Data Mining and Text Mining (Taylor's Version)

Utilizing PCA and Sentiment Analysis to Analyze Swift's Discography

Madison Yonash, Susan Nunez, Jaylyn Leatherwood, and Nicolas Garay

Methodology - Sentiment Analysis

- Utilize the lyrics column from the taylor_album_songs dataset

- Unnest the tibbles and remove stop words

- Lexicons:
  - Bing
  - NRC Emotion

- Packages used include tidyverse, tidytext, wordcloud

```
my_words <- c("oh","ooh","eh","ha","mmm","mm", "yeah","ah","hey","eeh","uuh","uh","la","da","di","r
a","huh","hu","whoa","gonna","wanna","gotta","em")

tidy_taylor <-
  taylor_album_songs  %>%
  unnest(lyrics)  %>%
  unnest_tokens(word, lyric) %>%
  filter(!word %in% my_words) %>% #Remove undesirables
  filter(!nchar(word) < 3) %>% #Words like "ah" or "oo" used in music
  anti_join(stop_words)
```


"RATATATATA TATATA DA DA!"
– Taylor Swift

Results - Text Analysis

Most common words

By Song

| track_name	word	n
<chr>	<chr>	<int>
Red (Taylor's Version)	red	107
Shake It Off	shake	78
Run (Taylor's Version) [From The Vault]	run	52
This Love	love	52
Call It What You Want	call	43
Daylight	daylight	40
Out Of The Woods	woods	38
I Knew You Were Trouble (Taylor's Version)	trouble	34
All You Had To Do Was Stay	stay	31
Stay Stay Stay (Taylor's Version)	stay	31

Results - Full Discography Sentiment Analysis

NRC Emotion Sentiment Analysis
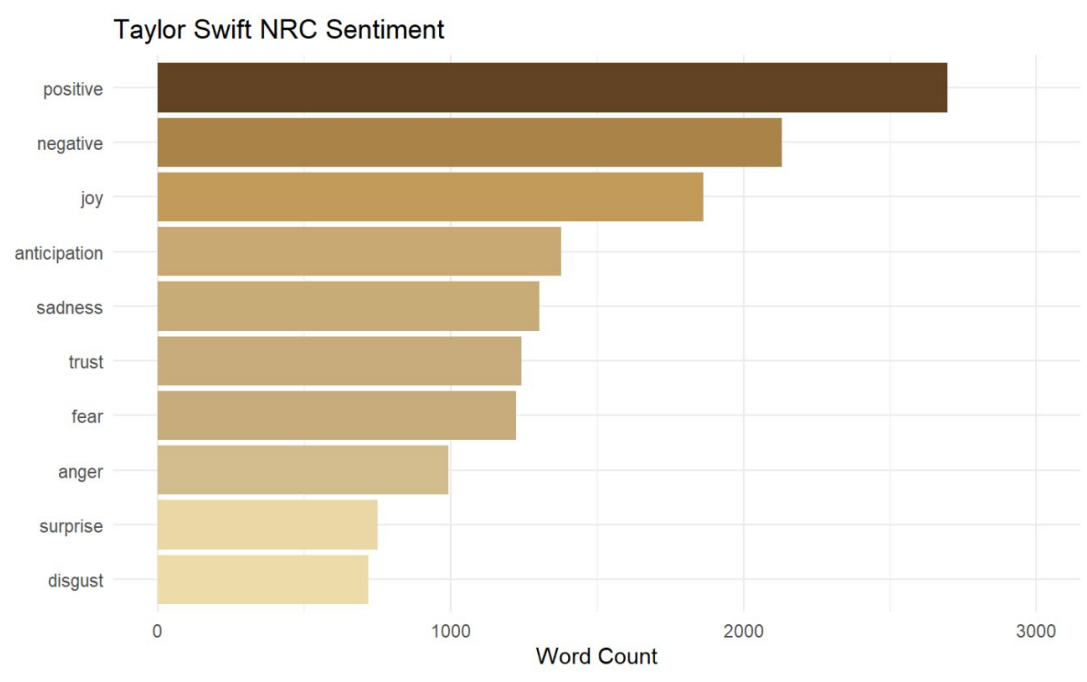
Bing Sentiment Analysis

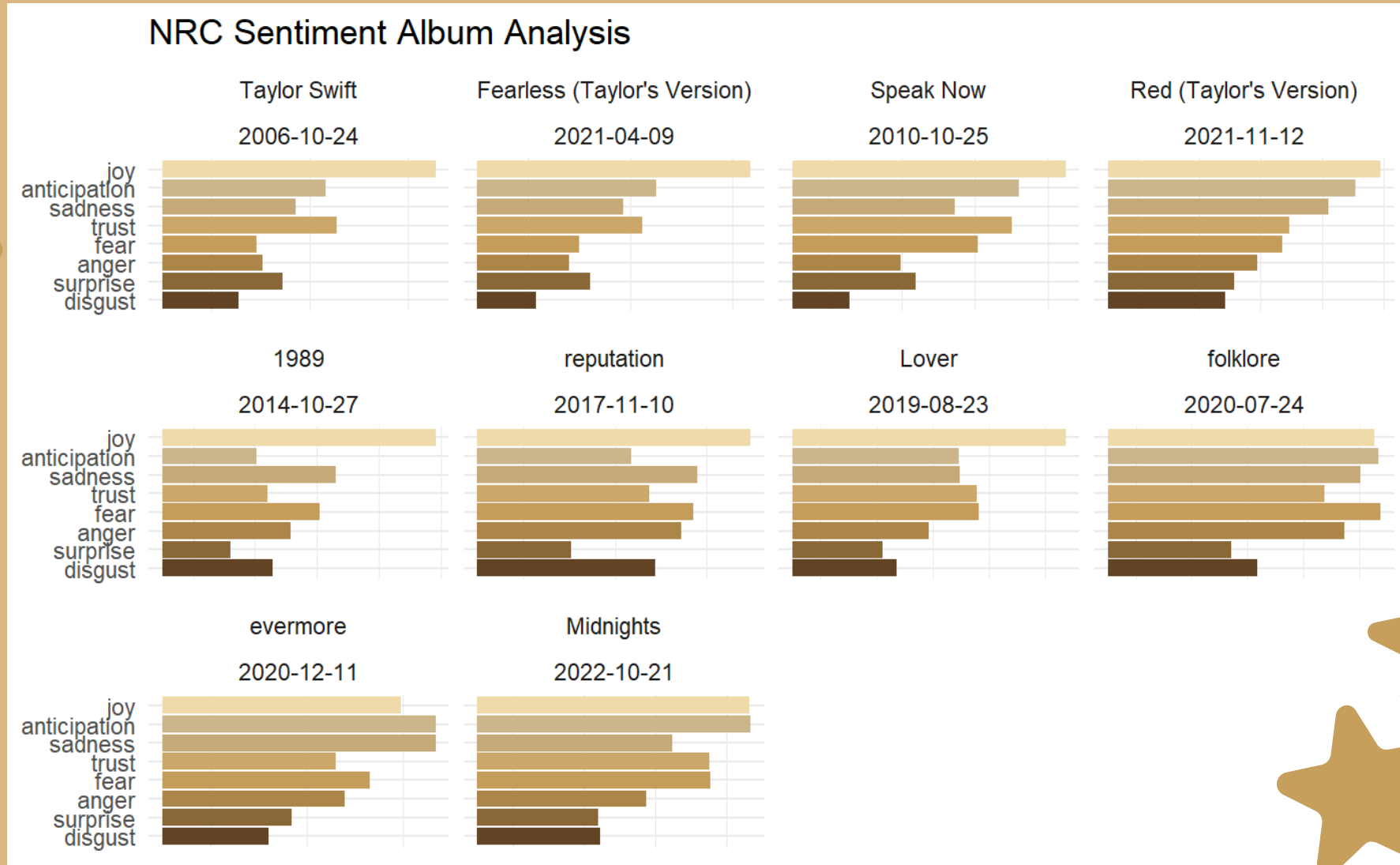
Taylor Swift NRC Sentiment

Taylor Swift Bing Sentiment
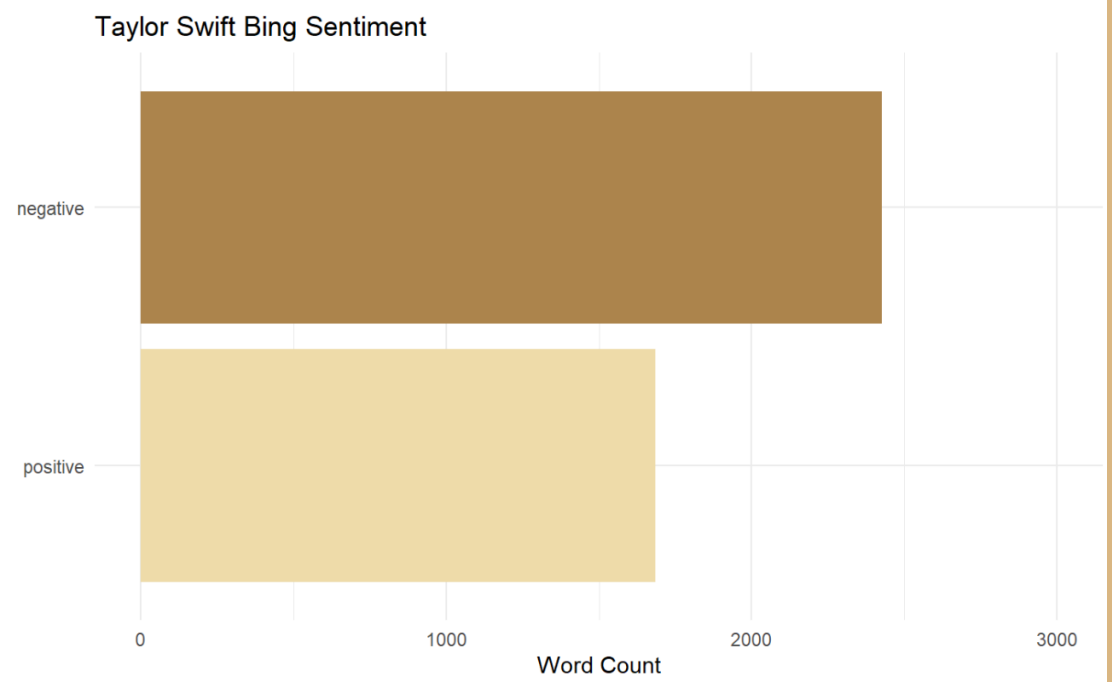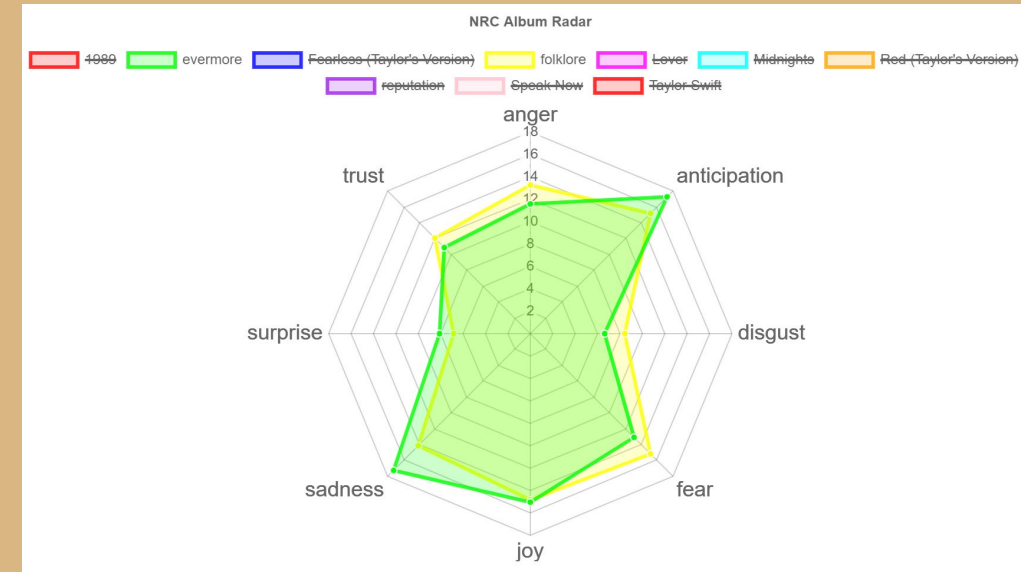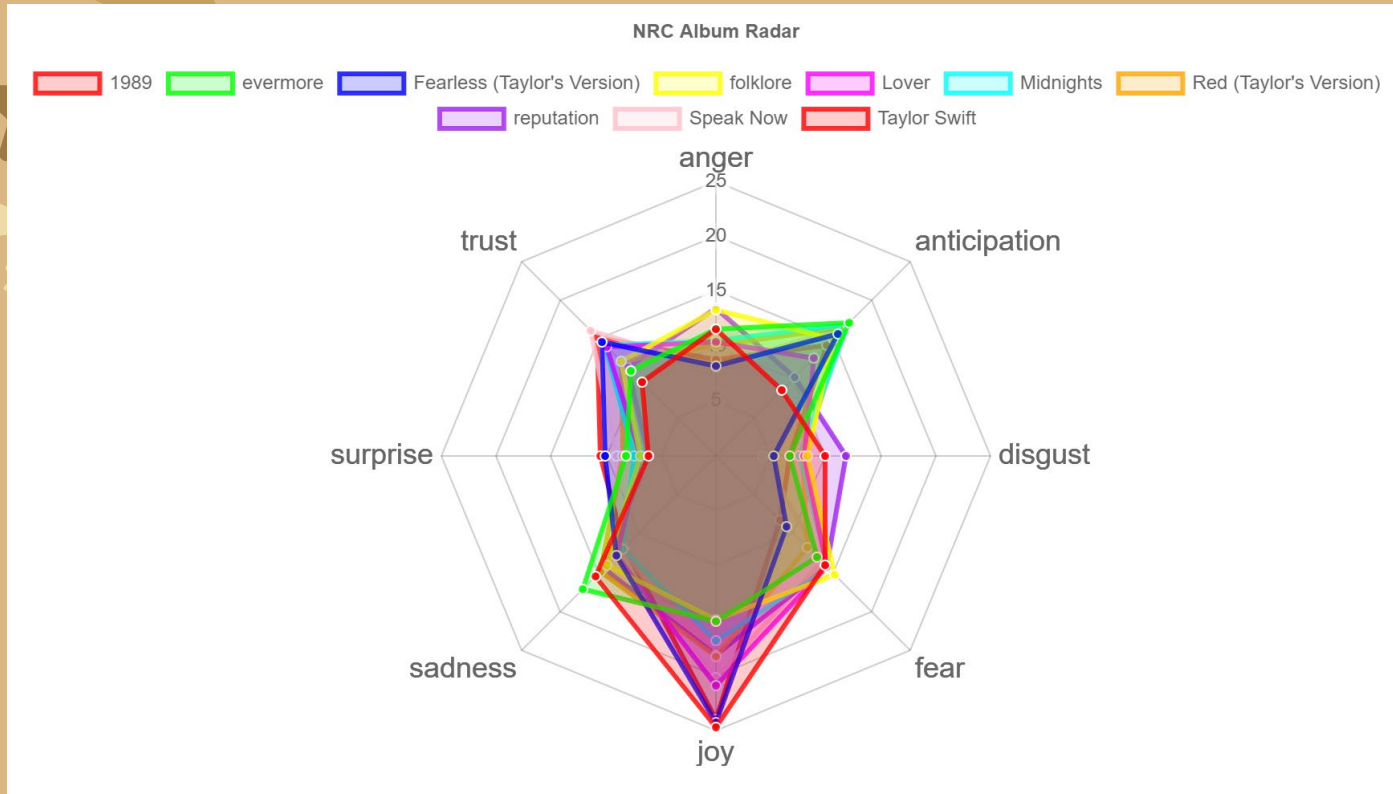
Results - Sentiment Analysis by Album

Results - Sentiment Analysis by Album
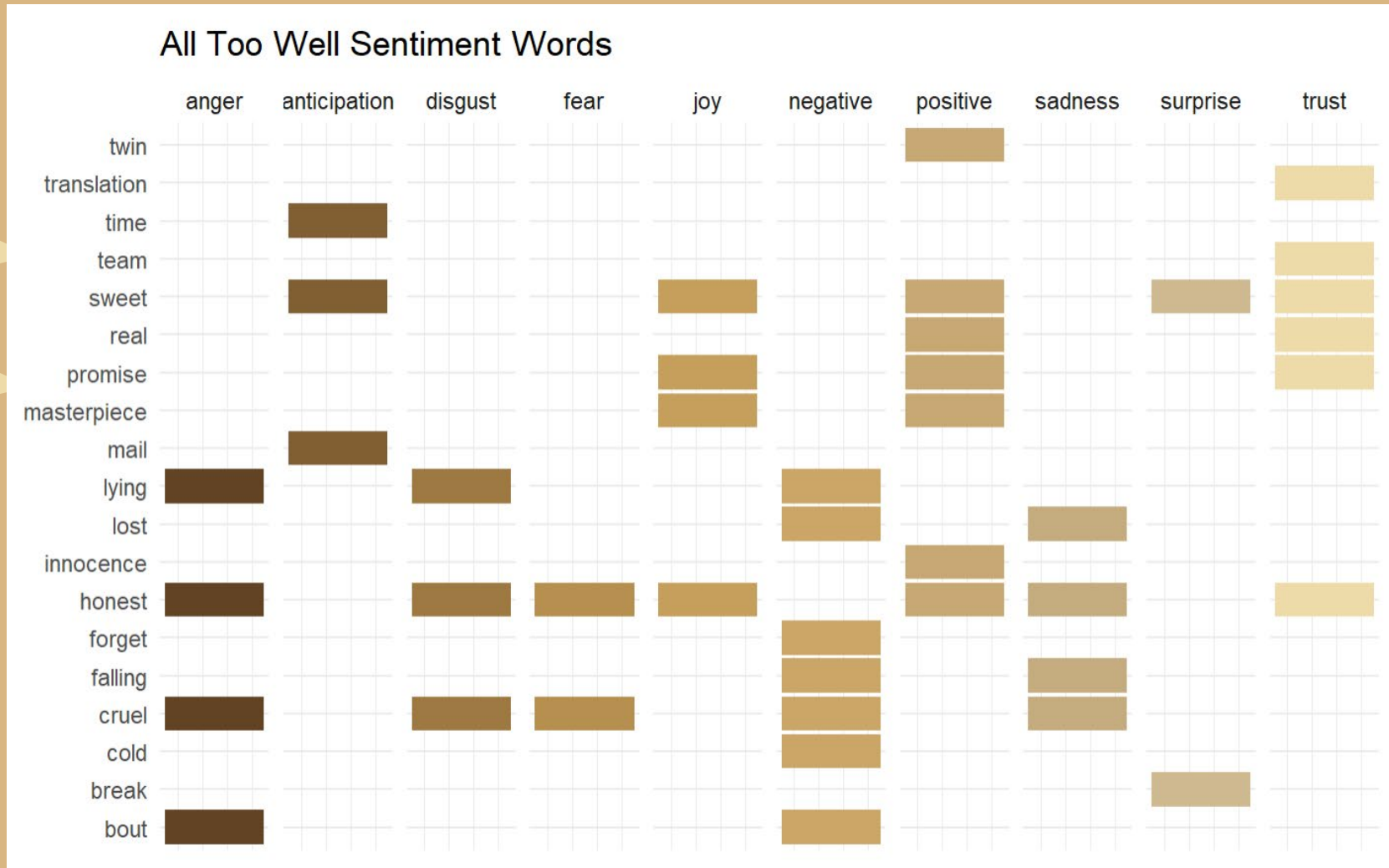

NRC Sentiment Album Analysis

Results - Sentiment Analysis by Album

Interactive Radar Chart to Compare Album Sentiments

Results - Sentiment Analysis of All Too Well

All Too Well Sentiment Words

Results - Sentiment Analysis of Track 5

NRC Sentiment Track Five Analysis