test

Amanda McDermott

12/5/2018

datasets before combining.

```
psq_texts <- read_csv("https://raw.githubusercontent.com/Glacieus/STAT-612-Final-Project/master/psq_tex
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   date = col_date(format = ""),
##   text = col_character()
## )
```

```
# Create clean version of article_texts - remove numbers at the start of text column
psq_texts2 <- psq_texts %>%
  filter(str_detect(psq_texts$text, "\"[0-9]+\",\"")) %>%
  mutate(text = gsub("\"[0-9]+\",", "", text)) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words)
```

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

```
# Nest words back into whole articles
clean_psq <- psq_texts2%>%
  nest(word) %>%
  mutate(text = map(data, unlist),
         text = map_chr(text, paste, collapse = " ")) %>%
  select(name, date, text)

# Add source column
clean_psq$source <- "PSQ"
```

More cleaning before combining…

```
#FINAL CSV WITH BOTH PSQ AND APSA
clean_article <- read_csv("https://raw.githubusercontent.com/Glacieus/STAT-612-Final-Project/master/arti
```

```
## Parsed with column specification:
## cols(
##   X1 = col_integer(),
##   name = col_character(),
##   date = col_date(format = ""),
##   text = col_character()
## )
```

```
clean_article$X1 <- NULL
clean_article$source <- "APSA"

clean_article2 <- clean_article %>%
  unnest_tokens(word, text) %>%
  nest(word) %>%
  mutate(text = map(data, unlist),
         text = map_chr(text, paste, collapse = " ")) %>%
  select(name, date, text, source)

# Combine rows
full_txt <- rbind(clean_psq, clean_article2)

# Take out year and make a separate column
full_txt <- full_txt %>%
  mutate(year = year(ymd(date)))
```