

CS3300 Project 2 Report

Net Ids: smh367, js3548, jrf268

Names: Sawyer Huang, Joanna Saikali, Joseph Fuentes

The Data

A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.

Our first data source consists of Billboard artists, songs and their hometowns, found from an online source and downloaded via Tableau.¹ There are many features in this dataset, but the ones of interest for our project were Artist Name, Song, City, Country, State, Genre, and Year. We used a python script to clean and aggregate the data into the format desired - one "hometowns" dataset with one unique row per artist and their hometown information, and one "billboard hits" dataset with one row per unique song that made the Billboard Top 100 per year. We also aggregated the Genres into 4 categories - pop, rap, country, and latin. The data prep script will be provided in our submission.

Our second data source was map-related. We used a dataset as seen in class that maps state names to state IDs, simply for the state abbreviations to map to state names.

Our final data source used was a Topology dataset for the United States, as provided in class, which was used for our topojson setup. We used this dataset to draw state boundaries. This allowed us to generate the map and color-code it based on how many Billboard hits were in each state.

Our python script used to clean and aggregate data is provided in the zip file.

Design

An overview of your visual design rationale. A good rule of thumb to follow is "every pixel must be justified." Instead of a 100,000-element breakdown, give us an overview of the design decisions you made and the trade-offs inherent in how you displayed the data. This part ought to include a description of the mapping from data to visual elements. Describe marks and channels you employ such as position, color, or shape. Mention any transformations you performed, such as log scales.

The color theme for the page is blue. This is reflected in the color of the states and of the title. The organization of the page is laid out for all the elements to fit without the user having to scroll. We didn't add too many elements to allow the user to focus on the simple information displayed, and interact with the visualization to understand insights conveyed through the data.

¹ <https://thedataface.com/2015/10/culture/mapping-hometowns-billboard-artists>

For the map we decided to remove any lines that were distracting for the user. These include a border on the canvas and latitude/longitude lines. This was included in feedback received saying that the additional lines made the visualization look too cluttered.

For our color scale, we used a log sequential scale. This is because the data is heavily skewed towards a handful of states (California, New York, Tennessee) that many top artists are from. With a log scale we are better able to see a spectrum of color indicating how many songs are attributed to various home states. For the coloring itself, we decided to use “d3.interpolateYlGnBu.” We chose a color scale with a non-white starting color so that it would be easy to distinguish between states with 0 billboard hits and states with a small number of hits.

Interactive Elements:

An overview of your interactive elements and their design rationale. Give us an outline of the design decisions that went into the interaction affordances you added to your visualization. What process did you use to choose the interactions you developed? How did you make them discoverable, usable, and interesting?

For our year scale, we decided to leverage a simple slider package as found in an online resource.² We considered using sliders as seen in class but felt that this was more visually appealing as tick marks are in the background and the user can easily see which year was selected.

We also allow the user to pick from 4 genres of music - rap, pop, country and latin - using checkboxes. The map updates according to what is selected. All four are selected by default.

We also decided to print the top 5 states and the number of billboard hits that came from that homestate. This updates as the user changes the years & genres, and allows a user to see a higher-level trend that may be more difficult to perceive via color alone.

The Story

What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

Our visualization relays the states with the most successful artists, according to the Billboard charts. The user can view the variability of success of states between the years 2000 and 2015. The main takeaway from our project is that several states are consistent winners for certain genres; scrolling the slider back and forth reveals little change in the dark blue of these states. There are 8 states with no artists on the Billboard in these 15 years.

Contributions:

At the end of your PDF file, include an outline of team contributions to the project. Identify how work was broken down in the group and explain each group member's contributions to the project. Give a rough breakdown of how much time you spent developing and which parts of the project took the most time.

² <https://bl.ocks.org/johnwalley/e1d256b81e51da68f7feb632a53c3518>

Joanna Saikali

- Data extraction, cleaning, and pre-processing (4 hours)
- Setup of the html file, creating the map, reading in relevant files, initial color coding based on number of hits (2 hours)
- Year slider functionality (2 hours)
- This writeup (1 hour)

Sawyer Huang

- Genre filtering
- Displaying top 5 states
- Hover-over map interactivity

Joseph Fuentes

- Overall design of visualization (3 hours)
- The story and part of design section of writeup (1 hour)