MusicMeter
A View of Music through its Instruments

Info 247 Final Project Report

Cristian Garay   Deepa Kalpathi   Pratibha Rathore

Introduction

Welcome to MusicMeter - a view of music through its instruments. For the final project we wanted to choose a domain that would be fun to explore (for the end user) and have a universal appeal and so, we opted for Music.

Initial exploration of visualizations of music data showed that there had been a lot of interesting data visualizations done on different aspects of music such as artists, tracks, collaboration between different band of artists etc. However, we did not find any visualizations on the instruments the artists used. Hence, our team decided to explore the music domain with particular focus on instruments played by artists.

The data set for all our analysis is taken entirely from www.musicbrainz.org. MusicBrainz is a community-maintained open source encyclopedia of music information. All the information about music from artists to their releases and work is stored in MusicBrainz database. MusicBrainz is a peer-produced music metadatabase and users from around the world develop and distribute musical knowledge through the community website. The data in the database is placed in public domain and is available to anyone to download and use for analytics, visualization or creating an application.

Project Goals:

Early user research (explained in a later section) helped us fine tune our project goals and determine the visualizations to be built. Based on this, following were our project objectives:
1. Explore a little bit of history about musical instruments and create a narrative
   a. Research on different categories of musical instruments
   b. Who and when were some of the famous instruments invented

2. Overview of types of instruments (or families) across all the recordings
   a. Number of instruments per instrument type
   b. Spread of types of instruments across all the recordings
   c. Use of instruments over decades
   d. Most commonly used instruments across all the recordings

3. Explore the use of instruments based on gender
   a. Is there a difference in the type of instrument played by the artist based on the gender of the artist?
   b. How does the above ratio compare to vocals?
   c. Is there a difference in this ratio across countries?

4. Visualize which instruments are played together

5. Visualize geographical spread of different musical instruments

6. Overall learn interesting insights about musical instruments and their role in the music industry.

Data

The data set we used for our project can be found at www.musicbrainz.org (Link to the database here). MusicBrainz is an open music encyclopedia that collects music metadata and makes it available to the public. As mentioned on their website, MusicBrainz aims to be the ultimate source of all music information, which is both reliable and unambiguous. The entire MusicBrainz database has information about 1M+ artists, 20M+ songs, 761 types of instruments (includes all roles played by artists, including vocals) and in 70 different formats.

Accessing the Data:
The database is available as a PostgreSQL database, and we accessed the database using a virtual server in Virtualbox which was running on Ubuntu and a PostgreSql server. We faced several issues during the setup and data access phase. However, the moderated community was a great help. Following are some screenshots of questions posed to the community.

MB virtual server connection to Tableau
■ MusicBrainz

cgaray — 21d

Hi!

I have downloaded and installed the virtual image of MusicBrainz server on VirtualBox without much problem, but now I'm trying to connect Tableau to this database to visualize this data, and that's not working.

The connection on Tableau is a standard connection to a PostgreSQL database (it gives you the option of that particular db so drivers should no be an issue), and I have use this tutorial to open connections on the virtual machine: http://www.cyberciti.biz/tips/postgres-allow-remote-access-tcp-connection.html.

But no connection, do anyone has been able to do this or something similar?

Thanks in advance!

Reply

| created | last reply | 5 | 91 | 3 | 2 | 2 | |
| C 21d | C 14d | replies | views | users | likes | links | |

How to see instruments in a track? ✎
■ MusicBrainz

dk29 — 13d

Hi,

I am trying to understand what instruments have been used in a song (release) for an analysis. My over arching objective is to study the evolution of different types of instruments over the years. How do I go about this?

What I have done:
I have tried to connect l_instrument_release along with instrument as entity0 and release as entity1, along with link, link attribute, link attribute type and link type (as specified in the schema). However, since l_instrument_release is empty, I get no output.

Could someone help me out with this?

Reply

| created | last reply | 5 | 72 | 3 | 3 | 1 | |
| D 13d | 12d | replies | views | users | likes | link | |

Exploratory Data Analysis

Post accessing the database through the virtual machine, we connected it to Tableau, for exploratory data analysis. Since the database had innumerable tables and multiple links between the tables, we explored subsets of data at a time. Following is the schema of the database.
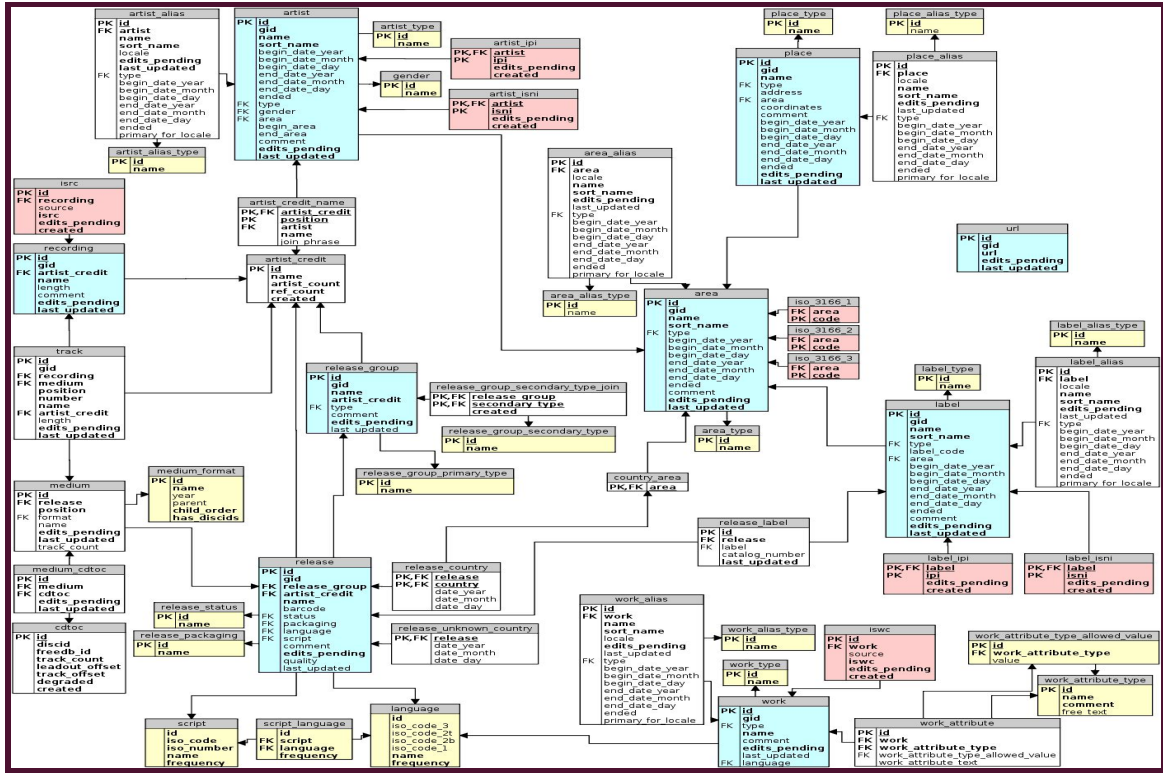
Challenges with respect to data:

Some of the challenges we faced were:

a. The data on hand was huge around 30 GB with information for more than 20 million songs. The biggest challenge we faced was how to scope it down and what variables to choose which would make an interesting story. Even though we knew we wanted to use only a subset of information, we could not create a subset of the database since we were accessing it using a virtual machine. So, we were limited to exploring a subset of the dataset. This resulted in long query times for every execution in tableau / pgAdmin3 (postgresql tool).

b. The schema and the variables were complex to understand and implement. The entire documentation had to be read to understand each entity and how they were related to others. For instance, the table l_artist_instrument is a link that connects artist table and instrument table and it needed 6 other tables to be linked to it, to get all the information (artist, instrument, link, link_type_attribute_type, l_attribute, l_attribute_type).

c. A deep understanding of the music metadata was required. For example, in one of the visualizations we had to aggregate the number of recordings at country level and we used *track_count* attribute in the *release_country* entity as the count of all the

recording for a country. However, the resulting visualization did not make sense and later we found that *track_count* was actually the track position and not count of tracks released in a country.

Data processing/Cleaning

Some of the pre processing is listed below:
1) Removing incorrect data points - Some of the fields had incorrect information and these had to be filtered out. For example: Under year of release, some of the years were 2, 200, 2018, 2088, etc. Clearly, these were incorrect entries and hence filtered out.
2) Creating groups - The database had individual information regarding each track. We had to create groups and add further comparisons to it at our end. For example, the instrument column only mentioned the name of the instrument, but the schema explanation had the type of instrument family it belonged to. We had to create these groups ourselves.
3) Categorical data - For some of the tables, especially with categorical data, our approach to cleaning the data was to go through the variable list and filter it manually. For instance removing *'guest'* from instruments table (shown below).

Tools

- VirtualBox - To access the virtual machine
- PgAdmin3 - PostgreSQL tool
- Tableau - Data analysis
- D3 - Visualization
- Illustrator - Infographic and icons visualization
- Excel - Working with some subsets for visualizations / exploration

Related work

We came across some of the works done using the same dataset. Although these were mostly related to artist, tracks, music collaboration, they helped in streamlining our focus.
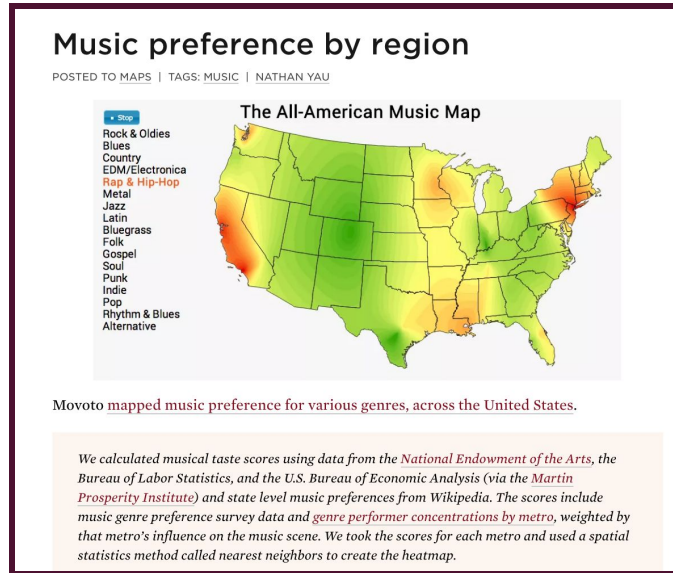
1. Flared music is a project by BBC which lets us visually research relationships between musicians and bands using the Musicbrainz database. The project uses a tree structure to display the results of search terms that you enter into it as a kind of musical family tree. For example submitting The Rolling Stones will grow a tree starting from the initial search term to show the individual band members such as Mick Jagger and Ronnie Wood and collaborators such as Billy Preston. We incorporated a similar idea in our of instrument collaboration visualization.
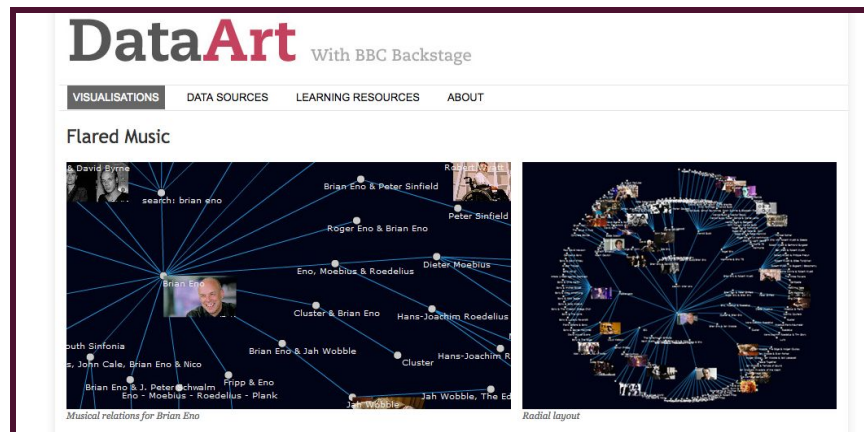

http://data-art.net/flared_music/

2. In the below graph, Flowing data uses a geographical heat map to plot music genre preferences across different states in United States, using music taste scores. We

used a similar idea for creating our geographical spread visualization, where in instead of music score we have used number of recordings released in a country with information about instruments overlaying on the map.


Movoto mapped music preference for various genres, across the United States.

We calculated musical taste scores using data from the National Endowment of the Arts, the Bureau of Labor Statistics, and the U.S. Bureau of Economic Analysis (via the Martin Prosperity Institute) and state level music preferences from Wikipedia. The scores include music genre preference survey data and genre performer concentrations by metro, weighted by that metro's influence on the music scene. We took the scores for each metro and used a spatial statistics method called nearest neighbors to create the heatmap.

3. We really liked the below Los angeles menu visualization from last year's Menu journeys project. The chord diagram that shows the relationships between two dishes -- the chords are thicker when the dishes occur more frequently on the same menu. This is similar to what we wanted to do for creating a visualization which can show which instrument are played together across all the recordings. This chart also offers greater interactivity and allows for user to play around with.

http://youarehere.cc/j/menu/losAngeles.html
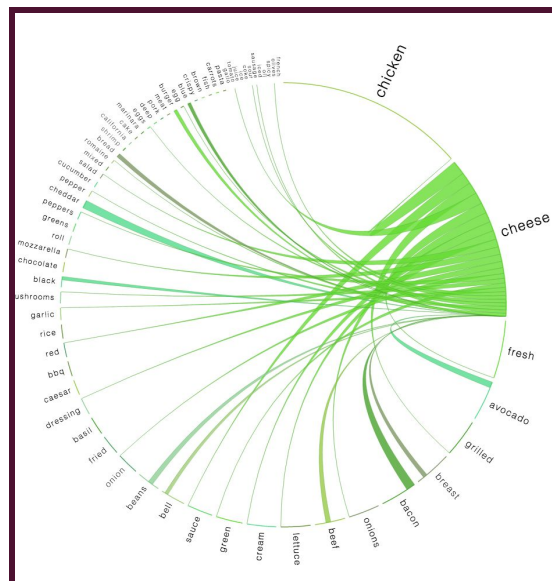
4. The below visualization shows history of music in UK. This chart combines elements for narrative infographic as well as bar charts in Tableau. We have used similar idea when creating visualization for Instruments by gender and overview of instruments across recordings.


https://public.tableau.com/s/gallery/history-music-1s-uk
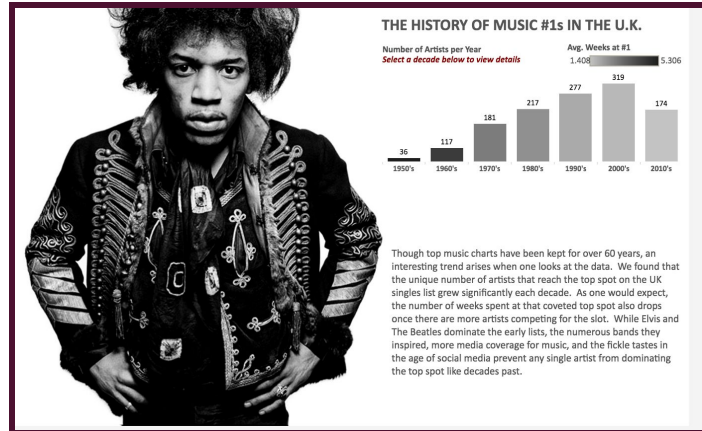
5. The below visualization is created using Tableau and describes how much did each individual member of the Beatles band contribute to the writing of each album. We used a very similar approach when creating visualization for Instruments by gender and overview of instruments across recordings.

https://public.tableau.com/s/gallery/beatles-albums
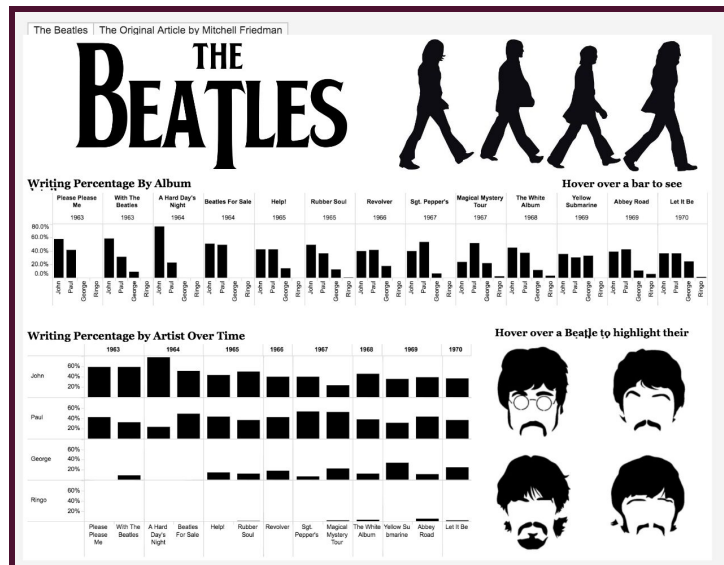
6. The below visualization is created using Tableau and shows how the prevalence of diabetes is geographically distributed. We used similar approach for creating our visualization that shows prevalence of different types of instruments across all the countries.
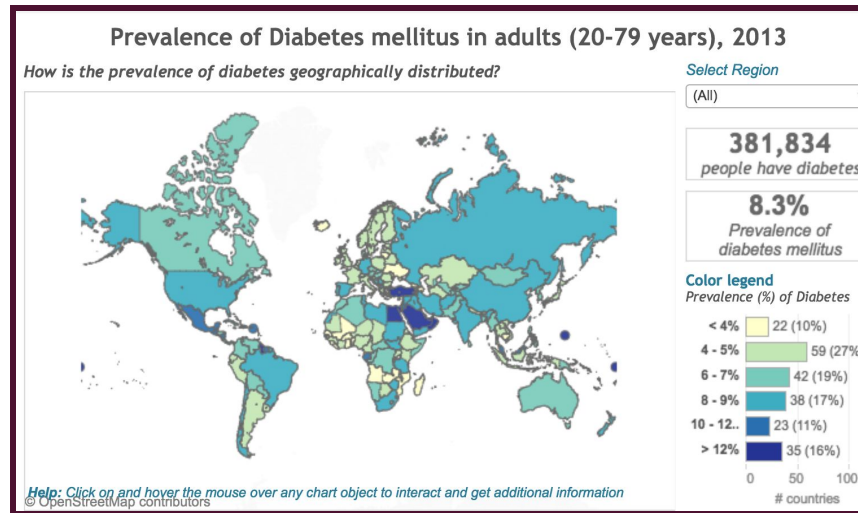

7. The visualization below is a popular one with music data. It showcases songs by genres and sub-genres. We explored using this visualization for families of musical instruments and the instruments within them. However, since there are not as many levels as in musical genres, it would have been a very flat visualization with only the family name and the instrument name under it.
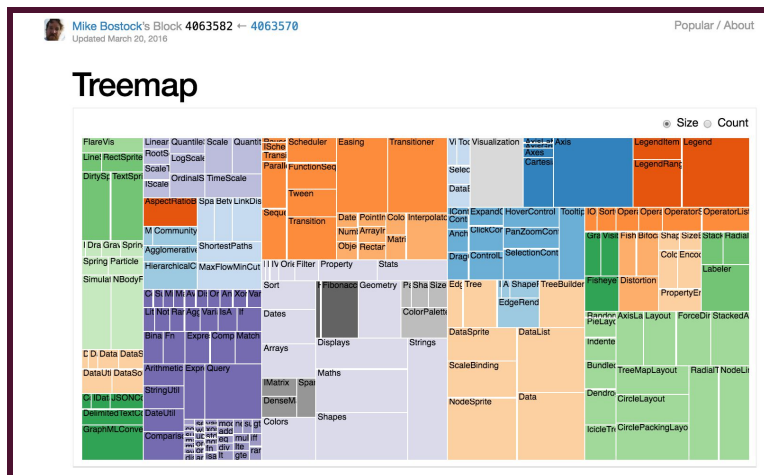

Source: http://research.google.com/bigpicture/music/
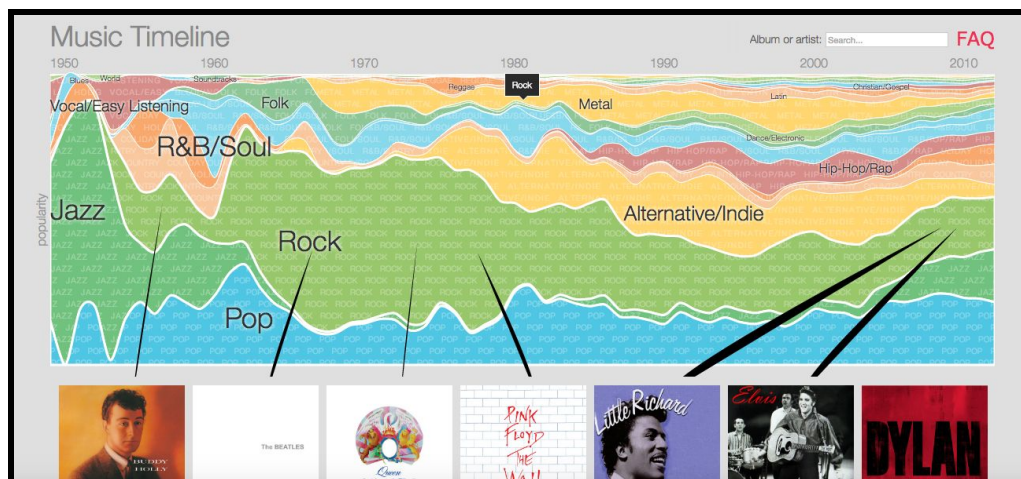
8. The below chart uses chord diagram to visualize frequency of rides between different neighbourhoods in San Francisco. The thickness of the curves explains the frequency of rides between two neighbourhoods. This is similar to what we wanted to do for creating a visualization which can show which instrument are played together across all the recordings. This chart also offers greater interactivity and allows for user to play around with.

https://bost.ocks.org/mike/uberdata/
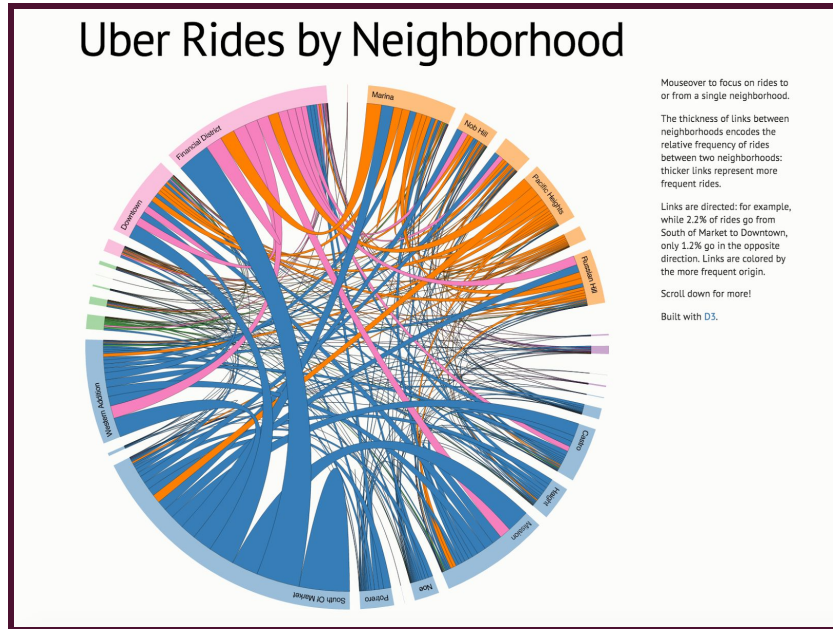
9. We also explored the possibility of using below Tree map from Mike Bostock's collection to show for each of the country, how many recordings were released and for each of the recordings what does the distribution of instruments use looks like. However, later we decide to use choropleth map as it provided a simpler and effective way of showing geographical distribution.

Our work

The website for our final project is
http://people.ischool.berkeley.edu/~cgaray/MusicMeter/

Github:
The code can be accessed on github at this link:
https://github.com/cagaray/MusicMeter.git

User research:
We knew that we are dealing with a massive data set and it was very important for us to not only scale down the data we should be using for our analysis, but also scale down the focus considering different aspects of the music. During our mid project presentation we got feedback from Prof. Marti and TAs that in order to choose a story line for our project with MusicBrainz data set, we should do a user research to get a sense of which variables or attributes might be interesting to the users.

Before going to users, we first needed to understand what the scope of the visualizations would be. Since visualizations with music had not been done, we chose this as the broad theme and decided to get user feedback on what aspects of instrument visualization would be interesting for users.
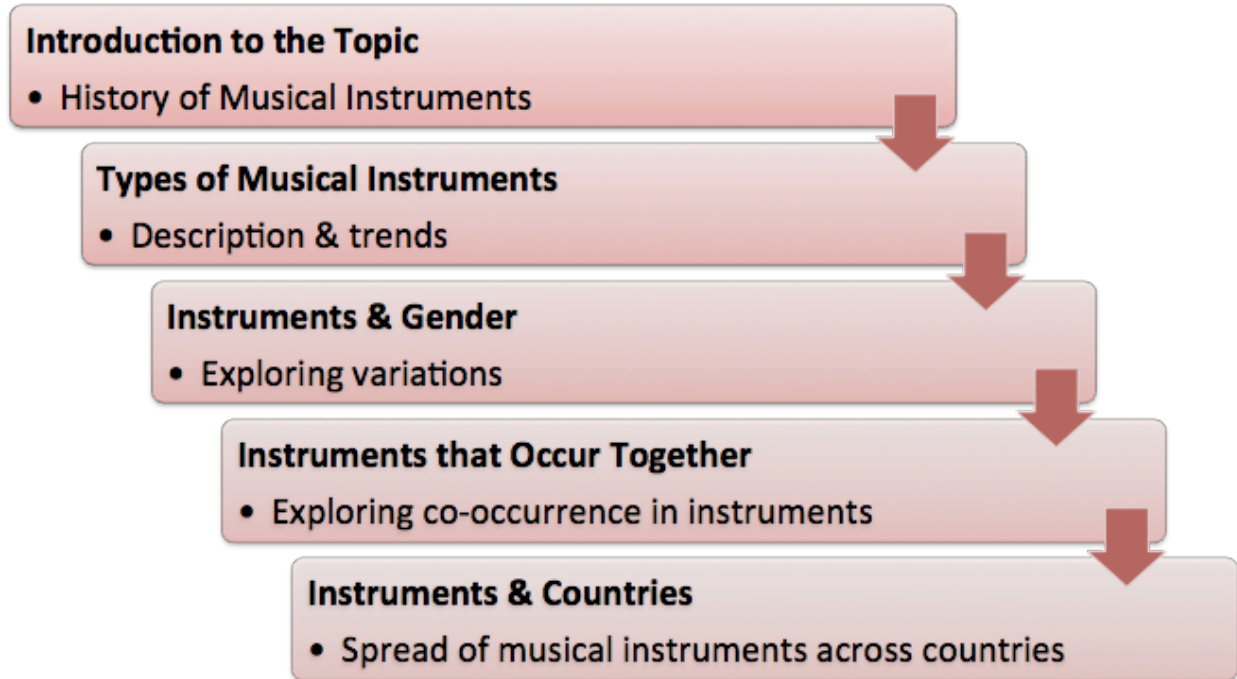
We did qualitative user research with 5 users and asked for feedback on possible routes to take with respect to music visualization. The broad routes that we had identified based on data exploration were instrument - gender, instrument - medium of release, artist playing the instrument's origin country, instrument - label of the track, instrument - country of release, multiple role of artists, types of instruments and instrument families and instruments that go together.

Of these, users were keen on exploring instrument - gender, types of instruments, instruments and countries and instruments that are played together. Also among the chosen options was multiple roles of artists (For example: An artist is a lead vocal in one track and plays the drums in another). However, since this would take the focus away from instruments, we decided to not go ahead with this visualization.

Hence, the four routes described above were chosen and visualizations were built on these.

Visualizations:

The overall flow of visualization is given below.

Introduction to the Topic
- History of Musical Instruments

Types of Musical Instruments
- Description & trends

Instruments & Gender
- Exploring variations

Instruments that Occur Together
- Exploring co-occurrence in instruments

Instruments & Countries
- Spread of musical instruments across countries

Each of these sections is explained in detail below.

An Understanding of Musical Instruments

First, we created a narrative infographic embedded in the web page which tells a brief history about different types and categories of musical instruments. We have included information about instruments such as who was the inventor, when was it invented etc. This information was important as a part of our story to build the context as well as give our audience a general sense of how these instruments evolved over time.

1830
Flute

It produces sound when a stream of air is directed against the edge of a hole, causing the air within the body of the instrument to vibrate. It was invented by the German instrument maker Theobald Bohm.

1850
Guitar

Guitar has a flat-backed rounded body that narrows in the middle, a long fretted neck, and usually six strings, played by strumming or plucking. The modern classical guitar is credited to Antonio Torres.

1909
Drum set

William F. Ludwig and his brother Theobald made history, revolutionizing drumming with the invention of the first spring driven pedal.

Overview of Instrument type across recordings

TYPES OF INSTRUMENTS

Musical instruments are grouped into different types primarily based on how sound is produced. However, not every instrument fits neatly in a group, as some play a dual role.

For example, a piano has strings that vibrate, and hammers that strike. Is it a string instrument or a percussion instrument? Some say it is both!

Wind

Musical instruments in which sound is produced by the vibration of air, typically by the player blowing into the instrument.
The notes are produced by manipulatng the air columns.
Broadly classified into brass and wind instruments based on how sound is produced rather than the material.

String

Musical instruments that produce sound from vibrating strings.
Main methods of playing - plucking, bowing, striking.
Broadly classified into 3 groups - lutes (support strings via a neck), harps (contain strings within a frame), zithers(strings mounted on a body).

Percussion

Musical instruments that are sounded by being struck or scraped by a beater / hand / another instrument.
Produce sound through vibration of entire body or when membrane is struck.
Broadly classified into pitched percussion instruments & unpitched percussion instruments.

Electronic

Musical instruments that produces sound using electronics.
The instrument might include a UI to control pitch, frequency & duration of each note. However, it is also common to separate it out into music controller & music synthesizer

Other

A class of various other devices that have been used as instruments in muscial compositions.
Some examples are vacuum cleaner and chainsaw

We start this section with a small explanation of the various types / families of musical instruments. Then, we move on to the data analysis done on the basis of these families of musical instruments.

There are 548 musical instruments captured in this data (The remaining 200 or so were classified as vocals, additional, guest performance, etc.), across 2.1 million recordings. As can be seen, the Strings family has the most number of instruments at 195. Even when we look at the number of recordings and the instruments used in them, we see that strings dominate. This is a natural result of their being more string instruments.

We wanted to explore the use of musical instruments across the decades. We considered data from 1950 upto 2015. We see that the trend across families of musical instruments is similar - more and more musical instruments are being used in recordings over the years. One interesting trend is that in the current decade, the number of musical instruments

being used is almost as much as the previous decade (with only 6 years of data). There could be two explanations for this. One, that the cost of releasing music has drastically gone down given the ease of streaming music online. With the dissolution of this barrier of entry to the music industry, more artists and using instruments in different ways to create music. The other explanation could be that data from the previous decades have not been captured in the database. We think both explanations contribute to this trend.

In analyzing the distribution of different musical instruments, we see that the distribution is a long tail. The top 25 instruments, which account for about 4.5% of all instruments, contribute to almost 50% of all recordings. Some of the more unusual instruments used are a vacuum cleaner and a chainsaw.
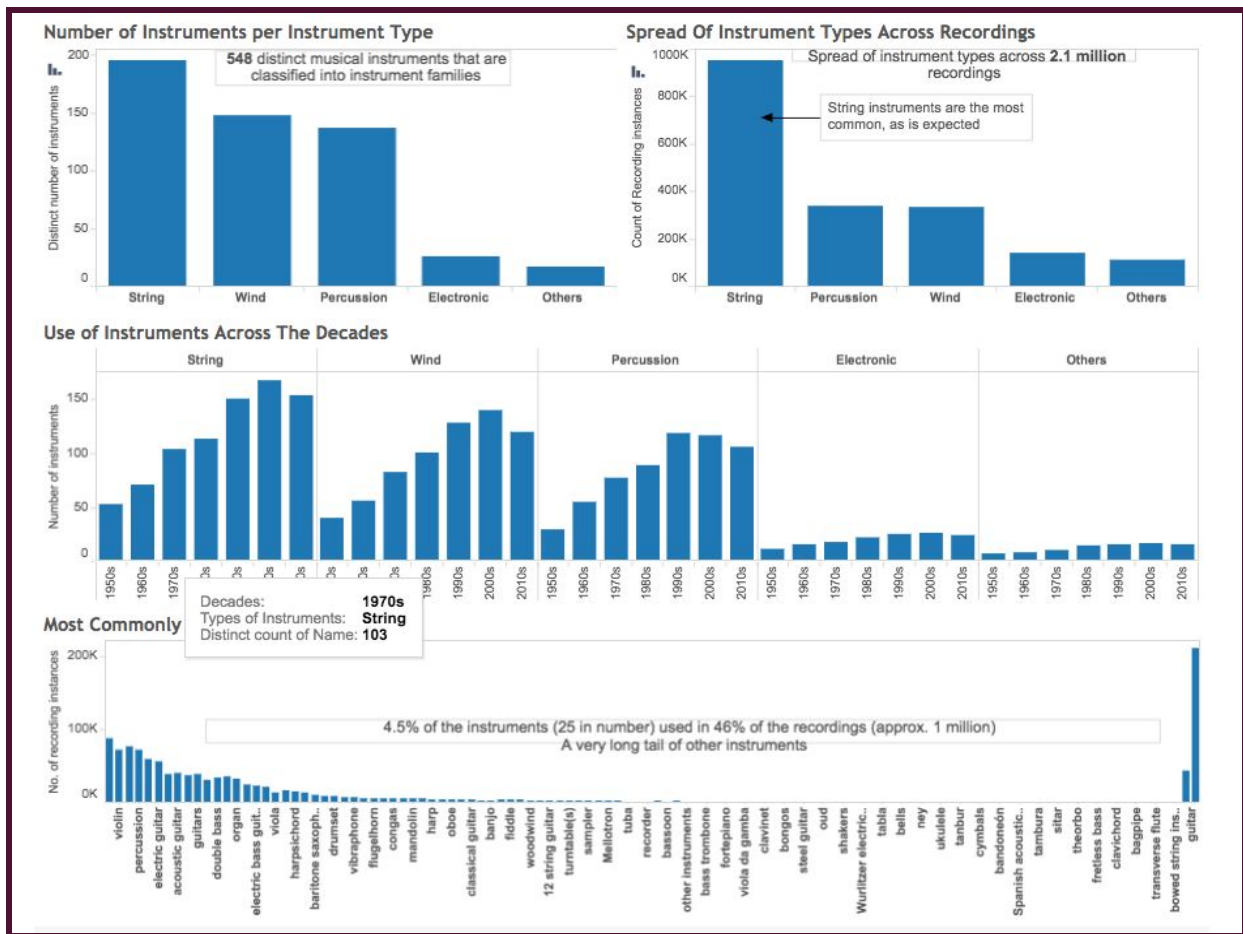
Instruments by Gender

In many societies, musical roles are divided along gender lines: women sing and men play. The dashboard below showcases the marginality of representation of women in the field of musical instruments. String instruments have the most representation at 10%. In contrast, women make up 30% of the vocalists, lending further credence to the claim "women sing, men play".
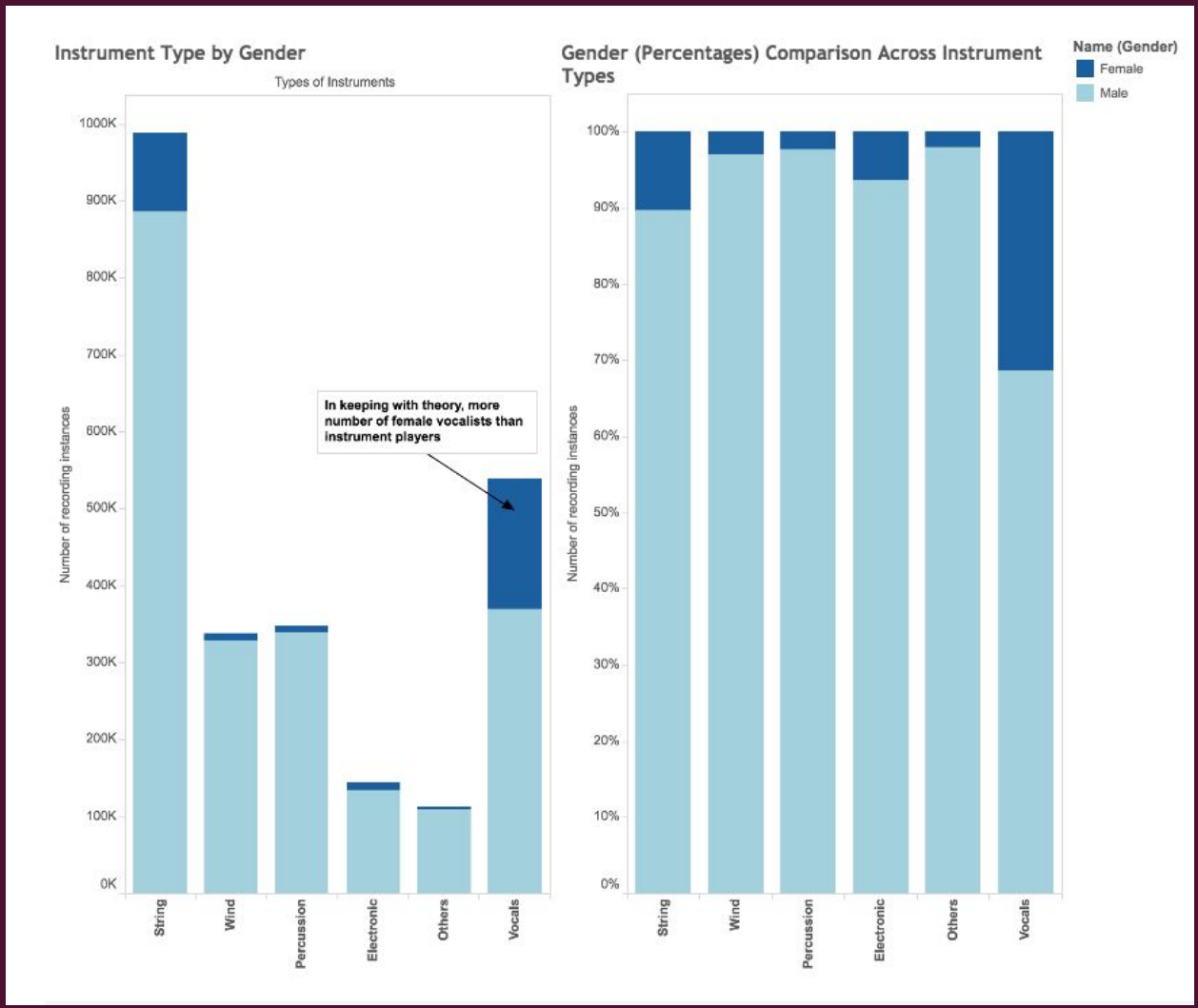
When comparing countries on the percentage of women playing musical instruments, some interesting insights can be seen. For instance, in many ways, equality for women in the developed world has been more than in the developing one and hence it could be expected that more women in a developed country would play musical instruments than in a developing one.

However, if we consider an example of US and India; while representation of women is marginal in both, a larger percentage of women in India play string instruments. While on the surface this seems to defy our expectation, this trend is probably due to the close association of certain string instruments and vocal music, thereby inadvertently supporting the claim that "women sing, men play".
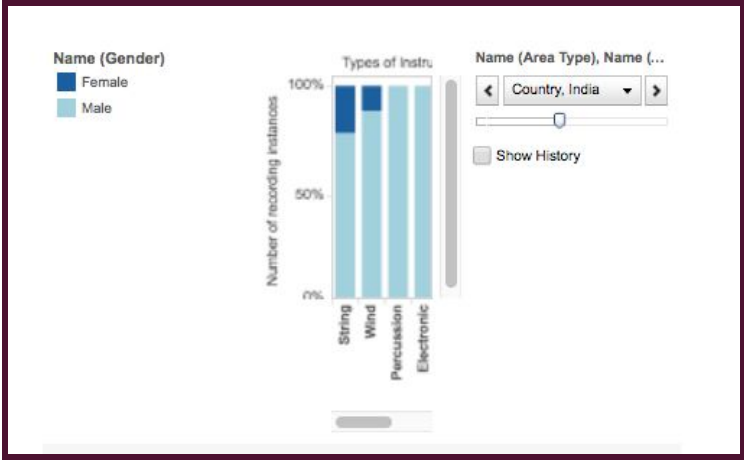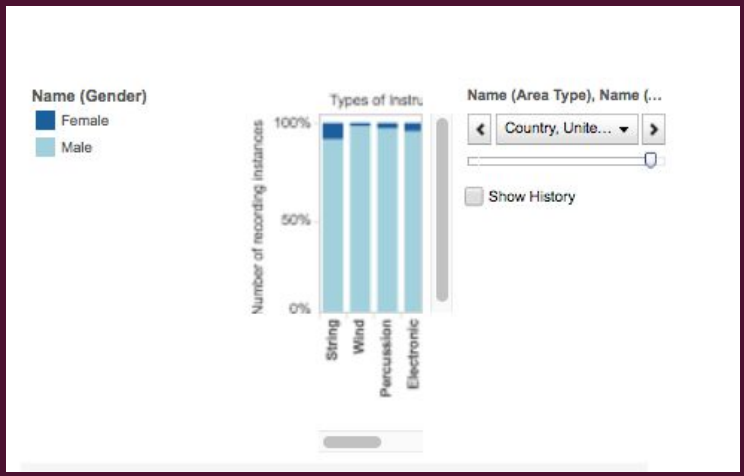
Instruments that play together sound good together

Be it a few friends jamming in the basement or a symphonic orchestra playing, musical instruments are often played together, with each musical instrument family complementing the other.

To understand trends among the top 25 musical instruments, we looked at the frequency of occurrence of each with the others. The visualization below is dense since it takes into account only the top 25 instruments. However, within these, we can see that some instruments are played more with others. For example, keyboard and acoustic guitar. Also, among the top 25 musical instruments, there are five categories of guitars. If these were to be combined, then there would be a stronger connection to keyboards.
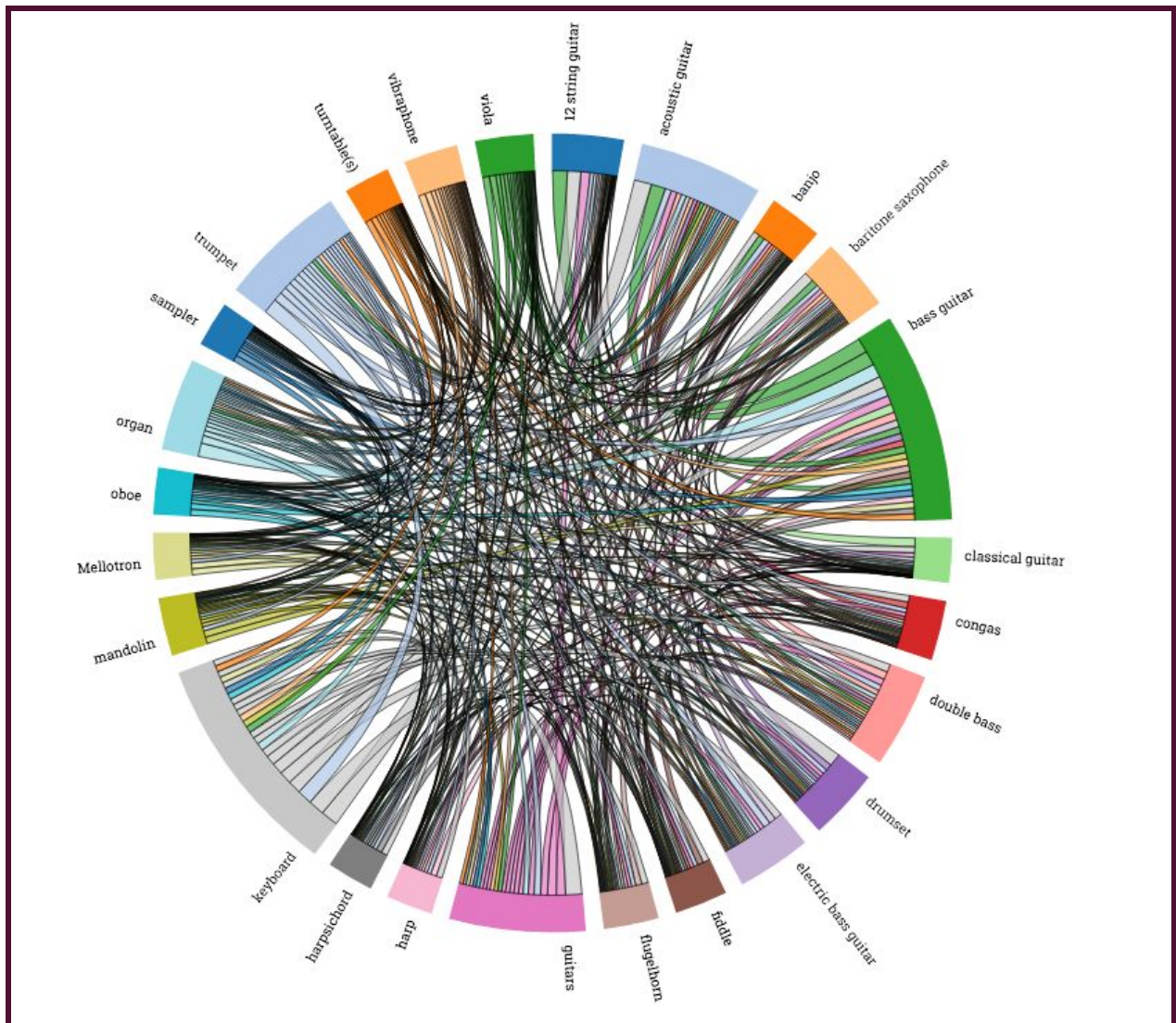
Geographical spread of musical instruments

We wanted to see the prevalence of different instruments in different countries around the world. The sql script to get the data for this visualization is attached in the project folder. We streamlined the query to fetch information only after 1950 onwards. This visualization was created using tableau. Each instrument is color coded.
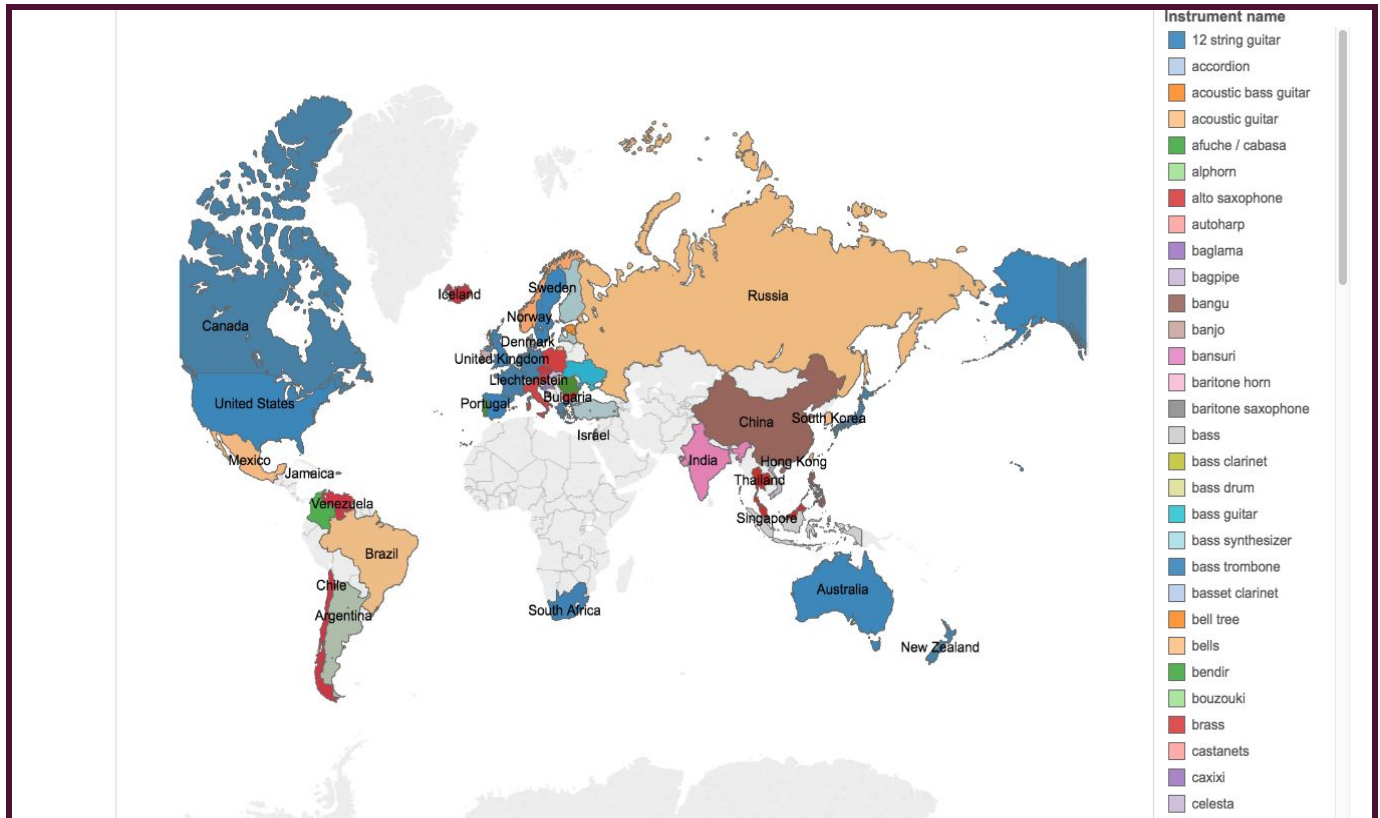
If we filter for example *'acoustic guitar'* from the instruments list, we see that it is prevalent in a larger geographical area especially Russia and Brazil. Saxophones are more prevalent in Italy, Poland and Bulgaria. The visualization also shows a local appeal of regional musical instruments, for example *'bansuri'* from India, *'bangu'* from China and *'samisen'* from Japan are limited to their places of origin.

We do not see any information about instruments in most parts of Africa and the Middle East, this is because we used a subset of the entire database for this visualization. We used did a random sampling of 10k records for every 100k records we retrieved from database. This was done around ten times and the above visualization was made using 100k records for it to be representative of entire MusicBrainz data we had. It is possible that music from this region did not get picked up.

Creating some Music

Having visualized the data as required, we decided to end the visualization with a fun visualization.

We explored a few different options - a memory game using musical instruments, a musical tone generator with the musical instrument that would show how each musical instrument sounded, or a piano visualization that would allow the user to create music.

We decided on using the piano visualization for this since it would afford the user more freedom to play with it. Created using D3, this produces sounds for each key pressed.
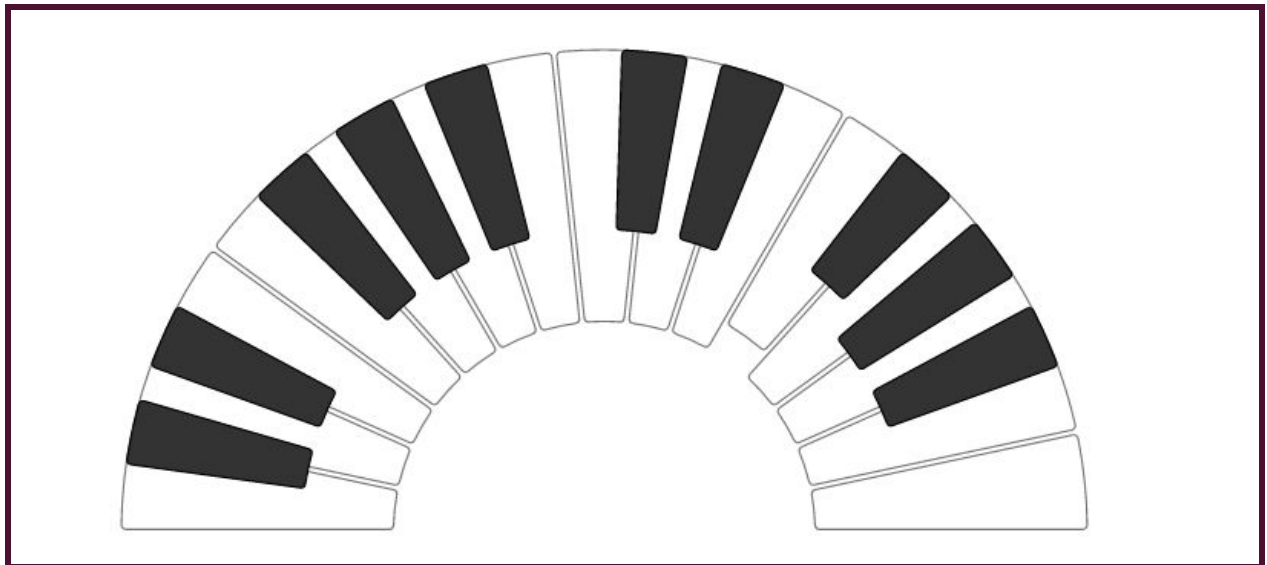
Usability Testing

Once the visualizations were made, we did usability testing with 5 users. We tested the design on various parameters.

- **Task Completion:** For each of visualizations, we had predetermined tasks or questions based on the data, that we got the user to complete. For example: How many instruments belong the percussion family?
- **Data Exploration:** We got the users to explore the data and come up with insights that were new to them. For example: One of the users compared instruments by gender - Sweden and Argentina, and saw that about 4.3% of string instruments as played by women in Sweden while the corresponding percentage in Argentina is about 34%. (Note: While we do not have an explanation for the increased percentage in Argentina, some of our research showed us that in Swedish folk music, women are not allowed to play certain instruments. 'Women don't play the fiddle, they play the horn' ([Source](#))- this could be a reason for the low percentage.)
- **Overall Look and Feel:** We initially tested using tabs for each visualization, but then opted for infinite scrolling based on the usability testing. Users felt this was an easier method to scroll. Another feedback we got was to create a distinction between the data exploration part and the fun part - the 'create music' part. Hence, we added a visual element to create this separation. We also added a black footer at the end of the webpage to signal that the page had ended. The users also enjoyed playing on the piano at the end.

Contribution of each team member:

All of us worked on finding a potential data set to work on for our final project. Deepa and Cristian picked up MusicBrainz data and Pratibha found immigration data for the city San Jose. As a team we decided to go for MusicBrainz data set, because of the possibility of having more options for playing with different variables and having multiple stories to explore. Below is the percentage contribution for each team member per task:

Project tasks	Deepa	Cristian	Pratibha
Finding data set for the final project	33%	33%	33%
Setting MusicBrainz data virtual machines and connecting to tableau	0%	100%	0%
Data exploration using Tableau	33%	33%	33%
Data cleaning and pre processiong for individual visualizations	33%	33%	33%
User Research	33%	33%	33%
Gathering Information about evolution and history of musical instruments	50%	0%	50%
Visualization: Overview of Instruments across recordings	100%	0%	0%
Visualization: Instruments by Gender	100%	0%	0%
Visualization: Instruments played together	0%	100%	0%
Visualization: Geographical spread of instruments across countries	0%	0%	100%
Usability testing	33%	33%	33%
Project Website and pushing visualizations to the site	20%	80%	0%
Final report	30%	10%	60%

References:

1. Hemerly, Jess. "Making metadata: The case of MusicBrainz." *Available at SSRN 1982823* (2011).
2. Bertin-Mahieux, Thierry, et al. "The million song dataset." *ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference, October 24-28, 2011, Miami, Florida*. University of Miami, 2011.
3. Roads, Curtis. "Early electronic music instruments: Time line 1899-1950."*Computer Music Journal* 20.3 (1996): 20-23.
4. Yim, Ji-Dong, Chris D. Shaw, and Lyn Bartram. "Musician Map: visualizing music collaborations over time." *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009.
5. Zhu, Jianjun, Xiangyang Xue, and Hong Lu. "Musical genre classification by instrumental features." *Int. Computer Music Conference, ICMC*. Vol. 2004. 2004.
6. https://blogsnatch.wordpress.com/2014/03/24/why-men-play-instruments-and-women-sin/
7. Hallam, S., Rogers, L., & Creech, A. (2008). Gender differences in musical instrument choice. *International Journal of Music Education*, *26*(1), 7-19.
8. O'Neill, S. A., & Boultona, M. J. (1996). Boys' and girls' preferences for musical instruments: A function of gender?. *Psychology of music*, *24*(2), 171-183.
9. O'Neill, S. A. (1997). Gender and music.
10. Delzell, J. K., & Leppla, D. A. (1992). Gender association of musical instruments and preferences of fourth-grade students for selected instruments. *Journal of research in music education*, *40*(2), 93-103.