Final_Project

Miriam Nwaru

12/7/2018

Load necessary libraries here.

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggformula)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: ggstance
```

```
##
## Attaching package: 'ggstance'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     geom_errorbarh, GeomErrorbarh
```

```
##
## New to ggformula?  Try the tutorials:
##  learnr::run_tutorial("introduction", package = "ggformula")
##  learnr::run_tutorial("refining", package = "ggformula")
```

```
library(readr)
```

Read file

```
#read in csv file
news = read_csv("FoxNews_CNN_clean.csv", n_max = 10500)
```

```
## Parsed with column specification:
## cols(
##   ID_String = col_double(),
##   Text = col_character(),
##   News_Org = col_character(),
##   SA = col_integer()
## )
```

View top few rows

```
head(news)
```

```
## # A tibble: 6 x 4
##   ID_String Text                                              News_Org    SA
##       <dbl> <chr>                                             <chr>    <int>
## 1   1.07e18 @TessatTys @davidaxelrod @ggreenwald @RahmEman~ CNN          1
## 2   1.07e18 @realDonaldTrump @TheRickWilson @CNN @TheRickW~ CNN          1
## 3   1.07e18 @DeeBlog @Dennis2Clive @CNN Not so- we like hi~ CNN          1
## 4   1.07e18 @BoSnerdley @CNNPolitics There is a wealth of ~ CNN          0
## 5   1.07e18 @AppleWithAFace @paperbackwrit3r @HuffPost @Fo~ FoxNews      1
## 6   1.07e18 @DIVINE_VISUALS @CNN It's all right. Everybody~ CNN          1
```

Find values for Two-Sample Test for a Difference of Proportions

```
length(which(news$News_Org == "FoxNews" & news$SA > 0)) #Number of FoxNews tweets that were positive
```

```
## [1] 559
```

```
length(which(news$News_Org == "FoxNews")) #Total number of FoxNews tweets
```

```
## [1] 2376
```

```
length(which(news$News_Org == "CNN" & news$SA > 0)) #Number of CNN tweets that were positive
```

```
## [1] 1765
```

```
length(which(news$News_Org == "CNN")) #Total number of CNN tweets
```

```
## [1] 6205
```

Run Two-Sample Test for a Difference of Proportions

```
prop.test(c(559, 1765), n = c(2376, 6205), alternative = "two.sided")
```

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(559, 1765) out of c(2376, 6205)
## X-squared = 20.793, df = 1, p-value = 5.116e-06
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.06988768 -0.02846965
## sample estimates:
##    prop 1    prop 2
## 0.2352694 0.2844480
```

Create df to show counts of SA grouped by News Organizations

```
news_df <-
  news %>%
    group_by(SA, News_Org) %>%
    summarise(numb = n())
```

Create df and Create bar graph

```
news_df%>%
  group_by(News_Org) %>%
    mutate(countT= sum(numb)) %>% #Count the total for each News Org
    group_by(SA, add=TRUE) %>%
    mutate(per=paste0(round(numb/countT,2)))%>% #Calc the prop for each SA divided by each news org tot
    filter(News_Org != "NA")%>% #Remove NAs
```

```r
gf_col(per ~ SA, fill =~ News_Org, position = position_dodge(),data = news_df)%>%
gf_labs(
  title = "Sentiment Analysis of FoxNews vs CNN",
  x = "SA Categories",
  y = "Proportion",
  fill = "News Organizations"
)
```