u01
David Cherne, Olivia Papotto
CPCS 4030, Fall 2020

Visualizing Spotify Data in North and South America

Project Report

Introduction

Our project consisted of gathering and visualizing data from spotifycharts.com and using Spotify's API to extract additional attributes for the songs. The website spotifycharts.com shows the Top 200 songs filtered by location (globally or within a specific country), time frame (daily or weekly), and the date (ranging from 01/01/2017 to the current date). After first discovering the data on the Spotify Charts website, we became curious as to how different genres trend across the Americas and overall wondered how different countries listen to music. In summary, our project focused on visualizing the different musical trends as captured by Spotify through North and South America.

Dataset

The dataset we used was originally obtained from the site www.spotifycharts.com. For our charts, we specifically focused on the countries that had data provided in North and South America on a random date 10/11/2020. When downloading the Top 200 songs track for the specific country on a specific date, it gives the user the following information/attributes:

- Song position in the Spotify Chart
- Track Name
- Artist
- Number of Streams
- URL

Since this only gives the basic information for each song, we then built a python script to pull additional attributes (genre in this specific scenario) from Spotify's API and update the CSV file with this additional attribute. To access our dataset and to see a more detailed description of each file, please visit our github page as follows,

https://github.com/OliviaPapotto/DataVisF20

Data Exploration

With Milestone 3 and the use of Tableau, we chose to initially explore our dataset by visualizing how genres were streamed over the entire year in the United States, what the most popular genre on specific United States holidays was, and the most streams in North and South America on our selected date 10/11/20. When using Tableau, we were able to explore the different types of visualizations that it offered to see how our dataset could be manipulated. However, our focus with this dataset was wanting to see how the genres were streamed around the world. This led to us focusing on the third graph we created in Tableau with the map of North and South America since we enjoyed the visual aspect of seeing what genres or streams come from where. Overall, this allowed for us to have a better idea for which visualization we wanted to take to Observable when we began coding with d3.

Design Solution

Graph 1 - Number of Streams on 10/11/20 in North and South America

For our first graph, we chose to show our data by constructing a map of North and South America to encapsulate our information, instead of an x-and-y graph. We visualize the attribute of the number of total streams that are emanating from a country, and we also display all of the countries in the Americas that Spotify Charts provides their data for. In order to complete the look of the map, we decided to include the countries that Spotify Charts does *not* include data for; for these specific countries, we simply display "no data provided" with their tooltip.

To display the data, each country is colored a shade of green, based on how many streams that country had contributed to the total stream count of North and South America. The shade of green is darker if more streams come from that country, and is lighter if otherwise. An even spread of streams would render the entire graph a lighter shade of green,

while an uneven spread would mean some countries would look green while the rest appear white or light-green. We thought that this was effective because you can see the entire scope of the data, while also easily discerning which countries provided more total streams over others.

Graph 2 - Most Popular Genre on 10/11/20 in North and South America

For our second graph, we again chose to show our data by constructing a map of North and South America to encapsulate our information, instead of an x-and-y graph. We visualize the attribute of the most streamed genre for the countries throughout the Americas. We decided to include the countries without any data provided by Spotify Charts to complete the look of the map again; we had decided to do this with all of our graphs.

To display the data, each genre is assigned a specific color that can be seen in the legend to show the prominent genre of our selected day. The genre/color was chosen by going through every individual (Top 200) song that was streamed in that country, tallying up the streams from each song, and grouping them with that songs' genre. At the end of the tally, the genre with the most streams was declared "most popular", and that country was colored based on that genre. We thought this was an effective visualization because each country displays their prominent genre in relation or contrast to their neighbors, making it quick to see which genres are popular geographically, or to see which specific country prefers what genre.

Graph 3 - Popularity of the Top 5 Genres on 10/11/20 in North and South America

For our third and final graph, we once more chose to show our data by constructing a map of North and South America. We visualize the attribute of the number of total streams that are emanating from a country for a specific genre. Instead of supplying the visualization for all genres, we chose the top 5 genres based on the total streams that genre had; in this case, it would be Latin, Pop, Rap, Hip-Hop, and Carioca, in that order. Each genre is given their own respective color -- if the color was used in the second graph, we had chosen the same color. To shift through the graphs, a radio button is provided above the current genre being visualized, to where the viewer can choose which genre to view.

To display the data, each country is colored a shade of the color of the genre chosen, based on how many streams that country had contributed to the total stream count of that genre. The shade of the color is darker if more streams come from that country, and is lighter if otherwise. An even spread of streams would render the entire graph a lighter shade of their color. We thought that this visualization was effective, because you can view the specific data coming from the genre you choose, rather than being overwhelmed by all of the data of every genre on one graph. The radio button provides a quick way to switch over to whichever genre the viewer wants to see visualized.