Syllabus

CSC 31167: Foundations of Data Science

Instructor: Filipa Calado

Email: fcalado@gradcenter.cuny.edu

Zoom address: https://nyu.zoom.us/my/filipa.calado

Office hours: https://www.bit.ly/calado_office

Bulletin Description: This course introduces the fundamental concepts and computational techniques of data science to all students, including those majoring in the Arts, Humanities, and Social Sciences. Students engage with data arising from real-world phenomena—including literary corpora, spatial datasets, and social networks data—to learn analytical skills such as inferential thinking and computational thinking. The competencies learned in this course will provide students with skills that will be of use in their professional careers, as well as tools to better understand, quantitatively and qualitatively, the social world around them. Finally, by teaching critical concepts and skills in computer programming and statistical inference, the class prepares students for further coursework in technology-dependent subjects, such as Digital Humanities. The course is designed for students who are new to statistics and programming. Students will make use of the Python programming language, but no computer science pre-requisites are required.

Instructor Description: This course is designed to introduce humanities students to the basics of data science using Python programming. The course will focus on the critical feminist approach to studying data, which emphasizes the importance of understanding and addressing the ways in which power and privilege in social systems shape the collection, analysis, and interpretation of data. The critical feminist approach centers marginalized identities and experiences and the intersectionality of gender, sexuality, race, and class as factors that shape data creation, our methods for analyzing data, and the conclusions we can draw from it. Students will explore the ways in which a feminist approach to data science can be used to reinforce or challenge existing power structures and promote social justice.

The critical feminist approach to data science in this course begins by contextualizing intersectionality as a historical movement and critical method of analysis. Students will then move to deconstructing the role that power and privilege have in shaping data collection and analytical methods, and the need to actively work to counteract these biases in the way we handle and interpret data. This course grounds discussion of intersectionality, power, and privilege with practical experimentation, introducing students to programmatic methods of data analysis with Python. As they learn to code with Python, students will examine how bias infiltrates computational processes, examining firsthand how the necessity for standards and rules that enable computation can stymie the expression of real-world and human complexity.

Prerequisites: [NA] Co-Requisites: [NA] Credits/Hours: 3 Credits/3 Hours

This course satisfies Pathways Math and Quantitative Reasoning requirement.

Course Learning Outcomes:

- Interpret and draw appropriate inferences from quantitative representations, such as formulas, graphs, or tables.
- Use algebraic, numerical, graphical, or statistical methods to draw accurate conclusions and solve mathematical problems.

- Represent quantitative problems expressed in natural language in a suitable mathematical format.
- Effectively communicate quantitative analysis or solutions to mathematical problems in written or oral form.
- Evaluate solutions to problems for reasonableness using a variety of means, including informed estimation.
- Apply mathematical methods to problems in other fields of study.

Required Texts/Readings:

There is only one book for the course, which is free online: Lauren Klein and Catherine D'Ignazio, Data Feminism. https://data-feminism.mitpress.mit.edu/

Technology:

In-class lessons and homeworks are done in Jupyter notebooks. The notebooks assume a Python 3 installation with the standard modules from an Anaconda installation such as NLTK, Pandas, Numpy and Matplotlib. If you have trouble installing python, there are backup solutions, and please reach out to me.

Assignments overview:

- Homework assignments (20%) - Short coding assignments meant to get you to demonstrate your comprehension of in-class lessons. Includes a short written component. Will be graded on effort rather than accuracy. Prompts are posted on the class website.
- Participation (30%) - Students are expected to be actively engaged in class activities. This means paying attention to lessons and participating in class discussions. Students are expected to come to class having done the reading and being prepared to give their opinions. For a student who comes to every class (not counting excused absences), is actively listening and engaged, and speaks up at least once during class, they will get 100% on participation.
- Final project (30%) - Group projects centered on posing a research question and doing exploratory analysis of a dataset. Includes a coding and written component, and groups will present their process and preliminary findings in the last week of class. Instructions will be distributed during the final unit.
- Exams (20%) - Midterm and final exam which will assess students' understanding of data analysis procedures as applied to their own research interests. Format will be jupyter notebooks.

Grade distributions:

- Homework assignments (20%)
- Exams (20%)
- Participation (30%)
- Final projects (30%)

Course overview:

- Unit 1: Introduction to Python programming
- Unit 2: Introduction to data analysis and visualization, *Data Feminism* by Lauren Klein and Katherine D'Ignazio
- Unit 3: Research methods and final projects