STAT 633 - Data Visualization - Final Project

Due date to submit presentation and accompanying files: Wednesday, 5/15/2024, 5:59pm EST

Presentations to be held: Wednesday, 5/15/2024, 6:00pm-8:00pm EST

Your presentation must be submitted in Moodle with the following files:

(1) RMarkdown presentation format file (Rmd)
(2) an html file
(3) if your presentation includes shiny apps, one R script for each app. In this case, please have a slide in your presentation that you submit that refers the reader to the appropriate R script. In the presentation that you show in class, you may try to embed the shiny app within your slides using the function `shiny::shinyAppFile()` within an RMarkdown code chunk. This sometimes works and sometimes doesn't, be sure to test this option thoroughly if you choose to use it for the in class presentation.
(4) Any data file you utilize.

The instructor should be able to recreate all of your plots and apps based on the code and data you submit.

Your task is to tell a convincing story that uses data visualization tools.

The presentation must include at least **one shiny app that uses materials about using `plotly` events as inputs in shiny apps**.

You **may** (after consultation with me) be able to reuse materials from Presentation #1 or #2. This may only be allowed if you convince me that you have a well structured question and plan for your project, one that thoughtfully incorporates the feedback you will have received for the second presentation, both from instructor and peers. For this to be approved you must enhance the original presentation much beyond its original scope - by adding thoughtful analyses, incorporating additional data, wise apps etc. This can be a great learning opportunity, but because you'll have the advantage of already having presented and receiving feedback, the expectations will be higher.

Content that must be addressed:

- Start your presentation with the interesting **question** you are trying to address.
- Then tell us why the question is important.
- Formulate an hypothesis (or a set of those) you are directly exploring using visualizations.
- And then tell us about the data you use.
  - Which data sets are you using?
  - What are the variables? what do they really measure? what are the sources for your data? what may be potential weaknesses in it?
  - Which of the variables are proxies? (recall the presentation of Prof. Miriah Meyer for the first lecture in this class)
  - If you combine several data sets, please explain briefly how you did so technically (join, pivot etc.)
  - In the presentation, you may use any of the visualization techniques that we learned (and more) as you see fit.
- The idea is to show that you use your technical skills wisely - more bells and whistles do not always serve your narrative. You must verify that your plots are legible when you share your screen via Zoom. All fonts, markers, colors, lines or any other aesthetic property must be clear enough to read easily and distinguish.

- Your goal in this presentation isn't necessarily to answer the research question. It is to show that you develop smart tools that may assist in finding meaningful answers to it. One such meaningful answer may be that the data you chose is limited in its capacity to provide useful information about your question.
  - E.g. an interesting question may be something like: "how is social inequality affecting life expectancy and how does this relationship evolve over the years?"
  - The gapminder data set may or may not contain information that can assist us in answering this question.
  - Perhaps combined with other data sets it can help.
  - How do we measure social inequality? Is gender a good proxy for that? Or is the relationship between gender and social inequality more complicated? Even if it is more complicated, what may we gain by conditioning the life expectancy on gender?
  - We have to think carefully if we want to examine the question by... continent? predominant religion in a country? fertility rates?
  - Perhaps we want to examine inequality by always looking at the difference between the countries with lower income vs. countries with higher income (in which case the identity of the countries that we discuss may vary from year to year, recall the shiny gapminder exercise where we filtered the data by quantiles)
- Some sources for free datasets:
  - Built into R: https://www.rdocumentation.org/packages/datasets/versions/3.6.2
  - Other free sources: https://r-dir.com/reference/datasets.html
  - Google public datasets: https://www.google.com/publicdata/directory#!
  - Datasets for data mining and data science: http://www.kdnuggets.com/datasets/index.html
  - World bank data: https://data.worldbank.org/
  - OECD data: https://data.oecd.org/
  - U.S. Department of Health and Human Services medical expenditures panel survey: https://meps.ahrq.gov/mepsweb/
  - Centres for Medicare and Medicaid Services: https://data.cms.gov
  - StatCrunch: https://www.statcrunch.com/datasets/shared
  - Bureau of Transportation Statistics: https://www.bts.gov/
  - http://www.stat.ufl.edu/~winner/datasets.html
  - Open Gov. Data: https://www.data.gov, https://www.data.gov.uk, http:/opengovdata.org/
  - US Government Spending Data: https://www.usaspending.gov/
  - Kaggle: https://www.kaggle.com
  - UCI Machine Learning Repository: http://archive.ics.uci.edu/ml/
  - StatLib: http://lib.stat.cmu.edu/datasets

Administrative stuff:

- There will be 9 groups, 10 minutes per presentation with 3 minutes Q&A for each presentation. **The time limit will be strictly enforced** (this time, for real). You must rehearse to make sure you meet the requirement.

Assignment to groups:

First name	Last name	Email address	Group for Final
Giana	Davila	gdavila@umass.edu	1
Minsu	Kim	minsu@umass.edu	1
Amani	Chehimi	achehimi@umass.edu	1
Shane	Fitzgerald	skfitzgerald@umass.edu	2
Rebekah	Kristal	rkristal@umass.edu	2
Bartu	Bingol	bbingol@math.umass.edu	2
Ping	Yang	pinyang@umass.edu	3
Kuan-Jung	Huang	kuanjunghuan@umass.edu	3
Jiacheng	Wang	jiachengwang@umass.edu	3
Ariel	Lutati	alutati@umass.edu	4
Benjamin	Goldberg	bhgoldberg@umass.edu	4
Ian	Knightly	iknightly@umass.edu	4
Dylan	Jacobs	dcjacobs@umass.edu	5
Richard	Heath	rlheath@umass.edu	5
Nikki	Lin	jiaqilin@umass.edu	5
Ziyan	Zhao	ziyanzhao@umass.edu	6
Lindsay	Knupp	lknupp@umass.edu	6
Xinyu	Zhou	xzhou@umass.edu	6
Paul	McDonnell	pmcdonnell@umass.edu	7
Derrick	Wu	derrickwu@umass.edu	7
Duc	Nguyen	dmnguyen@umass.edu	7
Jun	Pu	junpu@umass.edu	8
Chsristy	Yang	yuxyang@umass.edu	8
Jenny	Ding	junwending@umass.edu	8
Andrew	Hennessey	aphennessey@umass.edu	9
Paul	Shannon	pshannon@umass.edu	9
Owen	Gallagher	ogallagher@umass.edu	9

- **Set up a 30 minutes meeting with me to discuss your tentative ideas and methods for the project**.
- **Meetings with groups will be held on Friday 5/3/2024, 11:00am-1:30pm or 4:00-6:00pm**.

https://docs.google.com/spreadsheets/d/1tXSC230cWTOE8eT43LYWsbkOabNP3gU8H_gZbi2pQWo/edit?usp=sharing

- The order of the presentations will be posted later.
- Allow all members of the group to express themselves.
- Your grade will be full if you demonstrate thoughtfulness - if I see that you made a sincere attempt to provide us with a good story. Mistakes are welcome and instructive as long as they are not a result of pure laziness.
- Each student will be asked to provide short feedback in writing about the presentations of each of the other groups. This feedback will be collected (anonymously) and sent to the presenters.
- Each student will be asked to report which sections of the presentation they were doing individually, and which collaboratively.
- As a courtesy to your peers, you are required to turn your camera ON throughout the entire session.
- You are welcome to consult me anytime (via email or zoom) from now till the due date about anything

related to the presentation.