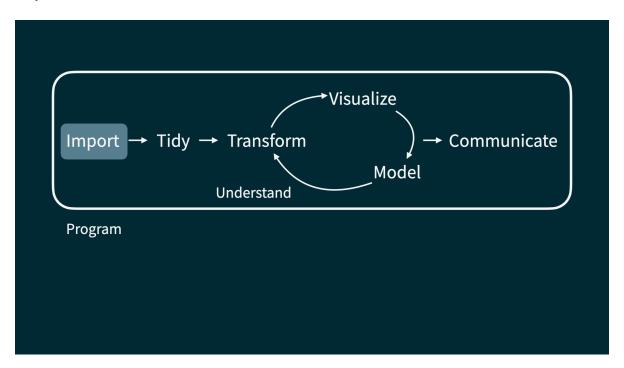
Software

Excel - not...

R

```
R version 4.2.1 (2022-06-23) -- "Funny-Looking Kid"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.79 (8095) x86_64-apple-darwin17.0]

[History restored from /Users/mine/.Rapp.history]

>
```

RStudio

Data science life cycle

Data science life cycle

Import

Tidy + transform

Visualize

Model

Understand

Google Trend Index

Trend index

Travel ▾

Date range

2012-01-01   to   2017-07-31

☑ **Overlay smooth trend line**

Smoother span:

0.05   1

0.01   0.11   0.21   0.31   0.41   0.51   0.61   0.71   0.81   0.91   1 ▶

Higher values give more smoothness.
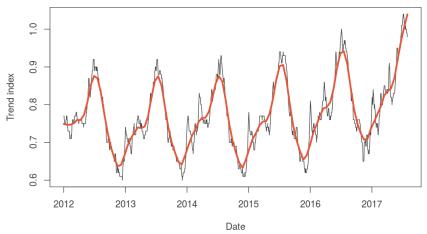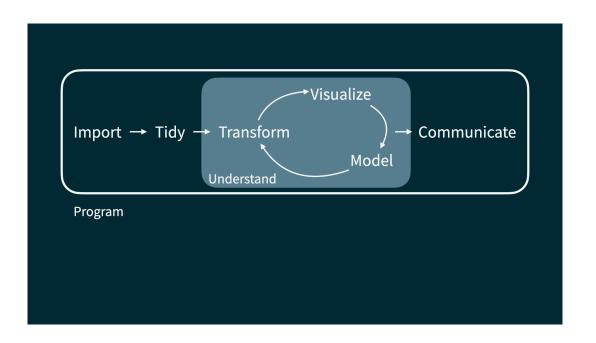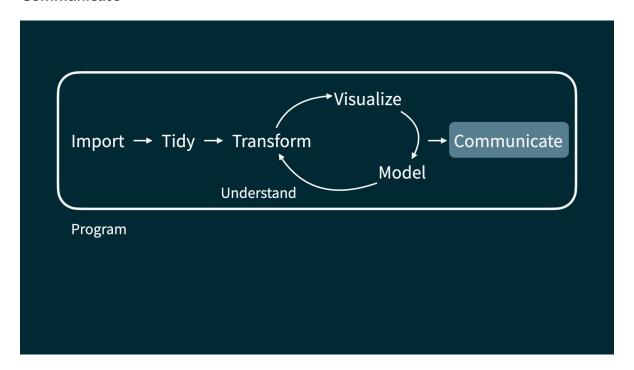

The Google Travel Index tracks queries related to airlines, hotels, beach, southwest, las vegas, flights, etc. The index is set to 1.0 on January 1, 2004 and is calculated only for US search traffic.

Source: Google Domestic Trends

```
# A tibble: 5 x 2
  date             season
  <chr>            <chr>
1 23 January 2017  winter
2 4 March 2017     spring
3 14 June 2017     summer
4 1 September 2017 fall
5 ...              ...
```
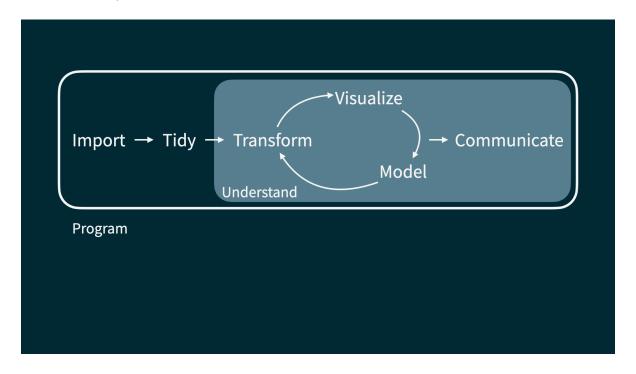
Communicate

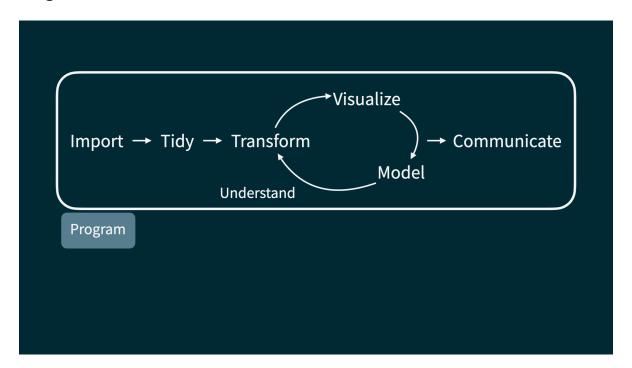**Understand + communicate**

Program

Let's dive in!

Application exercise

Or more like demo for today...

  github.com/sta199-s24/ae-00-unvotes

Course overview

Homepage

https://sta199-s24.github.io

- All course materials
- Links to Canvas, GitHub, RStudio containers, etc.

Course toolkit

All linked from the course website:

- GitHub organization: github.com/sta199-s24
- RStudio containers: cmgr.oit.duke.edu/containers
- Communication: Ed Discussion
- Assignment submission and feedback: Gradescope

Activities

- Introduce new content and prepare for lectures by watching the videos and completing the readings
- Attend and actively participate in lectures and labs, office hours, team meetings
- Practice applying statistical concepts and computing with application exercises during lecture, graded for completion
- Put together what you've learned to analyze real-world data

    - Lab assignments x 7
    - Exams x 2
    - Term project presented in the last lab session

Exams

- Two exams, each 20%

- Each exam comprised of two parts:

    - In class: 75 minute in-class exam. Closed book, one sheet of notes ("cheat sheet", no larger than 8 1/2 x 11, both sides, **must be prepared by you**) – 70% of the grade

    - Take home: The take home portion will follow from the in class exam and focus on the analysis of a dataset introduced in the take home exam – 30% of the grade

> 🔥 Caution
>
> Exam dates cannot be changed and no make-up exams will be given. If you can't take the exams on these dates, you should drop this class.

Project

- Dataset of your choice, method of your choice

- Teamwork

- Presentation and write-up

- Presentations in the last lab

- Interim deadlines, peer review on content, peer evaluation for team contribution

- Some lab sessions allocated to working on projects, doing peer review, getting feedback from TAs

> 🔥 Caution
>
> Final presentation date cannot be changed. If you can't present on that date, you should drop this class. You must complete the project to pass this class.

Teams

- Assigned by me
- Project
- Peer evaluation during teamwork and after completion
- Expectations and roles

    - Everyone is expected to contribute equal *effort*
    - Everyone is expected to understand *all* code turned in
    - Individual contribution evaluated by peer evaluation, commits, etc.

Grading

Category	Percentage
Labs	35%
Project	20%
Exam 1	20%
Exam 2	20%
Application Exercises	5%

No specific points allocated to attendance, but attendance will be recorded periodically throughout the semester, and this information will be used as "extra credit" if you're in between two grades and a minor bump would help.

See course syllabus for how the final letter grade will be determined.

Support

- Attend office hours
- Ask and answer questions on the discussion forum
- Reserve email for questions on personal matters and/or grades
- Read the course support page

Announcements

- Posted on Canvas (Announcements tool) and sent via email, be sure to check both regularly
- I'll assume that you've read an announcement by the next "business" day
- I'll (try my best to) send a weekly update announcement each Friday, outlining the plan for the following week and reminding you what you need to do to prepare, practice, and perform

Diversity + inclusion

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit.

- Please let me know your preferred name and pronouns on the Getting to know you survey.
- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me. I want to be a resource for you. If you prefer to speak with someone outside of the course, your advisors, and deans are excellent resources.
- I (like many people) am still in the process of learning about diverse perspectives and identities. If something was said in class (by anyone) that made you feel uncomfortable, please talk to me about it.

Accessibility

- The Student Disability Access Office (SDAO) is available to ensure that students are able to engage with their courses and related assignments.

- I am committed to making all course materials accessible and I'm always learning how to do this better. If any course component is not accessible to you in any way, please don't hesitate to let me know.

Course policies

Late work, waivers, regrades policy

- We have policies!
- Read about them on the course syllabus and refer back to them when you need it

Academic integrity

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.

Wrap up

This week's tasks

- Complete Lab 0

  – Computational setup
  – Getting to know you survey

- Read the syllabus
- Start readings for next week

Midori says...