R Workshop: Sentiment Analysis

3/25/2022

AGENDA

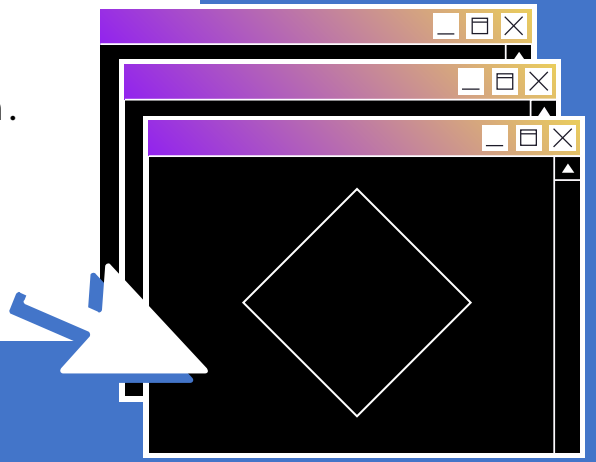
1. Overview of project
    a. How do we perform sentiment analysis?
    b. How do different techniques compare?
    c. How can we use it to make and publish an interactive graphic?
2. Searching for data
3. Coding in RStudio
4. Publishing your work
5. Further inspiration

SEARCHING FOR DATA

Start with sites dedicated to datasets:

- datasetsearch.research.google.com/
- kaggle.com/datasets
- public.opendatasoft.com/explore/

Use keywords and filters to narrow your search. Today, we'll be working with political data from Twitter, fetched using rtweet.



LINKS FOR TODAY'S WORKSHOP

- rstudio.cloud
  (Must create an account)
- https://github.com/SamanthaLLee/R_sentiment_analysis
- rpubs.com
  (Must create an account)

USING PACKAGES

If first time using,

```
install.packages("syuzhet")
install.packages("ggplot2")
install.packages("plotly")
install.packages("readr")
```

If already installed,

```
library(syuzhet)
library(ggplot2)
library(readr)
library(plotly)
```

LOADING DATA

```
# Load and view tweets
load("data/politicstweets.RData")
summary(politicstweets)
View(politicstweets)
```

```
  user_id            status_id            created_at                    screen_name
Length:2000        Length:2000        Min.   :2021-11-22 12:28:40   Length:2000
Class :character   Class :character   1st Qu.:2021-11-22 12:31:57   Class :character
Mode  :character   Mode  :character   Median :2021-11-22 12:35:10   Mode  :character
                                      Mean   :2021-11-22 12:35:07
                                      3rd Qu.:2021-11-22 12:38:13
                                      Max.   :2021-11-22 12:41:39
```

	user_id	status_id	created_at	screen_name	text	source
1	335854200	1462763203444674565	2021-11-22 12:41:39	praveenullal	Those who cannot match me in my service to the peo...	Twitter for Android
2	232371006	1462763199288205315	2021-11-22 12:41:38	LastFilmSeen	None of these women are protected in the way I am. ...	Twitter for Android
3	1222433033132953604	1462763197958692872	2021-11-22 12:41:37	mpotse20	Good Morning 💚 Reporting live from the Inaugural C...	Twitter for Android
4	1222433033132953604	1462760183835279369	2021-11-22 12:29:39	mpotse20	@errolbsk The ANC runs a patronage system, jobs for...	Twitter for Android
5	976331366102056966	1462763197048389640	2021-11-22 12:41:37	_liberal_Dank_	Crime, corruption, coercion are endemic to TMC's pol...	Twitter for iPhone
6	1034582706229731329	1462763196566183939	2021-11-22 12:41:37	NogoodChuck2	#NewProfilePic I ain'ts gots time to help you! I's busy ...	Twitter for Android
7	258734547	1462763196192890882	2021-11-22 12:41:37	alicherryanders	None of these women are protected in the way I am. ...	Twitter for iPhone
8	3036362195	1462763195978989580	2021-11-22 12:41:37	LmbBuckley	None of these women are protected in the way I am. ...	Twitter for iPad
9	3067400736	1462763195064471568	2021-11-22 12:41:37	spencersbrook	“On Twitter, Posobiec has repeatedly promoted ‘Foun...	Twitter for iPhone
10	1217985259075264513	1462763194737471491	2021-11-22 12:41:37	BillyKulikowsky	I don’t know if the person who drove through Waukes...	Twitter for iPad

C&P: https://raw.githubusercontent.com/SamanthaLLee/facebook-misinformation/main/data/facebook-fact-check.csv

Different Analysis Techniques

- The following are “bag of words” techniques that consider singular words and their sentiment classifications.
  - nrc
    - categorizes words in a binary fashion (“yes”/“no”) into categories of positive, negative, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust
  - bing
    - binary positive and negative categories
  - AFINN
    - assigns a numerical value in range [-5,5]
  - Syuzhet
    - custom dictionary developed in the Nebraska Literary Lab

- Pros
  - Intuitive, easy to get started with
- Cons
  - Overly simplified, does not consider how words work together

CALCULATING SENTIMENTS

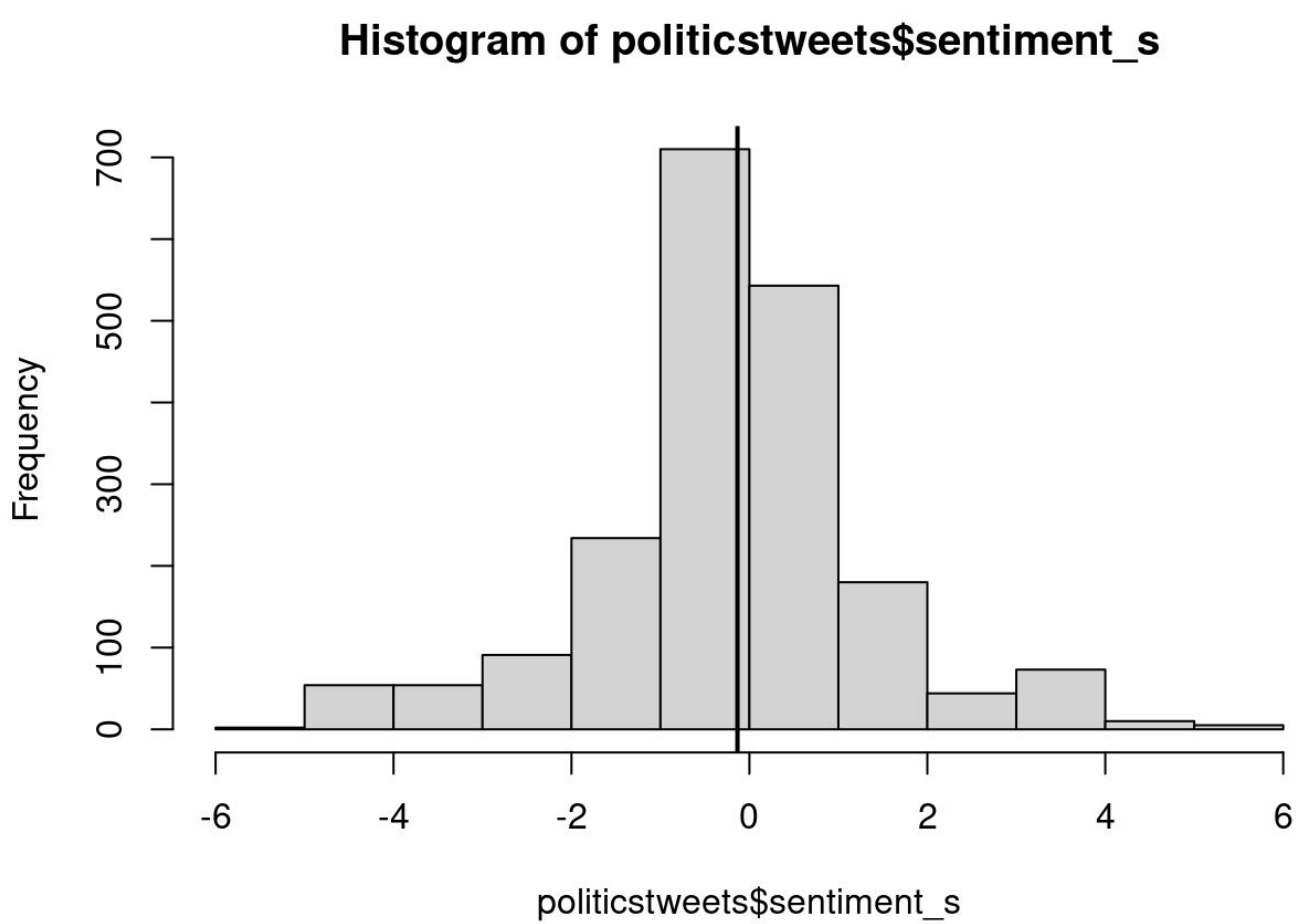
- Sum up shares, reactions, and

```
# Calculate sentiment scores
politicstweets$sentiment_s <- get_sentiment(politicstweets$text, method="syuzhet")
politicstweets$sentiment_n <- get_sentiment(politicstweets$text, method="nrc")
politicstweets$sentiment_a <- get_sentiment(politicstweets$text, method="afinn")
politicstweets$sentiment_b <- get_sentiment(politicstweets$text, method="bing")
```

```
> politicstweets$sentiment_s[1:10]
 [1] -1.25  0.40  5.10  0.55 -4.00 -1.00  0.40  0.40 -1.90 -3.25
> politicstweets$sentiment_n[1:10]
 [1]  0 -1  7 -1 -5 -1 -1 -1 -2 -2
> politicstweets$sentiment_a[1:10]
 [1]  -4  -5   8  -1 -18  -2  -5  -5  -6 -11
> politicstweets$sentiment_b[1:10]
 [1] -2 -3  4  1 -7 -2 -3 -3 -4  0
```

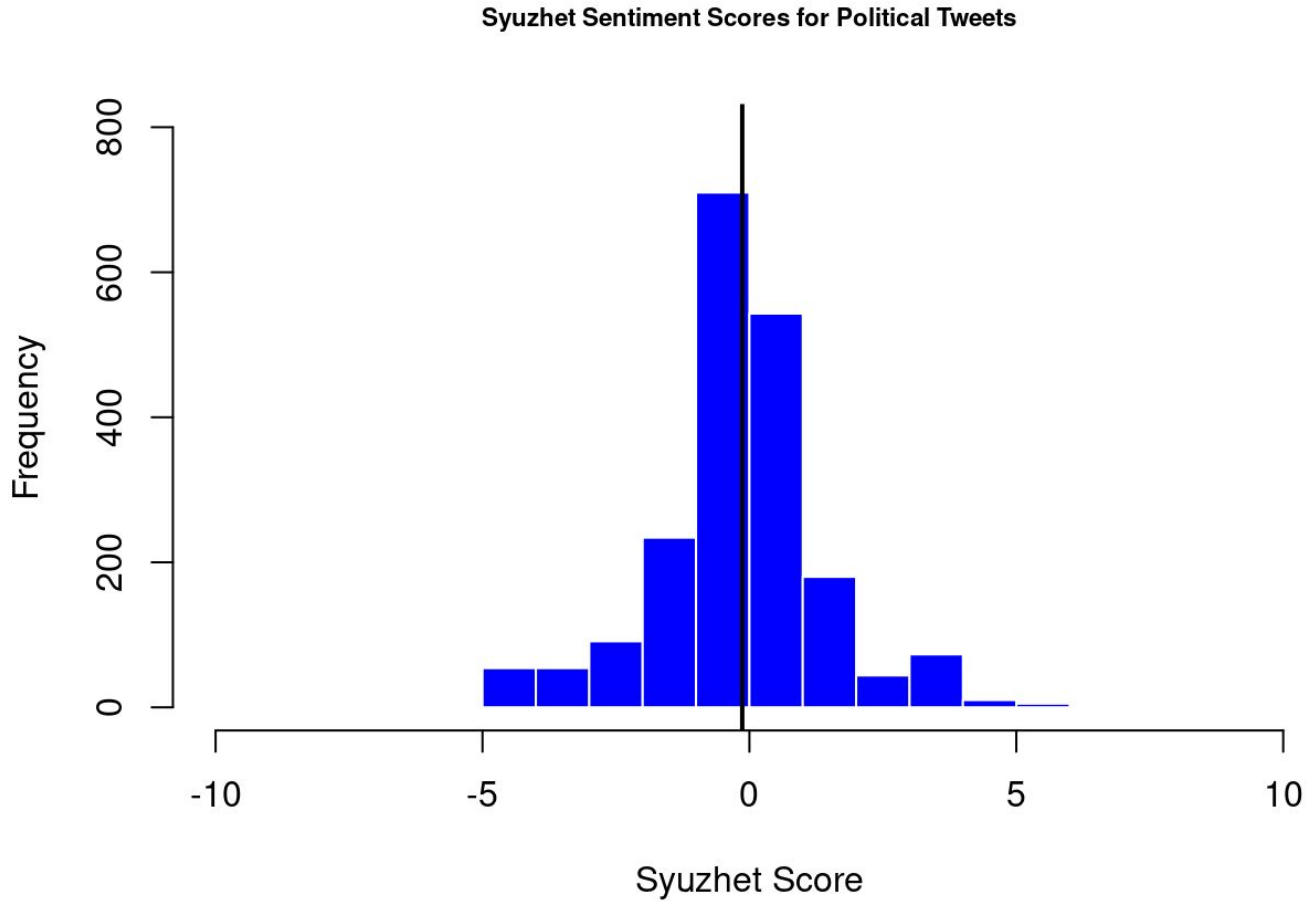
PLOTTING DATA

```
# First plot
hist(politicstweets$sentiment_s)
# Add line for the mean sentiment
abline(v=mean(politicstweets$sentiment_s), lwd=2)
```


Histogram of politicstweets$sentiment_s
Frequency
0
100
300
500
700
-6
-4
-2
0
2
4
6
politicstweets$sentiment_s

PLOTTING DATA, REVISED

```
# First plot
hist(politicstweets$sentiment_s,
     xlab="Syuzhet Score",
     main="Syuzhet Sentiment Scores for Political Tweets",
     cex.main=.7, col="blue",
     ylim = c(0, 800),
     xlim = c(-10, 10),
     border= F)
# Add line for the mean sentiment
abline(v=mean(politicstweets$sentiment_s), lwd=2)
```

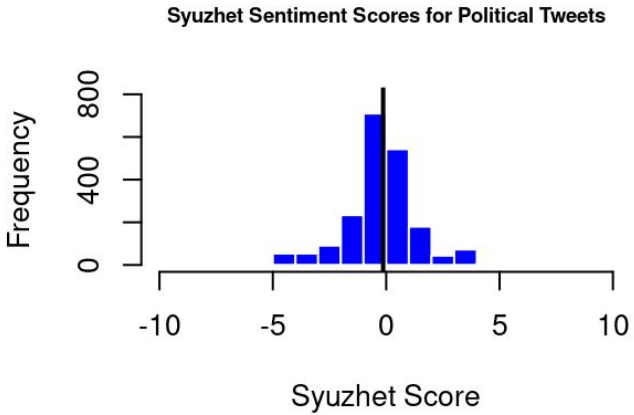

Syuzhet Sentiment Scores for Political Tweets
Frequency
0
200
400
600
800
-10
-5
0
5
10
Syuzhet Score

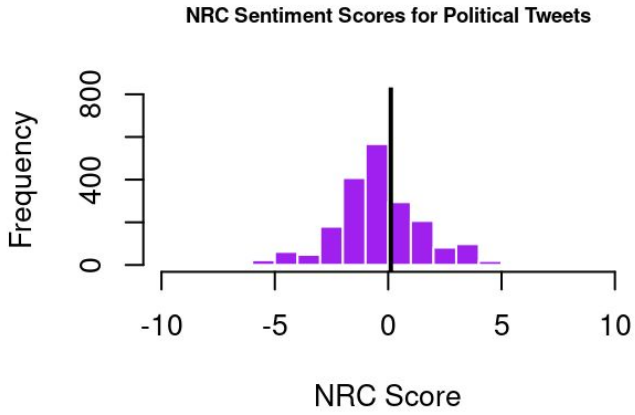
GRAPH ALL SENTIMENT SCORES

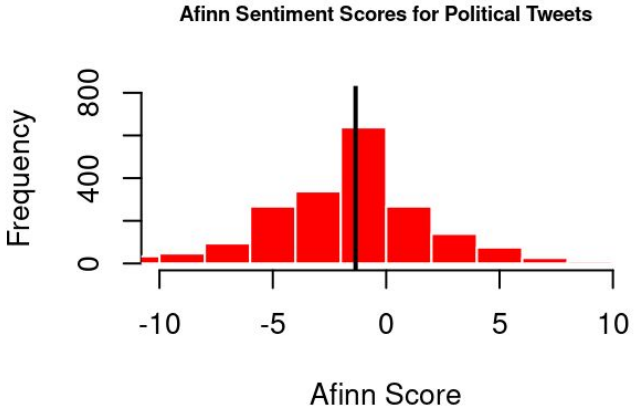
```
# Set par
par(mfrow = c(2, 2)) # 2 rows with 2 plots
```

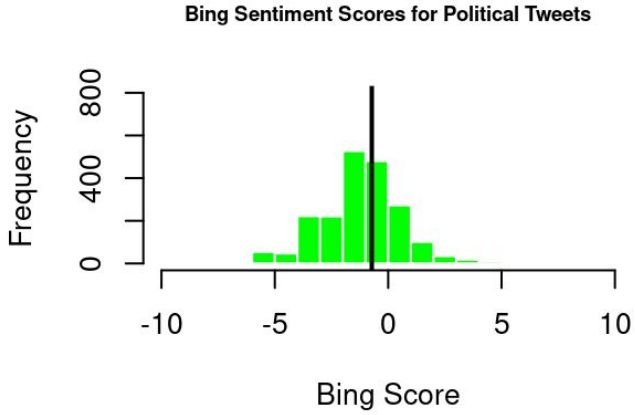
```
hist(politicstweets$sentiment_s,
     xlab="Syuzhet Score",
     main="Syuzhet Sentiment Scores for Political Tweets",
     cex.main=.7, col="blue",
     ylim = c(0, 800),
     xlim = c(-10, 10),
     border= F)
# Add line for the mean sentiment
abline(v=mean(politicstweets$sentiment_s), lwd=2)
```

change accordingly


Syuzhet Sentiment Scores for Political Tweets
Frequency
0
400
800
-10
-5
0
5
10
Syuzhet Score


NRC Sentiment Scores for Political Tweets
Frequency
0
400
800
-10
-5
0
5
10
NRC Score


Afinn Sentiment Scores for Political Tweets
Frequency
0
400
800
-10
-5
0
5
10
Afinn Score

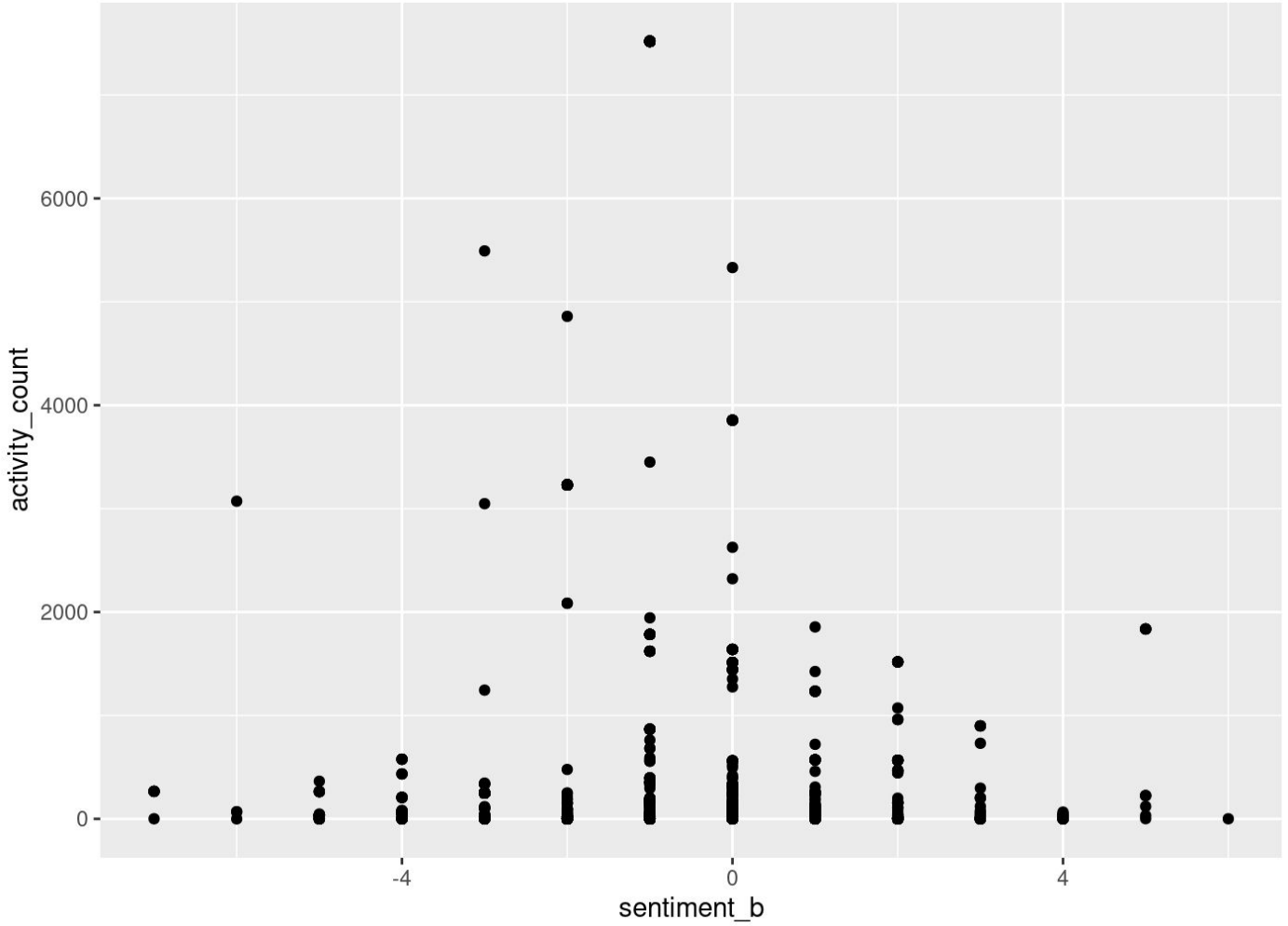

Bing Sentiment Scores for Political Tweets
Frequency
0
400
800
-10
-5
0
5
10
Bing Score

DO SENTIMENTS CORRELATE W/ ACTIVITY?

- Sum up retweets and favorites, plot against sentiment of choice

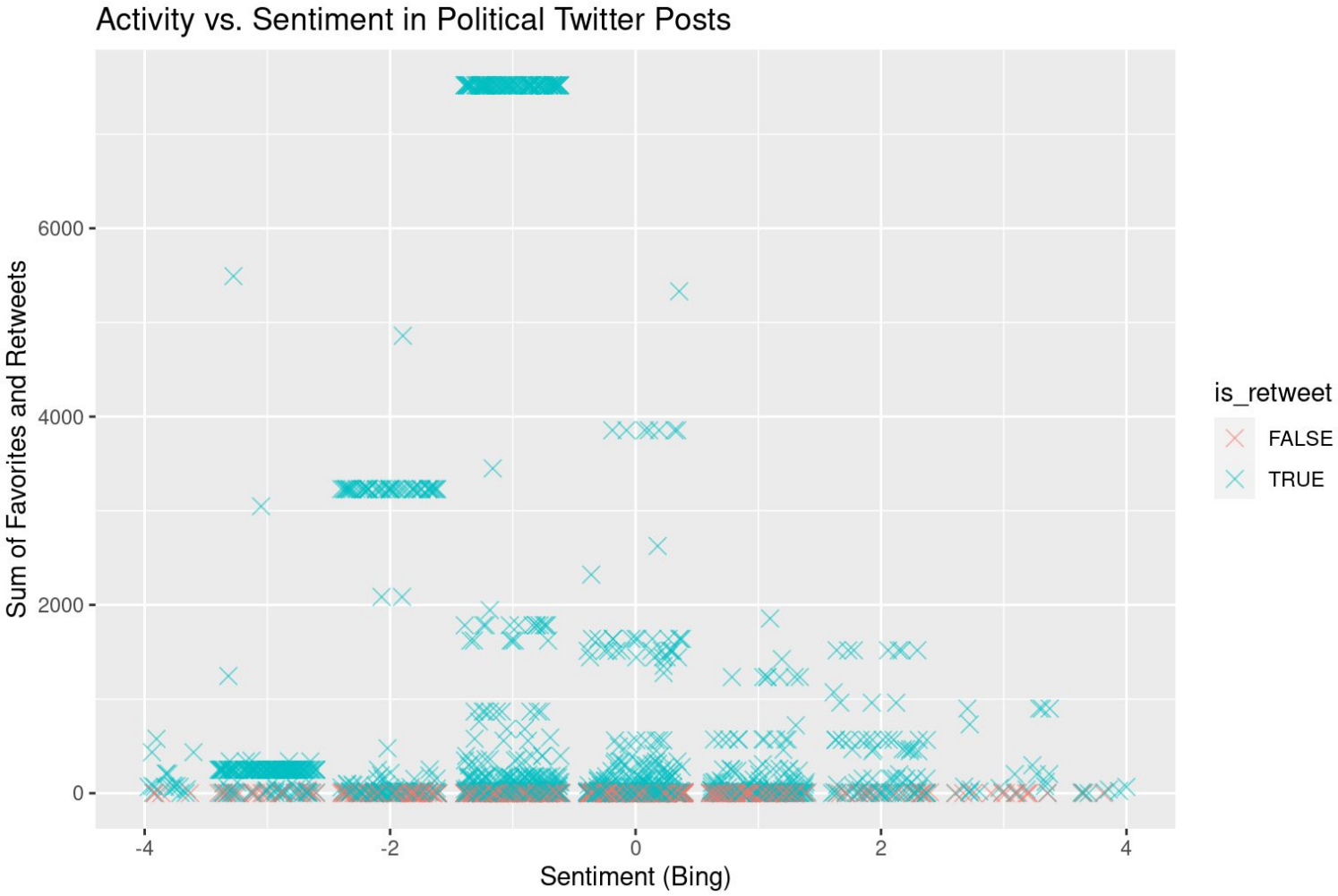
```
politicstweets$activity_count = politicstweets$favorite_count + politicstweets$retweet_count
```

```
p <- ggplot(data=politicstweets, mapping = aes(x=sentiment_b, y=activity_count)) +
  geom_point(na.rm=T)
```

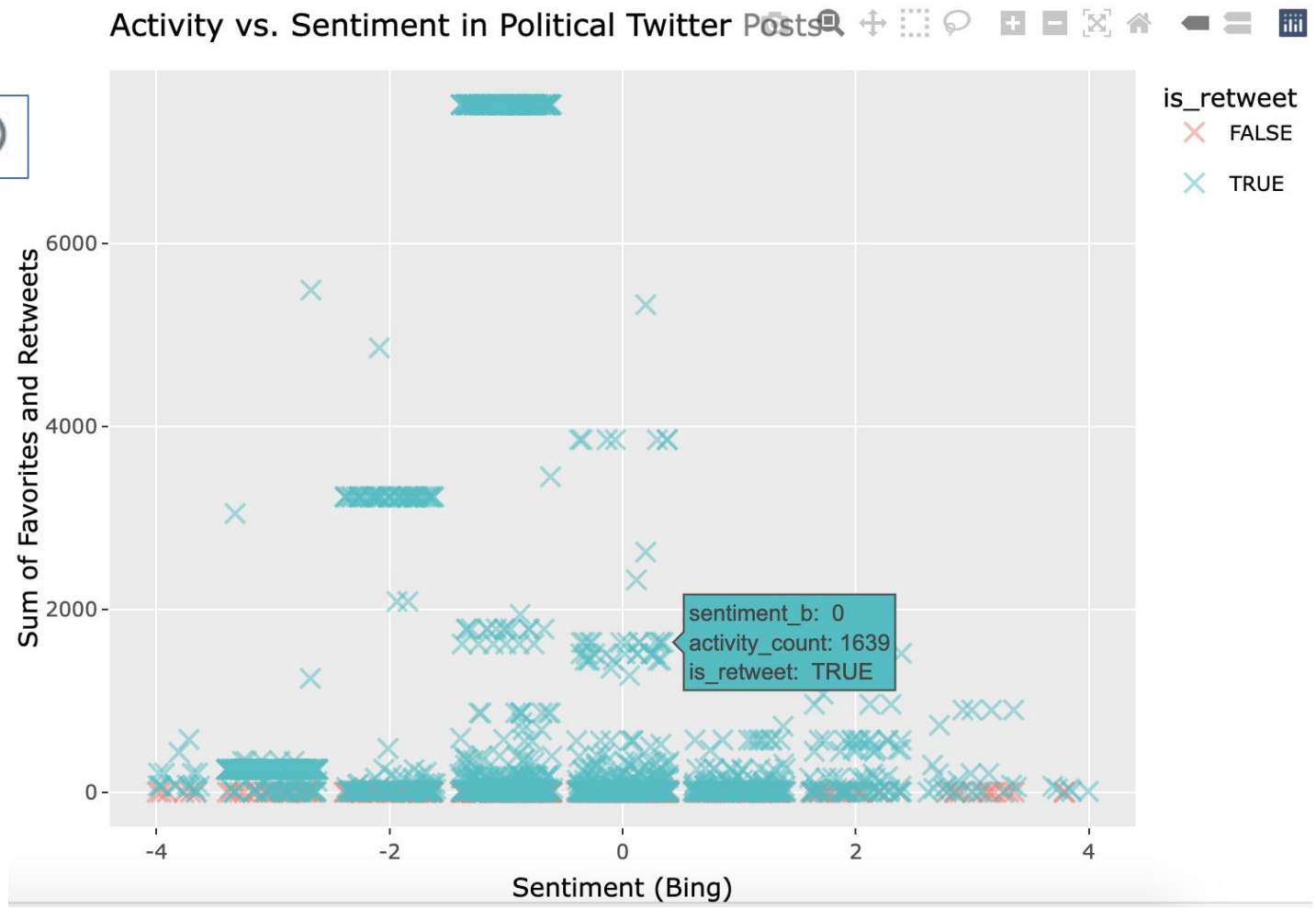

activity_count
6000
4000
2000
0
-4
0
4
sentiment_b

FLUFF IT UP

```
p <- ggplot(data=politicstweets, mapping = aes(x=sentiment_b, y=activity_count, color=is_retweet)) +
  geom_point(na.rm=T, size=3, shape=4, alpha = .5, position = "jitter")+
  ggtitle("Activity vs. Sentiment in Political Twitter Posts")+
  ylab("Sum of Favorites and Retweets")+
  xlab("Sentiment (Bing)") +
  xlim(-4, 4)
```

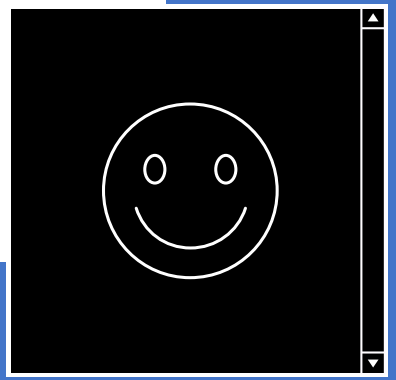

Activity vs. Sentiment in Political Twitter Posts
Sum of Favorites and Retweets
6000
4000
2000
0
-4
-2
0
2
4
Sentiment (Bing)
is_retweet
FALSE
TRUE

```
+ ggplotly(p)
```

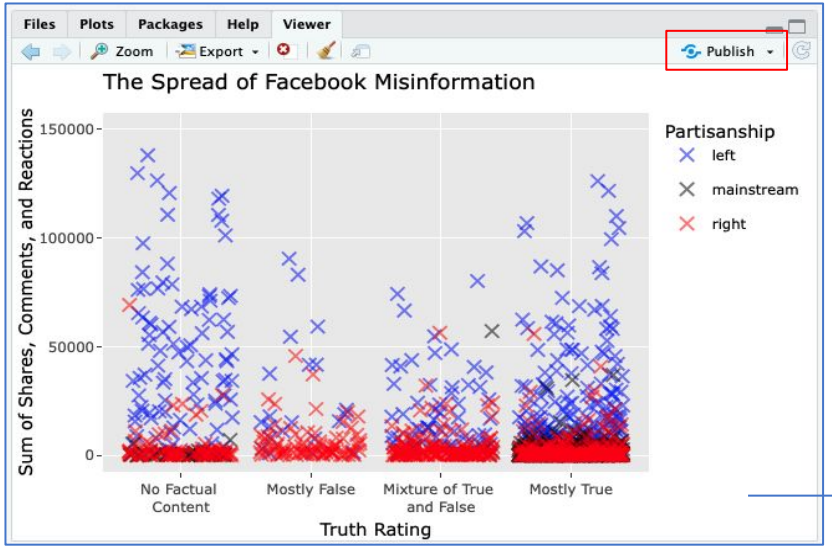

Activity vs. Sentiment in Political Twitter Posts
is_retweet
FALSE
TRUE
Sum of Favorites and Retweets
6000
4000
2000
0
sentiment_b: 0
activity_count: 1639
is_retweet: TRUE
-4
-2
0
2
4
Sentiment (Bing)

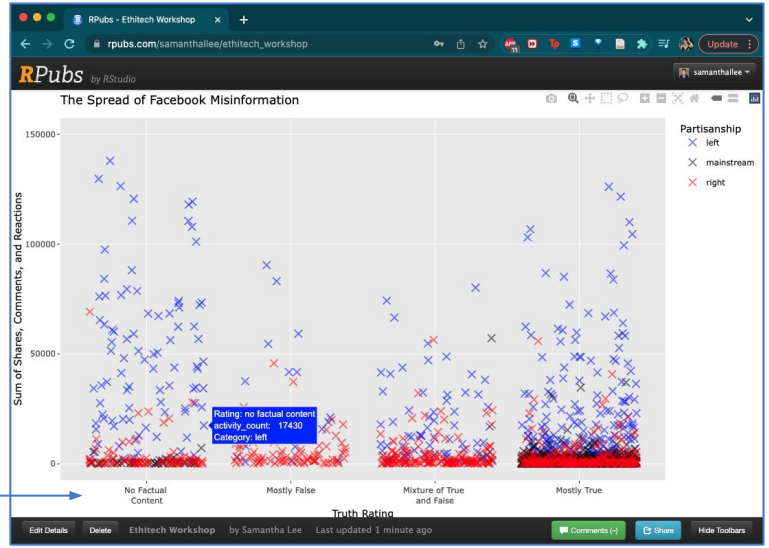
WHAT ' S NEXT?

You can directly publish your plots to https://rpubs.com/ to share your work.



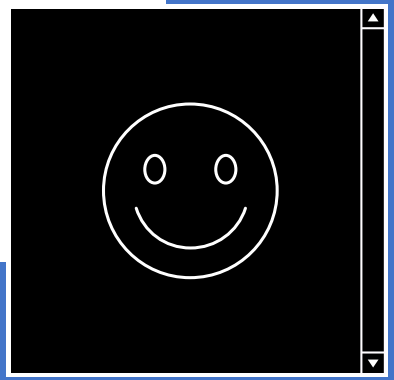
PUBLISHING DIRECTLY

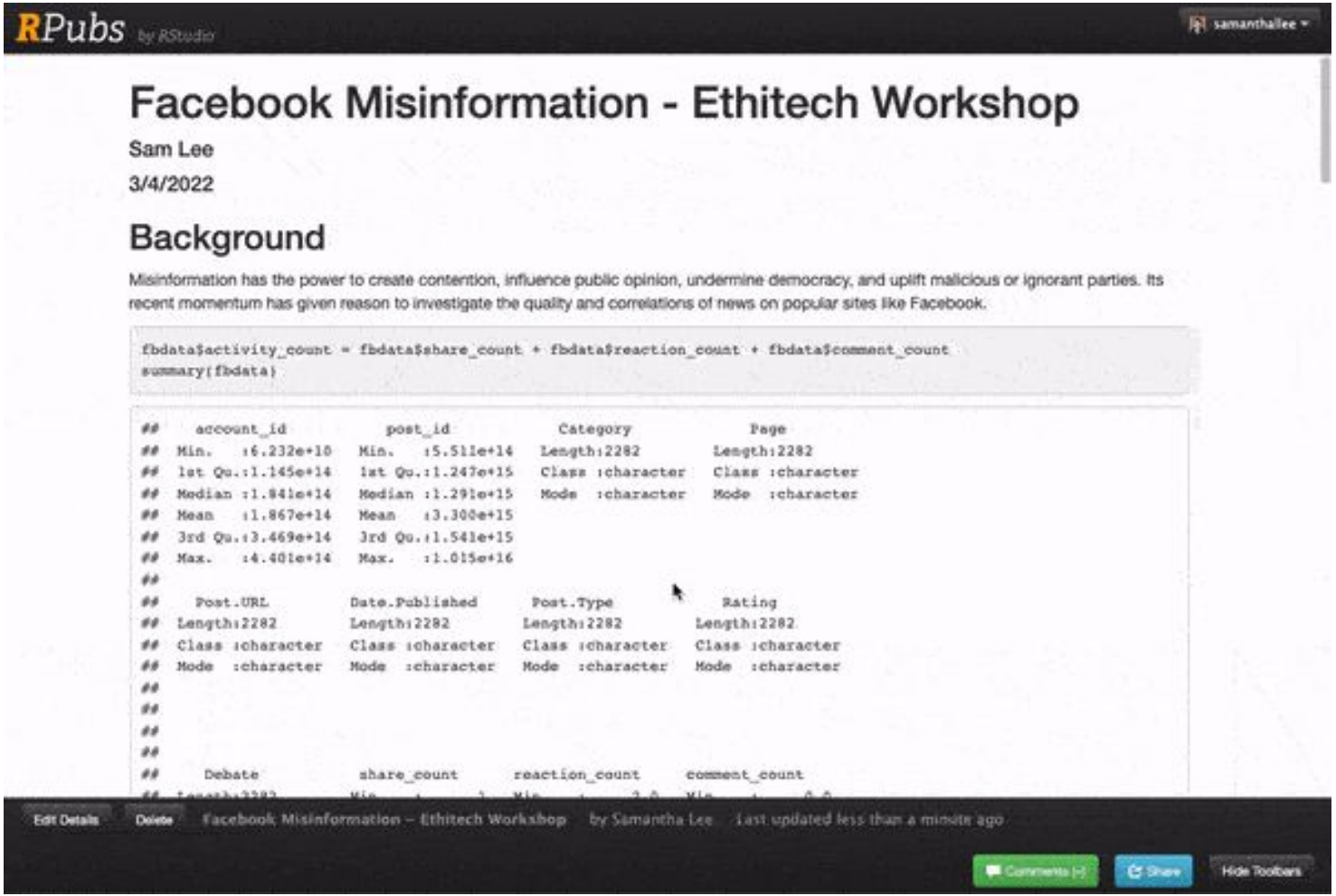




R MARKDOWN

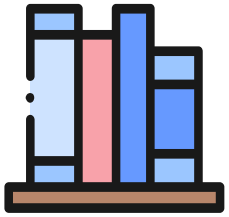
To have more control over what your RPubs page displays, you can create an RMD file to better format, provide context, arrange data and code, and more.





See @ https://rpubs.com/samanthallee/ethitech_workshop

FURTHER INSPIRATION

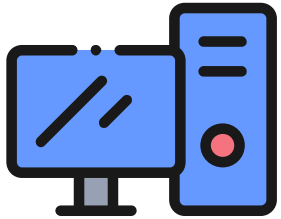


The R Graph Gallery

https://www.r-graph-gallery.com/

r/dataisbeautiful

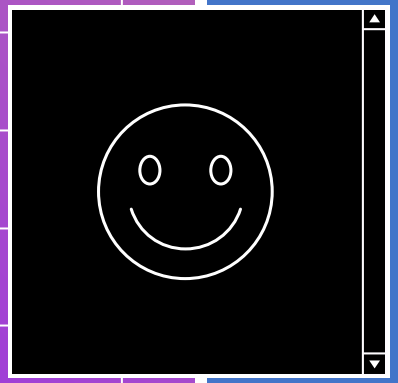
https://www.reddit.com/r/dataisbeautiful/



Information is Beautiful

https://informationisbeautiful.net/

THANKS!



Any questions?

Find all code & slides @
https://github.com/SamanthaLLee/R_sentiment_analysis

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, infographics & images by Freepik