Intro to Data Science (DSCI 401)

Fall 2024

Course Description

This course provides students with an introduction to using R and Python programming language for statistical use covering such topics as data wrangling, data visualization, principles of reproducible research, building simple statistical models/machine learning.

Prerequisites: None

Instructor

Mena CR Whalen, PhD
Assistant Professor
✉ mwhalen3@luc.edu
☎ 773.508.3580
📍 Loyola Hall (Office 108)

When: Monday & Wednesday 4:00 - 5:40 pm

Credit Hours: 4

Location: Cuneo Hall - Room 324

Office Hours: Loyola Hall (Office 108)
TBA Message on campuswire for appointment

Make individual appointment/meeting requests through Campuswire (discussed below) by selecting *Post to instructors*.

Course Structure

- Reading: Readings are posted in the tentative schedule that should be read *before class*. Come to class prepared to solidify the readings through lecturing and group activities.
- Class Time: Class will be composed of lecturing, discussions, collaborative activities, and coding practice. Please come to class having done the reading, a charged computer, and ready to discuss and learn in a collaborative manner.
- Assessments: Students will complete homeworks, a midterm, 1 project, and a final.
- Participation, Discussion, and Group Work: One important aspect of a Jesuit education is learning to respect the rights and opinions of others. Please respect others by (1) allowing all classmates the right to voice their opinions without fear of ridicule, and (2) not making objectionable (gendered, racial or ethnic) comments, especially comments directed at a classmate. Group work and discussion are vital to this class since no one student will understand everything, please lean on each other for help and learn to hear concepts and ideas from another perspective.

Textbook

REQUIRED Modern Data Science with R (2nd edition). Baumer, Kaplan, and
Horton
    PDF of Book: `https://beanumber.github.io/mdsr2e/index.html`
    REQUIRED Introduction to Data Science: A Python Approach to Concepts,
Techniques and Applications. Igual and Segui.
    PDF of Book: `https://github.com/DataScienceUB/introduction-datascience-python-book`
    Recommended Grolemund, G., & Wickham, H. (2017). R for Data Science.
O'Reilly Media.
    PDF of Book: `https://r4ds.had.co.nz/`

R, Posit (formerly RStudio)

WE WILL BE USING/INTRODUCING the free statistical software R. While R is
the engine, we will use the free and open source IDE (Integrated Development
Environment) RStudio to run it. R and RStudio are set up and available on all
library computers.

R: Version 4.4.1, "Race For Your Life", `https://cran.rstudio.com/`

Posit Version "2024.04.2", https://posit.co/

Python and Jupyter Notebooks

WE WILL ALSO BE USING/INTRODUCING the versatile programming language
Python, which is widely used in data science for its readability and extensive
libraries. To work with Python, we will utilize Jupyter Notebooks, a powerful
tool that allows you to write and execute code in an interactive, web-based
environment. Jupyter Notebooks are particularly useful for data exploration,
visualization, and sharing your findings in a dynamic format.
    In addition to Jupyter Notebooks, you will have the option to work with
Python directly in Posit (formerly RStudio). This integrated environment allows
you to use both R and Python in one place, streamlining your workflow and
making it easier to switch between the two languages.

Python: Version 3.10.11, https://www.python.org/

Jupyter: Version 6.5.4, https://jupyter.org/

Quarto

Quarto is an open-source publishing system that allows you to create dynamic,
reproducible documents by integrating code, text, and visualizations. In this
course, you'll use Quarto alongside R, Python, and Posit (RStudio) to generate
reports, presentations, and other types of documents. Quarto supports a
variety of output formats, including PDFs, HTML, and slideshows, making it an
excellent tool for both data analysis and communication.

Quarto Version "1.5.56", https://quarto.org/

Github

GitHub will be an essential resource in this course for accessing course materi-
als, such as labs and supplementary resources. We will use a dedicated GitHub

repository, STAT 401 GitHub, to share these materials with you. By cloning or downloading the repository, you'll have easy access to all course content in one place, and you'll be able to stay up-to-date with any updates or changes. Additionally, you'll learn how to use GitHub for version control, which is a valuable skill for managing your projects and collaborating on code.

Asking Questions & Course Communication

This term we will be using Campuswire as our preferred platform for questions about homework, reading checks, R questions and general course questions. The system is highly catered to getting you help quickly and efficiently from classmates and the instructor. Rather than emailing questions to the instructor, you should post your questions on Campuswire. You can ask and answer questions anonymously on the site.

Enrollment Code: 6370

Questions concerning individual grades should be addressed through email.

I will check Campuswire periodically and answer questions[1], but I strongly encourage students to answer each other's questions.

[1] Please do not expect answers during weekends and evenings.

Evaluation

STUDENTS WILL BE EVALUATED through (1) Homework; (2) Midterm; (3) Project (4) Final.

Exams

There will be one midterm exam and one final for the semester. Both exams may not be made up unless there is a serious reason for missing and arrangements are made prior to the test. The midterm will be cumulative up to that point in class. Final will be cumulative. Both exams would incorporate interpretations and outputs from both R and Python.

Homework

Homework is due approximately every other week. Discussion between classmates is encouraged; however, the final work should be independent. Homework must be submitted through Sakai. Homework turned in 2 days after the due date will receive no credit. To help your final grade, please avoid late homework.

Project

The individual project will require students to find a raw data set, wrangle the data into a useful format, perform some interesting analysis, and present results in a written report following the principles of reproducible research. All

code must be version controlled through github (or repository of your choice) and a link to the repository must be submitted along with the final report.

More specific details on the project presentations and reports will be given at a later date but note that the project will have multiple due dates throughout the entire semester.

Attendance

Attendance will not be taken and does not count for credit in this course but due to the small number of class periods missing classes should be avoided.

Grading

Category	Weight
Homework	25%
Midterm	25%
Final	25%
Project	25%

Grading Scale	
90 - 100%	A
80 - 89.9%	B
70 - 79.9%	C
60 - 69.9%	D
Below 60%	F

Student Academic Services

Tutoring

The www.luc.edu/tutoring embodies the mission of Loyola University Chicago by providing academic services and resources which foster development of skills and attitudes necessary to increase the knowledge and academic independence of all students. Through multiple learning services, the Tutoring Center helps to contribute towards student success and growth efforts that are made by Loyola University Chicago.

Accommodations

Loyola University provides reasonable accommodations for students with disabilities. Any student requesting accommodations related to a disability or other condition is required to register with \href{Student Accessibility Center (SAC), located in Sullivan Center, Suite 117. Students will provide professors with an accommodation notification from SAC, preferably within the first two weeks of class. Students are encouraged to meet with their professor individually in order to discuss their accommodations. All information will remain confidential. For more information or further assistance, please call 773.508.3700 or email ✉ sac@luc.edu.

Academic Integrity

Academic dishonesty can take several forms, including, but not limited to cheating, plagiarism, copying another student's work, and submitting false documents. Academic cheating is a serious act that violates academic integrity. Cheating includes, but is not limited to, such acts as

- Obtaining, distributing, or communicating examination materials prior to the scheduled examination without the consent of the teacher
- Providing information to another student during an examination
- Obtaining information from another student or any other person during an examination
- Using any material or equipment during an examination without consent of the instructor, or in a manner which is not authorized by the instructor
- Attempting to change answers after the examination has been submitted
- Unauthorized collaboration, or the use in whole or part of another student's work, on homework, lab reports, - programming assignments, and any other course work which is completed outside of the classroom Falsifying medical or other documents to petition for excused absences or extensions of deadlines
- Any other action that, by omission or commission, compromises the integrity of the academic evaluation process
- For more details on Loyola's Academic Integrity Statement please see here.

Loyola University Chicago takes seriously the issues of plagiarism and academic integrity. Below is an excerpt, quoted directly, of the university's statement on integrity.

"The faculty and administration of Loyola University Chicago wish to make it clear that the following acts are regarded as serious violations of personal honesty and the academic ideal that binds the university into a learning community:

Submitting as one's own: 1. Material copied from a published source: print, internet, CD-ROM, audio, video, etc. 2. Another person's unpublished work or examination material. 3. Allowing another or paying another to write or research a paper for one's own benefit. 4. Purchasing, acquiring, and using for course credit a pre-written paper. The critical issue is to give proper recognition to other sources. To do so is both an act of personal, professional courtesy and of intellectual honesty."

Read through Loyola's full statement on Academic Integrity here.

Intellectual Property

All lectures, notes, PowerPoints, and other instructional materials in this course are the intellectual property of the professor. As a result, they may

not be distributed or shared in any manner, either on paper or virtually without my written permission. Lectures may not be recorded without my written consent; when consent is given, those recordings may be used for review only and may not be distributed. Recognizing that your work, too, is your intellectual property, I will not share or distribute your work in any form without your written permission.

Recordings

In this class software will be used to record live class discussions. As a student in this class, your participation in live class discussions will be recorded. These recordings will be made available only to students enrolled in the class, to assist those who cannot attend the live session or to serve as a resource for those who would like to review content that was presented. All recordings will become unavailable to students in the class when the Sakai course is unpublished (i.e. shortly after the course ends, per the Sakai administrative schedule). Students who prefer to participate via audio only will be allowed to disable their video camera so only audio will be captured. Please discuss this option with your instructor.

The use of all video recordings will be in keeping with the University Privacy Statement shown below:

Privacy Statement

Assuring privacy among faculty and students engaged in online and face-to-face instructional activities helps promote open and robust conversations and mitigates concerns that comments made within the context of the class will be shared beyond the classroom. As such, recordings of instructional activities occurring in online or face-to-face classes may be used solely for internal class purposes by the faculty member and students registered for the course, and only during the period in which the course is offered. Students will be informed of such recordings by a statement in the syllabus for the course in which they will be recorded. Instructors who wish to make subsequent use of recordings that include student activity may do so only with informed written consent of the students involved or if all student activity is removed from the recording. Recordings including student activity that have been initiated by the instructor may be retained by the instructor only for individual use.

Diversity Equity and Inclusion

The diversity that students bring to this class, in all its forms, is viewed as a resource, a strength, and a benefit. It is my intent to invest in each student's success and attend to each student's learning needs, both in and out of class.

It is my intent to present materials and activities that are respectful of diversity, equity and inclusion, and that students from all diverse backgrounds and perspectives be well-served by this course. Students in this course are encouraged to participate freely and share personal opinions, perspectives, and stories. There may be diverse, and perhaps contradictory ideas shared, in class. This variety is a strength of the academic community. Students are asked to show respect and treat peers in a way that validates various experiences and opinions based on a range of identities, including ability, economic class, ethnicity, faith tradition or no faith, gender identity and expression, nationality, religion, sexual orientation, veteran status, and their intersections.

Acts of bias, harassment, abuse, discrimination, relationship violence, sexual violence (i.e. sexual assault, sexual harassment, etc.), gender harassment, and stalking are not tolerated at Loyola. If you or someone you care about has experienced any one of these crimes and/or violations of LUC Community Standards, please know that you have rights, reporting options, and other support services available to you. Please visit here for more information.

Notice of Reporting Obligations for Responsible Campus Partners

As an instructor, I am a Responsible Campus Partner ("RCP") under Loyola's Comprehensive Policy and Procedures for Addressing Discrimination, Sexual Misconduct, and Retaliation. While my goal is for you to be able to engage fully and authentically with our course material through class discussions and written work, I also want to be transparent that as a RCP I am required to report certain disclosures of sexual misconduct (including sexual assault, sexual harassment, intimate partner and/or domestic violence, and/or stalking) to the Office for Equity & Compliance ("OEC"). As the University's Title IX office, the OEC coordinates the University's response to reports and complaints of sexual misconduct (as well as discrimination of any kind) to ensure students' rights are protected.

As an instructor, I also have an obligation under Illinois law to report disclosures of or suspected instances of child abuse or neglect `https://www.luc.edu/hr/legal-notices/mandatedreportingofchildabuseandneglect/`.

The University maintains such reporting requirements to ensure that any student who experiences sexual/gender-based violence receives accurate information about available resources and support. Such reports **will *not* generate a report to law enforcement** (no student will ever be forced to file a report with the police). Additionally, the University's resources and supports are available to all students even if a student chooses that they do not want any other action taken. If you have any questions about this policy, you are encouraged to contact the OEC at `equity@luc.edu` or 773-508-7766.

If you ever wish to speak with a confidential resource regarding gender-based violence, I encourage you to call The Line at 773-494-3810. The Line is staffed by confidential advocates from 8:30am-5pm M-F and 24 hours on

the weekend when school is in session. Advocates can provide support, talk through your options (medical, legal, LUC reporting, safety planning, etc.), and connect you with resources as needed – without generating a report or record with the OEC. More information about The Line can be found at `luc.edu/wellness`.

Land Acknowledgement

As we come together as a learning community, we need to acknowledge the land we live and work on by naming the Muscogee Creek, Cherokee, and Chickasaw Peoples upon whose unceded and stolen territory the university stands. Also, we should acknowledge the enslaved peoples, primarily of African descent, whose labour built much of the university. Visit LUC's Faculty Center for Ignatian Pedagogy land acknowledgement page for more information.

Campus Support Services

- ITS HelpDesk ✉ `helpdesk@luc.edu` ☎ 773-508-4487
- Library

  – Subject Librarian Greer Martin

- Student Accessibility Center
- Writing Center
- Ethics Hotline ☎ 855-603-6988
- Center for Tutoring and Academic Excellence
- Bookstore
- Financial Aid
- Wellness Center

  – Mental Health Appointment First Steps
  – For urgent, non-life threatening mental health needs ☎ 773-508-2530 **option 3**

Tentative Course Schedule

Week	Date	Content	HW/Projects
Week 01	Aug 26 (Mon)	Syllabus and Installation	
	Aug 28 (Wed)	Introduction to R, RStudio, and Reproducible Research	
Week 02	Sep 2 (Mon)	Labor Day **No Class**	
	Sep 4 (Wed)	Introduction to Python, Jupyter Notebooks, and Version Control	
Week 03	Sep 9 (Mon)	Loops and Function (R)	
	Sep 11 (Wed)	Loops and Functions (Python)	HW 1
Week 04	Sep 16 (Mon)	Data Explore (R)	
	Sep 18 (Wed)	Data Explore (Python)	HW 2
Week 05	Sep 23 (Mon)	Data Visualization (R)	
	Sep 25 (Wed)	Data Visualization (Python)	HW 3
Week 06	Sep 30 (Mon)	Data Wrangling for a Single Table (R)	
	Oct 2 (Wed)	Data Wrangling for a Single Table (Python)	Project Approval
Week 07	Oct 7 (Mon)	Fall Break **No Class**	
	Oct 9 (Wed)	Data Wrangling for Multiple Tables (Python)	HW 4
Week 08	Oct 14 (Mon)	Data Wrangling for Multiple Tables (R)	
	Oct 16 (Wed)	Midterm Exam	
Week 09	Oct 21 (Mon)	Working with Tidy Data, Reshaping Data (R)	
	Oct 23 (Wed)	Working with Tidy Data, Reshaping Data (Python)	HW 5
Week 10	Oct 28 (Mon)	Strings and Dates (R)	
	Oct 30 (Wed)	Strings and Dates (Python)	
Week 11	Nov 4 (Mon)	Iteration (Vectorized Functions, Map Family) (R)	
	Nov 6 (Wed)	Iteration (Vectorized Functions, Map Family) (Python)	HW 6
Week 12	Nov 11 (Mon)	Statistical Modeling (R)	
	Nov 13 (Wed)	Statistical Modeling (Python)	Project Plan
Week 13	Nov 18 (Mon)	Interactive Data Visualization (R)	
	Nov 20 (Wed)	Interactive Data Visualization (Python)	HW 7
Week 14	Nov 25 (Mon)	Mapping with sf in R	
	Nov 27 (Wed)	Thanksgiving **No Class**	
Week 15	Dec 2 (Mon)	SQL	
	Dec 4 (Wed)	SQL	HW 8

Statistics may be defined as "a body of methods for making wise decisions in the face of uncertainty."   — W.A. Wallis