jbisbee1 /
DS1000_F2023

<> Code   ⊙ Issues   ⑂ Pull requests   ▶ Actions   ⊞ Projects   ⊘ Security   ⬚ Insights   ⚙ Settings

Intro to Data Science: Fall 2023

☆ 9 stars   ⑂ 2 forks   ⊙ 3 watching   ⟋ Activity

🌐 Public repository

⑂ main ▾

⑂ Branches   ⬚ Tags

jbisbee1 pset 6 ak   …   ✓ yesterday   🕑 126

View code

DS 1000-01: How Data Shape Our World (Fall 2023) 🔗

Course Material for DS1000: How Data Shape our World

Table of Contents 🔗

☰ README.md   ✎

Overview 🔗

The target audience is someone who is interested in Data Science, but who has no prior experience. The class is designed to be application-forward -- demonstrating what you can do with the tools of data science in the hopes of motivating and encouraging students to go deeper and further. As an introductory class with no prerequisites, the statistical and programming fundamentals behind what we do is only briefly mentioned; the goal is to provide a sense of what can be done with data science rather than to provide a comprehensive foundation on a smaller set of topics. The contents of this repository represent a work-in-progress and revisions and edits are likely frequent.

The main text for the course is "R For Data Science" which can accessed free online here.

Students are expected to have a free ChatGPT account which can be accessed [here](#)

Back to ToC

Required Applications 🔗

Brightspace 🔗

This is the course management software used at Vanderbilt to support course learning. I will use this to post readings, lectures, assignments, and news for the course. I will post announcements and changes to the home page of the site; though I will always announce changes in class, please keep an eye out. *If a change to the syllabus or requirements is posted in the announcements on this site, you are responsible for those changes.*

Don't forget to download the related app, which is called Pulse, to your phone and set it to alert you if there are new content or announcements for the course.

Campuswire 🔗

I have set up a Campuswire workspace for our use this semester to help us better communicate with each other and the TAs. You will need to create an account and join our workspace by following [this link](#). You are encouraged to adopt these [Slack etiquette tips](#).

Here is the list of channels you should see upon joining the Campuswire workspace:

- **Class feed**: A space to post questions and respond to other posts.

- **#announcements**: A space for all course announcements.

- **#general**: A space for you to share and discuss stories you've seen in the news or on social media that are relevant to our class.

- **Calendar**: A calendar containing all lectures, due dates, office hours, and labs.

- **Files**: A space for course materials (**NOT USED. VISIT BRIGHTSPACE INSTEAD.**)

- **Grades**: A space for grades (**NOT USED. VISIT BRIGHTSPACE INSTEAD.**)

GitHub 🔗

I have created a `GitHub` repository to prepare and share all course-related content. This very syllabus is available as the repository's README and all links below are connected to the appropriate folders, sub-folders, and files in this repository.

You are encouraged to adopt the following workflow for this class:

1. Prior to each lecture, download the appropriate `.Rmd` file, open it in `RStudio`, and read through it. **This is your primary homework assignment!** As you work through it, try to tweak some of the code and answer the toy examples where provided. Each time you make a change, click the `knit` button in `RStudio` to see if everything still loads.

2. During each lecture, extend and modify the `.Rmd` file to follow along with me as I present. You should think of each lecture's `.Rmd` file as your homework, your in-class notes, and your study guide for problem sets + exams.

3. After each lecture, tweak the `.Rmd` file further to test out new ideas that you come up with which were not covered in class. Each lecture's slides will be made available as `PDF` for you use to help you review. Thinking creatively about how to modify and extend what we do together might mean you already guess some of the exam questions!

Back to ToC

Evaluation & Responsibilities 🔗

Given the nature of the material, as well as the size of the class, participation is not part of your grade. Instead, you will be assigned weekly problem sets beginning with Topic 3 which will test your ability to apply what you've learned in class, and to think creatively about your own data science ideas. These problem sets are assigned on the Monday of each week and are due by **11:59PM Nashville time the following Friday**. You are welcome to collaborate on these problem sets, and are encouraged to ask questions on the **Class feed** on `Campuswire`.

There are also two exams. The first is a mid-term exam that is scheduled for October 25th (Wednesday) and will take place in class. The preceding Monday (October 23rd) will be an optional review session, where you should come prepared with questions. The second exam is a cumulative final that is scheduled for December 11th.

The final grade is calculated as a weighted average of these components with the following weights:

- **Problem sets:** 9 in total, each worth 5% of your final grade. You can drop the lowest-scoring problem set, resulting in a total value of 40% of your final grade.
- **Exams:** Midterm and final, both of which are in-class. These are each worth 20% of your final grade, comprising 40% of your final grade.
- **Quizzes:** There are 20 quizzes over the course of the semester which can only be taken if you attend class. Each consist of five questions plus an honor statement. 50% of the quiz grade comes from accurately answering all five questions, and 50% comes from signing the honor statement. Each quiz is worth 1% of your final grade, and combined comprise the final 20%.

Both problem sets and exams will have opportunities to earn extra credit, theoretically meaning you could score above 100%. All grades are curved without taking the extra credit results into consideration. We follow the standard Vanderbilt grading system, reproduced below:

- A: 94+
- A-: 90-93
- B+: 87-89
- B: 84-86
- B-: 80-83
- C+: 77-79
- C: 74-76
- C-: 70-73
- D+: 67-69
- D: 64-66
- D-: 60-63
- F: <60

Back to ToC

Course Policies 🔗

Late Assignments 🔗

Every problem set is assigned on a Monday, and due on Brightspace by **11:59PM Nashville time on the following Friday**. Problem sets should be submitted via Brightspace. The problem sets are designed to require no more than two hours in total to complete. Late submissions will be **penalized 1 point off for each day late**. After three days, problem sets will no longer be accepted and will be scored 0. (But also remember that your lowest scoring problem set will not be counted toward your final grade, effectively giving you one "freebie".) Answer keys for the preceding week's problem set are posted on Wednesdays after lecture.

Cell Phones, Laptops, Tablets, etc. 🔗

You are expected to bring your laptop to class in order to work through the `.Rmd` file during the lecture. These `.Rmd` files will be posted to the GitHub repository at least 24 hours prior to the lecture. Students are encouraged to download these files and work through them prior to class.

You are asked to silence your cell phone / tablet / smart watch before class begins.

Academic Honor Code 🔗

Students are assumed to have read and agreed with the Vanderbilt University Academic Honesty policy, found at URL: https://www.vanderbilt.edu/student_handbook/the-honor-system/

Collaboration is the heart of data science, but your work on your assignments should be your own. Please be careful not to plagiarize. The Undergraduate Honor Council has a very helpful guide to understanding plagiarism. **In particular, while students are invited to work on problem sets together, collaboration is prohibited on the midterm and final exams.**

Academic misconduct includes, but is not limited to, cheating, fabrication, plagiarism, altering graded examinations for additional credit, having another person take an examination for you, falsification of results, or facilitating academic dishonesty or as further specified in the university policy found at the website above. These and other forms of cheating are all potentially grounds for penalties including failure of the assignment or the course, as well as program- or university-level disciplinary action.

ChatGPT and related Large Language Models (LLMs) are essential tools in the data scientist's toolkit, and acceptable resources for completing the assignments and learning concepts at a deeper level. However, graded assignments cannot be generated purely by these tools. All assignments must include a log of the ChatGPT prompts and resulting output used to complete the assignment.

Accommodations for Learning or Access Disabilities 🔗

This class respects and welcomes students of all backgrounds, identities, and abilities. If there are circumstances that make our learning environment and activities difficult, if you have medical information that you need to share with me, or if you need specific arrangements in case the building needs to be evacuated, please let me know. I am committed to creating an effective learning environment for all students, but I can only do so if you discuss your needs with me as early as possible. I promise to maintain the confidentiality of these discussions. If appropriate, also contact the Vanderbilt Student Access office to get more information about specific accommodations; please visit https://www.vanderbilt.edu/student-access/ as soon as possible to become registered and ensure that accommodations are implemented in a timely fashion. Requests for academic accommodations are to be made during the first three weeks of the semester, except for unusual circumstances.

COVID-19 Policies 🔗

As per Vanderbilt's [F22 Covid Plan](#), we are conducting class in as normal an environment as we've had since before the start of the COVID-19 pandemic. In general, this means:

- We will return fully to in-person instruction.
- Restrictions have been lifted on meetings and gathering sizes.
- Mask mandates are no longer in place.
- Asymptomatic testing has been suspended.
- Isolation guidelines only apply to those who have tested positive for COVID-19.

As such, DS1000 will be meeting in-person every Monday and Wednesday from 1:15PM - 2:30PM. Lectures will not be hybrid or remote. For students affected by COVID-19, this means that you are responsible for obtaining any course material you missed. As you'll see below, I make all materials available on the course GitHub repository. In addition, I will be recording all lectures and posting these the evening following the lecture. **NB: this does not mean that students are permitted to skip the lectures.**

I will be holding my office hours in-person in The Commons Center Room #353. The TAs are free to determine how best to hold their office hours, and their choices will be communicated to the students as soon as they are decided.

If You Need Help 🔗

There are many things that you might be dealing with that can hinder your ability to succeed in this course, your college career, and your life. You might be struggling with illness, socioeconomic issues, or personal issues that make it hard to concentrate, to work, or to attend class. If any of these or other things begin to hinder your ability to do your best, you can reach out to the office of Student Care Coordination for programs, training, accommodations, and assistance (find more information or make an appointment here: https://www.vanderbilt.edu/carecoordination/). The Student Care Coordination can help guide you to whatever assistance you might need, whether it be short term or long term. *If you specifically need help or accommodation in this course due to your difficulties, please come meet with me so we can find a solution that allows you to succeed while being fair to others.*

Mandatory Reporting 🔗

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, etc. If you or someone you know has been harassed or assaulted, you can call the Project Safe 24-hour crisis/support hotline at 615-322-7233 and you can find a list of resources at Project Safe. You may also contact the University's Title IX Coordinator (615-322-4705) and you can find the appropriate contacts for resources and confidence here: https://www.vanderbilt.edu/title-ix/

As a faculty member, one of my responsibilities is to help create a safe learning environment on our campus, no matter their identity or circumstances. I also have a mandatory reporting responsibility. It is my goal that you feel able to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep information you share private to the greatest extent possible. However, I must note that I am a representative of an institution that we want to make safer for all people, therefore I am a mandatory reporter. University faculty, many staff members, and some student leaders are required to report incidents of sexual assault, sexual harassment, dating violence, domestic violence, stalking, and child abuse, as well as any suspected discrimination (about age, race, color, creed, religion, ancestry, national or ethnic origin, sex/gender, sexual orientation, disability, genetic information, military status, familial status or other protected categories under local, state or federal law) to the University's Title IX Coordinator (615-322-4705), as required by University policy and state and federal law. If you disclose an experience of interpersonal violence and/or child abuse to me or to classmates with mandatory reporting, whether in class discussion, through a course assignment, or in private communication with me, your disclosure will be kept as private as possible but may not be able to be kept confidential.

Diversity Statement 🔗

Data science is, at its core, about thinking creatively to answer challenging questions. Creative thinking requires exposure to different perspectives, which are themselves borne of diverse experiences. I value diversity in all its forms including age, ability or disability, ethnicity, national origin, race, religion, sex, gender, sexual orientation, and family and marital status. I expect that all students participating in this class will respect differences and strive to understand how other peoples' perspectives, behaviors, and worldviews may be different from their own.

Religious Holidays 🔗

The observance of religious holidays (activities observed by a religious group of which a student is a member) and cultural practices are an important reflection of diversity. As your instructor, I am committed to providing equivalent educational opportunities to students of all belief systems. At the beginning of the semester, you should review the course requirements to identify foreseeable conflicts with assignments, exams, or other required attendance. If at all possible, please contact me within the first two weeks of the first class meeting to allow time for us to discuss and make fair and reasonable adjustments to the schedule and/or tasks.

[Back to ToC](#)

Office Hours (OH) / Labs 🔗

- Prof. Bisbee: OH M & W at 3PM in Commons #353
- TA Brych: OH T from 1PM to 3PM on Zoom (see Campuswire for link)
- TA Fang: OH TH from 3PM to 5PM in FGH 201
- TA Gupta: OH W from 11AM to 1PM on Zoom (see Campuswire for link)
- TA Manda Li: Lab TH from 2PM to 3PM in Stevenson 2212
- TA Qifan Lin: OH W from 5PM to 7PM in DSI Building Suite 208
- TA Negi: OH T from 4PM to 6PM on Zoom (see Campuswire for link)
- TA Quagliata: OH TH from 8:30AM to 9:30AM, and 5:15PM to 6:15PM in Sarratt Student Center Lounge
- TA Zha: OH T & TH from 12PM to 1PM on Zoom (see Campuswire for link)

All these can also be found on the [Campuswire calendar](#), along with the Zoom links for those hosting their office hours remotely.

[Back to ToC](#)

Syllabus 🔗

Date	Lecture		Learn
8/23	[Intro 1](#) [- Recording](#)		Scier meth analy camp
8/28	[Intro 2](#) [- Recording 1](#), [2](#)		Chat Data
8/30	BREAK		
9/04	[Intro to R 1](#) [- Recording](#)		Obje funct %>%
9/06	[Intro to R 2](#) [- Recording](#)		Visua in R
9/11	[Intro to R 3](#) [Recording 1](#), [2](#), [3](#)		More visua
9/13	Review Session [- Recording](#)		
9/18	[Data Wrangling](#) [- Recording](#)		Repli and
9/20	[Univariate](#) [- Recording](#)		Sumr one v
9/25	[Multivariate 1](#) [- Recording](#)		Sumr multi varia
9/27	[Multivariate 2.a](#) [- Recording](#)		Visua of m varia
10/02	[Multivariate 2.b](#) [- Recording](#)		Visua of m varia
10/04	[Multivariate 3](#) [- Recording](#)		Unce and boot
10/09	Review Session [- Recording](#)		
10/11	[Regression 1](#)		Linea

Date	Lecture	Learr
	- Recording	regre
10/16	Regression 2 - Recording	Inter and evalu
10/18	Regression 3 - Recording	Mult regre categ
10/23	Review Session -Recording	
10/25	**Midterm Exam**	
10/30	Classification 1 - Recording	Logis Regre
11/01	Classification 2 - Recording	Inter and evalu
11/06	Classification 3 - Recording	Predi
11/08	Review Session - Recordinghttps://brightspace.vanderbilt.edu/d2l/le/content/442169/viewContent/3046369/View)	
11/13	Clustering [- Recording]	k-me clust
11/15	NLP 1 [- Recording]	Topi
11/20	BREAK	
11/22	BREAK	
11/27	NLP 2 [- Recording]	Senti analy
11/29	Review Session	
12/04	Ethics [- Recording]	Ethic scien
12/06	Review Session	
12/11	**Final Exam**	

Helpful Resources 🔗

[Rstudio Cheat Sheet: Data Wrangling](#)

[Rstudio Cheat Sheet: ggplot2](#)

[R-graphics Cookbook](#)

[... And the full list of Rstudio cheat sheets](#)

[Tidymodels Resources](#)

[Back to ToC](#)

Acknowledgements 🔗

The contents of this course and of my teaching pedagogy are influenced and inspired by:

- Emily Hencken Ritter, Vanderbilt University
- Andrew Princep, MarketWatch & Oxford University
- Josh Clinton, Vanderbilt University
- William Doyle, Vanderbilt University
- Matthew Salganik, Princeton University

This course is modeled on the course of the same name, taught by Professors Josh Clinton and William Doyle in the spring of 2022 at Vanderbilt University. The lectures on data science & ethics are inspired by content prepared for the Summer Institutes for Computational Social Sciences ([SICSS](#)), and by presentations on machine learning and ethics prepared by Andrew Princep ([@AJPrincep](#)). The syllabus is heavily inspired by [Emily Hencken Ritter](#)'s syllabi for PSCI 3270, Politics of Human Rights.

[Back to ToC](#)

Releases

No releases published
[Create a new release](#)

Packages

No packages published
[Publish your first package](#)

Deployments 78

✅ **github-pages** yesterday

[+ 77 deployments](#)

Languages

● **HTML** 95.1%   ● **JavaScript** 4.2%   ● **Other** 0.7%