Toronto marriage data*

David Qi

October 10, 2024

Practice document for week three. Includes a brief explaination of the toronto marriage dataset and a graph.

1 Introduction

In this paper, we make use of R (R Core Team 2023) and the tidyverse package (Wickham et al. 2019) with supporting package dplyr(Wickham et al. 2023) to analysis Spotify(An online music platform launched in 2008) around Elvis Presley(American celebrity, 1935-1977). The data set is from spotify API (Spotify 2024), gathered using spotifyr (Thompson et al. 2022).The graph is created using ggplot. We have used some information and codes form the book Telling Stories With Data(Alexander 2023) on accessing API.

The remainder of this paper is structured as follows: We talk about how the data is gained under Data section. Then, we provide a table that compares our estimation with actual data in the Results section. We then discusses reason of error in estimation Discussion section.

2 Data

In this section, we will discuss some aspect of the data, and outline how we will use the data.

As shown in Figure 1, the date of album release span across 1956 to 2024. Note that album release was suspended at 1977, when Elvis Presley passed away, and resumed from 2010 onward. This is due to Spotify's new licensing agreements allowed releases of albums of classic artists.

There were also many albums released in 1997, 20 years after Elvis Presley passed away. These albums were released to recognize this event.

*Code and data are available at: https://github.com/UTDQi/reflection5.

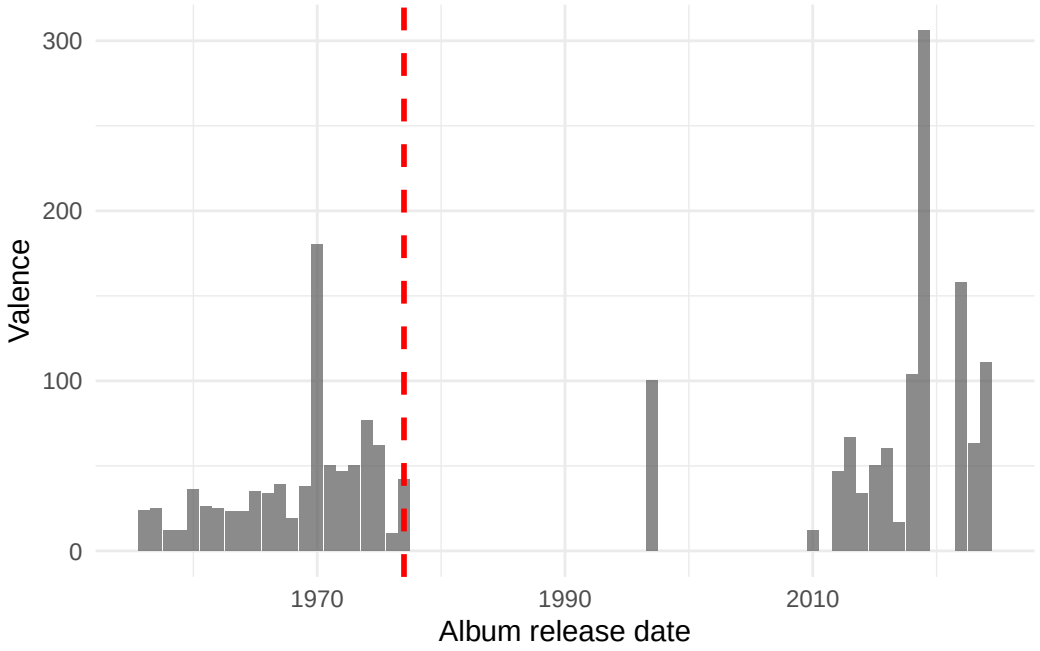

Figure 1

In this paper, we will analysis these two groups of albums to see how people's tastes have changed over these 30 years. We note that albums released reflect the general taste of the audience at the time, as albums makers tends to select and remix songs according to this taste.

To reflect the very abstract concept of "taste", we will use data gathered about the liveness, valence and tempo of the music. It is very important to note these data are form the results generated using a machine learning model that we do not have much information about. They may not fully reflect actual "liveness", "valence" or "tempo" of the music. Despite that, they do serve as a useful generalization of these factors. And the fact that data is being used by Spotify itself indicates that it is useful for data analysis.

3 Results

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Spotify. 2024. "Spotify Web API." https://developer.spotify.com/documentation/web-api/.

Thompson, Charlie, Daniel Antal, Josiah Parry, Donal Phipps, and Tom Wolff. 2022. *Spotifyr: R Wrapper for the 'Spotify' Web API*. https://github.com/charlie86/spotifyr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.