Data Visualisation report

Brijee Rana
br1e21@soton.ac.uk

1. DATA STORY SUMMARY

My data story is about exploring the changes in the way people have listened to music. Technology has continued to advance in all areas including the way we listen to music. Hence, currently more and more people are using music streaming services compared to cumbersome physical copies such as CD's and vinyls. I visualise the sales of each music format from 1973 to 2019 in the US and discuss the changes due to technology. I then move to discuss what currently, most people are using to listen to music which is music streaming services like Spotify. I further investigate the growth of music streaming servicing by showing the rapid growth in users and revenue Spotify generates. My data story has changed from what was initially proposed as the idea was not interesting and the graphs did not provide compelling arguments to what songs do people listen around the world. Some questions did overlap as I know that there is alot of data being collected in music streaming sites.

2. DATASET SUMMARY

The data for the US music sales in different formats is from Kaggle which used information from RIAA. I then used datasets about spotify, music streaming services and polls from Statista who link the primary source which I scope and check the usability.To show the number of spotify users that are non-paying and premium members in the 3rd visualisation set, I joined to csv files and subtracted the overall members to the paid members to find the number of members who use the free-version of the app. Some of the csv files also contained years with yearly quarters in them which made it difficult to read so I corrected it.

3. VISUALISATIONS

3.1 Sales of music in different formats over the years

1. As shown in figure 1, I used an area chart to show the sales of music in different formats between 1973 to 2019. The viewer can see the different changes in sales and can select a format to see its single performance or view the sales in a particular year. [1]

2. This colour-coded visualisation allows the viewer to see the general trends in sales of different music formats, the peaks, inclines and declines are very clear. As it is an area graph, it is easy to see the sale of each format by looking at the proportions of each area as well comparing between the formats at each time period. [2]

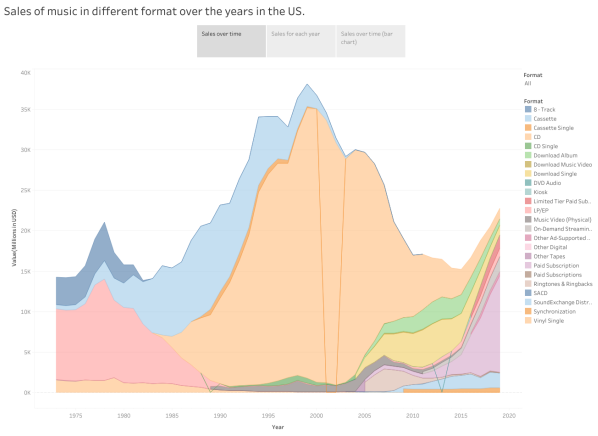

Figure 1: Visualisation 1

3. In terms of my story, it would be difficult to add narrative design patterns such as argumentation and emotion. Figure 1, is the first visualisation the viewer sees which is a timeline in the sales of different music mediums. This builds facts and events that frame the beginning the of the data story. [3]

4. Figure 1 shows a lot of great detail and information without complicating the information it is sending as you can clearly see inclines and increases in area and vise versa. However it does making trying to compare a single year of sales difficult as you would have to calculate the proportions from the area.

5. To improve I could add a scale which the viewer can adjust to see the difference in sales between 2 specific years and also add data from the rest of the world rather than just the US.

3.2 Sales of music in different formats in a year

1. As shown in figure 2, I use a pie chart than can be changed to each year which shows the proportion of distribution of sales for each format.

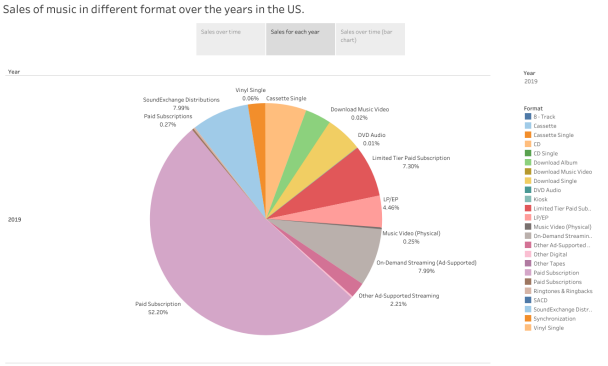

Figure 2: Visualisation 2

2. To cover the disadvantages of visualisation 1, visualisation 2 shows more in-depth information about the sales of music formats in each year. More specifically showing clearly the proportions of each sale of music format. The year can be moved around to observe the differences in sales. Furthermore, figure 2 is much easier to understand the data and does not require trends and patterns to be observed which can be sometimes misunderstood and missed in general.

3. Figure 2 is a natural flow to figure 1 as it helps strengthen the message of the data story. This is because figure 1 makes it difficult to analyse the sales of music formats in each year where as pie charts display relative proportions of multiple classes of data. Pie charts are also simpler and provide a better overview of the data in each year.

4. Pie charts are very easy to read and can show the distribution very well. It can be used to directly compare what is there in the data and make a instant point. However, it is more difficult to figure out the trends of the data. Piece charts are also less interactive to use and you cannot tamper with the classes as it distorts the data making it bias. Too much data can make the strength of the simplicity of a pie chart a weakness as it becomes difficult to read. Unlike visualisation 1, this visualisation does not show the total sum of sales, it cannot provide more insightful trends because a pie chart is harder to find patterns.

5. When there are a lot of different music formats being displayed in the pie chart it can get messy, I would change it next to remove negligible formats. As some of the values do matter perhaps I would change to a donut chart so that multiple values can be displayed. I could have 2 pie charts available at the same time so it would be used better for comparisons between 2 specific years.

3.3 Spotify's revenue throughout the year

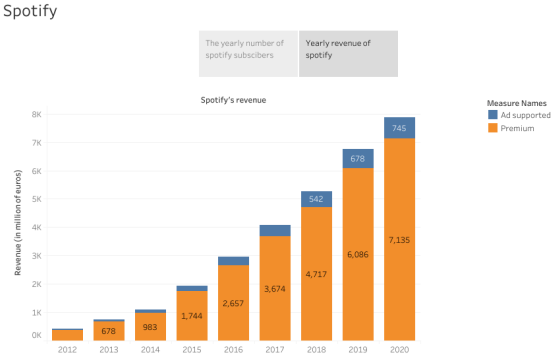

Figure 3: Visualisation 3

1. As shown in figure 3, I use a stacked bar chart to show the revenue of spotify with 2 sources of revenue. 1 is from premium subscriptions and the other from ads that non-premium users are given. In terms of interactivity, you can only choose to see one or all of the sources of revenue.

2. To continue to show how much music streaming services were growing I made figure 3 with also members that are premium subscribers and non-premium who is given ads for the streaming. That complements this graph which shows the revenue from ads and from premium members. This gave a lot of information as there are actually more non-premium members than premium members yet figure 3 shows that a significant portion of the revenue is from premium members. A stacked bar chart shows both groups as well as the total users in year while being a simple graph for no confusion or chart junk.

3. This graph is another structure to the data story that finally shows how rapidly music streaming has increased, especially in terms of revenue which is huge factor a music format being invested in more. The prior visualisation also helped figure 3 as it made the point that spotify is the most used music app in 2021.

4. It is far more readable and easy to understand than figures 1 and 2 in terms of showing changes over time. However, this visualisation requires further explanation or context. The strength of figure 3 is also its weakness as there is not a lot that the viewer can receive in terms of information but is simple and sends a clear message. When comparing, it is also difficult as each bar is not aligned.

5. To improve it, I would add more interactivity by adding more sources of revenue of spotify to the chart as well as costs which gives the viewer more information such as the yearly net profit. However I am aware that adding too much information to a stacked bar chart would make it difficult to compare as the values are all stacked in one, further access of interactivity to pick as choose certain variables must be implemented. Also, make all the labels visible or bars larger without warping them which adds bias.

4. CONCLUSION

To summarise, current technology has made music streaming the most common music format that people listen to and is exponentially growing(figure 3). Unfortunately, streaming sales do not reach the same high peaks as CD's at this moment in time (figure 1). Physical copies of music are being used less and less but some niche's like vinyls are still there(figure 2). Music streaming services will continue to be format our generation will use until a disruption in the market and supersedes it, perhaps in a couple of years.

For my audience, I wanted to start of with the initial question of how did we change the way we listened to music, that is where my first set of visualisations can be used to show the trends of each music medium. Then present how currently, music streaming is becoming very popular with figures on most number of app downloads. Finally, provide some insightful analytics on its members and how they account for the company's revenue as shown in figure 3.

This course has made me more actively scrutinize visualisations for any bias or chart junk. It has also made me appreciate people who can make data visualisations so aesthetically pleasing as I could not do that.

5. REFERENCES

[1] M. Islam and S. Jin, "An overview of data visualization," in *2019 International Conference on Information Science and Communications Technologies (ICISCT)*, 2019, pp. 1–7.

[2] C. Healey, "Choosing effective colours for data visualization," in *Proceedings of Seventh Annual IEEE Visualization '96*, 1996, pp. 263–270.

[3] M. Chen, D. Ebert, H. Hagen, R. S. Laramee, R. van Liere, K.-L. Ma, W. Ribarsky, G. Scheuermann, and D. Silver, "Data, information, and knowledge in visualization," *IEEE Computer Graphics and Applications*, vol. 29, no. 1, pp. 12–19, 2009.