**HOMEWORK 5: MODEL VISUALIZATION PORTFOLIO PROJECT**

**Instructions**

This homework consists of a portfolio visualization (or two) and an associated write-up.

You may prepare EITHER:

* A single interactive visualization, or
* Two static, non-interactive visualizations

Your single interactive visualization OR both of your two non-interactive visualizations must visualize the results of a clustering algorithm, dimensionality-reduction algorithm, topic model, prediction model, Bayesian model, OR deep learning model. For your write-up, please upload **a separate, well-formatted PDF or Word or other text document.** 750 words maximum. Do not include your write-up in your R code file!

**If You Choose Two Static Visualizations**

Please submit:

* **Your code file(s), such as .Rmd or .R file(s)**
* **Your write-up**, and **please include your static visualizations directly in your write-up document.** Please ensure you are satisfied with how the visualizations look in your write-up (e.g. are labels large enough?) since that is what we’ll grade.

**If You Choose to Submit a Shiny App**

If you choose to submit a Shiny app, you have a few choices as usual to demonstrate its functionality. If you choose a series of screenshots or link to a shinyapps.io site, please include that directly in your write-up. If you choose a screen recording, submitting that separately is acceptable.

Please submit:

* **Your code file(s), such as .Rmd or .R files that create the Shiny app,**
* **Your write-up**; if you choose to provide a series of screenshots of your app working or a link to shinyapps.io, that should go directly in this document!
* If applicable, **a screen recording of your app working**; only needed if you don’t have screenshots/shinyapps.io link in your write-up.

You do **not need to submit any data files.**

**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, AWS Data Registry, Wisconsin DNR, IPUMS Census Data, 538, Data.gov, Madison Open Data, Bioconductor Datasets, Awesome Public Data). You may start from the same data you used in a previous submission.

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?

**Rubric**

**Visualization [12 points]**

Design Choices [8 points]: The visual interface should support meaningful interactive queries and be appropriately annotated. Your design choices should go above and beyond defaults, and show critical thinking as well as attention to detail towards aiding the viewer in understanding the takeaway(s) of the visualization. Strive to make your graphs as polished as possible; especially with choices such as layout, labels, sizes, themes, and colors. Though it may build from course examples, the submission demonstrates independent and creative thinking.

Code Useability [2 points]: Techniques to improve the conciseness and readability of code are used, such as using reactive expressions to avoid unnecessary duplication, and extracting graphing or other helper functions outside of render\* commands.

Problem Formulation [2 points]: The context of the visualization is appropriately communicated, and all data are reported within context, rather than assuming prior familiarity with specific variable names or data collection methods, for example. (*You can still use field-specific terminology if your intended audience would know what it means, but any such terms should be explained in your write-up.*) The problem or question the visualization investigates should not have an obvious answer, and the visualization could have an audience beyond the course.

**Write-up [8 points]**

Discussion Quality [6 points]: The write-up is thorough with respect to each prompt but not overly wordy and avoids technical jargon. Writing demonstrates critical thinking about the author’s own workflows, decisions, and what information needs to be communicated to the viewer.

Discussion Formatting [2 points]: The write-up is free of grammatical errors and logically organizes the text into clear sections (e.g., with headers or other formatting devices).