A4

Start Assignment

Due Friday by 11:59pm   **Points** 40   **Submitting** a file upload

DS4200 Assignment A4

Total points **(40 points)**

Assignment due: **Due Friday 2/17 (Feb 17) @ 11:59 pm**

Notes:

- *This is an **individual** assignment.*
- *Make sure to read the full assignment before you begin (in most cases the most substantial part of the assignment is the last part!).*

Grading Notes:

- *You will be graded on the quality of your writing in all assignments. Points will be deducted for grammar and spelling mistakes.*
- *Points will be deducted for not following the submission instructions.*

Part 1: Visualization Categorizations & Graphs in Python (10 points)

In this part of the homework, we will see different ways to sort and categorize visualization (e.g., by marks, by data, by task, etc.).

Instructions:

First, go explore the following two collections of visuals that one can make in Python and Altair:

- **https://www.python-graph-gallery.com/** ↪ **(https://www.python-graph-gallery.com/)**

- ⤷ **(https://altair-viz.github.io/gallery/index.html) https://altair-viz.github.io/gallery/index.html** ⤷ **(https://altair-viz.github.io/gallery/index.html)**

1. What is a new type of visualization you learned about on one of the two gallery pages (or, if you are already familiar with all of the visuals, what is the most interesting one to you)? *[2 points]*

Pick one plot type presented on **https://www.python-graph-gallery.com/** ⤷ **(https://www.python-graph-gallery.com/)** that we have not covered already during our in-class Python tutorials and implement it into a Python notebook. As part of your notebook, include a text box that explains why you chose this visual encoding (i.e., why is it interesting to you). *[8 points]*

Part 2: Introduction to Altair (4 points - Homework)

Browse the examples gallery for Altair:

https://altair-viz.github.io/gallery/index.html **(https://altair-viz.github.io/gallery/index.html)**

Please address the following in your submission,

1. Which is your favorite example in the gallery, and why? Provide a link to your favorite with a screenshot, and a couple sentences stating why it is your favorite. *(1 point)*
2. Although the "Case Studies" are really interesting, the code is poorly commented thus at times challenging to understand the functionality of the code. Pick **one** of the following case studies, copy-and-paste the code, and add comments to explain how the code is generating the visualization *(3 points)*:

"Locations of US Airports" (**https://altair-viz.github.io/gallery/airports.html** **(https://altair-viz.github.io/gallery/airports.html)**)

"Natural Disasters" (**https://altair-viz.github.io/gallery/natural_disasters.html** **(https://altair-viz.github.io/gallery/natural_disasters.html)**)

"Seattle Weather Heatmap" (**https://altair-viz.github.io/gallery/weather_heatmap.html** **(https://altair-viz.github.io/gallery/weather_heatmap.html)**)

Part 3: Visualizing Boston Pollution Data (26 point - Homework)

(7 points per graph (5 for graph, 2 for write-up), and 4 points for paragraph summarizing the paper.)

In this final part of the assignment, you will build three Altair visualizations in a Jupyter notebook. We will be looking at data relating to Boston Harbor pollution and creating static visualizations to explore trends.

INSTRUCTIONS:

1. First, watch this short video about the Boston Harbor Clean-up project: "
   Boston Harbor Cleanup ⇲ **(https://youtu.be/ybmoVbO0omk)**

   (https://youtu.be/ybmoVbO0omk)
   "

2. Next, read the following journal article:

"Evaluating Boston Harbor Cleanup: An Ecosystem Valuation Approach" ⇲
(https://www.frontiersin.org/articles/10.3389/fmars.2018.00478/full)

Note: you do not have to understand every detail of the paper, but must understand enough to appreciate and interpret the figures/tables.

1. Summarize the article in one paragraph in "common" everyday language the public can appreciate. In your summary make sure to paraphrase a thesis statement for the journal article.
2. Next, re-create the following three tables from the article as grouped bar charts with Altair: **Table 3, Table 5,** and **Table 6**. Each of these tables, in my opinion, are challenging to appreciate and execute comparison tasks. Recall that grouped bar charts are an example of superimposed layers and the advantages of this method include saving screen real estate and being easier to compare subtle value differences (see Lecture 12). The alternative, which we are not doing in this assignment, would have been to make small multiples to juxtapose each condition as an individual bar chart. For simplicity, we have created three datasets for you to use in this assignment: **table3.xlsx (https://northeastern.instructure.com/courses/141299/files/folder/Datasets)** , **table5.xlsx (https://northeastern.instructure.com/courses/141299/files/folder/Datasets)** , and **table6.xlsx (https://northeastern.instructure.com/courses/141299/files/folder/Datasets)** .
3. For each new visualization created in #4 (i.e., new grouped bar chart), make sure to include a title, clearly labeled axes and a color legend if necessary. Include a **written summary** for each chart

that both states what the plot is showing as well as what the key insight/take-away is from the graph. (In my opinion this is a weakness of the original article's figure captions as they do not state anything insightful.)

Submission Instructions: Include in your PDF submission for Assignment A4 a link to your github repo. Make sure your repo is shared with the teaching staff (see Canvas for a list of teaching staff usernames).

Submission Instructions: *Include in your PDF submission for Assignment A4 a link to your GitHub repo where notebooks for Part 1, 2 and 3 should be. Make sure your repo is shared with the teaching staff (see Canvas for a list of teaching staff usernames).*