

Process Book - Streaming Wars

Members:

- Abbey Lasaster
 - **uID** u1326695
 - **Email** u1326695@utah.edu
 - **Github** abbey-lasater
- Kasidy
 - **uID** u0960696
 - **Email** u0960696@utah.edu
 - **Github** kasidyffernandes

Project Repository:

<https://github.com/dataviscourse2023/final-project-dimensioneers>

Background and Motivation:

In recent years, due to the growing popularity of online streaming and the shift away from traditional television, media companies and platforms have entered a highly intense market field. The motivation behind our project is to perform a competitive analysis to highlight content trends and platform performances between different streaming services.

Project Objectives:

We aim to reveal relationship patterns and use raw data to highlight trends, comparisons and insight into popular streaming services.

Questions:

- Which streaming service has had the most growth over the years? How has the revenue changed?
- Which streaming service has the most global reach?
- Which streaming service creates the most original content? What genres are the most common in their original productions?

We expect to have some typical data cleanup depending on the quality of the datasets found. This data processing and clean up can be done using Python Dataframes Library.

The following attributes will be used to build the visualizations:

- Platform (Netflix, Hulu, Amazon Prime, Disney+)
- Title
- Type (Movie, TV)
- Year of Release
- Rating (IMDb, Rotten Tomatoes)
- Age Demographic
- Genre
- Revenue

Assignment overview:

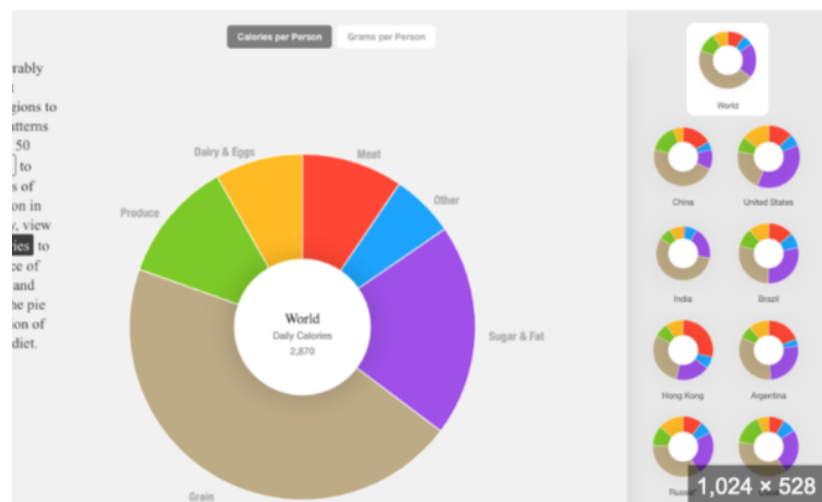
Develop your process book out of the project proposal. Your process book is the place you describe and document the space of possibilities you explored at each step of your project. Include many figures in your process book, including photos of your sketches of

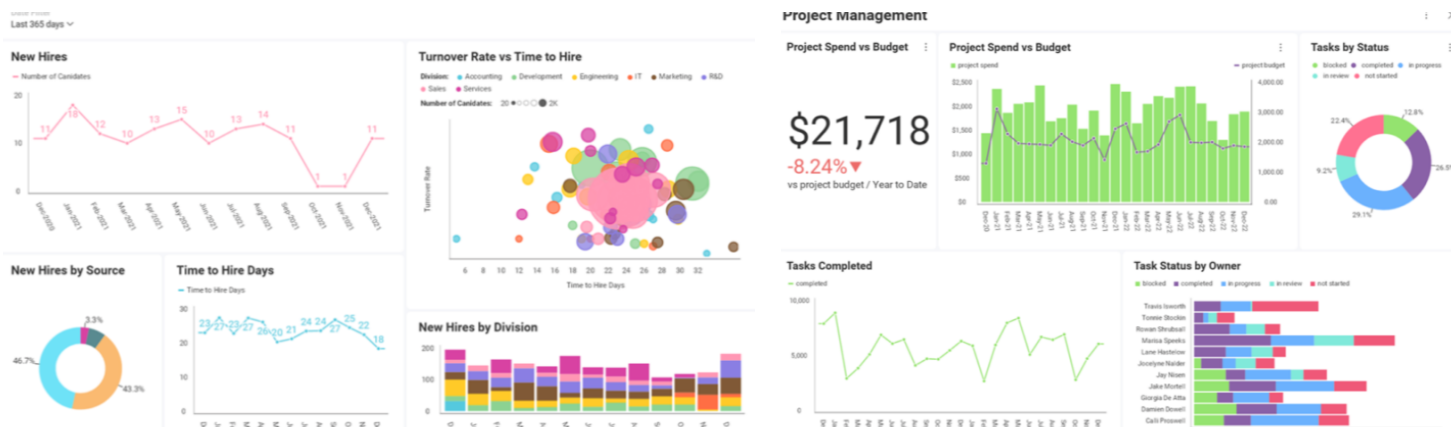
potential designs, screen shots from different visualization tools you explored, inspirations of visualizations you found online, etc. Use text to describe the rationale behind the evolution of your project.

Overview and Motivation:

In recent years, due to the growing popularity of online streaming and the shift away from traditional television, media companies and platforms have entered a highly intense market field. The motivation behind our project is to perform a competitive analysis to highlight content trends and platform performances between different streaming services.

Related Work:





A big inspiration for us to visualize streaming platform data was the recent uptick in the number of streaming platforms and the high volume of users that are moving away from conventional cable for streaming platforms. Based on the data we collected and what we wanted to portray, the following images above were a great starting point for how to do it. We ended up doing donut charts, cards, bar graphs, line-charts, and scatter-charts. Alongside the research we did, the assignments in class were a big help when it came to decide which visual would be best for what question we were trying to answer. A bar-chart to show average age, a scatter plot to show ratings, line-chart to show revenue over time, cards to visualize counts, and donut charts to show percentages.

Questions:

At the beginning we had some core questions we wanted to answer, like revenue and ratings, but as we gathered data and saw what was available to us and what would be relevant we discovered more questions to answer such as counts of tv shows vs movies and their corresponding percentages and average age of viewers per platform. These questions that

developed over the course of our analysis only helped to support our original questions thus making it very easy to keep a continuous flow throughout the topics of our visualizations.

- Which streaming service has had the most growth over the years? How has the revenue changed?
- Which streaming service has the most global reach?
- Which streaming service creates the most original content? What genres are the most common in their original productions?
- Which streaming services perform the best through ratings?
- Which streaming services focus on movies, which focus on TV shows?
- What is the average age of viewers for each streaming platform?
- Which streaming services provide the most content?

Decision Log:

- Bring in streaming platform data to analyze their performance over time:
Gathered data from various platforms.
- Task distribution: Kasidy – data collection, scatter plot, filter pane, barchart, cards, final design; Abbey – data collection, line chart, donut chars, final design,
README: Weekly meetings helped to keep task delegation fair and consistent.
Each person took on tasks each week based on progress from previous tasks.
- Flow of visuals: Best way to structure visuals in order to keep a continuous flow.

Data & Exploratory Data Analysis (All Team Members)

Since we based our visualization on different data sources we needed to create consistency across all the tables. We decided the best way would be to create two master data sets for Streaming Revenue and Streaming Catalog. To get a better understanding of the initial data sets we have performed data analysis to see what values we want to highlight.

IMDb	Min	Max	Average	
Year	1954	2021	2015	5667
IMDb	1.1	9.4	7.11	1876
Rotten Tomato	10	100	54.13	5660
HULU	Min	Max	Average	
Year	1923	2021	2011	2669
IMDb	1.5	9.3	7.08	1419
Rotten Tomato	10	100	55.80	2669
Prime	Min	Max	Average	
Year	1904	2021	2004	5945
IMDb	2.4	9.6	7.15	1183
Rotten Tomato	10	93	46.50	5945
Disney	Min	Max	Average	
Year	1922	2021	2001	1274
IMDb	3.2	9.6	6.9	316
Rotten Tomato	10	96	55.86	1274

For this analysis we initially split the data into movie and television data and used R to calculate the following statistical functions:

Movies	Min	Max	Average	Total
Netflix	1954	2021	2014	3696
HULU	1923	2021	2011	1048
Prime	1914	2021	2002	4114
Disney	1922	2021	1997	923

Movie	Min	Max	Average	Total
Netflix	10	98	54.44	3696
HULU	12	94	60.39	1048
Prime	12	93	50.39	4114
Disney	10	96	58.31	923

TV	Min	Max	Average	Total
Netflix	1960	2021	2016	1972
HULU	1931	2021	2010	1419
Prime	1904	2021	2009	1832
Disney	1955	2021	2012	352

TV	Min	Max	Average	Total
Netflix	10	100	53.55	1972
HULU	10	100	52.83	1419
Prime	10	93	50.39	1832
Disney	10	95	49.42	352

TV	Min	Max	Average	Total
Netflix	1.1	9.4	7.11	1876
HULU	1.5	9.3	7.08	1419
Prime	2.4	9.6	7.06	1832
Disney	3.2	9.6	6.97	352

This project will require an aggregation of different datasets. We plan on sourcing this data from credible sites like Kaggle and IMDb Developers API. Listed below are some example datasets.

- [IMDb Developer](#)
- <https://www.kaggle.com/datasets/ruchi798/tv-shows-on-netflix-prime-video-hulu-and-disney>
- <https://www.kaggle.com/datasets/ruchi798/movies-on-netflix-prime-video-hulu-and-disney>
- <https://www.kaggle.com/datasets/azminetoushikwasi/ott-video-streaming-platforms-revenue-and-users>
- Hulu Revenue: <https://www.businessofapps.com/data/hulu-statistics/>
- Netflix Revenue: <https://companiesmarketcap.com/netflix/revenue/#:~:text=According%20to%20Netflix%27s%20latest%20financial,that%20were%20of%20%2429.69%20B.>
- Amazon revenue: <https://www.businessofapps.com/data/amazon-statistics/>
- Disney+: <https://www.businessofapps.com/data/disney-plus-statistics/>

We put together line-charts and scatter plots to initially look over trends of our data. The websites we got our data from also had a few simple visualizations of the trend of the data.

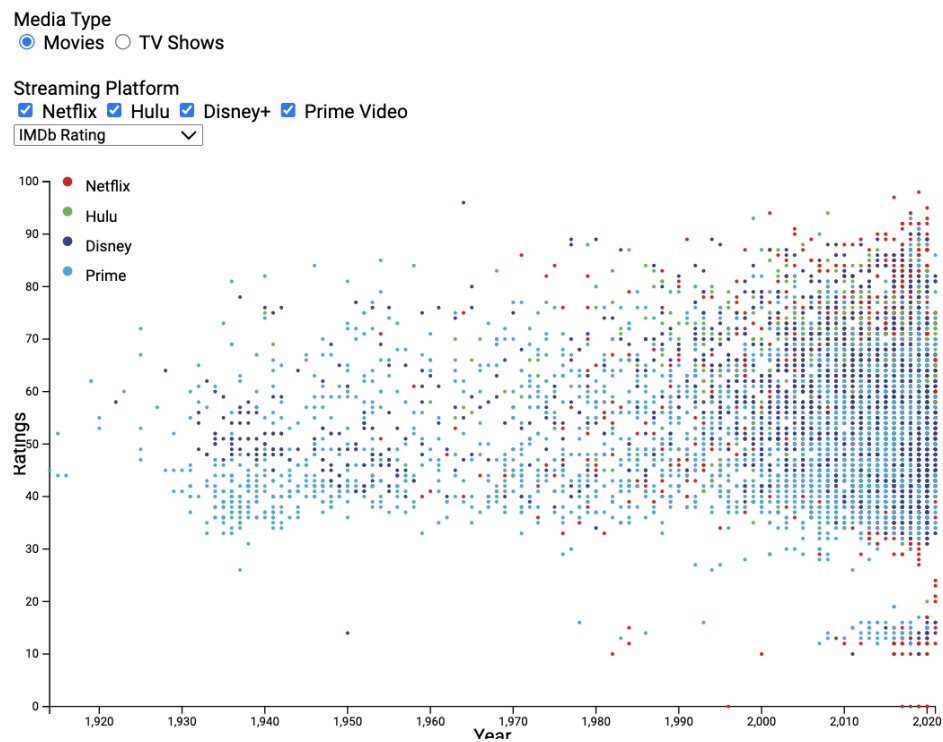
We pulled in data by bringing in the data files as CSV files then reading in the data from there.

- Line-chart.js: Creating a line-chart to show the trend of revenue for each streaming platform over time.
- Scatter.js: Creating a scatter plot to show the trend of ratings over the different streaming Platforms
- BarChart.js: Creates a bar chart to show
- Script.js: Loads in data.
- Main.html: Sets up the structure of our website
- Styles.css: Colors, formatting and designs.

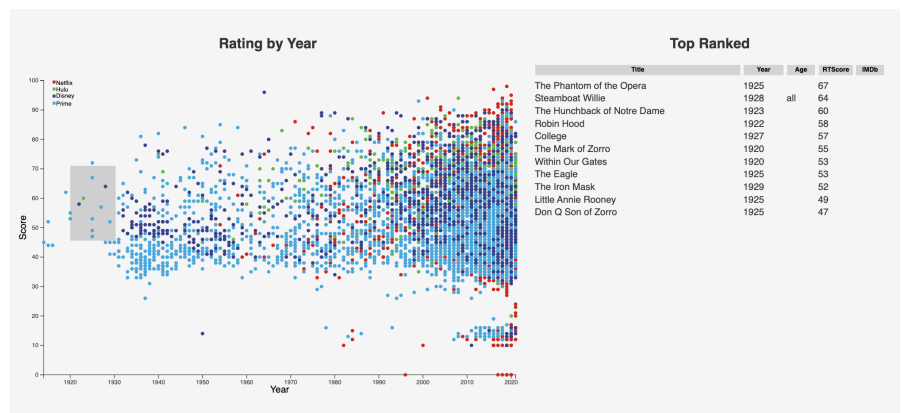
Design Evolution and Implementation:

The main graph is going to be a The evolution of this project is still consistent with our initial drafts. We mainly pulled our data visualizations from examples from class but we do intend on adding more complex concepts to draw a higher level of interactivity. We will be using a filter tool that interacts with different views by updating dynamically. The main graph shows the year and ratings of each individual show/movie while encoding the platform with color. This will update with transitions based on the input filters. For

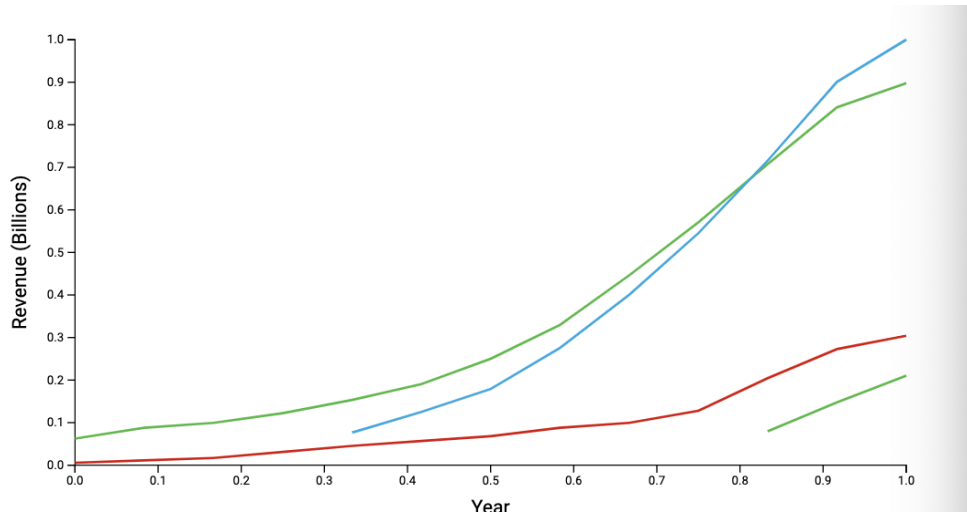
our final project we want to add tooltips and brushing to gain more specific information.



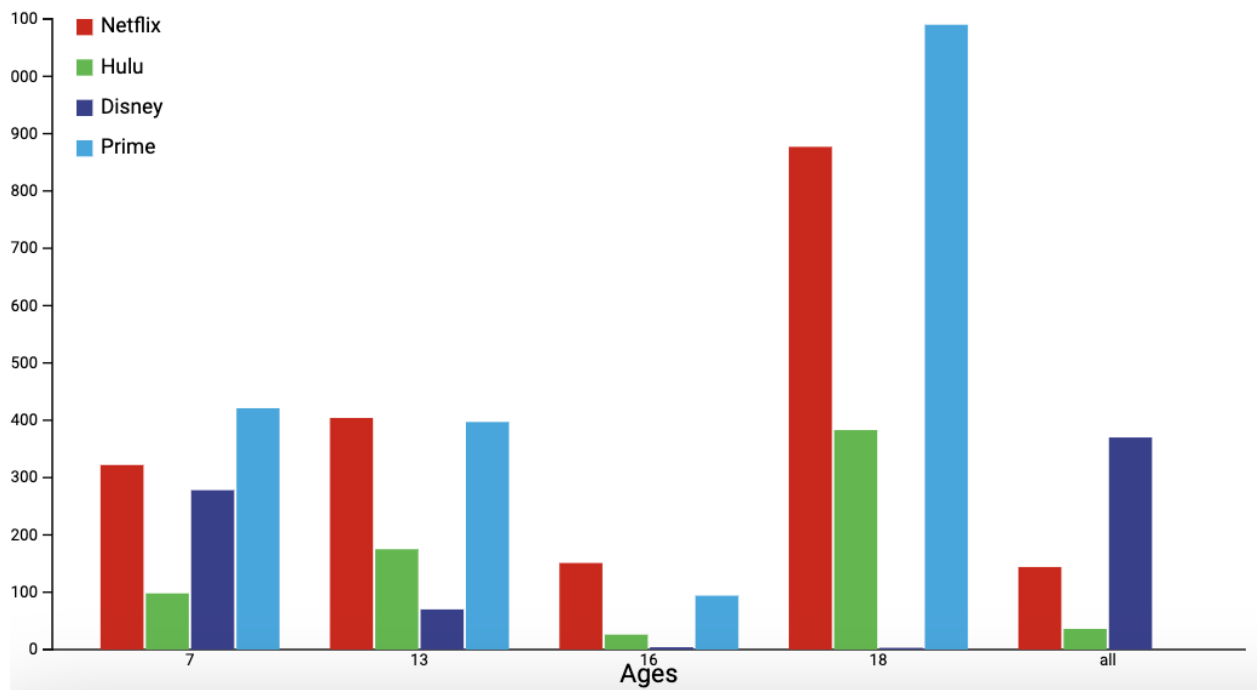
We've added a brushing feature to dynamically update the list view to highlight specific areas of the graph.



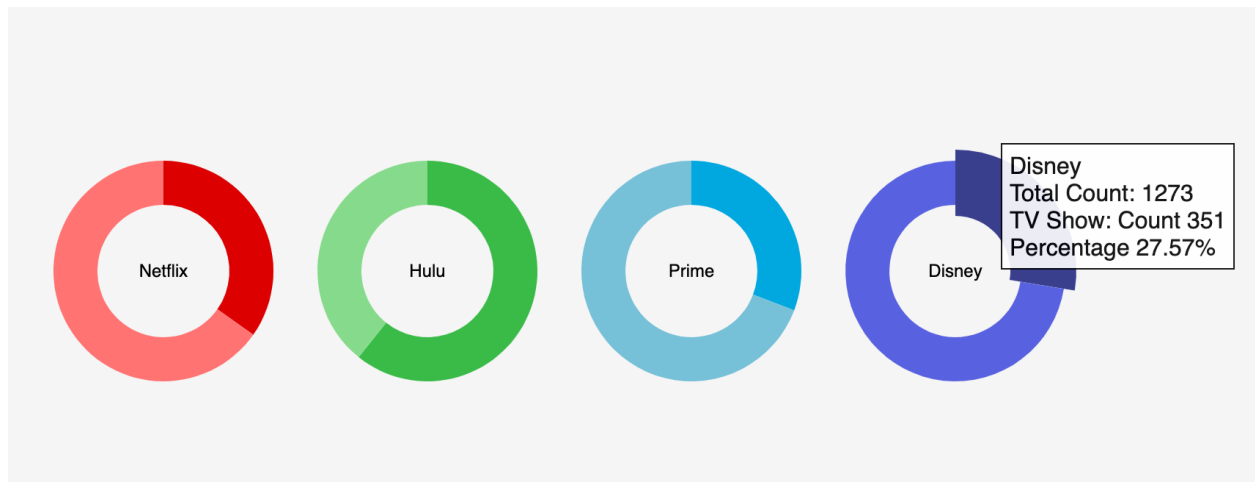
Next we have a line chart that visualizes the revenue by year of each streaming platform.



We also have a bar chart that highlights the age and demographics of each streaming platform.



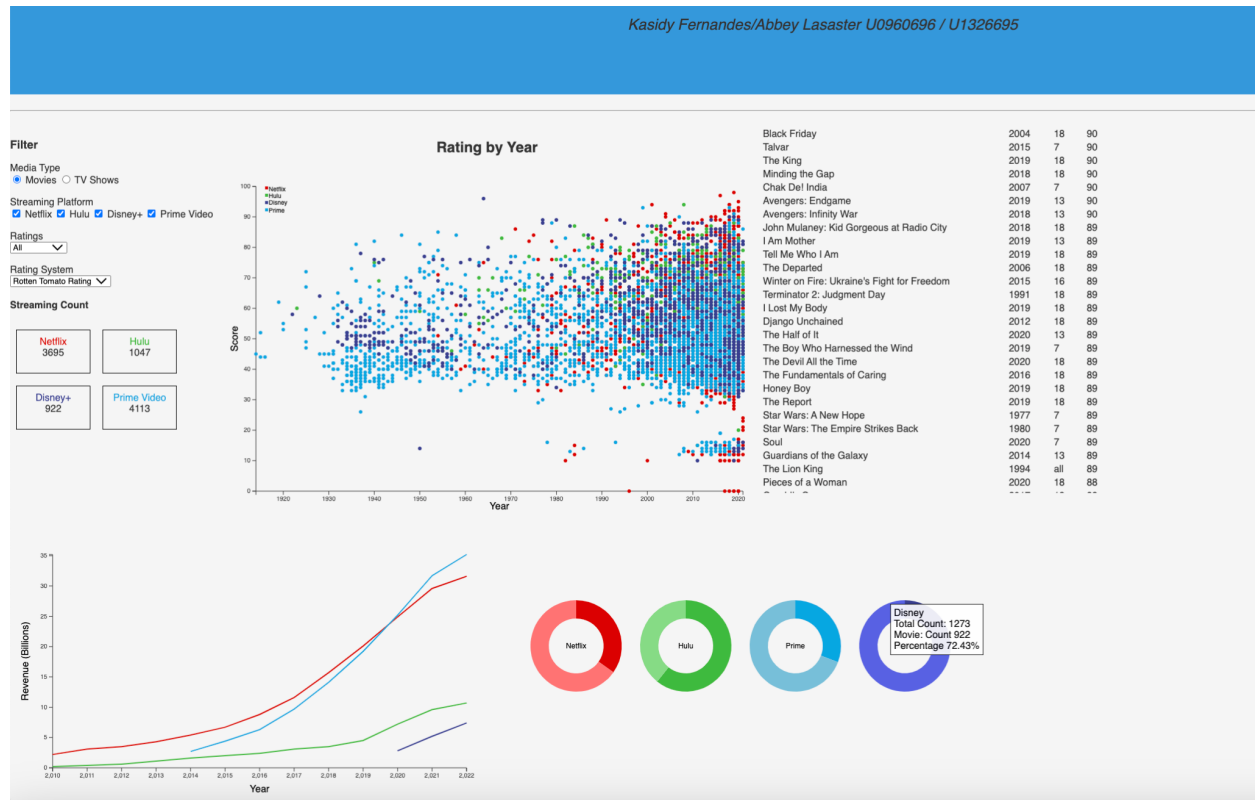
The addition of donut charts helps to point to what each streaming platform's main focus is, TV shows vs. Movies.



Final Design Implementation

For our final design we've cleaned up the filter tool for a more robust and accurate system. We've added cards displaying the count of each streaming platform to quickly summarize the data for the user and a listview which shows the top ranked film/shows

for the filtered options.



Current Project Schedule - Milestone One

This is something we will be working on throughout the project. The implementation will involve all team members, the documentation is divided into its essential sections and distributed evenly among team members for documenting our process.

Component	Who	Task Description
Overview, Motivation, Questions	All Team Members	We wrote this in our proposal...probably won't change it after the Staff Review.

Decision Log	Abbey	Adds important decisions to the process book, providing a short description/justification
Related Work & Resources	Abbey	Essentially manage a bibliography, but includes relevant or important screenshots/sketches
Data & EDA	All Team Members	Creates the data generation method and maintains a reasonably detailed account of what the code does, resources used, and the goals for each addition/script.
Design Evolution	Kasidy	Tracks a high-level narrative of our decision log, starting with our design process for the Project Proposal.
Implementation Description	Kasidy	Provides a high-level

		<p>write-up intent, functionality, key elements; essentially provides a simple user-introduction (what our project is how to make sense/interact with our visualization).</p>
Evaluation	All Group Members	<p>Tracks the challenges and solutions we encounter in the process of completing the project. Also writes a critique of the final submission (what did we learn/accomplish, what could be improved, etc).</p>

Functional Prototype Milestone

Due Date: November 3rd

Component	Who	When
Code	All Team Members	10/30
Process Book	All Team Members	10/30

Final Submission

Due Date: December 1st

Component	Who	When
Project Website (Github Page)	Kasidy (prepares site for submission, making any necessary final touches)	11/30
Code, Data, README	Abbey (prepares code, data and readme on github for submission)	11/30
Process Book (Clean up)	All Team Members	11/30
Video	All Team Members	11/30