Psychology of Aesthetics, Creativity, and the Arts

You Do Not Get to Tell Me About Sad: Swiftian Saudade in Taylor Swift's Lyrics

Jean Clipperton and Loizos Bitsikokos

Online First Publication, October 2, 2025. https://dx.doi.org/10.1037/aca0000810

AMERICAN PSYCHOLOGICAL ASSOCIATION
DIVISION 10
Psychology of Aesthetics, Creativity, and the Arts

ISSN: 1931-3896
https://doi.org/10.1037/aca0000810

You Do Not Get to Tell Me About Sad: Swiftian Saudade in Taylor Swift's Lyrics

Jean Clipperton[1] and Loizos Bitsikokos[2]
[1] MACSS, University of Chicago
[2] Brian Lamb School of Communication, Purdue University

Song lyrics can be powerful: conveying imagery, emotion, and connecting listeners to a song. Few, if any, songwriters offer a better case study than Grammy-winning pop icon Taylor Swift. Swift, known for her lyricism and shattering Billboard records, has 11 studio albums that she has penned either wholly herself or cowritten with a new album arriving in October 2025. We find that her work invokes what we term *Swiftian Saudade*—a mixture of sadness, anticipation, nostalgia, and joy. We combine a lexicon-based approach with principal component analysis of Swift's song lyrics to analyze her evolution as a songwriter, finding that Swift's lyrics can be categorized using a two-dimensional landscape of lyrically focused and emotional words. Her work has changed over time, moving from a more lyric-focused place toward one of deep and conflicting emotions, her signature *Swiftian Saudade*. This is particularly true within track five on her albums: these tracks exhibit the highest rate of saudade elements at nearly 50% and a significantly higher level of authenticity than her other tracks. Our analysis contributes to substantive work on vocal music and on Swift's songwriting, providing a nuanced approach to understanding the craft of songwriting and emotion in lyrics. We also contribute to methodological approaches to lyrical analyses, providing a novel combination of principal component analysis with lexicon-based analyses that can inform future text-based research.

Keywords: Taylor Swift, song lyrics, natural language processing, songwriting, emotion

Supplemental materials: https://doi.org/10.1037/aca0000810.supp

Taylor Swift's music is a lightning rod for attention, conversation (Sloan, 2021), and disagreement, even prompting Donald Trump to post "I HATE TAYLOR SWIFT" on social media (Cowan, 2024). Swift's rerecording and subsequent rerelease of her music as "Taylor's Version" (TV) was a watershed moment for the music industry, one that came after Swift's criticism of Spotify that also later shifted how artists were reimbursed for their work (Sweney, 2018). As Swift's fame grows, her work continues to inspire artists and fans alike, through conferences and symposia (Ellis, 2020; Swiftposium 2024, 2024) and courses (Dailey, 2023). Yet, the research on Swift is typically smaller in scope or scale, focusing on either specific albums (e.g., *folklore* and *evermore*; Harrison & Ringrow, 2022) or songs (e.g., "Look What You Made Me Do"; Alek et al., 2020), Swift herself as a celebrity (Fogarty & Arnold, 2021; Morris, 2024), Swift's economic impact (Smialek, 2021), or isolated linguistic or literary focus on her works (Kendong et al., 2023; Sloan, 2021). While some research does consider themes in Swift's work (Arnold, 2021; Santika & Syafryadin, 2023), linguistic analysis (Kendong et al., 2023), or imagery (Yastanti & Susilawati, 2020), few examine her as an artist from the perspective of her songwriting and lyrical content across her albums. We draw inspiration from these textual analyses of Swift's work, and we incorporate sentiment and emotional analysis across the entirety of Swift's catalog.

Few authors examine emotion within songs, although some include valence (e.g., Parada-Cabaleiro et al., 2024). Among those studies that consider emotion, researchers are frequently focused upon perception of emotion or mood rather than identifying how emotions are used within songs (Brand et al., 2019; Cook et al., 2019; Guo et al., 2020). In the case of Taylor Swift, we observe not only the use of different emotions but also the distinct combination of contrasting emotions such as joy and sadness, which we term *Swiftian Saudade*. *Swiftian Saudade* is persistent in her work, particularly within the fifth track on her albums. We develop a novel methodological approach to analyze the sentiment and emotional content of her text-based works, providing a two-dimensional approach to studying lyrics. Our methodological innovation incorporates lexicon-based analyses to provide insights not only into individual songs and albums but also into particular trends in Swift's lyrical construction of songs and paves the way for future study of emotionally charged texts.

Study of Music

Music plays an important role in the lives of listeners, offering social connection and a means of emotional regulation (Cook

Evangelia G. Chrysikou served as action editor.
Jean Clipperton https://orcid.org/0000-0001-8833-0052
Loizos Bitsikokos https://orcid.org/0000-0002-8201-0207
Jean Clipperton acknowledges Sanja Miklin, Zhao Wang, and Alejandro Sarria-Morales for their comments and support. Loizos Bitsikokos acknowledges the support of the Onassis Foundation through their scholarship program (Scholarship identification numbers: F ZT 010-1/2023-2024 and F ZT 010-2/2024-2025).
Correspondence concerning this article should be addressed to Jean Clipperton, MACSS, University of Chicago, 1155 East 60th Street, Chicago, IL 60637, United States. Email: clipperton@uchicago.edu

et al., 2019; Thoma et al., 2012). Lyrics become more important in sad music enhancing the emotion (Ali & Peynircioğlu, 2006), and individuals become more engaged with familiar music (Pereira et al., 2011).

To understand noninstrumental music, many scholars focus upon lyrics, and what those texts can tell us (Parada-Cabaleiro et al., 2024). Many of these focus upon a single artist. Artist-based analyses of music focus on an artist's canon, their evolution, and creative flow (Czechowski et al., 2016; Eaton et al., 2022; Kalichman & Smyth, 2023, 2024; Petrie et al., 2008; Rambsy, 2018). Lexicon-based approaches are popular for artist-centric work (Czechowski et al., 2016; Eaton et al., 2022; Kalichman & Smyth, 2023). Male artists (e.g., The Beatles, Bob Dylan, Tom Petty, and Bruce Springsteen) are the primary object of study, despite evidence that men and women write differently (Newman et al., 2008). Relatively few studies cover women and those that do are often from a lens of gender and image (Lister, 2001), lack coverage of more recent artists (Eaton et al., 2022; Kalichman & Smyth, 2023, 2024), and/or are limited in scope (DeWall et al., 2011; Dodds & Danforth, 2010; North et al., 2021).

Some scholars explore larger time periods and/or sets of artists (Brand et al., 2019; Parada-Cabaleiro et al., 2024) but not from the perspective of emotional content relative to linguistic framing. We incorporate a contemporary pop star, Taylor Swift, not only for her notability as a singer–songwriter but also because she is known for her lyric-forward approach to songwriting.

Sentiment Analysis and Emotional Detection Within Texts

Songwriters' work is typically researched through the lens of lyrical analysis, although there is no clear and universally agreed upon approach for linguistic study. We use a lexicon-based approach, in which emotion and/or sentiment are assigned to words from a predefined dictionary in the form of a score or a label (e.g., the word "bad" is typically labeled as negative in lexicon lists). In contrast, machine learning-based approaches are probabilistic and data driven: models learn patterns from annotated data during a training phase and can then provide predictions at the level of words, sentences, or other text chunks, depending on the model's training and design. For example, a machine learning model for sentiment analysis, when applied to "Honey, I rose up from the dead, I do it all the time," a lyric from "Look What You Made Me Do" on *Reputation*, might classify it as having negative sentiment because of its reference to death. On the other hand, a lexicon-based classifier would aggregate the sentiment of each word in the sentence and would potentially reach a more neutral or slightly positive classification depending on the words and their associations in the dictionary.

Machine learning-based approaches can outperform some lexicon-based approaches when appropriately fine-tuned (Srivastava et al., 2022). However, machine learning approaches are frequently sensitive to training data, data size, and quality (Taboada et al., 2011; Zhang et al., 2014). Many studies comparing machine learning and lexicon-based approaches focus on sentiment, commonly interpreted as whether a text is positive or negative. Sentiment studies use tweets and find that part of how lexicon-based approaches stumble is in the use of more informal elements, such as with emojis and emoticons (Dhaoui et al., 2017). Although Talaat (2023) demonstrates how the incorporation of emojis can affect model performance, and how limited we know lexicon-based approaches can be when dealing with less formal language, neither reason seems as compelling to rule out lexicon-based approaches for lyrical analysis. Lexicon-based approaches have similar (though slightly lower) levels of performance compared to fine-tuned machine learning approaches. Additionally, lexicon-based approaches can be implemented without the substantial demands for large-scale training data (Darwich et al., 2019; Dhaoui et al., 2017; Taboada et al., 2011; Zhang et al., 2014).

Within the realm of emotional classification, lexicon-based approaches are stable, perform well (Kušen et al., 2017), and are preferable when "simulating the effect of linguistic context" (Taboada et al., 2011, p. 269). Lyrics are a prime candidate for this approach (North et al., 2021). For these reasons, we use lexicon-based approaches to text analysis and incorporate an unsupervised machine learning-based approach (principal component analysis [PCA]) in our overall analysis, which identifies patterns in data without guidance (annotations, labeling, etc.). We focus on Swift because of her work as a songwriter and her frequent discussions about her songwriting process and "typology" of songs (Nicholson, 2022) and "checklists" for writing (Apple Music, 2020). Swift herself has been quite open about her writing process and creativity as a songwriter, making her work even more relevant for study (Alek et al., 2020; McGrath, 2023; Morris, 2024). Swift talks about her process and had been particularly open in documentaries, videos, and blogs. See, for example, the "Making of a Song" partnership with AT&T (Armstrong, 2017), her documentary "Miss Americana" (Wilson, 2020), any number of her released voice memos (e.g., memos on *1989* Deluxe Version [DE, Track 17; Coscarelli, 2019; Dailey, 2024), or of her conversations with fans (Swift, 2019).

Swiftian Saudade

In terms of other characterizations of Swift's lyrics, her songs are known for taking on heavy themes, particularly heartbreak, love, and sadness. Sadness within a song can have many levels of nuance, from pure sadness to one of aching or longing: a bittersweet yearning for what has or could have been. Swift uses nostalgia in her work (McGrath, 2023; Morris, 2024) and references it as an inspiring creative force (Rahmani & Nasution, 2019). This juxtaposition of emotions evokes the Portuguese concept of saudade, a sentiment both full of longing but containing joy, missingness, happiness, nostalgia, and love. Saudade is "characterized by its ambivalence; that is, it comprises both positive and negative connotations" (Neto & Mullet, 2014, p. 661). We formalize this into a concept we term *Swiftian Saudade* and operationalize using three contrasting elements: joy, sadness, and nostalgia. We code saudade as present when there are two of three contrasting elements, such as sadness and anticipation or happiness and anticipation as top emotions within a song. We use lexicon-based sentiment primarily for these measures as they provide greater depth of emotions than other lexicon-based approaches that might use only positive/negative or happy/sad.

Swift's Songwriting and Album Design

In her 2022 speech accepting the Songwriter–Artist of the Decade Honor at the Nashville Songwriter Awards, she mentions three "types" of her songs—those metaphorically written with a glitter pen, a fountain pen, and a quill (Nicholson, 2022). Each of these archetypes features a different combination of traits: "glitter pen" songs are

happy and upbeat, "fountain pen" songs are more modern narratives, while "quill pen" songs are similar to "fountain pen" songs but more rooted in the past. These distinctions offer insight into Swift's songwriting process and provide inspiration for how we anticipate her music to be structured along emotional and narrative lines.

In addition to the songwriting element of Swift's work, there is a strategic component to the design of albums, that is, where songs are arranged within an album. Fans recognized a pattern of track five being more vulnerable or emotionally honest. These tracks are known for touching on heartbreak, relationships, and longing. This trend of song placement became something Swift allegedly leaned into: tracks in the fifth spot on her albums are considered honest and biting, digging into the nuanced realities of life. For example, fans frequently discuss "Track five" songs on social media, ranking and debating their favorites (e.g., Browne, 2024). On her most recent album, *The Tortured Poets Department* (*TTPD*), the track list was leaked prior to release and friends eagerly awaited hearing Track five, "So Long, London" (Schultz, 2024). We expect these track fives to be the most saudade-heavy within Swift's catalog as they are places where we expect Swift to be most emotionally real. Here, realness comes with themes, the language (personal) she is using, and the emotions invoked in the writing. In particular, we expect that she will be leaning into the nostalgia she mentions in her songwriting.

Research Hypotheses

We bring together our definition of *Swiftian Saudade* with prior research on emotions, lyrics, text analysis, and songwriting to develop specific hypotheses around Swift's songwriting and album structure.

We anticipate finding clustering or "grouping" of songs that align with Swift's "pen"-based allusion of her songwriting process and hypothesize that key dimensions for Swift's process will be linguistic (in line with the narrative-heavy fountain and quill pens) and emotional (such as the glitter pens' happiness and the raw emotion present within her narratives).

Additionally, we expect to observe *Swiftian Saudade* (the presence of conflicting positive and negative emotions) within Swift's work, particularly within songs placed in the infamous "track five" location on albums.

Data

The data set contains album and song data originally pulled from Kaggle (Johari, 2023), then augmented with album data collected from Spotify using their application programming interface. The song data are subsequently linked with lyrics publicly available through Genius.

The final data set contains 386 songs, with 395 total tracks, including prologues, poems, and voice memos across 21 albums (and their respective variations). It covers a total of 153,390 words (tokenized). The songs vary from approximately 164 ("It's Nice to Have a Friend") to 1,009 words ("All Too Well [10 min version, TV]"; tokenized counts).

Note that we group the albums by the albums themselves and not only by release date. For example, *Fearless* was originally released in 2008. Swift then rerecorded and released *Fearless* (TV) after a dispute over the original masters. The rerecord is similar to the original in that the first 13 songs are the same as those in the original. The additional 13 songs (for a total of 26 songs) are "vault tracks" that were not included on the original recording. This general structure is true for all TV rerecords: the original is a subset of the later recording. Given that the songs are from that era, it makes sense to group by album rather than by release date. This choice makes the most sense for Taylor Swift's work but may not be applicable for other artists: In the case of TV, her rerecord was partially a reclaiming of her work and a release of work that she felt was authentically hers. For later work (e.g., the three different versions of *Midnights*), this pattern is likely weaker. We wanted to make a consistent choice across all Swift's catalog, and for her, the expansive version is most representative of her work that she put out. The bulk of our analyses, unless mentioned otherwise, include one representation of each album: we choose TV recordings over Deluxe albums over originals, as applicable. The list of albums used for most analyses is *Taylor Swift*, *Fearless* (TV), *Speak Now* (TV), *Red* (TV), *1989* (TV), *Reputation*, *Lover*, *folklore* (DE), *evermore* (DE), *Midnights* (TD), and *TTPD* (Anthology). As noted in Table 1, we use the following abbreviations for album editions: TV: "Taylor's Version," DE: "Deluxe Edition," 3E: "3am Edition," TD: "Till Dawn Edition." We also use the abbreviation TTPD for the album *The Tortured Poets Department*.

Throughout our analysis, we refer both to the albums (Swift refers to each album as its own "Era"; hence the "Eras Tour") and to album tracks (e.g., "track fives") because these are very important in both how Taylor Swift talks about her music and how her fans to connect to it. We provide three tables as the online supplemental materials to help readers. Each table contains the same three elements, albums, track numbers, and track titles but with each element prioritized. For example, Table SI1 in the online supplemental materials provides a list by album of all track numbers and titles, Table SI2 in the online supplemental materials provides a list by title of album sorted by song titles, and Table SI3 in the online supplemental materials provides a list sorted by track number of titles and albums. This can help readers identify relevant tracks.

Preprocessing

We conduct our analysis using the NRC Emotion Lexicon by first tokenizing the words in the song—separating out words as separate units of analysis. We performed preprocessing for these analyses and cross-checked tokenization across two packages within R. We compared additional functions within and across these packages, specifically whether to use "get sentences" on the lyrics or using "get nrc sentiment" directly on the words (both are presented as options within the package documentation). We found that for our corpus, grouping by album and title, deriving sentiment directly from tokenized and lemmatized words, then aggregating at the song level, worked best (Packages tidytext; Queiroz et al., 2024 and syuzhet; Jockery, 2023). For tokenizing, we found that syuzhet performed better with contractions (e.g., "it's" to "it is") and possessives (e.g., "Jehovah's"). We additionally performed additional preprocessing: we expanded informal contractions (e.g., "feelin" to "feeling," "don" [from don't] to "do not," and "wanna" to "want to") and reincorporated words which, when transformed to lemma form using the textstem package (Rinker, 2018), were not in the NRC dictionary (e.g., "disturbed" vs. "disturb"). This allowed us to return 643 ("ing" words), 2 ("don" and "wanna"), and 85 words (lost in lemmatization), respectively. Lemmatization also affected approximately

Table 1
Album Information on All Tracks in Taylor Swift's Album Catalog

Album	Release date	Track	Numbered tracks	Total words	Words per track
Taylor Swift	October 24, 2006	16	15	4,969	311
Fearless	November 11, 2008	14	13	5,176	370
Fearless (TV)	April 9, 2021	26	26	9,544	367
Speak Now	October 25, 2010	14	14	6,057	433
Speak Now (TV)	July 7, 2023	22	22	9,380	426
Red	October 22, 2012	15	16	5,555	370
Red (TV)	November 12, 2021	30	30	11,917	397
1989	October 27, 2014	14	13	6,604	472
1989 (DE)	October 27, 2014	19	19	8,788	463
1989 (TV)	October 26, 2023	21	21	9,478	451
Reputation	November 10, 2017	18	15	8,863	492
Lover	August 23, 2019	18	18	7,343	408
folklore	July 24, 2020	16	16	5,122	320
folklore (DE)	August 18, 2020	17	17	5,364	316
evermore	December 11, 2020	15	15	5,535	369
evermore (DE)	January 7, 2021	17	17	6,349	373
Midnights	October 21, 2022	13	13	4,591	353
Midnights (3E)	October 22, 2022	20	20	7,032	352
Midnights (TD)	May 26, 2023	23	23	8,358	363
TTPD	April 19, 2024	16	16	6,150	384
TTPD (Anthology)	April 19, 2024	31	31	11,215	362

Note. The information presented involves release dates, total number of numbered tracks, total number of album words after the tokenization process, as well as the average number of words per track for each album, including memos and other textual items. TV = Taylor's Version; DE = Deluxe Edition; 3E = 3am Edition; TD = Till Dawn Edition; TTPD = The Tortured Poets Department.

991 words, allowing them to be added when the original base was not included (e.g., castles were not in the NRC lexicon dictionary but castle was). Consult Supplemental Information Table SI4 in the online supplemental materials for the full list of transformed words. We then obtained a count of the number of NRC's eight specific emotions on the basis of songs.

We also removed most stop words from our NRC analyses using "stop words" from tidytext. After cross-comparison with NRC, we elected to add back in several from the stop words list that were frequently used in Swift's writing: good, important, kind, parting, problem, and wanting. We did allow the removal of the following stop words, despite having some emotional evaluation in NRC as the meaning could be ambiguous: case, differently, fact, facts, fully, general, highest, interested, long, older, present, show, thought, and young. We elected to not preprocess the linguistic inquiry and word count dictionary (LIWC) data to replicate similar approaches to lyrical studies and existing literature (Boyd et al., 2022; Tausczik & Pennebaker, 2010).

We include relevant variables from the NRC and LIWC lexicon-based sentiment classifiers, described in each of these sections, and information on cowriting sourced from Kaggle (Earhart, 2024) and cross-referenced in Wikipedia and Swiftpedia (Taylor Swift, 2009). Swift has a mixture of collaboration in her catalog: she is a solo writer on 65 of her 234 songs in our set of expanded albums and tracks. From her 169 songs with collaboration, she has collaborated with Jack Antonoff and Aaron Dessner more than any two other individuals at 56 and 41 times, respectively. The next collaborator would be Max Martin at 21. Their collaboration was not as long or productive and so we retain the top two collaborators, Antonoff and Dessner. We use this information to create three dummy variables: one for if Swift is the solo writer, one if Jack Antonoff is listed as a cowriter, and one if Aaron Dessner is listed as a cowriter.

Method

Following insights from Czarnek and Stillwell (2022), we pair two lexicon-based approaches, NRC and LIWC, to ensure a more balanced analysis. We combine insights from these into our PCA to show the landscape of emotional and linguistic dimensions. As far as we are aware, our work is the first to incorporate PCA in this context of lyrical analysis. Our approach can provide helpful text-based insights, exploring more than positive/negative analysis of sentiments and allowing for comparisons across artists, genres, and time.

Lexicon: NRC and LIWC

Lexicon-based approaches contain lists of words and their resulting sentiment or polarity: some systems (such as AFINN Lexicon; Nielsen, 2011) focus on whether a word is positive or negative while others (e.g., NRC) identify particular emotions from the text. Lexicon approaches can be beneficial because they are simple and straightforward. In the context of lyrics, they may be particularly apt as songs are quite short, and songwriters are choosing words within the limited space of a 3–4-min song (generally speaking, although there are certainly extended editions). Additionally, lexicon-based approaches are preferable when "simulating the effect of linguistic context" (Taboada et al., 2011). Generally speaking, lyrics are a prime candidate for this (North et al., 2021), and Swift's lyrics are often mentioned for how she writes in an engaging and accessible way.

NRC

NRC provides a dictionary of words and their associated emotions. The mapping is to eight emotions (anger, disgust, fear, sadness, trust, anticipation, surprise, and joy), crowdsourced, covering well more than 10,000 terms (S. M. Mohammad & Turney, 2013). In our analyses, we

focus primarily on these eight emotions (anger, disgust, fear, sadness, trust, anticipation, surprise), not including the general categories of "positive" and "negative," as we are looking for specific emotional content. We discuss preprocessing (specifically, the removal of stop words, tokenization, and lemmatization) within the Data section.

LIWC

LIWC refers to itself as the gold standard in text analysis software. It is used commonly within psychological research, covering 100 dimensions of text. It contains over 1,000 dictionaries and reports back percentages of words matching the relevant terms (e.g., positive tone). There are also summary measures for different aspects of writing, including authenticity, emotional tone, analytical thinking, clout, sadness, use of pronouns temporal focus, cognitive processes, and social references (Boyd et al., 2022).

Together, these elements can provide an image of the work and how the author communicates. We incorporate LIWC measures on relevant composite measures from the literature, those based on specific (or general) emotional states and linguistically focused elements relevant to Swift's work and reputation as a songwriter.

Table 2 contains summaries of emotion-related terms within LIWC. We sought to include relevant emotional elements to capture this aspect of Swift's work. There are differences between NRC and LIWC in included emotions (e.g., surprise, joy, and disgust are not emotional categories within LIWC). Affect is the feeling of an emotion (VandenBos, 2007), while emotion is overall use of emotions. We also include three emotion-specific measures connected to particular emotions: anger, anxiety, and sadness (LIWC does not have specific measures of joy or other positive emotions). Swear reflects the degree of explicit words present, and is something fans and critics have noted as being on the rise in Swift's work (especially within *TTPD*). Finally, tone is the overall tone (positive numbers are higher), akin to reflecting sentiment overall, although it also connects to emotional words (Boyd et al., 2022).

In addition to LIWC's emotion-related measures, we incorporate linguistic measures. We incorporate a composite measure of various linguistic dimensions, linguistic, in addition to something Swift is known for: the use of personal pronouns. We also selected measures that reflect descriptions of Taylor Swift as an artist, such her temporal focus (present, future, and past). This is particularly interesting as time was an important feature within Swift's songwriting (Sloan, 2021). Similarly, we incorporate the composite measure of authenticity, perceived genuineness and honesty, and analytic, logical thinking, as these are other traits recognized in Swift's songwriting and are often included in analyses of other songwriters (Eaton et al., 2022; Kalichman & Smyth, 2023, 2024). A summary of all these linguistic measures is presented in Table 3.

No one approach is without issue, and each approach has disadvantages (e.g., some neutral-seeming words may have emotional valence in the NRC; Zad et al., 2021), but NRC still performs relatively well (Tabak & Evrim, 2016). We use NRC in concert with LIWC to provide more balance to our analyses and expand to incorporate additional linguistic measures.

PCA

PCA offers the means to reduce data dimensionality using a nonparametric approach. PCA identifies variable groupings, clustering them along orthogonal dimensions that have the most variance

Table 2
Average LIWC Measures for Different Emotional Elements for Each Album in Taylor Swift's Catalog

Album	Affect	Emotion	Anger	Anxiety	Sadness	Swear	Tone
Taylor Swift	6.00	2.50	0.11	0.02	0.33	0.11	54.41
Fearless	5.06	1.88	0.06	0.11	0.19	0.00	44.17
Fearless (TV)	4.97	2.36	0.11	0.11	0.21	0.00	48.66
Speak Now	5.56	1.73	0.12	0.13	0.23	0.00	43.76
Speak Now (TV)	5.02	2.04	0.13	0.16	0.16	0.00	46.51
Red	7.12	2.94	0.36	0.09	0.53	0.02	51.10
Red (TV)	6.07	2.45	0.21	0.11	0.38	0.03	50.78
1989	7.04	2.81	0.36	0.02	0.13	0.07	40.40
1989 (DE)	6.73	2.76	0.33	0.04	0.16	0.05	41.60
1989 (TV)	6.68	2.59	0.28	0.05	0.14	0.11	39.42
Reputation	4.99	2.18	0.28	0.12	0.09	0.06	38.97
Lover	6.45	2.55	0.37	0.20	0.19	0.06	42.71
folklore	5.18	1.82	0.44	0.07	0.20	0.23	39.62
folklore (DE)	5.15	1.89	0.41	0.07	0.32	0.22	38.19
evermore	5.09	1.97	0.10	0.03	0.13	0.39	38.96
evermore (DE)	4.90	1.82	0.09	0.03	0.14	0.35	37.34
Midnights	6.26	2.23	0.29	0.17	0.20	0.70	47.26
Midnights (3E)	6.26	2.03	0.19	0.14	0.27	0.48	46.43
Midnights (TD)	6.65	2.07	0.20	0.16	0.29	0.60	46.38
TTPD	6.82	2.54	0.11	0.20	0.48	0.65	27.08
TTPD (Anthology)	6.04	2.13	0.20	0.17	0.37	0.46	26.80
Original	5.95	2.29	0.24	0.11	0.24	0.20	42.42
Taylor's Version	5.68	2.37	0.18	0.11	0.23	0.03	46.87
Grand mean	5.90	2.26	0.22	0.11	0.25	0.22	42.30

Note. LIWC = linguistic inquiry and word count dictionary; TV = Taylor's Version; DE = Deluxe Edition; 3E = 3am Edition; TD = Till Dawn Edition; TTPD = The Tortured Poets Department.

Table 3
Average LIWC Measures for Different Linguistic Elements for Each Album in Taylor Swift's Catalog

Album	Linguistic	Personal pronouns	Focus present	Focus future	Focus past	Authentic	Analytic
Taylor Swift	82.44	19.37	7.97	2.65	4.00	60.98	7.33
Fearless	81.45	19.39	9.32	2.49	4.82	92.63	4.06
Fearless (TV)	80.67	19.54	8.33	2.03	5.95	91.72	8.54
Speak Now	80.62	18.46	8.04	2.16	5.24	89.44	13.38
Speak Now (TV)	78.90	18.57	7.17	2.50	5.13	84.92	12.94
Red	80.65	18.99	6.79	1.83	5.71	79.08	12.14
Red (TV)	79.85	18.43	6.62	2.05	5.26	83.33	10.87
1989	78.47	17.46	6.38	2.29	6.00	81.65	14.17
1989 (DE)	78.09	16.97	6.41	2.12	5.60	75.85	19.15
1989 (TV)	77.37	17.35	6.66	2.52	5.59	81.07	17.02
Reputation	77.79	18.91	6.23	1.58	4.38	64.75	15.34
Lover	76.65	19.99	6.69	2.58	3.49	79.68	13.47
folklore	77.62	18.89	4.60	1.27	6.99	83.96	20.89
folklore (DE)	77.49	18.64	4.55	1.19	6.73	84.83	21.95
evermore	77.08	17.95	6.27	1.21	5.10	93.57	17.90
evermore (DE)	77.06	17.82	6.18	1.26	4.96	94.09	18.52
Midnights	73.29	16.52	6.55	1.09	5.79	74.27	26.96
Midnights (3E)	74.10	16.24	5.84	1.46	7.30	76.13	22.79
Midnights (TD)	73.44	15.85	6.15	1.28	6.73	76.49	25.87
TTPD	75.03	20.33	6.65	1.83	5.42	70.83	15.43
TTPD (Anthology)	74.91	18.78	5.79	2.04	6.06	73.89	19.99
Original	78.27	18.82	6.83	1.92	5.13	78.70	14.56
Taylor's Version	79.33	18.52	7.20	2.25	5.48	85.41	12.02
Grand Mean	77.68	18.30	6.60	1.90	5.58	80.53	16.13

Note. LIWC = linguistic inquiry and word count dictionary; TV = Taylor's Version; DE = Deluxe Edition; 3E = 3am Edition; TD = Till Dawn Edition; TTPD = The Tortured Poets Department.

(Abdi & Williams, 2010; Shlens, 2014). Each principal component is a representation or summary of the original contributing variables (Greenacre et al., 2022). Using PCA can not only reduce the number of variables in play but also provide an opportunity for connections and relationships to become clearer (Abdi & Williams, 2010).

The dimensions, or components, in PCA have loadings, associations with relevant variables, that can be used in interpreting the dimensions (Abdi & Williams, 2010; Greenacre et al., 2022). Researchers frequently name the dimensions, although those names are not defined by the model and usually arise through interpretation and observation (Greenacre et al., 2022). The practice of using PCA has precedence within music (e.g., Fricke et al., 2018; Park et al., 2019), although these individuals look at features of music rather than lyrics.

For our PCA, we use variables on authorship, emotion, text, and linguistic construction. We rescale variables to have a mean of 0 and a standard deviation of 1 to mitigate any undue influence from units of measurement. For our analysis, as we noted before, we select only one album per album grouping (e.g., *Fearless* [TV] but not *Fearless*; *Midnights* [Till Dawn] but not *Midnights* nor *Midnights* [3 am]). This ensures songs do not have undue influence in the analysis through being counted multiple times. We also incorporate all tracks for the analysis, including any liner notes or prologues. These additional elements are part of the work released to fans. We exclude them when looking at track-specific analyses if they are not assigned a track number on the album.

Swiftian Saudade

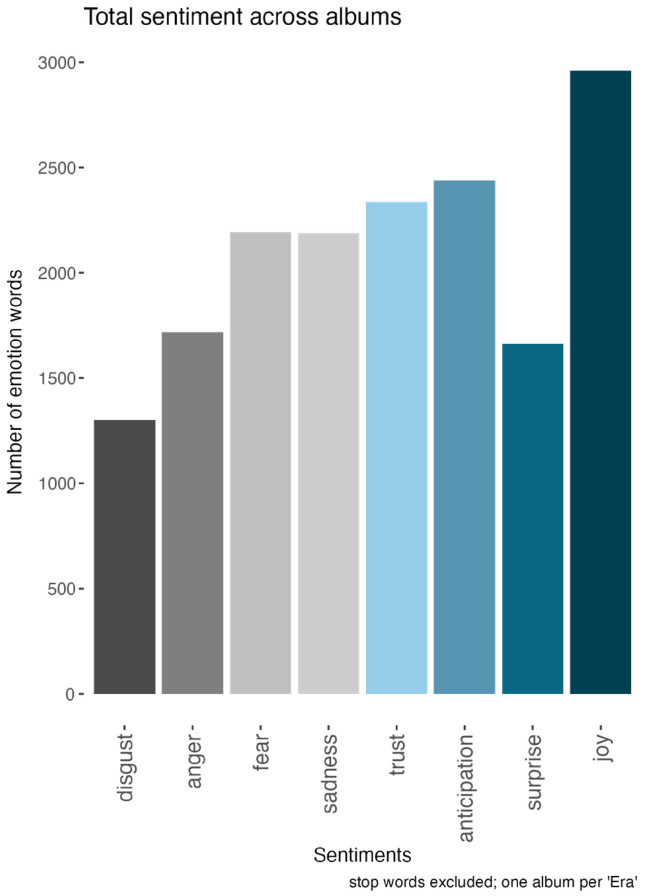
Figure 1 provides a basic summary of the eight emotions, organized from negative to positive (loosely): disgust, anger, fear, sadness, trust, anticipation, surprise, and joy. Surprise was a challenging emotion to place: it can be considered positive or negative, and Plutchik has it opposite to anticipation (Plutchik, 1982). However, in NRC analyses (S. Mohammad & Turney, 2010), surprise was rated by observers as more positive. As such, we kept it to a positive location. It is not an element in our operationalization of *Swiftian Saudade*.

We show first how Swift used these elements within her work in Figure 2. Here, we see a movement over time from more positive emotions (blues [bottom portions of bars]) to negative (grays [top portions of bars]). To develop our measure of Saudade, we look specifically at contrasting emotions that evoke nostalgia. In particular, we focus on the combination of joy with sadness or sadness with anticipation based on their location in Plutchik's wheel of emotion (Plutchik, 1982). As there are only eight emotions within the NRC, we wanted to develop a definition of *Swiftian Saudade* that captured the essence of the term without being overly expansive relative to the data. For this reason, we did not include surprise, despite its potential as a counterpart to anticipation, nor anger as anger is too far from the general concept of *Swiftian Saudade* we hoped to capture. It is this contrasting nature that we wanted to emphasize in our measure. As such, the top three emotions, including ties, present per song were used in the determination of saudade. Songs are coded as having "Swiftian Saudade" if they have either joy and sadness or anticipation and sadness within the top three emotions. We present the results on a by-album basis in Figure 3. We have similar results including ties or only including the top three elements, but including ties was better able to capture the definition of saudade.

Overall, within Swift's catalog, 21% of songs on the original release exhibit *Swiftian Saudade*, while 24% of her songs in the most expansive album version exhibit *Swiftian Saudade*. On average, across all songs (original, expanded, and most expansive), she

Figure 1

NRC Emotion Lexicon Sentiment Measures Across 11 Albums



Note. Only one recording is chosen per album (TV, DE., or Original), based on which is the most expansive version. TV = Taylor's Version; DE = Deluxe Edition. See the online article for the color version of this figure.

averages around 23% of her songs as having *Swiftian Saudade* present across her entire catalog.

Results

We discuss our findings from our three hypotheses regarding *Swiftian Saudade*, the composition of albums, and Swift's songwriting, finding support for all three. Within Swift's catalog, we observed that over 24% of her songs exhibit *Swiftian Saudade* with her wholly solo-authored *Speak Now* (original release), a notable exception.

Track-Based Analyses: Track Five

Fans have long regarded Swift's albums as having a secret agenda—primarily that track five is the more "personal" track (e.g., "Dear John," allegedly about her breakup with John Mayer and "So long, London" allegedly about her relationship with Joe Alwyn.). These tracks are thought to address themes of sadness and longing, precisely what we understand as *Swiftian Saudade*.

In Figure 4, we summarize Swift's work on the basis of track number—again, focusing upon one representative album for each track. For readability, we include the first 15 tracks of albums (all expanded Swift albums have at least 15 tracks so we use that for analyses to maintain consistency for track number inquires). We observe that track fives do have the highest proportion of saudade elements at nearly 50% of tracks. We also observe the presence of *Swiftian Saudade* in most track numbers aside from track ones—albeit not at the high levels of track five. A significance test of these values produces a p value of .102; thus, we cannot reject a null hypothesis that the proportion of songs with *Swiftian Saudade* within track fives (45%) equals that of *Swiftian Saudade* within any other track number (24%). This result is not statistically significant, which is not surprising given that there are currently only 11 track fives.

Another claim is that these tracks are more authentic and real, given their themes of breakup/heartbreak. Here, we consider the average authenticity of track five songs versus other track numbers. A paired difference of means reveals an average authenticity of 93.9 among track fives and an average authenticity of 80.2 among other tracks, which is statistically significant at $p < .001$.

Taken together, we plot authenticity against *Swiftian Saudade*. Figure 5 reveals a curvilinear relationship between *Swiftian Saudade* and authenticity: some songs are highly authentic but without *Swiftian Saudade* (e.g., Tracks 1, 4, and 7), while others are high in both *Swiftian Saudade* and authenticity (Tracks 5 and 9). Track five is particularly noteworthy in that it is the "emotionally real" track in the truest sense: it is highly authentic and contains a juxtaposition of emotions. There are some *Swiftian Saudade* tracks (10 and 12) that seem to be more about the emotions than the narrative or authenticity.

We have strong support that track fives are indeed special: they have significantly more authenticity and are more saudade laden (but not to a statistically significant level). There seems to be types of tracks in the emotions they contain (level of *Swiftian Saudade*) and how real or authentic they are in expressing them (authenticity). This sets up our PCA analysis to evaluate if there are trends on the song and album level.

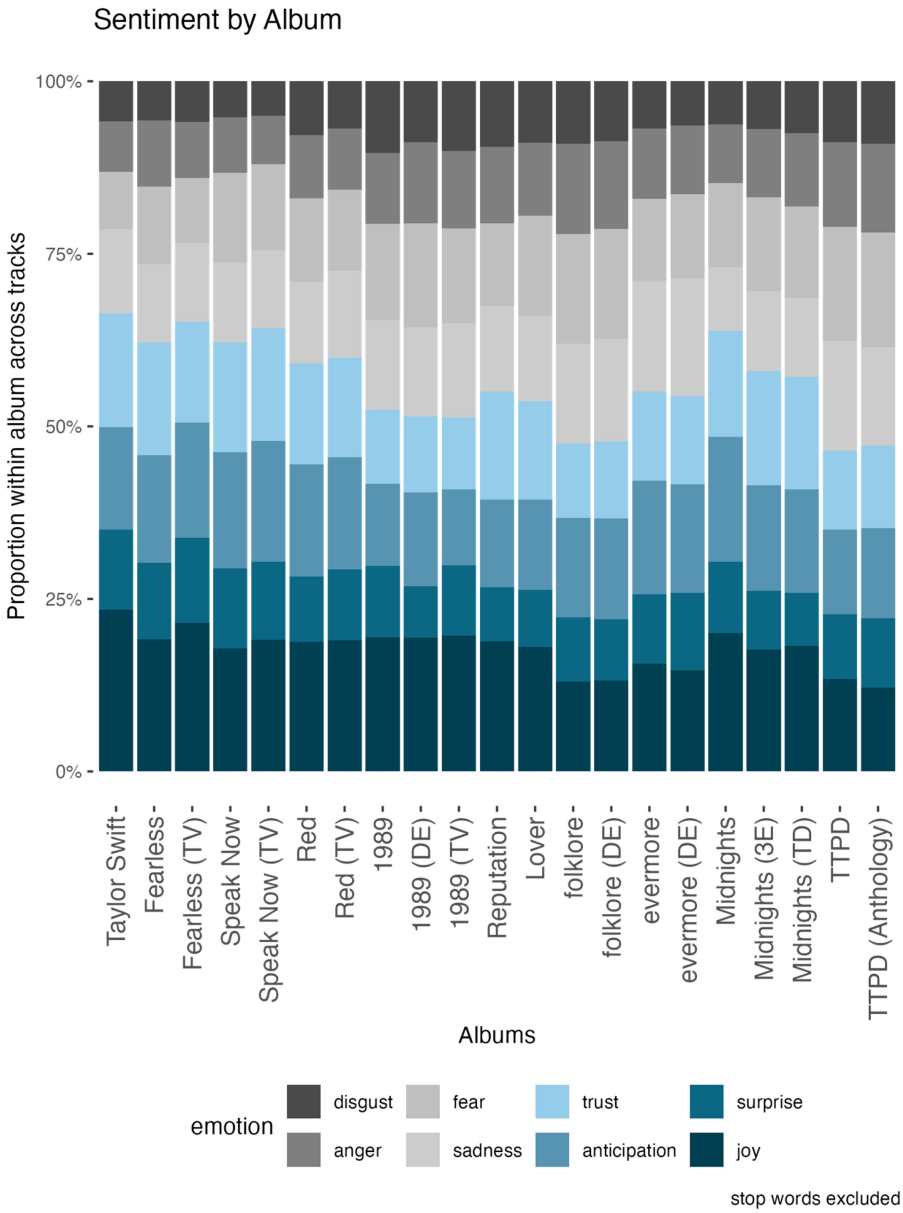
PCA: Linguistic and Emotional Exploration

We also hypothesized that Swift's work would have two clear dimensions: one of linguistic components and one related to emotion. Within PCA, there are two key steps: visualizing the number of dimensions and their relative loadings and plotting the individual observations within that space. For the analysis, we only incorporate the data of 11 total albums. Only data points (songs) in each of the representative albums (TV and DE versions whenever possible) for the 11 total albums are included in the analysis.

We combine variables related to emotion and linguistic construction into our PCA. We use NRC variables on emotion (fear, anger, sadness, trust, joy, and anticipation) and those from LIWC on time (past, future, and present), linguistic construction (linguistic, personal pronouns, authenticity, analytic) and emotional content (affect, emotion, tone), variables related to writing credits on the songs (solo, antonoff, dessner), and a dummy variable for whether *Swiftian Saudade* is present in the song. As all variables are rescaled to have a mean of 0 and a standard deviation of 1, we do not include their means or standard deviations. We do include their minimum, median, maximum, and number of observations. See Appendix Table A1 for more detail on these variables.

Figure 2

Emotional Content of Each Album Over Time (Albums Are Presented in a Chronological Order)



Note. TV = Taylor's Version; DE = Deluxe Edition; 3E = 3am Edition; TD = Till Dawn Edition; TTPD = The Tortured Poets Department. See the online article for the color version of this figure.

Dimensional Attributes of Swift's Work

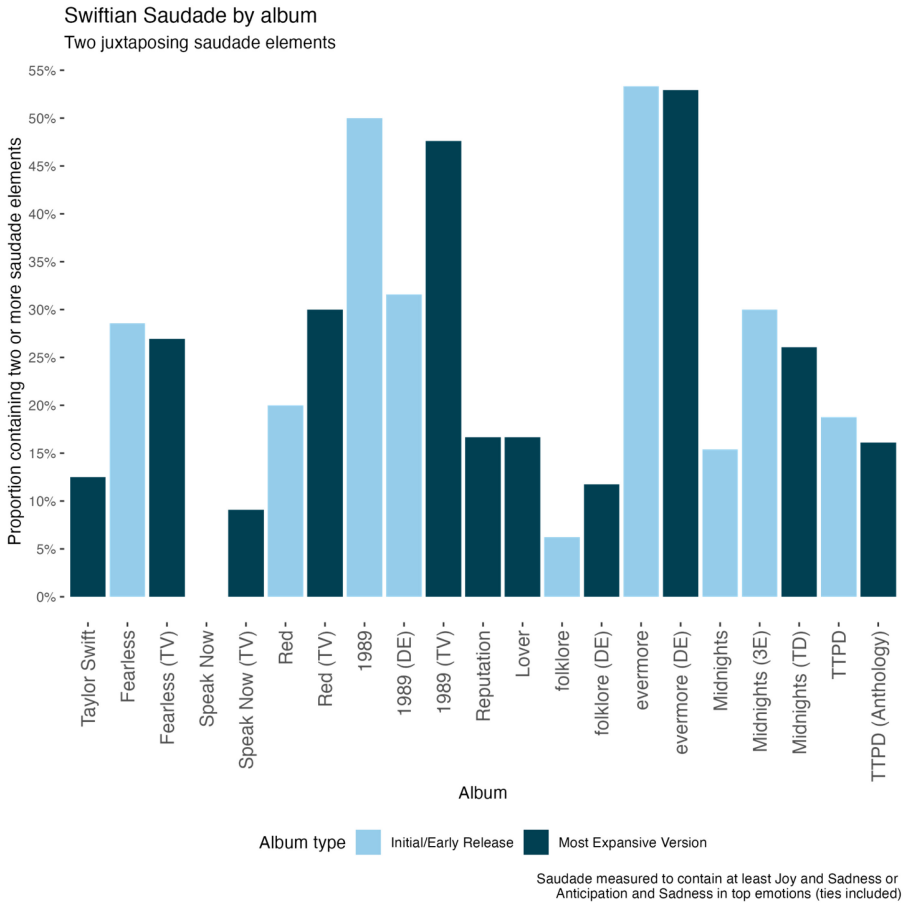
One of our prior hypotheses pertained to the dimensions and an expectation that linguistic or narrative-based elements would be important, as would emotional elements. PCA helps to reduce variation when there are many correlated variables. Within PCA, it is not the specific location (e.g., which quadrant) as much as the relative location that matters: variables that appear within the same dimension can be considered connected to one another. Thus, our expectation is that we will see emotion-related variables together and linguistic-related variables together. The dimensions of the model can explain overall variation. For ease of interpretability, we include two dimensions. In our overall results, the proportion of variation explained by the first two dimensions is approximately 30%.

Figure 6 provides the variables and their respective locations using the first two components. The terms are shaded by the amount they contribute to the dimension: darker values are stronger contributors, while lighter shades contribute less.

As we anticipated, we see a clustering of variable types. The lower left quadrant features more linguistically oriented variables, the

Figure 3

Presence of Saudade in Each Album



Note. For each album, the most expansive version is also available and shaded in a darker shade while prior versions are lighter shade. For example, *Midnights* and *Midnights 3am* are both in lighter shade, while *Midnights (Till Dawn)* is the most expansive album and is a darker shade. TV = Taylor's Version; DE = Deluxe Edition; 3E = 3am Edition; TD = Till Dawn Edition; TTPD = The Tortured Poets Department. See the online article for the color version of this figure.

lower right quadrant is primarily negative emotions, and the top center of the plot is more positive emotions. Thus, the three primary elements that capture the variation in Swift's work are the elements of positive emotions, negative emotions, and linguistic construction.

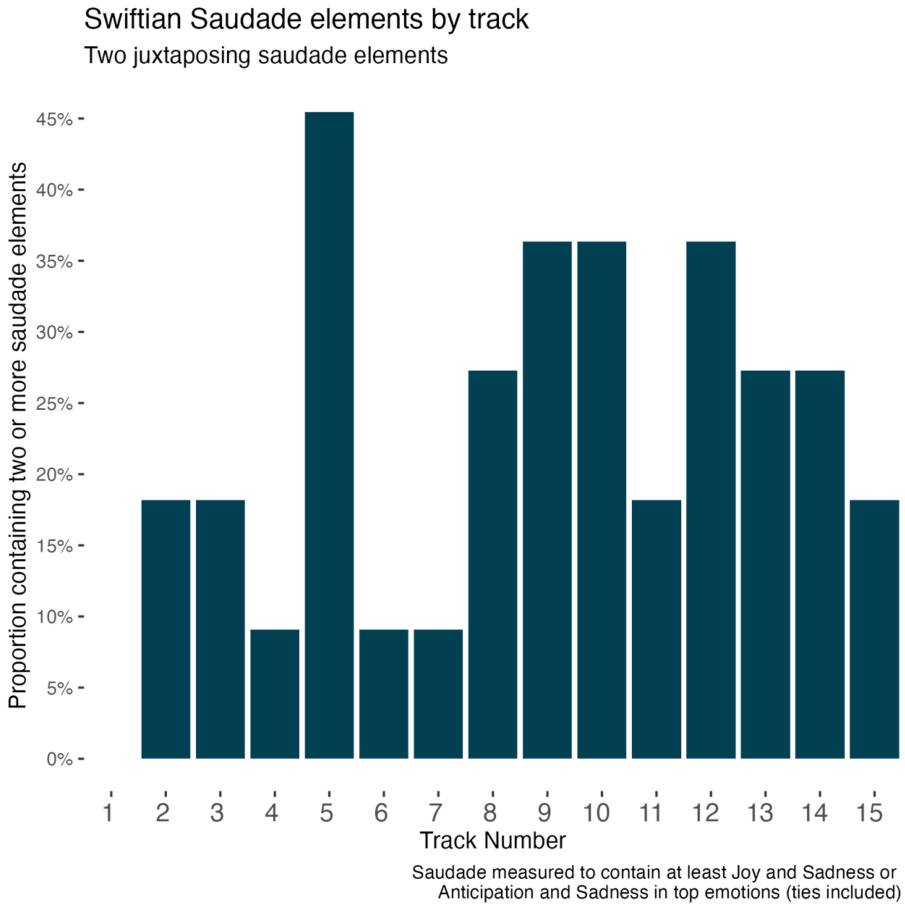
We include three dummy variables regarding the authorship of a song: one for Swift (solo) and one for each of her most frequent collaborators, Jack Antonoff (antonoff) and Aaron Dessner (dessner). The vectors for these three writers are in very different directions, suggesting that Swift seeks and incorporates her cowriters when wanting to write a different type of song.

Additionally, Figure 6 shows the values and original dimension names. To facilitate interpretation, we provide descriptive titles to the axes. For Dimension 1 (Dim 1), in the negative quadrant, temporal focus (past, present, and future), authenticity, and personal pronouns are present. These terms are in the domain of fountain pens and linguistic construction. From this, we label the negative values of Dim 1 as more literary or narrative-driven in focus. On the positive portion (moving to the right), the variables are related to emotion, with negative emotions on the negative side of Dim 2 and positive emotions on the positive side of Dim 1. For this reason, we label this portion of Dim 1 as more emotional. For Dim 2, the negative side is more negative of emotion (although linguistic elements are still present), moving to more positive emotions and tone. For this reason, we label Dim 2 as moving from negative to positive emotion.

From these initial dimensions, we incorporate the data points themselves. The resulting plot is a bit complex but captures the key findings from our earlier analyses. The left to right (x axis) can be thought of as one moving from a more literary focus to a more emotional, while the y axis moves from negative to positive emotion as we move to larger positive values. The gray points represent each song's location within this two-dimensional space. We then provide two sets of averages: averages for each album and averages for each track. The numeric points represent the average for a given track number (e.g., "6" represents the average location of all track six songs). Similarly, an album title represents the average of the album. These three elements (points, track averages, and album averages) allow us to see where these each exist in the space. Figure 7 shows a plot zoomed into the center for greater

Figure 4

Presence of Saudade in Each Track Number Across Albums



Note. Songs at track number 5 present the highest proportion of saudade. See the online article for the color version of this figure.

visibility and ease of readability. We then expand to include averages for the first 15 tracks in Figure 8. We include analysis on the full set of songs in Figure A1 in the Appendix.

We have applied labels to general areas of the plot to guide interpretation: starting from the far left, along the negative *x* axis is "Long live for the hope of it all": these songs are positive and express desire in some way (wanting something or someone, perhaps speaking to dreams, goals, or ambitions). From there, there are more purely joyful songs in what we term "glitter pen" in homage to the style Swift cites. Glitter pen is where Track 11 averages lie, which makes sense given prior analysis and findings. Next, we move toward negative emotions, into more intensely felt feelings, one we have termed "Sad, Beautiful, Tragic" after a song on *Red* that is heavily about feeling and being in one's emotions. As the negative feelings intensify, we move into "melancholy and anger." For this section, we add the title of "female rage: the musical," a title referenced by Taylor Swift in her first "Eras" show featuring the *TTPD* set. Finally, we move into a lower left quadrant we have called "fountain pen" in homage to one of Swift's famed song styles: these are more narrative-driven pieces, packed with personal pronouns and emphasizing time in some way. They are noteworthy in their construction and use of time and focus.

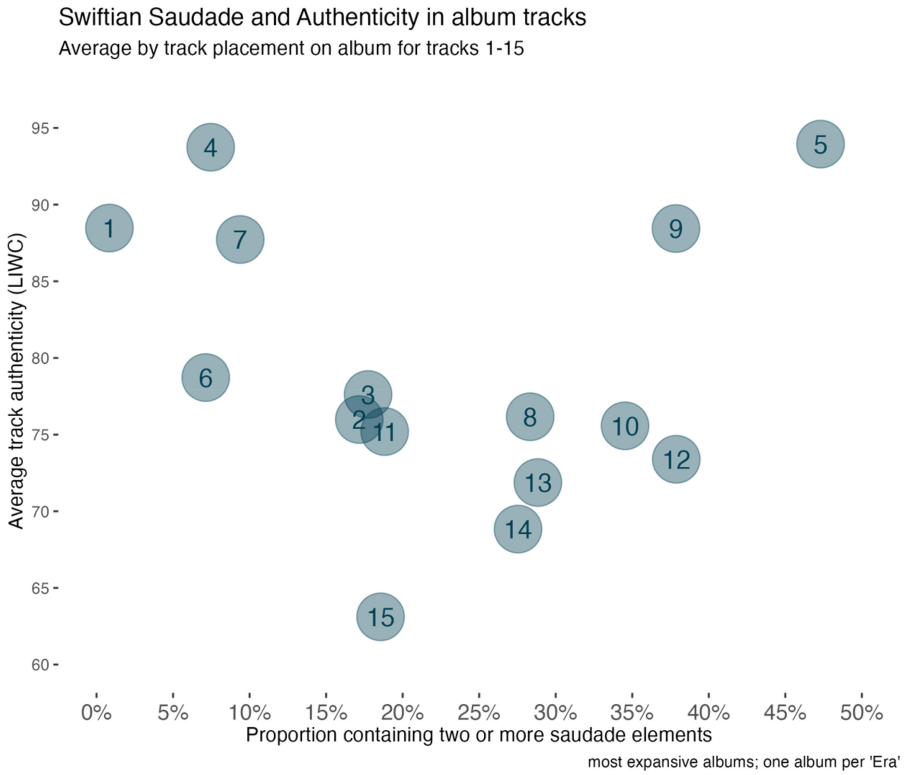
Trends in Swift's Work

Initially, we begin at the level of albums in Figure 7: we include the average by album in the space. Swift has repeatedly emphasized albums as units of analysis—from naming them as her "Eras" in her recent Eras Tour, to discussing them with fans as items that would have distinct personalities (Swift, 2019). As such, we focus on by-album differences. The PCA visualization makes Swift's songwriting evolution clear: Swift moves from the upper left quadrant, a side more linguistically focused, from an area we refer to as "Long live (for the hope of it all)" (this is a mashup of titles from Swift's *Speak Now* ("Long Live"), and a song "August" from *folklore*). She moves more or less along a line that becomes not only increasingly emotion focused but moves toward the negative emotion, "melancholy and anger (female rage: the musical)" portion of the plot.

This evolution in Swift's songwriting was noticeable before in Figure 3, but when focusing only on albums, the shift can be more apparent. To make this clearer for readers, we (a) color the albums from early to mid to current using lighter to darker saturated shades of blue and (b) include shaded groupings on the plot. In this plot, we illustrate the movement from Swift's earliest albums (Taylor Swift

Figure 5

Swiftian Saudade *and Authenticity for Each Track Number*



Note. Tracks 5 and 9 exhibit both high average authenticity as well as high proportion of saudade elements. LIWC = linguistic inquiry and word count dictionary. See the online article for the color version of this figure.

and Fearless) through *Speak Now* and *Red*, then *1989*, *Lover*, and Midnights to, finally, her most recent release, *TTPD*. We name these clusters based on when in her career they were released to more fully signal to the reader that there is some temporal aspect to the albums and their composition. After establishing general placement of albums, we look at more track-focused analyses.

We present this information in Figure 8. The figure presents both song values as points (circles for coauthored work and triangles with Swift as the solo author) and averages by track numbers (15 or lower) and albums.

Discussion

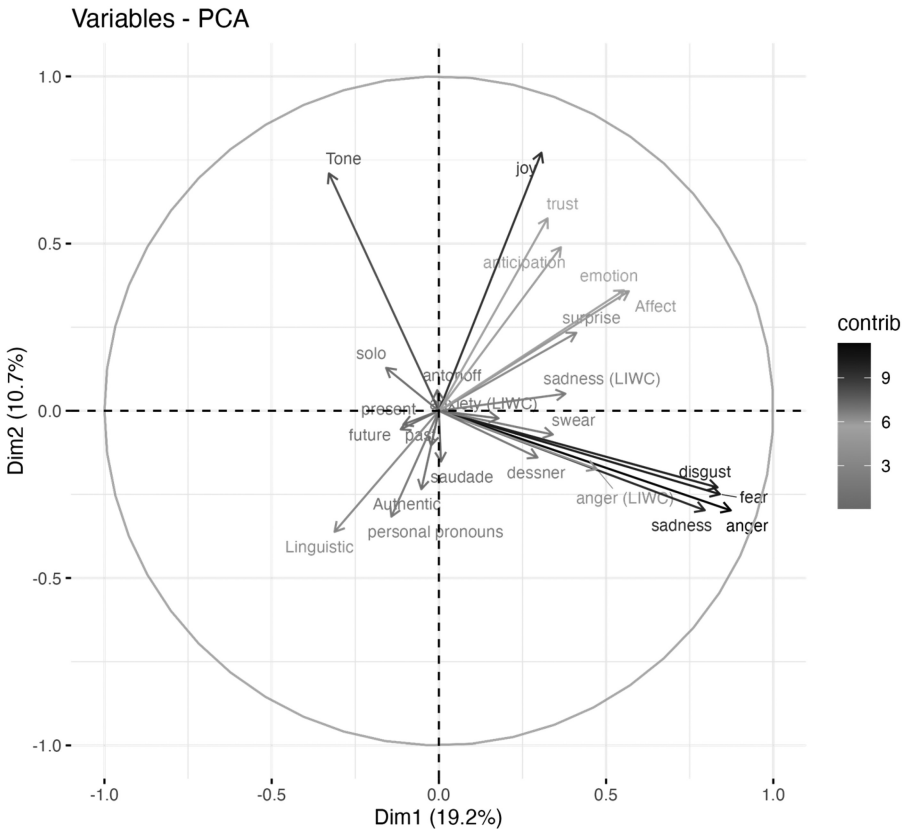
Our characterization of *Swiftian Saudade* raises questions about the incorporation of emotions within songwriting and music composition. For Swift, she does seem to anchor her songwriting in a space that is linguistic and charged by emotions (positive and negative). We want to underscore how interesting this is: from the perspective of PCA, there is no reason why any particular variable should be grouped with another unless they have relationships within the data. In the case of Swift, that these clusters make sense is in and of themselves informative and a great starting point for future research.

In addition to the dimensional attributes of Swift's work, we incorporated measures of authorship related to solo writing ("Solo"), collaboration with Jack Antonoff ("antonoff"), and Aaron Dessner ("dessner"). Swift's solo work is primarily earlier in her career (Swift famously wrote the album *Speak Now* entirely by herself). She also has solo-written songs throughout her catalog, including "Who's Afraid of Little Old Me" from her most recent release, *TTPD*. However, because Swift's early work was more heavily solo written, the "Solo" variable is in the upper left ("Long Live") quadrant. Curiously, Antonoff's vector is near the center, indicating that Swift may use him as a kind of counterbalance or amplification of her own "center." In contrast, Dessner's vector is in the melancholy/anger quadrant. There does not seem to be much difference regarding how Swift varies as a solo writer versus a cowriter regarding *Swiftian Saudade*. A significance test of her Saudade rate (.23 in solo work to .25 in cowritten) reveals a p value of approximately .78, far from any reasonable significance threshold. There is an evolution in Swift's work, perhaps because she is strategic in working to achieve certain songwriting goals through writing partnerships (see documentary *Miss Americana*; Wilson, 2020), in addition to voice memos and interviews; Apple Music, 2020 for more on Swift's songwriting process). In all her works, Swift's writing is active and participatory, and she is making important decisions regarding the song's construction and tone regardless of authorship.

Throughout Swift's work, *Swiftian Saudade* is consistently present: it appears in the lower quadrant, with other linguistic elements in Figure 6. This is particularly interesting in that it appears so "far" from other emotional terms in the plot and is instead aligned with

Figure 6

PCA: Variables in PCA Analysis and Their Relative Contributions to the Dimensional Components



Note. PCA = principal component analysis; Dim = dimension.

authenticity and personal pronouns. We believe this comes from how Swift pairs these contrasting emotions—not only is positive or negative emotion present but it also exists within a narrative space. This is similar to what we saw in Figure 5: *Swiftian Saudade* can stand alone as a way to express complex emotions or Swift can leverage in the process of a narrative, blending in authenticity. These variables, saudade and authenticity, are very near one another in the PCA.

Given this understanding, and how Swift employs Saudade, the location of track five is less surprising. Track five songs, on average, are on the negative emotion side but located in a more lyrically focused quadrant, indicating that the tracks are likely more narratively driven. This adds to our understanding of the role of track fives in Swift's work.

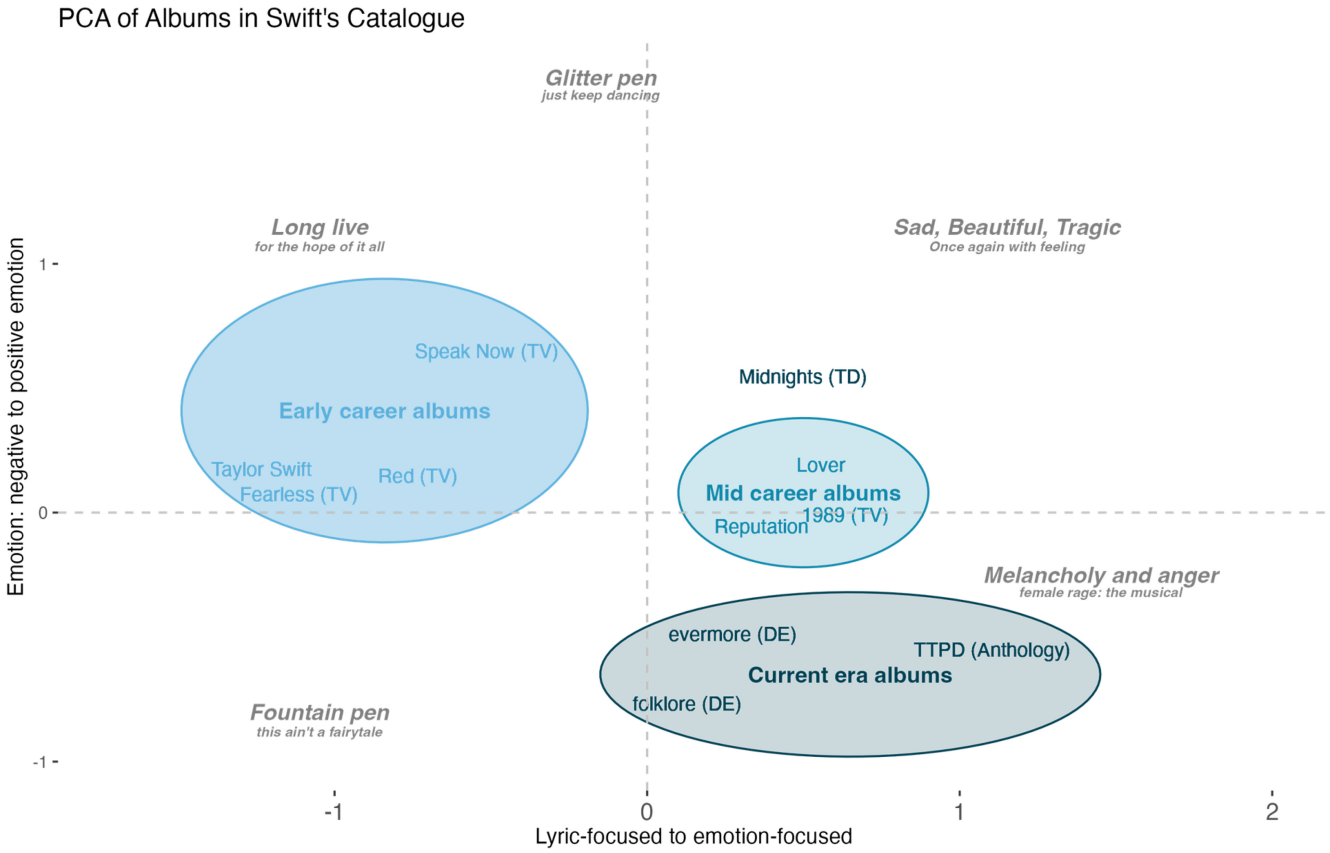
Although we did not formally analyze other track-focused features, leaving that to future work, we observe in Figure 8 that track elevens are particularly joyful in what we term the "glitter pen" position and Tracks 4 and 12 appear to also be particularly poignant, although to a lesser extent than track fives. This observation is particularly interesting given that the track five song on *TTPD*, "So long, London" is considered by fans a *Swiftian Saudade* song about a breakup with her "London Boy" ("London Boy" is a glittery Track 11 from her album *Lover*). We also observe that there are a cluster of tracks in the realm of heavy, dark emotion—nearly half of Swift's tracks have an average in this space. These tracks may be engaging and emotional, but they generally do not tell a story in the same way that a track five song does. Track five songs, in addition to being emotionally complex, usually tell a story of some kind, dealing with heartbreak and deep feelings. That track fives land near personal pronouns and are highly authentic fits their reputation.

In our analysis, we find that Swift's work has an interesting progression: her work has three distinct eras, excepting *Midnights*. She starts out yearning and ambitious in this first era, moves toward more emotionally engaged work in her mid-career era. The current era, and the pandemic, brought *folklore* and *evermore*, sibling albums that were heavy on storytelling and emotionally rich, with a negative emotional charge. They inhabit space near the middle of the plot. Swift's latest release, *TTPD*, has been recognized as dark, complex, and raw. It also inhabits the melancholy space, the area where intense negative emotions and linguistic construction are present.

Swift has alluded to "sister" albums in the past, and we wonder if Swift intentionally or unintentionally approached multiple albums that way. Swift's *folklore* is curious in that it was described as more narrative driven and also as a "sister album" to *evermore* and yet the two exhibit very different levels of *Swiftian Saudade* with *Swiftian Saudade* much more present on *evermore*. The emotions present within *folklore* are indeed a departure from Swift's usual

Figure 7

PCA Plot of Swift Albums Across the Lyric-Focused and Emotional Dimensions Extracted From Figure 6



Note. Albums are grouped based on three stages in Swift's career (early career, mid-career, and current era albums). TV = Taylor's Version; DE = Deluxe Edition; TTP = The Tortured Poets Department. See the online article for the color version of this figure.

style—which Swift herself mentioned both in interviews and throughout the Eras tour (emphasizing that she wanted to do things differently and develop characters; Linzi, 2024). Tracks on *folklore* feature emotions adjacent or near one another on Plutchik's emotional wheel (e.g., fear and sadness, surprise, and joy) rather than "opposites" (e.g., sadness and joy or trust and disgust; Plutchik, 1982). *Midnights* (TD), from Swift's "current era" period, is a second possible "sibling" album. It is a bit "outside" the grouping here as it is more positive and emotion based than its "current era" albums and is near *Lover*. Some fans consider *Midnights* and *Lover* sister albums in some regards, and the nearness of the two albums is indeed interesting. There is a clear album ("Era") movement: she is evolving as a songwriter, on average, leaning more into complex emotions like *Swiftian Saudade* yet always balancing emotion with lyricism.

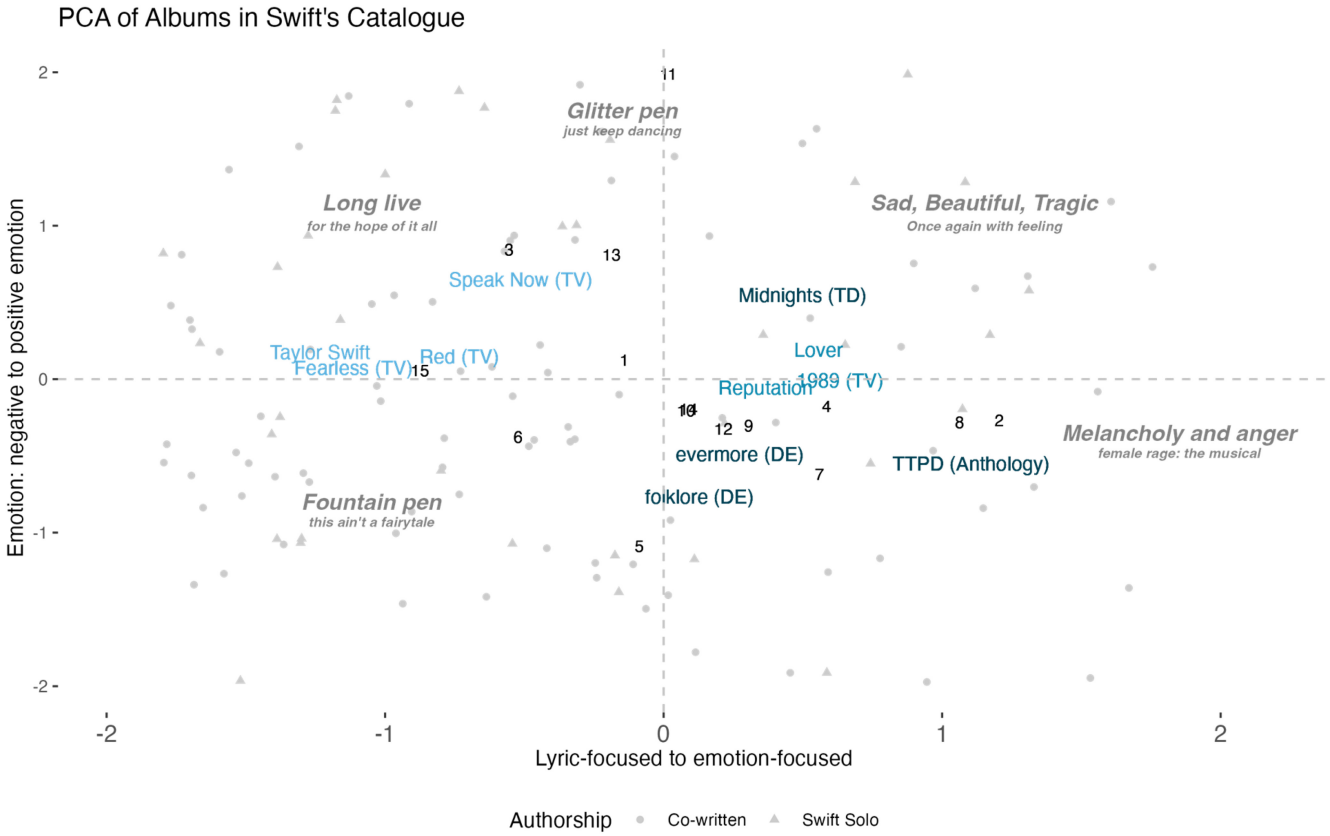
Our work builds on previous contributions using a contemporary artist, Taylor Swift, and expands upon prior work methodologically, incorporating multiple lexicon methods into machine learning-based analysis. Taylor Swift in particular has an intensely dedicated fanbase: Swifties love her albums and listen repeatedly when they come out (e.g., Swift's *TTPD* set records for most-streamed album in a single day, also racking up more than 300 million streams; the prior record holder had garnered that many streams in 7 days; Dailey, 2024). We noted in our survey of literature that lyrics are more meaningful in familiar sad music. It may be that Swift's unique happy–sad brand of *Swiftian Saudade* pulls listeners in and her narrative building is the invisible string tying it all together and making the storyline so compelling.

Our research also opens up many new questions about artists and songwriting in general, specifically how emotions can be amplified by and/or moderates the listening experience for readers when in concert with one another. There is also the question of whether Swift's approach to songwriting is distinct: are other contemporary artists taking this approach? Is this nothing new or well established?

Swiftian Saudade is a signature element within Swift's songwriting and the tension between lyrical composition and emotional content is one that she has increasingly incorporated in her work. Our selected approaches are accessible, common, and provide demonstrated insight. We addressed the challenge of capturing emotional and narrative nuance while incorporating emotional and lyrical aspects of songs in an easily interpretable manner. Our PCA of Swift's work provides a landscape in which we can explore additional artists to understand how emotions and lyrics can be balanced in the craft of songwriting. In the case of Taylor Swift, our work has revealed how Swift metaphorically wields her pens, empirically supports statements Swift has made about songwriting, provides evidence for fan theories about track fives, and reveals elements of Swift's songwriting edge and why she still has us singing along.

Figure 8

PCA Plot of Swift's Work Across the Lyric-Focused and Emotional Dimensions From Figure 6



Note. The figure displays both song-level values and averages by track number and album. At the song level, cowritten songs are represented by circles, while solo tracks by Swift are represented by triangles. For tracks, we include the average location by track number only for tracks numbered 15 or lower. PCA = principal component analysis; TV = Taylor's Version; DE = Deluxe Edition; TTPD = The Tortured Poets Department. See the online article for the color version of this figure.

Conclusions

How artists incorporate emotions within songs matters for the listener experience: when individuals replay songs, the lyrics increase in importance for some emotions but not others, particularly sad ones (Ali & Peynircioğlu, 2006; Pereira et al., 2011; Sizer, 2019). To focus on the lyrics, we employed a text-based approach to analyze emotional expression within lyrical work. We conduct our analysis on the work of Taylor Swift as she is known for the emotional depth of her work and her dedicated fanbase (Lansky, 2023; Morris, 2024). In our study, we evaluated three hypotheses: that Swift's work exhibits an emotional complexity we term *Swiftian Saudade*, that her albums' track five songs are distinctive and Saudade-laden, and that her work is characterized by linguistic composition and emotional content.

We then considered the arrangement of tracks within albums, as specific tracks are thought to have meaning, particularly in the case of the fifth song on an album. Track fives are at a distinct place in the landscape of our PCA analyses. Track fives straddle the line between the two key dimensions within Swift's work: lyrically focused and emotionally focused. Following our expectations, we also found that her track fives not only possess *Swiftian Saudade* but are highly Authentic.

In our PCA, we identified clear clusters of variables, where indeed, some of the familiar nomenclature from Swift's work appeared. In line with our expectations, there are groupings of variables around linguistic and narrative elements (desire, cognitive processes, authenticity, temporal focus, and personal pronoun usage), one for more positive emotions (joy, positivity, trust, and anticipation), and one for more negative emotions (anger, fear, sadness, and negativity).

Finally, we found that *Swiftian Saudade* is present across Swift's catalog. *Swiftian Saudade* exists even in Swift's early work, becoming more present as her catalog expanded. Our analyses reveal an artist whose work is complex, who has a structure within albums (with *Swiftian Saudade* intensified in track fives), and whose work has evolved over time.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101; https://onlinelibrary.wiley.com/do

Alek, A., Marzuki, A. G., Hidayat, D., & Sari, E. (2020). A critical discourse analysis of song "look what you made me do" by Taylor Swift. *Eralingua: Jurnal Pendidikan Bahasa Asing dan Sastra*, *4*(2), 154–161. https://doi.org/10.26858/eralingua.v4i2.11199

Ali, S. O., & Peynircioğlu, Z. F. (2006). Songs and emotions: Are lyrics and melodies equal partners? *Psychology of Music*, *34*(4), 511–534. https://doi.org/10.1177/0305735606067168

Apple Music. (2020). *Taylor Swift's Songwriting Process on "evermore"—Apple Music*. Retrieved March 20, 2025, from https://www.youtube.com/watch?v=CQacWbsLbS4

Armstrong, M. (2017, November). *Taylor Swift And AT&T Announce "Making of a Song" Behind-the-Scenes Series*. Retrieved March 7, 2025, from https://www.billboard.com/music/pop/taylor-swift-att-making-of-a-song-gorgeous-8022152/

Arnold, G. (2021). I dont give a damn about your bad reputation: Taylor Swift, Beyoncé Knowles, and performance. *Contemporary Music Review*, *40*(1), 27–40. https://doi.org/10.1080/07494467.2021.1945224

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. (2022). *The development and psychometric properties of LIWC-22*. University of Texas at Austin. https://www.liwc.app

Brand, C. O., Acerbi, A., & Mesoudi, A. (2019). Cultural evolution of emotional expression in 50 years of song lyrics. *Evolutionary Human Sciences*, *1*, Article e11. https://doi.org/10.1017/ehs.2019.11

Browne, E. (2024, April). *All of Taylor Swift's Famously Devastating Track 5's, Ranked*. Retrieved March 1, 2025, from https://www.vulture.com/article/taylor-swift-track5-songs-ranked.html

Cook, T., Roy, A. R. K., & Welker, K. M. (2019). Music as an emotion regulation strategy: An examination of genres of music and their roles in emotion regulation. *Psychology of Music*, *47*(1), 144–154. https://doi.org/10.1177/0305735617734627

Coscarelli, J. (2019). How Taylor Swift Writes a Love Song. *The New York Times*. Retrieved March 7, 2025, from https://www.nytimes.com/2019/12/24/arts/music/taylorswift-lover.html

Cowan, R. (2024, September). *Trump says "I hate Taylor Swift!" in Truth Social post—Reuters*. Retrieved September 26, 2024, from https://www.reuters.com/world/us/trump-says-i-hate-taylor-swift-truth-social-post-2024-09-15/

Czarnek, G., & Stillwell, D. (2022). Two is better than one: Using a single emotion lexicon can lead to unreliable conclusions. *PLOS ONE*, *17*(10), Article e0275910. https://doi.org/10.1371/journal.pone.0275910

Czechowski, K., Miranda, D., & Sylvestre, J. (2016). Like a rolling stone: A mixed-methods approach to linguistic analysis of Bob Dylan's lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(1), 99–113. https://doi.org/10.1037/aca0000045

Dailey, H. (2023, November). *Taylor Swift Is In Her College Era: Harvard, University of Florida Add Courses Dedicated to Pop Star*. Retrieved April 2, 2024, from https://www.billboard.com/music/music-news/taylor-swift-classes-harvard-university-offlorida-2024-1235513354/

Dailey, H. (2024, May). *Taylor Swift's 'Tortured Poets Department': Records Broken*. Retrieved April 26, 2024, from https://www.billboard.com/lists/taylor-swift-torturedpoets-department-records-broken/

Darwich, M., Mohd Noah, S. A., Omar, N., & Osman, N. (2019). Corpus-based techniques for sentiment lexicon generation: A review. *Journal of Digital Information Management*, *17*(5), 296–305. https://doi.org/10.6025/jdim/2019/17/5/296-305

DeWall, C. N., Pond, R. S., Campbell, W. K., & Twenge, J. M. (2011). Tuning in to psychological change: Linguistic markers of psychological traits and emotions over time in popular U.S. song lyrics. *Psychology of Aesthetics, Creativity, and the Arts*, *5*(3), 200–207. https://doi.org/10.1037/a0023195

Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: Lexicon versus machine learning. *Journal of Consumer Marketing*, *34*(6), 480–488. https://doi.org/10.1108/JCM-03-2017-2141

Dodds, P. S., & Danforth, C. M. (2010). Measuring the happiness of large-scale written expression: Songs, blogs, and presidents. *Journal of Happiness Studies*, *11*(4), 441–456. https://doi.org/10.1007/s10902-009-9150-9

Earhart, E. (2024). *Taylor Swift Dataset (TTPD included)*. Retrieved March 13, 2025, from https://www.kaggle.com/datasets/yaminimanral/taylor-swift-dataset-ttpd-included

Eaton, J., Sharma, A., & Law, D. (2022). "I may look fake but I'm real where it counts": Positivity and authenticity in the songs of Dolly Parton. *Psychology of Aesthetics, Creativity, and the Arts*, *19*(2), 205–217. https://doi.org/10.1037/aca0000527

Ellis, J. (2020, December). *Tis the Damn Season: Taylor Swift's evermore—Post45*. Retrieved October 11, 2023, from https://post45.org/sections/contemporaries-essays/tis-thedamn-season-taylor-swifts-evermore/

Fogarty, M., & Arnold, G. (2021). Are you ready for it? Re-evaluating Taylor Swift. *Contemporary Music Review*, *40*(1), 1–10. https://doi.org/10.1080/07494467.2021.1976586

Fricke, K. R., Greenberg, D. M., Rentfrow, P. J., & Herzberg, P. Y. (2018). Computer-based music feature analysis mirrors human perception and can be used to measure individual music preference. *Journal of Research in Personality*, *75*, 94–102. https://doi.org/10.1016/j.jrp.2018.06.004

Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., & Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, *2*(1), Article 100. https://doi.org/10.1038/s43586-022-00184-w

Guo, S., Lu, J., Wang, Y., Li, Y., Huang, B., Zhang, Y., Gong, W., Yao, D., Yuan, Y., & Xia, Y. (2020). Sad music modulates pain perception: An EEG study. *Journal of Pain Research*, *13*, 2003–2012. https://doi.org/10.2147/JPR.S264188

Harrison, C., & Ringrow, H. (2022). Disnarration and the performance of storytelling in Taylor Swift's folklore and evermore [Number: 1]. *International Journal of Literary Linguistics*, *11*(1). https://doi.org/10.15462/ijll.v11i1.140

Jockery, M. (2023). *Syuzhet*. Retrieved March 20, 2025, from https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html

Johari, I. (2023, June). *Taylor Swift All Lyrics (64 Albums)*. https://www.kaggle.com/datasets/ishikajohari/taylor-swift-all-lyrics-30-albums

Kalichman, S. C., & Smyth, J. M. (2023). "And you don't like, don't like the way I talk": Authenticity in the language of Bruce Springsteen. *Psychology of Aesthetics, Creativity, and the Arts*, *17*(5), 581–589. https://doi.org/10.1037/aca0000402

Kalichman, S. C., & Smyth, J. M. (2024). "There's no one as honest as those in pain": The language of Tom Petty's song lyrics. *Psychology of Music*, *53*(2), 176–191. https://doi.org/10.1177/03057356241235645

Kendong, F. A., Daud, A. S., & Joharry, S. A. (2023). A corpus-driven analysis of Taylor Swift's song lyrics. *International Journal of Modern Languages and Applied Linguistics*, *7*(2), 59–82. https://doi.org/10.24191/ijmal.v7i2.7489

Kušen, E., Cascavilla, G., Figl, K., Conti, M., & Strembeck, M. (2017, August 21-23). *Identifying emotions in social media: Comparison of word-emotion lexicons*. 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), Prague, Czech Republic. (pp. 132–137). https://doi.org/10.1109/FiCloudW.2017.75

Lansky, S. (2023). *Taylor Swift Is TIME's 2023 Person of the Year*. TIME. Retrieved December 14, 2023, from https://time.com/6342806/person-of-the-year-2023-taylor-swift/

Linzi, L. C. (2024, June). *Taylor Swift Eras Tour LIVE in Lisbon—Folklore & Evermore*. Retrieved March 12, 2025, from https://youtu.be/3bKqt8JWIEY?si=CLxhgeb6VuoXjGle&t=6

Lister, L. (2001). Divafication: The deification of modern female pop stars. *Popular Music and Society*, *25*(3-4), 1–10. https://doi.org/10.1080/03007760108591796

McGrath, J. (2023). The return to craft: Taylor Swift, nostalgia, and COVID-19. *Popular Music and Society*, *46*(1), 70–84. https://doi.org/10.1080/03007766.2022.2156761

Mohammad, S., & Turney, P. (2010). Emotions evoked by common words and phrases: using mechanical turk to create an emotion lexicon. In D. Inkpen & C. Strapparava (Eds.), *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and*

Generation of Emotion in Text (pp. 26–34). Association for Computational Linguistics. https://aclanthology.org/W10-0204/

Mohammad, S. M., & Turney, P. D. (2013). *Crowdsourcing a Word-Emotion Association Lexicon*. arXiv. https://arxiv.org/abs/1308.6297

Morris, A. (2024). Drew a map on your bedroom ceiling: Fandoms, nostalgic girlhood and digital bedroom cultures in the Swiftie-sphere. *Celebrity Studies*, *16*(1), 77–95. https://doi.org/10.1080/19392397.2024.2338540

Neto, F., & Mullet, E. (2014). A prototype analysis of the Portuguese concept of Saudade. *Journal of Cross-Cultural Psychology*, *45*(4), 660–670. https://doi.org/10.1177/0022022113518370

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236. https://doi.org/10.1080/01638530802073712

Nicholson, J. (2022, September). *Taylor Swift at Nashville Songwriter Awards: Read Her Full Speech—Billboard*. Retrieved December 14, 2023, from https://www.billboard.com/music/country/taylor-swift-nashville-songwriter-awards-full-speech-1235142144/

Nielsen, F. A. (2011). *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs*. Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages (Vol. 718, pp. 93–98). https://ceur-ws.org/Vol-718/paper16.pdf

North, A. C., Krause, A. E., & Ritchie, D. (2021). The relationship between pop music and lyrics: A computerized content analysis of the United Kingdom's weekly top five singles, 1999–2013. *Psychology of Music*, *49*(4), 735–758. https://doi.org/10.1177/0305735619896409

Parada-Cabaleiro, E., Mayerl, M., Brandl, S., Skowron, M., Schedl, M., Lex, E., & Zangerle, E. (2024). Song lyrics have become simpler and more repetitive over the last five decades. *Scientific Reports*, *14*(1), Article 5531. https://doi.org/10.1038/s41598-024-55742-x

Park, M., Thom, J., Mennicken, S., Cramer, H., & Macy, M. (2019). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*, *3*(3), 230–236. https://doi.org/10.1038/s41562-018-0508-z

Pereira, C. S., Teixeira, J., Figueiredo, P., Xavier, J., Castro, S. L., & Brattico, E. (2011). Music and emotions in the brain: Familiarity matters. *PLoS ONE*, *6*(11), Article e27241. https://doi.org/10.1371/journal.pone.0027241

Petrie, K. J., Pennebaker, J. W., & Sivertsen, B. (2008). Things we said today: A linguistic analysis of the Beatles. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(4), 197–202. https://doi.org/10.1037/a0013117

Plutchik, R. (1982). A psychoevolutionary theory of emotions. *Social Science Information*, *21*(4-5), 529–553. https://doi.org/10.1177/053901882021004003

Queiroz, G. D., Fay, C., Hvitfeldt, E., Keyes, O., Misra, K., Mastny, T., Erickson, J., Robinson, D., & Silge, J. (2024). *tidytext: Text mining using "dplyr," "ggplot2," and other tidy tools* (Version 0.4.2) [Computer software]. https://cran.r-project.org/web/packages/tidytext/index.html

Rahmani, D. D., & Nasution, S. S. (2019). Figurative language in song lyric "Red" by Taylor Swift. *LEXEME: Journal of Linguistics and Applied Linguistics*, *1*(2), 26–31. https://doi.org/10.32493/LJLAL.V1I2.5124

Rambsy, K. (2018). *#TheJayZMixtape*. Publishing Without Walls. https://doi.org/10.21900/pww.2

Rinker, T. (2018). *Textstem: Tools for Stemming and Lemmatizing Text*. https://cran.r-project.org/web/packages/textstem/index.html

Santika, P., & Syafryadin, S. (2023). An analysis of figurative language in song lyrics of the album "MIDNIGHTS" by TAYLOR SWIFT [Number: 1]. *Wiralodra English Journal (WEJ)*, *7*(1), 14–28. https://doi.org/10.31943/wej.v7i1.189

Schultz, K. (2024). Why Taylor Swift's new track five 'So Long London' holds extra meaning. *Washington Post*. Retrieved October 5, 2024, from https://www.washingtonpost.com/entertainment/music/2024/04/18/taylor-swift-track-5-so-long-london/

Shlens, J. (2014). *A tutorial on principal component analysis*. arXiv. https://doi.org/10.48550/arXiv.1404.1100

Sizer, L. (2019). Sad songs say so much: The paradoxical pleasures of sad music: Sizer *Sad songs* say so much. *The Journal of Aesthetics and Art Criticism*, *77*(3), 255–266. https://doi.org/10.1111/jaac.12659

Sloan, N. (2021). Taylor Swift and the work of songwriting. *Contemporary Music Review*, *40*(1), 11–26. https://doi.org/10.1080/07494467.2021.1945226

Smialek, E. (2021). Who needs to calm down? Taylor Swift and Rainbow Capitalism. *Contemporary Music Review*, *40*(1), 99–119. https://doi.org/10.1080/07494467.2021.1956270

Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative analysis of lexicon and machine learning approach for sentiment analysis. *International Journal of Advanced Computer Science and Applications*, *13*(3), 71–77. https://doi.org/10.14569/IJACSA.2022.0130312

Sweney, M. (2018). Taylor Swift strikes a blow for fellow artists as digital revenues soar. *The Observer*. Retrieved October 6, 2023, from https://www.theguardian.com/business/2018/nov/24/taylor-swift-blow-fellow-artists-streaming-revenues-soar-universalspotify

Swift, T. (2019, August). *Taylor Swift answers questions from fans from lover's lounge live*. Retrieved March 7, 2025, from https://www.youtube.com/watch?v=iyYhqYELWtA

Swiftposium 2024. (2024). *An academic conference on Taylor Swift. (2024)*. https://swiftposium2024.com

Tabak, F. S., & Evrim, V. (2016, October 13–14). *Comparison of emotion lexicons*. 13th International Symposium on High Capacity Optical Networks and Enabling Technologies, HONET (2016 HONET-ICT), Nicosia, Cyprus (pp. 154–158). https://doi.org/10.1109/HONET.2016.7753440

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, *37*(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Talaat, A. S. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data*, *10*(1), Article 110. https://doi.org/10.1186/s40537-023-00781-w

Tausczik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927X09351676

Taylor Swift. (2009). In *Taylor Swift wiki*. Retrieved March 14, 2025, from https://taylorswift.fandom.com/wiki/Taylor_Swift

Thoma, M. V., Ryf, S., Mohiyeddini, C., Ehlert, U., & Nater, U. M. (2012). Emotion regulation through listening to music in everyday situations. *Cognition and Emotion*, *26*(3), 550–560. https://doi.org/10.1080/02699931.2011.595390

VandenBos, G. R. (Ed.). (2007). *APA Dictionary of psychology* (p. xvi, 1024). American Psychological Association.

Wilson, L. (Director). (2020). *Miss Americana* [Film]. Netflix. https://www.netflix.com/title/81028336

Yastanti, U., & Susilawati, S. (2020). Imagery in song lyrics of Taylor Swift [Number: 2]. *LINGUA: Jurnal Bahasa, Sastra, dan Pengajarannya*, *17*(2), 135–156. https://doi.org/10.30957/lingua.v17i2.642

Zad, S., Jimenez, J., & Finlayson, M. (2021, August). Hell Hath no fury? Correcting bias in the NRC emotion lexicon. In A. Mostafazadeh Davani, D. Kiela, M. Lambert, B. Vidgen, V. Prabhakaran, & Z. Waseem (Eds.), *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)* (pp. 102–113). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.woah-1.11

Zhang, H., Gan, W., & Jiang, B. (2014). *Machine learning and lexicon based methods for sentiment classification: a survey*. 2014 11th Web Information System and Application Conference (pp. 262–265). https://doi.org/10.1109/WISA.2014.55

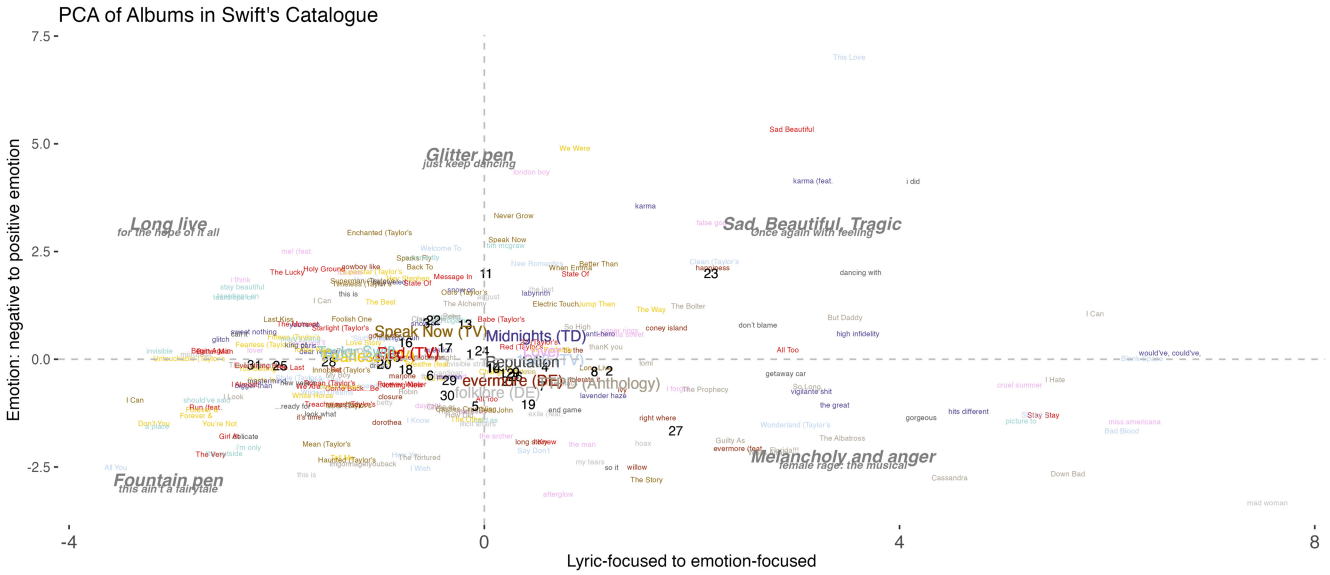
(*Appendix follows*)

Appendix

Full Plot of Taylor Swift Songs in Discography

Figure A1

Full PCA Plot of Swift's Work Across the Lyric-Focused and Emotional Dimensions



Note. PCA = principal component analysis; TV = Taylor's Version; DE = Deluxe Edition; TD = Till Dawn Edition. See the online article for the color version of this figure.

Table A1

Variables and Respective Summary Stats for PCA (Mean at 0)

Emotion	Min	*Mdn*	Max	*n*
Joy	−1.43	−0.27	6.01	234.00
Sadness	−1.33	−0.17	3.30	234.00
Anger	−1.13	−0.35	4.04	234.00
Authentic	−3.22	0.51	0.73	234.00
Trust	−1.48	−0.26	4.16	234.00
Disgust	−1.09	−0.29	3.73	234.00
Surprise	−1.40	−0.19	3.61	234.00
Anticipation	−1.62	−0.19	3.91	234.00
Fear	−1.33	−0.18	3.85	234.00
Present	−2.14	−0.06	3.08	234.00
Sadness (LIWC)	−0.42	−0.42	10.15	234.00
Swear	−0.38	−0.38	7.85	234.00
Tone	−1.40	−0.21	1.92	234.00
Future	−0.91	−0.18	7.51	234.00
Past	−1.36	−0.26	4.40	234.00
Linguistic	−2.79	−0.10	2.85	234.00
Affect	−1.57	−0.20	5.43	234.00
Anger (LIWC)	−0.35	−0.35	7.57	234.00
Anxiety (LIWC)	−0.41	−0.41	8.82	234.00
Emotion	−1.08	−0.26	6.62	234.00
Personal pronouns	−2.53	−0.08	4.37	234.00
Solo	−0.62	−0.62	1.61	234.00
Antonoff	−0.56	−0.56	1.78	234.00
Dessner	−0.46	−0.46	2.16	234.00
Saudade	−0.57	−0.57	1.76	234.00

Note. PCA = principal component analysis; min = minimum; max = maximum; LIWC = linguistic inquiry and word count dictionary.

Received October 22, 2024
Revision received July 24, 2025
Accepted July 25, 2025 ■