HW 5 - Presentation 1

Presentation time: 3/13/2024, 6:00-8:30pm 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.

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.
- And then tell us about the data you use.
  - Which datasets 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 story should incorporate 3 plots (static or interactive, but without animation) and one R Shiny app. No more. Only with instructor's approval you may add plots.**
- 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 5 minutes Q&A for each presentation:

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

- The order of the presentations will be posted later.
- You will be questioned thoroughly about your work during class time. Be prepared to explain each of the choices you make about your work.
- 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.
- 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 until the due date about anything related to the presentation.