Introduction to "Introduction to Data Science"

Erik Fredner

2024-08-26

Table of contents

1 Outline

1. Getting to know each other
2. Course overview
3. Syllabus
4. Install stuff
5. Next time

2 Getting to know each other

2.1 Introductions

1. Preferred name
2. Pronouns
3. Something you *dislike*

2.1.1 Example:

1. Professor Fredner
2. He/him
3. Gushers

2.2 Introduce *me* to UR

- This is my first semester at UR!
  - If you want to learn more about who I am: https://fredner.org
- How would you describe the culture of UR?
  - Discuss with the people next to you.

2.3 Anonymous survey

Please complete this brief **anonymous** survey: https://richmond.ca1.qualtrics.com/jfe/form/SV_8vsOr6mRZGypJVc

This will help me design the class.

3 Course Overview

3.1 What do we mean by "data science?"

- We will learn and practice a series of **methods for organizing, collecting, visualizing, manipulating, and exploring different kinds of data**.
- We focus on the creation of data and application of methods, not theoretical or foundational questions.
- This is not a mathematics course, nor will it resemble a traditional introductory statistics class. We will spend the entire semester writing code to apply data science concepts.

3.2 Why do data science?

- There is too much information in the world.
  - e.g., every minute, approximately 500 hours of video are uploaded to YouTube.
- People value *useful* information and new knowledge.
  - Almost none of those 500 hours are worth your finite time.
- Data science transforms data into useful information.

3.3 How do people do data science?

- **R** and **Python** are the two most popular programming languages for data science.
- We will be using R in this class.
- However, the main learning goal of this class **is not R**.
- The main learning goal is to understand:
  - what good data is
  - how to ask good questions of data
  - how we can use good data to answer good questions

3.4 An effective example

https://www.youtube.com/watch?v=Z8t4k0Q8e8Y

4 Syllabus

- Go to the course Blackboard site and click on Syllabus.

4.1 Questions about syllabus?

5 Install stuff

5.1 Step 1: R

Install the version for your operating system: https://cloud.r-project.org

If you already have R installed, install the newest version.

If you have a Mac that was made in 2020 or later, choose "Apple Silicon."

5.2 Step 2: RStudio

Install the free version of RStudio Desktop: https://posit.co/downloads/

If you already have RStudio installed, install the newest version.

5.3 Step 3: Test it

1. Open RStudio.
2. Type the following in the Console:

```
print("Hello, world!")
```

3. Press Enter

This should return:

```
[1] "Hello, world!"
```

5.4 Step 4: Install the `tidyverse`

Type the following in the console:

```
install.packages("tidyverse")
```

A bunch of things should happen.

You only need to let me know if you get an `Error`.

5.5 Step 4.5: Debugging

- If you ran into a problem during the installation, raise your hand.
- If you finished installing everything successfully and think you can help others, go to the closest person with their hand up.

5.6 Step 5: Download materials

- Go to Blackboard / Course Documents / 00 - Intro
- Download `DSST289.zip`
- Extract this `.zip` file
- Move it to a location you can find easily (e.g., `Desktop`)

6 Next time

1. Read `notes01.Rmd`. (All notes will be numbered sequentially.)
2. Complete the questions at the end of `notes01.Rmd`.
3. Complete the questions in `notebook01.Rmd`.
4. Upload your completed `notebook01.Rmd` to Blackboard before class.

Time permitting, you can start on this now.