Silent Spring Senitment Analysis

Shale

4/13/2022

Load & Clean Data

```r
my_files <- list.files(pattern = "news.DOCX", path = here("data"),
                       full.names = TRUE, recursive = TRUE, ignore.case = TRUE)

dat <- lnt_read(my_files) #Object of class 'LNT output'
```

```
## Creating LNToutput from 1 file...
```

```
##   ...files loaded [0.26 secs]
```

```
##   ...articles split [0.30 secs]
```

```
##   ...lengths extracted [0.30 secs]
```

```
##   ...headlines extracted [0.30 secs]
```

```
##   ...newspapers extracted [0.31 secs]
```

```
##   ...dates extracted [0.34 secs]
```

```
##   ...authors extracted [0.34 secs]
```

```
##   ...sections extracted [0.35 secs]
```

```
##   ...editions extracted [0.35 secs]
```

```
## Warning in lnt_asDate(date.v, ...): More than one language was detected. The
## most likely one was chosen (English 95%)
```

```
##   ...dates converted [0.36 secs]
```

```
##   ...metadata extracted [0.37 secs]
```

```
##   ...article texts extracted [0.37 secs]
```

```
##   ...superfluous whitespace removed [0.40 secs]
```

```
## Elapsed time: 0.40 secs
```

```r
# create separate dataframes for different types of analysis

meta_df <- dat@meta
articles_df <- dat@articles
paragraphs_df <- dat@paragraphs

# create more dataframes so it's easier to join text with metadata
heads <- data_frame(element_id = seq(1:length(meta_df$Headline)),
                    Date = meta_df$Date,
                    Headline = meta_df$Headline)
```

```
## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## Please use `tibble()` instead.
```

```r
paragraphs_dat <- data_frame(element_id = paragraphs_df$Art_ID,
                             Text       = paragraphs_df$Paragraph)

tidypars <- inner_join(heads, paragraphs_dat, by = "element_id")
```

Clean paragraphs

```r
# remove paragraphs that contain website links or are shorter than 21 characters
cleanpars <- tidypars %>%
  mutate(link = str_detect(tidypars$Text, "http", negate = TRUE)) %>%
  filter(link == TRUE & nchar(tidypars$Text) > 20)
```

Get sentiment for paragraphs using NRC Emotions

```r
e_sents <- get_sentiments("nrc") %>%
  filter(!sentiment %in% c("negative", "positive"))

emotions <- cleanpars %>%
  unnest_tokens(word, Text) %>%
  inner_join(e_sents) %>%
  na.omit() %>%
  group_by(Date, sentiment) %>%
  add_count()
```

```
## Joining, by = "word"
```

```r
emotions$Date <- emotions$Date %>% str_remove_all("-\\d\\d$")

Dprop <- emotions %>%
  group_by(Date) %>%
  summarise(prop = n / sum(n),
            ntot = sum(n))
```

```
## `summarise()` has grouped output by 'Date'. You can override using the `.groups`
## argument.
```

```r
date_prop <- emotions %>%
  cbind(Dprop$prop) %>%
  mutate(prop = ...8) %>%
  select(Date, sentiment, n)
```

```
## New names:
## * NA -> ...8
```

```r
sent_prop = xtabs(n ~ Date + sentiment, date_prop) %>%
  proportions(margin = 1) %>%
  as.data.frame()
```

Visualize results

```
ggplot(data = sent_prop, aes(x = Date, y = Freq, color = sentiment, group = sentiment)) +
# geom_point(alpha = .5, size = 2) +
  geom_smooth(se = FALSE) +
  scale_x_discrete(guide = guide_axis(angle = 45)) +
  labs(title = "Monthly proportion of sentiment in 100 articles (2017-2022) \n on the search term 'Siler
       y = "Frequency",
       x = "Date (yyyy-mm)") +
  scale_color_brewer(palette = "Paired")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

Monthly proportion of sentiment in 100 articles (2017–2022)
on the search term 'Silent Spring'
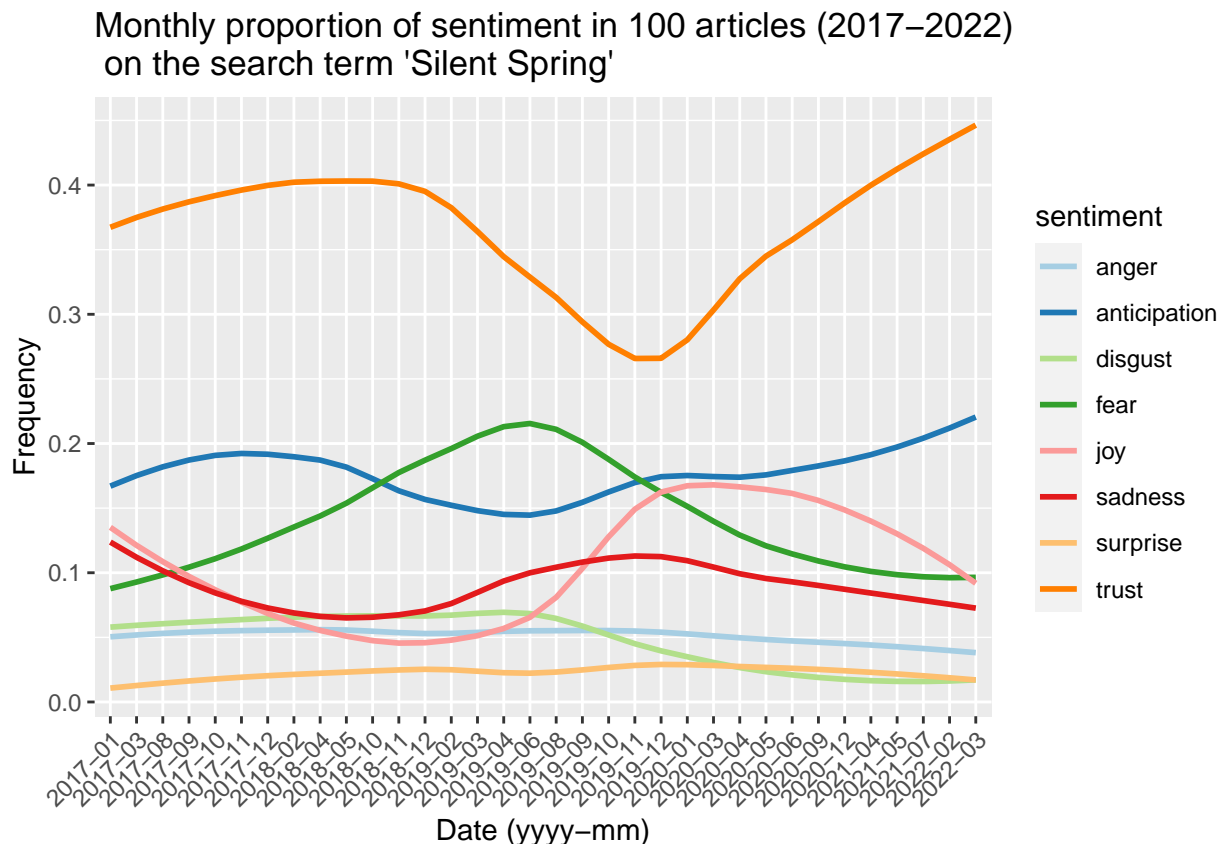
This sentiment analysis is based on a search for the term "Silent Spring" on the Nexus Uni database. The results are rather surprising, given the gloomy topic of that book; but of course the texts analyzed here are at best responses to Rachel Carson's famous book, and occasionally completely unrelated. Across the entire 5 year span from 2017 to 2022, trust was the most prevalent emotion detected, followed generally by anticipation and fear. Joy also became more prominent starting in late 2019. Consistently the least frequently coded emotions are surprise, anger, and disgust.

Interpreting these results in relation to expectations around *Silent Spring* yield no clear answers: why is trust so high? Perhaps people have high trust in the accuracy of reports around environmental advocacy that has stemmed from *Silent Spring*? Probably the most understandable result is the high levels of fear - makes sense, given the gravity of the topic.

Maybe it makes sense that surprise is low (after all, this isn't a new problem), but then why is anger so low? Is the high fear / low anger interaction a suggestion that people know this is a problem but don't feel like they have control over the massive agricultural corporations responsible for much of the problem? And what

the hell is going on with that sudden increase in joy in late 2019?

This basic analysis is not capable of answering these questions (there are a plethora of limitations both in the data itself and the quality of the text analysis methods), but hopefully by raising them I have given you more to think about in regards to contemporary public perspective on an historic environmental topic.

Notes on data presentation:

Instead of aggregating by day, it made more sense to aggregate by month because of the time span the articles are written over (about 100 articles over 5 years, which is about 1 article every 18 days). The graph above uses a smoothing function (different from Froelich et al.) to avoid jumpy / uninterpretable paths for each sentiment.