SI649 Scientific Visualization Project, Winter 2024

The second project in this course is about replication, redesign, and interaction. In it, you will be asked to choose an existing static visualization, and then design and implement an interactive version of it that effectively supports better understanding of the underlying data.

The background for this project is that many scientific papers make effective use of information visualization. However, for various reasons, the vast majority of these visualizations are only presented as static figures in PDFs, rather than interactive plots on the web. As such, we will ask you to use figures from scientific papers as your starting point (which we will provide examples of), and get you to imagine how you might extend or redesign those figures to better support user understanding of the data via interaction (and then implement those designs).

Due date: Sunday March 31st, 2024 at 11:59pm (Eastern time)

Group Size: This project can be done either **individually or in pairs**. However, if you work with a partner, you will not be allowed to work with the same person for the final project! If you would like to be randomly paired up, let me know ASAP and I will connect you to others.

1. Overview

For this project, we are providing pointers to several scientific papers (or unpublished reports) which include the necessary code and data to replicate their figures. Your job is to choose one or two of these figures, and then redesign and implement them as an interactive visualizations.

The overall guiding principle behind your design should be that an interactive figure can better allow the reader to understand the underlying data. Most static figures are not able to fully represent the full complexity of the data, but interactivity can dramatically enhance the potential of a figure to help the user understand and ask questions of the data. Imagine if the interactive figure could be placed back in its original context; what additional insight would the interactive version provide?

You are welcome to submit implementations for up to two figures; if you submit two, we will look at both and grade you on what we deem to be the best one. It is up to you to decide which figure to focus on. When doing so, you should be thinking about what figure (and underlying data) could most benefit from being made interactive, and what kinds of interactivity would allow for the deepest understanding by the reader.

2. Deliverables

1. Final visualization(s):
   - You can submit up to two interactive visualizations, which you should host somewhere online (e.g, Tableau public, Streamlit, HuggingFace, etc.). As part of your report, please provide **clear pointers** to where these interactive visualizations can be found (e.g., clearly marked URLs), and double check that these work as intended, even when you are not logged in (e.g., using a browser in incognito mode). Please also include static screenshots of your final visualization(s) in your report.

2. Implementation:
   - In addition to the online version of your visualization(s), please also submit your underlying implementation files, either in the form of code (e.g., Altair, etc.), a Tableau notebook, or similar.

3. Final report:
   - As with the first project, you should write a short informal report (probably around 5-6 pages of text in total), describing your process, due on the same day as the final visualizations. *For each of your visualizations*, your report should cover:
     - **Background**: What is the source of the original figure(s) you chose to redesign? What data was it showing, and what tasks did the original figure facilitate? What was it intended to communicate, in the context of that scientific paper?
     - **Tasks**: What comparisons is your redesigned visualization intended to support? What will the viewer learn and take away from it? Remember that a good visualization should be designed to effectively support specific domain tasks, and can typically support multiple tasks, not just one.
     - **Your (re)design process**: What did you get from the original (static) visualization? In what ways did you think about enhancing it? What did you try? (feel free to include sketches or screenshots). How does your design enable greater interaction with and understanding of the underlying data?
     - **Challenges:** Where did you run into the most difficulties in implementing your design? What sort of limitations did you encounter when trying to implement interactivity? Was there anything you wanted to support that you couldn't figure out how to do?
     - **Qualitative self-evaluation**: How would you judge your own design and creation? In what ways is it effective? How could it be improved? Please connect this to principles that you have learned in class.

3. Evaluation Criteria

For the final submission, we will use the following evaluation criteria:

1. **Relevance**: Are your visualizations relevant to the topic of the scientific paper? Would they fit naturally as part of the article if posted online?
2. **Insights**: How much does your visualization add to the reader's understanding, beyond the static version of the figure?
3. **Clarity, effectiveness, and functionality**: How effective are your visualizations for the tasks you have selected? How well does your interactivity function?
4. **Aesthetics**: Have you used good design principles and considered human perception and cognitive limitations in designing your visualizations?
5. **Creativity and effort**: How interesting or ambitious are your visualizations? Do they reflect a considerable amount of effort?
6. **Writing**: how well written, detailed, and reflective is the report?

4. Starting Points

To help get you started, we are providing the following pointers to scientific papers that have released the code and data necessary to replicate some or al of their figures. We suggest you look at these, and choose one or two as the basis of your work. **Note that some of these may require some effort to get working, such as installing relevant python packages, or even slightly modifying the code that has been provided.** Part of the point of this project is recognizing that replicating scientific results involves numerous hurdles, and using someone else's code provides advantages but also additional challenges.

Here are the starting points, in no particular order:

1. Bayesball: Bayesian Integration in Professional Baseball Batters
   - Paper: **https://www.biorxiv.org/content/10.1101/2022.10.12.511934v1.full.pdf**
   - Code and data: **https://github.com/KordingLab/Bayesball**
2. A Turing test of whether AI chatbots are behaviorally similar to humans
   - Paper: **https://www.pnas.org/doi/10.1073/pnas.2313925121**
   - Code and data: **https://github.com/yutxie/ChatGPT-Behavioral**

3. A public data set of spatio-temporal match events in soccer competitions
    - Paper: **https://www.nature.com/articles/s41597-019-0247-7**
    - Code: **https://figshare.com/articles/code/ Plots_replication_code_of_Nature_Scientific_Data_paper/11473365**
    - Data: **https://figshare.com/collections/Soccer_match_event_dataset/4415000/5**
4. SPARCS, a platform for genome-scale CRISPR screening for spatial cellular phenotypes
    - Paper: **https://www.biorxiv.org/content/10.1101/2023.06.01.542416v1**
    - Code and data: https://github.com/MannLabs/SPARCS_pub_figures
5. Visualization of the Challenges and Limitations of the Long-Term Sunspot Number Record
    - Paper: **https://arxiv.org/abs/2203.11919**
    - Code and data: **https://github.com/amunozj/NatAs_SN_Perspective**
6. Influence of cosmic-ray variability on the monsoon rainfall and temperature (plus a reply):
    - Original paper: **https://www.sciencedirect.com/science/article/pii/ S1364682614002697**
    - Reply paper: **https://arxiv.org/abs/1502.00505**
    - Data and code: **https://figshare.com/articles/dataset/ Comment_on_Badruddin_amp_Aslam_2014_/1299413**
7. Combien de pages pour une thèse ou un mémoire? [in French]
    - Blog post: **https://jhroy.ca/2016/11/07/combien-de-pages-pour-une-these-ou-un-memoire/**
    - Code and data: **https://github.com/jhroy/theses/tree/master**

If you would like to propose an alternative starting point for your project, you are welcome to do so. However, this must be done before **Monday March 11th**, and is subject to approval by the instructor. The general criteria are that it should be a static figure in a published scientific paper that does not already have an existing interactive implementation. You are welcome to use a figure from a paper that you yourself have written, as long as it meets the above criteria.

5. Advice

1. Start by looking at the PDFs of the associate figures, and try think about what figures you might want to re-implement, and how they could benefit from being made interactive. Using the principles we will cover in class, think about various types of interactivity that you could incorporate, and how those might help the reader to better understand the data.

2. Once you have found a few potential figures, download the associated data and code, and try to get a sense of how easy or hard it would be to a) get the code running; and b) reimplement them as interactive figures. You should also think about what data underlies the figure, and whether it supports adding the kinds of interactivity you had in mind.

3. Once you have sketched out an idea of what you want to make, start by reimplementing the figure in a package that supports interactivity (e.g., Altair or Tableau) and make sure you are able to first replicate the static version of the figure.

4. Once you have the static version finished (or close to finished), start working on the interactive parts of it. If time permits, get someone to test out your design, and use their feedback to improve it.

5. Once you have something working locally, figure out how you are going to host it online, and make sure that works, in advance of the deadline

6. Remember that if you submit two figures, we will only grade the best one. So, if you are short on time, submitting one design is sufficient.