


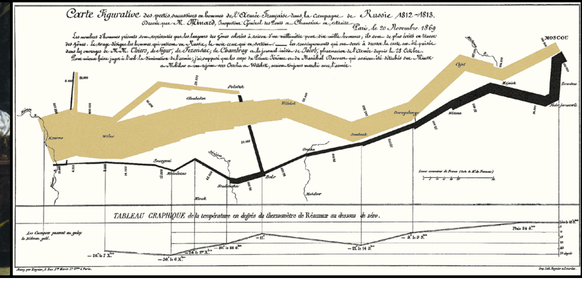

PSYC 81.06: Storytelling with Data (Spring 2019)

Meeting times: MWF 2:10 - 3:15 PM
X-hour: Th 1:20 - 2:10 PM
Classroom: Moore 302

Instructor: Dr. Jeremy R. Manning
Email: jeremy@dartmouth.edu
Office Location: 349 Moore Hall
Office Hours: During unused X-hours and by appointment

Course Description

In a world plagued by "alternative facts" but flush with "big data," how can we find truth? For example, can truth be objectively defined, or are there many equally valid truths? And does truth depend on the question we're asking, or is it a fixed property that we could somehow uncover with the right analysis? These sorts of question align with other deep questions about how we can really "know" something. For example, can we really ever hope to prove that the universe works in a particular way? If so, how? Or if not, what's the point of observing the world around us at all, or of becoming a scientist?

In this course we will define truth from a (somewhat cynical, but embarrassingly practical) psychological perspective: ***truth is the story about data that others find most convincing***. To that end, we will examine tools and strategies for finding patterns in complex datasets, crafting convincing stories about those patterns, and communicating them to others. One major focus will be on creating impactful data visualization and infographics. The second major focus will entail you "getting your hands dirty" by applying the tools we learn about to various datasets and telling stories about data to your classmates. Along the way, we will leverage findings from the psychological and neuroscientific literatures to create more convincing stories by exploiting the visual system's ability to extract patterns from the world around us, and by examining what distinguishes effective stories from ineffective ones.

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*. 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

We will reference miscellaneous figures and chapters from three classics in data visualization by Edward Tufte (purchasing these books is **optional**, but they are excellent for drawing inspiration from a true master of data visualization):

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

I will also provide various (free) PDFs for download via Canvas or GitHub throughout the term.

You will also need a laptop computer (Mac, Windows, or Linux) capable of displaying and outputting graphics and running a web browser. (A Chromebook will not work, nor will an iPad or similar.) We will use our laptops nearly every day in class to do demos, hands-on exercises, and presentations.

If obtaining these materials presents a financial or logistical hardship for you, please come speak to me as early in the term as possible.

Format and Overview

The overall format of the course will adapt to student interests and needs. My plan is for the first 3(ish) weeks of class to be primarily lecture-based, supplemented heavily by in-class exercises. I'll provide a general overview of programming in Python, data and code management, and different "classic" types of plots. (Some of the programming instruction will be redundant with CS 1, but students already familiar with programming will have the opportunity to explore further.)

During the remainder of the course, you (either in groups or individually) will download datasets, analyze them, produce figures, and informally present your results to the class. We will then constructively critique and make suggestions about each analysis and presentation, drawing on and/or exploring the relevant psychological and neuroscientific literature where applicable.

Will will approach the first of these projects as a class, using a dataset from a local nonprofit organization. We will meet with a representative from the organization, discuss what they are hoping to learn from their data, and then brainstorm as a class how we might meet their goals. Individual students and/or small groups will be responsible for exploring specific questions or tasks, and as a class we will put together our first "story" about the full dataset. (This is part of the Social Impact Practicum program at Dartmouth.)

After practicing and developing your data analysis and "story telling" (presentation) skills with several new datasets in "mini projects," the course will culminate in a more in-depth data storytelling project. For your final project, you will carry out an in-depth battery of analyses of a dataset of your choosing and present the results of your analysis in a more formal talk and final paper.

Jupyter

We will use Jupyter to develop and program analyses. This software provides an easy means of organizing notes, code, and graphics in a single cohesive format ("notebook"). We will also use Jupyter notebooks for the informal class presentations. The programming environment is accessed via a web-browser. (You will learn how to set them up and run the notebooks at the beginning of the course.)

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. You will need to create a (free) GitHub account at the beginning of the term. All assignments will be submitted via GitHub.

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 team at the beginning of the term by creating a (free) Slack account. 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.

Approximate Schedule

The schedule below is intended to give a rough sense of how the course will progress, and to indicate several days when X-hours will be used to replace regularly scheduled course meeting times. The content is subject to change.

Date	Description	Additional notes
Week 1	**Introduction and Overview**	
March 25, 2019 (Monday)	Introductory discussion, setting up computing tools, brief introduction to programming	Assignment 1 handed out: GitHub and programming basics
March 27, 2019 (Wednesday)	Approaches to data visualization	
March 28, 2019 (Thursday)	X-hour used for office hours	

Date	Description	Additional notes
March 29, 2019 (Friday)	Collaborative coding, presenting demos for Assignment 1	Assignment 1 due, Weekly snippet 1 due, Assignment 2 handed out: finding and wrangling data
Week 2	**Data wrangling**	
April 1, 2019 (Monday)	Numpy and Pandas	
April 3, 2019 (Wednesday)	Overview of scipy, scikit-learn, and hypertools	
April 4, 2019 (Thursday)	X-hour used for office hours	
April 5, 2019 (Friday)	Collaborative coding, presenting demos for Assignment 2	Assignment 2 due, Weekly snippet 2 due, Assignment 3 handed out: data visualization
Week 3	**Science hacking**	
April 8, 2019 (Monday)	Discussion of science hacking strategies	Watch science hacking tutorial prior to class, come prepared with ideas and questions
April 10, 2019 (Wednesday)	Hackathon focused on SIP data	
April 11, 2019 (Thursday)	X-hour used for office hours	
April 12, 2019 (Friday)	Social impact practicum dataset: discussion with community partner and overview of the data	Assignment 3 due, Weekly snippet 3 due
Week 4	**Exploring social impact practicum data as a class**	
April 15, 2019 (Monday)	Hackathon day	
April 17, 2019 (Wednesday)	Hackathon day	
April 18, 2019 (Thursday)	X-hour used for office hours	
April 19, 2019 (Friday)	Present stories about SIP data	In-class presentations and demos, Weekly snippet 4 due
Week 5	**Pitching hackathon ideas**	
April 22, 2019 (Monday)	Discussion: what makes a good pitch?	Discussions, brainstorms, and project pitches
April 24, 2019 (Wednesday)	Hackathon day	
April 25, 2019 (Thursday)	***X-hour used as make-up class***	
April 26, 2019 (Friday)	Hackathon day	Weekly snippet 5 due
Week 6	**Data hacking part 1**	
April 29, 2019 (Monday)	Hackathon day	

Date	Description	Additional notes
May 1, 2019 (Wednesday)	Hackathon day	Informal updates on each project
May 2, 2019 (Thursday)	***No X-hour***	Instructor out of town
May 3, 2019 (Friday)	***No class (can use classroom for hackathon projects)***	Instructor out of town, Weekly snippet 6 due
Week 7	**Collaborative science**	
May 6, 2019 (Monday)	Present data stories, discussion about building on prior work	In-class presentations, demos, and suggested future directions
May 8, 2019 (Wednesday)	Project pitches and hackathon	
May 9, 2019 (Thursday)	***X-hour used as make-up class***; hackathon day	
May 10, 2019 (Friday)	Hackathon day	Informal updates on each project, Weekly snippet 7 due
Week 8	**Data hacking part 2**	
May 13, 2019 (Monday)	**No class (can use classroom for hackathon projects)**	Instructor out of town
May 15, 2019 (Wednesday)	**No class (can use classroom for hackathon projects)**	Instructor out of town
May 16, 2019 (Thursday)	***X-hour used as make-up class***; hackathon day	
May 17, 2019 (Friday)	Present data stories	In-class presentations, demos, and suggested future directions; Weekly snippet 8 due
Week 9	**Final projects**	
May 20, 2019 (Monday)	Project pitches and hackathon	
May 22, 2019 (Wednesday)	Hackathon day	
May 23, 2019 (Thursday)	X-hour used for office hours	
May 24, 2019 (Friday)	Hackathon day	Informal updates on final project; Weekly snippet 9 due
Week 10	**Final project presentations**	
May 27, 2019 (Monday)	Final project presentations day 1	
May 29, 2019 (Wednesday)	Final project presentations (day 2), final discussion and recap	Last day of class; Weekly snippet 10 due; All course materials and assignments must be added to the central GitHub repository

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 (93–100); A- (90–92); 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).

Lab exercises (10%)

At the beginning of the course (during weeks 1—3), you'll build experience programming in Python, plotting and manipulating data, managing code, etc., via three problem sets. You'll start working on the problem sets in class, working through the problems in groups (with instructor supervision and assistance), and then you'll finish up the problem set outside of class as needed.

In class presentations and demos (50%)

In weeks 4—8, you'll start working on "mini projects" where you tackle several (2–3) small-to-medium scale analyses of datasets on a topic (or topics) of your choosing (aside from the first dataset, which we will explore together as a class). Each may be carried out individually or in small groups (or potentially as a class, depending on the number of interested students). For each analysis, you'll begin by brainstorming ideas for your mini project and discussing as a class. Then you'll carry out some of the proposed analyses. Finally, you'll turn your results into a succinct story that you'll present (briefly) to the class. Your total grade for this portion of the course will consider both the quantity and quality of the projects (and stories!) you produce during the course.

Weekly snippets (10%)

Each week you will write a brief (3–5 paragraphs total) reaction to the material and discussions from the preceding week. These "weekly snippets" will be made visible to the entire class, and we'll use them to help guide the topics we cover moving forward. You will answer the following three questions in a few sentences each:

1. What did you learn and/or work on over the prior week?
2. What are your goals for the coming week?
3. What are you stuck on and/or having trouble with?

Each reaction will be graded as "complete" (full credit) or "incomplete" (no credit). Reactions are due on the morning of the last class of each week.

Final project (30%)

During the last 2(ish) weeks of class you will tackle a larger analysis (or series of analyses) for your final project. You'll present your results at the end of the course in a (slightly) longer-format presentation where you integrate across the tools and techniques you acquired in the rest of the course.

The Academic Honor Principle

I expect you to abide by Dartmouth's Academic Honor Principle at all times. 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. However, it's also important that your weekly reactions are written by you (rather than copying from a classmate). In addition, 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. 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.

Scheduling Conflicts

This class requires you to be physically present (e.g. to take part in the in-class discussions, give presentations, participate in lab exercises, etc.). I expect you to attend *and be on time for* every class. ***If you know you will have a scheduling conflict with this course during the term, please meet with me before the end of Week 2 to discuss appropriate arrangements.***

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 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 to see me privately. Dartmouth's Student Accessibility Services Office can also help assist with setting up disability-related accommodations.