Assignment 3

The purpose of this assignment is two-fold: (a) to have you engaged in independent learning with respect to a novel visual exploration method, a novel python-based information visualization toolkit, or both, and (b) to enhance our knowledge overall in the class (including mine!) by engaging in peer instruction. More specifically, I want you to write a tutorial demonstration which shows a new visualization technique (e.g. pareto charts) and/or a new visualization toolkit (e.g. bokeh, seaborn, plot.ly). Feel free to do both a new technique and a new toolkit.

The purpose of using data visualization is to tell a story, and the purpose of writing up a demonstration is to teach someone else about what you have learned so that they can use it, too. Think of this as a worked example, something like a medium blog post that you have written up and are posting so that other data scientists can learn about the potential uses of this style of data exploration and visualization. You should be concise in your wording and clear in your demonstration.

You will need to include the write up and demonstrate in a new notebook which you will be submitting to the staff. Your notebook should be primarily narrative in nature, broken up by code cells that demonstrate the work that you're demonstrating. You should have the following main sections in your notebook (but feel free to use other subheadings as appropriate):

- Visualization Technique (25%)
    - A narrative description of the visualization you are planning to use, describing how it works
    - A discussion of in which circumstances this visualization should and should not be used (what is it close to? What else could you consider? How does it relate to specific aspects of data?
- Visualization Library (25%)
    - The library you are going to use, and a background on why the library is good for this visualization. For instance, who created it? Is it open source? How do you install it?
    - A discussion of the general approach and limitations of this library. For instance, Is it declarative or procedural? Does it integrate with Jupyter? Why did you decide to use this library (especially if there are other options)?
- Demonstration (50%)
    - The dataset you picked and instructions for cleaning the dataset. You should pick a suitable dataset to demonstrate the technique, toolkit, and problem you are facing.
    - The quality of your demonstration. First demonstrate the basics of this approach, then show a few of the edges of how the library might be used for other cases. This is the "meat" of the assignment.

Assignments should be submitted as both your source code as an .zip file of your notebook (.ipynb) and other assets (e.g. .csv) as well as a .pdf of this notebook which we can view in a browser.

Frequently Asked Questions

1. Q. This sounds daunting, can I have an example?
   A. Yes! Anthony Giove, a student from the Fall 2019 cohort, has provided his example for you in the

example_assignment directory in your workspace. This example is a stellar one, and is above and beyond what you need to do for this assignment. In particular, you don't need to find such an extensive set of data, nor do you have to go all the way into advanced features such as interactivity. Yet, I think many of you will find this motivating and I look forward to seeing what you came up with!

2. Q. This assignment is daunting! How big does it have to be?
   A. Think a couple of pages, one compelling visual, or a couple of less compelling visuals. Make sure the question you ask is well aligned with the technique/library you are using.
3. Q. So the visual and technique is the important part, where the insights are less so?
   A. Yes, this is a good interpretation. I care less about the outcome and more about your learning of a technique or library.
4. Q. Can I use tableau?
   A. No; it needs to be a python toolkit.
5. Q. What do you mean by declarative or procedural?
   A. Conceptually, Declarative programming is where you say what you want without having to say how to do it. With procedural programming, you have to specify exact steps to get the result.
   A more full discussion of declarative or procedural should be included in the lectures, however the assignment has been updated (see changes above) to clarify that these were to be example topics you might cover for each of the prompts given.
6. Q. Is a new visualization technique absolutely necessary? Could I use a new visualization package while using the visualization technique learned during the lecture?
   A. Yes! This assignment requires you to write a tutorial which shows a new visualization technique and/or a new visualization toolkit or package. This means you can use an old visualization technique while using a new visualization toolkit.
7. Q. Can I have more examples about what visualization package we can use for assignment 3?
   A. Yes! Apart from the packages mentioned above, we can use plotly, bokeh, dataviz, vega, altair, folium, geopandas, earthpy, etc. Those are wonderful packages to explore.
8. Q. I have some data that I want to explore. How shall I know which visualization techniques and visualization toolkits are suitable for such analysis?
   A. While the answer varies depending on the nature of your dataset and your visualization approach, a good way to research your problem is through looking at the gallery page for a couple of different visualization toolkits. If you look closely you will find at least some useful graphs to represent your data to the audience.
9. Q. Do we need to provide references for the narrative information relevant to the visualization technique or the visualization library? If yes, what citation style is recommended?
   A. A URL or APA citation style would be fine.