DATA VISUALIZATION PROJECT REPORT

Analysis of my Personal Spotify Dataset, Drawing Inspiration from "Spotify Wrapped"

Rishiraj Sinharay

Student ID:32435851

Tutor: Nina Sophia Dsouza

Monash University

FIT5147 Data Exploration and Visualization

Semester 2, 2022

Table of Contents

INTRODUCTION

The basic idea behind this project is to be able to make my version of "Spotify Wrapped". Spotify Wrapped is a Spotify marketing campaign that shows visualizations of persons streaming history over the year. The campaign is interesting in the sense that it helps users visualize how unique their streaming habits are compared to others around the world. The aim was to analyze my own Spotify streaming history from July 2021 to July 2022 and also compare it with the most streamed songs globally in the same period to understand my streaming choices over the specified period to come to a similar conclusion.

In this report, I will explain my thought process behind the narrative of my visualization and answer the following questions at the same time [4]:

1. What are my Top 5 streamed artists and top 5 streamed songs in the given period and how many minutes did I spend listening to each for the entire period? [4]
2. How unique is my 2021 streaming history compared to the global streaming history in 2021 (July to December)? [4]

The intended audience for this visualization is the end user (me in this case) who uses the Spotify application. The findings from the Data Exploration Project will be showcased similarly to the Spotify Wrapped feature but with a more analytical approach.

DESIGN

The top 5 streamed tracks and artists and the minutes spent listening to music on the app is standard information that is displayed in the Spotify Wrapped feature as shown in figure 1.

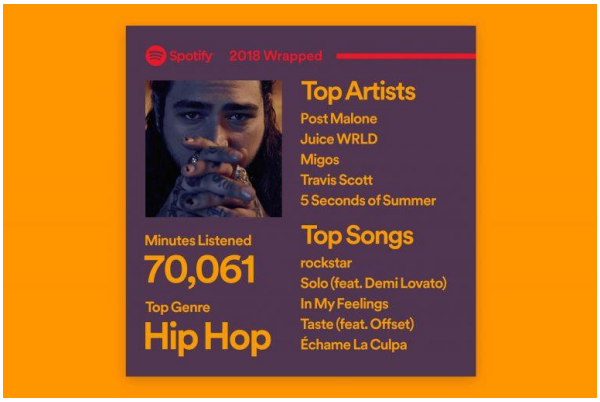


Figure 1. Ref: https://newsroom.spotify.com/2018-12-06/relive-your-year-in-music-with-spotify-wrapped-2018/

The idea is to narrate the user's streaming summary which would be **relevant** to the user concerning the application. My entire design focused on delivering the following information to the end user:

- Top 5 streamed tracks
- Top 5 streamed artists
- Minutes streamed monthly
- Top 5 tracks common with global streams
- Characteristic plots

I divided the above information into two tabs (**slideshow** format). The first tab would have the plots only concerning the user data in a **magazine-style** narration so that the user can **discriminate** and understand each type of data. The first tab would display the following information:

1. Top 5 streamed tracks and artists:
   These were both ranking plots that were going to be placed side by side. I went ahead with horizontal lollipop plots because I wanted to keep the plots simple for the end user to understand but at the same time different from the usual plots e.g., bar plots. Other plots like the Spider plot would have been hard to understand from the user's perspective. Horizontal because the information like track names and artist names could be comprehended better. Since the information was about the streaming history on Spotify, I choose Green (Hex color: #1DB954) and Black as the colors of most of my plots with a White background as green, black and white are the official Spotify colors [5].

   I also wanted to implement interactivity in my plots for the visualization so that the user gets to participate and interact. For the lollipop charts, the user gets to interact by hovering over the plot points which generates a tooltip containing information like track name, artist name, number of times the song was streamed, and number of minutes the song/ artist was streamed.

2. Minutes Streamed Monthly

For this section, Spotify just shows the total minutes streamed on the app in the entire duration. As shown in my Data Exploration Project [6], I divided it further into minutes streamed monthly to show a time series plot. I wanted it to look like a user's streaming journey over the month so I made it an animated time series plot. This way it looks more visually appealing using the **film/video/animation genre of narrative visualization**. The time-series plot will be an animated area plot so that it looks dense. The color green was chosen for this with a black outline for the line. The animation can be played by the user by interacting with the "Play" button. The kinks on the displayed plot also show details like the month, the number of streams in the month, and the total minutes streamed in the month. The user can also use the animation slider at the bottom to go to a particular time in the animation. This allows the user to **control a dynamic visualization**. The sidebar panel also shows the total minutes streamed in the entire duration.

The second tab showed plots for both the user data and the global streaming data in a **magazine-styled** narration with textual information explaining the plots. There were 2 plots with proper spacing for the user to easily **discriminate** and understand the information The following information was displayed:

1. Top 5 tracks common with Global Streams
   This was also a ranking plot but I wanted to make it different from the ones in the previous tab. I plotted various ranking plots and the circular bar plot looked good with text labels. Each track name was distinct and the bar heights could easily be used to compare between bars. There is no interaction in this plot. The color was again chosen as green to maintain the color coordination with the Spotify color scheme.

2. Characteristic plots
   This is designed with a **slideshow** style of narration where each slide is a different character being displayed. The plot is interactive where the user can choose to view a particular characteristic plot based on the selection from the radio buttons in the side panel. Each characteristic plot is a scatter plot with a trend line that infers the preference of the user. The points on each plot are also interactive and when hovered upon display

the track name, the number of times it has been streamed, and the measure of the characteristic. An inference statement based on the characteristic and the trend line is also displayed adjacent to the plot.

IMPLEMENTATION

The basic libraries installed before starting the implementation process were tidyverse, dplyr, and shiny. The datasets used were the wrangled personal streaming dataset and the wrangled combined dataset (personal and global streams). According to the Five Design Sheet, the entire visualization was supposed to be in a single tab but during implementation I wanted to divide the plots into different tabs, to encourage user interaction and also show information in relevant segments for better storytelling. The entire visualization was divided into two tabs using the tabPanel() function in the UI. The implementation task for each plot was done as follows:

1. Top 5 streamed Tracks and Artists

   The personal dataset was used for these two ranking plots. The imported dataset was grouped by the track name and artist name and summarized to get the count of each track name and the sum of the minutes played for each track to get a new data frame for top tracks. This data frame was sorted by the stream count and the top 5 rows were sliced out for the plot. Similarly, a data frame for top artists was obtained by grouping only by the artist's name and then summarizing in the same way as above.

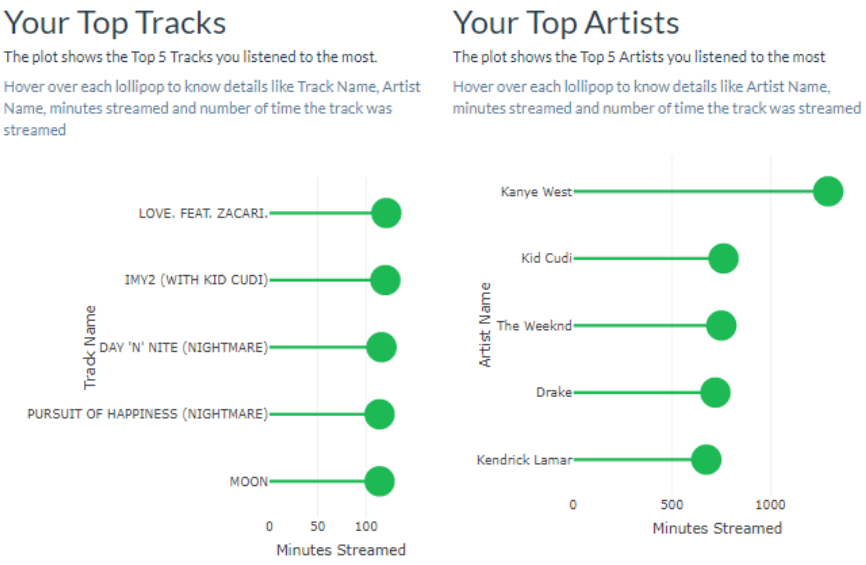


Figure 2. Top 5 Tracks and Artists

The next step was to build an interactive lollipop chart. In server.R I used the Plotly library to make the chart interactive. The difficult part was to make a lollipop chart as there was no function in plotly for the same. I made a scatter plot and added a bar plot with a small width in the same graph. The color was set to hex code #1DB954 for the Spotify green and the tooltip displayed details like Track name, artist name, streams, and minutes streamed. The plot was also arranged in ascending order of y-axis by the x-axis measures to show the arrangement from highest to lowest streams. In the UI I added text for the heading and help texts to guide the user. According to the five-design sheet, the plots were supposed to be vertical but I went ahead with horizontal lollipop plots so that the track names and artist names are visible clearly as they need to be conveyed properly to the user. The five-design sheet also mentions that this would be animated but the time-series animation looked better and animating everything would leave the tab blank with no plots. The current implementation looks filling and creates a better first impression.

A setback I faced was the part where I wanted to plot circular bar plots on tooltips for the lollipop chart for the top 5 artists showing their top songs which were streamed but was unable to do so.

2. Minutes Streamed Monthly

For this, I divided the row into a sidebar panel and a main bar panel in the UI.R file. The sidebar had a text with a calculation of the "total minutes streamed" pasted within the text which was calculated on the top of the UI file as the sum of all the rows of the minutes played column from the personal dataset. The UI.R file also has help texts to help the user with the animation and interaction of the time-series plot. The main panel had an animated time-series plot.

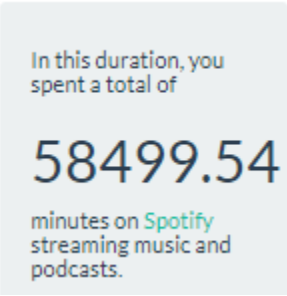


Figure 3. Total minutes streamed

In the server file, the personal stream dataset was grouped by month using the date column (converted to date-time format using the lubridate library) [11] and stored as a new data frame

with additional columns summarizing the monthly streams and monthly minutes played. A cumulative function was written to cycle upon different frames to make a continuous animation [7]. The cumulative function was used on the month column to make it continuous and an area plot was plotted with the month on the x-axis and minutes streamed on the y-axis. The animation was set to 100 frames. This was rendered to the UI file. As per the instructions in the side panel, clicking the "Play" button would run the animation plot.

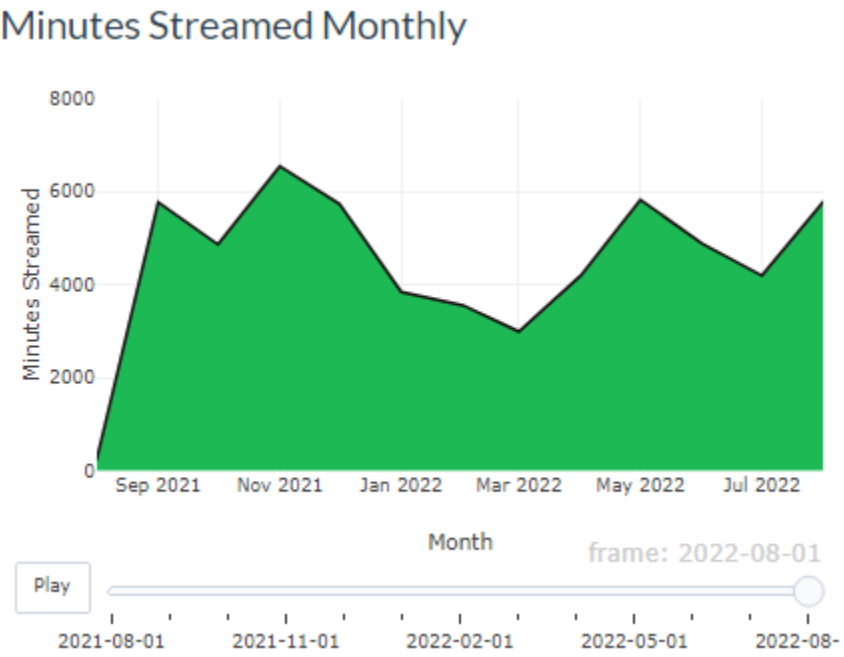


Figure 4. Minutes streamed monthly

3. Top 5 Tracks Common with Global Streams

   This was plotted in the "Listening with the World" tab. The wrangled combined dataset was used for the plots on this tab. The combined dataset was grouped by the track name and artist name to get the summary of the streams and minutes played and then was ordered from highest streams to lowest streams to get the top 5 common streams. In the UI, the row was divided into a side panel to explain the plot and a main panel for the plot. The side panel only had the text.

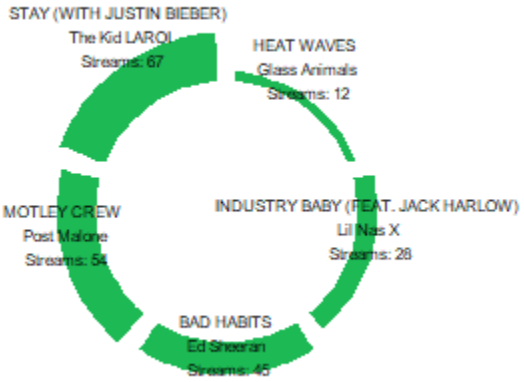


Figure 5. Top 5 common tracks

In the server file, the top 5 common streams were plotted into a circular bar plot using the ggplot2 library. This was done by first plotting an ordered bar chart and then making it circular using the coord_polar() function of the ggplot2 library. The theme was kept minimal to make the plot clearer and neater in the panel.

4. Characteristic Plots

This was plotted in the "Listening with the world" tab as well. The wrangled combined dataset was used to make different datasets for each type of characteristic. The combined dataset was grouped by the track name and artist name to summarize the average characteristic of the track, the minutes played, and the number of streams. In the UI file, the row was divided into a side panel and the main panel. The side panel had text regarding the characteristic being selected and radio buttons for each characteristic. Each radio button points to one characteristic. The main panel has the plot of a particular characteristic as selected in the sidebar panel with a description of the plot.

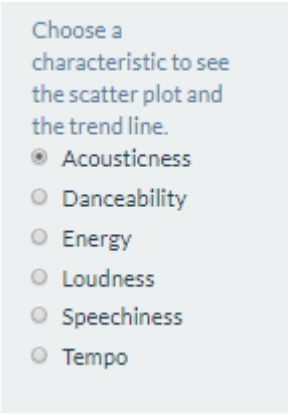


Figure 6. Characteristic plot radio input

In the server file, multiple if-else conditions are used to plot a scatter plot with a trend line using the ggplot2 library for the characteristic selected in the UI. Each plot is then converted to an interactive plot using the ggplotly library. Similarly, multiple if-else conditions are used to get a text output of the plot based on the radio button input selected in the UI.

According to the five-design sheet, I was supposed to implement a drop-down for selecting the characteristic to be plotted but that made the side panel look empty. The radio button selection took up a good amount of space and also informed the user of the different characteristic options available.

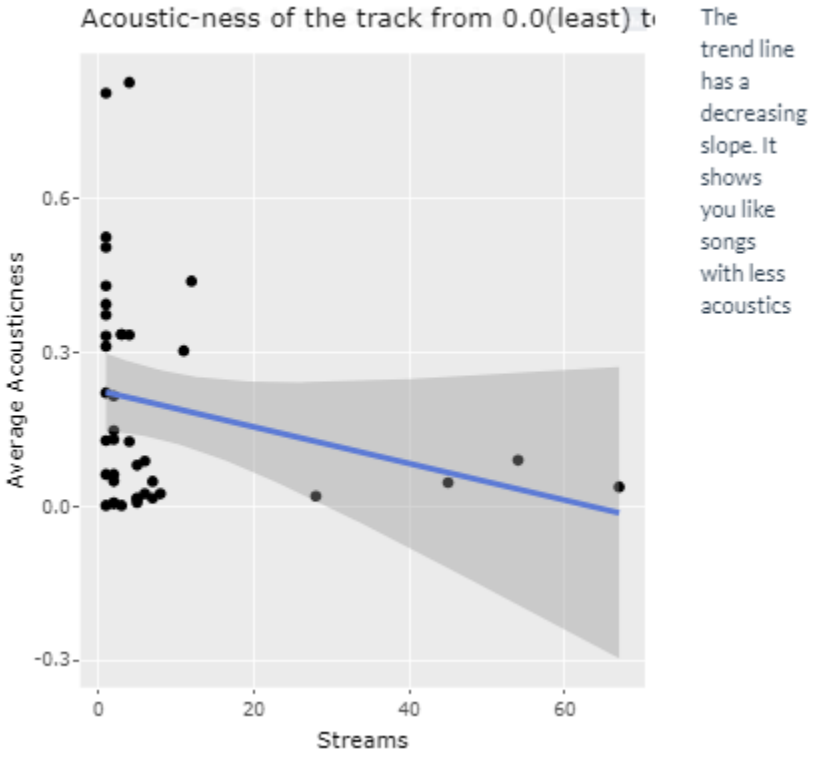


Figure 7. Characteristic Plot

Apart from this, each plot has a withspinner() function of the shinycssloader library piped to it in the UI file to show an animated spinner to show that a figure is loading in the application instead of a blank page.

USER GUIDE

The user should follow the instructions given in the app to navigate their way through the visualization. On running the shiny application, an Introduction page will pop up giving a summary of the visualizations (and the Spotify Logo [12]) and how to use the app. A **readme file**

has been provided in the zip folder with instructions about libraries to be installed and unpacked before running the shiny application. All the plots apart from the circular bar plot in the "What you had common with the world?" section are interactive and can be hovered upon to get more information. The "Minutes Streamed Monthly" time-series plot has a **Play** button at the bottom to run the animation.

CONCLUSION

With this narrative visualization, I was able to understand how one of the world's largest audio streaming providers created a marketing campaign using user data and visualization to engage an audience. I was also able to analyze my music streaming choices over the past year and compare them with Global streaming in the same period.

If given a chance, I would like to use different global datasets to compare with my streaming history or a dataset spread over a longer period and different genres of music. The goal would be to merge it into a larger dataset, big enough to create predictive models for music recommendation based on characteristics. Another thing I would like to add to the visualization would be user interaction to see the top tracks and artists streamed in different months.

BIBLIOGRAPHY

The following sources were used as a reference to successfully finish the Data Visualization Project:

[1] Curtis Joe, *How to see Spotify Wrapped results, stats, and playlists*, 30/05/2022. URL: https://www.androidauthority.com/how-to-see-spotify-wrapped-3069969/

[2] 2021 Wrapped, https://open.spotify.com/genre/2021-page

[3] Wikipedia, Spotify Wrapped, https://en.wikipedia.org/wiki/Spotify_Wrapped

[4] Rishiraj Sinharay, Project Proposal, FIT5147, Analysis of my personal Spotify dataset, drawing inspiration from "Spotify Wrapped"

[5] Spotify Colors, U.S. Brand Colors, https://usbrandcolors.com/spotify-colors/#:~:text=The%20official%20Spotify%20colors%20are,to%20visit%20the%20company%20website.

[6] Rishiraj Sinharay, Data Exploration Project Report, FIT5147, Analysis of my personal Spotify dataset, drawing inspiration from "Spotify Wrapped"

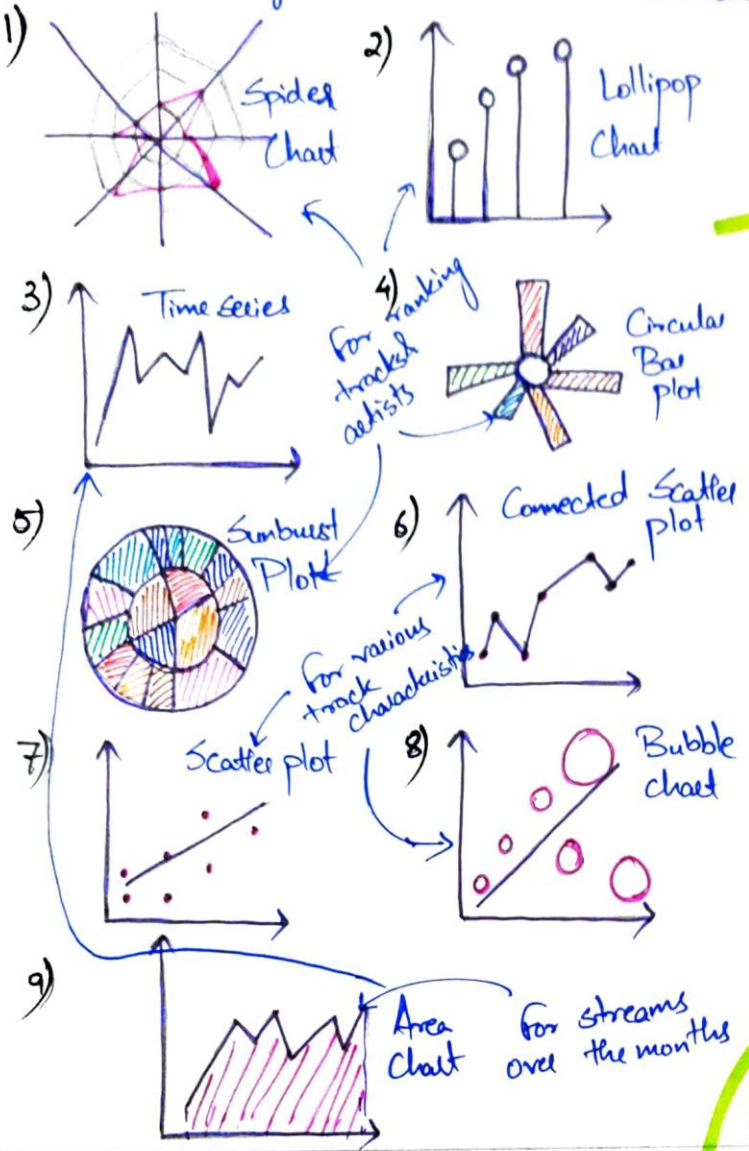
[7] Cumulative Animations in R, https://plotly.com/r/cumulative-animations/
[8] Shiny Themes, https://rstudio.github.io/shinythemes/
[9] Hover Text and Formatting in R, https://plotly.com/r/hover-text-and-formatting/#format-hover-text
[10] Different inputs & sidebars for each tab, https://community.rstudio.com/t/different-inputs-sidebars-for-each-tab/1937
[11] Zach, May 12, 2022, How to Group Data by Month in R (With Example), https://www.statology.org/r-group-by-month/
[12] Spotify Logo, https://www.pngitem.com/middle/mwTiwb_spotify-logo-small-spotify-logo-transparent-hd-png/

SHEET-1

1. Ideas

What to show?

- Rankings - Streams over time - Common characteristics



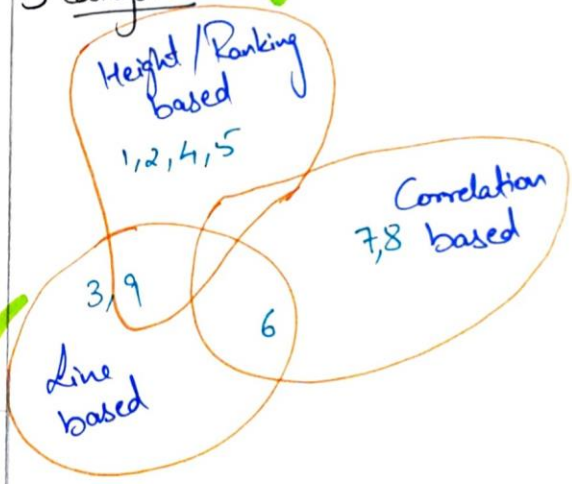
2. Filter

→ 2 & 4 seem more suitable for ranking. 1 can be used but might not quantify properly

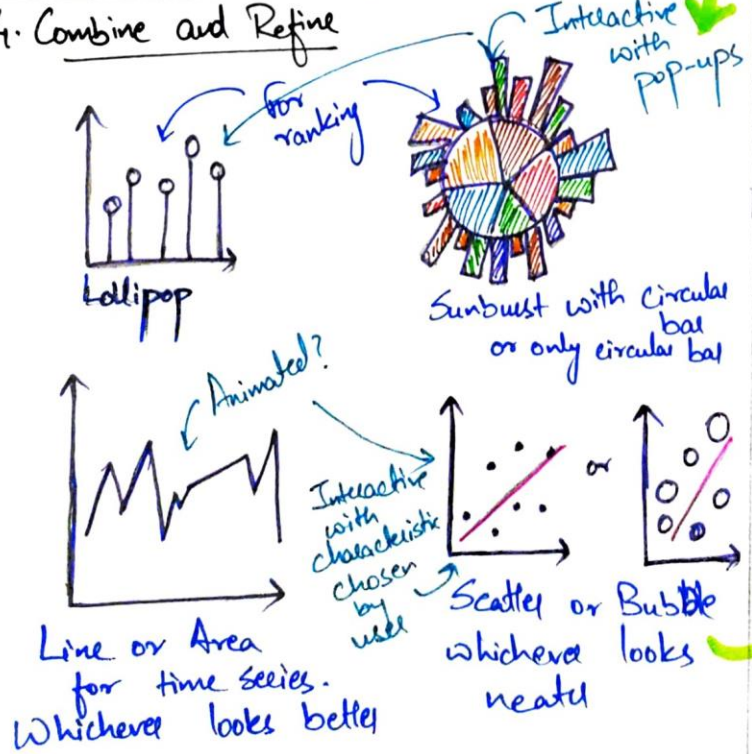
→ 3 & 9 for tracks streamed over the months. 9 would make the density more prominent

→ 7 & 8 for analysing track characteristic of common tracks

3 Categorize



4. Combine and Refine



5 Questions

- Which plots would compliment each other and give out the max. information?
- How to use colour and animation to make it more visually appealing?

Title : DVP

Author : Rishiraj Sinharay

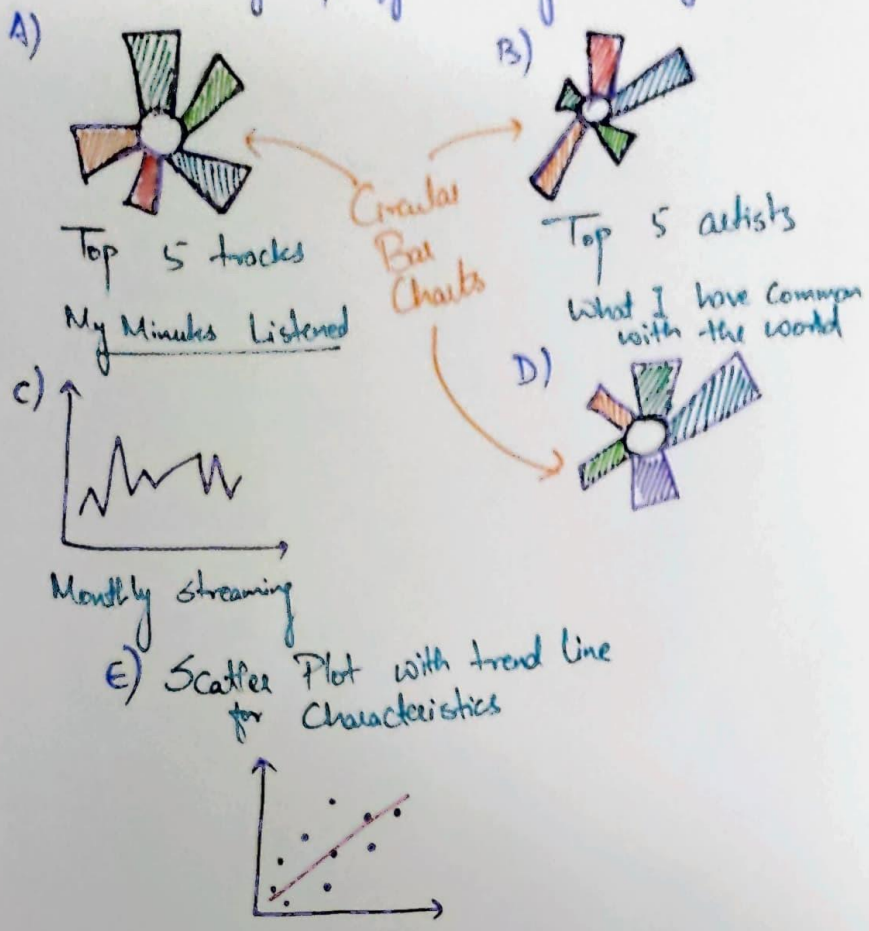
Date : 10/10/2022

Sheet : 2

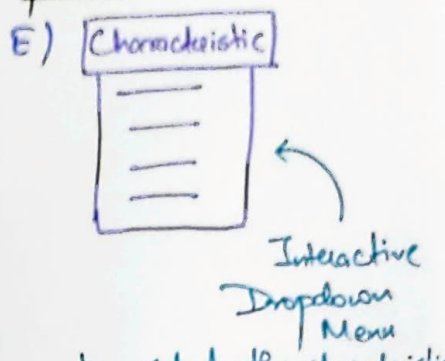
Task : Spotify Summary

Layout

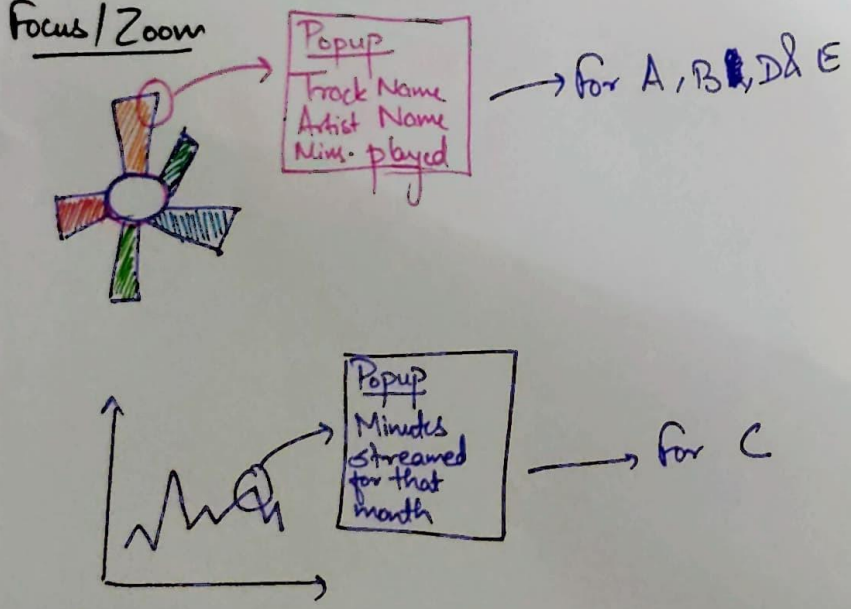
"My Spotify Streaming Summary"



Operations



Focus / Zoom



Discuss

Pros

- Gives the required information

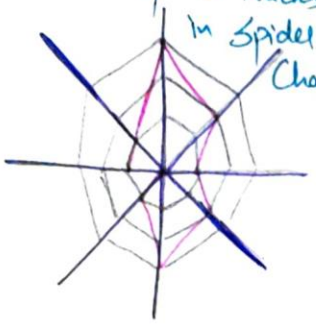
Cons

- Very basic
- No animation
- Not very visually appealing

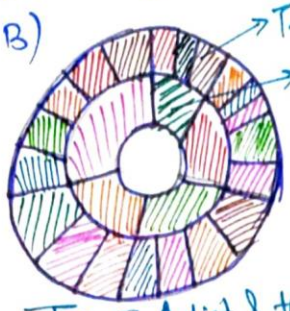
Layout

"My Spotify Summary"

A) Top 5 Tracks in Spider Chart

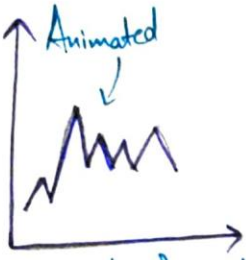


B) Tracks, Artists, Sunburst



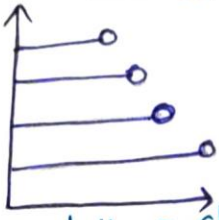
Top 5 Artist & their tracks streamed

C)



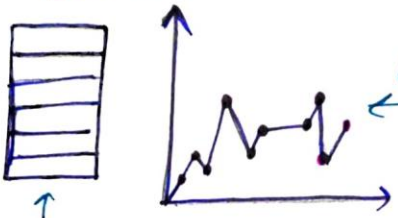
Streaming minutes & no. of streams over time

D)



Lollipop Chart showing Top 5 Common tracks

E)



Buttons for choosing characteristic to be plotted.

Title : DVP

Author : Rishiraj Sinharay

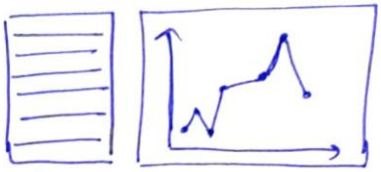
Date : 10/10/2022

Sheet : 3

Task : Spotify Summary

Operations

E)

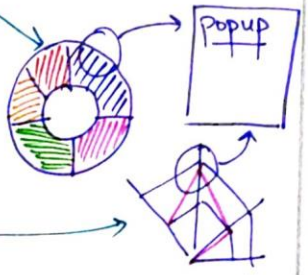


List of characteristics. Clicking one will give the corresponding graph

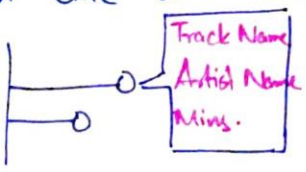
Focus / Zoom

B) → Hovering & clicking over each track and artist will give popup details like track name / artist name and minutes streamed

A) → Each point will give track name, artist name & minutes streamed.



D) → Each line will show popup details



E) → Each point will give the track name, artist name & characteristic value

Discussion

Pros

- Animated in C would look visually appealing

Cons

- A & B should look same as they are same type of ranking.
- Too many diff. graphs.
- The sunburst when showing the track names for an artist should have diff. sizes for minutes streamed

Title: DVP

Author: Rishiraj Sinharay

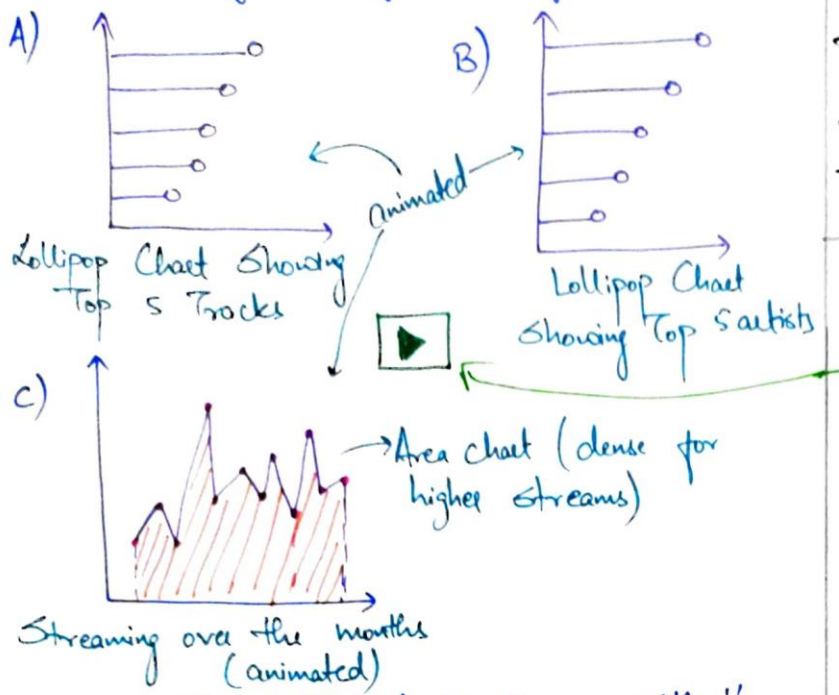
Date: 10/10/2022

Sheet: 4

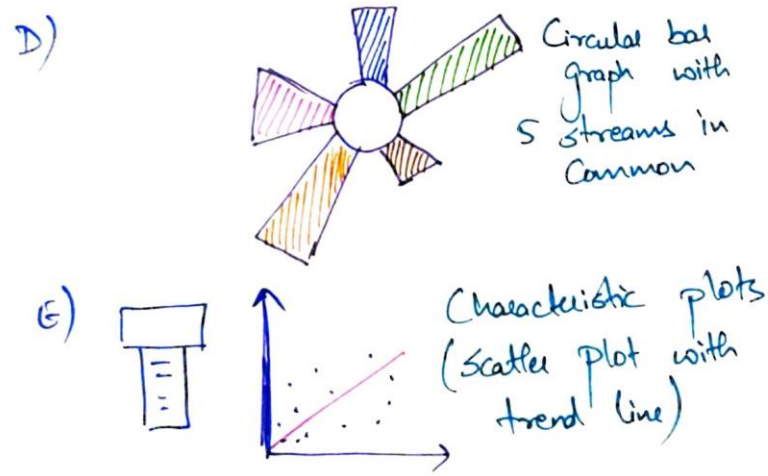
Task: Spotify Summary

Layout

"My Spotify Summary"



"What I had in Common with" the World



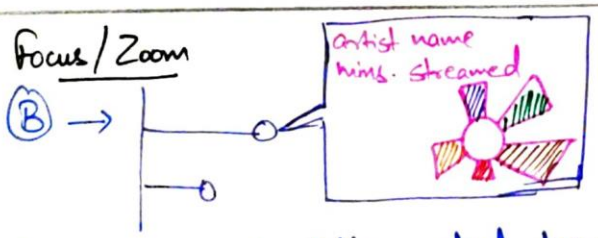
Operation

→ Will show animated monthly progress in C and top streams changing monthly in A & B.

E) → Dropdown to select the Characteristic.

Will plot the scatter/bubble plot with trend line.

Focus / Zoom



Clicking on each lollipop chart for artists will show a pop-up for all the tracks streamed of that artist (circular bar). Also show details like minutes streamed, track name, artist name.

A, C, D → Clicking on points will give popup details on minutes streamed and track & artist names.

Discussion

Pros

- Visually appealing because of the 3 animations
- Very informative

Cons

- Not sure between scatter / bubble plot for E.

Title : DVP
Author : Rishiraj Sivaraman
Date : 10/10/2022
Sheet : 5
Task : Spotify Summary

Layout

"My Spotify Summary"

A) Lollipop chart showing Top 5 tracks

B) Lollipop chart showing Top 5 Artists

Animated

C) → Area chart (denser for higher streams)

Streaming over the months

"What I had Common with the World"

D) Circular bar graph with 5 streams in common

E) Characteristic plots (scatter plot with trend line)

Operation

Play button → Will show animated monthly progress in C & top streams changing monthly in A & B.

E) → Dropdown to select the characteristic to plot.

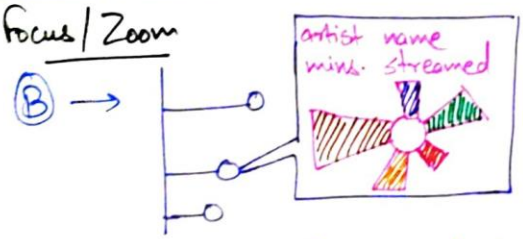
Will plot the corresponding scatter plot with trend line.

Detail

→ Animations show variations over time.

→ D and E depends on the intersection of 2 datasets to find common songs/tracks. Less data in common would give ~~less~~ visualisations which are not very accurate

Focus / Zoom



Clicking on each lollipop chart for artists will show a popup for all the tracks streamed of the artist (circular bar). Also show details like minutes streamed, track name, artist name.

A, C, D → Clicking on points will give pop-ups with details on minutes streamed & track & artist names.