SDS 192: Introduction to Data Science

Lindsay Poirier

Last updated on 2024-01-23

Table of contents

1 Syllabus

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.

Classes will be held on Mondays, Wednesdays, and Fridays from 9:25 AM to 10:40 AM in McConnell 404.

SDS 100: Reproducible Scientific Computing with Data is a co-requisite for this course and designed to help support you in coding in R. Please note that I walk into this course with the assumption that most students have never coded before. Coding for the first time can be intimidating, but I intend do everything in my power to support you through the learning curve and to make things both fun and relevant in the process. I personally picked up most of my data science skills through a lot of trial-and-error, practice, and curiosity. My hope is that, in this course, you will learn through experimentation, along with independent and collaborative problem-solving. Honing these competencies will serve you as you move on to other courses in the SDS program and/or at Smith.

1.1 Course Instructor

While you're welcome to refer to me as Professor Poirier, I would prefer it if you called me Lindsay. I am a cultural anthropologist that studies how public interest datasets get produced, how communities think about and interface with data, and how data infrastructure can be designed more equitably. My Ph.D. is in an interdisciplinary discipline called Science and Technology Studies - a field that studies the intricate ways science, technology, culture, and politics all co-constitute each other. I work on a number of collaborative research projects that leverage public data to deepen understanding of social and environmental inequities in the US, while also qualitatively studying the politics behind data gaps and inconsistencies. As an instructor, I prioritize active learning and often structure courses as flipped classrooms. You can expect in-class time to predominantly involve a mix of lectures and live problem-solving exercises.

1.1.1 Getting in Touch

1.1.1.1 Slack

I can best support students in this course when I can readily keep tabs on our course-related communication. Because of this, I ask that you please don't email me regarding course-related questions or issues. The best way to get in touch with me is via our course Slack. If you have course-related questions, I encourage you to ask them in the `#sds-192-questions` channel. When discretion is needed, feel free to DM. Please reserve more formal concerns like grades or accommodation requests for an in-person (or in-person virtual) conversation.

During the week, I will try my best to answer all Slack messages within 24 hours of receiving them. Please note that to maintain my own work-life balance, I often don't answer Slack messages late in the evenings or on the weekends. It's important that you plan when you start your assignments accordingly.

1.1.1.2 Office Hours

Office hours are a great opportunity for us to chat about what you're learning in the course, clarify expectations on assignments, and review work in progress. I also love when students drop in to office hours to request book recommendations, discuss career or research paths, or just to say hi! I encourage each student in the course to join office hours at least once this semester. If you're unable to attend my office hours at the regularly scheduled time, there is link on Moodle to book a meeting with me.

- Monday, 3-4, McConnell 212-213
- Wednesday, 3-5, McConnell 212-213

1.2 Course Texts

A number of excellent textbooks introducing data science concepts and methods have been written in the past few years, including a few from faculty in the Smith SDS department. To accompany the topics we will cover each week, I will be selecting my favorite chapters from these books and posting them to Perusall. However, all three books we will engage in this course cover almost every topic we will address, so feel free to supplement your reading with corresponding chapters in the other books: especially if you find yourself drawn to the teaching and writing style in a certain book. All books are available for free online.

- Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton. 2021. *Modern Data Science with R.* 2nd ed. CRC Press. https://mdsr-book.github.io/mdsr2e/.

- Irizarry, Rafael A. 2022. *Introduction to Data Science. Data Analysis and Prediction Algorithms with R.* https://rafalab.github.io/dsbook/.

- Ismay, Chester, and Albert Y. Kim. 2021. *Modern Dive: Statistical Inference via Data Science.* CRC Press. https://moderndive.com/.

Each week I will also list optional reading and resources in our course schedule that you may reference if you are struggling with a topic or if you wish to explore that topic further. I will update this list often throughout the semester.

1.3 Assessment

This course will be graded via a standards-based assessment system.

> 💡 Spinelli Center
>
> Smith's Spinelli Center offers a number of resources to support SDS students. Spinelli Center Data Assistants will visit our classroom regularly to support you through lab work.

The Center also offers drop-in tutoring hours Sunday through Thursday 7-9 PM. Finally, you can drop-in to Seelye 207D or schedule an appointment with the Data Research and Statistics Counselor (Kenneth Jeong). To schedule an appointment, email qlctutor@smith.edu.

1.4 Policies

1.4.1 Preparation

This is a 4-credit course with 4.5 hours per week of in-classroom instructions. Smith expects students to devote 7.5 out-of-class hours per week to 4-credit classes. I have designed the course assignments and selected the course readings with this target in mind.

1.4.2 Attendance

Attending class is not only important for your learning but also an act of community. Attendance is expected in this course. Many course assignments will be completed in-class. That said, you do not need to inform me when you will be absent. If you are sick, please stay home. If you must miss a class entirely, you should contact a peer to discuss what was missed. Please note that the SDS Program has adopted a shared policy regarding in-person attendance this semester:

> 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.

1.4.3 Extensions

There is an automatic 24-hour **grace period** on all *lab and project assignments*. There will be no penalties for submitting the project within this 24-hour period, and you do not need to inform me that you intend to take the extra time. You can also request up to a 72-hour **extension** on any project or lab assignment, as long *as you make that request at least 48 hours before the original assignment due date*. You can request an extension by filling out the Extension Request form on Moodle, and I will confirm your extension on Slack. Beyond this, late assignments will not be accepted.

1.4.4 Academic Honesty

> 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.

Any cases of dishonesty or plagiarism will be reported to the Academic Honor Board. Examples of dishonesty or plagiarism 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.

> **!** Deadlines for Quizzess and Course Advancement Assignments
>
> The standards you will be practicing in this course all build off of each other, and it's important that I know how students are doing on each standard in order to direct my teaching going forward. Because of this, the deadlines for quizzes are a bit more firm than for other assignments. On the course schedule and in Moodle you will see a suggested deadline for quizzes and a final deadline. I won't be able to accept quiz submissions after the final deadline, so you'll want to be sure to stay on top of these dates. Similarly, course advancements assignments are assignments that need to be completed to keep our course moving. Because of this, they need to be completed by the due date. I will give you an opportunity to work on many of these assignments in class.

1.5 Community

1.5.1 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.

1.5.2 Principles of Community

I hope that we can foster a collaborative and caring environment in this classroom: one that celebrates successes, respects individual strengths and weaknesses, demonstrates compassion for each other's struggles, and affirms diverse identities. Here are some ideas that I have for creating this environment in our course:

- Check-in with colleagues before starting collaborative work. "What three words describe how you're feeling?" "Name one challenge and one success from this week." "What are you doing for self-care right now?" Thank each other for sharing where they're at.
- 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 that there is much we don't know about how our colleagues experience the world. …but don't ask colleagues to speak on behalf of a social group you perceive them to be a part of.
- Cheer on colleagues as they give presentations or try something out for the first time.
- Ask questions often in our `#sds-192-questions` channel. Help each other out by answering questions when you can.
- Mistakes happen. I will certainly make mistakes in class. Admit mistakes, and then move on.

1.5.3 Pronouns

Using the proper pronouns for our students is foundational to a safe, respectful classroom environment that creates a culture of trust. For information on pronouns and usage, please see the Office of Equity and Inclusion link here: Pronouns

1.6 Support

1.6.1 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 Office of Disability Services in College Hall 104 or at ods@smith.edu. The Office 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 ODS, please contact me. At no point will I ask you to divulge details about your personal circumstances to me.

1.6.2 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. I can point you to these resources at any time throughout the semester.

1.6.3 Trigger Warnings

A trigger is a topic or image that can precipitate an intense emotional response. When common triggering topics are to be covered in this course, I will do my best to provide a trigger warning in advance of the discussion. However, I can't always anticipate triggers. With this in mind I've set up an anonymous form, available on Moodle, where you can indicate topics for which you would like me to provide a warning.

1.7 Infrastructure

1.7.1 Moodle

Grades, forms, handouts, and quizzes will be available on the course Moodle.

1.7.2 Perusall

All course readings and recorded lectures will be available on Perusall. You can access Perusall via our course Moodle page.

1.7.3 Slack

- `#general`: Course announcements (only I can post)
- `#sds-192-discussions`: Share news articles and relevant opportunities
- `#sds-192-questions`: Ask and answer questions about our course
- You can also create private Slack channels with your project group members.

1.7.4 GitHub

I will be using GitHub Classroom to distribute several course assignments, and 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.

> **ℹ RStudio/RStudio Server**
>
> This class will use the R statistical software package. In the first week of the course, I will help you install and configure R and RStudio. If you are using a laptop, you will install both on that computer. If you are using a Chromebook or Tablet, an account will be created for you on the Smith College RStudio Server so that you can access a cloud-based version of RStudio. You should let me know in the first week of the course if you are using a Chromebook or tablet.

2 Assessment

This course will be graded using a standards-based assessment system. In a more traditional grading system, your scores on a series of assignments are averaged over the course of the semester. In this course, rather than assessing and averaging your achievement on particular assignments, I will instead be assessing the development of your fluency in a set of pre-defined standards. You will have multiple opportunities over the course of the semester to showcase the depth of your understanding regarding these standards. A standards-based grading system carries the following benefits:

- Learning targets for the course are clearly defined from the outset, and almost every graded assignment that you receive will be directly tied to at least one standard. This

should make it abundantly clear what skills and competencies I'm assessing on each assignment. There is no "busy work" with a standards-based system.

- No one assignment will make-or-break your grade. You have multiple opportunities to demonstrate fluency in a standard. This rewards students that take the time to practice and learn from their mistakes. It prioritizes student growth throughout the course of the semester and allows for us all to have off-days.
- Assessments in a standards-based system are much clearer than in a point-based grading system. Saying that I've become proficient in data wrangling, joining, and visualizing means more than saying that I earned a 92.5 in my Introduction to Data Science course. Further, when approaching me about how to improve your grade in the course, we can focus our conversation more on how to deepen certain skills and competencies rather than how to hit certain numeric benchmarks.
- A standards-based grading system makes it easier to monitor your progress towards a certain grade.

This is my second iteration of teaching a standards-based course, and I've made several revisions from the first iteration based on what I've learned from experience and student feedback. I'm excited to refine this system this semester as I believe it aligns with my overarching goals for the course. In this course, I find it far more important that you come away with an understanding of the concepts behind core data science strategies (along with an ability to find and interpret reference materials) than it is to demonstrate memorization of `R` syntax. Developing this understanding will empower you to learn and apply new data science languages on your own.

2.1 What are the standards I will be assessed on in this course?

> **Data Visualization**
>
> This dimension refers to the development of your ability to produce multiple types of compelling and well-designed visualizations from data.
> VISUALIZATION AESTHETICS PLOTTING FREQUENCIES AND DISTRIBUTIONS MAPPING

> **Data Wrangling**
>
> This dimension refers to the development of your ability to transform datasets into new formats in order to prepare them for further analysis or visualization.
> TRANSFORMING DATA JOINING DATASETS TIDYING DATA

> **Data Science Workflow**
>
> This dimension refers to the development of your ability to apply data science best practices in your work.
> UNDERSTANDING DATASETS DATA RETRIEVAL GITHUB PROGRAMMING IN R

> **Data Ethics**
>
> This dimension refers to your ability to recognize and navigate ethical dilemmas that emerge in data science work. There is not a formal standard for data ethics because issues related to data ethics will be discussed in every unit of the course, and you will be assessed on data ethics issues related to other standards in quizzes and projects.

2.2 How will I be assessed on the course standards?

2.2.1 Informal Assessments

2.2.1.1 Readings

Each week, you will be assigned a section of the course texts to read prior to class. I expect that you will come having read this section in order to prepare for in-class exercise and labs. You do not need to complete the exercises in the course texts but may choose to do so if you wish. Please note however that we won't have time to go over the solutions in class, and I don't have a solutions manual for these texts (though I'm happy to go over them in office hours). All course readings will be available in Perusall, and you can post questions and comments in the reading for myself or your classmates to answer.

2.2.1.2 Lab Recaps

Lab solutions will be posted as video lectures in Perusall. I expect that you will check your lab answers by reviewing these videos. You may leave comments in Perusall at certain timestamps as questions come up.

2.2.2 Formal Assessments

> 💡 Tip
>
> 5PM (close of business day) will be the cut-off time for all assignments in this course. The reason I've set the deadlines to 5PM is that I'd like to discourage students from staying up late into the night to complete assignments. Note that there is a 24-hour grace period for submitting labs and project assignments. This means that you will still get full credit for these assignments as long as they are submitted by 5PM the day following the assignment due date. However, quizzes and course advancement assignments must be submitted by 5PM *on the due date* for credit.

2.2.2.1 Labs

In most weeks, you will be assigned a lab, which you will start in class and complete at home. There will be one lab per standard. Labs will be designed to help you practice applying the course standards towards the analysis of a dataset. You may work on labs in groups, but all group members should submit their own lab. Labs will be **graded for completion**, and you can earn 3 points per standard by completing the lab associated with that standard. All sections of the lab must be completed in good faith to earn these points.

2.2.2.2 Projects

There will be 3 projects, to be completed in groups of 3-4, assigned over the course of the semester. In each you will have an opportunity to demonstrate fluency in standards we have covered up to that point in the semester. I will provide prompts for each project, but you will have a lot of flexibility to demonstrate your own creativity and explore your own interests in designing a project around the prompt. You can earn up to 3 points towards 9 of the ten standards based on your project submission. If you don't earn full credit on a standard for a project submission, you may improve your score on that standard in the next project. The only standard that won't be covered in projects is **Data Retrieval**, which we will cover too late in the semester to work into projects. Projects will be **graded for fluency**.

2.2.2.3 Quizzes

There will be 3 quizzes administered throughout the semester - each assessing 3-4 course standards. There will be 3 questions per standard on each quiz, and you can earn up to 3 points towards each standard based on your quiz attempt: 1 point per question. In this sense, quizzes will be **graded for fluency.**

Quizzes will be taken at home, administered in Moodle, and are open book/open Internet. You may start a quiz at any time before its due date, but it must be completed by its due date in order to earn credit. Please note that extensions will not be granted for quizzes.

2.2.2.4 Course Advancement

There are a series of very short assignments on the syllabus that are designed to ensure that you are prepared for individual and collaborative work. This includes things like reviewing the syllabus, developing a course study plan, developing group collaboration plans, and completing peer evaluations. In total, there are 13 course advancement assignments, and you can earn 1 point towards your final grade per assignment.

Assignment	Points
Syllabus Quiz	1
CATME Survey	1
Problem Solving Lab	2
Study Plan	2
Study Plan Evaluation	2
Group Contract (x3)	3

These assignments will be **graded for completion**. For the most part, you will get out of them, what you put into them. Because these assignments are designed to keep our course running smoothly, please note that extensions will not be granted for course advancement assignments.

> **!** Reassessment
>
> This course will have an *optional* final exam. The final exam is an opportunity to reassess standards that you have not received full **quiz** credit on. If you choose to take the final exam, you only need to complete the sections of the exam associated with standards you wish to reassess, and the score you ultimately receive for a standard will be based on whichever is higher of your quiz score or your final exam score for that standard.

2.3 How will this system work?

If you've been adding, you may have figured out by this point that we have:

$(3 * 10) + (3 * 10) + (3 * 9) = 87$

You can earn the remaining 13 points by completing the course advancement assignments, for a grand total of 100 points. At the end of the semester, I will sum your scores on all standards and other assignments and assign final grades accordingly:

Letter Grade	Numeric Grade
A	92.5
A-	90.0
B+	87.5
B	82.5
B-	80.0
C+	77.5
C	72.5
C-	70.0
D+	67.5
D	62.5
D-	60.0
E	< 60.0

3 Schedule

> **ℹ Acknowledgements**
>
> Styling and infrastructure for this page inspired by related syllabi produced by Ben Baumer and R. Jordan Crouser.

> **💡 Tip**
>
> All readings for this course will be available in our course Perusall, which is linked in Moodle. I encourage you to complete the readings there so that you can leave comments and questions as they come up.

3.1 September 07, 2022

3.1.1 What is Data Science?

3.1.1.1 Due Today

3.2 September 09, 2022

3.2.1 Infrastructure Set-up

3.2.1.1 Due Today

Fill out First Day of Class Questionnaire in Moodle
Complete Syllabus Quiz in Moodle
Contact me if you will be using a Chromebook

3.3 September 12, 2022

3.3.1 Data Fundamentals

UNDERSTANDING DATASETS

3.3.1.1 Due Today

3.4 September 14, 2022

3.4.1 Introduction to R

UNDERSTANDING DATASETS

3.4.1.1 Due Today

(Read in Perusall; linked in Moodle) 2. R Basics , Irizarry, Rafael A. (2022). *Introduction to Data Science.* Data Analysis and Prediction Algorithms with R. URL: https://rafalab.github.io/dsbook/ (visited on Jan. 14, 2022).

3.5 September 16, 2022

3.5.1 Lab: Understanding Datasets

UNDERSTANDING DATASETS

3.5.1.1 Due Today

3.6 September 19, 2022

3.6.1 Grammar of Graphics

VISUALIZATION AESTHETICS

3.6.1.1 Due Today

(Read in Perusall; linked in Moodle) 2. Data Visualization , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R.* 2nd. CRC Press. URL: https://mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

3.7 September 21, 2022

3.7.1 Visualization Conventions

3.7.1.1 Due Today

Problem-Solving Lab Due
Lab 1 Due

3.8 September 23, 2022

3.8.1 Lab: Designing Effective Data Visualizations

VISUALIZATION AESTHETICS

3.8.1.1 Due Today

Study Plan Due

3.9 September 26, 2022

3.9.1 Frequency Plots

PLOTTING FREQUENCIES AND DISTRIBUTIONS

3.9.1.1 Due Today

(Read in Perusall; linked in Moodle) 2. Data Visualization , Ismay, Chester and Albert Y. Kim (2021). *Modern Dive: Statistical Inference via Data Science.* CRC Press. URL: https://moderndive.com/ (visited on Jan. 14, 2022).

Complete CATME Survey

3.10 September 28, 2022

3.10.1 Mountain Day!

PLOTTING FREQUENCIES AND DISTRIBUTIONS

3.10.1.1 Due Today

Lab 2 Due

3.11 September 30, 2022

3.11.1 Work on Group Projects in Class

PLOTTING FREQUENCIES AND DISTRIBUTIONS

3.11.1.1 Due Today

Project 1 Assigned

3.12 October 03, 2022

3.12.1 Lab: Collaborating via GitHub

GITHUB

3.12.1.1 Due Today

Read Lab 4 in Perusall
Watch Lab 2 Recap in Perusall

3.13 October 05, 2022

3.13.1 Boxplots

GITHUB

3.13.1.1 Due Today

Group Contract Due

3.14 October 07, 2022

3.14.1 Lab: Visualizing Data

3.14.1.1 Due Today

Lab 4 Due

3.15 October 10, 2022

3.15.1 No Class

3.15.1.1 Due Today

3.16 October 12, 2022

3.16.1 Subsetting, Aggregating, and Summarizing Data

TRANSFORMING DATA

3.16.1.1 Due Today

(Read in Perusall; linked in Moodle) 3. Data Wrangling , Ismay, Chester and Albert Y. Kim (2021). *Modern Dive: Statistical Inference via Data Science.* CRC Press. URL: https://moderndive.com/ (visited on Jan. 14, 2022).

Lab 3 Due

3.17 October 14, 2022

3.17.1 Lab: Aggregating and Summarizing Data

TRANSFORMING DATA

3.17.1.1 Due Today

3.18 October 17, 2022

3.18.1 Joining Datasets

TRANSFORMING DATA

3.18.1.1 Due Today

(Read in Perusall; linked in Moodle) 5. Data wrangling on multiple tables , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R.* 2nd. CRC Press. URL: https://mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

Watch Lab 3 Recap in Perusall

3.19 October 19, 2022

3.19.1 Ethics of Data Joining

TRANSFORMING DATA

3.19.1.1 Due Today

Project 1 Due

Quiz 1 Recommended Deadline
Lab 5 Due

3.20 October 21, 2022

3.20.1 Class will not meet in-person: Meet with project groups on Zoom

3.20.1.1 Due Today

Project 2 Assigned
Watch Project Intro Video in Perusall

3.21 October 24, 2022

3.21.1 Project 1 Shares

3.21.1.1 Due Today

Zoom link for today's class has been sent via email and is posted on Moodle.

Watch Lab 5 Recap in Perusall

3.22 October 26, 2022

3.22.1 Data Wrangling Problem-Solving

3.22.1.1 Due Today

Group Contract Due

3.23 October 28, 2022

3.23.1 Lab: Joining Datasets

JOINING DATASETS

3.23.1.1 Due Today

Quiz 1 Due

3.24 October 31, 2022

3.24.1 Tidying Datasets

TIDYING DATA

3.24.1.1 Due Today

(Read in Perusall; linked in Moodle) 6. Tidy Data , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R*. 2nd. CRC Press. URL: https: //mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

3.25 November 02, 2022

3.25.1 Pivoting Datasets

TIDYING DATA

3.25.1.1 Due Today

(Read in Perusall; linked in Moodle) 26. Parsing dates and times , Irizarry, Rafael A. (2022). *Introduction to Data Science.* Data Analysis and Prediction Algorithms with R. URL: https://rafalab.github.io/dsbook/ (visited on Jan. 14, 2022).

Lab 6 Due

3.26 November 04, 2022

3.26.1 Lab: Pivoting Datasets

TIDYING DATA

3.26.1.1 Due Today

3.27 November 07, 2022

3.27.1 Writing Functions

PROGRAMMING IN R

3.27.1.1 Due Today

(Read in Perusall; linked in Moodle) 7. Iteration , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R.* 2nd. CRC Press. URL: https://mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

Study Plan Assessment Due
Watch Lab 6 Recap in Perusall

3.28 November 09, 2022

3.28.1 Iteration

PROGRAMMING IN R

3.28.1.1 Due Today

Project 2 Due

Lab 7 Due

3.29 November 11, 2022

3.29.1 Lab: Programming in R

PROGRAMMING IN R

3.29.1.1 Due Today

Project 3 Assigned

3.30 November 14, 2022

3.30.1 Map Projections and Spatial Thinking

MAPPING

3.30.1.1 Due Today

(Read in Perusall; linked in Moodle) 17. Working with geospatial data (17.1-17.3) , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R*. 2nd. CRC Press. URL: https://mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

Watch Lab 7 Recap in Perusall

3.31 November 16, 2022

3.31.1 Lab: Mapping Point Data in Leaflet

MAPPING

3.31.1.1 Due Today

[Lab 8](#) Due

3.32 November 18, 2022

3.32.1 Polygon Mapping in Leaflet

MAPPING

3.32.1.1 Due Today

Quiz 2 Recommended Deadline

3.33 November 21, 2022

3.33.1 How to Lie with Maps

MAPPING

3.33.1.1 Due Today

(Read in Perusall; linked in Moodle) 17. Working with geospatial data (17.4-17.8) , Baumer, Benjamin S., Daniel T. Kaplan, and Nicholas J. Horton (2021). *Modern Data Science with R*. 2nd. CRC Press. URL: https://mdsr-book.github.io/mdsr2e/ (visited on Jan. 14, 2022).

Group Contract Due

Watch Lab 8 Recap in Perusall

3.34 November 23, 2022

3.34.1 No Class

3.34.1.1 Due Today

3.35 November 25, 2022

3.35.1 No Class

3.35.1.1 Due Today

3.36 November 28, 2022

3.36.1 Working with APIs

DATA RETRIEVAL

3.36.1.1 Due Today

Quiz 2 Due
Lab 9 Due

3.37 November 30, 2022

3.37.1 Advanced APIs

DATA RETRIEVAL

3.37.1.1 Due Today

3.38 December 02, 2022

3.38.1 Lab: Working with APIs

DATA RETRIEVAL

3.38.1.1 Due Today

Watch Lab 9 Recap in Perusall

3.39 December 05, 2022

3.39.1 Work on Group Projects in Class

3.39.1.1 Due Today

3.40 December 07, 2022

3.40.1 TBD

3.40.1.1 Due Today

3.41 December 09, 2022

3.41.1 Project 3 Shares

3.41.1.1 Due Today

Quiz 3 Recommended Deadline

3.42 December 12, 2022

3.42.1 Wrap-up

3.42.1.1 Due Today

Project 3 Due

Quiz 3 Due
Lab 10 Due