

## *BACKGROUND AND MOTIVATION*

---

We plan to develop an interactive visualization of 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 past.

## *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 over different periods of time
- 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 between music format types (e.g. physical, digital, streaming)
- being able to compare music sale dollar values accounting and not accounting for inflation
- being able to compare music sales of different time periods

## *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](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 values sold 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 values data include:

- the aggregate amounts over a specified time period
- the aggregate amounts of physical and digital format types
- the differences in amounts between two specified time periods
- the changes in amounts over each year

We plan to implement data processing by using the D3 CSV API to parse the CSV files and create an array of JavaScript object containing the following fields:

- formatName (CD, cassette, vinyl, download single, paid subscriptions, etc.)
- formatType (physical, digital, streaming)
- year
- metricType (units sold, dollar values sold, dollar values adjusted for inflation sold)
- metricValue

## ***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.

## ***MUST-HAVE FEATURES***

---

The must-have features of the visualization include:

- toggling the displayed sale metric data (units, dollar values, dollar values adjusted for inflation) through selection
- ranking the sales of each music format through sorting
- focusing on the sales of a subset of music formats through filtering
- focusing on the sales of a specified time period through brushing and context zooming
- viewing the changes in sales for each music format over time through sliding of the brushed selection

## ***OPTIONAL FEATURES***

---

The optional features of the visualization include:

- toggling between displaying raw sales data and derived data, which is the change in sales over each year either by percentage or actual value
- toggling between a line chart and stacked area chart
- comparing sales between different time periods through multiple brush selections
- adding a narrative storytelling component

## *PROJECT SCHEDULE*

---

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

### **Weekly Objectives**

#### **By 4/10/2015**

- dataset parsed and processed into JavaScript objects [Jason]
- initial HTML layout file with chart control elements and stubbed out placeholders for each chart []
- initial main line chart displaying the number of units sold for each music format from 1973 to 2014 [Jason]
- initial bar chart displaying the total units sold from 1973 to 2014 for each music format sorted in descending order []
- initial outline version of the Process Book []

#### **By 4/17/2015 (MILESTONE 1 DUE)**

- the main line chart displaying either units sold, dollar values sold, or dollar values adjusted for inflation sold through a toggle control [Jason]
- the main line chart displaying a subset of music formats through filter controls [Jason]
- the bar chart displaying a subset of music formats through filter controls []
- HTML layout file with chart controls bound and line & bar charts added []
- updated version of the Process Book []

#### **By 4/24/2015 (PROJECT REVIEW WITH TF)**

- an overview chart with the main line chart focusing on the sales of a specified time period through brushing and context zooming [Jason]
- the bar chart displaying the total sales of each music format for the time period specified by the brush selection []
- the main line chart displaying the aggregate values of each format type (physical, digital, streaming) through aggregation controls [Jason]
- the bar chart displaying the aggregate values of each format type (physical, digital, streaming) through aggregation controls []

**By 5/5/2015 (FINAL PROJECT DUE)**

- the main chart displaying a line or stacked area chart through toggle control [Jason]
- the main chart displaying raw sales data or derived data through toggle control []
- the overview chart allowing for a second brush selection []
- another bar chart displaying the differences in sales between two brush selections
- the final version of the Process Book []