

# RE03-Ukraine War

## Group 05

Xingwen Xiao  
Dingran Qi  
Lixiang Zhang

## Our tasks

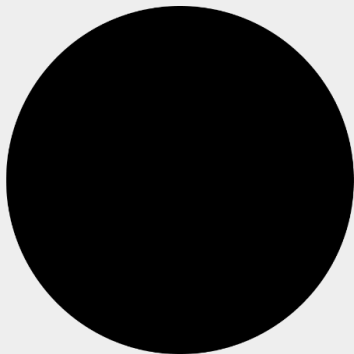
- Count the number of posts and posters on this topic overtime
- Sentiment analysis
- Political standpoints classification
- Display how political standpoints change overtime
- Summarize the topic with word clouds
- Find the most relevant subreddits and the most prolific posters overtime (classify based on political standpoint)

## Basic exploration on the dataset

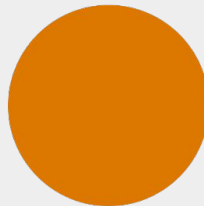
- Size of the dataset
- Remove useless attributes and write to parquet
- Find related submissions with YakeKeywordExtraction
- Join the filtered submissions with comments
- 2,373,520 submissions and 8,666,926 comments remain

**Time Span of More  
than 8 Years**

**Feb 2014  
to  
Now**



**2.0TB of Comments**

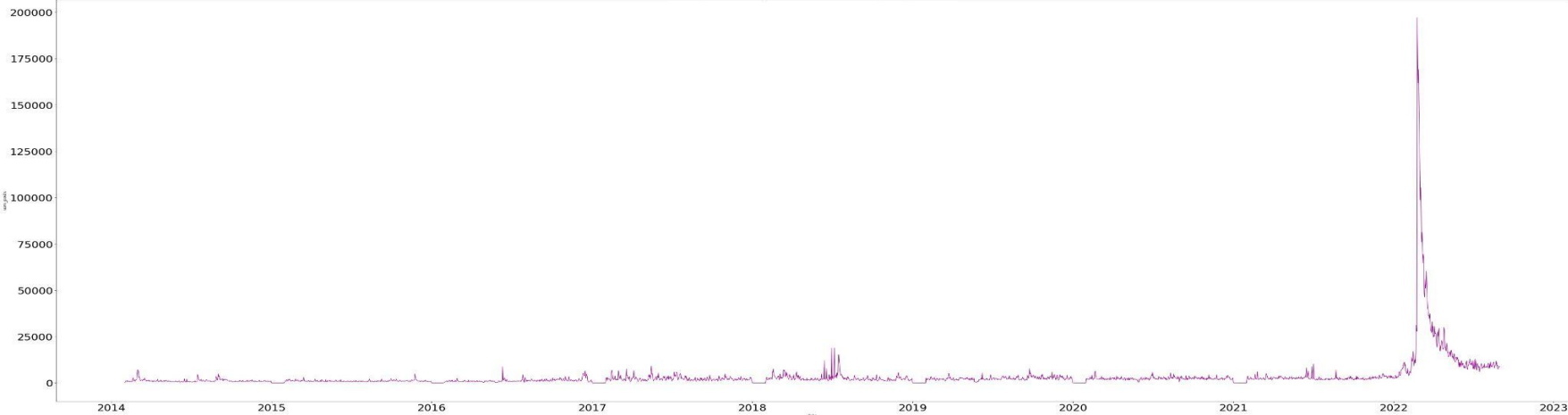


**0.8TB of  
Submissions**

## How to find related posts

- YakeKeywordExtraction:
  - Simple and easy to use, may include some of irrelevant information, but it is not prominent
  - Unsupervised keyword extraction algorithm
  - Check if the extracted keywords have intersection with our predefined keyword list
- BM25:
  - highly rely on keywords, rule out many useful information, complicated parameter calibration

Count of related posts in Reddit from 2014 to 2022

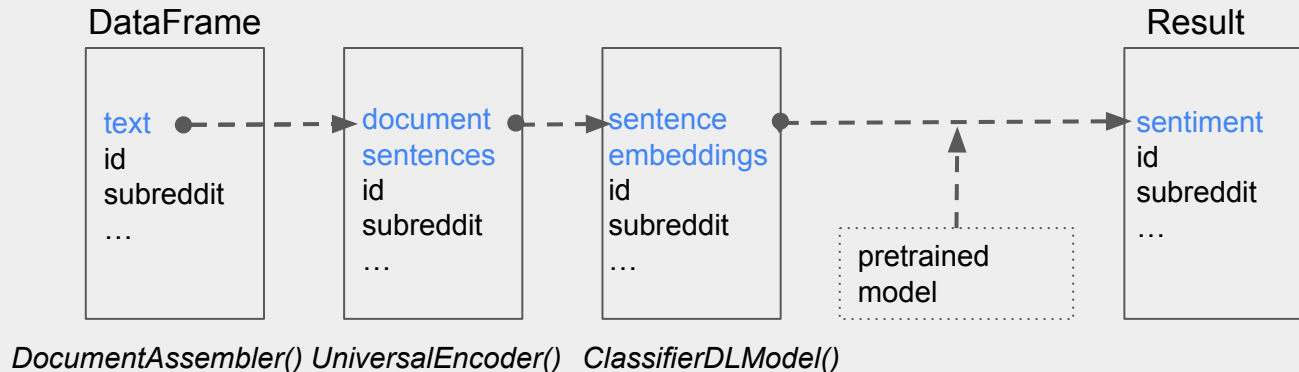


Count of related posters in Reddit from 2014 to 2022

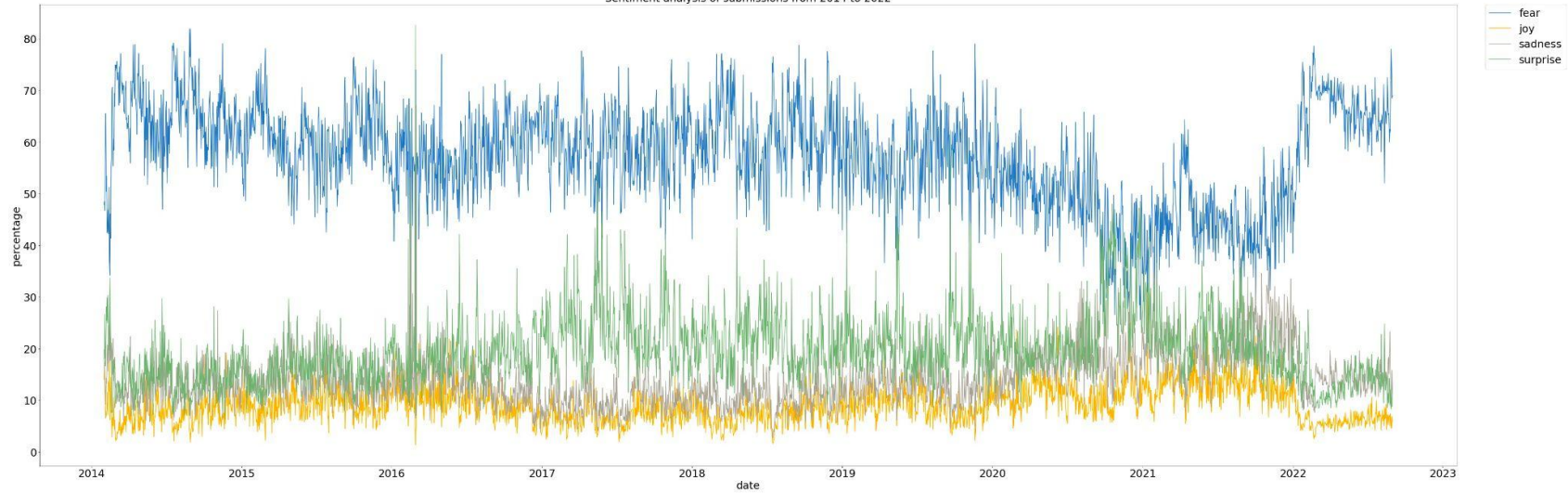


# Sentiment Analytics

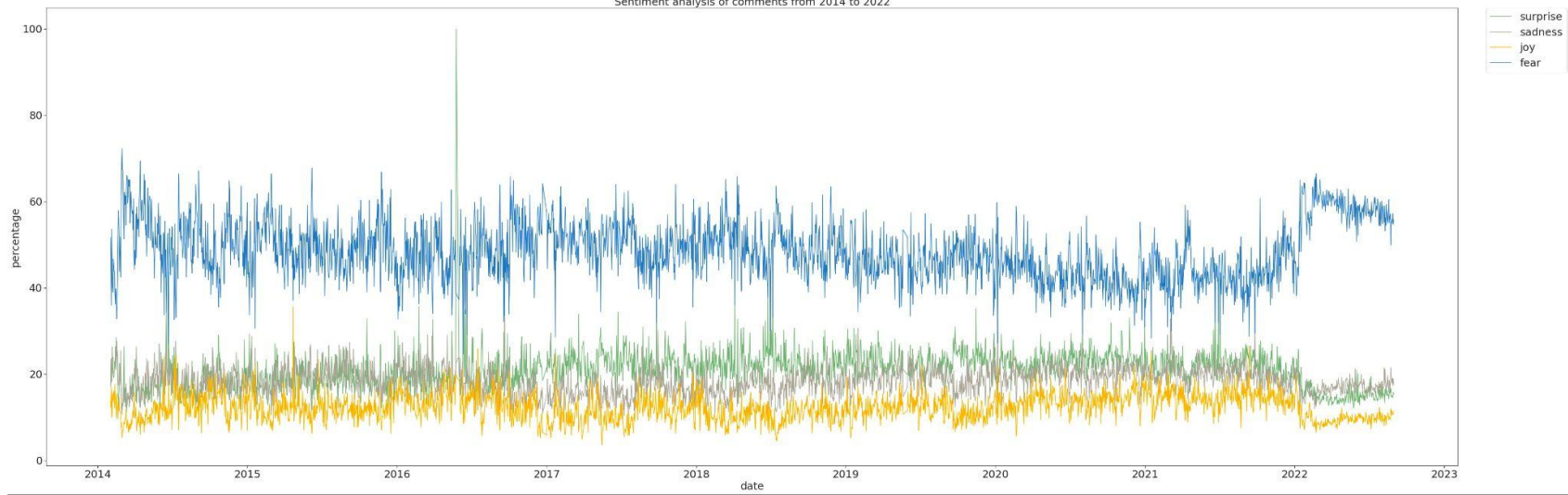
- Classified Comments & Submissions to **Fear, Joy, Surprise, Sadness**
- Pretrained DL classifier



Sentiment analysis of submissions from 2014 to 2022

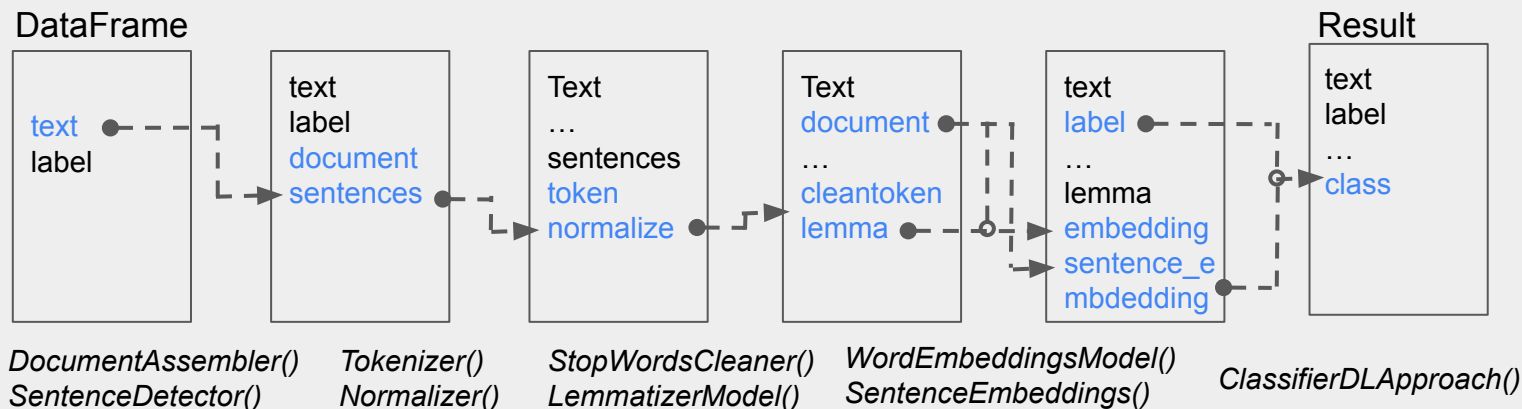


Sentiment analysis of comments from 2014 to 2022



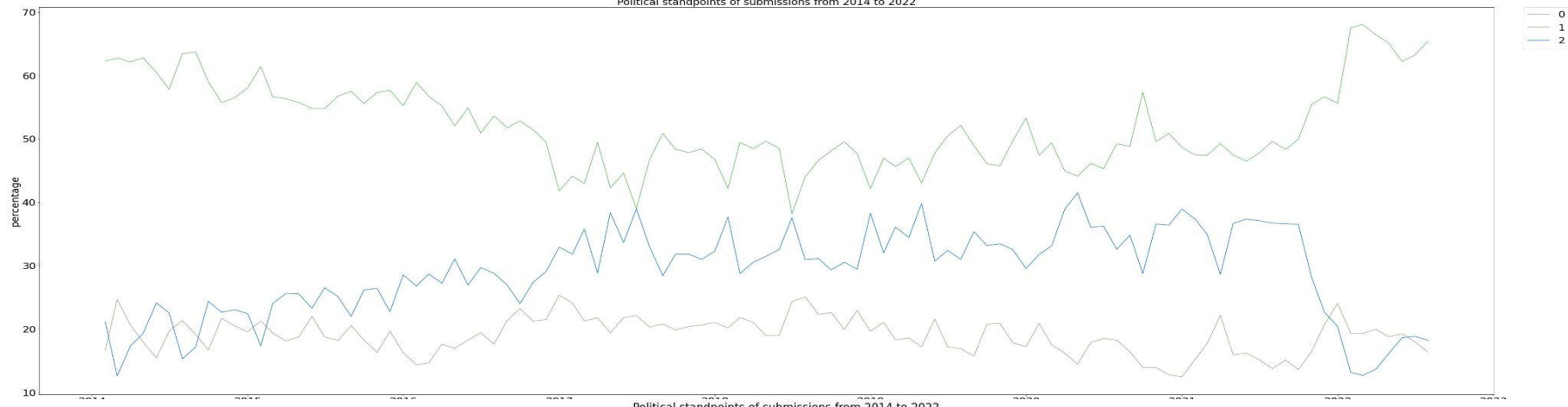
# Political standpoint classification

- Spark NLP pipeline model
- Manually label standpoints: **3540**
- Tripal classification with **Acc 76~78%**
- Pro-Ukraine: **15.4%** of comments & **54.7%** of submissions
- Pro-Russia: **8%** of comments & **19.3%** of submissions

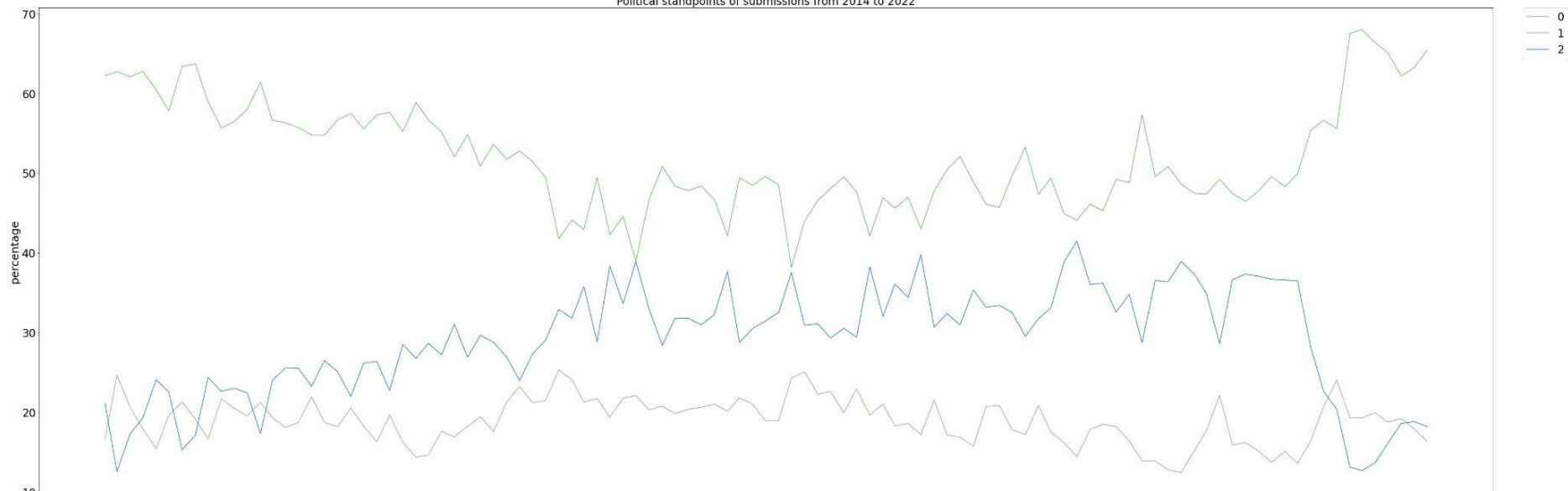




Political standpoints of submissions from 2014 to 2022



Political standpoints of submissions from 2014 to 2022



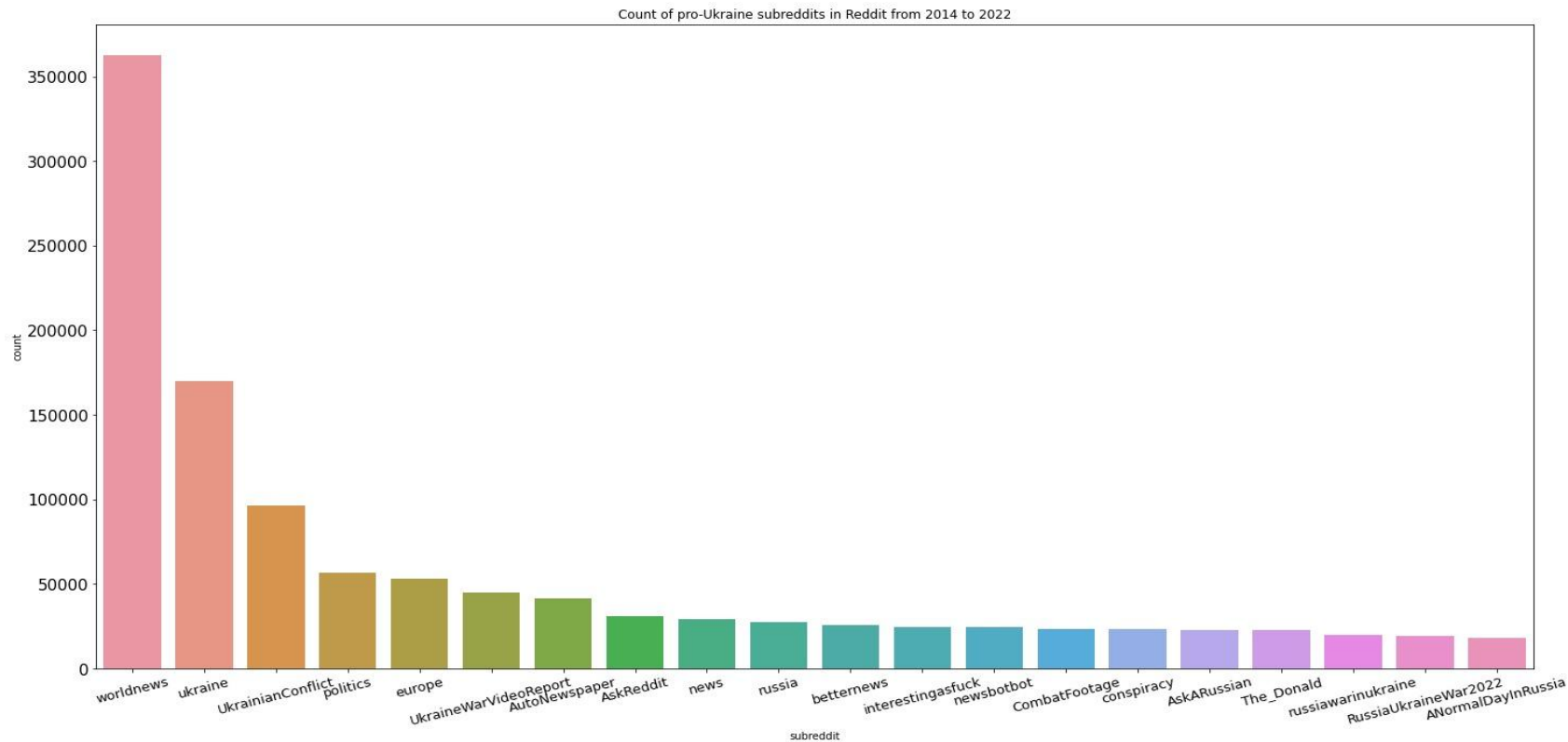
# Word cloud

- Classified by political standpoint
- For instance, word cloud of pro-Ukraine comments

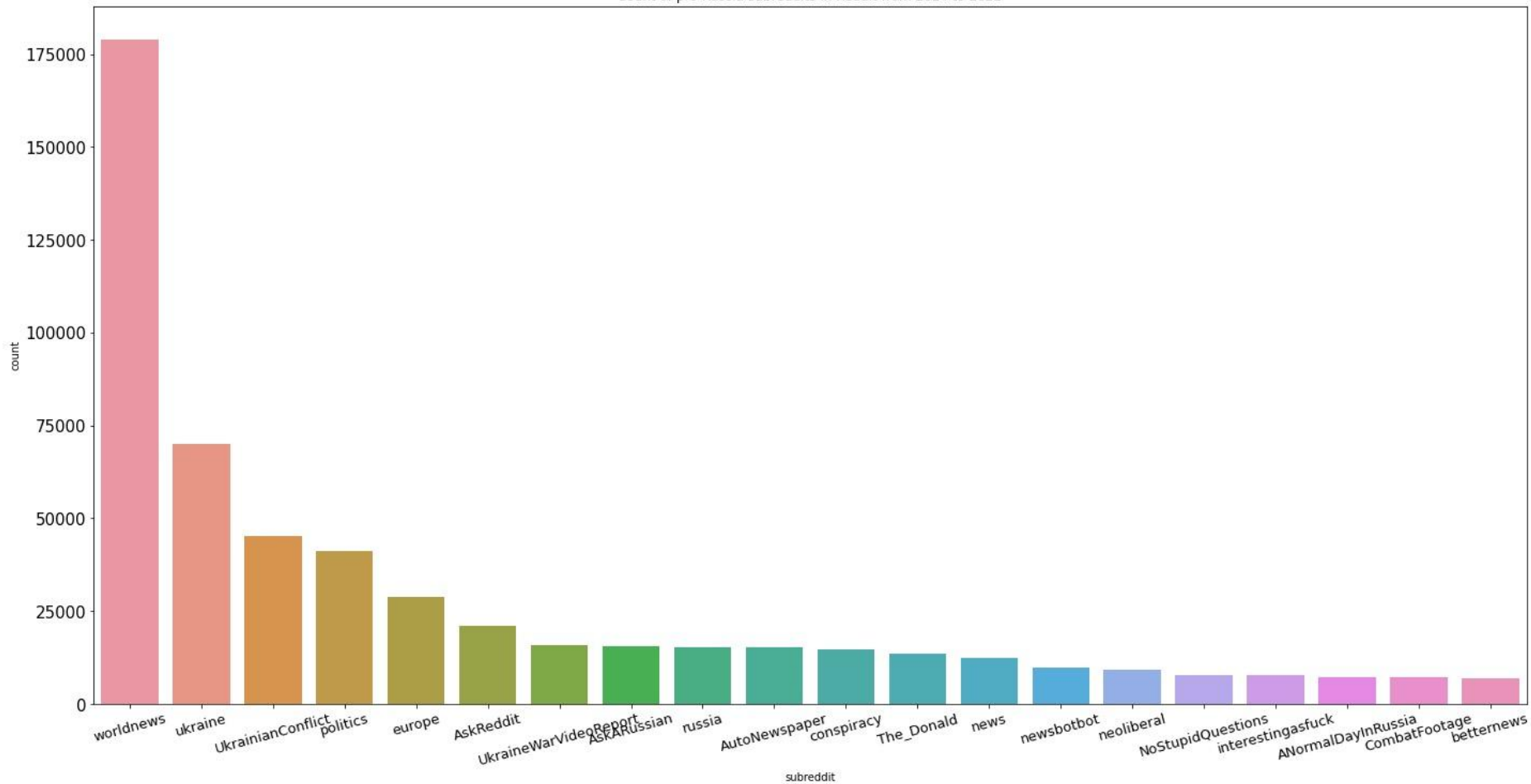


# The most relevant subreddits and prolific posters

- Also classified by political standpoints

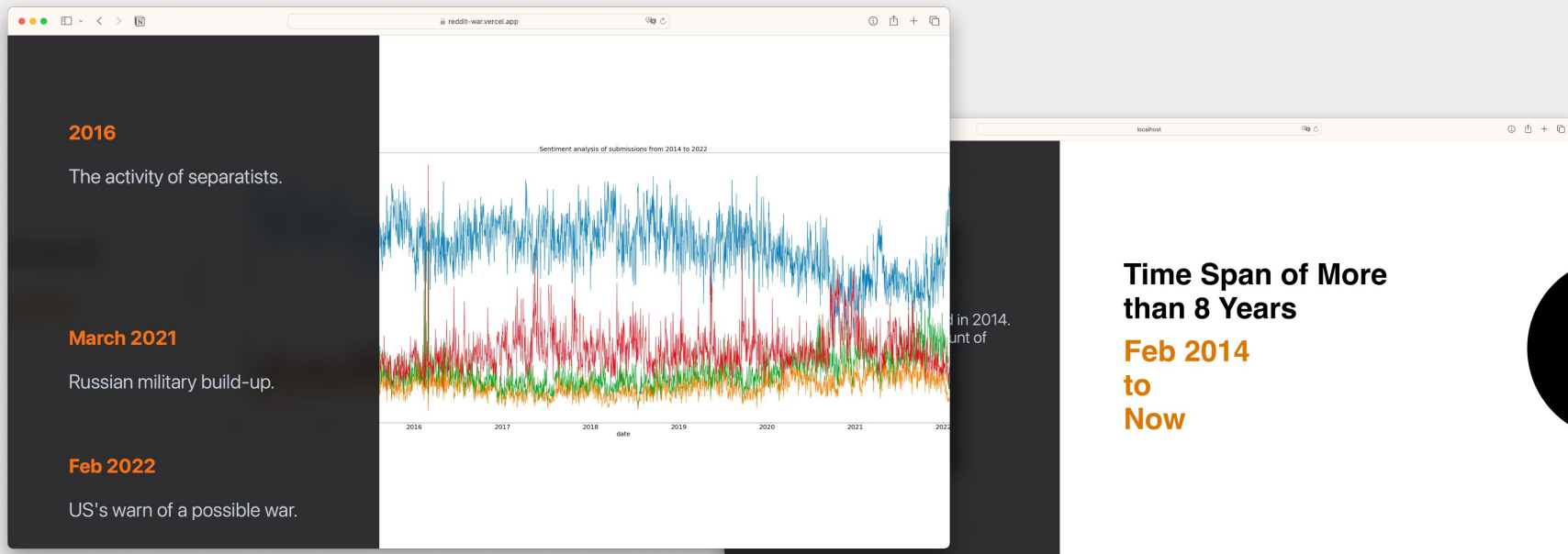


Count of pro-Russia subreddits in Reddit from 2014 to 2022



# Website

- Scrollytelling: Scroll web page to go through data and history events



**Thanks! Bedankt!**