Introduction to Data Science

Welcome!

Note: I have been asked to have a fixed seating chart for the semester. So, pick your seat today carefully!

<u>Introduction to Data Science</u>

Today we are going to get all of the administrative details dealt with.
Here is a quick outline:

   1. syllabus
   2. course content
   3. introductions
   4. install course materials

There should be plenty of time for questions throughout the class.

1. Syllabus

Questions & Office Hours

We will usually have a lot of time in class to answer any questions you have about the course material. I am usually around before and after class for additional questions.

Please also feel free to send questions by email. I typically respond within 24 hours.

Finally, I am of course happy to schedule an office hours meeting for extended questions or personal concerns. Just send me an email with your availability at least 1 day before you'd like to meet.

Coffee chat?

I get a lot of requests for longer discussions about graduate school, careers in data science, or research projects.

I love having these conversations with students. Just send me an email or ask after class and we can find a time to grab coffee/tea/whatever and answer any questions you have.

Masks and Such

As I mentioned in my first email to the class, policies regarding COVID-19 are constantly changing. In fact, they already have!

Following the UR guidelines, for the first three weeks we will wear masks while in class. We can re-assess the situation after that point.

Faculty are also asked to create a seating chart (kindergarten-style, see image), which we will do right now.

2. Course Content

Data Science?

This semester we will learn and practice a series of methods for **organizing**, **collecting**, **visualizing**, **manipulating**, and **exploring** different kinds of data. We are focused on the creation and application of **methods**, rather than 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.

Programming

There are several different programming languages for data science. By far the two most popular are R and Python.

We will be using R this semester but will learn a version that is easily adapted to other languages such as Python.

No specific experience with R or Python is expected!

An Example

If you are not familiar with the kinds of tasks that are common in data science and exploratory data analysis, an example can be very helpful. Here is a slightly dated but still the best concise example I know of:

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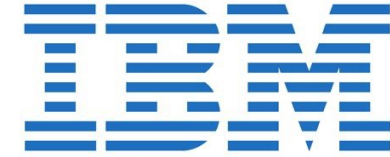
3. Introductions

About Me

- From New England: born in Maine, school in MA, ME, CT
- Moved to Richmond in 2016
- Research on large text and image datasets in linguistics and cultural studies

About Me

- Lots of industry experience in DS:
    - IBM (Healthcare)
    - Travelers (Insurance)
    - DARPA (social media)
    - AT&T (location analytics)
    - Telperian (pharmaceuticals)

About Me

- I have two Shih-Tzus: Roux and Sargent
- Roux is often in my office; please come say hello

4. Course Setup

[see other slides]