Data Acquisition
ITAO 40450

Office: 337 Mendoza

Email: seth.berry@nd.edu

Office Hours

M – 12:30 to 1:30

W – 10:00 to 12:00

F – 10:00 to 1:00

Class Days and Time

Tuesday and Thursday – 12:45 to 2:00

Location – L004

Course Objectives

While data is everywhere, you cannot always rely on having the appropriate data on hand. As an analyst, you need to be able to acquire data from many different resources to get answers. By the end of the class, I want you to be able to:

- Develop good surveys and refine existing surveys
- Create surveys for deployment in various online and mobile platforms
- Use appropriate statistical methods, like factor analysis, for analyzing survey data
- Programatically access data from the web, using APIs and scraping
- Articulate the different methods for analyzing streaming data

Class Activities

In addition to more "traditional" lecturing, we are going to engage in different activities during class time. We will try some socially-distanced, team-based coding challenges and concept quizzes. All of these activities are designed to practice the skills that we are learning in class, in a (hopefully) fun and low-stakes manner.

Readings

There is no official textbook for this course, but there are going to be a few assorted readings and resources for topics. I will also share resources for topics. At the start of each class, we will each be finding an article to share with the class. You will be expected to find an article that interests you and share a brief synopsis with the class. This is designed for you to see the wealth of research that exists within the topics that we are going to cover.

Homework

During the course of the mod, we will have 3 homework assignments (each worth 30 points). All homework assignments must be submitted in a compiled file (knitted from R Markdown or a Python-flavored notebook of your preference) – no other file types will be accepted and reminders won't be given.

Homework will be composed of three distinct parts – initial submission, individual feedback session, second submission – and each will count towards 1/3 of the points. The initial submission must be completed before any feedback can be offered. While you do have flexibility with regard to completing the homework assignments, pushing these until the end of the mod is not advisable.

Final Project and Presentation

As opposed to a final exam, we will be having presentations on our last day of class. You can work individually or as duos. The goal is to answer a question or solve a problem. Presentation guidelines will follow, but general creativity and appropriate technique use will figure heavily into your grade. This is a chance for you to do something interesting, not just go with what might be easy on Kaggle.

Engagement

Engagement is not just coming to class, but being an active participant. Throughout class, you will be given the opportunity to practice content. At the end of each class, you need to turn in your code (it does not need to be pretty and can just be any text-based file). Each submission is worth 10 points for up to a maximum of 100 points. Project touch points will also be included in this.

Grade Breakdown

Engagement – 100 points

Homework – 90 points

Project – 60 points

Presentation – 50 points

Total – 300 points

Schedule

Table 1: Tentative Schedule

Week	Date	Topic	Assignments
1	02/04 (R)	Introduction, Software, Programming	
	02/09 (T)	Regular Expressions	
2	02/11 (R)	Survey Design - Basics	
	02/16 (T)	Survey Design - Web	Homework #1 Assigned
3	02/18 (R)	Survey Design - Action	
	02/23 (T)	Survey Analysis	
4	02/25 (R)	APIs and Forms	Homework #2 Assigned
	03/02 (T)	APIs and Forms	
5	03/04 (R)	Scraping - Easy	
	03/09 (T)	Scraping - Not easy	Homework #3 Assigned
6	03/11 (R)	Automation	
	03/16 (T)	Streaming data analysis	
7	03/18 (R)	Project work	
	03/23 (T)	Presentations	Presentations