Syllabus

SDS 192 (02): Introduction to Data Science, Spring 2026

Overview

Contact

- Instructor: Jericho Lawson
- Email: jlawson01@smith.edu
- Office Hours:

    - Mondays, 3-4pm
    - Fridays, 10am-12pm
    - McConnell 207

Location/Time

- Mondays, 1:40 - 2:55pm in Sabin-Reed 301
- Wednesdays & Fridays, 1:20 - 2:35pm in Sabin-Reed 301

Description

Data science involves applying a set of strategies to transform a recorded set of values into something from which we can glean knowledge and insight. This course will introduce you to concepts and methods from the field of data science, along with how to apply them in R. You will learn how to acquire, clean, wrangle, and visualize data. You will also learn best practices in data science workflows, such as code documentation and version control. Issues in data ethics will be addressed throughout the course.

Learning Outcomes

By the end of this course, students will be able to do the following:

- **Data visualization**: Create informative and clean visuals through various types of plots. Appropriate statistics and metrics are displayed. Use of base R and advanced libraries to generate plots (i.e. ggplot2).
- **Data wrangling**: Develop techniques to properly use data in a workflow, which includes the transformation of data, data cleaning, and data joining. Students will also gather data in appropriate and ethical manners.
- **Workflow**: Practice common tasks to working through particular data science problems, such as data retrieval, R programming, GitHub, and light statistics.
- **Data ethics**: Identify best practices and ethical dilemmas that stem from data science work. This includes contextual thinking, environmental concerns, and intellectual property. Students will also develop ethical ways to use generative AI.

Materials

Textbooks

- Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2024). *Modern Data Science with R (3rd ed.)*. CRC Press. https://mdsr-book.github.io/mdsr3e/

Optional:

- Ismay, C., Kim A. Y., & Valdivia A. (2025). *Statistical Inference via Data Science: A ModernDive into R and the Tidyverse (2nd ed.)*. CRC Press. https://moderndive.com/v2/
- Irizarry, R. A. (2022). *Introduction to Data Science: Data Analysis and Prediction Algorithms with R*. CRC Press. https://rafalab.dfci.harvard.edu/dsbook-part-1/

Technology

- For coding: R, RStudio
  - For downloading/installing: Instructions
- For assignment turn-ins: GitHub
  - For joining GitHub classroom: To join
- For discussion assignments: Perusall
  - For joining Perusall course: To join

        ∗ Enrollment code: LAWSON-2BX49

- For course management: Moodle, but will be used seldomly
- For course materials: This website
- For discussions and announcements: Slack

Grading and Expectations

Breakdown

Formative		Summative	
Readings, Activities, Problems	10%	Projects (2)	30%
Labs	30%	Exams (2)	30%

A traditional grading system will be used here. Cumulative numerical averages of 90-100 are guaranteed at least an A-, 80-89 at least a B-, and 70-79 at least a C-. The exact ranges for letter grades will ultimately be determined at the end of the course.

MW Classes, Readings, Activties, and Exercises (10%)

Prior to the upcoming week, readings will be assigned through Perusall that prepare you for material that will be explored thereafter. The readings on Perusall will involve:

- Reading through the material
- Writing 2-4 annotations
- Highlighting one item you learned and one item that seems challenging

These annotations will allow me to identify which concepts to focus on in more detail during the week. Generally, these items will be due on **Sundays at 11:59pm**, with no acceptance of late work. Readings will be graded on **completion**.

During Mondays and Wednesday classes, we will go through mini-lectures, activities, and demos. Classes will be experiential, meaning that you will have the opportunity to apply concepts from class to activities, exercises, and investigations. Students are expected to participate during lecture by asking questions and working with other students. Additionally, it is required that you bring your laptop to class. Class activities will be administered often during lecture, and you will have until **the end of lecture** to turn them in through your GitHub submission repository. These activities will be graded based on **completion and/or correctness.**

On occasion, a reading exercise may be assigned on Wednesday for you to complete. Similar to the activities, these will be turned in through your GitHub submission repository by **Thursdays at 11:59pm**. These exercises will be graded based on **completion and correctness**.

Friday Classes and Labs (30%)

Friday's classes are dedicated exclusively to **labs** unless otherwise stated. In these labs, you will have the opportunity to compile (no pun intended) what you've learned and work on a experiential assignment that showcases those skills. This will involve strong usage of R and GitHub. In earlier weeks, we will get you all familiar with the technical programs along with the workflow.

Labs in this course introduce a data science skill by walking you through exploratory analysis of a dataset documenting a socially-relevant issue (such as racial profiling in policing, affordable housing, and pollution). Labs will be started in class on Fridays and completed for homework by **Wednesdays at 11:59pm**. If you finish a lab early, you are encouraged to help your classmates. Labs will be graded on both **completion and correctness.** Note that at the end of each lab, there will be a prompt asking you to consider some of the ethical considerations of the data analysis you just completed. You must respond to this prompt in Slack to earn full credit on the lab.

All lab assignments will be submitted through your GitHub submission repository. You will submit assignments by pushing changes to template documents to a private GitHub repository. I will provide guidance on how to do this early in the semester.

Projects (30% total)

Two projects will be assigned throughout the semester. The projects are designed to test your ability to successfully complete course objectives as described earlier. The following two projects, along with their tentative due dates, are seen below:

- Project #1: Data Visualizations and GitHub: **due Wednesday, 3/4 at 11:59pm**

  - Create meaningful, informative, and clear visualizations that communicate information from the data and real-world issue.
  - Configure an appropriate pipeline that funnels all relevant information, files, and summaries into a GitHub repository.

- Project #2: Workflow and Real-World Investigation: **due Monday, 4/27 at 11:59pm**

  - Complete a detailed, attentive workflow of an investigation related to a real-world issue with appropriate data and context.
  - Present relevant information and code to various audiences in multiple formats, including a report, presentation, and demonstration.

More details to come in later weeks.

Exams (30% total)

Two exams will be administered during the course of the semester, which will include written and oral parts. The first exam will be administered in-class on **Friday, 3/13**, while the second exam will be administered during finals week as a self-scheduled exam. More details to come in later weeks.

Policies

Preparation and Attendance

As a four-credit course that meets 4.5 hours per week, Smith expects students to dedicate at least 7.5 hours per week towards the course outside of class. The assignments, readings, and assessments are designed with this target in mind. Expect to read through lecture notes, examples, articles, and forums outside of class along with the assignments.

Attending class is imperative to your learning and taking part in an active community. Attendance will be taken for each class period. However, things come up and you may not have the capacity to attend. As such, you will be able to miss **3 classes** with no penalty. After the third unexcused absence, your overall grade will drop by 1% for each class missed. Additional absences may be excused due to family/personal difficulties, sickness, or school or career-related activities; however, I will require some form of documentation for these absences. Please speak with your class dean or the Accessibility Resource Center so that we can get documentation of your need.

Do make every effort to arrive to class on time. If you happen/plan to arrive more than 10 minutes late, please inform me ahead of time. Any unexplained tardiness will result in a marked absence. If you must miss a class entirely, you should contact a peer to discuss what was missed.

Extensions

You will be granted **2 free late days** to use for lab assignments, with a maximum of one day used on each lab assignment. Additionally, you will be granted **2 late days** to use for final project submissions, with a maximum of two days used at a single time. No need to inform me that you intend to take these late days. Beyond this, late assignments will not be accepted without an accommodation from a class dean or from the ARC.

Note that this policy does not apply to Perusall annotations, reading exercises, exams, or project checkpoints/proposals/presentations.

Academic Honesty

As a student at Smith College, the college expects all students to be honest and committed to the principles of academic and intellectual integrity in preparation and submission of all course work and examinations, as outlined by the Academic Integrity Board (AIB). The AIB provides an Academic Integrity Statement, which all students are expected to abide by. Any cases of academic dishonesty or plagiarism will be reported to the Academic Honor Board. Examples of these behaviors include:

- Submitting work completed by another student as your own.
- Copying and pasting words from sources without quoting and citing the author.
- Paraphrasing material from another source without citing the author.
- Failing to cite your sources correctly.
- Falsifying or misrepresenting information in submitted work.
- Paying another student or service to complete assignments for you.
- Submitting work generated by artificially intelligent tools such as ChatGPT without permission or instruction to do so.

You are encouraged to discuss course material, including assignments, with your classmates. All work you turn in, however, must be your own. This includes both writing and code. Copying from other students, from books, or from websites (1) does nothing to help you learn how to program, (2) is easy for us to detect, and (3) has serious negative consequences.

Generative AI

As mentioned in the Academic Integrity Board, the professor for each course decides whether and how students are allowed to use generative AI in a given course. For this specific course, any use of generative AI to complete assignments or produce content for this course is **prohibited**, unless otherwise stated in the assignment itself.

As a foundational course, it is critical that you are able to think like a data scientist, which includes producing meaningful code, developing logical solutions to problem, and critiquing good results from the bad ones. While the use of generative AI can be beneficial at producing base-level code and providing insights, it comes at the detriment of your own critical thinking. The human element will be critical in succeeding in data science. If you don't develop the skills to understand how the underlying code is composed/works, then you will not be prepared for this kind of work.

Prohibited forms of generative AI usage include but are not limited to:

- Summarizing course readings.
- Drafting, editing, and proofreading responses to written prompts on any assignment.
- Composing and/or formatting code and comments.
- Answering lab, quiz, or exam questions.

- Conducting analysis of any plots, methods, and results for any assignment.

Any unauthorized use of generative artificial intelligence in this course will be considered a case of academic dishonesty/plagiarism and will be reported to the Academic Honor Board.

Caveat: While the policy here is strict, do note that there will be assignments and lectures that go through the ethical and effective use of generative AI in data science as part of a unit. In these assignments, further directions will be given to showcase the use of generative AI. However, unless otherwise stated, generative AI should not be used in any assignment.

Community & Support

Code of Conduct

As the instructor for this course, I am committed to making participation in this course a harassment-free experience for everyone, regardless of level of experience, gender, gender identity and expression, sexual orientation, disability, personal appearance, body size, race, ethnicity, age, or religion. Examples of unacceptable behavior by participants in this course include the use of sexual language or imagery, derogatory comments or personal attacks, trolling, public or private harassment, insults, or other unprofessional conduct.

As the instructor I have the right and responsibility to point out and stop behavior that is not aligned to this Code of Conduct. Participants who do not follow the Code of Conduct may be reprimanded for such behavior. Instances of abusive, harassing, or otherwise unacceptable behavior may be reported by contacting the instructor.

All students and the instructor are expected to adhere to this Code of Conduct in all settings for this course: seminars, office hours, and over Slack. This Code of Conduct is adapted from the Contributor Covenant, version 1.0.0, available here.

Principles of Community

Whether in a class, college, or neighborhood setting, achieving a warm community is essential to your well-being. In this class, I hope we can foster a collaborative and welcoming environment: one that celebrates successes, respects individual strengths and weaknesses, demonstrates compassion for each other's struggles, and affirms diverse identities.

To establish this, consider the following:

- Check-in with colleagues before starting collaborative work.
- Consider when to step up and when to step back in class discussions, creating space for others to contribute. Listening is just as important to community-building as speaking.
- Acknowledge what we do and don't know, as well as how our colleagues experience the world.

- Support colleagues that may be stepping outside of their comfort zone (i.e. presentations).
- Ask questions often in our Slack workspace. Help each other out by answering questions when you can.
- Admit mistakes. They happen, and I will certainly make mistakes in class.
- Use pronouns. This provides a foundation to a safe, respectful classroom environment that creates a sense of trust. For information on pronouns and usage, please see the Office of Equity and Inclusion link here: Pronouns

Accommodations

It is my goal for everyone to succeed in this course. If you have personal circumstances that may impact your experience of our classroom, I encourage you to contact the Accessibility Resource Center in College Hall 104 or at arc@smith.edu. The Center will generate a letter that indicates to me what kind of support you need and how I can make your classroom experience more accommodating. Once you have this letter, you are welcome to visit my office hours or email me to discuss ideas about how we can tailor the course accordingly. While you can request accommodations at any time, the sooner we start this conversation, the better. If you have concerns about the course that are not addressed through ARC, please contact me. At no point will I ask you to divulge details about your personal circumstances to me.

Student Well-Being

College life is stressful, and life outside of college can be overwhelming. It is my position that attending to your physical and mental health and well-being should be a top priority. I will remind you of this often throughout the semester. I encourage you to schedule a time to talk with me if you are struggling with this course. If you, or anyone you know, is experiencing distress, there are numerous campus resources that can provide support via the Schacht Center.

Additional resources and support offered by the college are listed below:

- Accessibility Resource Center (ARC)
- Spinelli Center: Support for students doing quantitative work. Includes tutoring and resources.

    - For this class: Sun-Thurs, 7-9pm in Sabin-Reed 301.
    - Email qlctutor@smith.edu for specific request for help.

- Crisis Resources
- Counseling Services
- Wellness Resources
- Gender Identity and Expression
- Discriminatory Harassment

Course Outline

Week	Dates	Topics	Ch.	Assignments / Notes
1	1/26-30	Introduction to Data	1	
2	2/2-6	Intro to R and Workflows	1, B	
3	2/9-13	Data Visualization	2	
4	2/16-20	ggplot2	3	
5	2/23-2/27	Data Wrangling	4	
6	3/2-3/6	Data Joining	5	Project #1 due: Wed, 3/4
7	3/9-13	Statistics, Exam #1	9	Exam #1: Fri, 3/13
8	3/16-20	Spring Break!		*No class all week long*
9	3/23-27	Data Tidying	6	
10	3/30-4/3	Programming	7	
11	4/6-10	Ethics and Programming	8, C	
12	4/13-17	Spatial Thinking	17	
13	4/20-24	Spatial Thinking	17	
14	4/27-5/1	APIs and Review	18	Project #2 due: Mon, 4/27
Finals	5/2-9	Finals	NA	Exam #2: self-scheduled (5/6-9)

- Note that the assignments and topics covered are tentatively scheduled and may be altered slightly as the quarter progresses. The instructor has the right to modify the syllabus if needed. If this occurs, students will be notified before the change occurs.