

An Analysis of Taylor Swift ‘Through the Eras’

Donya Behroozi¹, Ethan Gutterman², Isabelle Phraner³, William Smithers³

¹Department of Statistics, California Polytechnic State University - San Luis Obispo

²Department of History, California Polytechnic State University - San Luis Obispo

³Department of Business, California Polytechnic State University - San Luis Obispo

May 31st, 2024
STAT 365, Dr. Emma Robinson

Abstract

This study delves into the relationship between the danceability of Taylor Swift's music and its critical reception, scrutinizing both user and critic ratings. Through comprehensive analysis, we reveal a noteworthy negative correlation between user scores and danceability, indicating that as user ratings ascend, danceability score tends to diminish. However, the correlation between critic ratings and danceability lacks significance, suggesting a lack of linear relationship between critic approval and the danceability of Swift's music. Discrepancies in rating criteria between user and critic populations are evident from the differing levels of significance in the models. Our findings, though insightful, are circumscribed by the scope and context of our data, which excludes singles and versions not owned by Swift. Additionally, the subjective nature of music ratings limits the generalizability of our results. Future research avenues could broaden the analysis by incorporating a wider array of musical attributes or exploring other artists and genres.

1 Introduction

Taylor Swift is an artist who has known that their path was world domination for years. She is one of the most streamed artists in the world, had the highest-grossing tour of all time, and has 14 Grammy's. In her own words, she is in her "imperial era". Born in the Pennsylvania suburban community of Wyomissing to a stockbroker father and an accountant mother had a very traditional upbringing. Her family wealth allowed her to self-finance her music journey and receive excellent tutelage from instructors across the state. "Swift had already decided to become a songwriter. When she was ten, her mother began driving her around on weekends to sing at karaoke competitions"¹.

This work explores how the danceability of Taylor Swift's albums, as quantified by their danceability scores, is associated with their critical reception as reflected in Metacritic ratings and audience scores. For our first analysis, we investigate whether there is a significant association between audience ratings and danceability scores of Swift's music. We follow a similar approach for analysis two, instead looking at whether there is a substantial association between critic ratings and danceability scores of Swift's music.

The major sources of data for this work are found on the October 17, 2023, Tidy Tuesday regarding Taylor Swift². The first dataset explores the site Metacritic, which acts as a compiler of reviews from across the internet and print media. They do not issue their reviews and just create a composite Metacritic score for films, television, and music³. On their site, they also host blogs in which Metacritic users can review media as well. This Metacritic dataset is simply the Metacritic and user scores for Swift's albums and songs. The other dataset for this study encompasses Taylor Swift's entire discography to date, from her debut album release in 2006 through to her most recent work as of 2022. W. Jake Thompson conducted the data collection process on October 17, 2023. sourcing information, including lyrical and audio features, for each Taylor Swift song from a Spotify Web API by obtaining an access token to request artist data for musical attribute scores of the songs. Spotify's API technology for analyzing musical attributes served as the primary specialized equipment used in the data collection. This tool allows for the extraction of information about song characteristics, including tempo, key, mode, and more, based on digital signal processing algorithms. Data collection was entirely digital and relied on advanced computational methods, eliminating the need for traditional physical equipment. Additional information, such as album release dates, track names, and user/Metacritic scores were compiled from authoritative music industry databases and chart records (e.g. Genius.com, Metacritic.com).

¹<https://time.com/6342806/person-of-the-year-2023-taylor-swift/>

² <https://github.com/rfordatascience/tidytuesday/tree/master/data/2023/2023-10-17>

³ <https://www.Metacritic.com/about-us/>

In one of Swift's first large interviews, a 2011 New Yorker Profile⁴, she is described as a twenty-one-year-old country-pop star. "She is tall and gangly, with porcelain skin, long butterscotch hair that seems crimped, as if from a time before curling irons, and smallish eyes that often look as if they were squinting. She loves to wear makeup, but it tends to resemble stage makeup: red lipstick, thick mascara."

Even in this early stage in her career, Swift knew the power of the media and intertwined her life and current actions into the music. She called it "human nature" for fans to want to be invested in her personal life. Flash forward to 2023, with the Era's Tour generating record revenue and being large enough for the Federal Reserve to notice its economic impact on the cities it visits. In the New York Times 2023 review of the concert⁵, the author calls it spiritual.

The mood was solemn — spiritual, even. I have prayed at dawn at the Temple Mount. I have stood among quivering supplicants at the graves of biblical forefathers. I have walked in trembling silence as I entered farther and farther into the inner sanctums of the Vatican. This was like that, except for girls.

The Eras Tour, and Swift's music, allow for her fans to celebrate themselves and center womanhood. The author notes how Swift combines quite sensual burlesque costumes and dances with jokes on stage and square dancing, reminding how they would act lip-syncing music in the privacy of their bedroom. The Era's Tour follows the eras of one's own life, with ups and downs, and celebrates them because they lead where you are in the current moment. She is called the leader of the postmedia celebrities. She hasn't given a major interview in years, because she does not need to. Technology has risen to the level that the press is no longer a necessary component to find success.

Despite the rarity of interviews, Time may have awarded Swift the 2023 Person of the Year just to secure an interview with the⁶, once country, but now just a star. The interview opens with quite a daunting line;

As a pop star, she sits in rarefied company, alongside Elvis Presley, Michael Jackson, and Madonna; as a songwriter, she has been compared to Bob Dylan, Paul McCartney, and Joni Mitchell. As a businesswoman, she has built an empire worth, by some estimates, over \$1 billion. And as a celebrity—who by dint of being a woman is scrutinized for everything from whom she dates to what she wears—she has long commanded constant attention and knows how to use it.

⁴ <https://www.newyorker.com/magazine/2011/10/10/taylor-swift-profile-you-belong-with-me>

⁵ <https://archive.is/QhDUR#selection-499.0-501.20>

⁶ <https://time.com/6342806/person-of-the-year-2023-taylor-swift/>

Swift has reached a status in the music industry that very few have attained. The prowess and ability to get streams had led to her forcing Spotify to change how they pay artists and receive a new deal with Universal Music Group.

The literature on cultural criticism, and even music criticism, is quite diverse in the topics it analyzes. For the use of this work, two articles analyzing popular music criticism will be discussed. The first article written by Bethany Klein in 2009 explores the difference between popular music critics and “high culture” critics, things like art exhibitions, architecture, or recitals of classical music⁷. The largest difference they find is that critiques of popular music have a different purpose. They are not being asked to amplify or platform smaller acts that the critique enjoyed. The thing that is being reviewed already is massively popular (pg 2)⁸.

Klein’s argument that this work makes is that there is a difference between a critic’s professional criticism and opinions. It is the job of a listener to decide what they like and do not like, it is the job of the critic to determine what is good and bad (6). They must interpret why artists become successful and interpret an artist through a standard set by the publication or critical precedent. Unlike the audience which can vary and not have any constituency.

Phillip Vannini’s 2011 article⁹ provides a case study for this dichotomy in which the rise and reception of Avril Lavigne is analyzed. An interesting point Vannini makes, which is entirely relevant to Swift’s career, is that of persona. It doesn’t matter what Avril Lavigne or Taylor Swift actually act like, or how they feel about punk music or country music receptively, only what they show in the real world. The persona is all that matters in pop music (50). Vannini disagrees with Klein’s arguments of critical standards. Vannini uses the rise of the internet and things like VH1 and MTV which means the artist’s lifestyle and aesthetics are now just as important as the music (51). With this being written in 2011, this argument has become even more true. With artists blowing up off on snippets posted across social media, where the audience usually knows more about the artists than just their music. Vannini disagrees further with Klein’s idea of the “expert critic”, especially about pop music. The reviews of pop music that get read the most are semi-amateur reviews posted on the internet, not highbrow reviews posted in coastal elite publications (52).

It has been said that the emergence of streaming platforms has permanently altered the practice of music consumption. As of 2018, streaming made up 47% percent of the music market. However, this share of the music market is made up of several individual streaming platforms, whose algorithms for popularity are not necessarily the same. Each platform has different key

⁷ https://doi.org/10.1207/s15405710pc0301_1

⁸ https://doi.org/10.1207/s15405710pc0301_1

⁹ <https://onlinelibrary.wiley.com/doi/full/10.1525/si.2004.27.1.47>

demographics, so popularity can vary between platforms (e.g., Spotify, Apple Music, YouTube). As such, what determines popularity can be affected by what an individual seeks to get from their listening experience. For instance, if they are seeking stimulation, or whether they're trying to bring themselves up or to come down from a high. Based on the existing literature on the topic, there are some clear trends in preferences for (and ergo, the popularity of) music. For instance, algorithms are influenced by broader cultural norms and emotional expressions, which can include gendered preferences and biases in music recommendation systems. This is consistent with Liew, et al's findings, "cross-cultural research on composite Spotify features [indicates that] users residing in Asia tended to play songs that had lower arousal than users residing in Oceania, Europe, and the Americas".

Liew et al also found that there was a relationship between danceability, a composite feature comprised of elements such as tempo, beat, and rhythm, and emotional arousal, which is the intensity of emotions experienced by the listener. High-arousal emotions are more intense, and potentially energizing, regardless of positive (excitement) or negative (angry) connotation. Liew et al found that the danceability of music is associated with high emotional arousal and that societies that have higher levels of high-arousal negative emotions such as anger or anxiety, also prefer more danceable music. As such, there are probably geographic differences in popularity based on the danceability of a song.

In regards to popularity, in her work "Organizing music, Organizing Gender: Algorithmic Culture and Spotify Recommendations," Werner found that within Spotify's recommender systems, more "male" genres tend to be valued higher for popularity than more "female" genres, indicating some amount of gender bias on a programmatic level. Spotify is also temporally sensitive, paying attention to its user's patterns across days, weeks, and months, potentially simplifying the listening experience. But what does this mean in the context of an international pop goliath such as Swift? As an artist whose work spans genres, and who has both high and low-energy songs and albums, with varying levels of danceability, and emotional arousal, Swift's work can be "suggestable" to many different demographics, from different geographic regions.

According to the Spotify for Developers documentation, danceability "describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A [danceability value] of 0.0 is least danceable and 1.0 is most danceable". Tempo is the speed of the track, measured in beats per minute (BPM), with faster tempos contributing to higher danceability. Rhythm Stability is the consistency and regularity of the rhythm, which can make a track easier to dance to. Beat Strength is the clarity and prominence of the beats. Strong, clear beats enhance danceability. To elaborate on this, we can use Dhruv Khurjekar's analysis of Spotify song characteristics, which finds that certain song keys are more danceable than others, namely F#. Further, he finds that decade and key appear to be associated, that explicit songs are likely more danceable than clean songs, and one can predict

average danceability from a given year. In layman’s terms, current tastes dictate the danceability of a song. No one wants to throw it back to an 1820 waltz. There are non-acoustic levels to danceability, such as preference for the artist, or novelty, but a “danceable” song frequently belongs to a genre with strong rhythmic elements, such as hip-hop, EDM, or Pop. These findings are confirmed by Gabriel Mäkelä’s research in their Thesis, “Hit or not? Predicting and explaining hit potential of songs in the Finnish music market using Spotify data”. Mäkelä finds that danceability increases decade by decade and that more explicit songs appear to be more danceable songs than non-danceable ones, confirming Khurjekar’s findings.

The existing literature is aligned on the use of Spotify’s Echo Nest, developer documentation, and native API for analyzing Spotify data. Because much of the literature uses the same sources, there is consensus regarding how danceability is defined, its subcomponents, and how it may be predicted.

A data and methods section follows which breaks down our two analyses, one looking at the association between audience rating of Taylor Swift albums and danceability, and the other looking at Metacritic ratings and danceability.

2 Data and Methods

2.1 Data Breakdown and Hypotheses

Hypotheses:

Analysis 1: Association between audience rating and danceability

- H_0 : Audience rating is not linearly associated with danceability score ($\beta_{\text{audience score}} = 0$)
- H_A : Audience rating is linearly associated with danceability score ($\beta_{\text{audience score}} \neq 0$)

Analysis 2: Association between critic rating and danceability

- H_0 : Critic rating is not linearly associated with danceability score ($\beta_{\text{critic score}} = 0$)
- H_A : Critic rating is linearly associated with danceability score ($\beta_{\text{critic score}} \neq 0$)

Each observation in this study represents a Taylor Swift song. The dataset includes all songs from Swift’s albums published from 2006 to 2022, excluding singles released separately from an album (e.g., "Only the Young," "Christmas Tree Farm") and non-Taylor-owned albums that have a Taylor-owned alternative (e.g. “Fearless (Taylor’s Version)” is included in this dataset instead of “Fearless”, which is owned by Swift’s previous record label, Big Machine Records).

The variables analyzed in this study include:

- **Album Name:** The name of Taylor Swift's albums for each observational unit, which includes albums from Taylor Swift (2006) to *Midnights* (2022).
- **Danceability:** A metric from Spotify's Web API describing how suitable a track is for dancing, with a scale from 0.0 (least danceable) to 1.0 (most danceable).
- **User Score:** The average rating (out of 10) given by users on Metacritic.com for Swift's albums.
- **Metacritic Score:** The average rating (out of 100) of critic scores on Metacritic.com, based on a weighted average of scores from respected critics, including those from Rolling Stone, The Guardian, and Variety.

The study aims to investigate whether there is a significant linear association between danceability score and both user and critic scores. Additionally, it explores how these associations differ between user scores and critic scores and the potential reasons for these differences, such as musical shifts in Swift's career.

2.2 Statistical Analyses

Before conducting our analyses, we combined album-level data of user and Metacritic scores with song-level data from Spotify metrics, aggregated by album name with R's tidyverse package. This merged dataset served as the basis for our subsequent analysis.

We examined the relationship between danceability and both user and Metacritic scores using simple linear regression with base R functions for both analyses. For analysis one, we created a regression model with user score as the explanatory variable, and danceability as the response. Similarly, for model two, we created a regression model with Metacritic score as the explanatory variable, and danceability as the response. We then conducted a two-sided t-test in R, using the p-value to assess significance. This method allowed us to assess whether there is a significant linear association between these variables in each analysis.

3 Results

In this section, we present the findings from our analysis of Taylor Swift's albums, focusing on the relationship between danceability and both user and Metacritic scores. Using album-level data, we aggregated song-level metrics from Spotify and combined them with user and critic scores to examine these relationships through linear regression and t-tests for significance.

Figure 1 presents the summary statistics for Taylor Swift's albums, detailing the mean danceability, standard deviation of danceability, mean user score, and mean critic score for each album. This table serves as a foundation for our analysis, providing an overview of the variability and central tendencies of these key metrics across her discography. Figure 2 and Figure 3 illustrate the associations between danceability scores and user and Metacritic scores, respectively.

Figure 1: Table of Summary Statistics

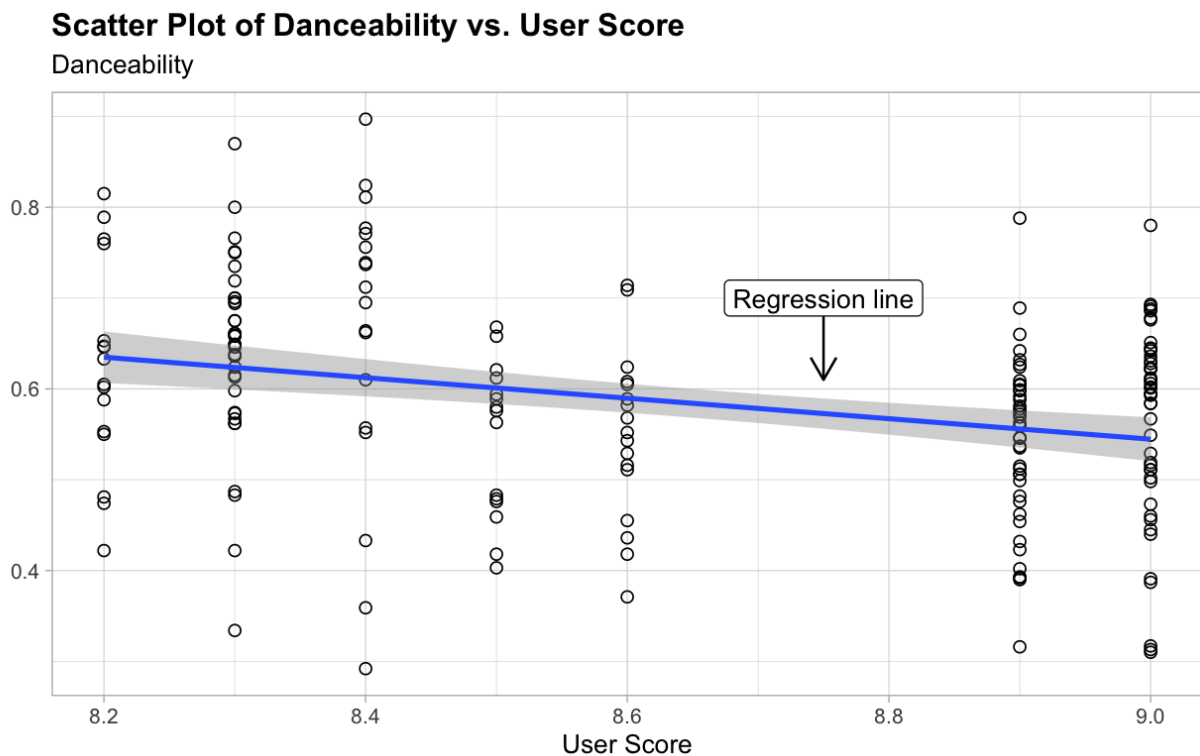
Album Name	Mean Danceability	Standard dev. Danceability	Mean User Score	Mean Critic Score
Lover	0.66	0.16	8.4	79
reputation	0.66	0.08	8.3	71
Midnights	0.63	0.12	8.3	85
1989	0.62	0.12	8.2	76
Red (Taylor's Version)	0.58	0.08	9.0	91
Fearless (Taylor's Version)	0.55	0.09	8.9	82
Speak Now	0.55	0.09	8.6	77
Taylor Swift	0.55	0.09	8.5	67
folklore	0.54	0.14	9.0	88
evermore	0.53	0.09	8.9	85

Summary statistics of Swift's albums, including average danceability score, a measure of variation (standard deviation) for danceability, average user score, and average critic score for each album.

This table provides an overview of the mean danceability, standard deviation of danceability, mean user score, and mean critic score for each of Taylor Swift's albums. The data highlights the variability and central tendencies of these key metrics, offering insight into both listener and critic reception, as well as the danceability of each album's tracks. Average danceability scores range on a moderate to high level, between 0.53 (evermore) on the low end to 0.66 (Lover and reputation) on the high end. Variability, as measured by standard deviation, in these scores tend to vary quite a bit across each album. Interestingly, Lover has the most variation in danceability scores, but still stands out with the highest average danceability score out of all the albums. User scores indicate strong positive reception from Swift's listeners, with all ratings above an 8.0 out of 10. On the other hand, critic scores tend to have more variation in their ratings. Notably, Swift's re-released "Taylor's Version" albums have higher ratings (scores of 91 and 82 out of 100 for Red (Taylor's Version) and Fearless (Taylor's Version) respectively) compared to most of the other albums released under Big Machine Records.

3.1 Analysis One: Danceability vs. User Score Results

Figure 2: Plot of danceability score vs. user ratings with simple linear regression line



Scatterplot of danceability (response variable) vs. user score (explanatory variable) with simple linear regression line, showing significant negative association between the two variables. Each dot represents a Taylor Swift song, with each “band” of data points representing a Taylor Swift album.

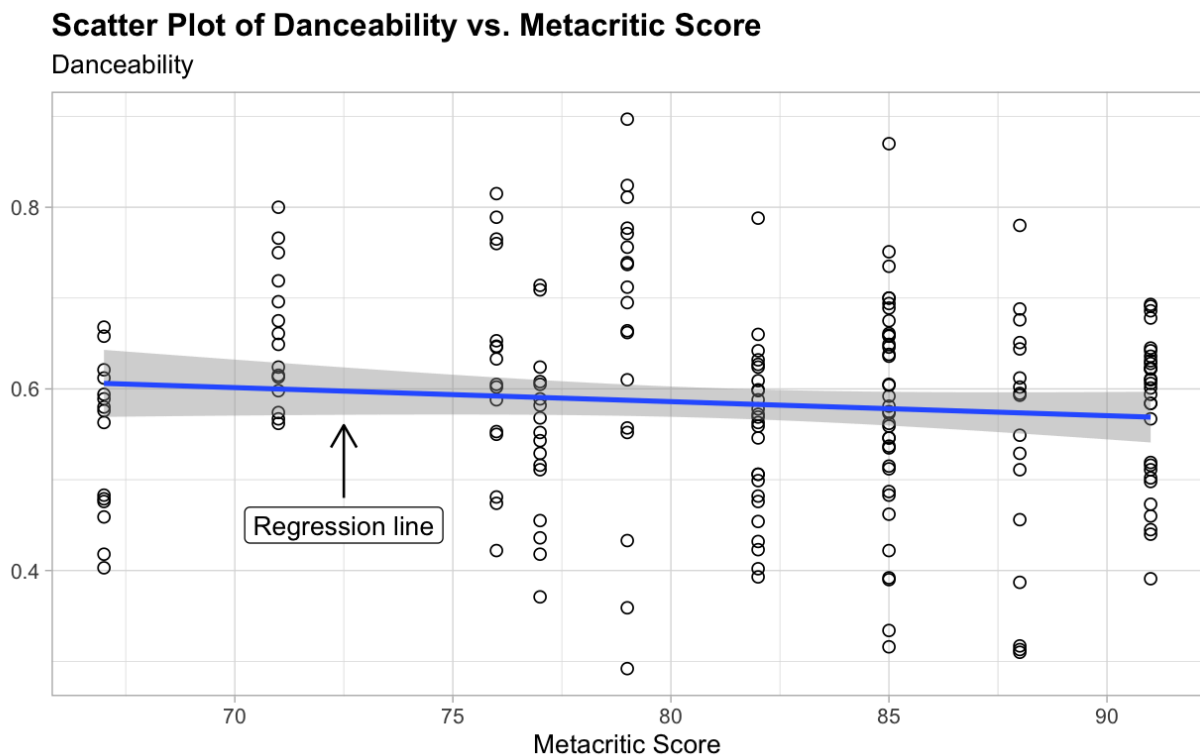
Figure 2 shows that as the user score values increase, danceability scores tend to decrease. There seems to be a somewhat weak negative correlation between danceability and user scores of Swift’s albums, given the correlation coefficient of -0.297 for this model.

Before conducting the two-sided t-test for significance, we ran full regression diagnostics for simple linear regression assumptions. All assumptions were met (see appendix, Fig. 4). After running the t-test, we obtained a large t-statistic of -4.271 (df = 189) and a very small p-value (<0.001). Based on these results, we have strong evidence to reject the null hypothesis.

Therefore, we have sufficient evidence to claim that user score is significantly linearly associated with danceability score. In particular, we are 95% confident that the expected decrease in danceability score for each additional increase in user score is between -0.165 and -0.061 points.

3.2 Analysis Two: Danceability vs. User Score Results

Figure 3: Plot of danceability score vs. Metacritic ratings with simple linear regression line



Scatterplot of danceability (response variable) vs. Metacritic score (explanatory variable) with simple linear regression line, showing weak, non-significant negative association between the two variables. Each dot represents a Taylor Swift song, with each “band” of data points representing a Taylor Swift album.

Figure 3 shows a similar negative correlation between danceability and Metacritic score, as we found in Figure 2. This signifies that as the Metacritic score values increase, danceability scores tend to decrease. However, the correlation between the two variables is much weaker in Figure 3 than Figure 2, given by the correlation coefficient for the model in Figure 3 being very low (-0.095) compared to this value for Figure 2.

Again, full regression diagnostics for conditions of this model’s simple linear regression assumptions were met (see appendix, Fig. 5) before conducting analysis two. After running the t-test, we obtained a small t-statistic of -1.317 (df = 189) and large p-value (0.189). Based on these results, we fail to reject the null hypothesis. Therefore, we do not have sufficient evidence to claim that Metacritic score is significantly linearly associated with danceability score. In particular, we are 95% confident the expected decrease in danceability score for each additional increase in Metacritic score is between -0.004 and 0.001 points.

4 Discussion

Our study aimed to investigate the relationship between Taylor Swift's songs' 'danceability' and two types of scores: user scores and Metacritic scores. Our initial hypothesis predicted a linear relationship between danceability and user scores, but our analysis found a weak negative correlation (correlation coefficient = -0.297). The significant t-statistic (-4.271) and the p-value (<0.001) indicate this relationship is statistically significant. Specifically, we found that for each additional point in user score, the danceability score is expected to decrease by approximately 0.165 to 0.061 points, with 95% confidence. This suggests that higher user-rated songs tend to be less danceable, although the practical significance of this finding may be limited. Additionally, we anticipated a stronger linear relationship between danceability and Metacritic scores, along with a smaller p-value from the t-test. However, the correlation between these variables was much weaker (correlation coefficient = -0.095) and not statistically significant (t-statistic = -1.317, p-value = 0.189), indicating no significant linear association. The negligible effect size, with an expected decrease in danceability of -0.004 to 0.001 points per unit increase in Metacritic score, supports this conclusion.

Despite the strength of our data collection and analysis methods, there are several limitations and potential biases to consider. The weak correlations indicate that other factors not included in our models may influence danceability scores. Furthermore, user scores may be influenced by factors unrelated to danceability, such as personal preferences or song popularity, resulting in variability in our findings. While our findings shed light on the perceptions of Taylor Swift's listeners and critics, the observational nature of our data prohibits causal inferences, and the focus on a single artist limits generalizability to a broader context outside of Taylor Swift. Our study adds to the existing knowledge base by emphasizing the nuanced relationships between musical qualities and audience and critic perceptions; however, more research into additional variables and broader contexts is required to improve understanding in this area.

5 Conclusion

In conclusion, our study comprehensively analyzes the relationship between the danceability of Taylor Swift's music and its critical reception. Our findings suggest a significant negative correlation between *user* scores and the danceability of Taylor Swift's albums, suggesting that as user ratings increase, danceability decreases. Conversely, the relationship between *critic* ratings and danceability was not significant, which suggests that there is not a significant linear relationship between critic approval and danceability of Taylor Swift's music. The difference in significance between the models utilizing user scores and the model using critic scores shows that the differences in criteria for ratings could be quite different between the two populations. While our findings are significant, they are limited by the scope and context of our data, as our data eliminates singles and some versions of Taylor's music owned by Big Machine Records, only utilizing Taylor's Version or more recently produced albums. Further, due

to the subjective nature of music ratings, the generalizability of our results is limited. Future research could expand on the topic by incorporating a broader range of musical attributes, or by exploring other artists and genres.

References

- R Core Team (2023). *_R: A Language and Environment for Statistical Computing_*. R Foundation for Statistical Computing, Vienna, Austria. <<https://www.R-project.org/>>
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *_Journal of Open Source Software_*, *4*(43), 1686. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.
- Xie Y (2023). *_knitr: A General-Purpose Package for Dynamic Report Generation in R_*. R package version 1.44, <<https://yihui.org/knitr/>>.
- Yihui Xie (2015) *Dynamic Documents with R and knitr*. 2nd edition. Chapman and Hall/CRC. ISBN 978-1498716963
- Yihui Xie (2014) *knitr: A Comprehensive Tool for Reproducible Research in R*. In Victoria Stodden, Friedrich Leisch and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC. ISBN 978-1466561595

Appendix

Figure 1: Table of Summary Statistics

Album Name	Mean Danceability	Standard dev. Danceability	Mean User Score	Mean Critic Score
Lover	0.66	0.16	8.4	79
reputation	0.66	0.08	8.3	71
Midnights	0.63	0.12	8.3	85
1989	0.62	0.12	8.2	76
Red (Taylor's Version)	0.58	0.08	9.0	91
Fearless (Taylor's Version)	0.55	0.09	8.9	82
Speak Now	0.55	0.09	8.6	77
Taylor Swift	0.55	0.09	8.5	67
folklore	0.54	0.14	9.0	88
evermore	0.53	0.09	8.9	85

Figure 2: Plot of danceability score vs. user ratings with simple linear regression line

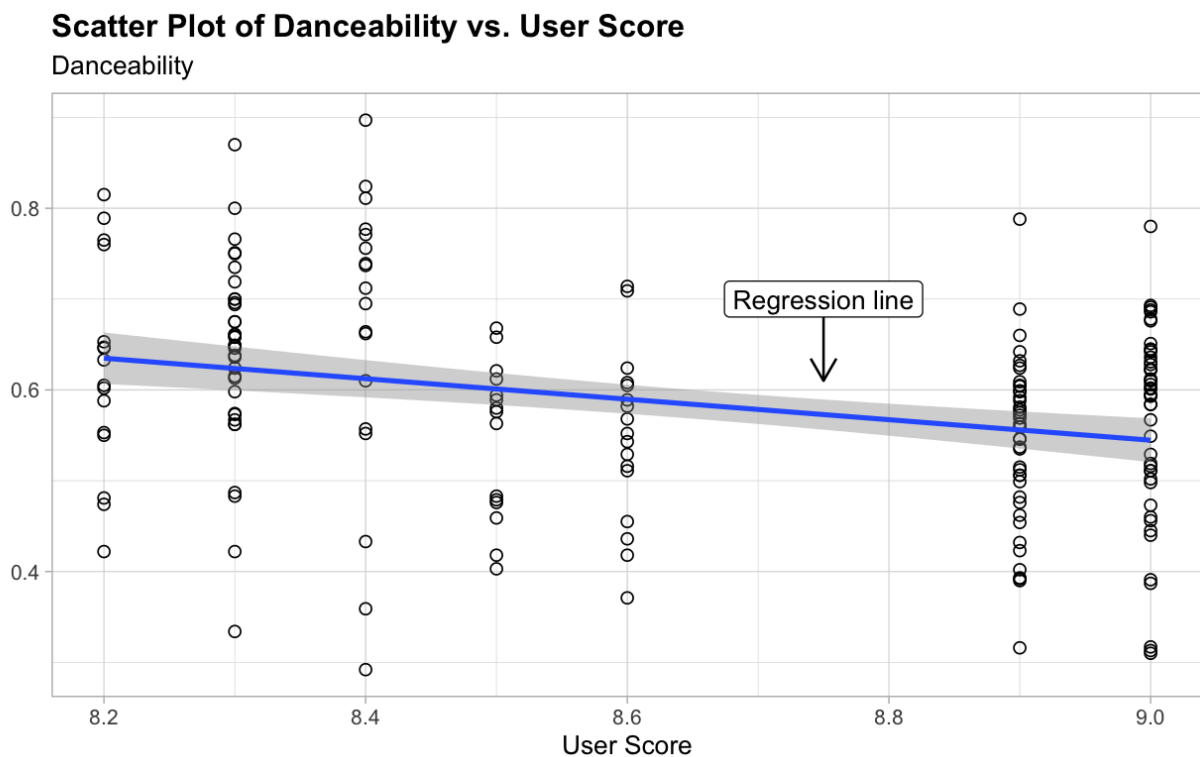


Figure 3: Plot of danceability score vs. Metacritic ratings with simple linear regression line

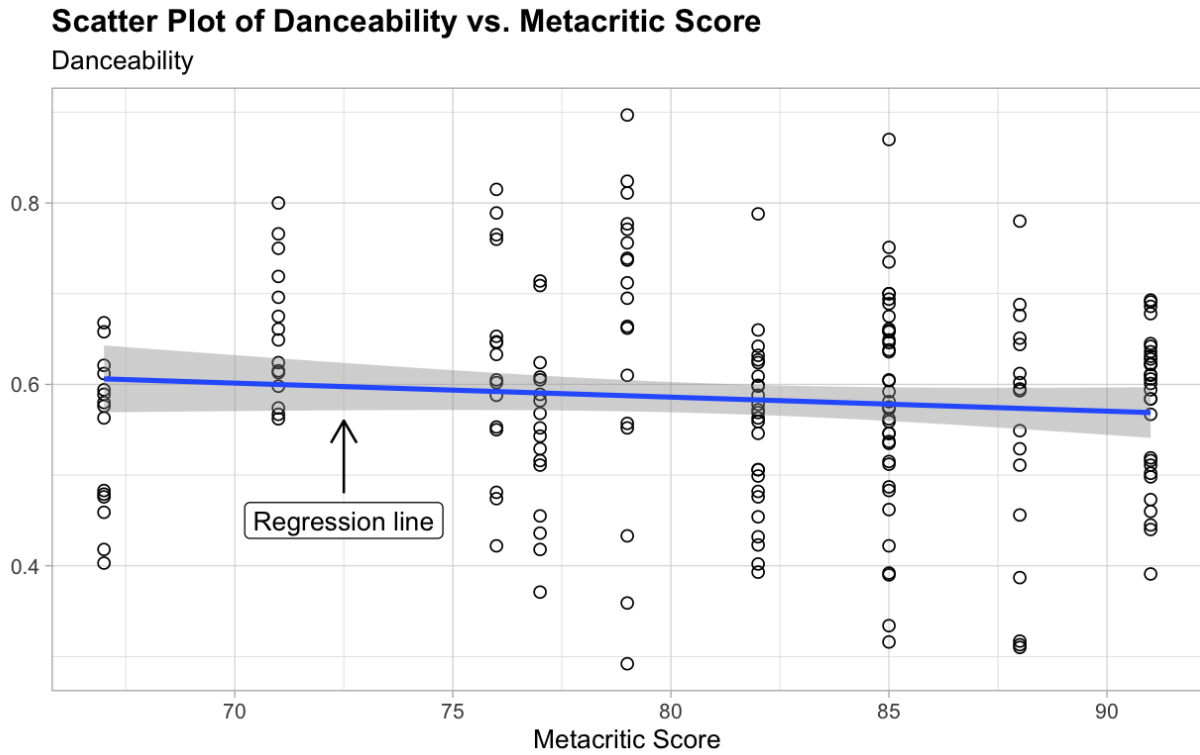


Figure 4: Simple linear regression diagnostics for model of danceability vs. user score

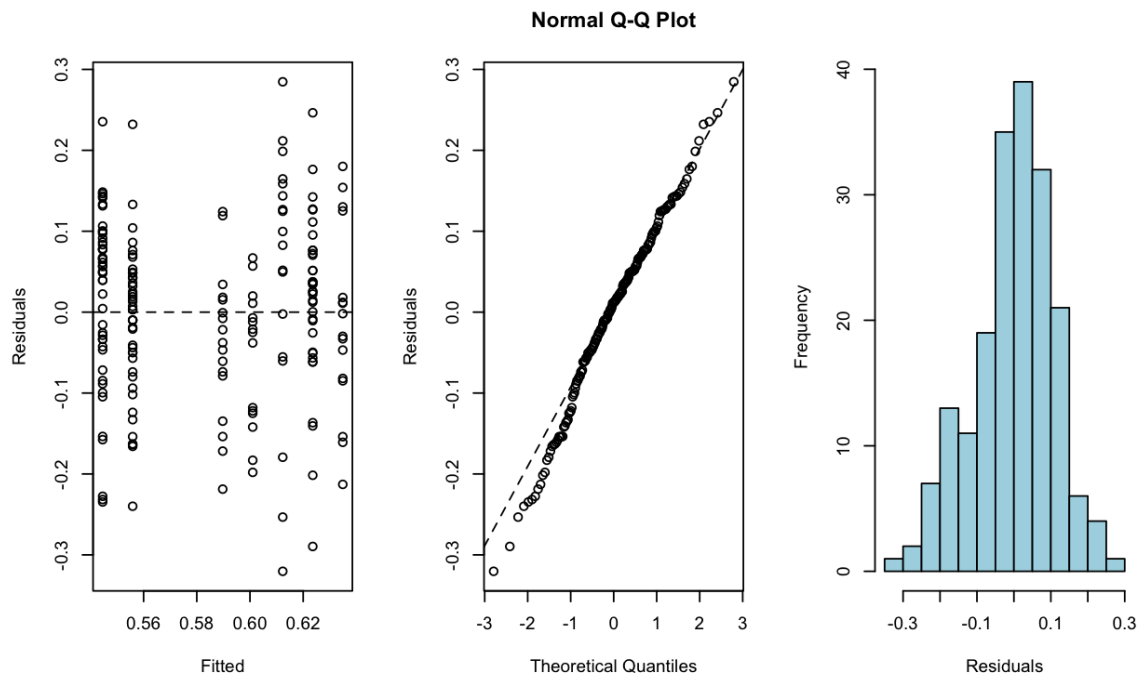


Figure 5: Simple linear regression diagnostics for model of danceability vs. Metacritic score

