Syllabus for ANLY503 Data Visualization

Georgetown University, Fall 2020

This syllabus is effective as of Wednesday, October 14, 2020 at 08:41 PM

Course Information

- **Instructors:**
  - Abhijit Dasgupta (abhijit.dasgupta at georgetown.edu)
  - Marck Vaisman (marck.vaisman at georgetown.edu)
- **TA's:**
  - Adam Imran (ai410)
  - Leslie Liu (hl755),
  - Kiwi Yu (ky240),
- **Class Schedule:**
  - Class meets Thursdays 6:30-9:00pm
  - Recurring Zoom Link (GU credentials required): https://georgetown.zoom.us/j/99864511820
  - Both sections will start together and break out from the main Zoom session (no separate Zoom Link)

Course Description

Data visualization is both art and science. More and more, the products of statistical analysis that we encounter in our everyday lives come in the form of data visualizations. Visualizations have the advantage of being easier to interpret for many people, but they also give the impression of being a form of absolute truth. As with any presentation method, though, visualizations can be manipulated by their creators to show the story they are trying to tell. In this class, we will explore the many methods of information visualization, and develop intuition for when data graphics are not telling the truth.

This course will draw upon methods from statistics, graphic design, and computer science. We will approach these ideas from the ground up, and they will all be framed in the context of visualization. As a result, you can expect to come away from this course with a basic understanding of concepts from all of the aforementioned disciplines, as well as one place where they fit together.

Course Objectives

By the end of the term, we expect you to be able to:

- Increase your data visualization vocabulary
- Understand what comes before and after creating visualizations
- Think critically about data
- Distinguish between using visualizations for data exploration or presentation
- Manipulate and arrange different kinds of data for the purposes of analysis and visualization
- Design effective visualizations for different purposes that are easily understandable by different types of audiences
- Apply a set of rules to create highly effective and engaging data visualizations
- Understand the role of data visualization within data science and for solving problems
- Use an array of tools (including R, Python, Tableau) to execute all of the above

Readings

There is no required textbook for the course. We have selected specific readings from many sources and these will be provided to you in PDF files on Canvas. Lectures may or may not follow the readings. **You must read assigned readings prior to the lectures.**

Software

We will mainly use two scripting languages in this course: R and Python. In particular, we will make heavy use of the tidyverse group of packages and its philosophy in R, and Python packages in the PyData ecosystem, primarily `pandas`, `matplotlib`, `seaborn`. Several other packages in both R and Python will be introduced as needed.

We will also use Tableau, a popular data visualization software.

Instructions for installing all of these, and the requisite packages, will be added here soon.

Learning Activities, Communication and Evaluation

This is a hands-on, practical, workshop style course that provides opportunities to use the tools and techniques discussed in class.

Lecture/Lab Format

The course is split into a lecture/lab format. During the lecture, we will discuss the concepts and techniques. During the lab sessions, you will be completing exercises and following examples which are designed to show you how to implement the ideas and concepts with different tools.

Some class meetings may be more lecture focused, some may be lab focused. Lectures may not cover all the material.

Assignments

Homework assignments are weekly and will focus on exercises that allow you to apply of what was discussed in class the previous week.

Important Note: We reuse problem set questions, we expect students not to copy, refer to, publish or make public, or look at the solutions in preparing their answers. Since this is a graduate-level class, we expect students to want to learn and not search online for answers.

Online Participation

We will host several asynchronous, time-bound, online discussions on Canvas:

1. You must make a *main post* that responds to specific questions noted. Please make your main post early enough so that others have time to respond to you.
2. You must also respond to **at least two other student main posts.**
3. Discussions are time bound and cannot be late.

Online Porfolio

You will create an online presence via a hosted website. In this portfolio, you will showcase the work you have done in this class. Several of the homework assignments will be a part of your portfolio, as well as the final project.

More details will be provided.

Final Project

You will create an R Markdown dashboard using the `flexdashboard` package to visualize data sets in at least 3 ways and to create a visual narrative, to show what your data looks like and what your analytic results look like. This dashboard will also be a part of the online portfolio, but it is an separate assignment.

We will provide information about datasets for the final project by week 10.

Grading

- Homework: 50%
- Online Portfolio: 20%
- Final Project: 20%
- Online Discussion Participation: 10%

Total is 100%. We have no plans to curve or adjust the final grade, so the letter grade will follow standard guidelines:

- A: >= 91.5
- A-: >= 89.5, < 91.5
- B+: >= 87.99, < 89.5
- B: >= 81.5, < 87.99
- B-: >= 79.9, < 81.5
- C: < 79.9

Course Calendar

This calendar is subject to change. We will make make any changes known in advance.

Class	Module	Date	Topics	What is Due
1	1 - Conceptual	Aug 27	History and purpose of dataviz, designing for an Audience	
2	1 - Conceptual	Sep 03	Picking the right visualization	A1 due Fri 9/4
3	1 - Conceptual	Sep 10	Making readable graphics, putting it all together	A2 due Fri 9/11
4	2 - Tools & Data	Sep 17	Tools overview	A3 due Fri 9/18
5	2 - Tools & Data	Sep 24	Data prep for visualization	
6	3 - Static Visuals	Oct 01	EDA Visualization, Static graphics with R/Python	
7	3 - Static Visuals	Oct 08	More Static graphics with R/Python	A4 due Fri 10/9
8	4 - Specialized Visuals	Oct 15	Maps and geospatial data	
9	4 - Specialized Visuals	Oct 22	Networks	A5 due Fri 10/23
10	4 - Specialized Visuals	Oct 29	Text data	A6 due Fri 10/31
11	5 - Dynamic Visuals	Nov 05	Dynamic graphs 1	A7 due Fri 11/6
12	5 - Dynamic Visuals	Nov 12	Dynamic graphs 2	A8 due Fri 11/13
13	6 - ML	Nov 19	Visualizing model dianostics and results	A9 due Fri 11/20

Class	Module	Date	Topics	What is Due
14		Dec 03	Wrapup	Final Project and Online Portfolio due Fri 12/11

Calendar notes:

- We will have a guest speaker, Jonathan Schwabish towards the end of the semester. Date TBD.
- * Readings will be posted a week before they are due, please check the calendar frequently

Policies & Expectations

General Policies

- **Online Zoom Rules**
  - Participate and speak while on Zoom. We know it's harder to do classes over Zoom, but we love participation. Ask questions. Make comments. Challenge us. Acknowledge us. If we speak for three hours to a silent classroom, it is a lot more boring and tiring for everyone.
  - Focus your attention to the class. Online classes require higher level of engagement for everyone. Don't multitask (you know what we mean. . .)
  - Turn on your camera. If you have a bad hair day, that's ok (we do too.) We want to see you. If you have bandwidth issues and can't use your camera, we understand as well.
- **Course Communication:** Please use the provided **Canvas Discussion Boards** (separate than the ones from required class discussion) for questions about the course, homework assignments, technical issues, etc. Staff will be monitoring them and providing answers on a regular basis. Individual emails will not be answered except for special circumstances.
- **Due Dates:** Homework assignments are posted on Thursdays and will be due on the Friday a week after they are assigned, at the end of the day Eastern Time. Due dates will not be extended.
- **Late Policy:** Late assignments or discussions will not be accepted and will get zero (0) points. Period. The two lowest assignment grades will be dropped. In lieu of a late policy this allows you to miss **up to two weekly assignments. This DOES NOT include discussions, portfolio or final project.**
- **Class materials are for class use only!:** Please refrain from making your private GitHub repositories or any other class materials public. A breach of this request is considered an Honor Code Violation.

Open Door Policy

Please approach or get in touch with us if something is not working for you regarding the class, methods, etc. Our pledge to you is to provide the best learning experience possible. If you have any issue please do not wait until the last minute to speak with us. You will find that we are fair, reasonable and flexible and we deeply care about your success.

Collaboration Policy

- **All individual work is by definition, INDIVIDUAL. We have a ZERO TOLERANCE POLICY for cheating and not doing individual work when required. Homeworks, Online Portfolio, and the Final Project are ALL INDIVIDUAL.**
- If an assignment deliverable looks too similar another one (or more) then the grade will be divided among all students whose work is similar.
- You can, and should, work together on the asynchronouise online discussions.

Grading Policy

Grading is holistic, meaning that there is no specific point value for individual elements of an assignment. Each assignment submission is unique and will be compared to all other submissions for the assignment. If

the submission meets or exceeds the requirements, is creative, is well thought-out, has proper presentation and grammar, and is at the graduate student level, then the submission will get full credit. Otherwise, partial credit will be given and deductions will be made for any of the following reasons:

- Instructions are not followed
- Poor and sloppy writing and/or presentation, including spelling and grammatical errors
- Code is not documented with comments
- Code files referenced are not in repository
- Absolute links included in repository
- Submitting something for the sake of submitting it, without thinking through and not providing analytic justification

Academic Integrity

All students are expected to maintain the highest standards of academic and personal integrity in pursuit of their education at Georgetown. Academic dishonesty, including plagiarism, in any form is a serious offense, and students found in violation are subject to academic penalties that include, but are not limited to, failure of the course, termination from the program, and revocation of degrees already conferred. All students are held to the Georgetown University Honor Code. For more information about the Honor Code see http://gervaseprograms.georgetown.edu/honor/

University Policies and Support Services

Accommodations for students with disabilities

Students with documented disabilities have the right to specific accommodations that do not fundamentally alter the nature of the course. Please alert us should you require accommodations.

Title IX Sexual Misconduct Statement

ho Please know that as faculty members we are committed to supporting survivors of sexual misconduct, including relationship violence and sexual assault. However, university policy also requires us to report any disclosures about sexual misconduct to the Title IX Coordinator, whose role is to coordinate the University's response to sexual misconduct.

Georgetown has a number of fully confidential professional resources who can provide support and assistance to survivors of sexual assault and other forms of sexual misconduct.

More information about campus resources and reporting sexual misconduct can be found at http://sexualassault.georgetown.edu