P2 - The Decline of the One Hit Wonder

Sidd Srinivasan ss2969, Brendan Coyle brc77, Kaitlyn Son ks772

A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection. (10 pts)

The data we use is an integration of two different data sets. The first dataset https://www.kaggle.com/rakannimer/billboard-lyrics, contains the lyrics of every billboard 100 song from 1965 to 2015. While we were not interested in the lyrics themselves, this data set was extremely valuable as each data point consisted of a song title, the artist, and the year the song made it onto the billboard 100.

As we were looking to see the effect of the one-hit-wonder, we wanted to examine the collection of billboard 100 songs that were related to each artist. In order to filter results, we wrote a python script to create elements for each artist. We then iterated across all the songs, and then added them to the artist elements.

As we wanted to examine the frequency of one-hit-wonders over different music listening mediums, we were concerned with a large time scale. This meant that we wanted to group songs by decade, rather than analyzing each artist's performance on a year to year basis. This when creating the artists elements, we kept counters for the number of billboard 100 hits they had in each decade.

For songs with featured artists, we attributed the song to the main artist, removing the features from the artist attribute of that song before processing the dataset. This ensured we didn't end up with artists such as "Anderson Paak feat. Ab Soul and James Blake" as we processed the entry to instead say "Anderson Paak."

The second variable we wanted to analyze was gender. We were curious as to whether a certain gender has a greater chance of being a one hit wonder, and whether the distribution of male and female one-hit-wonders in the billboard 100 varied over the decades as music shifted from records, to cds, and from radio to Spotify. To do this, we utilized a singers gender dataset: https://www.kaggle.com/rkibria/singersgender, which scrapped artists genders from wikipedia. After we processed the first dataset, we integrated the gender set by appending a column to our artist dataset with the gender if it existed in the singer's gender set.

For bands we decided that if the band was single-gendered, we would color it according to their gender, blue for males, pink for female. For mix gender bands, we assigned their color to purple.

We decided not to utilize the genre classification of the gender dataset, even tho we could have mapped artists to particular genres. We selectively decided that this would be overwhelming to users, as it would detract from the core component of the story. Part of this reasoning stemmed from the fact that the classifier created more genres than a normal user was familiar with, such as "Male American Singer Songwriter." We found the genres to be too specific, which would potentially detract from or not reflect the trend of the one hit wonder.

<u>A description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales. (10 pts)</u>

We decided the best way to visualize these elements was to populate the graph with bubbles and animate them using d3.force. This gives the graph a feeling of being alive and makes the interactivity fun to play around with. Each bubble represents a musical artist. Their radius shows how many hit songs that artist had. The larger the bubble is, the more popular the band was. We also decided it would be interesting to see the genders of these artists, so we colored the bubbles blue for a male artist or all-male group, light pink for a female artist or all-female group, and purple for a mixed-gender group or an artist whose gender was not specified.

The areas of the circles are scaled linearly based on their number of hit songs. This makes it easier for the user to directly and accurately interpret the true size of every artist in relation to another.

Through this scaling, we also had to modify the size of our svg windows. We wanted to let users compare different time periods against each other, so we had to play around quite a bit to determine which dimensions would create the least amount of clutter on screen but also be large enough to show the huge amount of data we had. We decided to make the windows 700 by 600 pixels to properly show our data.

Additionally, we added an animation reveal to our graph to put forward the idea of visual motion that we wanted for the project. It's something to make it feel like it's really responding to your input.

We decided to initialize the two comparisons on the 1960s and the 2010s. This fulfills the goal of walking the users "down the trail", as it is in line with our concept that one-hit-wonders were more prominent in the past than the current day. Then we give users the ability to “wander on the grass” and explore the different relationships between the other decades.

<u>The story. What does your visualization tell us? What was surprising about it? (5 pts)</u>

Do you remember the songs “Dynamite”, “Bad Day”, “You’re Beautiful”? “Somebody that I used to Know”, “Teach Me How to Dougie”. These songs were all songs in the Billboard top 40 in the last 10 years, yet the artists for these songs are virtually unknown, hence the term one-hit-wonder.

Billboard chart rankings are based off sales (physical and digital), radio play, and online streaming. Through the decades the number of one hit wonders decreases, as mainstream artists begin to monopolize the billboard charts. This may be attributed to said artists’ heightened brand and identity through social media and streaming platforms such as Spotify and Apple Music.

Though a portion of the data was ungendered it can be inferred from the data and genders shown that as time progresses, more female artists populate the billboards. As users can see, female artists in the later years have a better chance of gaining more traction and becoming bigger stars.