Lineup Dataset Visualizations Using RAWGraphs

Sita Robinson

6/30/18

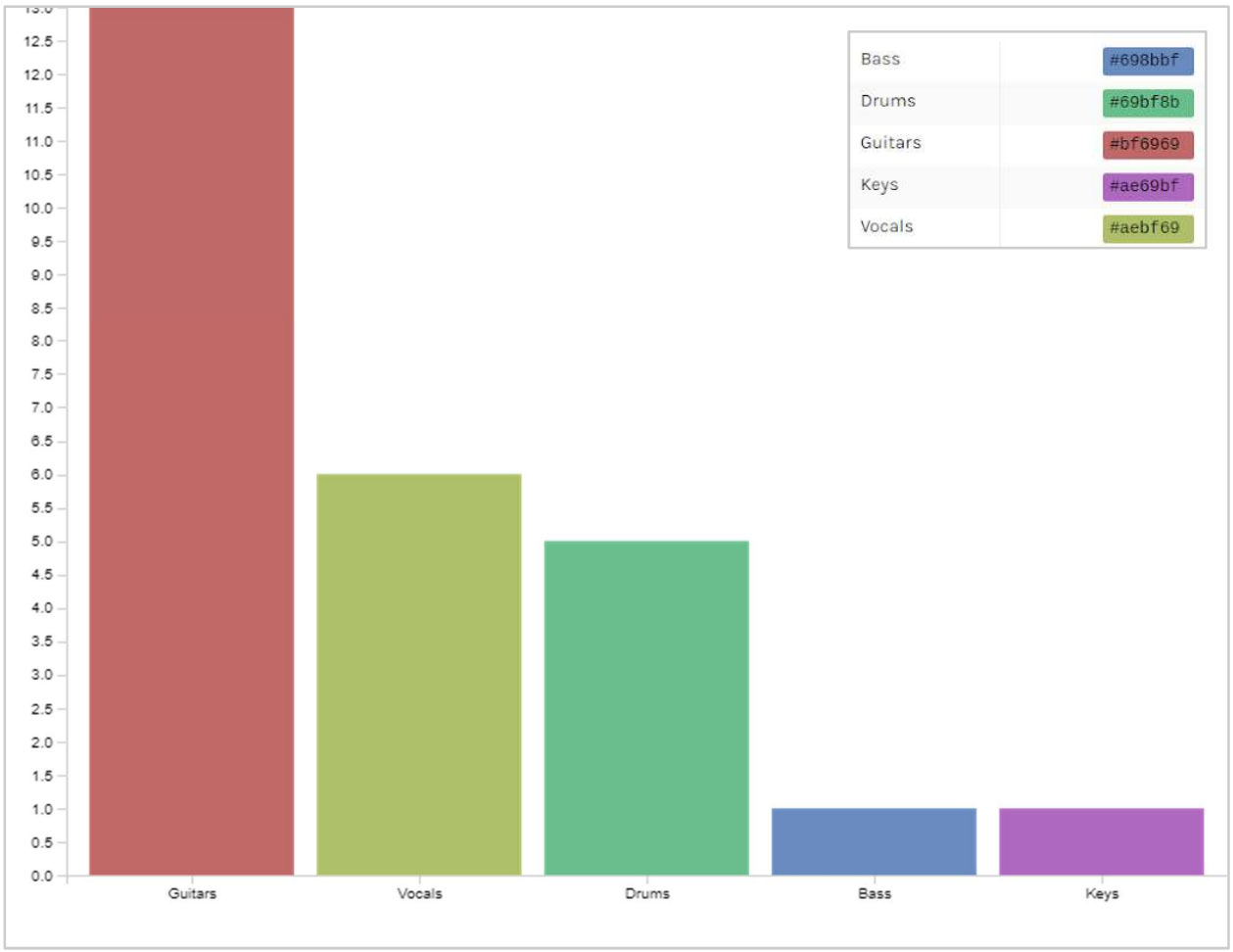
Introduction

The dataset chosen for this assignment was the "Lineup" dataset from the samples on the RAWGraphs website. RAWGraphs is a data visualization framework that is built on top of d3.js which is a javascript library for building data visualizations in web browsers. To pick this sample, I first went to the RawGraphs website, clicked on the "USE IT NOW!" button, and then clicked on the "Try Our Samples" button. The samples have the names of the dataset and the text under it describes the best way to visualize the data. By clicking on the Lineup dataset you will be able to see what the data looks like. I chose the "Lineup" dataset because it included columns with dates. A number of the visualization options RAWGraphs offers work best with date data. In this dataset there are four columns and 26 rows. In this dataset the four columns are Name, Begin, End, and Role. The data types are strings for Name and Role and dates for Begin and End. The data seems to represent different people with musical roles and the time they were in those roles. Since there is no description of the data and we do not know if this data is representative of the population, it is hard to make a decision on what the visualizations will tell us. However, we can try to make a general decision about what the data is trying to tell us. We also do not really know the source of the data since the RAWGraphs website does not make that available so it may not be the best. Out of twenty-one different visualization options, ten were picked to show in this report.

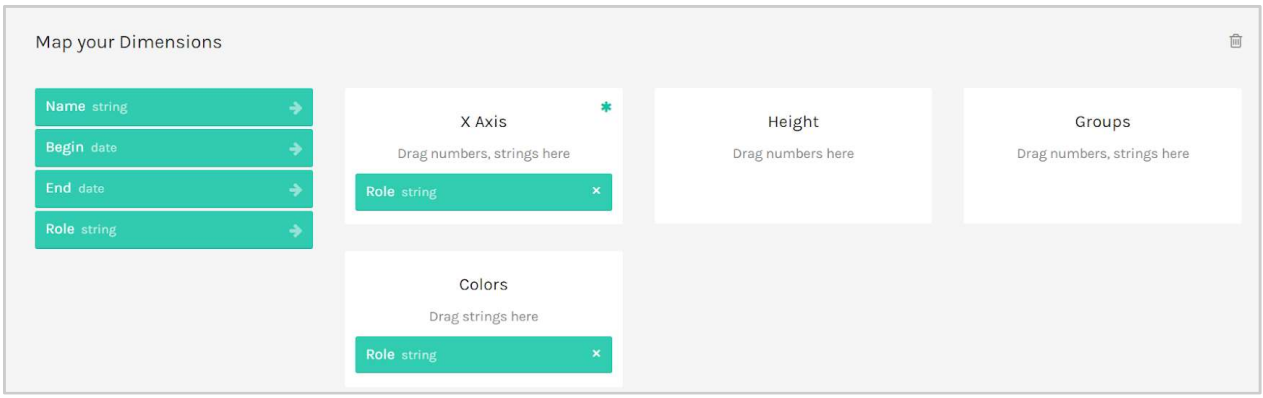
Visualizations

Figure 1: Number of Rows by Musical Role

First we can visualize the raw data to understand what we are looking at. This is a bar graph that shows the number of rows that are dedicated to each musical role. This graph was created by clicking the bar graph chart option and mapping the dimensions by dragging "Role" to the X Axis and the Colors option.



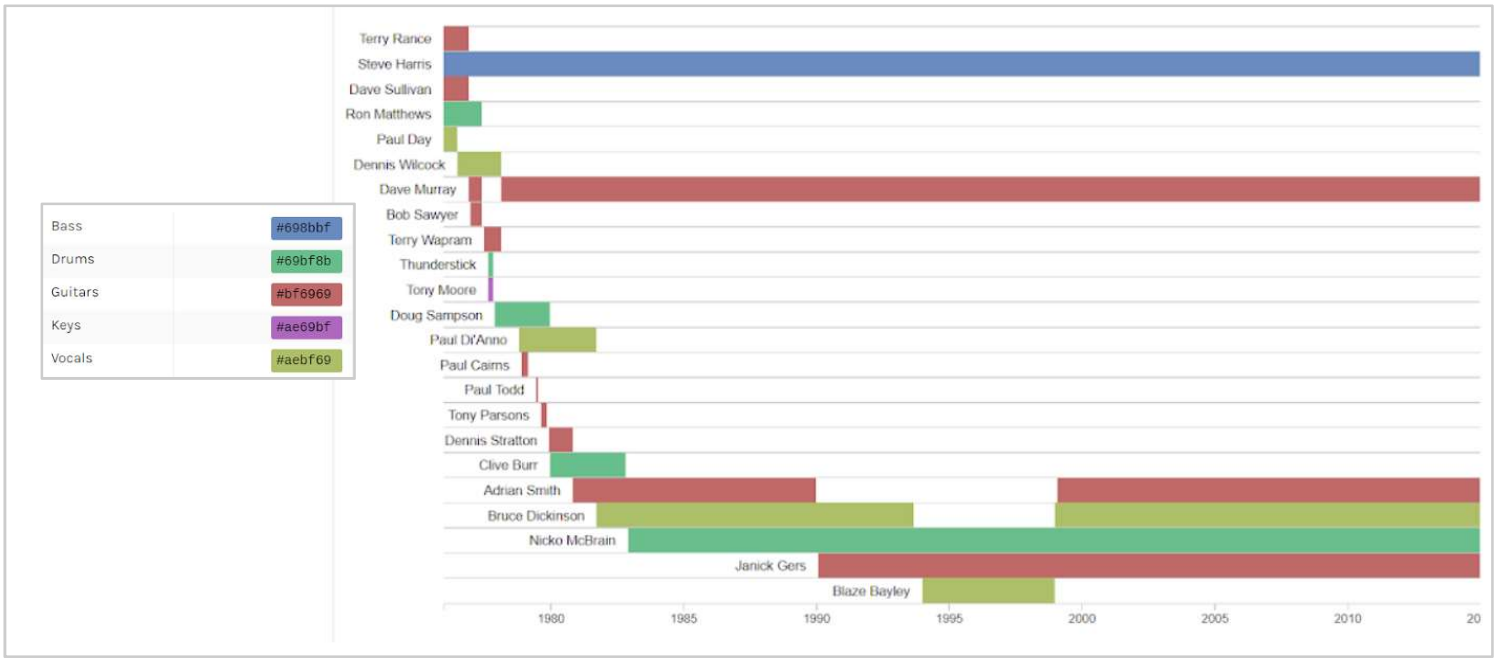
A bar graph using RAWGraphs allows for four filters X Axis, Height, Groups, and Colors. The Height option is not possible with the current data unless you were to add a numerical column which does not make sense to add. So in this case, the height is determined by the number of rows. The image below shows the options of the bar chart and what options were chosen to result in *Figure 1*.



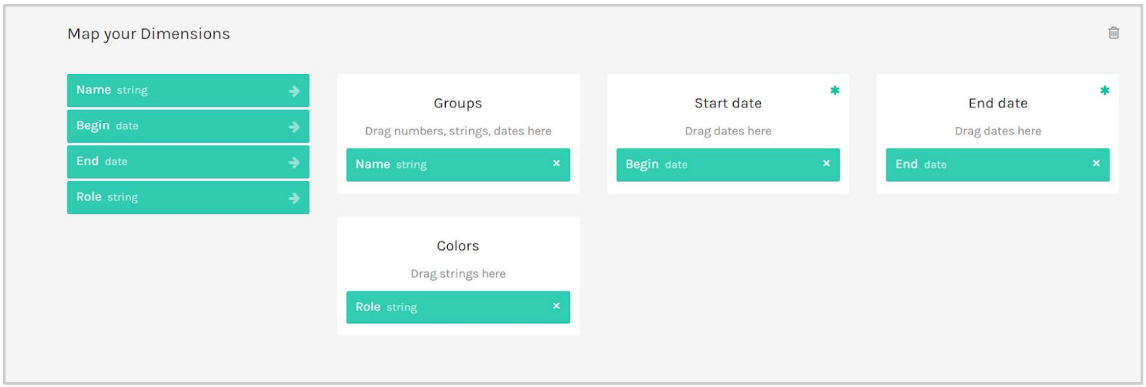
The largest role in this dataset is guitars followed by vocals, drums, bass and keys. There are thirteen rows of guitar, six who played vocals, five played drums, one played bass, and one played keys. However if you look at the data itself, it is not necessarily the number of people who played that specific role because there are people who took a break and started again. So there are multiple rows with the same people and the same rows. This graph does not account for that fact.

Key Findings
• The most number of rows in this dataset had the role of guitar (thirteen rows).
• The least number of rows in this dataset had the role of bass and keys (one each).

Figure 2: Gantt Chart Showing the Years When Musicians Were Doing Their Roles



The gantt chart seems to be one of the best methods to show the data since there is a start and end date. This graph was created by clicking the gantt chart option and mapping the dimensions by dragging "Name" for Groups, "Begin" for Start date, "End" for End date, and "Role" for Colors. The image below shows the options of the gantt chart using RAWGraphs and what options were chosen to result in *Figure 2*.

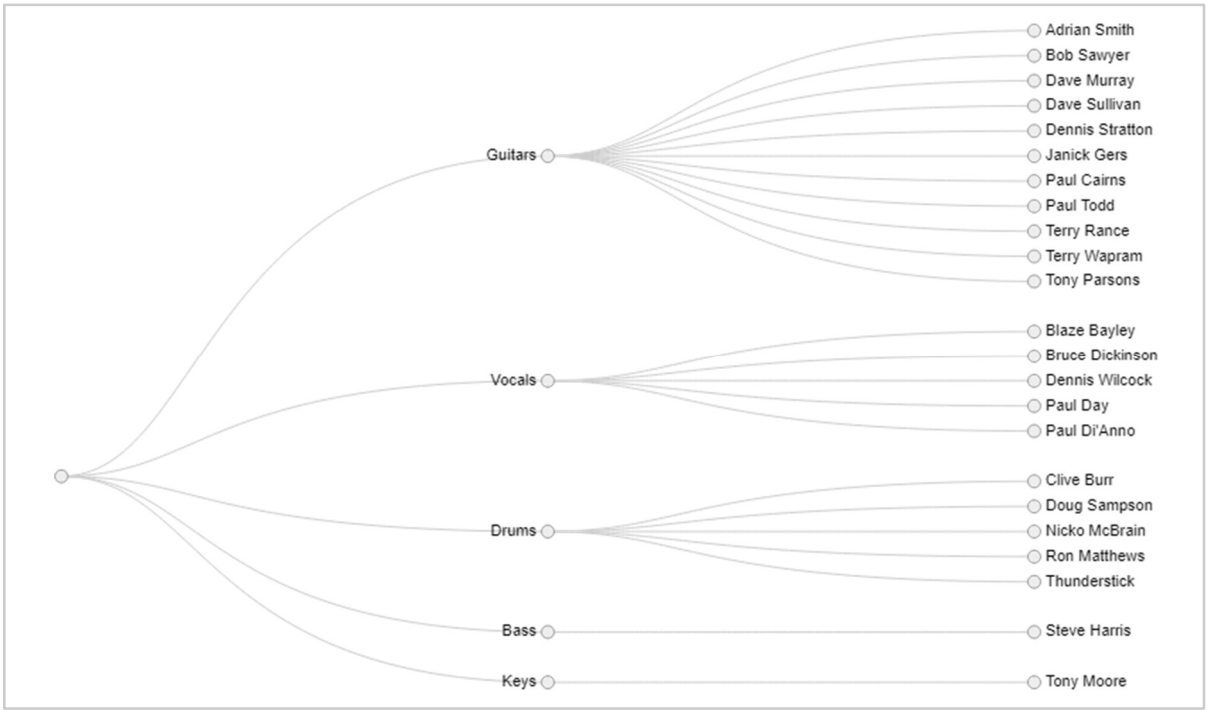


As you can see most of the people seem to have done their role continuously with the exceptions of Adrian Smith, Brian Dickinson, and Dave Murray. Paul Todd, Tony Moore, Thunderstick, and Tony Parsons seem to have short times in their roles.

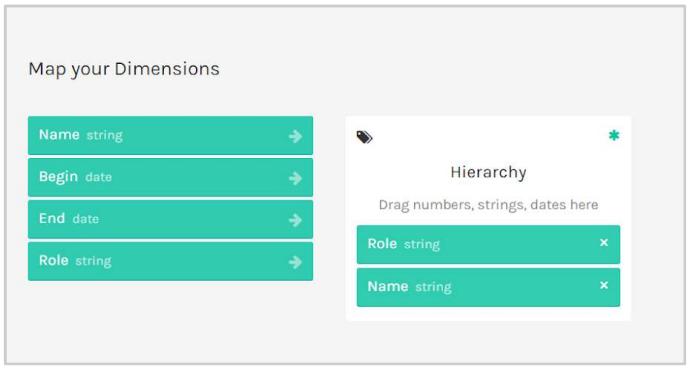
Key Findings
• A few musicians did not do their role continuously such as Adrian Smith, Brian Dickinson, and Dave Murray.

- Some musicians seem to have played their role for only a short period of time.

Figure 3: Distribution between Roles and the People Playing those Roles in the Dataset



This cluster dendrogram diagram was created by clicking the cluster dendrogram chart option and mapping the dimensions by dragging "Role" and "Name" for the hierarchy. The image below shows the one option of the cluster dendrogram and what options were chosen to result in *Figure 5*.

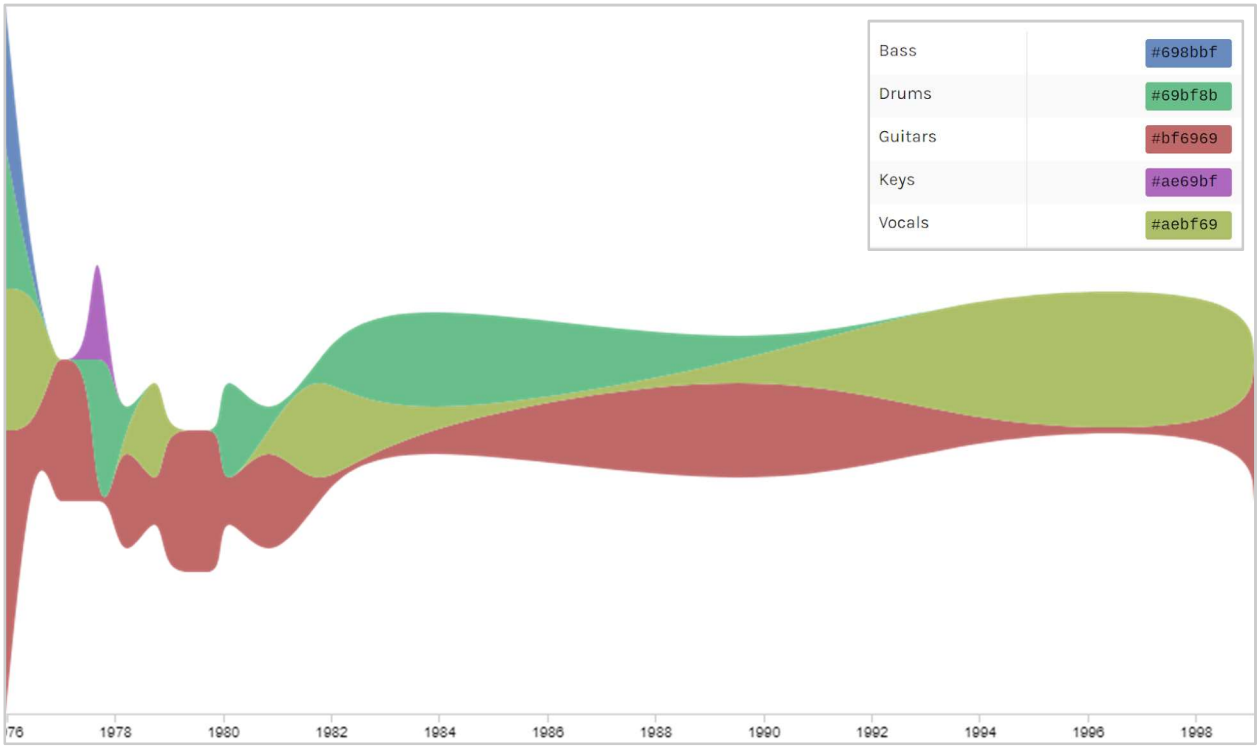


This cluster dendrogram diagram shows the distribution between the roles and people. Very few people played bass and keys in this dataset while guitars had the most number of people. Drums and vocals had five people each in this dataset.

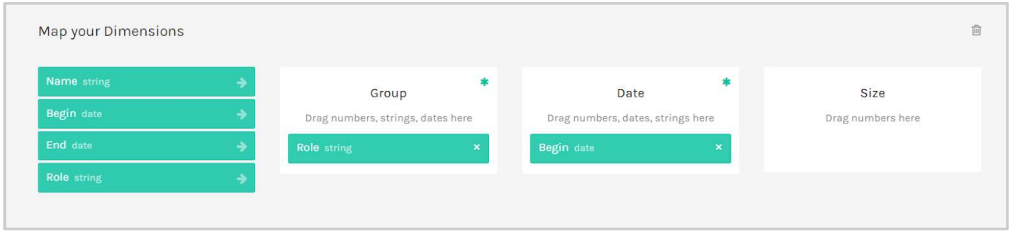
Key Findings

• Most people in this dataset played guitar (eleven people).
• The least amount of people in this dataset played bass and keys (one each).
• There are five people each with the roles of vocals and drums in this dataset.

Figure 5: Start Dates of Different Role Types



This visualization is called a streamgraph and there are three dimension options Group, Date, and, Size. To get *Figure 7* "Role" was chosen for Group and "Begin" for Date which can be seen in the image below.



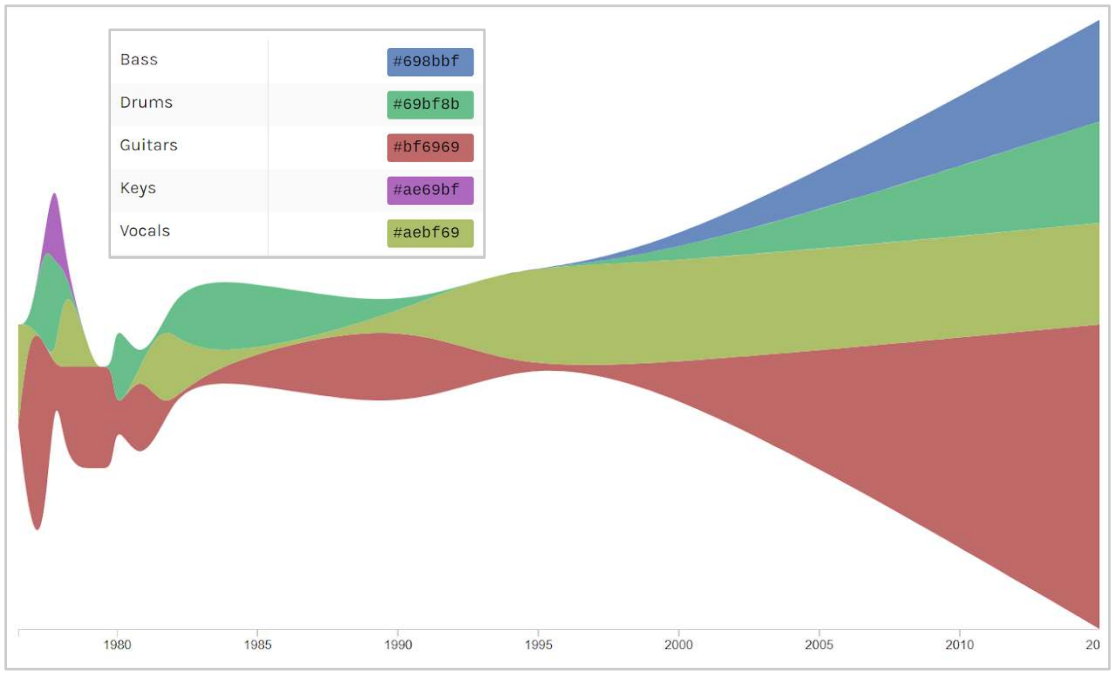
This graph shows the start dates of the musicians. The bands will become thicker as more people There seems to be a gap in musicians starting to play drums after 1994 in this dataset. Keys also

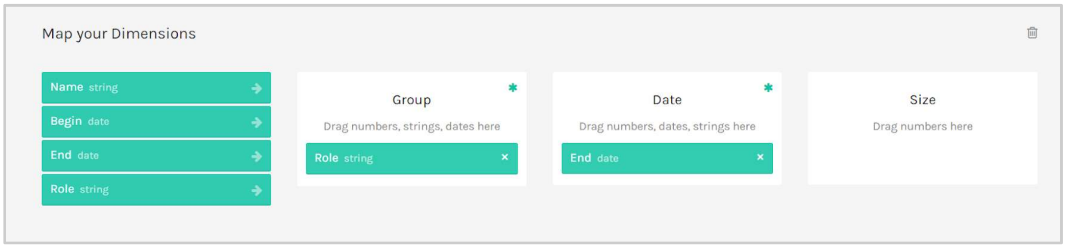
seems to be only have been started in 1977 but by looking at the raw data and other visualizations we can see it makes sense since there was only one person represented. The same thing happens with bass since there is also only one person in the dataset who played that. This graph is difficult to derive meaning from especially given the low volume of data.

Key Findings
• There does not seem to be any musicians starting to play drums after 1994.
• There needs to be more data overall but specifically for bass and keys since there is only one each.

Figure 6: End Dates of Different Role Types

This visualization is called a streamgraph and there are three dimension options Group, Date, and, Size. To get *Figure 7* "Role" was chosen for Group and "End" for Date which can be seen in the image below.



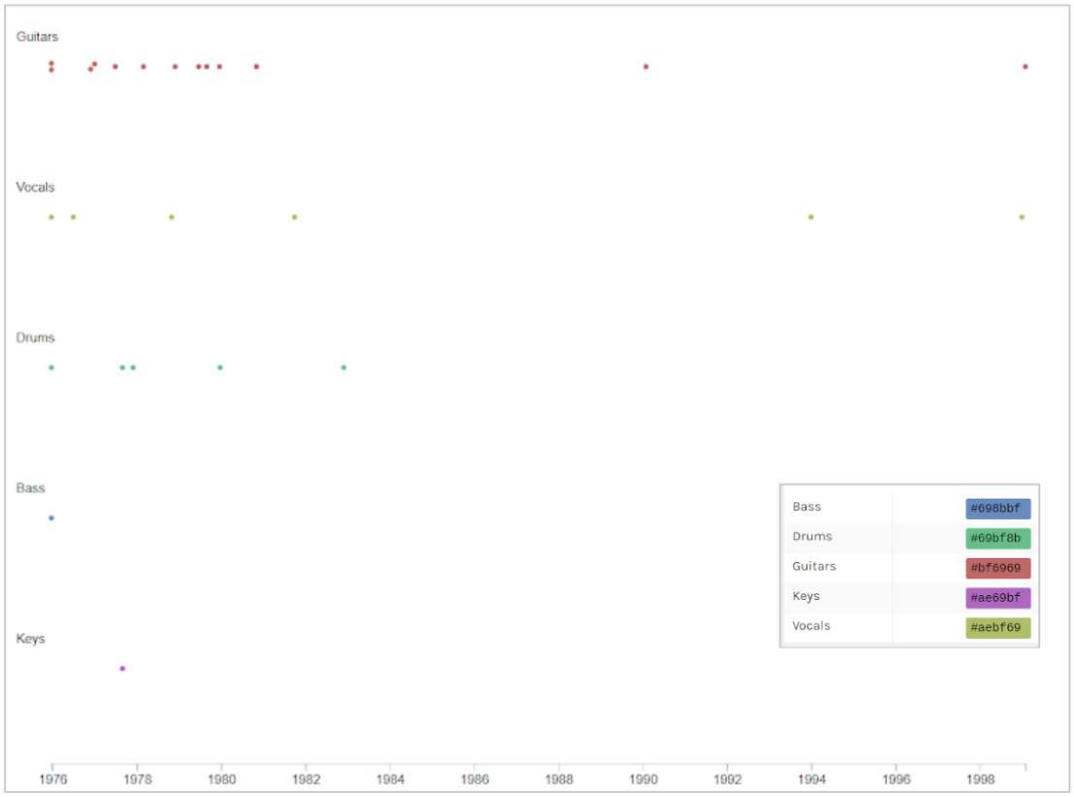


This graph is similar to *Figure 5*, but shows end dates of musicians instead of start dates. The most number of people in this dataset seemed to end their roles in the years 1995-2015. There doesn't seem to be many people ending the role of keys but that is because there is only one row with the role of keys so only one end date. Also there doesn't seem to be any musicians in this dataset ending bass before 1995 but that is probably because there is also only one musician.

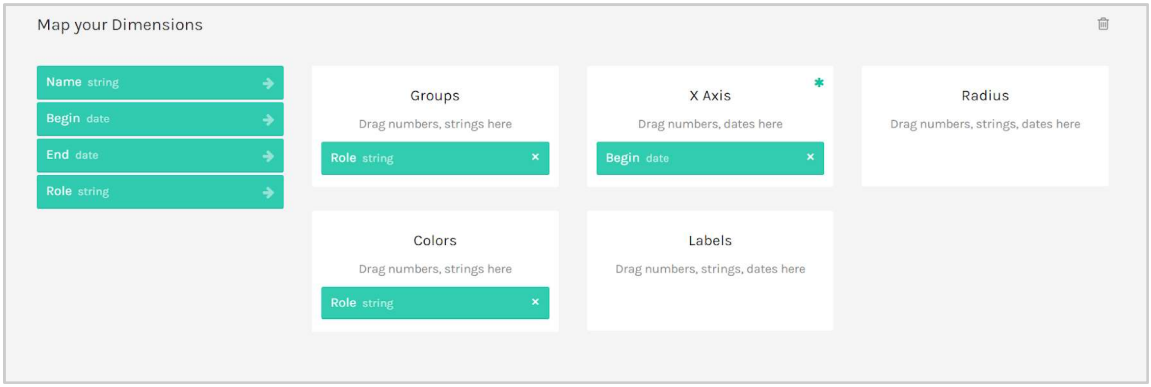
Key Findings
● There doesn't seem to be any musicians in this dataset ending bass before 1995. However, that is due to the lack of data for bass.
● There needs to be more data overall but specifically for bass and keys since there is only one each.

Appendix

Figure 7: Start Date Distribution of People By Role



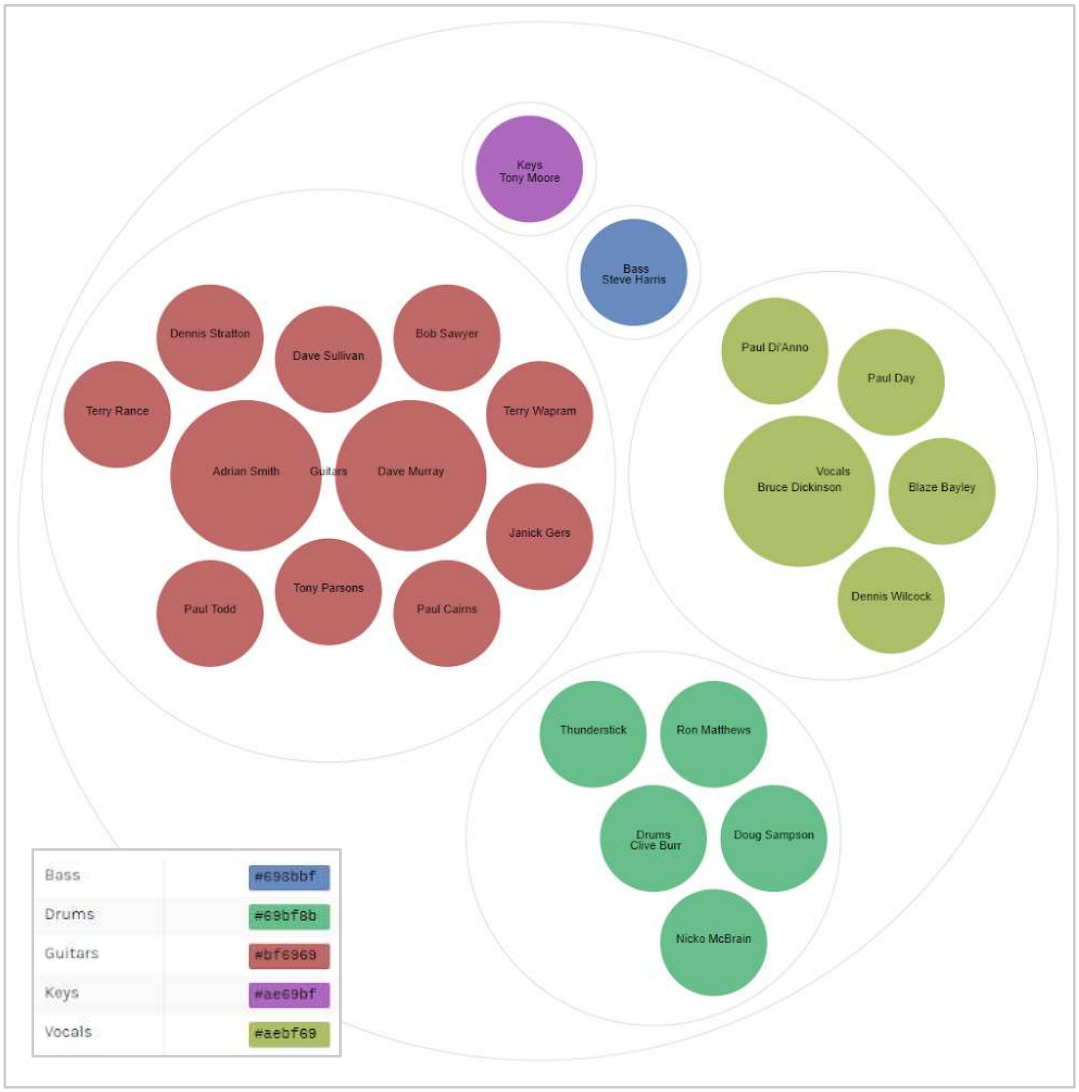
This graph is the beeswarm plot visualization. There are five options for the dimensions including Groups, X Axis, Radius, Colors, and labels. To make *Figure 7*, I dragged the "Role" column to Groups, "Begin" to the X Axis, and "Role" for Colors. The image below visually shows the options chosen.



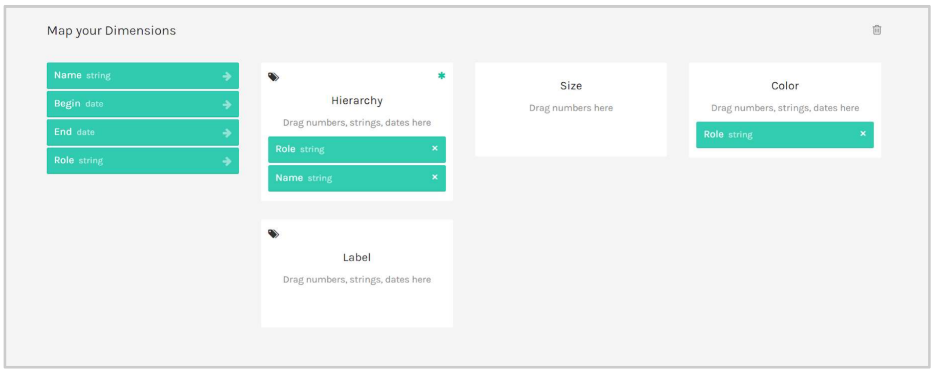
Most of the people in this dataset seem to have started around 1976-1983. However, there seem to be two vocalists and two guitarist rows that start musical roles later.

Key Findings
• Most of the people in this dataset started around 1976-1983.

Figure 8: People Per Musical Role



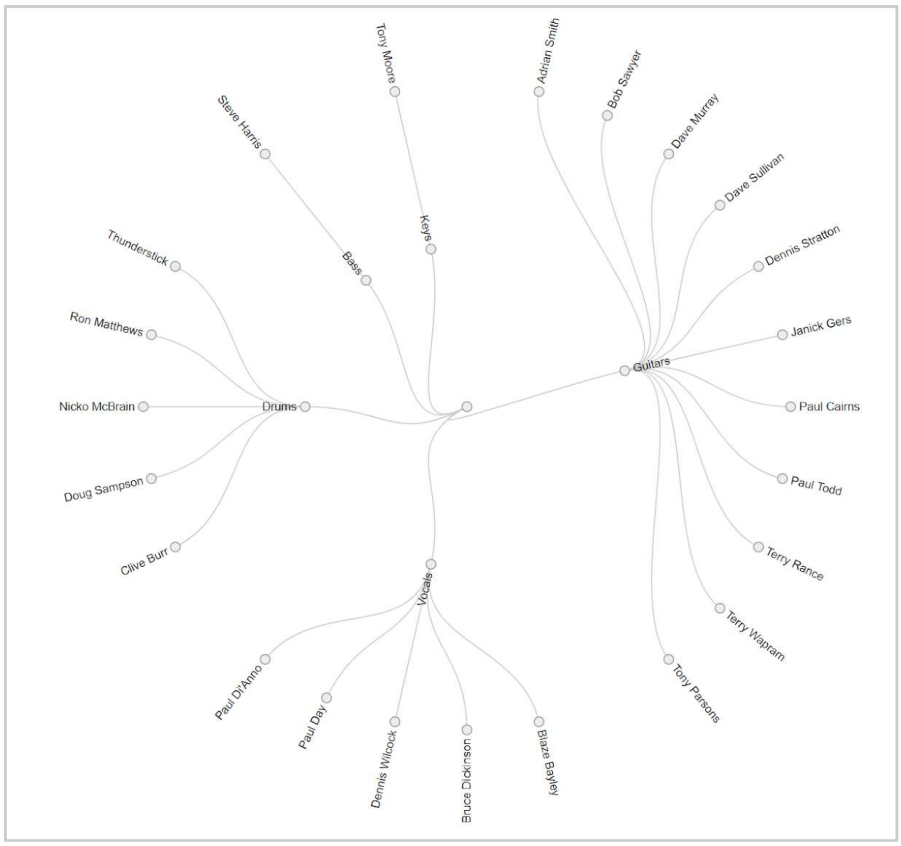
This graph was made using the circle packing option. This is another way to show the same information that is in *Figure 3* in a different format. The circle packing option has four ways to map the dimension of the visualization including Hierarchy, Size, Color, and Label. To get this visualization "Role" and "Name" were used for Hierarchy and "Role" for Color. The picture below shows the dimensions chosen to get *Figure 8*.



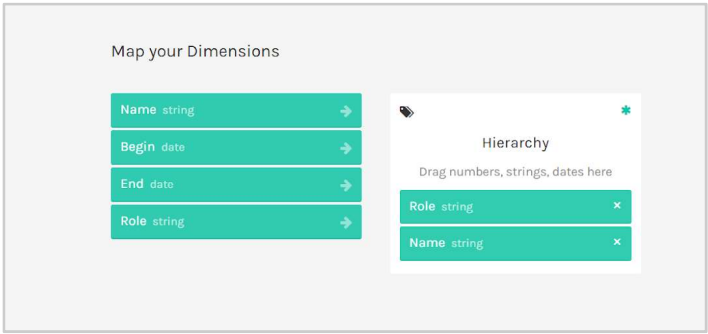
The most number of people in this dataset played guitars. While the least number of people played keys and bass. The circles are colored based on the role type and the size of the main circles are based on the number of people doing that role.

Key Findings
• Eleven people in this dataset played guitar, five people played drums, five played vocals, one played bass, and one played keys.

Figure 9: People Working Musical Roles



This graph shows the same information as *Figure 3* and *Figure 8* in a different format called a circular dendrogram. The circular dendrogram only gives one dimension option which is Hierarchy which I chose to drag "Role" and "Name" to get *Figure 9*. These options chosen can be found in the image below.



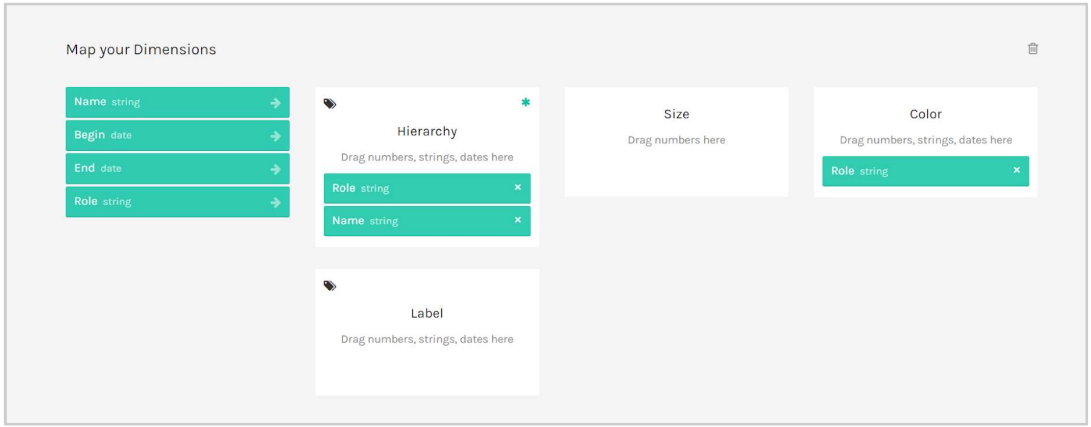
This is not the best format because the names are at different angles. However, this graph could be good in the situation where there are a lot of elements and hierarchy can be shown. The cluster dendrogram is nice but with too many elements the width might get too long. The visualization shows that the role with the most amount of people is guitar followed by vocals, drums, bass, and keys.

Key Findings

• Eleven people in this dataset played guitar, five people played drums, five played vocals, one played bass, and one played keys.
• A circular dendrogram can be used to take up less width space than a cluster dendrogram but is in a way harder to read because the text is at an angle.

Figure 9: Musicians working Musical Roles

Another way to show the same graph as *Figure 3, 8,* and *9* is by using a treemap. A treemap has four dimensional options which are Hierarchy, Size, Color, and Label. For Hierarchy "Role" and "Name" were chosen and it is colored by "Role". The image below shows visually the options chosen.



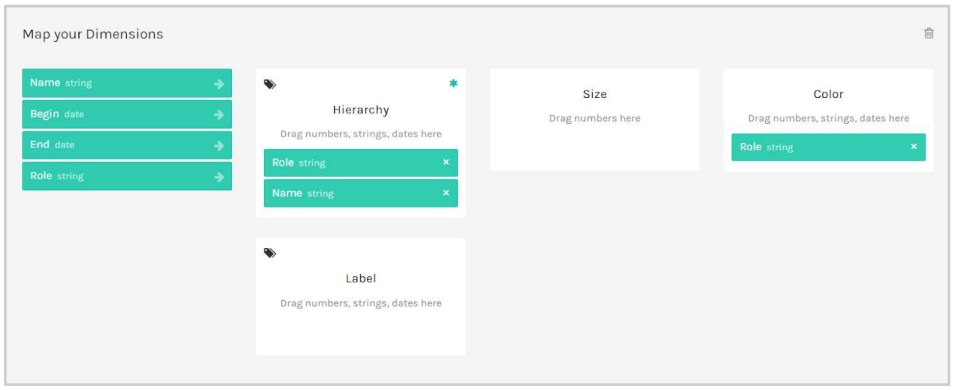
This does not seem like the best representation for this data. There is no telling the section the colors stand for without looking at the legend which is a flaw in the design. Restating from

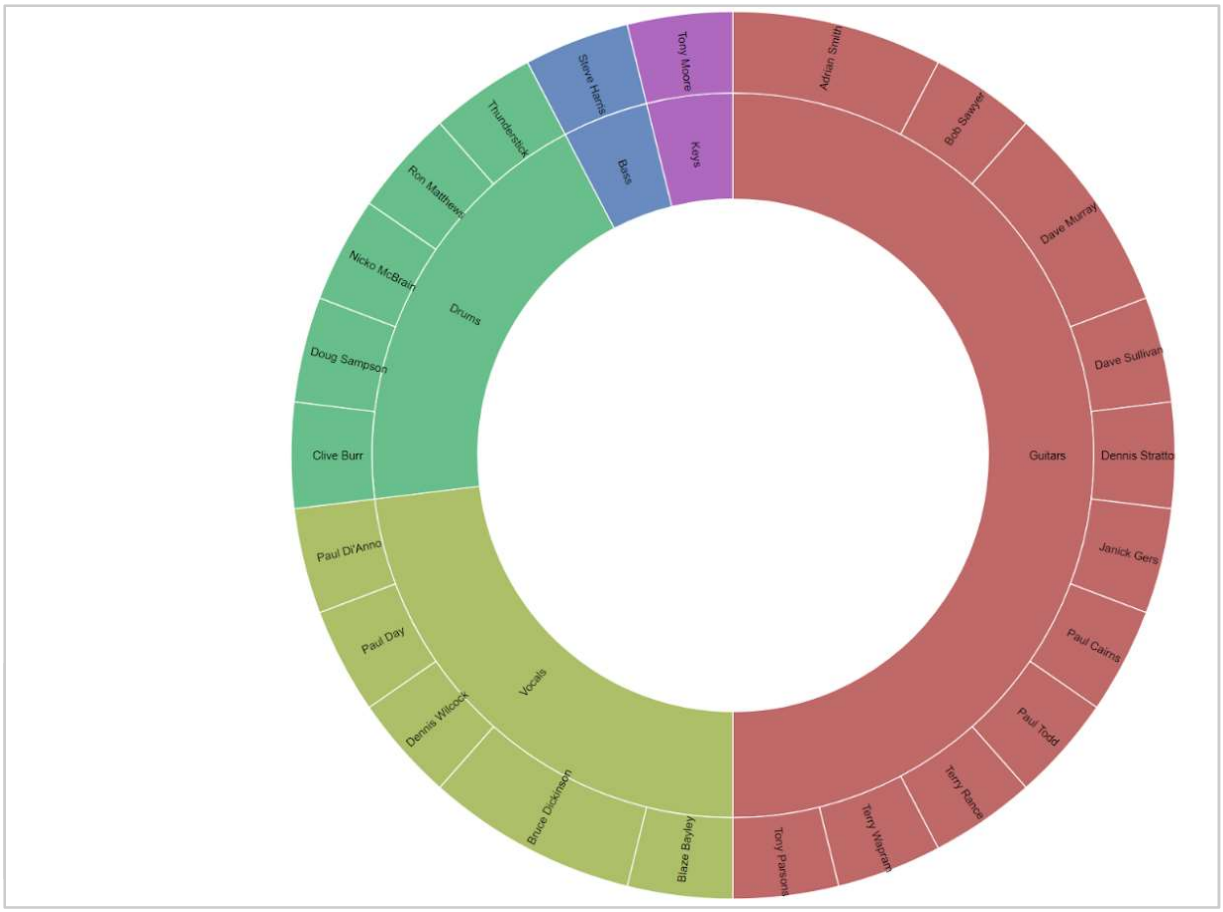
before it seems that in this dataset there are the most guitarists and the least amount of people played keys and bass.

Key Findings
• Eleven people in this dataset played guitar, five people played drums, five played vocals, one played bass, and one played keys.
• A treemap is hard to make sense of without the legend.

Figure 10: Groups for Each Musician

The same information as *Figure 3, 4, 8,* and *9* can be shown in a sunburst chart. The sunburst chart gives four dimension options which are Hierarchy, Size, Color, and Label. For Hierarchy the columns "Role" and "Name" were chosen and the Color is based on the "Role". There are no values that can be put for size since there are no columns that are numerical so the size is just based on the number of elements ("Names") for each "Group".

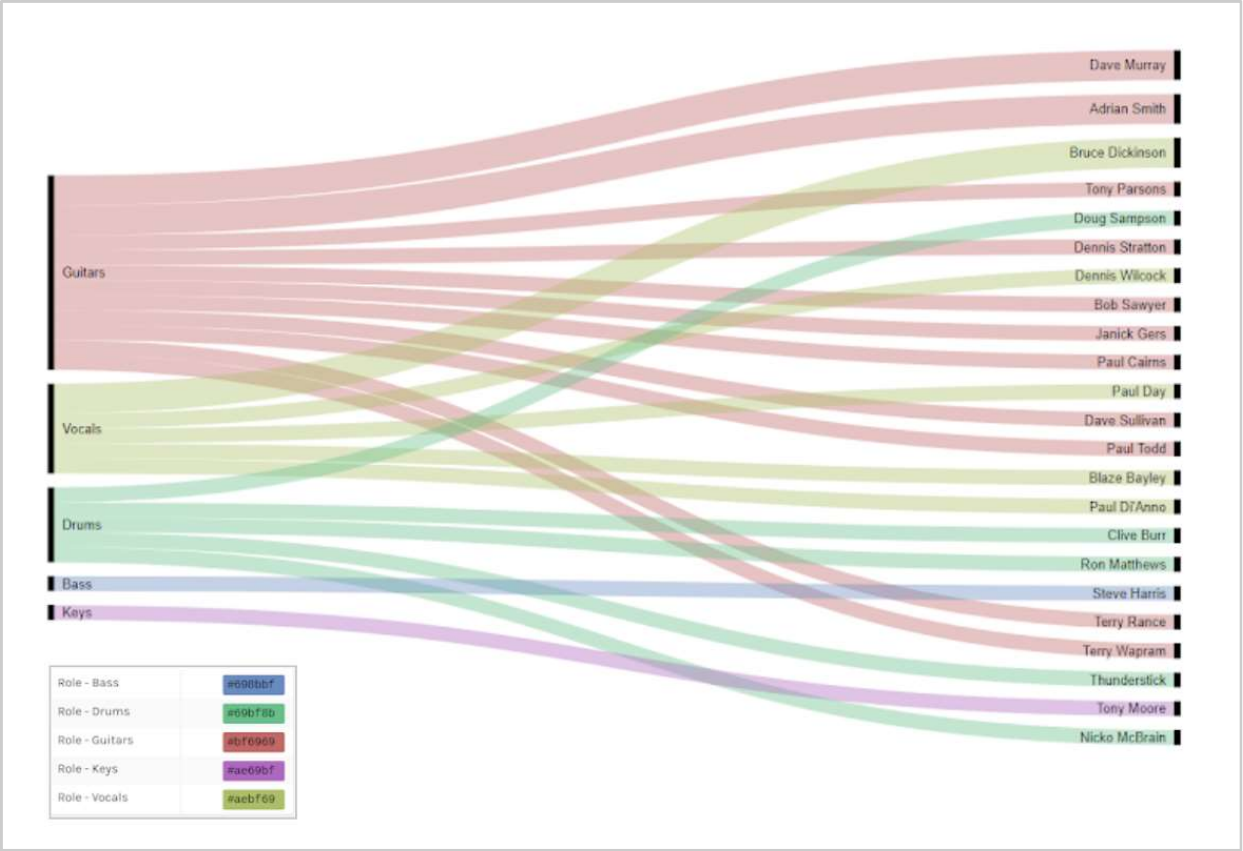




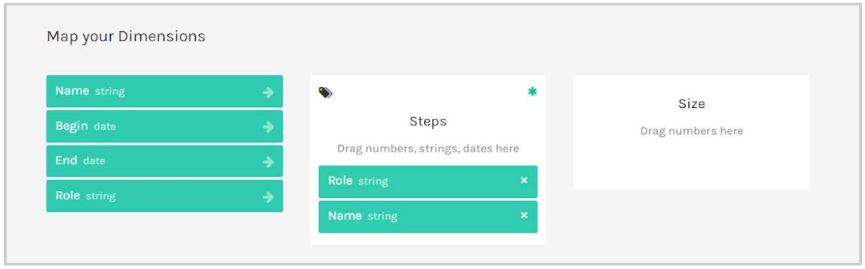
There are the most amount of people in this dataset who played guitars, the same amount (five) that played vocals and drums, and one each played bass and keys.

Key Findings
• Eleven people in this dataset played guitar, five people played drums, five played vocals, one played bass, and one played keys.
• A sunburst is hard to read because of the text angles. It is better than a treemap because you can tell the roles without looking at the legend.

Figure 11: Musical Groups for Musicians in this Dataset



This diagram is called an alluvial diagram. It shows similar information that the last couple graphs showed about roles that each musician performed. There are two dimensional options which are Steps and Size. For Steps I picked "Role" and "Name" and nothing for size since there are no numerical options.



Key Findings
• Eleven people in this dataset played guitar, five people played drums, five played vocals, one played bass, and one played keys.
• An alluvial diagram is a little hard to read with all of the overlapping lines.

Conclusion

The best visualizations that I found for the Lineup dataset was the gantt chart and the cluster dendrogram. Since the data includes start and end dates for each musician, the gantt chart is one of the best options. The cluster dendrogram is also good for showing hierarchy or in this case the musicians and the musical groups they were in. One problem I found with RAWGraphs is that you can't label the x and y axis. A big issue I found was the dataset since there is not enough data. Another issue sometimes the labels in a chart they tend to overlap by default making it difficult to read. In *Figure 8*, the circles with the text bass and keys were overlapping with their respective musicians. I tried to make the size of the graph bigger but that did not fix the problem so then I went into the HTML code to move the labels around so it was readable. There are limited options for legend formatting and placement. The colors can be changed but there are no preset color ranges.