feat. Visualizing the history of Hip Hop collaborations in the United States

Sonja Khan sonjak3@cs.washington.edu
Riley Porter rileymp2@cs.washington.edu
Kevin Quinn kchq@cs.washington.edu
Vinod Rathnam rathnam@cs.washington.edu

ABSTRACT
In this paper we describe 'feat.', a novel interactive visualization that facilitates dynamic temporal exploration of inter- and intra-regional hip hop collaborations in the United States. We provide a brief history of the hip hop genre and culture, discuss our methods for collecting and cleaning data, and the process of designing and refining the interaction.

Author Keywords
Data visualization, Hip Hop, collaboration, exploration, timeline

INTRODUCTION

Hip hop is a genre and culture born in the United States with a rich history and powerful sound. Our visualization is motivated by hip hop's emphasis on collaboration and regional distinction. Unfortunately, the availability of datasets centered around hip hop music are extremely limited, with most existing visualizations requiring hand compiled data focused on extremely specific aspects of the genre.

Brief History of Hip Hop

Hip Hop music originated in the late 60's in New York and has subsequently had widespread influence throughout the United States. From the original block-parties in the Bronx lead by DJ Kool Herc to the emergence of "Gangsta Rap" on the west coast led by Ice Cube, Dr. Dre and Eazy-E, Hip Hop has been a very dynamic culture. Since then, there has been wide-spread adoption of the genre throughout the country in the South, Midwest, and Northwest. Many pivotal events in the history of the genre have spurred conflict and unification in the Hip Hop community.

Motivation

Our project seeks to provide a tool for exploratory analysis of Hip Hop collaboration throughout its entire history in the United States. We wanted to offer a way for people with any level of background knowledge of the genre to explore the importance of collaboration and regional distinction in Hip Hop. Very few existing tools in the area provide an exploratory tool to analyze Hip Hop either regionally or temporally. This was evident by the lack of any unified or consistent data-sets in the field. By allowing users a fine-grained interactive tool, we hoped that anyone, regardless of prior knowledge, would be able to learn something new and become excited by a unique and powerful piece of American culture.

Another motivating factor in this project was to show how historically significant events such as the murders of Tupac Shakur and Biggie Smalls played a role in the regional collaboration trends in west coast and east coast Hip Hop, and also how these trends drastically influenced the distinctive sound of each region.

RELATED WORK

Hip Hop Collaboration Network [3]

This visualization encodes artists as nodes and collaborations as links, using data from Billboard's Rap Airplay Chart from 2002 through 2014. There is some size encoding for width of link and size of node, although it is not immediately obvious what this is encoding. Our best guess is that the size encoding is related to number of collaborations. There is also a hover interaction to allow users to explore the names of the artists in the network, and some interesting conclusions can be drawn, like which artists are separate from the main interconnected portion of artists.

The limitation of this visualization is that there are relatively few features incorporated besides the collaboration network. There is no sense of how long an artist has been active or where they are from. Also, the data set is limited since it is only from 2002-2014, and only includes tracks that reached the Rap Airplay Chart. Some of the artists are from genres outside of Hip Hop, like country and pop, and not including all tracks by the artists limit the links in the network.

J&M Hip Hop Graph [4]

This visualization also encodes collaborations between artists in a network, however, in order to not overwhelm the user, the visualization only shows collaborations with a single artist at a time. A user can select the main artist that they want to interact with, get some basic statistics about the artist, and see who that artist has collaborated with. The author used color and size to encode how many collaborations total each artist has. This provides more information than the previous visualization, because it gives a sense of which artists collaborate more and includes more data overall.

This visualization was a little bit difficult to interact with, because a user could only view how one artist fit in at a time. Also, the network was pretty cramped and difficult to search. There was no way to know which tracks artists had collaborated on, or when the collaboration happened. We wanted to encode these features into our final product, and also wanted to include regional data, which this visualization did not incorporate. We had to figure out how to show a network similar to this, that allowed for interesting discoveries, but also encoded regional, temporal, and more collaboration data.

METHODS

Our method for creating the visualization had two distinct phases – data collection, and interaction design and implementation.

Data Collection

Our data came from multiple different sources. Below we describe each step.

Selecting Artists

In order to create our collaboration data set, we first had to decide on the domain of artists that we would include. In making this decision, there were several limiting factors including time restrictions, and available space on our graph. We wanted to ensure that the most well-known artists were included in the visualization, so we started by selecting the 50 top selling Hip Hop artists of all time. We followed this up by trying to get a consistent cross-section of Hip Hop artists throughout history in order to provide a more representative sample of artists from the '70s and '80s (who tended to have less record sales). Beyond these criteria, we also tried to sample artists from varying geographical locations to ensure a better visual spread on our graph. There were also many artists that were excluded because their primary genre was not Hip Hop. See *Future Work* for more discussion of this.

MusicBrainz

MusicBrainz is a database of music metadata maintained by a global community of users [1]. We wrote a python script to query the *MusicBrainz* API for each artist in our set, looked through each artist's records, and determined tracks that included collaborations with other artists in our set. These tracks were added to our data-set, which we store as a JSON file. We were rate limited to 50 calls a minute, resulting in a runtime of close to 5 hours for the initial data collection.

Spotify

We incorporated the ability to play music into the visualization, using *Spotify*, a music service, as our data source for the audio playback. *Spotify* provides a JavaScript API [2] that allows requests for tracks as the user interacts with *feat.*, but one limitation with this approach is that there is a limit that the number of requests from *Spotify* cannot exceed. Retrieving entire albums could have solved this problem, but artists generally only collaborate on 1 or 2 songs per album, so again the limitation could occur.

Our solution to this problem was to write one script that retrievd all the unique *Spotify* URIs for the tracks in our JSON file. We then updated the JSON file such that each track includes the *Spotify* URI. We avoided the request limitation problem by giving some time in between each track request. To display tracks in our visualization, we loaded an HTML <iframe> tag with all the *Spotify* URIs needed. This solution greatly improved performance.

By using this approach and *Spotify* we were forced to make some sacrifices for smoother integration.. One, the script wasn't always able to successfully retrieve a URI for a given song. Two, *Spotify* does not host all the songs in our data-set. We believe that most of the missing tracks come from the latter reason.

Data Cleaning

Many of the tracks we pulled from *MusicBrainz* were unofficial releases that included a fair number of duplicates. In order to remove inconsistencies and clean-up the data, we had to manually spot check each collaboration that was generated. This involved using existing domain knowledge and manually looking up collaborations to verify consistency. One of the largest issues was 3rd party artists and producers creating mash-ups or unofficial tracks using samples. Unfortunately, the *MusicBrainz* API did not filter these results, so we had to manually determine their authenticity. While sampling is an interesting aspect of the genre, we decided to exclude these links from our visualization.

Visualization Design and Implementation

We iterated through many designs for our visualization before settling on the current interface seen in Figure 3. The map interaction had two main forms of interaction - regional collaborations within a map of the United States and artist to artist collaborations within a region, and two detailed panels - one showing the relationship between two artists, and another showing a summary of a specific artist. We implemented the entire visualization in html, css, and javascript using the d3 framework and SVGs. The region views presented the most challenges during implementation.

Region Views

The regions in the zoomed out view started out as colored circles that grew in size as artists became active (Figure 1). We decided this was confusing for the viewer because size was not used as an encoding in the zoomed in view. In the end we implemented the regions as closed b-splines that encompassed all of the artists for each region, and as artists became active we displayed small dots to indicate their existence. When inter-region collaborations began in the 90's, we showed these via links between the regions, where the link thickness encoded the volume of tracks.

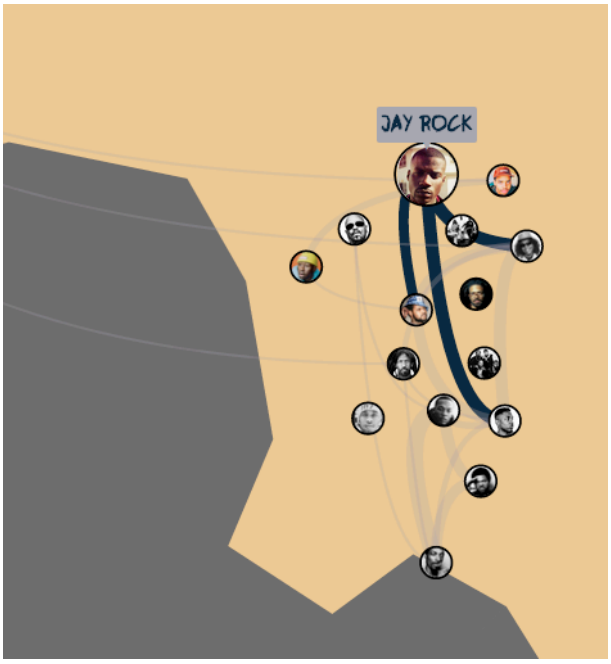

Figure 2. Single Artist View

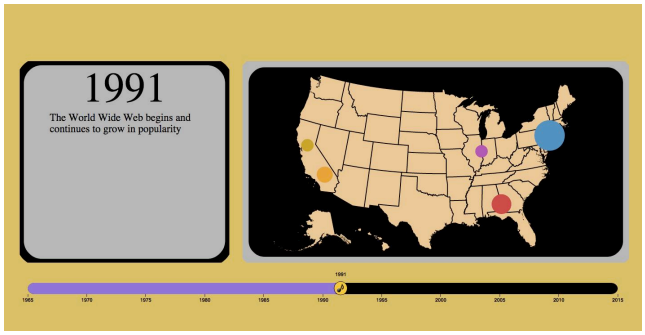

Figure 1. Initial interface design

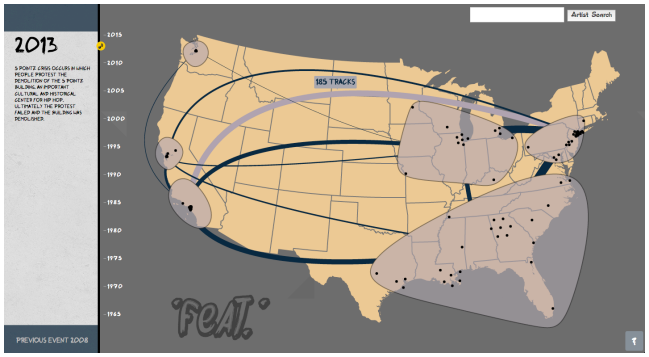

Figure 3. Overall view

The zoomed in view proved to be difficult because many artists originated in the same city, and some areas were densely populated (Figure 2). In order to display each artist without overlap and with enough space to place links between them, we had to hand-jitter their latitude and longitude. Active years were taken into consideration to maximize reuse of space. 64 of our artists were from New York State, and it was impossible to display them with other artists from the Northeast without completely disregarding the geography. We implemented a two level zoom in order to solve this problem, where a user would first zoom into the Northeast, then zoom into New York to access artists in the city.

RESULTS

Navigation Through Time

feat. stands out among Hip Hop collaboration visualizations due to the inclusion of temporal and regional encodings. The temporal encoding was vital in portraying the events in Hip Hop history and the effect on collaborations between artists. In the visualization, the artist nodes appear and the links between artists and between regions appear and get wider when a user advances through time to reflect how the landscape of collaborations is changing.

Although *feat.* is a visualization that provides detailed information on artist collaborations to those who already have a background in Hip Hop, we aim for it to be useful to an audience that is interested in learning more about the genre as well. Our visualization tells a story of Hip Hop collaboration through the data. We implemented a narration feature that provides cultural context. This narrations spans over time and reinforces the importance of temporal changes in how artists collaborate with each other.

The complete temporal navigation bar is situated on the left side of the visualization, as can be seen in Figure 3.

The user can use scrolling or click-and-drag to move the slider icon to a different year in Hip Hop's history. As this happens, the main view of the visualization updates to include the correct depiction of artists and collaboration links between them. But as the slider icon changes position, the narration text and year changes as well. This is a concise summary to avoid overloading text for the user.

We employed *scrollytelling* as another tool in our efforts to tell a story through our visualization. As the user scrolls, the slider position and year change to display updates in the narration and the map views. This is a somewhat hidden action that the user can take, so we included instructions in the navigation bar and a help dialog.

Another tool used to control the temporal variable is the shortcut buttons to skip to the next event. These are situated at the top and bottom of the navigation bar. These shortcuts provide another means for users to change the time setting of the visualization other than scrolling and sliding. Furthermore, this allows jumping straight to events; excluding years without events.

The events were accumulated through research and prior knowledge of our group members.

Spotify

The ability to play music through *feat.* tremendously improves the validity and usability of the visualization.

Sound integration allows users to audibly notice transformation and impact of artist collaborations throughout time. Having charts and text to show the collaborations between artists is worthwhile and informative, but being able to play and listen to those songs reinforces the validity of our data. Moreover, *Spotify* provides more usability scenarios; one can now also use this visualization as a an artist discovery tool.

The *Spotify* player demos tracks in both the artist to artist comparison view as well as the single artist view. The comparison view it provides some of the tracks both artists have collaborated on, and the single artist view it presents all collaborated for that artist.

Artist Views

The first artist view shows all tracks that two artists collaborated together on up until the year selected in the timeline. When *Spotify* tracks are available, an embedded player is displayed, allowing the option to listen to the tracks. The second artist view is a summary of a single artist, again only for tracks up until the selected year. This view gives links to websites related to the artist, including official websites, social media sites, and music sites, as well as a graph showing the breakdown of collaborations across all regions and a list of all tracks they have collaborated with others on.

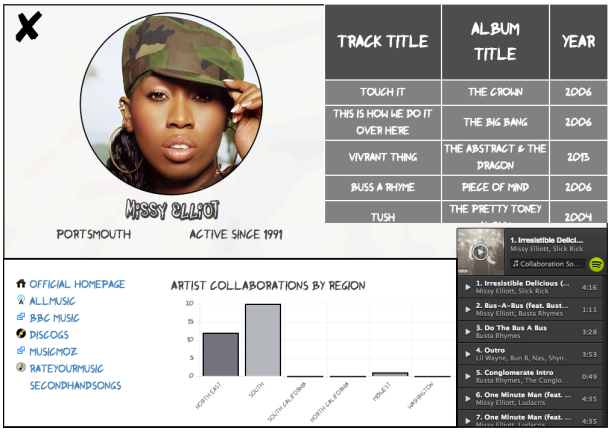

Figure 4. Single Artist View

Search Feature

We created a search bar to provide faster access to a specific artist's collaboration information. We also provided an auto-complete feature to assist users. Searching through this tool zooms into the region view the artist is located in. To indicate the specific artist among the others, the artist node will grow for a set amount of time before reducing back to its original size.

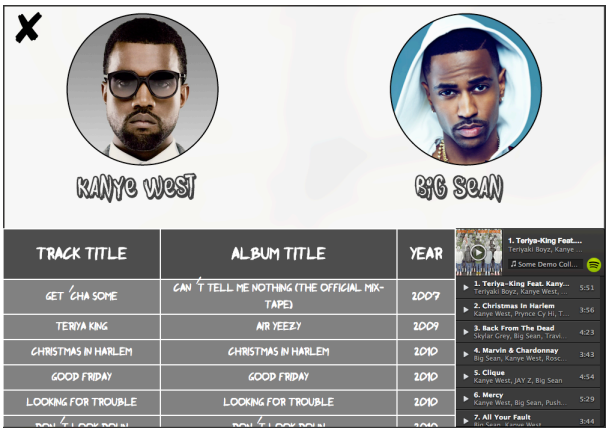

Figure 5. Multi Artist View

DISCUSSION

Our final visualization allows for novel interactions with a Hip Hop data set, which few of the existing works we explored have incorporated. A user exploring our system has some guided context through our events navigation bar, but also has full fine-grained control over exploring the data set in whichever direction they choose. Using our visualization, users can glean a variety of insight regarding 1) when collaborations start popping up, 2) which artists and which regions are collaborating the most, and 3) where artists are distributed across the country. With the addition of the temporal axis, a user can also learn about how the data is distributed through time. A user can explore how events have affected active artist concentration, when collaborations start to happen and which artists are doing these collaborations. With the addition of audio playback, another novel insight a user could make is how tracks evolve over time, and if there is a trend in sounds over time. While looking for related works, we could not find anything with a similar feature set or amount of comprehensive data.

While presenting our work at the poster session, we gained some valuable feedback on improvements that could be made to the system. Firstly, some of our users had issues with some of the basic interactions. Some additions that we think would help this are: add functionality to go back a view more easily, add zoom out buttons, allow clicking anywhere on the timeline to jump to that year, and adding a play button to allow a user to more intuitively figure out how to first interact with the visualization. Secondly, there was some confusion about the temporal axis and how time was incorporated into the various views. Two ideas to assuage this issue are to encode data on the timeline itself to give users a sense of where the data is clustered at a high level, and provide a separate timeline context when viewing one of the artist or linked views. Users at the poster session did not realize the views of detailed artist or collaboration information were based on the current year they where exploring. Overall, the feedback we received was overwhelmingly positive and none of the users mentioned confusion over visual encodings.

FUTURE WORK

Additional Data

There are many ways to represent the history of Hip Hop besides collaborations. It would be interesting to incorporate other data to show as an extension to the existing visualization, such as lyric analysis.

Extensibility

One of the challenges of this project has been collecting and cleaning the data, and finding space to display the artists. It would be helpful to explore options of how best to add additional data to our visualization without requiring manual labor. This also includes crowd sourcing the data gathering and correction.

Numerical Analysis

We have compiled a very comprehensive data set in the process of creating our visualization. This visualization does not analyze the data, but rather shows it in a way that enables exploration. We would like to see some analysis of our data set, and have made it publicly available in the hopes that it will inspire others.

ACKNOWLEDGMENTS

We would like to thank Professor Jeff Heer and the CSE512 staff, as well as Mike Bostock for the numerous d3 examples provided online. Thanks to the MusicBrainz community for taking the time to contribute data.

REFERENCES

1. MusicBrainz. https://musicbrainz.org/doc/python-musicbrainz2.
2. Spotify Web API. https://developer.spotify.com/web-api/
3. Bloom, Nick. "A Hip-Hop Collaboration Network 2002-2014." Hip-Hop Collaboration Network. N.p., n.d. Web. http://nickbloom.net/d3/rapnetwork/
4. Beach, Joshua. "J&M Hip-Hop Graph." J&M Hip-Hop Graph. N.p., n.d. Web. http://jmhiphopgraph.com/