Discovering Data Science

Summer 2022

Instructor Information

Lead Instructor: Paul C. Holaway
Email: paulch2@illinois.edu
Asst. Instructor: Abhi Thanvi
Email: athanvi2@illinois.edu

Class Schedule[1]

9:00 AM - 10:00 AM: Guest Speaker
10:00 AM - 10:05 AM: Break
10:05 AM - 11:15 AM: Lecture
11:15 AM - 11:20 AM: Break
11:20 AM - 12:30 PM: Lab/Group Discussion

Course Resources

Course Webpage: We will be using our course webpage to post notes, labs, and other documents such as this syllabus. Any course document will be posted here. It is the student's responsibility to regularly check the course webpage for new documents.

DISCOVERY PARTNERS INSTITUTE
PART OF THE UNIVERSITY OF ILLINOIS SYSTEM

Discord: We will be using Discord[2] as a course discussion forum and for posting course announcements. This is so students may ask additional questions outside of class and or office hours. Students are highly encouraged to post any questions they have in the Discord. Students are also highly encouraged to respond to another student's question if they can help out. However, most questions will receive a response from the instructors within 24 hours Monday through Friday. It is not appropriate to post answers to labs on Discord. Any student who does this will have their post deleted and given a warning by the instructors. Repeat offenses will result in a posting timeout and an academic integrity violation.

Gradescope: All labs and the final project will be submitted on Gradescope[3]. All grades and assignment feedback will be posted on Gradescope as well. It is the student's responsibility to check Gradescope regularly for updated grades and assignment feedback.

---

[1] Paul and Abhi will only be leading the 10:00 AM - 1:00 PM portion.
[2] Discord Link: https://discord.gg/GNJnDUyD
[3] Gradescope Course Entry Code: **PX56BR**

Course Description

Discovering Data Science is an introduction to topics that intersect between statistics, computation, and real-world applications. This 5-week-long course will be project-driven, which will give students opportunities to analyze real-world datasets and discover the impact of the data. Throughout each experience, students reflect on the social issues surrounding data analysis such as privacy and design.

Course Materials

Laptop/Desktop Computer: DPI has given you a laptop to use for this course that already has R and RStudio installed. Therefore you do not have to work on any kind of major software installation. If you have another computer you would like to work on that is also fine. However, if it does not have R and RStudio downloaded, please come to office hours so your instructors may help you with the setup. We will not be covering the setup of R or RStudio in class.

Lecture Binder/Folder: All notes will be posted on the course webpage the evening before the lecture/lab for students to look at in advance if they wish.[4] Students may either print out the notes sheet or download them and write/type in as the instructors go through them. There will be both handwritten and coding examples done during the lectures. It may also be useful to have a notebook or loose-leaf paper for handwritten examples. I highly encourage students to keep notes in a physical and or digital folder to aid in organization.

Assignments

Labs: We will have daily lab assignments Monday through Thursday. There will be three different types of lab problems; hand-written, coding, and short-answer. Hand-written problems will be to help students understand concepts and the mathematics of what the code is doing. Coding problems will be to help students learn how to use `RStudio` to complete their projects more efficiently. Short-answer problems will be to help students learn to interpret their results and to think outside the box. Some labs will include problems specifically designed to aid students in their Capstone Projects. All labs will be due on Gradescope the day they are assigned at 11:59 PM CDT. Students will be required to submit an .Rmd and .pdf file to Gradescope for each lab.

Feedback Surveys: At the end of every Thursday lab, we will have a feedback survey for the week. You may write as much or as little as you want. However, you must write something. It will factor into a portion of your lab grade.[5] If you have nothing to say, then just say that. This feedback will be useful for us to adjust how we are teaching and running the course as the

[4] We know this will not happen, but will do it anyway. :)

[5] We know students would give us feedback anyway, so they might as well get credit for it.

DPI DISCOVERY PARTNERS INSTITUTE PART OF THE UNIVERSITY OF ILLINOIS SYSTEM

summer goes along. We will consider all feedback, but cannot implement (most likely) all of it. Please be somewhat formal and if you have anything you do not like, be polite and offer constructive criticism. Any use of explicit language or disrespectful responses will be reported to the UIC Chance staff and the student(s) will be subject to the UIC Chance code of conduct for students.

Capstone Project: Students will begin working on their projects on day one. Students will gradually work on their projects throughout the course with the help of the in-class examples, lab assignments, and office hours. Students will complete a professional data analysis report and present a summary of their analysis and findings in a 5-10 minute presentation with slides.[6] The final project report will be due Sunday, July 17th on Gradescope at 11:59 PM CDT.

Late Submissions

No late submissions will be accepted. If you need help, please ask as soon as possible.

Course Grades

There are no letter grades assigned in this course. However, there will be point grading to track completion and performance for each student. Only lab assignments will be graded and each lab assignment will be worth 20 points. We will also grade and provide feedback on your projects at the end of the course. The project will be worth 200 points. You may also stop by during office hours to discuss your project and receive feedback during the semester.

Office Hours

Office hours will be held Monday through Wednesday except for the first day of class from 3:00 PM - 5:00 PM CDT. Additional office hours will be held on Friday, July 15th for final project questions from 3:00 PM - 5:00 PM CDT. All office hours will be held over Zoom.[7]

DISCOVERY PARTNERS INSTITUTE
PART OF THE UNIVERSITY OF ILLINOIS SYSTEM

Learning Collaboratively

Data science is a collaborative science. Do not try to tackle this course alone. We strongly encourage you to discuss all of your course activities with your friends and classmates. You will learn more through talking through the problems, teaching others, and sharing ideas. Continue to read "Academic Integrity" to understand the difference between collaboration and straight-up giving an answer to another student.

Technical Issues

If you are experiencing technical issues, please reach out to…

[6] See Capstone Project details on the course webpage.
[7] https://illinois.zoom.us/j/86715198632?pwd=fbtqcankNPoklpU2L23or7EgU_kRi_.1

- Your instructors if it is code related, RStudio related, or Gradescope related.
    - Please email both instructors simultaneously to receive the quickest response.
- A UIC Chance staff member if it is related to any other issue.

Academic Integrity

Collaboration is about working **together**. Collaboration is **NOT** giving the direct answer to a friend or sharing the source code to an assignment. Collaboration requires you to make a serious attempt at every assignment and discuss your ideas and doubts with others so everyone gets more out of the discussion. Your answer **must** be your own words and your code must be typed (not copied/pasted) by you. Academic dishonesty is taken very seriously. All violations will be reported to the UIC Chance staff and the student(s) will be subject to the UIC Chance code of conduct for students. Academic integrity also includes you protecting your work. If you share your work directly with others and they submit your work as their own, this will be considered an academic integrity violation as if you submitted someone else's work as your own.

Instructor's Notes

1. Please do not expect quick responses to emails or Discord posts after Friday 5:00 PM CDT. Any responses received after that may not be answered until Monday morning.
2. Paul will be attending his sister's wedding the weekend of July 8th - 11th. Any urgent matters should be directed to Abhi during this timeframe to receive a quick response.

Special Thanks

We would like to thank the following individuals for their aid in making this course possible.

- Dave Dalpiaz; University of Illinois Urbana-Champaign, STAT385
- Karle Flanagan; University of Illinois Urbana-Champaign, STAT107, Faculty Advisor
- Wade Fagen-Ulmschneider; University of Illinois Urbana-Champaign, STAT107
- Mark Harris; Discovery Partners Institute
- Christopher Kinson; University of Illinois Urbana-Champaign, STAT440
- Kay Monelle; Discovery Partners Institute
- Jonas Reger; University of Illinois Urbana-Champaign, Summer 2021 Instructor
- Alex Stepanov; University of Illinois Urbana-Champaign, STAT410

The following course material has been adapted from last year's Discovering Data Science course, the University of Illinois Urbana-Champaign's STAT107 (Data Science Discovery), with parts from STAT385 (Statistical Programming Methods), STAT410 (Statistics and Probability II), and STAT440 (Statistical Data Management) with permission from the instructors.

DISCOVERY PARTNERS INSTITUTE
PART OF THE UNIVERSITY OF ILLINOIS SYSTEM