MATH/CS 395: Statistical Learning, Fall 2017
Tuesday, Thursday 13:30-14:45   PURH G13

Instructor:	**Taylor Arnold**
E-mail:	tarnold2@richmond.edu
Office:	Jepson Hall, Rm 218
Office hours:	Tuesday, Thursday 10:30-12:00 or by appointment

Computing:

The focus of this course will be on applied statistics and data analysis over symbolic mathematics. To facilitate this, nearly every class assignment and exam will involve some form of computing. No prior programming experience is assumed or required.

We will use the **R** programming environment throughout the semester. It is freely available for all major operating systems and is pre-installed on many campus computers. You can download it and all supporting files for your own machine via these links:

`https://cran.r-project.org/`
`https://www.rstudio.com/`

I strongly recommend using your own machine for this course and bring a laptop or tablet to each class meeting. The lab computers in Jepson are available and contain some, though not all, of the required software.

Course Website:

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

`https://statsmaths.github.io/stat395`

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 week of class.

Labs:

Every course (other than this first one) through Thanksgiving will have an associated file named lab00.Rmd, with the appropriate class number replaced for the 00. By noon before the start of the next class, you must complete the questions contained within the lab notebook. Assignments will be submitted through GitHub; this process will explained in more detail during class.

During most class meetings, we will do some combination of presenting your results in small groups or to the class in general. Note that your presence and attention in class will be an important aspect of your lab grade.

Final Project:

There will also be a final project for this course consisting of both written and oral components. Details for this project will be given following Fall Break. You will have a wide flexibility in selecting an appropriate project.

Grades:

You will receive four letter grades in this course. Three of these will be aggregate grades covering approximately 1/3 of the labs each. The fourth will cover the written and oral components of your final project.
I want to make the grading extremely transparent. Your final grade will be computed by converting letter grades into numbers as follows (pluses increase the number by 0.33 and minuses decrease the number by 0.33):

Numeric Score	Final Grade
4	A
3	B
2	C
1	D
0	F

These numeric grades will be averaged according to the following weights:

- Final Project, 34%
- Labs, 66% (22% each)

And reading off of the following chart (grades are rounded to the second digit):

Numeric Score	Final Grade
3.84 - 4.00	A
3.50 - 3.83	A-
3.17 - 3.49	B+
2.84 - 3.16	B
2.50 - 2.83	B-
2.17 - 2.49	C+
1.84 - 2.16	C
1.50 - 1.83	C-
0.00 - 1.49	F

Exams:

This course has no exams, final or otherwise.

Weekly Topics:

The semester is broken up into roughly three parts, with the first week focused on setting up software and reviewing the required prerequisites. Weeks 2-6 offer a a basic introduction to machine learning covering principals such as complexity, over fitting, and regularization. In the second part, we spend 4 weeks covering topics in Natural Language Processing. In the final 4 weeks we cover topics from computer vision.

WEEK 01 - Introduction to R, RMarkdown, and Graphics

WEEK 02 - ML I: Linear Regression and Model Metrics

WEEK 03 - ML II: Incorporating Non-Linear Effects

WEEK 04 - ML III: Model Regularization

WEEK 05 - ML IV: Adaptive, Local Models

WEEK 06 - ML V: Using Dense Neural Networks

WEEK 07 - NLP I: Lexical Frequencies

WEEK 08 - NLP II: Sentiment Analysis

WEEK 09 - NLP III: Learning and Using Dependency Structures

WEEK 10 - NLP IV: Word Embeddings and Recurrent Neural Networks (RNN's)

WEEK 11 - CV I: Histograms, Filters, and Other Manual Features

WEEK 12 - CV II: Convolutional Neural Networks (CNN's)

WEEK 13 - CV III: Transfer Learning

WEEK 14 - CV IV: Style Transfer and Generative Models (GAN's)