# 2022 Russo-Ukrainian Crisis

**Audrey Lin**
**Donggyu Kim**
**Soojung Choi**

# Motivation / Background

- Social Movements and Political Discourse
  - How can we amplify important conversation
- What conversations are going on Twitter
- What topics encourage more engagement
- Audience sentiment regarding the Russo-Ukrainian Crisis

# Research Question & Hypothesis - Midterm

- RQ1 : Do high engagement tweets share a common sentiment compared to low engagement tweets? What topics do people care about?

- H1: There is a difference between topics discussed in high engagement tweets versus low engagement tweets.

# Research Question & Hypothesis - Final

- RQ2 : Does the source/agent's political leaning have correlation with engagement and sentiment regarding the Russo-Ukrainian Crisis?

- H2: Tweets created by users with stronger political leaning have **higher** engagement and **stronger** sentiment

# Approach - Dataset

- Wanted to look at #StopAsianHate (March 2021) tweets
  - Restricted to only most recent tweets
  - Not enough volume (less than 1k per day)

- Changed the topic to Russo-Ukrainian Crisis
  - Current event with lots of recent tweets
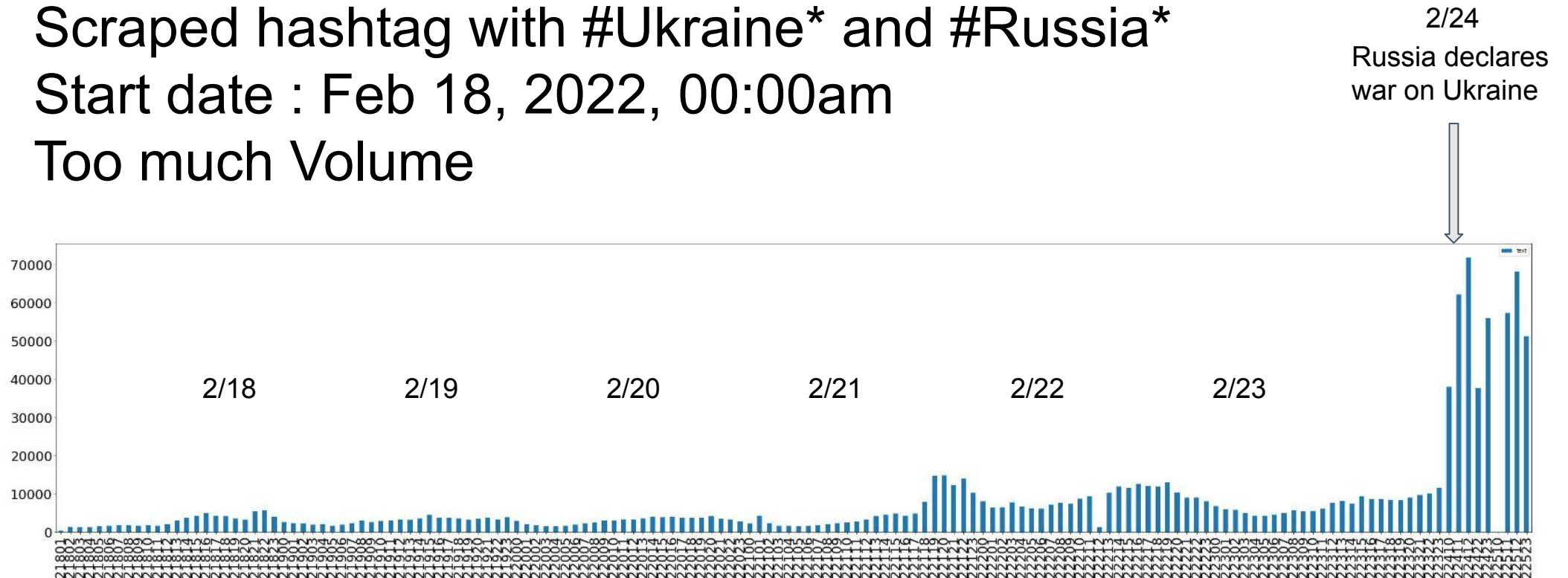
# Approach - Collection

- Scraped hashtag with #Ukraine* or #Russia*
- Start date : Feb 18, 2022, 00:00am

```
220226_125912 : 1002 tweets in dataframe at 220302_234741
220226_125823 : 2009 tweets in dataframe at 220303_000011
220226_125736 : 3009 tweets in dataframe at 220303_001234
220226_125649 : 4000 tweets in dataframe at 220303_002458
220226_125559 : 5004 tweets in dataframe at 220303_003731
220226_125509 : 6004 tweets in dataframe at 220303_005007
220226_125415 : 7000 tweets in dataframe at 220303_010249
220226_125320 : 8003 tweets in dataframe at 220303_011549
220226_125225 : 9001 tweets in dataframe at 220303_012843
220226_125128 : 10007 tweets in dataframe at 220303_014133
./Tweets/ukraine_220226_125128.csv saved
./Tweets/names_220226_125128.csv saved
:: Time stamp 2022-02-26T12:51:28.000Z with 10007 tweets
```

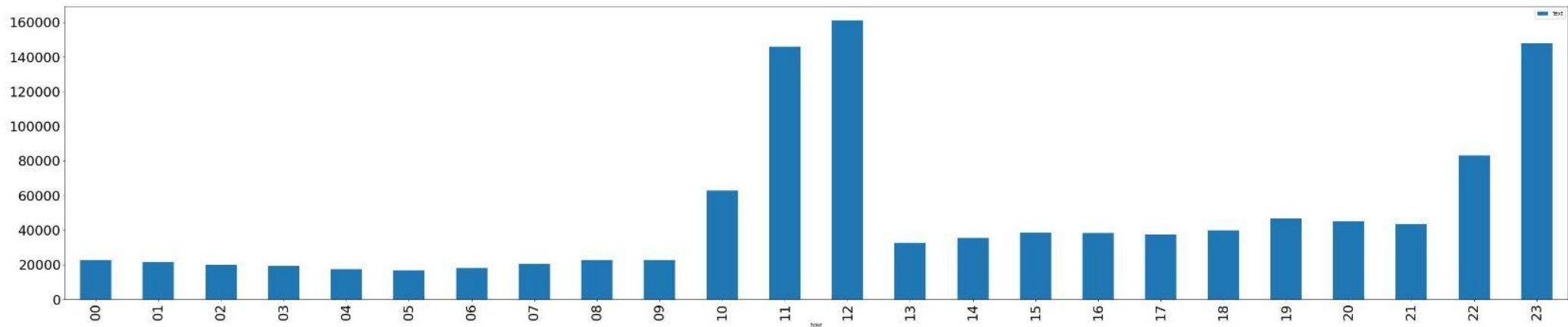# Approach - Collection

- Scraped hashtag with #Ukraine* and #Russia*
- Start date : Feb 18, 2022, 00:00am
- Too much Volume



2/24
Russia declares
war on Ukraine

# Approach - Collection

- Scraping could not follow the amount of tweet generated
- Decided to scrape 11am - 1pm & 11pm - Midnight (3hr / day)

# Approach - Collection

- Approx. 1M amount of Twitter data (Midterm Analysis)
- Currently around 1.5M Twitter data
- Challenge (Time Sensitive)

**Projects**

**Comm 557**                                                  ESSENTIAL

MONTHLY TWEET CAP USAGE ⓘ

                                                                          99%

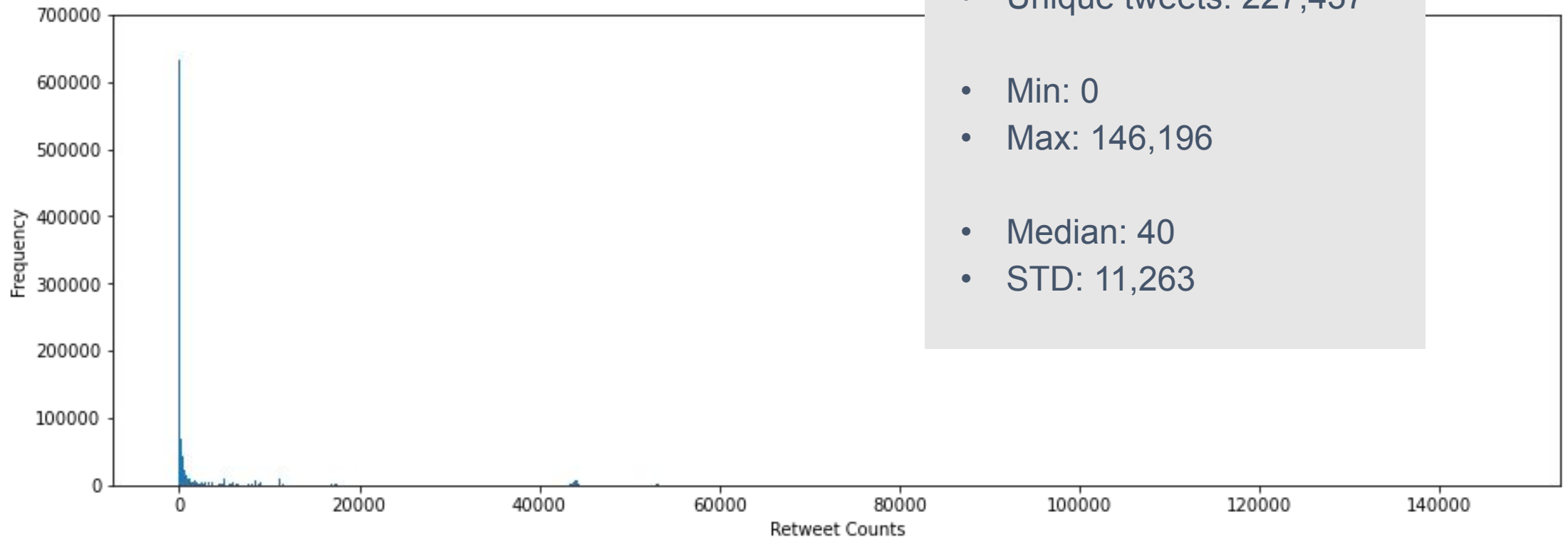493,486 Tweets pulled of **500,000**          Resets on March 11 at 00:00 UTC

# Data cleaning

- Remove hashtags, mentions, punctuation
- Reformat time feature
- Tokenize and stem

# Descriptive Statistics



- Total tweets: 995,282
- Unique tweets: 227,437

- Min: 0
- Max: 146,196

- Median: 40
- STD: 11,263

USC Annenberg
School for Communication
and Journalism

# Sentiment Analysis

| SentiStrength | Low engagement (≤40) | High engagement (40-10k) | Top (>10k) |
|:---:|:---:|:---:|:---:|
| -4 | 582 | 21 | |
| -3 | 10,121 | 457 | 1 |
| -2 | 38,512 | 1,691 | 2 |
| -1 | 25,236 | 968 | 4 |
| 0 | 107,418 | 3,737 | 10 |
| 1 | 28,861 | 1,153 | 4 |
| 2 | 7,861 | 310 | |
| 3 | 442 | 19 | |
| 4 | 23 | 2 | |
| **Total** | **219,057** | **8,359** | **21** |

USC Annenberg
School for Communication
and Journalism

# Sentiment Analysis

| SentiStrength | Low engagement (≤40) | High engagement (40-10k) | Top (>10k) |
|:---:|:---:|:---:|:---:|
| -4 | 0% | 0% | |
| -3 | 5% | 5% | 5% |
| -2 | 18% | 20% | 10% |
| -1 | 12% | 12% | 19% |
| 0 | 49% | 45% | 48% |
| 1 | 13% | 14% | 19% |
| 2 | 4% | 4% | |
| 3 | 0% | 0% | |
| 4 | 0% | 0% | |

# Sentiment Analysis

**Top tweets**

>10k retweets

55,908 tweets

21 unique tweets

Mean: -0.333

STD: 1.039

| Tweet | SS |
|---|---|
| Geography can be tough Heres a guide for Russian soldiers who keep getting lost amp accidentally entering | -1 |
| The recognition of the two separatist territories in is a blatant violation of international law the territorial integrity of Ukraine and the agreements The EU and its partners will react with unity firmness and with determination in solidarity with Ukraine | -3 |
| If youre gonna listen to any speech about  let it be this one The Kenya ambassador to the UNSC perfectly explains how people across Africa understand Ukraine and what the Kremlins acts of aggression mean in our postcolonial world | -1 |
| If youre an American and your not tracking on now is the time to start This is just not another foreign policy blip This is potentially bigger amp more deadly than | 0 |

# Sentiment Analysis

**Top tweets**

>10k retweets

55,908 tweets

21 unique tweets

Mean: -0.333

STD: 1.039

Geography can be tough. Here's a guide for Russian soldiers who keep getting lost & 'accidentally' entering #Ukraine

The recognition of the two separatist territories in #Ukraine is a blatant violation of international law, the territorial integrity of Ukraine and the #Minsk agreements.

The EU and its partners will react with unity, firmness and with determination in solidarity with Ukraine.

If you're gonna listen to any speech about #Ukraine 🇺🇦 , let it be this one.

The Kenya ambassador to the UNSC perfectly explains how people across Africa understand Ukraine, and what the Kremlin's acts of aggression mean in our post-colonial world.

# Sentiment Analysis

**Top tweets**

>10k retweets

55,908 tweets

21 unique tweets

Mean: -0.333

STD: 1.039

Geography can be tough. Here's a guide for Russian soldiers who keep getting lost & 'accidentally' entering #Ukraine

RUSSIA

NOT RUSSIA

👤 Canada in Ukraine

# Sentiment Analysis

## High engagement tweets

40-10k retweets

441,656 tweets

8,359 unique tweets

Mean: -0.475

STD: 1.242

| Tweet | SS |
|---|---|
| state media has been promoting this idea of genocide in for quite some time now thread Dont assume anything is off the table while the Kremlin continues to pursue this narrative as an attempted justification for its future plans | 0 |
| Good Morning  New   Thread | 1 |
| Fresh from no change in course according to latest info amp signs This is not encouraging We must be ready for all scenarios but remain in all our tracks diplomacy amp deterrence I cannot stress enough our utmost support to  we stand in unity 13 | 1 |
| Sanctions package ready to hit hard  it will hurt its economy will act fast if  escalates situation I believe it will be the mother of all sanctions   already loses part of GDP every year after 14 sanctions this time the cost will be much much higher 23 | -2 |

# Sentiment Analysis

**Low engagement tweets**

<40 retweets

497,718 tweets

219,057 unique tweets

Mean: -0.406

STD: 1.186

| Tweet | SS |
|---|---|
| The life that has passed Without finding a Satguru has been in vain Visit Sant Rampal Ji Maharaj YouTube channel | -2 |
| From my colleague in | 0 |
| Just to note Russia continues to say theyre removing troops from the Ukrainian border The US says they only continue to add The below statement has yet to be verified by US officials | 0 |
| PM ahead of todays meetings in Brussels the only one who can deescalate is Russia has created this situation amp can also take a step back you cant have negotiations at gunpoint | 0 |
| US Secy Of State Blinken agrees to meet Russian Foreign Minister Lavrov towards end of next week | 0 |
| has responded with proposed dates for late next | 0 |

# Sentiment Analysis

| | Low engagement | High engagement | Top |
|---|---|---|---|
| **Total tweets** | 497,718 | 441,656 | 55,908 |
| **Unique tweets** | 219,057 | 8,359 | 21 |
| **Sentiment: Mean** | -0.406 | -0.475 | -0.333 |
| **Sentiment: STD** | 1.186 | 1.242 | 1.039 |

# Topic Classification

- Building a model classifying the impression level of a text based on the terms used
  - Input: Tweet text
  - Output: Top, High, Low engagement

- Used models: Logistic regression, Random Forest Classifier
  - TFIDF Vectorizer is applied
  - Resampled to 40k tweets each for balanced classification

# Topic Classification

- Importance of features - Logistic regression
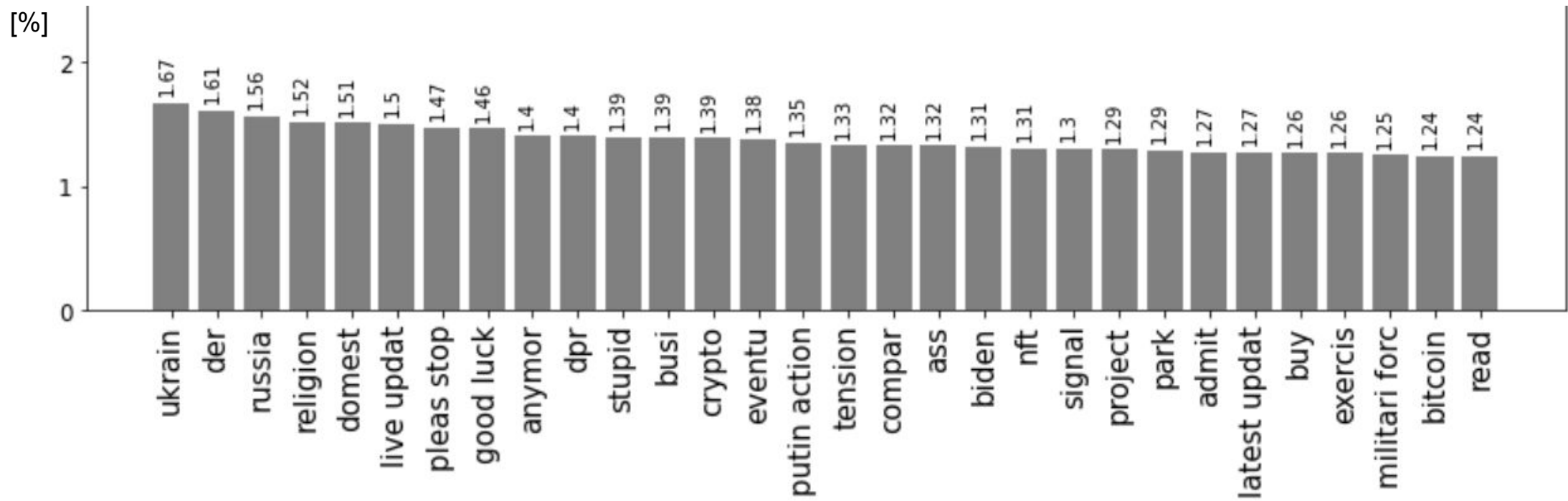
# Topic Classification

- Importance of features - Random Forest Classifier

# Topic Classification

- Logistic Regression train score: 0.9, test score: 0.85

Vector count in corpus

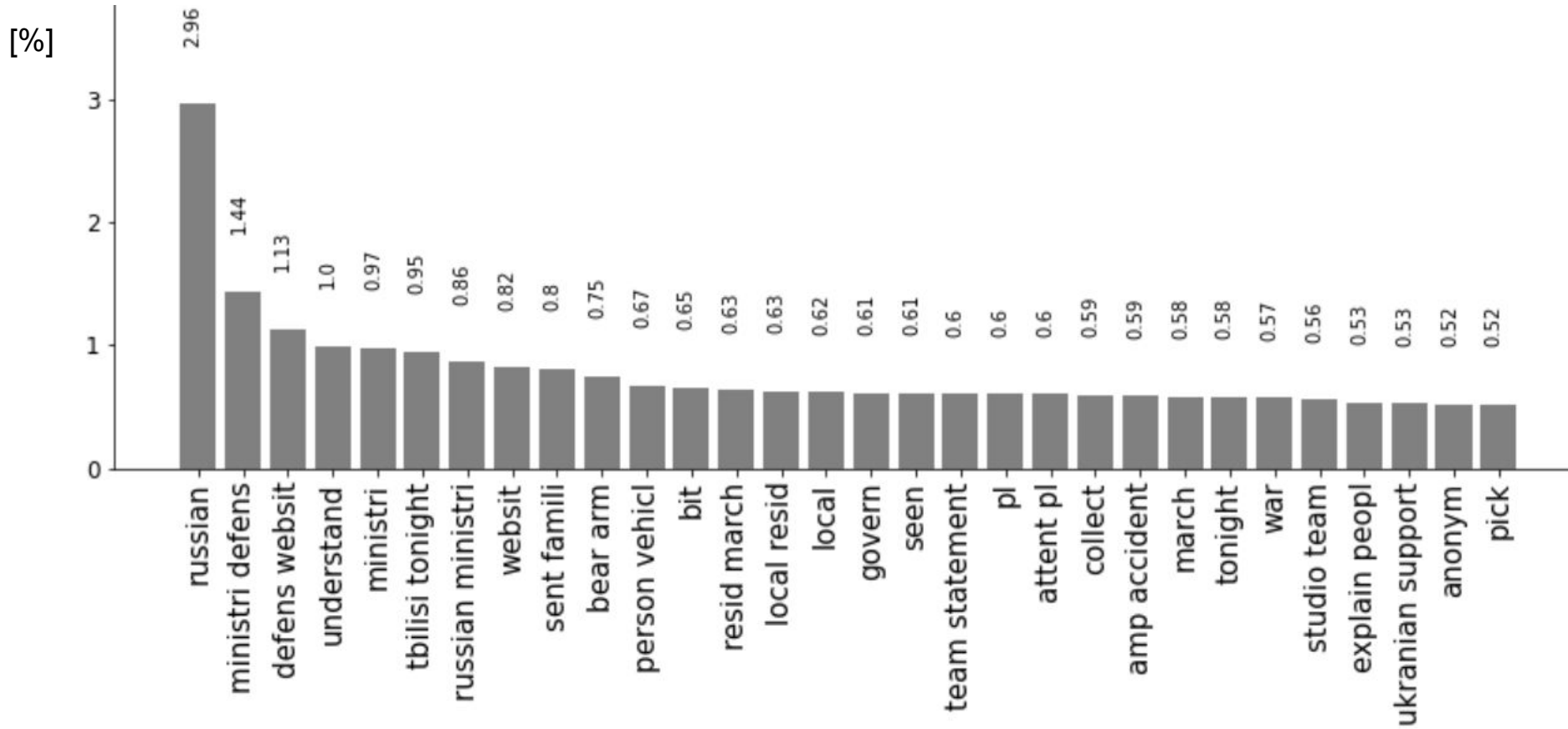| vector | ukrain | der | russia | religion | domest | live updat | good luck | anymor | dpr | stupid | busi | crypto | eventu |
|--------|--------|------|--------|----------|--------|------------|-----------|--------|-----|--------|------|--------|--------|
| top | 0 | 7702 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| high | 88 | 5825 | 172 | 0 | 7 | 9 | 0 | 7 | 0 | 24 | 87 | 35 | 6 |
| low | 150 | 5221 | 217 | 22 | 57 | 32 | 11 | 72 | 6 | 110 | 165 | 142 | 42 |

| tension | compar | ass | biden | nft | signal | project | park | admit | latest updat | buy | exercis | bitcoin | read |
|---------|--------|------|-------|-----|--------|---------|------|-------|--------------|-----|---------|---------|------|
| 0 | 0 | 1916 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 206 | 46 | 2869 | 0 | 2 | 19 | 91 | 16 | 24 | 10 | 104 | 81 | 3 | 1579 |
| 446 | 85 | 2352 | 16 | 11 | 68 | 194 | 62 | 65 | 25 | 193 | 132 | 23 | 1375 |

# Topic Classification

- Random Forest train score: 0.81, test score: 0.77

Vector count in corpus

| vector | russian | ministri defens | defens websit | understand | ministri | tbilisi tonight | russian ministri | websit | sent famili | bear arm | person vehicl | bit | resid march | local resid | local |
|--------|---------|-----------------|---------------|------------|----------|-----------------|------------------|--------|-------------|----------|---------------|-----|-------------|-------------|-------|
| top | 0 | 0 | 0 | 3793 | 0 | 0 | 0 | 3855 | 0 | 1882 | 0 | 1951 | 0 | 1865 | 1865 |
| high | 144 | 0 | 0 | 224 | 0 | 0 | 0 | 167 | 0 | 26 | 0 | 289 | 0 | 24 | 304 |
| low | 125 | 0 | 0 | 268 | 5 | 0 | 0 | 117 | 0 | 1 | 0 | 298 | 0 | 4 | 128 |

| govern | seen | team statement | pl | attent pl | collect | amp accident | march | tonight | war | studio team | explain peopl | ukranian support | anonym | pick |
|--------|------|----------------|------|-----------|---------|--------------|-------|---------|------|-------------|---------------|------------------|--------|------|
| 3770 | 3746 | 1951 | 5684 | 0 | 3832 | 1921 | 1865 | 1986 | 5715 | 0 | 0 | 0 | 0 | 1970 |
| 680 | 515 | 9 | 8975 | 0 | 114 | 0 | 131 | 703 | 5116 | 0 | 0 | 0 | 4 | 61 |
| 551 | 264 | 0 | 7553 | 0 | 109 | 3 | 40 | 236 | 4429 | 0 | 0 | 0 | 10 | 54 |

# Topic Classification

- Vectors with high importance according to the classification

**Logistic Regression: test score 0.85**

| top | der | sport | earn | san | big | ya | bs | und | new | movi | sit | et | | na | sanction | eve |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| high | ass | read | app | injur | fair | throw | highli | displac | bag | district | year ago | | ill | humanitarian assist | | lose | du |
| low | ukrain | russia | religion | domest | live updat | good luck | anymor | | dpr | stupid | busi | crypto | eventu | | tension | compar | biden |

**Random Forest: test score 0.77**

| top | understand | websit | bear arm | | bit | local resid | local | govern | seen | team statement | collect | amp accident | | march | tonight | war | pick |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| high | russian | pl | ukrainian | retweet tell | unsc | resid | | ive | defens | attent | day | | invas | arm sound | kremlin act | im | famili |
| low | ministri | anonym | russian war | russian govern | tbilisi | ukrain | | russian invas | protocol | | global impact | snapshot | start anoth | temporari | | ne border | russia | ne |

# Topic Modeling - LDA

- Building a model clustering the word vectors according to the similarities
  - Input: Tweet text, number of clusters
  - Output: Clusters and Vectors in the clusters

- Used model: Latent Dirichlet Allocation
  - TFIDF Vectorizer is applied
  - Observe topic changes along the time (Feb 18-25)

# Topic Modeling - LDA

- News breaks (Feb 18-25)

| **18** | **19** | **20** | **21** | **22** | **23** | **24** | **25** |
|---|---|---|---|---|---|---|---|
| Biden says the U.S. believes Putin has decided to **invade** Ukraine.<br><br>'A whistling sound, then an **explosion'**: Shelling hits a kindergarten in Ukraine. | President Vladimir Putin held a test of its **nuclear** readiness that included launching three missiles. | Shelling intensifies in eastern Ukraine and thousands of civilians board buses to **evacuate**. | The Kremlin said Russia would recognize two areas in Ukraine held by Moscow-backed separatists, setting the stage for a potential **invasion**. | President Biden is speaking. He is expected to announce new **sanctions** against Russia, joining E.U. leaders condemning Vladimir Putin.<br><br>Germany halted a key Russian natural **gas pipeline**, the West's strongest move yet to punish Moscow for recognizing separatist regions in Ukraine. | Russia's **invasion** of Ukraine has begun, the Ukrainian interior ministry said. **Explosions** were reported in Kyiv and other parts of the country. | Russia **declares war** over Ukraine<br><br>Ukraine's government vowed an "all-out **defense**" as Russia attacked more than a dozen cities and towns across the country **from land, air and sea.** | At the Polish Border, Tens of Thousands of Ukrainian **Refugees**<br><br>It's not just **China**: **India** is tiptoeing around Russia's invasion of Ukraine too. |

# Topic Modeling - LDA

- Top Vectors (Feb 18th) by 5 clusters
  - Topics: Evacuation, Invasion, Gas explosion, Video, Nuclear

| top_20 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|--------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| 0.076836 | evacuation | donetsk | shelling | civilians | eastern | separatists | region | ukraine | forces | separatist | ukrainian | dpr | republic | russia | russian | children | donbass | evacuate | residents | luhansk |
| 0.071517 | ukraine | invade | russia | war | putin | biden | president | says | russian | just | invasion | attack | crisis | convinced | world | amp | days | false | invade ukraine | going |
| 0.056428 | explosion | amp | ukraine | russian | border | security | media | military | pipeline | gas | russia | minister | donetsk | foreign | says | news | gas pipeline | reports | powerful | ukrainian |
| 0.053425 | video | watched | latest | watched latest | latest video | ukraine | russia | like | biden | amp | invasion | putin | crisis | time | lets | war | russian | thank | watch | right |
| 0.049866 | project | ukraine | russia | drills | russian | 2022 | amp | forces | attack | nuclear | defense | feb | analysis | 18 | day | february | results | ministry | missiles | strategic |

# Topic Modeling - LDA

- Top Vectors (Feb 25th) by 5 clusters
  - Topics: War, Shelters, India China vote, Zelensky, Invasion

| top_20 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.083687 | war | people | ukraine | world | stop | peace | pray | like | god | putin | amp | support | heart | stop war | love | russia | know | help | just | lets |
| 0.080910 | russian | forces | ukrainian | kyiv | sky | shelter | ukraine | capital | air | kiev | near | military | troops | people | close | city | destroyed | russian forces | close sky | shot |
| 0.067676 | ukraine | help | support | war | people | russia | resolution | council | need | invasion | security | security council | russian | amp | india | today | china | vote | time | solidarity |
| 0.066395 | ukraine | president | russia | ukrainian | people | stand | russian | zelensky | war | amp | nato | country | help | right | kyiv | fight | putin | dont | invasion | want |
| 0.061190 | russian | ukraine | invasion | russia | amp | putin | war | ukrainian | military | tank | russias | indian | video | kyiv | car | minister | students | army | new | foreign |

# Topic Modeling - LDA
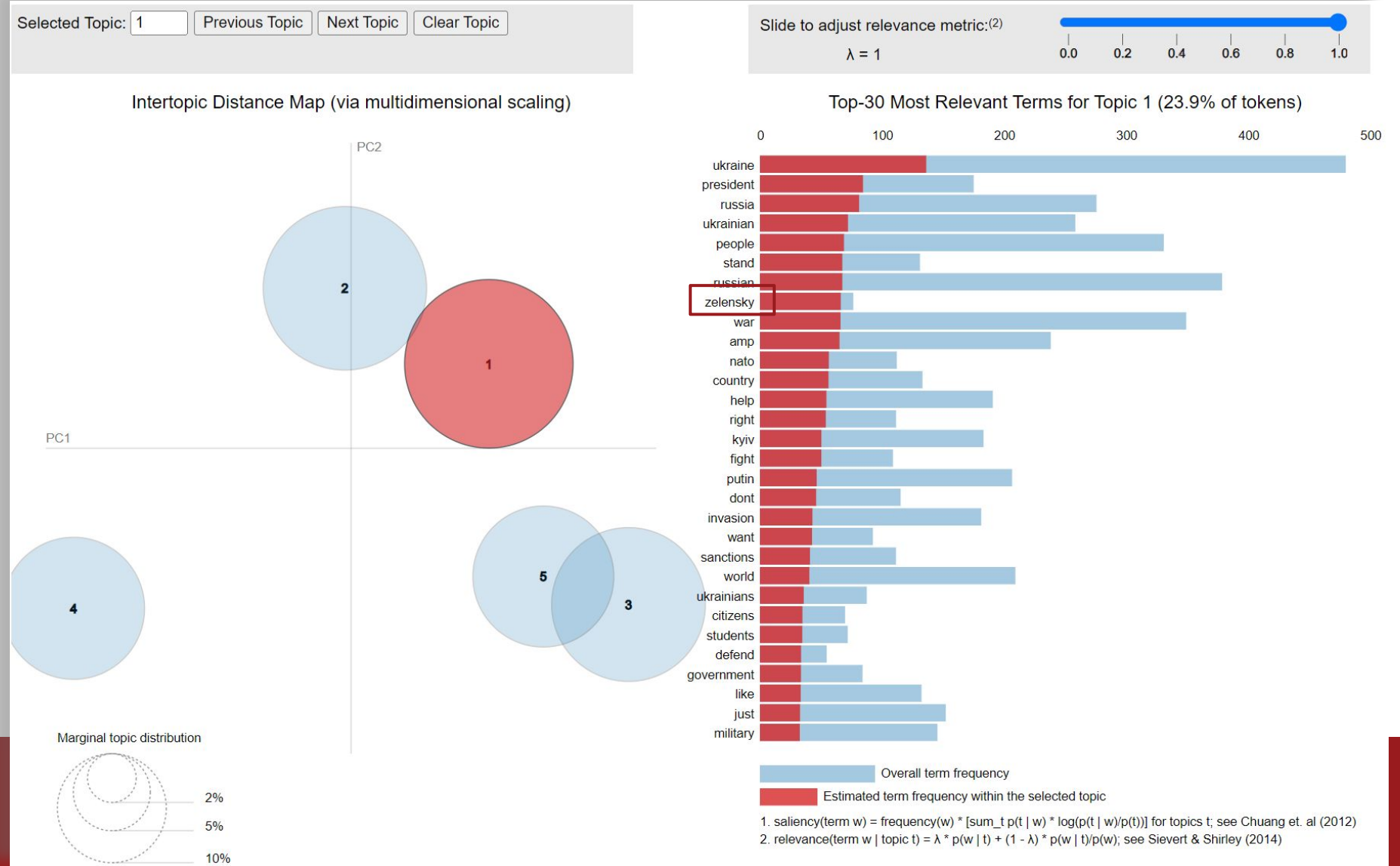
# Challenges

- Data collection
  - Recent tweets only (result in changing our topic)
  - High volume
  - Even with Elevated tier, Twitter didn't allow us get some features such as geolocation, media data
- Computing power
  - 24 hours to run SentiStrength on 1 million tweets
- Deciding on meaningful subset of dataset to analyze

# Next Steps

- Continue Scraping
  - systematic parameter for sampling
- Deeper understanding of Sentiment Analysis & Topic Modeling
- Network Analysis
  - user_id; name; user_name
  - @ Mentions
- Detect Political Leaning

# Thank You

## Any Questions?

USC Annenberg
School for Communication
and Journalism