A Textacular Take on Taylor Swift's Songwriting Style

Authors: Jethro R. Lee and Michelle Wang

Link to code repository: https://github.com/jethrorlee731/taylor_swift_textacular_analysis

The research data was obtained from various visualizations created by Python to analyze Taylor Swift's writing style. The data revolves around her songs' word counts, word lengths, and sentiment tones to make quantitative and qualitative conclusions about why her career has been so successful. The songs chosen for analysis encompass the entire span of Swift's career to emphasize any trends or discrepancies within her compositions over time.

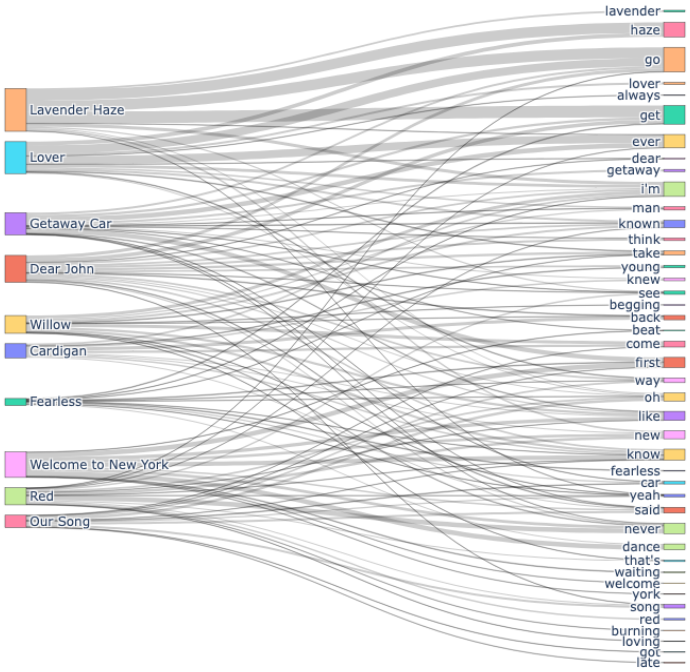
The files we used for this mini project contain song lyrics from a tune from each of Taylor Swift's ten albums. The world throughout the times when these albums were made, along with Swift's emotions, fluctuated a lot. These variations contributed to the distinctive moods that each album provides.

Taylor Swift, with hits such as "Our Song," was made during the early stages of Swift's career and contains a light-hearted and innocent tone. *Fearless*, which includes tracks such as "Fearless," displays Swift's confidence more brightly with a country style. *Speak Now*, which includes songs such as "Dear John," has fantastical songs that were also written when Swift experienced a breakup with artist John Mayer due to how supposedly abusive their relationship was. *Red*, with classics such as "Red," marks the time when Swift boldly started breaking away from country and used pop to express her increasingly fiery emotions. *1989*, with hits such as "Welcome To New York," was Swift's first full pop album and contains many bops that reflect the excitement she likely had moving into New York City. *Reputation*, with songs such as "Getaway Car," was made as a response to the unfair media backlash she received after Kanye West and Kim Kardashian tried to defame her. *Lover* completely contrasts *Reputation*, with songs such as "Lover" which emphasize Swift's desire to promote love and harmony. *Folklore*, with tracks such as "Cardigan," and *Evermore*, with songs including "Willow," were both written during the pandemic and hence contain a somewhat melancholic and subtle feel. Her most recent album, *Midnights*, with hits such as "Lavender Haze" is very autobiographical, with the premise of things that keep her up at night.

All the visualizations we created below use the songs referenced above in the form of text files. However, our framework is extensible and allows users to pass in JSON, CSV, and Excel files too. Also, our framework works for any type of text data, not just song lyrics. The framework includes different parsers that clean loaded files of different types of unnecessary whitespace, punctuation, and capitalization. The cleaned data is then passed to the visualization functions, each of which can produce the following: a Sankey diagram that maps words to texts, word clouds, a scatter plot comparing a file's positive score to its negative score, a bar chart comparing sentiment scores, boxplots summarizing word length distributions, a bar chart comparing average word lengths, and a boxplot presenting the distribution of word lengths for the words from all the registered files. Jethro mainly worked on the Sankey chart, word clouds, and sentiment bar charts while Michelle made the sentiment score scatter plot, boxplots, and bar charts. We contributed equally to the parsers and exception handling.

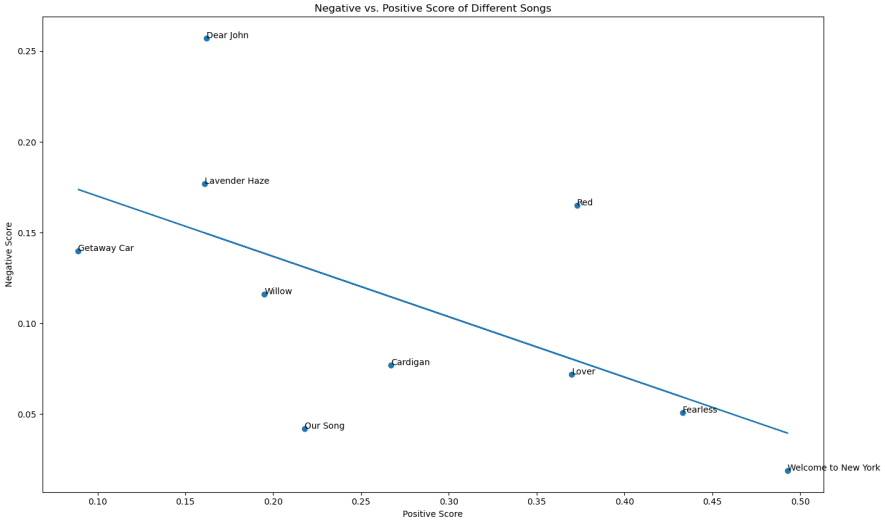
The Sankey diagram gives insight into Swift's writing style, particularly the words she uses the most often. Our diagram allows the user to specify what files or data are passed in, an optional list of words they want to appear in the diagram, as well as an optional integer representing the number of most common words they are interested in across each of the files. The Sankey diagram shows the degree to which the top 5 most frequented words in each song shown on the left nodes are included across those tunes. Thus, the chart shows that some of

the most popular words that Swift uses across each of the registered songs are “know,” “like,” “go,” “get,” “ever”, and “new.” Accordingly, upon closer observation, “know” appears throughout many of the repeating choruses in Swift’s work. “Know” appears in the choruses of “Our Song” (“‘Cause it's late and your mama don't **know**”), “Red” (“Forgetting him was like trying to **know** somebody you've never met”), “Fearless” (“And I don't **know** how it gets better than this… And I don't **know** why but with you, I'd dance”), and “Willow” (“The more that you say, the less I **know**). As for the word “like”, “Cardigan” frequently uses it for its many similes (“'Cause I knew you steppin' on the last train; Marked me **like** a bloodstain… I knew you leavin' **like** a father; Running **like** water, I”). “Dear John” also contains a simile (“So don't look now, I'm shining **like** fireworks over your sad, empty town”). Hence, the Sankey diagram confirms that Swift enjoys using figurative language such as similes to convey themes in her songs. It’s impressive how she has maintained this degree of artistry over a wide variety of songs since the start of her career. The artistic ambiguity she employs in her songwriting opens her works to diverse interpretations, which is why so many fans feel like they can resonate with her songs.



The word cloud arrays are exciting and demonstrate the variety in Swift's Songs. We allow the users to specify many criteria for the word clouds including what files are passed in, an optional list of color schemes, and a minimum font size used for the words. “Our Song” and “Willow” appear to be the least balanced songs. “Our Song” prominently features the words “song” and “late” in its lyrics, while “Willow” heavily features “man” and “take.” Interestingly, the only word clouds that prominently display the titles of the song they represent are for “Our Song,” “Fearless,” “Red,” “Welcome to New York,” “Getaway Car,” “Lover,” and “Lavender Haze.” Furthermore, “Red,” “Lover,” and “Cardigan” seem to be the most lyrically balanced songs since the word clouds that represent them don’t feature a small set of words that are significantly larger than those around them. Instead, they each have a lot of words around the same size. Swift has mentioned that she creates songs that typically fall into one of these categories: “fountain pen” songs, which are very detailed contemporary autobiographies, “quill pen” songs, which are old-fashioned and refined, and “glitter gel pen” songs, which are

not lyrically complex but are meant for people to jam to. “Dear John,” “Cardigan,” “Lover,” and “Lavender Haze” fall into the first category, which explains why their word clouds are so balanced and not skewed to a small set of words considering how Swift feels very connected with their lyrics. “Willow” and “Red” are considered “quill pen” songs, but whether a song has a 19th-century feel does not necessarily reflect in the counts for each of its distinct words. Finally, “Welcome to New York” is definitely an upbeat “glitter gel pen” song, confirmed by its lack of balance in its word cloud, which suggests it contains very repetitive lyrics. Thus, the word clouds demonstrate the expansiveness of the types of pieces that Swift can compose. She can make very intricate songs that convey great sentimental meanings while also forming bops with lyrics so repetitive they become classic earworms.



The scatterplot shows the relationship between the negative and positive scores of Swift’s songs. We allow the users to choose what files are passed in as well as the number of words analyzed from each file based on their frequencies in their respective files. Unsurprisingly, there is a moderate negative correlation between negative scores and positive scores, meaning songs with more positive words tend to have fewer negative words and songs with fewer positive words tend to have more negative words. However, it is worth pointing out the outliers, “Dear John”, “Red”, and “Our Song.” “Dear John” has a positive score between 0.15 and 0.20 and a negative score above 0.25. This suggests that “Dear John” is overwhelmingly negative as the negative score is **higher** than what it would be if its scores followed the general trend. Furthermore, Red has a positive score between 0.35 and 0.40 and a negative score between 0.15 and 0.20. Its negative score is also **higher** than what it would be if its scores followed the general trend. Both are heartbreak songs, which suggests why they are more noticeably negative than Swift’s other songs. On the other hand, “Our Song” has a positive score between 0.20 and 0.25 and a negative score of less than 0.05. The negative score is **lower** than what it would be if its scores followed the general trend, which makes sense because the song celebrates the fun things that a couple could do together. Interestingly, Swift doesn't really balance the negativity and positivity of a song. If that were the case, then as the positive scores of her songs increase, the negative scores would increase, and thus the points would follow a linear pattern. However, that is not the case. Therefore, she may purposely want some songs to be extremely negative or extremely positive depending on the story behind their lyrics.

The visualizations presented above are only a subset of the visualizations that we created to analyze Taylor Swift’s writing style. Other visualizations we made include sentiment analysis bar charts for each song, boxplots about the lengths of the words for each song, a bar chart comparing the average length of the words in each song, and a box plot summarizing the length of the words for all the files combined. The text files we chose in this project represent one from each of Swift’s albums. With each of the visualizations, we were able to draw different conclusions when analyzing word counts, word lengths, and sentiment tones to try and understand why the results are the way they are and any trends or discrepancies that are present in Swift’s style of composing.

References

AllMusic, Netaktion LLC. (n.d.). *Taylor Swift Albums and Discography*. AllMusic. Retrieved February 25, 2023, from https://www.allmusic.com/artist/taylor-swift-mn0000472102/discography

Roberts, K. (2022, October 16). *Taylor Swift curated three pen-themed playlists of her own music*. Cosmopolitan. Retrieved February 25, 2023, from https://www.cosmopolitan.com/entertainment/music/a41638221/taylor-swift-pen-themed-playlists-of-her-own-music/

Setiawati, Wilya & Maryani, Maryani. (2018). AN ANALYSIS OF FIGURATIVE LANGUAGE IN TAYLOR SWIFT'S SONG LYRICS. PROJECT (Professional Journal of English Education). 1. 261. 10.22460/project.v1i3.p261-268.

"Taylor Swift – Cardigan." *Genius*, 2020, https://genius.com/Taylor-swift-cardigan-lyrics.

"Taylor Swift – Dear John." *Genius*, 2010, https://genius.com/Taylor-swift-dear-john-lyrics.

"Taylor Swift – Fearless." *Genius*, 2008, https://genius.com/Taylor-swift-fearless-lyrics.

"Taylor Swift – Getaway Car." *Genius*, 2017, https://genius.com/Taylor-swift-getaway-car-lyrics.

"Taylor Swift – Lavender Haze." *Genius*, 2022, https://genius.com/Taylor-swift-lavender-haze-lyrics.

"Taylor Swift – Lover." *Genius*, 2019, https://genius.com/Taylor-swift-lover-lyrics.

"Taylor Swift – Our Song." *Genius*, 2006, https://genius.com/Taylor-swift-our-song-lyrics.

"Taylor Swift – Red." *Genius*, 2021, https://genius.com/Taylor-swift-red-taylors-version-lyrics.

"Taylor Swift – Welcome to New York." *Genius*, 2014, https://genius.com/Taylor-swift-welcome-to-new-york-lyrics.

"Taylor Swift – Willow." *Genius*, 2020, https://genius.com/Taylor-swift-willow-lyrics.