S55 MPH 5976: Skill Lab: Introduction for Public Health Data Analysis

Dominique Lockett

June 15, 2020 - June 19, 2020

E -mail: dlockett@wustl.edu

Synchronous Hours: M -F 2:00 -3:30pm

Web: dominiquelockett.com/

Lecture Hours: M -F Self -paced

Course Description

This course will allow students to grasp the fundamental tools necessary to use Python as an analytical tool for data analysis in Public Health. This course will introduce students to the fundamentals of the Python language, common Python modules for data manipulation and analysis, and Jupyter notebook environment. The course willith how to acquire data from publicly available sources and databases, cleansing and transformation of data, creation of descriptive statistics and graphics. The course will also introduce Python's natural language processing and machine learning modules for basic data classification and predictive modeling applications. Throughout the course, instruction and assignments will promote best practices for data analysis that can be shared and used for reproducible research.

Prerequisites

Familiarity with a similar analytic program such as R, Stata, or SPSS is preferred.

Course Objectives

The following skills will be taught to support data analysis in epidemiology and biostatistics:

- A brief overview of skills in basic probability and inference.

- Calculate basic measures of frequency and association

- Understand the appropriate methods for descriptive and inferential interpretation (continuous, categorical, count data).

- Use understanding of appropriate methods to conduct inference using appropriate methods (linear regression, logit regression, duration models).

- Create graphical interpretations and summaries of statistical analysis

Required Materials

Software: All synchronous meetings will occur on Zoom and installation of this tool will be necessary to participate live.

The university will provide a web -based Python programming environment called Jupyter for the purposes of instruction during class and use by students on exercises. Students will not be required to install any software on their personal laptops to participate in the course, but students may find it helpful to have a Python programming environment available on their personal laptops for further use outside of class. Jupyter can be downloaded and run locally using the instructions found here: http://jupyter.org/install.html

Recommended Materials: While there are no books required in this course, effective use of internet resources is fundamental in the use of analytical software. You are encouraged to review the following online resources to supplement the lectures provided in this course:

- https://realpython.com/

- Mastering Python Data Analysis

- I cannot understate the utility of doing web searches when you encounter a problem on Python. Whether it be the complicated installation processes or an inscutable error, applying a web search will reveal dozens of others who have encountered similar issues. Stack Overflow offers a priceless set of experts to help guide you in your troubleshooting.

University-Wide Policies

Accommodations for Sexual Assault
The University is committed to offering reasonable academic accommodations (e.g., no contact order, course changes) to students who are victims of relationship or sexual violence, regardless of whether they seek criminal or disciplinary action. If you need to request such accommodations, please contact the Relationship and Sexual Violence Prevention Center (RSVP) at rsvpcenter@wustl.edu or (314) 935-3445 to schedule an appointment with an RSVP confidential, licensed counselor. Information shared with counselors is confidential. However, requests for accommodations will be coordinated with the appropriate University administrators and faculty.

Mental Health Services
Mental Health Services' professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends and family, concerns about eating or drinking patterns, and

feelings of anxiety and depression. For more information, visit: www.students.wustl.edu/mental-health-services/

Relationship and Sexual Violence Prevention Center (RSVP)

The Relationship and Sexual Violence Prevention Center (RSVP) offers support for those who have experienced sexual violence, sexual misconduct, dating violence, domestic violence or stalking. RSVP can help those who are not sure what steps they wish to take to respond to their experiences. They offer confidential support and can help arrange for necessary classroom accommodations. To get help, contact the RSVP Center at 935-3445 or rsvpcenter@wustl.edu. Their office is located in Seigle Hall, Suite 435.

Bias Report and Support System (BRSS)

The University has a process through which students, faculty, staff, and community members who have experienced or witnessed incidents of bias, prejudice, or discrimination against a student can report their experiences to the University's Bias Report and Support System (BRSS) team. To report an instance of bias, visit www.brss.wustl.edu.

Center for Diversity and Inclusion (CDI)

The Center for Diversity and Inclusion (CDI) supports and advocates for undergraduate, graduate, and professional school students from underrepresented and/or marginalized populations, collaborates with campus and community partners, and promotes dialogue and social change to cultivate and foster a supportive campus climate for students of all backgrounds, cultures and identities. See: https://diversityinclusion.wustl.edu/

Organization of the course

Lectures for the course will be pre -recorded and will be distributed on Canvas and available 24 hours before the live meeting are to occur. Typically, these lectures will be split up into three thirty minute intervals. At the end of each interval you will be tasked with completing an activity to demonstrate your understanding of the materials at hand. Because of the unorthodox nature of this class, the in -class activities will serve as a measure of your participation in the course.

Synchronous meetings will serve as office hours for any questions of problems you may have while watching the lecture and completing the activities. These meetings will be no more than 90 minutes and will last as long as students need them. Participation in these meetings are not required but strongly encouraged. Links to the live meetings will be provided on Canvas along with a password to ensure the security of the meetings.

Assignments and Grading Criteria

Grading: For Brown School Students, final grades will be based on your performance on the four quizzes (40%), the final lab assignment (40% and participation (20%)..

   Quizzes: You will complete four quizzes in classes 2 -5 which cover the daily course material. These quizzes will focus on the skill obtained during in class activities and will aid the instructor and students monitor understanding of class material.

   Final Assignment: The final assignment be a problem set which asks the students to demonstrate their knowledge of course materials. The assignment is due June 26 at midnight (Canvas

Submission)

Participation: Completion of the lecture activities will account for 100% of the participation grade. If you are unable to execute an activity on your own, you are expected to participate in the synchronous session. To demonstrate participation, you will be required to provide one .py script per class meeting which includes all aspects of the activities provided on that day (uploaded to Canvas).

Course Outline*

Day	Date	Topics	Quizzes
1	6/15	**Introduction** **Getting started with Python and Notebook** **Python Basics** -IPython Notebook -Python types -Flow control -Arrays and Numpy -Panda data frame -Indexing	
3	6/16	**Data Cleaning and transformation** -Loading data -Merging datasets -Web scraping	Quiz 1
2	6/17	**Basic statistics and descriptive graphics** -Graphs and insights -Public health examples	Quiz 2
	6/18	**Data Visualization** -Single variable -Multi -variable -Time -	Quiz 3
	6/19	**Crash course in natural language processing and machine learning** -Frequency Analysis -Document classification	Quiz 4
	6/26		Final assignment due

*Content and order of this schedule are subject to change