1. **Four data analytics processes for data governance. List the features of R/tableau that can help you to provide visualisation according to data governance principles.**

Data governance is crucial for any business using data when it makes sure the information collected is handled consistently, securely, and effectively. Those actions are included in a set of principles and practices.​ There are many ways to go through the data analytics process for data governance. However, those practices can be sorted into 4 steps, in which one answers a big question.

Firstly is to answer the question why. The purpose for doing data governance can be to ensure business centricity (forbes), which are internal controls. According to Cheong, L, K and Chang . V (2007), those can include risk assessment, monitoring, communication,... (see table 1).

Second question is to find out what to focus on to reach effective data governance. The best possible Goals, success measures and funding can be found out by doing through data quality checking in R and tableau to make decisions on which aspects are more efficient and important to focus on.

Thirdly, When already deciding on which element to invest in. The company can divide roles based on types of governance mechanisms (according to Baijens, Jeroen; Helms, Remko W.) (see table 2). Based on that, the team can decide on when to do the steps based on the road map from develop statements to design the program and measure the report the insights found. R/tableau can be used in this stage effectively by tools to sort out and rearrange data (for example brom packages to clean data; pivot, select, group\_by to gather and modify data). Packages such as ggplot2 heavily aid the analyst in visualising the data needed to drive insight and impact business decisions.

Finally, the visualised data needs to be submitted to the stakeholders who are in need. The analyst job is to make sure the data are modified to that stakeholder preferences (for example retail departments need statistics on sales and KPI). R and tableau have helped the visualiser modify the data based on what have been done in part 3 conveniently and make sure no information is missed for the CEO, COO or CFO.

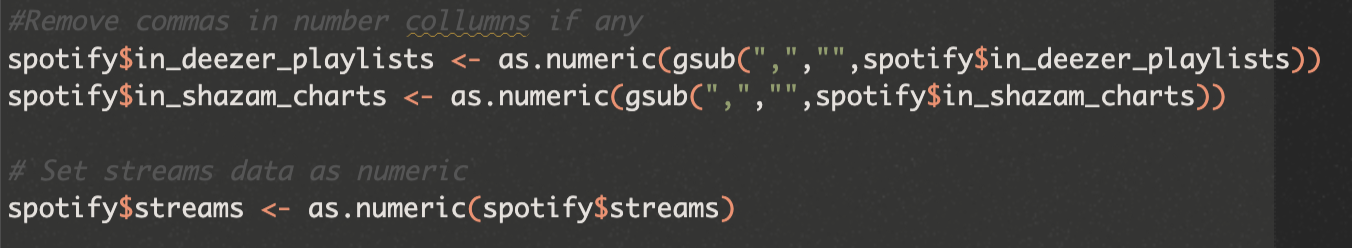
1. **Ethics is an important aspect of Data visualisation, discuss and apply any three ethics values that you have maintained.**

According to lecture notes, there are 4 ethical approaches: Rights, Utilitarian, Fairness, Common good. From a “Rights” and “fairness” perspective, all individuals have the right to access to the data without bias (not affecting business performance like confidential data). They also have the right for privacy when always being asked and known whether the shared data are under consent. With respect to that perspective, I always collected data from authentic sources to serve for my analysis knowing that no individual or organisation harmed.

Secondly, with the “common good” approach when collecting and sharing data, I always respect other works by understanding the ownership of data, information and knowledge. In that spirit, I always act on their contribution recognition by putting adequate reference and avoiding plagiarism.

Thirdly, I always ensure data transparency and honesty with respect to the originally collected data. The authentic data collected are clean and sorted by a way which caused no harm to the data original idea.

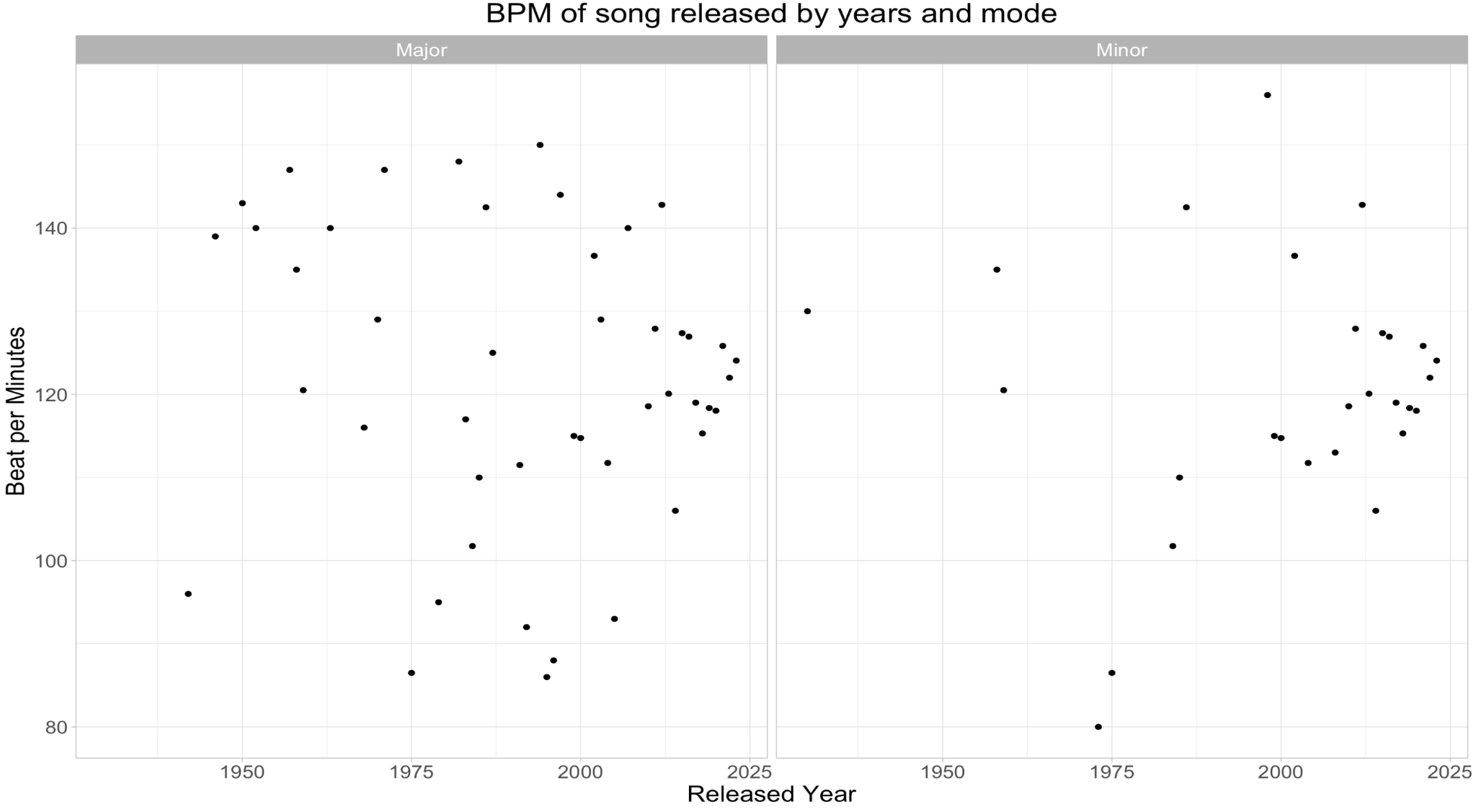
1. **Discuss Task A visualisation and provide sufficient justification on the prediction.**

****

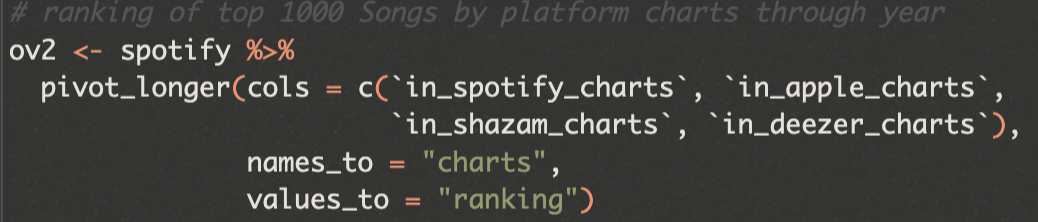
Data cleaning:

* Change blank data into NA if any
* Remove comma in numeric data if any
* Set number data to numeric data type.

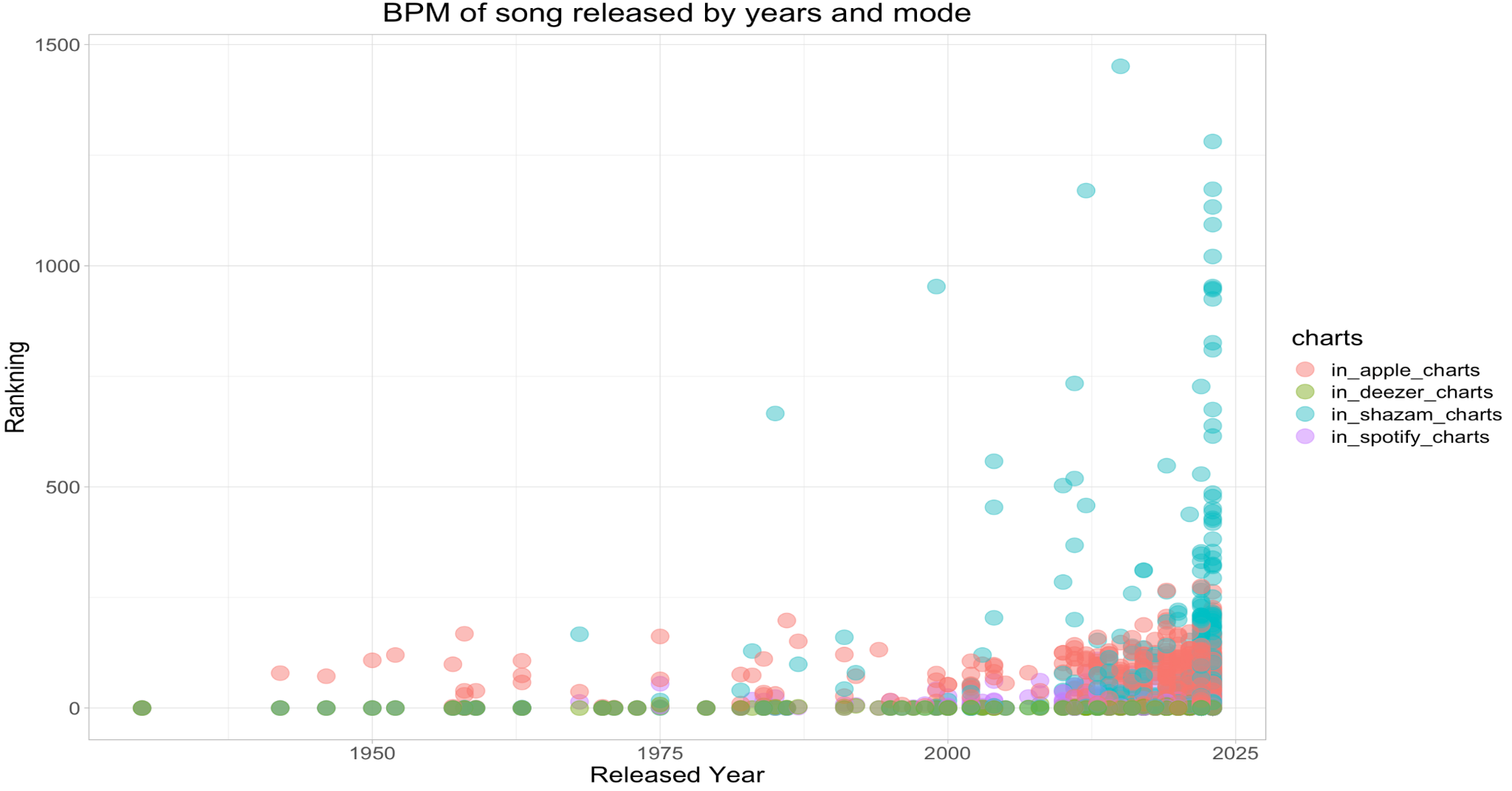
**Visualisation 1: Data overview**



**Graph 1\_1**

****

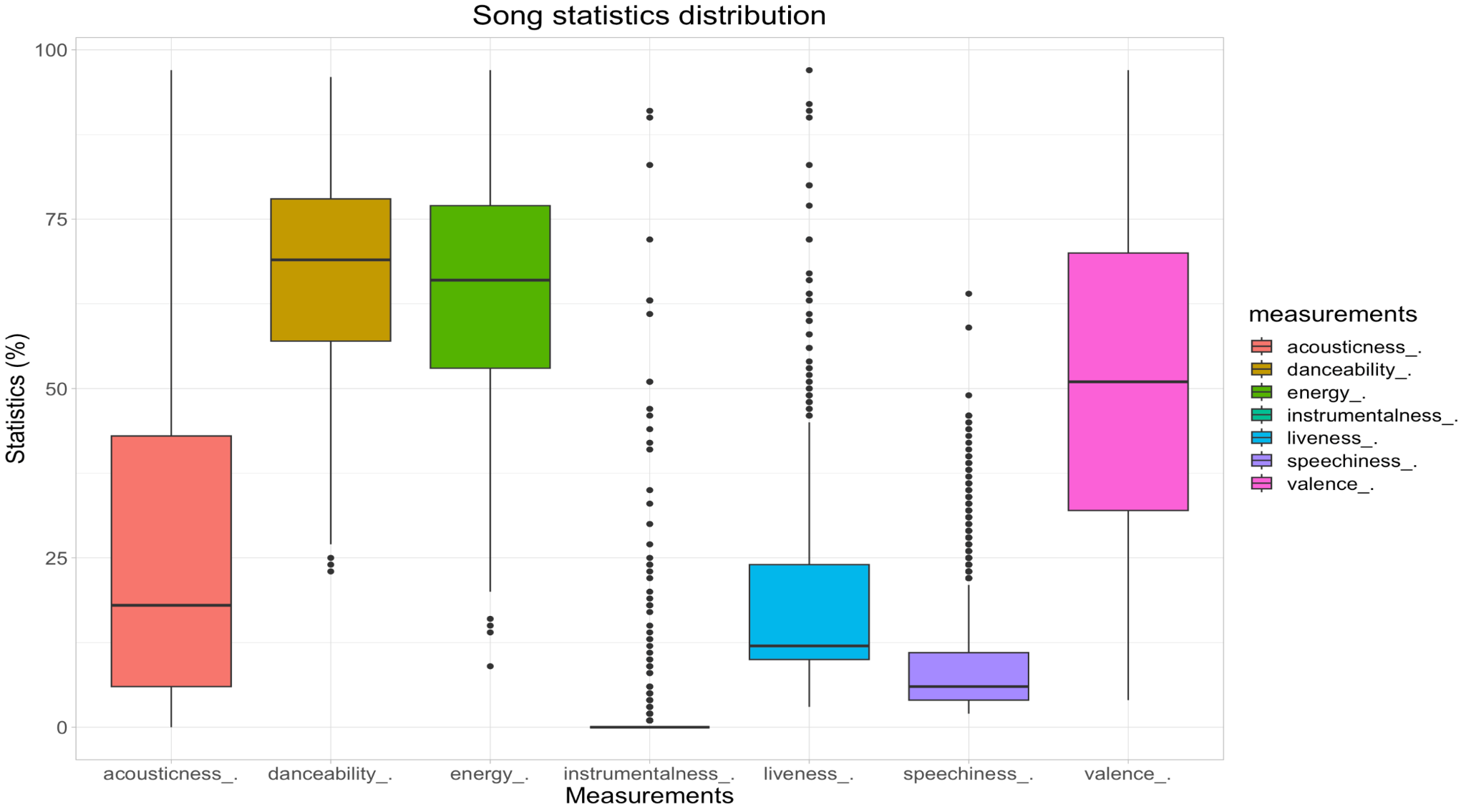
( convert 4 columns into 1 column with another one is value to help plot in the same graph)



**Graph 1\_2**

****

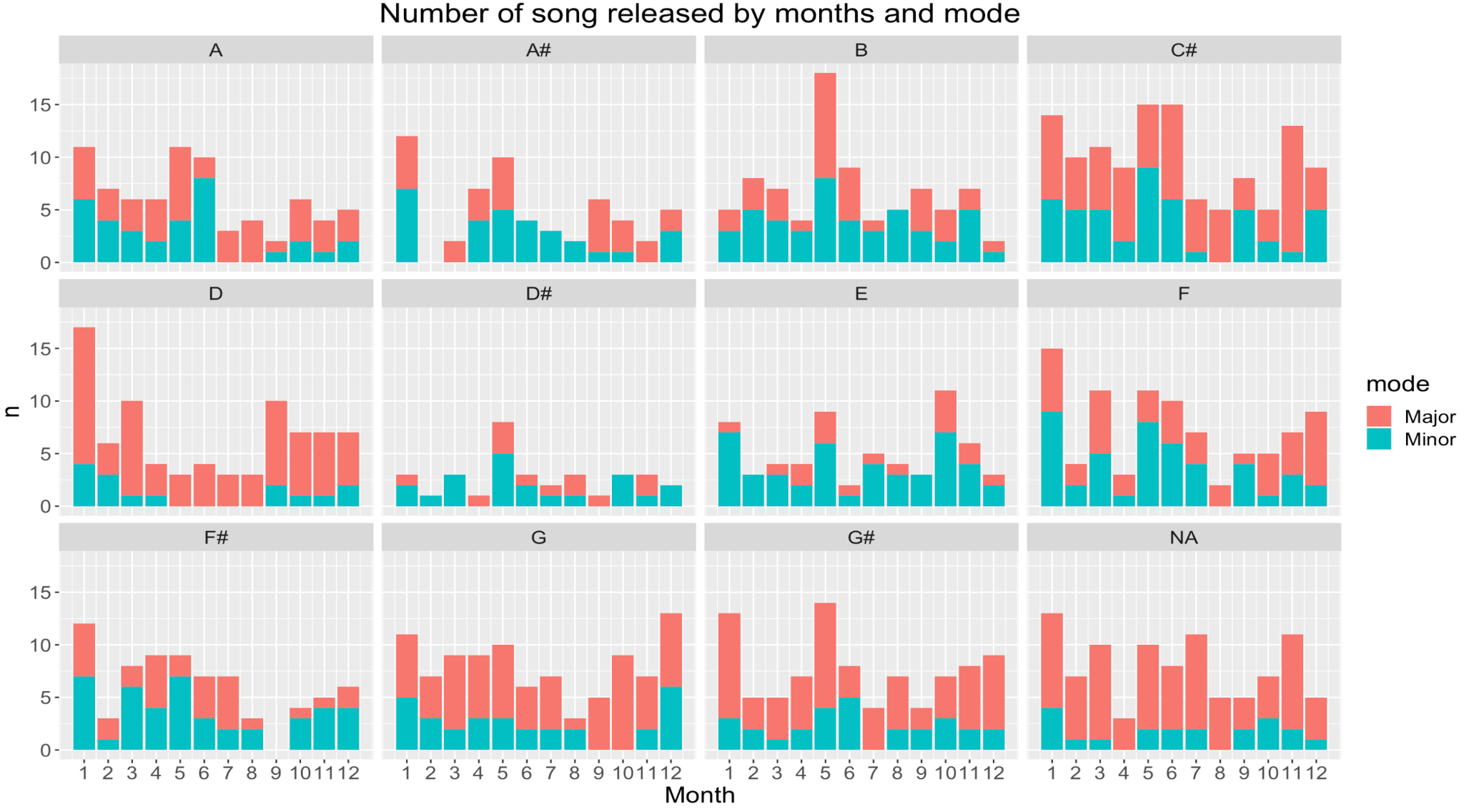
(convert columns from dancebility to spechhiness to 1 columns of measurement and one of statistics to show value to help plot in 1 graph)



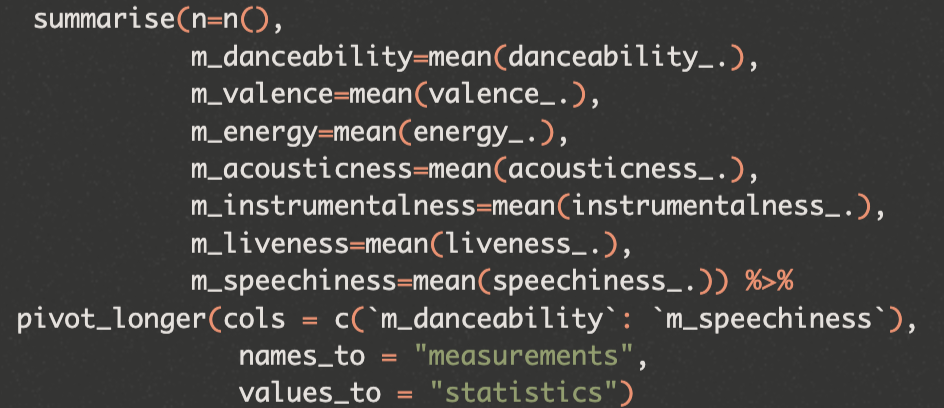
**Graph 1\_3**

Under visualisation 1, it can be seen that the beats per minute are dispersed all over the year with no specific pattern, while the major modes are more frequent than the minor (graph 1). In graph 2, the ranking shows the consistency between 3 charts except shazam, which gather at high rank. This indicates that the 3 consistent charts can reflect and predict well the preferences of users to popular songs. For song statistics, the danceability and energy always stay on top with 75%, which explain the factor causing song popularity while the widest variation comes from “acousticness” and “valence” mean that those variables do not significantly matters. Instrumentalness stay lowest saying the current taste on this is very low.

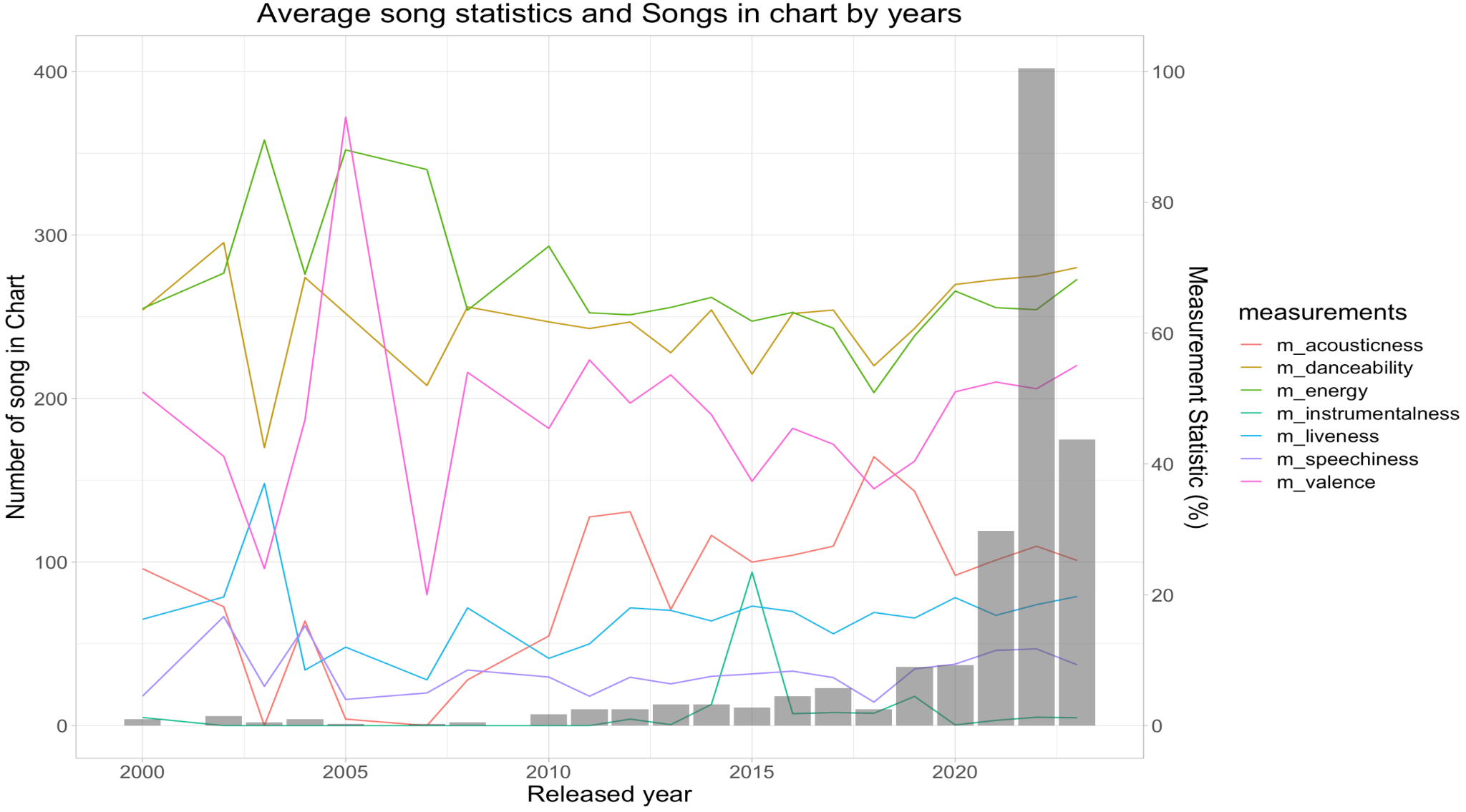
**Visualisation 2 : Song distribution and categories analysis**



**Graph 2\_1**



(summarise mean values and convert into 1 column)

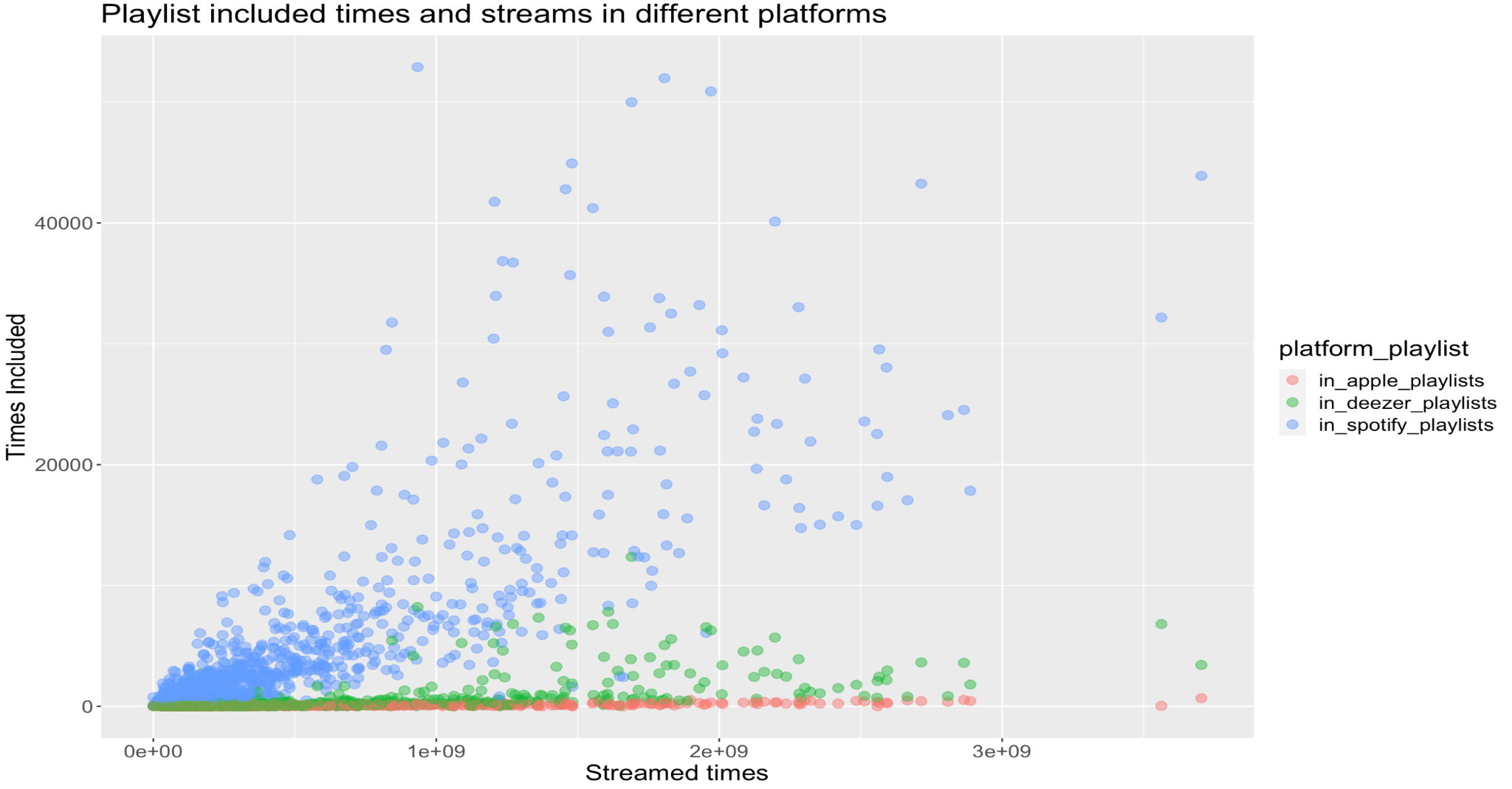


**Graph 2\_2**

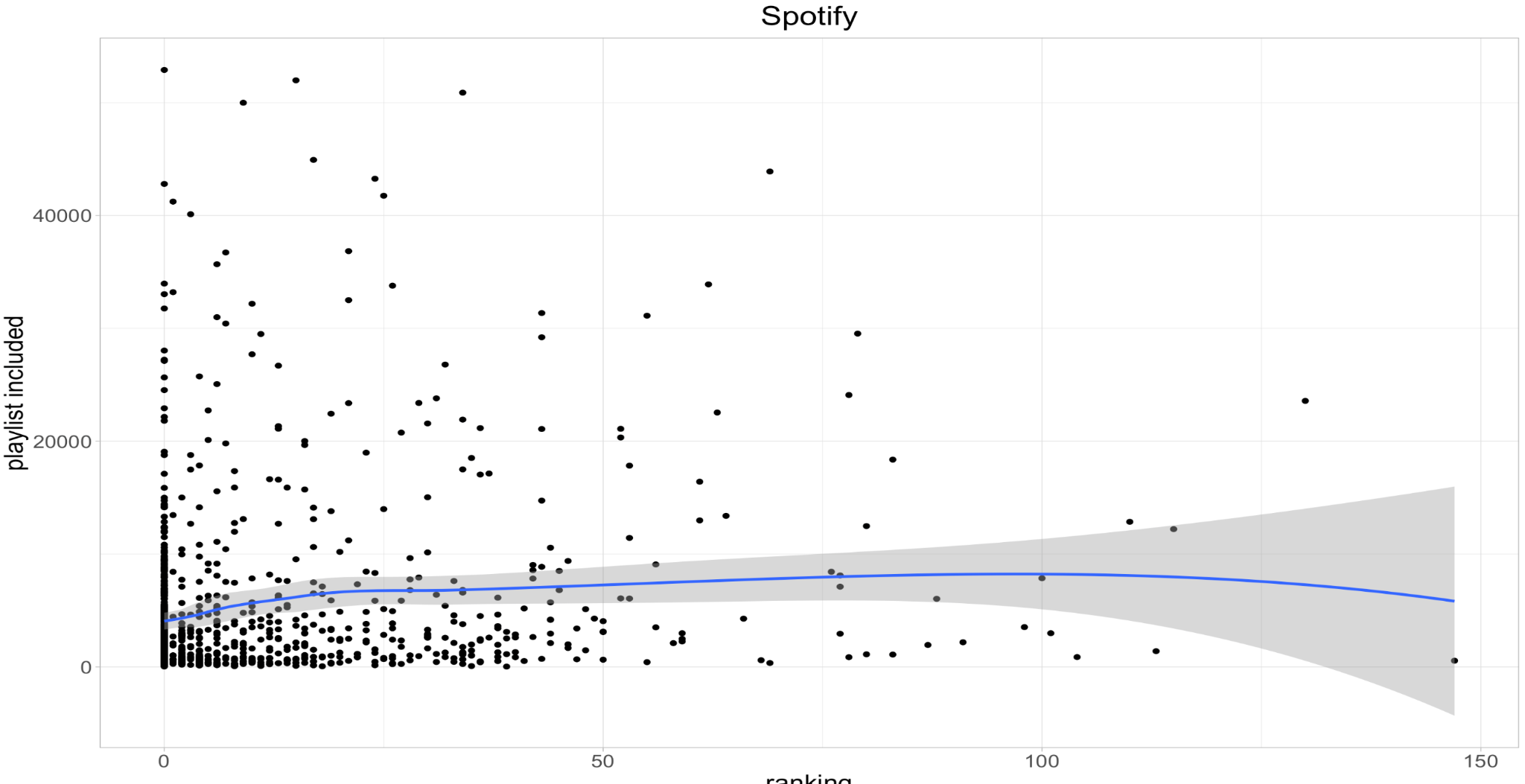
According to graph 1, most of the song are using key A, B, and C#. Those song releases focus into the first half of the year. While key D# and E are the least to be used where minor mode is dominant. The rest are all dominated by major keys.

In graph 2, we can see that the average statistic tends to remain over time its contribution (percentage) and less likely to change over the years except for the period between 2003 and 2007. Number of songs in the chart are dominant by 2022 and 2023, which is understandable are the data is of year 2023. It can be predicted that the number of song in chart will decrease the further the data getting away from that year

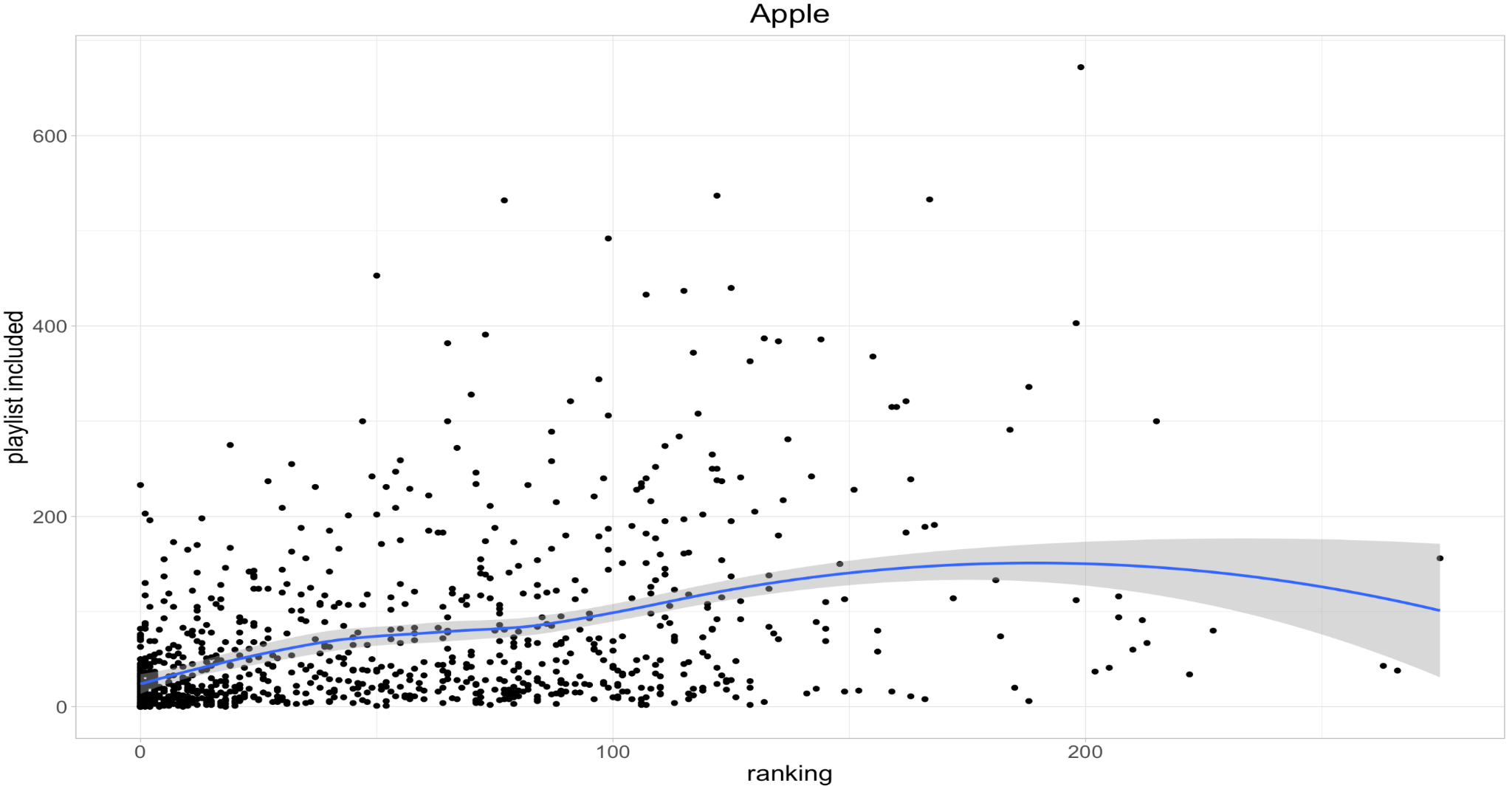
**Visualisation 3: streaming platforms analysis**



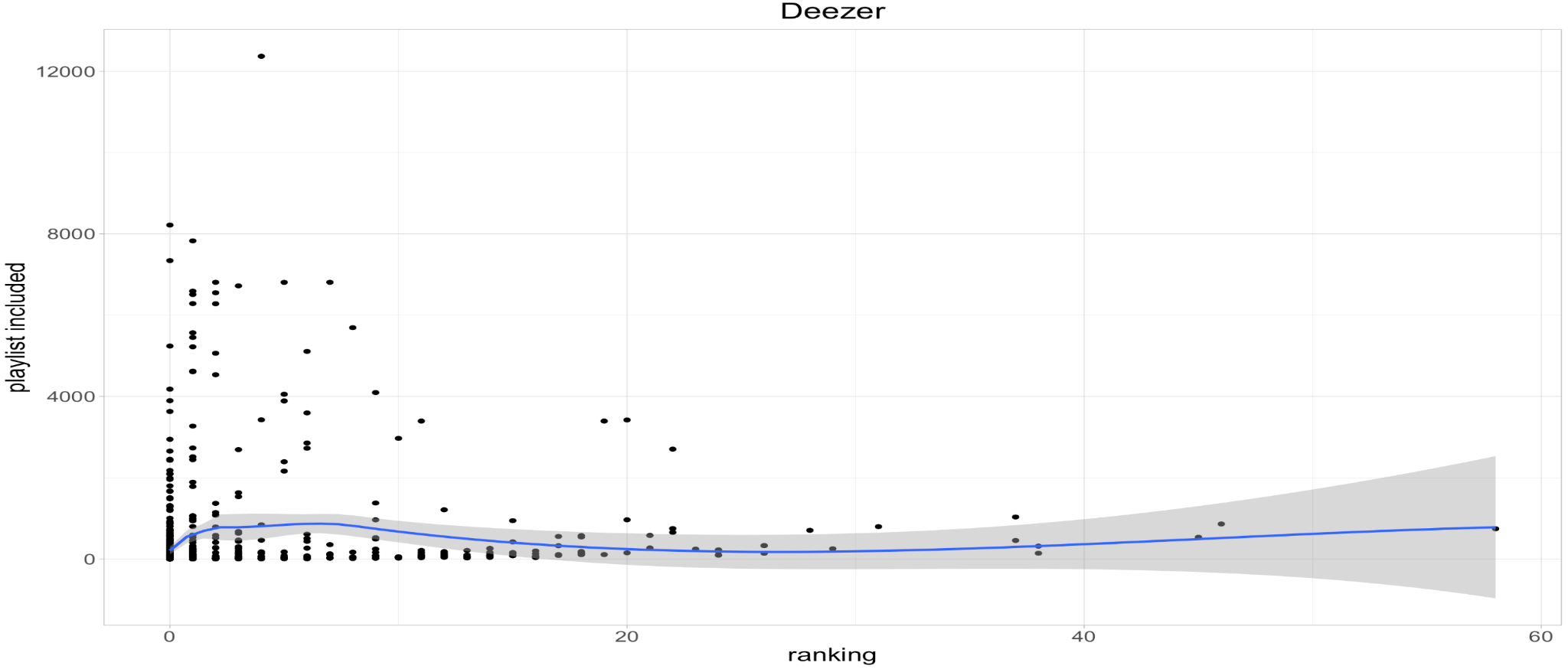
**Graph 3\_1**



**Graph 3\_2**



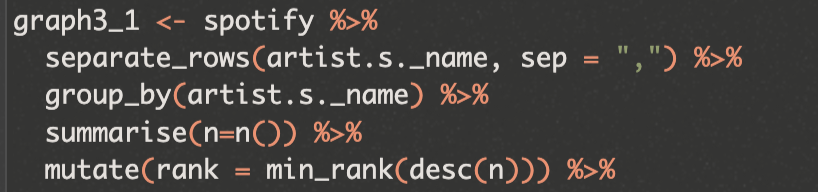
**Graph 3\_3**



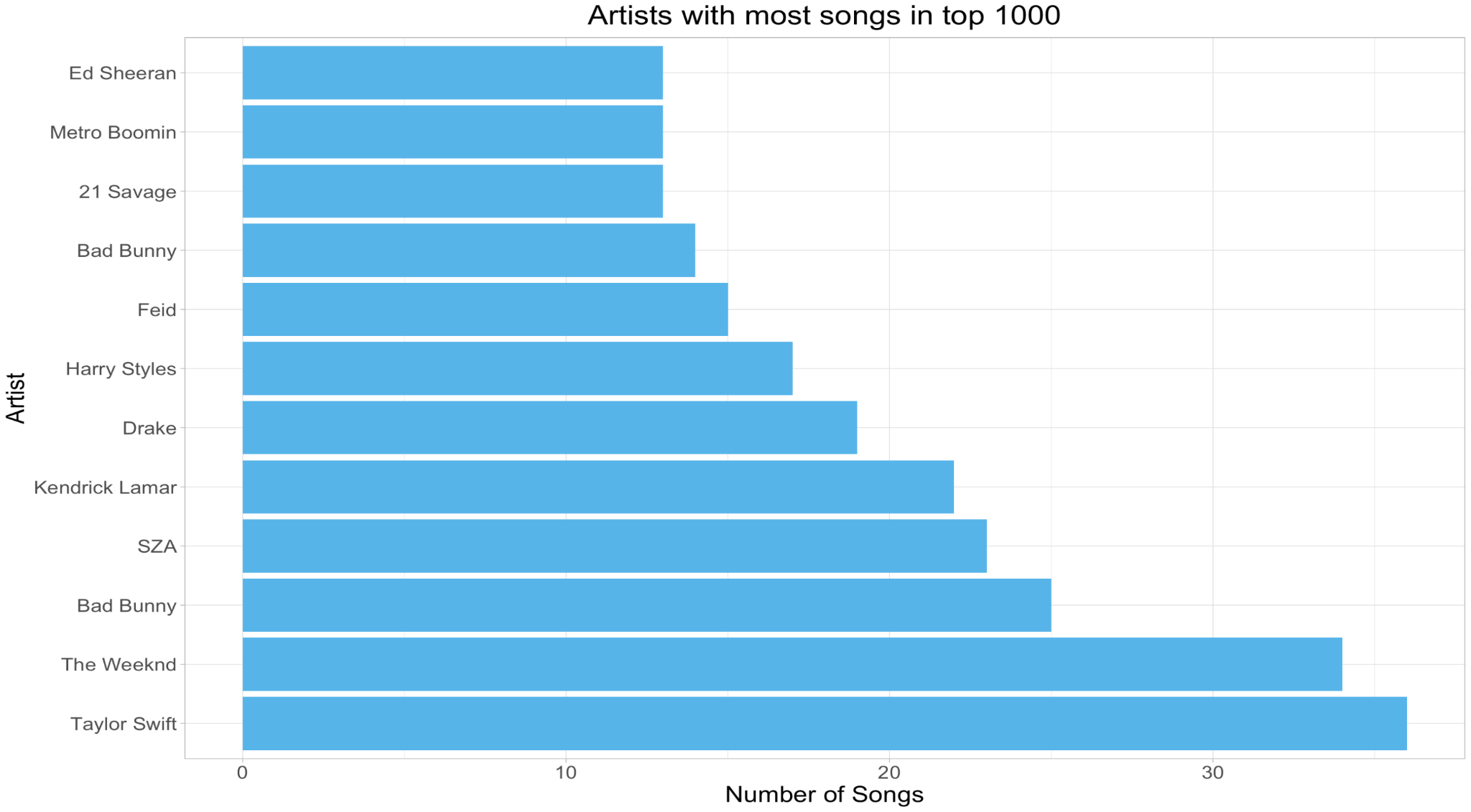
**Graph 3\_4**

It can be seen that the higher the streams the more time of the song is included in the playlist among different platforms. The trend is more significant in Spotify, then deezer, and lastly Apple Music (graph 3\_1). Regarding 3 other charts, we can see that the higher ranked the songs, the more playlists they are included in. Hence, there is a downward trend for the scatter plot.

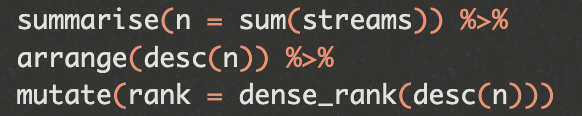
**Visualisation 4: Artist analysis**



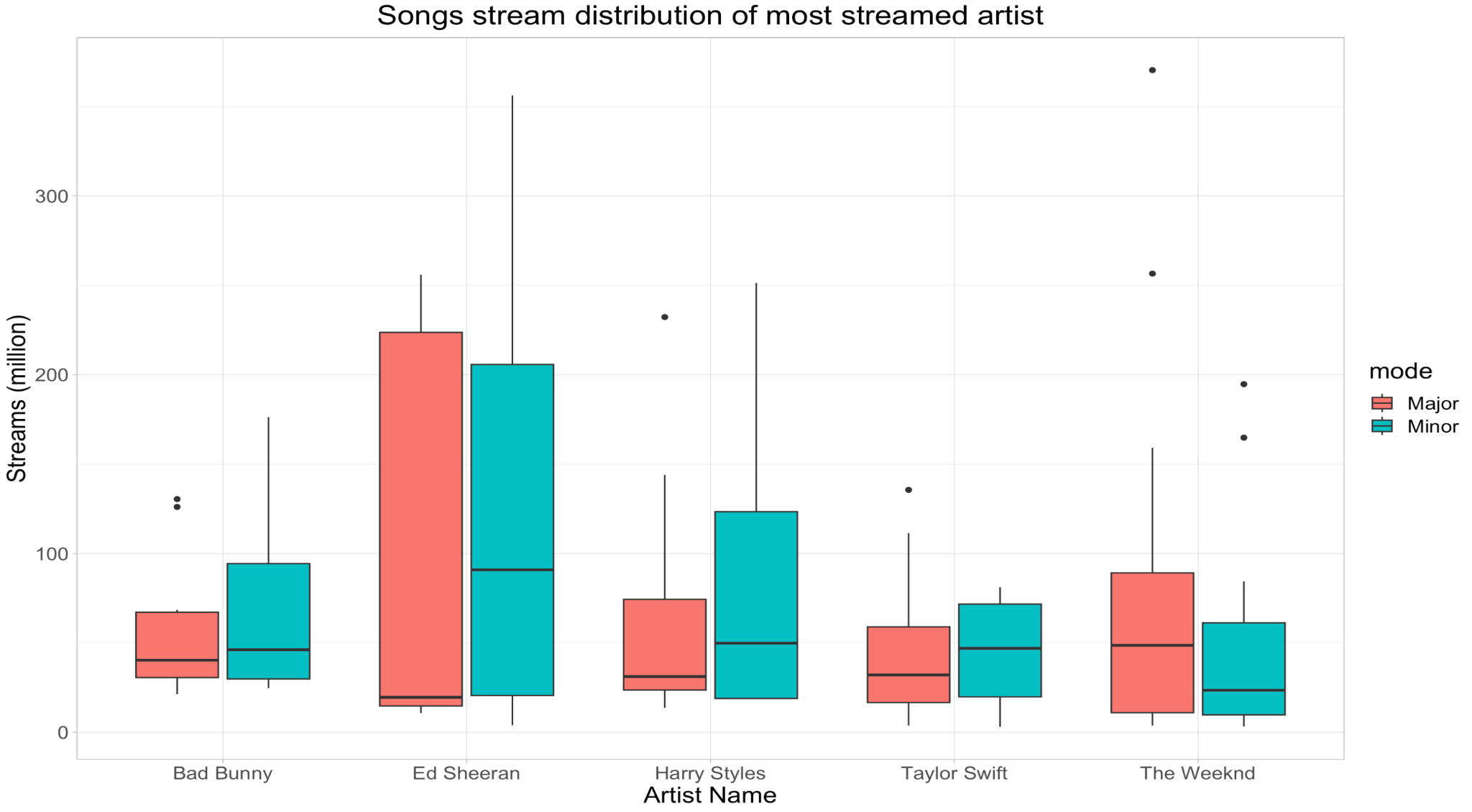
(separate rows of artist by commas and create rank column)



**Graph 4\_1**



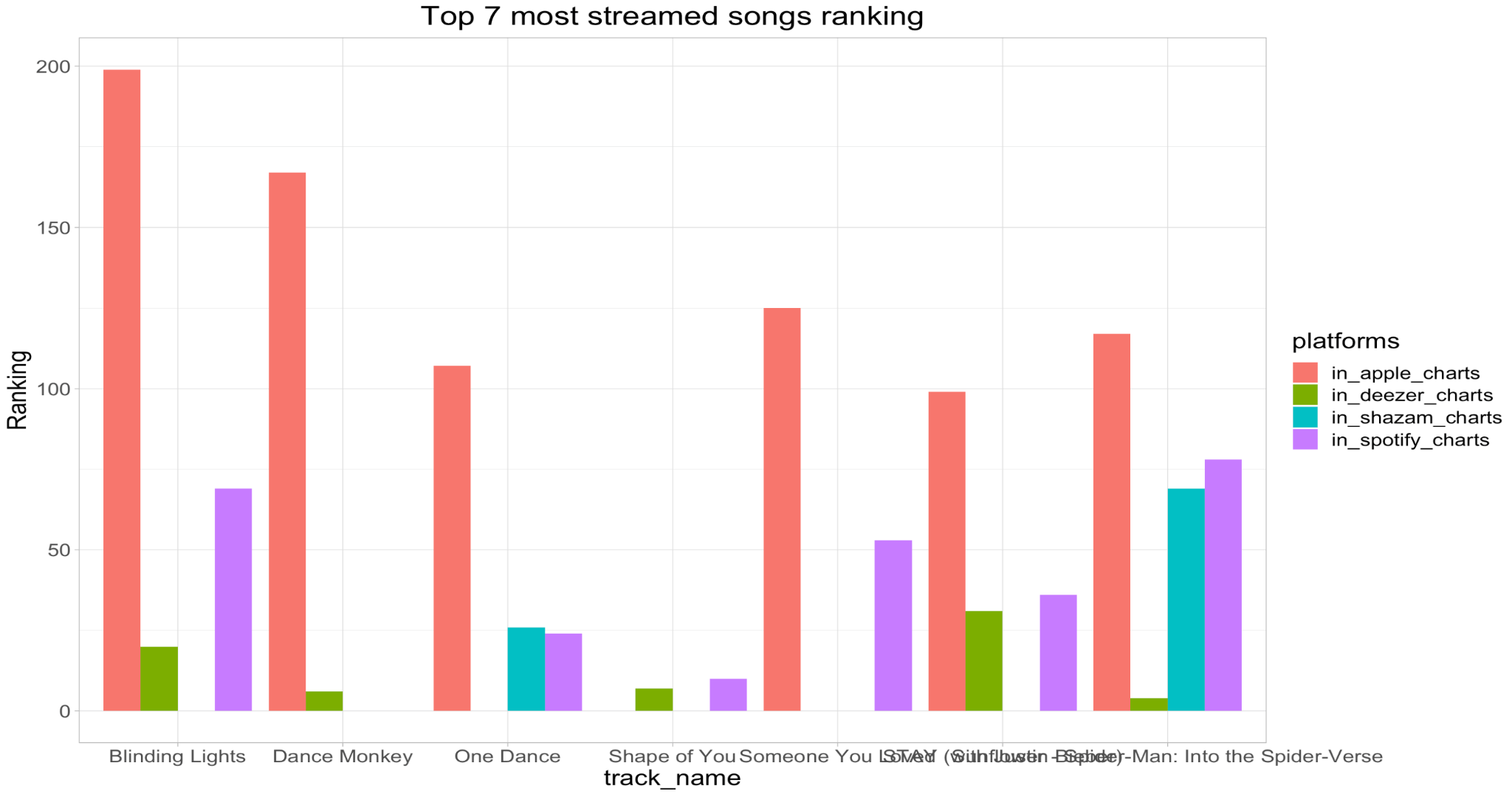
(create rank of most stream column for artists)



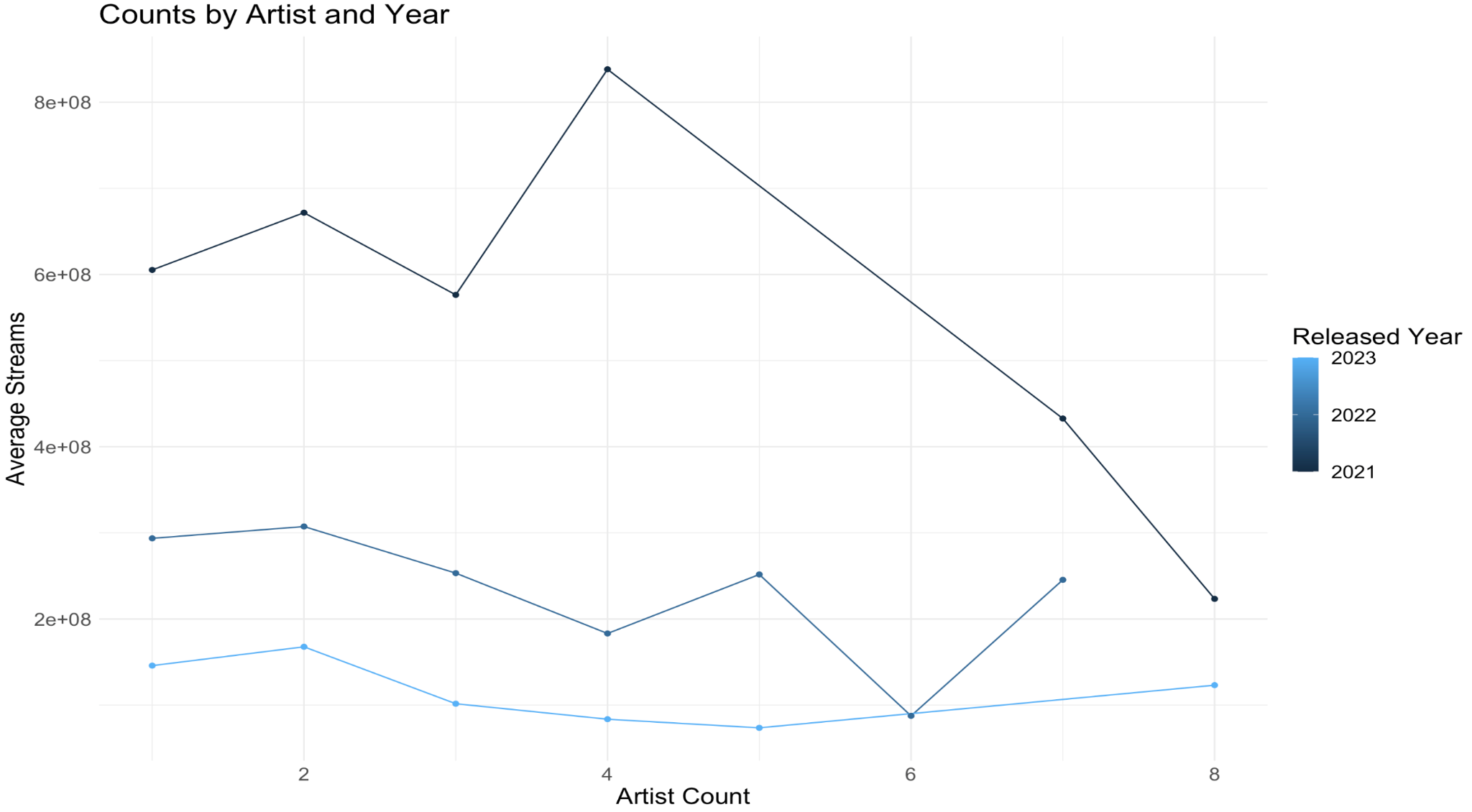
**Graph 4\_2**

From graph 4\_1, we can see that there is a large variety of artist inside the top 1000 data when there are only a few that have over 20 songs (5 artists). While graph 4\_2 show some differences in artists with highest total streames, which can conclude that the number of songs only partly reflect the streams number (some songs have significantly higher streams time)

**Visualisation 5: mixed variables analysis**



**Graph 5\_1**



**Graph 5\_2**

Graph 5\_1 shows differences between ranking of top songs on different platforms. From that information, we can understand the differences in users' tastes in music regarding the app they are using. Number of average streams by the number of artists show huge differences between recent years (graph 5\_2). Most of the songs from 2021 must have significant streams so that it can maintain such a high average number. The number of featured artist does not have any significant effect on streams

**Conclusion**

From 5 visualisations above, it can be inferred that the trend of focus on listening to popular songs will continue. Most of the songs will remain popular for 1 year and then significantly decrease. While the usage in song statistics of artists will be unchanged in the near future. Therefore, the CMO of Sony Music can use strategies to collaborate and create relationships with consistent and popular artists (according to graph) and boost the awareness of companies to music listeners. The danceability and energy of the song should be focused for best on-chart performance.

**REFERENCES LIST**

Baijens, Jeroen; Helms, Remko W.; and Velstra, Tjeerd, "Towards a Framework for Data Analytics Governance Mechanisms" (2020). ECIS 2020 Research Papers. 81. https://aisel.aisnet.org/ecis2020\_rp/81

Cheong, Lai Kuan and Chang, Vanessa, "The Need for Data Governance: A Case Study" (2007). ACIS 2007 Proceedings. 100. http://aisel.aisnet.org/acis2007/100

Cote, C. (2021). *5 Principles of Data Ethics for Business*. [online] Harvard Business School Online. Available at:<https://online.hbs.edu/blog/post/data-ethics>.

Knapton, K. (2022). *Council Post: Four Best Practices For Big Data Governance*. [online] Forbes. Available at:<https://www.forbes.com/sites/forbestechcouncil/2022/06/16/four-best-practices-for-big-data-governance/?sh=3f04198e6003> [Accessed 3 Nov. 2023].

**APPENDIX**

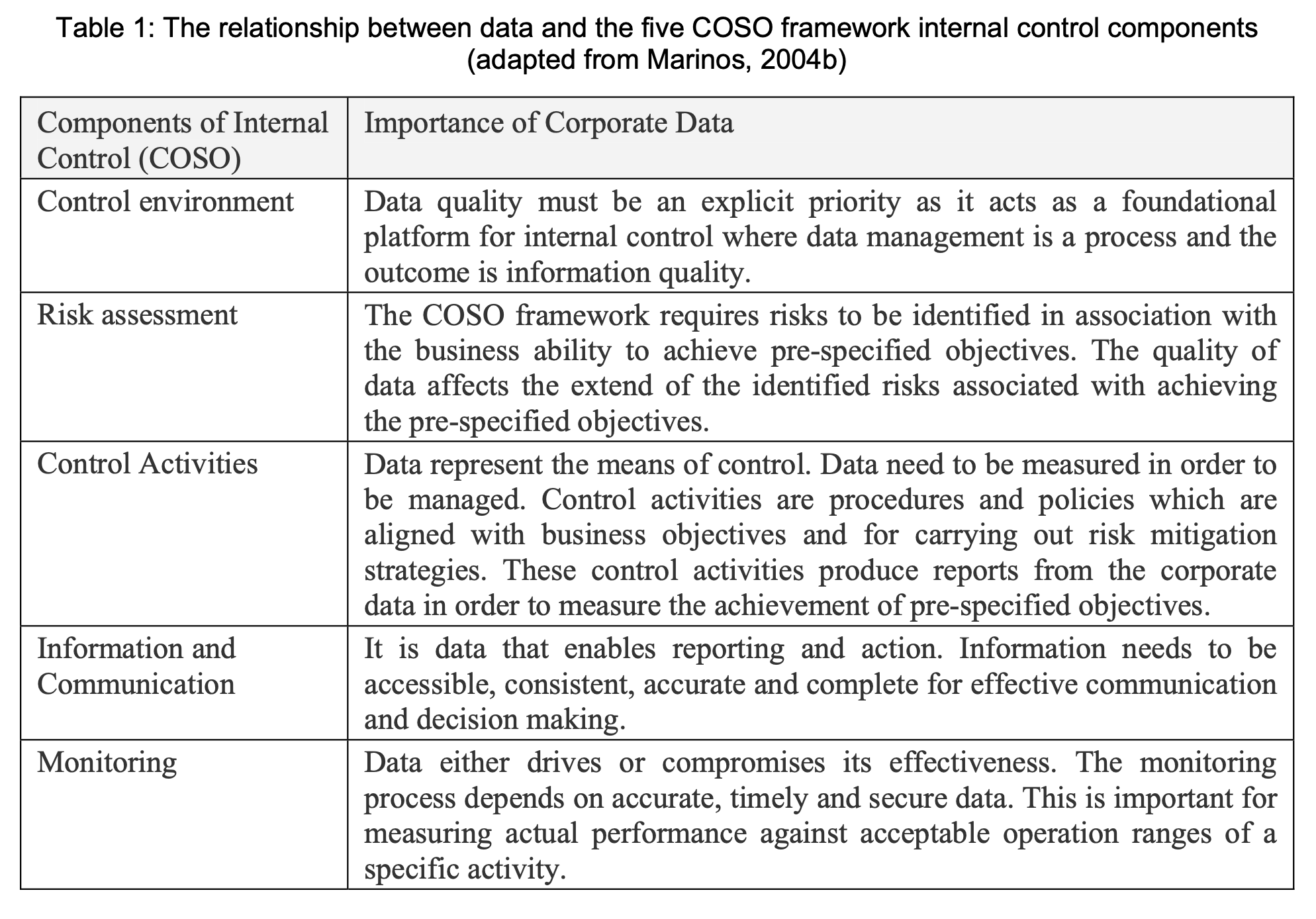


Table 1

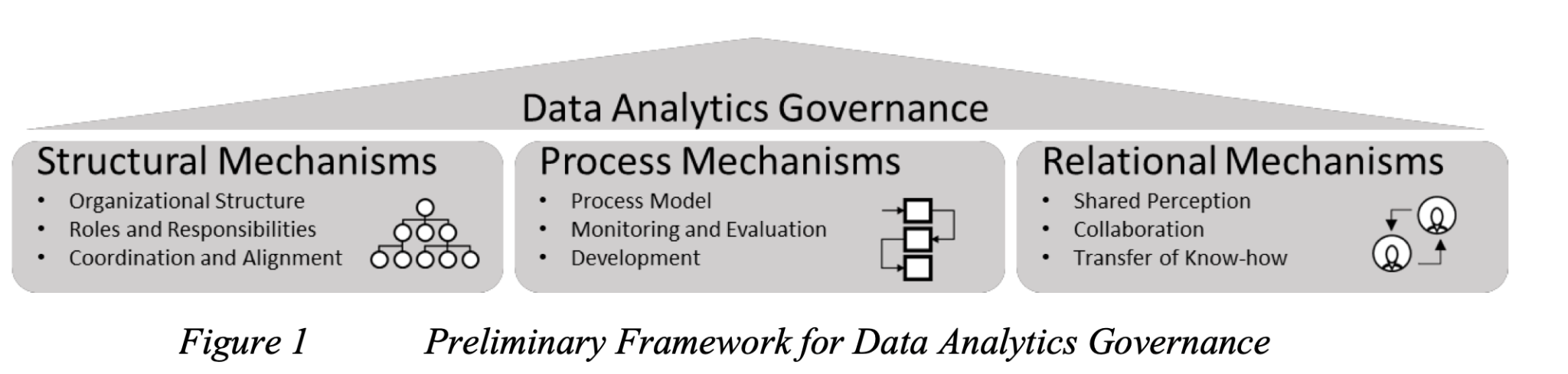


Table 2