

Process Book

Iris Yan, Lauren Wattendorf, Cindy Liu

Week 7 - Project Proposal

Project Title:

Seeing Sound - Visualizing Spotify Statistics

Project Abstract:

We will examine listener and music trends on Spotify using [Spotify Research's data](#) on playlists and streaming sessions. This project is motivated by our shared love for curating playlists and exploring new music, most often fueled by Spotify's database of songs, recommendations, and features such as "Spotify Wrapped" and its analysis of music preferences. In our project, we will be analyzing both trends on an individual level, e.g. visualizing common music genres and themes, a dashboard of minutes spent, top artists/songs, a color that corresponds to the overall music portfolio of a user, etc., and an interpersonal level, e.g., visualizing how these previously discussed individual attributes lend way to similarities and differences between multiple users.

Week 8 - Detailed Project Plan & Team Agreement

Project Title:

Seeing Sound - Visualizing Spotify Statistics

Team Members:

Lauren Wattendorf - lwattendorf@college.harvard.edu

Cindy Liu - cindyliu@college.harvard.edu

Iris Yan - irisyan@college.harvard.edu

Background and Motivation

Upon brainstorming project ideas, we discovered that our team had a collective passion for listening to music and curating playlists on Spotify. We talked about the excitement our friends and peers feel when Spotify Unwrapped is released each year. Due to the strong sense of identity music gives our generation, we believe there is an intrinsically gratifying experience to viewing a personalized review of your favorite music, listening habits, and music preferences.

We are motivated to create a personalized dashboard for Spotify users to interact with their favorite songs, albums, and genres while exploring graphs and figures of their own Spotify data. Additionally, we hope to provide a social aspect, where users can compare their data to a small group of friends or public Spotify data. We believe that this will provide a fun and rewarding experience for users to learn more about their musical identity and interests.

Related Work

We drew inspiration from our love of music (I love reading Rolling Stones articles), Spotify, and sharing music through social media platforms. Additionally, we learned more about the ideation of Spotify Unwrapped and why it is so successful:

[Spotify Unwrapped: How we brought you a decade of data](#)

- Spotify's blog on creating the internet famous "Spotify Unwrapped" feature for their users

[Spotify Wrapped, unwrapped](#)

- Article by Vox discussing the pop-culture relevance of Spotify Unwrapped

Audience and Questions

Audience: Our audience will be frequent Spotify users and music fans. We think our visualization will be interesting for those just getting into their music journey and die-hard fans who have a plethora of Spotify playlists to dissect and engage with.

Goals and Objectives: As outlined above, our goal is to bring a fun, personalized visualization that guides users through their Spotify journey. Additionally, we want to have the ability for multiple Spotify users (realistically 2-3) to compare their music preferences. The main features of our data visualization will be to view your own data (most listened to albums, artists, etc / most active time of day) as well as to be able to compare your data to other users. Some potential questions we want to answer include:

1. How do my listening preferences compare to my friends?
2. What songs/albums/artists do we share in common?
3. What are my (+ versus my friends) listening patterns? Do I listen to a certain type of music during a certain type of day? Did I have a favorite band during a certain month of this year? Etc.

Lastly, leveraging both our individual and publicly available data, we will also offer users the ability to compare their listening behaviors with aggregated public data as a proxy for many other Spotify users as a whole. We will look to answer questions such as:

1. How do my listening habits (e.g., time spent listening, number of songs, skips, etc.) compare to general Spotify users?
2. For my top artists, tracks, or genres, how do my habits compare with others who have these similar music preferences?
3. How do the songs I enjoy compare to generally popular or frequently listened to songs?

Data

We will be collecting our Spotify from two primary sources: 1) personal account data, requested from Spotify (relevant fields for us are listed below), and 2) public Spotify research datasets, particularly on playlist curation and streaming session behavior, linked [here](#). More details of relevant fields follow below, and preliminary steps in our data processing are included in the next section.

Account Data

- Playlists
- Streaming history for the past year
- A list of items saved in your library
- Search queries
- No. of followers, accounts you follow, and blocked accounts
- User data
- Inferences
- Spotify for Artists data

Public Spotify Dataset: Million Playlist Dataset

- Track list (ID's, metadata per track, e.g., artist name, song title, song length)
- Edit time
- Number of edits
- Collaboration

Public Spotify Dataset: Music Streaming Sessions Dataset

- Track list (ID's, duration, etc.)
- Acoustic attributes (beat strength, bounciness, energy, loudness, tempo, etc.)

Data Cleanup

We will not be required to do substantial data cleanup, however we will have to use Python to parse the 'StreamingHistory' json files from our Spotify data request. We will create a .csv with the relevant data entries (discussed above) that we will be able to use for our Tableau visualizations. Some specific quantities we plan to derive from the data include the title and artist of each song played, the number of minutes we've streamed the song, and the album the song belongs to. We will also create an unique identifier for each song, so we are able to easily visualize relationships between listening history among the three members in our group. We will use the Python library Pandas to create dataframes, which we can then save as csv files to load into Tableau for initial data exploration.

With regard to the public data obtained from Spotify, it should also require minimal cleanup. It can be downloaded as a JSON file that can be filtered and wrangled to suit our visualization purposes. Some specific quantities that are given in the data that would be useful are unique song and playlist identifiers, metadata on both the playlist level (e.g. number of edits,

length, number of songs, number of albums) and on the song level (e.g., track ID, song length, artist name, song title, etc.). Similar to the above, we can use these either directly to create dataframes and .csv files, or as JSON data for our JavaScript processing to create interactive data visualizations.

Team Agreement Document

- We will communicate via our iMessage group chat, and have weekly in-person team syncs on Monday afternoons at the SEC
- All code should be written individually, but collaborated on during in-person meetings and via a Git workflow
- We will split up work with actionable deliverables at the end of each team meeting to be finished by the next meeting. All team members will agree upon division of work to ensure a work is shared as equitably as possible
- When issues or emergencies arise, we will use iMessage group chat to inform other group members as early as possible
- Final design decisions will be discussed among all members; fair compromises should be made when necessary
- Any disagreements should be mediated either individually if possible, or among the entire group if necessary, as soon as possible
- A week-by-week rough timeline will be created and adhered to as closely as possible to ensure the project is spaced out evenly throughout the weeks leading up to the due date

Signatures:

Iris Yan, Lauren Wattendorf, Cindy Liu

Date: 10/23/2022

Week 9 - Map

Possible Target Audiences:

1. Businesses interested in marketing to or learning more about music listening habits of a particular audience or generation, e.g., if their business model centers around providing music recommendations, concert advertising, etc.
2. Avid Spotify users (like ourselves) who want to learn more about their listening habits and the relationships between their listening and their friends, or with the Spotify clientele in general
3. Producers and songwriters curious about what features are most common among popular music in order to better inform their work and understand music tastes

Going with target audience #2, avid Spotify users, we can assume strong knowledge of what kind of features and data Spotify has, but not necessarily high levels of visualization literacy. For these reasons, we will look to present detailed and interesting data to our audience in an easily palatable visualization format, using visualization types that are easy to understand and that together convey an intuitive, easy-to-follow narrative. This audience will be most interested in visualizing trends in Spotify usage regarding type of music, popularity, frequency of listening, and streaming/saving habits in Spotify. Thus, we will focus our data visualizations around these more detailed, individualized usage habits (versus a more aggregate focus for some of our other possible audiences), contextualizing them with our aggregate Spotify data at large as well to see how they measure up.

Questions for our Target Audience

1. How many playlists do Spotify users have? How do I compare?
2. How often do I listen to Spotify compared to the general population?
3. What are the most listened to albums/songs/playlists?
4. What does the normal Spotify user's playlist look like? E.g., how long are songs and how many songs are there?
5. What are the most popular artists that are listened to?
6. What is my most unique album/artist/song I listen to (i.e., least common among other listeners)?
7. What genre's do breakout songs usually fall within?
8. What genres are growing in popularity? What genres are decreasing in popularity?
9. When do I most often listen to music? When was the peak in my listening habits?
10. How have my favorite genres changed over time?
11. What are the artistic qualities of the music I listen to (ie acousticness, tempo, song length, etc)
12. How have the artistic qualities of the music I listened to changed over time? For example, do I listen to more upbeat music now than I used to listen?

Data Exploration

Below, we list the main attributes that are present in each of our data sources and provide a brief description, categorizing by data type.

Categorical	Ordinal	Quantitative
1) Playlists – playlists you've listened to	1) Pos - position in playlist (index in playlist)	1) Follower data – number of followers, accounts you follow, and blocked accounts
2) Streaming history for the past year – all songs streamed in the past year	2) Playlist Index - unique identifier for each track, based on order of input	2) Duration of song
3) A list of items saved in	3) Track Index - unique identifier for each track,	3) Number of tracks per playlist

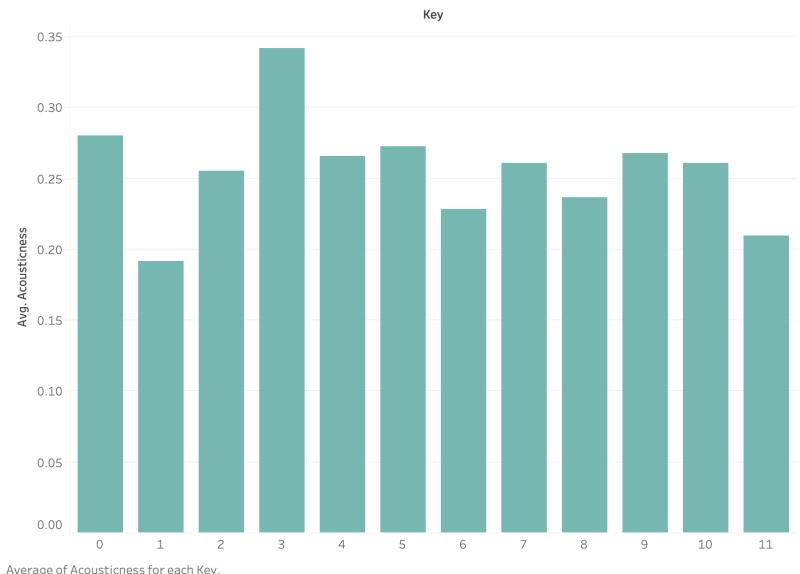
<p>your library – saved songs</p> <p>4) Search queries</p> <p>5) Album name - album for tracks included in playlists</p> <p>6) Artist name - artist name for tracks included in playlists</p> <p>7) Track name - song name for tracks included in playlists</p> <p>8) Names of accounts you follow & accounts that follow you</p> <p>9) Playlist name - Names of playlists, both for general data & personal data</p> <p>10) Saved items - Names of items saved in your library</p> <p>11) Streaming history - Names of tracks played throughout the past year</p>	<p>based on order of input</p> <p>4) Key – main key of the track, where pitches are mapped to integers using standard Pitch Class notation</p> <p>5) Time signature – notional convention to specify how many beats are in each bar (or measure). The time signature ranges from 3 to 7 indicating 3/4 to 7/4 time</p>	<p>4) Popularity (measured from 0 to 100) – based on the number of plays</p> <p>5) Dance-ability (0.0 to 1.0) – suitability for dancing based on a combination of musical elements</p> <p>6) Energy (0.0 to 1.0) – Perceptual measure of intensity and activity</p> <p>7) Loudness – loudness of a track in decibels</p> <p>8) Speechiness (0.0 to 1.0) – presence of spoken words in a track</p> <p>9) Liveness (0.0 to 1.0) – presence of an audience in the recording</p> <p>10) Valence (0.0 to 1.0) – musical positiveness</p> <p>11) Tempo – overall speed or pace of a track in beats per minute</p>
--	--	---

Individual Visualizations

Cindy:

Average Acousticness by Key

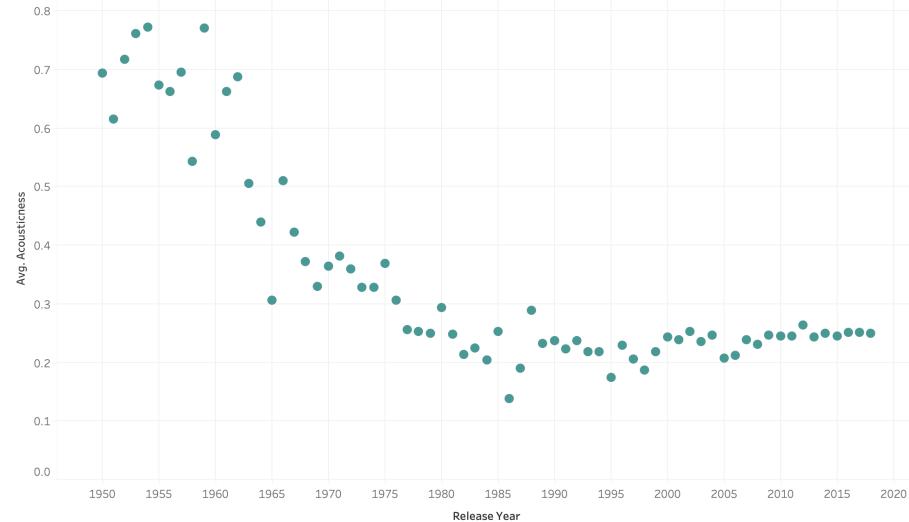
Pitches are mapped to integers using standard Pitch Class notation. E.g. 0 = C, 1 = C#, etc.



Average of Acousticness for each Key.

Cindy:

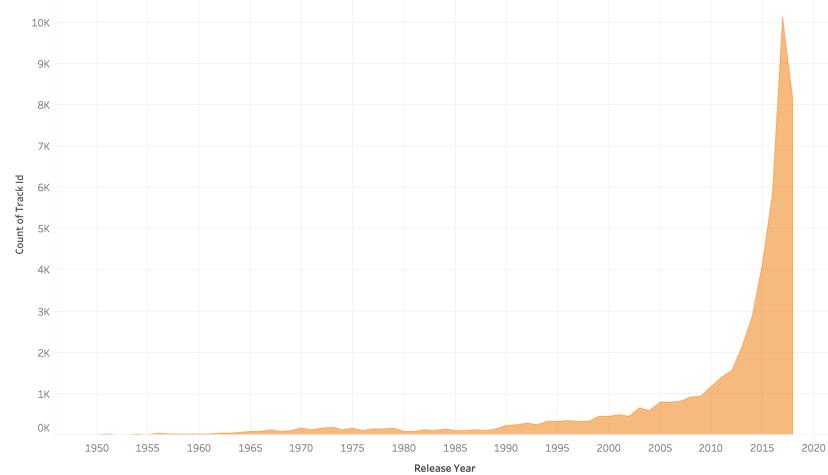
Average Acousticness by Release Year



The plot of average of Acousticness for Release Year.

Cindy:

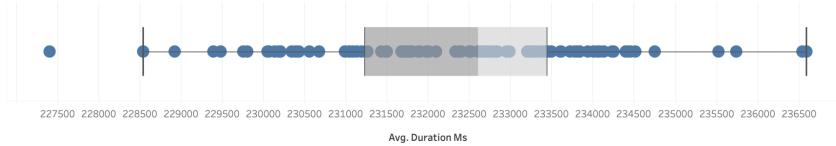
Number of Tracks Released Over Time



The plot of count of Track Id for Release Year.

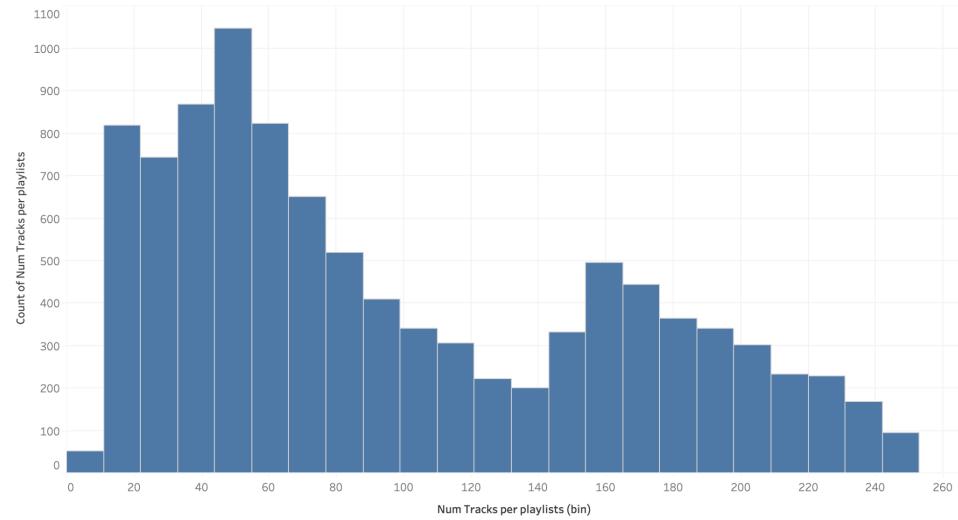
Iris:

Average Song Duration is Around 4 Min.



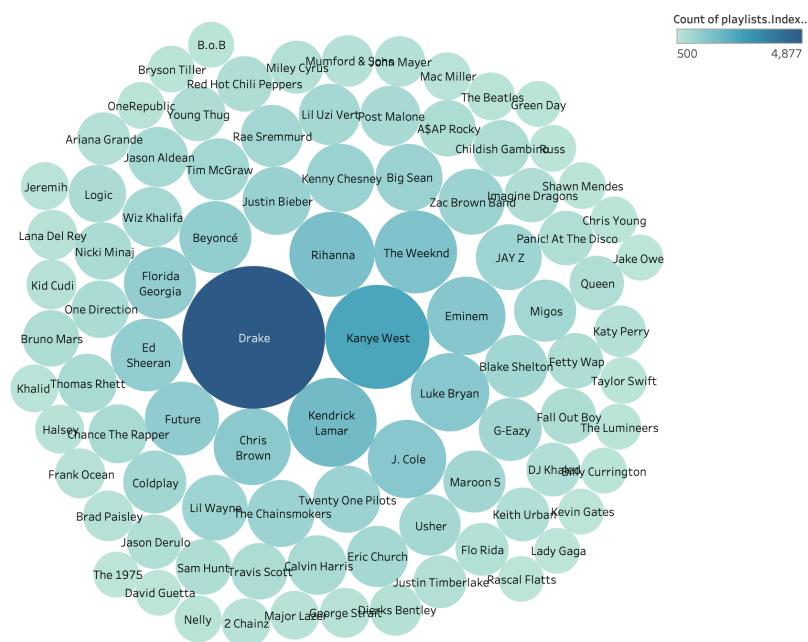
Iris:

The Number of Tracks per Playlist is Bimodal, Peaking at 50 Tracks



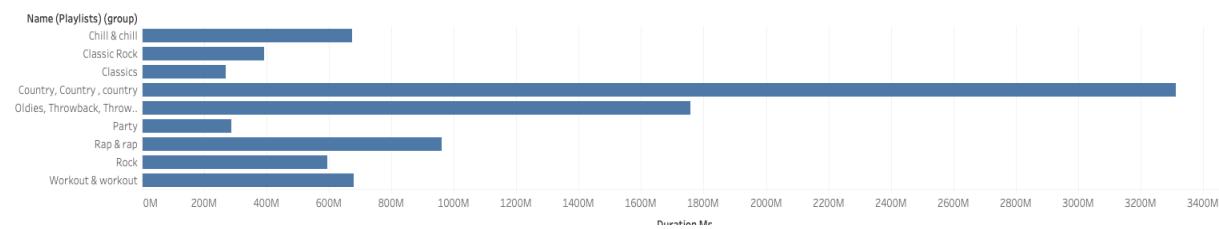
Iris:

Top Artists include Drake & Kanye West



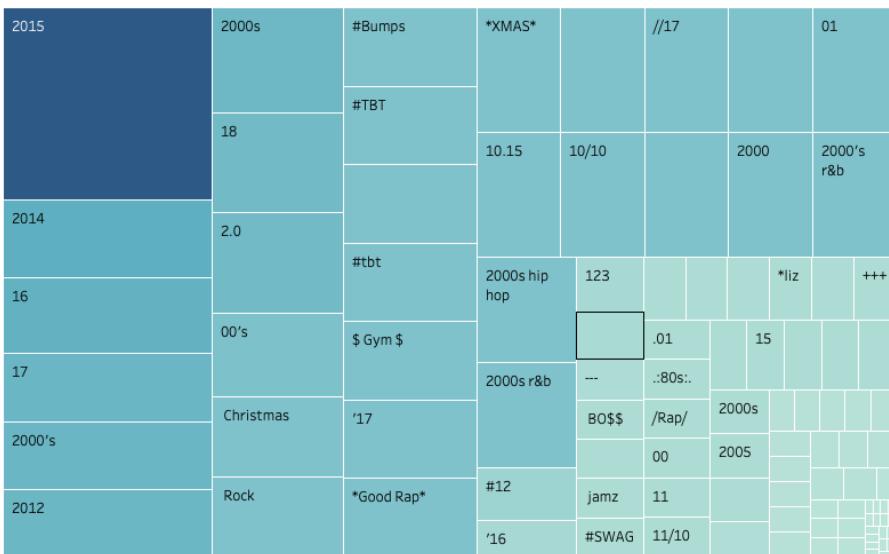
Lauren:

Average Playlist Duration per Top Playlist name



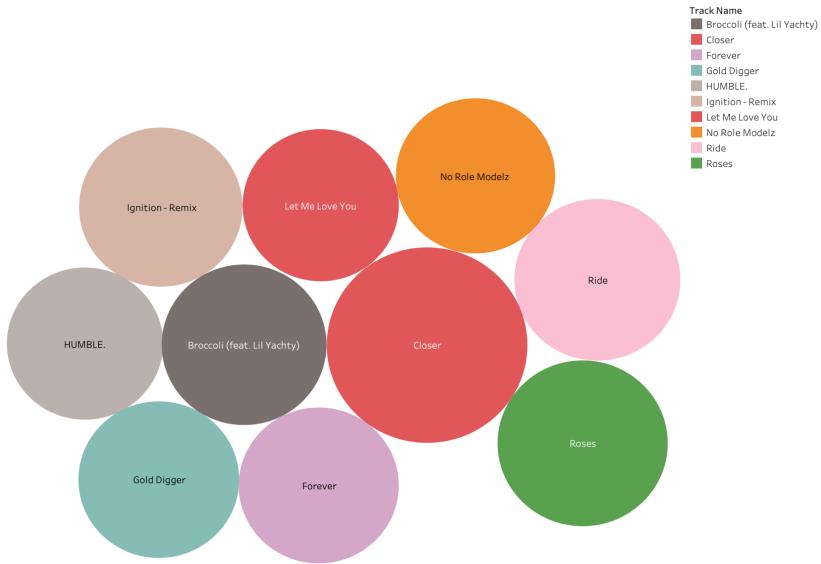
Lauren:

Top Playlist Names: Majority of people name their playlists after the year



Lauren:

Top 10 Songs added to Playlists



Track Name. Color shows details about Track Name. Size shows count of Pid. The marks are labeled by Track Name. The view is filtered on Track Name, which keeps 10 of 53,955 members.

Discussion:

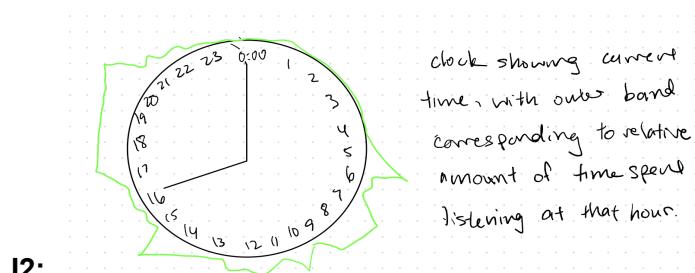
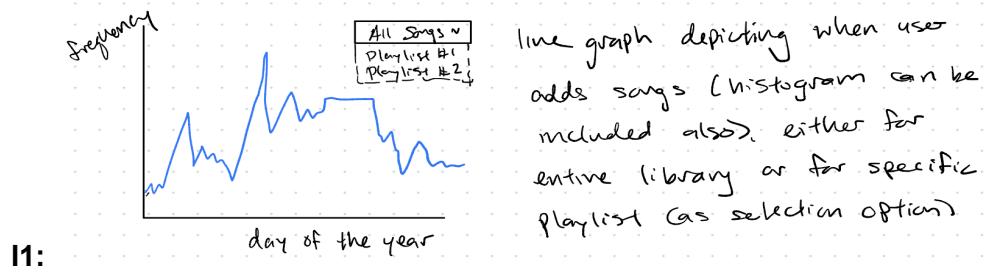
Our current visualizations describe various trends in aggregate Spotify data, including top artists, average track duration, top playlist names, etc.. We also analyze how the artistic qualities of music (e.g. acousticness, tempo, song length, etc) has changed over time, and how it varies based on other musical attributes such as the key of a piece. We will be able to answer

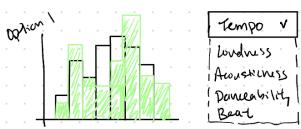
most of the questions we posed above, however we are lacking some key features for others, e.g., none of our datasets have had genre (as was previously expected), so it will be difficult to analyze this directly. Additionally, we are still waiting for our personal Spotify data pull requests (roughly 5 days), so we can compare our personal preferences with general Spotify listening preferences. Our initial data mapping visualized general listening habits, which we will enhance with personalized listening data at the individual level. The most interesting questions are those that compare the individual to the aggregate whole; however, these are also the more complicated ones to answer. We currently have the aggregate data, but will need to receive our individual Spotify data before making more headway in answering those ones. Some of the more preliminary questions about overall listening habits (e.g., playlist length and naming) are informative but may not be as engaging for our target audience, but these are answerable among the above visualizations and current data we have.

Week 10 - Data, Sketches, Decide & Storyboard

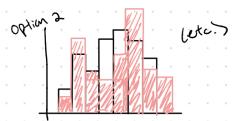
Sketch step:

Iris's Sketches:





drop down selection of different attributes compared between your music and the aggregate whole



Card - comparison of metrics, with option to explore top songs/artists in more detail

YOU

- 100 Hours listened
- 10 playlists
- 450 songs
- Top Artist: >
- Top Song: >

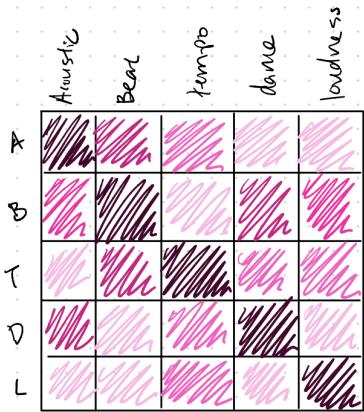
VS. WORLD

- 60 hours listened
- 12 playlists
- 506 songs
- Top Artist: >
- Top Song: >

TOP 5

- 1.
- 2.
- 3.
- 4.
- 5.

I3 & I4:

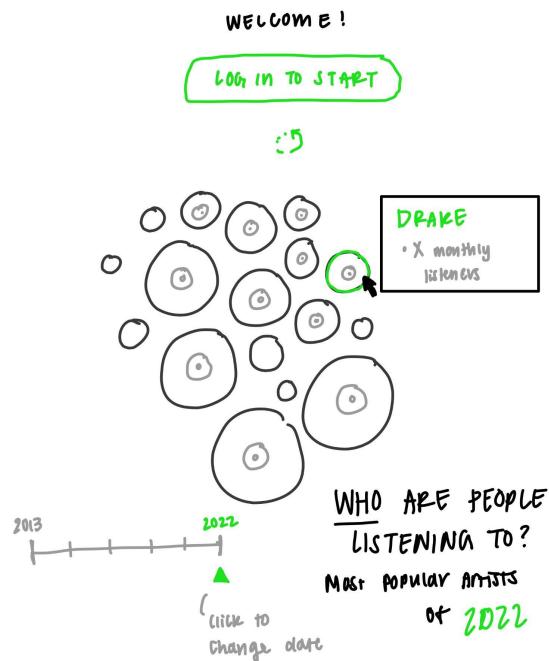


Heatmap of correlations between features in aggregate data, showing which one most related to each other.

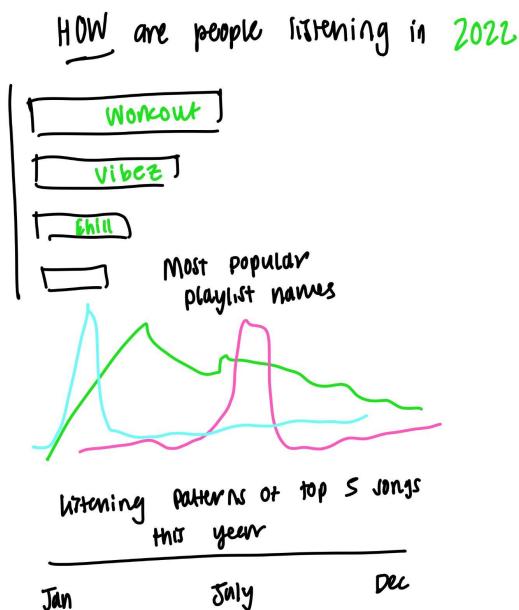
I5:

Lauren's Sketches:

L1

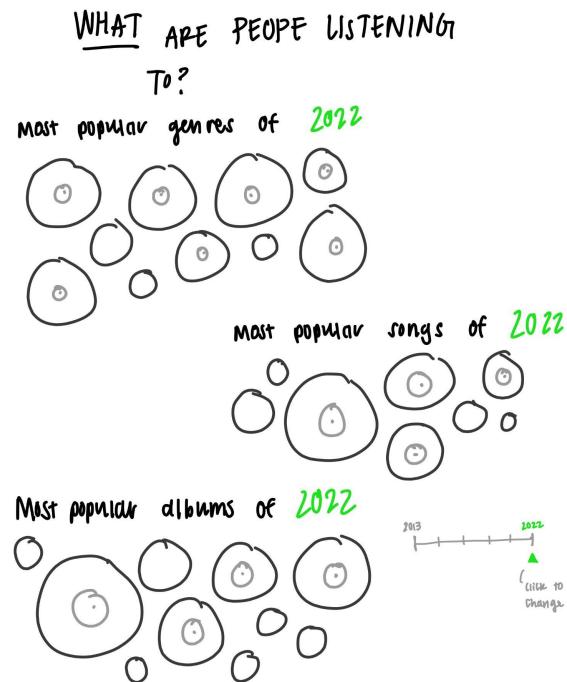


L2&3:

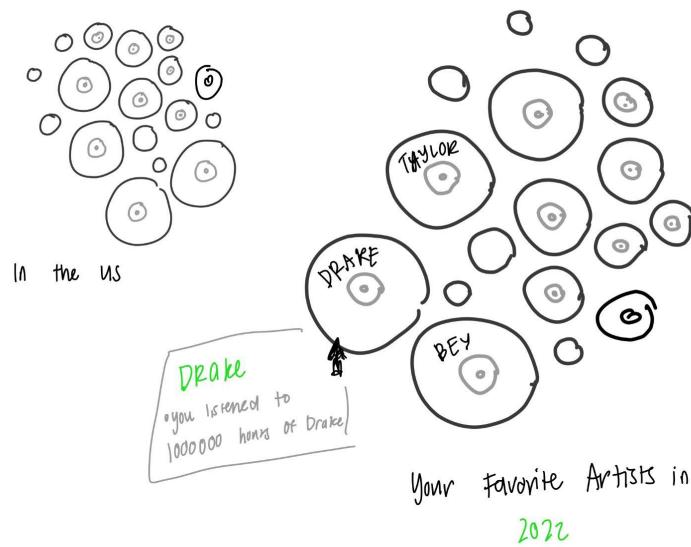


HOW DO YOU COMPARE

L4,5,6:



L7:



etc...

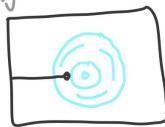
repeat all the aggregate graphs
w/ a small version next to
the personalized version

L8:

Creative illustration representing the acoustic qualities of your top song

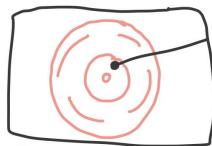


Beautiful
By Christina Aguilera



Drivers License
by Olivia Rodrigo

circles animate
by heating to
the tempo of
the song. The
size of the circle

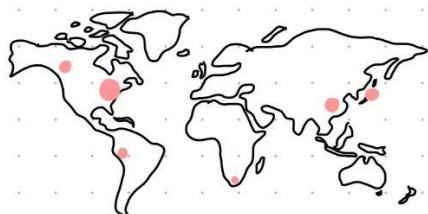


is based on loudness
and the colors is based
on dancability.

Cindy's Sketches:

C1:

Where in the world are your top artists from?



or

V2: interactive/scrollable globe

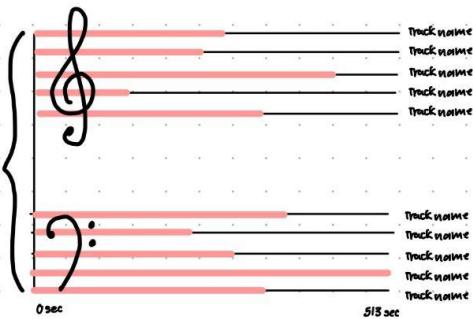


Legend:

- Map or globe to visualize geography/countries
- Pink dot (scaled to represent # artists) - from a certain country/region

C2:

How long are your top 10 songs?

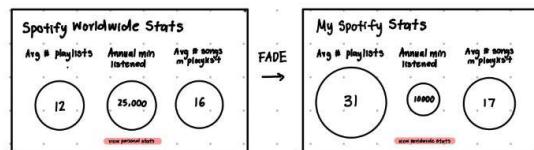


Legend:

- Using treble and bass clef to visualize length of user's top 10 songs
- Pink bars (overlaid on each line of the Staff) are scaled on track length, with "x-axis" label at bottom showing range of track length

C3:

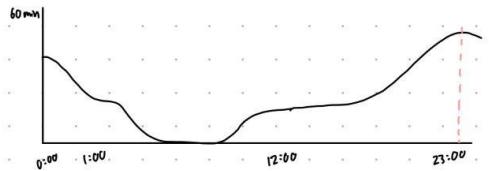
How do I compare?



Legend/description: use fade in/out circles to compare personal Spotify stats (e.g. # playlists, annual min listened, # songs) to aggregate worldwide listener stats.
Toggle buttons to change between personal + worldwide stats.

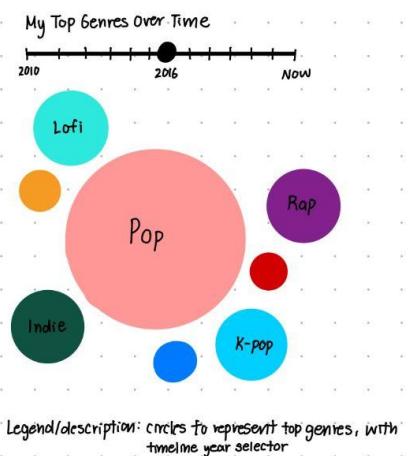
C4:

When do I usually listen to music?



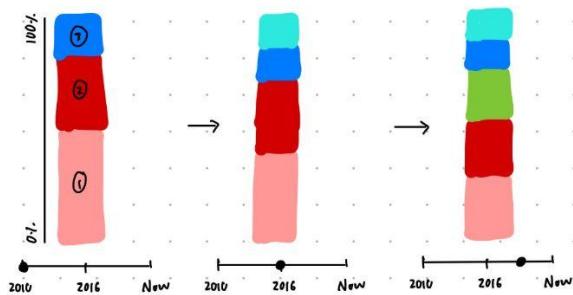
Legend/description: line graph showing trend of listening time and peak listening time

C5:



C6:

How are my songs distributed over playlists?



Decision Step:

Question IDs:

1. What are the most popular genres over time? How do my favorite genres compare?
2. How many playlists do Spotify users have? How do I compare?
3. How often do I listen to Spotify compared to the general population?
4. What are the most listened to albums/songs/playlists?
5. What does the normal Spotify user's playlist look like? E.g., how long are songs and how many songs are there?
6. What are the most popular artists that are listened to?
7. What is my most unique album/artist/song I listen to (i.e., least common among other listeners)?
8. What genre's do breakout songs usually fall within?

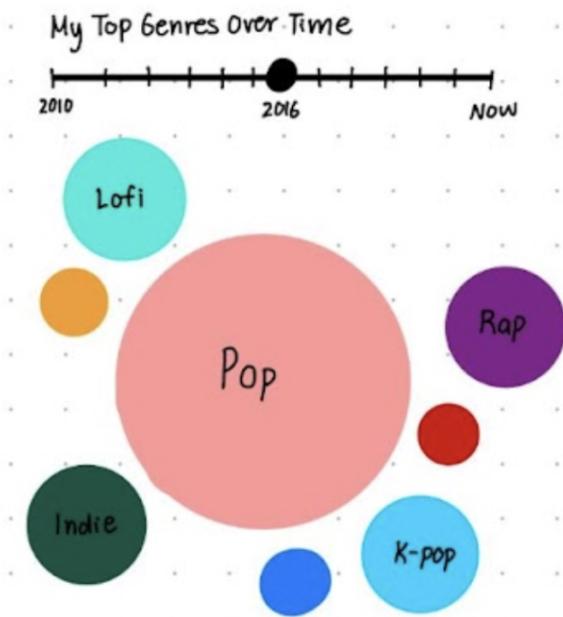
9. What genres are growing in popularity? What genres are decreasing in popularity?
10. When do I most often listen to music? When was the peak in my listening habits?
11. How have my favorite genres changed over time?
12. What are the artistic qualities of the music I listen to (ie acousticness, tempo, song length, etc)? How do artistic qualities relate to each other more broadly?
13. How have the artistic qualities of the music I listened to changed over time? For example, do I listen to more upbeat music now than I used to listen?
14. What regions/countries are my top artists from?
15. What are my playlist-building habits? I.e., when do I most often add music to playlists? How are my songs distributed among playlists?
16. When were the top songs of this year most popular? What was the listening trends of these songs?
- 17.

In this step, you will work with your group to **decide** which sketches to implement in D3 during the prototype phase. You should pick between 4-6 of your sketched ideas.

Affinity Mapping:

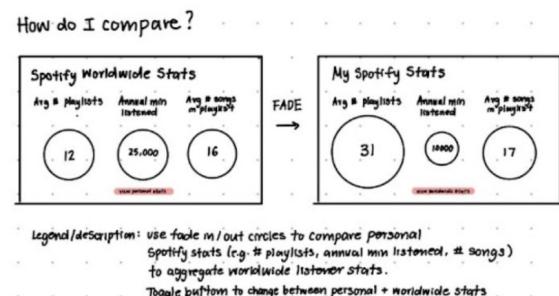
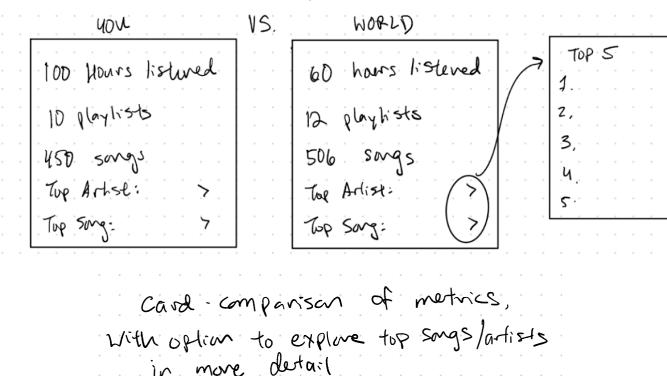
Sketch ID	Question ID	Author	Total Votes (discussed together)
C5, L4	1	CL & LW	3
I4, C3	2	IY & CL	
I4, I2	3	IY	
L7, L4, L5, L6, I4	4	IY & LW	
C2, C3, I4, L2, L3	5	CL & LW & IY	3
L1, L7, I4	6	LW & IY	
C4, I2	10	CL & IY	
C5	11	CL	3
I3, I5, L8	12	IY	3
C1	14	CL	
I1, C6	15	IY & CL	3
L3	16	LW	

Final Sketches (5 included here, descriptions below screenshots):

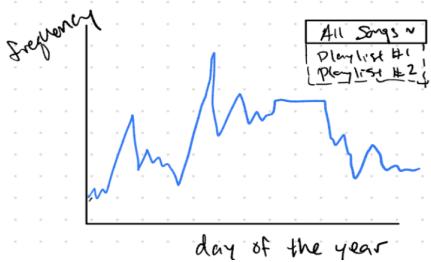


Legend/description: circles to represent top genres, with timeline year selector

1. Our favorite graph (which both Lauren and Cindy drew) were Bubble Graphs for top 10 genres, songs, albums, artists over time. We will include this as one of the main features of our data visualization.



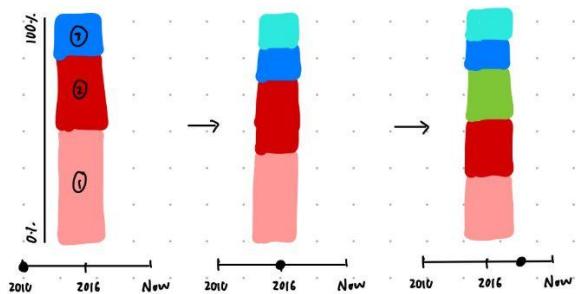
2. We decide to combine the 2 visualizations above to compare global listener stats with personal listener stats (e.g. time listened, number of playlists, number of songs per playlist, etc.). This will be visualized in data cards showing main statistics in a clear coherent way, highlighting the numbers via the size of the circles.



line graph depicting when user adds songs (histogram can be included also), either for entire library or for specific playlist (as selection option)

3. Line graphs that depict how frequently users listen over time with time on the X scale. Additionally, we would make the x scale interactive so that it can be adjusted from days of the year, to days in a month, to hours in a day.

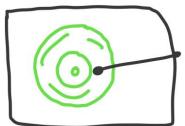
How are my songs distributed over playlists?



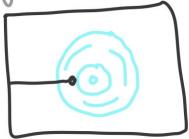
Legend/description: stacked bars to show concentration of songs in your playlists over time

4. Our fourth visualization will be a stacked bar chart visualizing the proportion of songs in each playlist (with each segment denoting a separate playlist). A slider over time indicates how these proportions change as new songs and playlists are added.

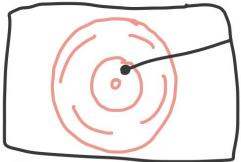
Creative illustration representing the acoustic qualities of your top song



Beautiful
By Christina Aguilera



Driver's License
by Olivia Rodrigo



Power
by Kanye West

Circles animate by beating to the tempo of the song. The size of the circle is based on loudness and the colors is based on danceability.

5. Our last visualization will be an image of a record with an animation corresponding to the different musical attributes of the song. For example, the frequency at which the record pulses will correspond to the beat of the song, and the color may correspond to the danceability, or the size of the pulses (i.e. how big or small the record gets) to the loudness, etc. The user will be able to toggle between their top songs to visualize these differences.

Rationale:

When deciding which of the sketches we would like to implement, we considered what would be a natural progression of a data story to tell the user. We decided that it would be most natural for a user to have the context of more global, aggregated data of “all spotify users” to then fit their listening data into that context and be able to notice what attributes of their listening history/preferences make them unique. To do this, we decided we first needed a clear but simple graphical representation of the major statistics we expect users to be wondering about (genre, artists, tracks). Two of us came up with the bubble graph (Final Sketch #1), to represent this, and we agreed that this graph clearly shows the users “top” in the form of size, and enables the user to quickly make observations across the visualization on which genre/artist/tracks were biggest/“top”. Second, we wanted a visualization that distilled the essential statistics of user listening data and summarized it all in one place, which is why we chose Final Sketch #2. Then, we wanted graphs that would allow the user to explore their own listening habits, which led us to Final Sketch #3 and #4, which are interactive visualizations for the user to explore their own personal data, and draw out conclusions on their own. Finally, we wanted a fun creative

visualization for the user to understand the artistic qualities of the music they listened to in a unique visual presentation. After brainstorming as a group, we settled on Final Sketch #5, which is an animated representation of the users top songs and using properties of the visualization to represent several different aspects of the artistic elements (temp, acoustics, etc). We believe these five visualizations accomplish our goals of summarizing key statistics and trends, enabling the user to quickly compare themselves to the general population, and providing unique exploratory experiences for the user.

Storyboarding:

Individual insights:

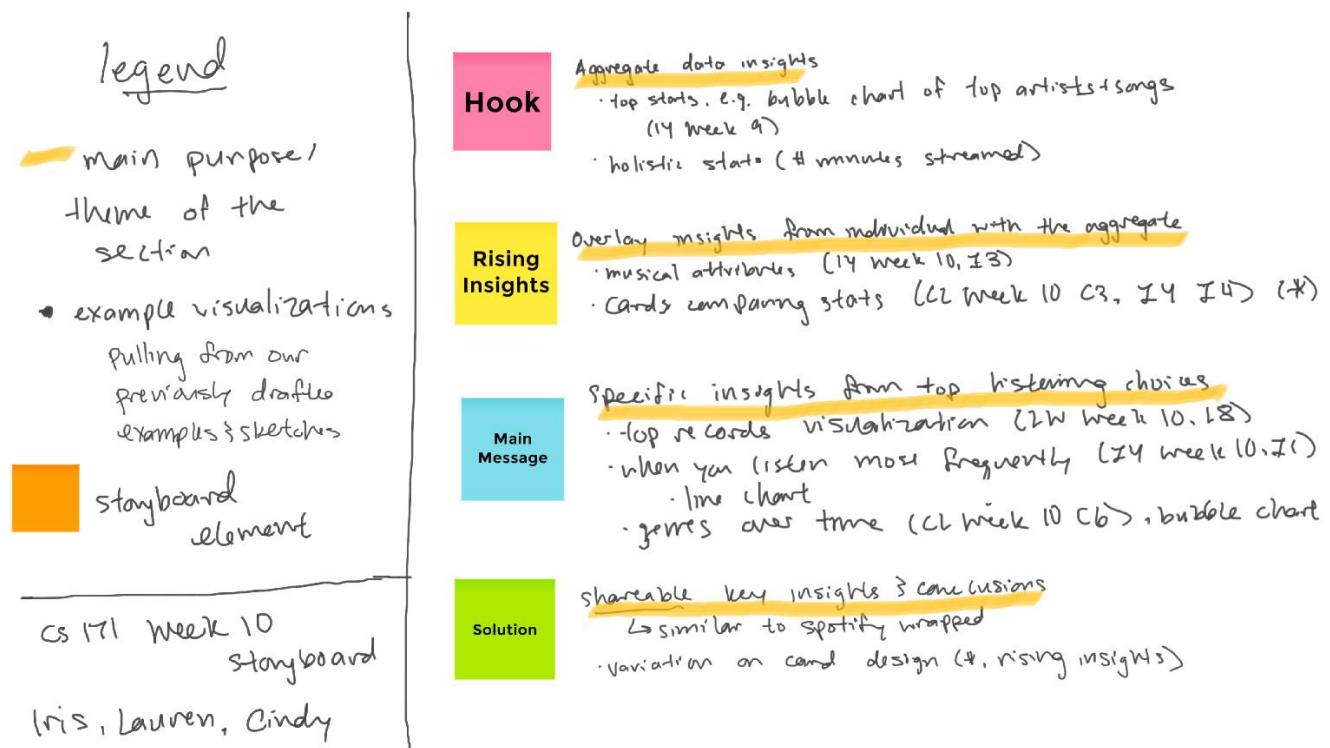
Iris	<ol style="list-style-type: none">There is a wide range of playlists lengths, with a peak around 50 songs and another around 160 songs; my playlists vary in size, with several around these peaks but also others that are much longer (~200 songs) and others that are much shorter (~20 songs).My main artists exhibit some amount of overlap with most popular artists saved to playlists, e.g. they also include Ed Sheeran, Rihanna, Taylor Swift, etc., but not all of my top artists, e.g. Sebastian Yatra, dhruv, etc. appear.
Cindy	<ol style="list-style-type: none">The average song duration is approximately 4 minutes, however the average duration of my top 10 songs is over 5 minutes and 20 secondsThe average acousticness of songs has decreased over time
Lauren	<ol style="list-style-type: none">The most common playlist names were based on the year the user created the playlist or the genre of musicMost of the top ten songs added to playlists were by rappers or in the rap genre

Main message:

After exploring our datasets for both aggregate Spotify users and individual listeners, we realized that we wanted our visualizations to be exploratory rather than persuasive. Our main message is that every individual listener's Spotify statistics / habits deviate from the norm in unique ways – whether it be genre choice, length of playlists, duration of top songs, or musicality (e.g. acousticness, danceability, loudness, etc.) of favorite songs. We want the visualizations on our app to help users explore how their own listening habits / choices compare to the broader Spotify audience. Having an open-ended main message will help us develop a captivating storyline that will bring people along the journey of self-exploration.

Storyboard Screenshot:

Note: Instead of sketching out visualizations here, we referenced visualizations we already created (initials, week, reference ID) from above that we believe would fit well into this narrative. As explained by the legend, each of the four storyboard elements has a main purpose/theme at the top, followed by some examples of visualizations that could follow in each of the sections. These would be presented sequentially (with a scrolling website) to the user.



Week 11 - Prototype V1

Process Overview

- Completed this week:
 - File structure, classes structure set up and outline of visualization website implemented in HTML
 - Playlist and listening data, both from individual and aggregate level, processed and used in visualizations (more notes below)
 - Three visualizations implemented, the line graph of how songs are added to playlists over time, (partially) the stacked bar of how songs are distributed

- across playlists over time, and cards depicting the statistics of an individual (Iris) compared to the aggregate group
- Draft of innovative visualization and the genre bubble chart over time completed with sketches and implementation notes
- To-do for next week:
 - Process data from musical attributes (from aggregate data set CSV and using Spotify Web API to find attribute data for individualized music tastes)
 - Implement interactions and dynamic changes on current visualizations, with some edits to currently implemented filtering and transitions
 - Specifically, we will add brushing and filtering for the line chart by playlist and over time
 - We will also work on making the stacked bar chart filtering by date of song added, not date of playlist update (current status) to make updates more granular
 - We will implement the “download card” functionality, as well as add additional statistics that may be interesting to the user
 - (Partially) implement 1-2 more visualizations, which are currently in the draft stage, and potentially others using the musical attribute data once it is processed
 - Add additional styling and storytelling elements for the current visualizations and website as a whole to increase cohesion

Contributors for Prototype V1: Lauren Wattendorf, Iris Yan, Cindy Liu

Notes on Cleaned Data:

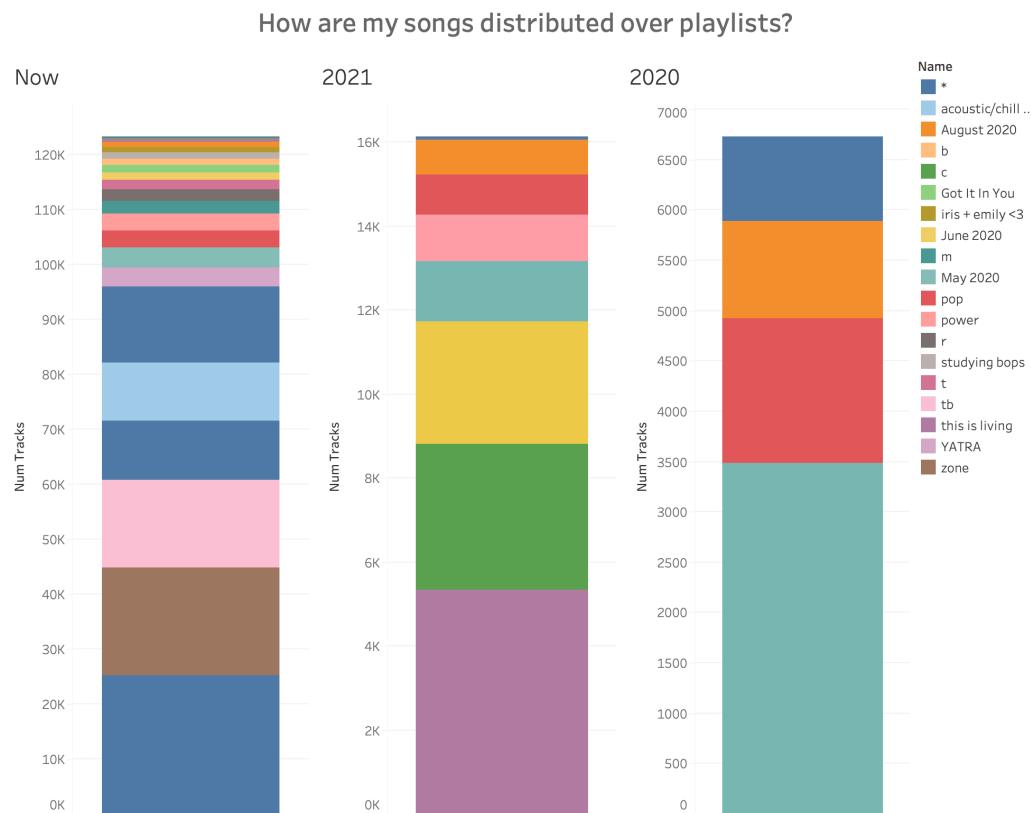
- The data processing this week was done on each of our individual datasets from Spotify - namely, our streaming history, our library of all songs saved, and our playlist data; these data were processed to contain only relevant fields (e.g. track names, lengths, URI's, etc.) and also for some overall statistics such as total number and length of playlists, duration of listening, and top artists and songs - from both aggregate and individual data.
- This processing allows us to compare the aggregate and individual statistics in parallel, as desired and outlined at the beginning of the project proposal
- The processing itself occurs in [initials]_dataprocessing.py and general_dataprocessing.py; this week ,the data that was used for the individual visualizations is Iris's (IY), so IY_dataprocessing.py is the relevant file.
- The cleaned data are in IY_all_songs.json, IY_playlist_songs.json, and IY_top_stats.json, for all song data, playlist data, and aggregate statistics respectively. The top statistics can be compared with the aggregate, in general_top_stats.json.

Implemented Visualizations

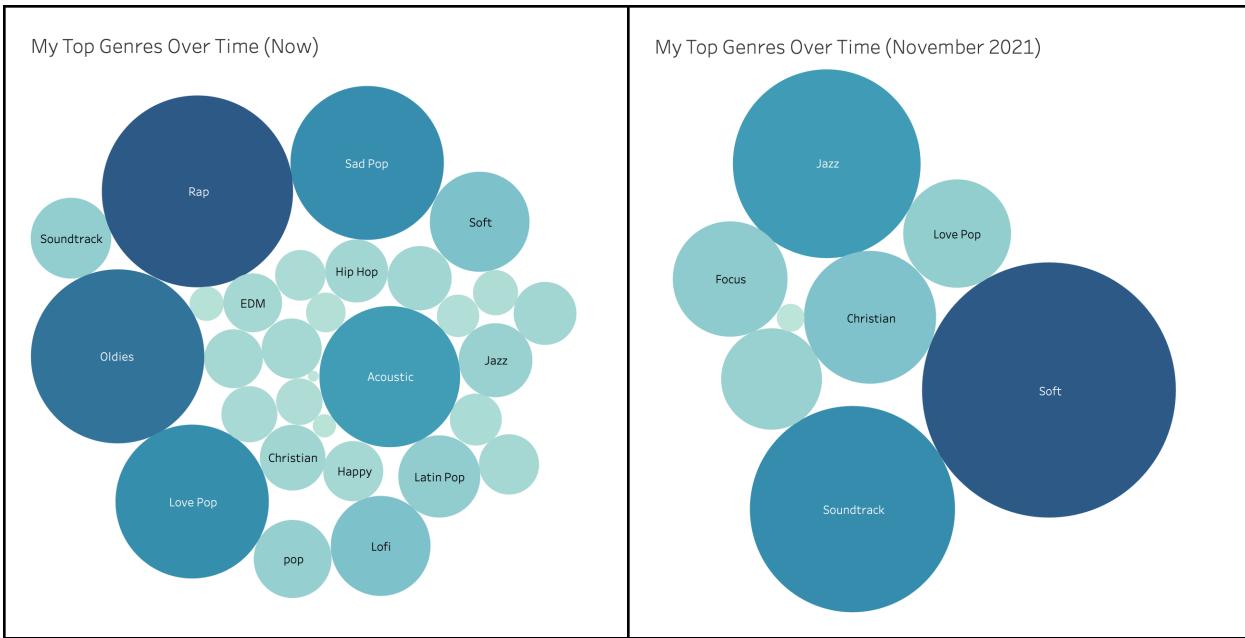
3 visualizations (the stacked bar chart, line graph, and summary statistic cards) are already implemented in Javascript/D3 and can be found in our submitted .zip file. There are some additional changes we would like to make (explained above).

Draft Visualizations

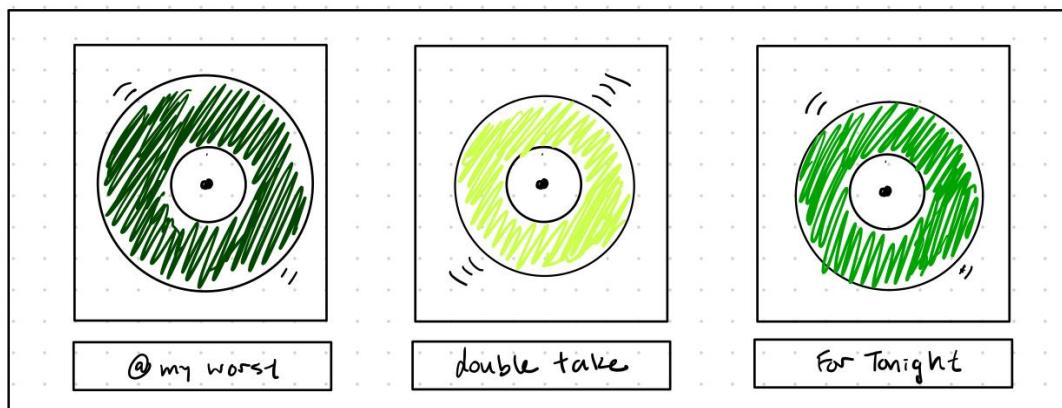
1. Stacked bar charts (with filtering) - Note that this is also implemented in D3/Javascript in our prototype v1, although without the dynamic changes over time; draft is left here for reference in Interactions section below & to document projected next steps



2. Bubble graph (with filtering) - Our next steps will include implementing the bubble graph below in D3, this visualization will show how users' top genres have changed over time. The size of the bubbles correspond to the number of songs that fall under a particular genre. In order to implement this visualization, we need to gather and merge genre data for the top songs in the library. The draft visualizations below show an user's genre data now (left) and in November 2021 (right). There will be a slider for users to select the period of time they want to visualize.



Innovative View



YOUR TOP SONGS

- For our innovative view, we will display top songs as records, with different properties of the records encoding different musical attributes of each of our top songs.
- These attributes are as follows:
 - The color of the record will correspond to the danceability and/or energy of the song, using a monochromatic scale that aligns with our color scheme; brighter hues will correspond to higher energy and danceability and dimmer ones to lower energy and danceability
 - The size of the record corresponds to the overall loudness attribute of the song

- The records will beat or pulse to the rate of the beat attribute in the song
 - Each of these attributes can be found using the Spotify Web API (publicly available and released by Spotify)
- This visualization draws on several different visual channels and concepts from lecture:
 - The repetition of the visualization illustrates the parallelism between each of the songs and groups them together
 - The alignment of the records will also help with this
 - The pre-attentive property of color both aligns with our color scheme and also helps users differentiate and process the different records

Interactions

Note: Text descriptions follow, and corresponding sketches can be found either above or in the working prototype.

- Stacked bar chart: As mentioned above, we will change this filtering to be by time based on how the playlists change over time (e.g. we might see the bar chart view advance from the one shown in “2020” through to the one shown in “Now” (all in one bar chart, changing with a slider that corresponds to the time)
- Line chart: Also as detailed above, we will add brushing, filtering, and zooming to our line chart so that the viewer can zoom in and out on certain parts of the line chart and filter by playlist or by time. This zoom and filter behavior will be modeled after the lab 11, with the dynamic bar chart.
- Statistic cards: currently planned to be static (since they are shareable and intended only to depict overall statistics)
- Bubble chart: as depicted above, the bubble chart will have a slider time selection where users can see the size of their bubbles grow and shrink relative to each other over time as the composition of their songs in different genres change.
- Records (innovative visualization): see above for description of planned interaction and dynamic properties

Week 12 - Prototype V2

Process Overview

Completed this week:

- Implemented all remaining visualizations
- Added brushing and zooming, filtering, interactivity
- Added downloadable button for data cards

To-do next week:

- Debug some small issues on the stacked bar chart
- Continue to tailor overall style and cohesiveness
- Address some small bugs (e.g. some spacing issues) and add additional small features (e.g. legends, some descriptions, etc.)
- Add togglable data between people
- Make small implementation changes and bug fixes based on user feedback

Line Chart (LW)

- ~~Add brushing / zooming on line chart~~
- ~~Add key to right side on click~~

Stacked Bar (LW)

- ~~Fix slider on stacked bar chart~~
- ~~Add key to right side on click~~

Cards (CL)

- ~~Un-hardcode cards~~
- ~~Make cards downloadable??~~
- ~~Flippable~~

Bubble Chart (IY)

- ~~Download and process data~~
- ~~Bubble chart for music attributes~~
- ~~One bubble for each attribute, then size of bubble depends on average value of that attribute from year to year~~

Creative visualization (LW)

- ~~Find data - API call? Need to pull music attribute data for top 5 songs (IY)~~
- ~~Implement animation~~

Misc

- ~~Make togglable to different users (IY - future)~~
 - ~~Clean Cindy and Lauren's personal data (IY - future)~~
- ~~Add descriptions / titles / subtitles (IY - future)~~

Order of Visualizations:

1. What does your music listening look like today? Creative
2. How has music listening changed over time? Bubble Chart
3. How has your music listening changed over time? Line Chart & Stacked Bar Chart
4. How does your music listening compare today? Cards

Week 13 - Think Aloud Study

Think-Aloud Study 1

Tester Name: Kathy Zhong

Test E-mail: kathyzhong@college.harvard.edu

General Observations from the think-aloud study:

- Likes the records, good hook!
- Cool color scheme
- Enjoys downloadable stats

What does the tester like about your data story?

- Hook is cool, likes the progression through the visualizations
- Likes how it narrows down throughout the website

What improvements does the tester point out?

- Do pulses match the tempo?
 - Have a legend about what color and pulse speed mean
- Add descriptions for the musical attributes and legend for the colors
- Axis for line graph is cut off

Was the intended key message clear to the tester? Why or why not?

- Breaking down how you and the general listener audience use Spotify
- Clear since that is what the visualizations focus on and narrow down to

Did the tester get your next steps or call to action? Why or why not?

- Don't really see call to action
- There's nothing in particular users are explicitly asked to do

Think-Aloud Study 2

Tester Name: Megan Cui

Tester Email: megancui@college.harvard.edu

General Observations from the think-aloud study:

- We should consider adding interaction to the records
- Make the line chart more obviously zoomable
- It is slightly unclear what the slider does on the stacked bar chart
-

What does the tester like about your data story?

- I really like the records at the top - exciting and eye catching
- I like the color scheme throughout

What improvements does the tester point out?

- I like the records, but I don't know what the colors mean
- Unclear what was me and what was global data
- Key for colors in bubble chart would be useful
- Title of bubble chart is confusing
- Scroll doesn't work when brushing on line chart, brush doesn't do anything
- What does each layer mean for the stacked bar chart and why is it stacked?
 - Is this my own playlist or something I listened to?
- Aggregate is not a clear word to use in the titles
- Would recommend changes to interaction and adding keys

Was the intended key message clear to the tester? Why or why not?

Look how interesting your Spotify listening is! Thought this because of personalization throughout the different visualizations

Did the tester get your next steps or call to action? Why or why not?

Not a clear call to action since no actionable message was used, but liked the download button at the end.

Discussion Questions

Based on the results of your ‘think aloud’ study, what would you improve in your data story?

Overall, we think we can improve our visualization by adding more clear sections and captions to explain each visualization and how they can connect. To improve this, we decided we will add subtitles to the prompting questions and captions for the graphs underneath each.

Additionally, there are a few changes we can make to the interactions, specifically on the record visualization and stacked bar chart to make it easier to understand what the data represents. Finally, there were several small suggestions to change wording to make it more clear what data is individual data versus global data.

Making the call to action and shareable insights at the end would be very helpful for our data story as well.

Are there any additional insights and visualizations you would use?

Would you amplify or change your message? Did your narrative work? Did the tester get your takeaways?

We decided it would be more intuitive to change the stacked bar chart to a regular bar chart where each bar is a different playlist and the user can drag a slider representing time to see the playlists grow over time. The stacked view was confusing to our testers as it was difficult to follow one bar (playlist) over time and compare the bars directly to each other.

Overall, the message and narrative was pretty clear, and each of our testers had no issue determining the takeaways and following the flow of the visualizations throughout the webpage. We can definitely make a few small tweaks here as explained above, with different subtitles, etc., but the flow is generally present.

Decide as a team which of these improvements you will implement and write down your decisions and why you made them in your process book as a numbered list.

Overall - toggle for different data

- ~~Process data (IY)~~
- ~~Implement HTML and data wrangling to use different data (IY) [in progress]~~
- ~~Spacing / height cut off (IY at least partially)~~
- ~~CSS: black background (at end LW) [in progress]~~
- ~~Captions for visualization (IY)~~
- ~~Improve data story - questions/headings (IY)~~
- ~~Fixed some spacing and centering (IY)~~

Vis 1 - Records:

- ~~Add button to play in spotify (CL)~~
- ~~Add legend for what colors, size and pulse speed mean (LW)~~
- ~~Fix height of label under records (at end) (LW)~~

Vis 2 - Bubble

- ~~Label spacing (at end) (IY)~~
- ~~Hover description in tooltip as well (IY)~~
- ~~Legend for colors (see LW^, formatted by IY)~~
- ~~Changing title → "Musical Attributes of Top Songs Over Time" (IY)~~

Vis 3 - Line chart

- ~~Add date and sum tracks when brushing (CL) [in progress]~~
- ~~Fix height (at end) (IY)~~
- ~~Fix clipping on left (IY)~~

Vis 4 - Stacked bar

- ~~Change to regular bar chart (LW)~~
- ~~Re-style slider and text displaying viewing year (IY)~~

Vis 5 - Card

- Download SVG to PNC (CL) [in progress]
- Make cards flippable (from week 12 to des) (CL)
- Top Tracks in aggregate alignment (IY)
- Add buttons to playing actual songs (CL)

Implement the intended changes and check them off your list (e.g., adding “done”). You can distribute the tasks among your team members.

Each change above is labeled with one of our initials for implementation this week.

Process Overview

Completed this week:

- Added new bar chart visualization
- Fixed narrative and visual bugs (e.g. spacing, new subtitles and titles, details to brushing on line chart, more detail to tooltips, etc.) to create a more informative visualization overall
- Added button functionality to play in Spotify
- Cleaned and processed all group members' data and altered structure to take all data, toggleable between people
- Added color legend for visualizations using color scale to make meaning more clear

To-do next week:

- Clean up small styling things, e.g. add background, section divisions, etc.
- Continue with small feature implementations (e.g. downloading on cards)
- Fix small bugs (e.g. spacing issues, tempo of songs doesn't match pulsing of records)
- Currently, visualizations update by person but does not write-over previous one; implement enter-update-exit sequences for visualizations so that each person's is not overlaid on the previous
- Record demo video & finalize project!

Week 14 - Final Submission

2 Minute Final Project Video

[Link to video](#)

Link to Project Website

[Link to website](#)

Project Implementation Code

Submitted via Canvas .zip file

Submission requirements:

- ~~Complete and clean process book~~
- ~~2-minute video (not longer)~~
- ~~Readme on how to run the code, used libraries, etc.~~
- ~~Data (submitted and with a file describing data fields)~~
- ~~Code (clean and documented)~~
- ~~Link to project website~~
- ~~Everybody in the team needs to fill out the peer evaluation form
(<https://forms.gle/YNxwWe6nht6FKE118>)~~
 - ~~++~~
 - ~~LW~~
 - ~~EE~~