Names/netIDs: Juliana Hong (jhh274), Alexa Bren (ab998), Kevin Tian (kt474)

A. 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)

Data Description

We pulled our data from an open-source dataset made available by FiveThirtyEight, available here: https://github.com/fivethirtyeight/data/tree/master/hip-hop-candidate-lyrics. This dataset was created using Genius as a data source. The original dataset includes information on the mentions in hip-hop lyrics for eight different political candidates (Trump, Clinton, Bush, Christie, Huckabee, Sanders, Carson, and Cruz) and eight different dimensions (candidate, song, artist, sentiment, theme, album_release_date, line_url). The theme and sentiment of each of the lyrics were determined by the informed opinions of FiveThirtyEight staff workers.

Variables

The variables that we visualize from our dataset include year (2008-2016), candidate mentioned (Donald Trump, Hillary Clinton), sentiment (positive, negative, neutral), and theme (Political, Sexual, Money, Hotel, Other) of hip-hop lyrics.

Data Filtering

We filtered this dataset down to a subset of the data that only included Hillary Clinton and Donald Trump, the only two candidates we were interested in comparing. We chose to focus on data from year 2008 and onward, as both candidates were consistently mentioned multiple times from that date onward.

Data Selection

Once we filtered the dataset, we grouped elements together to make the visual representation more clear. There were eight categories in total for the themes of the hip-hop lyric (Power, Personal, The Apprentice, Hotel, Sexual, Political, Money, Power, and N/A). We decided to group the categories of N/A, Power, Personal, and The Apprentice together into a group called "Other," while keeping the other categories the same. We made this decision because the categories of Power, Personal, and The Apprentice had the fewest number of mentions and were most ambiguous, so we thought that it would make the data clearer and more interesting to focus on these five categories, as opposed to the original nine. Although there were fewer total sexualy-related mentions in the whole dataset, we felt it was important to highlight that Clinton received disproportionately more mentions of the sort (and with negative sentiment) than Trump did. Although the data initially appeared in a random order, we sorted it by theme so that we could visualize the colors and themes in the same order for each year.

B. 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)

Scatter Plot

For the scatter plot, we mapped the data by treating each lyric data point as one circle shape on our plot. Each instance of a hip-hop lyric mention is visualized as one circle, and the color and icon of that circle represent data about the sentiment and theme (respectively) of that mention. The y-position of the icons correspond to the year that the mention was made (from 2008-2016). Because we are using discrete data, we did not use any scales. We decided to plot each year as a horizontal side-by-side comparison so that we could focus the viewer on comparing the candidates' mentions in each year. The order that the colors and themes appear is held constant across each year so that there can be a comparison made between common categories, and if any given category does not appear in that year, it will not be present.

Pie charts

For the pie charts, we mapped the data by creating four visualizations to represent pre- and post-2015 for the two candidates of Donald Trump and Hillary Clinton. These charts map the percentage of the sentiments of the mentions in each of those four categories. The color of the pie chart section corresponds to the sentiment, and the relative area of the section corresponds to the percentage of that sentiment as a part of the total of that category.

C. The story. What does your visualization tell us? What was surprising about it? (5 pts)

Cultural artifacts, such as music, often reflect key opinions and perceptions of political leaders. Our visualization tells the story of how the cultural perception of key political leaders has evolved over time, by displaying the types of mentions in hip-hop lyrics for Donald Trump and Hillary Clinton from 2008-2016. We decided to display the data by coupling two visualization types: a scatter plot and a set of pie charts. The pie charts are intended to show an overview of the sentiment of the hip-hop mentions pre- and post- 2015, while the scatter plot is intended to show a more granular look at the lyric data points.

Our scatter plot uses a side-by-side comparison to show how Trump and Clinton compare each year, in terms of the sentiment (positive, negative, neutral) and theme (Political, Money, Sexual, Hotel, Other) of the hip-hop lyric mentions. It is also interesting to notice the frequency of mentions across years for the two candidates. Trump has a lot more mentions in recent years, with especially large amounts post-2013, whereas Clinton has the most mentions in 2008, but she does not have as many in subsequent years. Thus, Trump has a much larger presence in the hip-hop world than Clinton.

As seen in the pie chart visualizations, overall, sentiment toward Trump in hip-hop was originally mostly positive, pivoting to mostly negative at 2015, the year he announced his run for presidency. Because we show the sentiment and theme of the mention at the same time, our graph visualization allows us to look at trends regarding the associations between the different variables. We found that Trump’s positive mentions were overwhelmingly related to money or the Trump Tower hotel, showing that the original positive sentiment toward Trump stems from using him has a symbol of wealth and success. Once the data shifts in 2015, most of his

negative mentions are politically related, showing the hip-hop community's critical opinions of Trump as a political figure.

Clinton had quite a variety of sentiments and topics in her mentions, but overall, moved from a negative to more positive direction over time. That said, something that did stand out to us about Clinton's mentions is that she received multiple sexually-related mentions, that were mostly with negative sentiment (Trump only received one, and it was of neutral sentiment). This shows that Clinton is more sexualized figure in hip-hop than Trump, perhaps because she is a woman.