Title: The Sentiment Analysis of the Ukraine War

Participants:

- Shiyun Li
- Peiyuan Huo
- Jun Ma (jm5309@columbia.edu)
- Kimberly Shan(xs2428@columbia.edu)

Techniques: tidyverse, ggplot2, ggmap, tmap

Brief description:

The purpose of this study is to analyze public reactions to the Ukraine war on social media. The Ukrainian armed conflict first erupted in early 2014 and quickly descended into a prolonged stalemate. On February 24, 2022, Russia launched a full-scale military invasion of Ukraine.

Under such background, this study was conducted to answer the following two questions. First, we want to explore the historical conflict between Russia and Ukraine. Based on our research to date, we decided to use the Tidyverse and maps in R to perform an exploratory analysis of the historical conflict between the self-declared Donetsk and Luhansk People's Republics and the Ukrainian government.

In addition, the second purpose of this project is to examine public reactions to the war in Ukraine on social media, using sentiment analysis. For example, we want to use NLP techniques to see the most common topics and words in the tokens. Also, we want to do a polarity analysis to determine the attitude or sentiment of the post.

Data:

- Data on Ukraine conflict events: https://data.humdata.org/dataset/ucdp-data-for-ukraine

  This data covers individual events of organized violence, and these occurrences are fine-grained enough to be geo-coded down to individual villages, with temporal durations broken down to single days.

- Weekly dataset on Ukraine conflict events: https://data.humdata.org/dataset/ukraine-acled-conflict-data

  The data provides the total number of reported political violence, civilian-targeting, and demonstration events in Ukraine.

- Reddit, try to extract text data on public's opinions relate to this conflict using sentiment keywords:

Posts about Ukraine-Russia conflict, 9.1k comments:
https://www.reddit.com/r/europe/comments/t00aon/ukrainerussia_conflict_megathread_6_live_thread/

Posts about Russia invades Ukraine, 14k comments:
https://www.reddit.com/r/europe/comments/t22yw6/russia_invades_ukraine_megathread_iv_posting/

Posts about Putin warns Europe will be dragged into military conflict if Ukraine joins NATO, 4.4k comments:
https://www.reddit.com/r/worldnews/comments/sn32xp/russian_president_vladimir_putin_warns_europe/

- Twitter
  Ukraine Conflict Twitter Comments:1.2M distinct tweets about the current ongoing Ukraine-Russia conflict.
  https://www.kaggle.com/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows/code

- GIS Data
  https://hub.arcgis.com/datasets/dd3afd55a8fc428daef8ae395e8cd582_0/explore

Types of visualizations:
Line graph: line chart is applied to show the GDP per capita of Russia and Ukraine in recent years. To make the compassion more clear, the Y axis on both sides represents different countries.

Tmap: We will use tmap to present the population distribution of two countries.

Histogram: This function mainly shows the numbers of Ukrainian and separatist forces killed in wars. We plan to use “dodge” to compare the two forces directly.

Wordmap: World cloud is commonly used to highlight popular or trending terms based on the frequency of use and prominence. We decided to create word clouds to show the most prominent words within the dataset.