Sentiment Analysis: "Pollinators"

Peter Menzies

2022-04-19

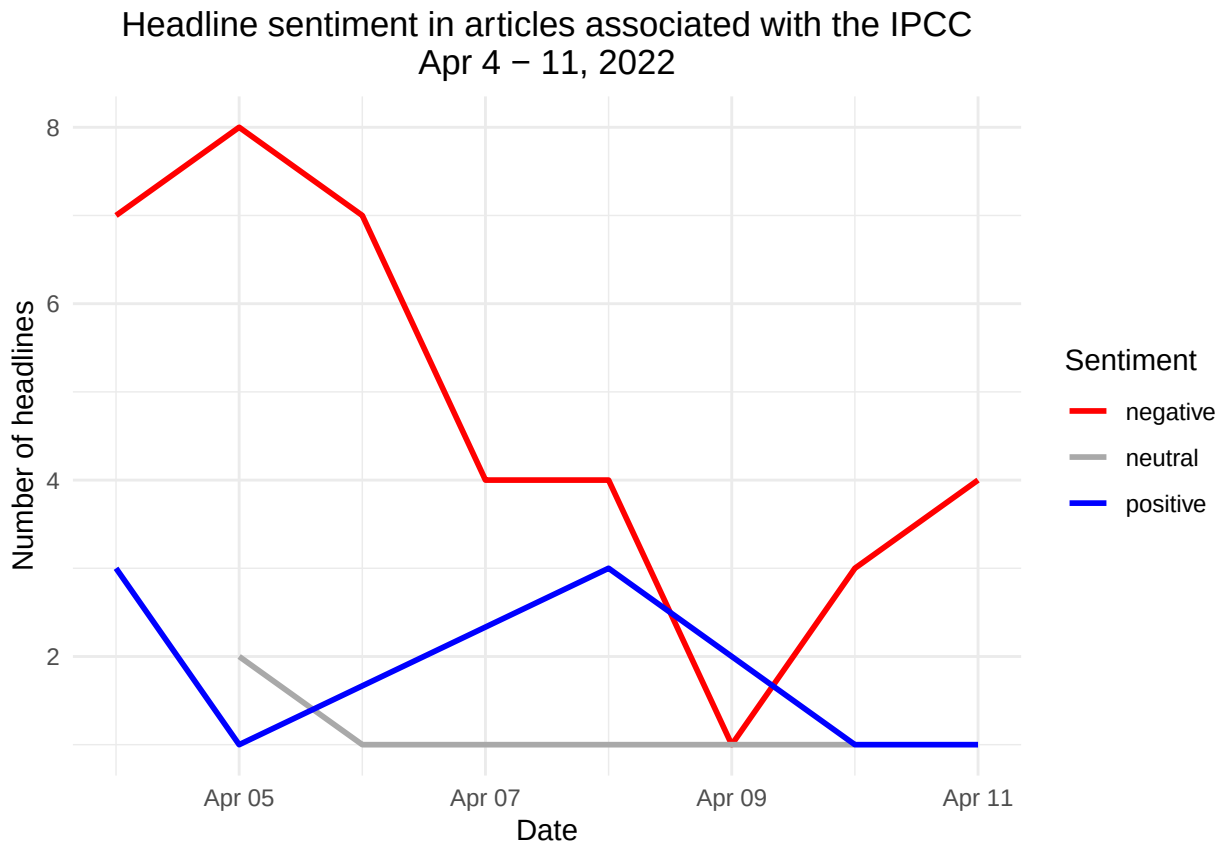
0. Sentiment polarity in "IPCC" articles

```
ipcc_lnt <- lnt_read(here("data", "Nexis_IPCC_Results.docx"))

ipcc_df <-lnt_convert(ipcc_lnt) %>%
  clean_names() %>%
  select(date, headline) %>%
  mutate(healdine = tolower(headline))

ipcc_df$headline <- gsub("[[:digit:][:punct:]]", "", ipcc_df$headline)

ipcc_words <- ipcc_df %>%
  select(date, headline) %>%
  unnest_tokens(output = word, input = headline, token = 'words', drop = FALSE)

bing <- get_sentiments('bing')

ipcc_sent <- ipcc_words %>%
  anti_join(stop_words, by = 'word') %>%
  inner_join(bing, by = "word")
```

```
ipcc_scores <- ipcc_sent %>%
  count(sentiment, headline, date) %>%
  spread(sentiment, n) %>%
  replace(is.na(.), 0) %>%
  mutate(raw_score = positive - negative,
  offset = mean(positive - negative),
  offset_score = (positive - negative) - offset) %>%
  arrange(desc(raw_score)) %>%
  mutate(sentiment = case_when(raw_score >= 1 ~ "positive",
                               raw_score == 0 ~ "neutral",
                               raw_score < 0 ~ "negative")) %>%
  count(sentiment, date)
```

```
ggplot(ipcc_scores, aes(x = date, y = n, color = sentiment)) +
  geom_line(size = 1) +
  scale_color_manual(values = c("red", "dark gray", "blue")) +
  labs(title = "Headline sentiment in articles associated with the IPCC\nApr 4 - 11, 2022",
       x = "Date", y = "Number of headlines", color = "Sentiment") +
```

```
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```

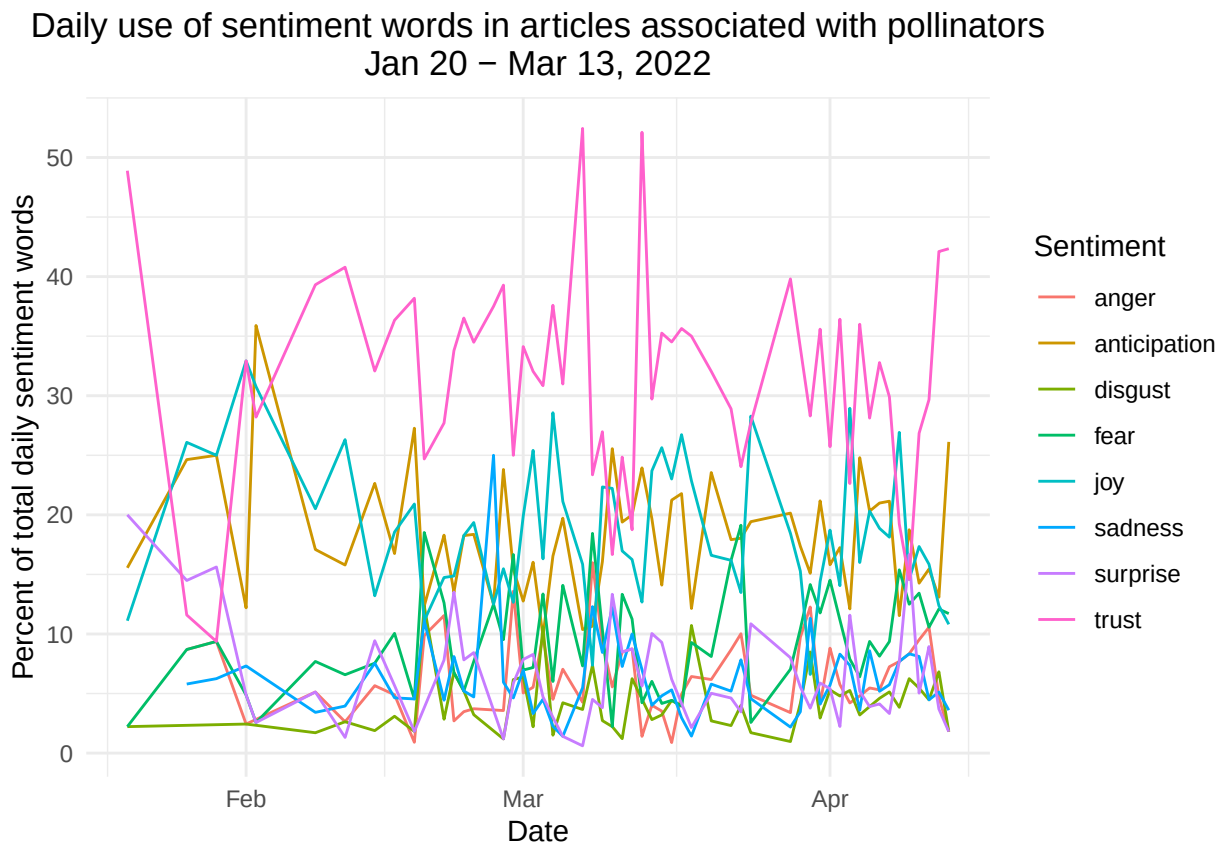


1-7. Emotion words in "pollinator" articles

```
# Reading in and cleaning up
text_lnt <- lnt_read(here("data", c("pollinators_nexis_1-100.docx", "pollinators_nexis_101-200.docx")))

text_df <- lnt_convert(text_lnt) %>%
  clean_names() %>%
  select(!c(section, edition, graphic)) %>%
  mutate(article = tolower(article))

# Removing as many URLs as possible
text_df$article <- gsub("(http\\S+\\s*)|(www\\S+\\s*)|(\\s\\S+.com\\S*\\s)", "", text_df$article)
text_df$article <- gsub("[[:digit:][:punct:]]", "", text_df$article)

# Splitting into individual words
words <- text_df %>%
  select(date, headline, article) %>%
  filter(!is.na(date)) %>%
  unnest_tokens(output = word, input = article, token = 'words')
```

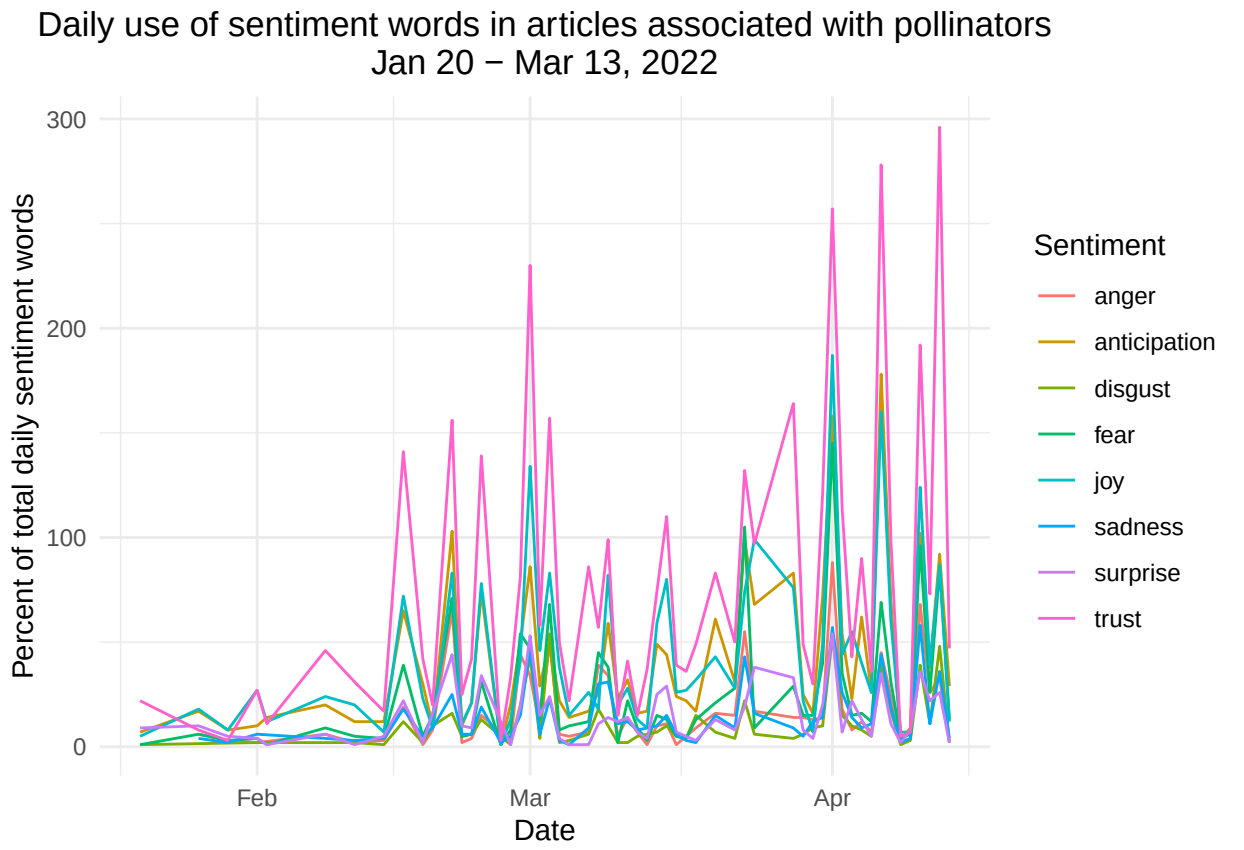
```r
# Joining with 8 emotion words
nrc <- get_sentiments('nrc') %>%
  filter(!sentiment %in% c("negative", "positive"))

nrc_word_counts <- words %>%
  inner_join(nrc, by = "word") %>%
  count(sentiment, date, sort = TRUE)

# Calculating percent of total emotion words used that day
nrc_word_pct <- nrc_word_counts %>%
  group_by(date) %>%
  mutate(percentiment = (n / sum(n)) * 100)
```

```r
ggplot(nrc_word_pct, aes(x = date, y = percentiment, color = sentiment)) +
  geom_line() +
  labs(title = "Daily use of sentiment words in articles associated with pollinators\nJan 20 - Mar 13, 2
       x = "Date", y = "Percent of total daily sentiment words", color = "Sentiment") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
ggplot(nrc_word_counts, aes(x = date, y = n, color = sentiment)) +
  geom_line() +
  labs(title = "Daily use of sentiment words in articles associated with pollinators\nJan 20 - Mar 13, 2
       x = "Date", y = "Percent of total daily sentiment words", color = "Sentiment") +
```

```
theme_minimal() +
theme(plot.title = element_text(hjust = 0.5))
```



At this time scale, there don't appear to be any explicit trends over time when looking at percentage of emotion words—things are pretty noisy. It is interesting, and uplifting, though that "trust" is the highest frequency sentiment throughout most of the time period. Hopefully people are starting to recognize just how much our food system and general welfare rely on pollinators.

When we look at raw counts, we do see a clear increase in frequency of most emotion words. This is to be expected as stories about pollinators are more prevalent around spring time, so the overall number of articles was generally increasing in this time period.