Milestone II - Project Report

Group: Full-Mark Musician

Group Member: Linda Huang, Jialu Jin, Minghao Wang

Project Name: Spotify Data Exploring

Cohesion

These visualizations are for audiences who are interested in understanding various aspects of music, such as trends, popularity, and relationships between different variables. These visualizations appeal to music industry professionals, music researchers, academics, enthusiasts, and general consumers. The visualizations provide valuable insights for decision-making, analyzing trends, and understanding the relationships between different musical elements, such as danceability, duration, mode, loudness, energy, key, and explicit content.

It will be beneficial to audiences because it enhances comprehension, and the interactive visualizations allow users to explore and understand complex relationships between various variables, such as track genre, duration, danceability, popularity, mode, loudness, energy, key and explicit content. This helps audiences gain a deeper understanding of the data and the patterns within it. By enabling users to select specific data points, filter data, or adjust parameters, interactive visualization empowers audiences to explore the data in a way that is most relevant and interesting to them. This makes the visualizations more engaging and useful for a diverse range of users. Advanced interactions like brushing and linking, uni-directional interactions, bi-directional interactions, and drop-down selections can help break down complex data into simpler and more manageable pieces. This helps audiences more easily grasp the relationships and trends within the data. It also facilitates the discovery of hidden insights and patterns that may be not immediately apparent through static visualizations. As users explore the data and adjust parameters, they can uncover new relationships and trends that can inform decision-making and drive further research. Most importantly, by providing users with the tools to actively explore and manipulate the data, interactive visualizations encourage deeper engagement with the content. This increased engagement can lead to better retention of the information presented and a more profound understanding of the subject matter.

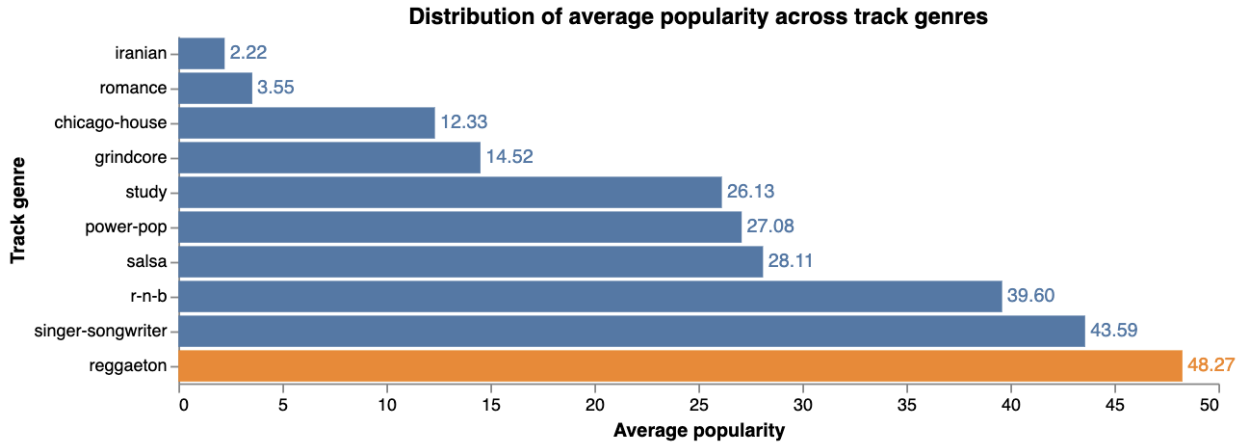
The tasks that were addressed are:

1.) Which track genre has the highest average popularity?
2.) For different intervals of duration, does the relationship between danceability and popularity change?

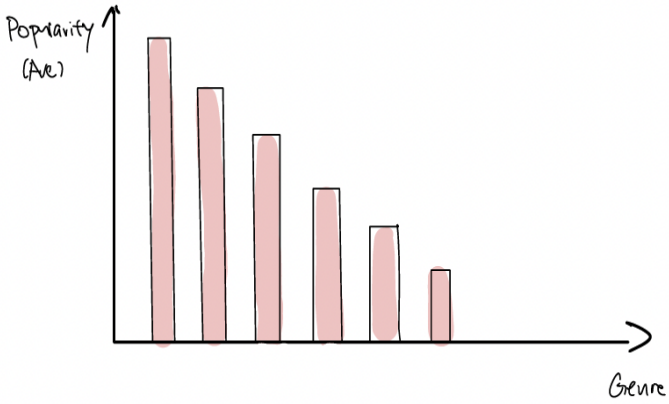
3.) Is there a difference in the relationship between popularity vs. danceability from different distributions of major and minor modes?
4.) Is there a relationship between danceability and loudness/energy in different genres?
5.) Does the key of the track influence the relationship between danceability and popularity?
6.) Which combination of the key and explicit types of the track sounds most positive?

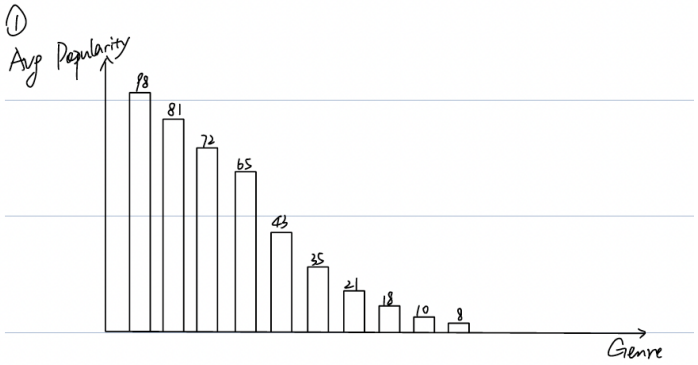
Task 1: Which track genre has the highest average popularity?

Final View Screenshot:



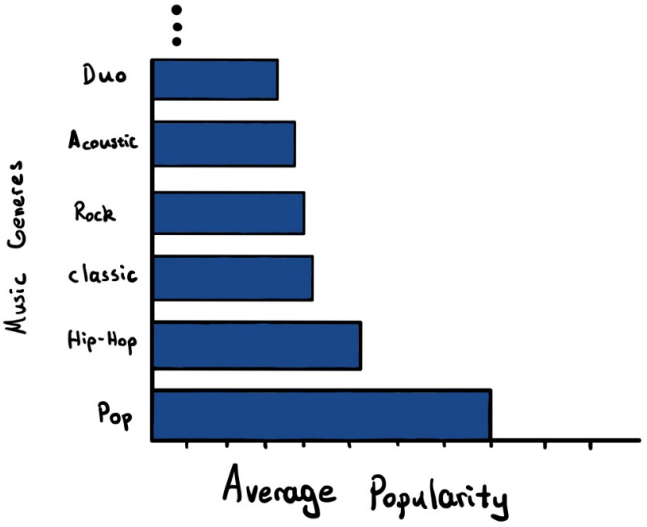
Previous Iterations:

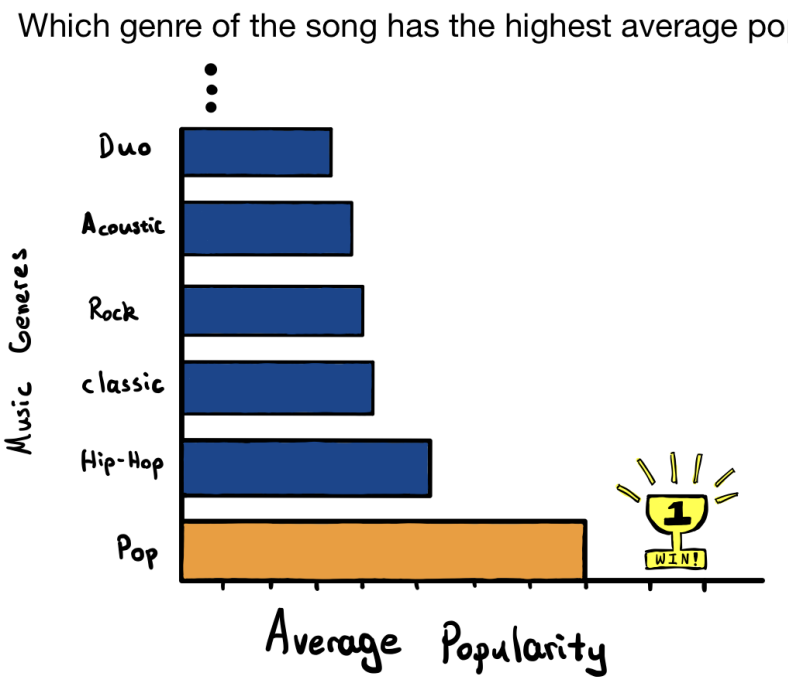
Iteration Stage	Visualization/Sketch Design
Draft	Linda:  Cindy:



Minghao:

Which genre of the song has the highest average popularity?



Preliminary	

Explanation of visualization choices from the theoretical principles you have been exposed to

Previous Iteration: In milestone I, we have decided to use a bar chart with color highlighting to visualize the task as shown in the preliminary sketch. The main reason to choose a horizontal bar chart was due to its practicality when dealing with lengthy category labels, making it easier for viewers to compare relative length bars. The chart has a y-axis that represents the various music genres, with the x-axis representing their respective average popularity scores.

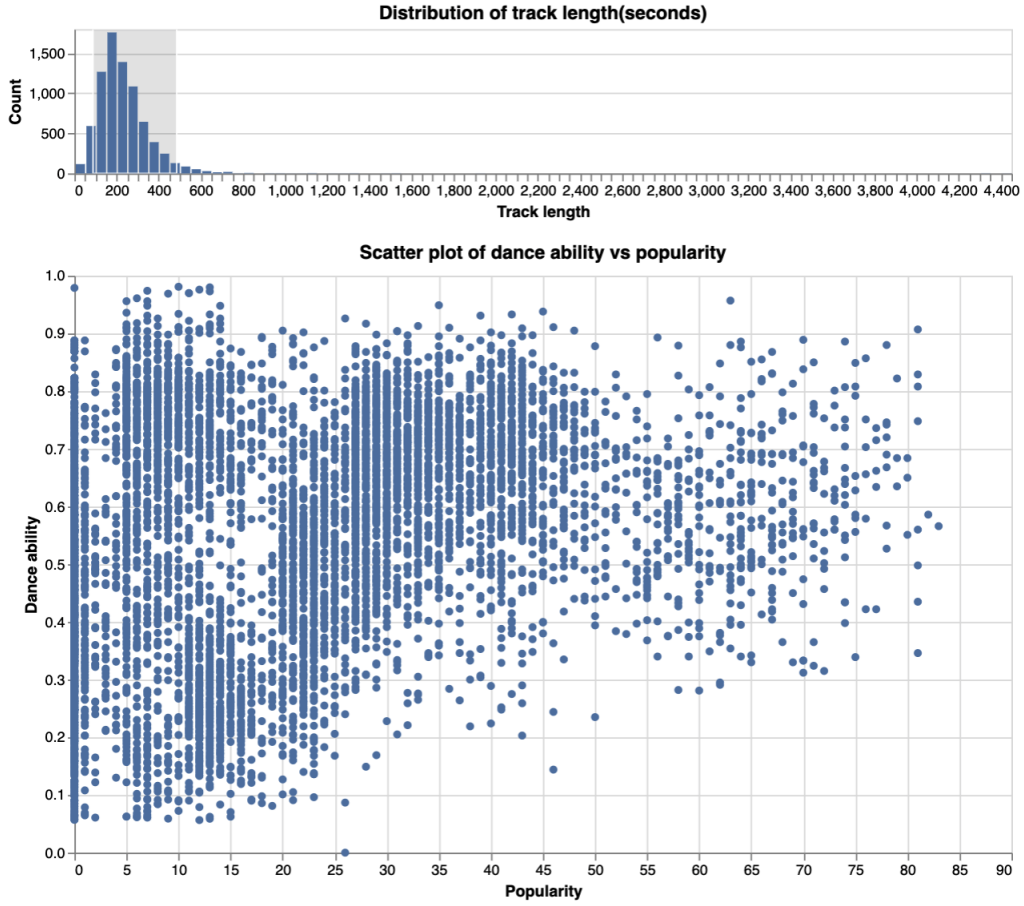
For the visualization we created above, we used color and position as the channel and line as the mark. The chart has a y-axis that represents the various music genres, with the x-axis representing their respective average popularity scores. In order for easier readability we have also colored the bar of the genre with the “highest average popularity” for popout. The additional color channel will also increase separability for readers to perceive the exact “highest” bar if the top two bars have similar numbers. In terms of discriminability, we only have 1 unique step to perceive the color channel. And to increase the accuracy of information read by views we’ve added the exact average number on top of each bar for accuracy, so that users can observe the info right away, being effective and efficient.

The horizontal bar chart is a highly effective visualization choice when determining which music genre has the highest average popularity. It provides a straightforward and efficient way to compare various categories. The chart's intuitiveness and ease of interpretation make it an

excellent tool for conveying data to a broad audience. According to the principle of proximity, the bars in the plot are grouped together. The bar chart above ensures to group of the “highest” bar is a different category from the rest by color. High continuity is achieved with the bars plotted continuously without gaps.

Task 2: For different intervals of duration, does the relationship between danceability and popularity change?

Final View Screenshot:



Explanation of visualization choices from the theoretical principles you have been exposed to:

- Scatter plot of Danceability vs. Popularity displays the relationship between Danceability and Popularity through a scatter plot with Danceability represented on the y-axis and Popularity on the x-axis.

- Distribution of Duration illustrates the distribution of Duration using a vertical bar chart that displays the total count for each time duration of the songs.

For this view, we chose to represent the connection between danceability, duration, and popularity using one scatter plot and a bar chart that allows for linked interactions. The advanced interaction, with brushing and linking, will allow users to find out the change of correlation with different ranges of song durations. When an interval of the song duration is selected on the bar chart (ex. 60s - 120s), the scatter plot will show the danceability and popularity of the songs within the time duration. The visualization is easily comprehensible and readable for users to explore the change of relationship when a different range is selected.

The visualization has uniform bar width and shape for all attributes and an optimal number of attribute levels for easy perception and comparison of data. No colors were used to group any classes, making the discriminability better with only 2 steps needed to find out about the x & y position when reading the scatter plot.

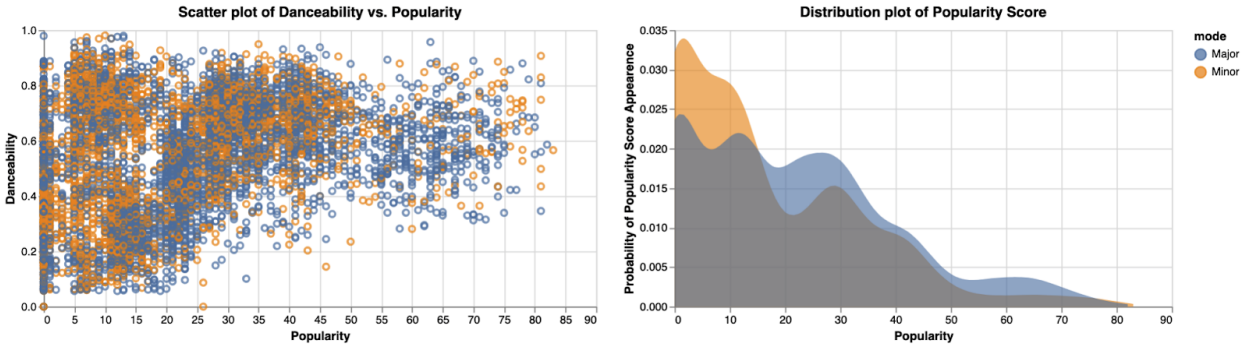
The interactive design empowers users with data control, enabling them to manipulate, explore, and prioritize aspects of their interests. The design also provides instant feedback, helping users comprehend the impact of their actions on the data. Additionally, interactivity caters to different learning styles and preferences, enabling users to interact with the data in a way that aligns with their understanding.

Critiques:
For the distribution plot, by applying the principle of proximity, bars are not grouped into categories, and a distinct color is not used to highlight any bar. To ensure continuity, the bars are plotted continuously without gaps. With the scatter plot, proximity was not achieved as no grouping was integrated for any classes. However, continuity and closure are satisfied as there are no gaps between points and it is not difficult to view each data point on the plot.

While brushing and linking advanced interactions can be useful in facilitating the analysis of complex data, it is necessary to provide clear instructions as some users are not familiar with the underlying data or the visualization tool. Users can explore on their own based on the information they want to obtain within certain intervals. However, users won’t be able to know if there’s an interaction without further instructions, since advanced mouse interaction is needed (like slide and drag).

Task 3: What is the difference in the distribution of popularity and the relationship with danceability between major and minor modes?

Final View Screenshot:



Explanation of visualization choices from the theoretical principles you have been exposed to:

In the view above, there are two different plots linked together, representing the distribution of popularity and relationship with danceability.

- The scatter plot of danceability vs. popularity displays the relationship between Danceability and Loudness through a scatter plot with Danceability represented on the y-axis and Loudness on the x-axis.
- The density plot of popularity shows the difference in the distribution of popularity score among minor mode and major mode.

For the scatter plot, we used point as mark, and position channel to represent the numerical attributes along with color channel to group between modes. The mark for the density plot is area and the channel used is the size (area). The color channel is also used in this plot to represent the two different modes. By observing the height of the density plot curve, viewers can determine the relative frequency of popularity values for each mode, and identify any differences in the density of values between the modes.

The use of color in both density plots and scatter plots will help with grouping and separability. Viewers will be able to easily distinguish every data point between modes with the color chosen for each mode. We have only one unique step to perceive the color channel and 2 steps to read the position of each data point. While the rest of the channels are well-used, the area channel in the density plot might have issues with accuracy as viewers will not be able to tell the exact density of the popularity for each score. Moreover, it is harder for users to interpret the meaning of “density”. On the other hand, it is a good visualization to compare the difference in distribution between two different classes of attribute “mode”.

The interaction used in this view will allow users to manipulate, explore and focus (or reduce) on the aspects they find most interesting. By integrating on-click-selection to the density plot, users will be able to click on the “mode” class they want to view, which will immediately change visual encoding for selection targets, reducing the unselected area’s opacity. At the same time the scatter plot will filter out any songs that are not in the mode selected. With the linked interaction, users will be able to easily compare and understand the relationship between variables and focus on the song mode that they are interested in.

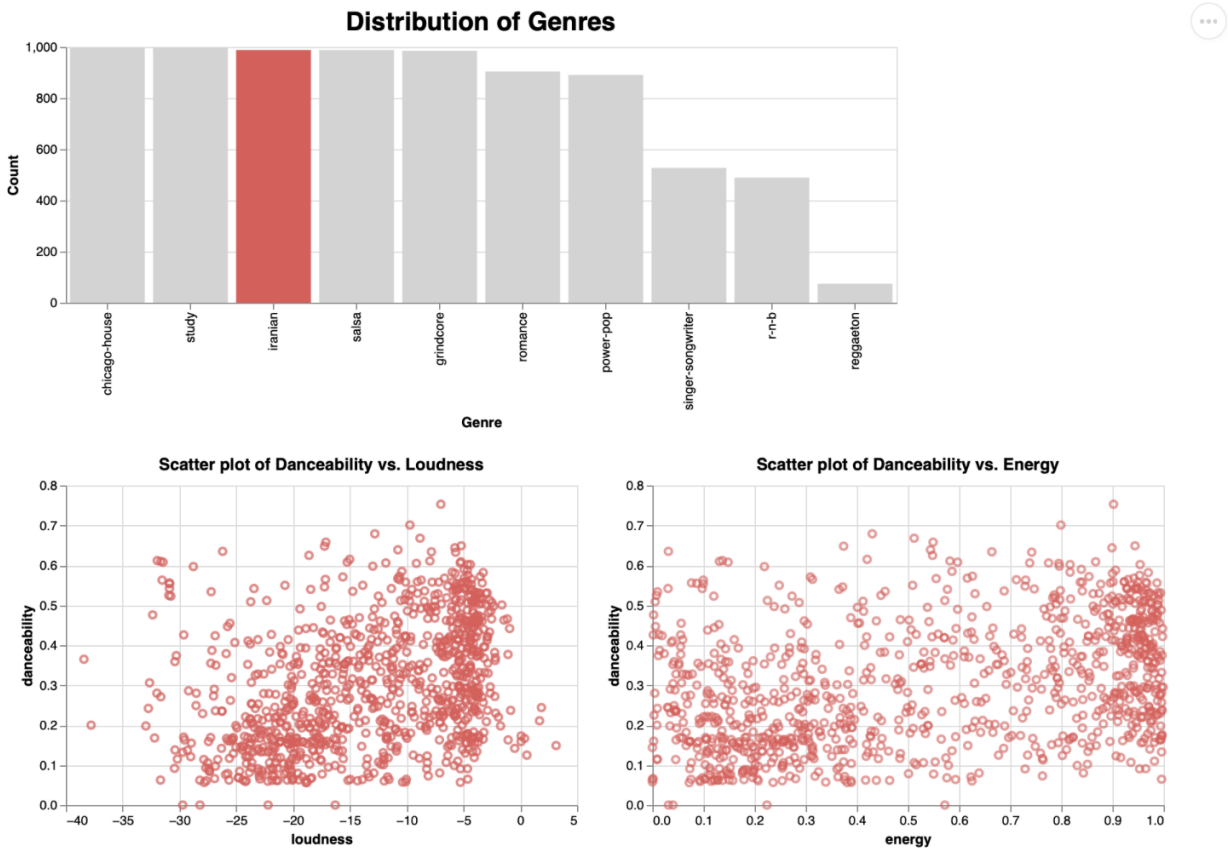
When it comes to density plots, the principle of similarity is relevant since the smooth curves that show the distribution of popularity for major and minor modes have a similar visual shape and form. This plot uses two different colors, one for each mode, which is within the recommended limit of five colors and ensures easy interpretation without any confusion. Furthermore, using unique color hues for each mode helps them stand out, enabling users to identify them quickly.

With the scatter plot, proximity was achieved to group related points together, and visual similarity is high with different colors representing different “mode” classes. However, continuity and closure could be improved as there are no gaps between points and it is difficult to view each data point with all of them clustered together.

While the linked interaction is useful when comparing between classes, is it important to also include information on the plot and how to read/compare as linked interaction with on-click-selection could be complex for first-time viewers. Users will be able to explore the visualization by click-interaction to select what information they would want to obtain.

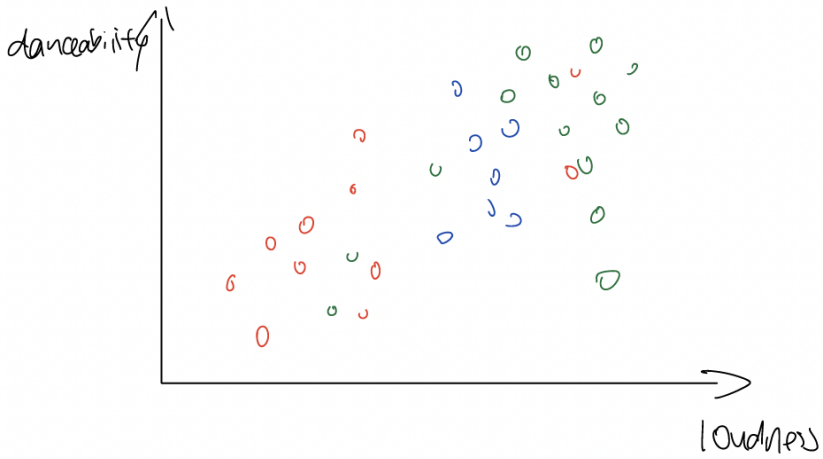
Task 4: Is there a relationship between danceability and loudness/energy in different genres?

Final View Screenshot:

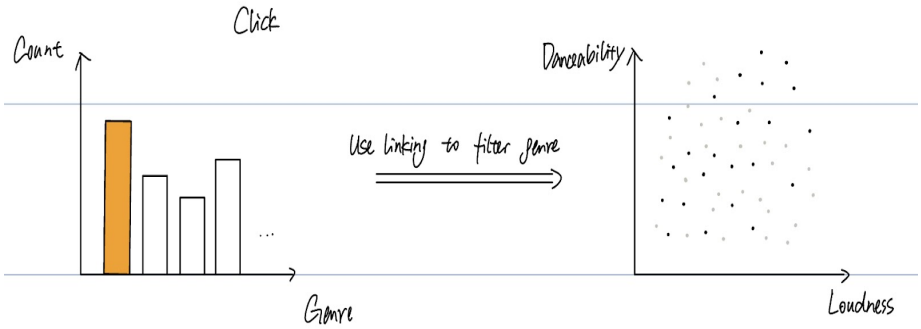


Previous Iterations:

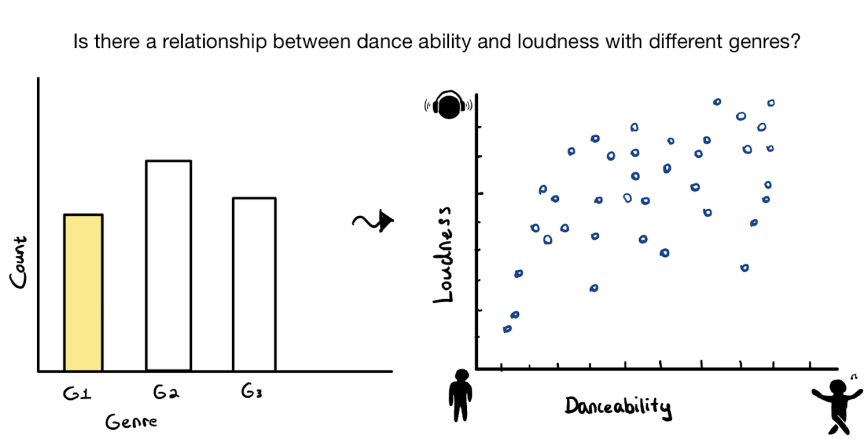
Iteration Stage	Visualization/Sketch Design
Draft	Linda:

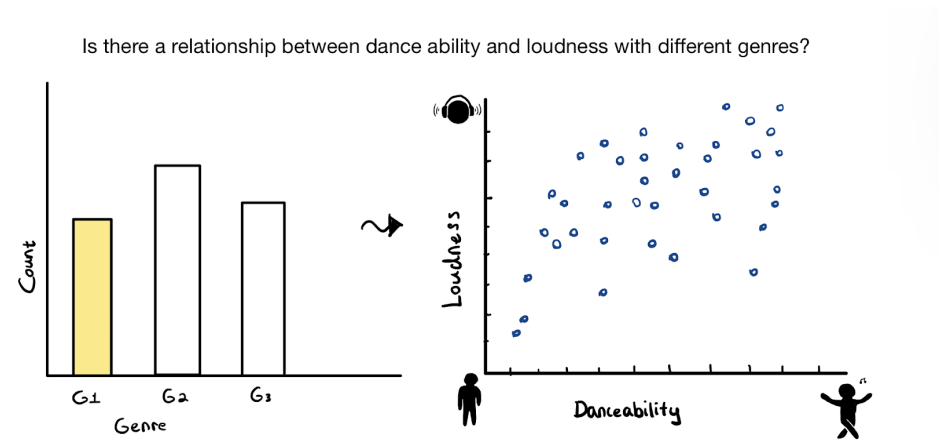


Cindy:



Minghao:



Preliminary	

- Scatter plot of Danceability vs. Loudness displays the relationship between Danceability and Loudness through a scatter plot with Danceability represented on the y-axis and Loudness on the x-axis.
- Scatter plot of Danceability vs. energy shows the relationship between Danceability and Energy through a scatter plot with Danceability on the y-axis and Energy on the x-axis.
- Distribution of Genres illustrates the distribution of genres using a vertical bar chart that displays the total count for each genre.

Similar to the preliminary visualization, we chose to represent the connection between danceability and loudness/energy using two scatter plots and a bar chart that allow for bi-directional interactions. A mouse-based interaction is linked to a filtering interaction that enables the data to be filtered by a selected music genre. This makes it simpler for audiences to identify any correlations between the two variables. The visualization is easily comprehensible and readable, making it a valuable tool for conveying the data to a diverse audience as it allows for quick comprehension of the relationship.

The visualization is designed with uniform bar width and shapes and an appropriate number of attribute levels to ensure that it is easy to perceive. The use of color hue and position makes the visualization separable and helps distinguish the differences between each genre, moreover, increases the accuracy. Upon selection, the rest of the genres are filtered out, making the selected genre stand out. This visualization also exhibits proximity and grouping.

To enhance communication with the audience, we integrated linking interactions along with mouse interactions like clicking and tooltips. A bar chart was utilized to display the count (y-axis) associated with each genre (x-axis). Two scatter plots were employed, one showing the relationship between danceability and loudness and the other, the relationship between danceability and energy. By using bi-directional interactions and mouse-based interaction, it

enables users to actively engage with the data. When a user clicks on a bar, the visualization highlights the selected genre and displays the relationship between danceability and loudness/energy on two different scatter plots. This allows users to explore the data and understand the relationships between the variables more effectively. This type of interaction allows users to focus on the genres they are most interested in, rather than being overwhelmed by all the information at once. They can choose to explore one genre at a time, making the experience more personalized.

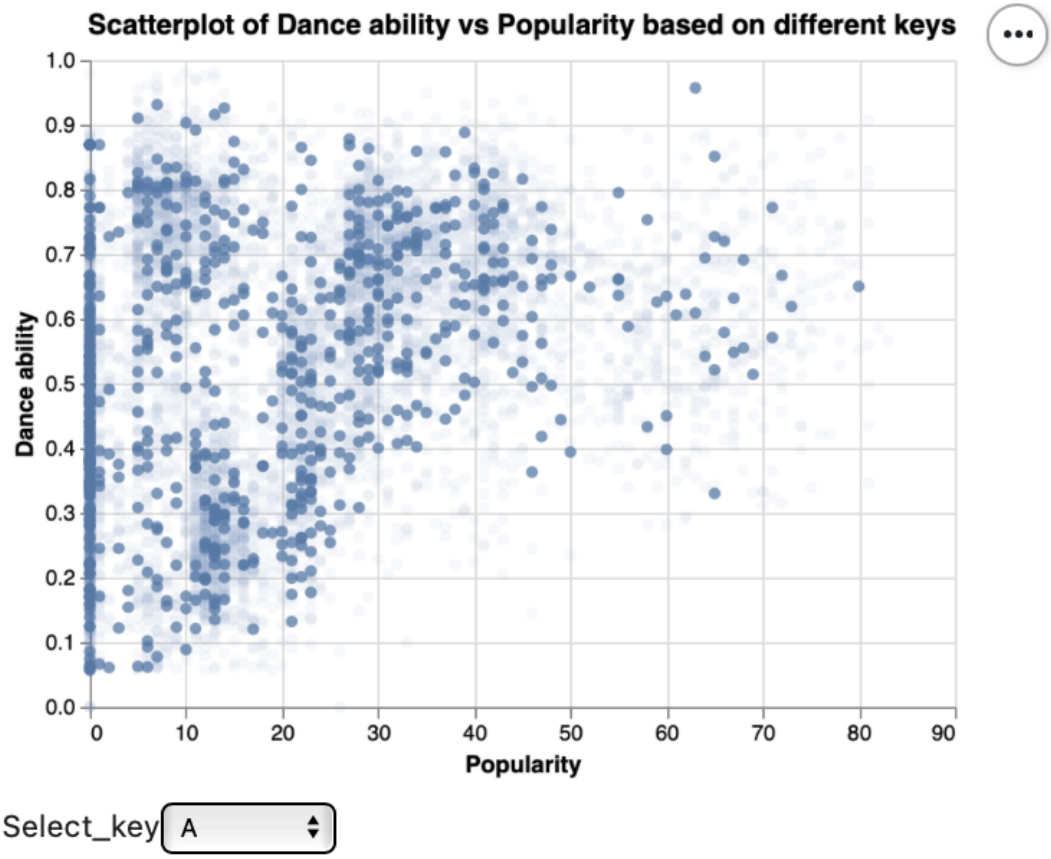
The interactive design gives users control over the data, allowing them to manipulate, explore and focus on the aspects they find most interesting. Interaction provides immediate feedback, helping users understand the effects of their actions on the data. Moreover, interactivity can accommodate different learning styles and preferences, allowing users to explore data in a way that makes the most sense to them.

Critiques:
The bar chart only shows the total count of each genre, which may not provide a comprehensive understanding of the relationship between danceability and loudness/energy within each genre, users must click to observe the relationships. Additionally, two scatter plots and multiple genres can induce a risk of overplotting.
The colors used to represent genres exceed the recommended usage of five, and this can cause distortion resulting in ineffective communication of the data. Moreover, the color hues are redundant on the bar chart. Having multiple representations, in this case, two scatter plots and a bar chart on the same visualization may overwhelm users, making it challenging to interpret the information effectively.
Proximity: Visualization must ensure that related elements are close enough to each other in the visualization to reinforce their relationship
Similarity: Use consistent colors to indicate that certain data points belong to the same genre, this helps users quickly identify patterns and relationships.
Continuity: Ensuring that the connections between the bars and the scatter plot points are smooth and continuous making it easier for users to follow the relationships between the data.

While bi-directional and mouse-based interactions can be useful, it is necessary to provide clear instructions and visual cues for users to know they can click on the bar to reveal more information. Additionally, users with visual impairments or color blindness, choose high-contrast colors and color-blind-friendly colors are necessary.

Task 5: Does the key of the track influence the relationship between danceability and popularity?

Final View Screenshot:



Previous Iterations: As the task was newly created after milestone I, there are no previous sketches available. However, the final screenshot has been provided above.

- Scatter plot showing the relationship between danceability and popularity.

We have decided to use a scatter plot, points are used as marks to represent individual tracks. Each point corresponds to a track and its position on the plot represents its danceability and popularity values. The x-axis represents danceability and the y-axis represents popularity. Positional channels are known for their high accuracy in depicting quantitative information. By using the x and y axes to represent danceability and popularity, it allows users to quickly and easily identify the relationship between these variables. Proximity, and similarity can be seen through this visualization as the same colored points that are closer together on the scatter plot imply a stronger relationship between danceability and popularity.

The drop-down interaction allows users to select a specific key from a list of available options. When a key is selected, the scatter plot updates to display only the tracks in that key. This interactive feature enables users to explore the data more thoroughly by examining the relationship between danceability and popularity for specific keys. The drop-down interaction is a dynamic way to allow users to explore the data more effectively. As users change their selection in the dropdown menu, the visualization updates in real time to reflect the new set of tracks in the chosen key. It promotes user engagement and facilitates the discovery of trends or patterns that might not be visible when examining all the data at once. By providing control over the displayed data, users can form a deeper understanding of the relationship between danceability, popularity, and key. Additionally, tooltips are being introduced, allowing further interaction.

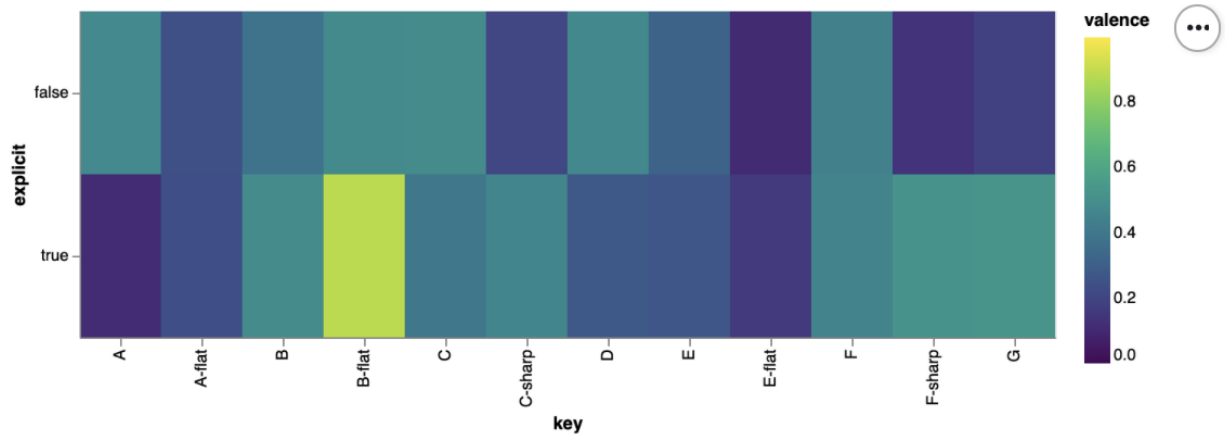
As mentioned above, the scatter plot with drop-down interaction is an effective visualization technique for exploring the relationship between danceability and popularity across different keys. It allows users to quickly identify patterns and trends in the data. By providing an option to focus on a specific key, the visualization helps users develop a deeper understanding of how the key might influence danceability and popularity. The scatter plot is an efficient way to represent large amounts of data in a compact and easily understandable format. Positional channels are one of the most efficient ways to convey quantitative information, as they allow for easy comparison and accurate perception of values.

Critiques:
The scatter plot is an effective way to visualize the relationship between danceability and popularity, as it enables users to quickly identify trends and correlations. The dropdown interaction empowers users to focus on specific keys, allowing for a more granular exploration of the data.

Overlapping points can make it difficult to discern the density of tracks in particular areas of the plots, which may be alleviated by incorporating transparency, and size. The visualization does not provide information about additional variables that could potentially influence the relationship between danceability and popularity. Since the interaction requires users to click and select a specific key, this may not be accessible for users with certain disabilities.
Proximity: Points close together in the scatter plot are perceived as a group, making it easier to see clusters or trends in the data.
Similarity: by using the same color for data points within a key, the plot reinforces the perception that the points belong to the same group.

Task 6: Which combination of the key and explicit types of the track sounds most positive?

Final View Screenshot:



- Heat map showing the relationship between keys, explicit types, and valence.

Previous Iterations: As the task was newly created after milestone I, there are no previous sketches available. However, the final screenshot has been provided above.

A heatmap is an effective way to visualize datasets with a color scale. In this case, it helps represent the relationship between keys, explicit types, and valence. Each cell in the heatmap represents a combination of key and explicit types.

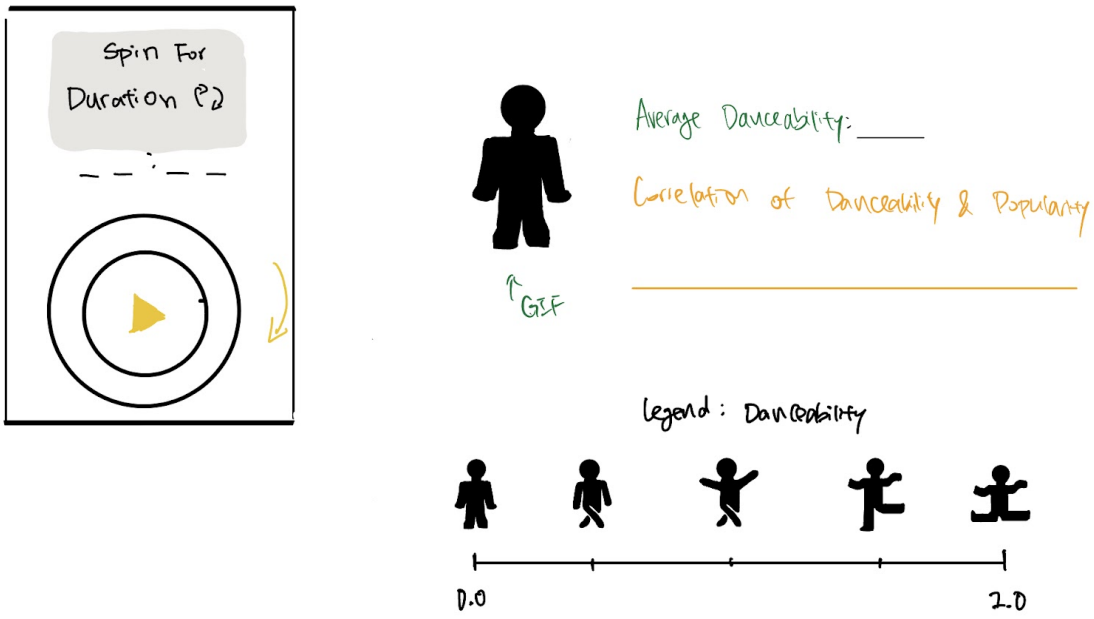
Spatial positioning allows for easy comparison between different combinations of key and explicit types. The color scale makes it easy to identify the most positive-sounding combinations based on valence.

When a user hovers over a cell, a tooltip provides detailed information about the combination of the key, explicit type, and the corresponding valence value. This allows users to interact with the visualization and get more information about the data trying to communicate.

Critiques:
The color hues selected are suitable for individuals with color blindness, making this visualization colorblind-friendly and separable.
The heatmap uses proximity and similarity(in color) to group cells with similar valence values, making it easy to identify patterns. It effectively communicates the relationship between key, explicit type, and valence, allowing users to identify the most positive-sounding combinations. Moreover, heatmap efficiently conveys a large amount of information in a compact and visually appealing manner.

Novelty Sketch:

Task 2: For different intervals of durations, does the relationship between danceability and popularity change?



Task 3: What is the difference in the distribution of popularity and the relationship with danceability between major and minor modes?(credit: Santa Ono)