

Visualizing Economic Effects on Popular Music (1958-2013)

David Price / Glen Barger

CS171: Spring 2014

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 1958-2013, 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

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

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 turned 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.

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.

5 Exploratory Data Analysis

6 Design Evolution

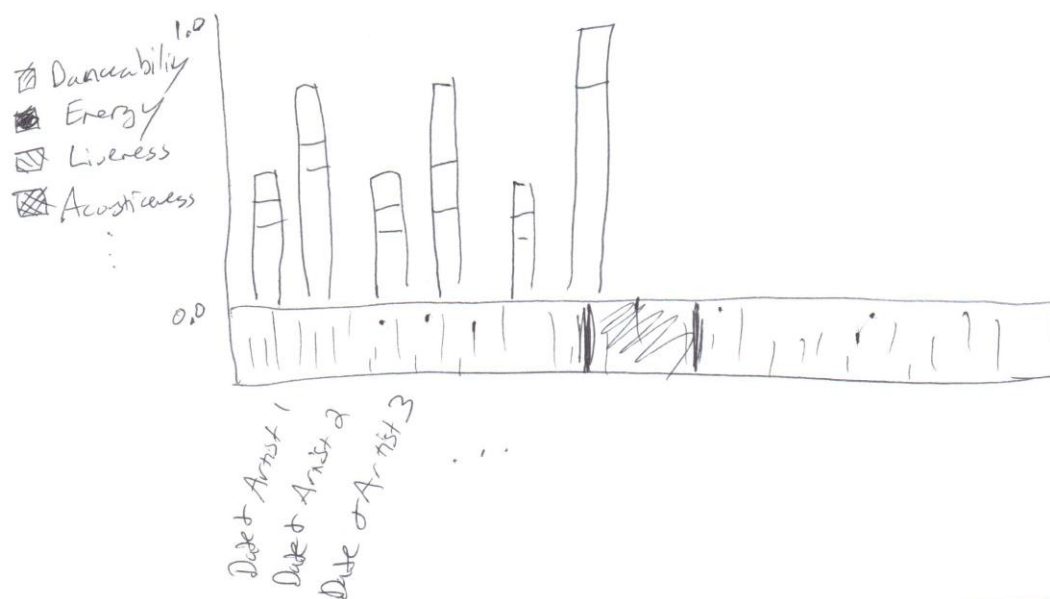
We went through numerous design sketches starting with basic sketches that we thought might be useful to display the details we were looking for, then after discussion started designing more detailed sketches getting to the heart of what we were looking for. Finally, after some pen and pencil sketches,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

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.

Dance	Energy	▼	Artist - Song	
1.0	1.0		Artist - Song	
1.0	1.0		Artist - Song	
0.99	1.0		Artist - Song	
0.90	1.0		Artist - Song	
0.80	1.0		Artist - Song	

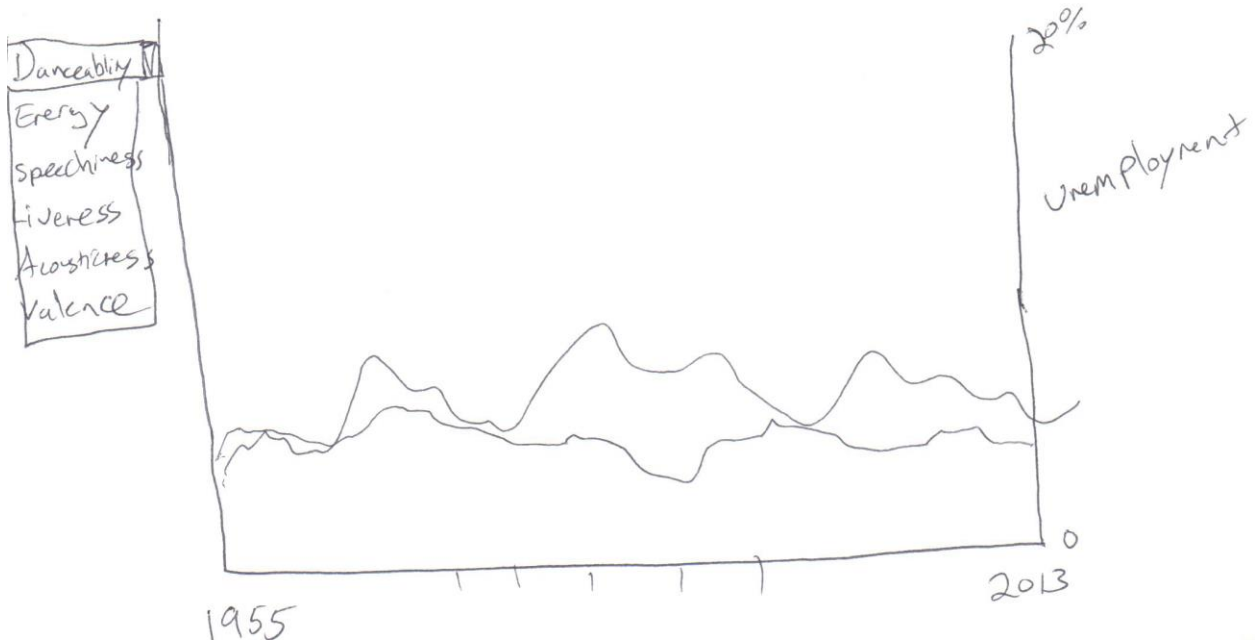
We liked the idea from the start of allowing the user to brush along a timeline to show data changing over time. The following is an initial sketch.

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.

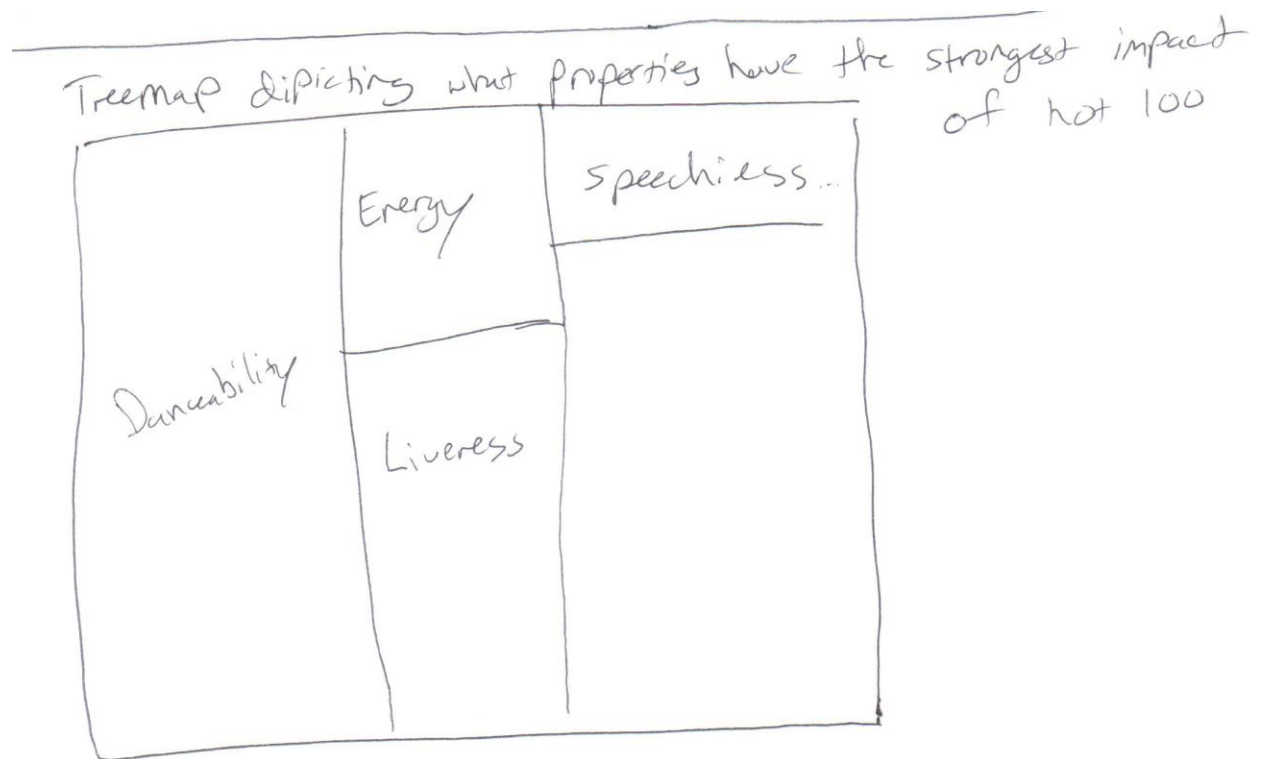


A graph where a user could select musical attributes to see, with that axis represented on the left, then the right axis to represent economic data.

Graphs are property of music over years

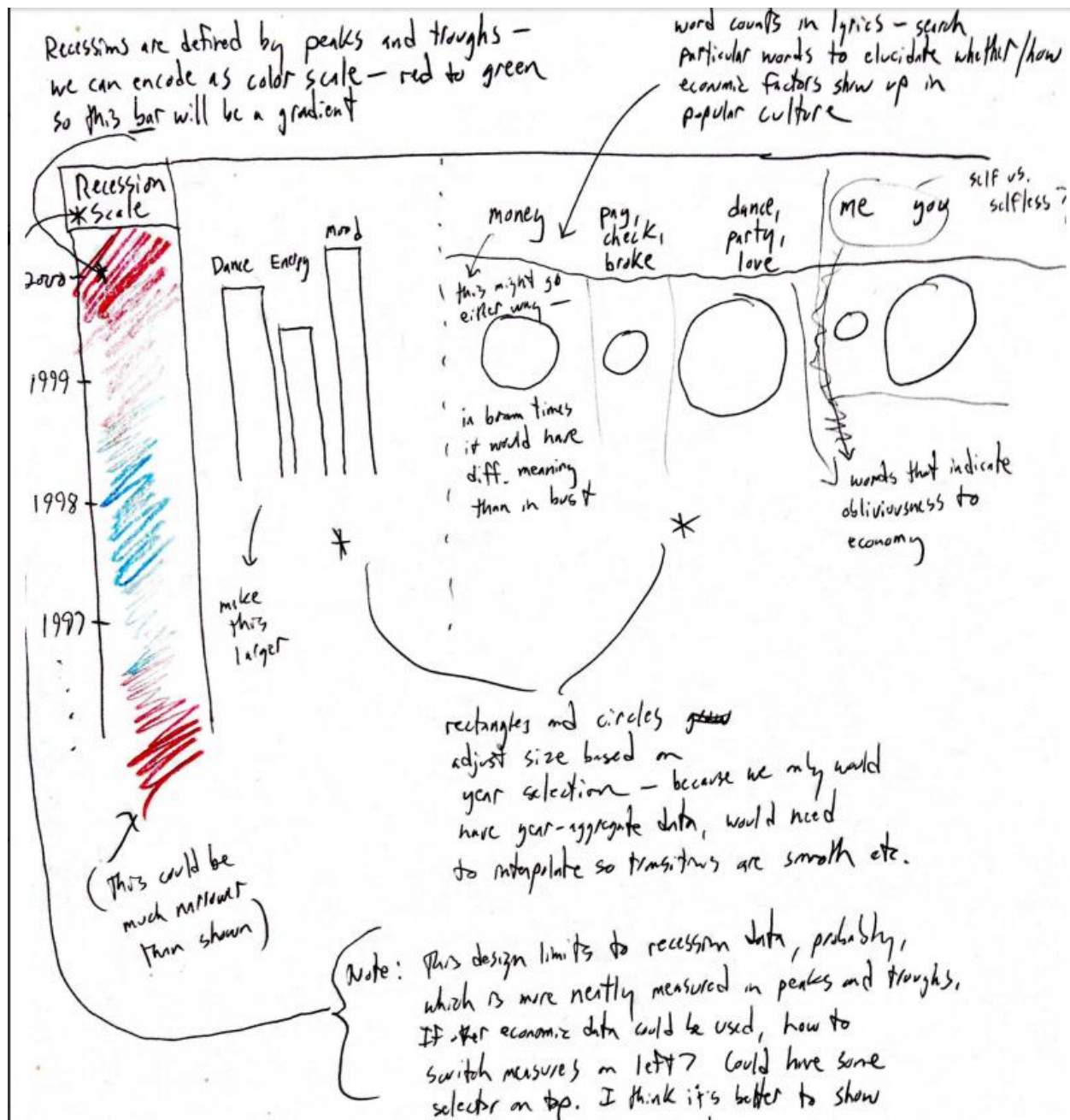


A possible tree map idea that could possibly be an animation to see how attributes have morphed over time.

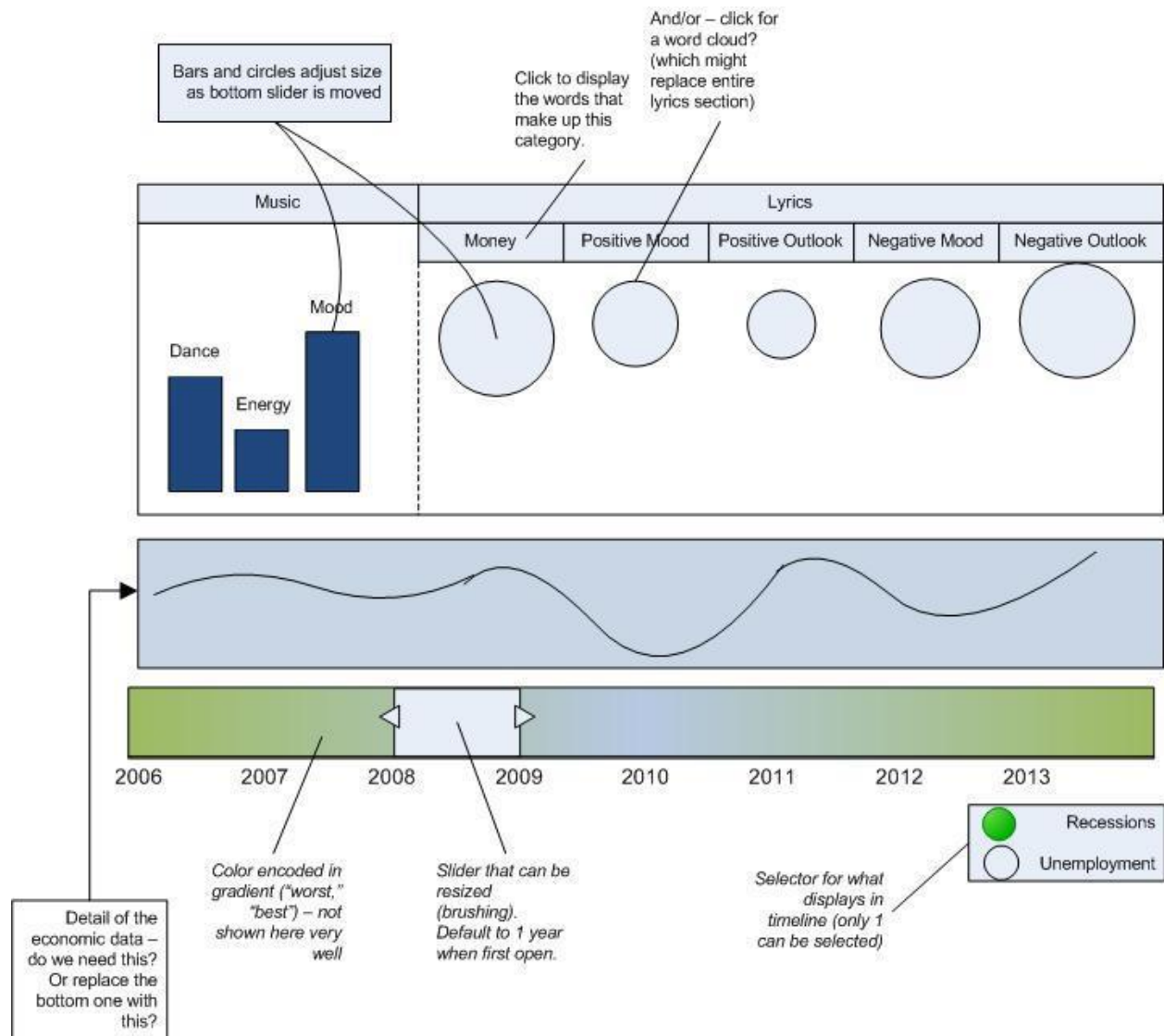


As we took stock of our initial ideas, we noticed (and the TF asked this question of us) that the diagrams were not connected at all. We wanted the user to have a more unified experience, without having graphs disappear or be disconnected.

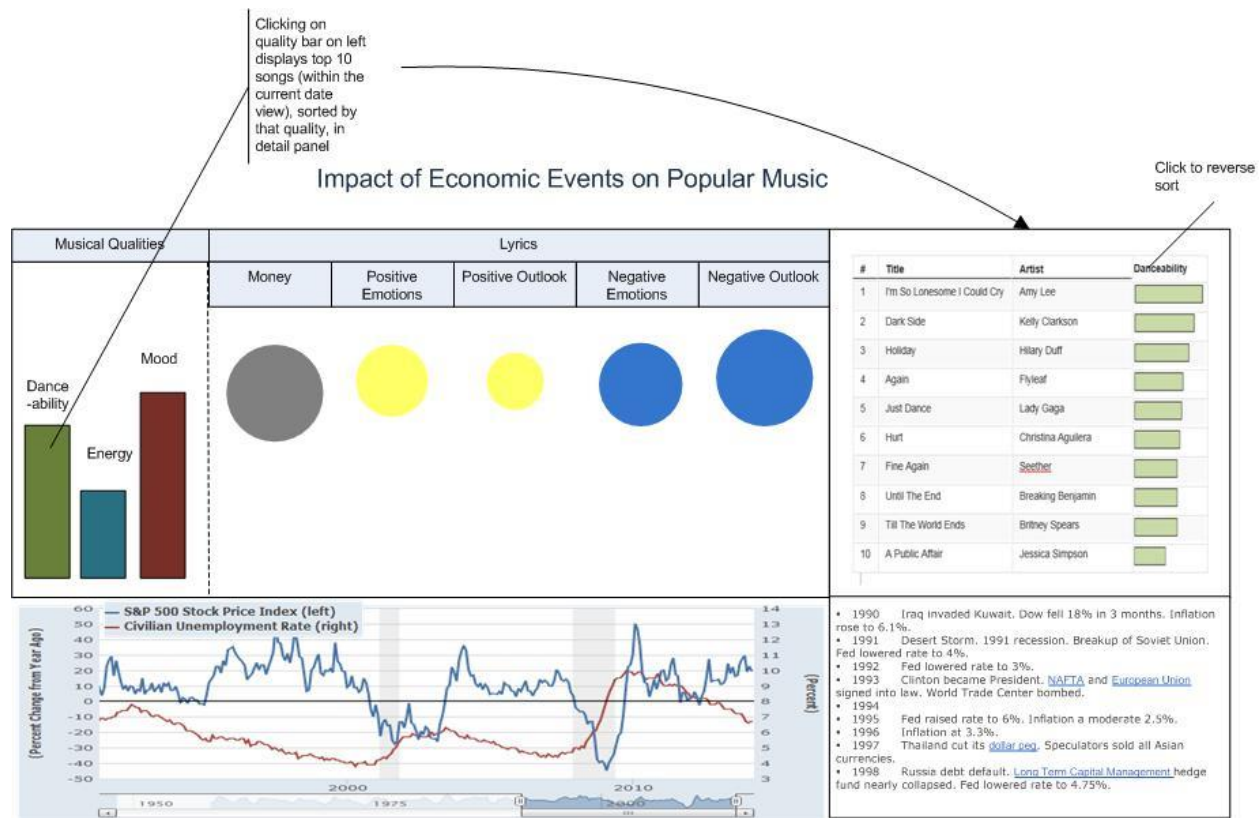
Next we started sketching out some details of what a brushing graph might look like.



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.



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.

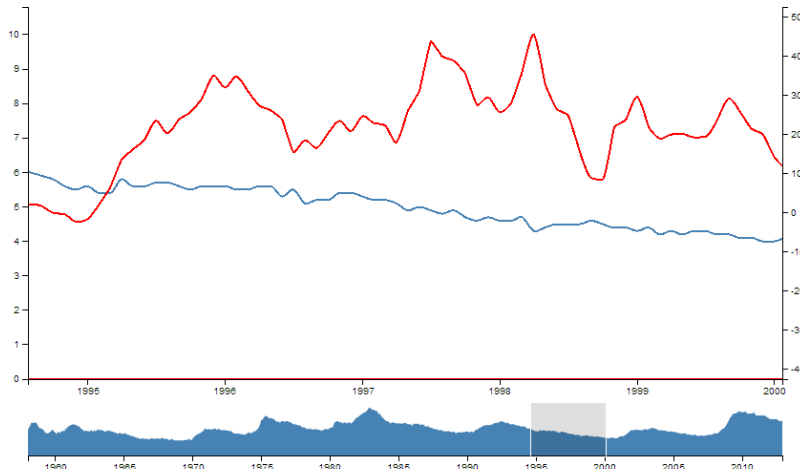
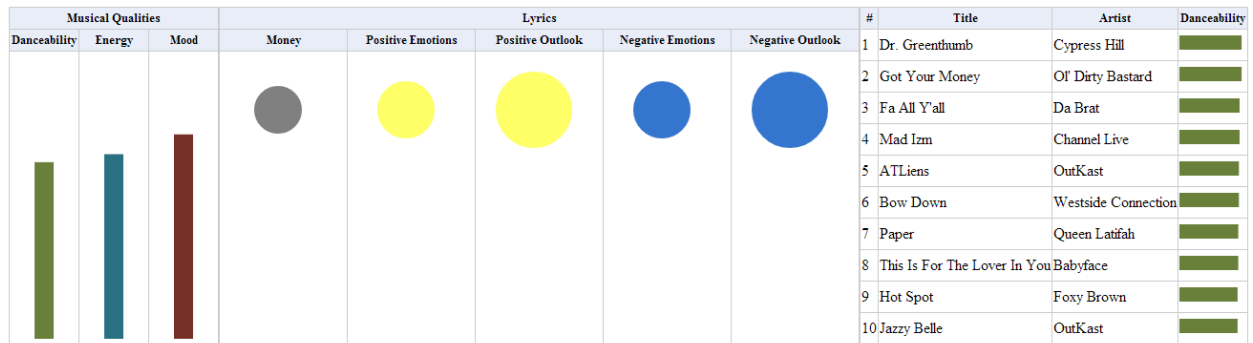


7 Implementation

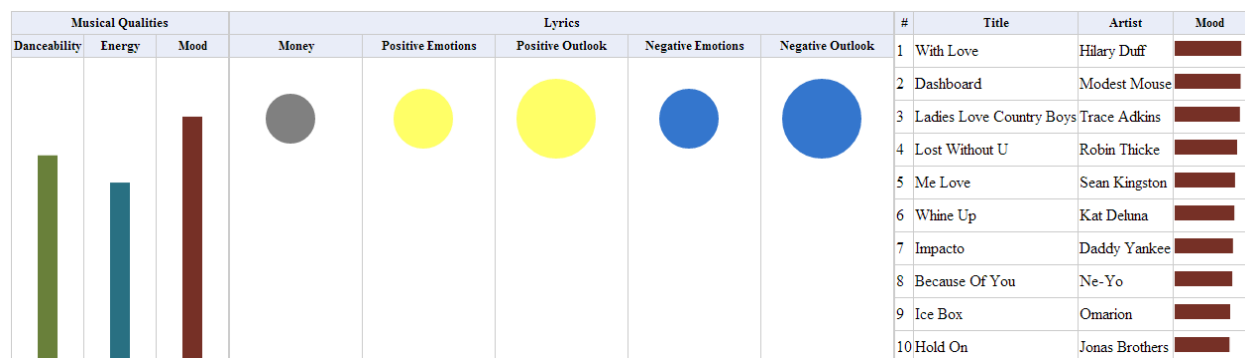
The implementation of our project happened by first building the brush graphs from economic data, which would be the timeline that all other data would be based on.

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.



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.



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.

