Data Acquisition
ITAO 40450

Office: 337 Mendoza

Email: seth.berry@nd.edu

Office Hours

Monday & Wednesday: 10:00 to 1:00

Most Fridays: 10:00 to 12:00

As of writing, I will be holding my M/W office hours on Zoom. On Friday, we will try Twitch.

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

All homework assignments must be submitted in an html file knitted from R Markdown.

A significant portion of your grade will come from the homework map (i.e., your path towards statistical enlightenment). The skills roadmap has different levels (bronze, silver, and gold). Each level must be completed in order (i.e., you cannot skip silver and go straight to gold from bronze). Silver and gold levels essentially act as modifiers to not only your grade, but also your understanding.

Final Project and Presentation

These will be given during the designated final day and time. These will be graded on the content of your presentation (general creativity, appropriateness of analysis, etc.), not necessarily on the scientific merit of your work! I really want you to be able to speak about your process and results in a defensible way.

Grade Breakdown

Participation – 50 points

Homework – 90 points

Project Touch Points – 50 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