

Light on Ukraine

Process book

Course : Data visualization

Spring semester 2022

Supervised by : Laurent Vuillon



Authors :

Alessandro Tavazzi 342501

Adam Narozniak 344874

Nicolas d'Argenlieu 276507

Motivation

One of our main motivations at the beginning of the project was to fight against disinformation in the context of the actual war between Ukraine and Russia. Indeed, a terrible conflict is going on between the two countries since the 24th of February. Through the technological progress of the past decades, information became numerical. This socio-political revolution played a significant role in the new version of an old idea: propaganda. Without diving into the political details, we remind the reader that information is the basis of any sound democracy. Peace itself might depend on the possibility of having access to good information. In fact, wars are often made possible through the manipulation of the local population blocking their access to objective and disinterested information. Of course, we must consider that we are ourselves probably influenced by the western propaganda, since we are Europeans in Europe. Nonetheless, we believe we can help promote our values, democracy, freedom and respect, amongst these tense times in our own way, i.e. by creating a website and providing insightful visualizations to fight disinformation.

The goal of this project is to give objective and reliable information based on the facts, that could help the visitor to understand the gravity and the local and global impacts of this war. Regarding the concrete aspect of the war, a part of the website is dedicated to casualties and the concrete consequences for the local population. The other parts aim to give a more global view of the situation. One is dedicated to the analysis of the sentiments throughout the population in the world and the political reactions of several western states. Information about the reactions of the international community with sanctions is also displayed as well as the amount of fleeing refugees. Finally, a part is also dedicated to the economy and to understanding how global and impactful are the consequences of this war for the world of financial markets.

Dataset

The data had to be found or created, and aggregated by ourselves. Indeed as the conflict is ongoing, finding trustworthy information for the chosen aspects can be a difficult task. Below we detail how we built our dataset. The data always focuses on the period from the 15th of February to the 24th of May.

The sentiment and emotion opinion was computed by hand. We applied machine learning NLP (natural language processing) techniques in order to extract the sentiment and emotion from a dataset of tweets. The tweets for these were fetched from Kaggle¹, and are updated daily. The models for interpreting these tweets are

¹ <https://www.kaggle.com/datasets/bwandowando/ukraine-russian-crisis-twitter-dataset-1-2-m-rows>

pretrained models made available through HuggingFace², a bank of ML models available through API interfaces. The sentiment model³ classifies tweets from their text according to 3 labels : positive, neutral and negative. The model was trained by the NLP division of Cardiff University. The emotion model⁴ assigns one label among the following to the content of the tweet: anger, disgust, fear, joy, neutral, sadness, surprise.

The refugees data comes from the official website⁵ of the United Nations High Commissioner for Refugees. The UN uses a special dashboard to track some statistics linked with the war in Ukraine. The data is constituted of the accumulated count of refugees fleeing out of Ukraine. Unfortunately, we did not get an answer to our emails and thus could not obtain statistics per country but only aggregated for most of Ukraine's neighbors, namely Romania, Hungary, Slovakia, Moldova, Poland and Belarus.

Then the sanctions dataset was obtained from Castellum.ai⁶, an international and independent sanctions watchlist company. We met them in order to get details and access to the dataset, which they kindly provided to us. The data is constituted of

For the economic part, we combined several datasets from several sources. One of them is based on European Central Bank⁷. It is the dataset with the exchange rate in US-dollars of the main currencies in the world and in Europe. The other datasets are taken from the US government and more specifically from the website of US Energy Information Administration⁸ which is an official website of the US government. We chose to visualize the data of the Europe Brent Spot Price FOB (in US-Dollars per Barrel) and Henry Hub Natural Gas Spot Price (in US-dollars per Million British thermal unit, which is a unit to measure the energy and therefore the volume of natural gas considered. Regarding currencies, we thought they represent a good indicator of the economic situation of a country. Regarding the resources, since Russia has a lot of natural resources (including brent oil and natural gas), we thought the war would have had a significant impact on the prices of these resources. This was indeed verified visualizing the data. Something interesting can be noted, that even if Henry Hub is in fact an American company located in the US, the war had a clear impact on the price of the natural gas (even in the US), which shows how global the consequences of this war are.

² <https://huggingface.co/>

³

<https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment?text=Urgent%0A500+Russian+tanks+approaching+the+capital%2C+Kyiv%2C+from+all+sides.%0A+%23Russia+%23Ukraine+%23Russia+%23Ukraine%0A%23Kyiv>

⁴ <https://huggingface.co/j-hartmann/emotion-english-distilroberta-base>

⁵ <https://data.unhcr.org/en/situations/ukraine>

⁶ <https://www.castellum.ai/>

⁷ https://data.humdata.org/dataset/ecb-fx-rates?force_layout=desktop

⁸ <https://www.eia.gov/>

The data on the casualties, in the beginning was supposed to be scrapped from the tweeter account KyivIndependent, albeit due to the lack of the ability to confirm the sources and the struggle to manually go through all the post, we change to the UNHumanRightsUA, which is official Ukrainian tweeter account from of United Nations. Besides the credibility, the information was easier to extract based on the text matching rules.

Pre-processing

Regarding the economic part, one of the most difficult tasks was to pre-process the data. Indeed, the visualization of the data was chosen to be day-dependent and the code we managed to create only allowed us json files as data input. We had to split the data and to transform it to obtain what we wanted. We also decided to produce data out of our data computing values such as the mean, the maximum and the minimum of the exchange of a specific currency over the period of time we wanted to consider. In addition to that, the dataset of the exchange rate of several currencies was clearly incomplete and we had to complete it using different sources. For example, the rouble and ukrainian Hryvnia were missing and choosing other sources allowed us to complete our dataset.

Moreover the usage of machine learning techniques on large quantities of data took a significant amount of time (several days) in order to get complete information out of the raw tweets. Then we can also mention that some data was incorrectly formatted and had to be repaired.

Therefore the preprocessing was a significant part of our work and we wanted to be certain to get relevant information out of our raw data collection.

Preprocessing of the data on casualties was based on text matching rules in python. Moreover, the missing points were interposated based on the surroundings.

Planification & Changes

When looking back at our discussions reported in the milestone, we can state a few points.

The website took a few different turns in terms of design, evolving from an initial single page website to a more complex and rich journey. This discovery of the website was initially planned to be guided by a global time selection. However, during the implementation, we acknowledged that all data was not interesting to be displayed singularly. Therefore, we kept the daily granularity but moved to an overall day-by-day visualization to a more data specific approach. We highlighted the data that are not common like the average number of casualties on a daily basis with emphasis on children.

Regarding the sanctions data, we hoped to get details on the sanctions to be able to cut them down to the entity issuing them, as we had seen this was available for some information during our exploration of Castellum's website. However, their team

asked us to get a premium package in order to get this granularity, which also limited us in the exploration.

Then finally, we planned originally to be able to cut the refugee flow down into a country-specific number for each day, but this data was not made available to us and we could only use the one scrapped from the public website.

Challenges

One of our challenges was to give a manifold and synthetic view of the ongoing war through an original visualization of objective and reliable data sources. This was difficult since we all chose to consider different aspects of the war and its consequences. Searching for good information about what is happening between Russia and Ukraine is a war. Nonetheless, we believe that we managed to find good data sources. Even if we consider our topic from different perspectives, the whole remains consistent with our original intentions.

Another significant challenge was to visualize our dataset related to the exchange rate in US dollars of several currencies in a meaningful choropleth map depending on the day we could have chosen interactively. Indeed, we had to split the data into several datasets to select a reference value to which we could have compared the actual exchange rate. We had to implement the choropleth map using a library called anychart⁹, which was also very challenging. We created a button that would have hidden and shown some divs, and the whole creation of the several functions involved took a lot of time. The loading was taking too much time, and therefore we had to find a way to load each map individually and not all of them simultaneously at the loading of the page. Another challenge was to give sense to the visualization itself. Indeed, the exchange rate is done in US-dollar, and this currency has not a stable value as all currencies. Therefore, we had to consider the inflation rate of the US dollar (which is very high at the moment). This was challenging, and we are happy we managed to do it.

The landing page that we wanted to focus on to draw the potential visitor's attention. We visualized the two sets of borders of Ukraine. The top one is in the color blue, and the lower one - is yellow- so they resemble the flag. The drawing of them is animated using the Vivius¹⁰ library. Also, additional requirements on the type of SVG image turned out to be very demanding (there was not possible to simply convert an image into an SVG).

Final Result

Even if the final result did not reach our imagination precisely, we built a quality standalone project that offered a comfortable experience and relevant information. Indeed, at the beginning of this project, we had a lot of ideas under the assumption

⁹ https://www.anychart.com/?utm_source=trial

¹⁰ <https://github.com/maxwellito/vivus>

that everything would go according to plan, and it took time to understand if those ideas were doable in practice and if they made sense. Therefore, we are satisfied that we managed to visualize several datasets from several data sources in an original and interactive way. All of this was done to realize a polyvalent website that would have allowed the visitor to read the website as a story having, at the same time, the possibility to explore the data on its own. As such, we encourage people to use the website as a starting point to get a good impression of the amplitude of this war. We feel that our original motivation to initiate thinking about the ongoing invasion is sufficiently grounded.

The refugee data growth and the trend can be observed through the curve of the bar plot. Similarly, we can quickly catch the differences in the involvement of the major countries that reacted by sanctioning Ukraine in the visualization of sanctions. Then, the viewer can more carefully go through the emotion and sentiments of each day of the conflict and get impregnated by the realization that most of the public opinion is negative. The shifts between fear, sadness, and disgust can also be felt during the exploration.

Moreover, the fascinating aspect of the website makes it easy to focus on one piece of information at a time. This allows us to think of all the consequences of each one of the various displayed data.

In the appendix (figure 1), you can find an example of how we visualized the data of the natural gas, displaying several objects such as text, images and an interactive line chart.

Intepretation

The war in Ukraine is a complex conflict. It involves an immense amount of information that requires complex, in-depth processing, and consequent thought to organize the data in a meaningful way. However, the project was a very insightful way to go through the entire pipeline from scratch through data processing and visualization. This project taught us the basics of the compelling D3.js library and its stack-like approach to visualization creation. We learned how to construct a website that displays data interactively and engagingly, and we are now prepared to do similar projects in the future.

We naturally don't pretend to give an exhaustive and complete visualization of the ongoing war between Ukraine and Russia. Instead, we provide insights about these terrible events that could show the visitor how terrible and essential this war is for the world. One of our goals is to fight against disinformation. We believe we have found reliable data sources that show accurate visitor information about facts related to the war between Russia and Ukraine.

Contribution (peer assessment)

Nicolas d'Argenlieu handled the data processing for the refugees, sanctions, sentiment, emotion, and economic data. He was in charge of developing the visualizations linked to the sanctions, refugees, and sentiment information.

Alessandro Tavazzi created the part dedicated to the visualization of economic data with the help of Nicolas d'Argenlieu for data processing. He managed to visualize several datasets from several data sources in an interactive way developing relatively complicated javascript code. He gave an interpretation of the impact of the war on the world's economy, making apparent how vital and impactful this war was for the world.

Adam Narożniak handled the landing page (self-drawing borders of Ukraine), the time it elapsed since the beginning of the war, and all data and statistics on the casualties. It includes creating tools for scrapping, then preprocessing, and in the end, all website functionality. I created pictogram plots to display the quantity in the fold on human icons. Also, an interactive plot with the division on civilians, and children, killed and injured with a hand-made tooltip to display the exact number for a day and with the possibility to display only certain parts of the plot and zoom it. I have also implemented the sketch of the slide-down library.

Appendix

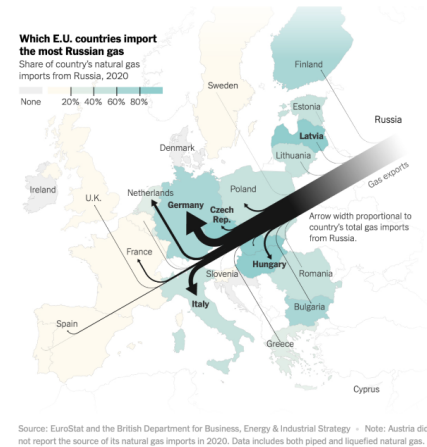
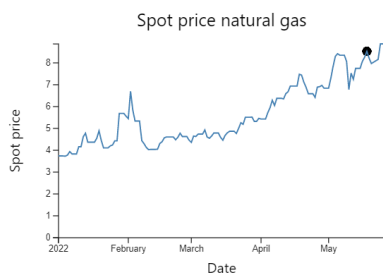
Figure 1

<< < Wednesday, 18-May-2022 > >>

Importance of the gas in russian economy with respect to Europe

The image on the right represents well the enormous importance of russian natural gas for Europe. That is why the war necessarily had an impact on the price of natural gas in Europe, but also simply in the world. Indeed, Russia is one of the biggest exporter of natural gas in the world.

In the line chart below, you can see the spot price of the natural gas from Henri Hub given in US-Dollar/Btu. The increase of the demand from Europe probably caused the increase of the price of the American natural gas.



4/4

Figure 2

Casualties

🇷🇺 Every day:

- 43 innocent civilians die, including 3 children
- 51 civilians get hurt, including 4 children.

🇺🇸 In total:

- 4169 innocent civilians died, including 268 children
- 4982 civilians got hurt, including 423 children.

The loss affects their families 🏠 friends, children and grandchildren's 🧓.

On the right :

total number of killed civilians



= 100 civilians

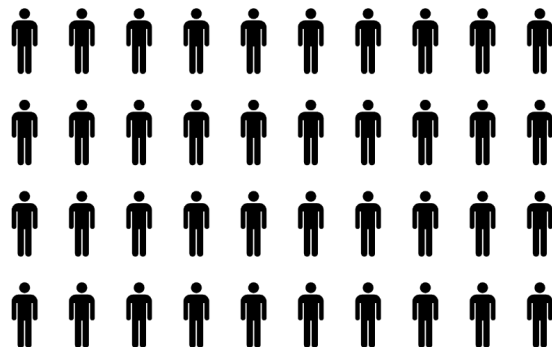


Figure 3

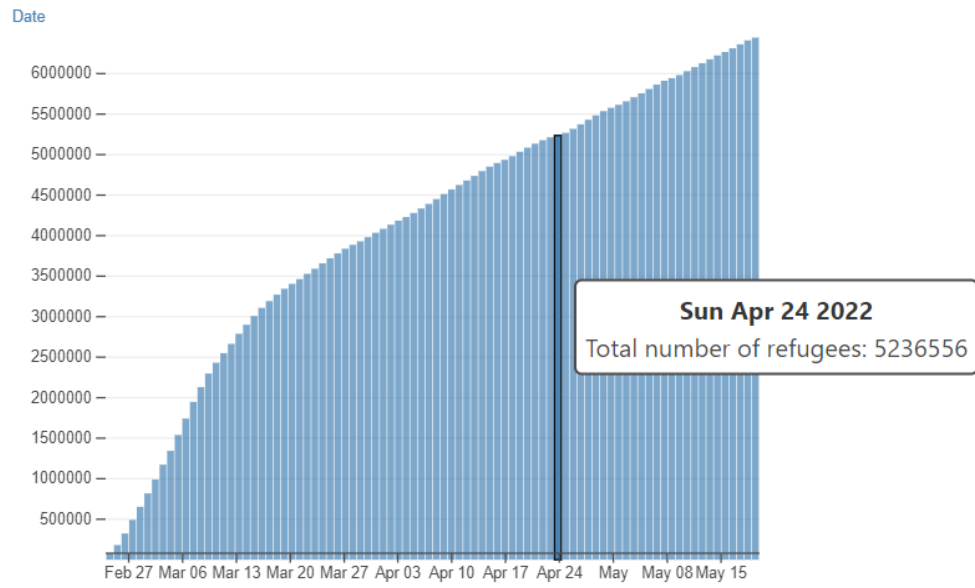


Figure 4

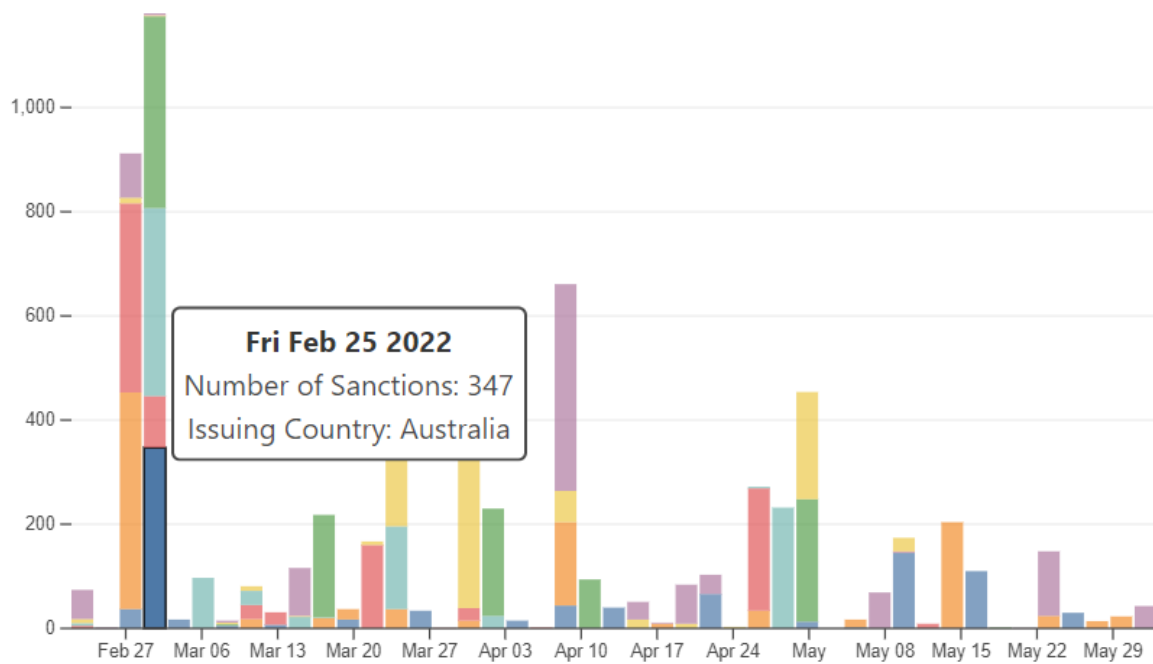


Figure 5

