OVERVIEW

This assignment provides you the opportunity to synthesize all of the concepts we’ve covered in the course to date. The basic framework is that you will create a COMPLETE data visualization BLOG post that is suitable as a showcase component of your Data Science Portfolio. The point is to **SHOW** people your skills.

STRUCTURE

The basic formatting guidelines for this assignment are:

1. Include code fold or code tools options (or both) that allow users to view and copy your code while maintaining overall readability of your post.
2. Suppress all output and warnings that might distract from your visualizations and writing.
3. Properly title your assignment. The main title should be **“BCB 520 - Midterm Portfolio Post”**, and the subtitle should be a descriptive title related to your question or topic.
4. Include author, date, categories, and a description in your YAML header.
5. Write clear, complete sentences for a target audience with some scientific background but little training in your specific discipline.
6. Include references if appropriate and use hyperlinks to external sources of data, inspiration, or examples.
7. Use the header hierarchy and create a sensible document outline with white space. Format for readability! Use **bold** and *italic* fonts to emphasize things! Use **color** by customizing your .css file!

**In addition to the above formatting guidelines, your portfolio post must contain the following sections:**

Preamble

Write a brief paragraph describing the primary question or purpose of the post. Ideally, the concept should be challenging enough that it requires at least two visualizations that use different idioms (ie. don’t just make two scatterplots with different variables). The concept should also be challenging enough that it captures the interest of the reader (i.e. a plot of height and weight that shows they are correlated is trivial and not appropriate). The best approach is to explore a topic or question in which YOU are very interested.

Data

Write a summary of the data sources you will use. Include a Data Dictionary table that fully describes each individual data file used. You may use your own research data or publicly available data from any source you like (with attribution). There aren’t any minimum or maximum data set size requirments, other than you need something big enough to be interesting and not so big that we don’t have a supercomputer capable of creating your visualization.

Visualizations

Create your visualizations and include text that explains any steps or design choices. Be sure to include clearly labeled axes and a concise but complete figure caption for each visualization. Make deliberate choices for color palettes, point marks, line types, etc. Demonstrate that you understand the concepts we have covered!

Conclusions or Summary

Answer your research question. Draw a conclusion or inference related to your topic. Summarize your results. What new questions have emerged as a result of your visualizations? What interesting next steps have emerged?

RUBRIC

I will evaluate the following for your portfolio post:

**1. Clarity of writing (15%):** Complete, clear sentences. Good Grammar. Understandable to target audience. Logical flow of ideas.

**2. Adherence to format (10%):** Did you follow directions?

**3. Topic suitability (15%):** Is the topic interesting? Are the visualizations challenging and interesting enough to showcase your skills?

**4. Viz Execution (40%):** Are the visualizations effective? Do they adhere to the principles of effectiveness? Are choices for idiom, marks, channels, etc made deliberately and well justified?

**5. Creativity (20%):** Did you push your boundaries and learn new techniques? Is the overall post compelling and interesting? Are the visualizations inspiring, creative, unique, and generally impressive? If I were recruiting a new data scientist (and I often am), would this portfolio post impress me, or would it damage your candidacy during review?