Visualizing Economic Effects on Popular Music (1958-2013)

David Price and Glen Barger

CS 171: Spring 2014

1. Overview and Motivation

We really wanted to do something fun for this project, preferably something in the entertainment realm. After looking around at some data sources we ultimately landed on 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

We've both been inspired by the type of technology that drives Pandora, unfortunately the Music Genome Project is owned by Pandora and that data isn't publicly available. We discovered a site called the Echo Nest, which provided public data for musical attributes, and we ultimately used that data as the foundational work for our project before discovering that we could access lyrical data as well.

3. Questions

The initial question we really wanted to to answer was how do musical attributes and the economy shift over time. We wanted to see if certain musical attributes like danceability, energy, and mood are similar across time when the economy is weaker or stronger. We came up with several design sketches based around these ideas, but then our questions grew to include possible lyrical attributes. We wanted to know if lyrics were more positive or negative, not just the attributes of the instrumentation. This adds another dimension to the dataset we decided we wanted to explore more after we discovered that we might be able to obtain that data.

4. Data

To start our data collection process we obtained a very large initial dataset from the Whitburn Project. The Whitburn Project is a project

made up of a number of music enthusiasts to try and document chart data on popular music starting in the 1890s. You can read more about the [project here](#).

The next step of our data gathering involved collecting the musical attributes of the songs, for this we utilized a website called [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. To retrieve the echo nest attributes we created an HTML document that utilizes d3 and javascript to loop through all the rows in a csv file and call the Echo Nest API once every 4 seconds for details of a single song and when complete output 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.

The final step for our data collection is gathering lyrical data from a project that processed several hundred thousand records from musicXmatch.com and processed the lyrics into word stems. This data can be mined to get a count of words for a specific song but avoids copyright issues because the full lyrics aren't provided, just the summary data. 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.

Once we had the list of songs with their Ids that could be matched up with the musicXmatch.com data, we created a json data set that could be used to quickly pull lyrical data by providing the track Id.

5. Exploratory Data Analysis

We did almost no exploratory data analysis because the data gathering and processing took much longer than we anticipated. We quickly built tools to gather the data, but the API limitations on making the requests put a huge limit on our initial gathering as well as the need to run the

gathering again when we discovered that we wanted more data from The Echo Nest api. Once that data was together, we quickly put together an initial visualization to see what our data looked like, which was in time for the first milestone. From the results of our project, the more interesting data turned out to be lyrical in nature as you can see different attributes increase and decrease through time. You can see religion becoming more prominent at certain times, along with violence and negativity during war times and after the assassination of President Kennedy.

6. Design Evolution

We went through numerous design sketches starting with really 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 on the computer for the discussion with Alain, our TF for the project.

Idea for a sortable table design.

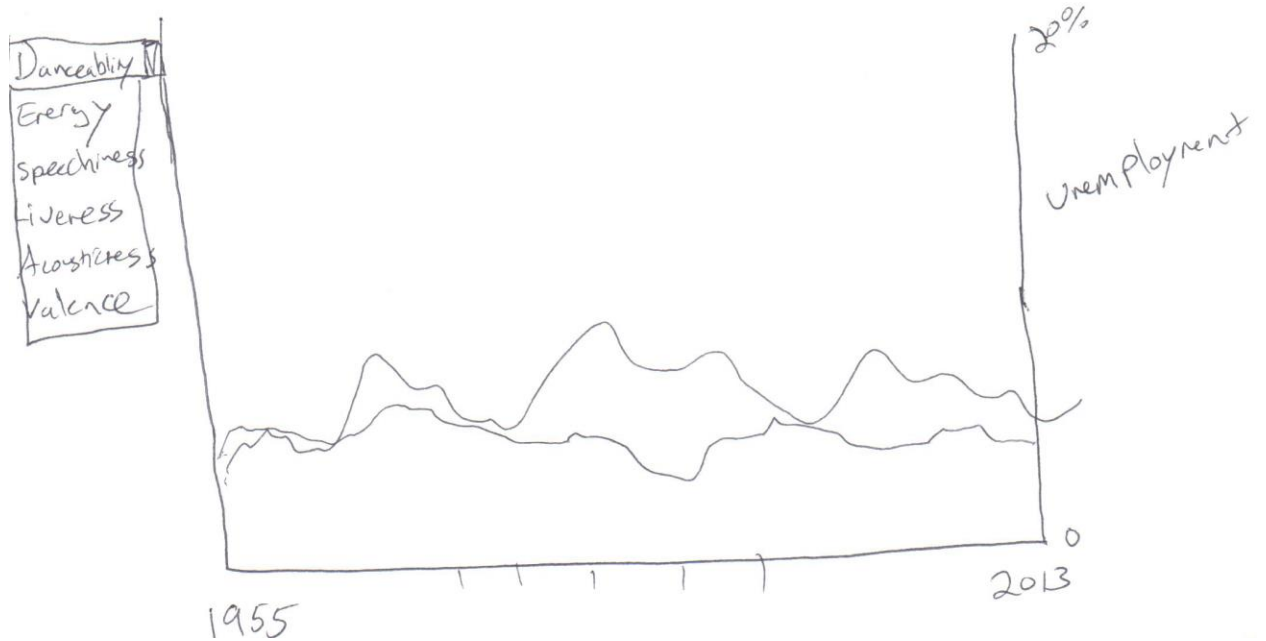Idea for a brushing graph where the user could slide along to see how musical attributes have changed over time.

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.

A graph where a user could select musical attributes to see, with that axis represented on the left, then the right axis to represent economic data.

Graph of ave property of music over years

Danceablity
Energy
speechiness
Liveness
Acoustitcess
Valence

20%

Unemployment

0

2013

1955

A possible tree map idea that could possibly be an animation to see how attributes have morphed over time.

Treemap dipicting what properties have the strongest impact of hot 100

Energy

Speechiess...

Danceability

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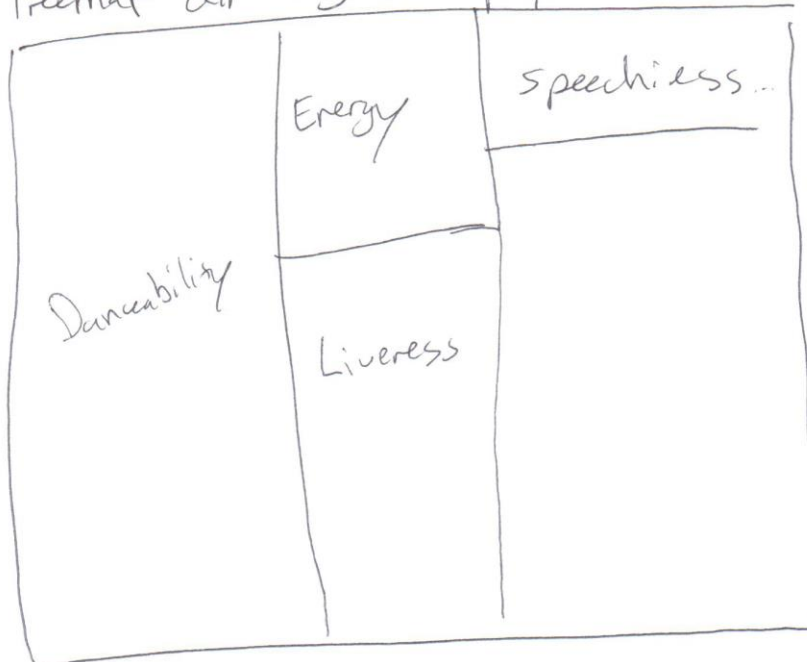
Next we started sketching out some details of what a brushing graph might look like.

Then we finally got down into a focused visualization that really spoke to both of us in terms of the ideas that we'd really like to display to begin telling our story.
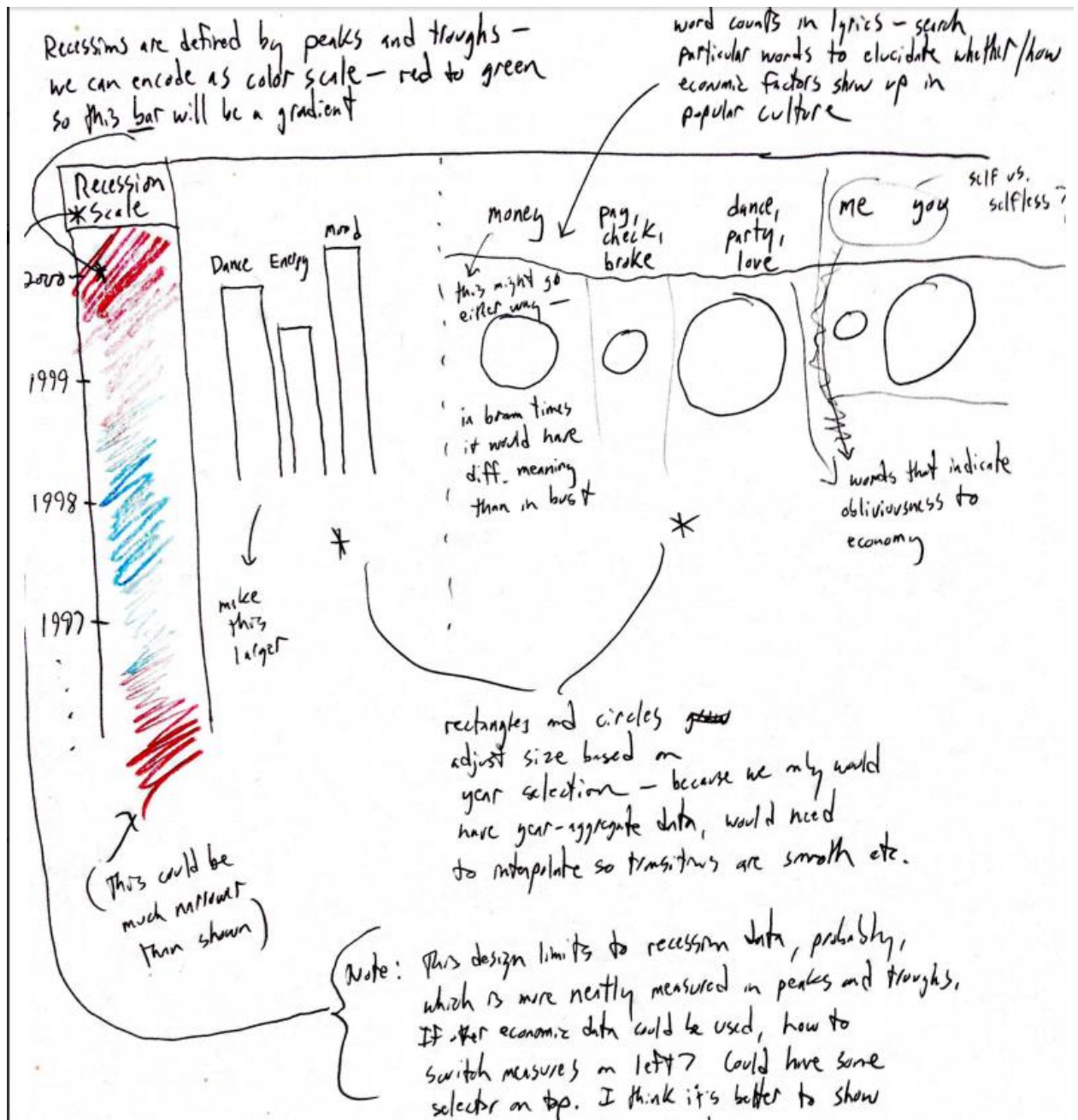
After discussing this visualization we decided on some changes we wanted to make before coming up with a design for our first discussion with Alain to prepare for our fp-1 milestone.

Bars and circles adjust size as bottom slider is moved

And/or – click for a word cloud? (which might replace entire lyrics section)

Click to display the words that make up this category.

Music		Lyrics				
		Money	Positive Mood	Positive Outlook	Negative Mood	Negative Outlook

Mood

Dance

Energy

2006   2007   2008   2009   2010   2011   2012   2013

Recessions

Unemployment

Detail of the economic data – do we need this? Or replace the bottom one with this?

Color encoded in gradient ("worst," "best") – not shown here very well

Slider that can be resized (brushing). Default to 1 year when first open.

Selector for what displays in timeline (only 1 can be selected)

After our project discussion we created a final project visualization that we'd like to try to create for our project with the added storytelling elements.

Clicking on quality bar on left displays top 10 songs (within the current date view), sorted by that quality, in detail panel

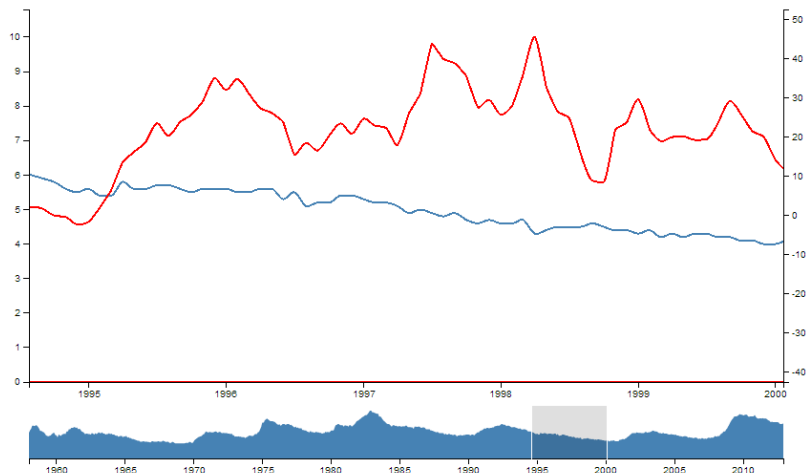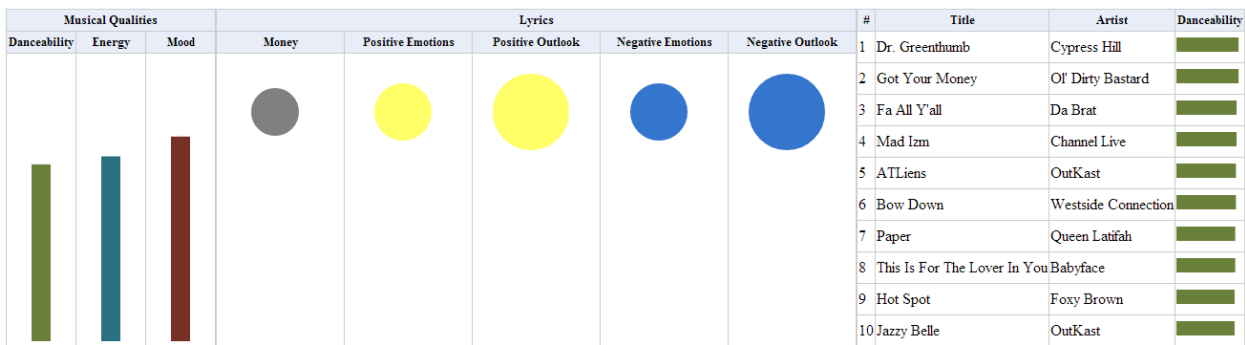Impact of Economic Events on Popular Music

Click to reverse sort
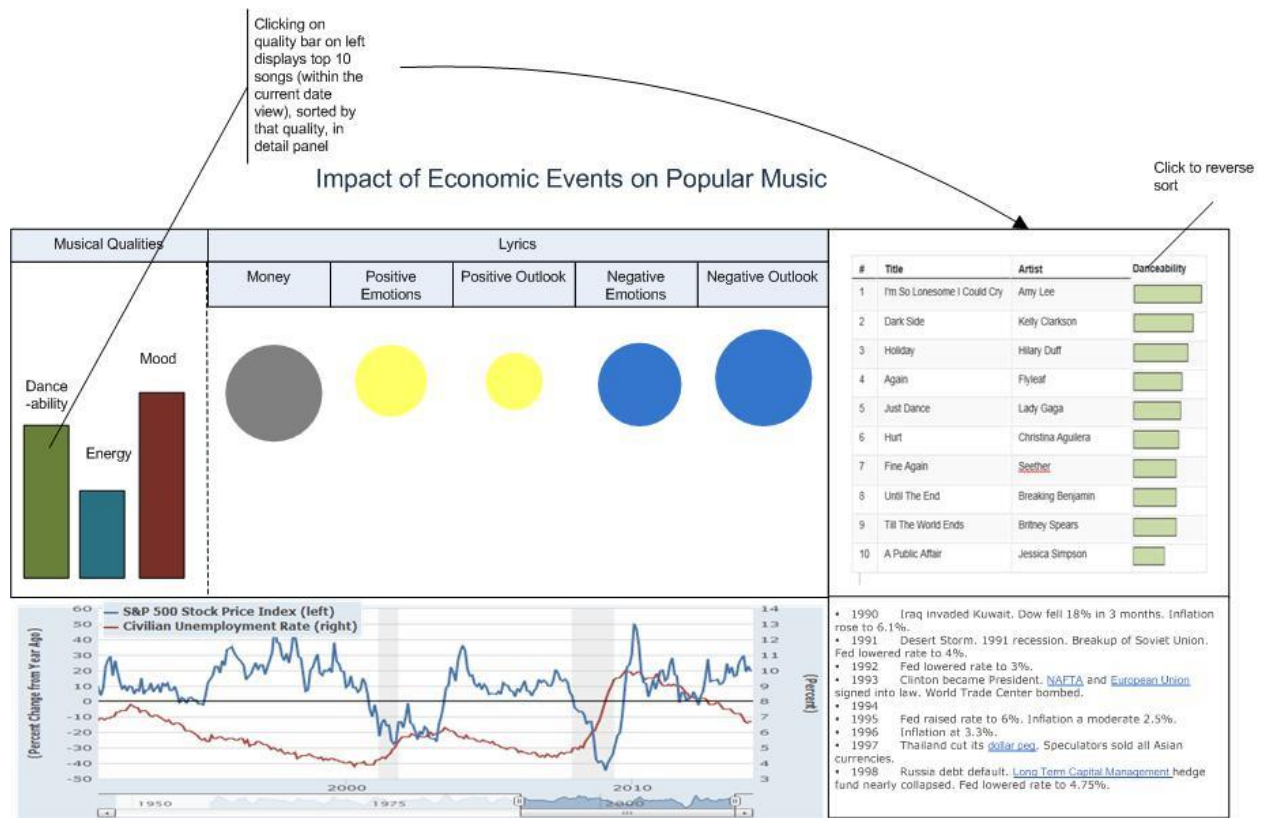
7. Implementation

The implementation of our project happened by first building the brush graphs from economic data, which would be the timeline that all other data would be based on.
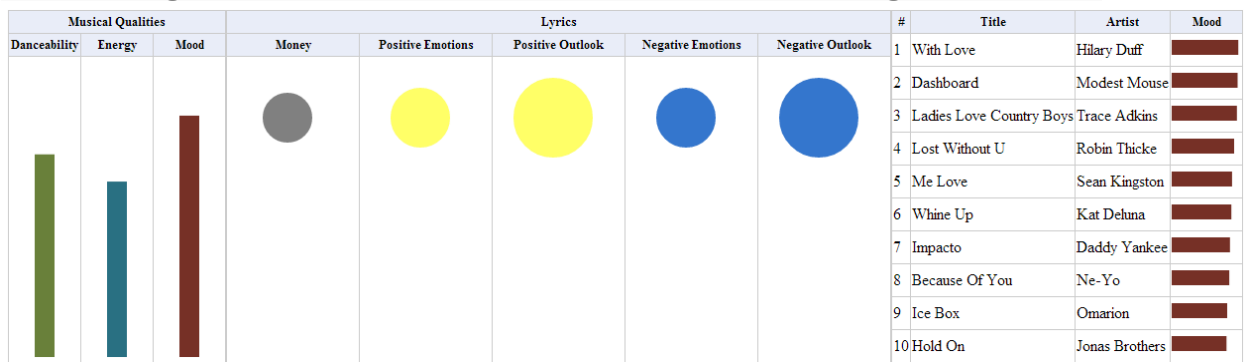
We wanted our graph to be extremely easy to use almost instantly, so we have one core functionality in our brush that will update every aspect of the page as the user interacts with the brush, and we added the more minor functionality of allowing a user to click the musical quality they're most interested in to update the far right side chart with the songs in the brush range that are the highest rated matches for those songs. These screenshots are from an early implementation so the visualization isn't completely interactive yet. The lyrical bucket data is static at this time. The following is an overview of the visualization in its current functional form.

The small graph at the bottom with the currently visible brush is where most of the interactivity happens. When a user drags that brush, the unemployment and S&P 500 data in the graph above it are updated along

with the musical qualities bars on the left and the top 10 songs that match the current musical quality on the right.

| Musical Qualities			Lyrics					#	Title	Artist	Danceability
Danceability	Energy	Mood	Money	Positive Emotions	Positive Outlook	Negative Emotions	Negative Outlook	1	Dr. Greenthumb	Cypress Hill	
								2	Got Your Money	Ol' Dirty Bastard	
								3	Fa All Y'all	Da Brat	
								4	Mad Izm	Channel Live	
								5	ATLiens	OutKast	
								6	Bow Down	Westside Connection	
								7	Paper	Queen Latifah	
								8	This Is For The Lover In You	Babyface	
								9	Hot Spot	Foxy Brown	
								10	Jazzy Belle	OutKast	

Here's an example of how the top section has changed after zooming the brush into the middle of 2007, then clicking on the mood bar. The songs with the highest mood are now listed in the chart on the right hand side.

| Musical Qualities			Lyrics					#	Title	Artist	Mood
Danceability	Energy	Mood	Money	Positive Emotions	Positive Outlook	Negative Emotions	Negative Outlook	1	With Love	Hilary Duff	
								2	Dashboard	Modest Mouse	
								3	Ladies Love Country Boys	Trace Adkins	
								4	Lost Without U	Robin Thicke	
								5	Me Love	Sean Kingston	
								6	Whine Up	Kat Deluna	
								7	Impacto	Daddy Yankee	
								8	Because Of You	Ne-Yo	
								9	Ice Box	Omarion	
								10	Hold On	Jonas Brothers	

Another piece of the interactivity we'll be adding for the final project launch is the automatic resizing of the circles, as well as adding storytelling elements to the right hand side below the top songs chart and to the right hand side of the economic chart.

We finally settled on a list of lyrical buckets with the bucket names changing sizes while brushing and displaying a list of the top words within those buckets. Those words can then be clicked on for a detailed list of songs that use those words.
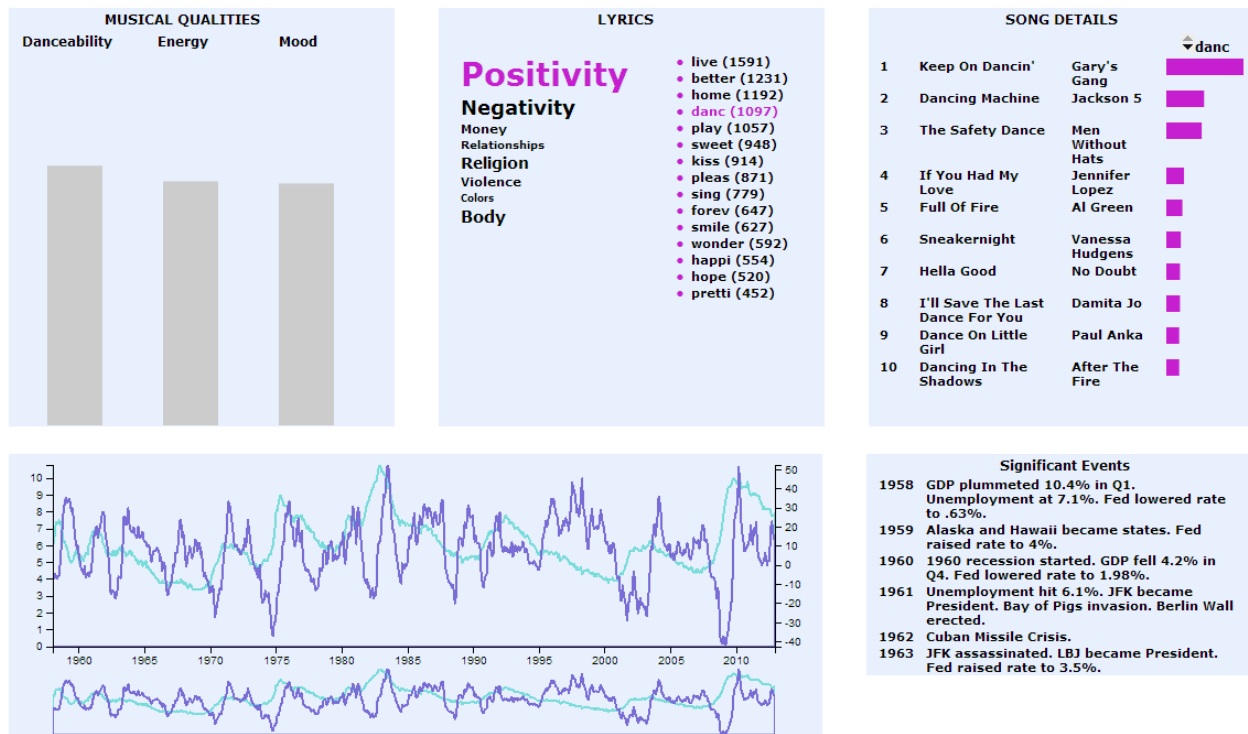
The final step in our visualization evolution was to add a help option that overlays a few minor details about the interactive portions of the visualization.
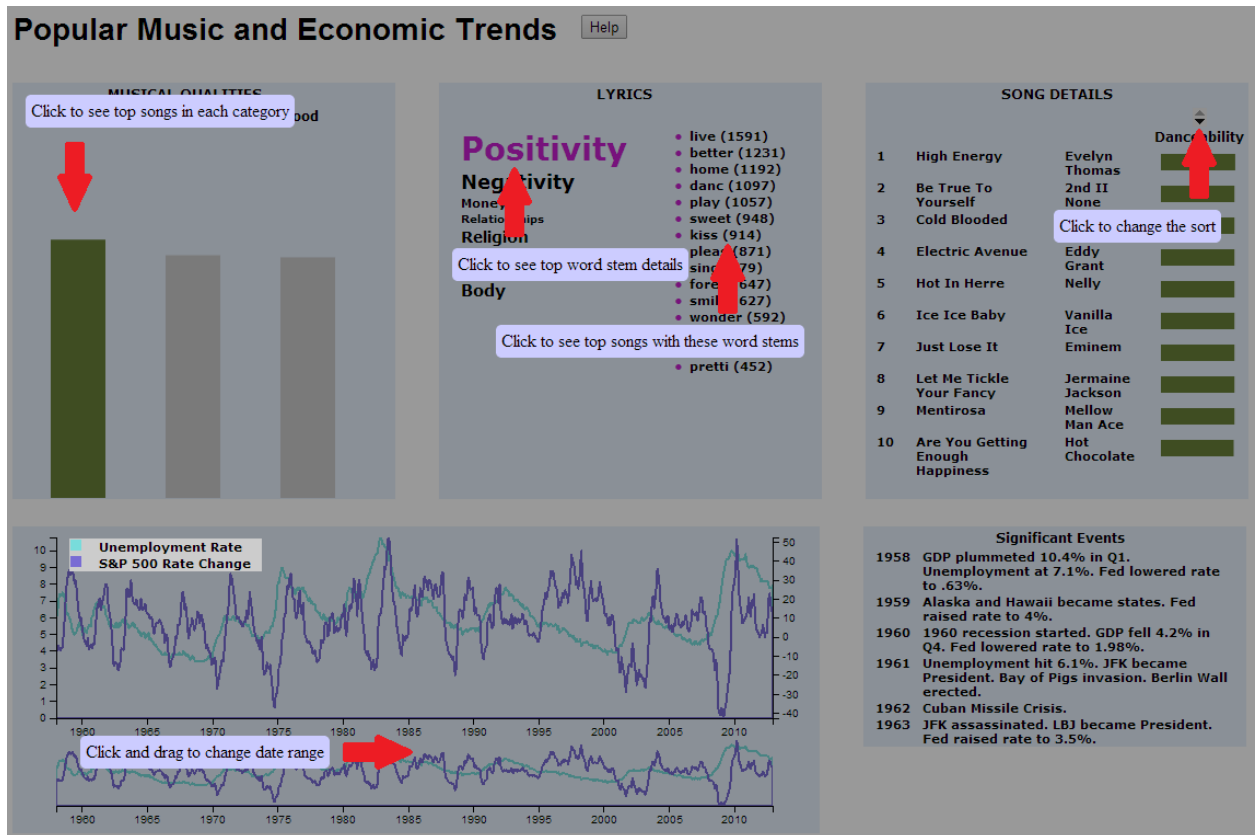
8. Evaluation

We didn't end up with the results we were hoping for with the musical attributes. Overall, the danceability and energy have reduced since the beginning of our data set, while mood increased to the highest values of the dataset in recent years. The lyrical data ended up being the most interesting aspect of our visualization. Positive words are always the most mentioned words, but the other words ebb and flow through time. When making the brush size small, there appears to actually be a decrease in money words when the economy starts sinking, then an increase when the economy starts picking back up.