

Milestone 2

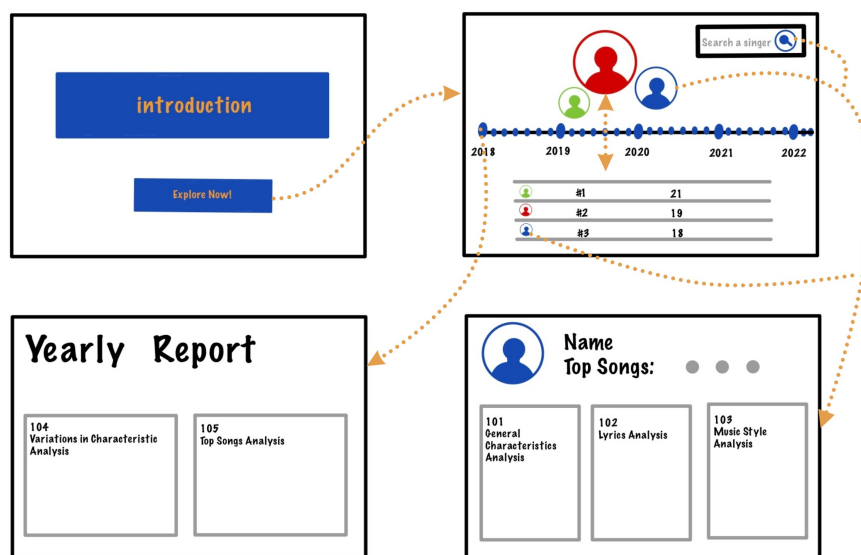
Please read the markdown version here: <https://github.com/com-480-data-visualization/datavis-project-2022-shazam/blob/main/README.md>, because the links are not rendered properly on the PDF

Core Visualization Sketches

Data set augmentation Our website is dedicated to analyzing billboard ranking data to make recommendations for new pop music lovers. The billboard weekly ranking data from 2018 to 2022 has been obtained.

The audio features and lyrics data for all top 20 tracks from each week, along with the singers' data that occurs within the time range, have been augmented to the dataset in preparation for milestone 3. The augmented data set includes around 280 distinct singers (in total, about 4000 tracks from all singers) and 300 distinct tracks from Billboard weekly top 20.

Due to the slowness and rating limiting of the Genius API, current data set stored on the website is mostly without the lyrics data.



Sketch

> Figure 1

As depicted in Figure 1, we attempt to make a Single-Page Application (SPA), which will have a fluid that guides the user to explore music with ease.

The main flow of our website is as followed: upon loading the website, the user will be presented with an introduction text to explain our goals and ideas (Figure 1, top left). With a click on the button to start the exploration, the

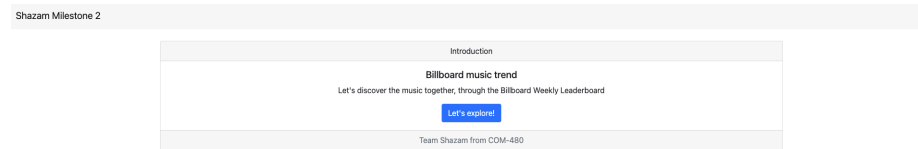
user will be presented with a time slider, in which the user can select a specific week between 2018 and 2022 (**Figure 1, top right**). Upon the selection of a week, the top singers of the week will show up, and the accumulated ranking leaderboard, presenting the most favored singer, will show up on the screen. The user can then click on the singer to view the singer's complete analysis (**Figure 1, lower right**), including general audio feature analysis, and lyrics analysis, with a breakdown with time, album, and all-time being available to the user to switch around.

The user can also select the years 2018, 2019, 2020, 2021, and 2022, to view the yearly music report (**Figure 1, lower left**), which will present the trends of the music, top songs, etc. so it's easy to see how the music evolves over time.

An alternative exploration can be done using the search bar by searching the singer's name. If the singer happens to be in the database, the same singer report analysis data (**Figure 1, lower right**), will be presented.

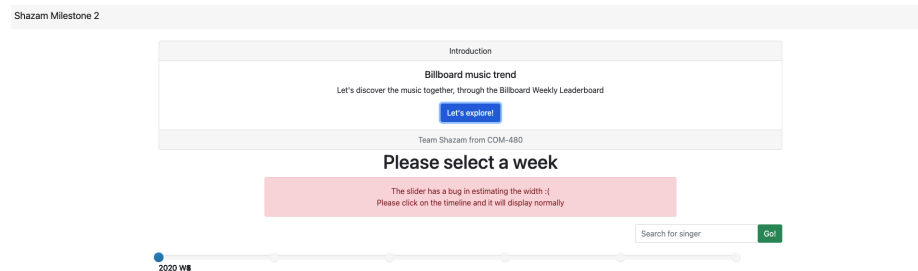
Website prototype

The prototype roughly presents our current experiment in the flow mentioned above. The style and functionalities will be heavily tuned and amended.



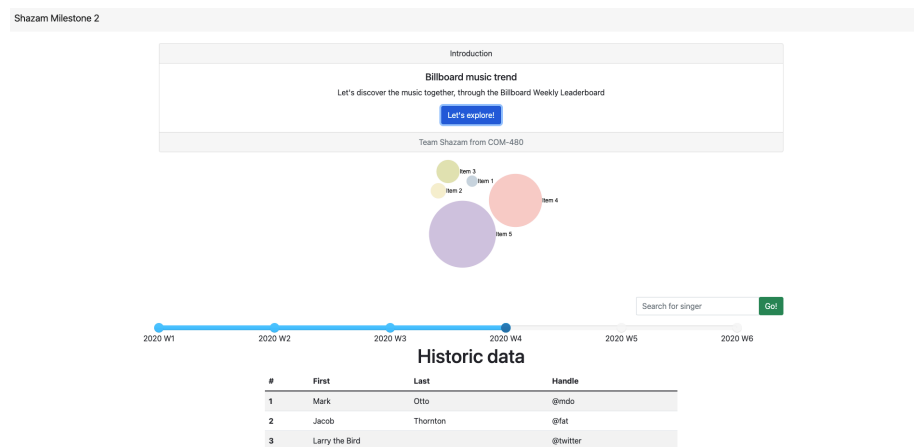
> Figure 2

Figure 2 presents the idea in Figure 1, top left.



> Figure 3

Figure 3 presents the idea in Figure 1, top right, before selecting a time on the time slider.



> Figure 4

Figure 4 presents the idea in Figure 1, top right, after selecting a time on the time slider.

The singer and yearly report are still work-in-progress at the time of writing, but it will be using the tools listed in the following section **List of tools in use for each visualization**. ### List of tools in use for each visualization

To showcase the singers who have the most popular songs during the selected week, we will use a bubble chart to represent the singers whose songs figured in the top 20. The size of the bubble represents the number of songs in the top 20 from the corresponding singer. We are planning to use d3.js to accomplish the task. (Reference links: [1] [2])

Radar chart will be used to present the musical characteristics (audio features), we take inspiration from the following links: [1] [2] [3]

To complete our exploration on audio features, we would like to visualize the tempo and key of each song for a given singer using a scatter plot with tempo on x-axis and key on y-axis, this will give us a sense about the artist's favorite keys and tempos, as well as the relation between them. (Reference link: [1])

The lyrics analysis will be carried out using a word cloud, which provides a direct view of the most frequently occurring words. (Reference link: [1])

The Lecture 6 slides are useful for us since they talked about the different marks and channels, and how to choose them to achieve high effectiveness. We understand that the position is the strongest mark, so we will make use of it to represent the timeline in our work.

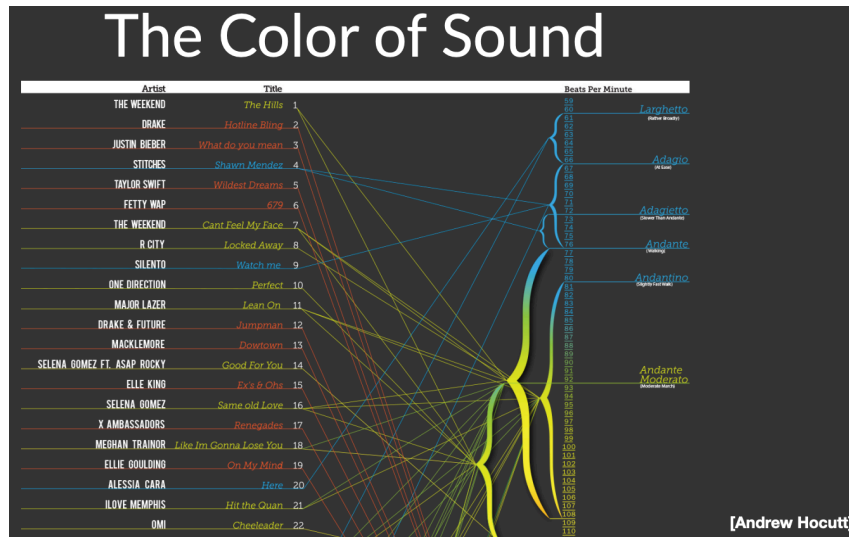
The Lecture 11 slides are particularly helpful because it gives us inspiration on visualizations of musical data, such as the color of sound and the evolution of music taste. We also learned about star plots and bubble plots in the Tabular data in this lecture.

Bootstrap is used to set up the main nav bar and miscellaneous parts of the website. It has decent CSS templates and alignment containers that we can use to structure our website. The current prototype is relying on Bootstrap heavily, especially aligning components using containers.

The interaction and data loading is accomplished using Vue.js since it supports convenient DOM manipulation. The website prototype is actually already leveraging using Vue.js, which allows us to animate the button actions and time slider actions, and easily load data and render the HTML code dynamically.

The data in use will be stored on Github course project repo, using Vue.js to fetch the raw files from it, and then parse the data and process it to feed it into the d3.js visualization. The data format being stored in the repo is always JSON, which is friendly to be operated on.

Potential Extra Visualization



- Reference: [Tweaking 1]
- We are considering replacing the name of singers in the bubble chart with their profile picture, which will be much more interaction-friendly and recognizable to the user.