THE RUSSIA-UKRAINE CONFLICT: A COMPREHENSIVE GUIDE TO NAVIGATING THE INFORMATION ENVIRONMENT

Integrated Capstone for Information Science
May 15, 2023

By

Sammy Rodriguez
Mirza Yaqub
Ian Mullinax
Shifa Nayab
Charles Smith

College of Information Studies

TABLE OF CONTENTS

Introduction

Reasons behind work

Our work aims to make info related to the War in Ukraine available and accessible to researchers at the iSchool so that they can easily use these datasets without any struggle. By creating this user guide, we are serving the academic community at the iSchool in a way that will help them understand the ongoing conflict in Ukraine.

Background

The iSchool - Open Dataset Research, Analysis, & Documentation (Project #1/3) project team members wrote this manual, working alongside Mary Ann Francis.

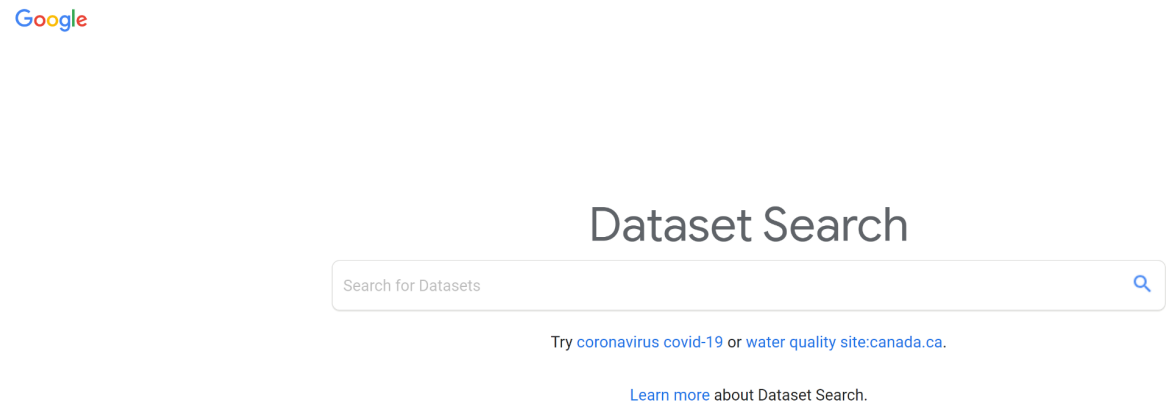
There are a lot of publicly available, free datasets online but the vast majority of them don't have documentation that would give the necessary context to the researchers. Our goal is to survey this information environment and analyze a wide range of datasets to provide clear navigational instructions for users (students, researchers, and other professionals at UMD). We hope that anyone using this user guide will get the most use possible out of these datasets and won't waste very much time.

Facts about War. What is it? Why is it important?

A significant issue at the moment is the war in Ukraine. Some Ukrainians disagree with the administration and so they want to separate from the country. Some other people want their nation to remain together. Many people's lives have been impacted by this war since they were forced to flee their homes and, regrettably, their families were torn apart. In order to help future scholars at the School better comprehend the conflict, we aim to carry out a trustworthy study of this array of datasets

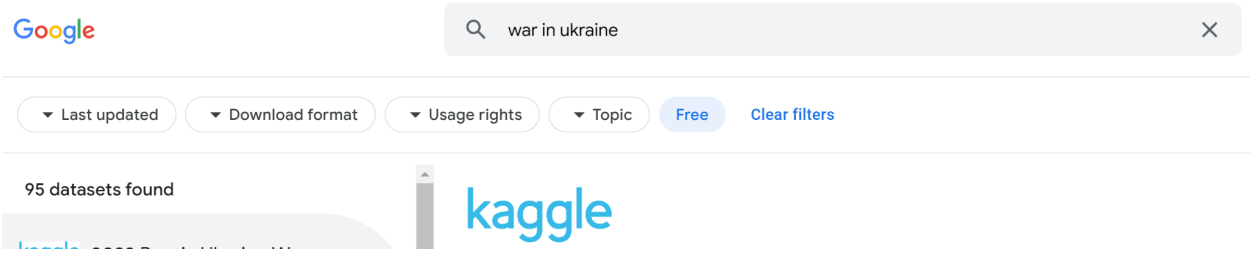
Description of Google Database search

We chose to use the online Google database. Google hosts structured and unstructured to make it easier to find, analyze, and access data. We found this source to be the best place to find data related to our topic.



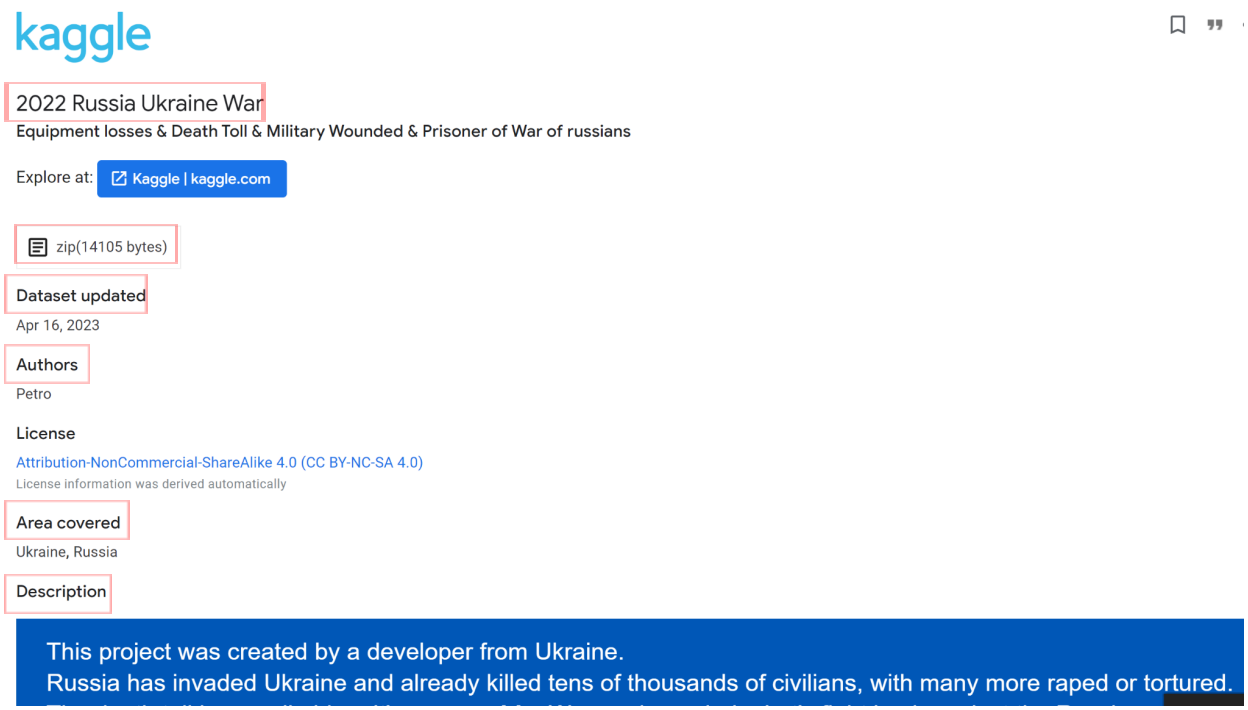
Filters

We started by typing “War in Ukraine” into the Google datasets search bar. This showed us all the datasets related to the War in Ukraine. Next, we clicked on the “Free” filter. We wanted to only include data that is free for public access. This helps us avoid any licensing issues.



Search Keys

Google provides a lot of information for each dataset it lists. For example, Google provides the authors, type of data, when the dataset was last updated, areas covered in the dataset, and a small description of the dataset. This lets you quickly analyze the data and pick the exact dataset the user needs.



The key elements are squared in red and are very easy to spot.

Web Scraping

We started by using ScrapeStorm to web scrape the data from Google datasets. This is a free AI tool built by an ex-Google crawler team. It is very intuitive and easy to use. This provided us with an Excel document that we could build upon and clean.

Data Cleaning

Once the original data was scraped we had to clean the data. We had 5 members in our group, so we split up the original 100 datasets into cleaning 20 datasets each. We started by filling in the missing information. The AI tool wasn't perfect and wasn't able to scrape everything in the files, therefore we had to fill in all the missing information. While we were filling in the missing information, we made sure the data was accurate and the AI tool correctly scraped the data.

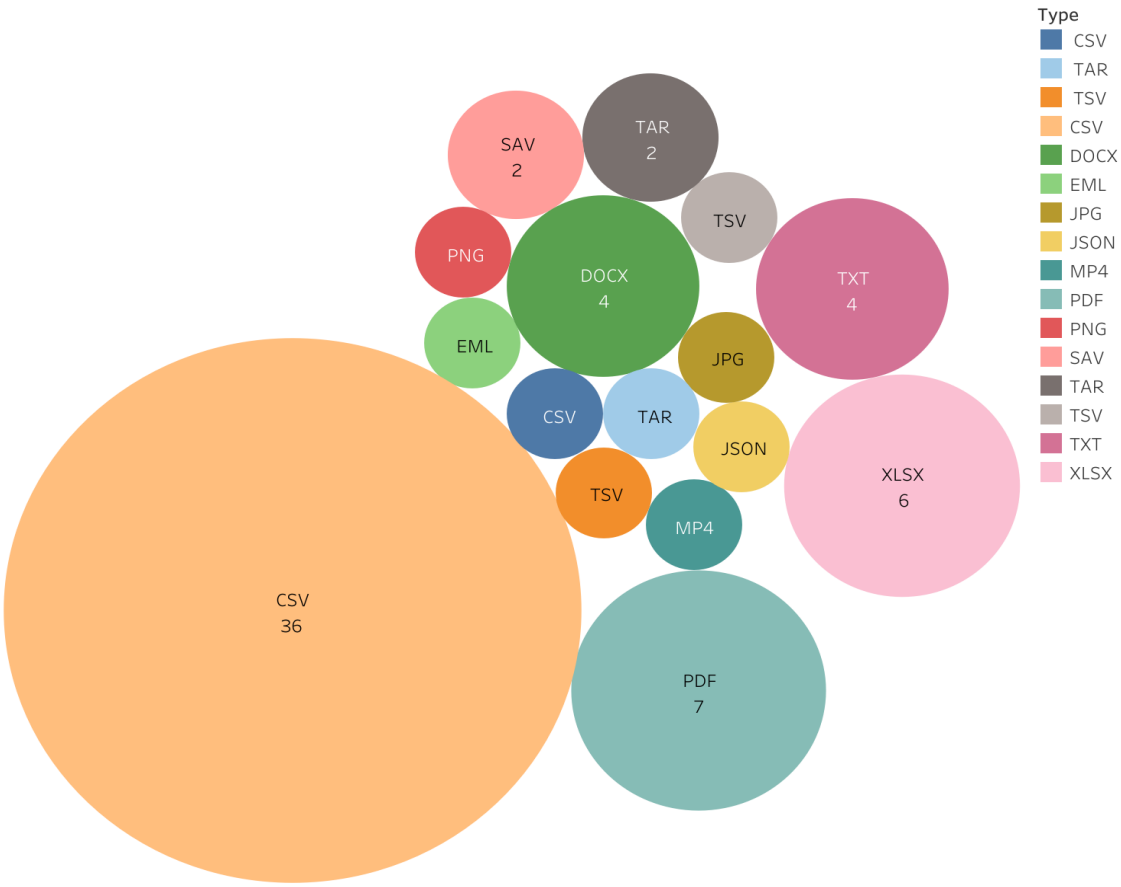
After filling in the missing information we deleted all information that was not relevant to the War in Ukraine or was not readable. For example, some of the data could not be downloaded or was not in English. We deleted all this information and only kept the relevant information. We were eventually left with approximately 70 datasets after cleaning.

Excel Spreadsheet Structure: linked here
We placed all our cleaned data into an excel spreadsheet. The outline is as follows and can be

Title of FIle	This is the title of the file. It is a searchable term used to find each file in Google Datasets. Example: *2022 Russia Ukraine War*
Original Source	This is where the original source of the data comes from. Biases may exist in the data. Example: Ukraine Govt.
Base Site	This is where the Google Datasets link takes you. It is the site that has the information. Example: Kaggle
Last Updated	This is the data the dataset was last updated. This may be different from the actual creation date. Example: 3/19/2023
Type	This is the format type the dataset is in. Example: CSV
Google Link	This link redirects you to the Google Datasets file. Example: 2022 Russia Ukraine War
Real Link	This link redirects you to where the actual datasets are. Example: 2022 Russia Ukraine War
Location	This is the location the files focus on. Example: Russia, Ukraine
Date Created	This is the original date the file was created. It may be different from the last updated date. Example: 4/18/2022
Qualitative/Quantitative	This is how to see if the data is qualitative or quantitative. Example: Quantitative
Author	The author or authors of the files. Example: Satyajit Ghosh
Notes	This is where we put general notes on the file. Example: Data refers to conflicts leading up to war

Full Data Collection Visualizations

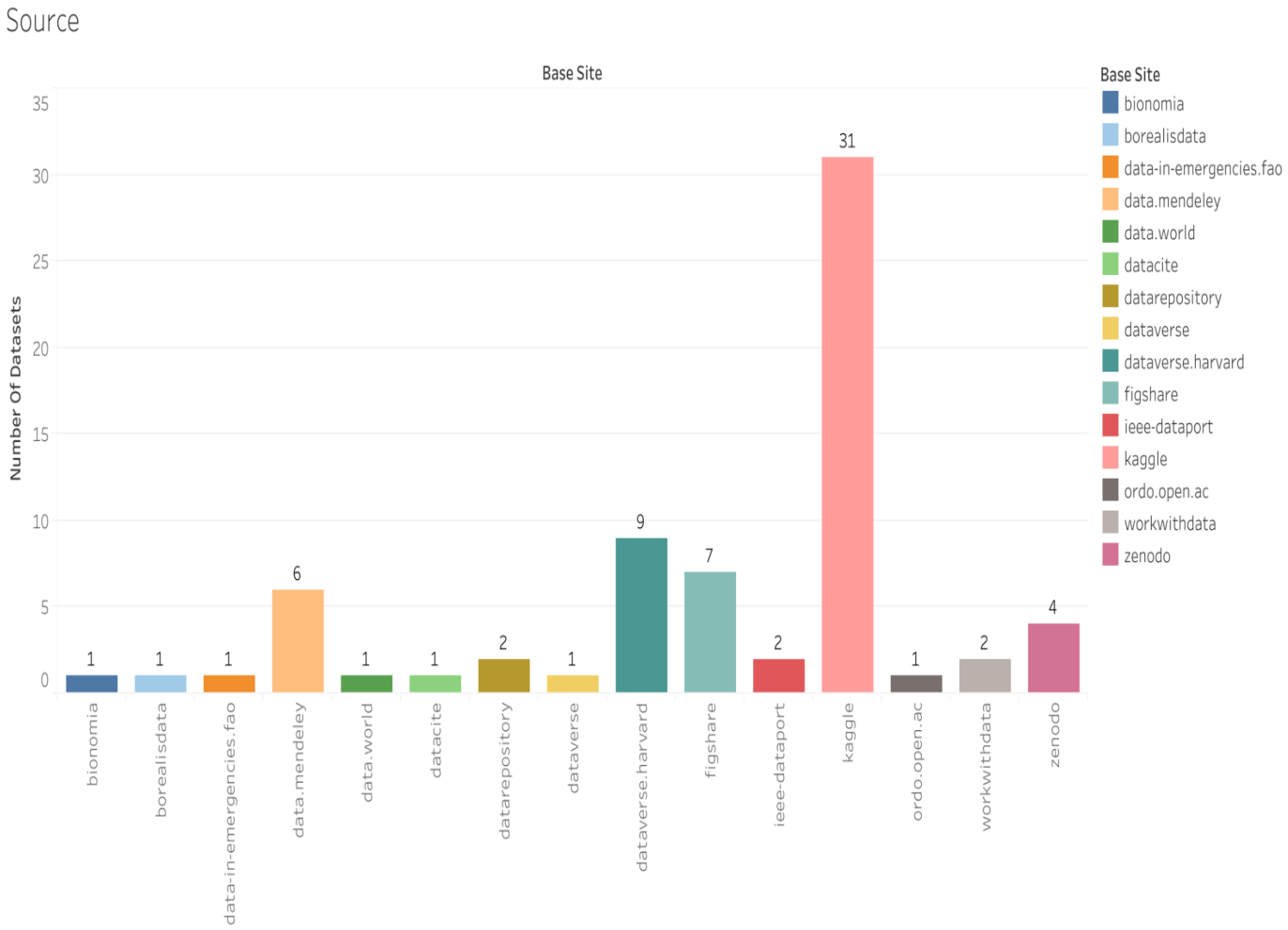
File Types



[Tableau Link]

This visualization shows the different file types of the individual datasets. Most of them are CSV files and can be easily accessed and analyzed. There are also some PDF files and some DOCX files. We can conclude that most of the datasets can be used for further research and analysis into the war on Ukraine.

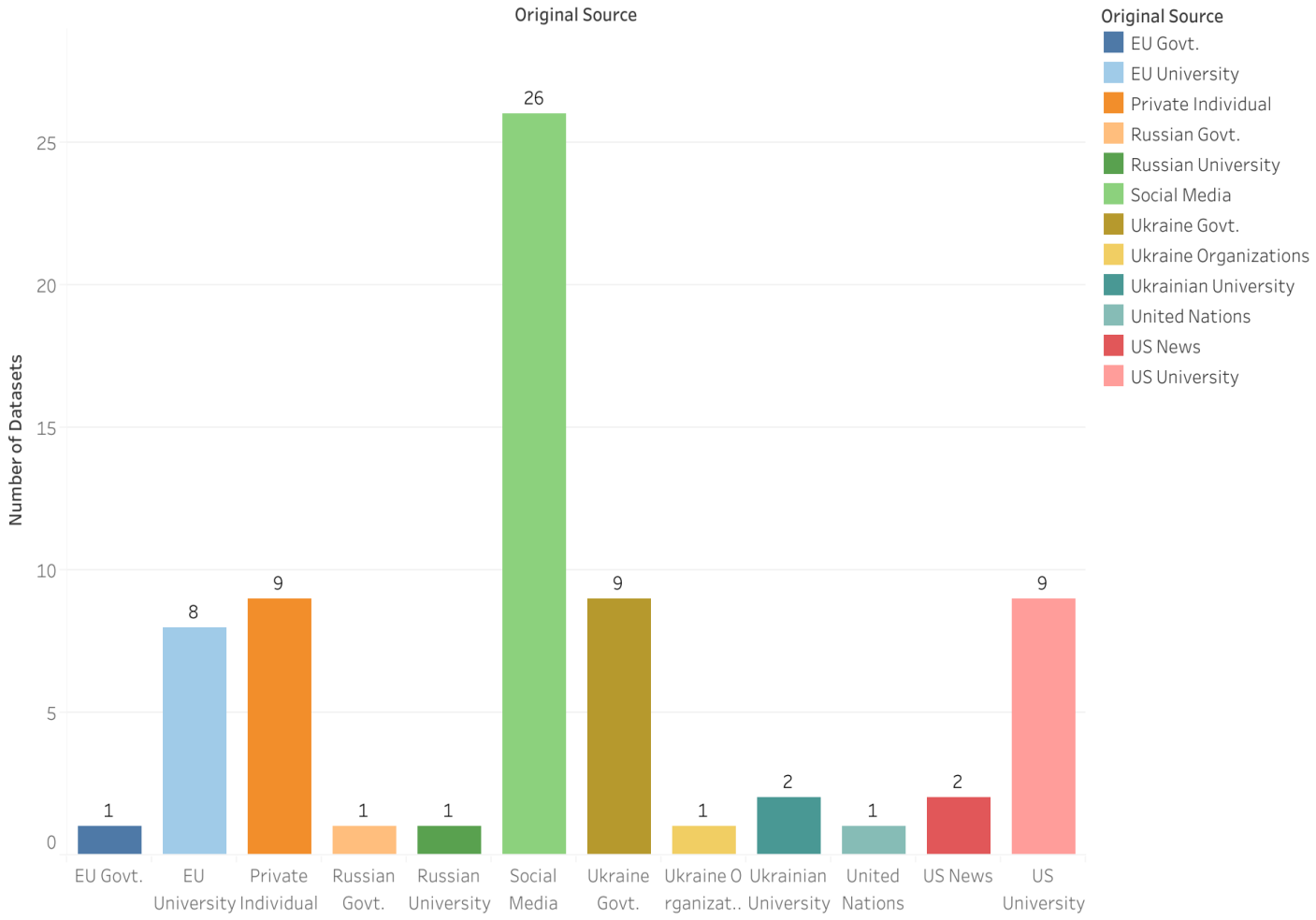
Source Site



[Tableau Link]

This visualization shows the website where the data was originally found in the Google database search collection. All of these websites are free to the public and anyone can view or use the data. However, a lot of the datasets in the collection had data that was difficult to use or not formatted properly. This is a common problem when using free data from sites like Kaggle, Data mendeley, etc.

Original Source



This visualization showcases the original source of the data. The datasets might be housed on different websites like Kaggle, Dataverse, and Zenodo but that does not tell us where the actual data comes from. So, we analyzed each dataset to find out the original individual or institution that published the data. As we can see, most of the data comes from social media. This can include platforms like twitter and reddit, where the data consists of scraped tweets or posts. Some data also originated from US universities like Harvard and Ukrainian government agencies.

Independent Dataset Visualizations

Russian Exports Dataset

This dataset consists of all exports by russia to the world in the last 20 years. This dataset contains over 4000 commodities and over 280 trade partners. It can be used to get an idea of how the sanctions placed upon Russia could have a global economic reaction all over the world. Here is a link to the dataset for the following visualizations regarding exports and commodities.

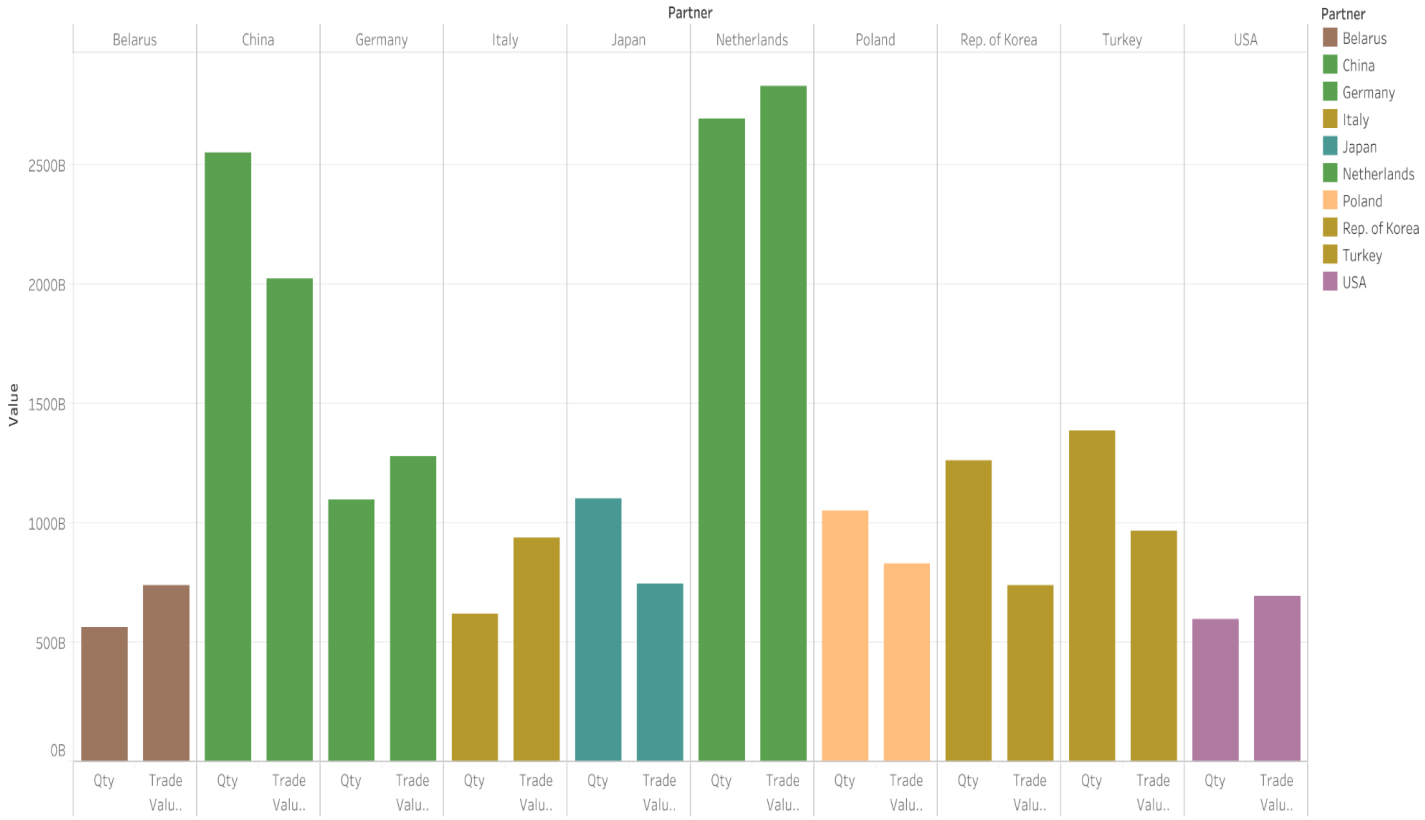
Top 10 Russian Export Partners

Partner	Qty	Trade Value (US$)
Netherlands	2,698,183,824,566	2,832,996,596,334
China	2,555,499,419,103	2,023,158,276,386
Germany	1,100,142,723,286	1,281,290,819,365
Turkey	1,386,431,861,639	964,200,823,854
Italy	617,537,038,972	936,478,026,840
Poland	1,050,996,559,059	828,736,173,364
Japan	1,104,048,859,099	742,760,953,564
Belarus	561,305,715,591	740,227,763,786
Rep. of Korea	1,261,969,978,281	737,164,221,640
USA	595,627,984,062	693,061,195,584

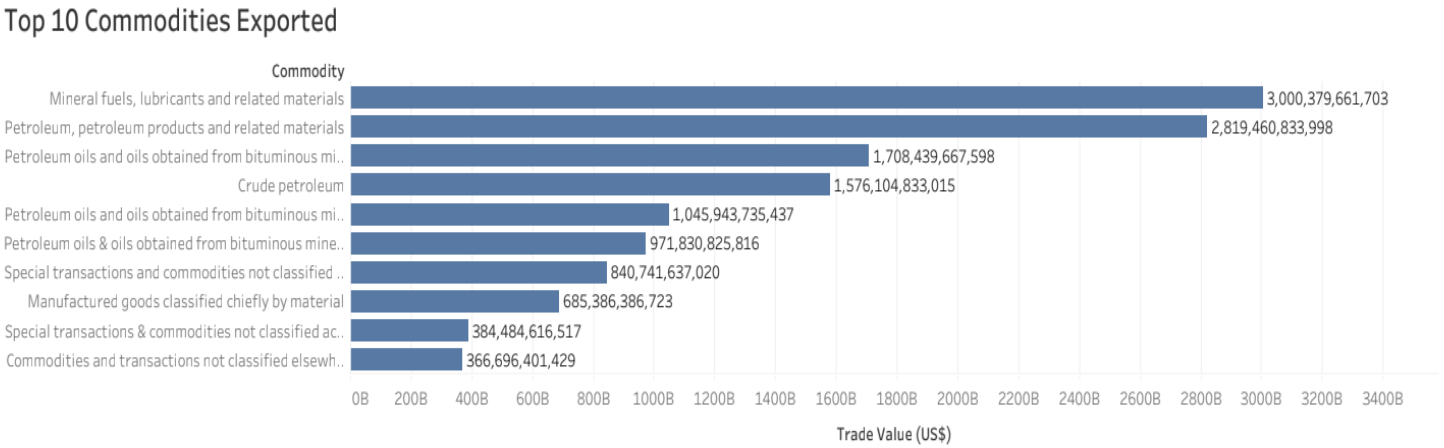
Tableau Link

This visualization consists of top 10 Russian export partners, meaning the countries that Russia has traded the most with in terms of quantity and trade value (US$).

Top 10 Russian Export Partners



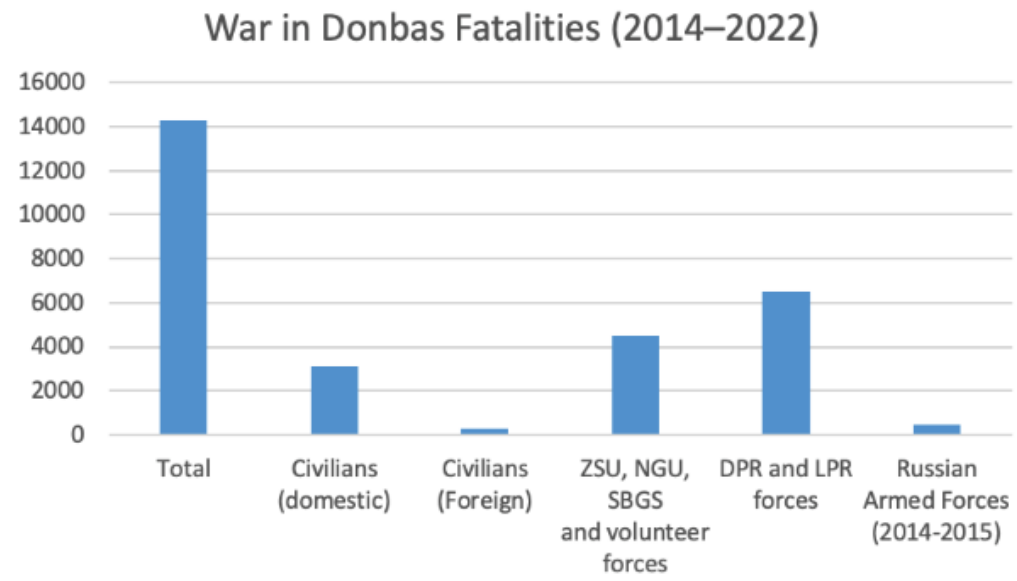
[Tableau Link]



[Tableau Link]

This graph shows the top 10 russian commodities that were exported based on trade value (US$). Mineral fuels is the highest commodity with a trade value of ~ $3 trillion, followed by Petroleum products, and oils at $1.7-2.8 trillion.

Fatalities/Prisoner Exchanges Dataset



This visual (above) shows the 14,200–14,400 civilians and military troops killed during the War in Donbas (2014–2022) - The dataset was authored by Johnny Shollaj, a Research Engineer at NUS, who sourced the data from the United Nations. Mr. Shollaj employed Power BI/DAX to gather the data

This visual is based on data from the United Nations and displays the prisoner exchanges that occurred in 2022. The dataset was also created by Johnny Shollaj, a Research Engineer at NUS, who used Power BI/DAX to collect the data. The visual indicates that a greater number of Ukrainian prisoners were exchanged in comparison to Russian prisoners

Version History

Version	Month and Year
1.0	May 2023