

Sentiment analysis of top 10 songs of all times

1902224

11 Nov 2019

Assignment

Figure out a use-case (like comparing sentiments of articles on different news sites about current topics) and publish a post on Medium.com about what you found.

Setup

```
#Have your SelectorGadget on Google Chrome: https://selectorgadget.com/
#Login to AWS account and generate access keys

#Loading packages
library(rvest)
library(aws.s3)
library(aws.comprehend)
library(dplyr)
library(ggplot2)
library(gridExtra)

# accessKeys.csv was downloaded from AWS
keyTable <- read.csv("accessKeys.csv", header = T)
AWS_ACCESS_KEY_ID <- as.character(keyTable$Access.key.ID)
AWS_SECRET_ACCESS_KEY <- as.character(keyTable$Secret.access.key)

Sys.setenv("AWS_ACCESS_KEY_ID" = AWS_ACCESS_KEY_ID,
           "AWS_SECRET_ACCESS_KEY" = AWS_SECRET_ACCESS_KEY,
           "AWS_DEFAULT_REGION" = "eu-west-1")
```

Scrape websites for songtext lyrics

Create table with URL pointing to the songtexts:

```
url_table <- data.frame(
  RANK = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25),
  URL = c(
    'https://www.songtexte.com/songtext/bob-dylan/like-a-rolling-stone-2bd5e8fa.html',
    'https://www.songtexte.com/songtext/the-rolling-stones/i-cant-get-no-satisfaction-5b979348.html',
    'https://www.songtexte.com/songtext/john-lennon-and-yoko-ono/imagine-5bb44f44.html',
    'https://www.songtexte.com/songtext/marvin-gaye/whats-going-on-43d64313.html',
    'https://www.golyr.de/aretha-franklin/songtext-respect-6802.html',
    'https://www.lyrix.at/t/beach-boys-good-vibrations-51b',
    'https://www.golyr.de/chuck-berry/songtext-johnny-b-goode-11504.html',
    'https://www.songtexte.com/songtext/the-beatles/hey-jude-73d292a1.html',
    'https://www.songtexte.com/songtext/nirvana/smells-like-teen-spirit-73d0c625.html',
    'https://www.songtexte.com/songtext/ray-charles/whatd-i-say-13da4991.html',
    'https://www.golyr.de/the-who/songtext-my-generation-193770.html',
    'https://www.songtexte.com/songtext/sam-cooke/a-change-is-gonna-come-2bd73c6a.html',
```

```

'https://www.songtexte.com/songtext/the-beatles/yesterday-6bd292f6.html',
'https://www.songtexte.com/songtext/bob-dylan/blowin-in-the-wind-2bd5e8d2.html',
'https://www.songtexte.com/songtext/the-clash/london-calling-53d23fd5.html',
'https://www.songtexte.com/songtext/the-beatles/i-want-to-hold-your-hand-6bd292d6.html',
'https://www.golyr.de/the-jimi-hendrix-experience/songtext-purple-haze-bbc-sessions-1538374.html',
'https://www.golyr.de/chuck-berry/songtext-maybellene-hail-hail-rock-n-roll-sountrack-version-87899',
'https://www.songtexte.com/songtext/elvis-presley/hound-dog-23d34c2b.html',
'https://www.songtexte.com/songtext/the-beatles/let-it-be-73d2920d.html',
'https://www.lyrix.at/t/bruce-springsteen-born-to-run-66e',
'https://www.songtexte.com/songtext/the-ronettes/be-my-baby-3d0b9af.html',
'https://www.songtexte.com/songtext/the-beatles/in-my-life-1bd28d64.html',
'https://www.songtexte.com/songtext/the-impressions/people-get-ready-5bd0fbbc.html',
'https://www.lyrix.at/t/david-bowie-god-only-knows-214'
),
LYRICS = NA,
stringsAsFactors=FALSE)

```

Function for scraping the songtexts:

```

get_songtext <- function(my_url) {

  xml_page <- read_html(my_url)

  if (length(grep("lyrix",url_table[i,2]))==1) {
    songtext <- xml_page %>%
      html_nodes('.songtext') %>%
      html_text()
  } else {
    songtext <- xml_page %>%
      html_nodes('#lyrics') %>%
      html_text()
  }
  return(songtext)
}

# get songtext for every song in the url_table
for (i in 1:nrow(url_table)) {

  url_table[i,3] <- get_songtext(url_table[i,2])
}

```

Sentiment analysis

Check language of every song and translate, if necessary, to English:

```

# simple language detection
url_table$LANGUAGE <- NA

for (i in 1:nrow(url_table)) {
  if (!is.na(url_table[i,3])) {
    url_table[i,4] <- detect_language(url_table[i,3])[1,1]
  } }

table(url_table$LANGUAGE)

```

```
##
## en
## 25

Check sentiment for eachs song:

sentiment_table <- data.frame()

for (i in 1:nrow(url_table)) {
  if (!is.na(url_table[i,3])) {
    sentiment_table <- rbind(sentiment_table, detect_sentiment(url_table[i,3]))
  }
  sentiment_table[i,1] <- i
}

# join tables
top25songs <- full_join(url_table, sentiment_table, by = c("RANK"="Index"))

# save file
write.csv(top25songs, "top25songs.csv")
```

Plot outcomes

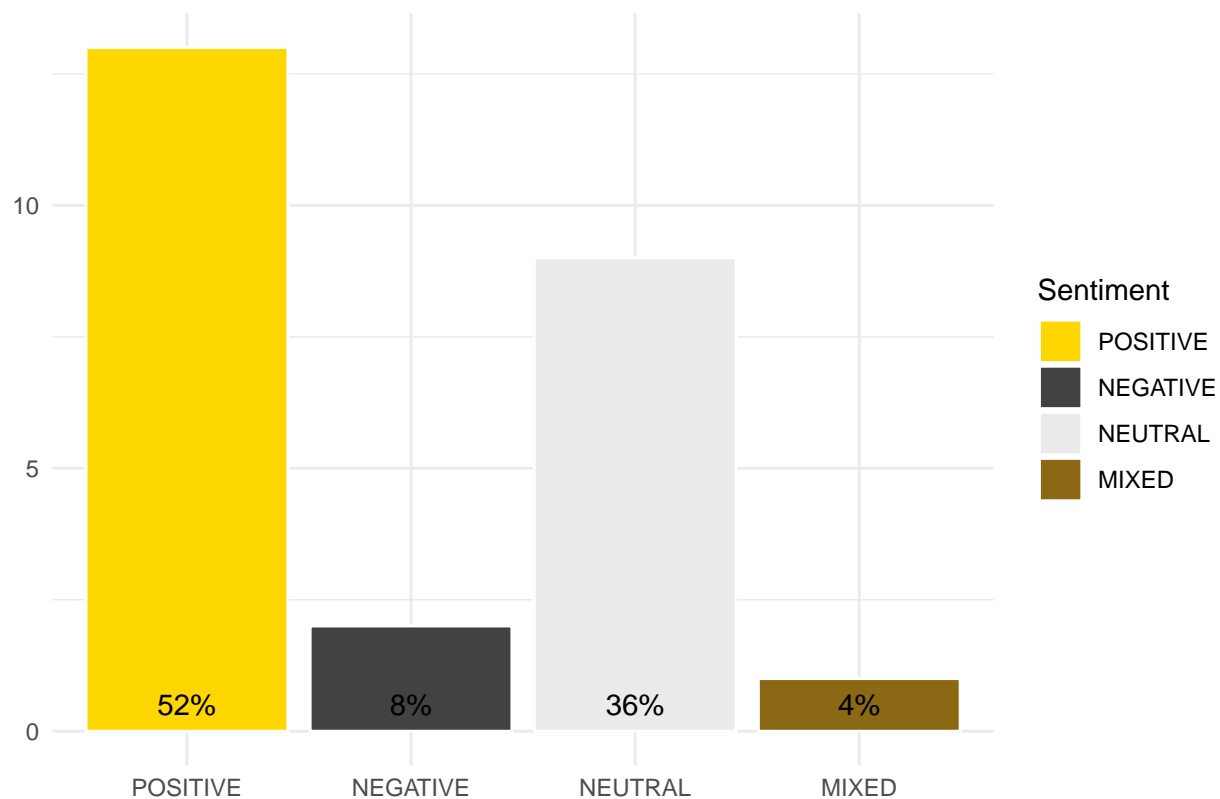
```
theme_set(
  theme_minimal() +
  theme(legend.position = "right")
)

# Summarize to get counts and percentages
sent.pct = top25songs %>% group_by(Sentiment) %>%
  summarise(count=n()) %>%
  mutate(pct=count/sum(count))

sent.plot <-
ggplot(top25songs, aes(Sentiment, fill=Sentiment)) +
geom_histogram(stat="count", colour="white") +
  geom_text(data = sent.pct, aes(label=paste0(round(pct*100,1),"%"),y=0.5)) +
  scale_fill_manual(values= c("gold", "gray26", "gray92", "goldenrod4")) +
  labs(x = NULL, y = NULL, title = "Sentiment Analysis for the top 10 songs of all times")

sent.plot
```

Sentiment Analysis for the top 10 songs of all times



```
# save plot
ggsave("sent.plot.png")

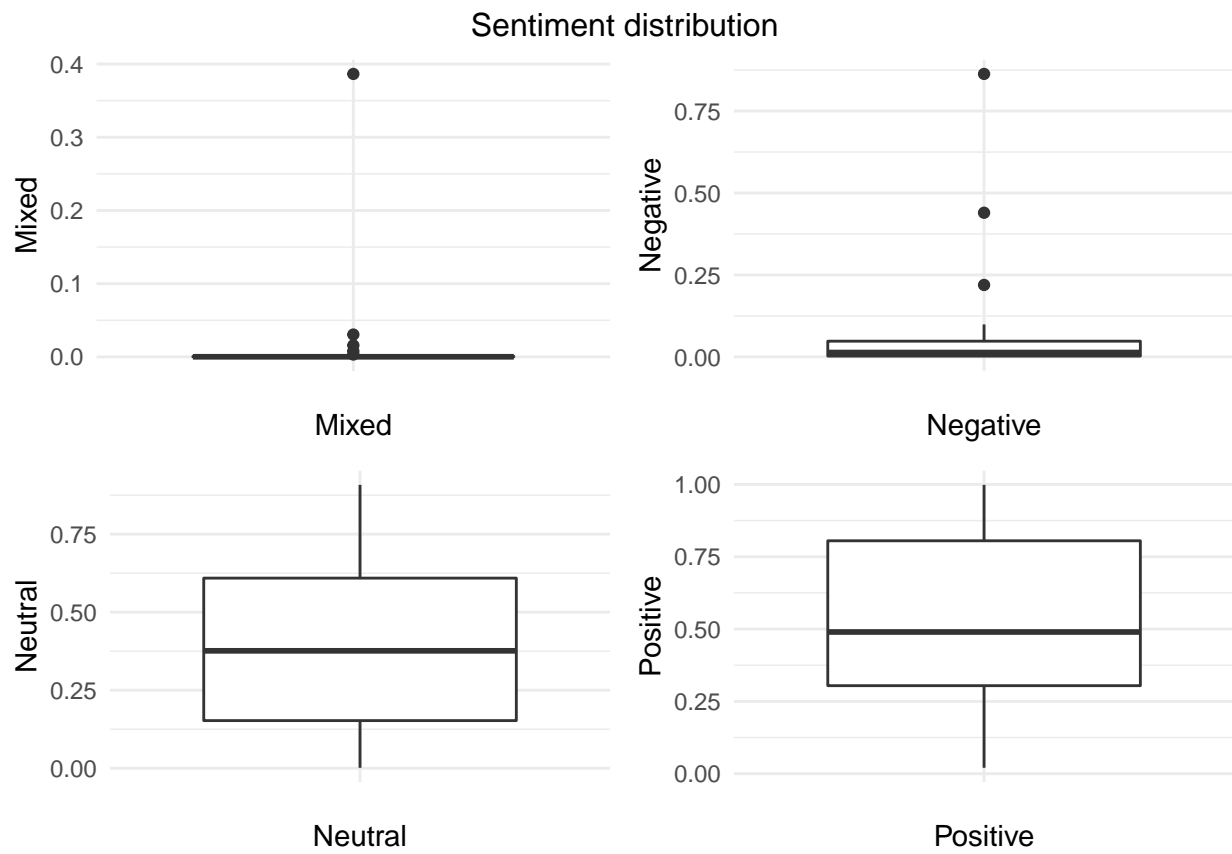
plot1 <- ggplot(top25songs, aes(x="", y=Mixed)) +
  geom_boxplot() +
  labs(x = "Mixed")

plot2 <- ggplot(top25songs, aes(x="", y=Negative)) +
  geom_boxplot() +
  labs(x = "Negative")

plot3 <- ggplot(top25songs, aes(x="", y=Neutral)) +
  geom_boxplot() +
  labs(x = "Neutral")

plot4 <- ggplot(top25songs, aes(x="", y=Positive)) +
  geom_boxplot() +
  labs(x = "Positive")

grid.arrange(plot1, plot2, plot3, plot4, ncol=2, top = "Sentiment distribution")
```



```
# save plot
ggsave("sent.plot.2.png",
  arrangeGrob(plot1, plot2, plot3, plot4, ncol=2, top = "Sentiment distribution"))
```