Process Book - Data Visualization - EPFL

Chiara Orvati, Khalid Omari, Skander Hajri

December 22, 2017

1 Introduction

Terrorism is one of the most covered topics by the media. However, it is often the case that only terror attacks on first world countries such as the USA are covered extensively, and the everyday horror in other countries is swept under the carpet. With our dataset we would like to show the global scope of terrorism and the countries which are most affected. Moreover, we would like to show details about individual terror groups, like what weapons do they use most often, or what target do they typically attack? Our target audience is the everyday citizen, who would like to gain some insight about terrorist attacks in the past 40 to 50 years, in order to complement the information shared by the mainstream media.

1.1 Related Work

The inspiration for this project comes from the work *An untold story of terrorism in India* by Kaggle user *BharathSivaraman* [uB]. The author claims that the media discourse of terrorism in India has always centered around terrorism in Jammu & Kashmir and has neglected to highlight the real hot bed of terrorism in India, the North east insurgency. The analysis of the author attempts to highlight extremism in the northeastern part of India in order to complement the biased coverage by the media. Instead of focusing on a single country, we want to bring to the user the global scope of terrorism: As media coverage of terrorism is often very local and only focused on the most recent attack, we want to create a visualization with a more complete view of terrorism across the world, which also attempts to summarize certain aspects of terrorism. In order to summarize nearly 50 years of terrorism in an engaging way, we drew inspiration from the plot shown in Figure 1, created by *BharathSivaraman*, to create our chord diagram.

1.2 Dataset Description

The Global Terrorism Database (GTD, [gtd]) is an open-source database including information on terrorist attacks around the world from 1970 through 2016. The database is maintained and published by the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland. Annual updates of the dataset are planned for the future with the year 2017 being in progress. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and includes more than 170,000 cases. For every incident, more than 100 variables are available, amongst which are basic information, like the date and location of the incident, but also more interesting variables, like the attack type (e.g. Bombing, Assassination, Kidnapping), the target (e.g. Military, Private Citizens & Property, Transportation), and the name of the group who claimed responsibility for the attack. For more information on the available variables, please refer to the official Global Terrorism Database Codebook [cod17].

1.3 Processing Steps

As the dataset covers more than 100 parameters for each incident, we first had to determine the ones that might be relevant to us and our specific problem. After a thorough examination of the variables' documentation, we kept 45 parameters; the rest was filtered out using Python's pandas library, and the result was stored in a new csv-file. For the final visualization even less variables were used; these are listed in the next two paragraphs.

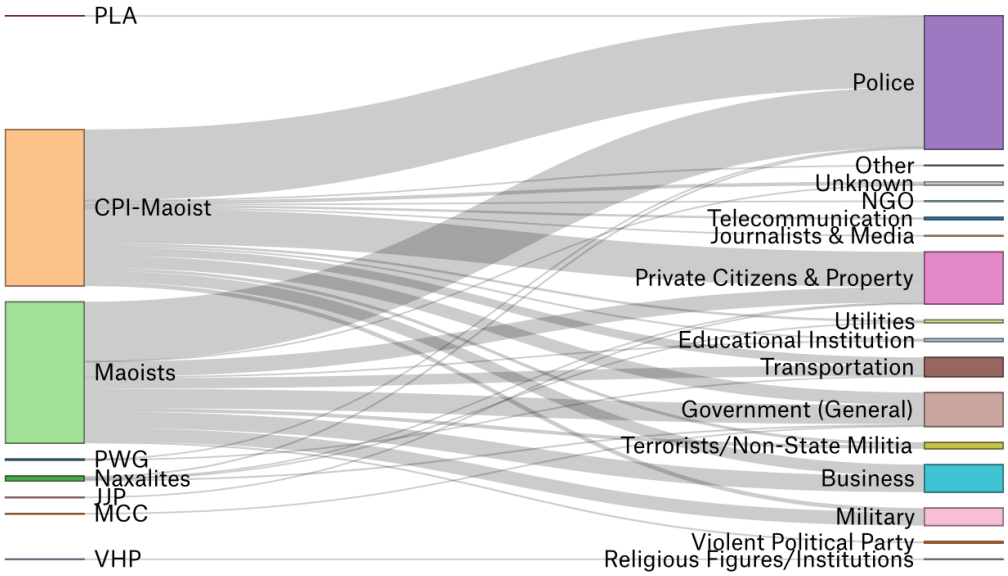

Figure 1: Inspiration for the chord diagram, courtesy of Kaggle user BharathSivaraman

Chord Diagram For our interactive chord diagram (see next section for more information), we computed adjacency matrices between pairs of variables using Python. Again, we used the pandas library and not d3 because the information summarized in the chord diagram is fairly static and does not change with every update of the database. Hence, it suffices to go through this process once - and not every time we want to see the chord diagram - which also benefits the speed of our visualization. With the chord diagram we focus on the 30 most common terrorist groups (i.e. the groups responsible for the most attacks), and show their relation to other variables. In order to guarantee the legibility of the diagram, we need to select discrete variables with not too many categories. Based on these observations, we chose to characterize the terrorist groups by the regions in which they're active, the weapons they use, their targets & preferred type of attack, how often they're successful in their attacks and which groups conduct suicide attacks. We thus computed adjacency matrices for the following pairs of variables:

- gname ↔ region_txt
- gname ↔ attacktype1_txt
- gname ↔ targtype1_txt
- gname ↔ weaptype1_txt
- gname ↔ success
- gname ↔ suicide

The values these variables can take are listed in Table 1.
Note that the group name 'Unknown' is actually the most common with about 46% of all incidents. As it makes no sense to display information for this category, we first filtered out all events where the attacker group was unknown. Finally, the 30 most common groups make out 31% of all attacks.

Table 1: Variables used in the Chord Diagram

Variable name	Values
gname	Name of the attacker group, taken from standardized list of group names. Can be "Unknown".
region_txt	North America, Central America & Caribbean, South America, East Asia, Southeast Asia, South Asia, Central Asia, Western Europe, Eastern Europe, Middle East & North Africa, Sub-Saharan Africa, Australasia & Oceania
attacktype1_txt	Assassination, Hijacking, Kidnapping, Barricade Incident, Bombing/ Explosion, Armed Assault, Unarmed Assault, Facility/ Infrastructure Attack, Unknown
targtype1_txt	Buisness, Governement(General), Police, Military, Abortion related, Airports & Aircraft, Government (Diplomatic), Educational institution, Food or water supply, Journalists & Media, Maritime, NGO, Other, Private citizens & Property, Religious figures/ Institutions, Telecommunication, Terrorists/ Non-state militias, Tourists, Transportation, Unknown, Utilities, Violent political parties
weaptype1_txt	Biological, Chemical, Radiological, Nuclear, Firearms, Explosives/Bombs/Dynamite, Fake Weapons, Incendiary, Melee, Vehicle, Sabotage Equipment, Other, Unknown
success	1 if the attack was successfull, 0 otherwise
suicide	1 if it was a suicide attack, 0 otherwise

Table 2: Variables used in the World Map

Variable name	Values
gname	Name of the attacker group, taken from standardized list of group names. Can be "Unknown".
country_txt	This field identifies the country or location where the incident occurred. In the case where the country in which an incident occurred cannot be identified, it is coded as “Unknown.”
longitude	This field records the longitude (based on WGS1984 standards) of the city in which the event occurred.
latitude	This field records the latitude (based on WGS1984 standards) of the city in which the event occurred.
eventid	Unique identifier of a terrorist attack which was mainly used to count the number of attacks per country and geographical coordinates.

World Map To create our map we first filter out the features listed in Table 2 and then group by longitude and latitude and then by gname to get every group of attacks from a specific terrorist organization per location. This will allow us to create the circles on the map and show the number of attacks in a certain place. We then produced a file that corresponds to all the countries listed in the dataset, alongside the number of events per country, and make them identical to the ones in the topojson file used to create the initial map. For example 'West Bank and Gaza Strip' corresponds to only 'West Bank' on the map, or 'Tanzania' corresponds to 'United Republic of Tanzania'.
Next, we choose to get rid of events with an 'unknown' tag for the attacker group as it isn't very informative and there's a very high amount of it, yet we count them in the total number of attack per country. As a last preprocessing step, we keep only the top five gnames (i.e. the attacker group who are responsible for the highest number of attacks) in each country. Otherwise the map would be too slow and not readable for countries with high/moderate number of attacks.

2 Visualization

2.1 Designs

Initial Designs At the beginning, our goal was to show a temporal evolution of the terror attacks from 1970 up to 2016. We wanted to show an evolution in the number and places of attacks, but also in the weapons used by the different terrorists groups, and many other aspects. Additionally, we wanted to display two different types of visualization: a historical (i.e. a timeline) and a geographical (i.e. a map), in order to bring more dimension to the data. A design for a multiview-type of visualization is shown in Figure 2. A first sketch for the timeline is shown in Figure 3.

Problems encountered and evolution towards final design The initial designs suffered from several issues. First, after starting the implementation of the timeline, we began to realize that a timeline can be very boring if not realized very carefully and in an inventive way. Also, as there are over 170'000 events registered in the database, it would take the user considerable time to explore the visualization from 1970 to 2016. Moreover, due to the huge amount of data, the initial visualization was very slow, thus not guaranteeing a pleasant experience to the user. As if this wasn't already enough, we realized that often the exact date of the incident was missing and that only the year was indicated. This practically forced us to summarize the timeline by years, although it could have been interesting to have a greater level of detail. For the reasons above, we then decided to not implement such a timeline at all.
With the timeline gone, the multiview-visualization (shown in Figure 2) was not possible anymore as such. However, we already feared from the beginning that this visualization could overwhelm the user, and would possibly not fit on one screen without needing to scroll. After some consideration, we decided to go with two separate visualizations (meaning on seperate screens), which are linked together by the different terrorist groups: First, we wanted to implement a map of the world where the attacks were indicated by circles. The size of the circles would show in logarithmic scale the number of attacks that happened in this location, and the color would indicate the terrorist group that conducted the attack.
Furthermore, we wanted to show some information for single countries, such as pie charts showing the different terrorist groups active in this country. But this would have been quiet out of tone, as it showed to be very unaesthetic and not so informative, as we can clearly see the dominant group in a country when selecting that particular group in the dropdown list.
Another difficulty was that some countries such as Somaliland is referred by the dataset as part of Somalia while it isn't in the map because of the status of Somaliland being a self-declared state. This causes that some events that occurred geographically in Somaliland are only shown when clicking on Somalia. Finally there's a problem with the fact that Russia's territory is present at both edges of the map, this causes weird zooming effect when clicking it.
The second visualization we agreed upon was a chord diagram relating the 30 most common terrorist groups to different variables such as the type of target they usually attack (see Section 1.3 for more information). The variables can be selected in a drop down menu. The decision for the chord diagram was made very quickly after we had seen the example shown in Figure 1, and how it could effectively show relationships between entities.

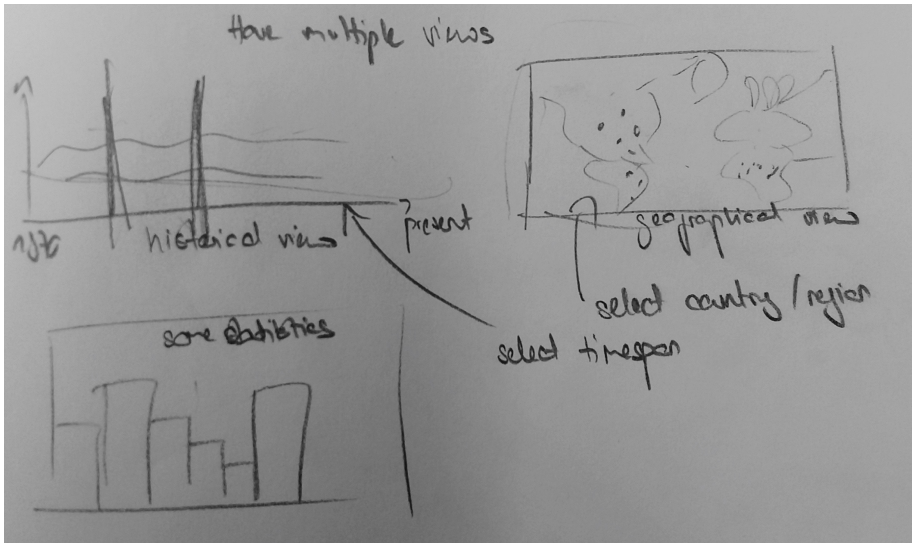

Figure 2: First design for a multiview: The user would select the timespan in the timeline (upper left corner) and the geographical area in the map (upper right corner), and some statistics would then be displayed at the bottom.

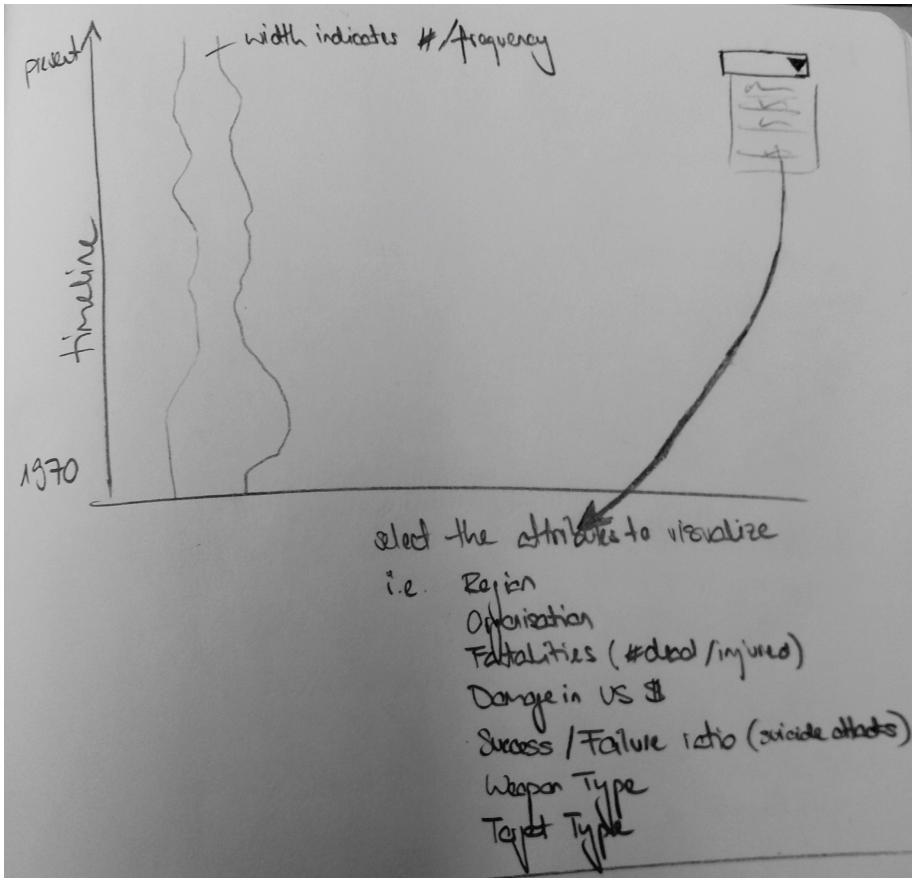

Figure 3: First design for a timeline. The user vertically explores the selected statistic, with the most recent placed at the top.

Figure 4: The initial color scheme for the chord diagram

Figure 5: The final - more visually pleasing - colorscheme for the chord diagram.

Final Designs The final designs are shown in Figure 8 for the chord diagram and in Figure 9 for the map.
Let's first discuss the chord diagram. In order to well differentiate the terrorist group names from the parameter selected by the user, we took care to keep the group names in a single, neutral color. As the user-selected parameter is of categorical type, therefore we need to have an according colormap as well. At first, we chose the `d3.schemeCategory20` colormap, which is displayed in Figure 4. This choice, however, turned out to be too bright and too colorful, especially in the case of the 'Target type' parameter, which has many categories. The result of this first colormap can be seen in Figure 4. As the design should be visually pleasing, we decided to make a compromise and chose the `d3.schemeCategory20b` colormap, displayed in Figure 5, which is a mix between categorical colors and color gradients.
In the case of the 'Target type' parameter, we don't use a specific assignment of colors to categories, as we need to use all colors. In fact, as this parameters has 21 categories, we repeat the colormap. The contrary is the case for the remaining parameters: the parameters 'Attack type', 'Weapon type', 'Attack region' have between 9 and 12 categories. If we just used the colormap as is, we wouldn't use all the main colors it presents, but only use the 3 main colors blue, green and brown and their 4 respective variations. Thus, we manually selected the colors to use in order to better distinguish the categories. A comparison between the default colors and the manually selected ones is demonstrated in Figure 7. Finally, for the two binary parameters 'Successful attacks' and 'Suicide attacks', we chose green if attacks belonged to this category and red otherwise.

Nevertheless, carefully chosen colors are not enough in order to allow the user to precisely perceive the connections between categories. For this reason we implement a 'on-hover' feature: The user can now hover over either a group name or a category of the selected parameter (example: the group name 'Islamic State of Iraq and the Levant' in Figure 8) and the chords connected to it are highlighted while the rest fade into the background. This allows for a better distinction of details by the user, who can now for instance easily learn what the typical targets of ISIS are, which terror groups are active in which regions of the world, and that actually very few terrorist groups use suicide attacks extensively (refer to Section 3 for more information on this). For big terror groups, like the Taliban, we can clearly see to which other categories they're connected to and we can even somewhat distinguish the proportions of incoming chord groups. As we want to provide the user with clear and easy-to-read information - even for small categories - we add an additional feature that is displayed when hovering over the category: we show the names and percentages of the three categories to which the hovered-on parameter is most connected to. In the example shown in Figure 8 we can now easily read that 40% of all attacks conducted by ISIL target private citizens and property. Now, we can obtain more precise information for the bigger categories and also extract meaningful information for smaller categories.

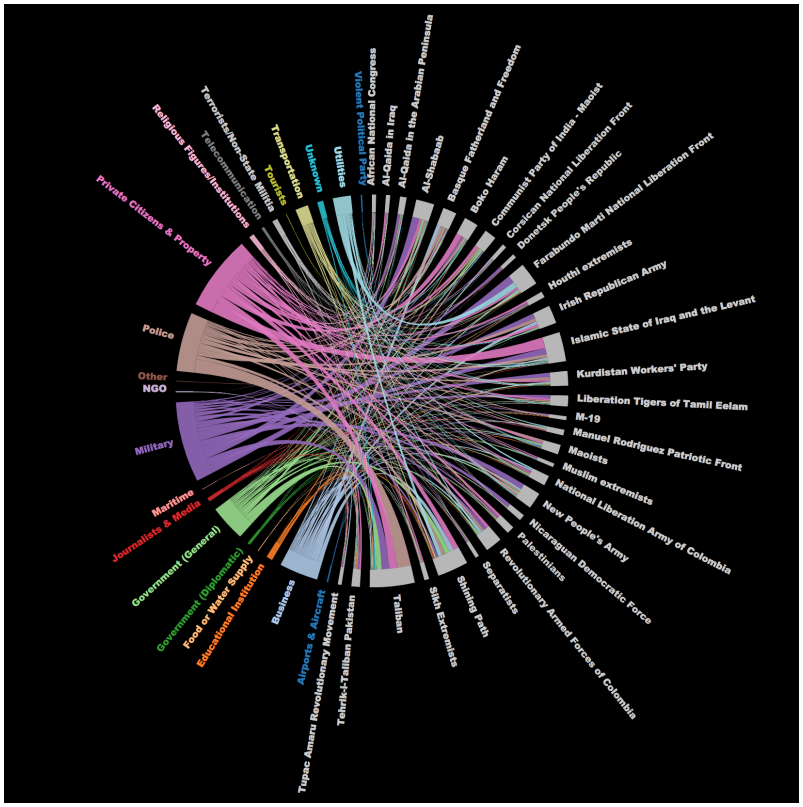

Figure 6: Chord diagram with the inital colorscheme.

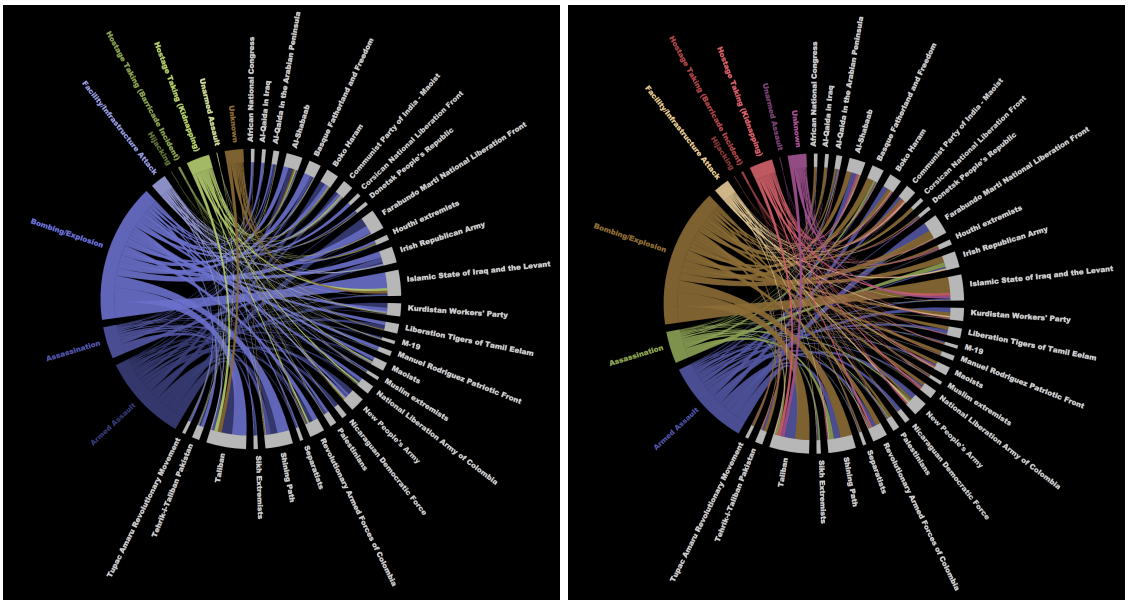

Figure 7: Comparison between default colors and manually selected colors

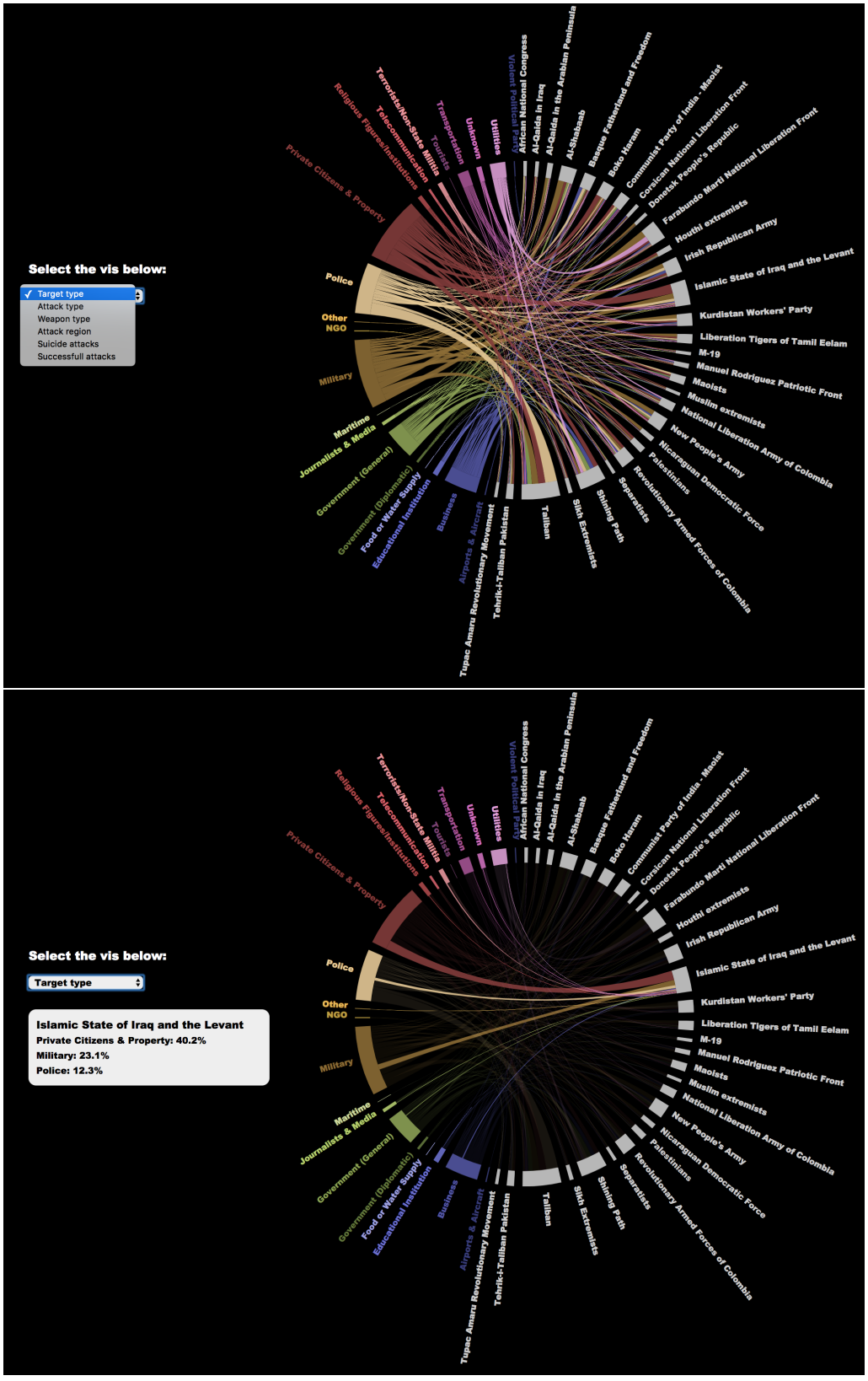

Figure 8: Final chord diagram: when hovering over the text or the arc of a category, only the chords connected to it remain visible. Additionally, we display the top 3 contributions of the categories of the selected parameter.

Let's now consider the map visualization shown in Figure 9. The first view that the user obtains is a choropleth map of the entire world, where the different shades correspond to the total number of attacks in the country over the entire timespan from 1970 to 2016. To create the scale we linearly interpolate the colors from blue to red, and then create color classes which correspond to very low, low, medium, high and very high number of attacks in a given country. These classes where created so we have a more or less balanced number of countries in all classes apart from the first one which contain all the outliers and countries with very few information. This allows the user to obtain a global view of terrorism and perceive the countries most subjected to terrorism. The user can then click on the map in order to obtain more information on specific countries. In a first time, the locations of the attacks in this country are indicated by circles which are colored by the terror group responsible for the attack. When hovering over such a circle, it is displayed how many attacks were perpetrated in this particular location and which group conducted the attack. Furthermore, the total number of attacks is displayed above the map. In a menu appearing in the upper left corner it is then possible to select single terror groups and to display only the attacks conducted by this group in the selected country. The groups listed in the drop-down menu are the five most represented terror groups in this country. This feature allows the user to learn more about local terrorism, i.e. which groups are active in this country and how much they're active. The user can also easily see whether the country is affected by 'globally' operating terrorist groups like ISIS who conduct attacks in many countries.

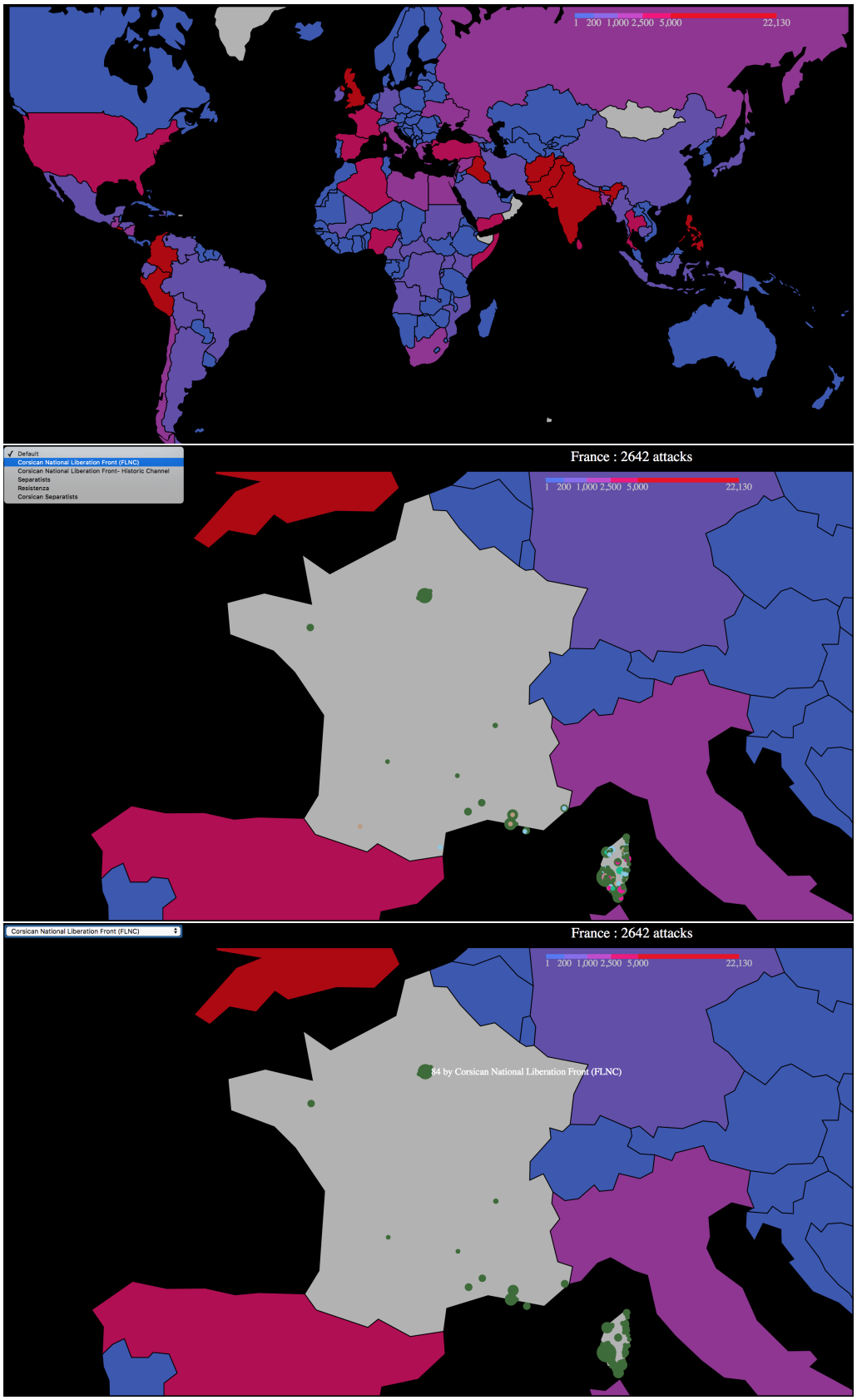

Figure 9: Final world map. top: choropleth map, middle: all the attacks in France, bottom: all attacks conducted by the Corsican National Liberation Front, with Paris being a major target (84 attacks).

3 Discussion

3.1 What we learned about the data

We will now list some examples of what we learned about our dataset using our visualizations.

Chord Diagram In Figure 10 we depict the proportion of suicide attacks for each terror group. We can see that only 10 groups out of the 30 use this type of attack, and also that the Taliban and ISIS are responsible for more than half of all suicide attacks. If we explore the visualization for the target type, we can detect some interesting details: we see that the three main targets of the Taliban are the police, private citizens & property, and the military. For the case of ISIS, the order is different: the main target is private citizens & property, followed by the military and the police, as it can be seen in the lower part of Figure 8. The Farabundo Marti National Liberation Front gives another interesting example: it alone is responsible for about one third of the attacks on 'Utilities', which in this context are facilities for the transmission or generation of energy. Finally, when selecting 'Attack Region', we can easily determine which groups are active in which regions of the world. For instance, in Sub-Saharan Africa, the active groups are the African National Congress, Al-Shabaab and Boko Haram, as it is shown in Figure 11. Equally, we immediately notice the regions most affected by terrorism, amongst which are South Asia and South America, also demonstrated in Figure 11. Usually, we don't necessarily associate these regions with terrorism. This is because the groups active in these regions are very local, i.e. their activities are restrained to this area and do not expand to Europe or America. Hence, as these groups don't pose an immediate threat to the everyday first world citizen, they are only rarely mentioned in news reports. This is a pity, as we can clearly observe that about one third of the worldwide terror attacks happen in South Asia and South America (Note: this proportion of one third is actually true for the entire dataset, and not only for the top 30 terror groups, as a quick manipulation in Python shows).

World Map From the choropleth map, which indicates the number of attacks per country, we can make out the countries most affected by terrorism: Besides the 'expected' ones like Iraq (22'130 events), Afghanistan (11'306 events) & Pakistan (13'634 events), we also notice some countries which are not typically covered by the mainstream media: We notice for example India with nearly 11'000 events, the Phillipines with 6212 attacks, and Peru and Columbia, which both have between 6000 and 9000 attacks. Another interesting example is El Salvador, which has more than 5000 attacks, which is a huge amount for such a small country. In fact, in the visualization, the whole country is filled with the circles as it can be seen in Figure 12. It is sad that such countries do not get more attention from the mainstream media, considering the fact that the United States of America, which cover almost 500 times the surface of El Salvador, have only suffered from half the number of attacks.
When looking at the situation in the USA in Figure 13, we can see that the most prominent group are Anti-Abortion extremists followed by Left-Wing Militants (although these attacks may date a while back).

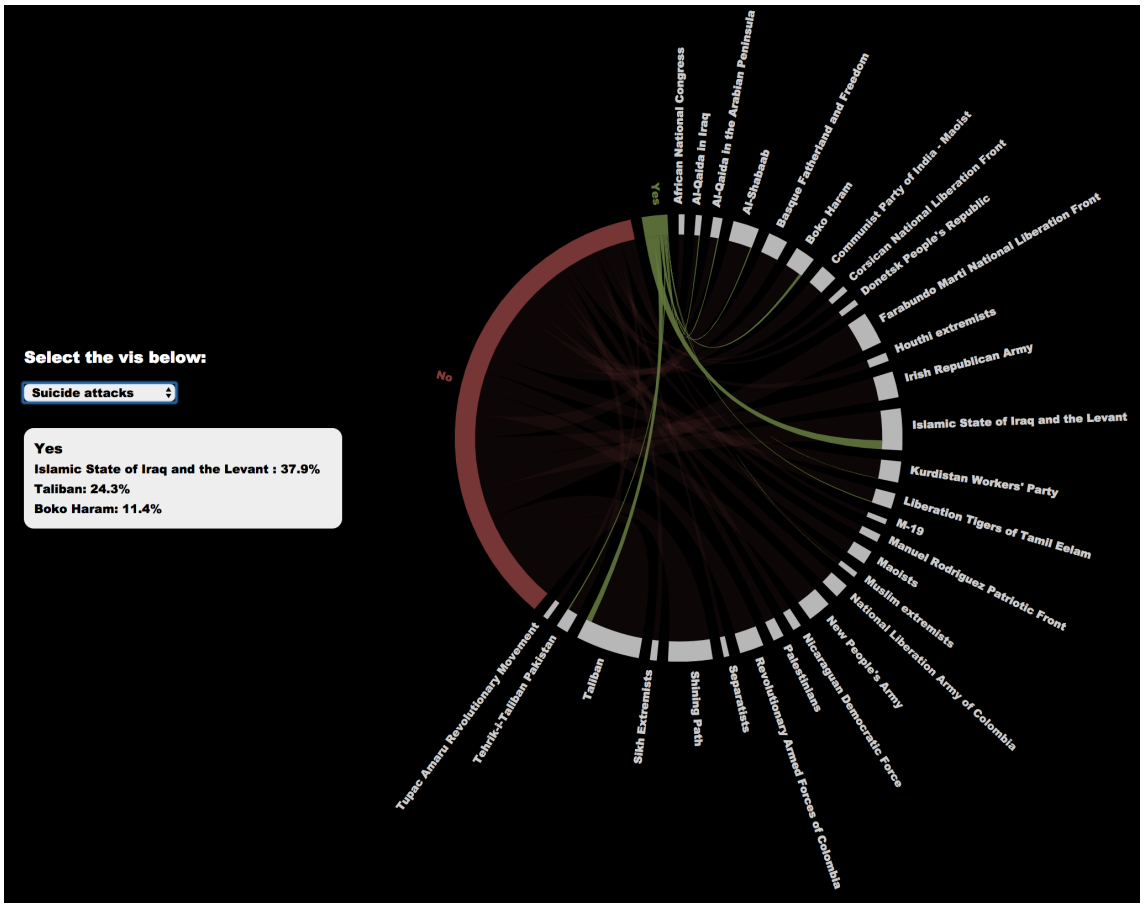

Figure 10: Proportion of suicide attacks per group. Suicide attacks are indicated by "Yes".

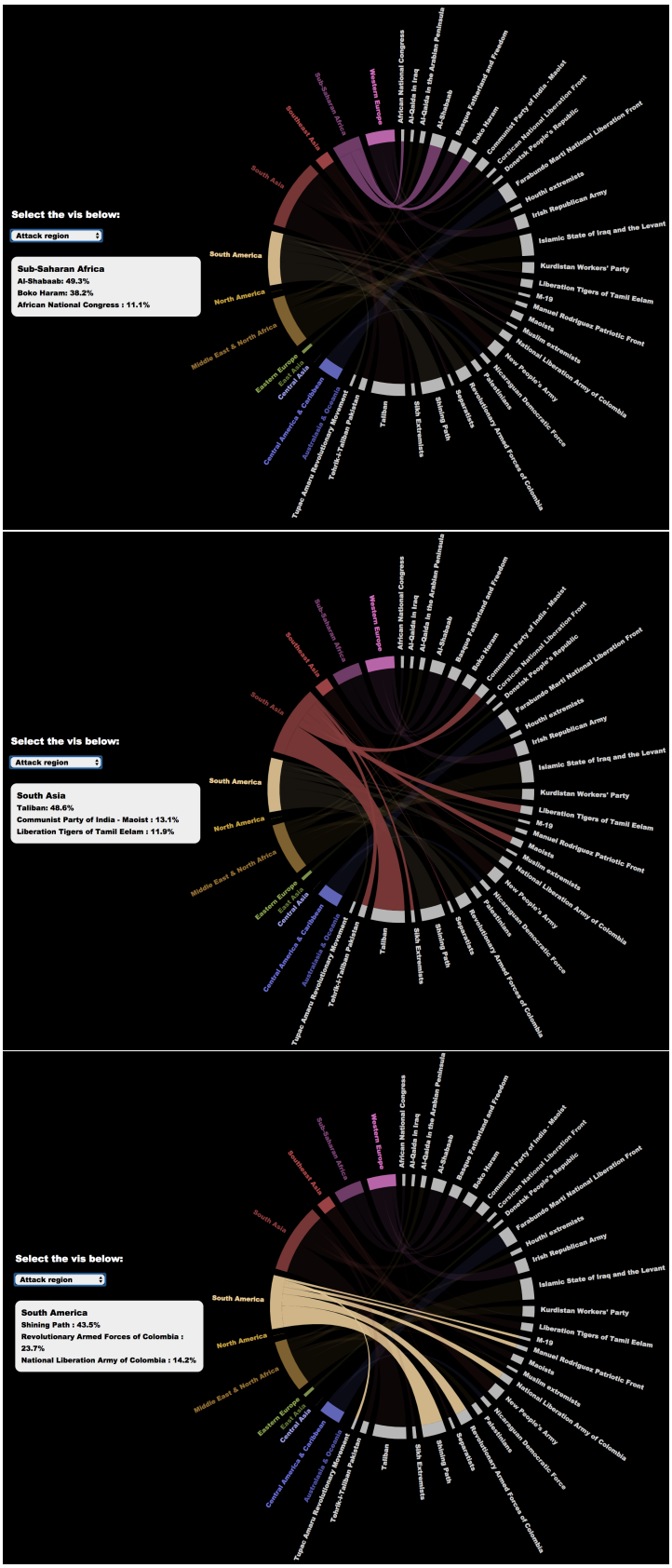

Figure 11: Chord diagrams for the parameter 'Attack Region'

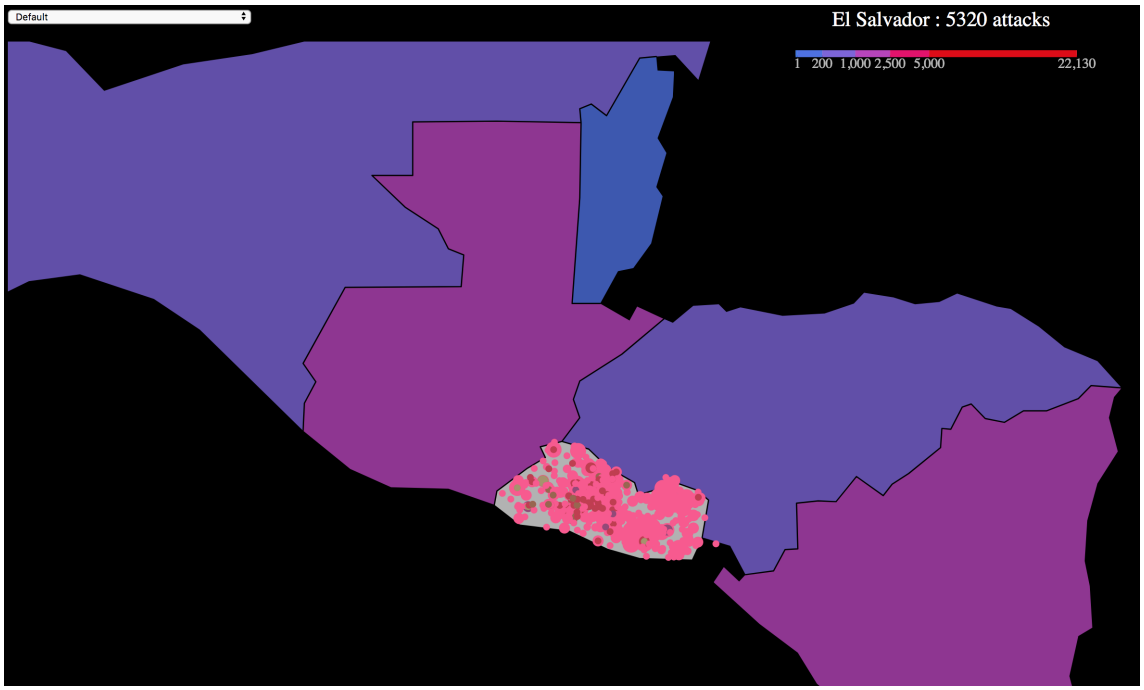

Figure 12: All attacks in El Salvador from 1970 to 2016.

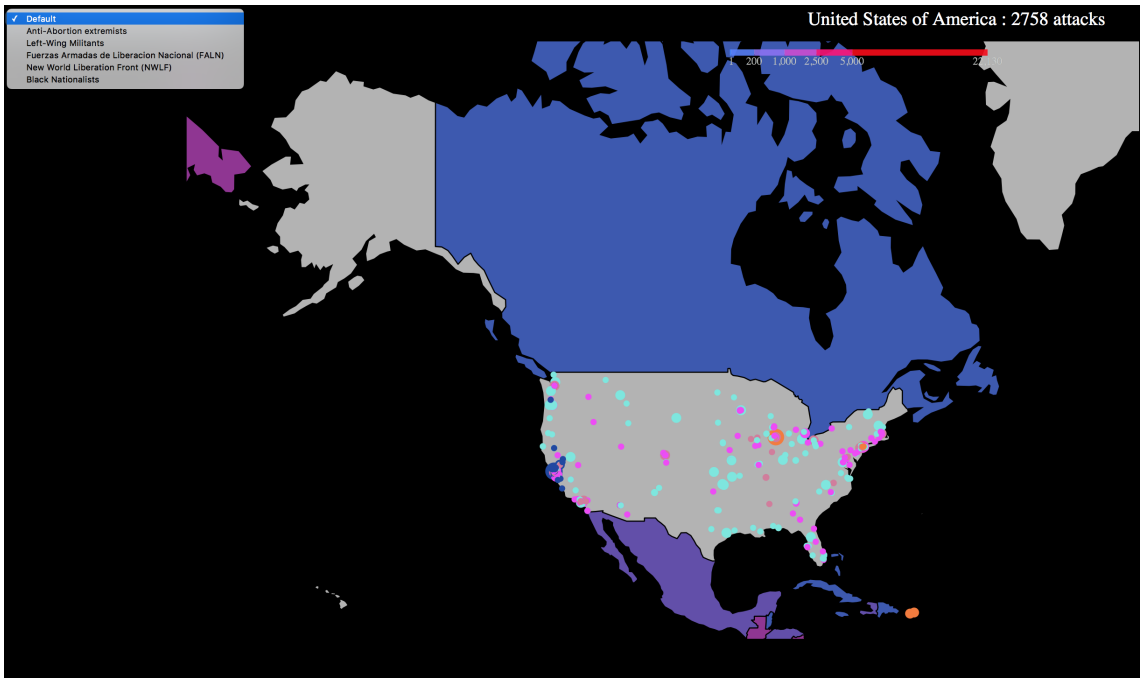

Figure 13: Most attacks in the USA are conducted by Anti-Abortion-Extremists.

3.2 Future Improvements

Chord Diagram A nice improvement here would be to generate the data to be displayed on the fly: this would allow for more flexibility in the shown content. The user could for example select specific terror groups, or dynamically select any two parameters for which he wants to show the co-occurrences. It would also be interesting to be able to select a timespan for the chord diagram. In this way we could see whether the relationships between variables changed over time. The main difficulty herein is to handle the huge amount of data in an efficient way, such as to guarantee a smooth and real-time experience for the user.

World Map A great new additional feature would be to have a timeline with which we can select the timespan. This would help to understand the data better: it may occur that a terror group is only active during a relatively short timespan. Take for example the case of the United Kingdom, which appears in bright red counting a number of 5100 events. Most of these attacks have been executed by the Irish Republican Army, as shown in Figure 14, which is no longer active and our current visualization does not represent this fact. Another extension would be to have for example a piechart, where we numerically show the proportions of terror groups in a country.

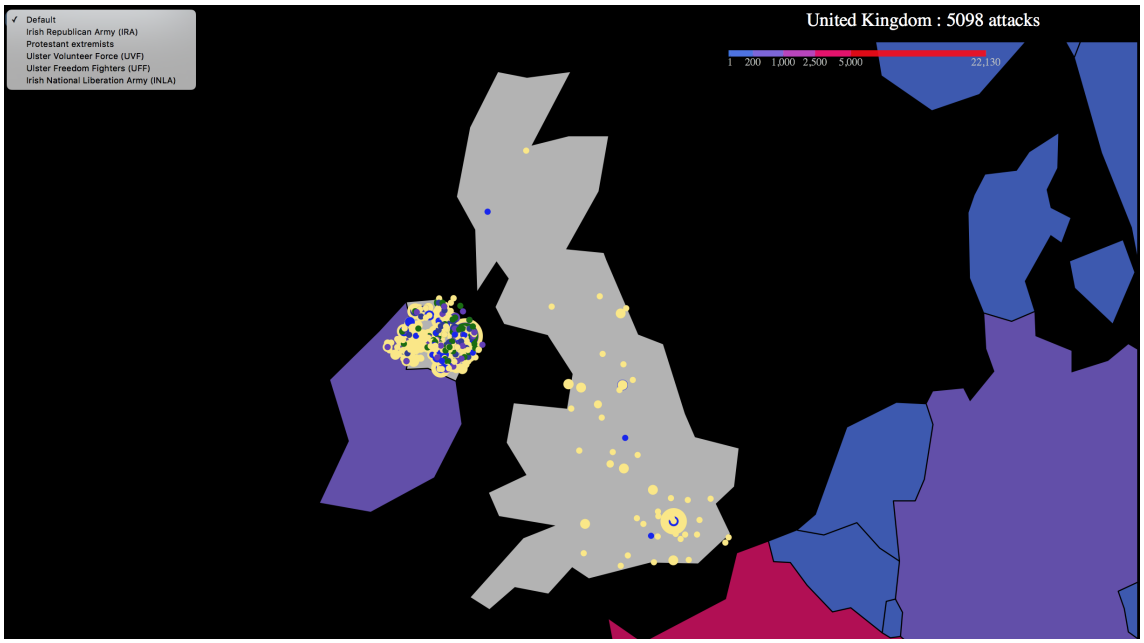

Figure 14: All attacks in the UK.

3.3 Peer evaluations

Our group worked really well and everyone was happy with the others work, as we can see from Table 3 below.

Table 3: Peer evaluations

Question	Khalid	Skander	Chiara
Preparation – were they prepared during team meetings?	Yes(Chiara), Yes(Skander)	Yes(Chiara), Yes(Khalid)	Yes(Skander), Yes(Khalid)
Contribution – did they contribute productively to the team discussion and work?	Yes(Chiara), Yes(Skander)	Yes(Chiara), Yes(Khalid)	Yes(Skander), Yes(Khalid)
Respect for others' ideas – did they encourage others to contribute their ideas?	Yes(Chiara), Yes(Skander)	Yes(Chiara), Yes(Khalid)	Yes(Skander), Yes(Khalid)
Flexibility – were they flexible when disagreements occurred?	Yes(Chiara), Yes(Skander)	Yes(Chiara), Yes(Khalid)	Yes(Skander), Yes(Khalid)

References

[cod17] National consortium for the study of terrorism and responses to terrorism (start). global terrorism database codebook: Inclusion criteria and variables. http://start.umd.edu/gtd/downloads/Codebook.pdf, June 2017. Last accessed: December 2017.

[gtd] Global terrorism database. https://www.kaggle.com/START-UMD/gtd/data. Accessed: November 2017; National Consortium for the Study of Terrorism and Responses to Terrorism (START). Global Terrorism Database - Full Data File 1970-2016, June 2017 Release.

[uB] Kaggle user BharathSivaraman. An untold story of terrorism in india. https://www.kaggle.com/bharathsivaraman/an-untold-story-of-terrorism-in-india. Last accessed: December 2017.