



SENTIMENT ANALYSIS IN R: THE TIDY WAY

Ranking pop songs through the years

Julia Silge

Data Scientist at Stack Overflow



Lyrics of pop songs

`song_lyrics`

- rank, the rank a song achieved on the Billboard Year-End Hot 100
- song, the song's title
- artist, the artist who recorded the song
- year, the year the song reached the given rank on the Billboard chart
- lyrics, the lyrics of the song.

Lyrics from the Billboard Year-End Hot 100, [sourced by Kaylin Walker](#)



Tidying song lyrics

```
> tidy_lyrics %>%  
+   count(word, sort = TRUE)
```

```
# A tibble: 42,156 x 2
```

	word	n
	<chr>	<int>
1	you	64606
2	i	56472
3	the	53451
4	to	35752
5	and	32555
6	me	31170
7	a	29282
8	it	25688
9	my	22821
10	in	18553

```
# ... with 42,146 more rows
```



Sentiment analysis of pop songs

```
> lyric_sentiment %>%  
+   count(song, sentiment, total_words)  
  
# A tibble: 39,564 x 4  
  song      sentiment total_words     n  
  <chr>      <chr>      <int> <int>  
1 0 to 100 the catch up      anger      894     29  
2 0 to 100 the catch up anticipation 894     14  
3 0 to 100 the catch up      disgust    894     33  
4 0 to 100 the catch up      fear       894      9  
5 0 to 100 the catch up      joy        894      5  
6 0 to 100 the catch up      negative   894     47  
7 0 to 100 the catch up      positive   894     34  
8 0 to 100 the catch up      sadness    894     12  
9 0 to 100 the catch up      surprise   894      8  
10 0 to 100 the catch up      trust      894     29  
# ... with 39,554 more rows
```



SENTIMENT ANALYSIS IN R: THE TIDY WAY

Let's practice!



SENTIMENT ANALYSIS IN R: THE TIDY WAY

Connecting sentiment to other quantities

Julia Silge

Data Scientist at Stack Overflow



Sentiment and...

- how far a song reached on the Billboard chart?
- when the song was released?

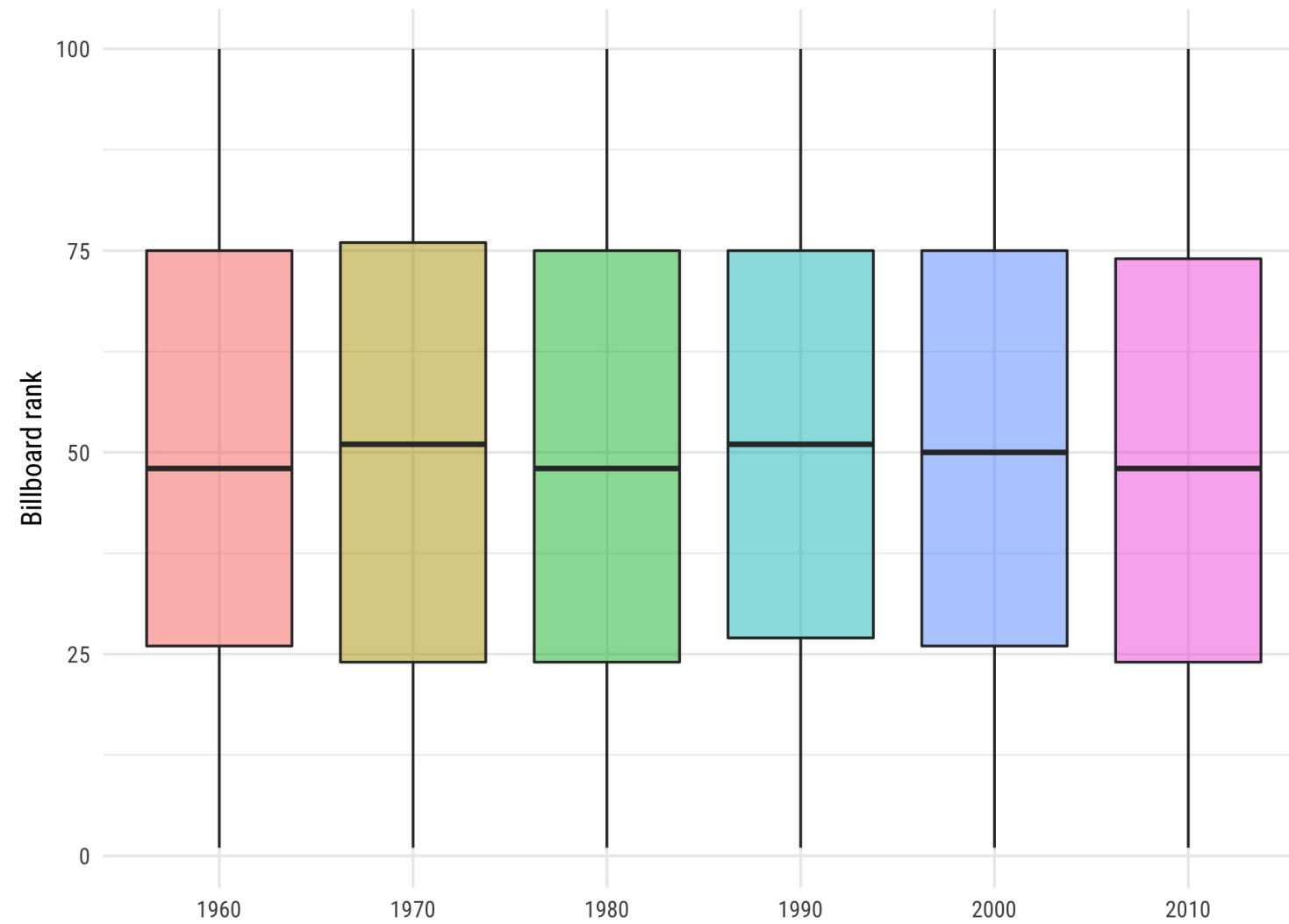


Sentiment and rank

```
> lyric_sentiment %>%  
+   filter(sentiment == "positive") %>%  
+   count(song, rank, total_words)  
  
# A tibble: 4,777 x 4  
  song      rank total_words     n  
  <chr> <int>      <int> <int>  
1 0 to 100 the catch up    97    894    34  
2 1 2 3 4 sumpin new    40    670    18  
3 1 2 3 red light    48    145     9  
4 1 2 step         5    437    20  
5 100 pure love    46    590    11  
6 100 pure love    82    590    11  
7 100 years    77    257     4  
8 123          62    220    15  
9 18 and life    61    285     9  
10 19 somethin    84    281     6  
# ... with 4,767 more rows
```




Exploring with boxplots





SENTIMENT ANALYSIS IN R: THE TIDY WAY

Let's practice!



SENTIMENT ANALYSIS IN R: THE TIDY WAY

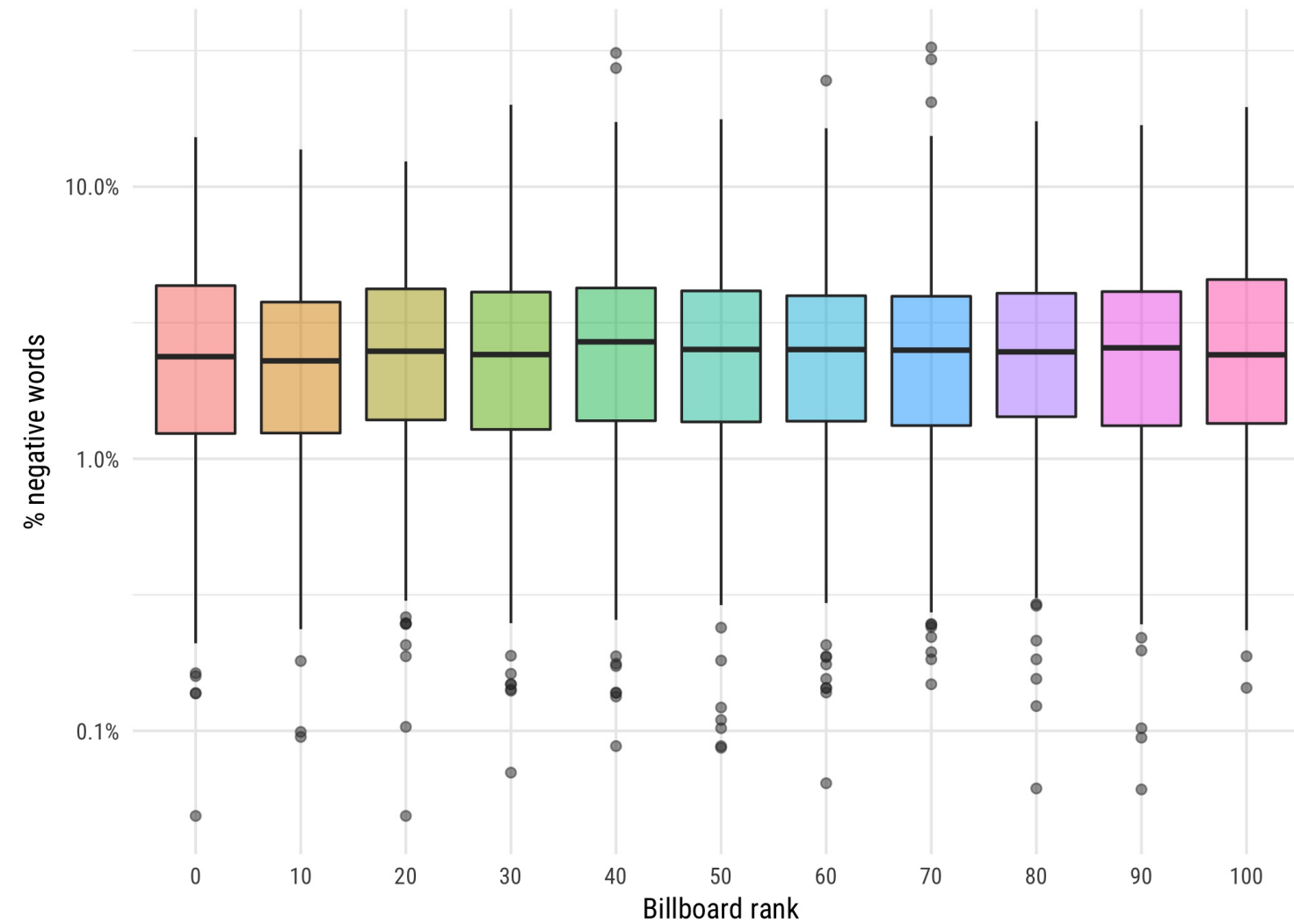
Moving from rank to year

Julia Silge

Data Scientist at Stack Overflow



Sentiment and rank





Pop songs over time

```
> lyric_sentiment %>%  
+   filter(sentiment == "positive") %>%  
+   count(song, year, total_words)  
  
# A tibble: 4,772 x 4  
  song      year total_words     n  
  <chr> <int>      <int> <int>  
1 0 to 100 the catch up 2014      894    34  
2 1 2 3 4 sumpin new 1996      670    18  
3 1 2 3 red light 1968      145     9  
4 1 2 step 2005      437    20  
5 100 pure love 1994      590    11  
6 100 pure love 1995      590    11  
7 100 years 2004      257     4  
8 123 1988      220    15  
9 18 and life 1989      285     9  
10 19 somethin 2003      281     6  
# ... with 4,762 more rows
```



Pop songs over time

- Define new columns using `mutate()`
- Visualize using `geom_boxplot()`



Modeling sentiment

```
> sentiment_model <- lm(percent ~ year, data = sentiment_by_year)
> summary(sentiment_model)
```



SENTIMENT ANALYSIS IN R: THE TIDY WAY

Let's practice!



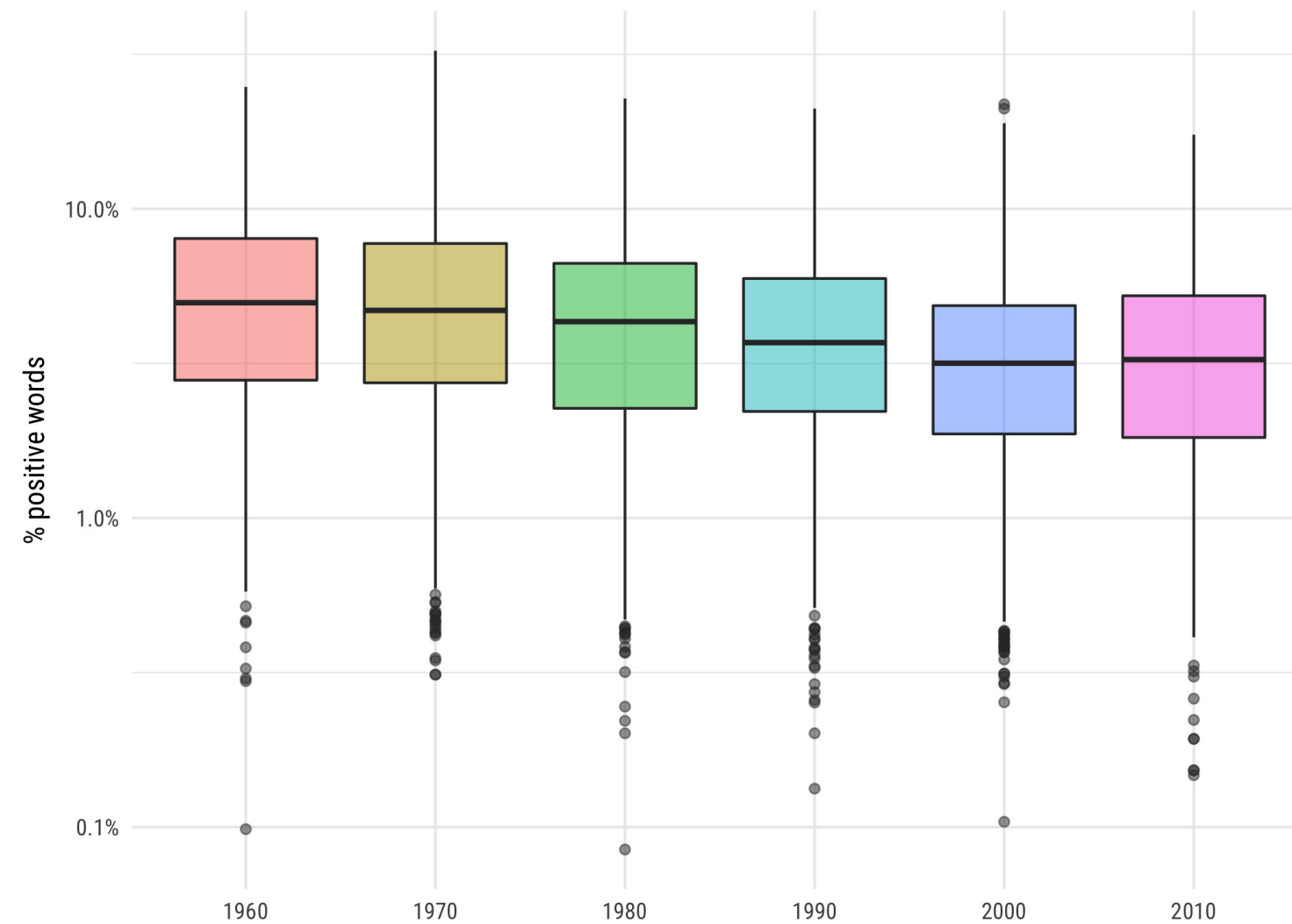
SENTIMENT ANALYSIS IN R: THE TIDY WAY

You made it!

Julia Silge

Data Scientist at Stack Overflow

Positive sentiment and year





Diverse texts, powerful techniques

- Social media text from Twitter
- Classic narrative text by Shakespeare
- TV news text sourced from closed captioning
- Lyrics from pop songs



Tidy text

Tidy data principles makes text mining easier and more effective

<http://tidytextmining.com/>



SENTIMENT ANALYSIS IN R: THE TIDY WAY

Thanks!