PLS 202: Introduction to Data Analytics

Spring 2024

Time: Tue Thu 12:40 PM - 2:00 PM
Location: Akers Hall 137

Instructor: Hyerin Seo
Email: seohyeri@msu.edu
Teaching Assistant: Ha Eun Choi (choiha3@msu.edu)
Teaching Assistant: Srinija Kondur (kondurs1@msu.edu)
Teaching Assistant: Stephanie Tolly (tollyste@msu.edu)
Teaching Assistant: Jolie Kretzschmar (kretzsc9@msu.edu)
Instructor Office Location & Hours: South Kedzie 230, Tue 11:00 AM - 12:00 PM

Course Slack: Click to join
Course Schedule: Click to view

Course Description

Data analytics is a growing field that draws on many disciplines including Statistics, Computer Science, and Graphic Design. This course will provide an introduction to modern data analytics, with a focus on practical skills for social science research. The main goal of this course is to introduce R, a programming language designed specifically for statistics and data analysis. The course will cover basic programming concepts such as functions and data structures, as well as tools for data analysis and visualization. We will also cover methods for uncovering relationships in data, starting simple with cross-tabulations and correlations, before moving into more advanced topics like regression and linear models that are seen in most published pieces of scientific research.

This course is designed to be fun (at least to people like me!) and interactive. Many times you will likely run in to problems and find yourself googling errors and searching on stackoverflow (or even ChatGPT or Github CoPilot) and other forums for ways to do things. This is good! This is how even experts spend a lot of their time. I always joke that writing code is 80% googling (and really, it's not a joke).

Inclusion and diversity are core values of Michigan State University and my classroom. As Spartans, we are dedicated to respecting people of all backgrounds, beliefs, and identity status. I am committed to creating a safe, supportive, and welcoming environment where all students can pursue academic and personal success. We all deserve each other's respect, support, recognition, and protection.

Maintaining a respectful and inclusive community requires vigilance. We must, therefore, all stand up against derogatory and discriminatory language or actions whenever we see them. It is essential that we all work together to foster an inclusive community where Spartans of all backgrounds can study, work, and thrive.

Required Materials

In order to complete this course, you will need a computer, internet access, and access to D2L. We will be using Desire2Learn (D2L) as our online course management software, which is found at: `https://d2l.msu.edu`. You will be able to access course lectures, readings, quizzes, assignments, and your grades on D2L.

Any problems you encounter with D2L should be reported to MSU's Distance Learning Services, which is available 24 hours a day, 7 days a week. Their local number is 517- 355-2345. Their toll-free number is 1-800-500-1554.

In addition to D2L, you need to install R (the language) and RStudio (a program for writing, testing, and publishing R code). To download R, go to `https://cran.r-project.org/` and click the link for your operating system under the section titled "Download and Install R". Then follow the instructions to install the version appropriate for your specific system version. To download RStudio, go to `https://www.rstudio.com/products/rstudio/download/#download` and click the link for your operating system. This will download an installer file. Once the installer is downloaded, double click it to run, and then follow the instructions to complete the installation. This will be covered in the first lecture, as well.

Last, we'll be using Slack so that you guys have a faster and more real-time way to interact with me and the TA. To join click on the following link: `https://join.slack.com/t/pls202/shared_invite/zt-29tjil6xz-jl4f1BsUhuxybbQ50TDlJA`.

There are no required textbooks for this course, but there will generally be 1-2 assigned readings for each module. These are listed when relevant in the syllabus. If additional readings are posted on d2l, I will let you know by email.

Course Set Up

This course is broken down into six modules. Each module typically includes 1-2 readings, and 2 lectures. At the end of each module, we will have a module wrap-up session, in-class quiz, and assignment review session.

There will be a final assignment due the last week of class, as well. Below is an overview of the topics for the class. See the end of the syllabus for a complete schedule with readings and assignments for each module. See also here for a more detailed course schedule with specific dates. Please note that the syllabus is subject to change. Any updates will be communicated well in advance.
In terms of homework, I will always let you know when things are due and if anything ever seems uncertain please reach out to me on Slack or via email.

Lectures

The class begins with a lecture to introduce topics, but will quickly move to the part where I demonstrate how to write code in RStudio. I expect not simply to listen to the lecture, but to work

Module	Topics Covered
Introduction	Class Organization, Installation of R & RStudio
Module 1	Basics of R: objects, vectors, functions
Module 2	Advanced R: variables, datasets, data wrangling
Module 3	Descriptive Analysis: significance, hypothesis tests, correlation
Module 4	Data Visualization: plots, graphs, R package: 'ggplot2'
Module 5	Advanced Advanced R: loops, functions, apply, tidyverse
Module 6	Advanced Analysis: spatial data
Final Project will be given to you via D2L on 4/16 (Tue) and is due by 4/23 (Tue) at 11:59pm	

in RStudio alongside me and have written your own copy of the script by the time the lecture is over.

Please note that learning to work in a statistical computing language like R involves doing more than listening. The biggest advice I can give is to be patient with yourself, start assignments early, and ask questions often. I am always more than happy to chat with you.

Grading

Your final grade will be based on the following:

- Module Quizzes (Mod 1-5) : 20%
- Assignments (Mod 1-5): 50%
- Final Project: 30%

The grading scale for the course is as follows:

- 4.0 (92%-100%)
- 3.5 (86%-91.9%)
- 3.0 (80%-85.9%)
- 2.5 (75%-79.9%)
- 2.0 (70%-74.9%)
- 1.5 (65%-69.9%)
- 1.0 (60%-64.9%)
- 0.0 (<60%)

Assignments

At the end of each module, a homework assignment will be released on D2L. The due dates can be viewed on the Google sheet here. You must work on these assignments independently. Assignments must be turned in as a PDF document on D2L. Your homework will be graded on the following criteria:

- Were you able to "compile" your script?
- Does the program answer the question(s) given in the assignment?
- Does the program make correct use of the skills covered in the relevant course material?

Quizzes

Each module also has a corresponding in-class quiz, which will be conducted in an open-book format. However, the use of external resources from the Internet or AI tools such as ChatGPT is not permitted. The quiz asks questions based on the lectures from the module and any assigned reading from that same module. Each quiz contains around 10 questions. You are allowed two attempts to take each quiz, and your highest score will be saved.

Final Project

Rather than a final exam, on the last day of class you will be given a final project. This project will ask you to apply what you have learned in class to a real-world application of data analysis. It will look a lot like an extended module assignment (think 1.5 to 2 assignments in one).

Course Policies

Course Communication

Course announcements will often be sent out via the email list provided by the registrar. It is your responsibility to make sure you can readily access any emails that are sent to your MSU email address. Announcements will also be posted in D2L when appropriate.

Instructor Availability

Questions may be communicated to me or the teaching assistant (TA) by email or Slack. When you are writing an email, do not forget to **include "PLS 202" in the subject line.**

Late Work

Work that is not handed in by 11:59 PM EDT of the due date is late. Late assignments will receive a:

- 10% penalty if submitted less than an hour late,
- a 20% penalty if 1-4 hours late,
- a 30% penalty if 4-12 hours late,
- and 50% penalty if 12-24 hours late;
- you will receive no credit for anything submitted later than 24 hours past the due date.

Grade Appeals

People make mistakes. If you think there is an error in your recorded grade, then you should email me as soon as possible.

Academic Dishonesty

All University policies on Academic Dishonesty apply to this course. Academic dishonesty includes but is not limited to, plagiarism, falsification/fabrication, tampering, cheating, or sharing work. More information on academic integrity and procedures followed for violation can be found here.

Learning Needs

Any student who may need an accommodation because of any disability should contact MSU's Resource Center for Persons with Disabilities (`http://rcpd.msu.edu/`) at 130 Bessey Hall within the first two weeks to provide me with information for proper accommodation. If you have any questions please feel free to contact them or ask me, all information and documentation will be kept confidential.

Children in the Classroom Policy (adapted from Dr. Melissa Cheyney's Syllabus)

Taking care of young children should not keep you from participating in class. If you are a caregiver of young children, those children are welcome when you cannot find or afford child care. In order to ensure an inclusive space and preserve a conducive learning environment for all enrolled students, caregivers of babies and children are encouraged to select seats closer to the door. This arrangement allows for easy exit if a child requires special attention or if their presence interrupts other students' learning. Finally, it is expected that all students will help in creating an environment that welcomes all forms of student diversity, including diversity in parenting status.

Artificial Intelligence (AI) Policy

AI tools can be pretty handy and learning how to use them is a valuable skill these days. You can definitely try using AI like ChatGPT in our class, but you don't have to go all out. Just remember, there are some things to keep in mind when dealing with ChatGPT:

- If you just throw in minimum effort prompts, you will get low quality results. You might need to play around with your questions to get good stuff. It might take a bit of effort.
- Don't trust anything it says. If it gives you a number or fact, assume it is wrong unless you either know the answer or can check in with another source. You will be responsible for any errors or omissions provided by the tool.
- AI is a tool, but one that **you need to acknowledge using.** Please include a sentence at the end of any assignment that uses AI explaining what you used the AI for and what prompts you used to get the results. Failure to do so is in violation of academic honesty policies.
- Be thoughtful about when this tool is useful. Don't use it if it isn't appropriate for the case or circumstance.

Religious Observance Policy

As stated in the University Policy on Religious Observance, "it is the responsibility of those students who need to be absent to inform their instructor at least two weeks before the observance day, make arrangements in advance with their instructors, and catch up on any material discussed and assignments given during that class period."

Final Caveat

Stuff happens and I like to try and tailor courses to how students are doing, so I reserve the right to make changes to the syllabus as needed.

Semester Schedule

Find below a detailed overview of the semester, including the topics of lecture and assigned readings.

Introduction

Learning Goals

- Understand setup and expectations and of class
- Download and install the R programming language
- Download and install the application RStudio

Readings

- "We are All Social Scientists Now"
- "History and Overview of R" (Chp 2 of Roger Peng's *R Programming for Data Science*).

Module 1

Learning Goals

- Learn the basics of R
- Create objects and vectors
- Create very simple plots

Readings

- Starting out in R (Click here to access Chp 1 of *Introduction to R* by Paul Harrison)
- Wickham Style Guide

Module 2

Learning Goals

- Import data and summarize variables
- Subset variables and datasets
- Create new variables
- Deal with missing values
- Plot variables

Readings

- Data Types (Click here to access Chp 2.4 from *Introduction to Data Science* by Rafael A. Irizarry)

Module 3

Learning Goals

- Understand measures of central tendency and dispersion
- Difference in means tests and correlations
- Tests for statistical significance

Readings

- Click here to access Statistical Significance – Introduction to Psychology

Module 4

Learning Goals

- Rules for effective data visualization
- Using package ggplot2 to make publication-quality graphics

Readings

- Click here to access A Brief Guide to Designing Effective Figures for the Scientific Paper
- Click here to access Do's and Don'ts for Effective Graphs

Module 5

Learning Goals

- Quickly make multiple calculations using loops
- Create new variables with ifelse statements
- Use the apply function
- Introduction to tidyverse

Readings

- Click here to access Intro to dplyr (Chp 3 from *A Gradual Introduction to the Tidyverse* by Ismay & Laderas)

Module 6

Note that we might not get this far.
Learning Goals

- Applications of techniques we covered in class to "real world" projects

Readings

- None

Final Project posted to D2L on 4/16 (Tue) and due 4/23 (Tue) by 11:59pm