CSE 5544 Lab 3 (15/60 pts):

Ethics in Data Visualization

Assigned 2/21, Due 3/7 8:59pm

Data Science and Computer Science students have the same assignment.

Learning Goal

It is tempting to think of data and data visualization as a neutral. An emphasis on a minimalist aesthetic — particularly through the use of clean, precise geometric lines — seems to make visualizations objective, transparent forms of communication that masks visualization's persuasive power. Given the growing ubiquity of visualization as a medium for recording, analyzing, and communicating data, we have a responsibility to examine how our design choices can influence the way a visualization is read, and what insights a reader walks away with.

In this assignment, we will grapple with these ethical concerns by visualizing a single dataset from two different perspectives:

- P1: "honest/ethical/truthful" and
- P2: "dishonest/unethical/deceiving"

For this assignment, we will consider a visualization from Perspective 1 to be one where:

- The visualization is clear and easy to interpret for the intended audience (often parts of the general population);
- Any data transformations (e.g., filters, additional computations, etc.) are clearly and transparently communicated; and
- The sources of the data, including potential bias, is communicated.

A visualization from Perspective 2, on the other hand, exhibits one or several of the following characteristics:

- The visual representation is intentionally inappropriate, overly complex and/or too cluttered for the audience;
- Labels, axes, and legends are misleading;
- Titles are skewed to intentionally influence the viewer's perception;
- The data has been transformed, filtered, or processed in an intentionally misleading way; or
- The source and provenance of the data is not clear to the viewer.
- For the intended questions, examine the visualizations to make sure
    - It includes a view of ALL data (overall); before a filter is applied (to produce partial or item-based display). In this way, viewers can see the context.

We might never imagine ourselves to be (nor aspire to be) to fulfill all categories in the "ethical/honest/truthful" (P1) category so we are going to temporarily ignore that we can't be perfect in order to better appreciate the extent of the rhetorical force of visualization, and build our critical reading skills.

Step 1: Choose Dataset

You will be working with a the same eye-tracking dataset from Lab 2 **or** one of two you can find below. These datasets are intentionally chosen to cover either politically charged topics or trust in science, for the simple reason that these are typically the type of data where ethical visualization is important. Note that you do not have to visualize the entire data from the chosen dataset (i.e., you may choose a subset of the data to visualize).

The two additional datasets are the following:

- The DEA Pain Pills Database. The Washington Post has published a significant portion of a database maintained by the Drug Enforcement Administration (DEA) that tracks every opioid from their manufacturer, through to distributors, and into pharmacies in towns and cities across the United States. This is an enormous database, and you can choose to work with it at any level of detail (e.g., state-wide, individual counties, or national summaries). You need to create a free account at the Washington Post to access this dataset.
- VIS30K dataset; ImageNavigator. This is also part of Rui's dissertation work! We have collected all images, including table, figures, captions, text, title, authors etc from the IEEE VIS publications. You can choose the show for example color use, image features (by mining textures), co-authorships, topic-progress overtime, e.g., using this result from Prof. Nelson Beebe.

Step 2: Visualizations

You will be visualizing your dataset from two perspectives: P1 and P2. As a result, you will be generating two static visualizations – one for each perspective. We construe "visualization" broadly (e.g., a single visualization may comprise several small multiple visualization representations).

(8 points) Deploy a web-application, in either D3 (CS students) **or** Streamlit/Python (Non-CS students) **or** your favorite tools (Tableau - all if you wish)

- To generate a web-based matrix visualization (also called heatmap) using two different color choices (rainbow and Extended Blackbody) **or**
- To provide an overview of the data.
    - Make sure that your Visualization includes a descriptive title, potentially subtitle and labels. You have the freedom to place annotations and labels to help best convey the message from a particular perspective,

      or
- To refine **one** of your Lab2 visualizations to improve the design. How to choose one? - choose the one you think that is most ethical from many design solutions you have and then iteratively refine it, **or**
- To combine two above. For example, if you have used heatmap in Lab2, try different colormaps available in D3 (https://github.com/d3/d3-scale-chromatic) or Python (https://matplotlib.org/stable/tutorials/colors/colormaps.html).

Note: you can use python or any of your favoriate tools for data processing. D3 or Python or Tableau can just be used for all or for your final stage of showing the data / insights from data.

(7 points) Describe your rationale in a short write-up (no more than 4-5 sentences per visualization) to justify why your current design either new or refined is better.

Extra credits

For 10% extra points, design a new visualization of some kinds (see below). Describe your rationale in a short write-up (no more than 4-5 sentences per visualization) to justify why your design is better for what tasks. You can try

one more from your Lab2 as well to earn this extra credits. Make sure that your Visualization includes a descriptive title, potentially subtitle and labels.

Rubric:

This part of the assignment will be scored using the following rubric. Note, rubric cells may not map exactly to specific point scores.

Component	Perspective	Excellent (+20%)	Satisfactory (100%)	Poor
Color Encodings	P1/2	Subtle colormap choices (such as binning the data or non-linear mapping) require close and careful reading to identify features.	Direct mapping from data range to color range.	Incorrectly mapped data to colors.
Data Transformation	P1/2	More advanced transformation (e.g., sorting, filtering, grouping) were used to extend the dataset in interesting or useful ways. This can be done as a preprocessing or at the interactive exploration stage.	Simple direct view were primarily used by year.	
Titles & Labels	P1	Titles and labels helpfully describe and contextualize the visualization.	Most necessary titles and labels are present, but they could provide more context.	Many titles or labels are missing, or do not provide human-understan dable information.
Titles & Labels	P2	Titles and labels subtly skew reading the visualization.	Titles and labels leave out important information, but an astute reader	Titles and labels are largely present, visible, and facilitate reading the visualization.
Write-up	Both	Your writeup is well-crafted and provides reasoned justification for new design choices	Design correct mapping of rainbow and heatmap.	Missing or incomplete. Several design choices are left unexplained.
Creativity & Originality		You exceeded the parameters of the assignment, with original insights or particularly creative visualizations or	You met all the parameters of the assignment.	You met most of the parameters of the assignment.

		transformations.		

Submission

Follow the same naming convention as your Lab2 to name your files properly.
Submit source code and report

Help:

<I cannot think of anything. Please let me know if I can add anything here.>