CPSC 4030 - Milestone 2: Project Proposal
Due: Sunday, October 17th

Feedback from Federico:

> The topics looks very good to me. The number of rows is low but the attributes are enough to guarantee some complexity.
> If you have run already some EDA on the dataset and feel confident that there is a need to study multiple attributes (3,4,5, or more) together that will enhance the complexity of your future visualizations/website that's fine with me.
> The first question you list there sounds cool (the second one too, btw) but it is one of the many questions your website should address (and with many I mean at least 2 or 3). Ideally you will have one viz per question.
> The first question (with all the sub questions you listed) is perfect. But be aware that that won't be enough. Grow a little bit the scope proposing other aspects that would be worth being explored.
> (That is something to be addressed in the next milestone. For now you are good)

Table of Contents:

Basic Info: 1

Background and Motivation: 2

Project Objectives: 2

Data: 2

Data Processing: 2

Visualization Design: 2

Must-Have Features: 2

Optional Features: 2

Project Schedule: 2

Basic Info:

Project Title: Tune Tracker
Link to the project repository
Team Members:

- Rebecca Kahn, rkahn@g.clemson.edu, C14523233

- Jake Liguori, jrliguo@g.clemson.edu
- Cierra Oliveira, colivei@clemson.edu, C73041981

Background and Motivation:

- *Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.*

- All three of us are users of a music streaming service, and we all regularly listen to music. Because listening to music is a part of our daily lives, we are interested in what drives us to listen to the songs we do. Are the popular songs we encounter a result of our region in the U.S.? Or are there features of popular songs that make them more likely for us to encounter them? We are motivated mostly by our own curiosity around the reasons why music becomes popular and where popularity extends.

Project Objectives:

- *Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits.*

- We have 2 primary questions, which can each be answered by several sub-questions such as the following:
    1) What makes a song popular?
        a) How do beats per minute affect popularity?
        b) How does danceability affect popularity?
        c) How does genre affect the length of popularity?
        d) What genres are the most popular and how has it changed over time?
        e) Of the top genres, what are the top artists?
    2) How does music popularity differ by U.S. region/city?
        a) Do different genres dominate different regions?
        b) Do specific artists top the charts in specific regions?
- We hope to use these questions to learn about the nature of music popularity in the US. Insights we gather can help us understand past and anticipate future trends in music. The knowledge we gather could be potentially useful for those in the music industry who are interested in reaching the charts. Our research could also serve fellow music appreciators who wish to learn more about music trends.

Data:

- *From where and how are you collecting your data? If appropriate, provide a link to your data sources.*

- For our first question, we are using a dataset from Kaggle: Spotify Top 200 Charts (2020-2021)
    - This dataset contains all songs that appeared on the Spotify Global Top 200 chart from January 2020 - August 2021. It is hosted on Kaggle and was originally scaped using the Spotify API and spotifycharts.com. From Kaggle, we are able to download a CSV with all 23 attributes and 1,556 items.
- For our second question, we are scraping city chart playlist data from Apple Music. To do this, Python is used along with the BeautifulSoup library to scrape the info from each playlist. Then, the Pandas library is used to combine the data into a table and export it to CSV to then be loaded in D3. We have created this process already, and will finish processing the rest of the data later if it is going to be used.

Data Processing:

- *Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?*

  We do not expect much data cleanup. This data comes directly from Kaggle and is already in a tabular format with clean values, making it easy for us to extract and use data. One instance of cleanup we need is with dates. The data set provides dates in a week format (e.g., 2021-07-23--2021-07-30) and we need to extract months. Additionally, the genre cell for each song contains a list of genres (e.g., ['dance pop', 'pop']), meaning we can clean this attribute by generalizing the songs into broader genres (pop, rock, hip-hop, etc.) or leave it as is and let each subgenre be its own category. Data processing will be implemented through D3, which will read and store the data. Afterwards, helper or accessor functions will be written to extract the relevant portion of the element and convert it to the right data type, but in almost all attributes the data is ready to be plotted.

  For our backup/supplemental regional data, this data needs to be scraped from Apple Music's city playlists. The scraping process is described above. After, the cleaning only involves some text cleanup (some unwanted characters, only certain portions needed, etc.). This ensures for each song, we can extract the name, artist, genre, city, and chart position only.

Visualization Design:

- *How will you display your data? Provide some general ideas that you have for the visualization design. Develop three alternative prototype designs for your visualization.*

Create one final design that incorporates the best of your three designs. Describe your designs and justify your choices of visual encodings.

- **Q1: What makes a song popular?**
    - Q1-1: How does genre affect the length of popularity?:
        - To display data about the length of popularity, we will sort our data by genre and use the “Weeks Charted” attribute to discover how long, on average, a genre remains on the charts.
        - We plan to use a bar chart for this visualization. Genre is a categorical variable and each genre type will use an area channel. The different genres will be distinguished by their horizontal position on the x-axis and color hue. The number of weeks a genre is charted will use a vertical position channel. In other words, the height of a bar for a given genre will indicate the number of weeks charted.
        - See sketch Q1-1 for a draft visualization

    - Q1-2: Of the top genres, what are the top artists?
        - To display data about the top artists per genre, we will sort the data by genre and collect the top 5-10 artists.
        - We plan to use lines to connect genres to its top artists. Because some songs are categorized with multiple genres, we have to use a visualization that allows for multiple genres to connect with a singular artist. We will distinguish genres by color hue and draw colored lines between a genre and its top artists. Multiple lines can lead to a singular artist, indicating that it’s a top artist for more than one genre.
        - See sketch Q1-2 for a draft visualization

    - Q1-3: Correlations between danceability, acousticness, etc and streams or top charting position.
        - For this visualization, we have designed a grid of six scatter plots which show danceability, acousticness, energy, instrumentalness, liveness, and valence. Every attribute is on the same 0-1 score scale on the x axis and the same streams scale on the y axis, allowing us to quickly analyze each attribute’s correlation with streams. The scatter plot points also provide an easy way to add interaction later to click and see song info. Additionally, above or below each scatter plot is a distribution plot. Some attributes may not be normally distributed among songs in the charts, so this part of the visualization will show if an attribute is generally high or low for charting songs.

    - Q1-4: See how top genre changes of time
        - To display data displaying the changes in the top genre over the designated time period in our data, we wanted to display it in a creative way that can easily distinguish between the various genres. We used various icons to differentiate between the genres with colors as well to be able easily tell which genre it is.
        - We observe the song with the highest charting position and compare it to the number of weeks the song was on the top charts.
        - Our timeline views the data every two months to observe any trends in genres during certain times of year.
        - See sketch Q1-4 for a draft visualization
- **Q2: How does music popularity differ by U.S. region/city?**
    - Q2-1: What are the top genres per region?
        - To display data about how genres differ by region, we will use the U.S. map and mark the top genre spatially. The region will use vertical and horizontal position channels. The top genre will be marked by a colored circle, where the genre type is distinguished by the color hue of the circle.
        - See sketch Q2-1 for a draft visualization
    - Q2-2: What are the top artists per region?
        - To display data about how top artists differ by region, we will use the U.S. map and mark the top artists spatially. The region will use vertical and horizontal position channels. The top artists for a given region will use a bar chart, where the vertical position of the artist name indicates the placement on that region's charts. The bar for each artist will use horizontal position to indicate the number of times charted in that region.
        - See sketch Q2-2 for a draft visualization

Must-Have Features:

- *List the features without which you would consider your project to be a failure.*
- ***Q1-3: Once you click a song/datapoint, its highlighted on the other charts***
- ***Q1-3: when you click a song/datapoint, the name and artist name shows up***
- ***Q1-3: sort by genre***
- ***Q1-3: hover over an attribute name like "danceability" and the definition shows up***
- ***Q1-2: filter by time***

One feature we need to have is for our scatter plot visualizations (Q1-3), clicking on a data point in one plot should result in the point for that song being highlighted (color changed) in every other plot. This will allow the user to see where a specific song lands in each section of the visualization. Similarly, clicking or hovering over a data point should display the song name, artist, and possibly additional information. For the same visualization, we would also like to

filter or sort the data by genre. This will allow the user to analyze one specific genre's correlations with the attributes. Finally, for this visualization we will also add a hover effect to show a description of the attribute. Some of these attributes are a bit vague, so this will give the user a better understanding of what they are viewing.

For our time based visualizations, such as Q1-2, changing the time period is a must-have feature. This will allow the user to focus on a specific month, year, season, or just all the data available.

Optional Features:

- *List the features which you consider to be nice to have, but not critical.*
- ***Q2-½: Once you click a song/datapoint, it brings you to that spotify song/artist***
- ***Q1-4: have the data points show up over time / play with movement***
- ***General song look up?***
- ***When you select a song, it begins playing within the website***

Some features we considered to be nice to implement in our project but are not essential for the visualization would be to be able to click on a song (datapoint) and that link takes you to that specific spotify song and artist (Q2-1&2). Another visualization that is not critical would to have the data points show up over time and display with movement rather than already having the chart displayed on the screen to show how the points come up over time (Q1-4). Another feature for the site that would benefit our visualization would be being able to do a general song lookup and view any information our website may have from our data regarding its popularity and position in the charts. A final feature we were considering is to be able to select the specific song/genre displayed, and having the song play within our website. While all of these may be great features to have with our visualization, we do not believe they are necessary for properly displaying our data.

Project Schedule:

- *Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.*

Deadlines

10/17 - Project Proposal

10/18 - 10/24

- *Jake - Begin working on designated questions in creating the visualization on the website (Q1-3 & Q2-1)*
- *Cierra - Begin working on designated questions in creating the visualization on the website (Q1-1 & Q1-2)*
- *Rebecca - Begin working on designated questions in creating the visualization on the website (Q1-4 & Q2-2)*

10/25 - 10/31

- *For all 3: Add more stylestic designs that align with our color palette, fonts, and organization on the website for the specific questions*

11/1 - 11/6

- *Jake - Add interactive components to designated visualizations*
- *Cierra - Add interactive components to designated visualizations*
- *Rebecca - Add interactive components to designated visualizations*

11/7 - Project Prototype

11/8 - 11/13

- *Jake - Critiques the prototype's visual encodings*
- *Cierra - Critiques the prototype's interaction techniques*
- *Rebecca - Critiques the prototype's design quality*

11/14 - Peer evaluation

11/15 - 11/21

- *All 3 of us: review our peer evaluations and make the necessary changes*

11/22 - 11/28

- *All 3 of us: finish making the changes and make sure all the other interactive components still work*

11/29

- *All 3 of us: tidy up any last minute changes and practice delivering the presentation*

11/30 or 12/2 - Oral Presentation

11/30 or 12/2 - 12/4

- *Jake - Check over the final website and make sure group members visualizations are working*
- *Cierra - Organize splitting up the project report among the group members*
- *Rebecca - Organize and prepare the screen-cast to be ready for the group to record*

12/5 - Final Delivery and Peer Assessment

For Skeleton:

- **Sketch of the layout**
- **Title of Project**
- **About the team**
- **The question we are asking, the subquestions**
- **A description of data**
- **SVG elements for visualizations**

Links:

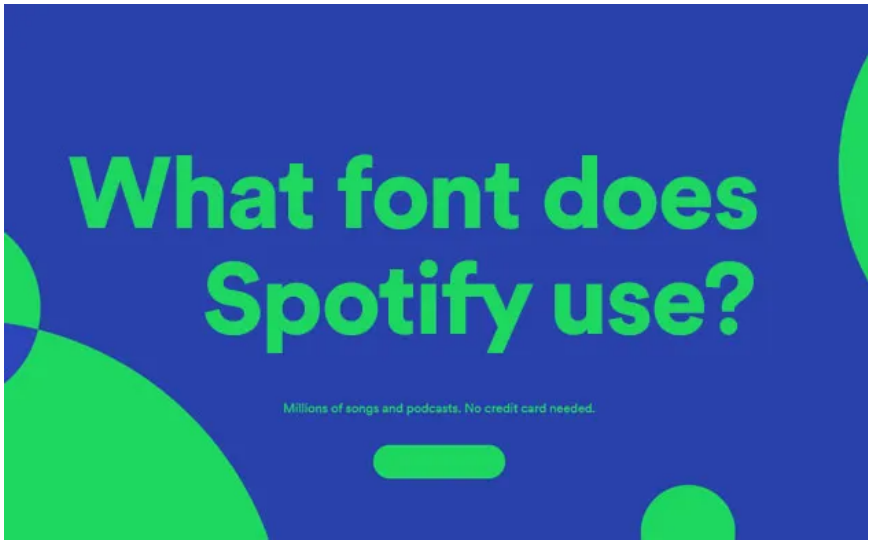
https://www.behance.net/gallery/47444741/Music-Listening-Data-Visualization
https://www.tableau.com/about/blog/2019/7/how-i-found-inspiration-data-behind-my-favorite-music
https://visual.ly/community/Infographics/entertainment/musical-data-visualization

https://www.tableau.com/community/music/spotify-top-tracks
https://searchbusinessanalytics.techtarget.com/news/252465871/Music-data-visualization-comes-to-life-in-Tableau-collection

Sketch Tasks

- 3 Layouts of the page for q1
  - Include visualization for each sub-question
    - Cierra: q1, 2
    - Rebecca: q4, layout rearrangement
  - Change visualization / layout arrangement for each layout
- 3 Layouts of the page for q2
  - Cierra: q1,q2

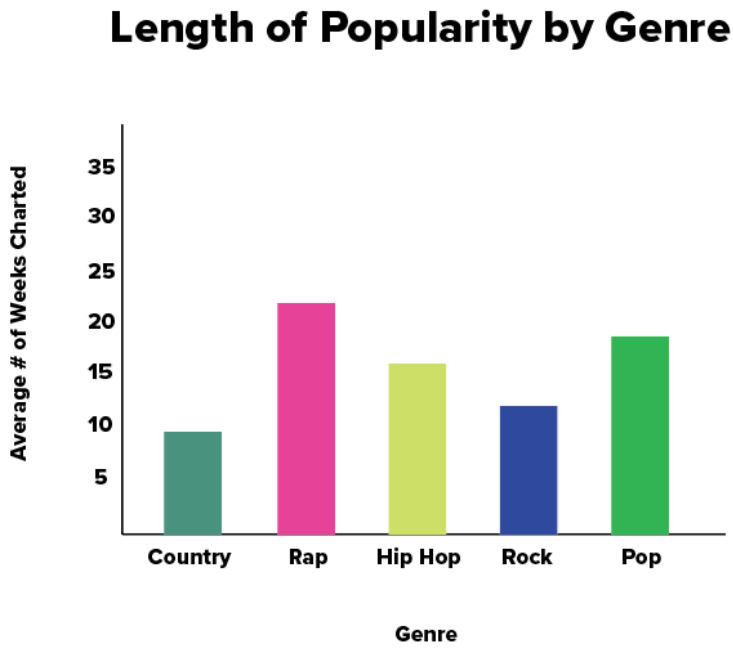
Font and Color Information



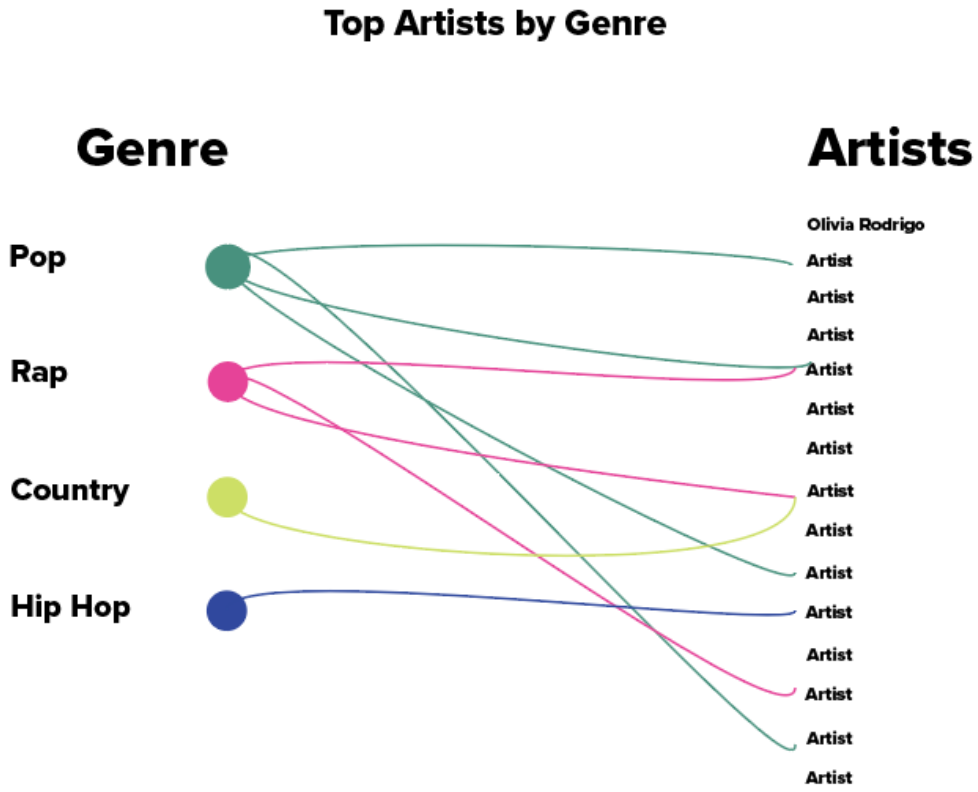
Info about font similar to spotify's
https://arturth.com/spotify-font/

Layout Inspo:
https://gabyyy.com/spotify

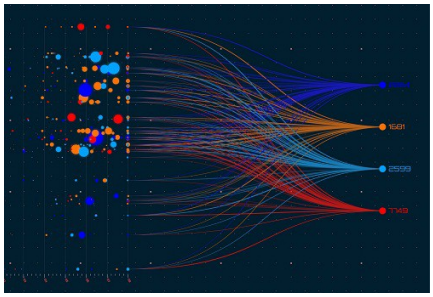
Sketches:
Q1-1

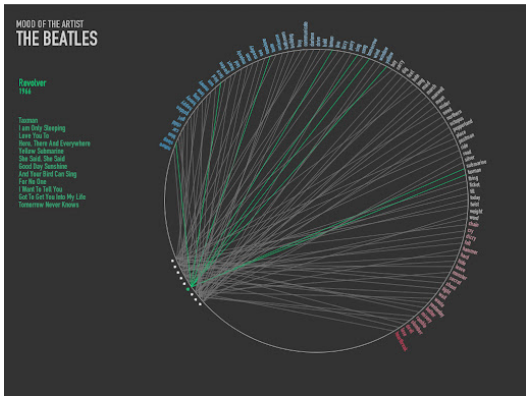


Q1-2



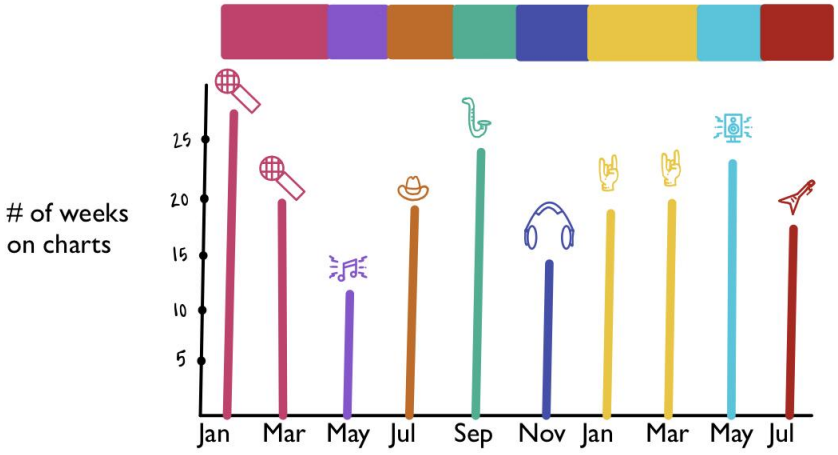
(Inspired by these)



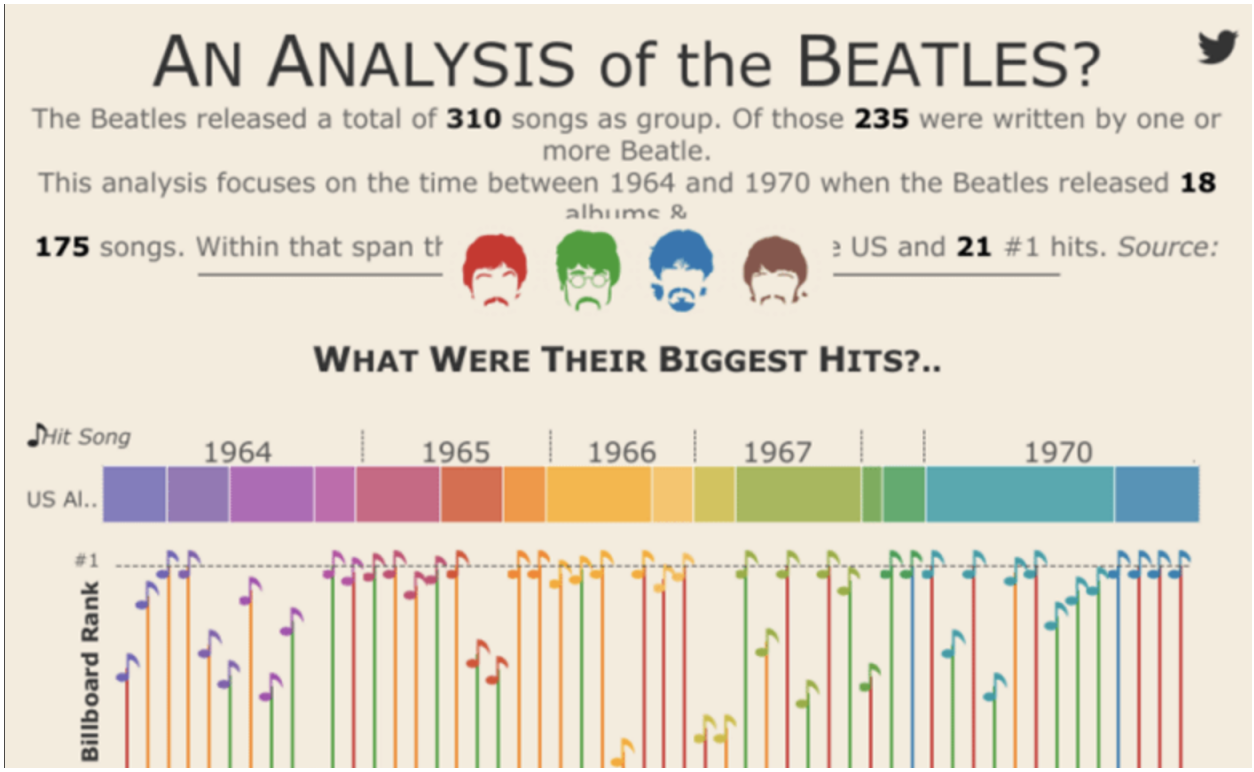


Q1-4

1. Pop
2. Hip-hop/rap
3. Rock
4. Dance/EDM
5. Latin
6. Indie/alt rock
7. Classical
8. K-Pop
9. Country
10. Metal



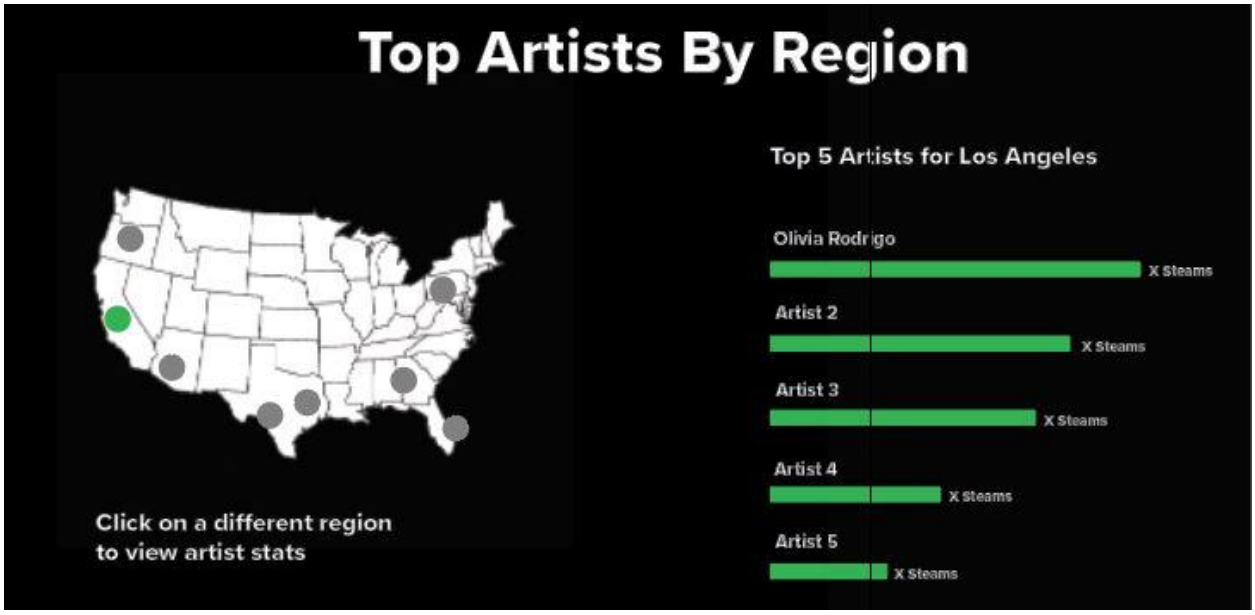
(Inspired by this):


AN ANALYSIS of the BEATLES?
The Beatles released a total of 310 songs as group. Of those 235 were written by one or more Beatle.
This analysis focuses on the time between 1964 and 1970 when the Beatles released 18
175 songs. Within that span th
US and 21 #1 hits. Source:
WHAT WERE THEIR BIGGEST HITS?..
Hit Song
US Al..
1964
1965
1966
1967
1970
#1
Billboard Rank

Q2-1

Top Genre By Region
Pop
Rap
Country
Hip Hop

Q2-2

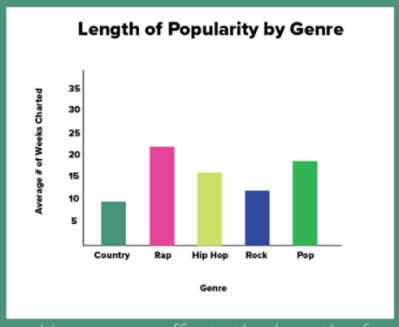

Top Artists By Region
Top 5 Artists for Los Angeles
Olivia Rodrigo
X Steams
Artist 2
X Steams
Artist 3
X Steams
Artist 4
X Steams
Artist 5
X Steams
Click on a different region
to view artist stats

Layouts

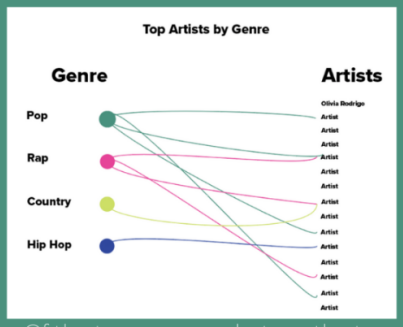
Tune Tracker

Rebecca Kahn, Jake Liguori, Cierra Oliveria

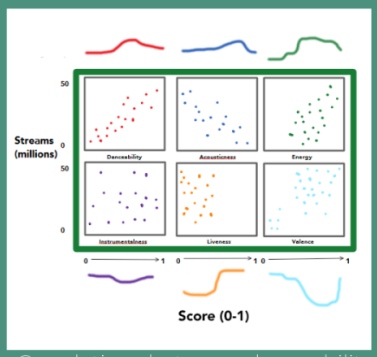
What makes a song popular?



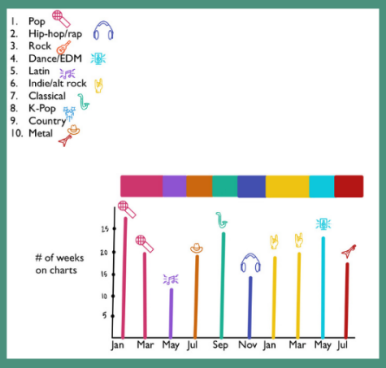
How genre affects the length of popularity



Of the top genres, what are the top artists



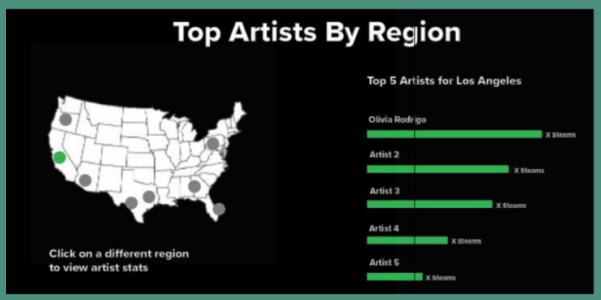
Correlations between danceability, acousticness, energy, instrumentalness liveness, and valence and streams



How does top genre change over time and how long do they stay on the charts

How does music popularity differ by U.S. region/city?

What are the top genres per region

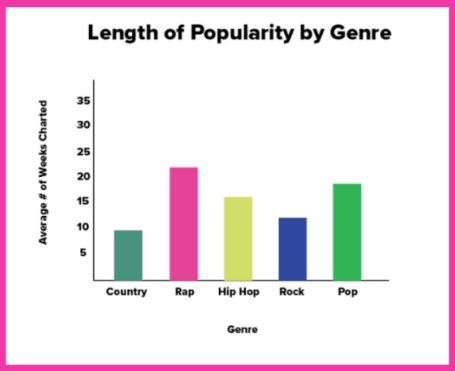


What are the top artists per region

Tune Tracker

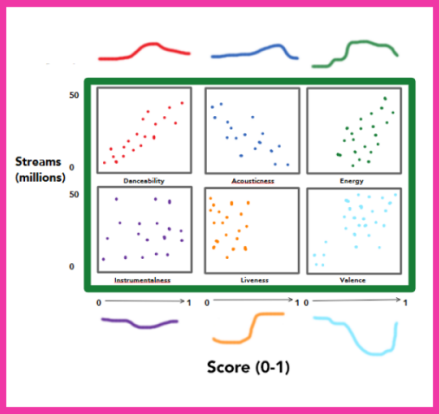
Rebecca Kahn, Jake Liguori, and Cierra Oliveria are all seniior Computer Science students in Data Visualization 4030. All three are advid users of a music streaming service and regularly listen to music. They are interested in what drives them to listen to the songs they do. From their curiosity in music popularity, they created these data visualizations to display their findings.

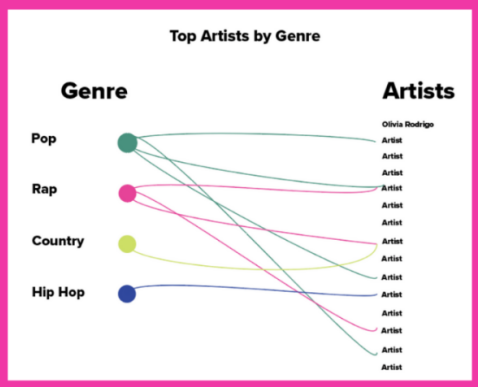
What makes a song popular?



How genre affects the length of popularity

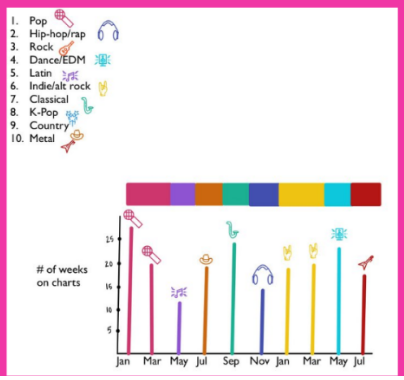
Correlations between danceability, acousticness, energy, instrumentalness liveness, and valence and streams





Of the top genres, what are the top artists

How does top genre change over time and how long do they stay on the charts

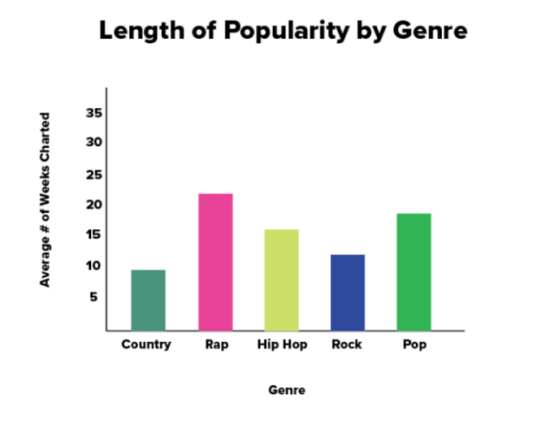


Rebecca Kahn, Jake Liguori, and Cierra Oliveria

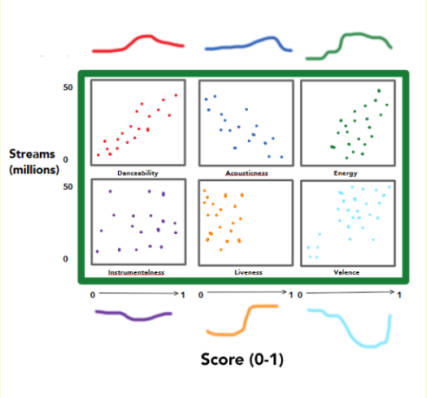
♪ Tune Tracker

What makes a song popular?

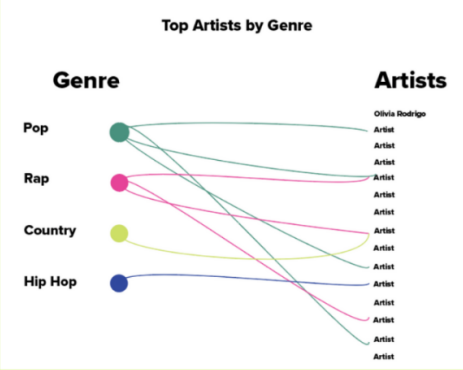
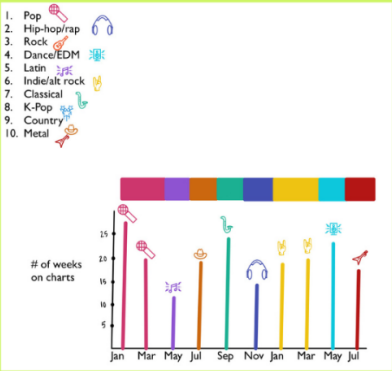
How genre affects the length of popularity



Correlations between danceability, acousticness, energy, instrumentalness liveness, and valence and streams



Of the top genres, what are the top artists



How does top genre change over time and how long do they stay on the charts