MATH 289: Applied Regression Analysis, Fall 2017
Tuesday, Thursday 09:00-10:15 JPSN, G30

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. The lab computers will be available as well, though I find them to be quite slow and poorly maintained. We will devote a substantial amount of the time learning how to work within the R programming framework.

References:

There is no required textbook for the course, but we will pull from a number of sources for material. The following texts are freely available online and highly recommended:

- David M Diez, Christopher D Barr, Mine Çetinkaya-Rundel. *OpenIntro Statistics*. OpenIntro, Inc., Third Edition, 2015. Online: `https://www.openintro.org/stat/`.
- Garrett Grolemund, Hadley Wickham. *R for Data Science*. O'Reilly Media, First edition, 2016. Online: `http://r4ds.had.co.nz/`.

Other free references for particular sections will be mentioned throughout the course.

Course Website:

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

`https://statsmaths.github.io/stat289`

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.

Grades:

All grades in this course will be given as either a letter grade or on a 4-point scale. While occasionally possible to receive pluses / minuses or fractional points, these will usually be given a whole letter or number grade.

I expect most students to get full marks (A) for labs and participation. Students found to be delinquent in either will first receive a written warning, followed by an initial 50% reduction (C) in the respective grade, and finally a 100% reduction (F).

Your final grade will be determined by converting all grades into a numeric scale 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

I want to make the grading extremely transparent, so your final grade will simply consist of taking your weighted numerical average using the following weights:

- Labs, 10%
- Participation, 10%
- Quizzes, 30% (drop lowest 2 grades)
- Data Reports, 50%

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

Labs:

Most class meetings, particularly in the first half of the semester, will have an interactive lab associated with it. These consist of a set of questions that must be answered with either small snippets of code or short descriptive answers. Your solutions must be uploaded to your GitHub page.

Quizzes:

There will be short quizzes given during the semester. These will be at the end of class on Tuesdays. All will consist of entirely objective questions, such as True/False, multiple choice, and matching. You will be able to drop your two lowest grades; therefore, make-up quizzes will be offered only in extreme circumstances.

Data Reports:

Data reports are short written documents that mix code, graphics, and prose to provide a comprehensive analysis of a data set. You will have 3-4 data reports due throughout the semester. These must also be uploaded to GitHub. The format for these reports will be described latter in the course.

Exams:

This course has no exams, final or otherwise.

Attendance and Late Policy:

You are expected to submit work on-time. You should aim to attend all class meetings, however I am fully aware that through the course of the semester various issues – illness, sports, and family emergencies – will prevent many of you from attending every class. As long as you come prepared to most meetings of the course you will receive full credit for participation. Likewise, attending every class but continually failing to do the labs or engaging in the classroom activities will not earn full course credit.

Weekly Topics:

These topics are subject to change based on the pace of the course, but give a good sense of roughly what we are going to cover:

WEEK 01 - Introduction to R and RMarkdown

WEEK 02 - Basic Graphics

WEEK 03 - Variable Types

WEEK 04 - Data Collection

WEEK 05 - Data Manipulation

WEEK 06 - Simple Linear Models

WEEK 07 - Multivariate Linear Models

WEEK 08 - Spatial Data

WEEK 09 - Theories of Data Visualisation I

WEEK 10 - Theories of Data Visualisation II

WEEK 11 - Tidy Data

WEEK 12 - Relational Data

WEEK 13 - Strings and Dates

WEEK 14 - Penalized Regression and Text Processing