STAT 4380 | Data Science

Section 001 | Spring 2022

Instructor Contact

Instructor: Jacob Rozran
E-mail: jacob.rozran@villanova.edu

Meeting Times

Meeting Time: 8:30 – 9:45 AM on Tuesdays and Thursdays
Meeting Place: Mendel Hall G87 (subject to change)

Office Hours: Office Hours by Appointment
Office Location: Contact me to set up an appointment; I don't have a formal office.

Important Class Information

Course Description: Data-savvy professionals are in high demand in businesses, public agencies, and nonprofits. The supply of professionals who can work effectively with data at scale is limited and is reflected by rapidly rising demand and salaries for data scientists. This course explores how real-world data from a variety of disciplines are gathered, managed, and used for making decisions or predictions. This course will introduce students to the statistical software R, one of the most popular in data science practice.

Course Objectives: At the conclusion of the course, the students will be able to use the R programming language to do the following. I have broken this into objectives by specific topics.

- Data Science
  - Understand the role of data science and its demand in society
  - Discuss how choices about data collection, extraction and processing affect bias and generalizability
  - Identify and describe ethical issues related to data access and use
  - Communicate results to technical and non-technical audiences
- Data Visualization
  - Summarize data using appropriate data visualization techniques
  - Use aesthetics to explore multivariate relationships
  - Use layers to augment visualizations

VILLANOVA UNIVERSITY

- Data Wrangling/Reading
    - Import/export data using various formats
    - Scrape data from web sources
    - Transform data into a tidy format
    - Process data to clean it and summarize over groups
    - Work with dates and strings
- Statistical Thinking
    - Reinforce concepts of bias and multivariable thinking (confounding)

Textbook: Modern Data Science with R, Second Edition, by Baumer, Kaplan, and Horton. You can either access for free on the MDSR website or pay for a printed copy from CRC Press or Amazon. The book also has a main website and a GitHub repository.

Course Format: This class will be a mix of pure lecture (where I am usually introducing a topic), in-class exercises (where we then practice the topic using code), and in-class project time (where the class is working together and I am available for assistance).

Course Content: Will be posted in a Google Drive folder and on GitHub (though using GitHub is entirely beyond the scope of this course and meant only as a supplemental resource for those that feel it will benefit them). This will include the syllabus, code, and any materials I present.

Blackboard: I will likely use Blackboard very little in this class.

Zoom: I will send out Zoom invites as necessary (appointments or any class changes).

Grading: Your grade will be determined by your participation (10%), homework average (25%), your score on two mid-term projects (20% each), and your score on the final term project (25%). Grades will be assigned either according to the following scale or according to one more favorable to you:

A	92–100	B-	79–81	D+	66–68
A-	89–91	C+	76–78	D	62–65
B+	86–88	C	72–75	D-	59–61
B	82–85	C-	69–71	F	< 59

Computing: We will use the R programming language and RStudio Interactive Development Environment (IDE) almost exclusively for in-class work, homework assignments, and projects. You can find instructions on installing these in our Google Drive and GitHub site; we will spend time in class ensuring everyone is up and running.

Homework: Homework will be assigned as material is covered. Some assignments may be collected, some assignments will not be collected. I encourage you to work with other students in solving the homework problems. However, it is to your advantage to make sure that you can do all the problems on

your own. Some solutions will be made available in the course Google Drive. I will drop your one lowest grade when computing your homework average.

How to Do Well: Attend every class meeting and be an active participant. Begin doing the homework as soon as we cover the relevant material in class. Take advantage of my office hours, and ask questions as soon as you have them. Since the later material builds on the earlier material, waiting to ask questions may not be a good strategy.

Attendance: Attendance is essential if you wish to do well in this course, and it is expected that you will attend each class meeting. If you do miss a class, it is your responsibility to find out what was covered and what was assigned. A good first step is to get the notes from one of your fellow students.

Class & Semester Dates

Tentative Dates Project Completion:

- Individual Project -
- Group Project - April 12
- Final Project - Dates TBD

Other Important Dates:

- March 1 – No class due to Spring Recess.
- March 3 – No class due to Spring Recess.
- April 14 – No class due to Easter break.

Tentative Class Schedule:

- Week 1: Introductions + Chapters 1 + Appendix B
- Week 2: Appendix C
- Week 3: Appendix D
- Week 4: Chapter 2 & 3
- Week 5: Chapter 3 & 4
- Week 6: Chapter 4 + Introduce Data Viz Recreation Project
- Week 7: Chapter 5 + Introduce Final Project
- Week 8: Semester Recess
- Week 9: Chapter 5 & 7
- Week 10: Data Viz Recreation Project Presentations
- Week 11: Chapter 7 + Introduce Data Wrangling & Viz Project
- Week 12: Chapter 6
- Week 13: Chapter 6 & 19 + Data Wrangling Project Due
- Week 14: Chapter 19 + Easter Break
- Week 15: Chapter 13
- Week 16: Chapter 8

- Week 17: Final Project Presentations

Other Resources

Masks & Health: All faculty, staff and students—regardless of vaccination status—are required to wear masks when inside classrooms and any public campus building including dining halls (except while eating or drinking) that is open to the public.

Academic Integrity: All students are expected to uphold Villanova's Academic Integrity Policy and Code. Any incident of academic dishonesty will be reported to the Dean of the College of Liberal Arts and Sciences for disciplinary action. For the College's statement on Academic Integrity, you should consult the Student Guide to Policies and Procedures. You may view the University's Academic Integrity Policy and Code, as well as other useful information related to writing papers, at the Academic Integrity Gateway web site: https://library.villanova.edu/research/subject-guides/academicintegrity

Office of Disabilities (ODS) and Learning Support Services (LSS): It is the policy of Villanova to make reasonable academic accommodations for qualified individuals with disabilities. Go to the Learning Support Services website http://learningsupportservices.villanova.edu for registration guidelines and instructions. For physical access or temporarily disabling conditions, please contact the Office of Disability Services at 610-519-3209 or 610-519-4095, or email ods@villanova.edu. Registration is needed in order to receive accommodations.

Absences for Religious Holidays: Villanova University makes every reasonable effort to allow members of the community to observe their religious holidays, consistent with the University's obligations, responsibilities, and policies. Students who expect to miss a class or assignment due to the observance of a religious holiday should discuss the matter with their professors as soon as possible, normally at least two weeks in advance. Absence from classes or examinations for religious reasons does not relieve students from responsibility for any part of the course work required during the absence. https://www1.villanova.edu/villanova/provost/resources/student/policies/religiousholidays.html