PSYC 81.09: Storytelling with Data (Winter 2024)

Meeting times: T/Th 10:10 — 12 PM
X-hour: F 3:30 — 4:20 PM
Classroom: Moore 302

Instructor: Dr. Jeremy R. Manning
Email: jeremy@dartmouth.edu
Office Hours: By appointment on Tuesdays

Stats help: Dr. Jianjun Hua
Email: jianjun.hua@dartmouth.edu
Office Hours: Wednesdays from 2—4 PM (Rocky Center Atrium or via Zoom)

Course Description

Humanity creates nearly 350 TB of data each day, and much of it is freely available for us to download. But making sense of that staggering amount of text, video, audio, and other forms of data is like drinking from a firehose: the sheer quantity of data makes it tricky to make sense of what it *means.* In this course we will study how to find, clean, analyze, make sense of, and communicate about complex data.

In addition to helping you to build up your data science toolbox, the course is centered around telling engaging, compelling, (often) entertaining, and, above all, *clear* "stories" about data. This will require considering your audience's mindset, goals, motivations, and backgrounds, and then leveraging those considerations to develop effective ways of conveying the major plot points.

Course Goals

The primary goals of this course are to help you develop intuitions and tools for (a) finding meaningful patterns in complex datasets and (b) communicating your findings in an intuitive way. To that end, you will be learning to write basic computer programs (in Python), analyze and create plots, and present your work effectively. I see data interpretation and visualization as a deep psychological issue: creating compelling figures and stories is just as much about theory of mind of the receiver as it is about accurately reporting the deeper truths underlying the dataset. We will therefore motivate the tools and skills we develop from a psychological and neuroscientific perspective. My hope is that you will draw heavily on what you learn in my course in "real life," in whatever fields, professions, and/or hobbies, you choose to pursue.

Pre-Requisites

You should have taken a course on statistics or probability (e.g. PSYC 10, AP Stats, or similar). An online statistics course is fine as a stand-in, but I will expect you to know about hypothesis testing, probability distributions, and have some intuitions about how to design and interpret statistical analyses *before you start this class*. You will also benefit from prior work in experimental design, such as PSYC 11 or similar. Additional prior course work or experience in Psychology, Neuroscience, Computer Science, Math, Engineering, Creative Writing, Theater, and/or Graphic Arts will also be useful.

In defining these pre-requisites somewhat vaguely, I hope we will end up with a group of students with a diverse array of backgrounds, interests, and skills. Some students may be more "quant" minded (they might have an easier time learning and applying the programming and analysis components of the course) and others will be more "humanities" minded (they might have an easier time crafting and presenting compelling stories). However, although you don't need to know how to program before you start this course, every student will eventually need to both program analyses and present results to the class. Ultimately we'll all build on our existing skills and experiences to help each other grow intellectually and learn new practical skills.

Course Materials

All required text, tutorial, and audiovisual materials will be freely available to you throughout the term. You do not need to purchase any additional materials to fully participate in this course. A complete outline, with links to all materials, assignments, and lecture recordings, may be found <u>here</u>.

You will also need an internet-enabled computer or tablet capable of displaying and outputting graphics and running a standard web browser (e.g., Firefox or Chrome). You'll use your computer to participate in Zoom meetings, engage with self-paced learning materials, and complete your assignments.

In addition to the above required materials, you may **optionally** find it useful to read through one or more of the following classics in data visualization by Edward Tufte:

> <u>The Visual Display of Quantitative Information</u> (ISBN: 0961392142)
> <u>Envisioning Information</u> (ISBN: 0961392118)
> <u>Visual Explanations: Images and Quantities, Evidence and Narrative</u> (ISBN: 0961392126)

Purchasing these books is optional, and we will not even be referencing them directly in the course. However, I am including them in our syllabus because I have found the books to be particularly excellent for drawing inspiration from a true master of data visualization. If you are an aspiring data scientist, I highly recommend investing in these books at some point in your career. In the meantime, I suggest checking the books out of the library and glancing through the examples as a reasonable way of determining whether they'll be worth the investment for you.

Format and Overview

The first 4(ish) weeks of the course are primarily instructor-directed, supplemented by in-class discussions and assignments. We'll cover four "modules," with each module lasting for roughly a week: (1) What makes a good story?, (2) Visualizing data, (3) Python and Jupyter notebooks as a medium for data storytelling, and (4) Data

science tools.   These modules will introduce you to the fundamentals of storytelling with data.   Our course incorporates a Social Impact Practicum component as part of module 4.   This year we'll be partnering with Climate Interactive to examine (and tell stories about!) climate data.

New in the 2024 offering of this course, we will also engage with the Hopkins Center for the Arts, including a "field trip" to a performance and an in-class discussion with the the author to explore ways of conveying meaning and information through sound.

During the second phase of the course, you (either in groups or individually) will download datasets, analyze them, produce figures or other representations, and informally present your results as "data stories" to the class each week as YouTube videos.   We will then constructively critique and make suggestions about each others' stories, helping each other to become better data storytellers.

Slack

We will use Slack (a tool for organizing notes, files, and conversations) to provide a forum for asking and answering questions, posting demos, etc.   You will need to join the class workspace at the beginning of the term by following the link.   We'll set up a series of channels (one for each topic we decide to explore) and we will use the tool during and outside of class to keep track of our thoughts and ideas.   Your first two assignments will be submitted using Slack.

GitHub

We will use GitHub to manage and share data and code.   GitHub provides an easy way of managing multiple versions of data and code that may be easily shared and tracked.   To use GitHub, you will need to create a (free) GitHub account.   After the second assignment, all other coursework will be submitted via GitHub.   (You will learn how to use GitHub as part of Module 3.)

Google Colaboratory

We will use Google Colaboratory to develop and program analyses.   This tool provides an easy means of organizing notes, code, and graphics in a single cohesive format ("notebook").   We will also use Google Colaboratory notebooks to tell our data stories; they combine analyses, reports, and presentations into a single document. The programming environment is accessed via a web-browser.   (You will learn how to set them up and run the notebooks in Module 3 during the first part of the course.)

Grading

All course assignments will be assigned a point value, added together, and converted to the nearest equivalent letter grade as follows (all scores in parentheses are percentages of the total possible number of points): A (95–100); A- (90–94); B+ (87–89); B (83–86); B- (80–82); C+ (74–79); C (57–73); C- (50–56); D+ (44–49); D (37–43); D-(30–36); or F (0–29).

Most of the course assignments will take the form of a YouTube video presenting a story, often with an accompanying "notebook" containing the associated code and data.   You and your peers will provide feedback to each other (both formally and in class discussions).   Your self assessments, peer evaluations, and my feedback will all contribute to your grade for each "story" assignment.

Any assignment may be re-done at any point in the course (up to the last minute of the last day we'll meet this term— 11:59pm on Tuesday, March 5). If you choose to hand in a new version of any assignment, only the latest version will "count" towards your final grade.

Scheduling Conflicts

This class works best when students are participating synchronously. When you are able, I expect you to attend and be on time for every class meeting.   However, I also understand that your personal circumstances (e.g., health concerns, religious observances, other commitments, emergencies, etc.) may limit your ability to participate synchronously.

If you are generally able to attend the synchronous course meetings, I would appreciate a heads up (as far in advance as you're able to tell me) if you know you will be missing class. That will help me prepare (and, if necessary, reconfigure) group activities, etc.

If you are generally *not* able to attend the synchronous course meetings, please reach out to me as soon as possible (and before the end of Week 2 at the latest) so that we can figure out a plan that will help you to get as much as you can out of the course.

Importantly, however, **no part of your grade will depend on attendance**. I trust that you will make the best decision for yourself, based on your personal goals and circumstances, about whether or not to attend each class meeting. By the same token, if you do miss a class meeting, it is your responsibility to make up the material.

Student Needs

I strive to maintain a welcoming and accessible classroom environment. I want you to be an active participant and contributor to ongoing discussions and activities, and that means that every student should feel comfortable in my (in-person or virtual) classroom. If you would like me to be aware of any issues that arise during the term, or any personal needs that may require adjusting how I run my class or how you participate, I encourage you reach out to me privately. Dartmouth's <u>Student Accessibility Services Office</u> can also help assist with setting up disability-related accommodations.

The Academic Honor Principle

I expect you to abide by Dartmouth's <u>Academic Honor Principle</u>. This course is a bit different from some others, in that I encourage (and expect) you to discuss your assignments with your classmates. The class will be heavily collaborative, and I encourage group presentations and collaboration on assignments. You can also use "AI" tools like ChatGPT to help analyze your datasets, come up with story ideas, flesh out your writing, and so on. In short, you are encouraged to use any and all tools and resources at your disposal to enable you to do **your** best work.

The above said, note the emphasis on "your" in the previous sentence. It's important that you contribute something substantive to each project that you participate in, and that **you** actively engage with the material. Copying and pasting from another student, or from ChatGPT (et al.) will not fulfill your assignment requirements. Similarly, you cannot "re-use" projects from other courses without modifying them, although some projects will allow you to (optionally) build on prior work.

Put simply, you should hand in your own (new) work, even if you collaborated or discussed your assignment with a classmate, used other resources or tools, an so on. If you have any questions about the Academic Honor Principle and how it applies generally to this course, or specifically to a particular assignment, please ask me.