Class Overview

Contact Info

- Email: jcampbell32@fordham.edu
- Office Hours: 6-8 PM Tuesdays (*BY APPOINTMENT ONLY*)

Intro

Welcome to R statistical programming! This is a course focused on the R programming language. It's developed for beginner to advanced-beginner programmers.

R can do a lot, but it's mainly used for statistical applications. Because of this, we'll be going over some statistical concepts. The focus of those lessons will be using R as a tool and not so much an in-depth analysis of the concept itself.

Course Learning Goals

- Gain functional R programming knowledge
- Gain functional statistical knowledge using R
- Be able to express R-focused work in an easy to understand manner (i.e., going beyond "I typed this in and got this")

Additional Readings

I don't have a set book for the course, but if you are interested in expanding on ideas we go over in class, I suggest the following, both available for free online:

- R for Data Science; Grolemund and Wickham
- Hands on Machine Learning with R; Boehmke and Greenwell

Grading

The letter grade system is as follows:

```
## # A tibble: 12 x 2
##    Grade Number
##    <chr> <chr>
##  1 A     95-100
##  2 A-    90-94
##  3 B+    87-89
##  4 B     84-86
##  5 B-    80-83
##  6 C+    77-79
##  7 C     74-76
##  8 C-    70-73
##  9 D+    67-69
## 10 D     64-66
## 11 D-    60-63
## 12 F     Below 60
```

I do not guarantee a curve, but if I feel one needs to be used at the end of the year, I will use one.

Your grade is made up of three assignments, your final project, and participation.

- 15% Assignment 1

- 15% Assignment 2
- 15% Assignment 3
- 50% Final Project
- 5% Class Participation

Homework

You will have three homework assignments throughout the semester. These assignments will cover multiple weeks of lessons. Each homework is due **on a class day prior to that class**. You should be submitting two things for homework:

- Your script with code
- A write-up that explains output

If you submit your homework late or don't include one of the mentioned parts of it, a penalty will be added to your grade.

The write-up is the most important part; it allows me to see that you are able to understand the material beyond just typing stuff into R. I'll post an example of a good write-up on Blackboard, but think of it as an opportunity for you to explain your thoughts and methods I can understand your process beyond just code.

You are allowed (but not forced) to partner with someone else to do homeworks. Your partner can change from assignment to assignment. All I ask is that you let me know who you are working with via email.

Final Project

Your final project will be a group project in which you can choose to work on one of three projects I provide for you. You will be expected to both write a paper and give a 10-15 minute presentation the last week of class. More details will be provided later on in the semester.

Class Participation

This is more of a lecture style course than open discussion. However, we still have a class participation grade. It's mostly related to the following:

- Asking questions in class
  - Specifically, I'm interested in questions on the topic we are going over that everyone can learn from; if you have a technical difficulty-related question, save that for a break in class or after class
  - I'm looking for these interesting questions to be asked **in class**. Coming to me after class with a question on the topic is cool, but the whole class won't benefit from it.
- Being respectful
- Showing up

Academic Accomodations

Under the Americans with Disabilities Act and Section 504 of the Vocational Rehabilitation Act of 1973, all students, with or without disabilities, are entitled to equal access to the programs and activities of Fordham University. If you believe that you have a disabling condition that may interfere with your ability to participate in the activities, coursework, or assessment of the object of this course, you may be entitled to accommodations. Please schedule an appointment to speak with someone at the Office of Disability Services (RH -OH Hall, Lower Level, x0655). Please inform me ASAP if you are eligible for any accommodations.

Topics

- Week 1: Workflow Overview
  - Go over syllabus
  - Basics of R Studio

  - Basic coding structure
  - Installing and loading packages
- Week 2: Data Objects
  - Understanding of numeric, string, factor, etc. data
  - Working with tibbles
  - Basic necessities
- Week 3: Introduction to the Tidyverse
  - Data transformation with `dplyr`
- Week 4: Visualizing our Data
  - Learn to plot with `ggplot2`
  - **Homework 1 (Weeks 1-3) Due**
- Week 5: Exploratory Analysis
  - Summary statistics
  - What is our data telling us?
  - How to interpret plots
- Week 6: Linear Regression Pt. 1
  - Simple linear regression
  - Regression assumptions
  - Making predictions
- Week 7: Linear Regression Pt. 2
  - Multiple linear regression
  - Multi-collinearity
  - Polynomial regression
- Week 8: Logistic Regression
  - Modeling a binary response
  - Differences with linear regression
  - **Homework 2 (Weeks 4-7) Due**
- Week 9: Modeling Metrics and Feature Engineering
  - What is a good model?
  - What do different metrics mean?
  - Developing new features
  - Transforming variables
- Week 10: Cross-Validation
  - Using the `caret` package
  - Bootstrapping, K-folds, repeated K-folds
- Week 11: Regularized Regression
  - Modifying the bias-variance tradeoff
  - Feature selection
- Week 12: Looping
  - Using the `purrr` package
  - How to loop functions over multiple columns, objects, etc.
  - **Homework 3 (Weeks 9-11) Due**
- Week 13: Re-hash or Presentations
- Week 14: Presentations
  - **Final Presentations Due**