

OVERVIEW AND MOTIVATION

We plan to develop an interactive visualization of the sales data for key music formats (CD, cassette, vinyl, download single, paid subscriptions, etc.) in the United States over the years. We think it would be interesting to explore how music consumption has changed from generation to generation and how evolving music format technologies have influenced that. Lastly, with many business articles reporting that the music industry is in decline, it would also be interesting to verify if this is the case or not based on trends and patterns in the dataset.

PROJECT OBJECTIVES

The primary questions we are trying to answer with our visualization are:

- How has the popularity of various music formats changed throughout the years?
- Have overall sales in music been decreasing gradually over time?
- Have new and evolving music format technologies hurt or helped music sales?

What we would like to learn is if any trends or patterns exist in the music sales dataset. In particular, the things we would like to accomplish include:

- ranking the various music formats to find out which ones dominated during different time intervals
- finding the peaks and valleys in music sales over time
- discovering when formats increased and decreased in popularity
- finding the greatest changes in sales throughout each year
- discovering any resurgences of older music formats (e.g. vinyl, cassette)

The benefits include:

- being able to compare sales between various music formats over time
- being able to compare sales in dollar value accounting and not accounting for inflation
- being able to compare sales between music medium types (physical, digital, streaming)
- being able to compare music sales between different time intervals

DATA

The United States music sales dataset comes from the Recording Industry Association of America (RIAA) Shipment Database. The database is accessible through the following login page: https://www.riaa.com/keystatistics.php?content_selector=riaa-shipment-database-log-in. The RIAA has provided us access to their database for academic use. The database provides an option to export the dataset as an Excel spreadsheet, which is how we are collecting the data.

DATA PROCESSING

We do not expect to do substantial data cleanup. We plan to save the Excel spreadsheet exported from the RIAA database as CSV files and process the data using the D3 CSV API. The dataset includes year-end sales statistics from 1973 to 2014 for the recorded music industry in the United States. The statistics comprise of both the number of units and dollar value sold (in millions) for key music formats (CD, cassette, vinyl, download single, paid subscriptions, etc.).

The quantities we plan to derive from the number of units and dollar value data include:

- the cumulative amounts over a specified time interval
- the aggregate amounts of physical, digital, and streaming medium types
- the changes in amounts over each year
- the differences in amounts between two specified time intervals

We plan to implement data processing by using the D3 CSV API to parse the CSV files and create three arrays of JavaScript object for each sales metric type (units, dollars, dollars adjusted for inflation) with each object containing the following fields:

- format (CD, cassette, vinyl, download single, paid subscriptions, etc.)
- medium (physical, digital, streaming)
- year - year of the sales metric value
- value - value of the sales metric

VISUALIZATION

The visualization will comprise of multiple coordinated charts that link together so that when a user interacts with one of the views, the others dynamically update through animated transitions. We will display the music sales data using line, stacked area, and bar charts. Please refer to the sketches below.

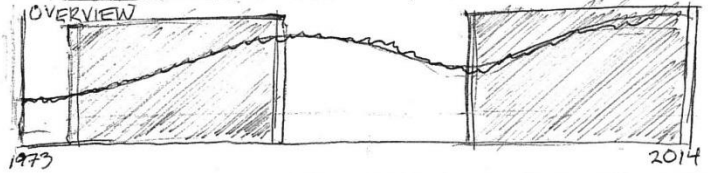
On the top of the visualization is the toggle control for the sales metric type (units, dollars, dollars adjusted for inflation) encoding of all the charts. To the right of the Sales Metric Type Control is the Aggregate Medium Types Control that toggles aggregation of the sales data by medium type (physical, digital, streaming).

On the left side of the visualization is the Context Overview Chart. This area chart shows the total sales of all music formats for each year over the entire time interval of the dataset 1973 to 2014. A user can brush this chart to select a time interval to focus on through context zooming. The Primary and Secondary Focus Charts below the Context Overview Chart display the sales of the various music formats in the selected time intervals. The Chart View Controls above the Primary Focus Chart toggle between the different view types of the focus charts, which are line, stacked area, or derived data (i.e. the change in sales over each year by either percentage or actual value) charts. The slider controls to the left of the focus charts allow the user to rescale the vertical axis of the chart in a nonlinear fashion.

METRIC: ○ — ○ — ○ —

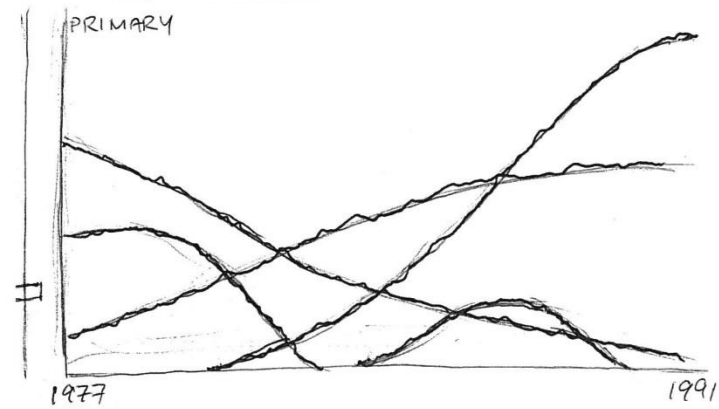
AGGREGATE MEDIATYPES: □

OVERVIEW

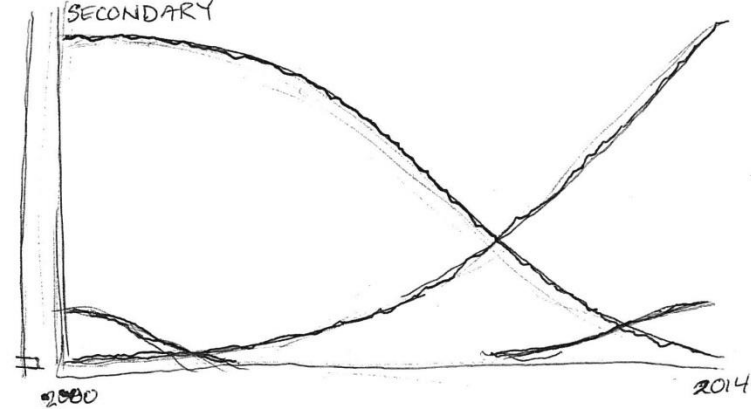


VIEW: ○ — ○ — ○ — ○ —

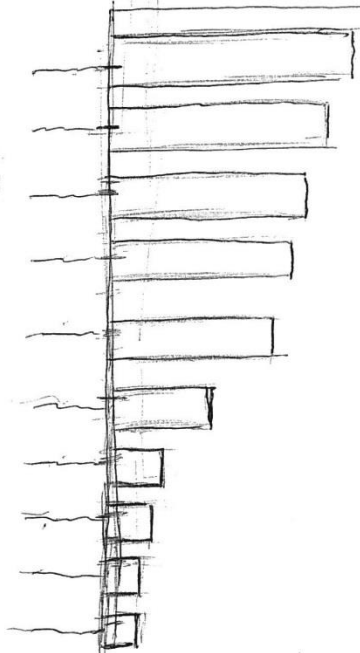
PRIMARY



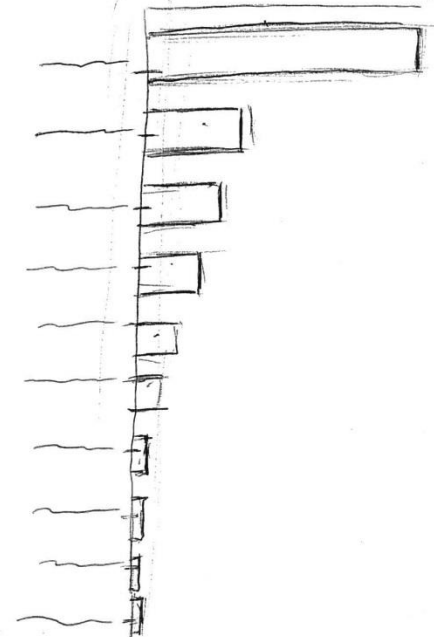
SECONDARY



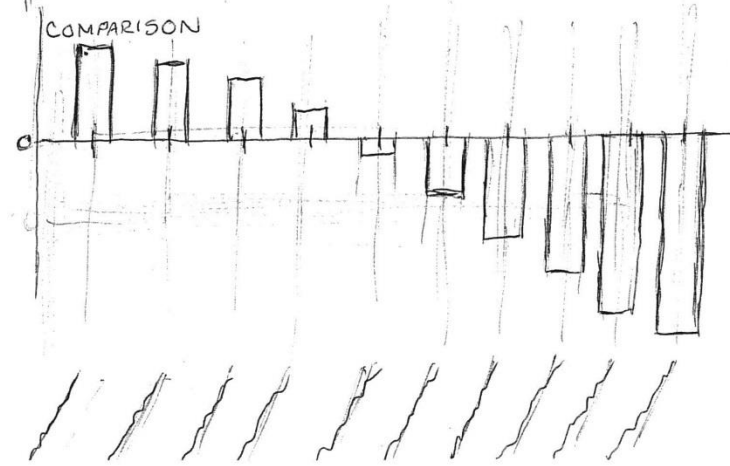
RANKING (PRIMARY)

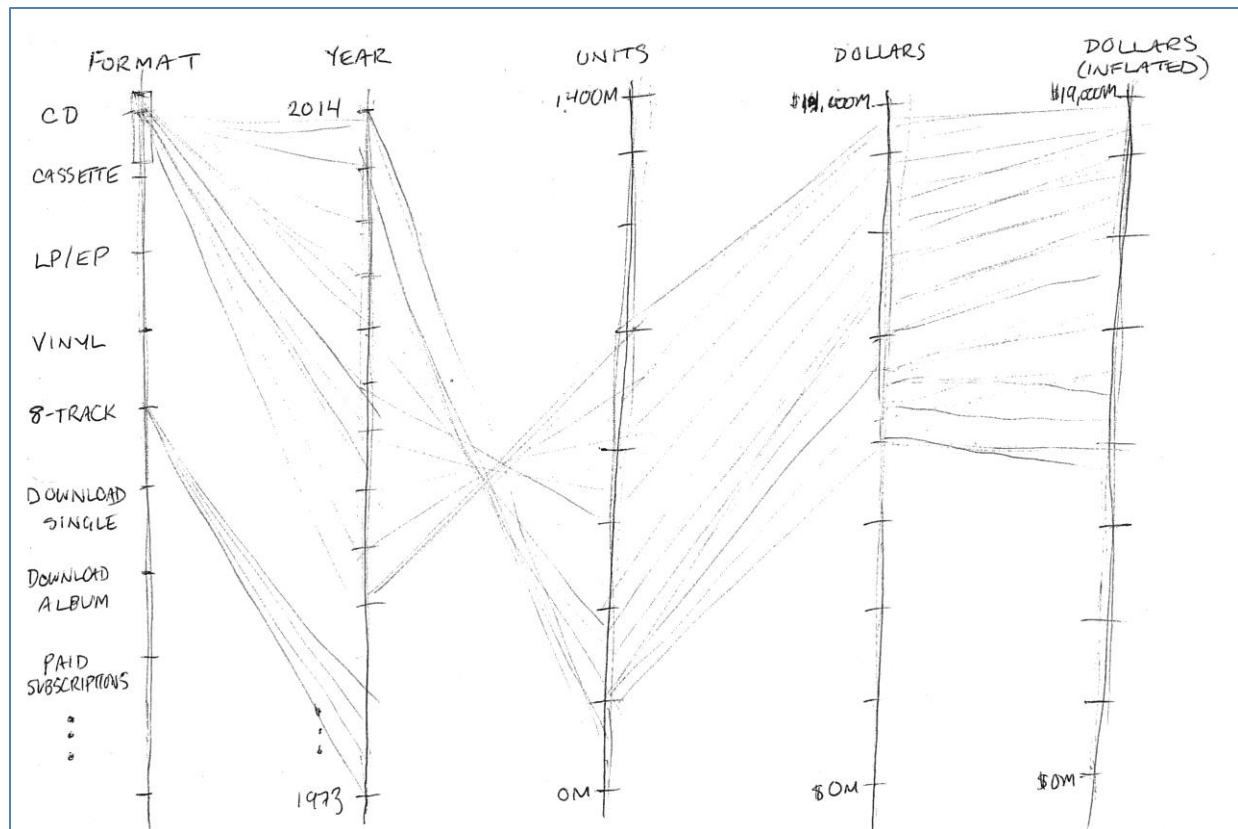


RANKING (SECONDARY)



COMPARISON





On the right side of the visualization are the Primary and Secondary Ranking Charts. These sorted bar charts show the rankings of the cumulative sales for each music format over the selected time interval. Under the ranking charts is the Comparison Chart that displays the difference in sales for each music format between the secondary and primary time interval selections. Lastly, on the bottom of the visualization is the Parallel Coordinates Chart that shows the relationship across multiple attributes in the data along the parallel axes.

The user can hover over the bars to highlight a particular music format of interest and fade the other music formats in both ranking charts as well as the focus charts. The user can do the same thing by hovering over the lines in the focus charts. The user can also click on a music format label in the ranking charts to toggle filtering of a format of interest in the focus charts. Finally, each format has a distinct color so the user can distinguish between different formats easier.

MUST-HAVE FEATURES

The must-have features of the visualization include:

1. toggling the displayed sales metric data (units, dollars, dollars adjusted for inflation) through selection
2. focusing on the sales of a specified time interval through brushing and context zooming
3. ranking the sales of each music format through sorting
4. viewing the changes in sales for each music format over time through sliding of the brushed selection

5. highlighting and fading of music formats through mouse hovering
6. focusing on the sales of a subset of music formats through filtering
7. having a storytelling guide of music milestones synchronizing with the context overview
8. showing the relationship across multiple attributes through parallel coordinates

OPTIONAL FEATURES

The optional features of the visualization include:

9. comparing sales between different time intervals through multiple brush selections
10. aggregating sales by medium type (physical, digital, streaming)
11. toggling between displaying raw sales data and derived data, which is the change in sales over each year by either percentage or actual value
12. toggling between a line chart and stacked area chart
13. rescaling the vertical axis of the focus chart in a nonlinear fashion
14. showing the differences in sales between specified time intervals in a comparison chart

PROJECT SCHEDULE

The following project schedule lists the objectives for each week leading up to the final project deadline and the team member(s) responsible.

Objectives for 4/10/2015

1. an initial dataset parsed and processed into JavaScript objects [Jason]
2. a JavaScript object for the music format coloring scheme metadata [Jason]
3. an initial HTML layout file with chart control elements and stubbed out placeholders for each chart [Kyle, Jason]
4. an initial focus chart displaying the number of units sold for each music format for the entire time interval of the dataset 1973 to 2014 [Jason]
5. an initial ranking bar chart displaying the total units sold from 1973 to 2014 for each music format sorted in descending order [Shadi]
6. an initial outline version of the Process Book [Jason]

Objectives for 4/17/2015 (MILESTONE 1 DUE)

7. derived dataset parsed and processed into JavaScript objects [Jason]
8. the focus chart displaying either units sold, dollar value sold, or dollar value adjusted for inflation sold through toggle control [Jason]
9. the ranking bar chart displaying either units sold, dollar value sold, or dollar value adjusted for inflation sold through toggle control [Shadi]
10. a context overview chart with the focus chart displaying the sales of a specified time interval through brushing and context zooming [Jason]
11. the ranking bar chart displaying the cumulative sales of each music format for the time interval specified by the brush selection [Shadi]
12. an initial version of the parallel coordinates chart [Kyle]
13. HTML layout file with context overview chart added and its controls bound [Jason]

14. HTML layout file with focus chart added and its controls bound [Jason]
15. HTML layout file with ranking bar chart added and its controls bound [Shadi]
16. HTML layout file with parallel coordinates chart added and its controls bound [Kyle]
17. an updated version of the Process Book [Jason, Kyle, Shadi]

Objectives for 4/24/2015 (PROJECT REVIEW WITH TF)

18. the focus chart highlighting a music format of interest and fading the other music formats through hovering [Jason]
19. the ranking bar chart highlighting a music format of interest and fading the other music formats through hovering [Shadi]
20. the focus chart displaying a subset of music formats through filter controls [Jason]
21. the ranking bar chart displaying a subset of music formats through filter controls [Shadi]
22. the context overview chart allowing for a second brush selection [Jason]
23. an initial version of the music milestone guide added to the HTML layout file [Jason]
24. another focus chart displaying the sales for the second brush selection interval [Jason]
25. another ranking bar chart displaying the sales for the second brush selection interval [Shadi]
26. the parallel coordinates chart allowing for brushing and filtering [Kyle]
27. an initial version of Project Website [Jason]

Objectives for 5/5/2015 (FINAL PROJECT DUE)

28. the final version of the music milestone guide [Jason]
29. the focus chart displaying the aggregate values of each medium type (physical, digital, streaming) through aggregation controls [n/a]
30. the bar chart displaying the aggregate values of each medium type (physical, digital, streaming) through aggregation controls [n/a]
31. the focus chart displaying a line or stacked area chart through toggle control [n/a]
32. the focus chart displaying raw sales data or derived data through toggle control [Jason]
33. the vertical axis of the focus chart rescaling in a nonlinear fashion [Jason]
34. the comparison bar chart displaying the differences in sales between the two brush selection intervals [n/a]
35. the final version of the Process Book [Jason, Kyle, Shadi]
36. the final version of the Project Website and Screencast [Jason, Kyle, Shadi]
37. README file [Jason]

PROGRESS AND MILESTONES

The following is a timeline of progress made and milestones achieved.

3/26/2015

- contacted RIAA to request access to their shipment database for academic use

3/29/2015

- RIAA grants us student access to their database

4/2/2015

- created project repository

4/7/2015

- completed Objective #1

4/8/2015

- revised data structure
- completed Objective #3
- completed Objective #4

4/9/2015

- revised design based on feedback from our Teaching Fellow project advisor to include a parallel coordinates chart
- revised data structure format due to confusion expressed by our TF
- completed Objective #2

4/11/2015

- completed Objective #6

4/12/2015

- completed Objective #7
- completed Objective #8
- completed Objective #13

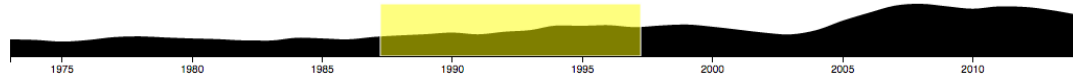
4/13/2015

- completed Objective #10
- completed Objective #14
- completed Objective #18
- below is a screenshot of the visualization at this point in time:

RIAA Music Sales

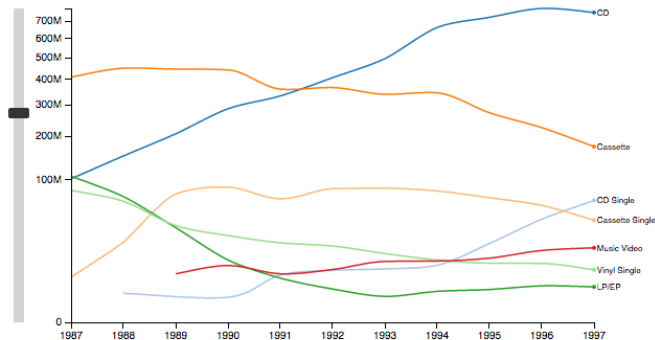
Sales Metric: ☒ Units ☐ Dollars ☐ Dollars (inflated) ☐ Price Per Unit ☐ Price Per Unit (inflated)

context overview



focus selections

rankings



parallel coordinates chart

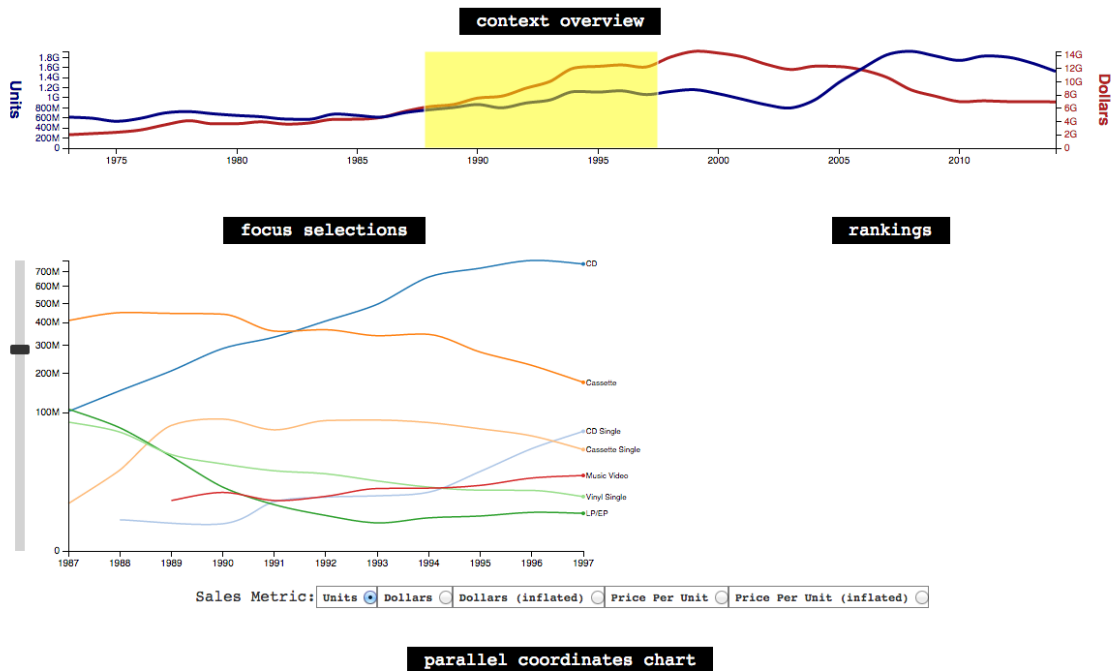
4/14/2015

- revised design based on feedback from our peer review session to include a storytelling guide of music milestones synchronizing with the context overview

4/15/2015

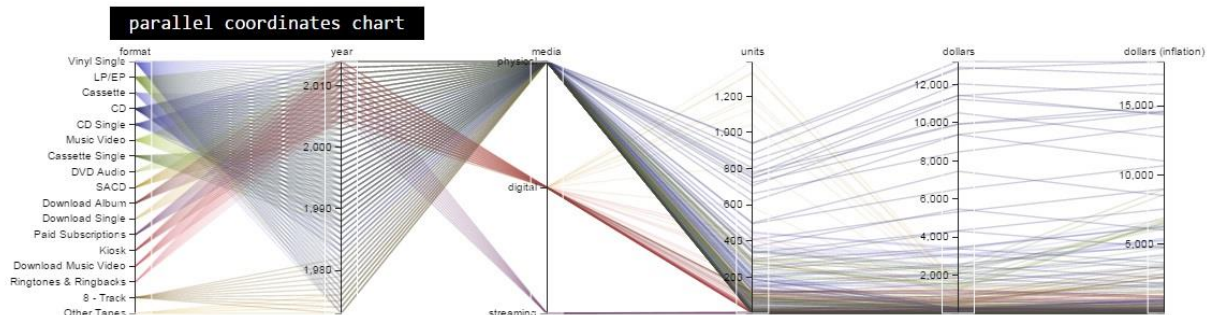
- replaced area chart with dual y-axes line chart in Context Overview
- below is a screenshot of the visualization at this point in time:

RIAA Music Sales



4/16/2015

- scraped music milestones data from the Pearson Education website (<http://www.infoplease.com/ipea/A0151192.html>) using the Kimono Labs API (<https://www.kimonolabs.com/apis/4olw8hoa>)
- completed Parallel Coordinates Chart and added it to the HTML index file. Below is an image of the Parallel Coordinates Chart at this time:



4/17/2015

- changed radio buttons to dropdown menus
- completed Objective #23
- completed Objective #24
- below is a screenshot of the visualization updates at this point in time:

RIAA Music Sales

milestone guide

Click the buttons to toggle through some music milestones.



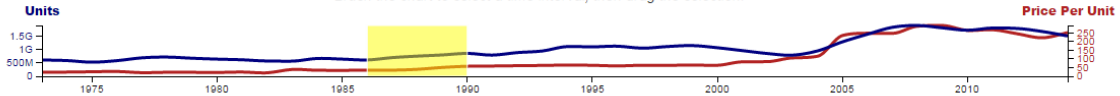
1988: CDs outsell vinyl records for the first time.

Sales Metric #1:

Sales Metric #2:

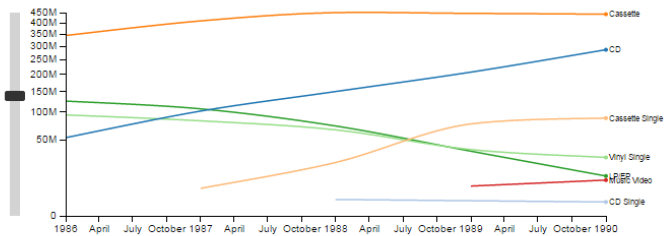
context overview

Brush the chart to select a time interval, then drag the selection.



focus selection #1

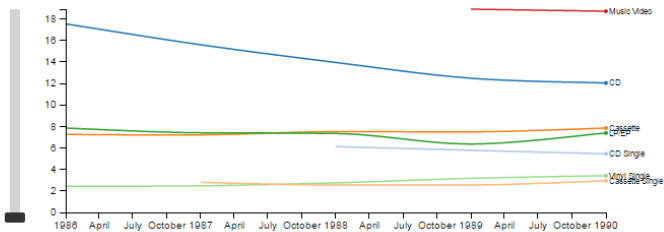
Hover over a format to highlight. Drag the side slider to rescale.



rankings #1

focus selection #2

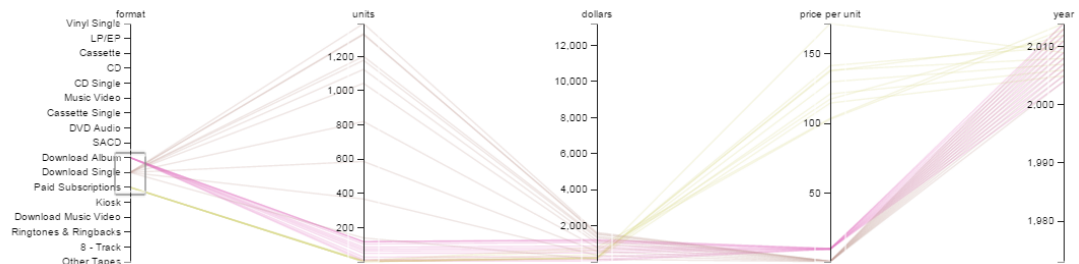
Hover over a format to highlight. Drag the side slider to rescale.



rankings #2

parallel coordinates chart

Brush or reorder the axes.

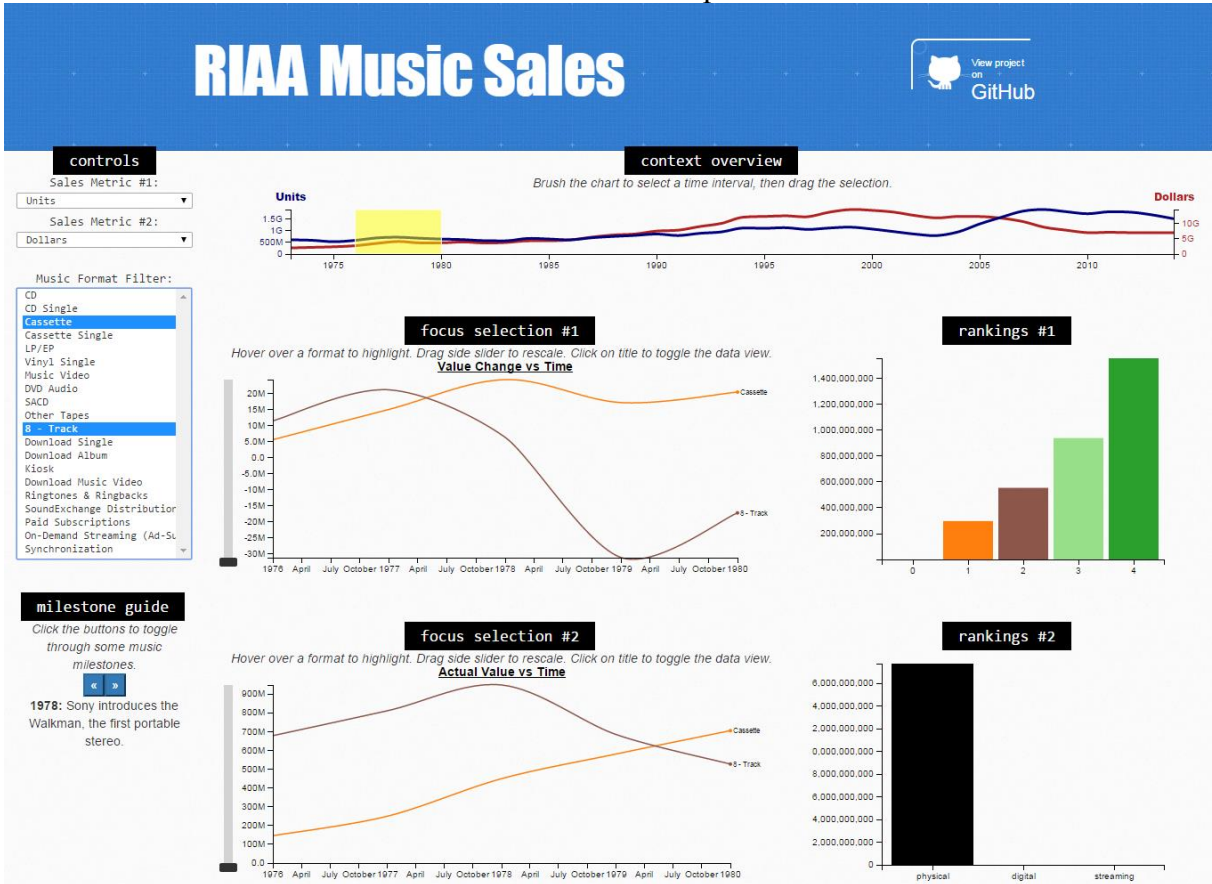


4/18/2015

- completed Objective #5
- completed Objective #9
- completed Objective #11
- completed Objective #15

4/21/2015

- updated HTML page styling
- reorganized HTML layout by moving controls and Milestone Guide to left of the charts
- completed Objective #20
- completed Objective #32
- below is a screenshot of the visualization at this point in time:



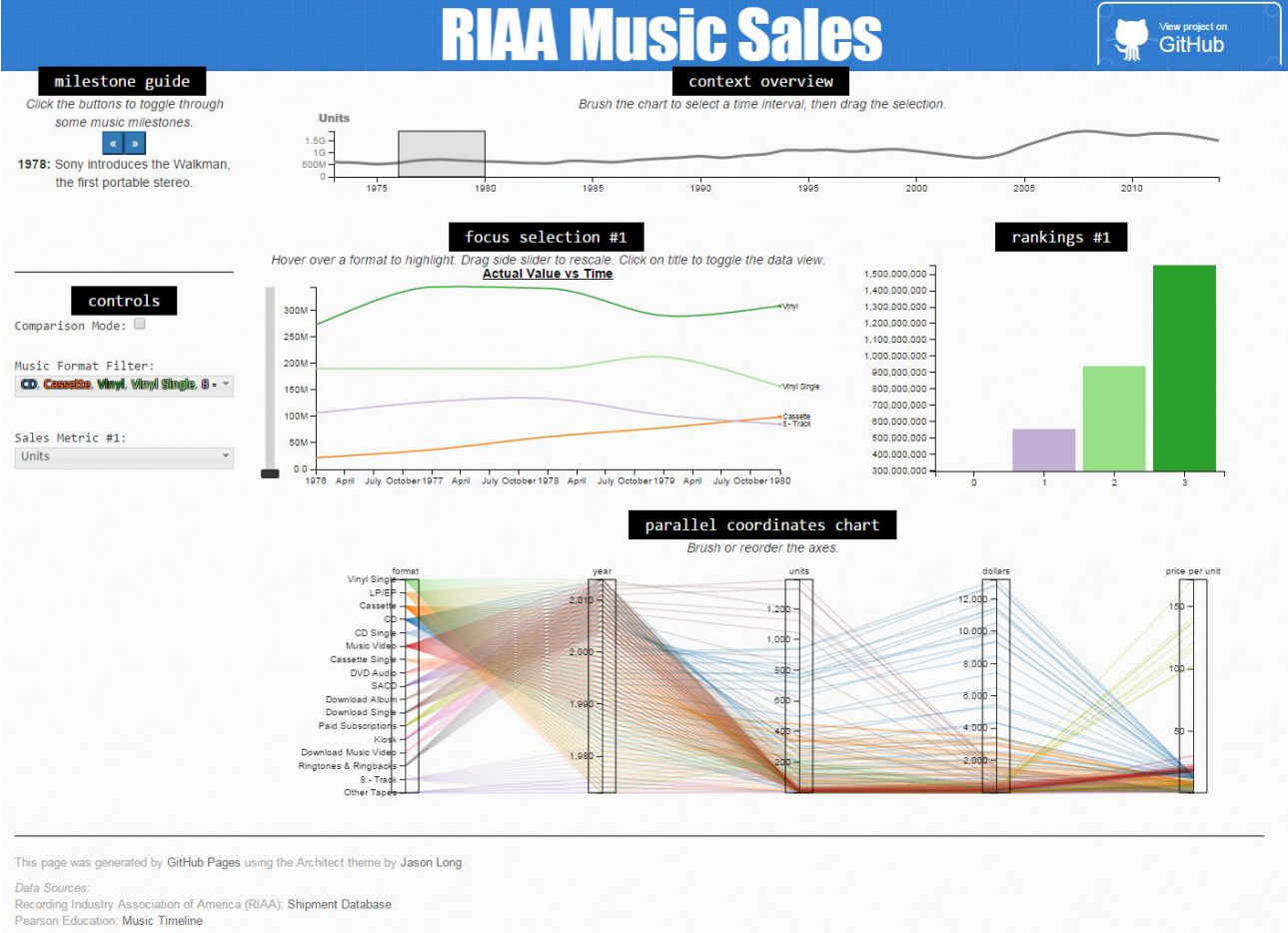
4/24/2015

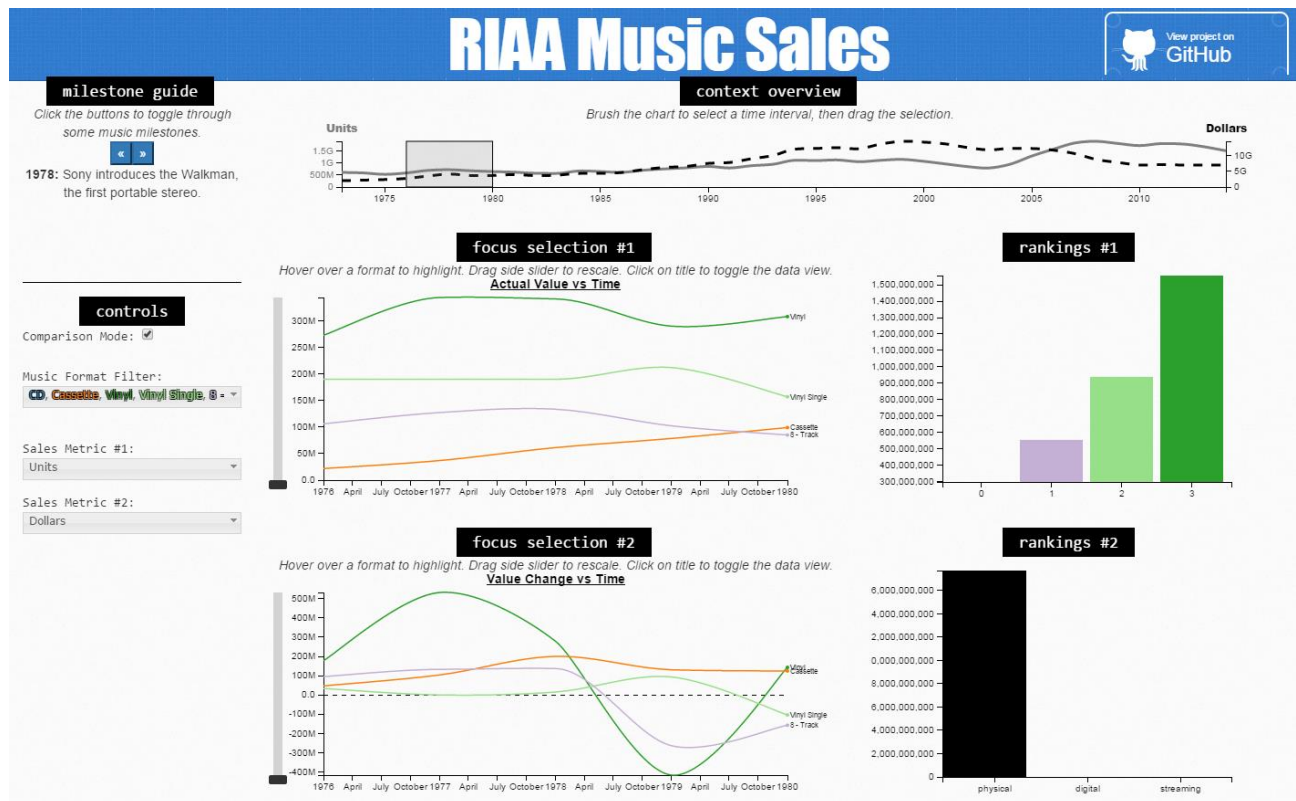
- transformed format filter select menu into a dropdown checklist menu using jQuery libraries
- colored format texts and grouped formats by medium type in Filter Control



4/26/2015

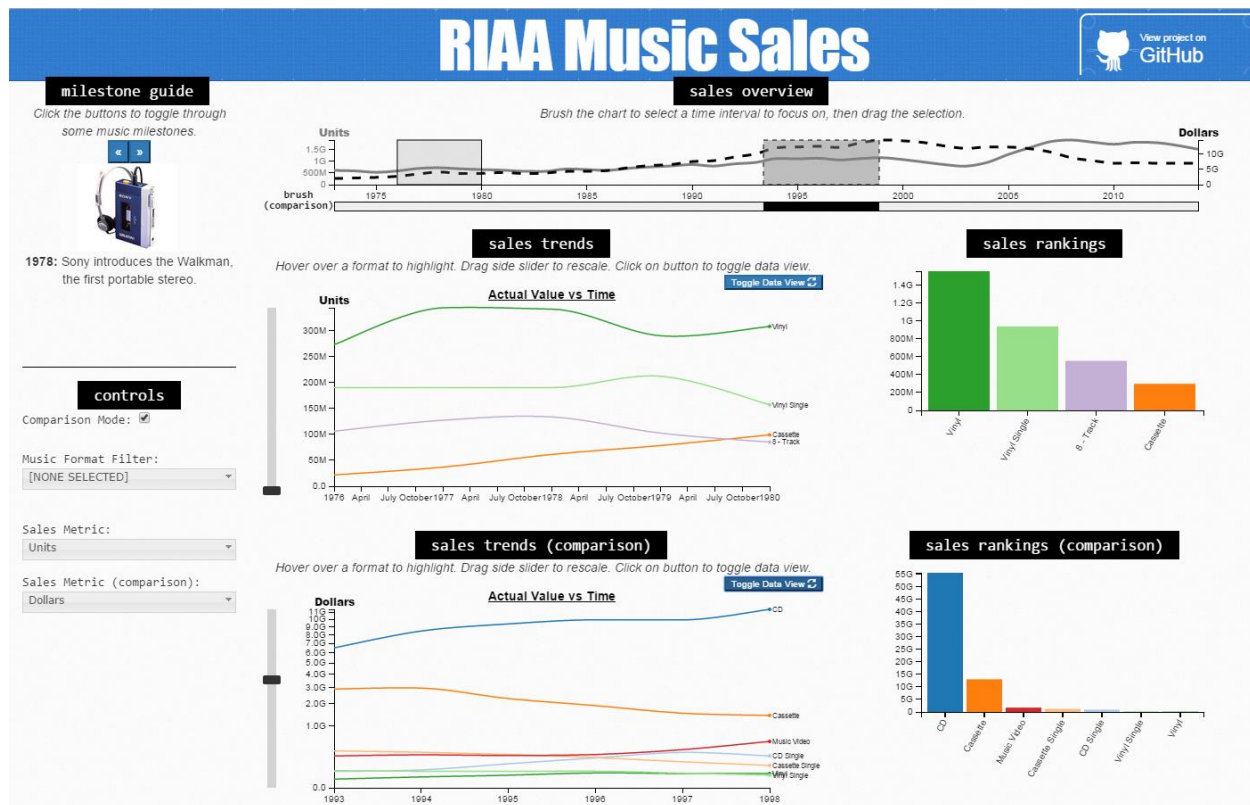
- added footer section with references and data source information to HTML index file
- added a zero y-axis dashed line to the Focus Charts
- updated HTML layout to use 'absolute' positioning and added a toggle checkbox to switch between the Parallel Coordinates Chart and the Focus & Rankings Comparison Charts
- below is a screenshot of the visualization at this point in time:





4/29/2015

- added some images to the milestone guide
- completed Objective #22 (second brush for comparison)
- below is a screenshot of the visualization in “Comparison Mode” at this point in time:

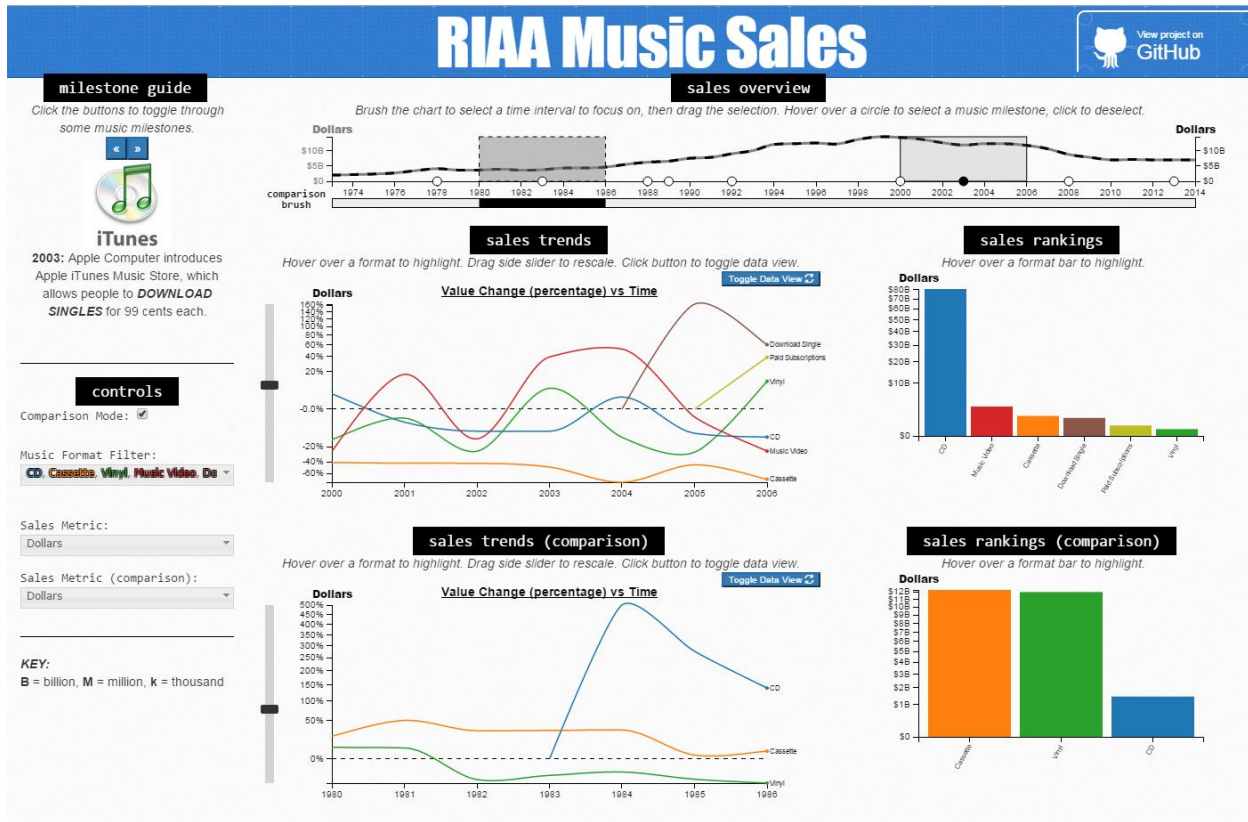


5/2/2015

- added a default subset of formats for the user to focus on initially
- added hoverable milestone markers in Context Overview Chart
- implemented brush to snap to each year on the axis in Context Overview Chart making brushing more intuitive and CPU efficient
- modified milestone descriptions to add bold and italicized text to put emphasis on the format types
- modified time axis to not show months, but only years in Focus Charts

5/3/2015

- implemented filtering, highlighting, axis rescaling and transition animation for the Rankings Charts
- added '\$' and '%' symbols to axis labels to make the value clearer to the user
- below is a screenshot of the visualization in "Comparison Mode" at this point in time:



RELATED WORK AND REFERENCES

The following lists code examples and references that inspired this project; such as a paper, a web site, visualizations we discussed in class.

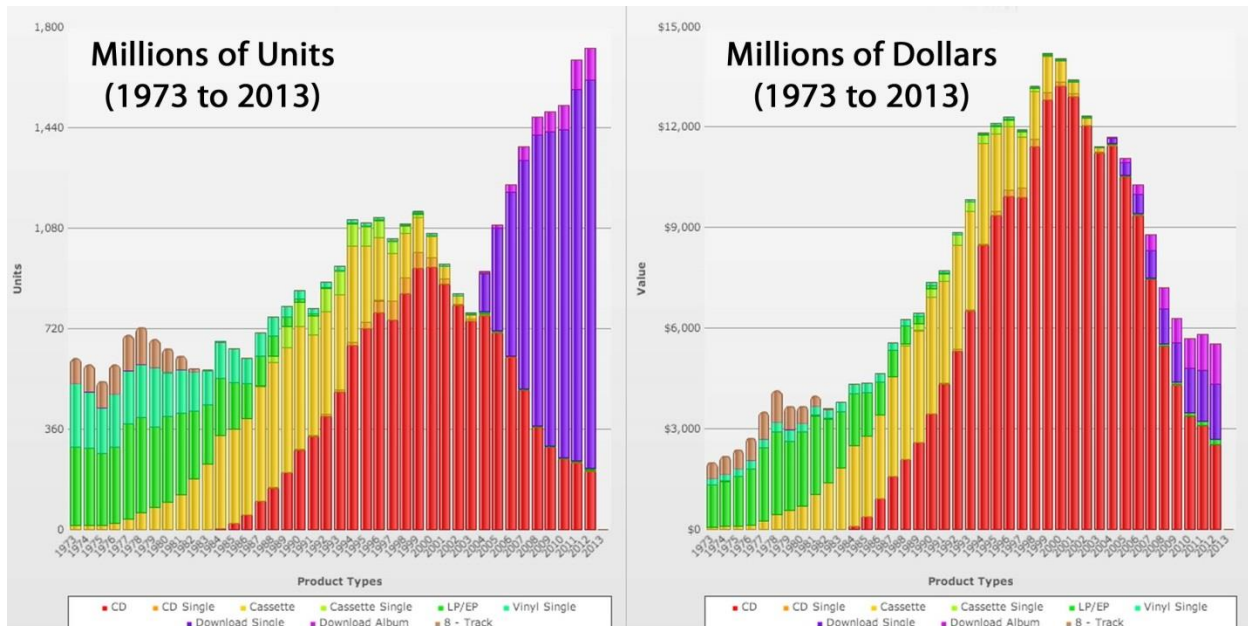
- News and Notes on 2014 RIAA Music Industry Shipment and Revenue Statistics
<http://riaa.com/media/D1F4E3E8-D3E0-FCEE-BB55-FD8B35BC8785.pdf>
- Music Timeline (Pearson Education)
<http://www.infoplease.com/ipea/A0151192.html>
- Spotify
<http://en.wikipedia.org/wiki/Spotify>
- 40 Years of Album Sales Data in Two Handy Charts
<http://blog.thecurrent.org/2014/02/40-years-of-album-sales-data-in-one-handy-chart/>
- D3 Code Examples
 - Multi-Series Line Chart
<http://bl.ocks.org/mbostock/3884955>
 - Multi-Line Voronoi
<http://bl.ocks.org/mbostock/8033015>
 - X-Value Mouseover
<http://bl.ocks.org/mbostock/3902569>
 - Focus+Context via Brushing
<http://bl.ocks.org/mbostock/1667367>
 - Line Graph With Dual Y Axes

- <http://bl.ocks.org/d3noob/e34791a32a54e015f57d>
 - Parallel Coordinates
 - <http://bl.ocks.org/mbostock/1341021>
 - <http://bl.ocks.org/jasondavies/1341281>
 - Multi-Brush
 - <http://bl.ocks.org/bollwyvl/8463137>
- Homework 3 – MyWorld 2015 Data Timeline
- jQuery Event Handler Attachment
 - <https://api.jquery.com/category/events/event-handler-attachment/>
- Parallel Coordinates and the parcoords.js library
 - <https://syntagmatic.github.io/parallel-coordinates/>
- jQuery Dropdown Check List
 - <http://dropdownchecklist.sourceforge.net/>
- GitHub Pages - Architect theme by Jason Long
 - <https://pages.github.com>, <https://twitter.com/jasonlong>

EXPLORATORY DATA ANALYSIS

The Recording Industry Association of America (RIAA) has a web form interface for their Shipment Database (https://www.riaa.com/keystatistics.php?content_selector=riaa-shipment-database-log-in) that allows a user to generate static bar and pie charts of the data. We used this initially to look at the music sales data. The following are sample charts the web form interface can create.

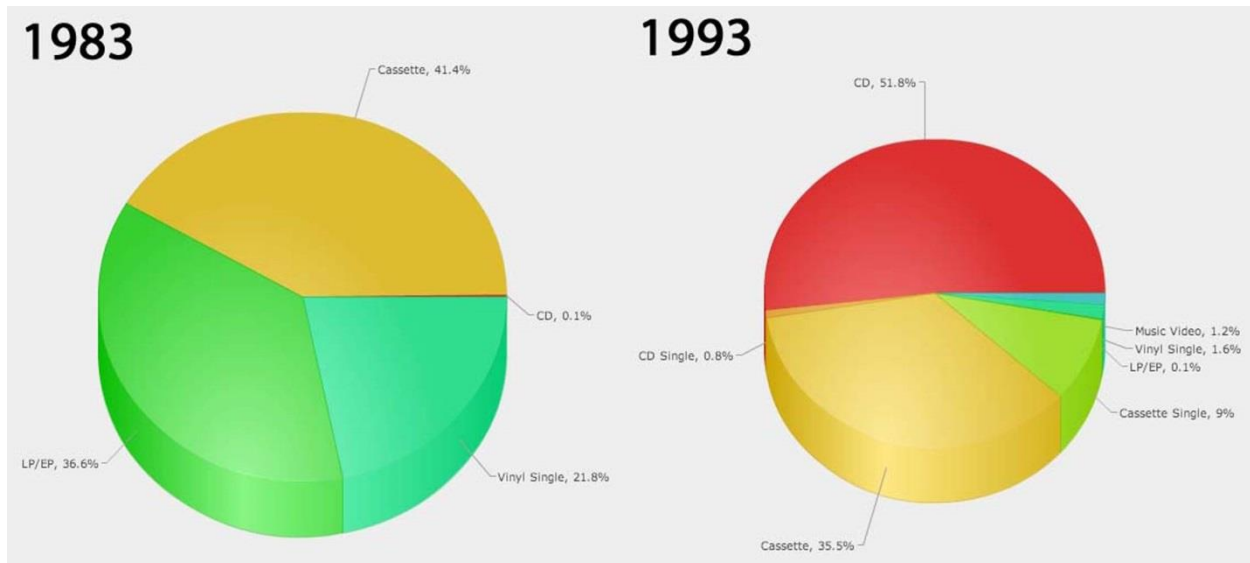
RIAA Stacked Bar Charts (Units vs. Dollars)



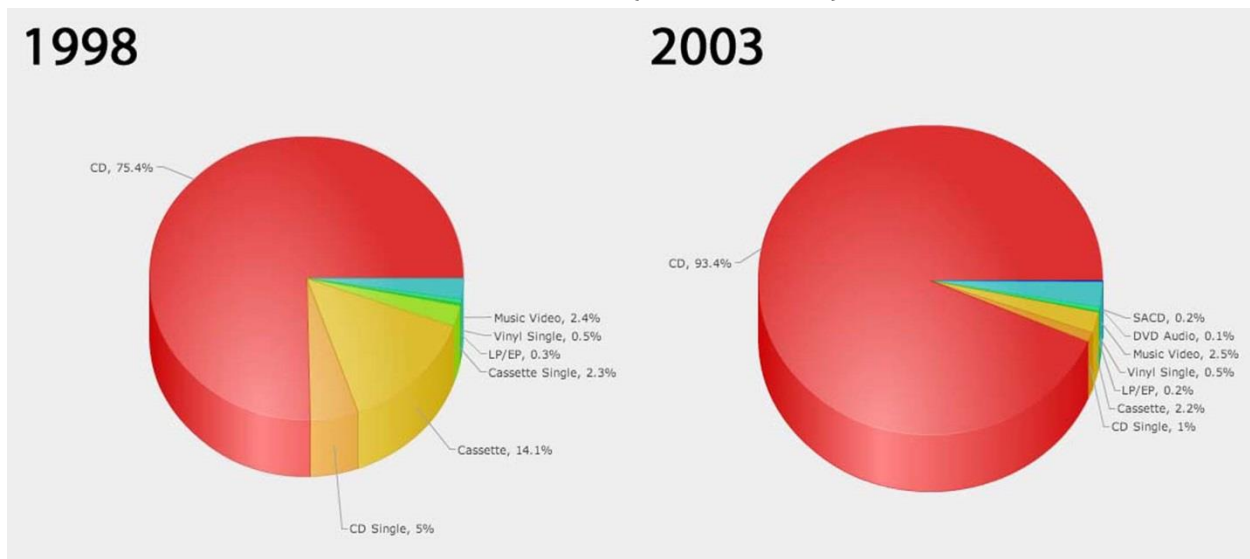
This gives insight into the overall dominance of the CD format. In recent years, however, the CD format has been decreasing in sales while the digital download formats have been increasing. Moreover, it appears the digital download format is currently dominating in terms of units, but

not as much in terms of dollar amounts. This insight informed are design by leading us to wonder how price per unit has changed over time.

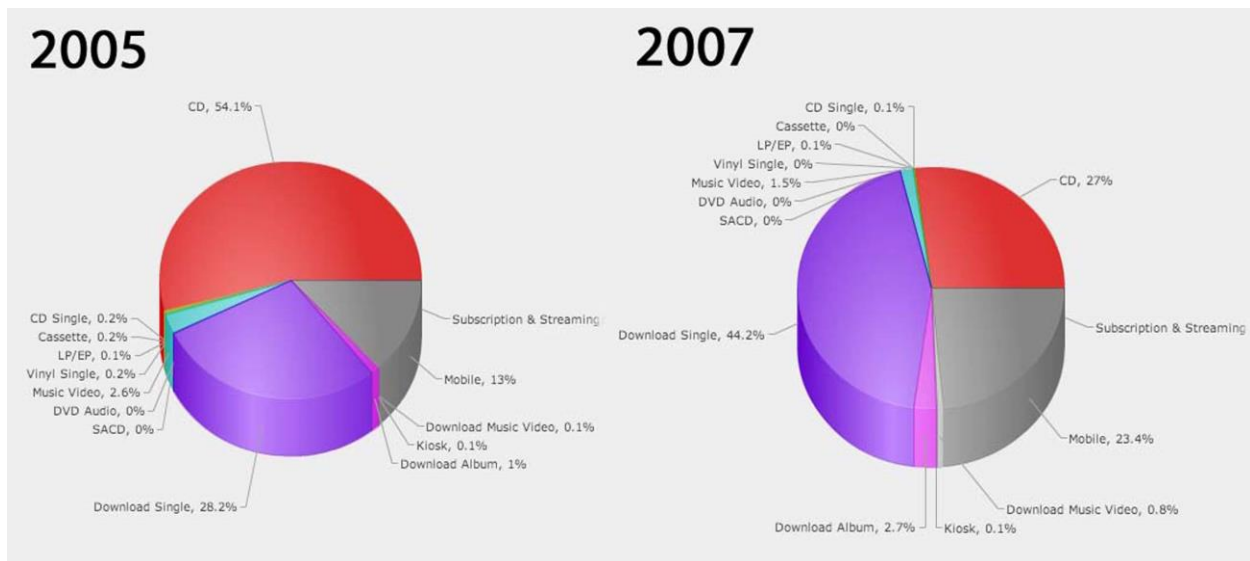
RIAA Pie Chart (1983 vs. 1993)



RIAA Pie Chart (1983 vs. 1993)



RIAA Pie Chart (2005 vs. 2007)



These pie charts provide insight into how the shift of music formats has been changing more quickly over time. The first pie chart shows the shift in popularity from the Cassette to the CD format over a period of 10 years. The second pie chart shows the CD format almost completely dominating after a 5-year period. While the third pie chart shows a drastic change to the digital download formats in only two short years.

The use of 3D pie charts in this visualization makes it somewhat difficult to judge the value of each portion although the labels help. For instance, the portions of the pie for the Cassette and LP/EP formats in the 1983 chart look almost the same visually, but one is 41.4% and the other is 36.6%, respectively. This influenced our design by using a Ranking Bar Chart in place of something like these 3D pie charts so that a user can better distinguish the portion values of different music formats.

DESIGN EVOLUTION

Our initial design did not include the Parallel Coordinates Chart. We decided to add it after a suggestion from our Teaching Fellow project advisor so that we could encode more than one sales metric attribute at the same time. We also decided using dual y-axes to accomplish this. Additionally, we decided to remove the Comparison Chart because we believed the Parallel Coordinates Chart would be more useful.

Another suggestion our Teaching Fellow gave us was to visualize the price per unit over time for the different music formats. Therefore, we added this derived data to our data processing logic. Moreover, we renamed some data fieldnames after our TF expressed some confusion. As a result, the fieldname 'formatName' became 'format' (CD, cassette, vinyl, download single, paid subscriptions, etc.) and 'formatType' became 'medium' (physical, digital, streaming). In addition, we revised the data structure from a single array to three arrays, one for each sales metric, to make the data easier to deal with in the code.

Based on feedback from the peer review session, we decided to incorporate a storytelling feature that would guide a user through various music milestones significantly influencing the sales of the different music formats. This Music Milestone Guide would synchronize with the timeline in the Context Overview Chart.

Based on the feedback from the project review session with our TF, we updated the HTML layout to use absolute positioning to fit the charts better. We added a “Comparison Mode” toggle control to switch between the Parallel Coordinates Chart and the Focus & Rankings Comparison Charts. This way all the charts could fit on the screen. In addition, we revised the Format Filter dropdown menu by coloring the format texts and grouping the formats by medium type.

We also completed integrated brushing and highlighting between the Parallel Coordinates Chart and the Comparison Charts, to improve the image that these charts are working in tandem to tell a story, rather than being two disjoint and separate visualizations.

We have been compiling iconic images to display along with our Context Overview Chart, per our TF’s suggestion, to catch the viewer’s eye and give them things to remember and explore within our charts. We are also in the process of developing a clickable timetable with those images on it, possibly overlaid on or below the Context Overview Chart, to give the viewer a more guided process in which to interact with the data.

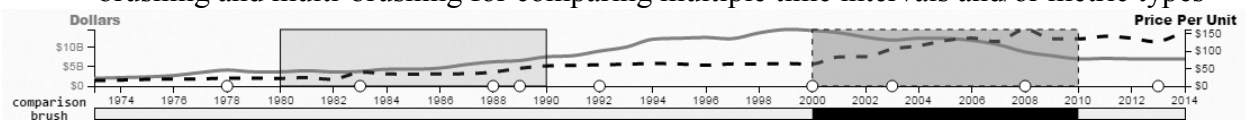
In addition, from our TF’s feedback, we have done some cleaning on our axis, making sure that numbers and units are clearly marked along all necessary axis. Also, we have been working on making some clickable text, which wasn’t quite obvious that it was clickable, stand out more by turning them into more obvious buttons. Finally, we have improved the general color scheme of our visualization, by removing bright distracting colors from less important objects like brushes, and using almost all colors solely for data representation, so that viewers’ attention is focused on the important part of the visualization, the data.

IMPLEMENTATION

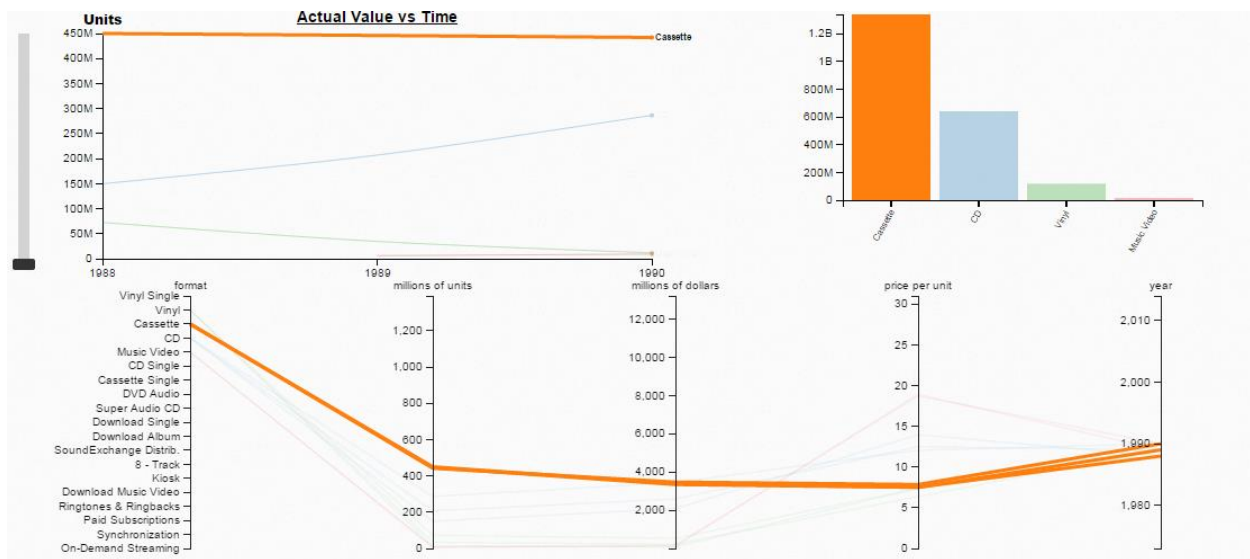
We implemented our visualization using the jQuery Events Handler library to coordinate data updates between all the different charts. We leveraged a Parallel Coordinates library developed by syntagmatic to create our Parallel Coordinates Chart. Last, we utilized a jQuery Dropdown Check List library to create our dropdown menu controls.

The interactive visualization functionalities we implemented include:

- format filtering
- selecting multiple sales metric types for comparison
- brushing and multi-brushing for comparing multiple time intervals and/or metric types



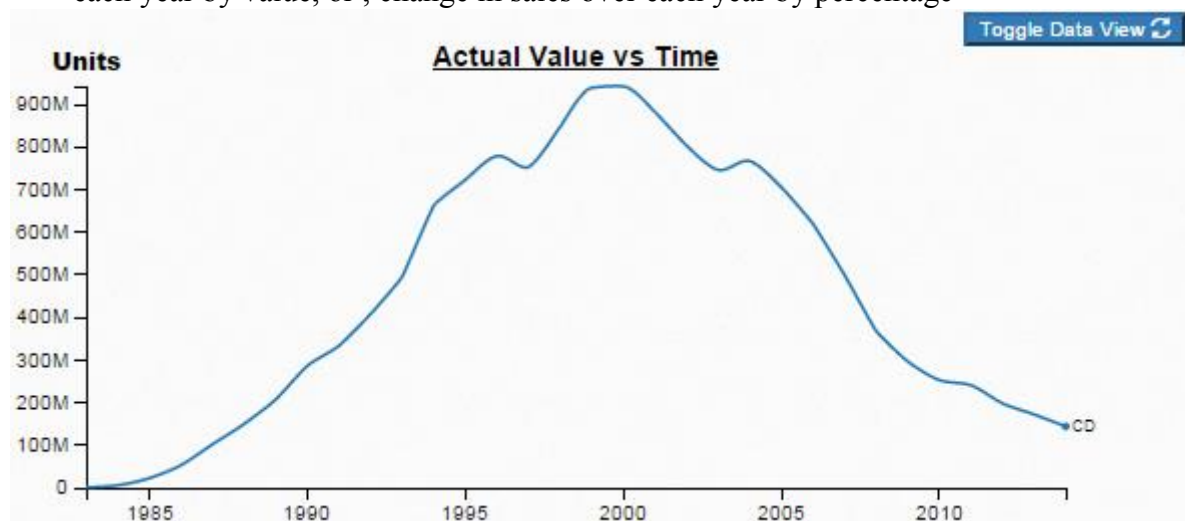
- format highlighting coordinated between multiple charts

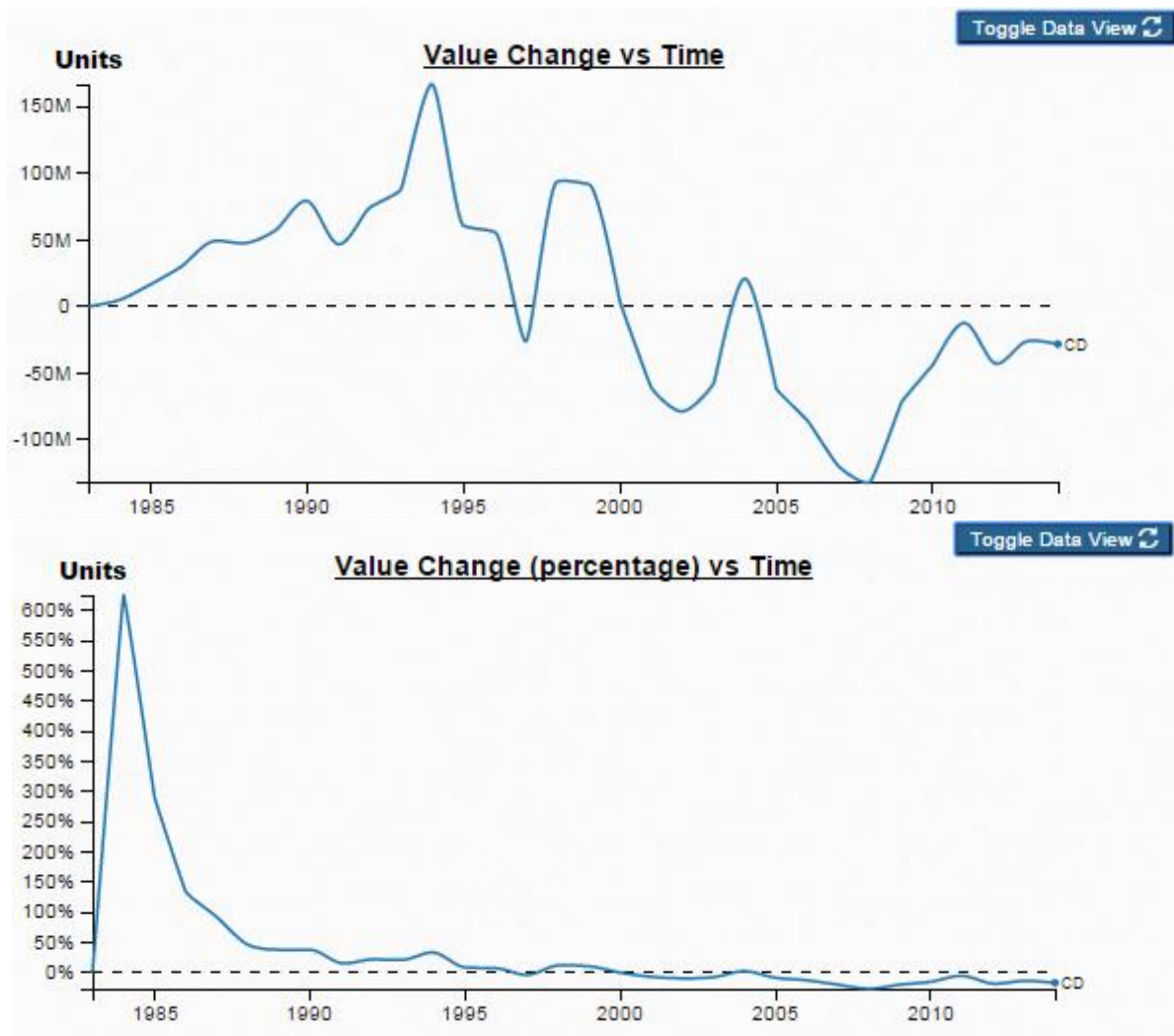


- toggling through music milestones pertaining to format sales using the buttons on the Milestone Guide or hovering over markers on the overview timeline



- toggling through data views to display either the actual sales values, change in sales over each year by value, or , change in sales over each year by percentage



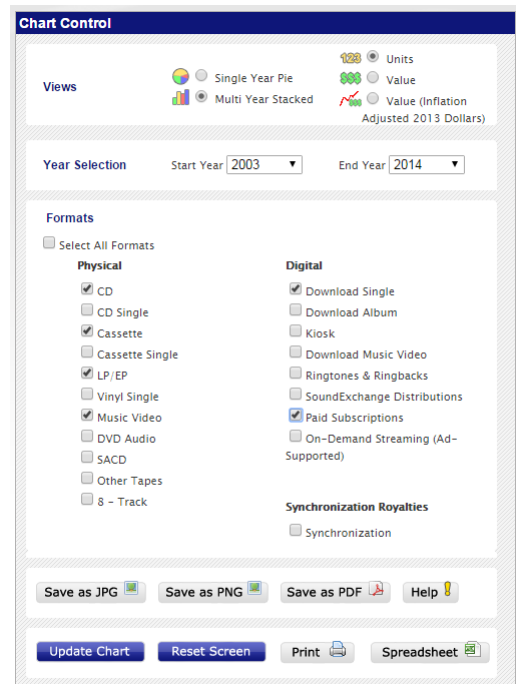


EVALUATION

There was a bit of concern during the implementation process whether having both the Sales Trends Chart as well as the Parallel Coordinates Chart in the same picture was necessary. While there may be a slight amount of overlap – via the similar brushing methods of the sales overview timeline, and the year axis on the Parallel Coordinates Chart, for a multitude of reasons we felt that it was still necessary to have both charts included. The Sales Trends chart provides a strong overview of the value/units of the various formats over time, as well as these formats compare to others at the same time. However, this chart is hindered by its inability to explore the dataset via brushing other attributes of the data. For example, if some were curious what the cheapest formats per unit were, which were also sold in high units, it would be difficult to see in the Sales Trends Chart. On the other hand, in the Parallel Coordinates Chart, one could simply brush the bottom half of the price per unit axis, as well as the top of the total units axis, to receive a good answer to this question. In this way, the ability to compare multiple variable constrictions makes the Parallel Coordinates Chart a useful addition for deeper exploration into the visualization of the dataset.

Improvements over the Original RIAA Visualizations

While the original stacked bar charts and pie charts found on the RIAA website were enough to inspire us to create this project, the shortcomings their setup also spurred us to create our project. For example, while the stacked bar charts on their page may be aesthetically pleasing, due to the nature of stacked bar charts, it is virtually impossible to compare changes within the same format between different years. In addition, even within a single year, while we could tell which formats were larger or smaller, it was quite difficult to say precisely whether for example CD sales were three, four or even five times larger than other formats. The pie charts also suffered from this downfall, albeit a little less so since that also gave the exact sizes of the wedges on the outside. However, once again there was no easy way to compare changes in the data between years in any easy manner. A user could ask for a pie chart of another year, but this would involve clicking a series of buttons, and there was no simple brushing and instant interaction within the program. For this reason, we believe our use of line plots, as well as interactive brushing, has greatly improved the interpretation of this data. Rather than simply showing how music formats were distributed in a single year, we have helped show in better detail the process of how the trends in music format purchases have changed over time. Additionally, our second comparison brush allows a user to compare two time intervals and/or metric types simultaneously. And our scale slider allows a user to see small data values more clearly than the static RIAA charts.



RIAA Chart Control

We learned several interesting things about the data from using our visualization. First, CD sales had great dominance over all other formats in the late 90s, while vinyl sales did in the 70s and cassette sales did in the 80s. But, the difference in magnitude of the latter two's dominance were not as great as that of CDs. We discovered this using the multi-brush comparison and Rankings bar chart feature of our visualization.

Another thing we learned is that the price per unit for CDs decreased over time and flattened out while the price for vinyl records steadily increased after 1992 and still continues to rise. And the overall most expensive prices are the Super Audio CD and DVD Audio formats. We discovered this using the filtering features of the Focus line chart and Rankings bar chart.

Moreover, we observed large spikes in value change in terms of percentage for newly introduced formats (e.g. CD, Download Single) the year directly following the format release. We

discovered this using the data view toggle feature of our visualization to display the change in sales over each year by percentage.

In addition, we noticed an inverse relationship between unit and dollar sales for the Download Single format using the Parallel Coordinates chart in our visualization.