COURSE AND CONTACT INFORMATION

Course: **STAT 6289 Computing Environments**

Lectures: 12:45PM - 03:15PM MON 350
Instructor: Dr. Tatiyana V. Apanasovich

Office Address: 729 Rome Hall
Telephone Number: NA
E-mail: apanasovich@gwu.edu

Office Hours: R, 3:30am-4:30pm and by appointment.

COURSE DESCRIPTION

The course " Computing Environments" is an engaging and comprehensive exploration of data analysis, statistical techniques, and the practical use of computing tools for effective data-driven decision-making. This course is designed to equip students with the knowledge and skills required to harness the power of statistics and programming languages in the modern data-driven world.

Through a hands-on and interactive approach, students will be introduced to various programming languages, including Python, R, and others, which are extensively used in data analysis and statistical modeling.

Module 1: Introduction to Computing Environments for Data Analysis

Role of computing environments in data analysis
Overview of programming languages (e.g., Python, R, C, Fortan,Jason,Julia)
Introduction to statistical software packages (e.g., RStudio, Jupyter Notebook)

Module 2: Data Manipulation and Cleaning:

Learning how to import and export data in different formats using programming languages and statistical software.
Techniques for cleaning and preprocessing data to handle missing values, duplicates, and outliers.
Hands-on exercises to practice data cleaning and manipulation tasks.

Module 3: Exploratory Data Analysis (EDA):

Understanding the importance of EDA in comprehending the characteristics and patterns in datasets.
Utilizing computing environments to calculate summary statistics and visualize data distributions.

Module 4: Data Visualization:

Exploring the principles of effective data visualization to communicate complex insights.
Using visualization libraries and tools in programming languages to create various types of charts and plots.
Customizing visualizations for better data representation and storytelling.

Module 5: Programming for Data Analysis:

Covering the basics of programming languages used for data analysis, including variables, loops, and functions.
Hands-on coding exercises to manipulate data and perform basic calculations.
Introduction to file input/output operations for reading and writing data files.

Module 6: Statistical Analysis with Programming Languages:

Implementing statistical models using programming languages, such as regression, t-tests, and ANOVA.
Utilizing libraries and packages specific to each programming language for statistical computations.

Module 7: Big Data Analysis

Introduction to big data concepts
Distributed computing frameworks (e.g., Apache Spark)
Handling and analyzing large datasets

Module 8: Reproducible Research and Reporting

Version control and collaboration using Git
Creating reproducible analysis workflows
Reporting and sharing results (e.g., R Markdown, Jupyter Notebooks)

Module 9: Applications and Case Studies

Applying computing environments to real-world datasets
Analyzing data from different domains (e.g., healthcare, finance, social media)
Ethical considerations in data analysis

COURSE PREREQUISITE(S)

Introductory Statistics: Familiarity with introductory statistics topics such as hypothesis testing, confidence intervals, and simple linear regression will provide a good foundation for the course.

Programming Fundamentals: Basic programming skills and familiarity with any programming language (e.g., Python, R, or similar) will be advantageous.

TEXTS

There is no required textbook. We will make use of several freely available textbooks and other materials. All course materials will be provided. We will use the R and Python software for data analysis, which is freely available for download. Some textbooks for R and Python

1. "Python for Data Analysis" by Wes McKinney

This book provides a comprehensive guide to using Python for data analysis, data manipulation, and visualization. It covers essential libraries like NumPy, pandas, and matplotlib, which are commonly used in data analysis tasks.

2. "R for Data Science" by Hadley Wickham and Garrett Grolemund

This book is an excellent resource for learning data analysis using the R programming language. It covers data manipulation, visualization, and modeling with packages like tidyverse, dplyr, and ggplot2.

LEARNING OUTCOMES

By the end of this course, students should be able to:

- Explain the significance of computing environments in data analysis workflows and statistical modeling.
- Utilize programming languages such as Python, R, to handle diverse data formats and perform data manipulations.
- Apply statistical software environments for data cleaning, analysis, and visualization.
- Calculate summary statistics, create data visualizations, and perform exploratory data analysis (EDA) to comprehend data characteristics and patterns.
- Implement statistical models, regression analysis, ANOVA, and t-tests using programming languages for data-driven decision-making.
- Utilize distributed computing frameworks like Apache Spark for big data analysis and handling large datasets.
- Create reproducible analysis workflows using version control tools like Git and present results through R Markdown and Jupyter Notebooks.
- Apply computing environments to analyze real-world datasets from diverse domains such as healthcare, finance, and social media.

GRADING

Your final grade will be a weighted average of your homework average (20%), midterm take-home exam (40%), and take home project (40%).

CLASS POLICIES

Homework: There will be 6-8 homework assignments, with greater frequency in the first half of the course. No late homework will be accepted, but the lowest score will be dropped

Midterm Exam: There will be a take-home exam consisting of a few problems, where you will be using the software to do the problems(posted on Feb, 29 and due March 08). Must work independently.

Take Home Project: There will be a take-home exam consisting of a few problems, where you will be using the software to do the problems (posted on April, 18 and due April, 26). Must work independently.

University policies:

University policy on observance of religious holidays
In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance. For details and policy, see: students.gwu.edu/accommodations-religious-holidays.

Academic integrity code

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For details and complete code, see: studentconduct.gwu.edu/code-academic-integrity

Safety and security
In the case of an emergency, if at all possible, the class should shelter in place. If the building that the class is in is affected, follow the evacuation procedures for the building. After evacuation, seek shelter at a predetermined rendezvous location.

Support for students outside the classroom

Disability Support Services (DSS)
Any student who may need an accommodation based on the potential impact of a disability should contact the Disability Support Services office at 202-994-8250 in the Rome Hall, Suite 102, to establish eligibility and to coordinate reasonable accommodations. For additional information see: disabilitysupport.gwu.edu/

Mental Health Services 202-994-5300
The University's Mental Health Services offers 24/7 assistance and referral to address students' personal, social, career, and study skills problems. Services for students include: crisis and emergency mental health consultations confidential assessment, counseling services (individual and small group), and referrals. For additional information see: counselingcenter.gwu.edu/