Final Project

Paloma Cartwright, Allie Cole, Wylie Hampson, Ben Moscona-Remnitz, Julia Parish

2022-06-01

Sentiment of Sustainable Agriculture over 30 year period

This text sentiment analysis was completed as an assignment for the course, Environmental Data Science 231: Text and Sentiment Analysis for Environmental Problems. The data was sourced from ...

Original assignment instructions can be found here

Load Libraries

```
#install packages as necessary, then load libraries
if (!require(librarian)){
  install.packages("librarian")
  library(librarian)
}

librarian::shelf(here,
                 igraph,
                 kableExtra,
                 ldatuning,
                 LDAvis,
                 LexisNexisTools,
                 lubridate,
                 pdftools,
                 quanteda,
                 quanteda.textplots,
                 quanteda.textstats,
                 readr,
                 reshape2,
                 sentimentr,
                 tidyr,
                 tidytext,
                 tidyverse,
                 tm,
                 topicmodels,
                 tsne)
```

Nexis Uni Data - Sustainable Agriculture Sentiment past 30 years

Data consists of news articles, law reviews, and journals using the Nexis Uni database, and published between the years 1990-2021 using the search terms "Integrated Pest Management" and "sustainable farming." Due to Nexis Uni limiting article returns to 100 documents per search, we collected the 100 most relevant articles

from each year. When a year returned less than 100 articles for the search terms, all returned articles were collected. Articles were not geographically restricted, but they were all published in English.

```
nu_folder <- here("data/text_data")

nu_data <- list.files(pattern = ".DOCX", path = nu_folder,
                      full.names = TRUE, recursive = TRUE, ignore.case = TRUE)

# lnt_read = read in a LexisNexis file
nu_data <- lnt_read(nu_data)
```

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

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

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

ipm_data <- inner_join(headline_df, paragraphs_df, by = "element_id")
```

Global FAO data on farming practices over time

Data was collected from the Food and Agriculture Organization (FAO) of the United Nations on global farming practices since 1960. The data was collected from 1990 to 2021 from the FAO data service website, and combined into one csv file.

```
farming_practices_full <- list.files(path = "data/farming_practices/", full.names = T)
farming_practices <- list.files(path = "data/farming_practices/", full.names = F)
practice_names <- str_replace(farming_practices, "fao_", "") %>% str_replace(".csv", "")
practices_df <- map(farming_practices_full, read_csv) %>%
        map(~ select(., Area, Element, Item, Year, Unit, Value)) %>%
        reduce(rbind)
```

```
summary_practices <- practices_df %>%
        group_by(Item, Year, Unit) %>%
        filter(Unit != "%") %>%
        summarize(value = sum(Value))
```

```
graph_practices <- function(topic) {

filtered_practice <- summary_practices %>%
        filter(Item == topic)

filtered_practice %>%
        ggplot(aes(x = Year, y = value)) +
        geom_line() +
        expand_limits(y = 0) +
        ylab(paste(filtered_practice$Unit[1], topic))
}

topics <- summary_practices %>%
```

```
        ungroup() %>%
        select(Item) %>%
        distinct() %>%
        pull()

practices_graphs <- map(topics, graph_practices)
```

```
filenames_practices_graphs <- paste0(topics, ".pdf")
ggsave_med <- partial(ggsave, device = "pdf", width = 10, height = 6, units = "in")
map2(filenames_practices_graphs, practices_graphs, ggsave_med)
```

```
## [[1]]
## [1] "Cropland area certified organic.pdf"
##
## [[2]]
## [1] "Cropland area under organic agric..pdf"
##
## [[3]]
## [1] "Nutrient nitrogen N (total).pdf"
##
## [[4]]
## [1] "Nutrient phosphate P2O5 (total).pdf"
##
## [[5]]
## [1] "Nutrient potash K2O (total).pdf"
##
## [[6]]
## [1] "Pesticides (total).pdf"
```

Let's adjust the summary we made earlier by adding a country grouping

```
summary_practices <- practices_df %>%
        group_by(Item, Year, Unit, Area) %>%
        filter(Unit != "%") %>%
        summarize(value = sum(Value))
```

Now, let's filter our data before we graph to only include the U.S.

```
graph_practices <- function(topic, country) {

filtered_practice <- summary_practices %>%
        filter(Item == topic, Area == country)

filtered_practice %>%
        ggplot(aes(x = Year, y = value)) +
        geom_line() +
        expand_limits(y = 0) +
        ylab(paste(filtered_practice$Unit[1], topic))
}

topics <- summary_practices %>%
        ungroup() %>%
        select(Item) %>%
        distinct() %>%
        pull()
```

```
practices_graphs <- map2(topics, "United States of America", graph_practices)
```

```
filenames_practices_graphs <- paste0(topics, "us_only", ".pdf")
ggsave_med <- partial(ggsave, device = "pdf", width = 10, height = 6, units = "in")
map2(filenames_practices_graphs, practices_graphs, ggsave_med)
```

```
## [[1]]
## [1] "Cropland area certified organicus_only.pdf"
##
## [[2]]
## [1] "Cropland area under organic agric.us_only.pdf"
##
## [[3]]
## [1] "Nutrient nitrogen N (total)us_only.pdf"
##
## [[4]]
## [1] "Nutrient phosphate P2O5 (total)us_only.pdf"
##
## [[5]]
## [1] "Nutrient potash K2O (total)us_only.pdf"
##
## [[6]]
## [1] "Pesticides (total)us_only.pdf"
```