CICS 197R Syllabus: Spring 2022

Course Details

- CICS 197R: Special Topics - Introduction to Data Analysis in R
- 1 credit. No prerequisites. Pass/fail grading for undergraduates
- Meets Tuesdays and Thursdays from 2:30 to 3:45 in Lederle GRT room 177
- Twelve sessions, January 25 to Mar 8 (not Feb 22)

An introduction to data analysis in the open-source R language, with an emphasis on practical data work. Topics will include data wrangling, summary statistics, modeling, and visualization. Will also cover fundamental programming concepts including data types, functions, flow of control, and good programming practices. Intended for a broad range of students outside of computer science. Some familiarity with statistics is expected.

Instructor

Jasper McChesney - jmcchesney@cs.umass.edu

Regarding email: Please include "197R" in the subject line of all emails. I typically will not answer messages outside of normal business hours, and will not respond instantly even during them.

Learning Objectives

After completing this course, I expect you will have a solid foundation to pursue practical data work in R. You will be able to load, clean, and reshape data; calculate summary statistics and create statistical models; and design useful visualizations. You will be able to discuss programming concepts such as data types and structures, functions, and flow of control. You will be able to write basic functions to perform data work, following good coding practices. Most importantly, you will know enough to be able to continue your education in R, programming, and data analysis on your own, in whatever discipline you work.

Diversity and Accomodations

I intend this course to be a welcoming environment for all kinds of students to learn programming and data analysis, regardless of previous programming experience, academic background, or personal characteristics.

If you have a documented disability on file with Disability Services, you may be eligible for reasonable accommodations in this course; please notify me as early as possible so we can make the proper arrangements.

Format and Requirements

We will meet in-person 12 times, in 75-minute sessions. These will include lecture and in-class exercises, potentially in small groups. I expect you to generally attend, though I will not take attendance. Bring a laptop in case we are coding in class.
Homework

After each class there will be problem sets for you to attempt. These are not graded, and instead intended for you to develop skills by doing what you've seen in the lectures. Answer keys are provided for each: you're encouraged to look at the answer to each question after you believe you've addressed it. We'll discuss these problem sets in class.

Video lectures may also be assigned on some days. These are of a similar format to normal lecture, and problem sets may relate to them. Typically each is less than 10 minutes long.

Quizzes

I may deploy short quizzes at my discretion. These will primarily assess your basic skills, and will help you know if you're keeping up.

Final Project

A final "portfolio" project will let you demonstrate some of the skills and approaches you've learned. You will have some choice about what it addresses: a topic or dataset of your choosing, research you're involved in elsewhere, or a project I provide (recommended for most undergraduates).

Basis of Grading

For undergraduates, the course has a mandatory pass/fail grading basis. Graduate students will receive a normal letter grade unless auditing.

Half the grade determination will hinge on the final project. It does not have to be perfect to earn a pass (or A), but must show good understanding of each major area we've covered during the course, command of the R syntax, an ability to write functions and scripts, and the ability to think about data tasks through coding.

Quiz attempts and in-class participation will count equally toward the other half of the grade.

Collaboration and Academic Honesty

Homework may be completed with help from class-mates, the internet, or elsewhere -- though I suggest attempting it on your own first to get the most out of it. The final project should be done on your own, but you are free to consult general R references as needed. Quizzes are of course done without aid, except for any help sections available inside RStudio itself.

Illness and Quarantines

It may of course occur that you or I will need to quarantine, or otherwise need to miss class. In the case that I cannot physically run the class, I will email you as soon as possible, and we will switch to a Zoom call instead, with the link provided by email and on the Moodle page.

In the event you are individually not able to attend, you should be able to make up the key material remotely: the Moodle page will have videos or book chapters that cover roughly the same material as in class; and I also recommend you find a friend or two in class to get notes from. One missed quiz will be entirely discounted from the grade. You don't need to contact me if you're out a single day, but please do so if you miss more than
one, and we will make arrangements for work to replace quizzes. (I'll presume there's a good reason. If you were just hung over, don't lie...just don't tell me the reason!)

**Materials**

**Course Sites**

Most materials, including video lectures and assignments, will be posted to Moodle.

The course web page provides certain overview documents: [https://people.cs.umass.edu/~jmcchesney/197R/](https://people.cs.umass.edu/~jmcchesney/197R/)

**Software**

You will need to install RStudio (available for Windows, Mac, and Linux): [https://rstudio.com/products/rstudio/download/#download](https://rstudio.com/products/rstudio/download/#download)

**Schedule of Topics**

The first third of the course will be devoted to more basic programming skills, while the middle has us looking at true more realistic datasets, and the final third has us diving back into some details to address realistic analysis questions.

<table>
<thead>
<tr>
<th>#</th>
<th>Date</th>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/25</td>
<td>Introduction, Variables, Vectors, Functions</td>
</tr>
<tr>
<td>2</td>
<td>1/27</td>
<td>Indexing and Logical Comparisons</td>
</tr>
<tr>
<td>3</td>
<td>2/1</td>
<td>Ordering and Sorting</td>
</tr>
<tr>
<td>4</td>
<td>2/3</td>
<td>Tabular Data</td>
</tr>
<tr>
<td>5</td>
<td>2/8</td>
<td>Summarizing Data Frames</td>
</tr>
<tr>
<td>6</td>
<td>2/10</td>
<td>Time Series</td>
</tr>
<tr>
<td>7</td>
<td>2/15</td>
<td>Working With Text</td>
</tr>
<tr>
<td>8</td>
<td>2/17</td>
<td>Linear Models</td>
</tr>
<tr>
<td>9</td>
<td>2/24</td>
<td>Data Wrangling</td>
</tr>
<tr>
<td></td>
<td>2/22</td>
<td><strong>No class: Monday schedule is followed</strong></td>
</tr>
<tr>
<td>10</td>
<td>3/1</td>
<td>Matrices and Data Visualization</td>
</tr>
<tr>
<td>11</td>
<td>3/3</td>
<td>Writing Functions</td>
</tr>
<tr>
<td>12</td>
<td>3/8</td>
<td>Approaches to Modeling</td>
</tr>
</tbody>
</table>
Before the First Class

Before we meet for the first time, you should download R Studio (see above), and make sure it runs. Search online or email me with any problems.