Markdown presentation

Loris Jeitziner, Olivia Fischer and Nieves Schwab

11/24/2020

Emotional Arcs in TED Talks

What does a great commercial, a sales pitch and a scientific presentation have in common?

They all try to tell (or sell) a compelling story.

Stories are appealing to us, as they use a tight and logical structure, they help us organize information. It has a logical progression, the classic story structure always looks the same pattern, first the context is set, and the hero is introduced. The hero has a problem or a challenge he has to solve. He struggles, then struggles more and finally he resolves the crises. Happy end.

This ups and downs in the stories are an important element, they capture the audience's interest and take them on an emotional journey. Kurt Vonnegut, proposed that there exist six "universal shapes" of stories, or emotional arcs, that follow specific pattern in stories.

We propose that emotion plays an important role in storytelling. And we wanted to analyze to which extent. To this purpose we have been looking at data from the masters of storytelling, the speakers of TED Talks.

According to Wikipedia: TED conferences (Technology, Entertainment, Design) is a media organization that posts talks online for free distribution, under the slogan "ideas worth spreading". They address a wide range of topics within the research and practice of science and culture, often thorugh storytelling. The speakers are given a maximum of 18 minutes to present ideas in the most innovative and engaging way as they can.

The dataset that we analyze contains 2,386 TED Talks from the years 1972 to 2017. It contains variables like the speaker, headline of the talk, duration, year of publication and... the number of views!

In the last two weeks we have been analyzing this dataset in a data-driven manner, exploring if emotion/valence has an influence on the number of views a talk has. For this purpose we have defined the variable of views as our DV, defining talks with more views are more successful.

Additional variables contained in the dataset stem from LIWC. They are, for example: word count, tone, number of pos (e.g. ppron), affect,...

#number of emotional words in the headline or description of the video

Data preparation

```
#load data
data = read_csv("data/ted.csv", locale = locale(encoding = "UTF-8"))
```

```
#exclude NA talks
data = data %>% filter(!transcript=="N/A", .preserve = TRUE)

#rename columns with strange names
data = data %>% rename(views=views_as_of_06162017)

#Selecting all the talks that were filmed from 2000 to 2007 and divide views by 1000.
talks <- data %>% filter( year_filmed >= 2000) %>%
                          mutate(views_thousand = views/1000)
```

Let's look at the data:

Actually, there are 45 TED Talks that contain the word "story" in their title. Coincidence? We don't think so! Stories are everywhere, and the emotions conveyed in a story are the essence of a talk.

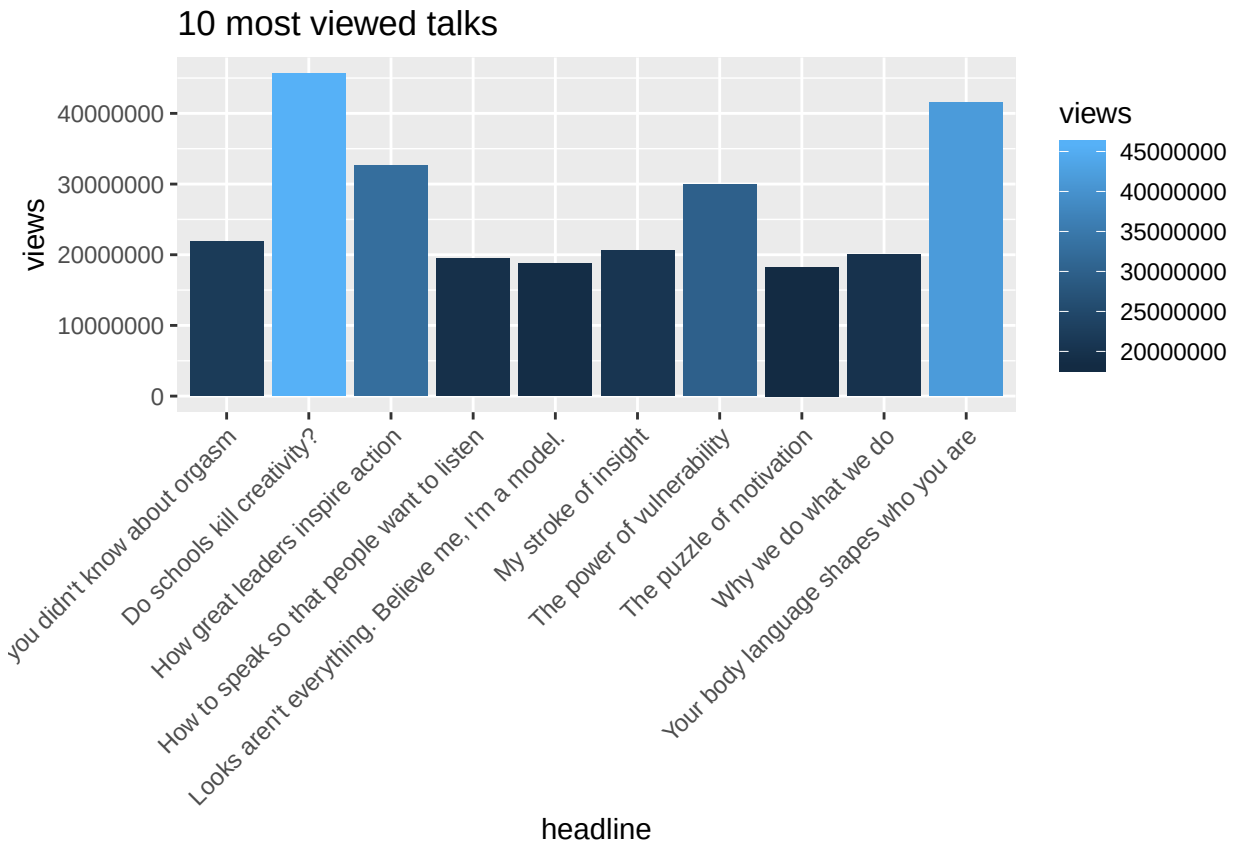
```
storytelling <- talks %>% select("headline", "id") %>%
  filter(str_detect(headline, "story"))
```

What are the 25 most popular/viewed talks? Ken Robinson's talk on "Do Schools Kill Creativity?" is the most popular talk of all time with impressive 45'622'906 views. There are 6 talks that have crossed the 20 million mark.

```
summary(data$views)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    138915    756080   1125707   1695917   1702883  45622906
```

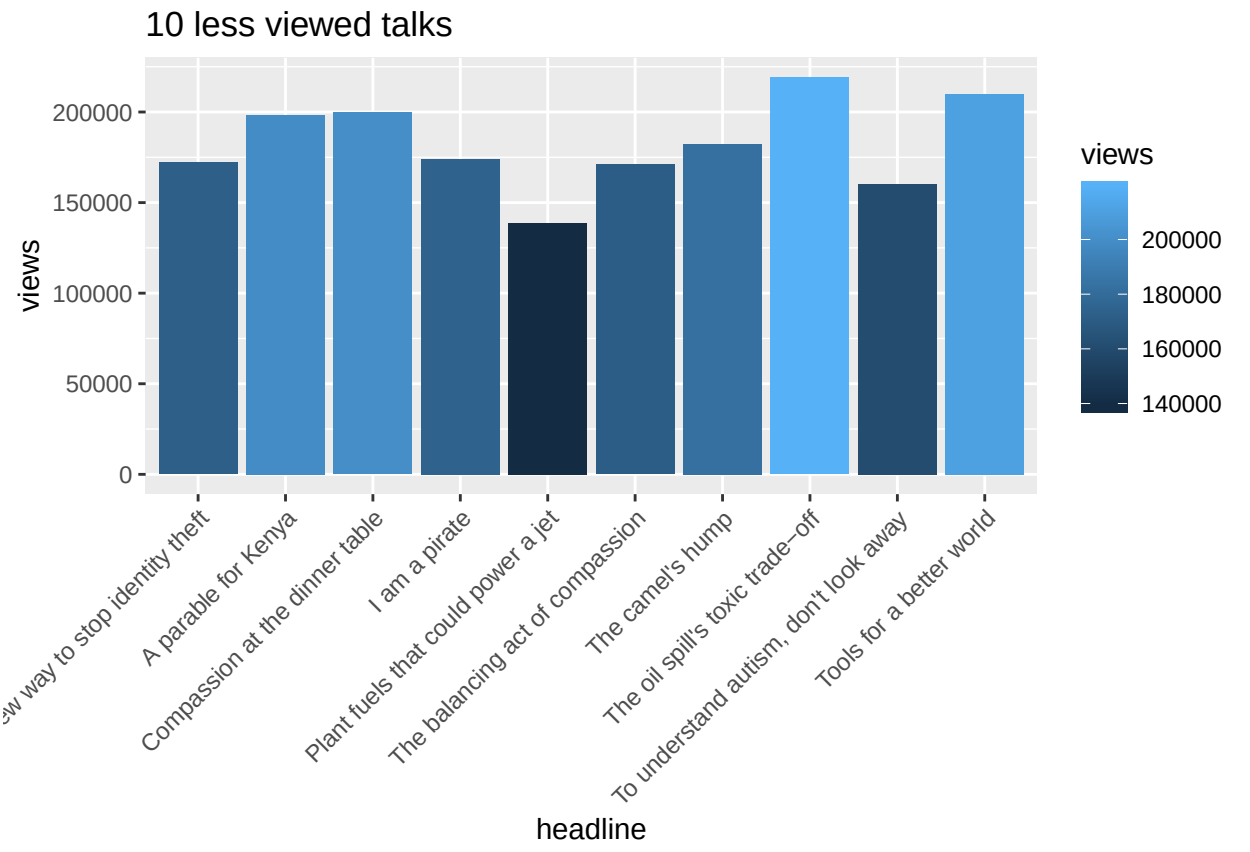
```
top_10talks<- talks %>% select("id", "headline", "speaker", "year_filmed", "duration", "views") %>% arra

tilt_theme <- theme(axis.text.x=element_text(angle=45, hjust=1))

top <- ggplot(top_10talks, aes(headline, views, fill= views)) +
  geom_col() +labs(title= "10 most viewed talks") + tilt_theme

top
```



The least viewed talk "Plant fuels that could power a jet" has 138'915 views.

```
last_10talks<- talks %>% select("id", "headline", "speaker", "year_filmed", "duration", "views") %>% ar
```

```
last<- ggplot(last_10talks, aes(headline, views, fill= views)) +
  geom_col() +labs(title= "10 less viewed talks") + tilt_theme
last
```



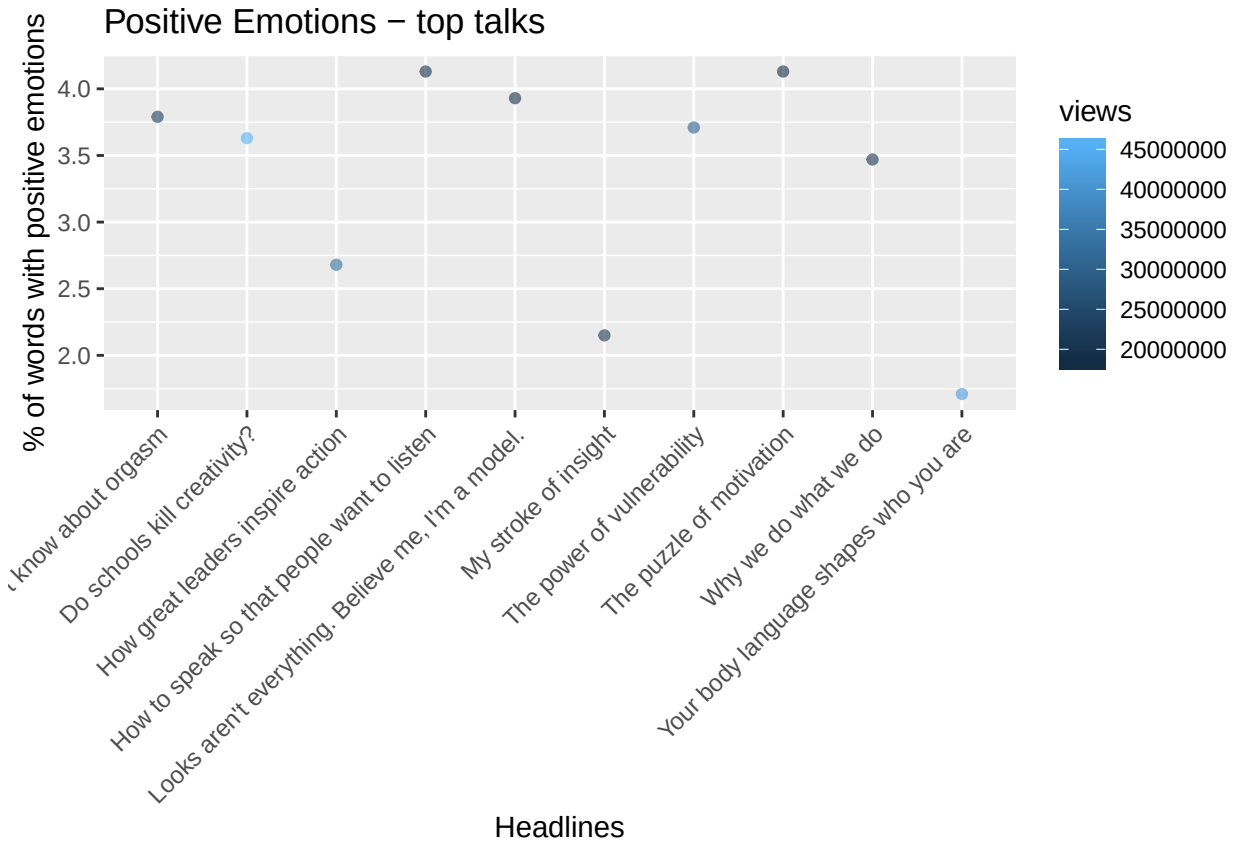
Does emotion matter? Emotions in TED Talks

Select the 30 most viewed and the 30 less viewed talks

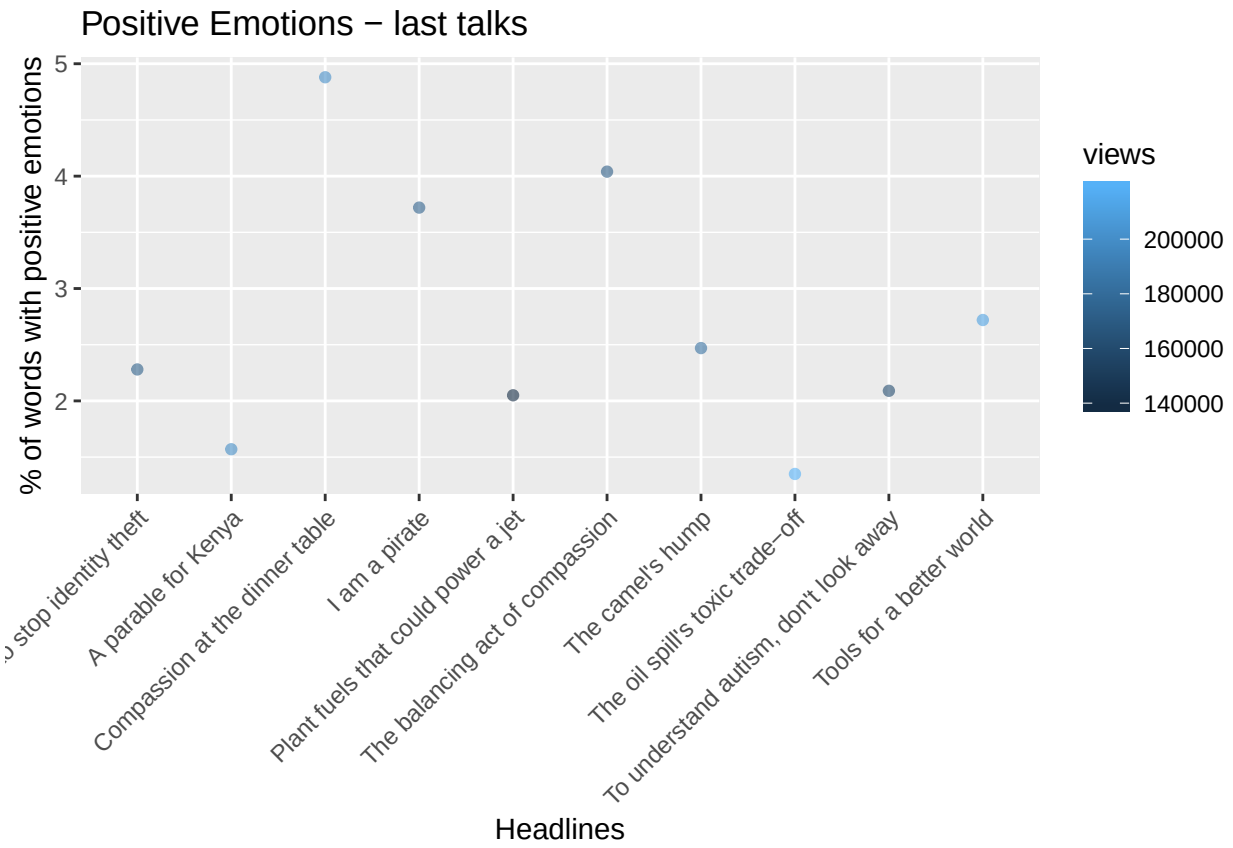
```
viewed_talks <- talks %>% arrange(desc(talks$views_thousand)) %>% head(10)

not_viewed_talks <- talks %>% arrange(talks$views_thousand) %>% head(10)
```

Positive Emotions (love, nice, sweet) –> The value of positive emotion of a talk is calculated as a percentage of all the words in that talk. For example, 4.2 % of all words in the talk were positive emotions.

```
posemo_toptalks<- ggplot(viewed_talks , aes(headline, posemo, color=views))+ geom_point(alpha=0.6)+  lab
posemo_toptalks
```

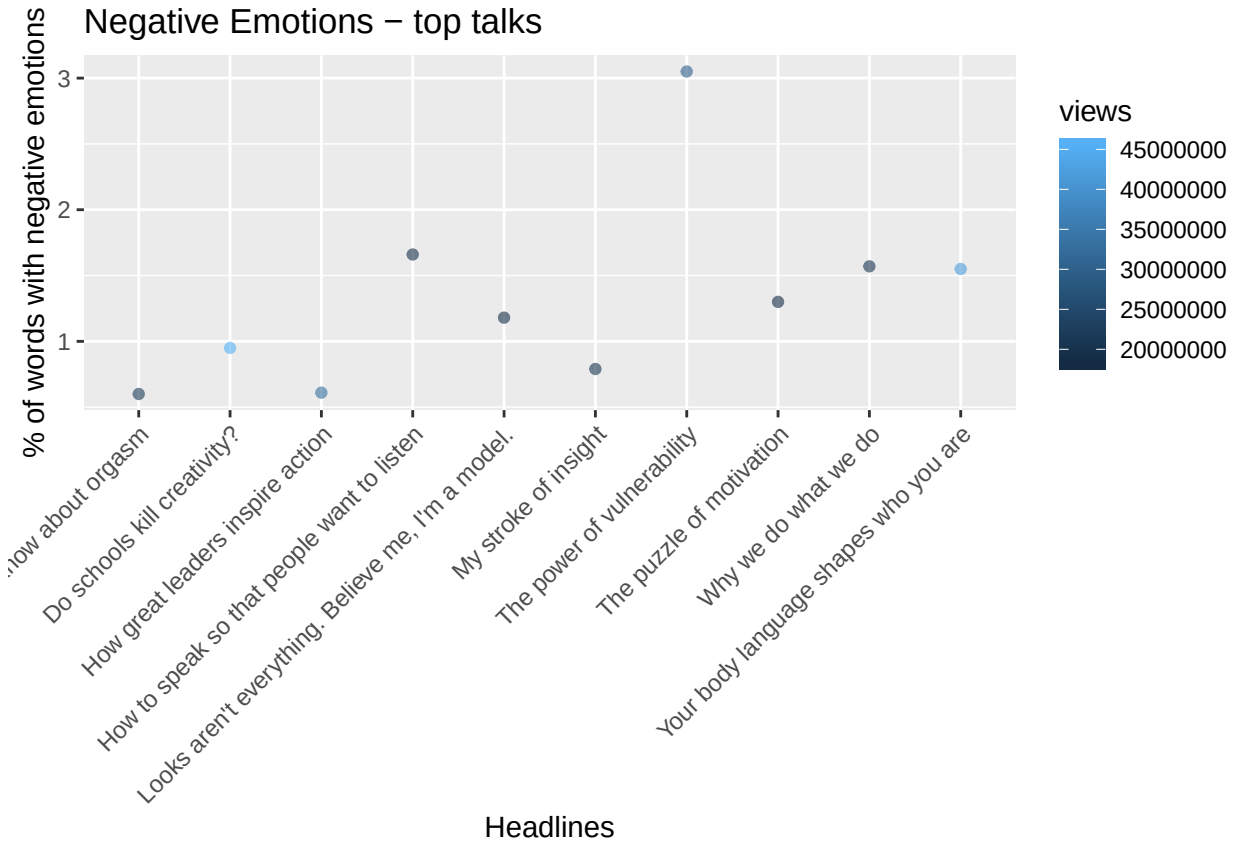



```
posemo_lowtalks <- ggplot(not_viewed_talks, aes(headline, posemo, color=views))+ geom_point(alpha=0.6)+
posemo_lowtalks
```

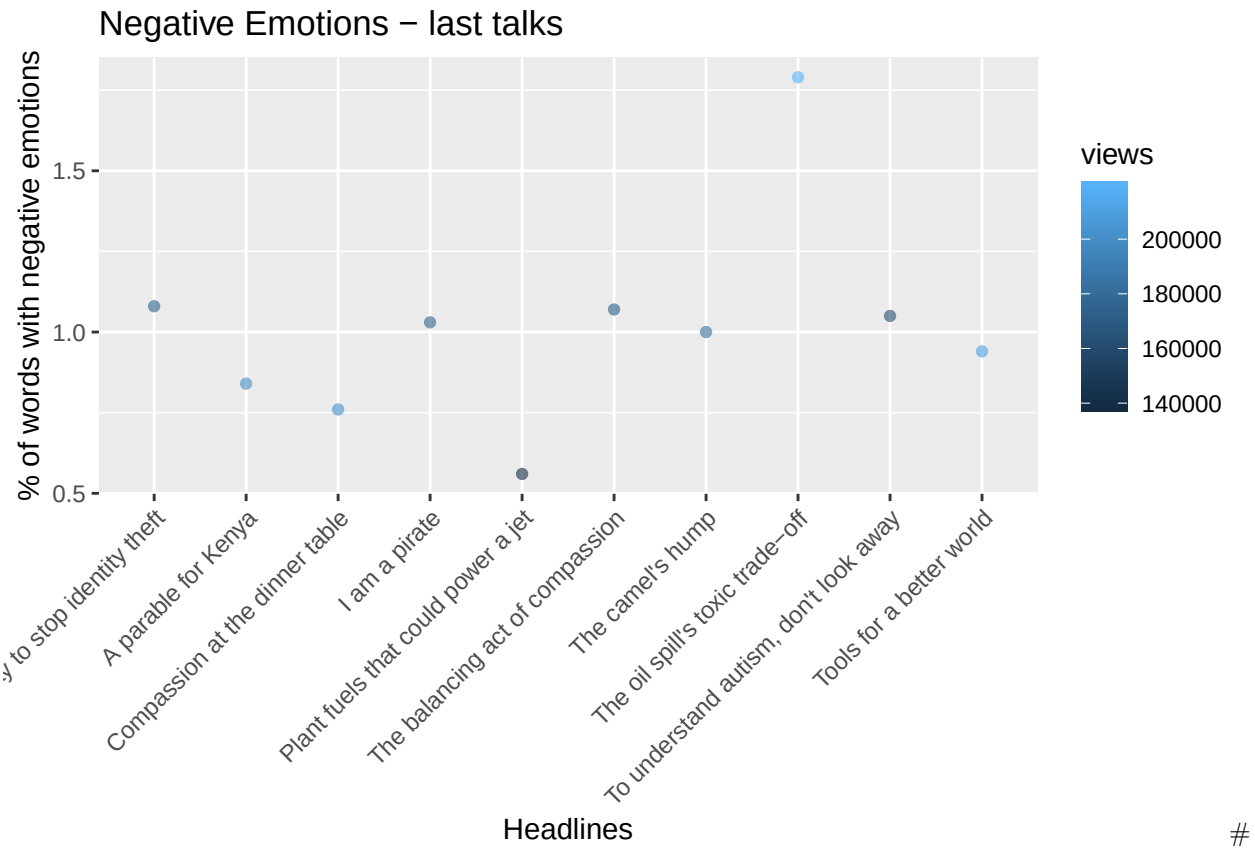


Negative Emotions (hust, ugly, nasty)

```
negemo_toptalks<- ggplot(viewed_talks , aes(headline, negemo, color=views))+ geom_point(alpha=0.6)+  lab
negemo_toptalks
```




```
negemo_lowtalks <- ggplot(not_viewed_talks, aes(headline, negemo, color=views))+ geom_point(alpha=0.6)+
negemo_lowtalks
```



#

Emotional Arcs in TED Talks

We are interested in discovering if there are distinct emotional arcs and if they can predict the number of views a TED Talk will have. To plot the sentiment arcs first we have to tokenize the transcripts of the talks and add the sentiment values.

Tokenize the data and add sentiment

```
#tokenize the transcripts
token_tbl = unnest_tokens(data, input = "transcript", token = "words", to_lower = TRUE, output = "word")

#relative position of the word for the sentiment analysis -> arc
token_tbl = token_tbl %>% group_by(id) %>% mutate(
  pos = 1:n(),
  rel_pos = (pos-1)/max(pos-1)
)

#load the sentiment word from the "afinn" repository
afinn = get_sentiments("afinn")

#join the tokens with the sentiment words
token_tbl <- inner_join(y = afinn,x = token_tbl)
```

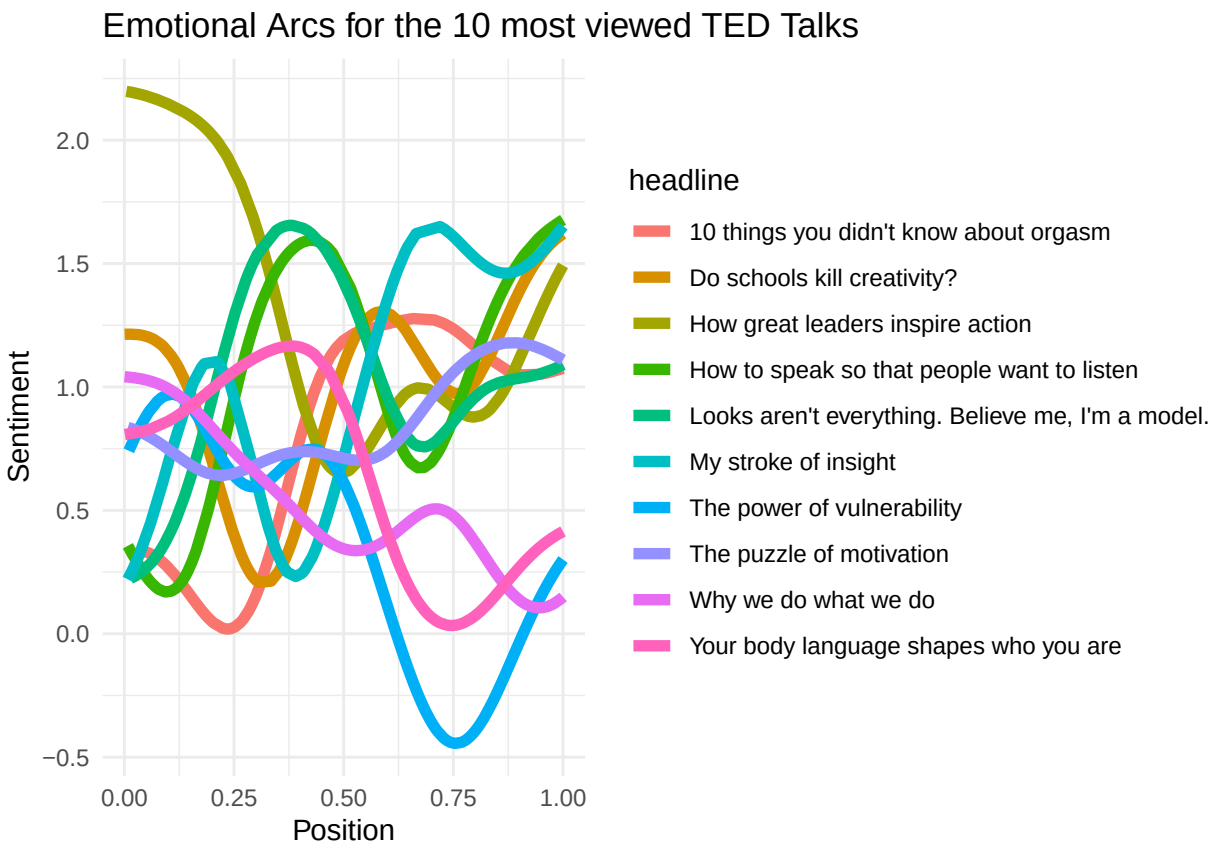
Smooth sentiment arcs

```
# smoothing function for smoothing the sentiment values relative to the position
smooth = function(pos, value){
  sm = sapply(pos, function(x) {
    weights = dnorm(pos, x, max(pos) / 10)
    sum(value * (weights / sum(weights)))
```

```
    })
  }

# define the sentiment values as sent_values per talk
token_tbl = token_tbl %>% group_by(id) %>% mutate(
  sent_values = smooth(pos, value)
)
```

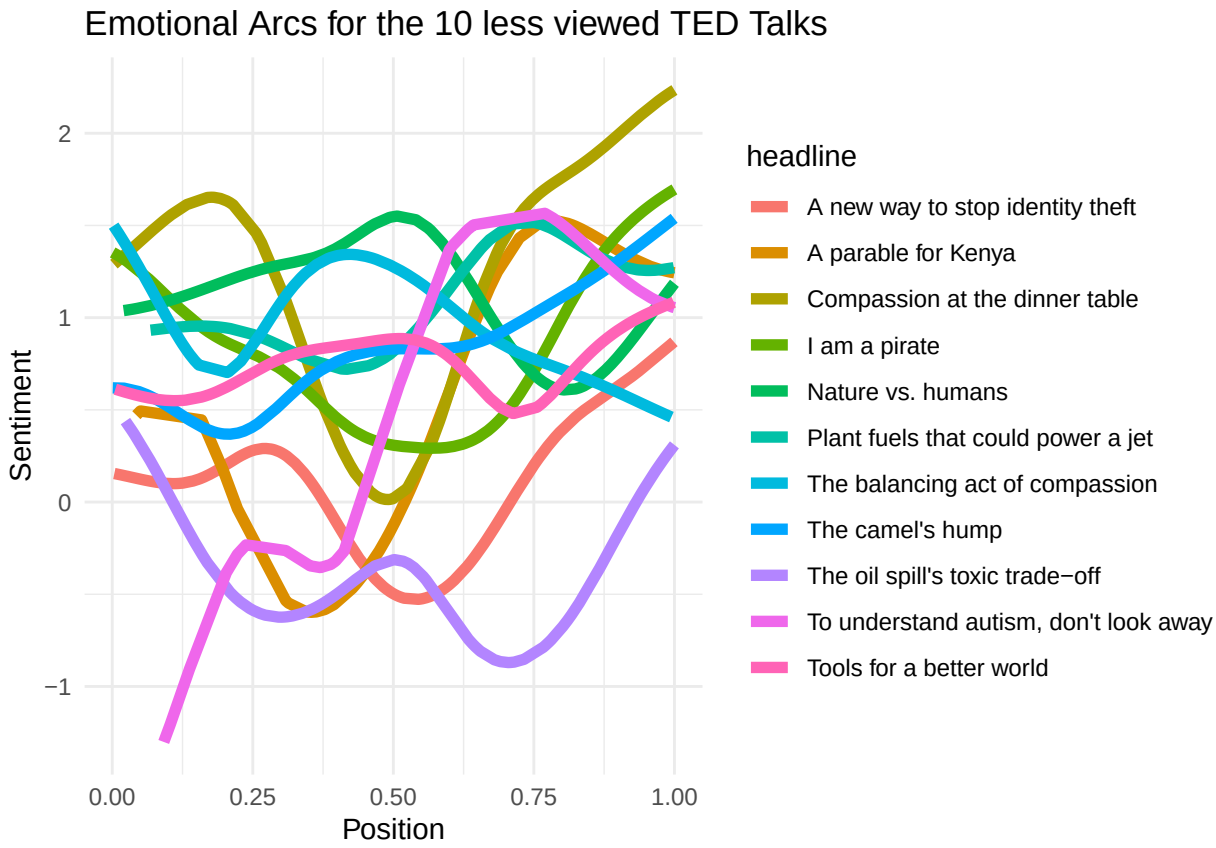
Sentiment Arcs for the most viewed Talks

```
# plot sentiment arcs not z transformed to compare, color by title
sent_top_views <- token_tbl %>% filter(views >= 18253944) %>% ggplot(aes(rel_pos, sent_values, color= he
sent_top_views
```



Sentiment Arcs for the less viewed talks

```
sent_last_views <- token_tbl %>% filter(views <= 218993) %>% ggplot(aes(rel_pos, sent_values, color= hea
sent_last_views
```


Emotional Arcs for the 10 less viewed TED Talks
Sentiment
2
1
0
−1
0.00
0.25
0.50
0.75
1.00
Position
headline
A new way to stop identity theft
A parable for Kenya
Compassion at the dinner table
I am a pirate
Nature vs. humans
Plant fuels that could power a jet
The balancing act of compassion
The camel's hump
The oil spill's toxic trade−off
To understand autism, don't look away
Tools for a better world