MATH/LING 289: Introduction to Data Science, Fall 2018
Tuesday, Thursday 12:00-13:15   MRC LL1

Instructor:	**Taylor Arnold**
E-mail:	tarnold2@richmond.edu
Office:	Jepson Hall, Rm 218
Course Website:	`https://statsmaths.github.io/stat289-f18`

Overview:

Data science is an interdisciplinary field concerned with drawing knowledge from data and communicating those results to various audiences. Data science needs to be learned *by doing* data science. There are no short-cuts and the process cannot be learned by simply working our way methodically through a textbook of disconnected topics.

Therefore, this course will be taught using a problem-based learning model. As a class, we will be addressing an open ended research question and discussing various methods that will assist our inquiry. Both individually and in small groups, students will learn how to address subtasks that drive our research forward. At the end of the semester students will have acquired a toolkit of methods, and the knowledge of how to use them in practice, to address important social, cultural, and scientific questions with data-driven techniques.

Because statistics is currently housed in the mathematics program, this course is registered under a MATH designation. The topics of this course, however, do not fall within the traditional disciplinary boundaries of mathematics.

Computing:

We will make heavy use of computing throughout the semester. Nearly all of our work will involve some amount of writing and understanding code. While there are no formal programming prerequisites, the pace of the course assumes you have had some prior exposure to manipulating data with code (i.e., R, Python, or Java). If you would like a class on applied statistics that moves at a more user-friendly pace please consider taking Introduction to Statistical Modeling (MATH209). We conveniently offer a section of MATH209 in the same time-slot as this course, making it easy to switch.

Course Website:

All of the materials for the course will be posted on the class website:

> `https://statsmaths.github.io/stat289-f18`

At the end of the semester, this version of the course will be archived and available for your reference.

GitHub:

All of your work for this semester will be submitted through GitHub, the same platform that hosts our website. You'll need to set up a free account, which we will cover during the first week of class.

Course Structure:

Throughout the semester, we will work through a collection of Jupyter notebooks. These notebooks contain both code and text. We will have two types of files:

- **Tutorials**: These tend to be shorter files that introduce a new programming or data science topic. There will typically be a number of exercises to fill in at the end of the notebook.
- **Projects**: These will be longer, and more open ended, than tutorial files. They may ask you to chain together a number of concepts in order to solve a data science task.

We will typically start both types of notebooks in class. You are expected to complete them prior to the next class meeting (some projects may extend over multiple class periods; these will be clearly marked with a due date). In some cases you may be allowed or required to work in groups on an assignment; please make sure that all group members remember to submit the notebooks within their own GitHub repositories to receive proper credit.

Final Grades:

Your final grade will be determined by your work on the tutorial and project notebooks. The tutorials are graded based only on completion; you will get a pass/fail mark for each one. Each project will be labelled with a total number of points and you will receive a grade indicating how many of these points have been earned. The final grade will be determined by weighting these as follows:

- **Tutorials**: 20%
- **Projects**: 80%

To pass the course, you must also miss no more than four class meetings. Attendance requires that you arrive on-time, complete any out of class assignments for the day, and fully engage with the course material. Failing to fulfil these attendance requirements may result in a failing grade of 'V' or a reduction to your final course average at the instructor's sole discretion.

Class Policies:

The following class policies address some of the most common questions and concerns that students have. If anything is unclear, please feel free to contact me for clarification at any point in the semester.

- **Academic honesty:** Cheating and plagiarism are grave scholarly offenses and potential grounds for expulsion; they are also a major barrier to your intellectual development. You are expected to familiarize yourself with the entirety of the University of Richmond's Honor Code. If you are confused or unsure about appropriate citation protocol or any other aspect of the Honor code, please consult me before turning in an assignment.
- **Special approval:** If you have special approval forms for extra time on exams or any other circumstances I should know about, please speak with me as early as possible so that we can best accommodate your needs.
- **Collaboration:** You are always allowed to collaborate on Tutorial notebooks, though you should physically type out all of your own solutions unless I have given you instructions otherwise. For projects, I expect you to work on your own with no external assistance unless the instructions indicate that the project is to be worked on in groups. Groups will be established or assigned ahead of time in class.
- **Late work:** You are expected to submit all work on-time. Late work will be accepted after the due date with a point deduction for each day late (rounded up for each 24-hour period).
- **Attendance:** You are expected to both attend and participate in most class meetings. If you must be absent due to illness or other pressing need, please let me know by email as soon as possible. A habit of arriving late, failing to participate, or failing to accomplish any out of class assignments is considered equivalent to an absence.

- **Make-up work:** In instances where students have a valid excuse for missing an assessment due date, please get in touch with me within 24-hours of missing class to make alternative arrangements.

- **Class conduct:** During class I expect you to refrain from checking email, being on phones, or working on assignments for other classes.

- **Computers:** I except you to bring a working laptop with Python 3 installed. If this poses a challenge, please speak with me at the start of the semester or as new problems arise.

- **Office hours**: Rather than fixed weekly office hours, I will provide blocks of open times to meet with me particularly focused around exams and project due dates. If you find me in my office, poke your head in and I am usually happy to meet on the spot. Otherwise, please email me to make an appointment so that we can chat. Please note that appointments should be booked at least 24 hours ahead of time.

- **Email:** I will also answer questions by email (it can, in fact, be much faster than scheduling an appointment for small issues). During the week, I aim to respond within 24 hours, with emails sent over the weekend responded to by Monday morning. If your question involves code, please attach your current lab or report as that will expedite my answering your question(s).

Notice:

I reserve the right to modify this syllabus, with advanced warning, throughout the semester. If necessary, I will email the class list and post an updated version of the document on the course website.