Assignement # 4
CS-E4450 Expl. information visualization
Olivier Moitroux

Master 2 computer engineering (intelligent systems)
Exchange student - 784928

Friday 15th November, 2019

Visualization of worldwide terrorism

Useful links

My git repository: contains the latest version of the report and ipython notebook.

1. Motivation

According to many historians, humanity has never seen such long peaceful period[1]. Since the end of world war 2, western citizens used to feel safe in their home and didn't want to see any kind of war conflicts in their land.

Then, suddenly, the demolition of the world trade center on the 11th Septembre 2001 shook the world. While the western society got suddenly back to reality, this raised a wave of fear and draw our attention and media coverage on other part of the world where civil war was going on. We realized that some countries were developing a spirit of *american and european lifestyle* rejection. That's, at least, the stereotype. Suddenly, we were speaking all the time about the middle-east conflicts, *daesh* and all of other kinds of radical activist organization that nourished themselves from fear and conflicts in politically or economically weaken countries. From our peaceful view of the world, we started to see more and more news covering worldwide conflicts and attacks targeted towards the western culture.

But what is the real picture, *the truth*, hidden behind what our media let us know ? How many terrorists events were recorded these recent years ? Is there, indeed, a global increase of such events, suggesting that local conflicts (with sometimes very broad targets) have replaced conventional wars for political or cultural claims?

This is the context in which we will develop, step by step, an *interactive* visualization based on the worldwide terrorism database. Our objective will be to represent, on a world map, "terrorist" events and spot their evolution across time in an clear manner, so that the viewer can also get an idea of the general trends ruling this topic. Let us write a story about time, space and social issues.

2. Approach

2.1 Workflow

Now that the topic and the broad picture have both been fixed, let us construct the roadmap of our journey to build the visualization.

The first obvious step is to look for (good quality) data. Other very interesting subjects have been explored on my own (see git) but, unfortunately, the lack of data or resources (time, machine learning models, computing power, ...) are quite often limiting factors. In this way, it is pointless to overthink about good visualization if the data is missing.

The second step will be to think of a spatial representation of the data. Looking online for inspiration and an overview of what the tools can offer is helpful at this step. Then, due to the quite large number of entries, it is obvious to split the dataset according to a time dimension. In this way, the third step will be to think about a temporal representation of the data. Lastly, a final choice of tool should be made to design a first draft. After the peer-feedback, the final visualization can finally be produced.

2.2 Data

Many data sources are available nowadays. Some of which are listed on the forum or my git. Among them, *Kaggle*, one the of most well known dataset and competition platform in the data scientist community, used to offer quite a lot of good quality datasets with open-access. In our use case, we will more specifically make use of the global terrorism database. This datasets consists of more than 18K rows (worth 155MB of data), each records representing a terrorist event according to the following definition:

> "The threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation."

This dataset has quite a lot of interesting properties:

- Geography: Worldwide
- Time period: 1970-2017, except 1993
- Geography accuracy: latitude, longitude, city name and country name
- Unit of analysis: attack
- Variables (features): >100 variables on location, tactifs, perpetrators, targets and outcomes
- Source: Unclassified media articles (but Global patterns are driven by diverse trends in particular regions, and data collection is influenced by fluctuations in access to media coverage over both time and place so time interpretation should be considered with care.)

Such feature-full data can thus, of course, be used to design several impactful visualizations. The list of relevant features is discussed in the git.

Furthermore, one could imagine extending the data by adding a new layer on top of the visualization: worldwide happiness or other socio-economical (gdp, ...) on the map. This dataset can be a good start to evaluate happiness (Europe only). Countries of the World (or this raw one) dataset can be used to get mortality rates of countries as well as gdp per capita.

2.3 Tools

Many tools can be used to develop geographical visualizations. *Tableau, highcharts, 3djs, and plotly*, just to cite some of them, are quite well suited for interactive spatial visualization. My

taste for Javascript being not very pronounced and since I prefer having full control on my visualization (>< Tableau), I will give a try to the open source *plotly* package. More specifically, I will rely on the subplot feature to expose several visualization in one and with *Scatter geo* class for building the interactive 3d globe. The scripting language will thus be *Python* and *Pandas* library will be used to. The visualization step by step guide will be in the form of an ipython notebook. However, due to the *very poor* performance of such tool, I will certainly move to my local setup rapidly.

3.   Discussion about space

Space plays a very important role in this context. Every event has its precise coordinates which can be plot on a world-map. By doing so, we will highlight some trends in the data. It is indeed expected that some part of the world would pop more than others.

Several visualization tool can be used: interactive choropleth Inset Maps, colorful World Choropleth Map with legend and tip on hover, 3d globe of the world with markers and heatmaps are some among many. Since the data is quite geographically accurate (gps location), it would be sad to loose this information by using choropeth maps. Indeed, these maps assign a color to each country while it would be more interesting to track terrorist events trends inside a country. For example, if events take place in the capital, it is perhaps a sign of internal conflict while, at the border, it would certainly highlight a conflict with a neighbour country.

Markers for events is thus the way to go. Should we consider a flat map or 3d visualization of the world ? Both have pros and cons. The first one has the benefit that the viewer can see, at one glance, all the countries but it also introduces some projection errors while the second option is accurate. This isn't critical in our use case. However, since I'm seeking for some interactiveness with the viewer, I will prefer a 3d visualization, eventhough, from a pure scientific point of view, the more classical approach would be preferred.

Each marker can be used to reveal additional information on tip hover. Such information can be simply displayed in text like number of killed people, target nationality and so on. The frontier of the countries should also be drawn since they represent key information to understand conflicts. This even allow us to use the filled area within the border of each country to represent another piece of information. This could be the number of people killed, the proportion of people that dies due to terrorism, hapiness indices, gdp or any other kind of metrics but this hasn't been decided yet at time of writing. An "heatmap" would be relevant but I don't think the plotly library allows that. Alternatively, we could simply personalize the color of each marker to be proportional to the number of killed people for this event.

To balance "explorative" and more "scientifically presentable" visualizations, some histograms can be plot beside the main 3d globe visualization. We can either aggregate countries by region (continent level) or display the top X most dangerous countries.

4.   Discussion about time

The data features more than 18K rows and span over 37 years. This information is very precious to understand the evolution of conflicts around the globe and how strong they evolve when time goes by. Moreover, it would allow us to answer to question raised in the introduction: is the number of conflicts in the world increasing or decreasing ? A nuance could also be introduced. Maybe, yes, the number of terrorist events increase but what if we normalize by the population

growth ? This could perhaps highlight that the aggressiveness of humanity has not especially increase/stabilize/decrease.

A first attempt will be done by splitting the data by years. However, it is likely that the number of markers would be too high to plot on the globe. In such a case, the data can be split by month: a decision that can make sense considering that it could highlight time of the year when more conflicts emerge !

More precisely, I'd like to have a cursor beneath the visualization that would allow the viewer to set the time as he wishes. An autoplay button also make sense to play the animation automatically.

5. Expected results

The expected result is a small interactive dashboard, featuring, in the left column, a large 3d globe representation of the world with markers on top, each one representing a single event. If the tool library allows it, the marker color will be proportional to the number of killed people. An experiment with the filled region inside the countries borders will also be done to represent either the happiness index, the total number of killed people or the proportion of the population killed by terrorist events.

The right column of the dashboard can be split in two. On top, an histogram can represent the top X countries/continent with the most events. Below, a bar plot with the number of killed people for the top X countries/continent can be shown. If the information turns out to be redundant with the previous one, I would consider using a simple multi-trace plot (one coloured line/continent) showing the evolution of the number of events across time.

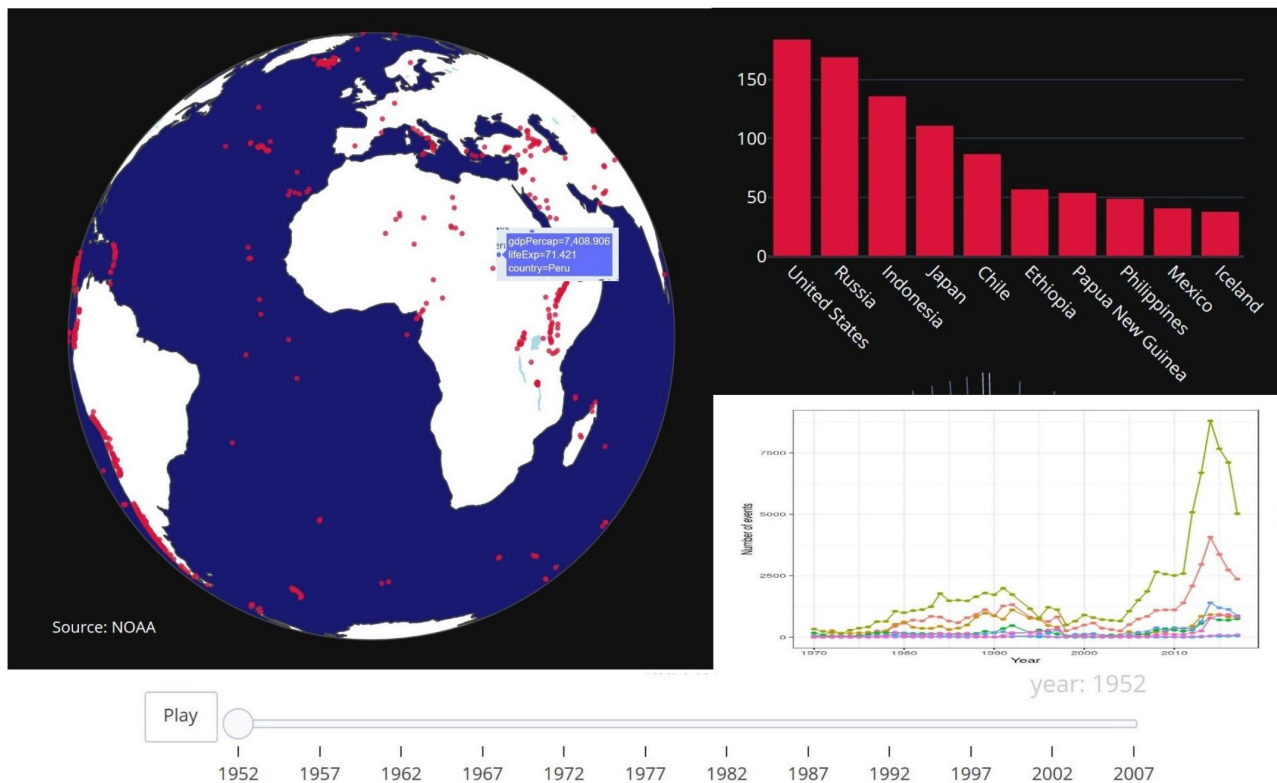Finally, at the very bottom of the dashboard, a slider for time will be added.

Figure 1: DRAFT of the visualization.

References

[1] Y.N. Harari. *Sapiens: A Brief History of Humankind*. Harper, 2015.