Eliot Huang, Daniel Frier, Eric Porter, Albert Caldarelli
Pitchfacts
INFO 4310
Final Project Design Document

Project Goals & Motivation

Our goal for this project is to explore how Pitchfork reviews of different genres and artists vary linguistically in both word choice and word usage. Our main goal is to attempt to create a "Genre Wheel" that spatially separates music genres much like the Color Wheel separates colors. This was motivated in part by this project in which someone used a chord diagram to represent the relationship between different genres as given by the Spotify API. Although we are not using a chord diagram, our goal would be to create a clear shape, ideally a circle, that defines particular music genres based on words and language used in reviews.

Intended Audience

The intended audiences of our design fit into the two broad categories of music enthusiasts and a more technical audience interested in linguistic analysis. As will be described in later sections, we have decided to focus our design most intently on its linguistic components, however we recognize that a design intended only for a technical audience would have limited reach. Because of this, we intend to provide considerate background on Pitchfork and our visualizations in order to attract more casual music listeners as well, however, our main target is still geared towards advanced listeners.

We hope to build a narrative from the analysis about the review culture of different genres of music in an understandable enough way for music enthusiasts or Pitchfork readers to enjoy. This also represents one potential use case for the design. Individuals can investigate the lexical differences in style that lend themselves to different genres, or whether these differences exist at all.

At the same time, we recognize that language analysis is a concept that can be rather abstract except for those who have practiced it. We intend to make use of this possible familiarity within our audience and allow them a high amount of control over the formulas that generate the visualization. This technical control represents a second form of a use case, wherein the audience members can reach their own conclusions about the data through manipulation of its parameters.

Related Materials

Some related materials that helped inspire this project were reviews from online sites like Pitchfork, Genius, and Pigeons and Planes. As avid music lovers, we as a team decided to investigate if certain people and demographics who listen to music genres expect a particular type of review, and we wanted to predict if the sentiments of that review are predictive of that particular genre. These types of music reviews made us think about which one was most influential to a particular genre and how one review can differ. We also questioned whether Pitchfork's writers changed their language used depending on what they were reviewing.

We chose Pitchfork in particular because of its quirkiness and long-time established influence within the alternative and less mainstream music communities. For example, one interesting fact is that from 2011-2015, Pitchfork reviewed 4 separate Meek Mill albums and gave every single one the same score of a 7.4. Another interesting thing is Pitchfork's infamous review of Jet's *Shine On*, where they gave the album a 0.0 and the review entirely consisted of a Youtube video of a monkey peeing on itself. Pitchfork's eccentricity as a reviewer made us question and want to analyze how the organization's actual reviews had established its amount of musical credibility in the industry.

Another thing that made us interested in examining Pitchfork reviews was the rise of internet review culture and the growing influence of music reviewers. For example, there have been many documented articles that credit the rise of now popular artists like Bon Iver, Sufjan Stevens, and Arcade Fire solely to a positive Pitchfork Review, including an article in the Washington Post titled Giving Indie Acts A Plug, or Pulling It which explains how after Pitchfork's gave then-unknown Arcade Fire's *Funeral* a score of 9.7, sales for the record exploded and the band reached mainstream success. With the rise of viral music reviewers like Anthony Fantano and the prominence of established publications like Pitchfork, we grew more interested in seeing what these reviewers were actually saying in their reviews and how their words could, literally, change the lives of artists.

Data Source

For this project, we will be using a dataset of over 18,000 Pitchfork reviews found on Kaggle at this linked here. This dataset contains many valuable features for this project, including review ID, genre of review, author, publication date, score, etc. We came across this data set initially while browsing on Kaggle's list of datasets looking for a project idea at the beginning of the brainstorm.

Possible Designs We Considered

Due to the multi-dimensionality of our dataset, there were many different kinds of approaches we considered taking when choosing our visualizations. One approach we considered was visualizing the distribution of Pitchfork scores, subsectioned by genre and artist. With this approach we would have wanted to show how the average score for certain genres has increased or decreased over time and make some broad claims about relative cultural life and death of genres in the music mainstream. Analyzing the global average Pitchfork score would also be interesting to see if Pitchfork has gotten more or less harsh with their scores as the site's prominence has risen with the rise of the Internet and online review culture. We also would report some interesting statistics on specific artists that have had the most variable or consistent score distributions from Pitchfork authors. Ultimately, we decided to move away from this type of approach to the data because it seemed rather basic and we thought it might be hard to shape a specific and connective story for the visualizations we intended.

Another approach we considered was to create a playlist generator. Essentially, users would be able to enter in their preferred genres or artists, and we would find Pitchfork reviews or authors that most closely matched the preferences that they entered. Then, we would retrieve the highest scores of albums within those genres and create a small playlist for the user, as well as some visualizations on similarity between the user and Pitchfork authors' tastes. During the poster session where we mocked up this idea, we quickly realized that this playlist generator not only seemed too complicated to explain to the user, but it also didn't seem to be in the scope of the project's data visualization intention and felt more like a loose information retrieval tool.
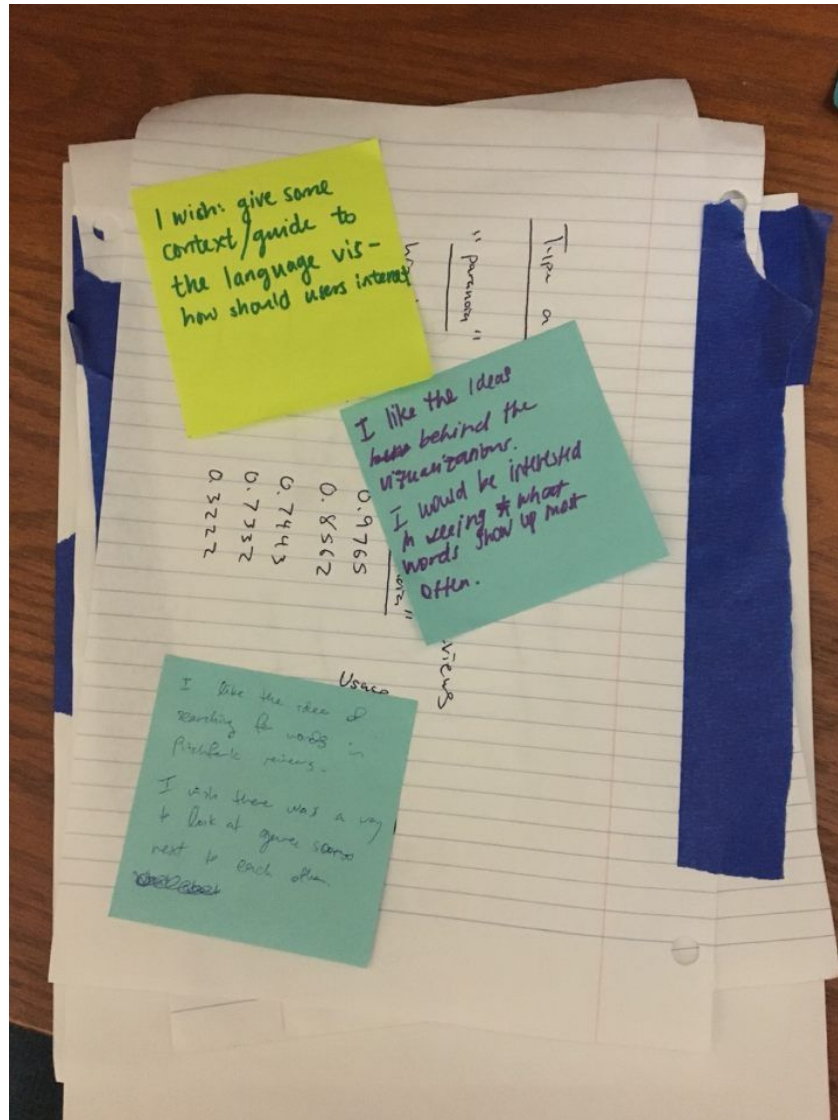
Some of the feedback we received from the poster session

Moving away from the Pitchfork scores and the writers, we began to think of how we could visualize the Pitchfork reviews themselves. This led us to wanting to approach this dataset from a more linguistic and text analysis standpoint. Taking the feedback that we got from the poster sessions, we thought of ways to combine the inter-genre score analysis we had originally and apply that to the word content of the reviews. It also seemed like, from the poster feedback we got, people were most interested in our text analysis ideas. Because of this, we decided to investigate how word usage in reviews of different genres varies, and what words happen to be used most frequently for reviews of each genre. This approach felt more easily explainable to users than our playlist generation idea, and felt more interesting as a whole.

Final Design

Our final design for our two visualizations is based on our more linguistic and textual approach to our dataset of over 18,000 Pitchfork music reviews. Our first visualization will consist of a large graph of reviews generated using LSA on the text of each Pitchfork review, then with certain interesting points highlighted. This will ideally generate fairly separable review clusters based on which genre that review was taken from so that users can see a general shape to the language used in reviews, that will hopefully correspond to genre differences. Users would be able to select specific areas to zoom in on, and genres to highlight, after which the visualization would filter out other genres and only show that genre's reviews, or that specified area, with a text display of rankings next to the cluster to indicate which reviews are most representative of that genre, and information about them such artist, tagged genre, and the words within the review that are strongly correlated with their genre. By drilling down like this we are going to be able to see what it is that defines how a reviewer discusses a certain genre.

By hovering on one of the interesting points, a small window will appear, then the user can click to hold it in place. The window will contain information such as album, artist, rating, Pitchfork labeled genre, linguistically guessed genre (i.e. which cluster it belongs to), and a link to the review.

Our second visualization is similar to a product comparison chart. Users can select two genres out of a list to be displayed side-by-side. Each display for a genre would contain words in Pitchfork reviews that are used most or least frequently compared to the other selected genre. For example the word "yo" might extremely high for hip-hop if compared to a genre like country. At the bottom of these two displays would be a shared word pool, which would consist of words that are most commonly used between the two selected genres. Ideally, users would be encouraged to experiment with different combinations of genres to see interesting similarities in vocabulary across genres.

Through our redesign efforts, some trade-offs had to be made. Originally, as described in our possible designs section, there were a few competing philosophies and we attempted to accommodate all of them in our first efforts. The nature of the dataset as a collection of music reviews lended itself, naturally, to investigations about music. Over time, however, we decided that much more interesting analysis could be done on the dataset's lexical properties with some only tangential references to the music-specific fields. We recognize that the direction we have chosen to take the final visualization will make it difficult to build a narrative that will interest a section of our intended audience, music enthusiasts, but we believe it is still possible to tell that story.

Implementation

Data Scraping
Before actually implementing these visualizations, we first have to retrieve the text data of the Pitchfork reviews. Because this dataset only contains a link to the actual Pitchfork review, we will probably be using some sort of online automatic scraper to extract the text data for each of our given reviews. This will involve creating another dataset that we will then be analyzing to create our text analysis visualizations.

Data Cleaning
After scraping and accumulating our data, we are going to clean the data by removing common stopwords and terms such as artist names, and genre words like 'rap' or 'pop' so that they don't skew the data unfairly.

First Visualization
Our first visualization will be implemented with the use of several text mining algorithms, primarily LSA (Latent Semantic Analysis) to convert the 'bag of words' in the reviews to lower, 2-dimensional coordinates that we can then plot. We may also include a slider that lets users change the specificity of the algorithm by removing more and more common words from the vocabulary, thereby 'forcing' the algorithm to generate clear clusters. We are hoping that a point will become clear at which the data forms clear clusters without being manipulated too much.

We are going to make the general points we plot semi-transparent, and tied to a hue generated from their coordinates so a general trend can emerge (sample image 1), then as the specificity is increased, more clear shapes will emerge (sample image 2). The hover-window from selecting, and the zoom feature can be seen in sample image 3.

Sample Image 1

Sample Image 2

Sample Image 3
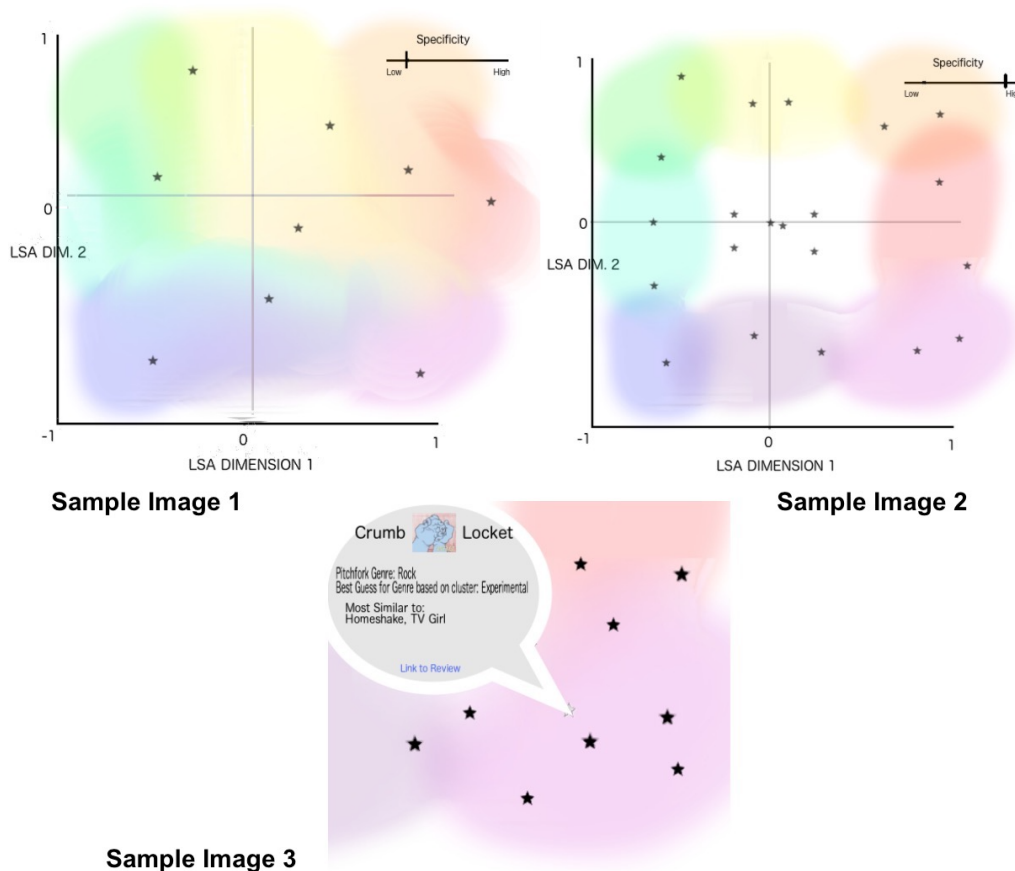
Seeing as how there are over 18,000 reviews, we are going to have some of the reviews show up with lower opacity and not be 'hoverable', and the more significant ones which we will handpick, as stars or some other symbol. We are going to also have the dimensions correspond to hue so that the channel of separation, or similarity, between them is doubly salient (position and color). Applying a clustering algorithm to the points will hopefully give us several distinct genres represented by the clusters, which we will then create a story for and explain.

Hovering over the points will reveal 'interesting' reviews, and the further you drill down into the data set by zooming in, new 'interesting' reviews will show up for only that subsection of the graph. This will let users explore the linguistic data behind the reviews fully.

Additionally, due to the large nature of the dataset, real time processing of the data would be far too costly and lagging, so we are going to apply the LSA technique and the clustering algorithms beforehand, generating a list of coordinates and points that we will

base the graph generation on. This way the visualization can perform fluidly and will have considerably shorter load times. We are going to do this by creating coordinate files for each level of specificity (in steps of 10) and create the visualization from these so that the user can still interact with the visualization, and it can load the points quickly.

Second Visualization

For the second visualization, we intend to show the greatest differences and similarities in particular words used between two genres that the user can select. For similarity metrics, we are considering using cosine similarity or tf-idf scores. These metrics will be computed on a subset of all reviews for each particular genre. We will then retrieve the top scoring words for each genre and provide a side-by-side comparison of the most similar and most different scores for particularly salient words.
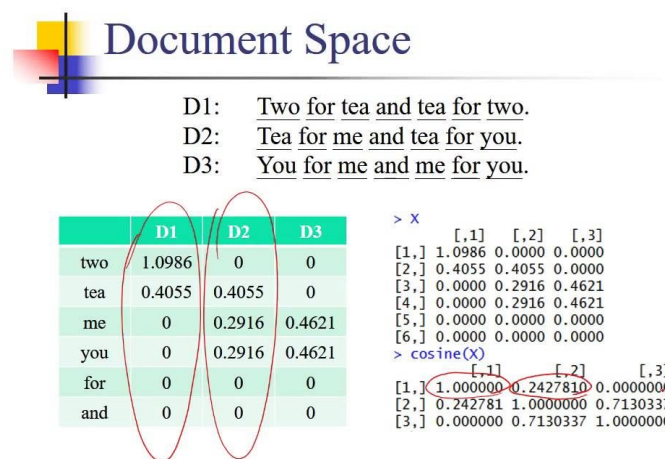
An example of text similarity comparisons between documents
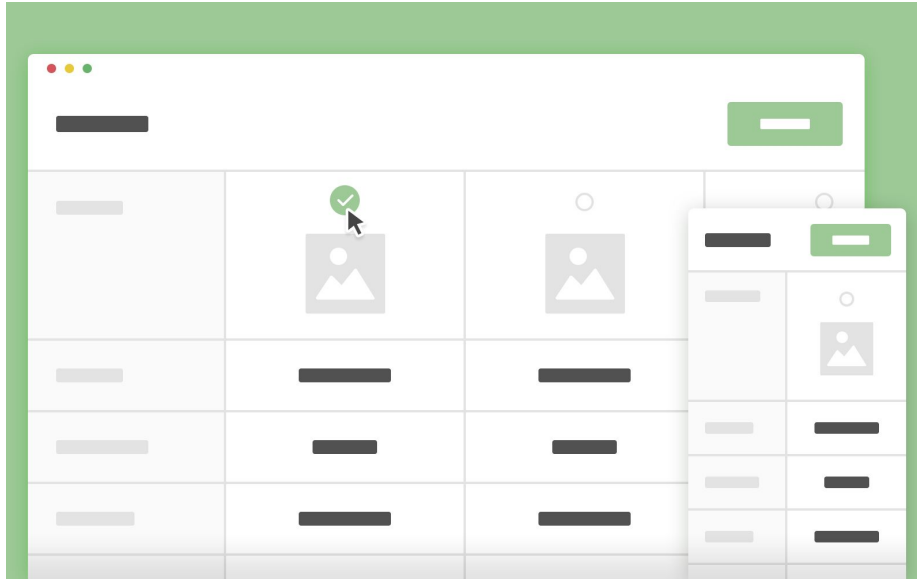
An idea for our genre-word comparison chart, where each column is a genre and each row represents a similar or different word

Team Roles

Eric
I anticipate helping with the flow of the document's timeline/storyline. I will also be responsible for helping scrape the text data from Pitchfork given the links in our dataset. I also anticipate making part of one of our main visualizations and potentially supporting static visualizations that are going to be on the final project.

Albert
I expect to help with the scraping of data as well as the implementation of the visualizations. I also will help with the design and flow of the web page as a whole.

Eliot
I will be in charge of helping implement the visualizations, specifically the second main visualization and computing the various text similarity metrics like cosine similarity and tf-idf. I will also be helping with the general design and story behind the news article. I will also probably be creating the final document and video for the project.

Daniel
I'm going to responsible for much of the data cleaning and manipulation. This includes applying text mining algorithms to the data, primarily LSA and perhaps topic modeling and word vectors as well for the first main visualization.