

Earthquakes and Tweets: Turkey

Leslie Medina Felix Wu Merve Dumlu
Stacy Deng Leo Ng Mianbo Hu
University of California, Los Angeles Department of Statistics

Introduction

Background:
On February 6, 2023, a devastating earthquake of magnitude 7.8 struck southern and central Turkiye(Turkey) and northern and western Syria. Many aftershocks happened not long after, leaving thousands of buildings collapsed and tens of thousands buried under the rubble.

Hypotheses:

Null hypothesis: The number of retweets and likes doesn’t have relationship to the use of the most frequent words:

Alternative hypothesis: The more retweets + likes (visibility) you get, the more likely you are to have used the most frequent keywords(Inside Turkey):

Goal:
To identify tweets that get a better reach/visibility for earthquake relief.

Cleaning up the Data

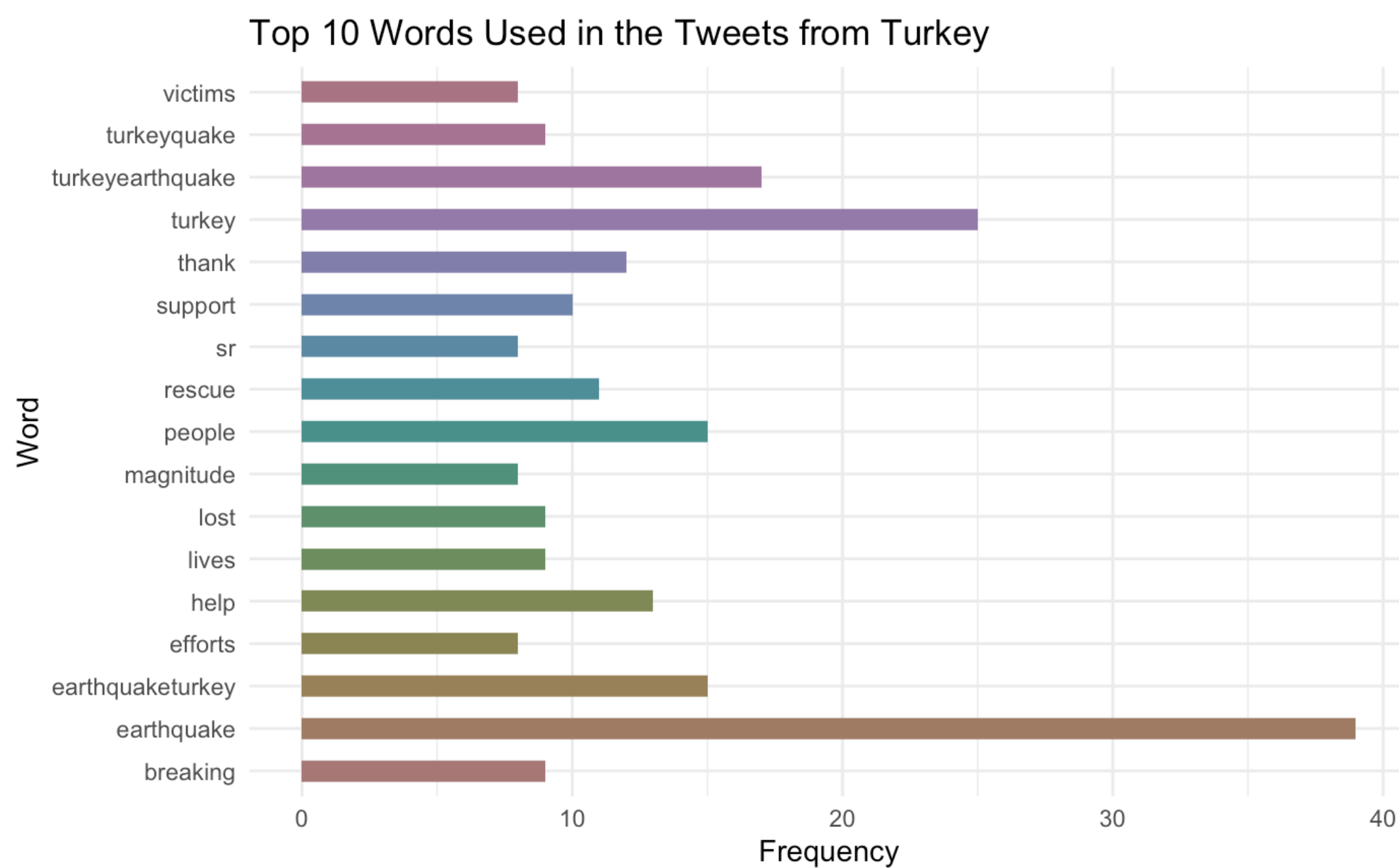
- We found this data from Kaggle.
- We create a new variable for tweet time (in 24 hour format) for a specific date in February 2023. And all the days are in February 2023, with a date range of Feb 7 to Feb 21.
- We also remove unnecessary columns and create 3 different variables (Day , hour , country)
- Created new dataset for all countries identified from user location (1705 rows)
- Created new dataset for only user location in cities from Turkey (133 rows)

Tweets Frequency

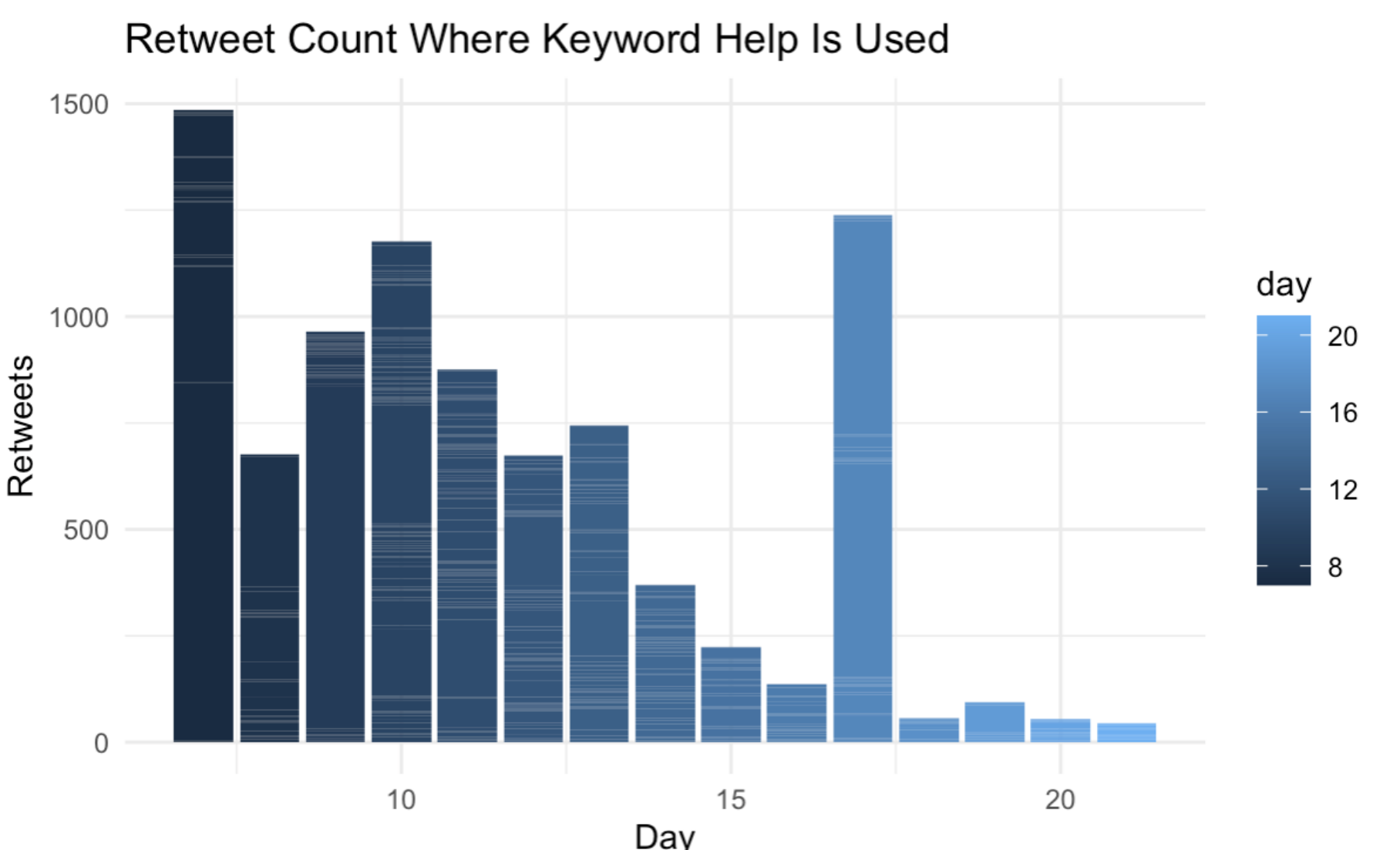
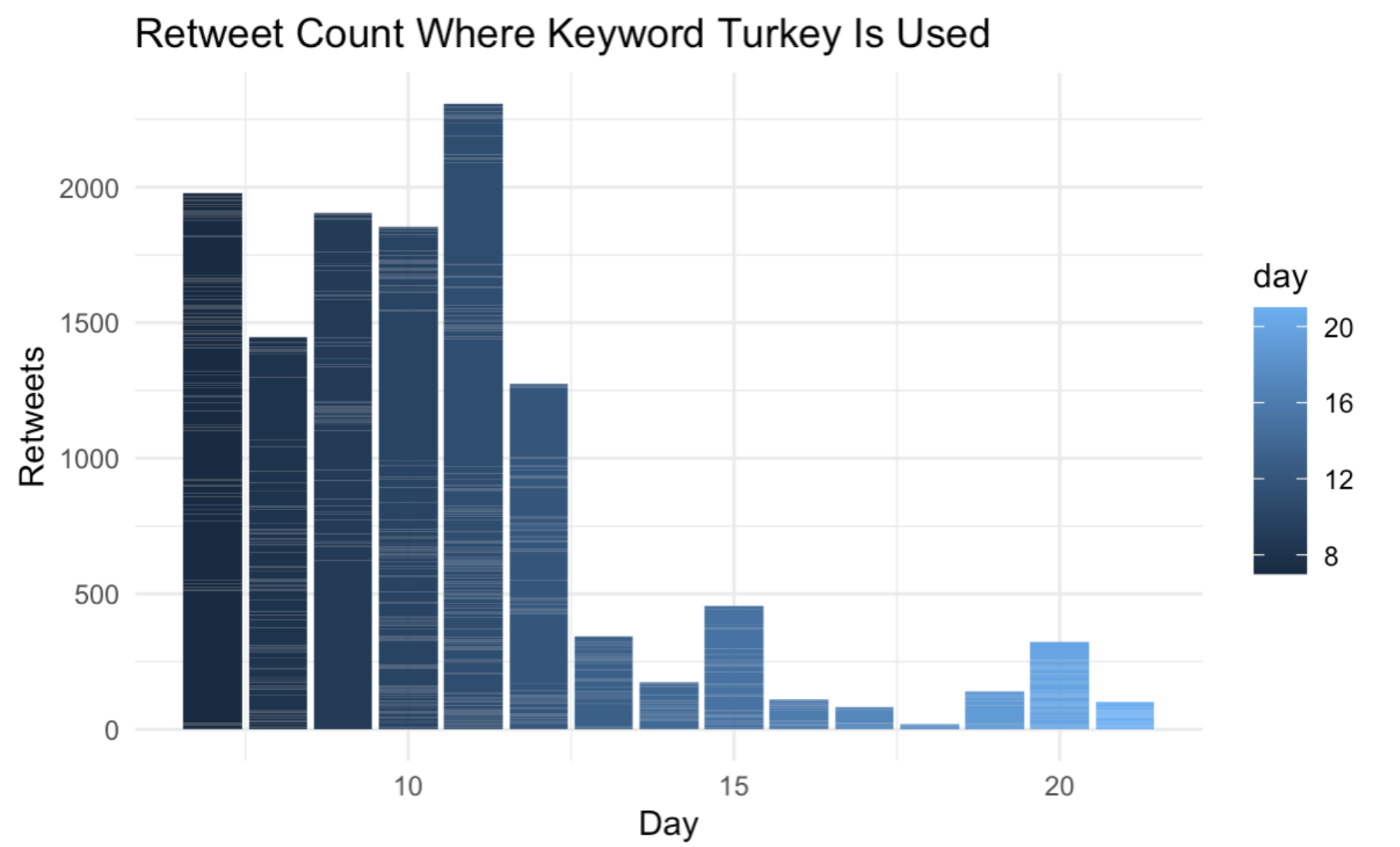
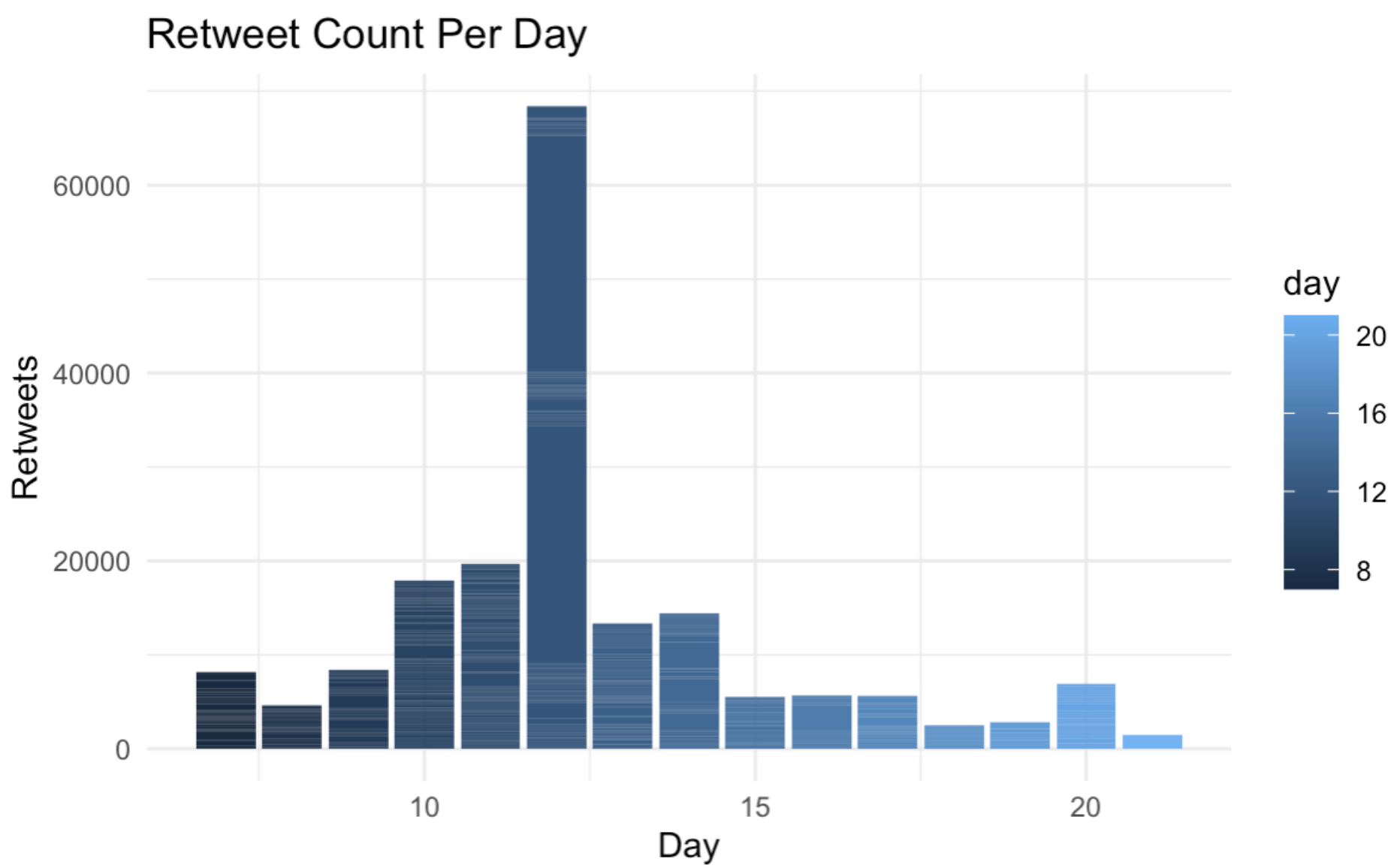
About table: The table below shows the most frequently used words, the city that tweets the most, and the city with the most favorites.

Table 1: Turkey Tweets: Most Frequent

Word	City:Tweets	City:Favorites
Earthquake	Istanbul	Istanbul

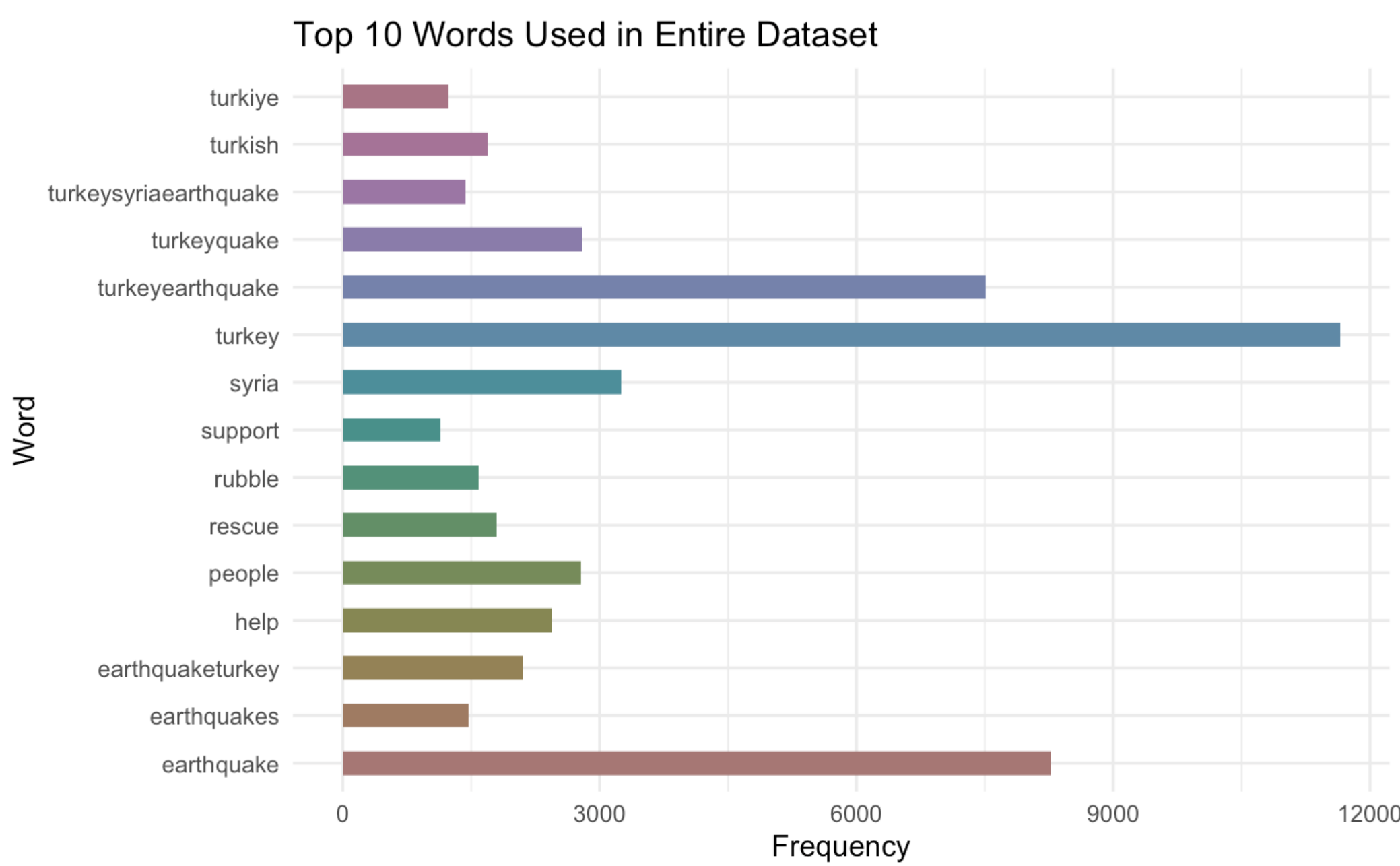


Plots



About the Plots

- Using library(qdap), we found most frequent terms used in all datasets removed unnecessary terms (i.e. the, a, in, ...)
- Using data[grep(“keyword”, data\$text) == TRUE,]
- we subset to find all texts where the keyword was used. Then compared retweet counts of each row where the keyword was used



T-tests

Welch Two Sample t-test

data: data\$retweets by data\$country_2
t = 3.8378, df = 1650.1, p-value = 0.0001288
alternative hypothesis: true difference in means between group Other country and group Turkey is not equal to 0
95 percent confidence interval:
2.176748 6.727545
sample estimates:
mean in group Other country 5.692748 mean in group Turkey 1.240602

Conclusion

In Our Project

According to our t test, we fail to reject, which means that we cannot conclude that there is a relationship between the most frequently used keywords and getting more retweets and likes. However, a result of ”Fail to Reject” does not necessarily mean that there is no relationship between the variables of interest. Such a relationship may exist, but is not large enough to be detected by the sample size or the statistical tests used.

References

Preda, G. (2023) Turkey earthquake tweets, Kaggle. Available at: <https://www.kaggle.com/datasets/gpreda/turkey-earthquake-tweets?resource=download> (Accessed: March 16, 2023).