Difference between Days X, and Y

Anon

2023-08-25

Load the needed libraries.

```
library("ggplot2")
library("tidyverse")
library("dplyr")
```

Load the CSVs from Day 1 and Day 2, using code.

```
df1 <- read.csv("~/Documents/Stats/4Chan Scraper/Aug 24 2023 18:11:19.csv")
df2 <- read.csv("~/Documents/Stats/4Chan Scraper/Aug 25 2023 10:51:42.csv")
```

Merge data frame, and take difference between day 2 and day 1, subtracting data frames from each other.

n.x = df1

n.y = df2

```
df_merged <- merge(df1, df2, by="word", all=TRUE)
df_merged$result <- df_merged$n.y - df_merged$n.x
```

Feel free to add more "non-words," or "noise" to this list as you see fit. The pattern should be simple to expand on.

```
df_difference_filter <- df_merged %>%
  filter(!word == "de"
         & !word == "je"
         & !word == "een"
         & !word == "dat"
         & !word == "en"
         & !word == "eu"
         & !word == "te"
         & !word == "tu"
         & !word == "niet"
         & !word == "van"
         & !word == "niet"
         & !word == "ik"
         & !word == "ze"
         & !word == "om"
         & !word == "met"
         & !word == "uk"
         & !word == "qt"
         & !word == "wat"
         & !word == "bb"
         & !word == "op")
```

Assign NA values to zero so the packages can function normally.

```r
df_difference_filter$result[is.na(df_difference_filter$result)] <- 0
```

Get bottom 20 (negative) numbers.

```r
df_bottom <- df_difference_filter %>%
  top_n(-20)
```

Get top 20 (positive) numbers.

```r
df_top <- df_difference_filter %>%
  top_n(20)
```

Bind into new data frame

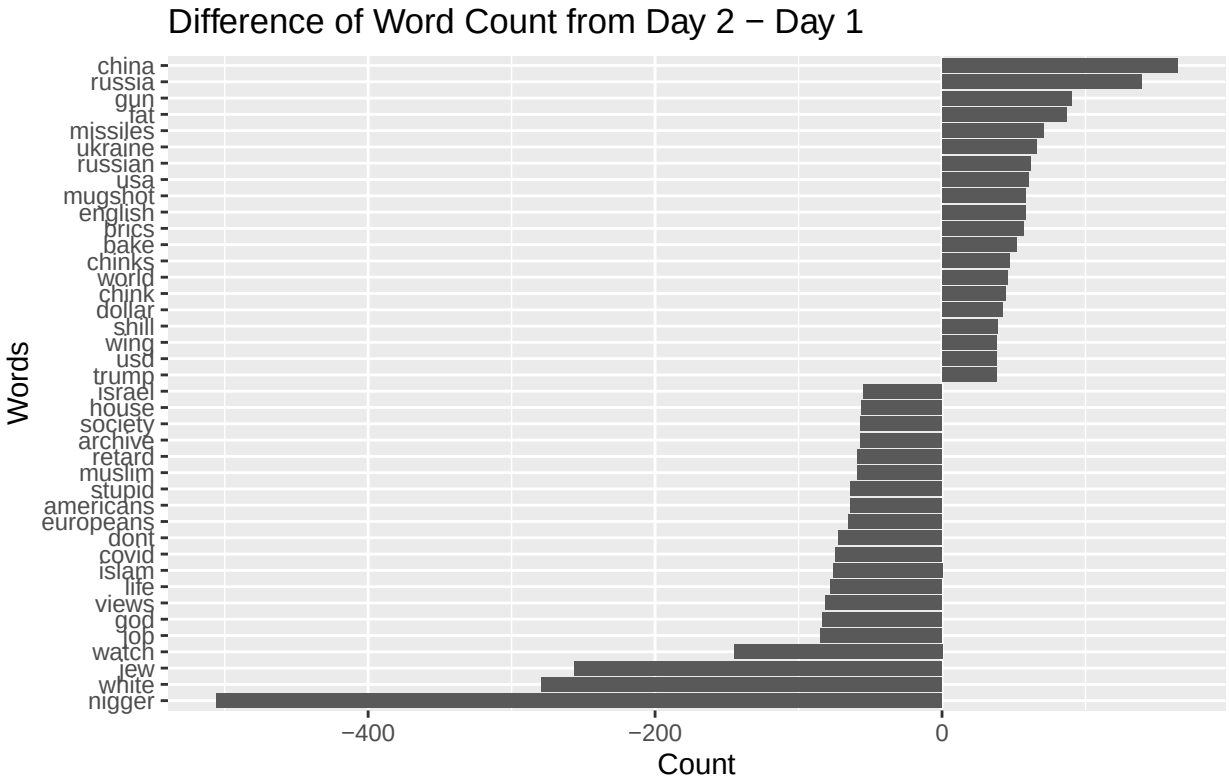
```r
df_merged2 <- rbind(df_top, df_bottom)
```

Bar graph of difference between Day 2, and Day 1.

Positive integers means there are more mentions on day 2, than what was observed on day 1.

Negative integers means there are more mentions on day 1, than what was observed on day 2.

```r
df_merged2 %>%
  top_n(40) %>%
  mutate(word = reorder(word, result)) %>%
  ggplot(aes(word, result)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Difference of Word Count from Day 2 - Day 1",
    x = "Words",
    y = "Count",
    caption = "Positive integers = More mentions on day 2
     Negative integers = Less mentions on day 2.") +
  coord_flip()
```



Positive integers = More mentions on day 2
Negative integers = Less mentions on day 2.