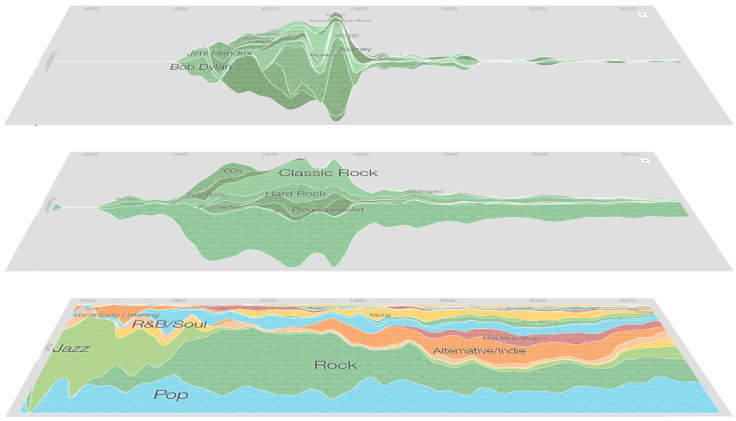

Music Timeline

Visual Analytics for Big Data
Visualization Method Theory Assignment

Eser Comak (s3432548)
Hichem Bouakaz (s2525763)

1.0
Groningen, February 7, 2019

Contents

List of Tables

List of Figures

Overview

Information Visualization is a valuable tool for analyzing data and discovering facts about the data that cannot be seen by just looking at the plain data, however, visualizing high dimensional data is not always an easy task. The purpose of this report is to evaluate the performance of Google Music Timeline visualization. The visualization can be accessed here [Mus] In the following sections, we will go over a brief introduction section section 1 in which we talk about the motivation behind using visualizations in general. Next, we will present the Google Music Timeline visualization in detail section 2 and then speculate about the potential user base in section section 3. It will be then followed by the data characteristic section section 4 in which we will present the variables. Then the discussion of problem domain section 5 will follow in which we inspect what kind of questions that the visualization tries to answer. Later we will evaluate how good Music Timeline Application is at answering these problems in section section 6. In section subsection 7.1, we will present the visual encoding strategies used within the visualization and then suggest potential improvements in subsection subsection 7.2. Finally, we will act as an end user and try to answer three questions using the visualization in section section 8.

1 Introduction

Over the past 70 years, humanity accumulated a gamut of musical creations and genres associated with them. Starting from the classical era to African roots of blues music. From the bloom of New Orleans jazz to 70s rock era. All leading up to the use of synthesizers in pop music that planted the seeds for dance music which has wide public appeal today. If you were to stop for a second and think about the number of genres and subgenres, bands, album tracks and most notable songs there are, you will immediately realize the vast amount of information that you are looking at. Sure all this information can be collected into an excel file with many features describing each observation of a musical entity, be it artist, genre, album, etc. Next, we can calculate descriptive statistics such as mean, range and quartile information and make sense of our data distribution. However, these calculations yield numbers, and number representations alone are limited in their ways of representing data for fast and intuitive understanding.

Luckily mankind has evolved to develop a fine visual instrument which has a good colour perception that made us very good at detecting edges in our visual field. Along with these capacities, our visual field is modulated by top-down influence that helps us to make sense of the visual input presented in front of us. All of these capacities can be exploited to enable high bandwidth transfer of information with the use of visuals. Infographics are the simplest example of this kind of uses. They convey large amounts of information by using different shaped geometrical shapes. By manipulating the size of these shapes sizes and colours, they provide quite intuitive snapshots of big data aggregated appropriately for human consumption(figure Figure 1)

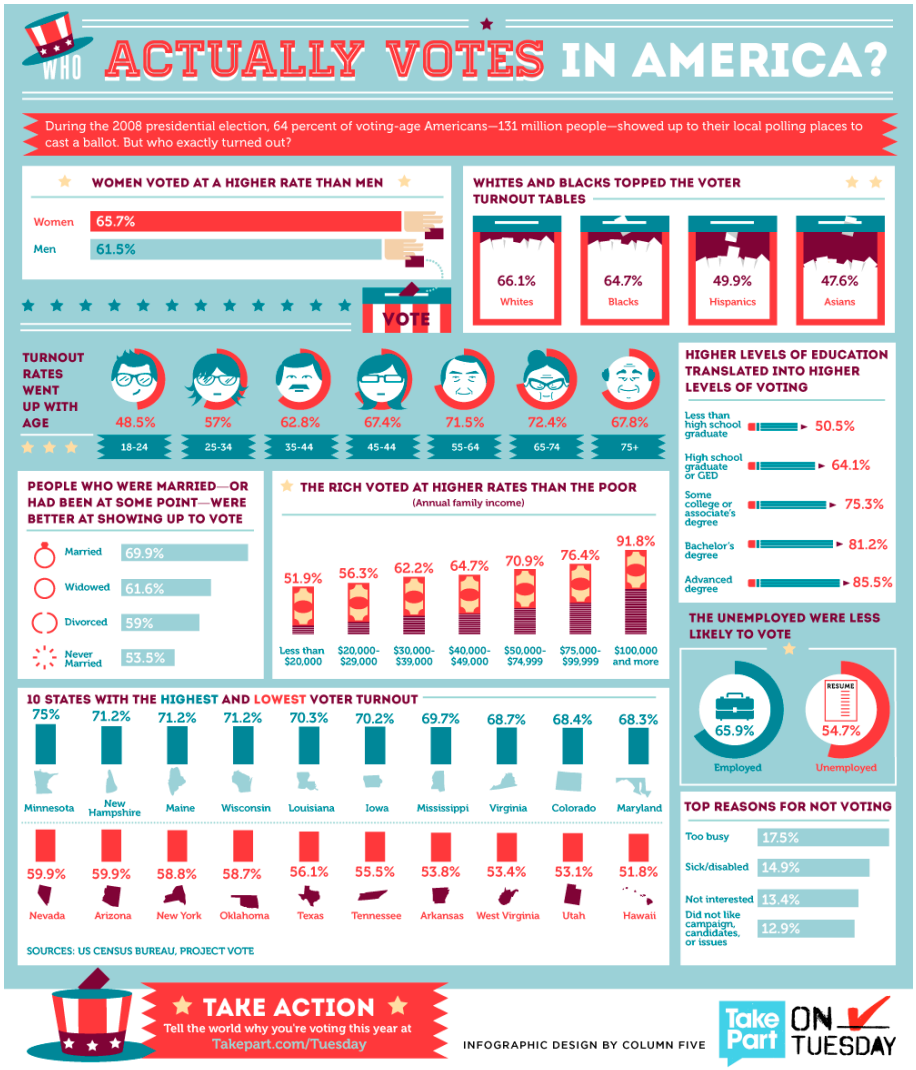

Figure 1: Infographic showing the demographics of the voter base. The data involves many dimensions to represent such as state distribution, marital status, employment status, etc. Infographics are powerful ways to represent many dimensions in one compact form.

While designing this kind of visuals seems to be a pretty straightforward process, there are challenges that one must be aware of to improve any visual design. A roadmap has to be determined before the design process starts to produce visuals that speak to the target audience and highlights the important or the features that the target audience wants to be focused on. Wrong evaluation of target audience can guide the design in focusing irrelevant aspects of the data. In this report, we are going to evaluate the capacity to which Google Music Timeline fulfils the aforementioned traits of a successful visualization.

2 What is the Music Timeline?

The Music Timeline [1] represents the popularity of various genres of music within a time frame between 1950 and 2010, the visualisation is based on the data gathered from Google Play Music users. The popularity timeline is based on the artist and album collections of the user database. Each area field represents the popularity of the genre, subgenre or artist at different levels of detail. The thickness of these fields gives an indication of how popular these musical entities were over the timeline. For example, the Rock field dominates the popularity measure between 1960 to 1985. (Figure 2)

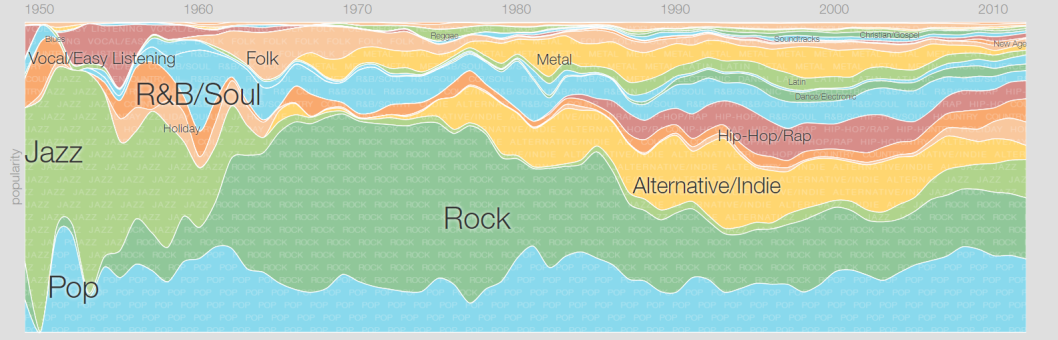

Figure 2: The popularity of each genre is depicted by the thickness of the colored stripes at any year.

Let's look at the navigational capabilities of the current visual design. Navigation is quite intuitive and enables fast switch between different level of details. Search box provides a direct search of the keyword and zooms out button(minus sign) sends user at a one step lower level of detail(figure Figure 3).

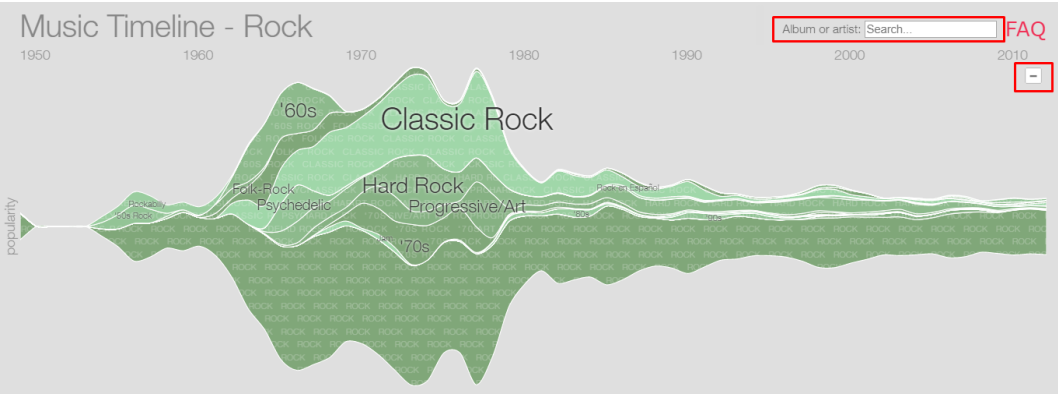

Figure 3: The Search box and zoom out button provides a fast navigation

Furthermore clicking any of the keywords marked by red triangles provide even faster navigation. (figure Figure 4)

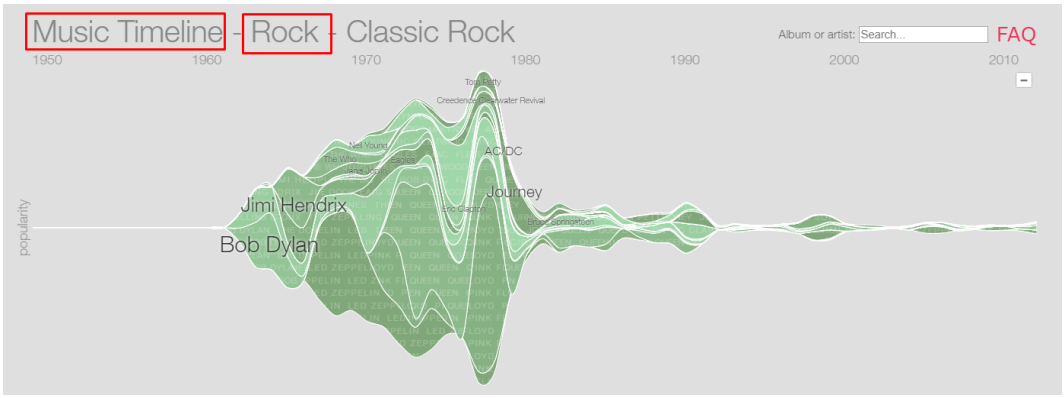

Figure 4: Clicking on genre/subgenre on the title above the visualization sends the user to that level of detail

Selection is enhanced by the use of highlighted areas and brightness changes. Clicking on any of the fields, separated by white borders, reveals the subgenres associated with the genre. The visualization uses one color with different brightness levels to create separation between sub-genres. (figure Figure 5) For our current example, we chose Rock genre.

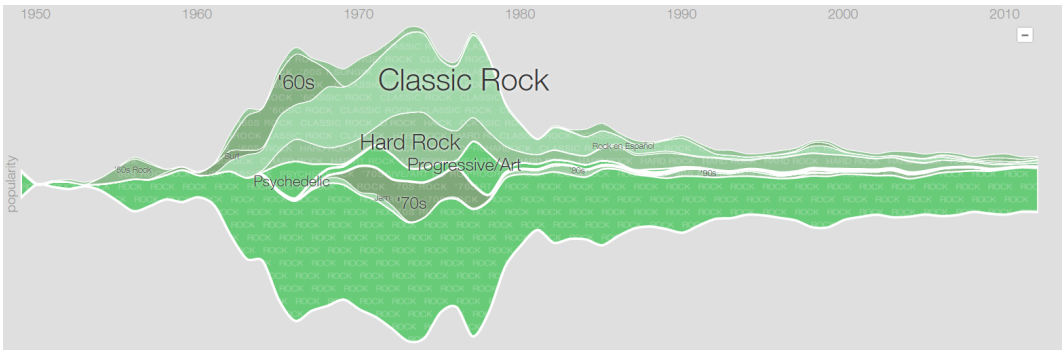

Figure 5: Subgenres of rock genre is depicted by stripes with varying stripe thickness. The upper part is mirrored to the bottom part to represent the total popularity of rock genre through the years.

When users hover the mouse around the visualization, the hovered area copies the brightness of the area labelled with letter A in Figure 6 and the white borders surrounding selected the area becomes thicker, also a label pops up right next to the field. Comparison of Figure 5 and Figure 6 shows the brightness change applied to the classic rock area field. The area labelled with letter A is the mirror image of the upper part of the visualization. It represents the collective popularity of all subgenres that span between 1950 and 2010. This functionality is only available for this level of detail. When users click on this bottom section, it reveals the popularity timeline of all of the artist belonging to all of the sub-genres(figure Figure 7).

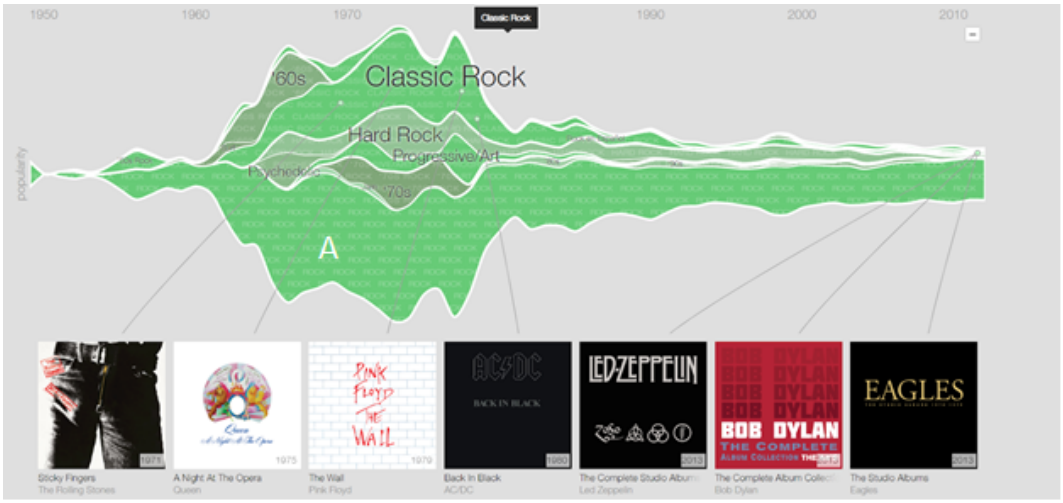

Figure 6: Hovering over classic rock genre highlights the section surrounded by white borders.

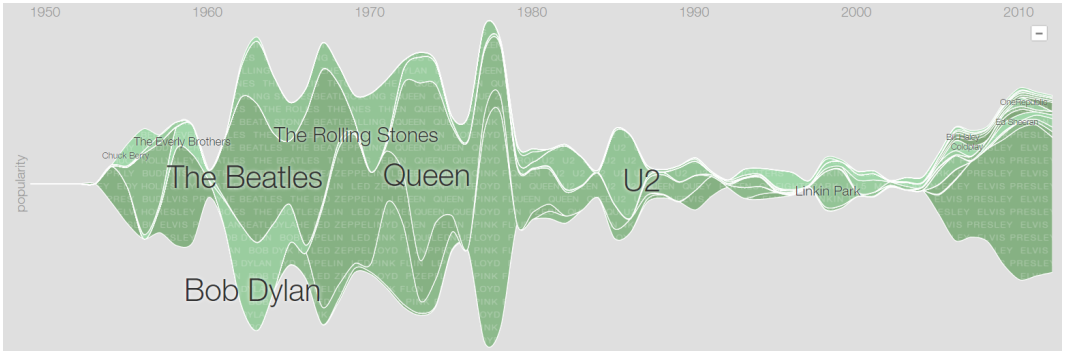

Figure 7: Selecting rock genre reveals the artists associated with all subgenres. Here, there is no mirroring applied as the total popularity of artists is already depicted at a higher level

Furthermore, clicking on any sub-genre reveals the artists only associated with that subgenre(hard rock) (figure Figure 8).

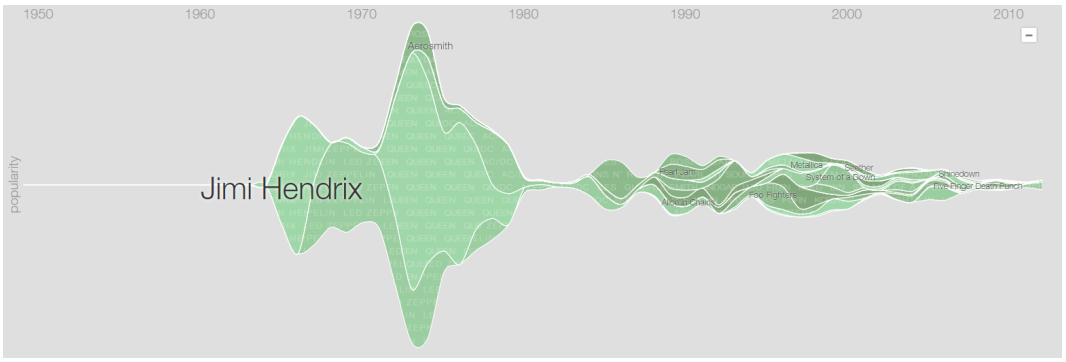

Figure 8: Selecting classic rock genre reveals the artists associated with the subgenre. Again, there is no mirroring applied as the combined popularity of artists is irrelevant and is already depicted in one level higher in the visualization

Selecting any specific artist then reveals the short biography associated with the artist(figure Figure 9)

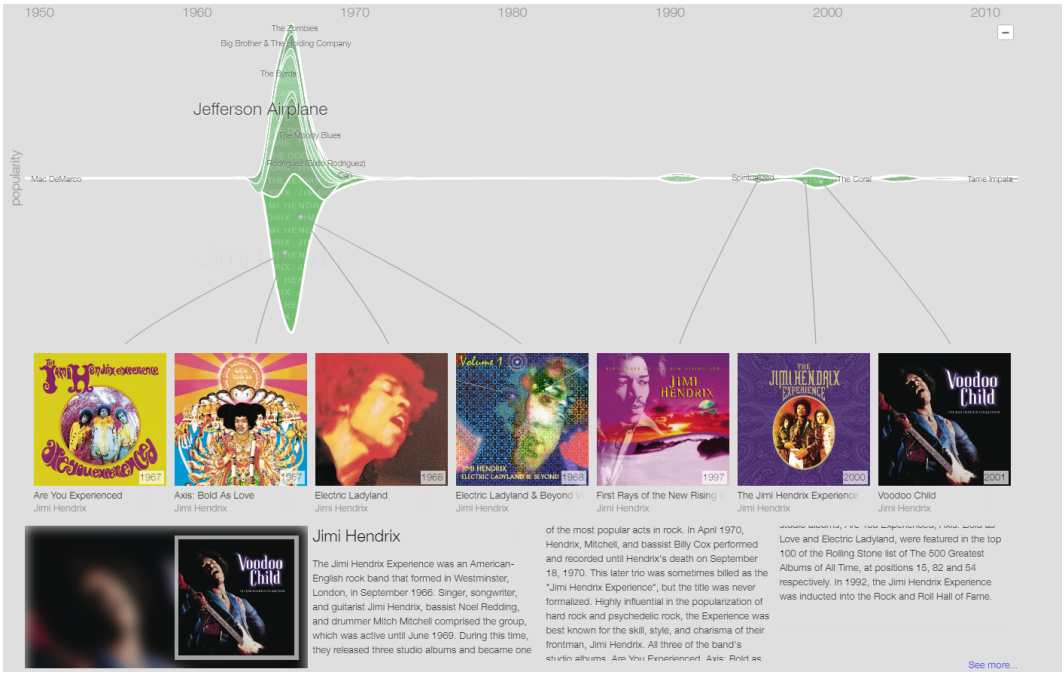

Figure 9: Selecting Jimi Hendrix reveals the short bio about his music career and impact within the genre.

Finally, hovering over each area field at any level of detail, separated by white borders, reveals the albums that best define the genre, subgenre and artist. The below figure shows the albums that define rock genre in total.(figure Figure 10)

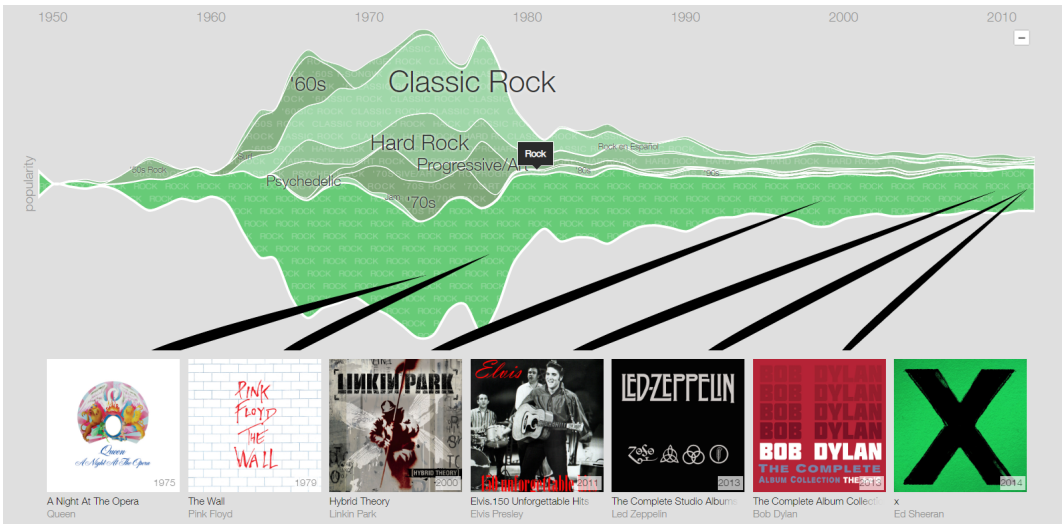

Figure 10: The most popular albums defining the genre, subgenre and artist can be accessed by hovering over the target. The albums will be shown under the visualization and sorted by album release year.

3 The user base

Music timeline is a useful tool for various types of users that range from regular music fans to professionals in the music industry. In this section, we will list the potential users and mention their interests and how each group of users might use the visualization differently.

3.1 Music fans

This group consists of people who listen to music (ir)regularly. They are often interested in discovering albums or artists that are similar to their taste, therefore, their main use of the music timeline is to find popular albums of their chosen genre, and probably purchase that album since clicking on an album will take them directly to google play store.

3.2 Music Students and Educators

They are specialized in the music field for either to become musicians or music educators. Their interest in the music timeline is to study the development of genres through time in that which genres were popular at a given period of time, and learn how genres influence each other.

3.3 Musicians

This group include musicians who are instrumentalists, composers, conductors, singers, etc. Basically, the population that shows interest in music for both occupational and personal reasons. Their interest in using the music timeline is to discover which genres were trending and what were the most popular albums that made an impact at that era so that they can find songs that can give them inspiration for their music.

3.4 Record companies

While this group presumably shows a very specific interest to the application, it is worth mentioning. Sometimes the record companies ask their signed artists to produce singles or specific projects around a specific genre or incorporate genre-specific elements into their music. Partly, this is motivated by the popularity of that incorporated genre at a certain time. By inspecting the popularity timeline, producers can make informed decisions about which genre-specific elements to add to their next production based on their past popularity.

4 Data characteristics

Arguably, the main purpose of the current visualization under the study is to visualize music popularity in a timeline which mainly includes genres, artists and albums that allows for an intuitive exploration between different levels of details. In this section, we will go over the list of variables that the visualization tool uses and their variable types. Before we list the variables, lets first describe the variable characteristics which we will use to categorize our variables. Looking at table Table 2, in the “Characteristics” column we have three adjectives used to define properties of our variables. For the first characteristics, we have either “continuous” and “discrete”. Continuous means that the variable is numerical in nature and the increment made to the variable can be continuously presented on a value axis. Conversely, discrete variables can only be assigned to certain values. For example, a time variable in the format of 1900, 1910, 1920 with 10 years separation between each recorded value is discrete while a time variable recorded as 10:23:34 21/07/1989, 12:41:27 12/03/1990, etc. is continuous. The second characteristics are either “qualitative” or “quantitative”. If we can express the recording of the variable in numerical values, it is quantitative, otherwise, rank, such as gold, silver and bronze medals are considered to be qualitative measures. Last, we used “interval”, ”nominal” “ordinal” and ”ratio”. These are the level of measurements. Nominal variables refer to names and there is no order. Ordinal measurements refer to the variables that can be ranked such as high socioeconomic status vs low socioeconomic status. Interval variables are numeric and the distance between values are meaningful but there is no absolute zero value and their ratio between two values. Last, the ratio measurement, there is absolute zero. An example would be a number of employees in a company.

Dimensions	Details	Characteristics
Popularity metric	Aggregated from Google play music	Continuous, quantitative, interval
Artist	Artist Names	Discrete, qualitative,nominal
Genre	Jazz, Rock, Pop...	Discrete, qualitative, nominal
Sub-genre	Bop, Hard rock, 80's pop	Discrete, qualitative, nominal
Album	Kind of Blues, Are You Experienced?, Thriller	Discrete, qualitative, nominal
Album release date	Years between 1950 and 2010	Discrete, quantitative, ordinal
URLs	albums and song purchase	Discrete, qualitative, nominal
Biographies	Short biographies about the artists	Discrete, qualitative, nominal
Number of albums in the music library	Aggregate count of the number of albums from each year collected from users	Continuous, quantitative, ratio
Number of occurrences for artists in the music library	Count of albums and separate songs that is attributed to an artist for each year collected from users	Discrete, quantitative, ratio

Table 1: Variables in music timeline data. 3 column table lists the variables, examples of the instances of those variables and the variable characteristics.

Popularity metric defines the thickness of each stripe as we have seen in, starting with **??**, at all levels of detail. In the visualization, popularity metric is represented on the y-axis as a dependent variable. Popularity metric is based on the number of albums and the artist counts found in the user's personal playlist. Instead of using absolute values, percentage popularity score is used to avoid any white space that might have occurred otherwise. However, when we go to the visualization for (sub)genre we notice that the genre expands to the height of the y-axis, in this case, the designer wants to keep the popularity representation of the genre which made avoiding the white space not possible. Looking at different levels of the detail in our visualization, we believe the popularity metric is probably calculated for dimensions genre,sub-genre, artists and album separately and by using different metrics. Since the details of complete data are not available we can only speculate about the data structure. Genre, sub-genre, artist and album are the main elements in the visualization which forms the stripes at differing levels of detail. Year (date of release) is used as an independent variable that enables users to observe how the aforementioned variables change over time. URLs provide a link to google play store for any potential purchase at the highest level of detail. The album covers give us a visual separation between each album, Biographies are provided to inform potential users about the artists and their impact within the sub-genre. Last two variables in table 2 are collected from the google play music user base.

When it comes to the number of observations, we don't really have any indication of the magnitude. However if we assume the entries in our data table is based on the albums, we can say that the number of rows will be at least equal to a number of distinct counts of albums. Some users may have songs that are not a part of an album but a live recording, studio session, etc. in their libraries. To account for such cases there might be an additional dataset which includes non-album tracks of artists. In the end, both album based and non-album based scores might have been combined to reach a final popularity metric.

5 Hypothesized needs of the user base

The problem domain corresponds to the needs of the user base. The demands made by the user base guide the designer to select the right medium and data simplification techniques. In this section, we will list the possible questions that the user base would be potentially interested in finding answers to. We think it is best to separate our question into two categories. First, potential users can either observe the change of popularity over a course of time or focus on a specific year to compare the popularity among variables. This makes the popularity to be both assessed vertically

and horizontally over a time course. Second, the level of detail provided by three different views depicting an overview of genres, sub-genres and artists can be compared in terms the way the popularity metric is used at each level. We will first go over the questions that look at each level of detail and then continue with the popularity assessment questions. In the below (Figure 11), each level of detail in the visualization is stacked on top of each other along the y-axis while the time component enables the visual search through the x-axis which describes the change of popularity of variables(sub-genre, genre and artist) over time.

Popularity assessment
Popularity metric is the dependent variable in our visual design which enables comparison between many variables. Users can either focus on individual years and make popularity comparison between genres, subgenres of a genre, and artists, thus vertically comparing multiple popularity scores or they can observe how these variables change over a time period, performing a horizontal visual search. Thus we further segment this part into two subsections, namely vertical and horizontal assessment.

Vertical assessment

1. How is the total popularity of (sub)genres distributed for a given time?
   (a) How much does each genre contribute to the total popularity across all genres for a given time?
   (b) How much does each sub-genre contribute to the total popularity of a genre for a given time?
2. How is the total popularity of artists distributed for a given year?
   (a) How much does each artist contribute to the total popularity of a subgenre for a given year?
   (b) Which albums are the most popular at a given year?

Horizontal assessment

1. Which (sub)genres are popular over a period of time?
   (a) Is the current color mapping appropriate for observing the trend of (sub)genres changing over time?
   (b) Which albums are popular at a given period of time for the within (sub)genres?
2. Which artists are popular over a period of time?
   (a) How is the popularity trend of artists changing over time?
   (b) List top 3 most popular albums for an artist

Assessment of the three level of details

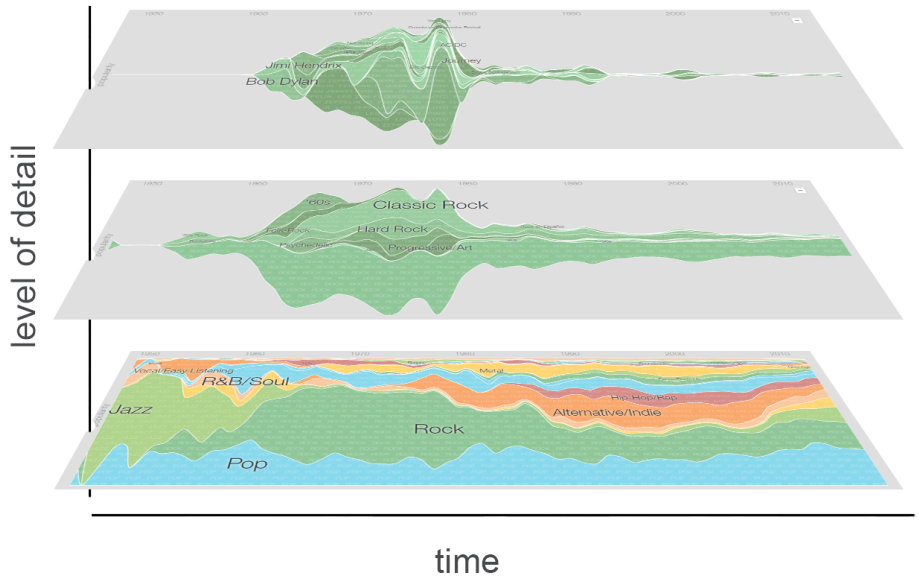

Figure 11: Music Timeline visualization laid out in different levels of detail on Y axis while the temporal attribute is presented on X-axis. In the figure at the highest level of detail, we can see the artists that belong to the classic rock genre.

In this part, we are first going to list a group of questions pointing out the current limitations of Music Timeline visualization. Later, we will list the popularity assessment questions in the following subsection.

1. Why relative normalization is used for the lowest level of detail?
   (a) Why the representative albums listed only by date of release? This also applies to remaining levels of detail
   (b) Why the most popular albums of all times such as Thriller by Michael Jackson is not within the representative albums list? This also applies to remaining levels of detail.
   (c) What determines the order of stripes from top to bottom?
2. Why relative normalization is not applied to the second highest level of detail?
   (a) How one can compare the popularity of sub-genres across genres?
3. Why relative normalization is not applied to the highest level of detail?
   (a) How one can compare the popularity of artists across (sub)genres?

6 Design performance evaluation

In this section, we will have a look at the questions listed in the previous section, discussing the inherent limitations and evaluating how good the current visualization tool is able to answer the listed questions. In our attempt to dissect the questions, we will consider the multiple levels of detail provided by the visualization. First, we will start out with the common issues through the visualization that can negatively impact the visualization's ability to be an efficient source of information. Second, the potential users can either observe the change of popularity over a course of time or focus on a specific year to compare the popularity among variables. This makes the popularity to be both assessed vertically and horizontally over a time course. Finally, the level of detail provided by three different views depicting an overview of genres, sub-genres and artists will be evaluated. The potential design improvements addressing the problems laid out in this section will be discussed in section 7.2

6.1 Common problems for each level of detail

- There is no zoom in functionality
  - It would enable a finer popularity measure exploration throughout the visualization. For example, a user who needs to inspect years between 1960 and 1965 can't figure out the finer rates of change that are squeezed between the 10-year axis tick marks
- The borders surrounding the (sub)genre and artist fields don't create clear separation.
  - It is hard for users who would like to observe the full temporal popularity change of (sub)genres and artists.
- Once a user clicks on one of the subgenres, artists related to that sub-genre are revealed, however, the design makes it very hard to select your artist of choice.
  - Fast hover over the subgenres and artists reveals the variable names so with some luck one can find the targeted variable, however, for some artists, it is really hard to find the area that represents it.
- The data is normalized by the total number of albums for each year yet there is no way the user can see how the data is normalized since the lack of scales in the visualization.
  - Users who are scholars within the field of music would be interested in more numerical representation in terms of z values. This could have been provided.
- The use of labels as texture to visually separate between the (sub)genres and artists is only useful when there is enough space to display the texture, otherwise, it is not visible.
  - Visual discrimination between mid to low popularity variables cannot be made accurately by the user
- Labels are missing for the least popular genres/ subgenres/ artists the user have to hover over the desired area to see the label.
  - Visual discrimination between mid to low popularity variables cannot be made accurately by the user
- Due to the cluttering of data, it is not easy to find how many subgenres each genre has, or how many artists does a subgenre have in the visualization. Same goes for the number of albums.
  - If a user wants to compare a number of subgenres a genre has or artist a subgenre has, it cannot be done easily. This kind of knowledge might be interesting for a student user base.

6.2 Popularity assessment questions

Vertical

For items 1 and 2(sub items included):
Although a visually inspecting vertical stack of stripes immediately reveals the element that has the highest popularity, the (sub)genres and artists with low popularity get lost in the stack. Since the visualization is trying to emphasize the most popular (sub)genres and artist, this is not a major problem. Currently, the time axis tick marks are placed apart by 10 years, it is hard to visually select a certain year to perform a vertical comparison between (sub)genres or artists. This limitation can be addressed by implementing a zoom in functionality.

For items 1.b and 2.a:
If the user were to compare two different imaginary lines vertically placed on the visualization in terms of their relative popularity at the level of sub-genre, the individual contribution of each subgenre would not be visually differentiated. To make this clear let's have a look at the following (Figure 12).

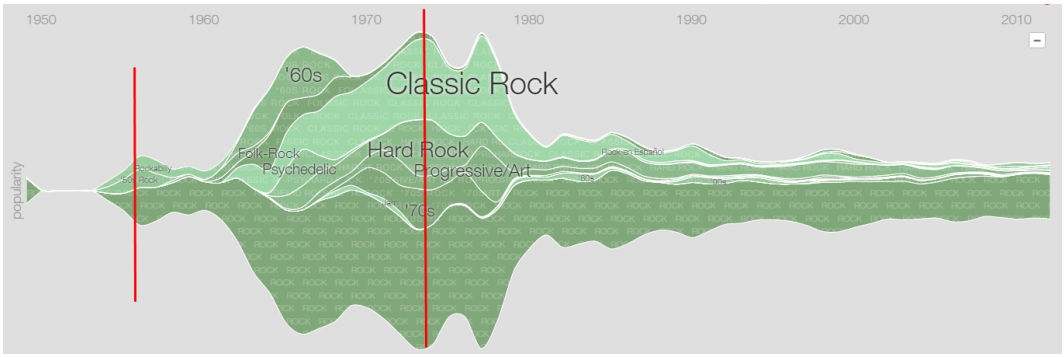

Figure 12: The red vertical line on the left side passes over the regions of rockabilly and 50's rock. The red line that is placed between years 1970 and 1980 is covering regions such as folk rock, classic rock, hard rock, etc...

A fast visual comparison between two lines in above figure would deceive a potential user into believing that rockabilly and 50's were not that popular within rock genre. While it is known that between 1970 and 1980, it was the golden era for rock music, the popularity of rockabilly and 50s rock was also quite popular back in the day. Due to use absolute values of the popularity measure, on this level of detail, we actually observe the absolute contribution of each sub-genre to the combined popularity of rock genre. In other words, while the rockabilly and classic rock subgenres almost account for the half of the total popularity in mid to late 50's and early 70's respectively, the area designated for classic rock is far bigger. The same effect is also present for the artist popularity detail. Eventually, the users cannot inspect the relative contribution of each (sub)-genre or artist. While this is a design choice, the addition of toggle switch between relative and absolute scale representation can make the visualization more flexible.

For item 2.b:
Music timeline uses the data for contemporary releases of old songs and greatest hits albums to account for the popularity of successful albums released in past. While there is no popularity comparison is provided for the current visualization. Even if there was one, it would be based on the contemporary releases of old albums.

Horizontal

For item 1 (sub items included):
To understand which genres are popular over a period of time, a user has to follow the stripe representing the genre and compare its height to the total height of the visualization, although this can give a user an estimate about how popular certain genre was at a given period of time it is not easy to follow genres with lower popularity. Since the main purpose of the visualization is to reveal variables with top popularity, this is not a major problem, however, with the same train of thought, we could argue against the inclusion of genres that have such low presence in the visualization while creating a stuffed look. This could have been avoided by using a longer y-axis which would have stretched the band's height, providing a more relaxed view overall. Furthermore, the color mapping used here do not give any indication of the genre similarities. To combat with the problems arising from using the same color for multiple genres, designers used hashing to provide separation between the genres. This solution is not practical in the case of low popularity genres because there is not enough area symbolizing the genre so that the hashing is not detectable by eye.

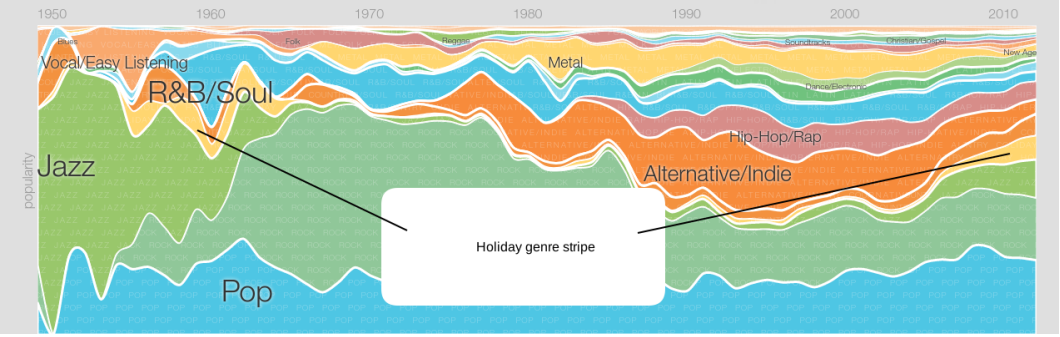

Figure 13: Popularity trend of genres over time. Holiday genre and many other genres that have even low popularity creates a stuffed feeling that needs to be addressed.

For item 2 and 2a:
To find out which artists are popular over a period of time the user has to go down the sub-genre level otherwise the artists are not visible. As mentioned before, the visualization only provides the popularity of an artist compared to other artists in a subgenre. Therefore, it is not possible to compare the popularity of two artists belonging to different genres/ subgenres. Even if the user compares artists in the visualization in different screens that will give him the wrong idea about the popularity of each artist in general, since a popular artist in an unpopular subgenre could be less popular than another artist that is less popular within his popular subgenre. One could argue that this comparison is beyond the goals of this current visualization, however considering the obvious design choices to highlight most popular instances among variables for genres and subgenres, the inclusion of cross subgenre and artist comparison as to find who are the most popular artists or subgenres could have been a useful addition for potential users.

For item 2b:
To list top albums of an artist, one has to travel between a couple of sub-genres to eventually find all the albums produced by the artist. For example, under 70's pop, Michael Jackson has a couple of albums however these are not the most popular ones for among all albums Figure 14. Looking at Figure 15, we can finally see the most popular albums of the artist. However, the fact that Thriller is the most popular album of him cannot be obtained by looking at the visualization. Generally, when a user hovers over the visualization, the albums will be shown below the visualization and arrows will be drawn to show which album belongs to which genres, subgenres, or artists, and also in which year the album was released. However, as mentioned before, there is no indication for the popularity of albums. An easy fix would be including an "order by popularity" option right next to the album images right under the visualization timeline.

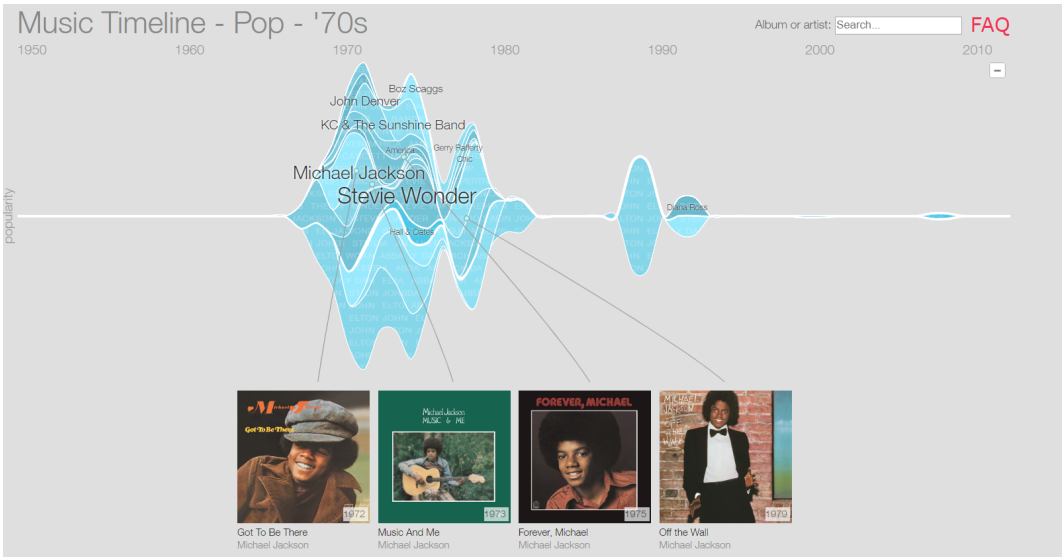

Figure 14: Popular albums of Michael Jackson under 70's Pop.

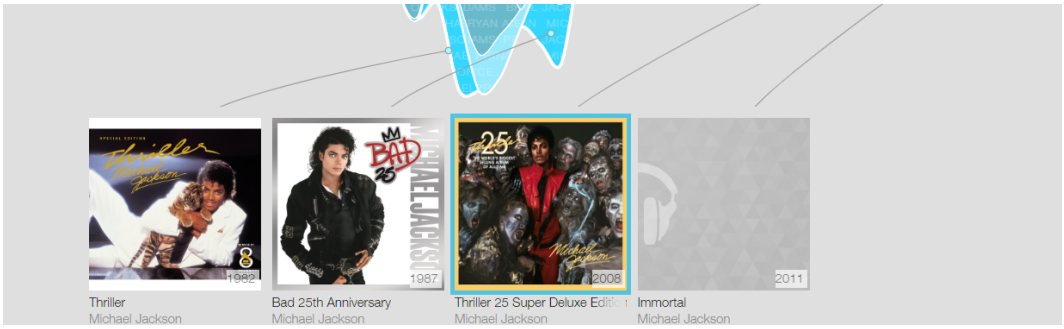

Figure 15: popular albums of Michael Jackson under 80's pop.

6.3 Assessment for three level of details

Normalization

When it comes to normalization of choice, the designers opted for a relative normalization only for the lowest level of detail which is the genre overview. Compared to that, absolute values are used for the remaining levels of detail. This is a design choice implemented for emphasizing the popularity difference between sub-genres and artists. Additionally, if they were to use an absolute normalization, the visualization couldn't have employed stacked area charts. However, having fewer counts of sub-genre for each genre makes it easier for using the absolute scale in the remaining levels of detail. Let's keep in mind though using percentile stacked area charts do not always give the best visual representation if the ordering of area stripes are not optimal. In order to get most out of this kind of visualization, the variable(in this case genre) with the highest variability should be placed on top. This emphasizes highly popular genres without any interruption caused by genres with very sparse data.

Most popular albums

The representative albums that appear right under the timeline are not the ones that are most popular. For example, for the pop genre, we have Lady Gaga- Fame Monster at the first place. While the Thriller album released by Michael Jackson is not even in the list. Considering the number of albums sales, Thriller by 33 million sales towers over the Fame Monster by 1.6 million. This is a single example in which all genres have the same problem. Basically, the popularity albums are not included in the design thus there is no order as well. However the arrows pointing to example albums sometimes have variation in the thickness of the arrow, however, this is not consistent. Whether the thickness has to do with the popularity is not clear. It can be also indicative of the

sample size for each album.

Across sub-genre and artist comparison Current visualization does not provide a comparison between sub-genres and artists across the genres. Therefore questions such as "Who is the most popular artist/sub-genre at a certain time?" cannot be answered. Such questions could be of interest for students who are asked to write reports using the music timeline

Summary of evaluation
A strong aspect of the current visualization tool that it is quite compact. Within a second, the user can acquire nicely aggregated music popularity information. However, it comes with a side-effect. Stacking stripes on top of each other create a stuffed look, especially towards the low popularity items. Furthermore, there is no visual encoding for the popularity of albums. This damages the ability of the visualization to allow for a more comprehensive information search.

7 Visual Encoding Analysis

In this section, we will talk about the visual encoding that is used in the Music Timeline which is clearly an infoviz since the data treated in the visualization do not have a natural representation like activations of fMRI scanned regions on the brain. Music Timeline combines categorical, quantitative and relational data and encodes them in an interactive area chart.

7.1 Visual encoding:

To encode the music data the designer of Music Timeline chose area chart. Area charts are usually a good tool to represent quantitative data (popularity in the case of Music Timeline). In this section we will go through each data and study what visual encoding was used for it. We will also discuss the optimality of the choice of the designer for each data type.

Data	Visual Encoding
Release date	X axis
Popularity	The thickness of the bands, label font size
Genres	Color hue, texture, label
Subgenres	Color luminance, texture, label
Artists	Color luminance, texture, label
Albums	Shapes, album cover
Biography	Text

Table 2: Variables used for data encoding in the visualization.

- The choice of encoding the release date on the x-axis is a good choice since it is a standard convention to plot the independent variable on the x-axis. It also gives the user a clear distinction between different years in the visualization. The user can easily compare the given data of each year.
- The choice of encoding the popularity into the band thickness is a good choice; each area (stripe) represents a genre. As the user easily differentiate between the popularity of the genres by just comparing the thickness of the bands, the lack of scale on the y-axis can cause problems since the selection of a genre expands the subgenre overview to the full height of the visualization area regardless of the general popularity of that genre in the lower level visualization. Figure 16 shows the amplification of the popularity of Alternative/ Indies on the left when it is selected compared to the actual popularity of the genre as shown on the right side. Using the label font size to represent the popularity of the genres does not work well in the visualization since there are many genres which are hardly visible because of their low popularity. The font for those genres is not visible.

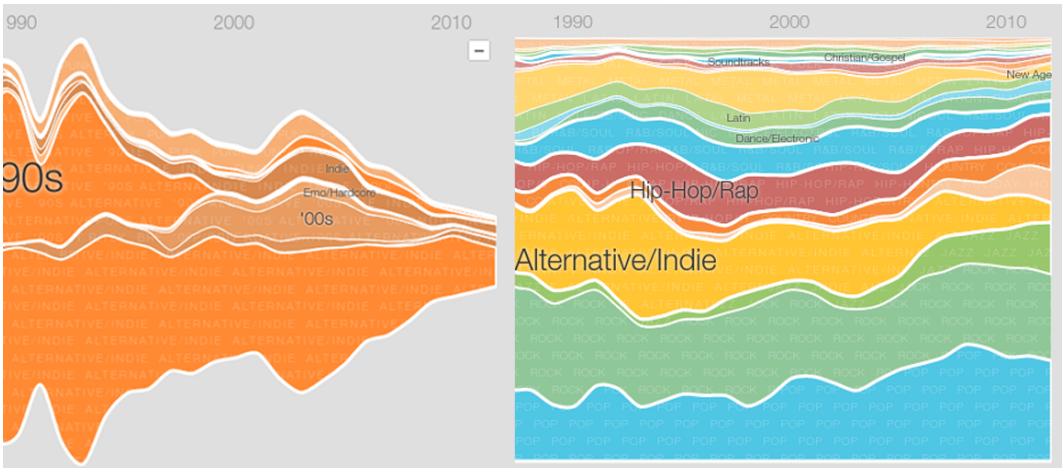

Figure 16: difference in areas size selected genre Alternative/Indies on the left and main visualization on the right.

- Using the color hue to represent the genres would have worked better if the colormap was rich enough to contain different color for each genre (which is very hard) because the number of genres is large. The visualization designer tried to solve this by including a second type of visual encoding to enhance the distinguishing between genres by placing the name of each genre in the texture of the area representing that genre. The main problem with this is that the solution works only for the genres with high popularity where there is enough space to display the texture, but when we have genres with low popularity the texture is hardly visible because there is not enough space to display the texture. Also choosing white as the color of the texture is not a good choice since it is hardly visible on light colored backgrounds, however we have noticed that when using the search option the colors are assigned randomly which can be very confusing; see Figure 17.

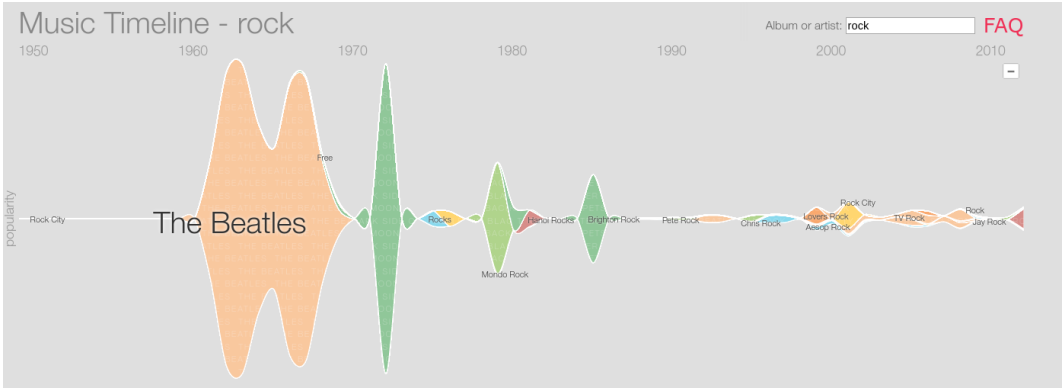

Figure 17: Search results of the term rock.

- Encoding subgenres using color luminance is a poor choice (using color luminance to represent nominal data is bad in general) for the visualization since using color luminance will wrongly indicate to the user that the subgenres are somehow ordered. As in fact the lighter subgenres have fewer samples while the darker subgenres larger sample size. In general, the color luminance is just used to visually differentiate between the subgenres and has no other meaning than that. The reason for this poor choice is most likely that the designer wanted to keep the same color as the parent genres and avoid giving subgenres colors similar to genres other than the parent genre.

- Encoding artists using color luminance have the same problems as mentioned above.

- Encoding the albums using shapes is good but it does not indicate the popularity of each album. The shapes are presented as small circles and the circle location shows when the album was released, and to which genre/ subgenre/artist the album belongs. Other than that the circle does not have any indication about the popularity that specific album.

- Using lines between the circle and the album cover is a good idea in case the amount of albums presented is not large. Otherwise, all the lines will combine to form a thick line and it is impossible to know where each album is in the visualization.

- Using white lines to visually separate between genres is a bad idea since those lines do not represent actual data.

- The visualization uses a grey background. This a good choice although white could have been a better choice. Using grey is good color since it is not used as a part of encoding the visualization elements which produces good visual contrast between the data and the background. The background takes a most of the space of the visualization of the genre/ subgenre which is not very practical since the space is not used for anything.

- There are ambiguities in the visualization; the designer used white lines to visually separate between the stripes. The user might mistake those white lines as a representation of data variables which is not the case. The use of the white lines would not have been necessary if the designer had a better choice of colors that are visually distinguishable. The other ambiguity caused by the poor choice of colors as mentioned above is the same (or very similar) color is used to represent different data as sometimes the thickness of the lines between the genres is quite big as shown in Figure 18.

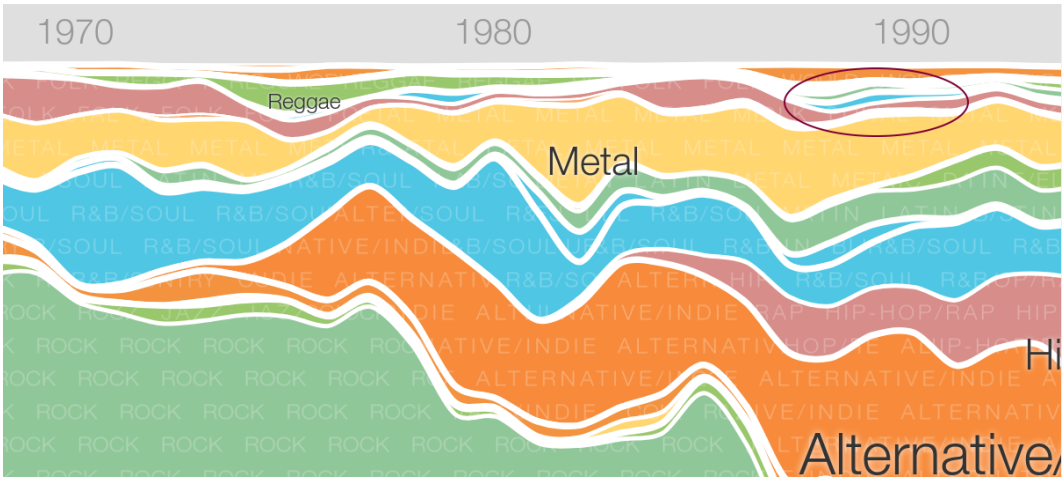

Figure 18: White lines can be thicker than the data encoded in between those lines.

- The user can clearly compare the popularity of genres either in general or at a given period of time, the same thing goes for subgenre comparison within the genre and the artists within the subgenre. The lack of scale on the y-axis causes another ambiguity since the change in the area height of the stripe to show the change of popularity in the subgenre overview is much bigger in comparison to genre overview. Using brighter color for both visually separate the highlighted area, and visually encode certain genres/ subgenres is a bad choice. In Figure 19, the user can be easily misled by the stripe encoded in light green as a part of the hovered data. We can see that hovered over Classic Rock area has the same color as Progressive/ Art rock area.

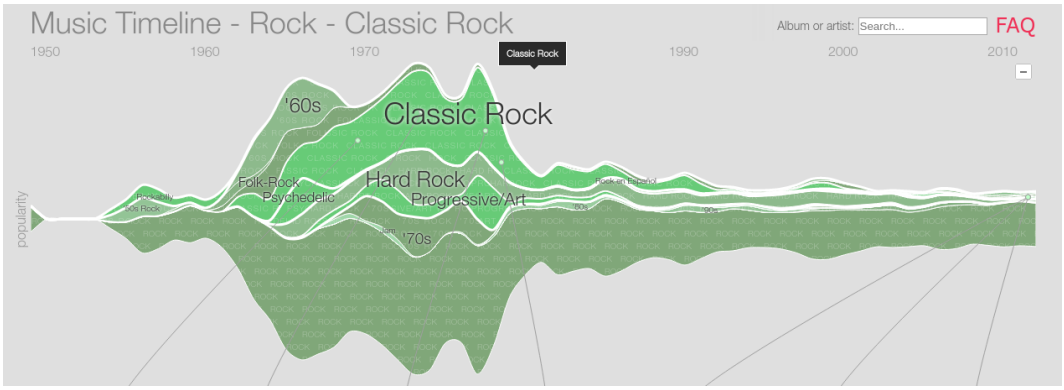

Figure 19: Hovering over Classic Rock area changes its color as the same color of Progressive/Art Rock.

- Data on time axis represented as years with 10-year tick marks on the axis. Assuming the aggregation made on year basis, we could say rational spline interpolation is used to calculate the popularity values corresponding to the data representations between each year. Thus these lines may not correspond to actual reality. However, if the user library information is collected monthly then, the smooth curves will be more close to the reality. Again, this is a design choice and the focus is always on providing the higher level information without being very specific for each visualization component.
- The stripes of the visualisation are smoothly curved, the designer might have chosen to smoothen the curves for an aesthetic purpose, since the popularity is computed at the year level . this smoothness of the curves can have two effects on the visualization, first it looks good visually and it can give us a suggestion of the trend line, interpolation for the future, however it is miss leading since the user will perceive this smoothness of line curves are actually the representation of the real data and not a design choice if the user unknowingly try to compare the beginning of the year with the end.

7.2 Improvements/Suggestions for the design issues

In this section, we will go through the limitations of the Music Timeline visualization and list improvements we propose for each limitation that was mentioned in the previous section.

- Color encoding of the genre overview can be improved by be even more distinguishable since the grey color will pop out between the non-grey colors.
- The texture that is used to enhance the encoding of genres is hardly visible, another texture can be used to enhance the colormap since as mentioned above having 20 colors that are equally different is not possible. Therefore using a texture which is visible will definitely help the user to differentiate between the genres.
- Introducing a percentage y-axis will help the user estimate the popularity percentage of each genre by comparing the stripe thickness and the scale.
- Since the data is not ordinal, instead of using color luminance to encode subgenres/artists, the designer can simply use color hue. A better approach is to order the subgenres by their popularity, this will allow the designer to use the color luminance which will be in this case useful because we have made data are ordinal.
- A possible solution for solving the cluttering of the data is to offer a zoom in functionality where the users can zoom into a cluttered area to see more details that are not visible by the viewer without the zoom.

- To solve the color problem when hovering over a genre, visualization can grey out the non-hovered genres to create visual contrast for the selected genre without having to change the luminance of the hovered genre. This will solve the problem of having to use the same color for both the hovered area and to present other genres.
- In the current visualization there is no option for the user to filter genres that one is not interested in or select multiple genres for comparison, adding this feature will allow the user to have more insight, for example, the user can isolate and compare the popularity of two genres within time to see if there is a possible correlation between the two genres.
- Instead of using white lines between the stripes the designer can use shading cushion profiles this will allow visual separation between the stripes (the use of shading is only necessary if we have colors that are hard to separate visually) , this will eliminate the unnecessary use of white lines that can cause confusion for the user, and saves some pixels.
- Instead of using labels to show the genre name in the visualization the designer could have added a legend below the visualization to specify what data every color and texture is encoding, by doing this the designer does not have to worry about genres that have lower popularity and does not have enough space to display the label and it also solves the overlap with other genres area problem.
- Another improvement is instead of showing all the albums when hovering over a genre (which have missing arrows linking the visualization to the album covers) the visualization should take not only the hovered genre but also the hovered year and show only albums of the hovered year, this will resolve the clutter of showing many albums that belong to the genre.
- Year variable could have been used to filter data so that we can avoid the horizontal clutter and to allow the user to compare genres only in the year he is interested in.
- In regards to zoom, a selection can be provided so that users can select a specific year range.
- Adding an extra level of selection for artists to show the contribution of each album the artist popularity quantitatively would be a good idea, in this level the albums should be encoded in the area same way that artists are encoded in the subgenre level, this makes the it easy to know how much did each album contributed to the popularity of that specific artist.
- To provide popularity information for albums, the thickness of the stripes pointing out to each album can be used as an additional visual encoding element. Moreover, a simple popularity ordering functionality can be provided when sorting representative albums beneath the timeline visualization.

8 Evaluation as an end user

In this section, we will select three questions from the fifth section[Hypothesized needs of the user base] and try to find relevant information on the visualization which answers our questions.
1. List top 3 most popular albums for an artist
For this question, we will try to find the top 3 most popular albums of the artist Michael Jackson. We proceed by selecting pop genre in the genre overview. Then we are left with many subgenres to select from. For users, who don't have prior knowledge regarding Michael Jackson's activity over the years, a fast hovering over action over all of the sub-genres reveals the most representative albums for the sub-genre. If the user is attentive enough to catch the MJ's albums, this approach might work otherwise users have to check each subgenre one by one. Assuming that the user knows that MJ was active between early 70's and late 90's, it is much easier to spot the albums like in Figure 20 and figure 1. However looking only to these albums, we cannot conclude which album was the most popular.

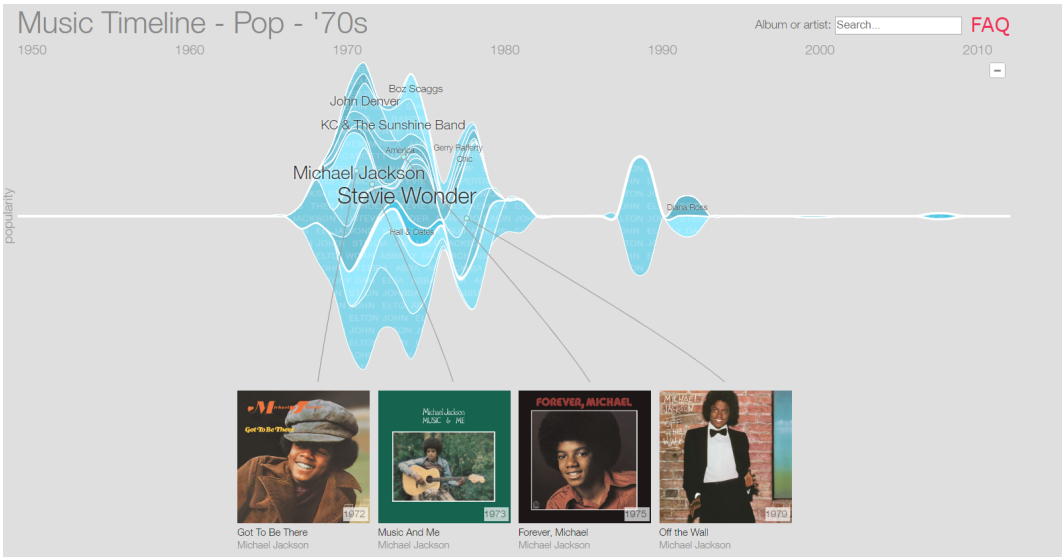

Figure 20: MJ albums under the 70's pop subgenre artist overview

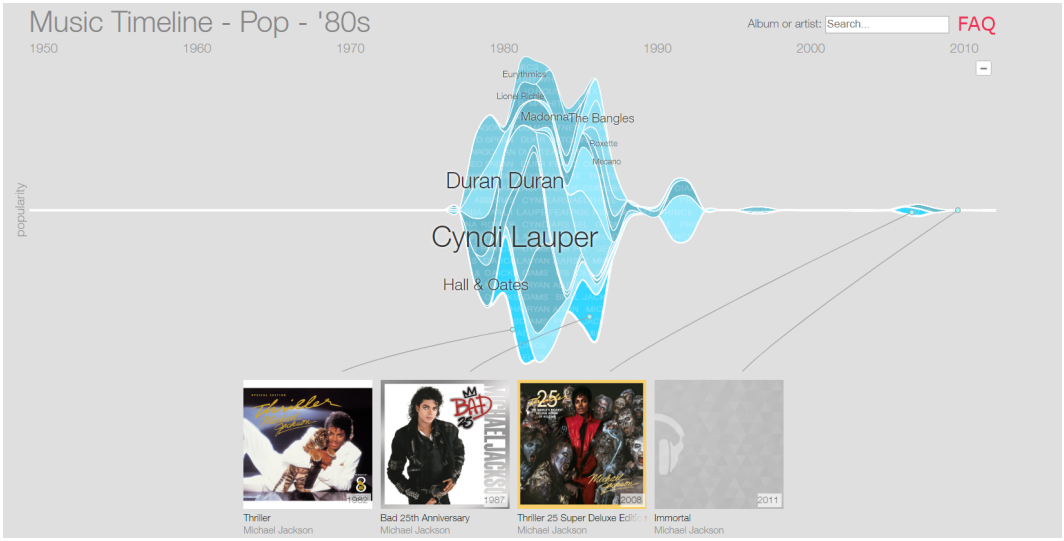

Figure 21: MJ albums under the 80's pop subgenre artist overview.

A second approach is to input "Michael Jackson" into the search box provided by the visualization. Once we request the search, the service returns the following figure

Figure 22: The resulting page that returns when a user searches for "Michael Jackson" key.

In this view, we can easily see the top 2 albums that were the most popular which are "Thriller " and "Bad" in the 80's. The third album could be either "Off the wall" or "Dangerous". In reality, it is the latter one however since the stripes accounting for these albums on the visualization have quite similar areas, it is hard to make a conclusion. Finding albums in this way by searching eliminates the need for discovering all the albums of an artist one by one. However, it would have been much more intuitive if we could access all of his albums from the visualization itself.

2. How much does Led Zeppelin contribute to the total popularity of the subgenre Classic Metal for the period between 1960 and 1980? To be able to answer this question we have to select the first the genre, then the desired subgenre, in the subgenre level each artist is represented by a stripe and the thickness of each stripe,Although Led Zeppelin has the most contribution to the Classic Metal during 1960 and 1980 the label of the band is not visible therefore we have to check the texture of the stripe to find out which stripe represents Led Zeppelin, we can estimate the how much Led Zeppelin has contributed to the popularity of Classic Metal in the period between 1960 and 1980 which is around 50% but it is impossible to know the exact percentage of the contribution check Figure 21. The steps we had to take to answer to this question were pretty long, the non-existence of Led Zeppelin Label made it hard to find the stripe since the white texture is almost invisible on the yellow background, although we can estimate the percentage of the contribution of the popularity it remains just an estimation.

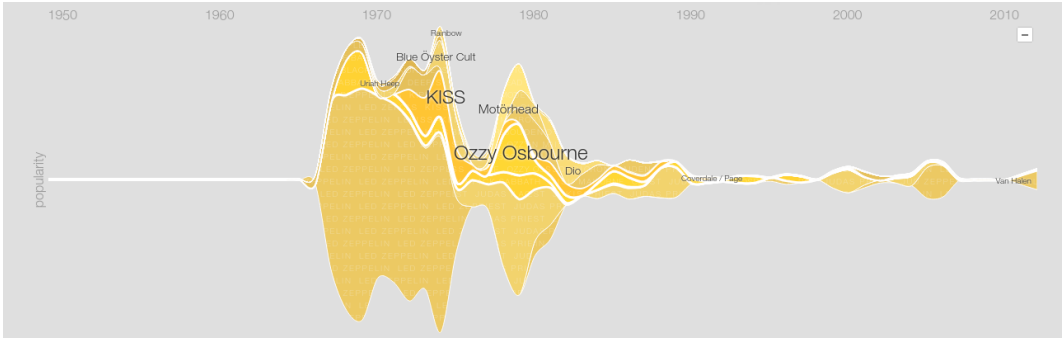

Figure 23: Contribution of Led Zeppelin to the popularity of Classic Metal subgenre in the period between 1960 and 1980.

3. How is the popularity trend of Korn changing over time: To find out the popularity trend of Korn over time we have tried two methods:

- **Method one:** in this method, we have to first know the genre and subgenre that the band

Korn belongs to, first we have to select the Metal Genre then select the subgenre Metal (it looks like Metal is a subgenre within the Metal genre in the visualization). Unfortunately, in the current visualization, we cannot isolate one artist from the other artists so the user has to try to follow the stripe of the popularity of the Korn band over the years which is very hard to do see Figure 23.

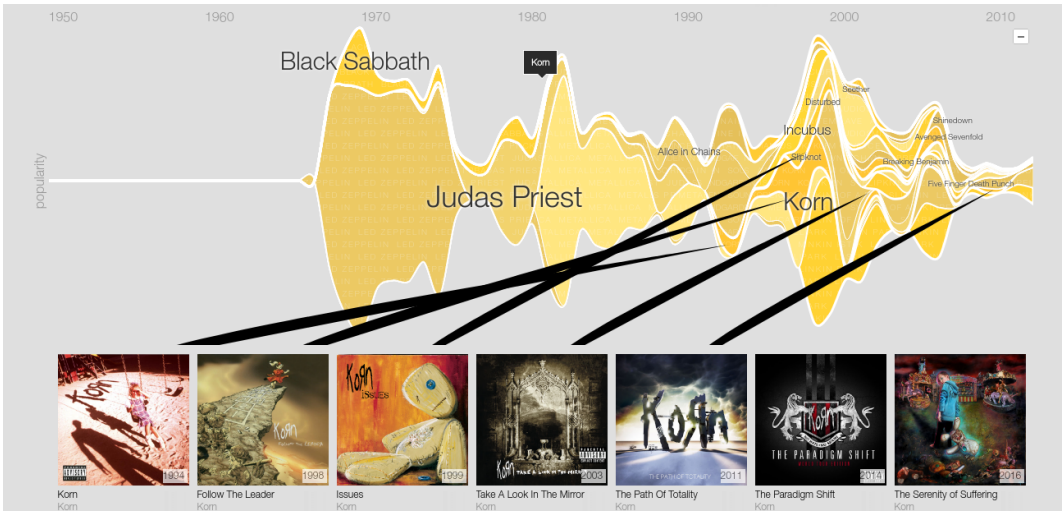

Figure 24: Popularity trend of Korn over the years

- **Method two:** the second method to answer the question is to use the search option to search for Korn, the problem with using the search option, is that we will get not only the Korn band but also albums and artists that have Korn word in them it such as Jiří Korn, Marc Korn ...etc which makes it not a good way to find the answer see Figure 24.

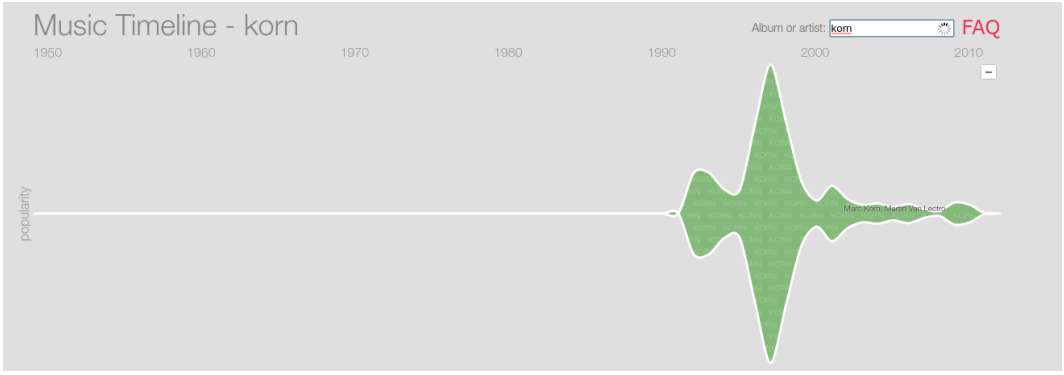

Figure 25: Popularity trend of Korn over the years using the search option.

We conclude that to answer to the popularity trend of Korn over the years we had to go through complex process going from the main visualization down to the subgenre visualization, find the right stripe representing Korn was hard since the label of Korn was overlapping with other stripes in method one, method two was straightforward and simple we only had to search for the word Korn.

The answer through the visualization using both methods is unsatisfying since we cannot get an accurate answer to the question. It very hard visually for the user to separate the stripe representing Korn band using the first method, and the stripe representing the popularity of Korn is not pure in the second method, but we can give less accurate answer that Korn was trending in the late 90's namely around 1997 then the trend line kept dropping after that.

9 Conclusion

Google Music Timeline does a fantastic job representing a huge amount of data in a compact tool. The compactness also comes with trade-offs in which the visualization has to sacrifice for finer level of details. These trade-offs are discussed and potential solutions are offered throughout the report.

Bibliography

[Mus] Music timeline. http://music-timeline.appspot.com. Last accessed on October 27, 2018.