HOMEWORK 4: MODEL VISUALIZATION PORTFOLIO PROJECT

Instructions

1. Upload your source code as either a .Rmd or .R file on Canvas.
2. Upload your write-up as a PDF or HTML document on Canvas.
3. Simultaneously attach your PDF or HTML within the portfolio folder on Piazza. This is how submissions will be shared for peer reviews.
4. Write-ups should not exceed 750 words.

Description

The portfolio exercises are opportunities for you to design and share visualization without the constraints of in-class exercises. The third portfolio assignment focuses on dimensionality reduction or model visualization. You may prepare either (i) a single interactive visualization or (ii) two static visualizations related to a topic that is interesting to you. Ensure that high-dimensional or model-based elements are clearly encoded in your final views. Accompany your visualization(s) with a discussion of how your proposed views inform your understanding of the broader problem. You may continue to use the same datasets as previous assignment, or you may choose a new dataset. As before, you may choose data from public sites (TidyTuesdays, Data is Plural, Kaggle Datasets, Google Dataset Search, Data.Gov, Madison Open Data, Bioconductor Datasets, Awesome Public Data), or you may generate and download data about your own life, taken from an app that you use regularly.

Code implementing your visualization should be modular and readable. For example, variables and functions should be named clearly, code duplication should be minimized, and commented code should be removed. To support reproducibility, ensure that the uploaded .R or .Rmd files can be run by directly sourcing or knitting the document. No directory structures should be assumed, and any required data should be downloaded from the web.

For each of your visualizations (or your single interactive one), prepare a discussion of the visualization and the process used to create it. Please address the following components,
- What is the essential question that your visualization is supposed to inform?
- How do aspects of your design support exploration of the essential question? Were there trade-offs you had to make so that certain features were more clearly visible?
- What are your key findings? How do they relate to your prior understanding?
- How did you create the visualizations? Were there any data preparation steps?

At the end of the course, I will ask you to choose your favorite submission from the portfolio exercises to include within a publicly visible end-of-course book / website. You will have a chance to revise your submission based on peer reviews before it is included in these. Examples from previous years' submissions can be found here.

Rubric

Discussion Quality [5 pts]: The write-up is precise, well-developed, and engagingly written. Paragraphs and / or headers are used to organize the text, and superfluous code outputs are suppressed.
Design Choices [6 points]: The visual interface is easy to use, appropriately annotated, and supports meaningful dynamic queries. Data are not unnecessarily summarized, and the views

have high information density. The design does not rely on visualization defaults and demonstrates attention-to-detail. Though it may build from or synthesize course examples, the submission demonstrates independent and creative visual design thinking.

Problem Formulation [5 points]: The focus of the application is on a broader, independently interesting problem domain. All data are reported within context, rather than assuming prior familiarity (with specific variable names or data collection methods, for example). The questions asked do not have obvious answers, and the visualization could potentially find an audience beyond the course.

Code Useability [4 points]: The code to generate the figures is readable and can be run easily.