Aalto University
School of Science

Assignement # 4
CS-E4450 Expl. information visualization

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

FRIDAY 20TH DECEMBER, 2019

Visualization of worldwide terrorism

Useful links

My git repository contains the latest version of the report and python code to generate the visualization. However, the generated visualization (html) can be directly accessed here.

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 and especially what are the geographical trends? Is there, indeed, a global increase of such events, suggesting that local conflicts 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 in the early stages 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 is to think about 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 is to think about a temporal representation of the data. Lastly, a final choice of tool is made to design a first draft. After the peer-feedback, the final visualization can finally be produced. Finally, we can discuss shortly the final results and propose some ideas of improvements.

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 or to adapt mine in many ways. The list of relevant features/columns is discussed in the git with greater details.

Furthermore, one could imagine extending the data by adding a new layer on top of the visualization like worldwide happiness or other socio-economical metrics (gdp per capita, ...) 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.

At a later stage of the project, I came to the observation that, depending of the filtering criterion, the observations could be really impacted. This was particularly striking when filtering out failed attacks. For example, a country like USA was number one in terms of number of attacks[1] but by filtering out failed attacks, it went out of the top-10 and UK was first.

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 Javascript skills being very limited and since I prefer having full control on my visualization (>< Tableau), the open-source *plotly* package will be tried. More specifically, I will rely on the *subplot* feature to combine several visualizations into one and on *Scatter geo* class for building an interactive 3d worldmap. The scripting language will thus be *Python* and *Pandas* library will be used for data consistency check and cleaning. The produced output is in html format and should be open with a webbrowser.

3. Discussion about space

Space plays a very important role in this context[2]. 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 neighbor country. Markers for events is thus the way to go and more specifically simple circles for clarity.

Each marker can be used to reveal additional information on tip hover. Such information is simply displayed in text and include the country name, the number of killed, injured people and the total as well the attack type for this specific event. Other information are present in the dataset (target nationality, ...) but adding all of them in such a small hovering text box would not be very user friendly.

It has finally been decided that the color of each marker would be proportional to the number of casualties (injured+killed people) for each event. This has the advantage of being easy to visualize at a glance, in opposition to make the radius of circle vary. Indeed, the later approach gives messy results in very dense regions due to overlapping. After some tests, it turned out a linear scale wasn't producing satisfying results as most markers were aggregated to the same color, while only major attacks were popping in another color. Changing the scale to a logarithmic one proved to be successful.

[1] Some care should be made here, since the dataset may be more accurate and complete for more developed countries.

[2] One of my choice when designing the visualization was to focus mainly on the geography.

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 at all in our use case though. However, since I'm really 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 obviously preferred. In the same manner, the visualization is made so that the user can zoom and focus on some regions easily. By zooming, the 3d map turns into 2D, which offers a good trade-off.

Obviously, 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 if one wants to extend my visualization in the future. This could be the proportion of the population that died due to terrorism, hapiness indices, gdp per capita, etc.. Giving a try to an "heatmap" representation would be relevant but the plotly library doesn't allow it on the *scattergeo* object used.

To balance "explorative" and more "scientifically presentable" visualizations, an histograms of the number of casualties can be plot beside the main 3d globe visualization. However, for readability purpose, it is better to constrain the horizontal axis with fewer elements. We can either aggregate countries by region (continent level) or display the top X most dangerous countries by using a bar plot instead. The later option was chosen with X= 10.

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 first attempt have been done by splitting the data by years. It turned out that the number of markers was not too high to plot on the globe. If this hadn't been the case, the data would have been split by month: a decision that can make sense considering that it could also highlight time of the year when more conflicts emerge.

Since the visualization strives to be interactive and more explorative than traditional scientific reports, the time dimension can be incorporated by adding a cursor beneath the visualization. This later one would allow the viewer to set the time as he wishes. An autoplay and pause button for the animation have also been added naturally.

The top right area of the visualization is used to display the top 10 countries with the *most frequent* tentative and succesful attacks combined. This nuance is motivated in order to not make double use with the globe visualization while also displaying some of the most risky areas in expectation (in a sense of probability of an attack, not its number of casualties).

It has finally been decided to use the remaining space of the visualization to plot a pie-chart representing the attack types in terms of % (for a given year). This can indeed reveal some trends in terms of evolution of the type of attacks when time goes by. It should be noted that by clicking on the legend, on the right, removes this element from the total proportion.

Finally, the total number of attack for the selected year is displayed in the bottom right.

5. Summary: final result

The final 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. The color of each marker is logarithmically proportional to the number of casualties. Hovering the mouse cursor on a marker reveals information linked to the attack.

The right column of the dashboard is horizontally split in two. On top, a barplot represents the top X countries with the *most frequent* events (number of events per year). Below, a pie-chart gives an overview of the which methods were the most used.

At the very bottom of the dashboard, a slider can theoretically be used to manually set a year. A play button besides the slider can move the slider in chronological order automatically. However, the callbacks were not fully implemented (see improvements) and thus the year is fixed to 2001.

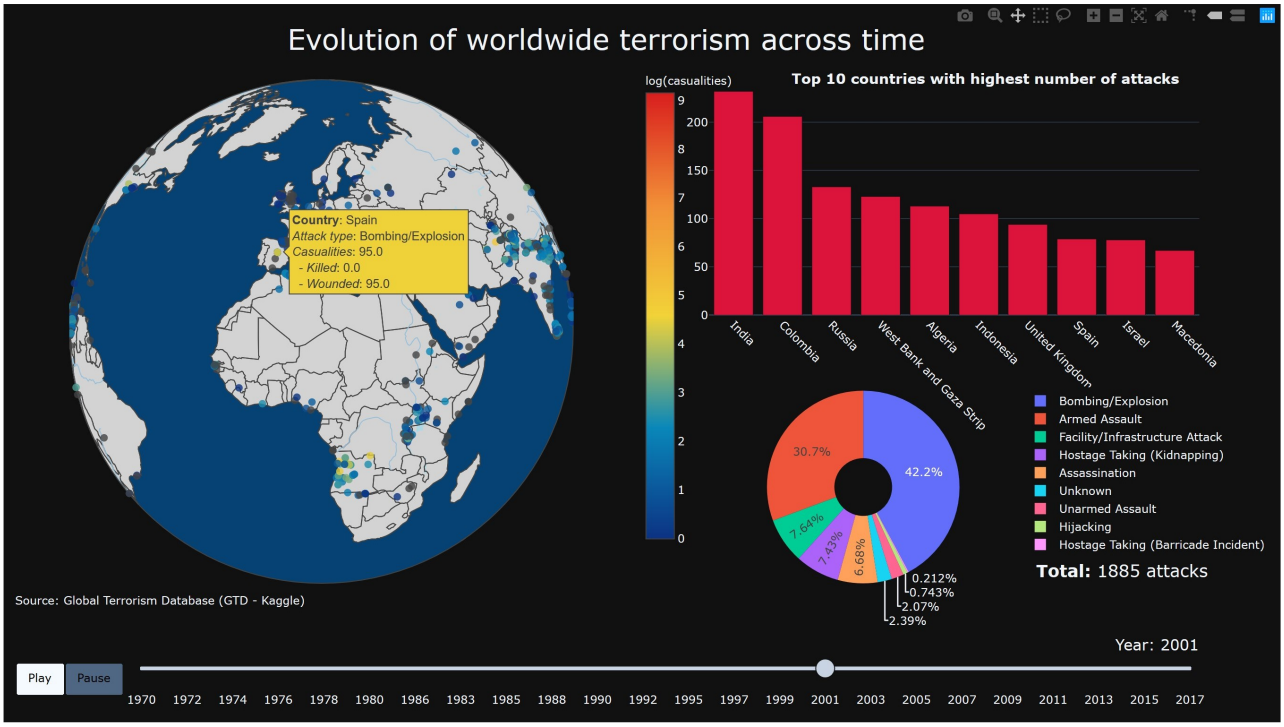

Figure 1: A screenshot of the final visualization.

By playing a bit with the visualization, it is quite funny and easy to spot part of the world where civil war or conflicts are really present. For instance, the Black Spring took place in Algeria in 2001, something that is clearly visible on the map. The region of Cachemire, in the Himalayan mountain in India is also clearly visible (especially close to the frontier) and is known for its conflicts with Pakistan and even now China. The same year, Angola turned out to be a very dangerous place with several attacks reaching more than 100 killed people.

By generating the visualization for different years, we can see how conflicts slowly build up and disappear like, for example the Lebanese Civil War that started 1975 and ended up in 1990 or some agitated protestations from basque country in the 80's.

The methodology didn't change too much on expectation but bombs attacks have becoming more frequent and assassination less frequent over time.

The reader is also recommended to take a look at the appendix page 7 where we discuss some ideas of improvements, and other alternatives that could be investigated in the future. Figure

2, in the appendix, is an alternative to the barplot in the form of a summary line plot exposing the evolution of worldwide number of attacks. This number increased steadily from 2005 up to its peak in 2014 before starting to decrease. Some care should be taken since the quality of data acquisition may have improved over time.

6. Conclusion

With an iterative workflow supported by peer-review, we demonstrated how to design an *interactive* visualization that enable the user to discover trends within terrorist events. By using the visualization, we proved it was possible to detect, easily, known civil wars or sparse significant attacks as well as to track their evolution and methodology, thus demonstrating how useful the visualization can be as an informative tool.

7. Appendix

The following sections summarizes all my ideas of improvements, some of which have also alreay been proposed by one or two people in the peer-review as the cherry on top. Naturally, this visualization can be adapted in an almost unlimited way but some choice needs to be made. Some alternatives have been discussed in the report and some are added here too.

7.1 Adding more data to find correlations

As discussed in subsection 2.2, it would be interesting to integrate other socio-economical analytic in the visualization. We may ,for example, suppose that countries with lesser gdp per capita or smaller happiness index to have a higher number of attacks within their countries. We could also analyze the proportion of deaths whose resulting from terrorist attacks for each country. In this way, we would not biased by the population of the country anymore.

A simple approach would be to integrate this new dimension by filling the area covered by a country with a color following a colorscale. For visibility reasons, this would obviously imply to visualize the number of casualties for a given attack not with a colorscheme anymore but by playing with the radius of the dots.

In order to analyze the evolution across time of the number casualties in the world, it would also make sense to normalize by the population growth. This could perhaps highlight that the aggressiveness of humanity has not especially increase/stabilize/decrease.

7.2 Adding more interactiveness

7.2.1 Toggle between 2d/3d visualization

Eventhough the choice of a 3d visualization has clearly been motivated in this report, it would be a nice add-on to let the user be able to change the map to a more "scientific-friendly" 2d projection if he wants to. This could be done with a drop-down list. However, implementing the callback would require a lot of code modifications for few benefits. This was also suggested by one person in the peer-review.

7.2.2 User input filtering criterion

As discussed in section2.2, letting the user choose if he wants to account for failed attacks and assassinations (cfr. peer-review), would be a nice add-on.

7.3 Using the last tick of the slider to provide a summary

It would be nice if the final step of the slider could provide a summary view of all the dataset. We can imagine removing all the marker and using an heatmap (eventhough it is not possible with the current scattergeo object) for the globe. The barplot would naturally be converted into a line plot (x-axis: a given year, y-axis: number of attacks, all countries combined) and the pie-chart would simply be adapted to include all the data rather than just a given year.

7.4 Using another API

As of right now, the visualization has been implemented with the *subplot* object of the *pyplot* pyhton api. However, I discovered at the end of the project how much it constrains the use of animation and more specifically, the callbacks required to synchronize all the plots together. Furthermore, this part turned out to be very badly documented. If I had to restart, this project, I would use instead the Python *dashboard* api.

7.5 Fixing the time slider callback

After long research on the web, I finally found out how to implement the callbacks for the time slider with the use of *frame* objects. However, using them would require deep and long code modifications of the existing code, without any true guarantee of results. A simple "hello world" with frames has been tested though but as of right now, the slider is just visual but unfortunately not effectively working.

7.6 Alternative visualization to the bar plot

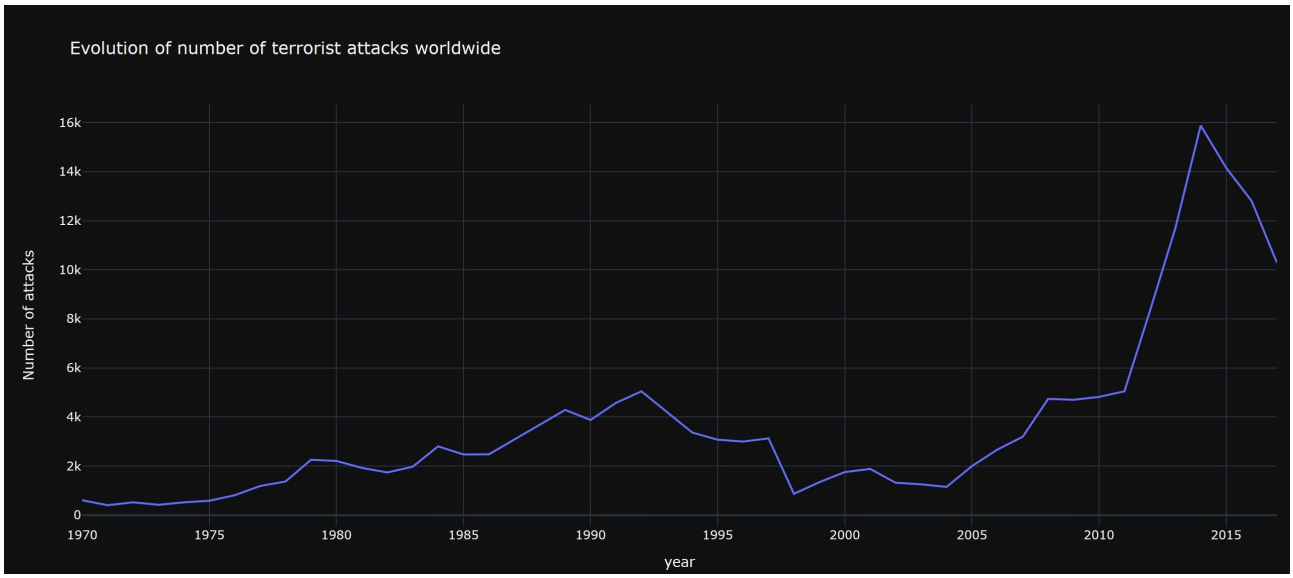

Figure 2: Visualization with evolution of number of attacks instead.

References

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