CS171 Project Book

Michelle Hewson michellehewson@g.harvard.edu
Soline Boussard sboussard@g.harvard.edu
Pluto Zhang tzhang1@hsph.harvard.edu

Week 7: Process Book & Project Proposal

You will choose a team leader and set up your process book. Make sure all your group members can edit it. You also need to add a project proposal to your process book. In your project proposal, you will let us know what topic you are interested in exploring, including a project title and abstract. The team leader also needs to submit the project proposal as well as the team members' information through the project proposal form.

Project Title:

- **TikTok's Influence on Music**
- 2023 through the Eyes of Spotify Soundtracking 2023: Spotify's Perspective
- Spotify's Take on 2023
- Spotify's 2023 Year in Review

Project Abstract:

- The influence of music on society is an extremely transformative force. Furthermore, the recent influence of TikTok on society has proven to be an even more impactful force. We are interested in exploring and visualizing how TikTok and Spotify data from 2022 correlate. Two datasets on Kaggle contain a wide range of features about trending music including track name, artist name, popularity, danceability, and more about music on TikTok and Spotify. We plan to create a cohesive visualization to understand the top songs, genres, and other relevant features from both the Spotify population's listening habits and the TikTok crowd's most-viewed songs in 2022. We want to visualize these songs in a way that incorporates a song's danceability, genre, and their evolution throughout the year. Additionally, we want to include user engagement, where a user could select certain songs or artists to find out more information. We also want to include the popularity of songs in conversation with current events. Some of these might include the Taylor Swift Era's tour or looking at the difference in the population's listening habits between summer and winter. Our website will offer an exciting opportunity for music enthusiasts, data aficionados, TikTok and Spotify users, or anyone curious about the dynamic relationship between social media, music, and society. By contextualizing the music of 2022 within the framework of the popular app TikTok, we aim to offer a visually appealing journey through the music of 2022.

Project Data:

- https://www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022
- https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2022

Week 8: Team Agreement & Detailed Project Plan

Create a team agreement and a detailed project plan and add both to your process book (i.e., your shared google doc).

- We will communicate by the iMessage group chat we have set up, and have weekly Zoom meetings on Sundays at 1pm ET
- Although code will be written by individuals, all team members should be involved with the technical aspects of the project. All code should be documented well.
- We will split up the project based on different designs we want to have implemented. Each person will have a set amount of designs (graphs, charts, etc.) to contribute to the project.
- Final design decisions will be discussed among all members; fair compromises should be made when necessary.
- Work hours should be split as evenly as possible (actual task output may differ based on an individual's ability / previous experience). This ensures not only fairness but also learning opportunities for everyone. We will keep each other accountable so that one person does not work too much / too little.
- We will use a Git workflow to aid our progress as a team and help us split up the work of coding.
- Work will not necessarily be done together in person, but good communication via iMessage is expected in a timely manner. Work may be done remotely as long as collaboration and communication are done well.

TODO List

- Set up GitHub repository with starter code
- Brainstorm what kinds of visualizations we want
- Create map of what visualizations we want and where they should go
- Make sure the data is clean
- Decide if we are going to merge it with any other datasets

Signatures: Michelle Hewson Soline Boussard, Pluto Zhang
Date: 10/29/23

Target Categories

Michelle:

1. Target
2. Implement
3. Evaluate
4. Design
5. Data wrangling

Soline:

1) Data wrangling
2) Implement
3) Target

4) Evaluate
5) Design

Pluto

1) Data Wrangling
2) Target
3) Evaluate
4) Implementation
5) Design

Project Plan

- TikTok's Influence on Music in 2022
    - Michelle Hewson michellehewson@g.harvard.edu
    - Soline Boussard sboussard@g.harvard.edu
    - Pluto Zhang tzhang1@hsph.harvard.edu
- In the past few years, the app TikTok has exploded in popularity and has influenced trends across a variety of different fields (music, fashion, etc.). We are interested in understanding how TikTok has influenced overall music trends in 2022. If a song was popular on TikTok, does this mean it was popular on Spotify? Do artists purposefully create songs that they know will be popular on TikTok to increase their sales? We will investigate what it takes to create a song that will be popular on TikTok *and* Spotify– a combination that can absolutely change the direction of an artist's career.
- Personally, all three of us use Spotify to listen to music (and occasionally indulge in TikTok scrolling binges at night). We were interested in investigating whether or not there is a correlation between a song being popular on TikTok and being popular in the overall music charts.
- We are using data sources from Kaggle. One dataset covers the most popular music in 2022 based on Spotify trends and the other dataset covers the most popular music on TikTok in 2022. Most of the variables are the same across both datasets so we will have little cleanup work when merging these two together.
    - https://www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022
    - https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2022
- We do not expect a lot of data cleanup to be necessary for this project. The Kaggle datasets are very well-kept and clean and a lot of the variables overlap in the 2 datasets we have chosen. If we find that we need more data as we complete the project, we will prioritize already-clean data. We want to focus on features like song popularity, artist popularity, danceability, TikTok popularity, and other metrics relating to how popular a song is on TikTok vs on Spotify.

Week 9: Map

As a team, your discussion should focus on the following questions:

1. Who is your audience? Come up with **at least three** options and pick one target audience.

2. Describe your target audience in more detail. What do they know? What are their interests? What visualization literacy do they have? At what level of detail will you present information to them?
3. What questions about your data will be interesting for your audience? Come up with a list of interesting questions that your audience may have about your data. The more, the better, but your team should come up with **at least ten questions**.
4. What data do you have? Download the data you picked from the website linked in the PDF that describes the data (available on Canvas, week 2). Look at it in Excel or Google spreadsheet and briefly describe each attribute and its data type (categorical, ordinal, or quantitative) in your process book. It's OK if you are unsure about the data type for some attributes - you can simply describe them (e.g., geographic location).

1. Audience options:
    1. Music labels/production companies/musical artists
    2. Users of Spotify and TikTok that are interested in trends and the traits of the songs over these platforms
    3. Music consumers and enthusiasts who are interested in the current trends of music and how they are influenced by outside sources (like TikTok)
2. Target audience:
    1. Music consumers and enthusiasts are people that are passionate about all aspects of music. People in this audience will have a wide range of knowledge about music, whether it be from casual listening or deep research. This audience may be interested in the musical analytics surrounding current music trends like what ranges of acousticness, valence, tempo, or energy are prevalent in the top songs. Additionally, this audience may be interested in artist insights including statistics like how many artists are dominating the charts, what genres these artists specialize in, and what genres we may see dominating the charts in the future. Overall, the interests of this audience encompass a broad spectrum of insights about music, but we believe that a common main interest that everyone in this audience shares is what the current trends are for top music and how these trends may be affected by an outside influence like TikTok. In order to cater towards this audience, the visualizations that we will include in this project will not be too technical so that they are easily accessible by this audience, however we will try to include as much information about the top songs as possible in each visualization since that is what the audience is most interested in. For example, this could mean including tooltips that contain extra information about a specific song's energy/key/mode.
3. 10 questions:
    1. Which artists have the most songs in the top Spotify/TikTok charts?
    2. What kinds of songs are most popular on Spotify?
    3. What is the danceability of the most popular songs on Spotify? Is the average danceability of the most popular songs on TikTok different from the most popular songs on Spotify?

4. Do TikTok songs with happy or sad undertones have higher popularity? Is this trend the same for Spotify songs?
5. What is the relationship between an artists' popularity and their songs' popularity?
6. What is the relationship between a songs' weeks on top charts vs popularity?
7. Are the faster songs also considered louder on Spotify?
8. Do the top songs on Spotify with higher acousticness also tend to have more danceability?
9. What are the durations of the songs on Spotify?
10. What key are the most top songs on Spotify and TikTok in?

4. Data
    1. TikTok data: The TikTok dataset found on Kaggle includes the top songs of 2022 on TikTok. There are 18 features ranging from track name, artist name to danceability. The TikTok dataset and Spotify dataset share a majority of the same features, so can easily merge the datasets and analyze them together. The track name, artist's name and album are all categorical strings. The artist popularity column assigns an integer value between 18 and 95 for the artist's popularity. The popularity of the track is an integer value on a scale of 0 to 97. The danceability is a float that describes how suitable a track is for dancing, ranging from 0.31 to 0.96. Energy is measured from 0.0 to 1.0 and represents the song's intensity. Loudness measures the track in decibels ranging from -13.5 to -2.63. Mode indicates if the track is major or minor, and is a categorical variable. Key is another categorical variable, describing the pitches of the track. Speechiness , acousticness, instrumentalist, liveness, valence, and tempo are additional features to describe the track on continuous scale.

    https://www.kaggle.com/datasets/sveta151/tiktok-popular-songs-2022

    2. Spotify data: The Spotify data found on Kaggle summarizes the statistics of the top songs in Spotify. Many of these variables overlap with the TikTok dataset and have the same attributes and datatypes. In this dataset, there are 16 variables ranging from track name, artist name to duration. Track name and artist name are the **categorical** variables that consist of strings; peak_rank, weeks, key, mode, and time_signal are the **numerical ordinal** variables; danceability, energy loudness, speechiness, acousticnes, instrumentality, liveness, tempo and duration are the **numerical quantitative** variables that are continuous.

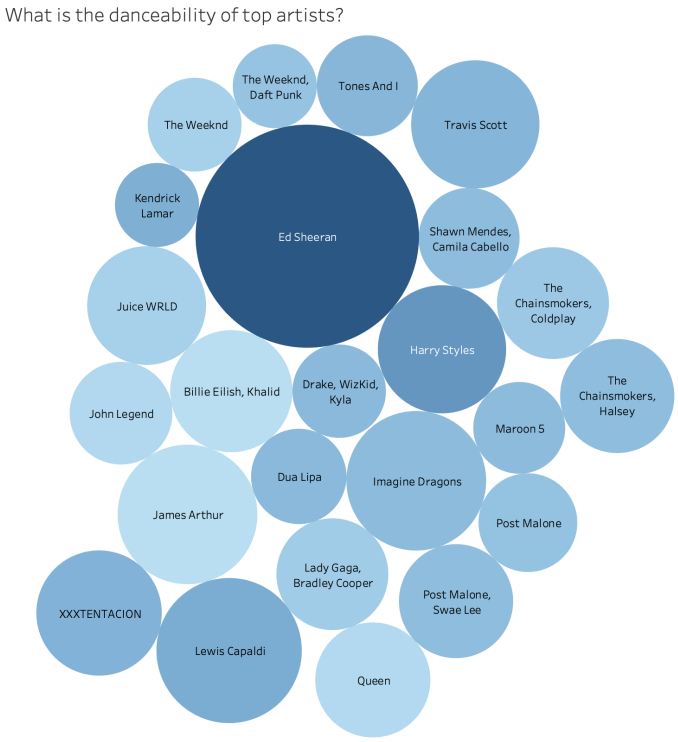
    https://www.kaggle.com/datasets/sveta151/spotify-top-chart-songs-2022

Individually at home, you should then

1. Load your data into Tableau and generate **at least three** initial visualizations **per team member**. Each visualization should use a new worksheet.
2. Save an image of each worksheet by right-clicking and selecting Copy -> Image... in Tableau. Upload the images into your process book. Make sure that the titles (i.e., the questions) are visible. Put the name of each team member before the screenshots that they uploaded.

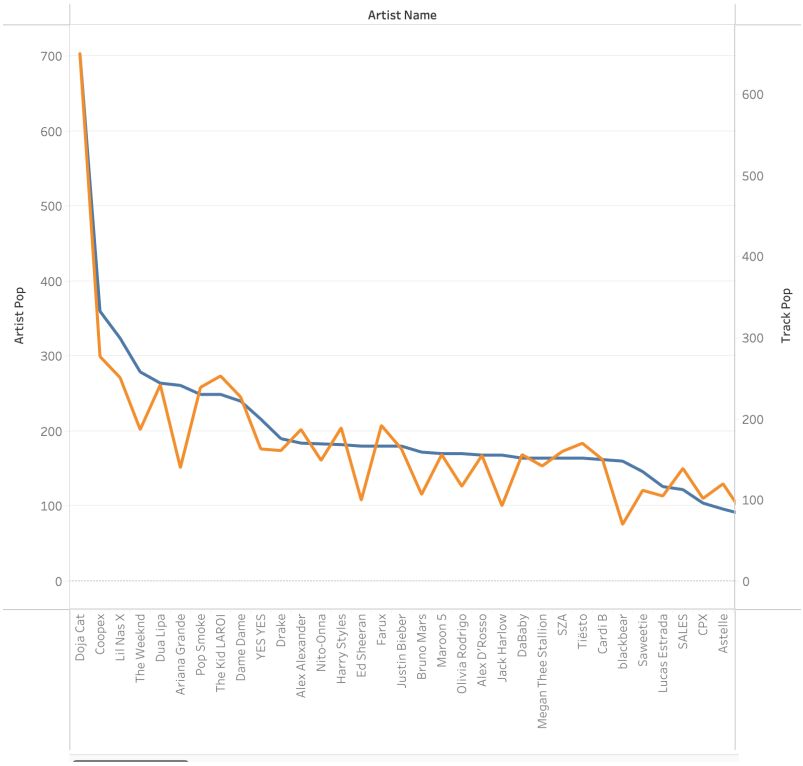
3. At the end of your screenshots, add a brief one-paragraph discussion of how the questions you answered in Tableau differed from the questions you came up with as a team. Why do they differ? Are some of them better than others, and why? Or why were you sticking to the original questions your team came up with?

Soline's Visualizations



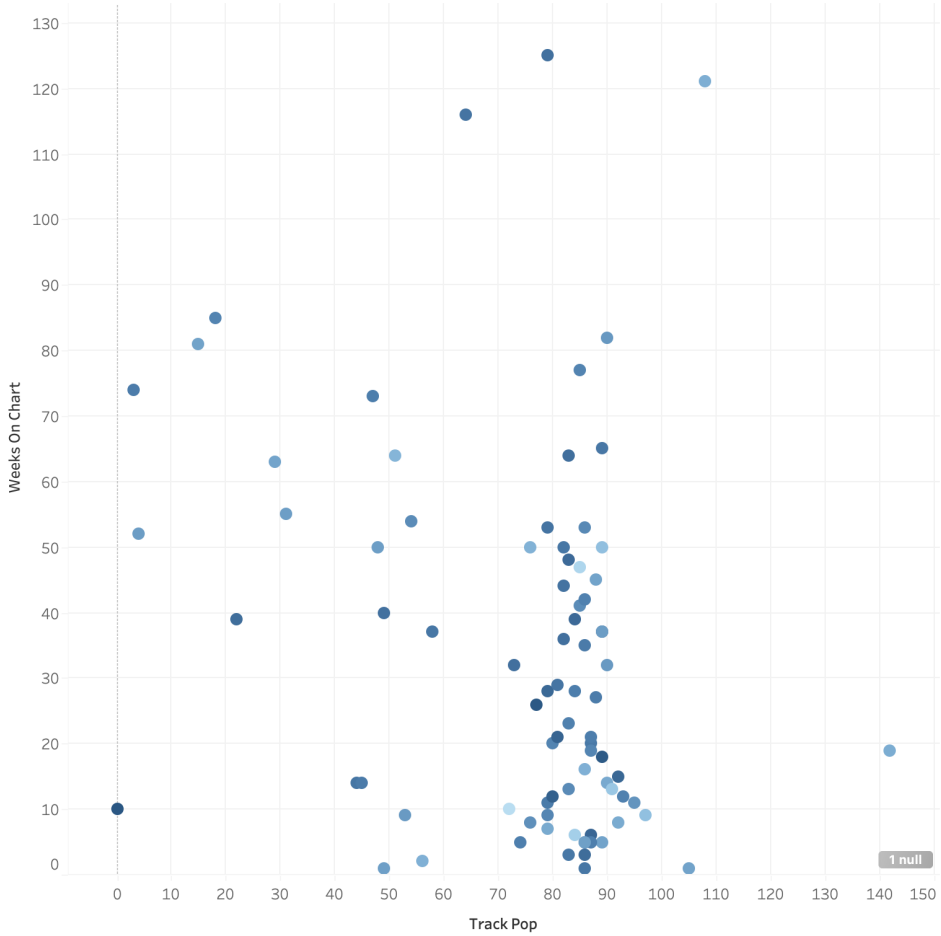
1. In this visualization, I am using the database on Spotify data to visualize the top artists in 2022. The size of the bubble reflects the total weeks they were on the top charts. The color of the bubbles is the danceability of each artist. This visualization explores our question of whether the top charts are influenced by dancing videos on TikTok. I was surprised to see that most of the top artists have a low danceability. This could suggest that popular songs on TikTok do not necessarily need to have high danceability, or there is little correlation between songs in the top charts and songs popular on TikTok.

Does an artists' popularity match their tracks popularity?



2. The next visualization I created was Artist popularity and track popularity. This attempts to answer the question of whether an artists' popularity is correlated with their songs' popularity. I sorted the artists by their popularity, and the visualization shows that as an artist's popularity decreases, so does their song's popularity.

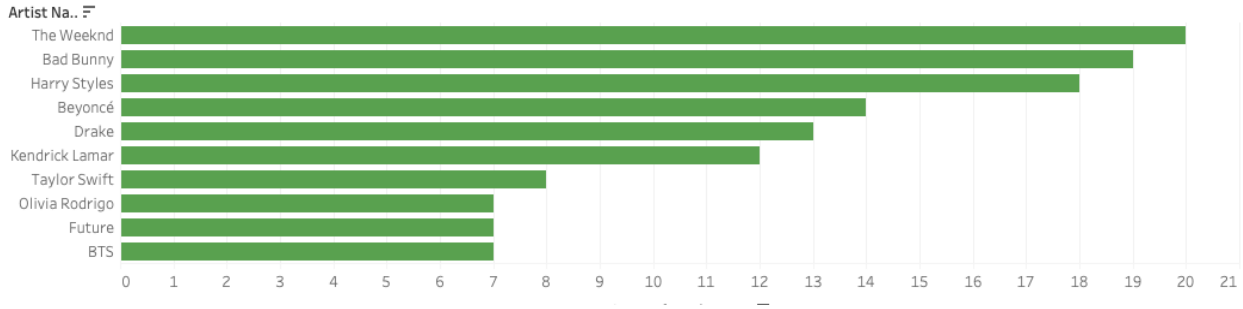
How does a tracks popularity effect the tracks weeks on the top charts?

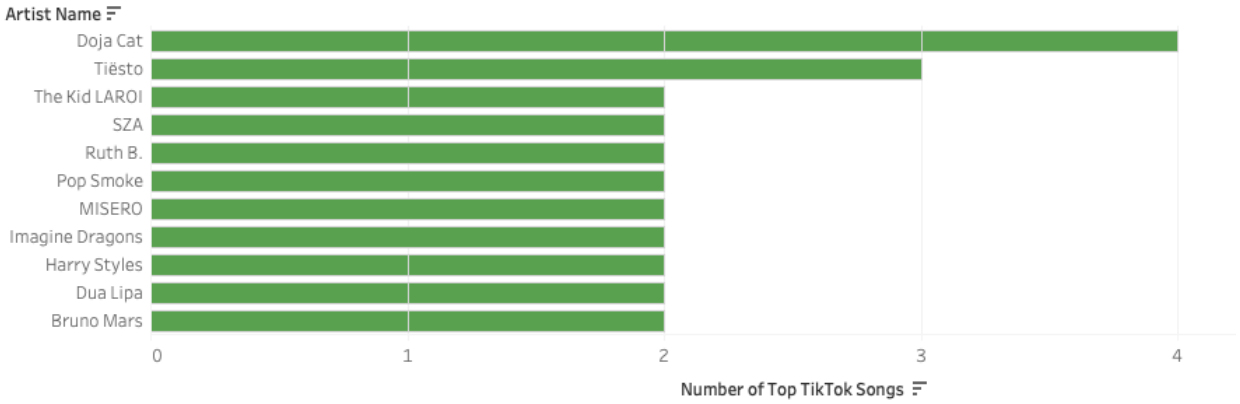


3. This plot explores the relationship between a song's popularity from the TikTok database and weeks on charts from the Spotify database. There appears to be no significant relationship between track popularity and artist popularity. Most of the track's popularity are clustered between 70-90 while the weeks on chart vary significantly.

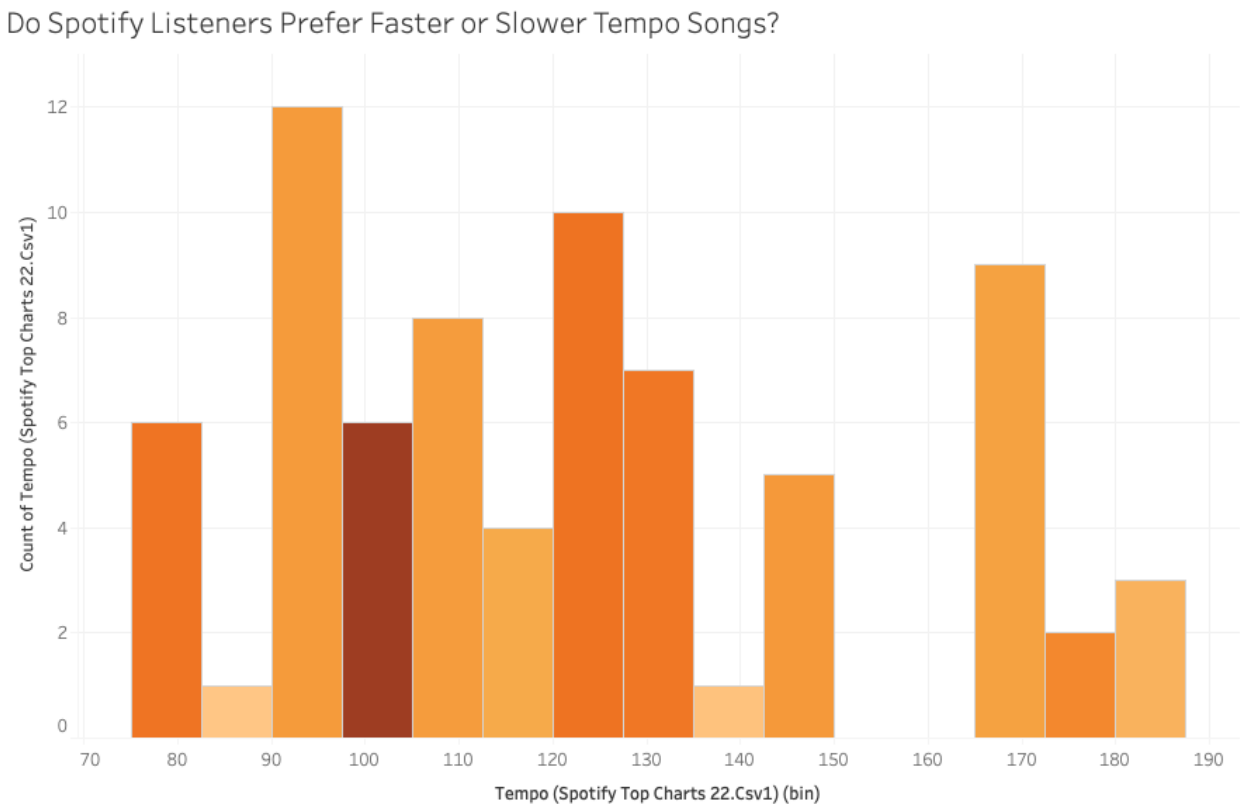
Michelle's Visualizations

Who Has the Most Songs in the Spotify/TikTok Charts?



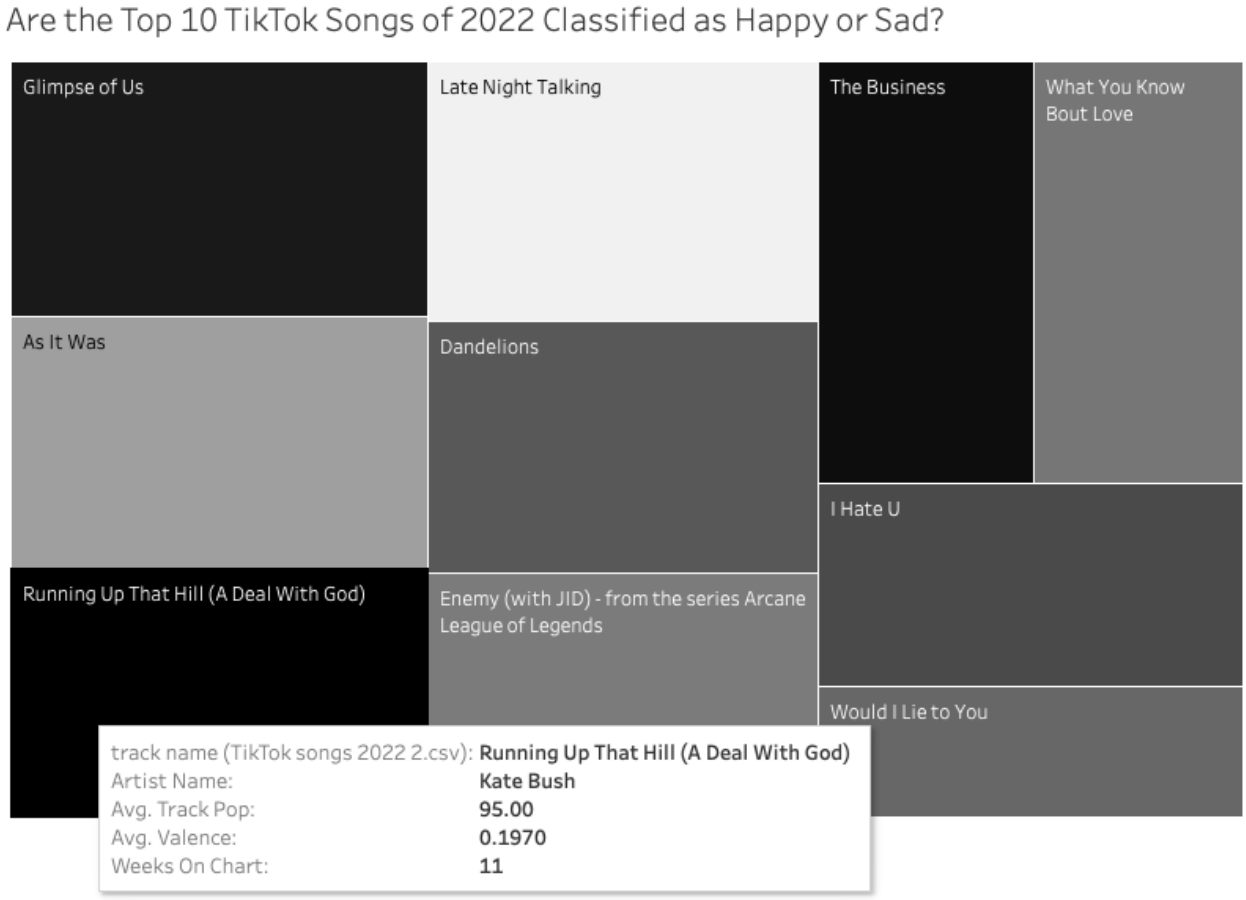


1. These two bar charts (intended to be one visualization) demonstrate how many songs an artist has in the "Top Spotify Songs of 2022" and "Top TikTok Songs of 2022", and it is filtered to show the top 10 artists who have the most songs on each chart. This addresses questions that may arise about which artists dominate the Spotify charts, which represent the population's listening habits, versus which artists dominate the TikTok song charts, a much smaller and unique portion of the population. "Who has the most songs in the top Spotify/TikTok charts? How many songs do popular artists have that are considered to be 'top?'" These types of questions behind this visualization are important because they can lead to investigations surrounding which artists/what types of songs/etc. tend to dominate the charts.



2. This visualization specifically investigates the type of music that is prevalent in the Spotify dataset. This is a

histogram that counts how many top 2022 Spotify songs are in each tempo bin. This visualization addresses questions such as “What kinds of songs are most popular on Spotify? Do listeners prefer faster/slower paced songs? What impact does mood or emotion have on song popularity?" Again, these questions are very important because they open the floor to discussion/investigation about what kinds of songs tend to dominate the top charts. Can artists use this data to create songs that appeal to the average person’s listening habits?

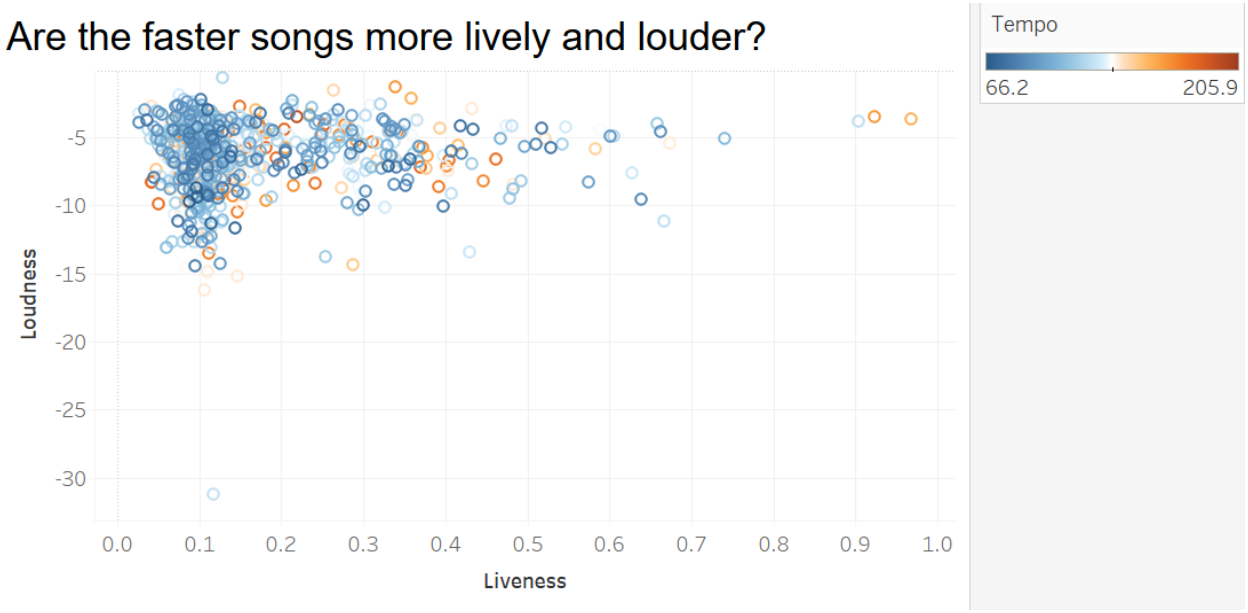


3. This last visualization illustrates the top 10 songs on TikTok in 2022 based on a combination of several different attributes. The darker the color, the lower the valence attributed to a song (valence classifies how happy/sad a song is perceived to be. Happy songs have valence closer to 1, sad songs have valence closer to 0). This visualization investigates similar questions that the visualization above addresses, such as “Which kinds of songs are most popular on TikTok? Do TikTok users prefer to watch videos that have happy or sad undertones? How many weeks on the Spotify charts do these topTikTok songs have?” Again, these questions are important because it gives us an idea of what an average TikTok user’s watching and listening habits are, and it provides a little insight as to how well a popular song on TikTok does in the Spotify charts. This visualization informs us that a majority of the top 10 TikTok songs of 2022 were classified with a lower valence. Lots of sad TikToks are being made!

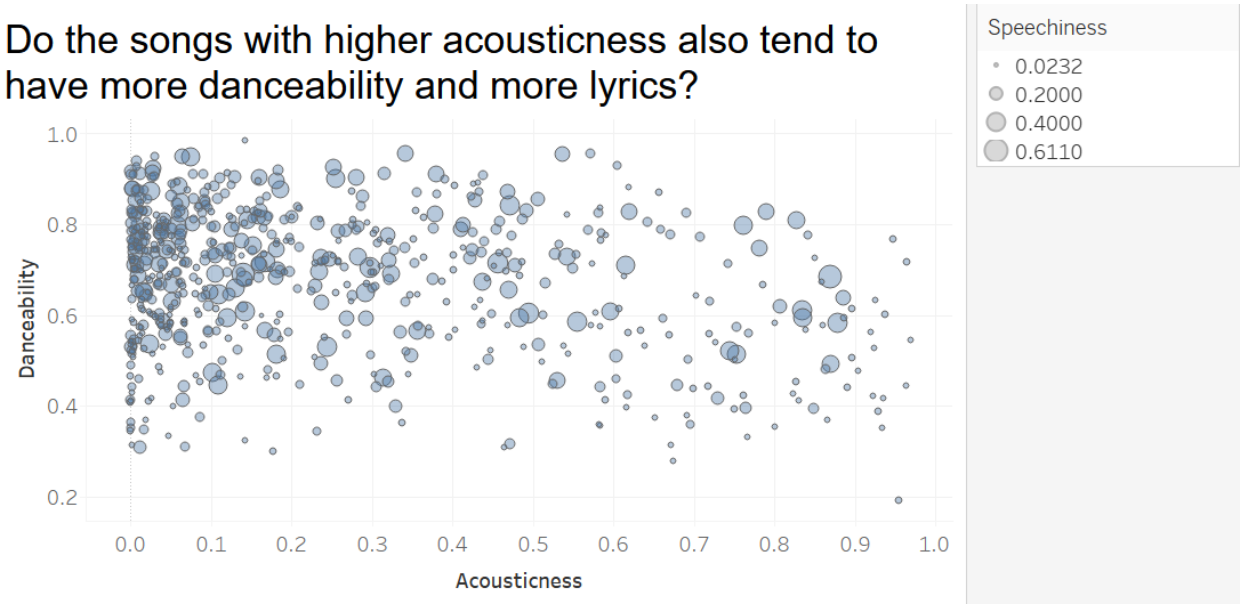
Pluto’s Visualizations:

1. Liveliness vs Loudness of spotify songs colored by tempo

Using colored scattered plot, this visualization demonstrates the relationship of liveness of the songs and loudness of the songs grouped by tempo. If the dot that represents the song locates further on the x axis, that means the song has more liveliness, while if the dot locates higher on the y axis, that indicates the song has more loudness. The color gradient helps visualize the tempo of the song, and if the color of the dot is more orange, the song is likely to be faster, while if the dot color is more on the dark bluer side, the song is slower. The question we can ask is, "are the faster songs more lively and louder?"
Looking at the plot we can see, there's no obvious trend between the tempo of the song and loudness, but songs that are more lively tend to be louder and faster.



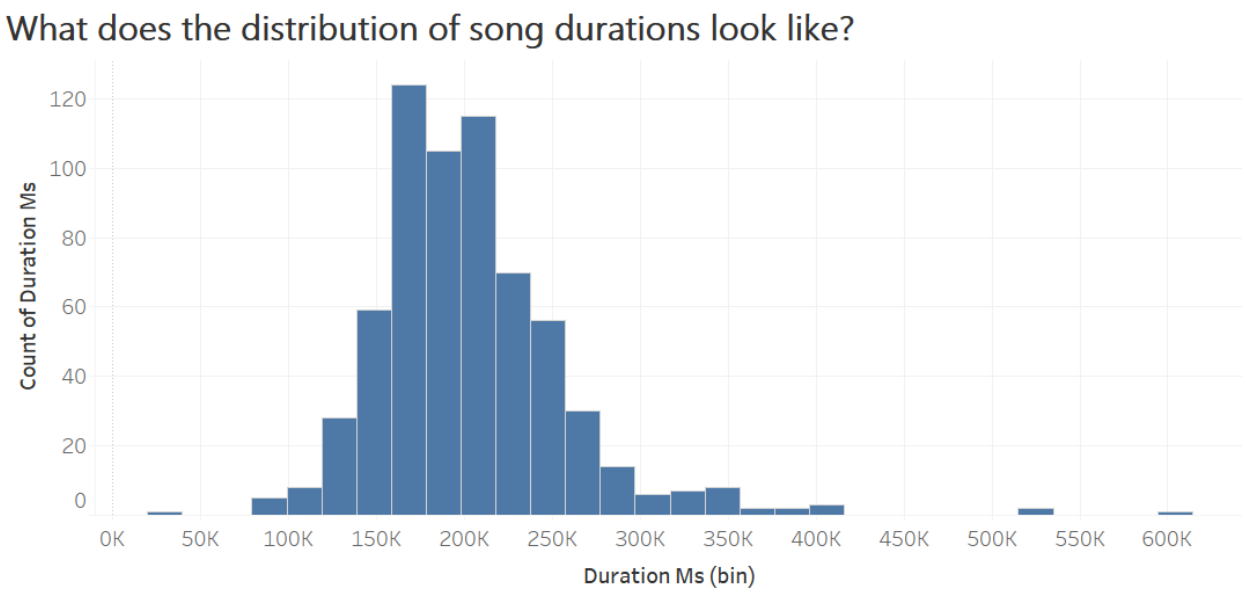
2. Acousticness vs Danceability of songs grouped by speechiness



From this scatter plot we are plotting the acousticness and danceability of the songs, marked by the speechiness. The further the dot is on the x axis, the more melodious the song is, and the higher the dot is on the y axis, the more danceable the song is. The dot size demonstrates the speechiness of the songs. Question to be ask: “Do the songs with higher acousticness also tend to have more danceability?” “Do the songs with more danceability also have a lot of lyrics?”

From the plot we can see, the songs’ acousticness and danceability doesn’t seem to have correlation with speechiness, but songs with higher acousticness tend to have lower danceability.

3. Distribution of song durations

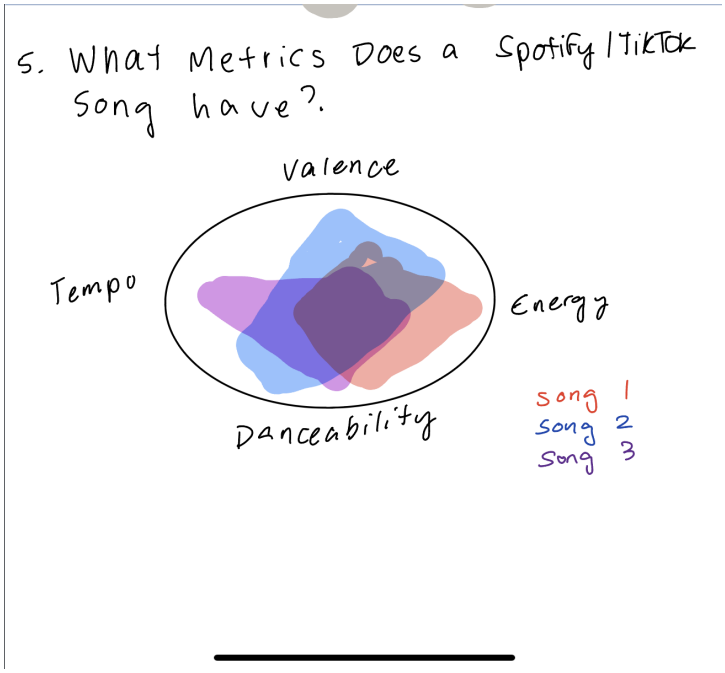


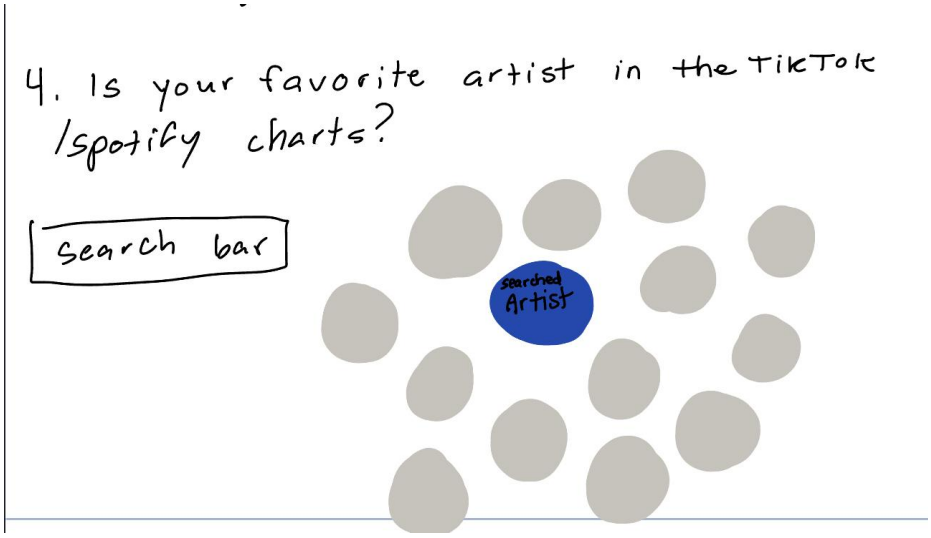
This visualization is a histogram that shows the count of songs within each duration range. The length of each bar demonstrates the number of songs that fall within each duration range, and the x axis position demonstrates the duration length of the song. We can ask this question: "do the count of songs durations follow a normal distribution, such that most of the songs have a duration that's in the middle, and fewer on the extremes?" The plot answers this question "Yes", such that the distribution follows the pattern of normal distribution, and most of the song lengths are in the middle, while few songs are extremely short or extremely long.

Week 10: Data, Sketches, Decide & Storyboard

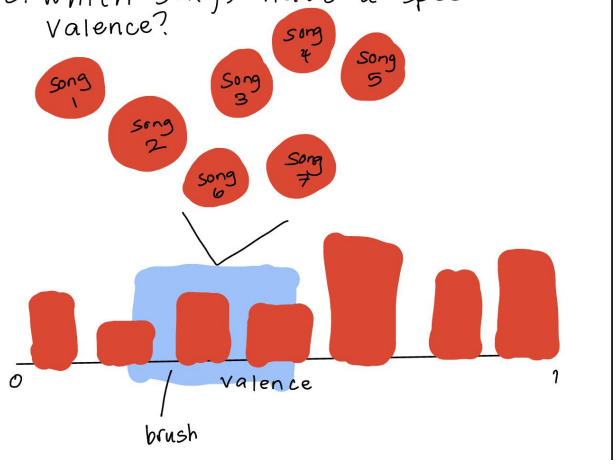
In this step, you will individually sketch visualizations to answer the questions you proposed during the 'map' phase. You do not need to create a single visualization that can answer all the questions. Instead, think of visualizations that answer each question individually. Be open-minded and sketch whatever comes to your mind. Please do not limit yourself to what Tableau or D3 might be capable of.

Michelle's sketches:

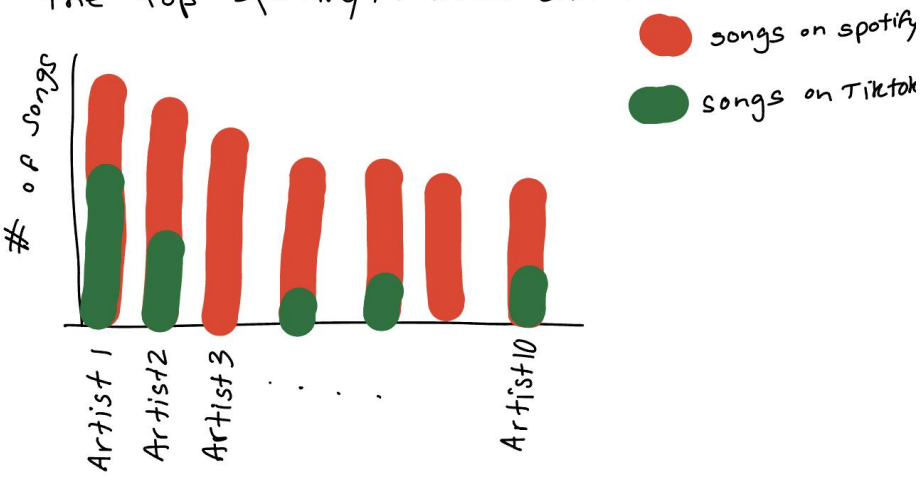




3. Which songs have a specific valence?

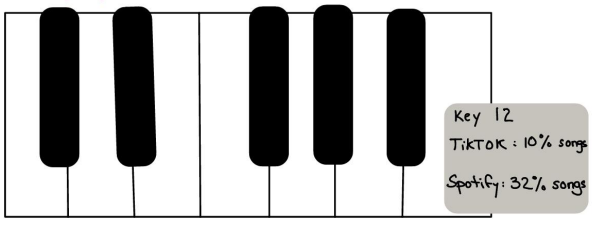


1. Which artists have the most songs in the top Spotify / TikTok charts?



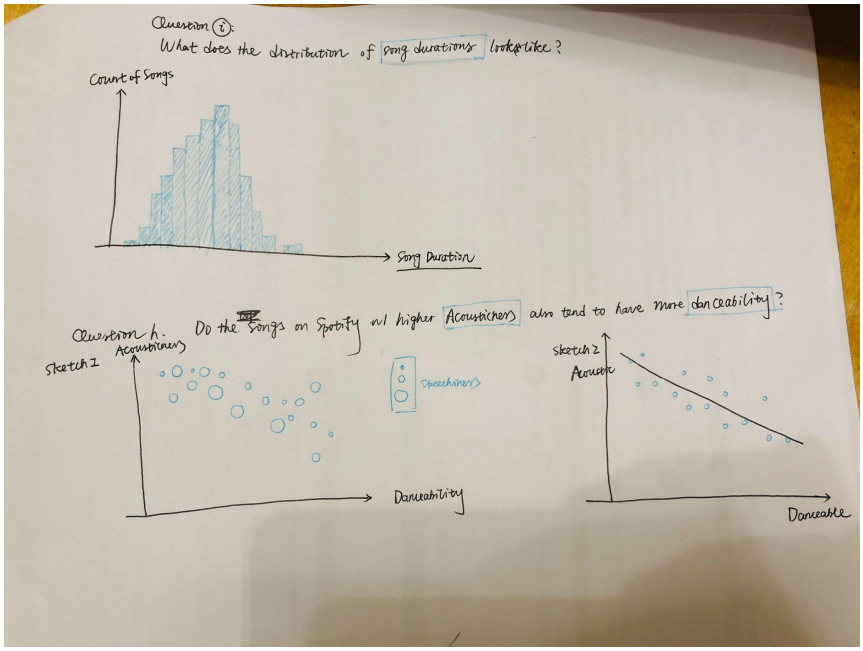
2. What keys are the most popular songs on TikTok/Spotify in?

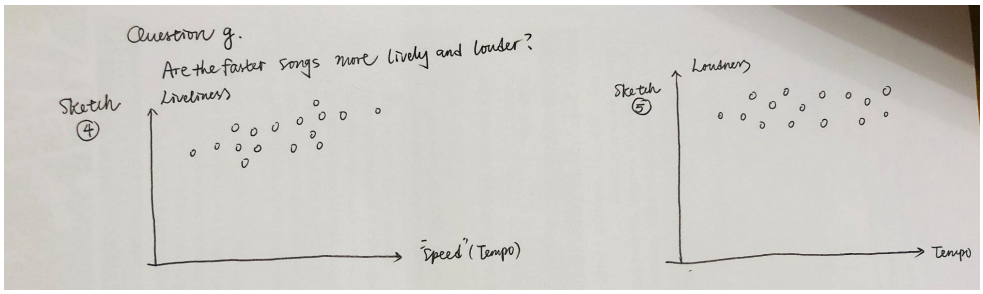
(Keys on Piano)



Tooltip will show how many songs are in each key category

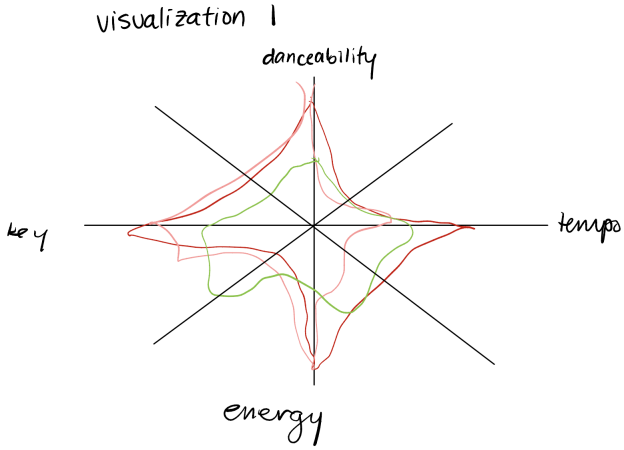
Pluto's sketches:



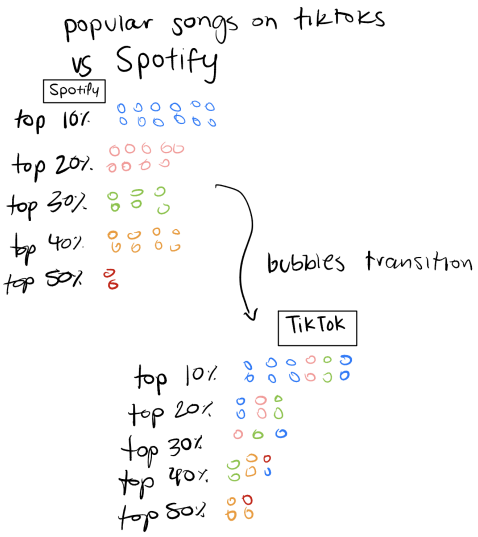


Soline's Sketches:

What kinds of songs are most popular on Spotify?

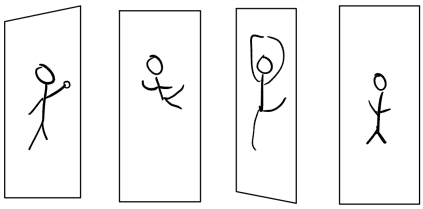


What is the relationship between a songs' weeks on top charts vs popularity?

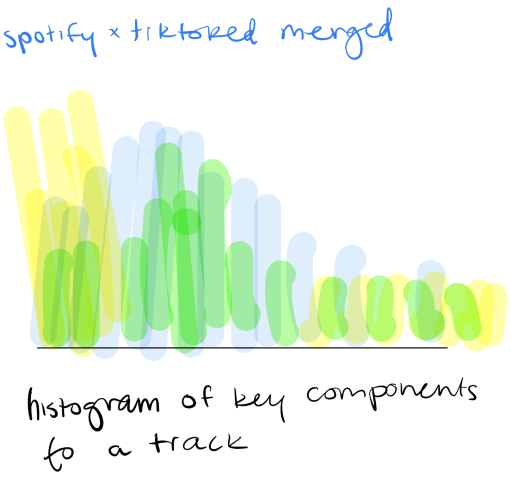


What TikTok trends are associated with the top songs?

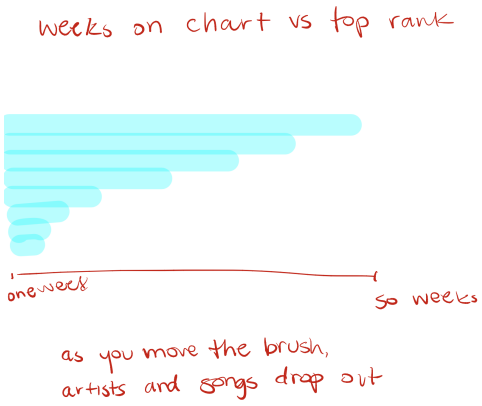
tiktok trends with the top 5 songs



* hover over for the video to play



What kinds of songs are most popular on Spotify?



Which artists have the most songs in the top Spotify/TikTok charts?

Decide step:

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.

Our questions:

1. Which artists have the most songs in the top Spotify/TikTok charts?
2. What kinds of songs are most popular on Spotify?
3. What is the danceability of the most popular songs on Spotify? Is the average danceability of the most popular songs on TikTok different from the most popular songs on Spotify?
4. Do TikTok songs with happy or sad undertones have higher popularity? Is this trend the same for Spotify songs?
5. What is the relationship between an artists' popularity and their songs' popularity?
6. What is the relationship between a songs' weeks on top charts vs popularity?

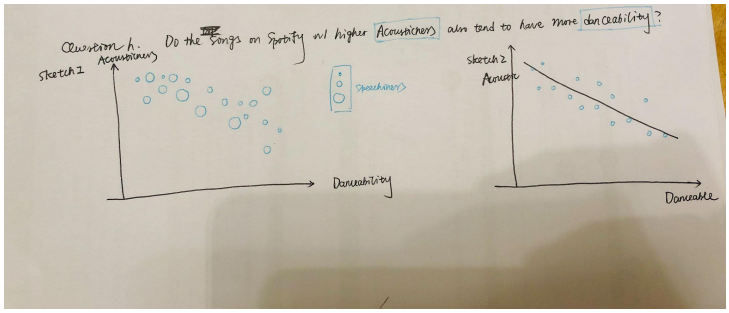
7. Are the faster songs also considered louder on Spotify?
8. Do the top songs on Spotify with higher acousticness also tend to have more danceability?
9. What are the durations of the songs on Spotify?
10. What key are the most top songs on Spotify and TikTok in?

Our visualization table:

*Sketch ID : Initials for the Author followed by her sketch ID

Sketch ID	Question ID	Author	Voting Score
PZ3	9	PZ	0
PZ4, PZ5	7	PZ	0
PZ1, PZ2	8	PZ	3
MH1	1	MH	2
MH2	10	MH	3
MH3	4	MH	1
MH4	1	MH	0
MH5	3	MH	1
SB1	2	SB	2
SB2	5	SB	1
SB3	4	SB	1
SB4	2	SB	0
SB5	1	SB	2

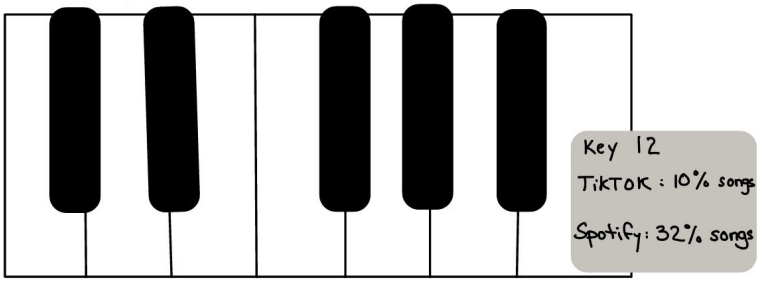
The visualizations we voted on:



-----use carousel tool to switch between these two sketches

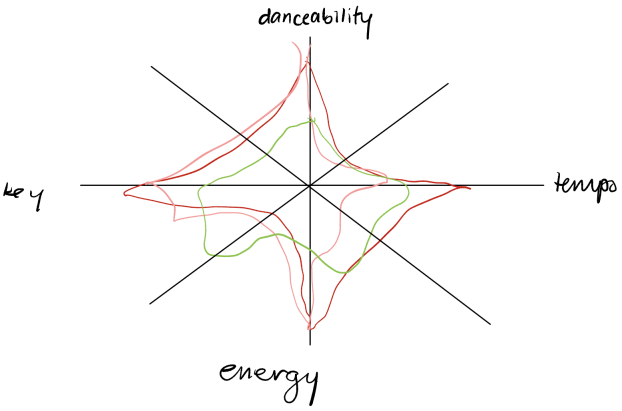
2. What keys are the most popular songs on Tiktok/Spotify in?

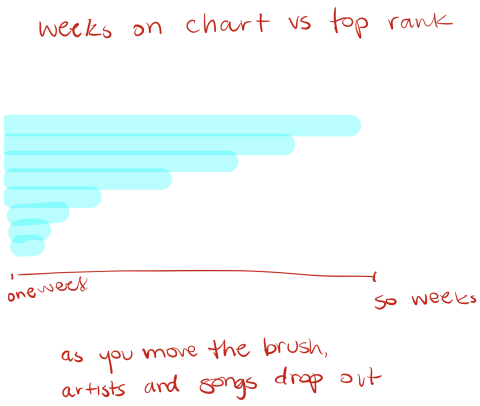
(Keys on Piano)

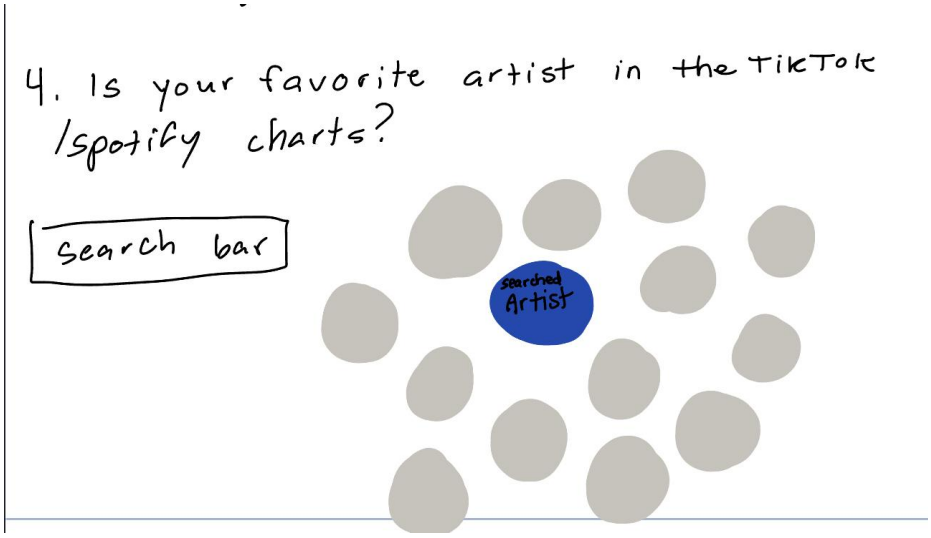


Tooltip will show how many songs are in each key category

visualization 1







Add a one-paragraph explanation at the end summarizing your decisions and rationale for choosing the sketches you plan to implement.

In this stage of the project, we are focusing on visualizations that we are hopeful that we can implement and are not too basic. Thus, a lot of the visualizations that we voted on are more difficult extensions of visualizations we have made in class/homework. We want our visualizations to include interactive features so the viewer can customize their experience, so we prioritized that while deciding which visualizations to vote on and include. We also wanted to examine a variety of different metrics in our dataset, ranging from the key/energy/valence/danceability of some songs to the popularity of specific artists in both sets of data.

1. Pick your main message.
    a. First, we encourage you to individually use Tableau to explore your dataset. Identify several insights (**at least 2-3 per team member**) that you think are important or interesting. Record these insights in your process book with each team member's name.

b. As a team, spend some time looking at the list and discussing the various insights. Pick **one main insight** that you think is most important or interesting to your audience. Formulate the insight as a message ("so what") in one sentence.
c. Record your main message and why your group chose it in your process book.

Michelle:

1. The Weeknd is the artist with the most songs in the Spotify top songs (20 songs), but he is not one of TikTok's top 10 artists (artists with the most songs in the top TikTok songs data).
2. Several of the top artists in the TikTok songs data are DJs that produced remixes of already popular songs.

Soline:

1. **There is not a clear trend between a track's popularity on TikTok and their weeks on Spotify charts, suggesting that artists who are popular on TikTok are not the same artists who are popular on Spotify.**
2. An artists' popularity and track popularity follow the same trend (if an artist's popularity is high, their track's popularity is likely high).
3. Ed Sheeran has the greatest sum of weeks on chart, but an average level of danceability when compared to other top songs.

Pluto:

1. There's no clear trend between the songs' tempo and their liveliness; however, the more lively songs tend to be louder.
2. The durations of the songs over spotify tend to be distributed normally, with most of the songs falling within the range of 2 to 5 minutes.
3. More acoustic songs tend to be less danceable. And more danceable songs tend to be less acoustic.

Main Insight:

There is not a clear trend between a track's popularity on TikTok and their weeks on Spotify charts, suggesting that artists who are popular on TikTok are not the same artists who are popular on Spotify.

Why our group chooses it as the main insight: first, it combines the information from our two datasets, corresponding to TikTok and Spotify, and offers a comparison between the two apps that we focus on; second, it suggests that the popularity of the songs over the two applications are different, which opens up the discussion on whether the audiences and people's preferences are similar across the two apps.

Sketch your data storyboard. Following the four steps, you will create a data storyboard to communicate your main message. Instead of physical sticky notes and pens (our preferred option if you meet in person).

1. Please make sure that your storyboard points are lined up with the four elements of a data story (hook, rising insights, main message, and solution).
2. Take a screenshot of the final data storyboard (or a photograph of your sticky notes, in case you meet in person) and put it in your process book.

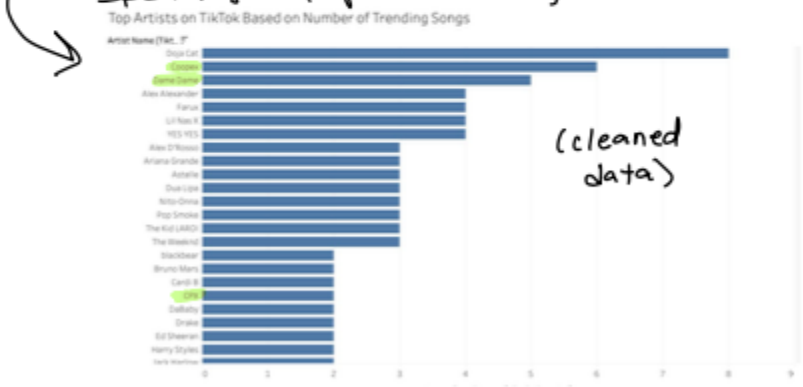
Hook

- 755 million active users on TikTok.
- In 2021, 430 songs surpassed 1 billion views on TikTok.
- Does TikTok have the power to influence music trends?

Rising Insights

- TikTok gives smaller artists the opportunity to gain traction.

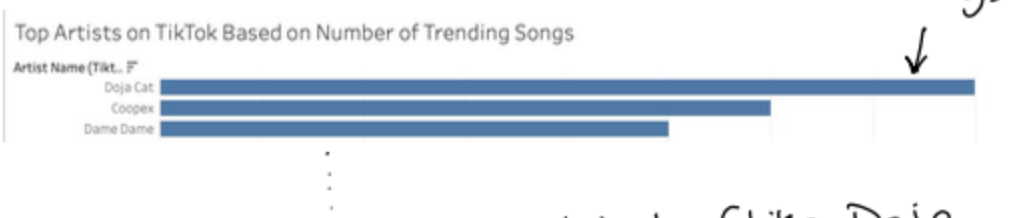
1. Lots of the top artists on TikTok are DJ's that make remixes of artists that are not well known (no songs on Spotify top charts).



2. Some artists (Doja Cat) can leverage their talents and have songs that are popular both on TikTok and Spotify. She

has 8 songs on both the TikTok and Spotify top songs. She has the highest number of songs on the top TikTok songs.

8 songs



3. What makes an artist (like Doja Cat) successful on TikTok?

Artist Name	Avg. Danceability	Avg. Energy	Avg. Instrumentalness	Avg. Tempo
Doja Cat	0.8	0.7	0.0	119.5
Coopex	0.8	0.8	0.0	126.2
Dame Dame	0.7	0.8	0.0	123.4
Alex Alexander	0.7	0.6	0.1	136.0
Farux	0.7	0.7	0.0	125.7

a. Popular artists on TikTok average high danceability, high energy, low instrumentalness, and high tempos.

Artist Name	Avg. Danceability	Avg. Energy	Avg. Instrumentaln..	Avg. Tempo
Bad Bunny	0.8	0.7	0.0	119.4
The Weeknd	0.6	0.7	0.0	121.2
Kendrick Lamar	0.6	0.6	0.0	120.6
Drake	0.7	0.5	0.1	126.7
Harry Styles	0.6	0.6	0.0	128.0

b. Popular artists on Spotify have ~~songs with very similar stats.~~

*Maybe it's not the music itself that gauges whether or not it will be popular on TikTok / Spotify? Let's investigate...

main Idea

The averages for every Spotify song:

```
Average Spotify Stats
Average of Danceability: 0.6744256965944272
Average of Energy: 0.6408356037151702
Average of Loudness: -6.356034055727553
Average of Tempo: 121.09157120743033
```

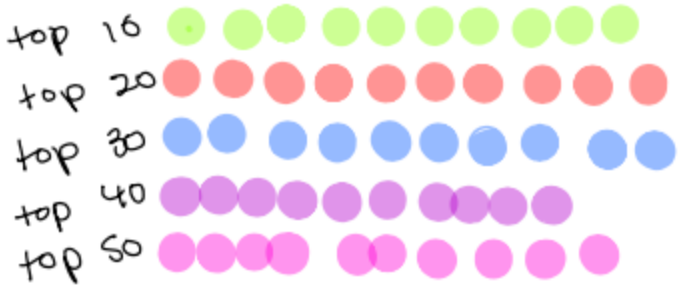
The averages for every TikTok song:

```
Average TikTok Stats
Average of Danceability: 0.7153384030418252
Average of Energy: 0.6826311787072242
Average of Loudness: -6.300513307984791
Average of Valence: 0.5061490494296578
Average of Tempo: 122.88369581749048
```

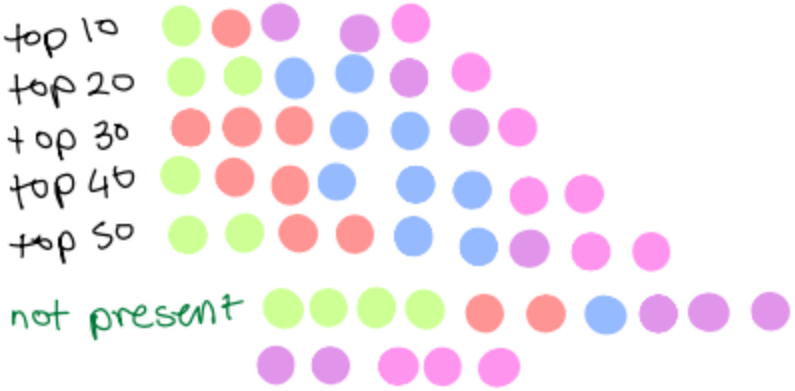
* Danceability average is higher on TikTok!
(we'll use some danceability visual here)

Only household artists' names appear in both TikTok and Spotify top charts

TikTok data



↓ reorder to top Spotify data



↑ This shows artist overlap in both datasets

Solution

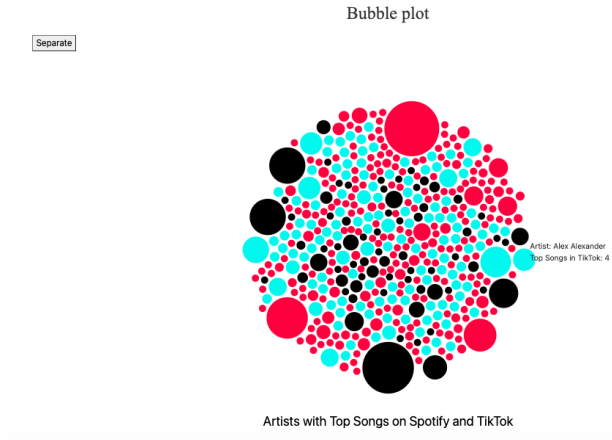
- For a song to be popular on TikTok, higher danceability and tempo stats seem to be necessary.
- To also be popular on Spotify, an artist needs to be considered a 'household' name.
  - 'Household' name artists dominate the Spotify charts
  - Lesser known artists appear more frequently in TikTok charts.

Week 11: Prototype V1

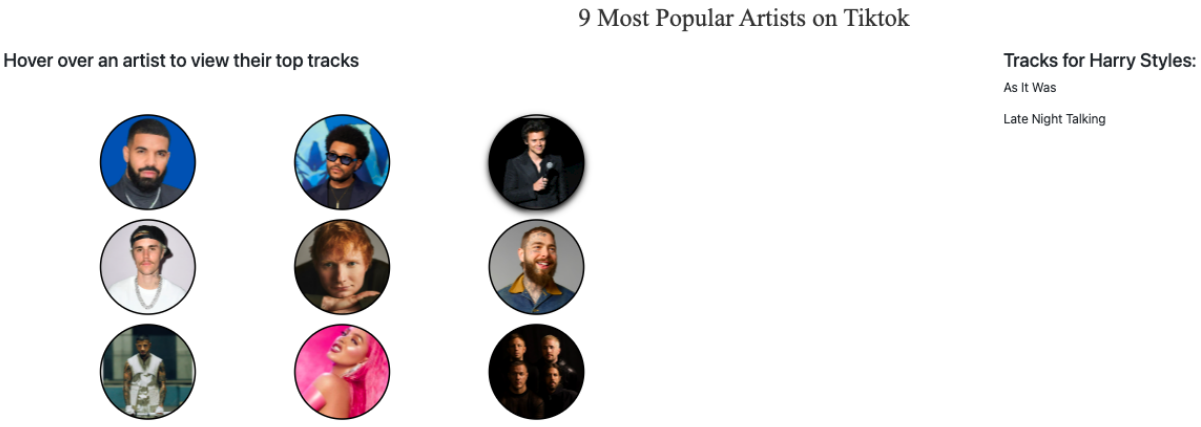
The exact requirements are:

- Name of students that worked on prototype V1 submission.
- Data scraping and cleaning complete (using the real data sets)
- **At least two D3 visualizations already partly implemented (including data loading and the basic vis, filtering does not have to work yet)**, and detailed drafts for 2-3 more visualizations
- Rough webpage design and structure has to be done and implemented (placeholders for visualizations, text, and images allowed)
- Storytelling is clear
- **The first design of an innovative view**
- Interactions (e.g., filtering, brushing, etc.) have to be designed (at least in a textual description and some sketches)
- Up-to-date process book

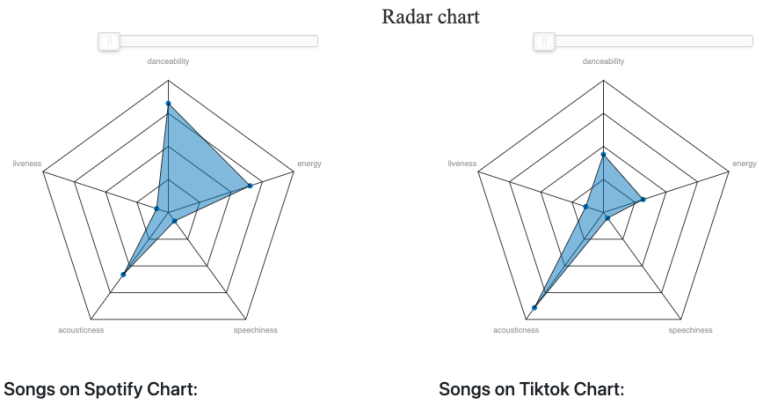
Honestly, we've been working on this project so much and forgot to document a lot of our milestones for prototype 1 as we worked on it! Thus, we had to download a zip from one of Github commits that accurately captured our progress that would best fit into a "Prototype V1" version of the project. If something's missing, please let us know– it was most likely implemented before the Prototype V1 due date and we just forgot to capture it in our zipped Git commit.

- Names: Michelle Hewson, Soline Boussard, Pluto Zhang
- Data Scraping: Lots of files (on Github: https://github.com/michellehewson/cs171_finalproject/tree/main/data)
- At least 2 visualizations partly implemented & drafts for 2-3 more visualizations::
    - 

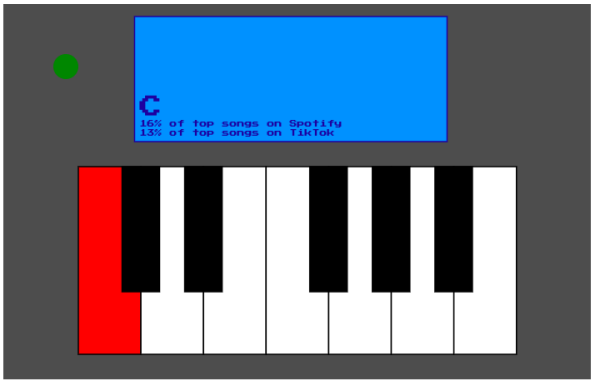
bubble chart for different artist filters, radio buttons will be used to switch between TikTok and Spotify data, one-hit-wonders, top artists, etc. Also want to implement a search bar to see if your favorite artist is in the top charts



- chart with faces of artists to show average artists statistics (danceability, tempo, etc.), buttons will be used to switch between TikTok and Spotify data

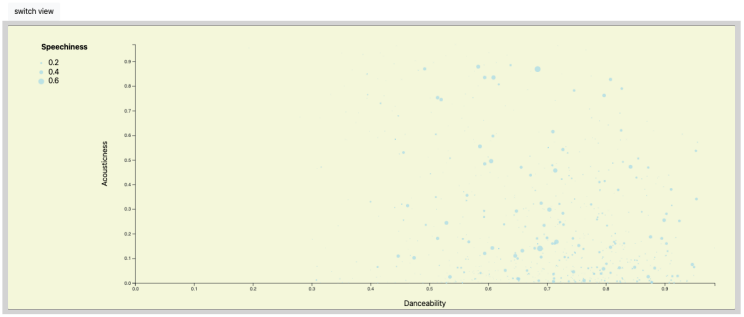


- radar chart to explore the different characteristics of the top songs on both TikTok and Spotify charts

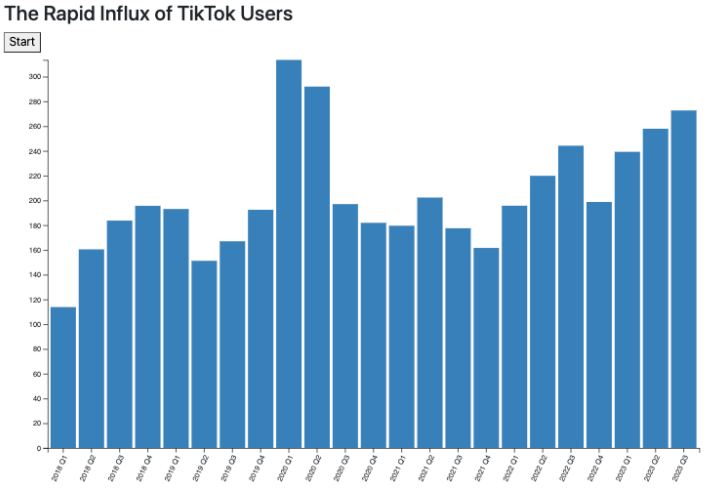


- interactive piano that shows the distribution of the keys that TikTok/Spotify songs are recorded in

scatter plot



- scatterplot for exploring trends in song characteristics (danceability, energy, tempo, valence, etc.)



- bar chart that will be animated so that the user can see how fast TikTok's popularity exploded by watching how fast the bars grow

- Innovative View:
  - We will be guiding the user through exploring the similarities and differences between the top songs and artists on the TikTok and Spotify charts. This will be done with several interactive features and visualizations, including:
    - A piano visualization that plays the sound of the key you press and displays the percentage of songs that are recorded in that key
    - A bubble chart visualization that allows the user to search whether or not their favorite artist is featured in the top charts
    - Yes/No question about whether the user uses TikTok and then relevant facts will generate based on the user's choice
    - We also plan on embedding some TikTok videos into our overall storytelling so that if a user is unfamiliar with the app, they can become acquainted with some of the trends/flow of TikTok
    - Please refer to our webpage layout (index.html) in the zip or in Github for more information about our innovative design!
    - Github: https://github.com/michellehewson/cs171_finalproject

Week 12: Prototype V2

Prototype V2 Features to add

- [x] ~~Title page~~
  - [x] ~~Design for title page~~
- [x] ~~Bubble page~~
  - [x] ~~Fix glitches~~
- [x] ~~Transition slides~~
  - [x] ~~Between bubble chart and face plot~~
  - [x] ~~Takeaway from bubble chart and face plot~~
- [x] ~~Face plot~~
  - [x] ~~Scroll bar~~
  - [x] ~~Bar chart~~
  - [x] ~~Look into adding music~~
  - [x] ~~Prettify~~
- [x] ~~Radar chart~~
  - [x] ~~Make svg not scrollable~~
  - [x] ~~Scroll bar~~
  - [x] ~~Labels~~
  - [x] ~~Prettify~~
- [x] ~~Histogram~~
  - [x] ~~Change text to match other pages~~
  - [x] ~~Transitions~~
  - [x] ~~Lower Svg height so it's not scrollable~~
- [x] ~~Scatterplot~~
  - [x] ~~Lower svg height so it's not scrollable~~
  - [x] ~~Change text to match other pages~~
- [x] ~~Conclusion~~
  - [x] ~~Write it~~
- [x] ~~Reference page~~
  - [x] ~~Write it~~
- [x] ~~Overall~~
  - [x] ~~Pick a font and stick with it~~

Week 13: Think Aloud Study

CS 171 – Think-Aloud Study

Please copy this form into your process book first!
Tester Name 1: Eli Dabkowski
Tester Email 1: elijahdabkowski@g.harvard.edu

Tester Name 2: Sydney Levy
Tester Email 2: sydneylevy@g.harvard.edu

General Observations from the think-aloud study:

Tester 1:

- Likes the minimalistic intro
- Thinks it's great that we have videos embedded
- Great intro slide, provides background about tiktok
- Sick button– great animation. Likes how the colors are tiktok colors. No complaints. Good info. Lots of effort put in. Doesn't like scroll on the 2nd slide
- Likes the colors of the bubble chart. Bad Bunny! Super satisfying that the bubbles moved. Type in your favorite artists instead of search. Post Malone is there! Make the highlight color not yellow, make it a bolder color (blends in with the white background). Bubbles cut off.
- Pretty great bar chart with the face plot. Clarify whether it's tiktok or spotify. Lots of white space
- Good explanations for song characteristics.
- Good transitions on the histograms. Explain what the histogram is telling us because not everyone might know
- Scatterplots– nothing to say. Looks good.
- Radar chart- need to fix the songs showing up. Doesn't really understand how to use radar chart. Needs more explanation.
- Piano – great piano. Keep it in.

Tester 2:

- Likes the style of the title page
- Loves the embedded tiktok videos, clicked on the videos which brought him to an external TikTok page
- The text might be too dense and weird looking because it's centered
- What does "together" mean on the bubble plot
- Artists with less featured songs → calculations are incorrect
- When searching an artist the light up feature more clear
- The text is too small in the tool tip
- Blinding lights is listed twice for the weekend
- Not clear that you can click the button
- Make the same x axis and y axis

- Didn't realize you could scroll through the songs

What does the tester like about your data story?

Tester 1:

- Tester 1 really enjoyed how different elements of TikTok were built into the visualization, including the embedded videos, TikTok logo, and use of TikTok colors throughout the website.

Tester 2:

- Tester 2 really liked the flow of the story and how we specifically looked into individual artists and then individual songs. The tester the transition pages with text that guide you through the website.

What improvements does the tester point out?

- Tester 1 believed that there is a lot of white space throughout the project. Some ways to improve that would be to add more elements, more text, or just make things bigger to take up the white space. There were also some small stylistic changes that they pointed out including keeping the font style consistent.
- Tester 2 said he wishes there were more hover effects. Additionally, he mentioned that the formatting, font and coloring should stay consistent throughout the project

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

- The intended message definitely was clear to Tester 1. They were able to scroll throughout the visualizations and come to the conclusions that we were hoping users would walk away with – that there are lots of similarities between songs on TikTok and songs on Spotify.
- Yes, the intended message was clear to Tester 2. The story led him through what the main takeaways were for each visualization. The instructions on each page guided him through how to use each visualization.

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

- Tester 1 understood the next steps throughout the visualizations. They understood where to click, what to click, and how to progress smoothly throughout the page. The only thing that may have troubled them was what to type in the search bar for the bubble page. The “Was your favorite artist included?” may have been too vague, for the tester typed “Yes” instead of their favorite artist. We plan to change the wording of this search bar thanks to the tester pointing out how it could be interpreted inaccurately.
- Tester 2 was able to get through all the next steps because the text on the slides told him which buttons to click. He was confused with the search bar to search up your favorite artist, similar to tester 1, however his confusion was due to the fact that his favorite artist lit up yellow and he didn’t notice the change. In the future we will make the light up more obvious.

- Based on the results of your ‘think aloud’ study, what would you improve in your data story?
    - We decided that we will try to lessen some of the text on our pages because the testers either complained that there was too much or got too distracted by the amount of text we had (and small font).
- 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?
    - The testers definitely got our takeaways! They all commented on how much insight was provided throughout the story, especially in the visualizations where the user was able to explore on their own (bubble chart, face plots, scatterplots).
- 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.
- Implement the intended changes and check them off your list (e.g., adding “done”). You can distribute the tasks among your team members. If you are unable to implement specific changes, please explain why and describe the expected results in your process book.

Features to implement after Tester Feedback

- ☑ ~~Title page background P~~
- ☑ ~~Text formatting (left align on page 2+3) M~~
- ☑ ~~Speed up transitions page 2+3 M~~
- ☑ ~~Larger button container page 3 M~~
- ☑ ~~Fix height and width of pages P~~
- ☑ ~~Center bubble plot M~~
- ☑ ~~Fix bubble buttons text M~~
- ☑ ~~Fix top artist calculation M~~
- ☑ ~~Hover over bubble M~~
- ☑ ~~Fix buttons for faceplot S~~
- ☑ ~~Add artist top song S~~
- ☑ ~~Add more bars to faceplot S (not sure if we need it now that we have music, your call Soline)~~
- ☑ ~~Fix the Weeknd’ 2 songs~~
- ☑ ~~Histogram bins need to be the same on both plots P~~
- ☑ ~~Labels for axes scatterplot P~~
    - ☑ ~~Make the axes the same scale P~~
- ☑ ~~Make text bigger for descriptions P~~
- ☑ ~~6 buttons instead of dropdown P~~
- ☑ ~~Remove initial songs for radar plot S~~
- ☑ ~~Add labels S~~
- ☑ ~~Make scrolling container clearer S~~
- ☑ ~~Box around text S~~
- ☑ ~~Clean up code~~

Final Features to implement

- [x] ~~More bouncing icons P~~
- [x] ~~Remove title background? P~~
- [x] ~~Fix font sizes P~~
- [x] ~~Page 3 phone fix M~~
- [x] ~~Condense text M~~
- [x] ~~Logo of what people are looking at on the face plot page S~~
- [x] ~~Sizing should be width of the screen for song characteristics M~~
- [x] ~~Adjust column sizes for histogram and scatter plot P~~
- [ ] Put all buttons into one button class M
- [x] ~~Change colors on radar plot S~~
- [x] ~~Change the opacity S~~
- [x] ~~Same font throughout (everyone)~~
- [x] ~~2 columns instead of scrolling for radar S~~
- [x] ~~Fix conclusion text → less dense M & S~~
- [x] ~~Add to sources list (everyone)~~

Needs fixing:

- [x] ~~Center song characteristics~~
- [x] ~~Make transitions more interesting~~
- [x] ~~Fix conclusion~~
- [x] ~~Add bouncing resource~~