CMSE 830 Midterms
Author: Hongzhi Wen - Predicting Gene Expression from Chromatin Openness: Background and Challenges

1. The first two figures focused on missingness or zero values, which were considered missing for this dataset. Hongzhi visualized the distribution of data values and found that the data is high dimensional and sparse, but dominated with unexpected zeros. He also used a umap algorithm to project high-dimensional features into 2D to demonstrate a batch effect of additional variables involved. The project also utilized SVD for feature reduction and feature normalization.

2. This project sought to measure openness of cromatin and predict gene expression in cells using multivariate regression. I believe this app would likely be used by geneticists to predict targeted genetic expression. The app did work completely without error using defaults, but after selecting other options from the drop down, it did not work as well. This could have been contained with conditional statements or try-except statements. However, I will add that the app was not completely clear in use case after the first paragraph, it was difficult to follow the story. I'm not sure if this should be attributed to a lack of background in the subject material, or to the web app itself, as I am not the target audience.

3. The provided explanation was heavily technical, and without discussion, I would have struggled to navigate this web app. Without acknowledgement of errors in this question, the narrative is still difficult to navigate as the goal and journey are not inherently clear.

4. Altair was used for his interactive visualization, it appeared on the first load but then disappeared due to a prior visualization's error and I am now unable to comment on that figure, even with a refresh. The streamlit interactive options actually made the web app less effective, as they were not contained within try-catch statements and errors were frequent.

5. It's difficult for me to answer whether the goal was achieved as I can no longer view the end of the file as I am not sure where the error stems from. I think interactivity should have been limited. By overcomplicating the application, it lost effectiveness. I also could not understand the purpose of many figures.

Author: Thejesh Mallidi - Global Crises Data by Country

1. Thejesh visualized the raw data and feature information in expandable drop downs on the web app's first tab, something I really appreciated. There were no missing values in this dataset. The only pre-processing required was the extraction of the year from the time series information.

2. The target audience spans financial analysts, bankers, investors, and others involved in the finance sector. The goal being the ability to visualize historic crises in Africa, I think this app was well accomplished and executed with clear visualizations. The app worked without error.

3. The app narrative was well-contained within the text, but there could have been more explanation behind the units of some variables. The crisis scatter plot by nation on the third tab does feel a bit meaningless but perhaps I did not find a meaningful nation to visualize. However, that would be my only complaint.

4. Thejesh used Plotly and Altair and HiPlot and many feature selection options with sliders and select boxes. On the second tab, the geographic plot also had a time series slider that could function on its own with a timestep/sleep strategy, or be filtered by the user. This was my favorite visualization.

5. For the geoplot, I would have recommended limited the frame to just the African continent, since that is the focus of the project, as well as moving the slider above the figure so it does not get missed. The index was not very clear for this figure as to what it meant overall. Appreciated the time step/sleep (auto-updating) figure. I loved the organization of this app into tabs, something I'd like to incorporate into my own.

Author: Vishal Kumar – The Weeknd Discography

1. The data, sourced from Spotify, came with no missingness. The first figure shows top hits, and the second plots popularity over time by tracking streaming.

2. There is an empty subplot on the grid of bar plots, something I may have tried to avoid. Vishal says his target audience is fans of The Weeknd or music and data nerds to track The Weeknd's popularity and music features. The app worked without observable issue.

3. This app was simple in a good way, it was extremely intuitive, and the plots were familiar. I felt that it satisfied the Tufte rules well, why overcomplicate. The context was well explained at the beginning and throughout the app.

4. All interactive visualizations appear to be created with Plotly. I appreciated the level of interactivity. Even the simplest plots allowed the user the ability to zoom or hover over data points to collect more information. The project also allowed the user to filter various features including a visualization organized by song and song features.

5. I would've picked either one correlation plot or the scatter correlation plot, not both/ For the Audio Features for Track figure, I would've liked the track to be a multi-select, to be able to compare tracks, because the figure appears sparse otherwise. However, the narrative was well executive and explained within the app, and the consistency of The Weeknd's musical artistry is very clear. My last request would have been the use of more columns or subplots to consolidate with the narrative.

Author: Max Gregg – Michigan Commuting Data

1. The data was merged with a geojson file to display map information with re-encoding of variable names to clean the columns. The estimates shown came with margins of error. This dataset has geographic features with records per geographic point. There were no missing values.

2. The target audience would be urban planners, legislators focused on commuting, economists, and automotive stakeholders in Michigan. The app worked without observable error or issue, but the second visualization did not always load. The app was meant to visualize which commute strategies were most impactful per county.

3. There was not a clear narrative in this app, and unfortunately, Max's presentation being so freeform (only open Q&A, no structured presentation) created difficulties in demonstrating a narrative impact.

4. The sidebar allowed for filtering the two plots. The app used PyDeck for the map and st.bar_chart for the first figure. There was hover-over information readily available for these subgroups.

5. The project goal was achieved, commuting options were visualized in Michigan, but the impact felt a bit undemonstrated. I would have added more informative visualizations to compare county by county with the distribution of commuter types. It felt extremely sparse. While the map was fantastic, two visualizations weren't enough to architect a full story. I also would have added hover information to the PyDeck visualization.