

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 media 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 media 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.)
- media (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.

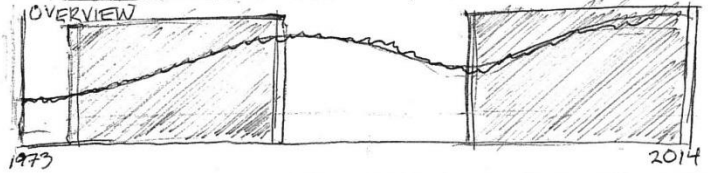
On the top of the visualization is the toggle control for the sales metric type (units, dollars, dollars adjusted for inflation) encoding of all the charts. To the right of the Sales Metric Type Control is the Aggregate Media Types Control that toggles aggregation of the sales data by media type (physical,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

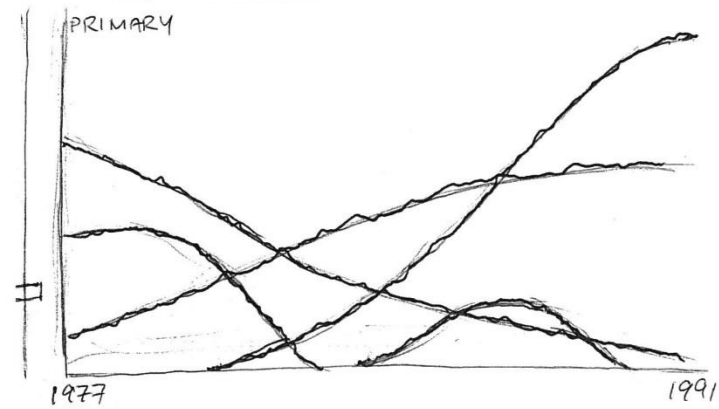
AGGREGATE MEDIATYPES: □

OVERVIEW

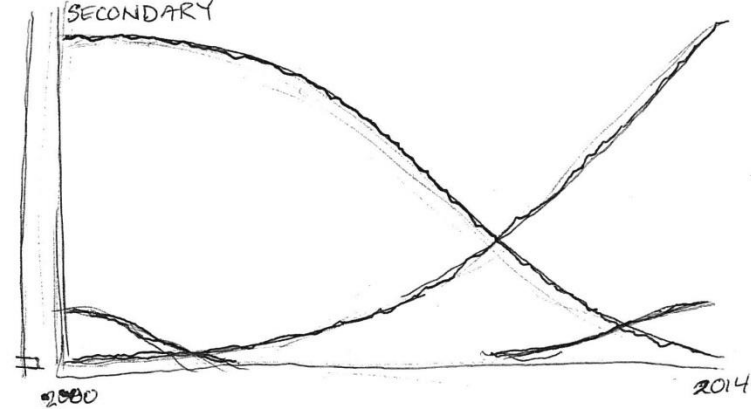


VIEW: ○ — ○ — ○ — ○ —

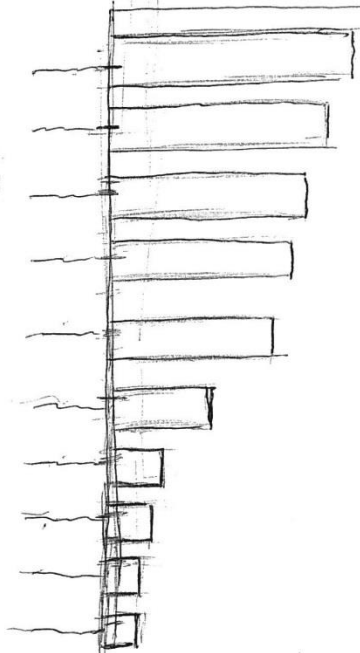
PRIMARY



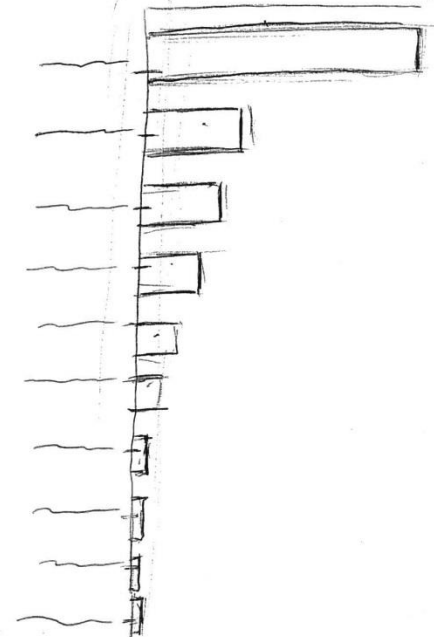
SECONDARY



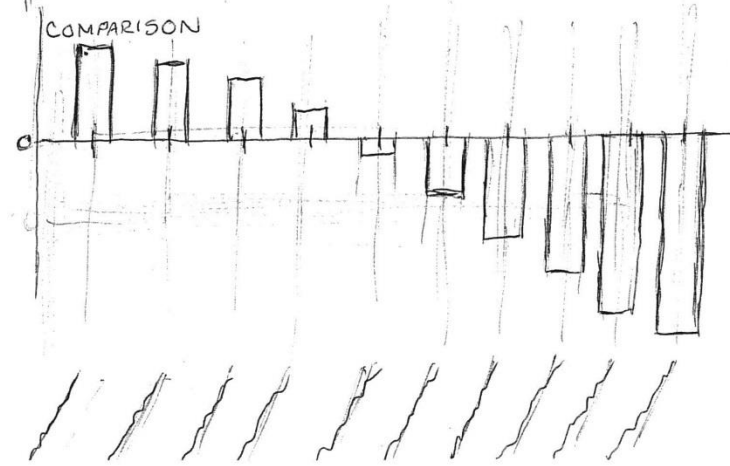
RANKING (PRIMARY)

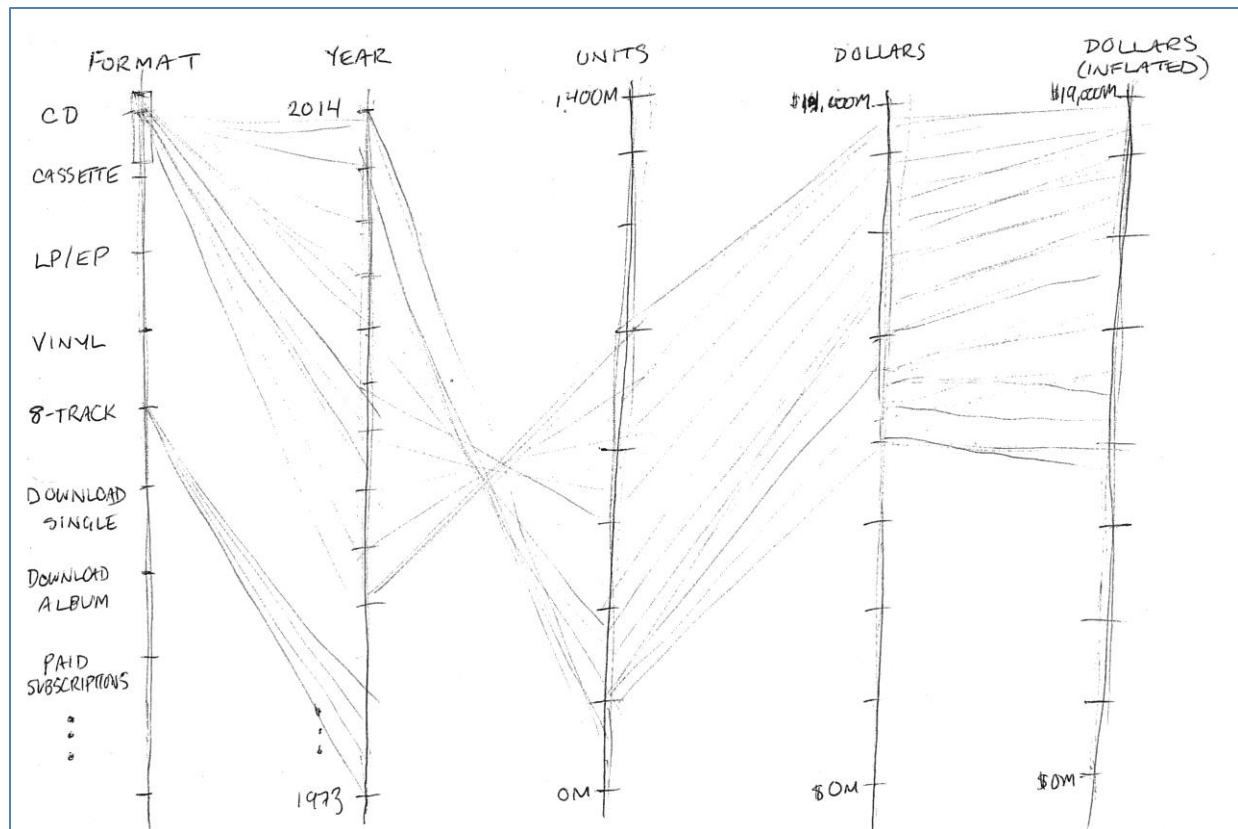


RANKING (SECONDARY)



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. comparing sales between different time intervals through multiple brush selections
8. showing the relationship across multiple attributes through parallel coordinates

OPTIONAL FEATURES

The optional features of the visualization include:

9. aggregating sales by media type (physical, digital, streaming)
10. toggling between displaying raw sales data and derived data, which is the change in sales over each year by either percentage or actual value
11. toggling between a line chart and stacked area chart
12. rescaling the vertical axis of the focus chart in a nonlinear fashion
13. 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 []

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 []
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 using opacity highlighting/fading through filter controls []
22. the context overview chart allowing for a second brush selection [Jason]
23. another focus chart displaying the sales for the second brush selection interval [Jason]
24. another ranking bar chart displaying the sales for the second brush selection interval []
25. the parallel coordinates chart allowing for brushing and filtering []
26. an initial version of Project Website []

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

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

PROGRESS AND MILESTONES

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>
- 40 Years of Album Sales Data in Two Handy Charts

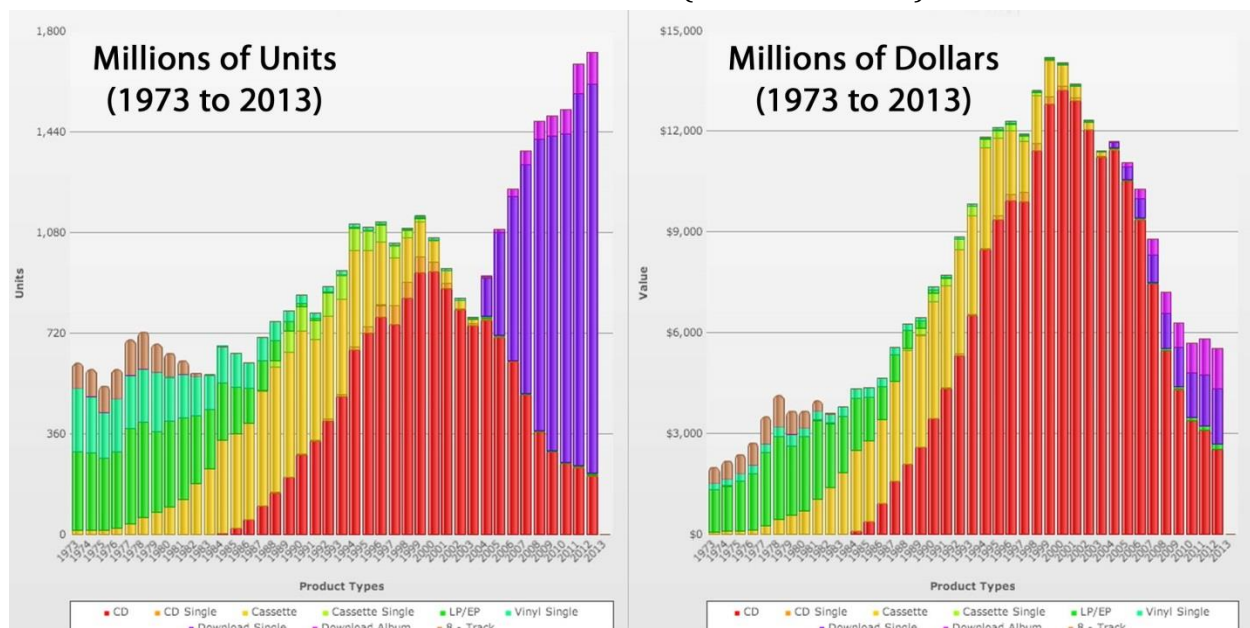
<http://blog.thecurrent.org/2014/02/40-years-of-album-sales-data-in-one-handy-chart/>

- 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

EXPLORATORY DATA ANALYSIS

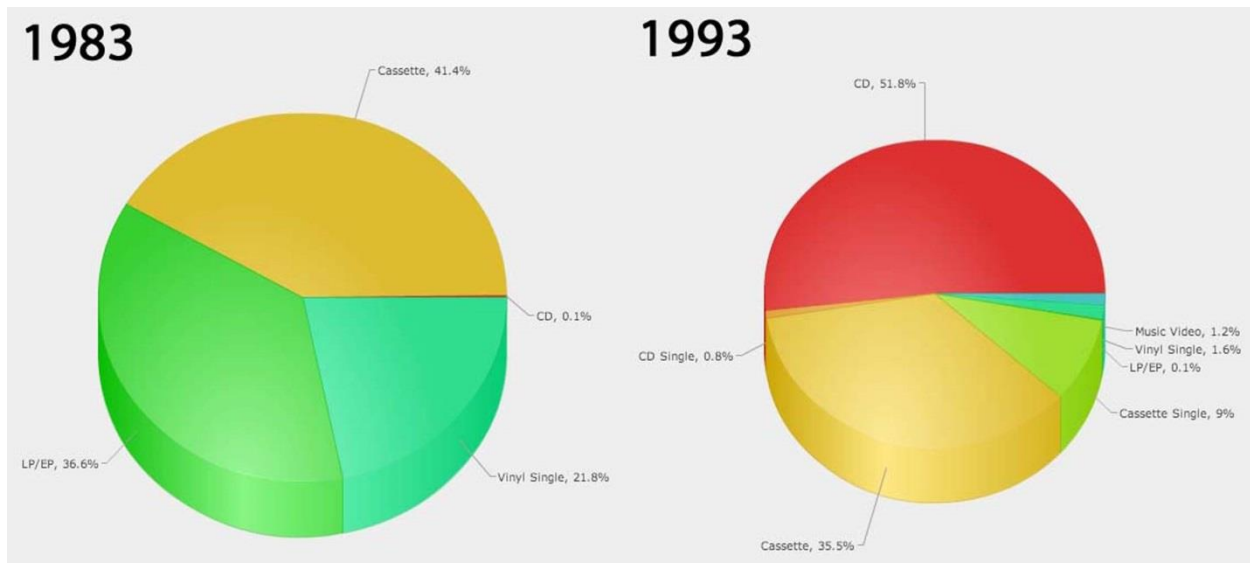
The Recording Industry Association of America (RIAA) has a web form interface for their Shipment Database (https://www.riaa.com/keystatistics.php?content_selector=riaa-shipment-database-log-in) that allows a user to generate static bar and pie charts of the data. We used this initially to look at the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

RIAA Stacked Bar Charts (Units vs. Dollars)

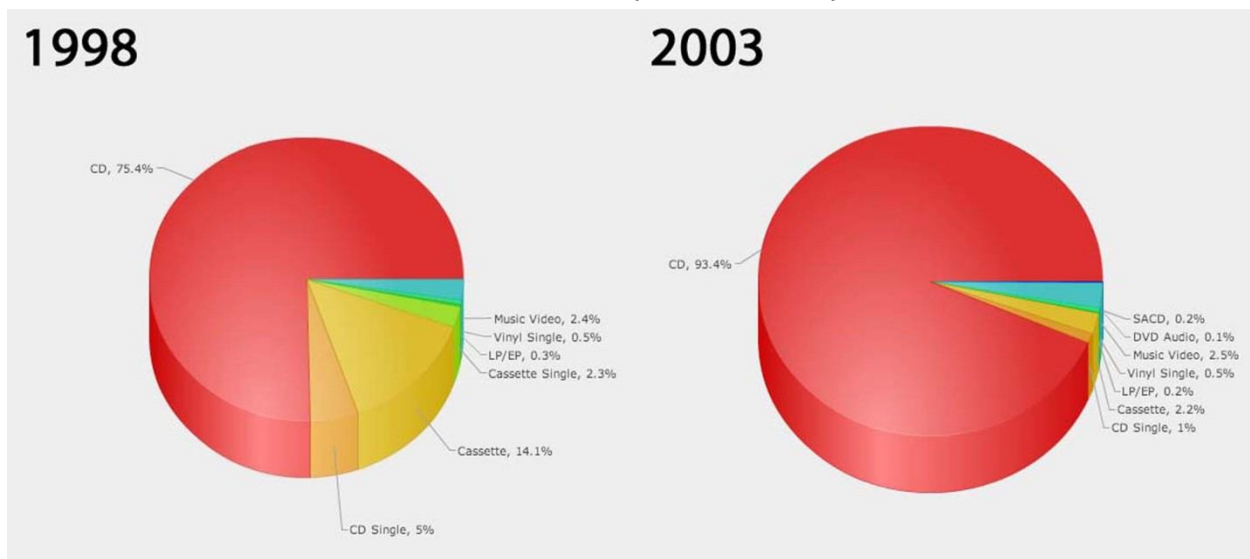


This gives insight into the overall dominance of the CD format. In recent years, however, the CD format has been decreasing in sales while the digital download formats have been increasing. Moreover, it appears the digital download format is currently dominating in terms of units, but not as much in terms of dollar amounts. This insight informed are design by leading us to wonder how price per unit has changed over time.

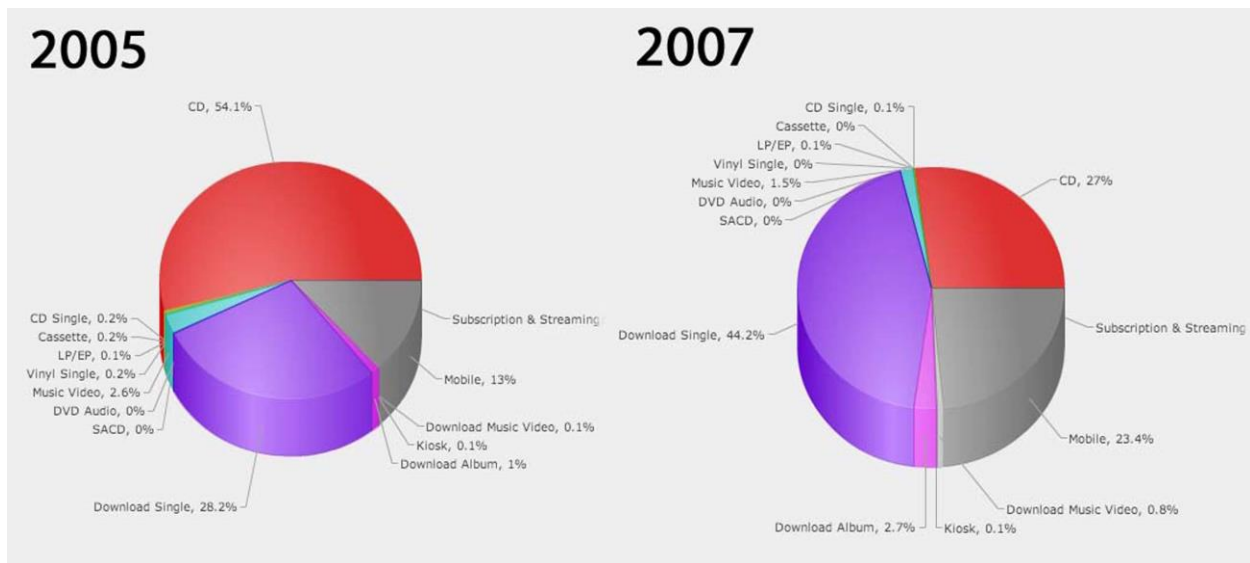
RIAA Pie Chart (1983 vs. 1993)



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 considered using dual y-axes to accomplish this. In addition, 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 'media' (physical, digital, streaming).

IMPLEMENTATION

EVALUATION
