INFO/CS 3300 Project 1

Michael Gingras, Charles Yu, Niranjan Ravi

March 7, 2017

The Data

The main inspiration for this project was Kanye West's infamously inflated ego. We thought it would be an interesting idea to see how many times he references himself in his songs, and compare that to the number of times other rappers reference themselves in their own songs to truly see who has the biggest ego. In addition to this "ego size" variable, we wanted to see if other factors could have influenced a rapper's ego, particularly the number of singles and albums they have created and the millions of albums they have sold. Ideally, the more successful the artist, the bigger their ego would be. The final variable is the rappers themselves, which we selected based on the datasets that we found to make things more consistent.

To find the number of references for each artist, we used two large datasets from Kaggle that contained lyrics for songs by a number of artists, and wrote a Python script to filter the data to only look at the lyrics for the artists that we were interested in. Some artists were missing lyrics for some of their more recent songs, so we then manually added those lyrics in order to appropriately represent each of the artists.

In order to actually count the number of references, we wrote another Python script to parse through all of the lyrics and place the counts in a JSON file, also ignoring lyric annotations such as "[Kanye:]" and "(Chorus)". Many of these rappers are known by a multitude of nicknames, such as "Yeezy" for Kanye West, and "Six God" for Drake. This made it more complicated to parse through the lyrics, as we had to account for all of these nicknames. Fortunately, we were able to find large collections of rappers' nicknames on Google or through fan forums, so we were able to turn them into regular expressions and count the number of times they appeared in each rapper's songs. However, due to the inconsistency of documentation on rapper nicknames, we couldn't find the same number of nicknames for each rapper. For some, we could barely find any at all. To account for this, we took the top 10 referenced nicknames for each rapper after counting all their nicknames. To account for differences in the number of songs, we normalized the values by dividing the number of references by the number of total words sung by the rapper.

For the rest of the data, we looked through each of the rappers' Wikipedia pages for their discography, and manually added the total number of songs, as well as the total number of albums sold. For the rappers' "Lifetime Discography", we added the total number of studio albums, EPs, singles, and so on, which can conveniently be found in the top right corner of their discography Wikipedia pages.

Visualization

For plotting the data, we looked at how the data was spread out before determining the axes scales. The discography numbers were spread out quite evenly, so we left the x-axis as a linear scale. The album sales, however, had a number of data points clumped towards the bottom, with a few outliers at the top. For that reason, we thought a square root scale would be the best scale option, since it gives preference to lower values.

One issue we had with plotting the data points was that 2 Chainz and Future both had very similar numbers for both discography as well as album sales, so their data points would overlap when we added pictures of their heads. To solve this, we had to move 2 Chainz' head up by a few pixels, and included an asterisk in the graph to represent where he would actually be.

Originally, we had planned to have the size of a rapper's head correspond with the size of their ego, but we found that to be too messy and many heads would overlap with one another. Instead, we decided to include a simple dashed circle centered at the rapper's head, with the radius of the circle corresponding to the rapper's ego. We used the median ego, Desiigner, to determine the size of the rest of the radii, scaling them linearly. If an ego was larger than Desiigner's, the circle would be larger by a certain amount, and so on.

In terms of our color scheme, we wanted it to be simple. Many rap album covers have a grayscale design, so we decided to keep the black-and-white design when creating images of the rappers' heads. We also decided that pink was a nice accent color for our other visual elements, so everything else is some shade of pink.

For our second visualization, we thought it would be interesting to see how many words a rapper says between their self-references. This time, instead of "references per word", we would look at "words per reference". For example, Kanye has a value of 98, so one possible example of his songs could look like "Yeezy...[98 words]...The Louis Vuitton Don", which are both nicknames of his.

We were originally going to do a simple horizontal bar chart, but soon noticed that it was beginning to resemble a pyramid. We expanded upon this, and got rid of the bar chart concept, instead opting for a pyramid with each level representing a rapper. The sizes of the levels also correspond linearly with the "words per reference" variable.

After creating the basic pyramid, we realized that the design was a bit unintuitive, since it seemed to imply that rappers with larger numbers had larger egos, despite the opposite being true. To counteract this, we added an

image of a crown at the top of the pyramid to represent the "King of Egos", and further playing off of the pyramid theme. Furthermore, we increased the opacity of the bar as we go up the pyramid to emphasize the bigger and more important egos.

Interpreting Results

As we can see by the sizes of the circles, it is quite apparent that Kanye West still has the biggest ego of these 10 rappers. This is very similar to our initial thoughts before starting this project, and is enforced by the second visualization, which shows that on average, he makes a reference to himself once every 98 words. Clearly, Kanye's love for himself is backed up by the data. One surprising thing, however, was that the number of songs and number of albums sold do not necessarily create a bigger ego as we initially speculated. For example, Gucci Mane's ego is much larger than Future's, despite releasing a similar number of songs and selling a similar number of albums.

Sources for Data Set

https://www.kaggle.com/viccalexander/kanyewestverses
https://www.kaggle.com/gyani95/380000-lyrics-from-metrolyrics
https://www.kaggle.com/mousehead/songlyrics

Sources for finding rapper nicknames

Nicknames found through Google Search's relevant info or through one of the following sites:
http://www.kanyetothe.com/forum/index.php?topic=245210.0
https://www.reddit.com/r/Kanye/comments/4e16nv/complete_list_of_kanye_nicknamesmonikers_what_am/
http://www.fuse.tv/videos/2012/01/what-2-chainz-nicknames-mean
https://www.reddit.com/r/ChiefKeef/comments/565uz0/all_chief_keef_nicknames/
https://genius.com/Rap-genius-the-many-nicknames-of-lil-wayne-lyrics
http://www.futurefanz.com/biography/

Image Sources

https://s-media-cache-ak0.pinimg.com/originals/cf/08/d0/cf08d07c40abe81de0464ec465882599.png
http://www.howwe.biz/uploads/articles/howwe_00b266f94556d12b65299f287a160d7f_1444032696.jpg

https://images-na.ssl-images-amazon.com/images/I/41s5fUzMbvL._SY355_
.jpg
http://cimg.tvgcdn.net/i/r/2013/09/14/ab2c0bc4-e9f7-4229-87f0-1ddf84c1634d/
thumbnail/660x960/abf9ecfd4bf4d173418cb3c70735182d/130914gucci-mane1.
jpg
https://pmchollywoodlife.files.wordpress.com/2013/06/2-chainz-bio-photo.
jpg?w=330
http://news.independence-card.com/wp-content/plugins/RSSPoster_PRO/
cache/7ceb9_2014Future_Getty172029035070414.jpg
http://images.musictimes.com/data/images/full/39741/fetty-wap.jpg
http://cdn01.cdn.justjared.com/wp-content/uploads/headlines/2016/04/
desiigner-panda-dethrones-rihanna-work-on-hot-100.jpg
http://assets.nydailynews.com/polopoly_fs/1.2239654.1432863272!/img/
httpImage/image.jpg_gen/derivatives/article_750/cent29n-1-web.jpg
http://www.freeiconspng.com/free-images/crown-icon-23690