

## A quick and easy way to show off the fun stuff

```
# read in subtitle files and create a `movie` column for identification
```

```
animated <- rbindlist(list(  
  data.table(read_srt(path = "MovieSubtitles/Animated/Aladdin.srt", collapse = " "))[, movie := "Aladdin"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Beauty and the Beast.srt", collapse = " "))[, movie := "Beauty and the Beast"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Brave.srt", collapse = " "))[, movie := "Brave"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Cinderella.srt", collapse = " "))[, movie := "Cinderella"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/The Princess and the Frog.srt", collapse = " "))[, movie := "The Princess and the Frog"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/The Little Mermaid.srt", collapse = " "))[, movie := "The Little Mermaid"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Moana.srt", collapse = " "))[, movie := "Moana"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Mulan.srt", collapse = " "))[, movie := "Mulan"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Pocahontas.srt", collapse = " "))[, movie := "Pocahontas"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Sleeping Beauty.srt", collapse = " "))[, movie := "Sleeping Beauty"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Snow White and The Seven Dwarfs.srt", collapse = " "))[, movie := "Snow White and The Seven Dwarfs"],  
  data.table(read_srt(path = "MovieSubtitles/Animated/Tangled.srt", collapse = " "))[, movie := "Tangled"]  
)  
animated[sample(1:nrow(animated), size = 6)]
```

| ##    | n    | start    | end      | subtitle                           | movie      |
|-------|------|----------|----------|------------------------------------|------------|
| ## 1: | 756  | 4150.104 | 4153.148 | -No! -Come on !                    | Mulan      |
| ## 2: | 594  | 2933.724 | 2936.143 | What's gotten into you, fella? Oh. | Mermaid    |
| ## 3: | 1218 | 3899.354 | 3901.396 | For you, it's gonna be tough       | Frog       |
| ## 4: | 775  | 3263.601 | 3264.521 | Thank you.                         | Cinderella |
| ## 5: | 1075 | 4589.520 | 4591.238 | You can't fix it!                  | Moana      |
| ## 6: | 372  | 1966.048 | 1968.259 | That's why the wisps led me here.  | Brave      |

```
# generates longer data using each of the three dictionaries
```

```
# each word is on its own row
```

```
aniAfinnWordLonger <- animated %>%  
  unnest_tokens(word, subtitle) %>%  
  .[afinn, on = "word", nomatch = 0]  
aniAfinnWordLonger[sample(1:nrow(aniAfinnWordLonger), size = 6)]
```

| ##    | n   | start    | end      | movie      | startScale | endScale  | song  | word     | value |
|-------|-----|----------|----------|------------|------------|-----------|-------|----------|-------|
| ## 1: | 748 | 3154.393 | 3155.121 | Cinderella | 0.7403409  | 0.7393216 | FALSE | please   | 1     |
| ## 2: | 541 | 3556.015 | 3559.477 | Snow       | 0.7036807  | 0.7035562 | FALSE | goodness | 3     |
| ## 3: | 449 | 1503.585 | 1506.129 | Frog       | 0.2602189  | 0.2599959 | FALSE | poor     | -2    |
| ## 4: | 723 | 3838.752 | 3841.962 | Mulan      | 0.7818999  | 0.7821767 | FALSE | yeah     | 1     |
| ## 5: | 741 | 3142.387 | 3145.936 | Aladdin    | 0.6359557  | 0.6361464 | FALSE | laugh    | 1     |
| ## 6: | 284 | 1567.858 | 1570.569 | Mermaid    | 0.3235100  | 0.3233132 | FALSE | pressure | -1    |

```
aniBingWordLonger <- animated %>%  
  unnest_tokens(word, subtitle) %>%  
  .[bing, on = "word", nomatch = 0]  
aniBingWordLonger[sample(1:nrow(aniBingWordLonger), size = 6)]
```

```
##      n      start      end  movie startScale endScale song      word
## 1:  675 3314.021 3318.108 Mermaid 0.70217454 0.7020031 FALSE    easy
## 2:  356 1846.910 1850.475 Tangled 0.33054471 0.3305995 FALSE  triumphant
## 3: 1026 5434.158 5435.857 Tangled 0.99191696 0.9916546 FALSE    beloved
## 4:  401 2069.026 2073.656  Brave 0.36142879 0.3617203 FALSE  attractive
## 5:   30 127.547 131.096 Aladdin 0.01995815 0.0199483 FALSE    dead
## 6: 1244 3989.652 3993.655  Frog 0.70613136 0.7062234 FALSE    right
##      sentiment
## 1: positive
## 2: positive
## 3: positive
## 4: positive
## 5: negative
## 6: positive
```

```
# in this case, words can be repeated if they have more than one sentiment
aniNrcWordLonger <- animated %>%
  unnest_tokens(word, subtitle) %>%
  .[nrc, on = "word", nomatch = 0]
aniNrcWordLonger[sample(1:nrow(aniNrcWordLonger), size = 6)]
```

```
##      n      start      end  movie startScale endScale song      word
## 1:   5   71.593   76.509 Tangled 0.003233807 0.00352433 FALSE  heavens
## 2: 1115 4356.446 4359.643 Sleeping 0.965943683 0.96500697 FALSE   marry
## 3:  566 3658.826 3662.079   Snow 0.725038192 0.72486177 FALSE  grumpy
## 4:  783 3914.872 3917.625 Mermaid 0.832472201 0.83191786 FALSE  lovely
## 5:  429 2073.747 2076.181 Aladdin 0.417609271 0.41750100 FALSE    hit
## 6:  909 3801.600 3803.556  Moana 0.591670299 0.59172568 TRUE    hide
##      sentiment
## 1: positive
## 2:   trust
## 3:  sadness
## 4: positive
## 5: negative
## 6:    fear
```

```
# this version is .001 seconds slower...
# aniNrcLineWider <- aniNrcWordLonger[, groupID := .GRP, by = list(n, movie)] %>%
#   unique() %>%
#   select(-c(word)) %>%
#   pivot_wider(names_from = sentiment, values_from = 1, values_fn = length) %>%
#   arrange(movie, start) %>%
#   mutate(across(nrcSentiments, ~ifelse(is.na(.x), FALSE, TRUE))) %>%
#   data.table()
# aniNrcLineWider[sample(1:nrow(aniNrcLineWider), size = 6)]

# creates a wider version of aniNrcWordLonger, but aggregates by line
# lines are identified with a `groupID` found by grouping by line number and movie
aniNrcLineWider <- aniNrcWordLonger[, groupID := .GRP, by = list(n, movie)] %>%
  unique() %>%
  .[, !c("word")] %>%
  pivot_wider(names_from = sentiment, values_from = 1, values_fn = length) %>%
  setorderv(c("movie", "start")) %>%
```

```
mutate(across(nrcSentiments, ~ifelse(is.na(.x), FALSE, TRUE))) %>%
data.table()
```

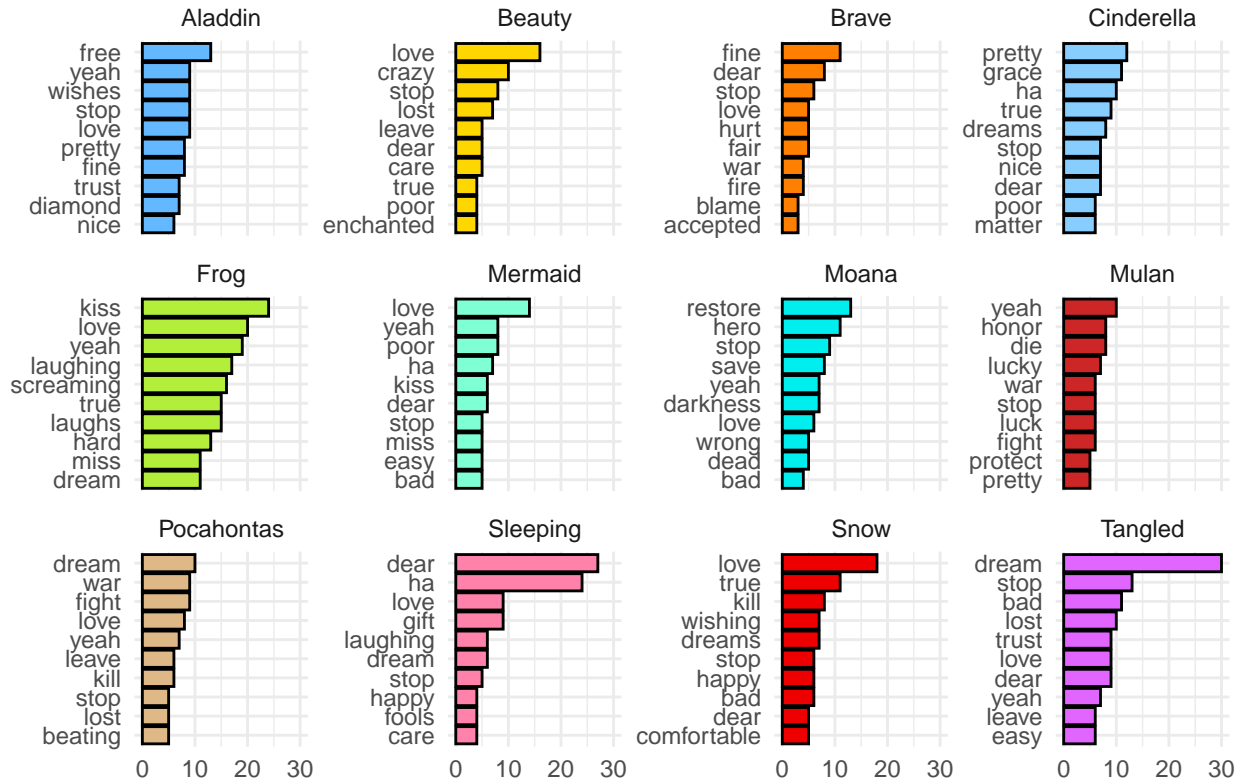
```
## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(nrcSentiments)' instead of 'nrcSentiments' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```

```
aniNrcLineWider[sample(1:nrow(aniNrcLineWider), size = 6)]
```

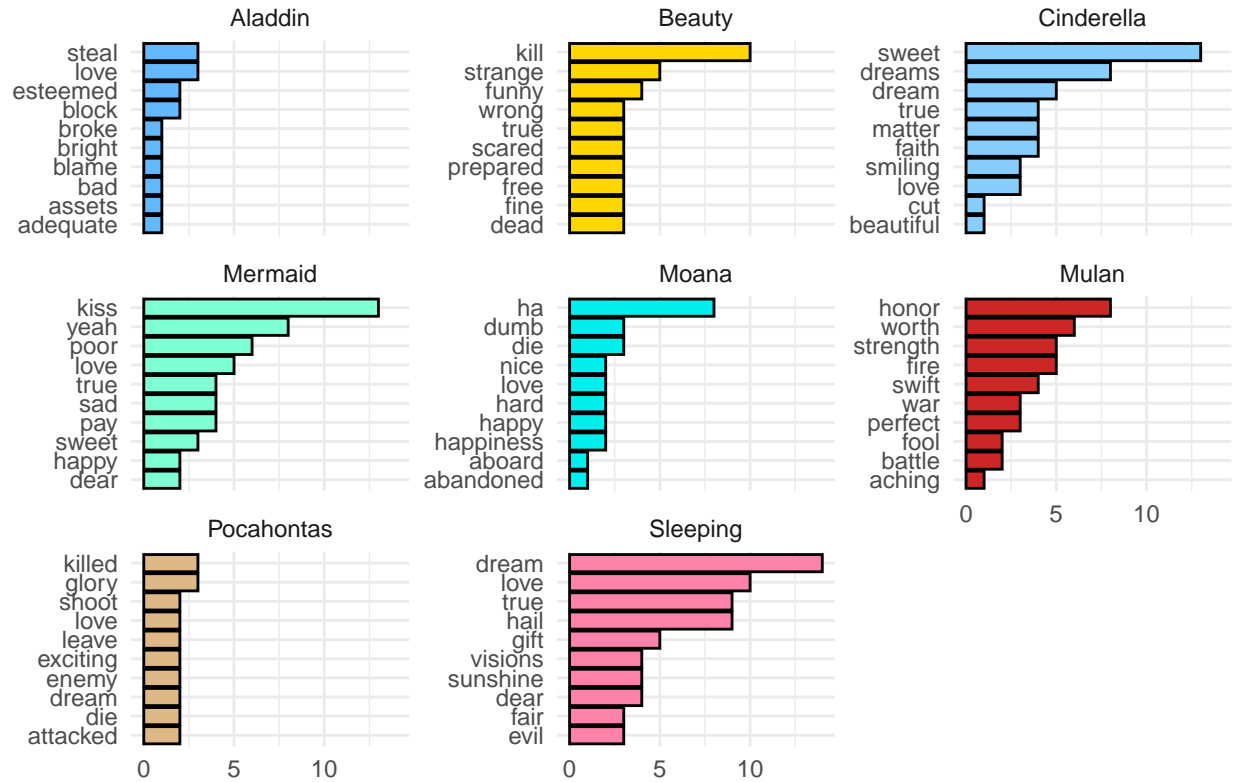
```
##      start      end      movie startScale  endScale  song groupID anger  fear
## 1:  438.137  445.441   Sleeping  0.0910155 0.09196511  TRUE      24 FALSE FALSE
## 2: 2810.891 2812.976     Frog  0.4947034 0.49442603 FALSE     3827 FALSE FALSE
## 3: 1530.362 1534.283   Mermaid  0.3153788 0.31545003  TRUE     3933 FALSE FALSE
## 4:   897.146  899.731     Frog  0.1514452 0.15121653 FALSE      994 FALSE FALSE
## 5: 4565.978 4570.064     Frog  0.8095038 0.80962313 FALSE     4447 FALSE FALSE
## 6: 4139.200 4143.400 Pocahontas 0.8564291 0.85632292 FALSE     2114  TRUE  TRUE
##      negative sadness positive disgust surprise  joy anticipation trust
## 1:   FALSE   FALSE    TRUE   FALSE   FALSE  TRUE      TRUE  TRUE
## 2:   FALSE   TRUE    TRUE   FALSE   FALSE  TRUE      FALSE FALSE
## 3:   FALSE   FALSE    TRUE   FALSE   FALSE  TRUE      TRUE  TRUE
## 4:   FALSE   FALSE   FALSE   FALSE   FALSE FALSE      TRUE  FALSE
## 5:   FALSE   FALSE    TRUE   FALSE   FALSE  TRUE      FALSE FALSE
## 6:    TRUE    TRUE    TRUE   FALSE   FALSE FALSE      FALSE FALSE
```

```
# I spent so much time on this but made it better above :(
# it also showcases the some of the tidyverse version of the nonsense above
# %>%
#   left_join(aniNrcWordLonger %>%
#             group_by(n, movie) %>%
#             mutate(groupID = cur_group_id()) %>%
#             select(-c(n, word, sentiment)) %>%
#             unique(),
#             by = "groupID") %>%
#   select(movie, n, start, end, nrcSentiments) %>%
#   arrange(movie, n)
```

## Top ten most common spoken words by movie

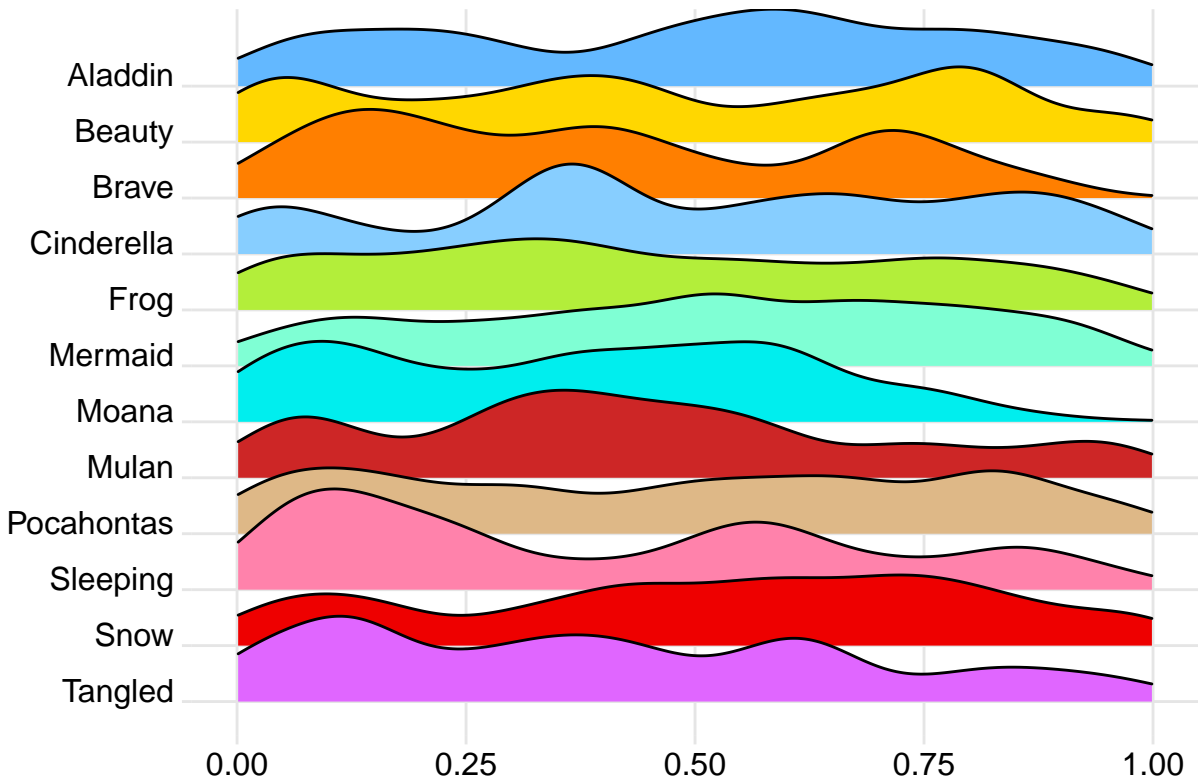


## Top ten most common sung words by movie

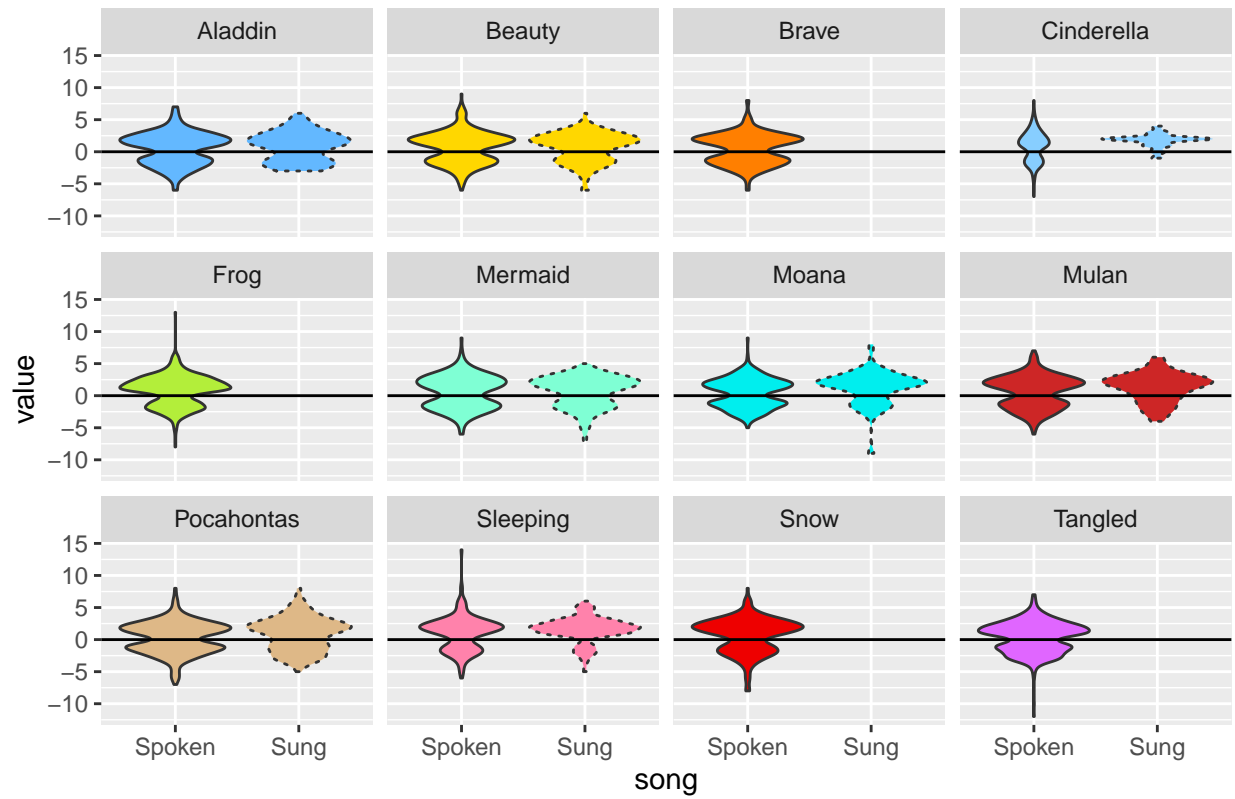


## Picking joint bandwidth of 0.0575

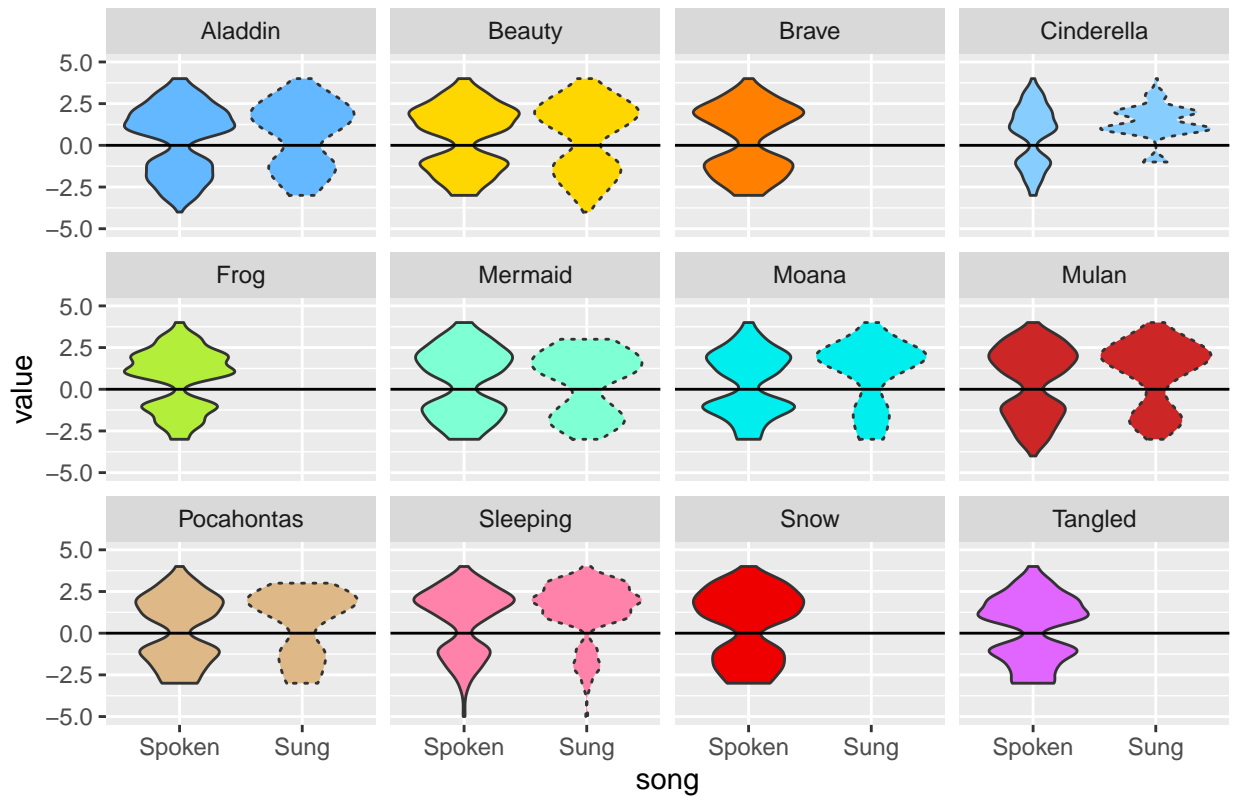
### NRC emotions over the course of the movie



Variation in positive and negative emotions of lines by type



Variation in positive and negative emotions of words by type



```
##      movie totalSentiment
## 1:   Frog           498
## 2: Sleeping           457
## 3: Cinderella        362
## 4:   Beauty           355
## 5: Aladdin           304
## 6:   Mulan            260
## 7: Mermaid           214
## 8:   Moana            213
## 9:   Snow            192
## 10: Tangled           162
## 11: Pocahontas        160
## 12:   Brave           113
```

```
##      movie negative positive totalSentiment
## 1: Cinderella    106     271        165
## 2:   Frog       263     424        161
## 3: Sleeping     107     257        150
## 4: Tangled      185     324        139
## 5:   Beauty     264     402        138
## 6: Pocahontas   221     358        137
## 7: Aladdin      270     397        127
## 8:   Mulan      168     291        123
## 9:   Moana      157     258        101
## 10: Brave       109     208         99
```



|        |         |     |     |    |
|--------|---------|-----|-----|----|
| ## 11: | Mermaid | 208 | 258 | 50 |
| ## 12: | Snow    | 162 | 203 | 41 |

| ##     | movie      | anger | anticipation | disgust | fear | joy | negative | positive | sadness |
|--------|------------|-------|--------------|---------|------|-----|----------|----------|---------|
| ## 1:  | Aladdin    | 107   | 207          | 97      | 193  | 223 | 253      | 475      | 119     |
| ## 2:  | Beauty     | 118   | 208          | 53      | 191  | 191 | 289      | 381      | 122     |
| ## 3:  | Brave      | 84    | 142          | 53      | 163  | 101 | 219      | 235      | 70      |
| ## 4:  | Cinderella | 36    | 221          | 36      | 54   | 208 | 153      | 373      | 81      |
| ## 5:  | Frog       | 94    | 271          | 101     | 131  | 320 | 251      | 542      | 115     |
| ## 6:  | Mermaid    | 70    | 177          | 57      | 119  | 174 | 178      | 334      | 120     |
| ## 7:  | Moana      | 76    | 152          | 30      | 107  | 132 | 170      | 260      | 78      |
| ## 8:  | Mulan      | 90    | 157          | 53      | 146  | 125 | 204      | 318      | 75      |
| ## 9:  | Pocahontas | 106   | 176          | 78      | 152  | 153 | 238      | 366      | 103     |
| ## 10: | Sleeping   | 47    | 187          | 40      | 83   | 227 | 119      | 425      | 56      |
| ## 11: | Snow       | 66    | 115          | 57      | 99   | 133 | 154      | 195      | 86      |
| ## 12: | Tangled    | 77    | 203          | 72      | 96   | 203 | 225      | 343      | 146     |

| ##     | surprise | trust | totalSentiment | totalFeeling |
|--------|----------|-------|----------------|--------------|
| ## 1:  | 128      | 238   | 222            | 1312         |
| ## 2:  | 126      | 205   | 92             | 1214         |
| ## 3:  | 62       | 126   | 16             | 801          |
| ## 4:  | 92       | 220   | 220            | 948          |
| ## 5:  | 136      | 253   | 291            | 1421         |
| ## 6:  | 107      | 133   | 156            | 957          |
| ## 7:  | 56       | 128   | 90             | 759          |
| ## 8:  | 81       | 206   | 114            | 933          |
| ## 9:  | 104      | 216   | 128            | 1088         |
| ## 10: | 124      | 168   | 306            | 932          |
| ## 11: | 69       | 106   | 41             | 731          |
| ## 12: | 72       | 203   | 118            | 1072         |

