Syllabus

Nic Schwab

Basic information

- **Course title:** MTH 190: Introduction to Data Science
- **Instructor:** [Nicholas Schwab (he/him)] - Assistant Professor of Mathematics. Please call me "Schwab" or "Professor Schwab"
- **Office location:** Frost 362
- **Email:** nschwab@hcc.edu
- **Meeting locations/times:**

    – MF: 11:00 AM - 12:15 PM / Kittredge Center 420

- **Getting help**:

    – For non personal/sensitive questions, ask on the `#questions` channel on Slack. For personal/sensitive matters, Slack DM Professor Schwab.
    – Prof. Schwab's hours:

        * MF 10:00am - 10:50
        * W 1:00 - 1:50pm
        * W 4:00 - 5:00pm by appointment only.
        * FR 362 or set up a zoom appointment.

    – AJ's office hours

        * Office hours: MW 7-9pm
        * See the slack chanel for AJ's zoom link

Instructor work-life balance

- I will respond to Slack messages sent during the week within 24h. I will respond to Slack messages sent during the weekend at my own discretion.
- If possible, please only Slack me with briefer and administrative questions; I prefer having more substantive conversations in person as it takes me less energy to understand where you are at.

- I will do my best to return all grading as promptly as possible.

How can I succeed in this class?

Ask yourself:

- **When I have questions or don't understand something:**
  - *"Am I asking questions in class?"*
  - *"Am I asking questions on Slack in the `#questions` channel?"* Even better: *"Am I answering my peers' questions on Slack?"*
  - *"Having I been going to the Spinelli tutoring center for help on R and the tidyverse?"*
  - *"Have I been coming to office hours?"*

- **Lectures and readings:**
  - *"Am I staying on top Slack notifications sent between lectures?"* If you need help developing a notification strategy that best suits your lifestyle, please speak to me.
  - *"Am I attending lectures consistently?"*
  - *"During in-class activities, am I actually running code line-by-line and studying the outputs, or am I just going through the motions?"*
  - *"During in-class exercises, am I taking full advantage that I'm in the same place at the same time with the instructor, the lab assistants, and most importantly your peers, or am I browsing the web/texting the whole time?"*
  - *"Have I been doing the associated readings for each lecture?"*

Course Description & Objectives

The world is awash with raw data, and students will benefit in the job market from learning to properly analyze it. The purpose of this course is to introduce students to the concepts of data science early in their college education. In this course, students will learn to code in R or Python, popular programming languages commonly used in Data Science. This course can also be seen as an introduction to programming languages and coding. It is designed to be accessible to students from a wide range of backgrounds and interests, including those without advanced mathematics or computer programming experience.

Prerequisite(s): MTH 085 or MTH 099 or MTH 011/MTH 012 with a grade of C- or better; or SM08, or adequate score on the Mathematics Placement Examination. No computer language experience necessary but student must have a willingness to learn to code.

3 Class Hours 3 Credits

BASIC OUTLINE OF COURSE CONTENT:

- Introduction to Programming Language for Data Science (R or Python) and Computing Environment (RStudio, Jupyter Notebook, etc.)

- Data Science Life Cycle

- Grammar of Graphics

- Data Visualization

- Data Wrangling

- Data Collection

- Web Scraping and Importing Data

- Data Management

- Data Ethics

- Introduction to Database Querying with SQL

- Further Topics based on time and student interest may include:

    - Iteration

    - Dynamic and Customized Data Graphics

    - Geospatial Data and Maps

    - Text Mining

    - Introduction to Machine Learning (to motivate interest in an Intro to Data Science 2 course that would be devoted to this.)

Student Learning outcomes

Students who complete this course will be able to:

- Develop skills in using modern technology to conduct data analysis.

- Draw meaningful conclusions and insights from real world data.

- Understand data collection from various sources and appropriate data management.

- Effectively organize, clean, prepare data for analysis

- Communicate clear, meaningful data visualizations.

- Develop understanding and appreciation of ethics in Data Science.

Lecture schedule

The lecture schedule and associated readings can be found on the main page of this course webpage.

Book and software

The Modern Data Science with R **textbook** is accessible on the navigational bar of this webpage as well as via this link. Students will use the HCC RStudio workbench to do computations in R find it here.

Class norms

- You are expected to stay until the end of lecture. If you need to leave early, please confirm with me at the beginning of lecture and sit somewhere where your departure will be minimally disruptive.
- *Attendance policy.* You may be dropped from the course for missing 3 or more days of class.
- Attendance will not be explicitly taken and occasional absences are excused. However, extended absences should be mentioned to me.

Readings

We'll be reading chapters 1-6,8,15,and 17 of Modern Data Science with R.

Evaluation

All due dates can be found on the main page of this course webpage.

This class is project based. As such your evaluations will be based on projects you create.

Projects 40%

There are four projects to complete during the semester.

Final project 20%

There will be a final project due the last day of classes. You'll do this final project in groups of 2-3 where you can choose your groupmates.

Problem sets 25%

Homework will be assigned weekly, its important to complete and stay on top of homework.

Engagment 15%

It is difficult to explicit codify what constitutes "an engaged student," so instead I present the following rough principle I will follow: *you'll only get out of this class as much as you put in.* That being said, here are multiple pathways for you to stay engaged in this class:

- In particular: Peer evaluations for all projects.
- Asking and answering questions in class.
- Working and staying focused in class.
- Coming to office hours.
- Posting questions on Slack.
- Even better: Responding to questions on Slack.

Tentative Assignment schedule

Homework will be assigned weekly and due by midnight the following Monday.

Project 1- due after reading chapter 4 of MDSR roughly the 10th day of class

Project 2- due after reading chapter 5 of MDSR roughly the 18th day of class

Project 3- due after reading chapter 15 of MDSR roughly the 22nd day of class

Project 4- due after reading chapter 17 of MDSR roughly the 26th day of class

Final Project/ Presentation will take place during our final time.

For a more up to date schedule see the course's website.

Policies

1. **Collaboration**: While I encourage you to work with your peers for problem sets, you must submit your own answers and not simple rewordings of another's work. Furthermore, all collaborations must be explicitly acknowledged in your submissions.

ACCESSIBILITY ACCOMMODATIONS:

Holyoke Community College values inclusion and equal access to its programs and activities and is committed to fostering an environment of mutual respect and full participation. Our goal is to create learning environments that are equitable, inclusive and welcoming. If you are an individual with a disability and require reasonable academic accommodations you are advised to contact the Office of Disability Services (ODS) at any time to discuss your accommodation needs and options. The ODS will work collaboratively with students with disabilities to develop effective accommodation plans for implementation in the classroom. The ODS is located in DON 147. For an appointment, please call 413-552- 2417. The Video Phone (VP) number is 413-650- 5502.