Data Analysis for Ed Sheeran*

Zitong Guo

October 10, 2024

Introduction

Ed Sheeran has become one of the most influential artists in contemporary pop music, known for his diverse range of musical styles and lyrical storytelling. Analyzing trends in his song durations over time can provide insights into the evolution of his artistic approach and how it reflects broader trends in the music industry. This study explores the changes in the duration of Ed Sheeran's songs using data retrieved from the Spotify Web API through the `spotifyr` package Thompson et al. (2022). The analysis focuses on how the length of his songs has varied over the years, particularly between 2015 and 2024, and seeks to identify patterns or shifts that may indicate changes in his creative process or adaptations to industry trends.

The `spotifyr` package allows for easy access to Spotify's extensive music catalog, facilitating data collection on various aspects of artists' discographies Thompson et al. (2022) . R R Core Team (2023), a statistical computing environment, and its accompanying packages, such as `tidyverse` for data manipulation Wickham et al. (2019) and `usethis` for project Wickham et al. (2024), provide the tools necessary for robust data analysis. Furthermore, the study applies storytelling principles to data analysis, as outlined by Alexander Alexander (2023), emphasizing the importance of clear communication in presenting findings.

By examining how Ed Sheeran's song durations have evolved, this study aims to contribute to the understanding of artistic growth and the impact of industry changes on music production. Such analysis not only sheds light on Ed Sheeran's career but also serves as a case study for broader shifts in musical trends over the past decade.

*Code and data are available at: [https://github.com/jennygzt/Spotify.git]

Data and Methods

The data for this analysis was retrieved using the `spotifyr` package in R, which allows access to Spotify's API for detailed information about artists' discographies. The dataset includes information on Ed Sheeran's songs, such as their release dates and durations (in milliseconds). The analysis focuses on how the duration of songs has varied across Ed Sheeran's albums, released between 2015 and 2024.

Data cleaning involved removing duplicates and missing values. The cleaned dataset was then visualized using `ggplot2` to create a scatter plot, illustrating song durations over time. The vertical axis represents the duration of songs in milliseconds, while the horizontal axis displays the release date of each song.

Results

The scatter plot in Appendix A shows that Ed Sheeran's song durations have generally become more varied over time. While his earlier songs around 2015 tend to cluster around a similar duration, more recent releases show a broader range. Notably, some tracks released after 2020 have durations significantly longer than 400,000 milliseconds, suggesting a shift towards longer compositions in his recent work. This could reflect changes in his creative process, experimentation with different genres, or adaptations to trends in the music industry.

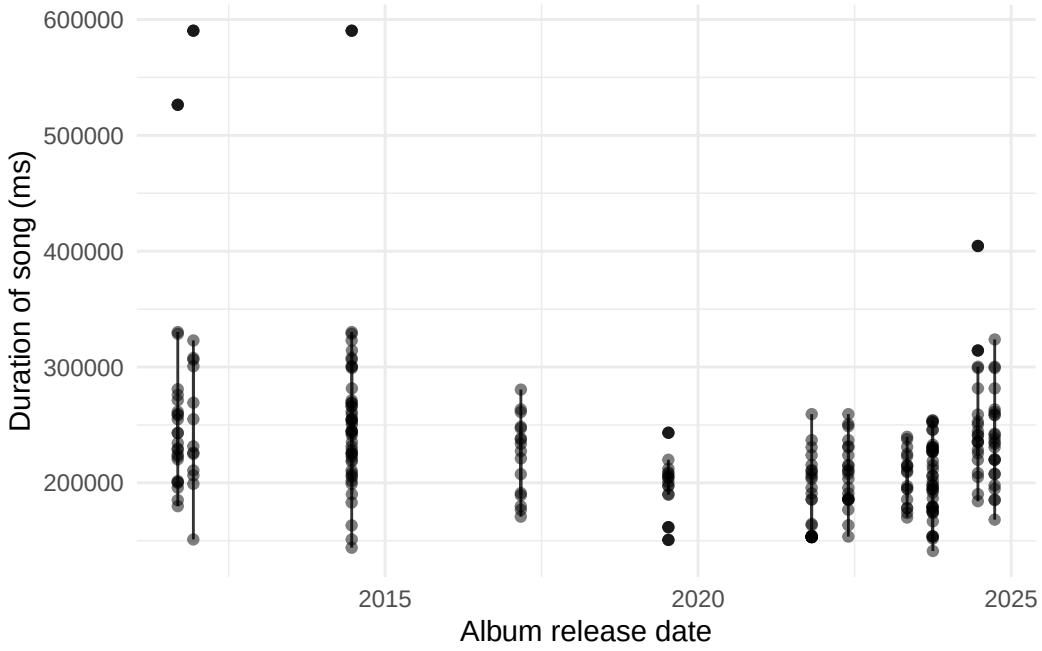
Discussion

The observed trend of increasing variation in song lengths aligns with Ed Sheeran's evolving style as he experiments with different musical forms. Longer tracks may reflect his deeper engagement with storytelling or a move towards more immersive compositions. Additionally, the music industry's shift towards streaming services has altered how artists release and structure music, which could contribute to the broader range of song durations.

These findings provide a foundation for further research, such as comparing Ed Sheeran's trends with those of his contemporaries or analyzing how song characteristics (e.g., tempo, key) have evolved alongside duration.

Appendix

Appendix A: Scatter Plot of Song Duration Over Time



References

Alexander, Rohan. 2023. *Telling Stories with Data: With Applications in r*. Chapman; Hall/CRC.

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

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, Jennifer Bryan, Malcolm Barrett, and Andy Teucher. 2024. *Usethis: Automate Package and Project Setup*. https://CRAN.R-project.org/package=usethis.