Visualizing Economic Effects on Popular Music (1958-2012)

David Price / Glen Barger

CS171: Spring 2014

Contents

1 Overview and Motivation 2

2 Related Work 2

3 Questions 2

4 Data 3

4.1 Chart and Music Quality Data 3

4.2 Lyric Data 3

4.3 Economic Data 4

5 Exploratory Data Analysis 4

6 Design Evolution 4

6.1 Initial Sketches 4

6.2 Adding the Lyric Data 8

6.3 Telling a Story 10

6.4 The Circle Question 11

6.5 Word Stems 12

6.6 Word Stems and the Song Details 12

7 Implementation 13

8 Evaluation 16

1 Overview and Motivation

Our project is a visualization of how popular music might be influenced by the economic environment of the time in which it became popular. We show US economic indicators together with the music that made the charts during the period 1958-2012, visualizing musical qualities and lyrics.

We really wanted to do something fun for this project, preferably something in the entertainment realm. We've both been exposed to Pandora's Music Genome Project through Pandora itself, which categorizes music in standard ways, allowing an automatic way of finding music that you like. We were interested in somehow mining that information, but upon learning that this data is proprietary, we investigated other sources. During the course of this investigation, we looked at multiple data sources and became drawn to the idea of seeing how music and economic indicators might correlate. We thought this would be a good opportunity to build a visualization that allows for interactive perusal and storytelling.

2 Related Work

For chart data, the following site had some very interesting insights using the Whitburn data we used.

http://rstudio-pubs-static.s3.amazonaws.com/9904_148209607fb84e9cac366ba009faf8e6.html

While we liked the depth of their analyses, we were looking to provide the ability to explore the data in different dimensions, especially qualitative measures.

For lyric data, the following visualization in the NY Times was inspiring:

http://www.nytimes.com/interactive/2011/06/10/education/commencement-speeches-graphic.html?_r=1&

This gave us the idea for grouping the lyric data so that we could shape it in order to tell a story, rather than leave it to a random word cloud kind of implementation. It also gave us the idea for using circles to represent counts, an idea which we later abandoned.

Another article was quite convincing about the need to tell a story when doing visualizations about word data, was primarily responsible for turning away from the word cloud direction:

http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/

3 Questions

While looking at music data, we initially wondered if we could visualize how musical attributes might shift over time, especially in response to economic events. For example, during a recession, does the mood of music become more, or less, energetic? Is music more "danceable" during good economic times?

We then began thinking about what song lyrics could tell us, and wondered how we could mine and visualize that information. Do the lyrics contain more, or fewer, references to things like money, paychecks, and work during a recession, or does the opposite occur – during recession, do people want to hear about distracting subjects that have nothing to do with the economy? In good times, do words with happy, positive emotional connotations predominate? Are words that are associated with a positive outlook more prevalent? Is there any correlation at all?

4 Data

4.1 Chart and Music Quality Data

We began our data gathering with the "Million Song Database," a dataset curated by The Echo Nest. The Echo Nest is a company borne out of the MIT Media Lab in a project to catalog musical attributes. Echo Nest is now used to help drive recommendation engines behind popular online music applications like Rdio and Spotify, the latter of which recently announced the acquisition of The Echo Nest. Right now Echo Nest provides a very robust and free API for developers to call and retrieve their catalogued attributes.

The Million Song Database contains "pure" music data. In other words, it doesn't include attributes like popularity, or even indicate the year that the music was released. We needed a way to be able to select only the most popular music. For this, we first turned to the Billboard charts, but obtaining this data required screen scraping, and they had broken out their ratings into numerous genres. We then tured to another dataset, the Whitburn Project.

The Whitburn Project is maintained by a group of music enthusiasts whose goal is to document chart data on popular music starting in the 1890s. The data details all music that reached the top 100 (not only restricted to Billboard), and includes how long the song was in which position for each week it charted, as well as its overall rank during the year. You can read more about the project here.

With these two datasets in hand, we were able to filter the data that we retrieved from the Million Song Database to include only songs that made the charts. That is, we looped through the Whitburn data (approximately 27K records) and found the matching songs from the Million Song Database. We chose to begin with 1958 because that's when music charts really took off.

To retrieve the data from Echo Nest, we created an HTML document that utilizes d3 and Javascript to loop through all the rows in a csv file generated from the Whitburn project, then call the Echo Nest API once every 4 seconds. This call returned details of a single song. The result of this process was a new csv file with the attributes. The data from this csv file was then opened in excel and vlookups were created to merge the original data in with the Echo Nest data.

4.2 Lyric Data

For the lyrics data, we began with a dataset that was created by a collaboration between Echo Nest and MusicXMatch. The latter company offers access to full lyrics, but at a commercial price. The dataset that we used, however, is free, because it doesn't contain full lyrics, due to the numerous copyright issues involved. Instead, it offers word counts (using word stemming) for each song in a "word bag" form. In all, there are some 237K songs for which there are word bags in this dataset. This data can be found here.

To match this information with the songs we have, we took our existing song list and used a table of 779,000 records to match up the artist and title to get a foreign key that could then be used to match up our song list to the lyrical list. Before getting the lyrical counts though, we picked a set of words that we could use to fill the buckets for our design.

Separately, the MusicXMatch data includes a list of the top 5000 words in all the songs. We went through those words and looked for ways to aggregate them such that we'd get more meaningful data. We assigned a group to the words (only using those words that fit into one of our groups). Initially, we looked at grouping words into a small set of categories. At the end, we had the following groups:

- positive emotions and outlook
- negative emotions and outlook
- money-related words
- relationships (nouns)

- words with religious connotations
- violent words
- colors
- parts of the body

The larger number of groups provides some interesting dimensions into the lyric data, and turns out to be fun to explore.

4.3 Economic Data

We're using data from the Federal Reserve Bank of St. Louis (http://research.stlouisfed.org/fred2). This date includes the following two series, covering the period 1958-2012.

- Civilian Unemployment Rate, Percent, Monthly, Seasonally Adjusted
- S&P 500 Stock Price Index, Percent Change from Year Ago, Monthly, Not Seasonally Adjusted

5 Exploratory Data Analysis

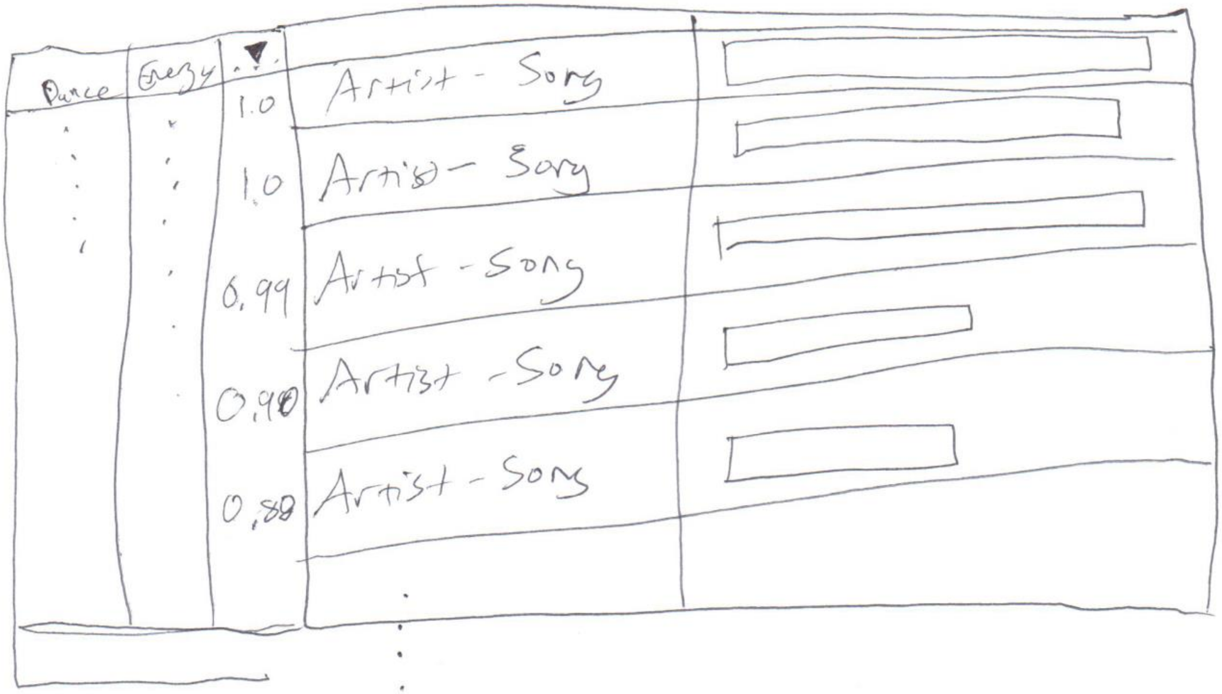
Because the data required such extensive manipulation, we could only look at the original data sources, but these sources, in isolation, didn't tell us much. In fact, one could say that the entire exercise of producing the final visualization was because there was no other way to see this data. We tried using Tableau, but it seemed that it was more trouble than it was worth, especially considering that we really couldn't get to a useful visualization without doing some extensive work in the product.

6 Design Evolution

We went through numerous design sketches starting with basic sketches that we thought might be useful to display the details we were looking for, then after discussion started designing more detailed sketches getting to the heart of what we were looking for. Finally, after some pen and pencil sketches, we started adding colors and mocking up a visual design using Visio for the discussion with Alain, our TF for the project.

6.1 Initial Sketches

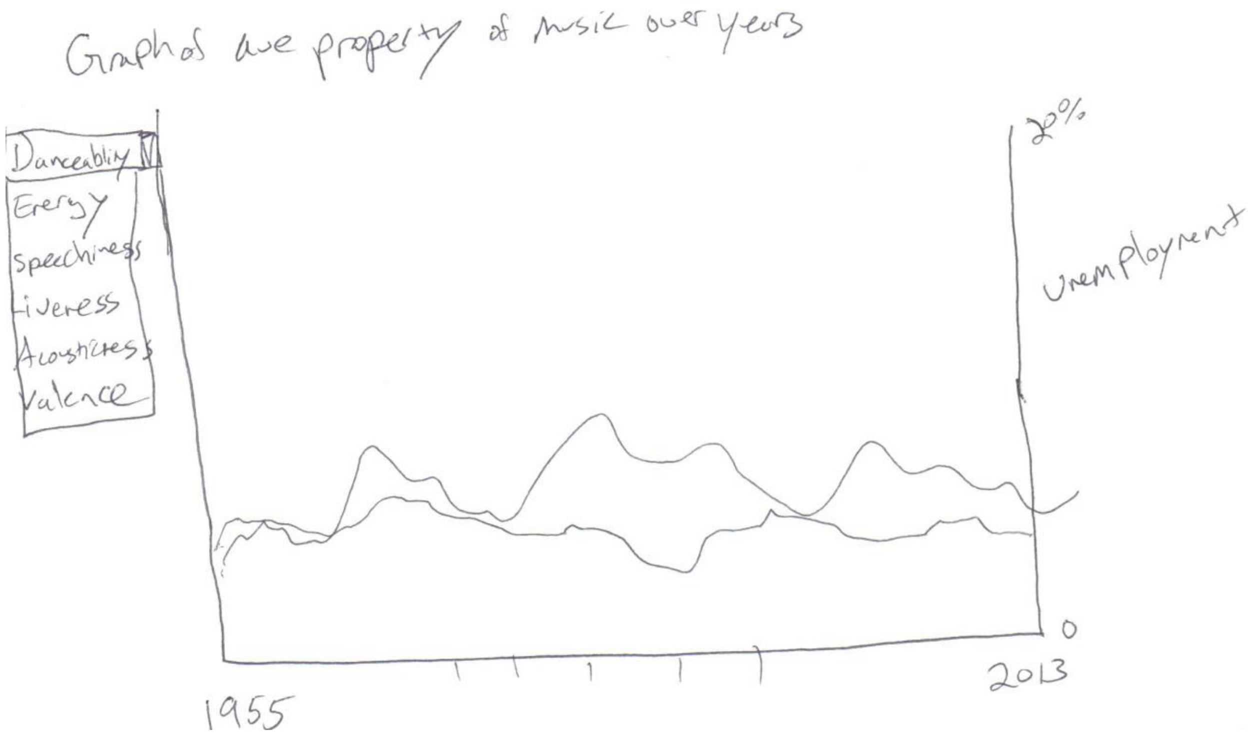
An idea for a song table, sortable by musical attribute, is shown here. We eventually incorporated this basic idea into the current iteration of the design as the Song Details panel.



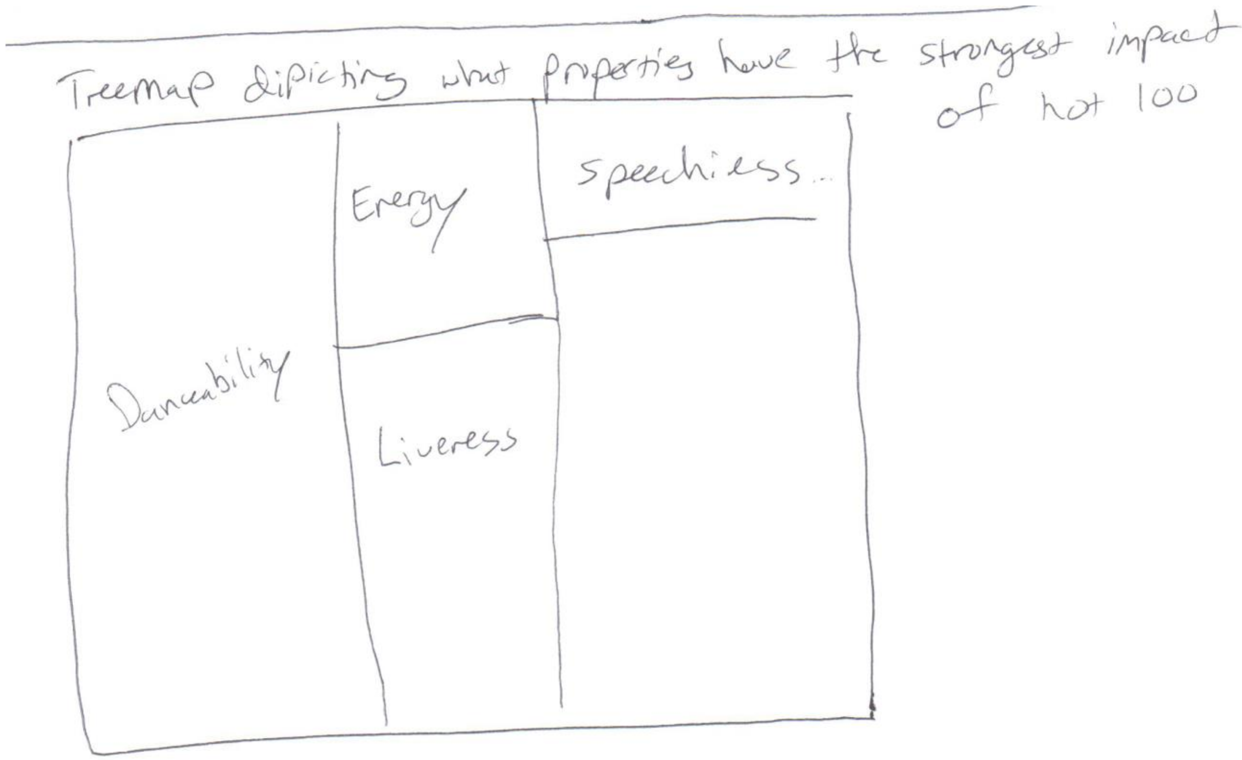
We liked the idea from the start of allowing the user to brush along a timeline to show data changing over time. The following is an initial sketch.

Each date is also represented by an artist and song, it would be interesting to see how a single artist might quickly change the attitude about a type of music.

We looked at having a graph in which a user could select musical attributes to see, with that axis represented on the left, then the right axis to represent economic data.

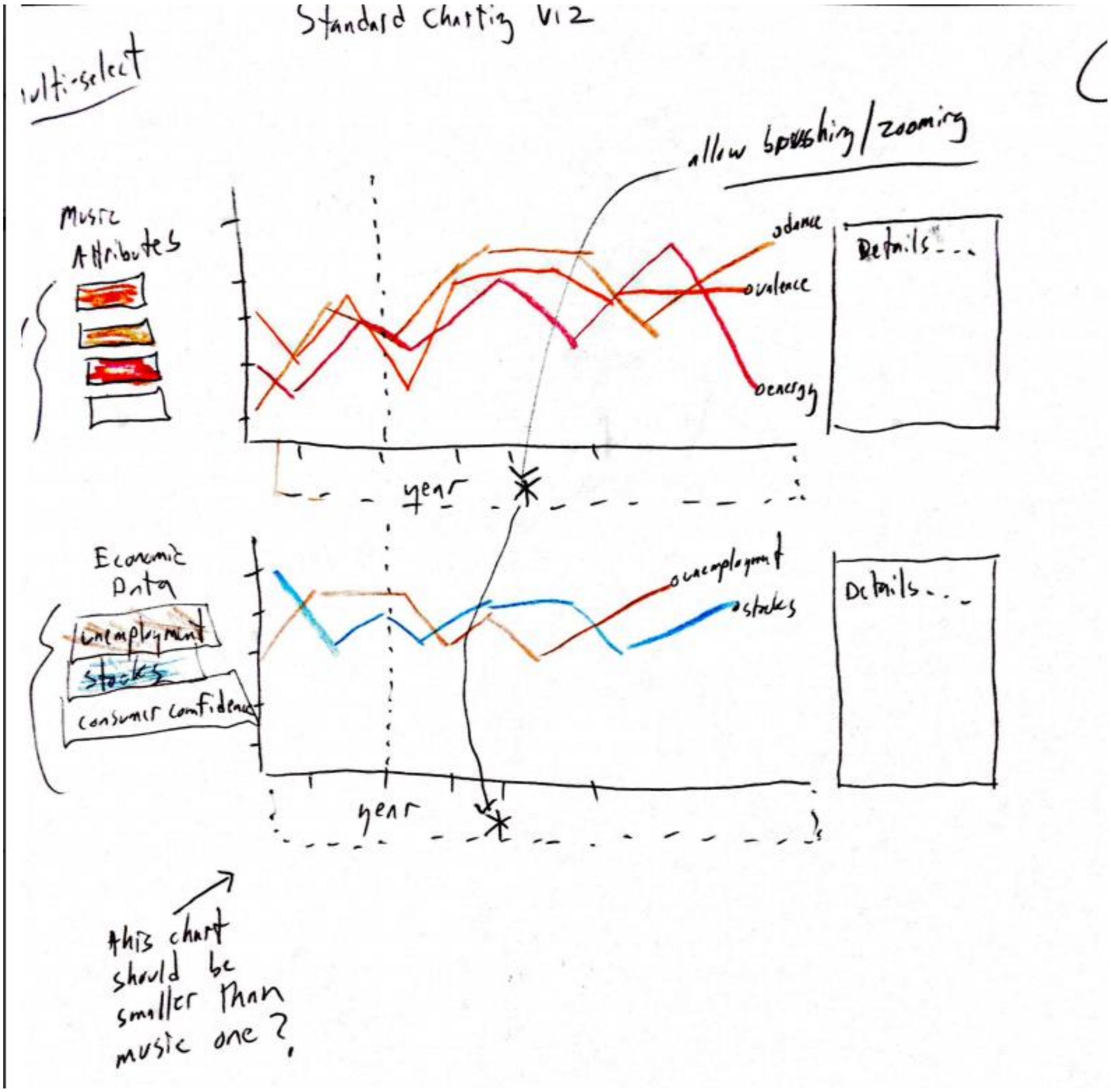


We also looked at a possible tree map idea that could possibly be an animation to see how attributes have morphed over time.



As we took stock of our initial ideas, we noticed (and the TF asked this question of us) that the diagrams were not connected at all. We wanted the user to have a more unified experience, without having graphs disappear or be disconnected.

Next we started sketching out some details of what a brushing graph might look like.

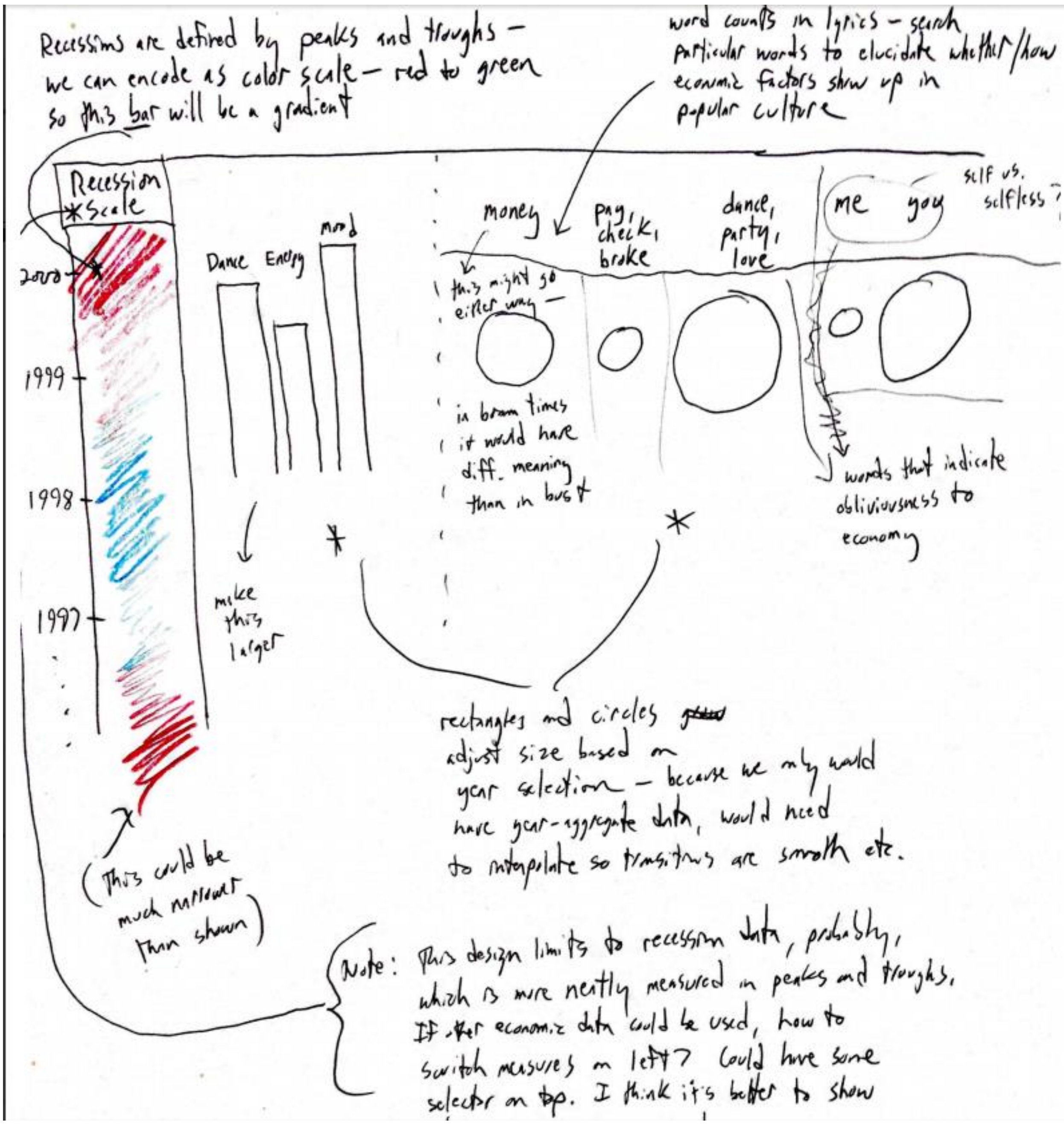


6.2 Adding the Lyric Data

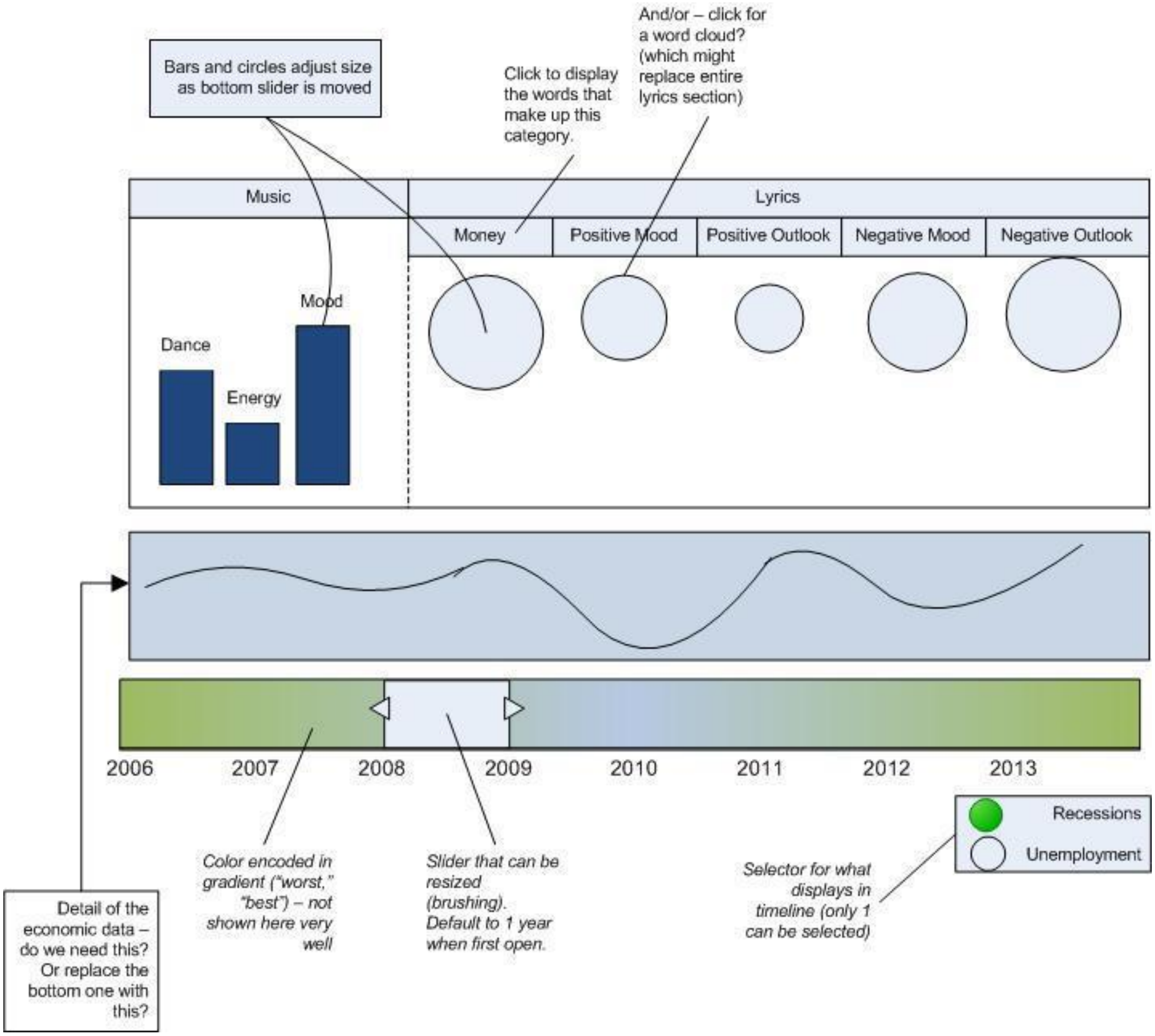
This sketch got most of what we wanted in one visualization, but there were a few issues. First, we weren't crazy about simply having some line graphs (purely from an aesthetic point of view). But mainly, this is the point at which we realized that maybe musical attributes would not be enough to explore.

The next sketch added lyric data to the mix. This began to really spoke to both of us in terms of the ideas that we'd really like to display to begin telling our story.

While word clouds seem to be all the rage these days, we didn't think that it would tell a story. We wanted a selective look at lyrics, so we created "buckets" that accomplish two things. First, doing this focuses the visualization and leads the viewer into the story. Second, it allows us to aggregate data in a way that can't be done with a word cloud.

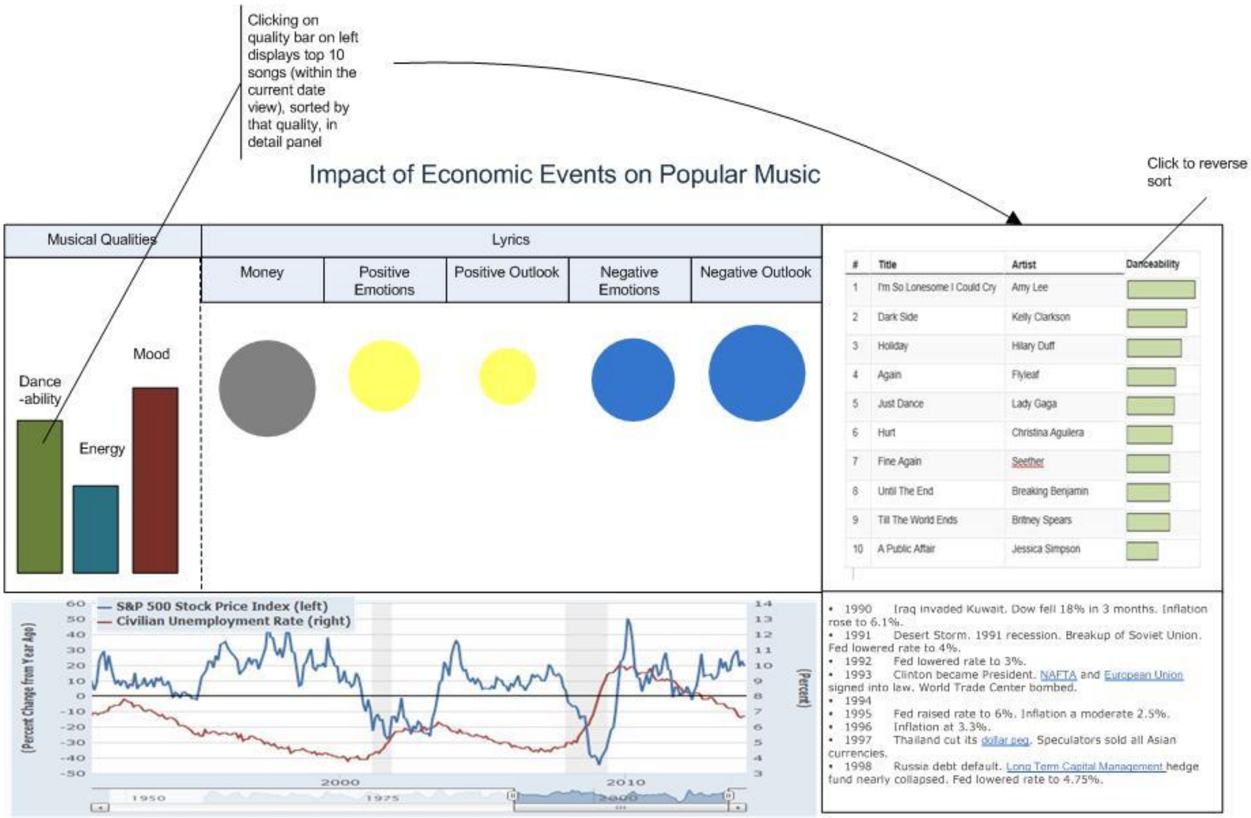


For this visualization, we realized that the time scale naturally belonged on the bottom, in a horizontal layout, because that's what people are accustomed to using. In addition, this particular version only looked at recession data, and used color to visualize it, which as we've learned in class, is not an effective way to encode numeric data.



6.3 Telling a Story

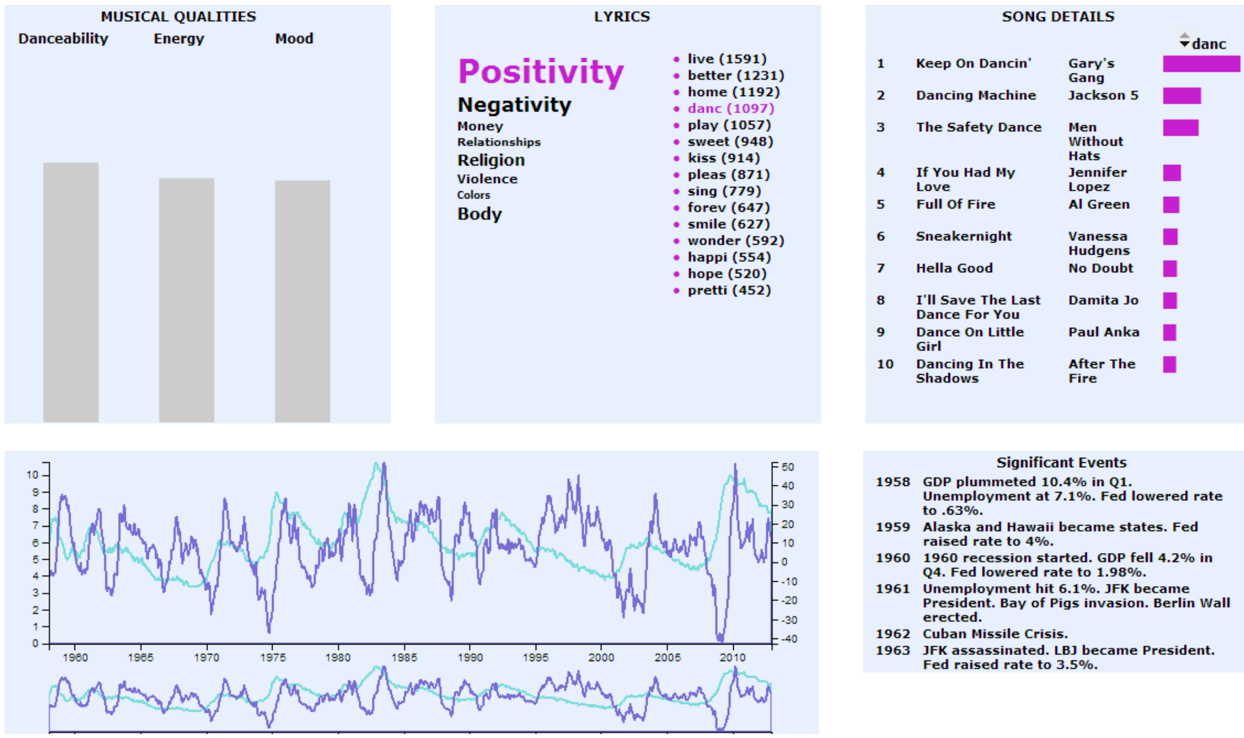
After our project discussion, we decided that we needed a little bit more "story" behind the economic data. We decided we would add another panel next to the graph that would show highlights (economic or historical) that might influence music. This panel provides a view based on what is selected for time scale, showing bullet points of the period's highlights.



6.4 The Circle Question

At a meeting with our TF, he noted that using circles for counts made it hard to make useful comparisons. Plus, it left lots of white space below.

We used his suggestion to switch from circles to text, and realigning from horizontal to vertical. We would indicate the counts using the font size of each word that represented a group.



6.5 Word Stems

The question arose, how would the user know which word stems we selected for each group? Knowing this would allow the user to better evaluate what he/she was seeing. We considered a tool tip or other overlay, but then we realized that we had all kinds of interesting data right in front of us – counts of individual word stems.

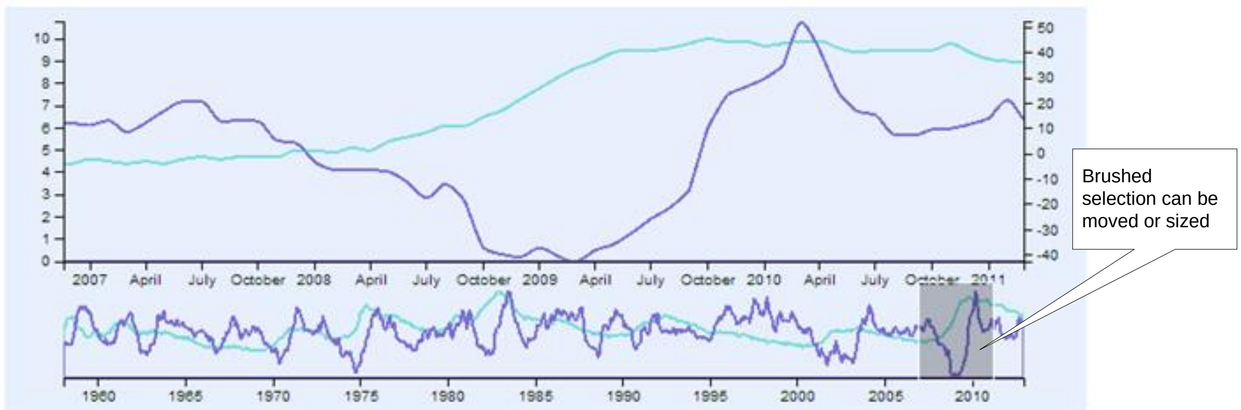
So we decided to add a list of the top 10-15 word stems for each group, which would be displayed when that group is selected. A count is shown next to each stem indicating the number of that stem within the current date range.

6.6 Word Stems and the Song Details

Immediately upon seeing that visualization, we noticed that we wanted to know which songs had the stem words in them. We knew that the final step would be to show, in the Song Details panel, a list of those songs. This would tie together the entire visualization by integrating both the Music Qualities panel and the Lyrics panel with the Song Details panel.

7 Implementation

We wanted our graph to be extremely easy to use almost instantly, so we have one core functionality in the brush graph at the bottom of the page that will update every aspect of the page as the user interacts with the brush. The small graph at the bottom with the currently visible brush is where most of the interactivity happens. When you drag the brush or resize it, the unemployment and S&P 500 data in the graph above it are updated, along with all three of the music-related panels above it.

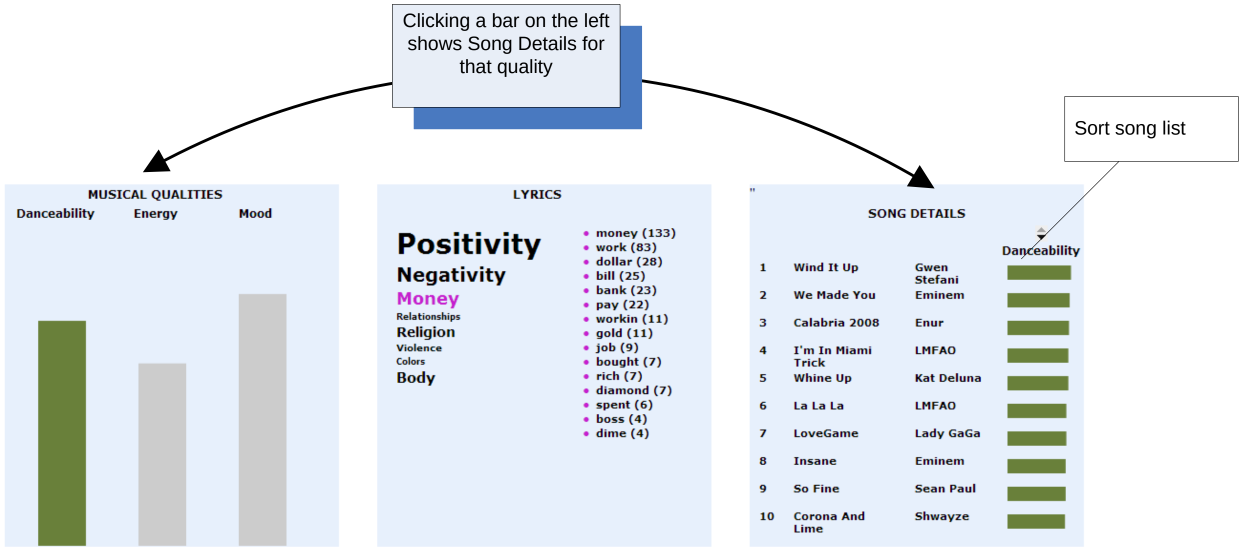


The Significant Events panel to the right provides additional context to the brushed section.

Significant Events

2004	**Fed started raising rates.**
2005	**Hurricane Katrina cost $250 billion in damage.**
2006	**Fed funds rate raised to 6.75%. Swine flu epidemic.**
2007	**Dow reached new high of 14,164.43. Inflation at 4.1%. Fed dropped rate 3 times, to 4.25%, to ease banking liquidity crisis. LIBOR rose to 5.6%.**
2008	**Stock market crash of 2008 led to global financial crisisand $350 billion spent on bank bailout bill. Fed lowered rate 7 times to 0%.**

While the brush determines the range of data in the visualization, other interactive features allow you to specify the details of that view. You can click any of the bars on the left to see song details for the top 10 songs, based on that musical quality.

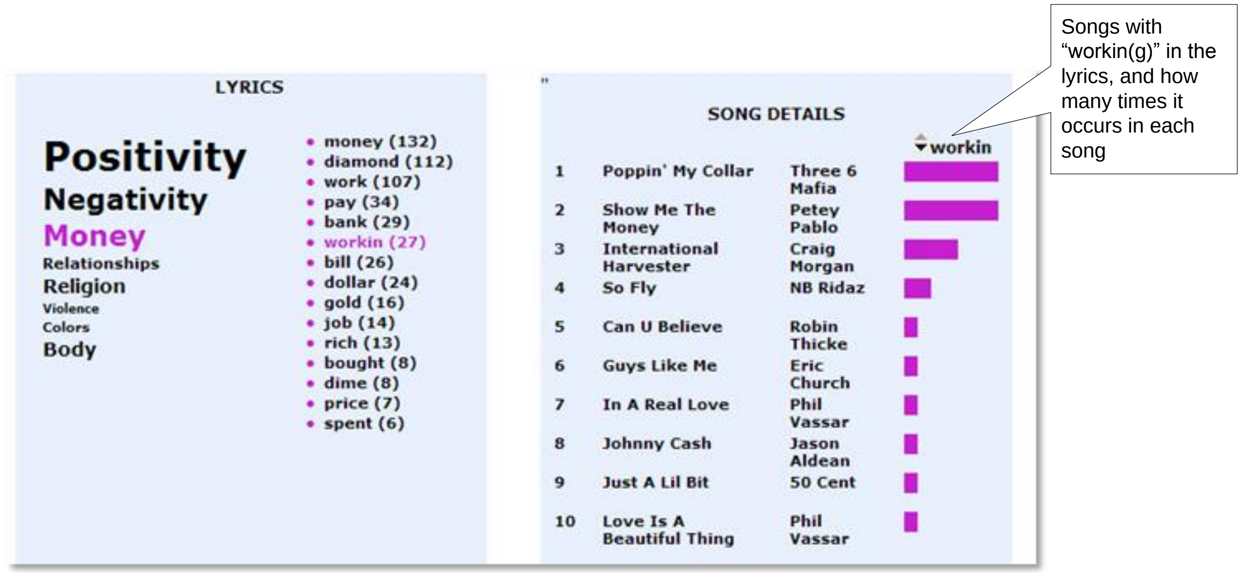


Notice that the colors of the bars on both sides match, providing a visual cue as to the connection.

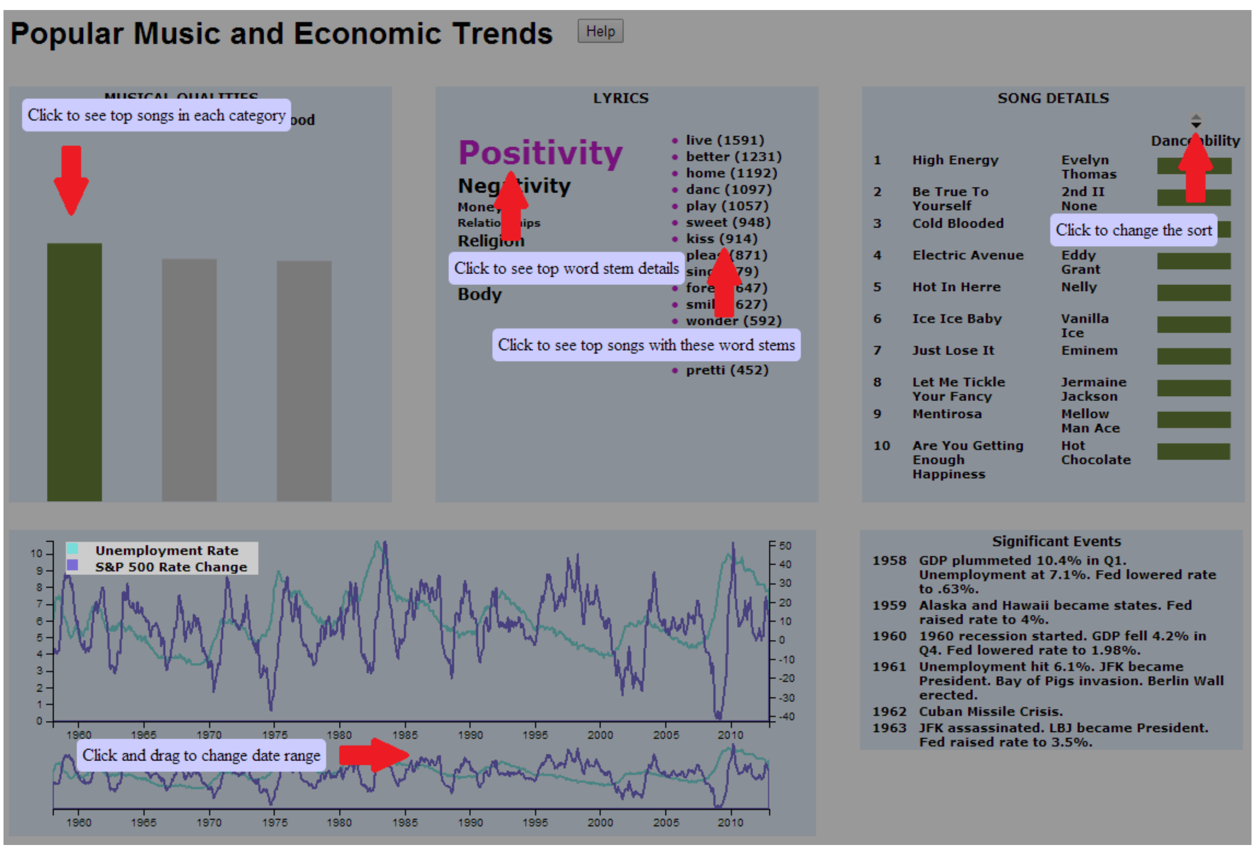
You can also explore lyric data using the middle panel. Click any of the grouping names ("Positivity," etc.) to see the list of word stems associated with that group, along with their counts. The counts represent the current brushed date range.

Notice that we use color to tie the selected group to the list (via the bullet points). Also note that the data is comprised of word stems (e.g., cri, alon, etc.).

You can select any of the word stems to see the list of songs that have that stem. In the Song Details panel, we repeat the color to tie that panel to the Lyrics panel.



Finally, you can click the Help button to show an overlay of tips about the interactive features. We used an overlay to keep the visualization as uncluttered as possible.



8 Evaluation

We had expected there to be a stronger correlation between economic trends and music, but the visualization shows this to be a somewhat ambiguous connection.

For the musical qualities, we notice that overall, the danceability and energy values have decreased since the beginning of the data set, while mood increased, hitting its highest values in recent years. The visualization calls into question the scoring system used by Echo Nest – not that it's somehow wrong, but that it doesn't provide enough information to become a story without additional data. It may be that further analysis that takes into account specific genres may yield more definitive results, or it may be that using measures more sophisticated and elaborate, such as Pandora's Music Genome Project, might be a more fruitful direction of study for non-lyrical music data.

The lyrics data, on the other hand, became the most interesting aspect of our visualization. We found that positive words are always the most mentioned words. The other words ebb and flow through time, however. For example, if you make the brush size small, there appears to be a decrease in money words when the economy starts sinking, then an increase when the economy starts to pick up again.

It's worth noting that it's possible that we're using the wrong economic data for this visualization. The Consumer Confidence Survey, put out by the Conference Board, is widely viewed as a good barometer of the national mood in the context of the economy. It tracks consumer sentiment, rather than measures that may or may not have a direct impact on the people who actually make popular music popular. This measure was unfortunately not freely available.

The visualization also shows how the lyric counts can be skewed by a single song. There are many cases where the top song has a count of a particular word that's many times higher than the second-place word. This leads us to wonder if there might not be a more statistically fair way to look at the data that eliminates these outliers.