A Causal Analysis of Determinants that Impact a Song's Success

Victoria Ontiveros, Rishika Shah, Randy Truong, Alex Zhong, Christine Zhou

QTM 220 Group 6

Dr. Seunghwa Rho

Word Count: 3,937

Abstract

The entertainment industry is in high demand, and continues to soar beyond the obstacles of the pandemic. As genres and songs become increasingly complex, a topic of interest is whether any song formulae are determinants of success, using a sample of the Million Song Dataset. We describe its creation process, its content, and its possible uses. This project examines song elements across a variety of genres using pop and non-pop songs as a starting point using regression analysis, and seeing whether it is worthwhile to run new analyses. We concluded that some elements of music are important to determine whether a song is considered a hit. Songs that were released more recently have faster tempos and are longer. Pop songs are loud, have more beats, and start earlier. Pop songs have slower tempos, and tend to be shorter.

Introduction and Literature Review

With the rise in digital streaming services, listening to music is becoming more accessible and popular in our world. Listeners can access their favorite songs or discover new hits through platforms like Apple Music, Spotify, Soundcloud, and YouTube; however, as new songs are recorded and shared each day from a wide range of genres and artists, listeners must sample their own dataset of music– a playlist– from the available population, which can be either enjoyable or tedious. One feature that Spotify offers to facilitate this process is through curated playlists based on songs' rising popularity as well as listeners' music tastes: "New Music Friday" and "Discover Weekly." Now, modern recommender systems can predict whether a listener will like a given song based on similarities in song metrics, such as the tempo or artist origin, but one problem that persists is determining the thresholds at which users will accept songs different enough from their usual listening habits (Bohra et al., 2015; Maillet et al., 2009). Given the diverse and dynamic landscape of music, there are a variety of factors at play that influence a song's visibility and popularity among listeners, which may be valuable to unravel for improving listener satisfaction (Suh, 2019).

This paper presents a causal analysis for investigating the elements of music that influence a song's popularity based on song metadata and extracted audio analytics data from the Million Song Dataset. To analyze the factors, we modeled three research questions using linear regression for cross-sectional data: (1) How do loudness, duration, and tempo influence a song's popularity? (2) How do tempo, duration, fade-in length, and fade-out length relate to a song's release year? (3) How do we predict the genre from beats time, tempo, loudness, and duration? Although we focus on only three questions, the results from these models will help ground tangential exploration in other factors, such as the impact of an artist's presence on other artists or the effect of a song release in one genre on other genres (Mehrotra et al., 2020; Ordanini et al., 2018). We decided to use multiple linear regression (MLR) as our primary tool for analysis since we are executing a preliminary empirical research study:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_{2} + \ldots \beta_k x_k + u$$

Linear regression is useful for explaining how a regressand– our y variable, *e.g.* song popularity– varies as the regressors– our x variables– change with other factors held fixed. We deemed linear regression modeling as the most suitable method. We utilize robust linear regression to minimize sensitivity to violations of the Gauss-Markov MLR assumptions for BLUE OLS estimators.

Many prior works related to music trends, song popularity factors, and related recommendation or prediction systems have been studied. Two works used the same data from the Million Song

Dataset (Bertin-Mahieux et al., 2011; McFee et al., 2012). One reveals initial positive results for predicting the year of a song based on factors, such as song creation process and content, and presents further use cases and explorations for the data (Bertin-Mahieux et al., 2011). A related study explores how songs may change in popularity over certain years (Fossi et al., 2021). Similarly, through subsetting a sample of 10000 songs and taking a cross-section of the data by year–namely from 2000 to 2010– our research expands on the initial conclusions for predicting song year and also attempts to predict genre of songs. We also aim at segmenting years to see which genres were the most popular for a given duration. One study uses a diffusion-regression state-space model to predict outcomes in a counterfactual world, comparing the impact of new track releases on other tracks and finding evidence that new track releases can positively impact the popularity of a track from a given artist as well as for other artists (Mehrotra et al., 2020). Our study better establishes a foundation for the previous study in examining the *elements* of songs that predict hotness, and then seeing whether these particular songs with certain characteristics have a larger impact on other songs. Based on top Billboard charts, another factor that can influence a song's popularity is artist features on a song, especially if the artist collaborations are culturally different (Ordanini et al., 2018). Our study will examine elements that contribute to a song's popularity, which could be used in combination with this study to see if the elements are also jointly and positively correlated with other artists' features and song popularity. Although our data consists mostly of Western commercial music, the previous study's results about cultural influences will be valuable for our future extensions. Further, our study distinguishes between pop and non-pop songs; researchers have found that "involuntary musical imagery" or "earworms" are associated with pop songs' common melodies and faster tempos

(Jakubowski et al., 2017). Through this study, our results about elements related to pop songs can be validated.

As our work pertains more to causal inference and analysis, we emphasize less about prediction; however, predictive analysis lends itself as a natural extension to our work. Some related works highlight how prediction can be used with song-related data and music-based systems. One study attempts to predict a listener's music tastes according to song type, suggest song recommendations, and pool users into specific profiles (McFee et al., 2012). Our study also attempts to predict certain characteristics in songs, which allow us to group songs based on similar elements. Our work would naturally pipe into future work on song prediction and recommendation systems. In another study, researchers recognize that there are a lot of factors that impact a song's popularity based on modeling Spotify's trending music. As a result, they analyze and diminish the errors associated with erroneous factors using principal component analysis (PCA) and model blending (Ge et al., 2020). This study is closely related to our own, however we take a subset of elements to analyze not only popularity but also year and genre. We believe that these factors need to be separately examined to create the best analysis.

Theory, Hypotheses, Data, and Methods

Our analysis used R and RStudio. We used the Million Song Dataset, acquired from CORGIS (The Collection of Really Great, Interesting, Situated Datasets), which contains song metadata and audio analytics about one million popular songs; however, due to big data constraints, we used a randomly sampled subset of this dataset containing 10,000 observations. We cleaned and wrangled the dataset to remove missing or improper data and outliers.

We selected columns from the dataset accordingly for each question. As the dataset contains songs between 1940 and 2010, we filtered out a specific time frame– from 2000 to 2010– to create a cross-section for the linear regression model. We chose this range because it coincided with the onset of music streaming services, and because it is relatively static in terms of pop music trends. Our research questions are as follows:

1. How do loudness, duration, and tempo influence a song's popularity?
2. How do tempo, duration, fade-in length, and fade-out length relate to a song's release year?
3. How do we predict the genre from a song's beats, tempo, loudness, and duration?

Question 1

In our first question, we tested the relationship between song popularity ("hotness") and various elements of the song–loudness, duration, and tempo–to evaluate their potential for song hotness prediction. Our hypothesis was that loudness matters because trending songs tend to have strong beats that would increase loudness. As trends may change over time, we filtered songs to be within a particular short time-span (2010-2015) to minimize time as a confounder. Since it may be likely that popularity is maximized at the middle for tempo and duration (songs that are too fast/slow or too long/short may not be as attractive), we conducted a linear regression that included the quadratic terms for these regressors: hotness ~ loudness + duration + tempo + duration^2 + tempo^2. Finally, since hotness is likely to vary differently for songs at different tempos and durations, heteroskedasticity-robust variances and test statistics were used to account for any uneven distributions in the error term. Before conducting this and all analyses described below, we filtered out observations with NA/missing/improper data using R. Our null hypothesis is $\beta 1 = 0$, and our alternative hypothesis is anything different to the null.

Question 2

In our second question, we examined temporal trends for certain time-oriented song elements: tempo, duration, fade-in length, and fade-out length. Our hypothesis is that tempo and duration will be significant. Since certain levels of these elements (e.g. a longer fade-out or a faster tempo) may go in or out style over time, it would be reasonable to test for these trends as a time series with the regression: element (tempo, duration, end_of_fade_in, or fade_out_length) ~ year. It would be likely for songs from later years to contain a larger variety of songs from different locations/cultures, so we would expect a greater variance in the elements for later years, and thus heteroskedasticity in the regressions. We thus used robust variances to calculate the test statistics for the coefficients. Our null hypothesis is $\beta 1 = \beta 2 = 0$, and our alternative hypothesis is anything different to the null.

Question 3

Thirdly, we examined whether song genres can be predicted by beats, tempo, loudness and duration. Our hypothesis is that beats, tempo loudness and duration are all relevant factors when predicting a genre, as songs with similar characteristics are categorized in the same genre. For example, pop songs tend to be similar in length and beats. Since these factors may all contribute to the genre categorization, we ran the regression: is_pop_num ~ song.beats_start + song.tempo

+ song.loudness + song.duration. Since genre varies for different songs at different beats, tempo, loudness and duration, heteroskedasticity-robust variances and test statistics were used to account for any uneven distributions in the error term. Before conducting this and all analyses described below, we filtered out observations with NA/missing/improper data using R. We would filter out songs to be within a particular year, then since all variables are continuous, numerical variables, we would conduct a heteroskedasticity-robust linear regression. Genre ~ beats + tempo + loudness + duration. Our null hypothesis is $\beta1 = \beta2 = \beta3 = \beta4 = 0$, and our alternative hypothesis is anything different to the null.

Results and Discussion

The results from running tests in RStudio are as follows:

Question 1

We explored the relationship between song popularity and their attributes: duration, tempo and loudness. Below are summaries of the four variables included in the regression model.

1) **song.hotttnesss**

This is a measure of a song's popularity, measured on a scale from 0 to 1 (least popular to most popular). From the summary statistics (Table 1a): the range is 0.1939-0.9843. The mean is 0.5112. The distribution of the song.hotttness values slightly right-skewed (Figure 1a).

2) **song.tempo**

A song's tempo refers to its pacing; it is measured in beats per minute (bpm). From the summary statistics (Table 1b): the range is 30.83-258.68. The mean is 126.57. The distribution of song.hotttness is slightly left-skewed but normally distributed (Figure 1b).

3) song.duration

This is the length of the song measured in seconds. From the summary statistics (Table 1c): the range is from 31.32-798.20 seconds. The mean is 236.81 seconds. Figure 1c shows the distribution of the values, which seems normally-distributed but slightly right-skewed.

4) song.loudness

The loudness of a song is measured relative to a chosen standard value of 60 dB. A more negative value indicates a softer song and a value closer to 0 indicates a louder song. From the summary statistics (Table 1d): the range is -31.982–0.566 decibels. The mean is -8.051 decibels. The first quartile, median, and third quartile are -9.736, -6.825, and -5.095 decibels respectively. Figure 1d shows the distribution of the values, which is left-skewed.

The equation we get from running the regression (Table 1e) is:

$$song.hotttnesss = 0.5824 + 0.00001130\, song.duration + 0.00003259\, song.tempo + 0.009356\, song.loudness - 0.00000004762\, song.duration^2 + 0.0000001813\, song.tempo^2 + u$$

The **β** estimates for song.duration and song.tempo are very close to zero, and the corresponding p-values are large, indicating that there is no discernable relationship between the length of the song or its tempo and how popular the song is. Additionally, the small **β** estimates of our quadratic terms (song.duration^2 and song.tempo^2) are extremely small and insignificant, signaling that it is unlikely that there is a nonlinear, quadratic relationship between the length or tempo of a song and the song's popularity. The **β** estimate for song.loudness is statistically significant; however, since the **β** estimate is essentially zero (0.009356), song loudness *realistically* does not have much influence on its popularity.

Question 2

For our second research question, we wanted to explore the temporal trends of song attributes, including tempo, total length, fade-in length, and fade-out length. There are seven variables that would be included in this regression model, listed below with a summary for each:

1) **song.year**

This is the year that the song was released. After removing songs with null or improper years (e.g. year = 0), we have the following summary statistics (Table 2a): the range is from years 1926-2010. The mean is 2001. Figure 2a below shows the distribution of the values, which is heavily left-skewed and almost resembles an exponential curve.

2) **song.tempo**

This is the tempo of the song in bpm. After removing songs with null or improper tempos (e.g. tempo <= 0), we have the following summary statistics (Table 2b): the range is from 30.8-258.68 bpm. The mean is 125.49 bpm. Figure 2b shows the distribution of the values, which is slightly right-skewed but appears normally distributed.

3) **song.duration**

This is the length of the song in seconds. After removing songs with null or improper lengths (e.g. duration <= 0), we have the following summary statistics (Table 2c): the range is from 7.131-1598.197 seconds. The mean is 238.387 seconds. Figure 2c shows the distribution of the values, which is right-skewed due to some outliers with very high durations, but appears normally distributed. This is the distribution after removing outliers (> 800).

4) **song.end_of_fade_in**

This is the time (in seconds after the start of the song) that the song's fade-in ends. Since the start is at t=0, this variable represents the length of the fade-in. After removing songs with null or improper fade-in lengths (e.g. fade-in length < 0), we have the following summary statistics (Table 2d): the range is from 0-38.249 seconds. The mean is 0.8007 seconds. Figure 2c shows the distribution of the values, which is heavily right-skewed, since most songs have either a very short fade-in or no fade-in.

5) fade_out_length

This is the length of the song's fade-out, in seconds. This variable was calculated using the equation: $fade\_out\_length = round(song.duration, 3) - song.start\_of\_fade\_out$, where song.start_of_fade_out is the time after the start of the song that the song's fade-out begins. Since this variable had a lower precision than song.duration, we chose to round song.duration to the precision of the fade-out variable. After removing songs with null or improper fade-out lengths (e.g. fade-out length < 0), we have the following summary statistics (Table 2e): the range is from 0-50.375 seconds. The mean time is 8.590 seconds. Figure 2c shows the distribution of the values, which is right-skewed.

Regression

We ran the following series of regressions, using robust variance-covariance and test statistics:

$lm(song.tempo \sim song.year)$,

$lm(song.duration \sim song.year)$,

$lm(song.end\_of\_fade\_in \sim song.year)$, and

$lm(fade\_out\_length \sim song.year)$.

The regression results above show that **β** estimates in all regressions was very close to zero, with **β** being positive for song.tempo and song.duration, and negative for song.end_of_fade_in Using robust variance, the only significant (α=0.05) song.year **β** estimates is in the song.duration regression (Tables 2f-2i), which indicates a positive trend for song duration over time, and likely no significant trend for tempo, fade-in length, and fade-out length.

From our analysis, we conclude that songs released in later years tend to be longer.

Question 3

For our third research question, we wanted to explore whether the genre can be predicted from the beats, tempo, loudness, and length of each song. Listed below is a summary for the five variables included in this regression model:

1) **is_pop_num**

This is a binary dependent variable created based on the artist.terms variable. If artist.terms is "pop" (it is a pop song), then we labeled it_pop_num=1 and and 0 otherwise. This was derived from whether values in 'artist.terms' contained the substring "pop."

2) **song.beats_start**

This is the start time of each beat, measured in beats. We have the following summary statistics (Table 3a): the range is from 0.02065-9.17234. The mean is 0.41737 seconds. Figure 3a shows the distribution of values, which is highly right-skewed.

3) **song.tempo**

This is the tempo of the song in bpm. After removing songs with null or improper tempos (e.g. tempo <= 0), we have the following summary statistics (Table 3b): the range is from 30.93-258.68 bpm. The mean is 125.82 bpm. Figure 3b shows the distribution of the values, which is slightly right-skewed but appears normally distributed.

4) song.loudness

This is the general loudness of the track, measured in dB. After removing outliers, NA's and other incorrect values, we have the following summary statistics (Table 3c): The range is -38.148-0.566 decibels. The mean is -8.492 decibels. Figure 3c shows the distribution of the values, which is highly left skewed.

5) song.duration

This is the length of the song in seconds. After removing songs with null or improper lengths (e.g. duration <= 0), we have the following summary statistics (Table 3d): -7.131-1598.197 seconds. The mean is 239.646 seconds. Figure 3d shows the distributions of values, which appears highly right-skewed, but if remove the outliers, the distribution will appear slightly less right-skewed.

Regression

We ran the regression:

$$lm(is\_pop\_num \sim song.beats \ + \ song.tempo \ + \ song.loudness \ + \ song.duration)$$

The equation we get from running the robust regression (Table 3e) is:

$$is\_pop\_num \ = \ 0.2215 \ + \ 4.816 * \ 10^{-3} song.beats \ - \ 2.194 * \ 10^{-4} song.tempo \\ + \ 3.713 * \ 10^{-3} song.loudness \ - \ 1.856 * \ 10^{-4} song.duration \ + \ u$$

The **β** estimates for beats and loudness are positive, while the **β** estimates for tempo and duration are negative, indicating that there is a positive relationship between pop songs and beats and loudness, and a negative relationship between pop songs and tempo and duration. The corresponding p-values for beats and tempo are non-significant (>0.05).

From this preliminary analysis using the LPM, we conclude that songs that are louder and shorter are more likely to be pop songs.

Discussion

Question 1

While regression results indicated that only the coefficient for loudness was significant, the value for the coefficient, 9.356e-03, is so small that realistic changes in song loudness (<10 dB) are unlikely to largely impact on hotness. Duration and tempo are not significant. Duration may not matter because in streaming, premium accounts on Spotify count one stream after 30 seconds of listening to the song (Spotify for Artists, n.d.). Therefore, if the song is skipped after 30 seconds, it may still count towards hotness, if measured by the Billboard streaming charts. Some popular songs like ballads and rap, have different tempos. Loudness affects how the song is heard, and on which platform, meaning that if a song is louder, it may be more memorable.

Question 2

Only song duration significantly increased over time, while tempo, fade-in, fade-out did not significantly change. Since this dataset encompasses a time range from the 1940s to 2010, this trend might be explained by improving music player technology introduced over the 1960s and 1970s, allowing for longer songs and albums to be played on a single record, then CD (McKinney, 2015). While we found a significant increase over time with our dataset, outside analysis on song durations found that low streaming payouts may cause decreases in song length over the short-term as artists are paid per-song instead of per-minute (Sanchez, 2019).

Question 3

Loudness and duration are significant for predicting whether a song is of the pop genre. The regression coefficients are close to zero, so they may not be of practical value. Loudness is significant for predicting song popularity based on our analysis in Question 1, and many pop songs are “hot”, so there is no surprise. However, duration is significant because it was not found to be significant in predicting song hotness. Pop songs may tend to be shorter because they are designed to be simpler and easier to remember. Beats start and tempo were not found to be significant in predicting pop songs. Beats measure when the beats start, which may not impact whether it is catchy or popular compared to song tune. Tempo varies between pop songs, which may change beats, so tempo may not be a good indicator of pop.

For further study, we could look into different years, since we only looked at a subset of songs from 2000-2010 in Questions 1 and 3. Additionally, it would be worthwhile to examine songs in shorter timespans, such as over a 5-year period, and re-run these regressions to see any changes in significance or direction.

Limitations

Our dataset, curated in 2011, may be outdated, but as our study focuses on 2000 to 2010, the data is sufficient for our purposes. Since the music landscape is dynamic, music trends have changed. Our research provides an initial, empirical foundation for determinants that impact song visibility or popularity, but may not be applicable or relevant to current music trends. Additionally, the dataset may be outdated in its sampling of the music environment. The creators of the Million Song dataset noted that the dataset largely lacked songs from non-Western or traditional music genres, a large gap in the wide range of genres that streaming services can expose listeners to (Bertin-Mahieux et al., 2011). With an over-representation of American music genres, analyses and predictions from the dataset may not accurately predict recommendations for the ever-broadening market share of listeners.

For Question 3, we divided our sample into two broad genres: “pop” and “non-pop.” As a result, the genre distribution became uneven with drastically more “non-pop” songs in our sample. Given that music is diverse and elicits a multitude of aural experiences, we wonder if we can rigidly label each song by these broad categories. This categorization may disregard the different and fluid subgenres of music, *i.e.* pop-rock, pop-rap, and Canto-pop. These niche factors may influence the results of our models. Our study does not account for edge cases, such as songs that experimentally bend multiple genres in itself.

References

1. Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B., & Lamere, P. (2011). The million song dataset. *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*.
2. Bohra, T. S., Kumar, V., & Ganesan, S. (2015). Segmenting music library for generation of playlist using machine learning. *IEEE International Conference on Electro Information Technology*. https://doi.org/10.1109/EIT.2015.7293429
3. Fossi, J., Dzwonkowski, A., & Othman, S. (2021). Analyzing Music Genre Popularity. *Lecture Notes in Networks and Systems*. https://doi.org/10.1007/978-3-030-80126-7_22
4. Ge, Y., Wu, J., & Sun, Y. (2020). Popularity Prediction of Music Based on Factor Extraction and Model Blending. *Proceedings - 2020 2nd International Conference on Economic Management and Model Engineering, ICEMME 2020*. https://doi.org/10.1109/ICEMME51517.2020.00214
5. *Help—How we count streams – Spotify for Artists*. (n.d.). Retrieved May 4, 2022, from https://artists.spotify.com/help/article/how-we-count-streams
6. Jakubowski, K., Finkel, S., Stewart, L., & Müllensiefen, D. (2017). Dissecting an earworm: Melodic features and song popularity predict involuntary musical imagery. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000090
7. Maillet, F., Eck, D., Desjardins, G., & Lamere, P. (2009). Steerable playlist generation by learning song similarity from radio station playlists. *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*.
8. McKinney, K. (2014, August 18). *A hit song is 3 to 5 minutes long. Here's why.* Vox. https://www.vox.com/2014/8/18/6003271/why-are-songs-3-minutes-long
9. McFee, B., Bertin-Mahieux, T., Ellis, D. P. W., & Lanckriet, G. R. G. (2012). The million song dataset challenge. *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web Companion*. https://doi.org/10.1145/2187980.2188222
10. Mehrotra, R., Bhattacharya, P., & Lalmas, M. (2020). Inferring the Causal Impact of New Track Releases on Music Recommendation Platforms through Counterfactual Predictions. *RecSys 2020 - 14th ACM Conference on Recommender Systems*. https://doi.org/10.1145/3383313.3418491
11. Ordanini, A., Nunes, J. C., & Nanni, A. (2018). The featuring phenomenon in music: how combining artists of different genres increases a song's popularity. *Marketing Letters*. https://doi.org/10.1007/s11002-018-9476-3
12. Sanchez, D. (2019, January 18). What's the Average Length of a Song? It's Shorter Than You Think. *Digital Music News*. https://www.digitalmusicnews.com/2019/01/18/streaming-music-shorter-songs-study/
13. Suh, B. (2019). International Music Preferences: An Analysis of the Determinants of Song Popularity on Spotify for the U.S., Norway, Taiwan, Ecuador, and Costa Rica. *CMC Senior Theses*. 2271. https://scholarship.claremont.edu/cmc_theses/2271

Appendix

Question 1

Table 1a: song.hotttnesss

Minimum	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
0.1939	0.3868	0.5183	0.5112	0.6316	0.9843

Figure 1a

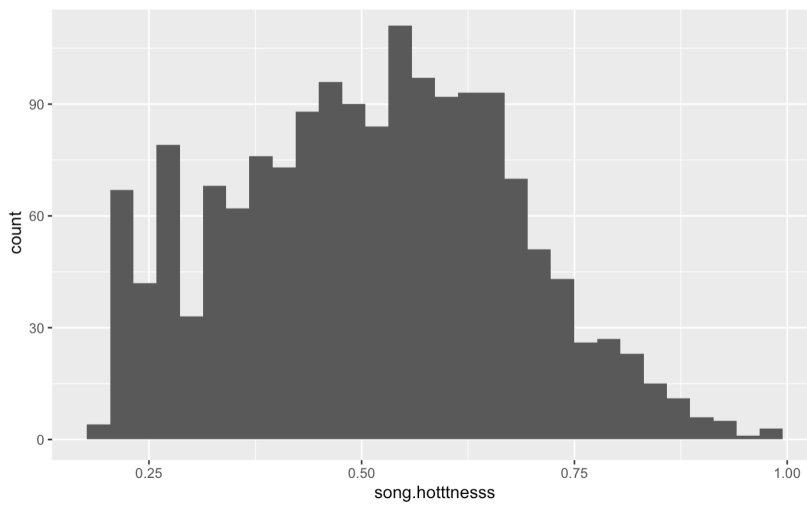

Table 1b: song.tempo

Minimum	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
30.83	99.38	123.94	126.57	148.50	258.68

Figure 1b

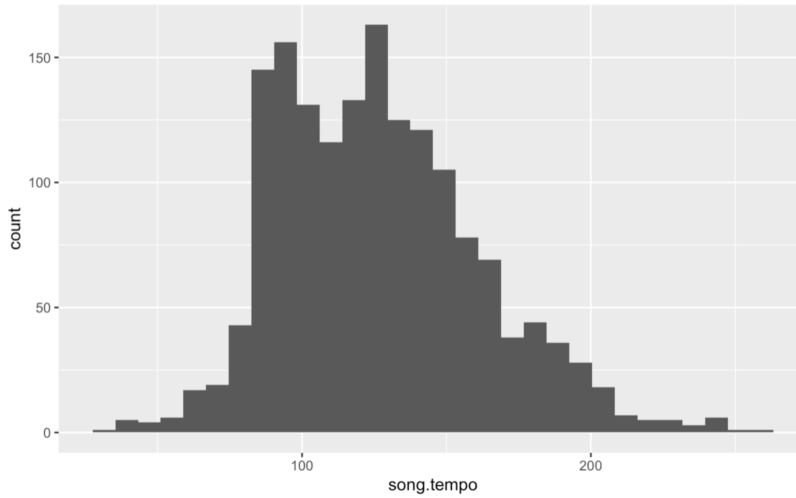

Table 1c: song.duration

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
31.32	186.46	229.64	236.81	274.42	798.20

Figure 1d

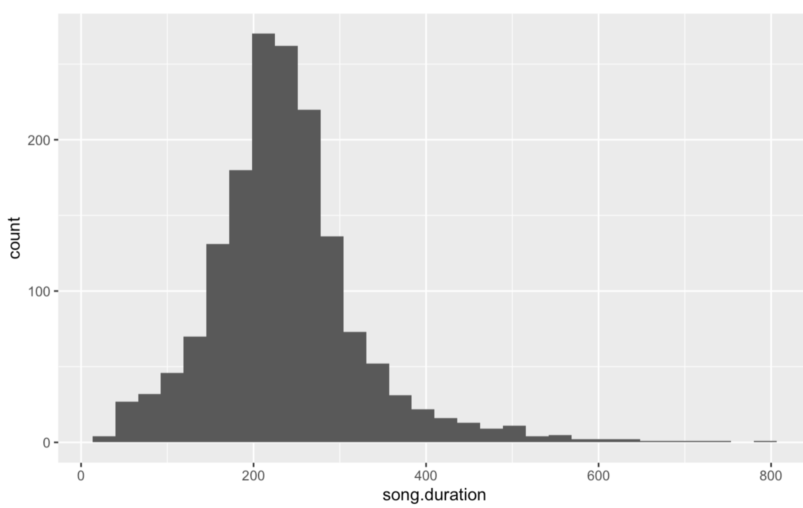

Table 1d: song.loudness

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
-31.982	-9.736	-6.825	-8.051	-5.095	0.566

Figure 1d

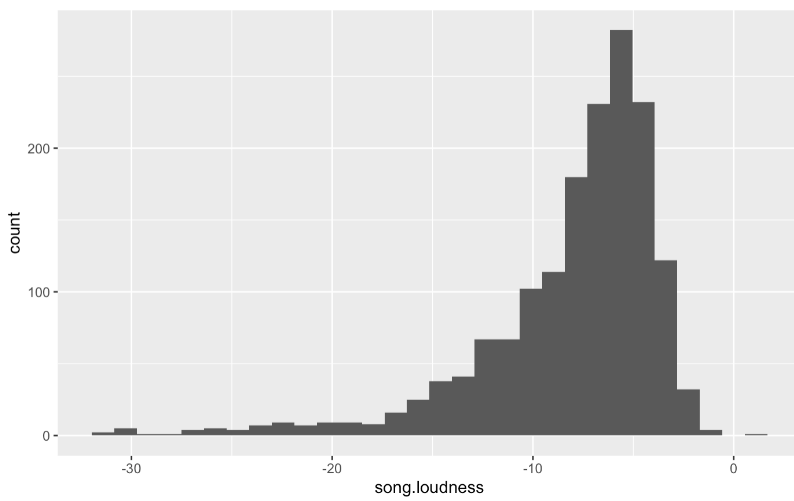

Table 1e

	Coefficient Estimate	**p-value**
Intercept	5.800e-01	<2e-16
song.duration	1.130e-05	0.8728
song.tempo	3.259e-05	0.9575
song.loudness	9.356e-03	<2e-16
song.duration^2	-4.762e-08	0.4170
song.tempo^2	1.813e-07	0.9335

Question 2

Table 2a: song.year

Minimum	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
1926	1993	2001	1997	2006	2010

Figure 2a

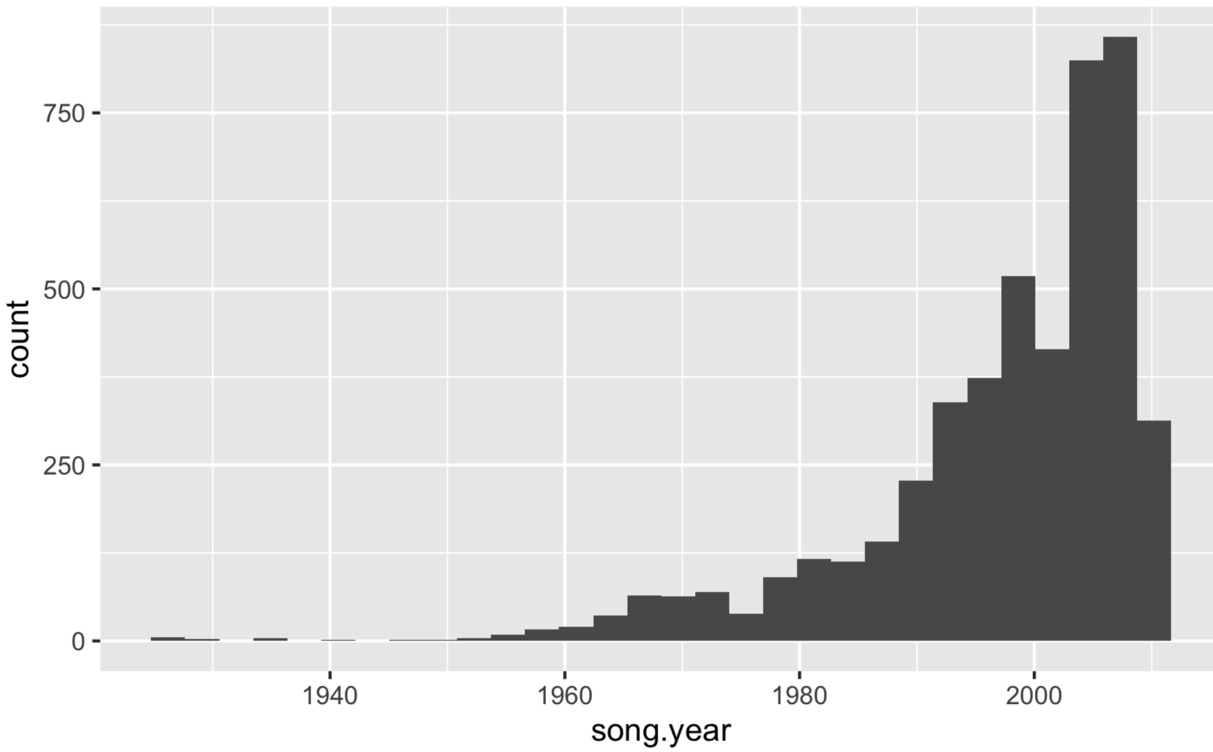

Table 2b: song.tempo

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
30.8	98.93	122.40	125.49	146.65	258.68

Figure 2b

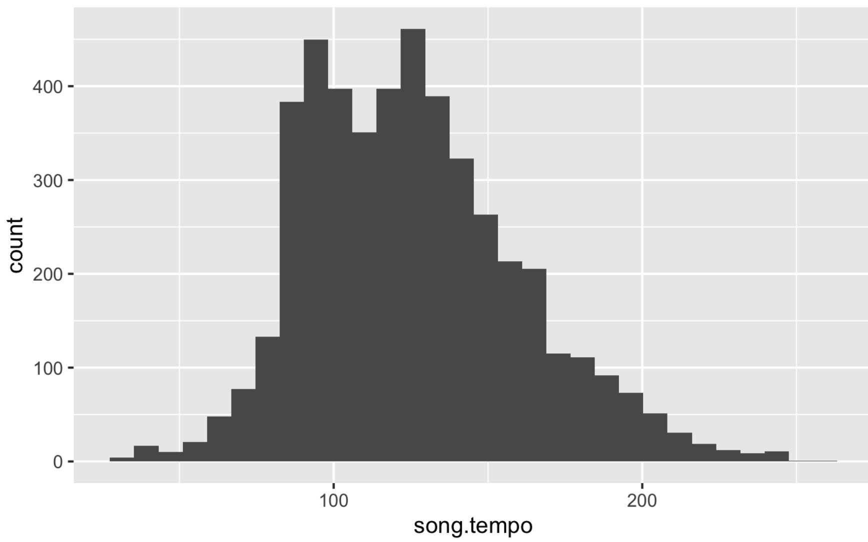

Table 2c: song.duration

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
7.131	181.178	227.369	238.509	278.387	1598.197

Figure 2c

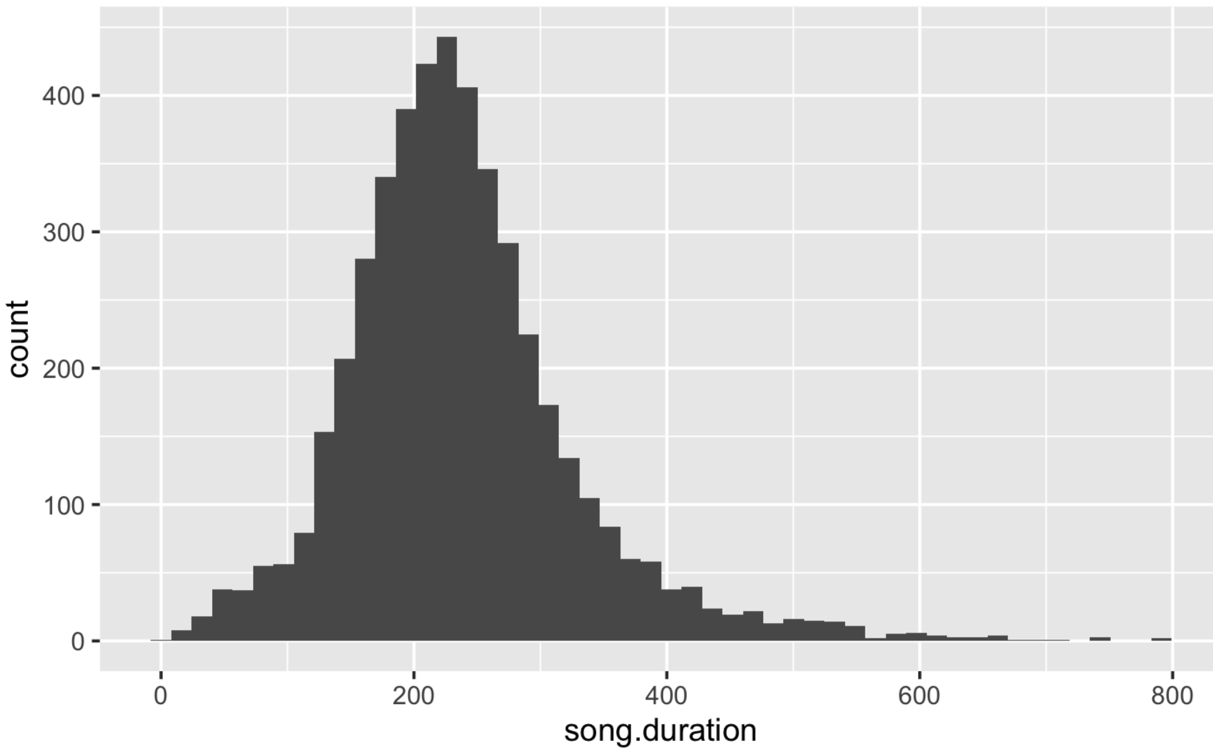

Table 2d: song.end_of_fade_in

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0	0	0.1935	0.8007	0.4193	38.2490

Figure 2d

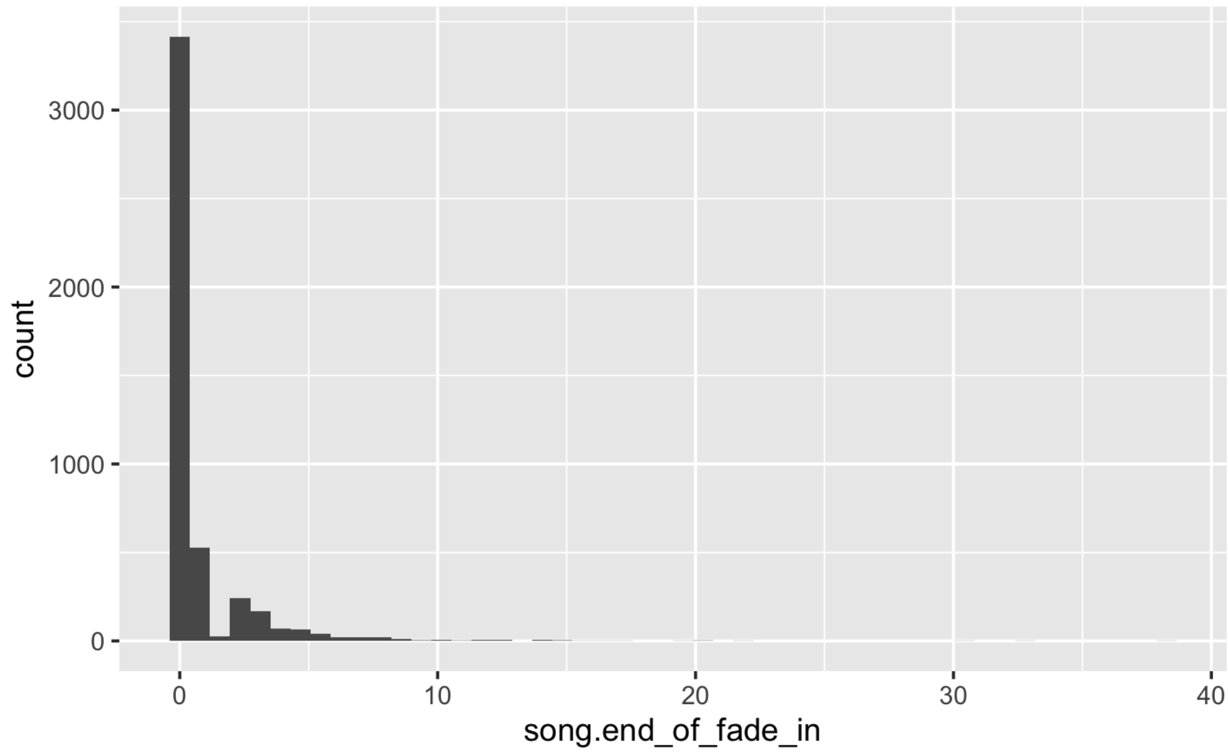

Table 2e: fade_out_length

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0	4.307	7.779	8.590	11.701	50.375

Figure 2e

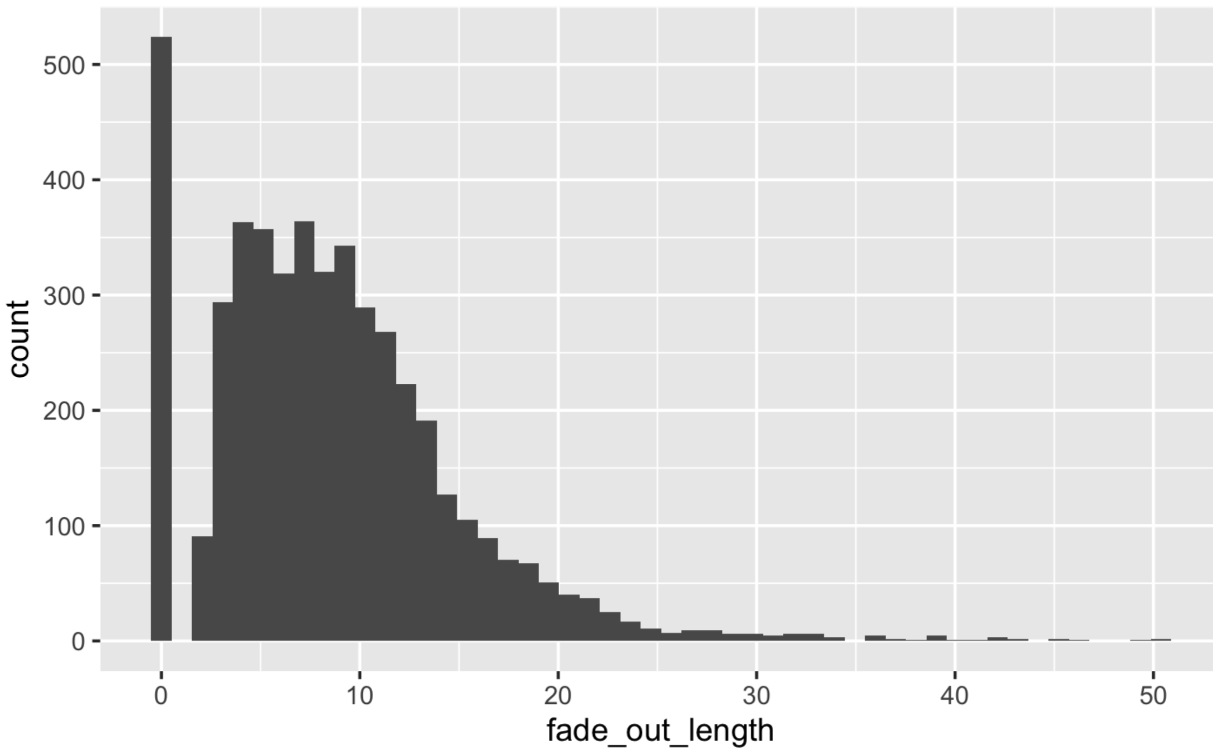

Table 2f: song.tempo

	Estimate	P-value
Intercept	-1.846295	0.9824
song.year	0.063754	0.1286

Table 2g: song.duration

	Estimate	P-value
Intercept	-764.09293	0.0007536
song.year	0.50200	9.913e-06

Table 2h: song.end_of_fade_in

	Estimate	P-value
Intercept	1.36793415	0.741
song.year	-0.00028401	0.891

Table 2i: song.fade_out_length

	Estimate	P-value
Intercept	11.4946802	0.4102
song.year	-0.0014542	0.8353

Question 3

Table 3a: song.beats_start

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
0.02065	0.18732	0.31611	0.41737	0.48531	9.17234

Figure 3a

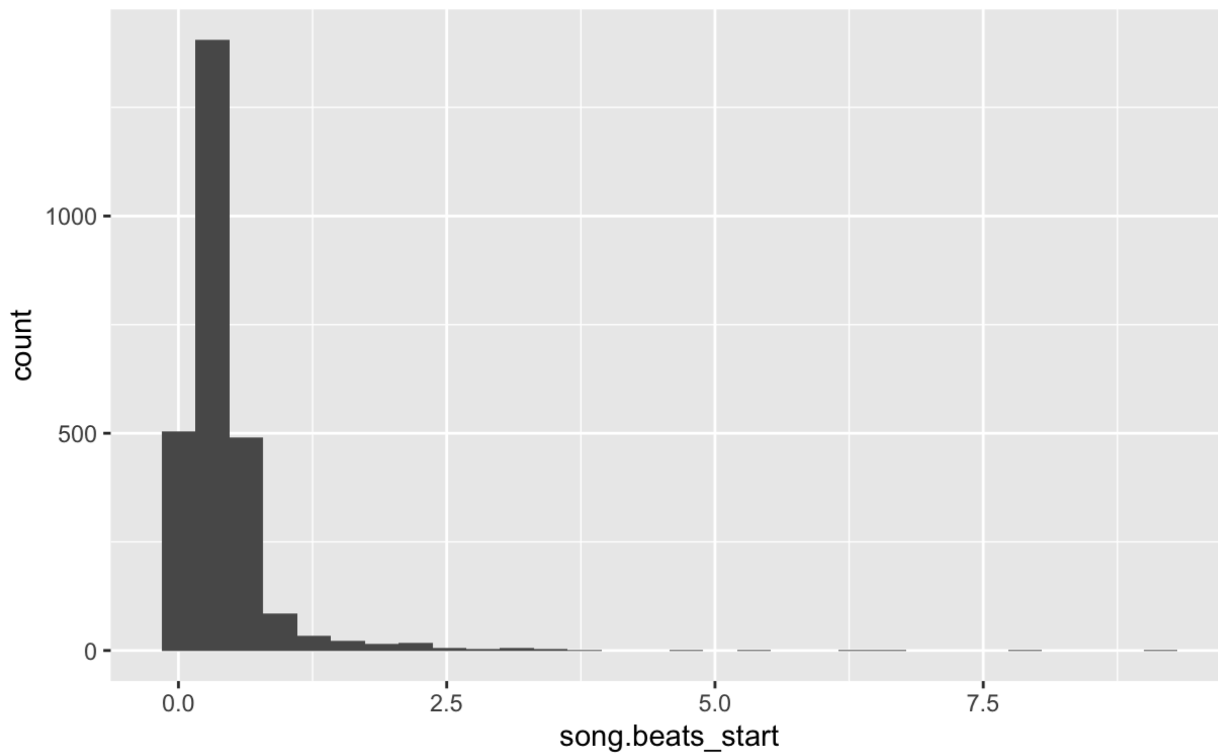

Table 3b: song.tempo

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
30.83	98.03	123.00	125.82	147.26	258.68

Figure 3b

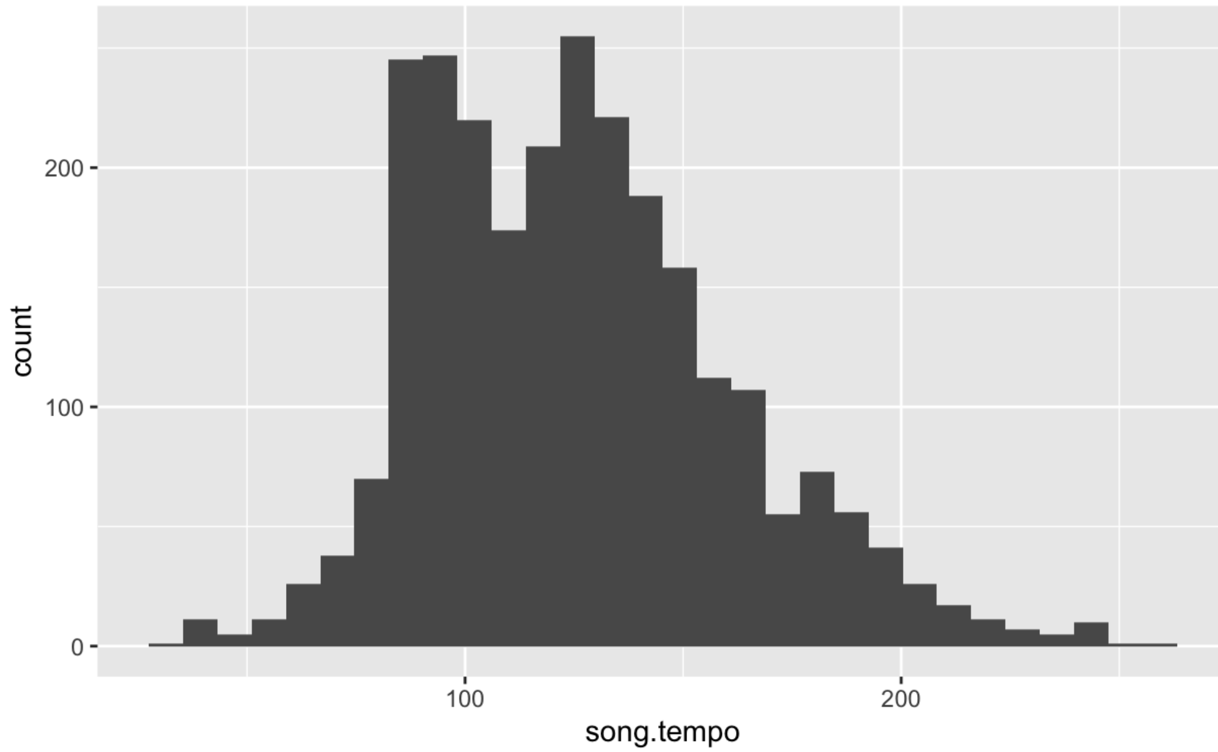

Table 3c: song.loudness

Minimum	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
-38.148	-10.377	-7.087	-8.492	-5.210	0.566

Figure 3c

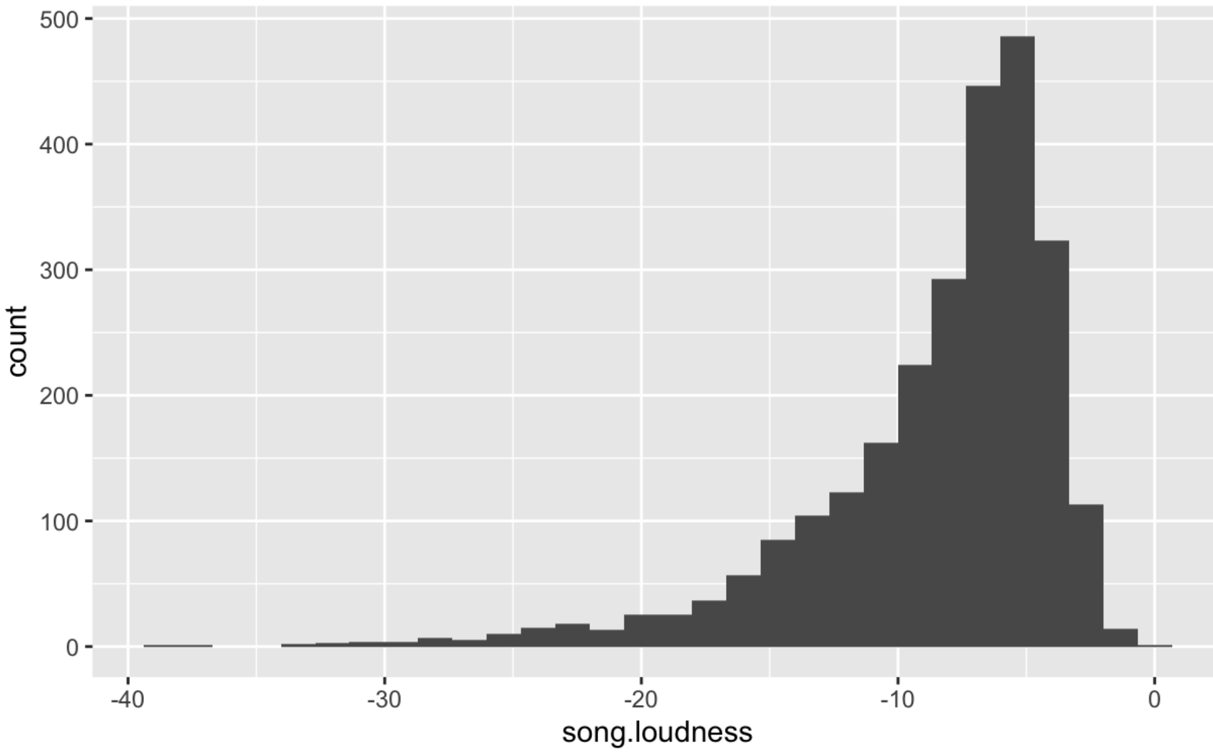

Table 3d: song.duration

Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
7.131	186.070	228.963	239.646	276.375	1598.197

Figure 3d

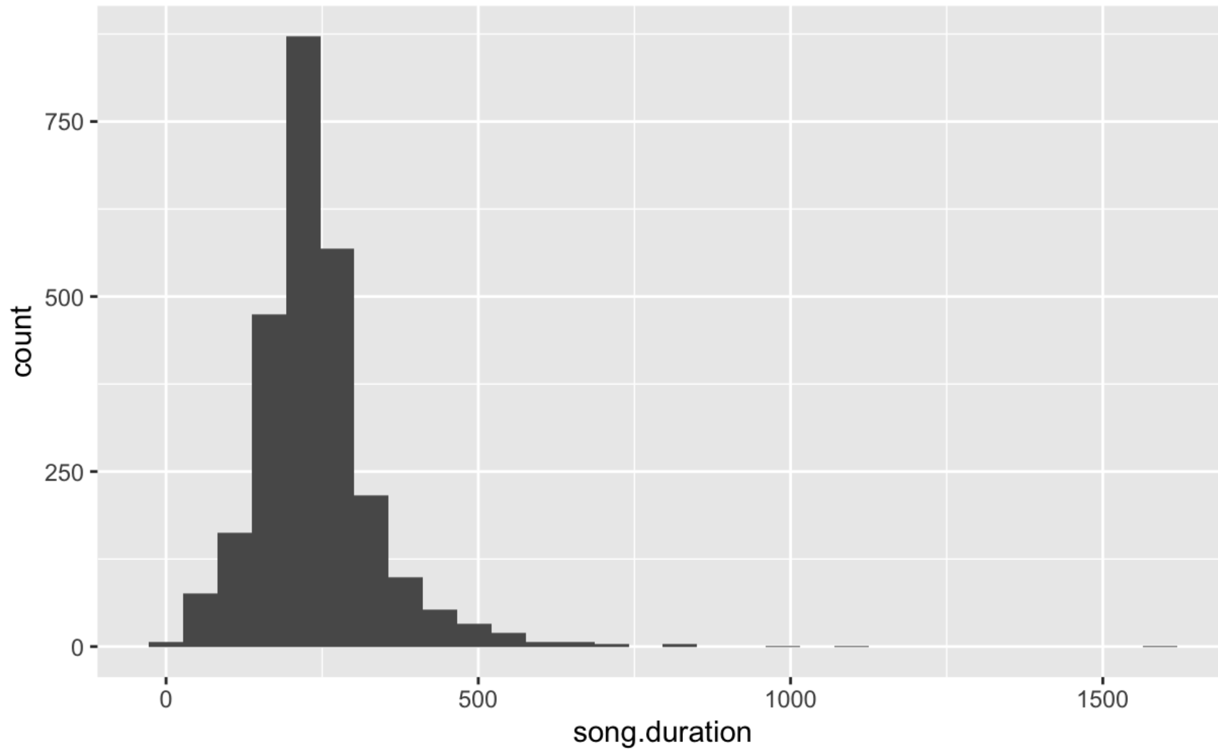

Table 3e

	Estimate	**P-value**
Intercept	2.215e-01	1.298e-13
song.beats	4.816e-03	0.6437959
song.tempo	-2.194e-04	0.1828949
song.loudness	3.713e-03	0.0021985
song.duration	-1.856e-04	0.0001709