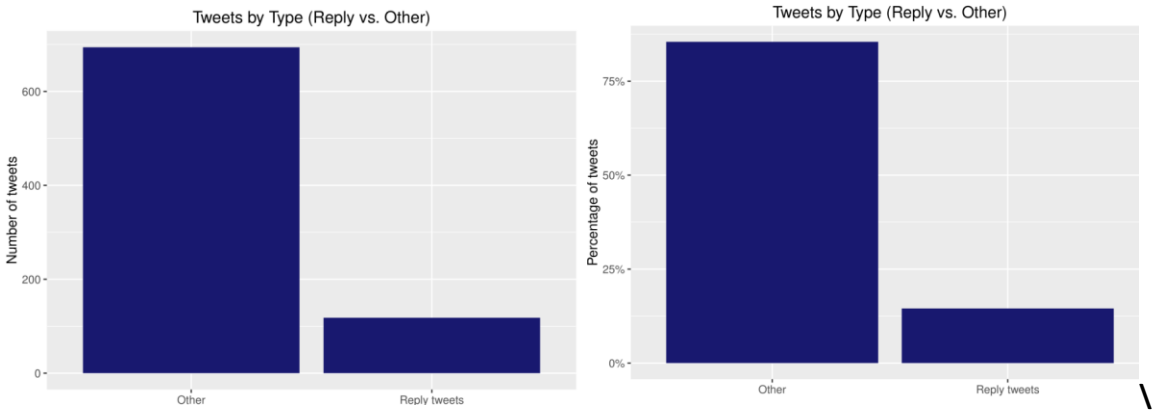
DS8006: Lab 3 "Twitter Basic – Analyzing Metadata"
Student's name: NAJLIS, BERNARDO (#500744793)

1. Briefly explain how you modified the scripts in Step 1 to load data directly from Twitter.

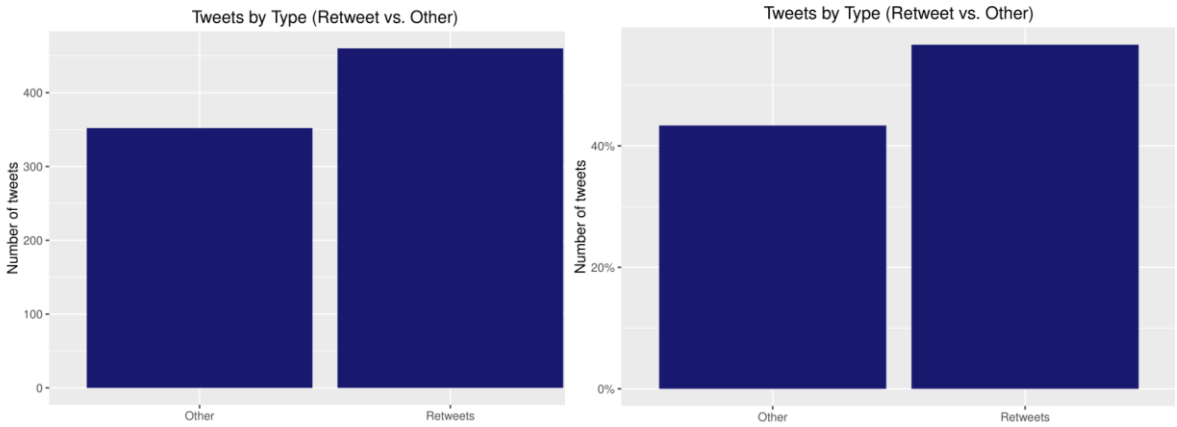
*The change was fairly straightforward, instead of using **read.csv()** to load tweets into a data frame, I used **searchTwitter("@RyersonU", n=1000)**. There are also a couple of additional steps to get to call searchTwitter (oAuth setup using **setup_twitter_oauth**) and return a data frame (using **twListToDF()** to transform the output obtained from **searchTwitter()**).*

2. Include screenshots for the four new charts (Task/Chart 1 – 4).

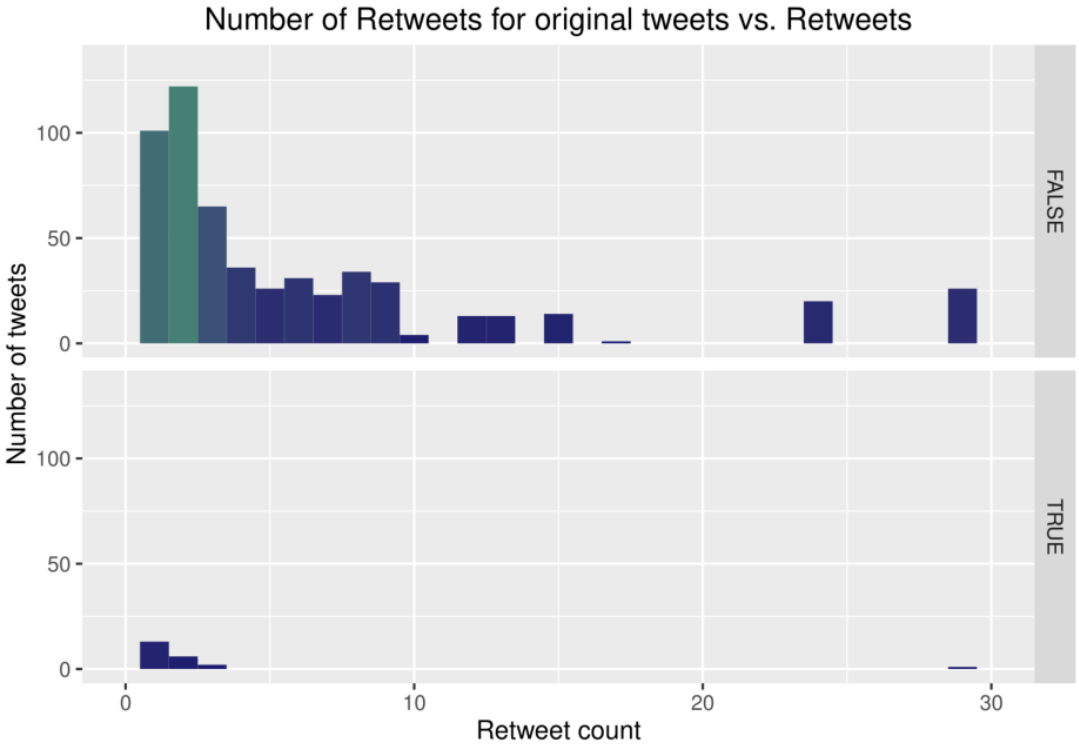
Task 1 – Tweets by Reply Type (as number and as percentage)



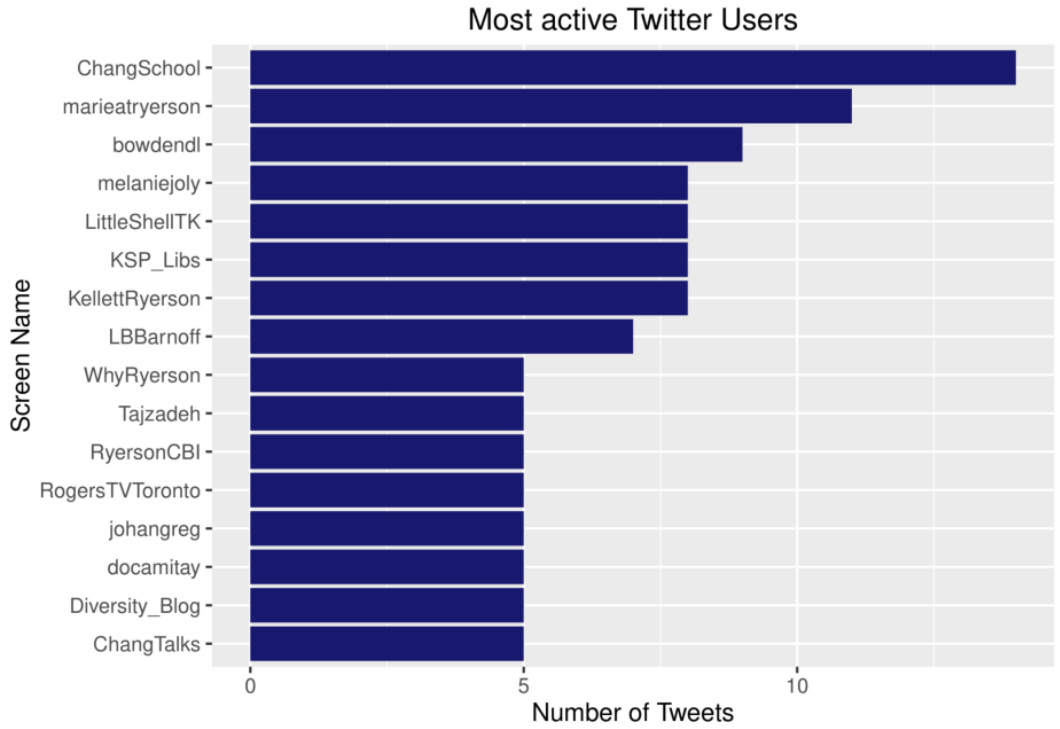
Task 2 – Tweets by Retweet type (Retweets vs Other) (as number and as percentage).



Task 3 – Comparison of retweet distribution for original vs retweet



Task 4 – New chart proposed to visualize other metadata fields: Number of tweets per Screen Name.



3. How would you change the script to display percentages instead of the actual counts in Charts 1&2?

Here is an example of the change required, compared side by side with differences highlighted.

Counts	Percentage

`ggplot(ryeTweets, aes(ryeTweets$isReply)) +`
`  geom_bar(fill = "midnightblue") +`
`  theme(legend.position="none", axis.title.x = element_blank()) +`
`  ylab("Number of tweets") +`
`  ggtitle("Tweets by Type (Reply vs. Other)") +`
`  scale_x_discrete(labels=c("Other", "Reply tweets"))`
	`ggplot(ryeTweets, aes(ryeTweets$isReply)) +`
	`  geom_bar(aes(y=(..count..) / sum(..count..)), fill = "midnightblue") +`
	`  theme(legend.position="none", axis.title.x = element_blank()) +`
	`  scale_y_continuous(labels=percent) +`
	`  labs(title = "Tweets by Type (Reply vs. Other)", y="Percentage of tweets") +`
	`  scale_x_discrete(labels=c("Other", "Reply tweets"))`

4. Briefly explain why someone would be interested in (meta)data represented in Chart 4.

Chart 4 helps bringing context who are the top 10 users more interactive with the account. This helps in understanding who are the users interacting the most and influenced by the @RyersonU account tweets.

5. What was the most challenging part of this lab?

The most challenging part was getting in the right mindset to understand ggplot2 concepts.