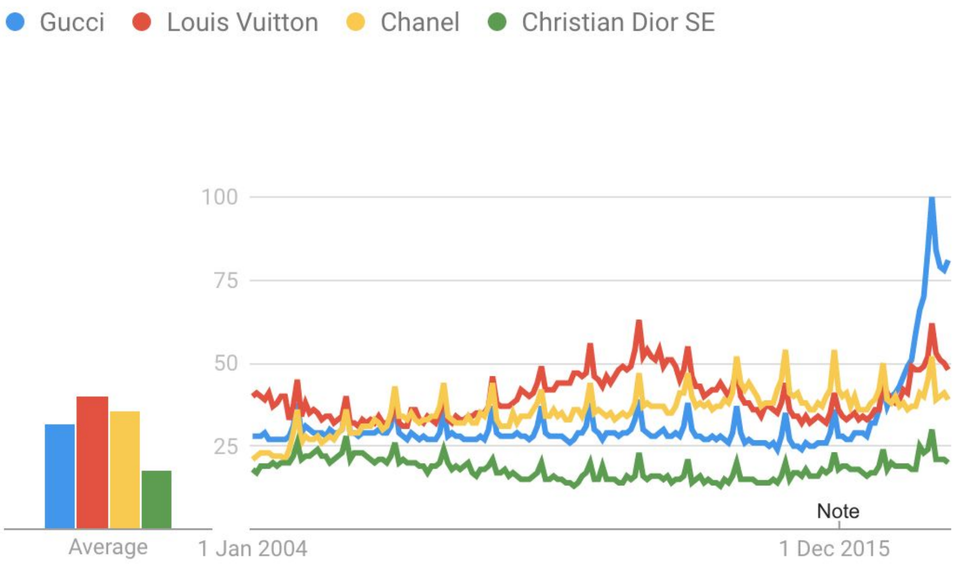
Basic Info.

Project title: “The Influence of Rap Lyrics on Pop Culture.”
Name: Steven Worrall, Daiya Masuda
Email: sjworrall@dons.usfca.edu, daiya.masuda@gmail.com
Github UID: Stevenworrall, dymzd
Github Link: https://github.com/StevenWorrall/DataVisFinalProject
Website Link: https://stevenworrall.github.io/DataVisFinalProject/

Background and Motivation:

Music industries have been growing exponentially; every year there’s more and more popular musicians/bands collaborating with major companies to advertise new products in songs. In particular, Rap music has become very popular over time, especially since the appearance of Drake and Kanye West. Both of them have collaborated with major fashion brands like Nike and Adidas along with many other big brands. After analyzing the lyrics in their songs, this becomes quite apparent. Given this, we wanted to know how true this actually was and how deep it actually went. The question we want to answer is: what effect have rappers had on search trends? More to the point, can we better describe a possible niche to market products in? After reviewing data, we think it might be possible to correlate the popularity of search trends with songs used in popular rap music throughout the last 10-15 years.

One example of a possible correlation is with the brand Gucci. Google Trend shows that it had a major outbreak in interest over the past year. This correlated with other similar search words such as: “Gucci gang lil pump” and “Gucci mane.” This possible correlation drove us to believe that it is possible there is a correlation between the behaviours of the consumers and trends in the marketplace.



Related Works:

A story board that we may slightly base ours off:
https://www.trulia.com/vis/tru247/
Stream graph example
https://bl.ocks.org/mbostock/4060954

Questions:

- What effect does pop culture have on search trends?
- Can we better understand and quantify the effects of marketing via songs and lyrics?
- Are we better able to understand why things become popular and can we predict popularity by cross referencing how songs become popular?

Benefits:

- Ability to better understand the tactics of marketings.
- Ability to better quantify the effect of marketing via rap and music in general.
- Ability to better understand the effects music has on trends, specifically search trends.
- Ability to possibly predict if new songs are going to gain popularity by analyzing the lyrics.
- Ability to possibly create a more profound space in which to market via music.

Over the course of the project these questions pretty much stayed try to what we have down. We really doubled down on the idea of being able to quantify whether or not it's a viable decision to market products or companies via song lyrics and if that led to a correlation with search trends and increased brand awareness.

We believe that by showing the history of products and ideas in songs and correlating that with increased trends, we can begin to show the marketability and reach of rap songs in modern day. One new question that arose via the project was how we could further quantify an increase in awareness and an uptick in trend for a specific product or idea.

We thought that a good way to do this would be to mix in actual data from increases in sales from publicly traded companies (something I'll touch on later in the Evaluation portion of the process book).

Data:

In order to get the data, we had to implement web scraping in Python. Our plan was to get data from the top 50-100 songs per year for the last 10-15 years and parse all the words in these songs. From there, we could put the data into a dictionary in python and write it out as a JSON file. The original plan was to read JSON directly into D3 but we ended up converting these JSON files into CSVs so they would work with our existing code base better. This data will be matched with google trend data to test correlation with popular words searched vs popular words in music at the time. We tried to eliminate as many standard words as possible and focus more on material objects like alcohol or high end luxury goods.

The trend data was gathered by hand by downloading the CSV from Google's website. We did this by first finding the word we wanted by hand from our data gathered by the Python Script and the manually searching the word through their database. Another thing I'll discuss later in the the evaluation section is the adding of a live API to pull trend data so that the graphs will be more up to date and more data (words) would be serviceable to the user. All data collection and cleaning was done in Python by scripts we made.

Data collection, by far, was the most tedious and time consuming part of the whole project. We kept getting IP banned from the site we were scraping from, even after putting in long breaks and using different IPs. We even tried rotating through proxies for our requests. Eventually though, we just put a long delay on the script and did it over a few different IPs. (coffee shops, school, home, etc).

Below is a list of links and references we used for data:

Google trends data for correlation to increased searches by popular words
https://trends.google.com/trends/explore

Google trends APIs:
Javascript:
https://www.npmjs.com/package/google-trends-api
Python:
https://github.com/GeneralMills/pytrends
https://pypi.python.org/pypi/pytrends

Song data:
To get top 100 songs per year
https://www.billboard.com/charts/year-end

Some starting python code to get lyrics from azlyrics:
https://www.quora.com/Whats-a-good-api-to-use-to-get-song-lyrics

Exploratory Data Analysis:

Initially, we saw the data while looking at trend graphs on the google trend data website. You can see an image of this in the introduction section of the process book. The way of looking at the trend data is a simple line graph but we were able to see huge peaks in interest. This was especially apparent in the trend graph for Gucci. We then started to slowly connect the dots and looked up the release date of the song “Gucci Gang”, which correlated with the trend graph surprisingly well.

We then started to get the idea that some trends may have been influenced by songs, so we started to manually line up release dates with big spikes in interest. We saw that multiple songs made peaks in interests in search trends so we brought the idea forward to our professor.

This is where we started with the idea of what we have in terms of the album selector and the stream graphs. The stacked bars were a perfect representation and way to show all of the different bars at once while still being able to distinguish when a bar had a big spike. The album selector came from the idea of wanting to be able to have a visual representation of the songs on the website and it was a good way to also make the user interact with the visualization.

The insights from the data showed us that we needed to make sure users could easily see when data should correlate with each other. This is why we ended up putting two stream graphs side by side. That way, the user can easily see across the two graphs and see where the spikes in both graphs are and see correlation easily.

Design Evolution:

Our idea was to display our data in a way that would mix in features from both a scrollytelling visualization and dashboard. We wanted to show the reader correlations that we’ve already found and then give them an interactive dashboard so they can look and explore the data for themselves. We wanted to make the data tell a story for itself and make it interactive so that the user would have to actually explore it for themselves and make their own conclusions.

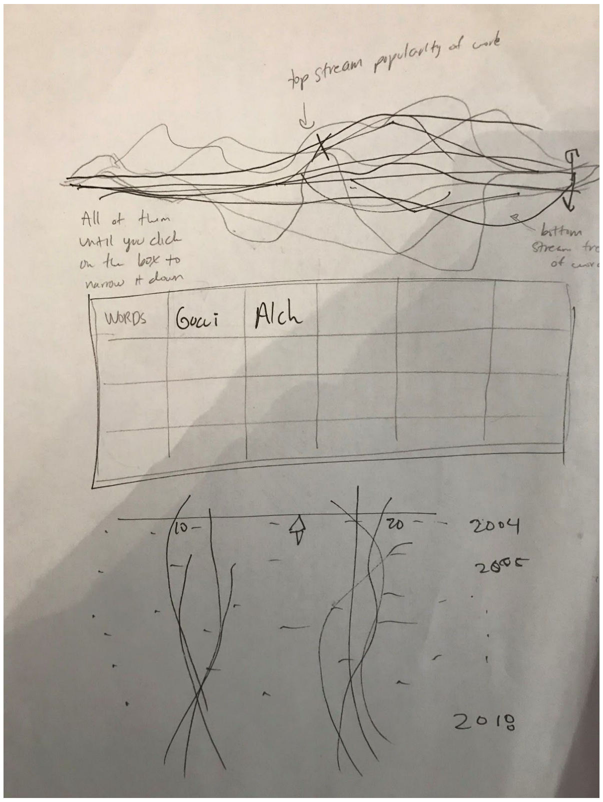
The reason we wanted to present the data this way is because we believe there are stories within the data they we won’t see / there’s just too much there for us only to do a story on it. This way, people can find stories within the data and can add that to the other findings. We also believe that there’s a lot more to search trends than just rap music lyrics. Search trends are based off pop culture in general and it would be cool to be able to add other pop culture references in the future and we think this design allows for that.

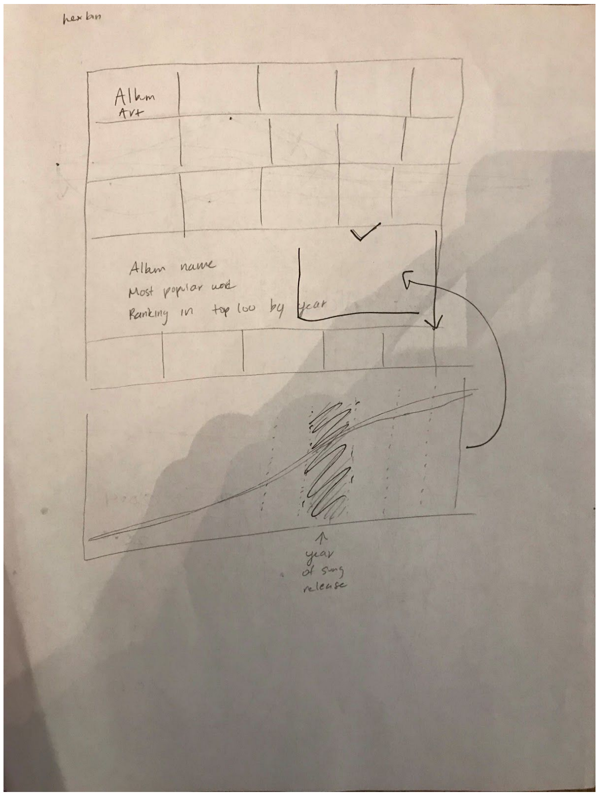
We considered multiple types of visualizations mixing in different features, but ultimately they all revolved around the stream graphs as a main way to show our big data and big idea. In the sketches below you can see that from the start, our idea really is what it is now. We had an idea and really thought it through and thought through how it would work with the data and through that process, we found something that worked the first time we tried it.

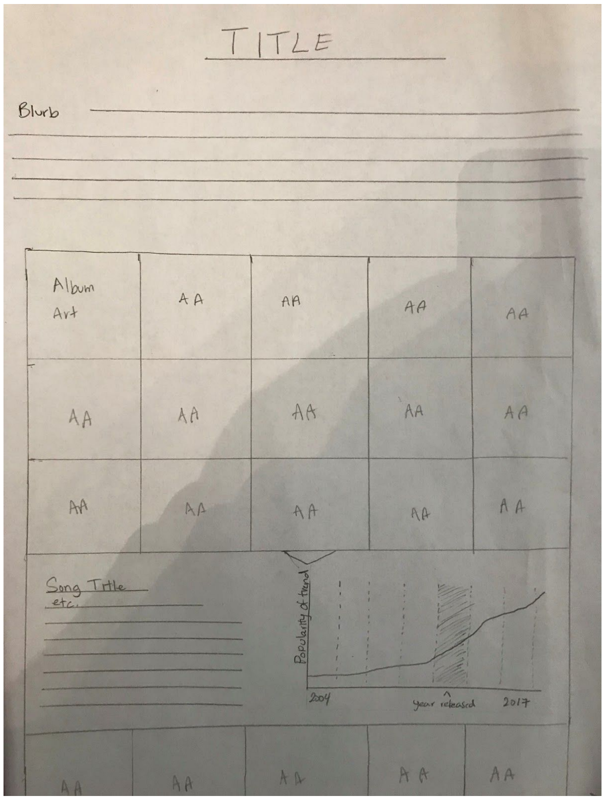
The main dashboard is based on a streamgraph. We have two stream graphs that are interactive viz tool tips and a pull down menu. The streamgraph changes based on which option is chosen in the menu and shows the popularity of the search term over time and the popularity of the word over time. The stream graphs are be displayed in order for the user to better see all the data on one page and easily draw conclusions on correlation.

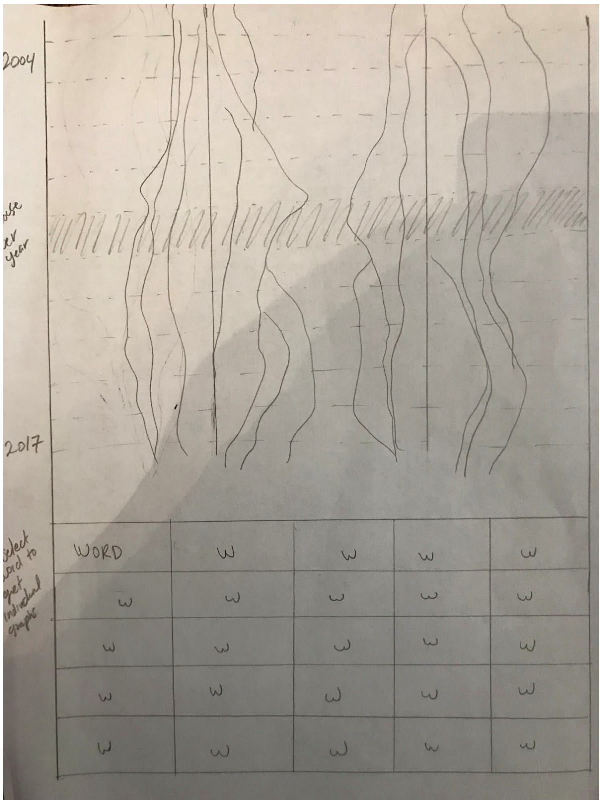
The other vis we have is our wall. This compares the most popular word in a particular song to the trend graph of that word. This is be displayed by clicking on an album artwork. It both displays information about the song and word and a small line graph showing trend data for the word. All of this can be seen in our design sketches below.

Sketches:

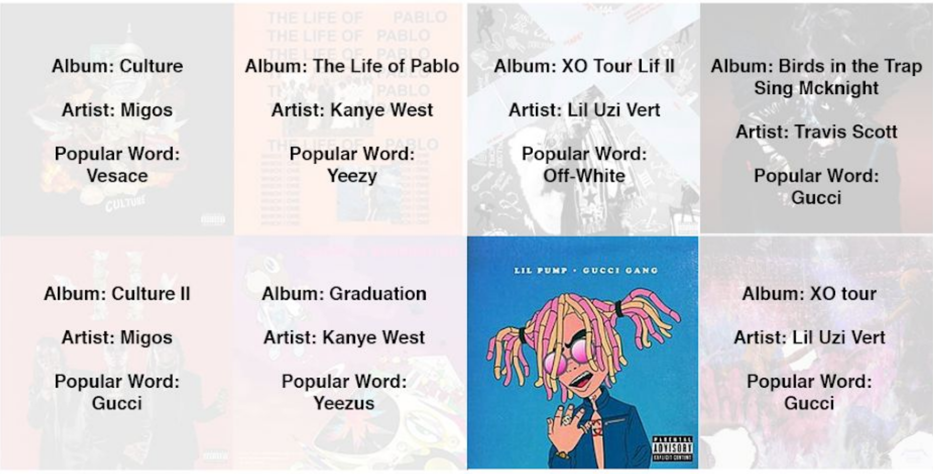

top stream popularity of word
All of them until you click on the box to narrow it down
bottom stream tre
of word
WORDS
Gucci
Alch
10
20
2004
2005
2018




TITLE
Blurb
Album Art
AA
AA
AA
AA
AA
AA
AA
AA
AA
AA
AA
AA
AA
AA
Song Title etc.
Popularity of trend
2004
year released
2017
AA
AA
AA
AA
AA


2004
ose
er year
2017
elect
ord to
get
individual
graphs
WORD
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W
W

Initial Design:



Album: Culture

Artist: Migos

Popular Word: Vesace

Album: The Life of Pablo

Artist: Kanye West

Popular Word: Yeezy

Album: XO Tour Lif II

Artist: Lil Uzi Vert

Popular Word: Off-White

Album: Birds in the Trap Sing Mcknight

Artist: Travis Scott

Popular Word: Gucci

Album: Culture II

Artist: Migos

Popular Word: Gucci

Album: Graduation

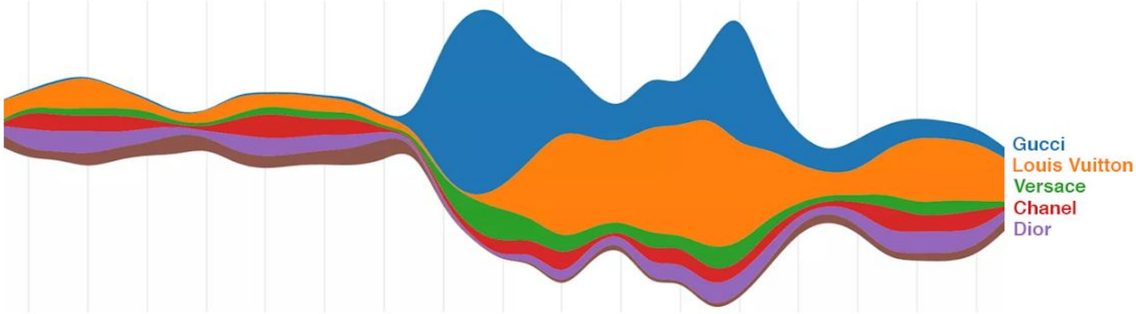
Artist: Kanye West

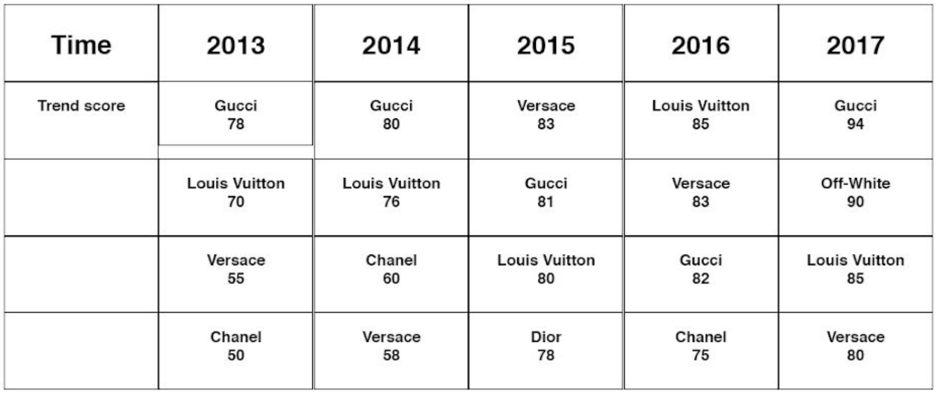
Popular Word: Yeezus

Album: XO tour

Artist: Lil Uzi Vert

Popular Word: Gucci



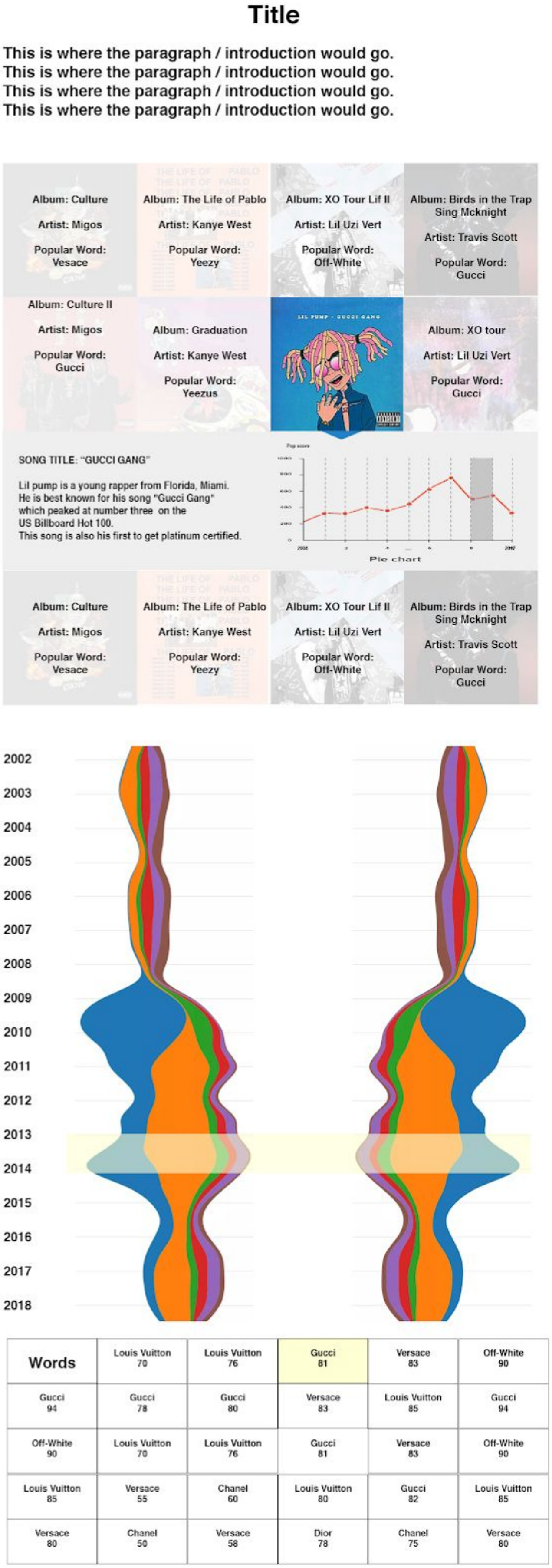


Time	2013	2014	2015	2016	2017
Trend score	Gucci 78	Gucci 80	Versace 83	Louis Vuitton 85	Gucci 94
	Louis Vuitton 70	Louis Vuitton 76	Gucci 81	Versace 83	Off-White 90
	Versace 55	Chanel 60	Louis Vuitton 80	Gucci 82	Louis Vuitton 85
	Chanel 50	Versace 58	Dior 78	Chanel 75	Versace 80

Who is “Lil Pump” ?

Lil pump is a young rapper from Florida, Miami.
He is best known for his song "Gucci Gang" which peaked at number three on the US Billboard Hot 100. This song is also his first to get platinum certified.
Story continues...

Final Design:



Title

This is where the paragraph / introduction would go.
This is where the paragraph / introduction would go.
This is where the paragraph / introduction would go.
This is where the paragraph / introduction would go.

Album: Culture Artist: Migos Popular Word: Vesace	Album: The Life of Pablo Artist: Kanye West Popular Word: Yeezy	Album: XO Tour Lif II Artist: Lil Uzi Vert Popular Word: Off-White	Album: Birds in the Trap Sing Mcknight Artist: Travis Scott Popular Word: Gucci
Album: Culture II Artist: Migos Popular Word: Gucci	Album: Graduation Artist: Kanye West Popular Word: Yeezus	LIL PUMP • GUCCI GANG	Album: XO tour Artist: Lil Uzi Vert Popular Word: Gucci

SONG TITLE: "GUCCI GANG"

Lil pump is a young rapper from Florida, Miami.
He is best known for his song "Gucci Gang"
which peaked at number three on the
US Billboard Hot 100.
This song is also his first to get platinum certified.

Album: Culture Artist: Migos Popular Word: Vesace	Album: The Life of Pablo Artist: Kanye West Popular Word: Yeezy	Album: XO Tour Lif II Artist: Lil Uzi Vert Popular Word: Off-White	Album: Birds in the Trap Sing Mcknight Artist: Travis Scott Popular Word: Gucci

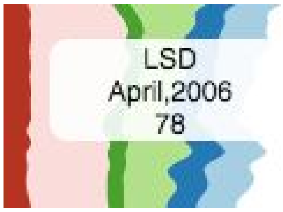
Words	Louis Vuitton 70	Louis Vuitton 76	Gucci 81	Versace 83	Off-White 90
Gucci 94	Gucci 78	Gucci 80	Versace 83	Louis Vuitton 85	Gucci 94
Off-White 90	Louis Vuitton 70	Louis Vuitton 76	Gucci 81	Versace 83	Off-White 90
Louis Vuitton 85	Versace 55	Chanel 60	Louis Vuitton 80	Gucci 82	Louis Vuitton 85
Versace 80	Chanel 50	Versace 58	Dior 78	Chanel 75	Versace 80

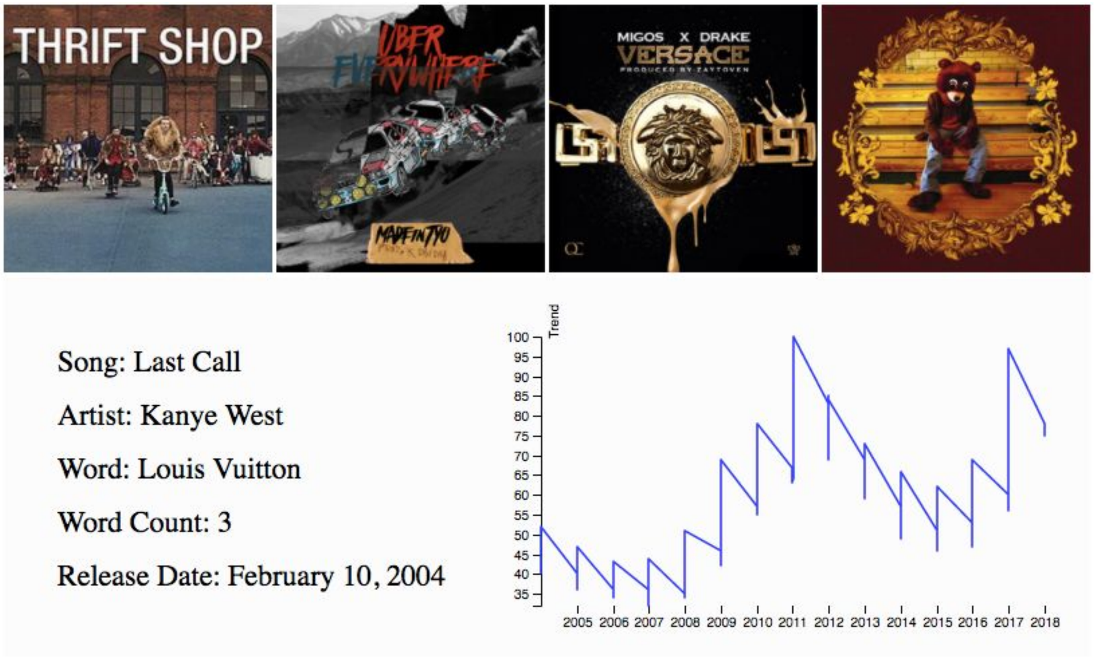
Implementation:

Our final design consisted of two main elements that helped to show our intent and story within context and with visual representation. The first is what we like to call our “wall”. It’s a grid of album artwork that is interactive. The second are the two stream graphs that boast the ability to interact with individual songs to see their correlation with search trends.

The “wall” is interactive in the way that when the user selects and clicks on an album artwork, a box is displayed below the wall giving details on demand. These details include: all the information about the specific song and popular word within that song. It also shows a line graph with the Google search trend data that corresponds with the popular word in that song.

The stream graph is actually two, side by side graphs that show the correlation between trend data and the most popular words overall. Overall meaning, throughout the years that we collected data on lyrics, we put together a CSV that holds the most popular words for tangible products and services. The data can be further refined to just see one word at a time, you do this through use of the drop down box at the top. Click on the word you want or reset to bring the original stream graph up for your viewing.





Song: Last Call

Artist: Kanye West

Word: Louis Vuitton

Word Count: 3

Release Date: February 10, 2004

Evaluation:

The data visualization really helped to solidify my belief that music is a good way to market products and ideas. If a song becomes really popular, people are constantly singing and saying the word so whatever it is, becomes a household name. Through the use of our vis, one can see for themselves that there really is a correlation there a lot of the time and it's quite interesting to look at different words with songs and different combinations.

As far as answering our questions, I think what we have gave us an indication that we are going in the right direction. By no means can we say the questions we had were answered, but I do believe that we can say the questions were definitely affirmed for being the right questions for the data and circumstance.

Because trend data doesn't really show the full picture of the impact a song had on pop culture, I think it would be almost necessary to bring in other data points to prove the point of it having an impact. The problem lies in the fact that it is extremely hard to find any quantitative data on influence of pop culture. The only thing possible would be scraping stuff like pop culture magazines etc.

One idea we did have was being able to feed in data for sales that a company has. That way, businesses could use it as a way to see if they did gain anything from the exposure via songs. It would also give a better picture of if the company actually experienced growth as a result of the exposure via music. This would also be possible to do via publicly traded companies.

Going forward, I think there's a lot of room for both improvement and an increase in functionality. We could add in the google trend API to allow for live data to be fed on demand, that way our searches would be more flexible. Instead of having data that is fixed, we would make a whole database full of a bunch of different words and the user could search for a term instead of having preset ones. That way, a user really would be able to draw their own conclusions from the data.

I think this visualization was a really good start into a really complex problem. The problem itself is actually quite a cool one in that the project could really keep expanding into something much bigger than what it is right now. It could really be expanded to be a tool for marketing with the use of songs.

Two things to note about the project are in the word count stream graph and the line chart for the album selector. In the stream graph, in order to get the axis to match up, we had to add months to the word count data. This is because you can't get a simplified version of trend data that only goes by year instead of also by month. Because of this, you’ll notice that our stream graph has become boxy. The only real way around this is to add in release dates for individual songs or somehow change the trend data to go by year instead of by month. We didn’t change the trend data because we didn’t want to manipulate the data. We didn’t change the word count data because we didn’t have data to show what songs released in which months. Along with this, changing the data collection process would’ve taken much more time than the time we had remaining.

As far as the line graph is concerned on the album selector, we also attempted to scale this to be larger, thus taking away the sharp movements in the line, but because there is so much data packed into the line chart, in order to make a difference in the appearance, the chart would have to be scaled to larger than the size of our actual canvas. Which is possible, but it would end up breaking the rest of our project because it’s all set to one canvas width.