Final Group Project – Poster Presentation

In the final group project for the *Data Visualisation* course, you will a prepare a poster presentation where you explore, analyse, and "tell a story" on a data set of interest to you. Your goal should be to create an original project that you would be proud to show off to your friends, a potential employer, etc. Posters will be displayed in the style of a professional conference, scheduled for the last day of classes. **A proposal is due on 08-Nov-2022**.

Data from the internet

The web is a vast source of datasets on almost any subject, such as demographics, disease, economics, geography, entertainment, science, etc. Check the links at the end of this document, but you can start with Google's Dataset Search https://toolbox.google.com/datasetsearch You should choose whatever dataset you are most interested in or will have the most fun with.

Requirements

The outline below is only a suggestion. If you have a completely different idea for project that does not fit neatly with the following requirements, please come talk to me.

- The data that you analyze should be complex to collect or to clean in some way.
- You should create informative visualizations to "tell a story", to explain the data and your findings.
- You should try to use some statistical and/or ML modelling techniques.

Instructions

Download a dataset and explore it. Many of these datasets are large and will not open (well) in Excel, so you'll need to load the CSV file into R with **vroom().** Most of these datasets have categorical, or factor, variables that you can use for grouping and summarizing, and many have time components too, so you can look at trends.

Find a story in the data. Explore that story and make sure it is insightful.

Use R (preferably, as then your analysis would be reproducible), Tableau or coloured pens [1] to create multiple plots to tell a story and communicate your findings. You can make as many plots as you want, but you must use at least three different chart types (i.e. don't just make three boxplot or three maps). You may find https://www.data-to-viz.com/ a useful tool to help you choose the appropriate type of a graph for a given dataset or question.

Deliverables

1. *Your final group presentation* which you should upload on canvas
2. *A github repo*: Theis should be organized as an RStudio project, and should be added to the class table on https://github.com/kostis-christodoulou/am10.mam2023
3. Create a new public repo for your final group project and initialize it with a "readme" file.
4. Edit the readme.md file to be a short, one-page description of what your final project is about. Add the data sources and question/issue you want to address
5. Organize your work under an *R Markdown* to introduce, frame, and describe your story and findings.

---

[1] I am not kidding. Look at some of the great work by Mona Chalabi, someone who used to work at fivethirtyeight.com https://www.instagram.com/monachalabi/ https://monachalabi.com/about-2/

6. Commit and push your work as you go along; the more commits, the merrier.
7. At the end of your project, I should be able to clone the repo locally and reproeduce all your work, visualisations, etc.

What should the presentation include?

In the pre-covid era, I use to run the final session as a poster presentation; the poster itself could be anything from 8—10 slides printed up and glued to a bristol board to an artistic rendition of the main research findings. "Presentation" is a misnomer. There is no time for the presenter to launch into a monologue; instead, there is a dialogue between the presenter and the viewer "What is your research about?", "What does this show?" etc.

Your poster/presentation should include the following components:

(1) A statement of the question or purpose. What problems or questions did you set out to analyse? What were the key issues raised?

(2) The background and preparation for conducting the project. How did you prepare for the project? What sources or background readings did you consult? What information did you use in developing your ideas from the conceptual stage to the finished project?

(3) Methodology. What did you do, and how did you do it? What statistical techniques did you use — for instance, scatterplots, correlation, confidence intervals, linear/logistic regression?

(4) Results and conclusions. This is where you summarise and present your data analyses and communicate your main results. What did you find out? This might include tables, graphs, or verbal summaries. What did you learn about the problem or question you set out to investigate?

(5) Discussion and critique. What did you learn about the process of carrying out your project? What went wrong, and how could you improve it next time? For instance, did any sources of bias creep into your survey or experiment? What advice would you give future students?

How will the project be graded? All group members receive the same grade; the project counts for 50% of your overall course grade and there is no final exam on this course. You will be graded on the attached rubric.

AM10 Data Visualisation and Story Telling
Autumn 2020

Final Group Assignment		Group 1			
Technical issues	**Excellent**	**Good**	**Needs work**	**Comments**	**Points**
20 points	20 19 18 17 16 15	14 13 12 11 10 9	8 7 6 5 4 3		
	You included all your code and output in an R Markdown document. Multiple commits in your github repo, from all group members. I can clone your repo and knit your Rmd and the output does not include warnings or messages. The code is well commented and easy to read/understand	You included your code and output in an R Markdown document, but the output includes warnings and package loading messages. Your code has a few comments and its readability is ok. Few commits on Github	You did not include your code in an R Markdown document. There is no way for someone to reproduce your results. The Github repo is essentially empty		
Visual Design	**Excellent**	**Good**	**Needs work**		
60 points	60 55 50 45	40 35 30 25	20 15 10 5		
	Your visualisation expertly demonstrates the best practices of visual design, including contrast, repetition, alignment, proximity, color, typeface choice, and is generally easy for your audience to understand. Great use of colour and proximity/similarity to focus attention. Used consistent colour scheme and different types of plots in each panel.	Your visualisation generally demonstrates best practices of visual design. There may be some elements of the visualisation that may be confusing to your audience. Some aesthetics could be simplified and your work would gain from use of contrast/highlight	Your visualisation generally lacks best practices of visual design. Elements of the visualisation are confusing to your intended audience.		
Truth and Beauty	**Excellent**	**Good**	**Needs work**		
60 points	60 55 50 45	40 35 30 25	20 15 10 5		
	Your visualisation expertly demonstrates a solid understanding of Alberto Cairo's five qualities of great visualisations and is truthful, functional, beautiful, insightful, and enlightening	Your visualisation generally demonstrates an okay understanding of the five qualities of great visualisations and is lacking in some aspects.	Your visualisation generally lacks any of the five qualities of great visualisations		
Story	**Excellent**	**Good**	**Needs work**		
60 points	60 55 50 45	40 35 30 25	20 15 10 5		
	Your visualisation explores a deeper story. You provide specific explanations of your design choices and demonstrate a deep understanding of the needs of your audience. You 'personalise' your story by seeing how it affects hypothetical people/context. You combine plots to build a cohesive narrative, building from simpler to more complex plots.	Your visualisation explores a slightly more complicated story. Your explanations of design choices are cursory and demonstrate an okay understanding of the needs of your audience. With more time and effort, this could be an excellent work	Your visualisation does not tell a compelling story. Your description offers only surface-level insights into your design choices and audience		
TOTAL					
200 points					

Deadlines:

8 Nov 2022. What is your topic? What issues or questions are you addressing? What is the source of the data you will be using? What statistical techniques do you think you may be using? Be as specific as possible; your github repo readme.md should be about a couple of paragraphs.

Final Course Day*:* Project day: posters/presentations due. No late sumbissions accepted. You should upload your presentation through Canvas and all your coding work should be pushed on to the github repo, so everyone can see your work.

A final word: I encourage innovation, creativity, and humor in your projects. I want you to have fun with this assignment, and to create a piece of work that you will be proud of. As always, I am happy to give feedback and help along the way— just not a few hours before the deadline! Please don't hesitate to email and/or come and get help. My goal is for you to learn and have a project you'll want to show off to all your friends.

Data Sources:

Google Dataset Search: Google indexes thousands of public datasets and you can search for them here
https://toolbox.google.com/datasetsearch

TidyTuesday datasets: https://github.com/rfordatascience/tidytuesday

Data is plural: a weekly newsletter of useful/curious datasets https://tinyletter.com/data-is-plural

You may also want to have a look at

https://mam2023.netlify.app/reference/world_bank_data/

https://mam2023.netlify.app/reference/eurostat_data/

https://mam2023.netlify.app/reference/other_data_sources/