

# EDS 231: Assignment 2

Scout Leonard

04/15/2022

## Objective

### Load Libraries

```
# load packages
packages=c("tidyr",
           "lubridate",
           "pdftools",
           "pdftools",
           "tidytext",
           "here",
           "LexisNexisTools",
           "sentimentr",
           "readr",
           "textdata",
           "dplyr")

for (i in packages) {
  if (require(i,character.only=TRUE)==FALSE) {
    install.packages(i,repos='http://cran.us.r-project.org')
  }
  else {
    require(i,character.only=TRUE)
  }
}
```

## Part 1

\*\*[Access the Nexis Uni database through the UCSB library] (<https://www.library.ucsb.edu/research/db/211>)\*\*

Got it!

## Part 2

**\*\*Choose a key search term or terms to define a set of articles.\*\***

Done! I chose the term, “school lunch.” My MEDS cohort knows I love talking about the USDA National School Lunch Program. . .

## Part 3

**\*\*Use your search term along with appropriate filters to obtain and download a batch of at least 100 full text search results (.docx).\*\***

Sweet! All downloaded.

## Part 4

**\*\*Read your Nexis article document into RStudio.\*\***

Now for some coding...

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

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

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$Headline)
```

## Part 5

**\*\*This time use the full text of the articles for the analysis. First clean any artifacts of the data collection process (hint: this type of thing should be removed: “Apr 04, 2022( Biofuels Digest: <http://www.biofuelsdigest.com/> Delivered by Newstex”).\*\***

```
paragraphs_dat <- data.frame(element_id = paragraphs_df$Art_ID,  
                             Text = paragraphs_df$Paragraph)  
  
dat3 <- inner_join(dat2, paragraphs_dat, by = "element_id")
```

## Part 6

**\*\*Explore your data a bit and try to replicate some of the analyses above presented in class if you'd like (not necessary).\*\***

## Part 7

**\*\*Plot the amount of emotion words (the 8 from nrc) as a percentage of all the emotion words used each day (aggregate text from articles published on the same day). How does the distribution of emotion words change over time? Can you think of any reason this would be the case?\*\***