2022 Russian invasion of Ukraine

Project of Data Visualization (COM-480)

Adam Narozniak
adam.narozniak@epfl.ch

Alessandro, Tavazzi
alessandro.tavazzi@epfl.ch

Nicolas, d'Argenlieu
nicolas.thierrydargenlieu@epfl.ch

March 2022

1 Introduction

On the 24th of February, the Russian Federation launched an invasion of Ukraine that Vladimir Putin called a "special military operation". In response to the attack, outsider countries, organizations, companies, and individuals stood with the decision of who they would support. Various and consequent actions have been taken to support Ukraine on-field or condemn Russia internationally, including military support, economic sanctions, and humanitarian aid. Unfortunately, due to the unstructured, unverified, and possibly manipulated information spread, some people may still be misinformed or underestimate the seriousness of the situation.

Therefore, the objective of this project is to propose insightful visualizations from various perspectives about the conflict. Our goal is to fight disinformation by expressing data about multiple aspects of the war (e.g. economy-, human-, material-wise) in an insightful way. At the same time, we aim to give the possibility to get an overview of the conflict concisely.

We will tackle three areas:

1. **Economy** - impact on the international economy,
2. **People** - damage to the people of Ukraine and a try to identify trends in behavior,
3. **Foreign implication** - actions and reactions of the international community.

The authors of this report would like to express their full support to the Ukrainian people and condemn all the atrocities that are taking place in the context of Russia's invasion.

2 Dataset

Because of the current and rapid evolution of the war, there does not exist a single dataset containing all the information. Therefore we aggregate information from various sources of different types and create our datasets in order to fulfill our objective. *Most of these datasets will be made available on Kaggle at the end of the project.*

Economy

We have found data about the daily evolution in US dollars of the most important currencies in the world and in Europe (34 in total) since 1991. This information lays in the Foreign Exchange Rates dataset (and Investing.com RUB/USD for missing data).

Due to Russia's role as an exporter of oil and natural gas, we explore the price fluctuation of these commodities. The evolution of the price of oil (dataset and notebook) in the world and in Europe account for oil. And the aggregation of the daily price of natural gas in Europe and in the world (Spot market data & Natural Gas Yahoo Finance; see 1 to understand the importance of Russia's natural gas for Europe) accounts for gas.

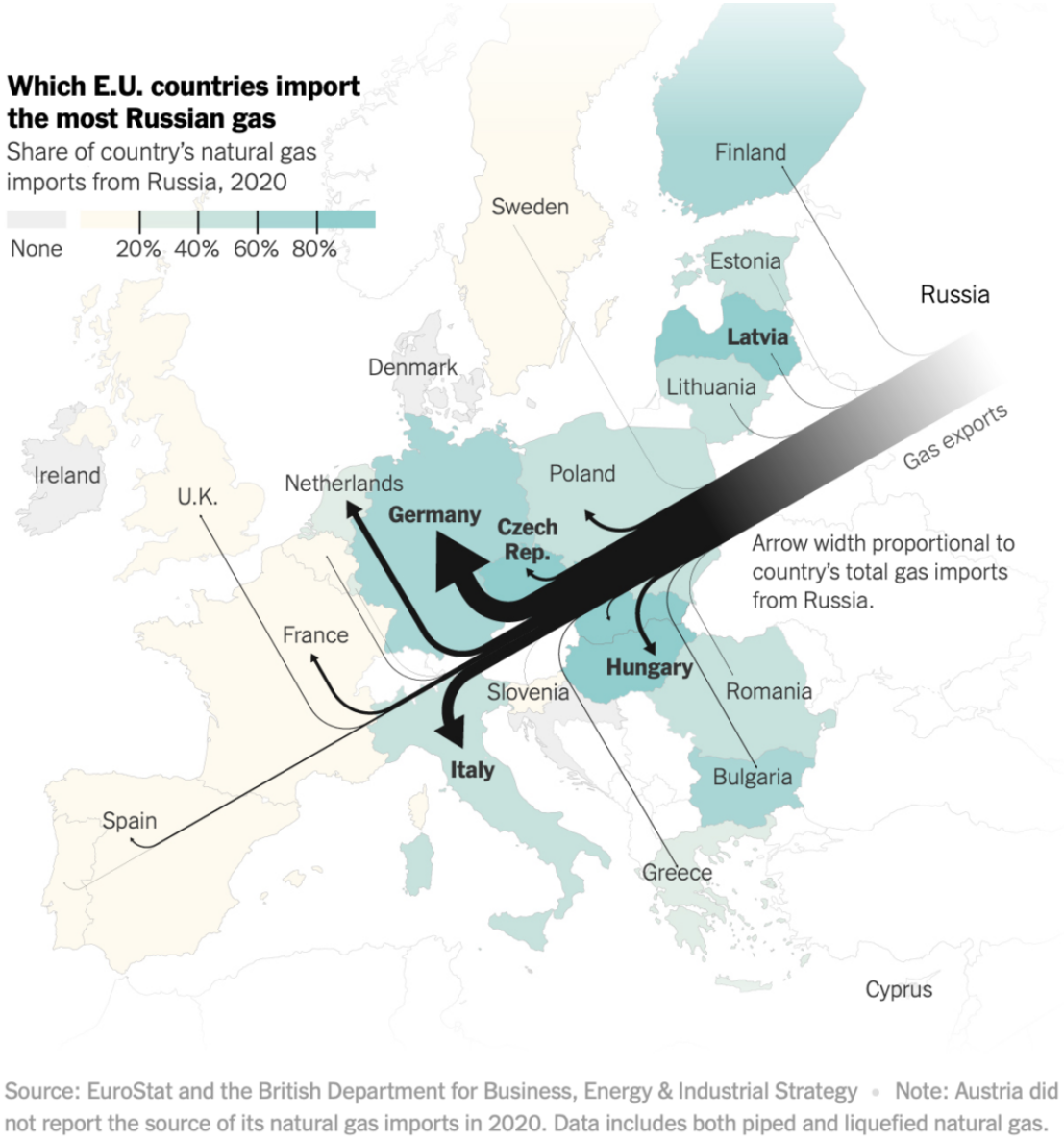


Figure 1: Map of Russian gas exports to Europe

People

We focus on the primary metric which is the **number of deaths and injuries** of **soldiers and civilians**. The source of information is the Twitter account **Kyiv Independent** due to the reputation and the number

of reported cases. The dataset was created by shortlisting lemmatized tweets filtered on specific keywords (`dead`, `death`, `kill` and `toll`), followed by human verification.

We also track **patterns in google searches** as a proxy of what the people of Ukraine and Russia think and feel. The data is gathered using gtab and pytrends. We expect some bias due to heavy censorship in Russia.

Foreign intervention

First, we focus on the **sanctions** issued towards entities, vessels, aircraft, individuals and locations related to the Russian Federation. The **sanctions** dataset is in the process of being obtained from Castellum AI. Then the **flow of refugees** exiting Ukraine towards neighboring countries is obtained from the UN Refugee Agency. The **public sentiment and emotion** dataset is created using an emotions model and an sentiment model, both applied to a dataset of daily tweets. Finally the **support and opinion** of surveyed people across the world is gathered in a collection of surveys aggregated from Statista, featuring the opinion and type of support towards or against Ukraine.

3 Problematic

This project is articulated around the wide subject of the war in Ukraine. More specifically, we focus on the scales of the conflict i.e. the numbers of different metrics exposed in the data. By this, we mean to guide the visitor into a structured exploration of the collected data and help it have a global overview on the conflict.

The **visualizations** will therefore depict the evolution of various metrics with time e.g. economic indicators, human counts and evolution of sentiment. They will illustrate key moments of the conflict since its beginning on the 24th of February 2022. The visualizations will also try to relate the information that we managed to extract regarding more uncertain information such as on-site casualties and injuries.

More specifically, we are thinking of a representation axed around a **timeline** that the user would animate with the **scroll** of its mouse. This would change the content of various information on the screen. If possible, we wish to integrate a map in order to facilitate linking the data to its origin.

With this task, we aim at making people **more aware of the situation through factual data**, and hope to help inform any opinion that the visitor may have. The idea behind it is to underline that this event has multiple and various consequences, both nationally and internationally, and how rapidly they are evolving. We aim at informing any visitor willing to get a synthetic and objective view on the situation in Ukraine from a general perspective.

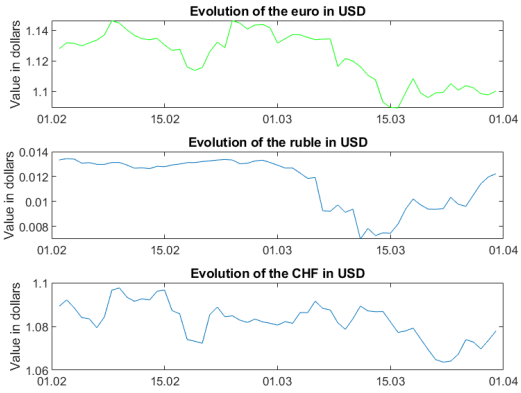
4 Exploratory Data Analysis

In this part, we detail and explore the data. On a general note, we hope to get the data up to the most recent date possible but we will be forced to decide of an end date in our visualizations.

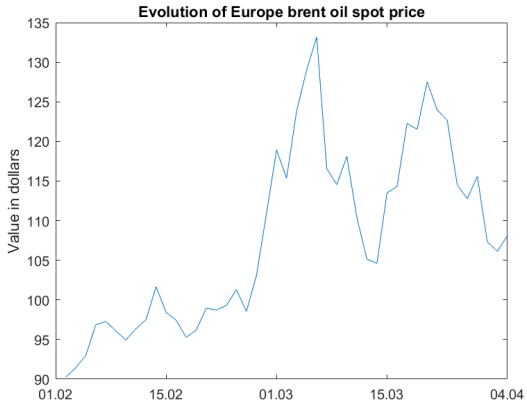
Regarding the economic aspect of this war, we can see the economic impact of the war in 2a and 2b. Indeed, the values of the displayed currencies drops and the price of a barrel of brent oil in Europe increases significantly, at the beginning of March.

The data from the Twitter features besides tweets also likes, retweets and number of comments. Additionally, the dataset created by us contains information about the number of injuries, deaths, their location and whether it concerns civilians or soldiers. Google trends data is a scalar representing the relative popularity of the topic at a specific time. You can find an example in 3 featuring the popularity of the Ukrainian and Russian president in Russia.

The data on the sanctions features the `type`, `issuing date`, `issuing entity`, `denomination of sanctioned`, `citizenship` and `list source` information for all sanctions targeting the Russian Federation between the 1st of January 2022 and present day. The sentiment data features each tweet's `text`, `userid`, `language`, `location` (if available) and `sentiment` (one of `[Positive, Neutral, Negative]`). Sentiment is computed



(a) Evolution of currencies (RUB,CHF and EUR) from 01.02.2022 to 04.04.2022



(b) Evolution of the barrel of brent oil spot price from 01.02.2022 to 04.04.2022

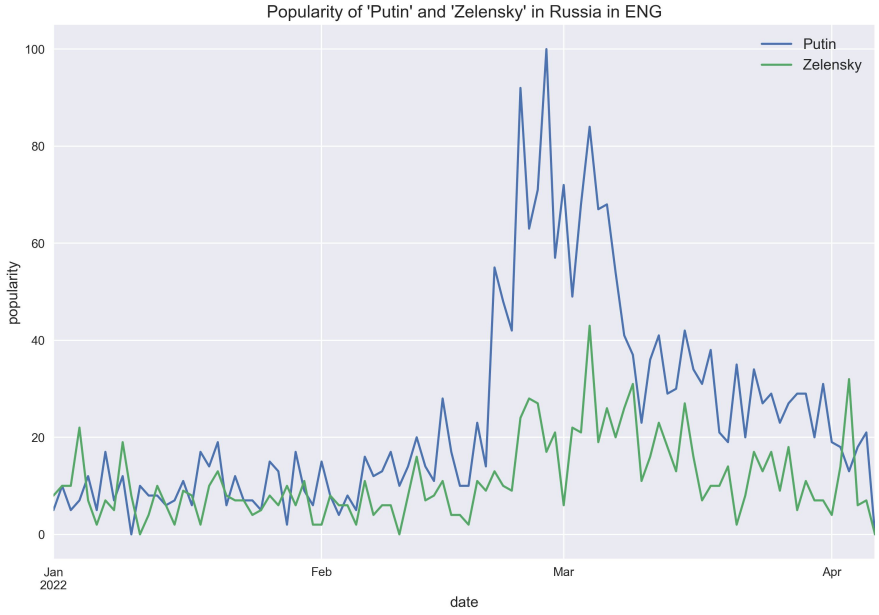


Figure 3: Popularity of 2 english terms in Russia according to Google Trends.

on tweets going from the 27th of February until now. The opinion data covers the period September 1st 2021 to April 1 2022 and contains information about the `queried country`, `size of group`, `sentiment` (one of `[ACTIVE POSITIVE, PASSIVE POSITIVE, NEUTRAL, ACTIVE NEGATIVE, PASSIVE NEGATIVE]`) and `comment` (one of `[help, resistance, support, business]`).

5 Related Work

Regarding the economic data, some statistics and line graphs (as done in this milestone) can be found easily through various sources.

Regarding the human and foreign aspects, as most datasets are being created for the project, we did not find any such work elsewhere. However, some of the information can already be represented on the websites of the sources that we gathered. Namely, we can note an interactive map depicting the show of refugees or graphics on the number of sanctions.