Jason Zheng (jz449)
Tiffany Song (ts547)

Project 2 Writeup

Description of the Data

Our data consisted of two datasets, both of which are from Kaggle. The first is a dataset containing the top Billboard songs on Spotify from the years 2010-2019. It contains the title of the song, artist name, genre, year, bpm, decibel rating, energy, danceability, liveliness, valence, duration, acousticness, speechiness, and popularity of the song. The last eight variables are taken from Spotify, which has its own way of calculating those aspects of a song we used the energy, danceability, decibel, liveliness, and bpm features of the songs because they seemed to be the most interesting variables provided and we believed that they would provide more insights into the overall features of a certain artists' work compare to the other features.

We used the artist name as the main plot point in our bar chart and used the song name and their popularity in the popout chart so that the songs could be ranked in popularity.

We did not use the genre of the songs because Spotify characterizes their genres much more in-depth than what we are normally used to. For example, rather than just 'pop' as a genre, pop is divided into many subcategories, so some songs are pure 'pop', while others are 'dance pop' and 'hip pop'. Additionally, there were many genres we had not heard of before, such as 'baroque pop', 'metropopolis', and 'escape room'. We had difficulty generalizing these into the commonly known genres such as 'pop', 'rap', 'edm' and realized that displaying the genres that Spotify provided would not be as effective in our visualization. Since we wanted filters on our chart, it would be confusing for the user to be presented with genres they may not be familiar with, so filtering by such genres would not be the most helpful.

The second dataset we used contains information about artists on Spotify. The variables include artist, gender, age, type, country, city1, district1, city2, district2, city3, and district3. The type variable corresponds to whether or not an artist is a solo person or a group. We chose only to use the artist's name, gender, and country. We did not include age because it was unclear when these ages were calculated. Since we have a range of songs from a range of years, it would be unclear what the age of the artist corresponded to. It could be deceiving that an artist was a certain age from this dataset, but they had one hit song from 10 years ago, so their age did not align. Additionally, the dataset was slightly incomplete so all the city and district variables were inconsistently filled out. Finally, we had to do some cleaning on the dataset because some popular artists had city variables but no countries. We manually edited the CSV so that popular artists had the proper countries associated with them. Although there were many countries represented in the dataset, we only took artists from countries that had more than one other artist from that country represented in the dataset. We felt it would be more informative to have countries with multiple artists so that the user could be able to make comparisons with other artists/genders.

Overview of the Design Rationale

Our main visualization is a bar chart. We chose this visualization over others because we wanted the viewer to be able to directly compare the number of popular songs to those of other artists. We considered using circles whose radius changed in proportion to the number of popular songs they had, but we decided that a bar chart would display the comparison much more accurately compared to other types of graphs. Additionally, our y-axis does not change to scale if there are bars on the graph that don't reach a large percentage of the graph. For example, the maximum number of songs artists from Australia have is six, by the artist Sia. The bars therefore only take up ~⅓ of the vertical space. We actively chose to make our y-axis this way because it conveys how popular artists of these filters are compared to those overall. Thus, we can see that although Sia is the most popular Australian artist, she is not that popular overall, since she only reaches one-third of the y axis. Had we scaled the y-axes proportionally, it would have seemed like Sia was widely popular. This way, we can both convey the popularity comparison amongst the artists that fit the filtered parameters, and also gauge their popularity amongst the top ten artists overall.
The marks for the bar chart are rectangles. The channels used are vertical alignment, horizontal spacing, and color.

We chose to make the checkboxes look like buttons because it encouraged the user to click on them more. The plus buttons on the checkboxes tell the user they can add it as a filter. Although it is not the traditional checkbox format that we are used to, it adds more polish to the overall visualization and is still intuitive to use.

We chose to put the extra information about the artists underneath the bar chart as opposed to right next to it. One rationale we had for putting the pop out to the right of the bar chart was so that all the information could be viewed without extra scrolling by the user. However, given the width of our bar chart and taking into consideration that different users have different width screens, we thought it would be better to put the popout below the chart, to eliminate the risk of having to scroll left, something less intuitive than scrolling down. Additionally, we implemented our filters so that the popout would close on a new filter. The popout does not need to be within the eyesight of the filters because the filters don't directly change the output of the popout. The popout only changes on click of the bar charts, so it made more sense for us to ensure that the popout and the bar chart both could comfortably fit on the screen together.

We chose to use a spider/radar chart to plot the aspects of the song data. We averaged the values for all of the songs for each artist to make the chart more clear so there would not be excessive plots on the chart. We chose this visualization over a bar chart because it makes it easier to make an analysis of the features of the song. By having all the features on a circular chart, it is easier to

make comparisons of the features across the board. The marks used in this chart are lines while the channels used were connection

Overview of Interactive Elements & Design Rationale

Our first interactive element is our filter buttons. Since we wanted to make comparisons between different groups of artists, we wanted a way for the user to filter between different categories. We chose select buttons so that the user to select multiple filters, to allow them more freedom in what they wanted to display and the amount of interaction they could have. Additionally, our bar chart has a hover interaction to it. When hovering over, the opacity and outline of the bar changes and a label appears to make the information more clear to the user. We chose this because we realized on a bar chart that is 10 elements long, it could be difficult to read the y-axis and we wanted the user to be able to access the number of songs readily. Additionally, we have a click interaction with the bars so that there can be more information about the artist. In order to make this interaction discoverable, the cursor becomes a selector shape as opposed to the default pointed cursor. The selector cursor is a universal indication that something can be clicked, so we ensure that the user would know that clicking could lead to something/more information. Although the click may not be so easily discoverable, we chose to hide the additional information as to not overwhelm the user with things they may not necessarily be interested in viewing. Additionally, having the click for more information makes the interaction more interesting.

Story

Our visualization tells us that the liveliness of songs does not matter much in making an artist popular. Most artists had relatively large amounts of danceability and energy, which indicates that artists with energetic and more upbeat songs tended to be more popular than those who did not. However, one thing that was surprising was that Adele had relatively low values for energy and danceability, yet she is one of the most popular female artists in this dataset. This shows that one does not necessarily need loud, energetic songs to make them popular, but energetic songs have a higher chance of doing well. The top ten artists overall have very similar spider charts, with nearly identical average BPMs. This further reinforces the idea that the songs with those levels of danceability, energy, and BPM are those that are popular. Something else interesting that the visualization revealed is that popular Colombian artists make songs with higher energy and danceability levels compared to those of other artists. While there is some variation in the values of the aspects of the songs, we can see that most popular songs are very similar and artists whose average song values follow the pattern of those in the top ten tend to have many more popular songs than those who don't.