assignment2

Jake Eisaguirre

04/13/2022

```
my_files <- list.files(pattern = ".docx", path = here(),
                       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 [0.17 secs]

##  ...articles split [0.19 secs]

##  ...lengths extracted [0.19 secs]

##  ...headlines extracted [0.19 secs]

##  ...newspapers extracted [0.20 secs]

##  ...dates extracted [0.21 secs]

##  ...authors extracted [0.21 secs]

##  ...sections extracted [0.22 secs]

##  ...editions extracted [0.22 secs]

## Warning in lnt_asDate(date.v, ...): More than one language was detected. The
## most likely one was chosen (English 97%)

##  ...dates converted [0.22 secs]

##  ...metadata extracted [0.23 secs]

##  ...article texts extracted [0.23 secs]

##  ...superfluous whitespace removed [0.24 secs]

## Elapsed time: 0.25 secs
```

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

dat2<- data_frame(element_id = seq(1:length(meta_df$Headline)), Date = meta_df$Date, Headline = meta_df$
```

```
## Warning: `data_frame()` was deprecated in tibble 1.1.0.
## Please use `tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

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

dat3 <- inner_join(dat2,paragraphs_dat, by = "element_id")
```

```
mytext <- get_sentences(dat3$Text)

sent <- sentiment(mytext)
```

```
sent_df <- inner_join(dat3, sent, by = "element_id")

sentiment <- sentiment_by(sent_df$Text)
```

```
## Warning: Each time `sentiment_by` is run it has to do sentence boundary disambiguation when a
## raw `character` vector is passed to `text.var`. This may be costly of time and
## memory.  It is highly recommended that the user first runs the raw `character`
## vector through the `get_sentences` function.
```

```
sent_df %>%
  arrange(sentiment)
```

```
## # A tibble: 2,788 x 7
##    element_id Date       Headline     Text      sentence_id word_count sentiment
##         <int> <date>     <chr>        <chr>           <int>      <int>     <dbl>
##  1         65 2022-04-01 Complexity ~ "Texas A~           1         17    -0.740
##  2         65 2022-04-01 Complexity ~ ""The de~           1         17    -0.740
##  3         65 2022-04-01 Complexity ~ "The Can~           1         17    -0.740
##  4         65 2022-04-01 Complexity ~ "Prepare~           1         17    -0.740
##  5         65 2022-04-01 Complexity ~ "Wildfir~           1         17    -0.740
##  6         65 2022-04-01 Complexity ~ "This we~           1         17    -0.740
##  7         65 2022-04-01 Complexity ~ "Over th~           1         17    -0.740
##  8         65 2022-04-01 Complexity ~ "More th~           1         17    -0.740
##  9         65 2022-04-01 Complexity ~ "Due to ~           1         17    -0.740
## 10         65 2022-04-01 Complexity ~ ""We wan~           1         17    -0.740
## # ... with 2,778 more rows
```

```
nrc_sent <- get_sentiments('nrc')

text_words <- dat3  %>%
  unnest_tokens(output = word, input = Text, token = 'words')
```

```
sent_words <- text_words%>% #break text into individual words
  anti_join(stop_words, by = 'word') %>%
  inner_join(nrc_sent, by = 'word') %>%
  filter(!sentiment %in% c("positive", "negative")) %>%
  mutate(Date = as_date(Date))

sent_word_count <- sent_words %>%
  group_by(Date, sentiment) %>%
  count(sentiment) %>%
  ungroup() %>%
  group_by(Date) %>%
  mutate(n_max = sum(n),
         percent = round((n / n_max) * 100, 2))
```

```
ggplot(data = sent_word_count) +
  geom_line(aes(x = Date, y = percent, fill = sentiment, color = sentiment)) +
  theme_classic() +
  labs(title = "Sentiment Surrounding Wildfires for 2022-03-16 to 2022-04-13",
       y = "Percent",
       x = "Date")
```

```
## Warning: Ignoring unknown aesthetics: fill
```

```
## Warning: Removed 8 row(s) containing missing values (geom_path).
```