

Spotify Data

Yanzun Jiang, Siyuan Lu, Yi Tang, Jin Zhang

November 21, 2024

This study explores the musical evolution of David Tao using data from Spotify, analyzing features like tempo, energy, and valence (happiness) across his songs from 1997 to the present. The analysis reveals that tempo and valence do not show a straightforward positive relationship, while energy levels in his songs decreased over time before rising again, potentially reflecting personal or artistic shifts. These findings show how an artist's musical characteristics can change over decades, indicating the relationship between creative expression and personal or cultural contexts. By examining the dynamics of music over time, this study highlights how data can deepen our understanding of artistry and its evolution.

Table of contents

1	Introduction	2
2	Data	2
2.1	Overview	2
2.2	Preview of the Dataset	3
3	Discussion	3
A	Appendix	4
A.1	Data Cleaning	4
A.2	Summary Statistics of the Data	4
	References	6

1 Introduction

Music is a powerful medium that reflects the personal journeys of artists and the cultural landscapes of their times. With the advent of music streaming platforms, unprecedented amounts of data on songs and artists are now accessible, enabling detailed analysis of musical features such as tempo, energy, and valence. However, while much research has focused on broad trends across genres or artists, there is a gap in understanding how the creative expression of individual artists evolves over time. This study addresses this gap by examining the works of David Tao through a dataset sourced from Spotify.

Our analysis highlights two key findings. First, there is no simple positive correlation between **tempo** and **valence**, indicating that faster songs are not always happier in Tao’s discography, likely due to variations in instrumentation and lyrical themes. Second, the energy levels in his songs exhibit a notable decline before rising again, possibly reflecting shifts in his personal experiences or creative priorities. These findings contribute to a deeper understanding of the complex interplay between an artist’s personal evolution and their musical output.

To conduct the analysis of songs on Spotify, a dataset downloaded from Spotify (2024) was utilized, as described in Section 2. Based on the initial findings, it was observed that **tempo** and **valence** are not positively correlated in a simple (linear) way, while energy falls then rises with time (Section 3). Also, supplementary insights are provided in Section A.

2 Data

2.1 Overview

The dataset used in this analysis is about the songs of David Tao, sourced from Spotify (Spotify (2024)). It was accessed by an API. It records various aspects of the songs by David from 1997 to present.

The variables analyzed in this study are listed below:

album_release_date: the date that the album was released.

energy: the energy of the song.

valence: the degree of happiness of the song.

tempo: the pace of the song.

The dataset was accessed using the **spotifyr** package (Thompson et al. (2022)). For the analysis, the R programming language was employed (R Core Team (2023)), utilizing the **tidyverse** (Wickham et al. (2019)) package for data cleaning, transformation, visualization, the **knitr** package to visualize the data, and the **here** package (Müller (2020)) accessing the

data in this analysis. Afterward, the cleaned dataset was processed and tested using additional functions from the `tidyverse` package (Wickham et al. (2019)).

2.2 Preview of the Dataset

This (Table 1) is the preview of the dataset.

Table 1: Preview of Data

album_release_date	energy	valence	tempo
2017-06-08	0.693	0.293	115.875
2017-06-08	0.770	0.480	110.940
2017-06-08	0.726	0.637	94.983
2017-06-08	0.717	0.650	187.903
2017-06-08	0.661	0.411	96.011
2017-06-08	0.807	0.533	121.067

3 Discussion

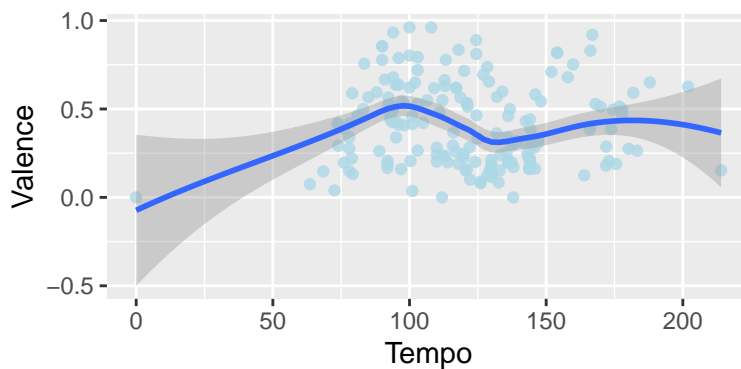


Figure 1: Relationship between Pace and Key in David's Songs

From Figure 1, we can see that `tempo` and `valence` are not strictly positively correlated as we expected. This might due to the lyric and usage of different musical instruments.

From Figure 2, we can find that the energy for David's Songs falls then rises with time, which might be associated with his personal experience.

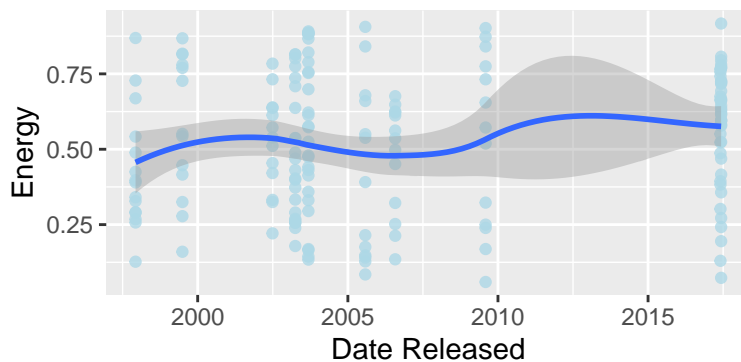


Figure 2: Changes of Energy with Time in David's Songs

A Appendix

A.1 Data Cleaning

The data cleaning process involved tidying the dates, filtering out useless columns from the raw dataset, and filtering out observations with NAs.

A.2 Summary Statistics of the Data

Table 2: Summarise Statistics of Energy in David's Songs

Mean	Median	Variance	Min	Max	IQR
0.5	0.5	0.1	0.1	0.9	0.4

Table 3: Summarise Statistics of Key in David's Songs

Mean	Median	Variance	Min	Max	IQR
0.4	0.4	0.1	0	1	0.4

Table 4: Summarise Statistics of Pace in David's Songs

Mean	Median	Variance	Min	Max	IQR
121.3	120.6	992	0	214	45.7

The summary statistics provides key statistical insights into the variables of David's songs. For each variable, the summarize tables (Table 2, Table 3, and Table 4) reports the mean, median, variance, minimum value, maximum value, and interquartile range of those variables, offering a concise overview of range, central tendencies, and variability in the data.

References

- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- 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 Music Data.” Data retrieved from <https://open.spotify.com>.
- 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 Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.