COM-480 DATA VISUALIZATION

JUNE 2022

MILESTONE 3

TEAM MEMBERS

ALBERT AILLET
ALEXANDER BROKKING
SHRIRANG BAGDI

COMING UP WITH THE IDEA

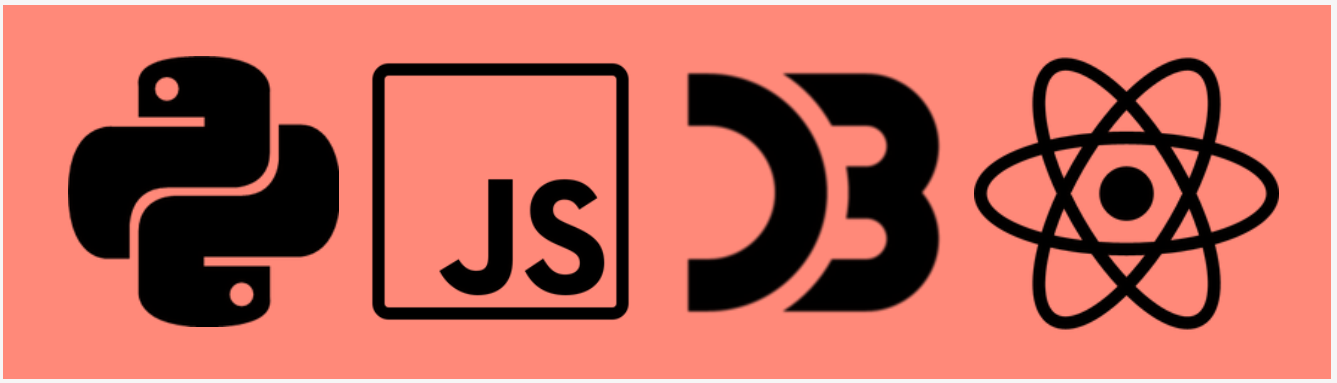
The process of developing this visualization has taken many twists and turns. Starting in mid March we began brainstorming around interesting options for our visualization project. We were looking for something contemporary that could interest our peers and which also provided enough depth to let us create a visual story.

We explored a range of different data sources including Kaggle, Amazon, opendata.swiss and scb.se. As this project was focused on visualization, a relatively clean and processed dataset was one of our criterias.

Since our arrival as exchange students, we had been impressed with the convenience of Publibikes, the bike-sharing service available in Switzerland and offered with a discount for EPFL-students. Hence, our first visualization idea was to explore the movement of Publibikes throughout Lausanne and how it changed through time. Perhaps identifying patterns and interesting anomalies that could be shown in creative ways. We ended up reaching out to the Publibike team who responded that they were excited to hear from us but would be limited to giving us data of the station occupancy and not data of individual trips. Considering this, we felt it would severely restrict our visualizations and thus we continued our search for the right idea.

Finally inspiration came. Through looking at previous examples and identifying Twitter as an interesting data source together with serpendiously reading news articles about Elon Musk, the idea of exploring his Twitter activity and its correlation to asset markets arose. The potential story met the criteria of being contemporary and provided enough ambiguity to be explored in depth.

After deciding to further pursue this idea, we initiated a phase of data gathering and exploration. One challenge that we faced was obtaining a dataset of Elon Musk's tweets. We were able to find a finished dataset of Elon Musk's tweets between 2011 and March 2021 on Kaggle. However, as the story was currently developing we wanted his latest tweets as well, this inclined us to apply for a Twitter development account to access their API. After approval, a Python script was developed to fetch his latest tweets using the Twitter API and enrich our dataset to include tweets up to March 2022. The assets of interest we decided to explore were chosen from the frequency in which they appeared in news articles we found. Thus, Tesla Stock, Bitcoin and Dogecoin were chosen. Datasets for these assets were uniformly created by developing a Python script that fetched the historical prices from the public yahoo finance API.



TECH STACK

PYTHON

Used to create the datasets and perform the initial data exploration. Main libraries used:

- Yahoo finance: Access the yahoo finance API for asset prices.
- Pandas: Handling the data for the data exploration
- Plotly: Quickly create interactive graphs as part of the data exploration.

HTML/CSS/JAVASCRIPT

Foundational web development technology stack to create our interactive visualization.

VITE

Used as tooling, providing a development server and build bundling.

D3.JS

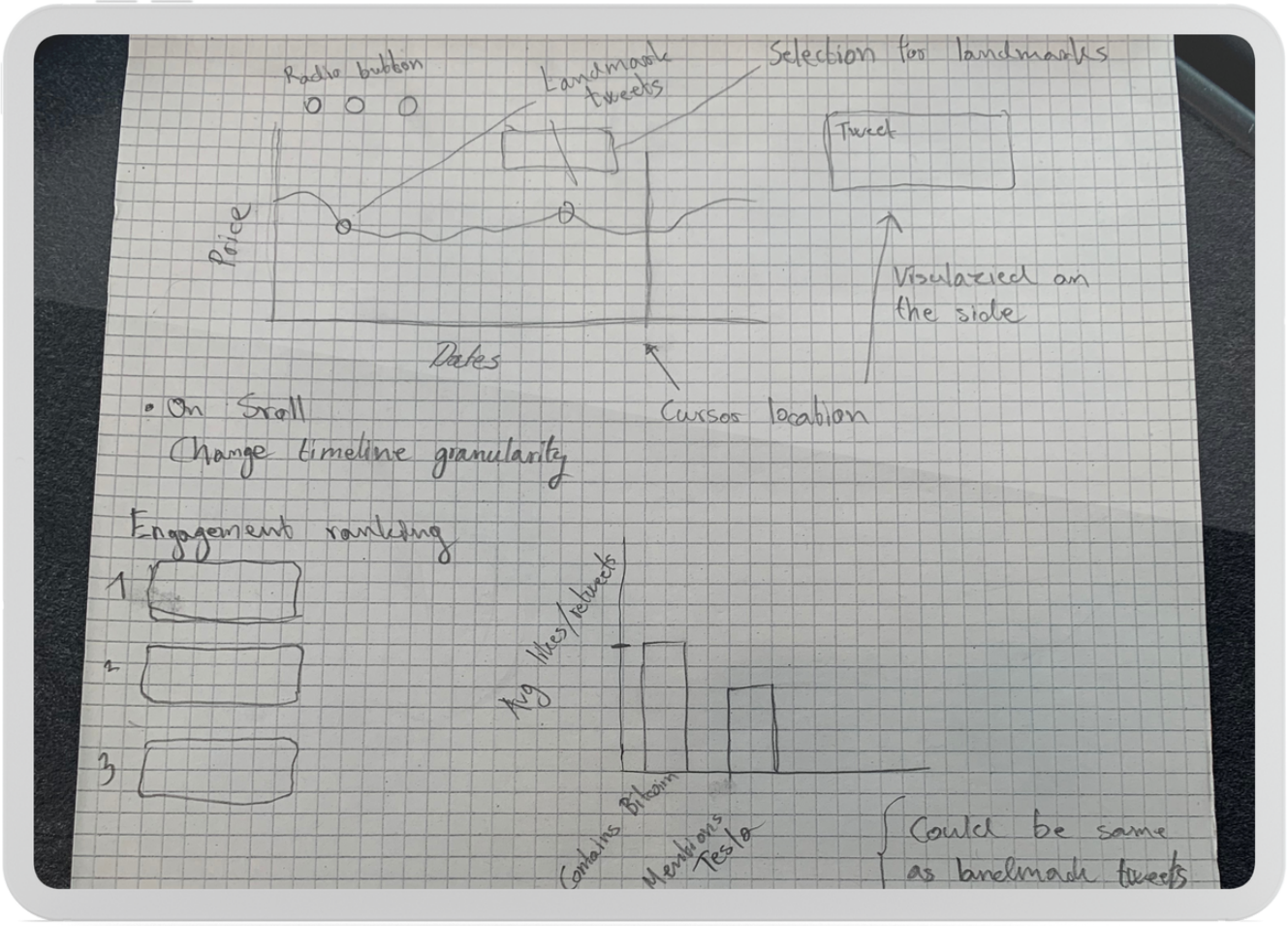
Used as the main tool for DOM manipulation. It has allowed us to create our interactive visualizations through binding data to the DOM.

REACT.JS

Used for more structured and declarative code with reusable components. Also enables integration of dependencies from the large react ecosystem. We ended up rewriting our codebase from plain Javascript to React as the project was growing and the added structure helped keep the project manageable.

- React-Router-Dom: This dependency is part of the react ecosystem and enables easy routing and conditional component rendering depending on the url path.

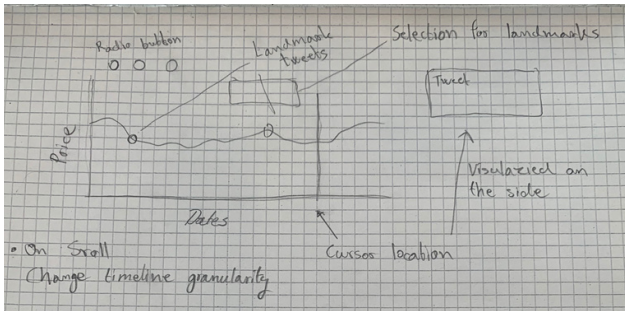
PROCESS

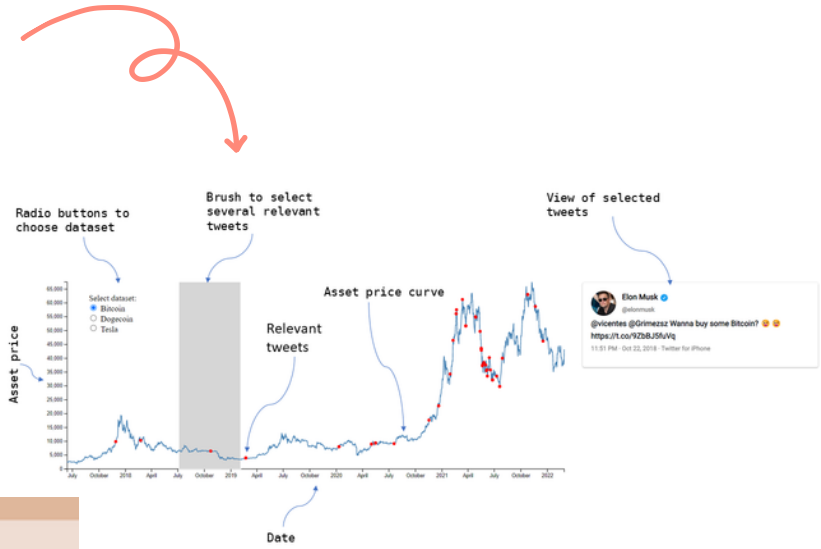


FIRST SKETCHES

Above is the very first sketch of our three main visualization ideas. This was drawn the first of March and served as a guide throughout our project. While the implementations of these sketches changed quite a bit since the first outline, much of the original ideas ended up being developed into our final visualization.

PROCESS - MAIN PRICE CHART



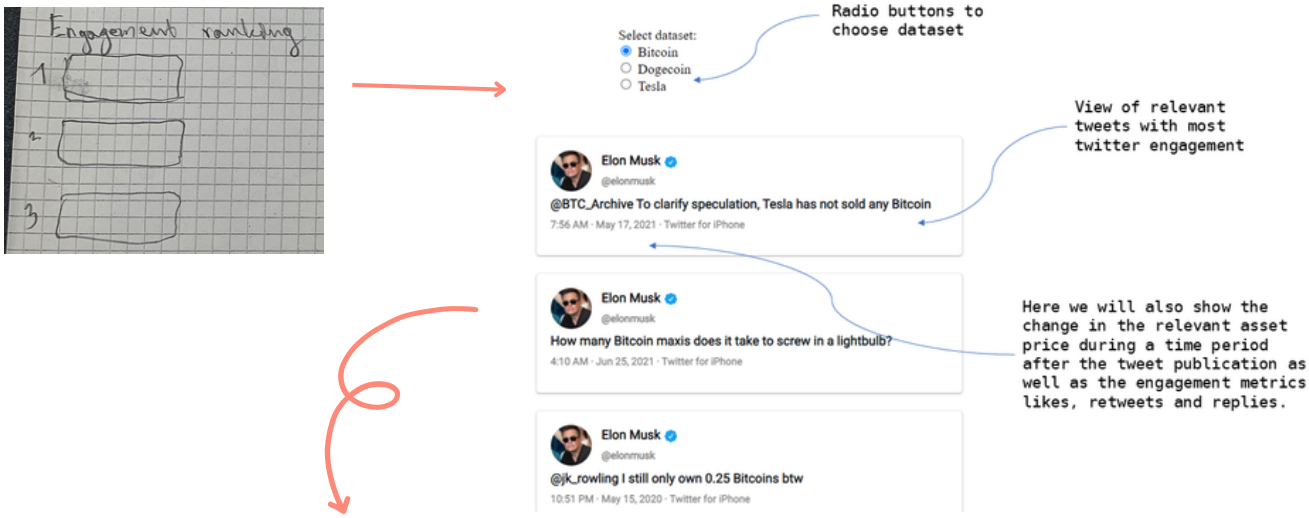


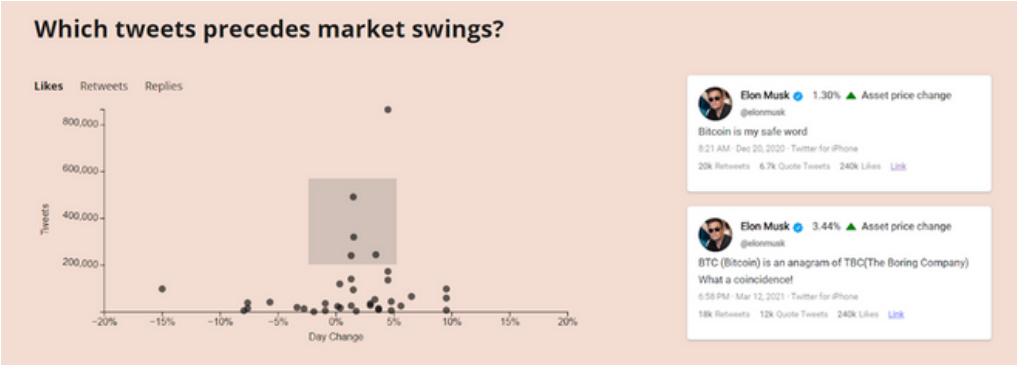
For the first visualization, the overarching design remained quite similar but with new ideas for more interaction and further developed design.
In the initial sketch the selection was a cursor limited to one dimension, it was then developed into a two dimensional brush for the subsequent milestones, allowing the user to see more tweets simultaneously. After milestone two this was combined with transforming the vertical list of selected tweets to a horizontal one. Thus, more tweets could be viewed and the graph could take up a larger portion of the screen. By having the tweets organized in chronological order horizontally, the user could naturally connect selected tweets with markers in the graph.

The selection of which tweets were shown, referred to as “Landmark Tweets” in the initial sketch, ended up not implemented. In combination with the asset selection, this would create a nested selection which did not feel intuitive and distracted from the content of the chart. The idea of scroll to change timeline granularity “zoom” ended up implemented as first conceptualized while the radio buttons for changing the current asset moved from being part of the chart to the navigation where it instead controlled the current asset of the entire website.

Some changes were also introduced from feedback we received. Only after the suggestion of our professor did we get the idea of implementing a rolling average of the asset prices with a color depending on the price change of the rolling average, a feature that ended up in the final design.

PROCESS - ENGAGEMENT RANKING



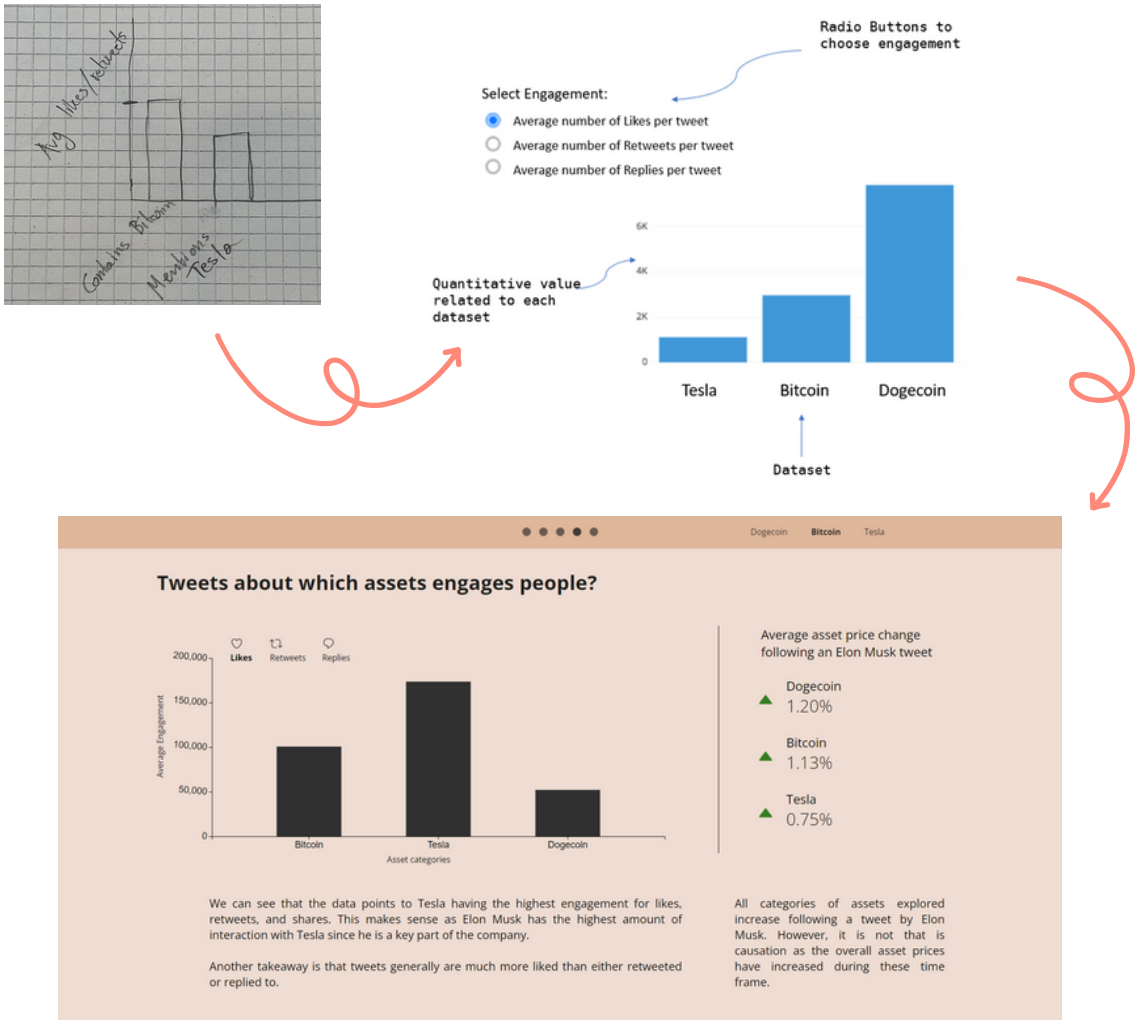


The second visualization referred to as the "engagement ranking" was originally conceptualized as a simple list of tweets that for each asset class ranked them in order of engagement. Moreover, the type of engagement would be able to be chosen between "Likes", "Retweets" and "Replies".

However, we felt that the simple list of tweets would not let the users explore the full extent of the data and the presentation made it hard for the user to get an overview of the relation between engagement and following change. Following this, we ultimately changed the visualization to show all the tweets for each asset category in a scatter plot where the y-axis is the selected tweet metric, and the x-axis is the asset price change 48 hours after the publication of the tweet. This more clearly illustrated the relation between the engagement of the tweet and how the market changed subsequently.

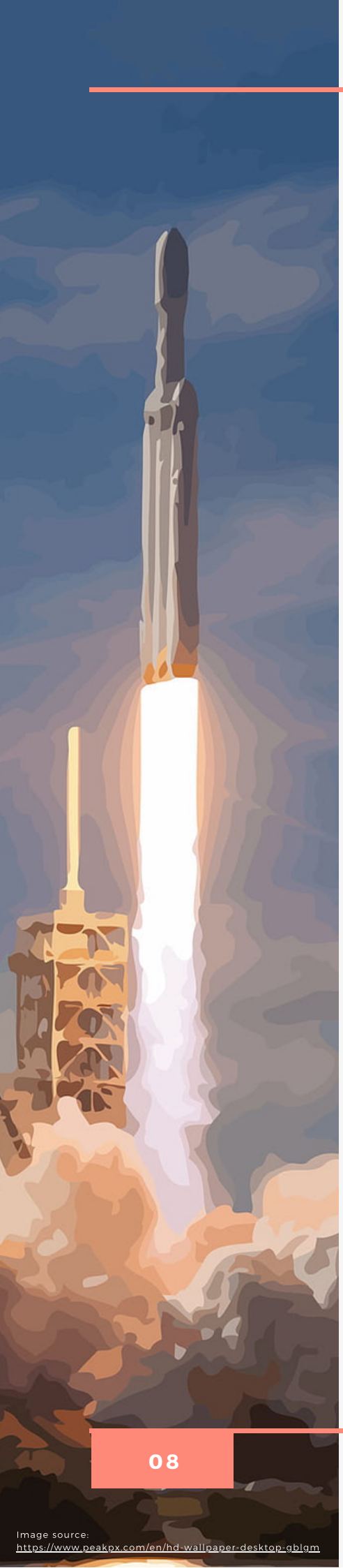
Compared to the simple list, this way of presenting the data enables the viewer to find outliers as well as see any potential trendlines or visible patterns.

PROCESS - HISTOGRAM



The main part of our last visualization ended up closest to our initial sketch. It shows the average engagement for different tweet metrics across the three different assets. While less interactive than the previous visualizations it is unique in showing the aggregate relation in terms of engagement between the three assets.

As the page still contained a lot of white space after implementing the histogram diagram, we continued to develop it between milestone 2 and milestone 3 to exploit the full screen real-estate. Firstly, text was added to guide the viewer around the visualization. Secondly, an additional column showing the average price change 48 hours after Elon Musk sent a tweet mentioning the asset class was added. Not only did this give an overall fuller look of the visualization, but it also enhanced the viewers ability to gain perspective on the asset classes on an aggregated level.



PEER ASSESSMENT

THIS SECTION INCLUDES A BREAKDOWN OF THE PARTS OF THE PROJECT COMPLETED BY EACH TEAM MEMBER.

In this first part we give a general outline of the different tasks carried out by the different team members.
Some of these tasks were developed jointly and in that case the main contributor is assigned to it.

DATA COLLECTION AND EXPLORATION:

Alexander: Data collection and formatting.
Albert: Data Exploration and preprocessing.

WEBSITE AND VISUALIZATION:

Alexander: Price line, introduction, conclusion, legends, tweets visualization, layout, styling and website deployment.
Albert: Brushing, zoom, tweet filtering, rolling average, axis labels.
Shrirang: Histogram.

All the Git history can be seen on the repository and is representative of member contribution:

Link to Git History

A specific breakdown of all individual completed tasks can be viewed in this spreadsheet.

Link to Spreadsheet

Image source:
https://www.peakpx.com/en/hd-wallpaper-desktop-qblqm