**Sentiment Analysis from Twitter for March 2022 of Ukraine War**

**Author: Dayvonn Jones, Nilay Vinchhi**

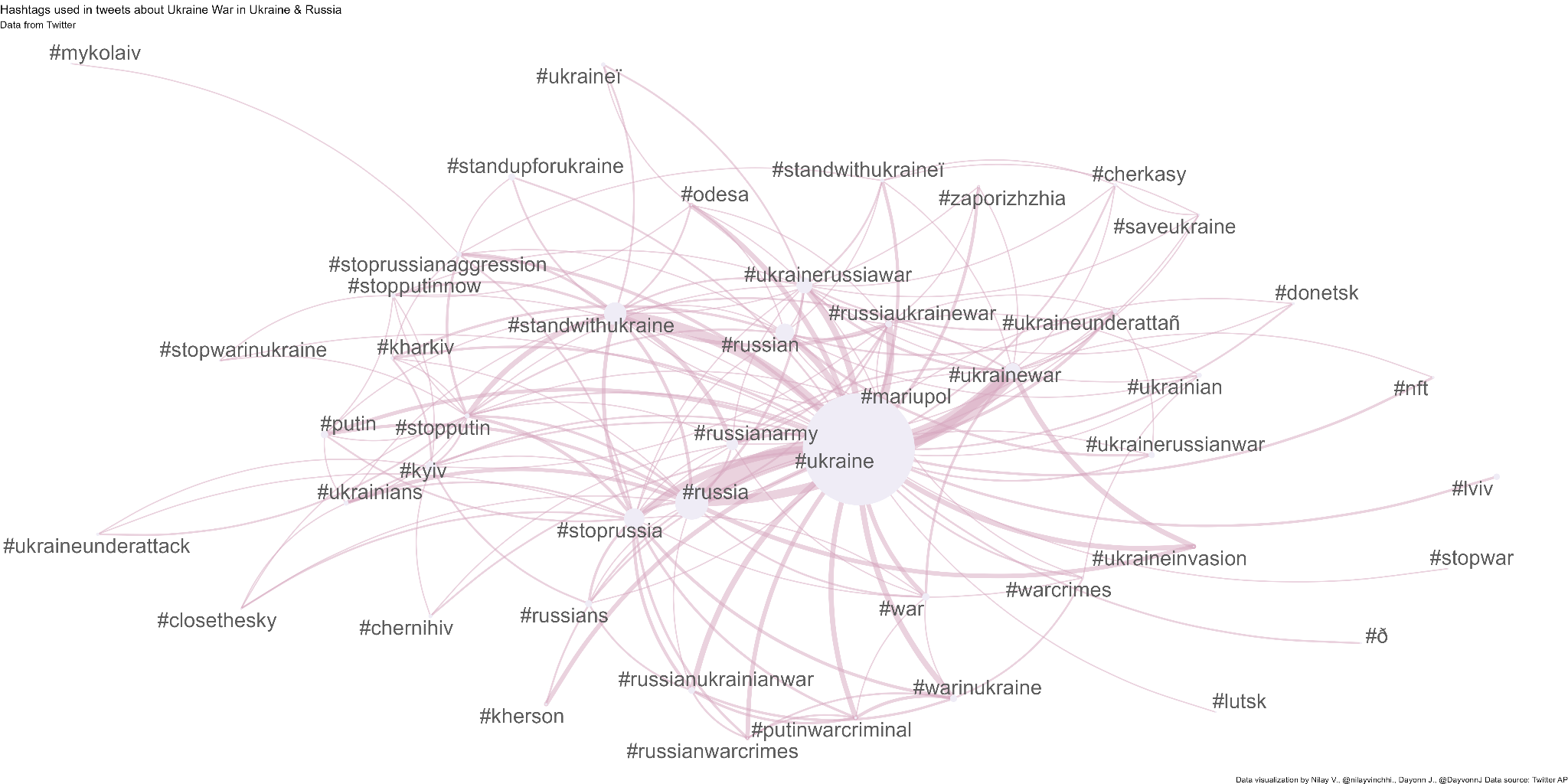
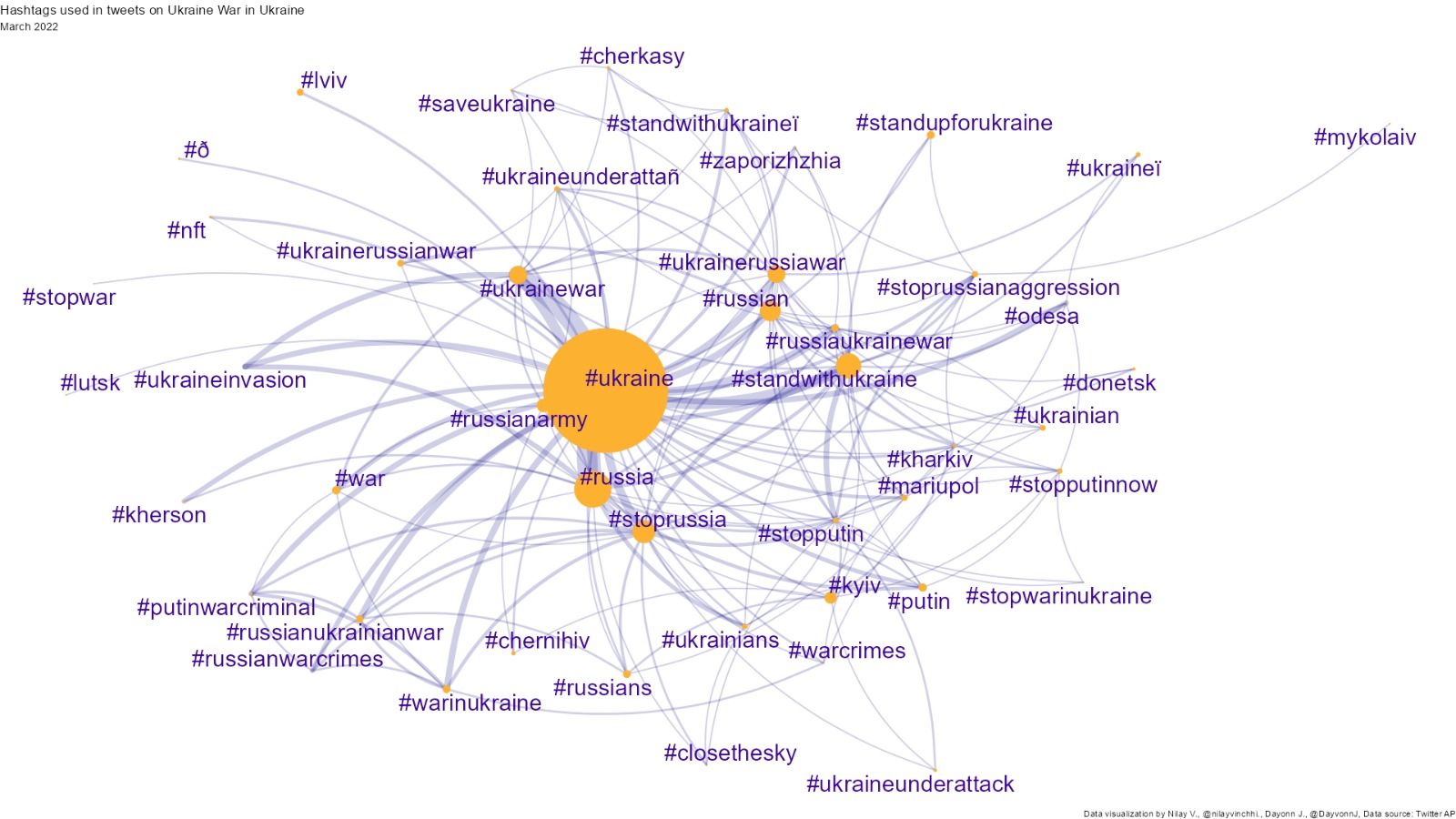
**Introduction/Description**: Social media has documented a significant amount of activity in the Russian-Ukrainian Conflict which began in February 2022. Multiple news outlets across the world have reported that misinformation in journalism and cybersecurity vulnerabilities have been exploited by entities within the country of Russia. There is one overwhelming theme and it associates organizations with origins in Russia, that have organized and targeted organizations and groups in other countries to spread misinformation and retrieve sensitive information. Twitter is one of the many organizations and social media platforms that have been exploited by and misused by pro-russian supporters. If there is a positive sentiment tweeted about the war in Ukraine it will most likely be from an account that has little engagement obscure history; a bot account.

**Objective**: Collect data from Twitter and report on the creation of the account and view the history of the engagement. The sentiment through Twitter accounts regarding the war in Ukraine will be analyzed with textual data by scraping tweets with an Twitter API (*rtweet*) and sorted using R Studio. In R Studio a user can view what tweets were created and the engagements (replies, retweets, quote retweets, and likes) with the tweet to understand the sentiment about the war between two moderately developed nations. After that the user account will be sorted by creation date, frequency (times per month) the user interacts with public tweets, and the number of followers they are following. Preparing a wordcloud of most frequent terms used in the tweets. Using textual analysis, we analyze which hashtags are strongly correlated in the month of March 2022.

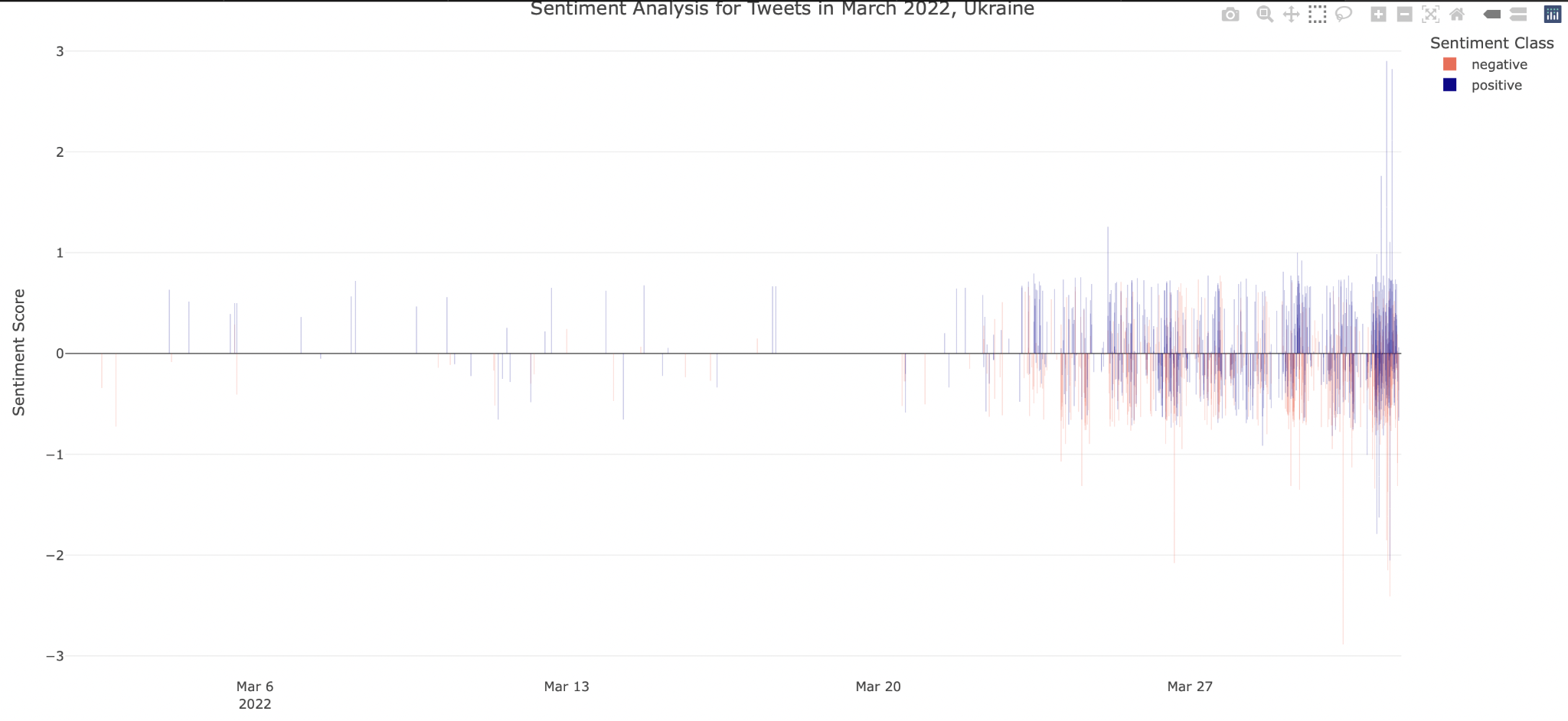
For the engagement statistics part we collected the statistics from the dataset that has the number of retweet counts and the number of replies and they were sorted from verified and unverified accounts. We have the variable that signifies if the account was Twitter by a verified account by Twitter or an account that was not recognized verified by Twitter.

**Methods**:

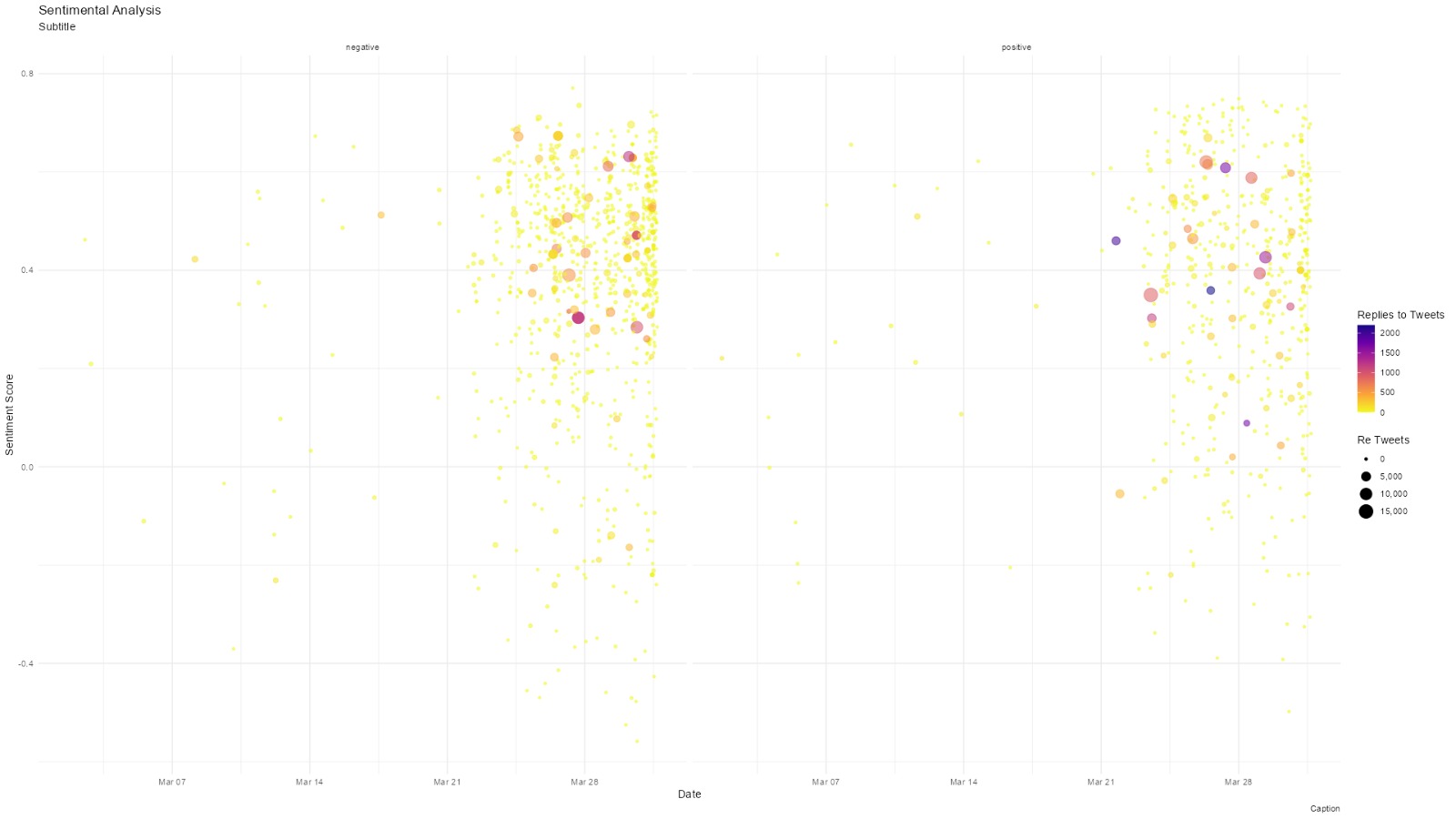
1. Collecting data using Twitter API (*rtweet*), the variable we are interested in is Full Text, (And other key variables names), which includes the tweet.
   1. Specifying the queries, for example cleaning the data for geolocation in in Ukraine. (For geolocation, we used the user’s provided geolocation data in that was record by Twitter). We merged the collected tweets and filtered the datasets, Version of R, using R Studio. (Also mention about the libraries we used to do it and what quereis, this will be in 1.a.)
   2. Give the hashtags used to fetch the data
2. Created a wordcloud using the *tm,* *wordcloud* package and library.
3. Network Analysis (*quanteda*) for hashtags (Result)



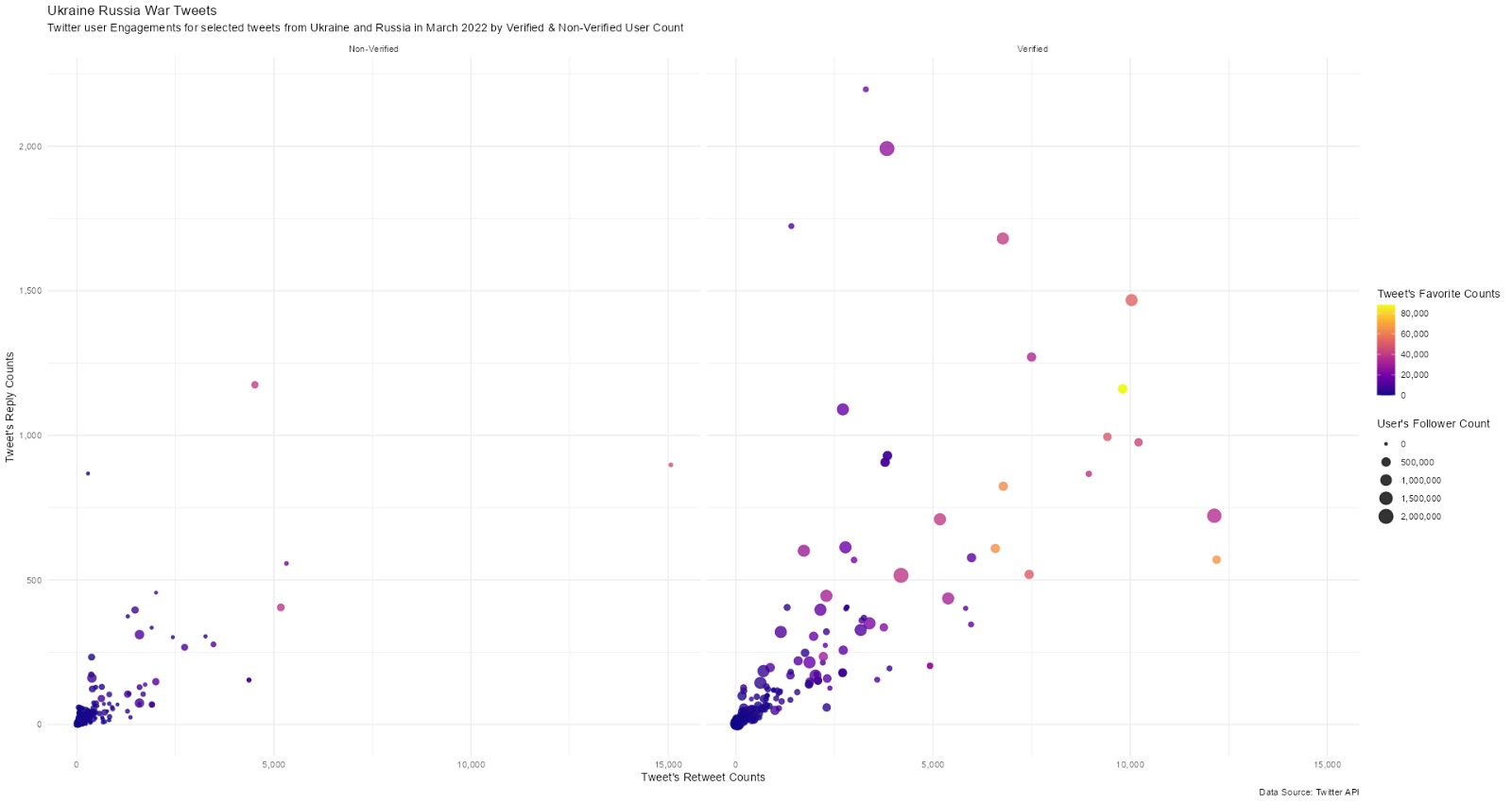
1. Sentimental Analysis using *Sentiment.ai* (Results)

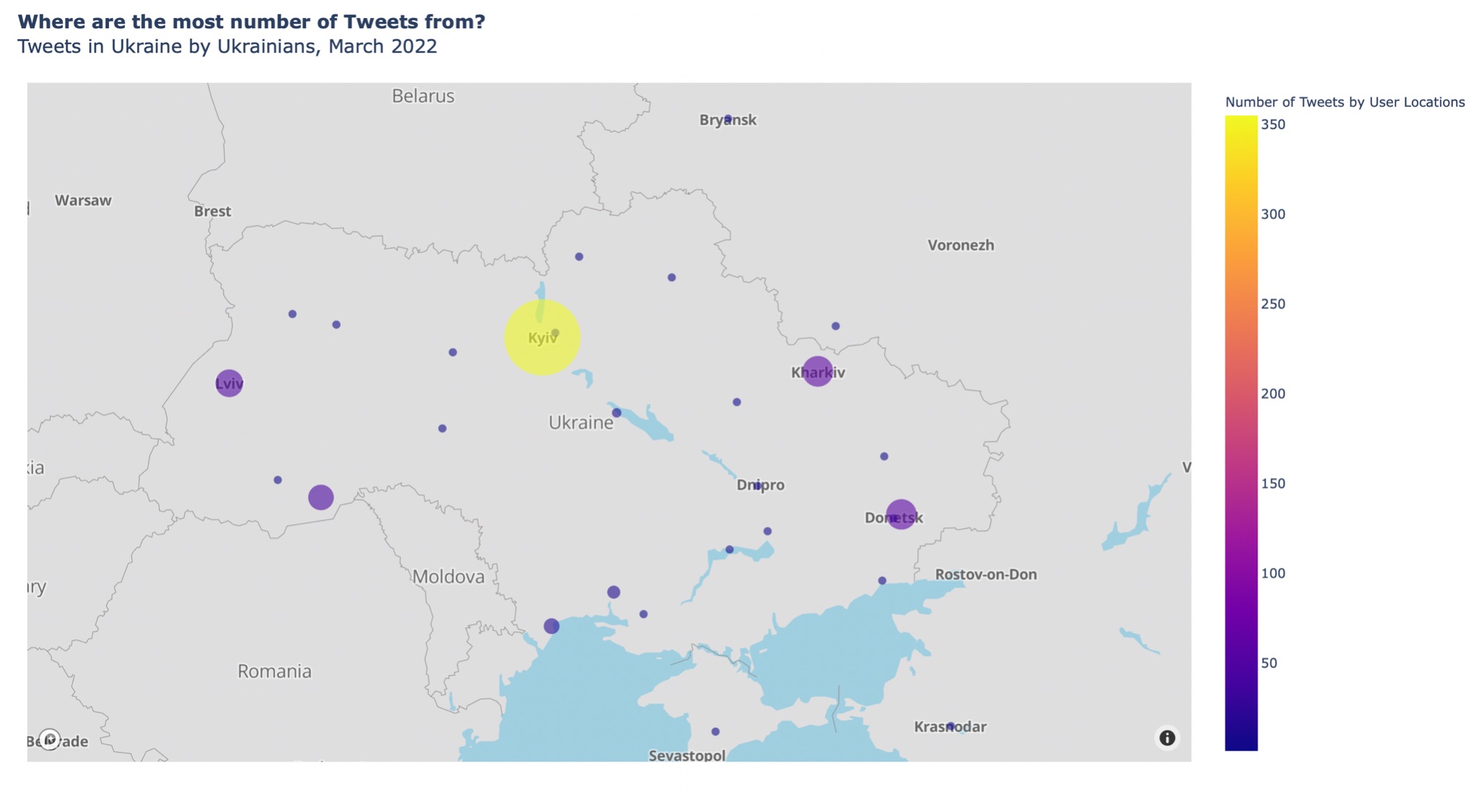
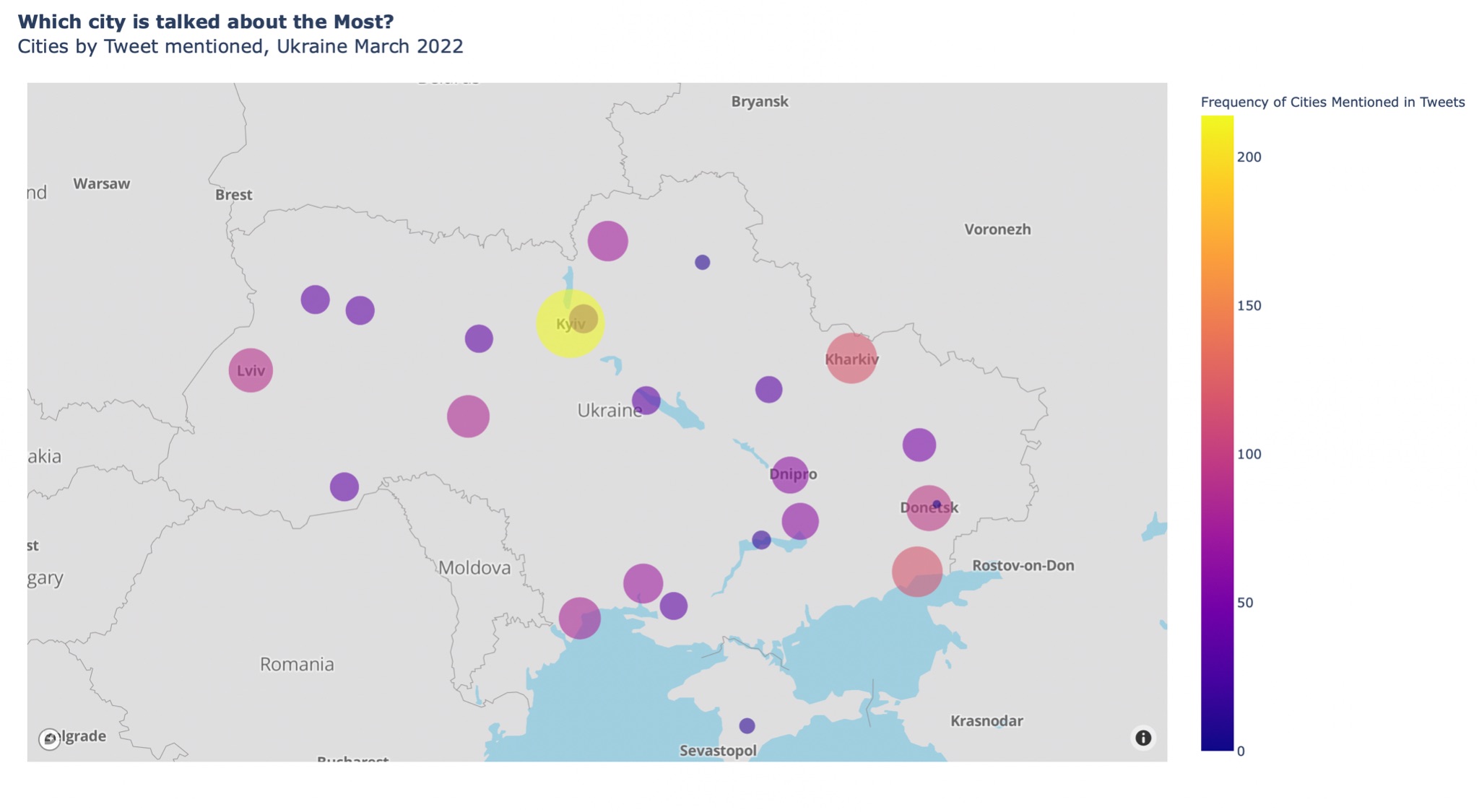


1. Timeline of Tweets by sentiment. (*QR*) (Results)



1. Engagement Statistics using (*ggplot2*) (Results)



1. Maps
   1. Tweet by city - Bubble chart map
      1. Taking User Location Data: This approach requires data cleaning. For location data, mention we are taking user’s provided location, and we are trusting that data.
      2. Fetching dataset again with Geolocation: Run a code to fetch data from the full data code that you ran in twitter. Will be lengthy but will get geocoordinates. (No guarantee to work, since we have a standard twitter API)[Adds single-point latitude and longitude variables to tweets data. — lat\_lng • rtweet (ropensci.org)](https://docs.ropensci.org/rtweet/reference/lat_lng.html)
2. 
   1. City mentions in Tweets - Bubble chart map
      1. Using string, we can run code on all texts in tweets and get a count of matching city names in all the tweets, plot the count as bubbles in the map. [Detect the presence or absence of a pattern in a string — str\_detect • stringr (tidyverse.org)](https://stringr.tidyverse.org/reference/str_detect.html). We have two methods to do this:
         1. Total times mentioned in the dataset
         2. Total number of tweets in which it is mentioned in dataset (limiting the mentions to be one per tweet) (Lengthy process)

To manage figure size in rmarkdown: [Different ways to set figure size in RMarkdown – Sebastian Sauer Stats Blog](https://sebastiansauer.github.io/figure_sizing_knitr/)

**Observations**:

Data containing languages that were a part of the data and not English were removed.

Results:

Future Research:

**References**:

Conger, K., & Raj, S. (2022, March 29). Pro-Russia Sentiment on Indian Twitter Draws Scrutiny. *The New York Times*. <https://www.nytimes.com/2022/03/29/technology/twitter-russia-india.html>

Drieger, P. (2013). Semantic Network Analysis as a Method for Visual Text Analytics. *Procedia - Social and Behavioral Sciences*, *79*, 4–17. <https://doi.org/10.1016/j.sbspro.2013.05.053>

Lang, S. (2020). Writing Analytics: Expanding the Body of Knowledge. The *Journal of Writing Analytics*, *4*(1), i–iv. <https://doi.org/10.37514/jwa-j.2020.4.1.01>

Redondo, T., & Sandoval, A. M. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, *3*(Special Issue on Big Data and AI). <https://www.ijimai.org/journal/bibcite/reference/2540>

Xing, F. Z., Cambria, E., & Welsch, R. E. (2017). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, *50*(1), 49–73. <https://doi.org/10.1007/s10462-017-9588-9>