Basic information

- **Course title:** SDS 192: Introduction to Data Science
- **Contact Info:**
  @ nschwab@smith.edu
  413-585-3440
  Nics-Github nschwab
- **Office location:** McConnell 213
- **Email:** nschwab@smith.edu
- **Meeting locations/times:**

Morning Section

```
-   MWF: 10:50 AM – 11:05 PM / Sabin–Reed 220
```

```
Afternoon Section
```

```
-   M: 1:40 AM – 11:55 / Sabin– Reed 301 WF: 1:20 AM – 12:35 / Sabin– Reed 301
```

- **Getting help**:
  - For non personal/sensitive questions, ask on the `#questions` channel on Slack. For personal/sensitive matters, Slack DM Professor Schwab.
  - Prof. Schwab's student hours:
    - W 12:15 - 1:15 pm
    - F 12:15 - 1:15 pm
    - Request a zoom appointment.
  - Spinelli Center

Instructor work-life balance

- I will respond to Slack messages sent during the week within 24h. I will respond to Slack messages sent during the weekend at my own discretion.
- If possible, please only Slack me with briefer and administrative questions; I prefer having more substantive conversations in person as it takes me less energy to understand where you are at.
- I will do my best to return all grading as promptly as possible.

How can I succeed in this class?

Ask yourself:

- **When I have questions or don't understand something:**
  - *"Am I asking questions in class?"*
  - *"Am I asking questions on Slack in the* `#questions` *channel?"* Even better: *"Am I answering my peers' questions on Slack?"*
  - *"Having I been going to the Spinelli tutoring center for help on R and the tidyverse?"*
  - *"Have I been coming to office hours?"*
- **Lectures and readings:**
  - *"Am I staying on top Slack notifications sent between lectures?"* If you need help developing a notification strategy that best suits your lifestyle, please speak to me.
  - *"Am I attending lectures consistently?"*
  - *"During in-class activities, am I actually running code line-by-line and studying the outputs, or am I just going through the motions?"*
  - *"During in-class exercises, am I taking full advantage that I'm in the same place at the same time with the instructor, the lab assistants, and most importantly your peers, or am I browsing the web/texting the whole time?"*
  - *"Have I been doing the associated readings for each lecture?"*

Course Description & Objectives

An introduction to data science using Python, R and SQL. Students learn how to scrape, process and clean data from the web; manipulate data in a variety of formats; contextualize variation in data; construct point and interval estimates using resampling techniques; visualize multidimensional data; design accurate, clear and appropriate data graphics; create data maps and perform basic spatial analysis; and query large relational databases. SDS 100 is required for students who have not previously completed SDS 201, SDS 220, SDS 290 or SDS 291.

4 Class Hours 4 Credits

BASIC OUTLINE OF COURSE CONTENT:

- Introduction to Programming Language for Data Science (R or Python) and Computing Environment (RStudio, Jupyter Notebook, etc.)
- Data Science Life Cycle
- Grammar of Graphics

- Data Visualization
- Data Wrangling
- Data Collection
- Web Scraping and Importing Data
- Data Management
- Data Ethics
- Iteration
- Introduction to Database Querying with SQL
- Further Topics based on time and student interest may include:
    - Dynamic and Customized Data Graphics
    - Geospatial Data and Maps
    - Learning with AI

Student Learning outcomes

Students who complete this course will be able to:

- Develop skills in using modern technology to conduct data analysis.
- Draw meaningful conclusions and insights from real world data.
- Understand data collection from various sources and appropriate data management.
- Effectively organize, clean, prepare data for analysis
- Communicate clear, meaningful data visualizations.
- Develop understanding and appreciation of ethics in Data Science.

Lecture schedule

The lecture schedule and associated readings can be found on the main page of this course webpage.

Book and software

The Modern Data Science with R **textbook** is accessible on the navigational bar of this webpage as

well as via this link. Students will use a local install or RStudio workbench to do computations in R find it here.

Class norms

- You are expected to stay until the end of lecture. If you need to leave early, please confirm with me at the beginning of lecture and sit somewhere where your departure will be minimally disruptive.
- *Attendance policy.* You may be dropped from the course for missing 3 or more days of class.
- Attendance will not be explicitly taken and occasional absences are excused. However, extended absences should be mentioned to me.

Readings

We'll be reading chapters 1-6,8,15,and 17 of Modern Data Science with R. Probably other texts.

Evaluation

All due dates can be found on the main page of this course webpage.

This class is project based. As such your evaluations will be based on projects you create.

Projects 40%

There are four projects to complete during the semester. They will be completed collaboratively with a team of students using SCRUM.

Final project 20%

There will be a final project due the last day of classes. You'll complete this final project over the course of the semester using a process called SCRUM. The Final Project will also have in groups of 2-3 where you can choose your groupmates.

Problem sets 25%

Homework will be assigned regularly, its important to complete and stay on top of the problem sets. They will be graded on the following scale:

- 5 points: Assignment is complete, mostly correct, rendered and submitted
- 4 points: Assignment is nearly complete, mostly correct, rendered and submitted

- 3 points: Assignment is around 50% of the problems have been completed, are correct, the assignment is rendered
- 2 points: Assignment is at most 50% complete, the problems are incorrect and is not correctly rendered
- 1 point : Assignment is largely unfinished and is not rendered
- 0 points: No assignment is turned in.

Engagement 15%

It is difficult to explicit codify what constitutes "an engaged student," so instead I present the following rough principle I will follow: *you'll only get out of this class as much as you put in*. That being said, here are multiple pathways for you to stay engaged in this class:

- In particular: Participation in all SCRUM Sprints
- Asking and answering questions in class.
- Working and staying focused in class.
- Coming to office hours.
- Posting questions on Slack.
- Even better: Responding to questions on Slack.

Tentative Assignment schedule

Homework will be assigned regularly and due by midnight the following Monday.

Project 1- due after reading chapter 4 of MDSR roughly the 10th day of class

Project 2- due after reading chapter 5 of MDSR roughly the 18th day of class

Project 3- due after reading chapter 15 of MDSR roughly the 22nd day of class

Project 4- due after reading chapter 17 of MDSR roughly the 26th day of class

Final Project/ Presentation during our final time.

For a more up to date schedule see the course's website.

Policies

Accommodation

Smith is committed to providing support services and reasonable accommodations to all students with disabilities. To request an accommodation, please register with the Disability Services Office at

with disabilities. To request an accommodation, please register with the Disability Services Office at the beginning of the semester. To do so, call (413) 585-2071 to arrange an appointment with Laura Rauscher, Director of Disability Services.

Policies

Inclusion

Inclusion policy

We are committed to fostering a classroom environment where all students thrive. We are committed to affirming the identities, realities and voices of all students, especially those from historically marginalized or underrepresented backgrounds. We are dedicated to creating a space where everyone in the class is respected, is free from discrimination based on race, ethnicity, sexual orientation, religion, gender identity, disability status, and other identities, and feel welcome and ready to learn at your highest potential.

If you have any concerns or suggestions for how to make this class more inclusive, please reach out to your instructor.

We are here to support your learning and growth as data scientists and people!

Attendance

We expect you attend class in person. If you choose to come to class, we expect your full attention. Please put your phone away during class unless otherwise directed.

In keeping with Smith's core identity and mission as an in-person, residential college, SDS affirms College policy (as per the Provost and Dean of the College) that students will attend class in person. SDS courses will not provide options for remote attendance. Students who have been determined to require a remote attendance accommodation by the Office of Disability Services will be the only exceptions to this policy. As with any other kind of ADA accommodations, please notify your instructor during the first week of classes to discuss how we can meet your accommodations.

If you are unable to attend class for any reason, please follow the materials on the course website and check with another student about what happened in class.

Collaboration

Warning

Much of this course will operate on a collaborative basis, and you are expected and encouraged to work together with a partner or in small groups to study, complete labs, and prepare for exams. However, **all work that you submit for credit must be your own**. Copying and pasting sentences, paragraphs, or *blocks of code* from another student or from online sources is not acceptable and will receive no credit. No interaction with anyone but the instructors is allowed on any exams or quizzes.

Academic Honor Code Statement

All students, staff and faculty are bound by the Smith College Honor Code, which Smith has had since 1944.

> Smith College expects all students to be honest and committed to the principles of academic and intellectual integrity in their preparation and submission of course work and examinations. Students and faculty at Smith are part of an academic community defined by its commitment to scholarship, which depends on scrupulous and attentive acknowledgement of all sources of information, and honest and respectful use of college resources.

Cases of dishonesty, plagiarism, etc., will be reported to the Academic Honor Board.

Code of Conduct

As the instructor and assistants for this course, we are 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, deliberate misgendering or use of "dead" names, trolling, public or private harassment, insults, or other unprofessional conduct.

As the instructor and assistants we 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.

> **Important**
>
> All students, the instructor, the lab instructor, and all assistants are expected to adhere to this Code of Conduct in all settings for this course: lectures, labs, office hours, tutoring hours, and over Slack.

This Code of Conduct is adapted from the Contributor Covenant, version 1.0.0, available here.

Resources

Moodle and course website

The course website and Moodle will be updated regularly with lecture handouts, project information, assignments, and other course resources. Grades will be recorded on Moodle. Please check both regularly.

Communication

- Slack is the primary mechanism for course-related discussions of all kinds. *Please do not email the course instructor with course-related questions!* Instead, post these on `#questions`

on Slack. If discretion is absolutely necessary, private message the instructor on Slack.

Writing

Your ability to communicate results (which may be technical in nature) to your audience (which is likely to be non-technical) is critical to your success as a data analyst. The assignments in this class will place an emphasis on the clarity of your writing.

Writing Enriched Curriculum

This course is part of Smith College's Writing Enriched Curriculum. As such, the course supports the Writing Plan of the Program in Statistical & Data Sciences.

Please read the SDS Writing Plan for more information.

The Spinelli Center

The Spinelli Center (now in Seelye 207) supports students doing quantitative work across the curriculum. In particular, they employ:

- SDS Tutors available from 7:00–9:00pm on Sunday–Thursday evenings in Burton 301. These students are trained to help you with your statistics and R questions.
- A Data Research and Statistics Counselor (Osman Keshawarz) who keeps both drop-in hours and appointments. Students are welcome to email qlctutor@smith.edu to make an appointment.

Your fellow students are also an excellent source for explanations, tips, etc.

---

credits