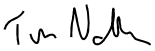
THE UNIVERSITY OF SYDNEY

Assignment Coversheet – GROUP ASSIGNMENT

Please fill in your details below. Use one form for each group assignment.

Personal Details of Students

Group Name/Number	Assignment 2 Groups 7				
Family Name	Given Name (s)	Student Number (SID)	Unikey	Contribution + percentage	Signature
Colino	Ignacio	480201931	icol5839	Undertook the design, implementation and evaluation, and discussion for Task 1. Joint evaluation other tasks(16.67%)	
Kam	Michael	480103233	mkam9179	Wrote the meeting minutes, conclusion and undertook the visualisation for Task 1. Joint evaluation other tasks(16.67%)	
Nallaiah	Timothy Rajkumar	198848633	tnal5652	Undertook the design, implementation and evaluation, and discussion for Task 2. Joint evaluation other tasks (16.67%)	
Ngo	Bao Chau	470157446	bngo5284	Undertook the design, implementation and evaluation, and discussion for Task 3 and 4. . Joint evaluation other tasks (16.67%)	
Gomez	Ereina	410025257	egom8541	Co-wrote the analysis, visualisation, and evaluation for Task 3 and 4 (16.67%)	
Pierre	Josh	430272420	jpie1166	Wrote the Introduction section. Joint evaluation other tasks (16.67%)	

Assignment Details:

Assignment Title		**Final Report**			
Assignment number		**2**			
Unit of Study Tutor		**Jialu Chen, Supraja Sridharan**			
Group or Tutorial ID		**Group 7**			
Due Date	**01/11/2018**	Submission Date	**01/11/2018**	Word Count	**5857**

Declaration:

1. I understand that all forms of plagiarism and unauthorised collusion are regarded as academic dishonesty by the university, resulting in penalties including failure of the unit of study and possible disciplinary action.
2. I have completed the **Academic Honesty Education Module** on Canvas.
3. I understand that failure to comply with the Academic Dishonesty and Plagiarism in Coursework Policy can lead to the University commencing proceedings against me for potential student misconduct under Chapter 8 of the ***University of Sydney By-Law 1999*** (as amended).
4. This work is substantially my own, and to the extent that any part of this work is not my own I have indicated that it is not my own by acknowledging the source of that part or those parts of the work.
5. The assessment has not been submitted previously for assessment in this or any other unit, or another institution.
6. I acknowledge that the assessor of this assignment may, for the purpose of assessing this assignment may:
   a. Reproduce this assignment and provide a copy to another member of the school; and/or
   b. Use similarity detection software (which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking).
7. I have retained a duplicate copy of the assignment.

Please type in your group number here to acknowledge this declaration:	7

COMP5048 Assignment 2
Final Report

Group 7
Michael Kam - 480103233
Timothy Nallaiah - 198848633
Ignacio Colino - 480201931
Ereina Gomez - 410025257
Bao Chau Ngo - 470157446
Josh Pierre - 430272420

1. Introduction

1.1 Dataset

The Global Terrorism Database (GTD) contains information on terrorist events from 1970-2017 (not including 1993), this includes 180,000 separate attacks from around the world ("Global Terrorism Database", 2018). Records in the GTD are sourced entirely from publicly available, open sourced information: mainly news articles, journals and legal documents. The GTD has been managed by different parties over time with the main contributors for each period listed below.

1. 1970-1997: A private security agency, established in the US.
2. 1998-2008: Centre for Terrorism and Intelligence Studies
3. 2008-2011: University of New Haven
4. 2011 to present: University of Maryland, the Study of Terrorism and Responses to Terrorism (START)

During each phase of new ownership of the GTD, significant changes were made to the dataset or selection criteria of the dataset. From 1970-1997 each record was sourced in real time, meaning if a newspaper article reported on a case of terrorism it would be added to the GTD immediately. The data from 1998-2007 was collected retrospectively, the lack of real-time sourcing is a potential loss of data, as newspapers or journals that may have reported on terrorist attacks at the time might have gone out of business and the reported attack might be lost. Additionally, in 1998 a new method for inclusion into the GTD was defined and to synthesize the data many records pre-1998 were deleted to match the criteria. Also in 1998, 84 additional fields (to the existing 44) were added, meaning all records pre-1998 have 2/3 of its fields empty. From 2007 to present all records were added in real time once again and from 2012 onwards the methodology for collecting data was significantly improved which skews the number of records found in the most recent years (8,500 separate attacks in 2012, followed by 12,000 and 17,000 attacks in 2013 and 2014 respectively).

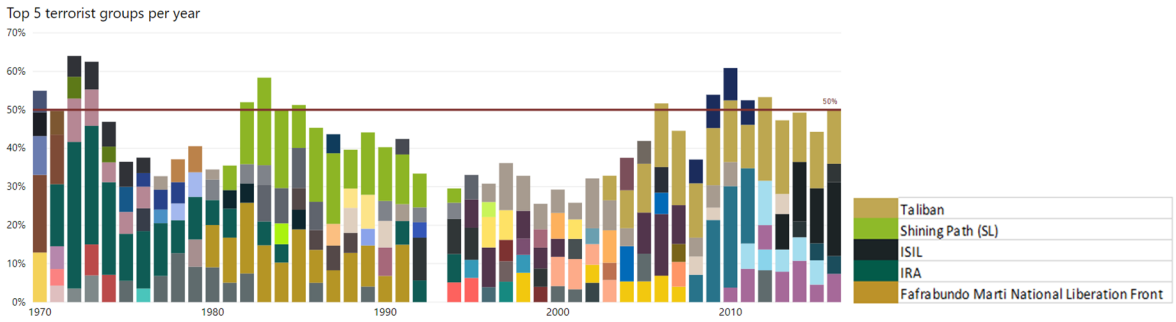
In summary, while the GTD is very extensive in the number of records and in-depth data it collects from around the globe, due to the duration of which the data has been collected and the nature of its sources, it has experienced major shifts in the number of records in each period which could influence our results.

1.2 Tasks

1. Show the trend of terrorist activities over time in different countries and regions
2. Find the insights from the 20 deadliest events, consisting of who, when, where, and what.
3. Show the relationship of the origin of the most active terrorist groups to their distribution of attack.
4. Analyse the pattern of the terrorist attacks based on the month, day, and public holidays on the related country.

1.3 Aims and Contribution

The GTD contains 132 fields, ranging from the number of deaths, ransom note text, amount of monetary damage done, etc. As a group we decided to focus on information that we have learnt about or events we could recall and then further analyse those, these highly impactful attacks were generally the highly fatal attacks carried out by well known terrorist groups e.g. 9/11, several attacks by ISIL & the Taliban in Iraq and the Middle East, the dominance of the Shining Path group in Peru, IRA in Ireland. Further exploratory analysis on the impact of terrorist groups in general, showed that in recent years the top 5 groups contributed to ~50% of attacks in any one year. That is, despite the GTD showing information of over 3000 different groups, in any one year 5 groups still make up a large percentage of the data. Because of these considerations, as a group, we decided to further analyse the impact and reach of terrorist groups.



Tasks 1 to 3 explicitly name and highlight different groups attack the region they attacked, the period of time in which they were active and the number of deaths/wounded in each attack. While task 4 does not explicitly name the different terrorist groups, analyzing the calendar date in which attacks have occurred and the importance of certain days in different regions, which can show the trend or likelihood of attacks to occur on certain days.

2. Task 1: Show the trend of terrorist activities over time in different countries and regions

2.1 Design

2.1.1 Analysis

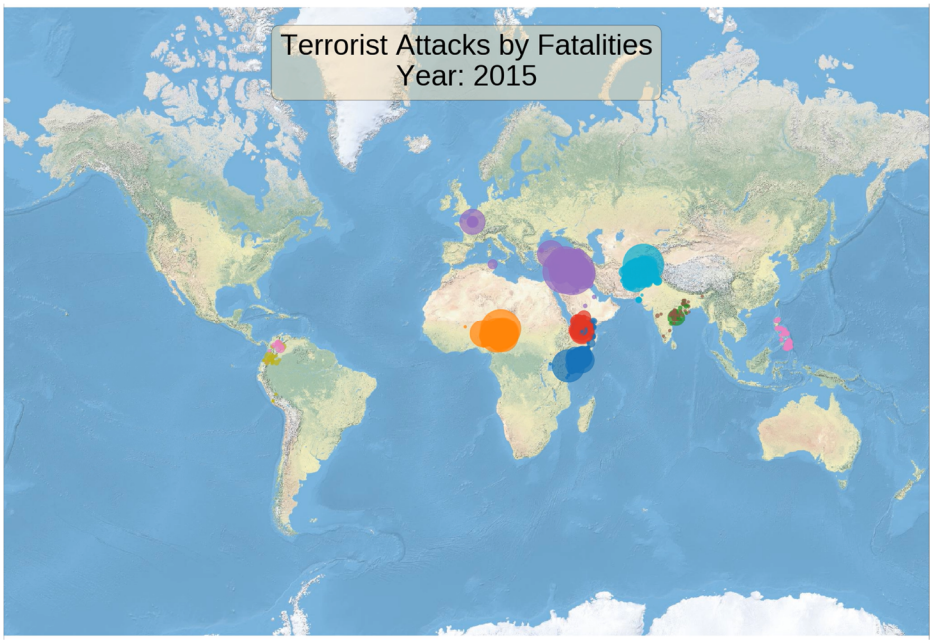
The first task of this analysis consists of getting a comprehensive visualisation of the evolution of terrorism through time and space. For this task, we aim to visualise the events in the dataset on a map that reflects the occurrence of the terrorist attacks by their location, as specified in the coordinate fields. With this method, we got a solid grasp of which countries and regions have been most affected by terrorism by the number of killed people, represented by the size of the scatters. We decided to focus on this metric, in contrast to others like material damage and wounded people because we believe it is the most serious consequence of terrorism by far.

This visualisation serves as a trigger for further analysis as it helps to get a network representation of the dataset through time that narrows interesting areas and groups. Since, in general, the groups tend to focalise their attacks in specific regions or even countries, the visualisation reveals which areas are affected by the top 20 groups. We also worked in getting an interactive visualisation, so that the users can explore regions and time at free will and discover aspects of interest to them.

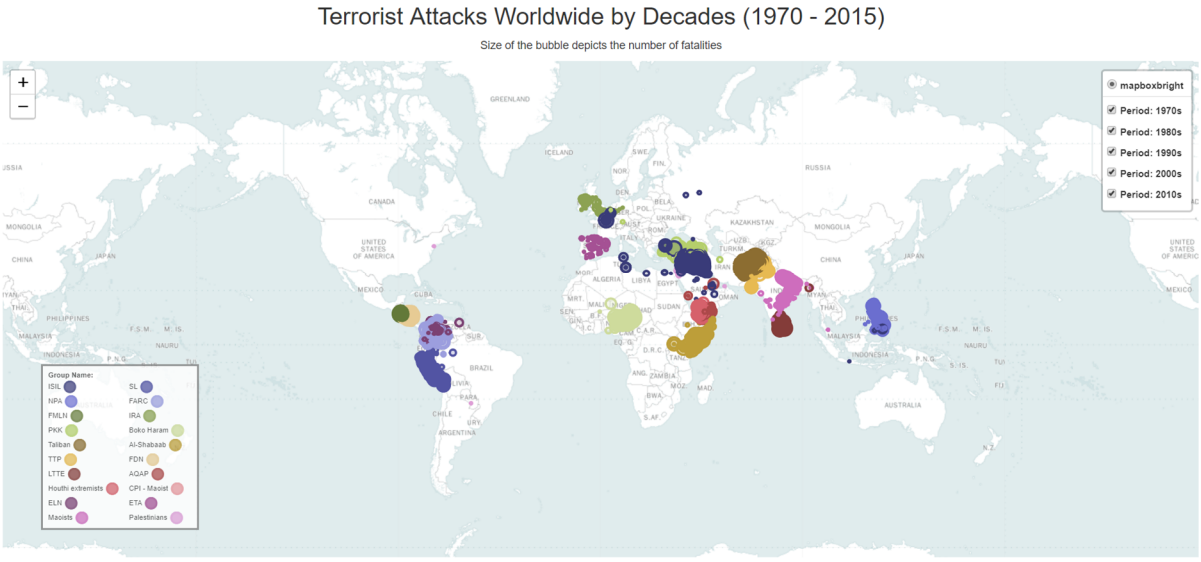
2.1.2 Visualisation

For the visualisation of this task, we implemented two approaches using world map where each data-point or record in the dataset is a scatter in the graph. The first approach involves animated map while the second one employs an interactive version.

The animated map allows us to visualise the evolution of the terrorist attacks through time. The characteristics of the scatter represent different attributes of the data. The damage produced by the size and the perpetrating group by the colour of the circle. For example, the image below shows the cluster and spread of the groups in 2015.



Meanwhile, the interactive map allows users to zoom in regions and also focus on different decades and particular event. This helps to get a granular capability to let the user go from a general picture of attacks to focusing on very specific regions that may be of interest. As we can see from the image below, we can confirm that most attacks are clustered based on the group and geographical location. Selecting different decades will show the spread and different active groups.



2.2 Implementation

For this task, we used Google Colaboratory for Python-based development. Pre-processing for the data is done using pandas library in order to organize and manipulate attributes and features as needed. We selected the top 20 groups based on the number of fatalities.

To implement the animated map, we used the Basemap Python library which provides the maps and cartography capabilities. Using Matplotlib and FFMPEG libraries we created an MP4 video animating the plots iteratively based on the year. Meanwhile, we used Folium library for the interactive map in the form of HTML. Each decade is represented as a layer that can be selected by users to filter the relevant attacks.

2.3 Evaluation

2.3.1 Results

As suspected, the groups mostly attacked locally up to neighbouring countries and regions. Meanwhile, some international groups like ISIL and Taliban have impacts on different regions. Using both interactive and animated map, we can see the different regions affected by terrorist attacks over the years and decades.

To evaluate the task, we performed a Heuristic Evaluation based on Nielsen's principles and severity rating (1995a, 1995b) to compare and contrast both approaches. We found the strengths and weaknesses with each method and summarised them as follows.

Animated Approach

Strengths

Description of Strengths	Heuristics Related
User can see the trend over year gradually	Visibility of system status

User can view the cluster and groups	Visibility of system status

Weaknesses / Problems

Description of Weakness	Heuristics Related	Severity Rating				
		Josh	Tim	Ereina	Chau	Michael
The legends are not shown in the plot	Visibility of system status	1	1	2	2	3
User needs to remember previous data	Recognition rather than recall	1	0	0	1	2
Overlapping of the scatters	Aesthetic and minimalist design	0	0	0	1	2

Interactive Approach

Strengths

Description of Strengths	Heuristics Related
User can see the trend over the decades	Visibility of system status
User can view the cluster and groups	Visibility of system status
User can inspect each decade and event	User control and freedom
Legends are shown	Visibility of system status

Weaknesses / Problems

Description of Weakness	Heuristics Related	Severity Rating				
		Josh	Tim	Ereina	Chau	Ignacio
Colours of the group look similar	Recognition rather than recall	1	1	2	1	1
Overlapping of the scatters	Aesthetic and minimalist design	0	0	0	0	0

2.3.2 Discussion

We realised from the evaluation that the issue with similar colours needs to be addressed. Since we focus on 20 groups, finding 20 contrasting colours might be inefficient. Combination of shape and colour might be able to be considered. In the animated map, the legend could not be shown. However, the problem is solved using the interactive approach. Other weaknesses are cosmetics and rated low by the experts.

For further work regarding this task, we believe clustering could help reduce the number of points in the map and still help to visualize the impact for groups that attack and focus on very specific areas. We focused on the top 20 groups by the number of

killed people but it would also be of interest to explore relevant groups for historically less affected regions and countries.

3. Task 2: Insights from the 20 Deadliest Events - Who? When? Where? What?

3.1 Design

3.1.1 Analysis

Task 3 focuses on the 20 deadliest events. The analysis focuses on obtaining insights from these events - Who? When? Where? And What?

- The idea to understand any patterns or trends in place, time, the perpetrator (actor) or event/conflict type within the deadliest events.
- The 20 most fatal events were grouped by time (decade) and region (with respect to the number of fatalities), with additional fields for the country involved and analysis of the event itself.
- Additional research was done on each event to augment the GTD database, including categorising the event in the context of a broader ongoing conflict.

3.1.2 Visualisation

The aim is to highlight the relationship and show the "flow" between the events and the key features of interest (time, actor, location and conflict). Sankey diagrams were chosen for their visual impact and ability to effectively convey flow patterns.

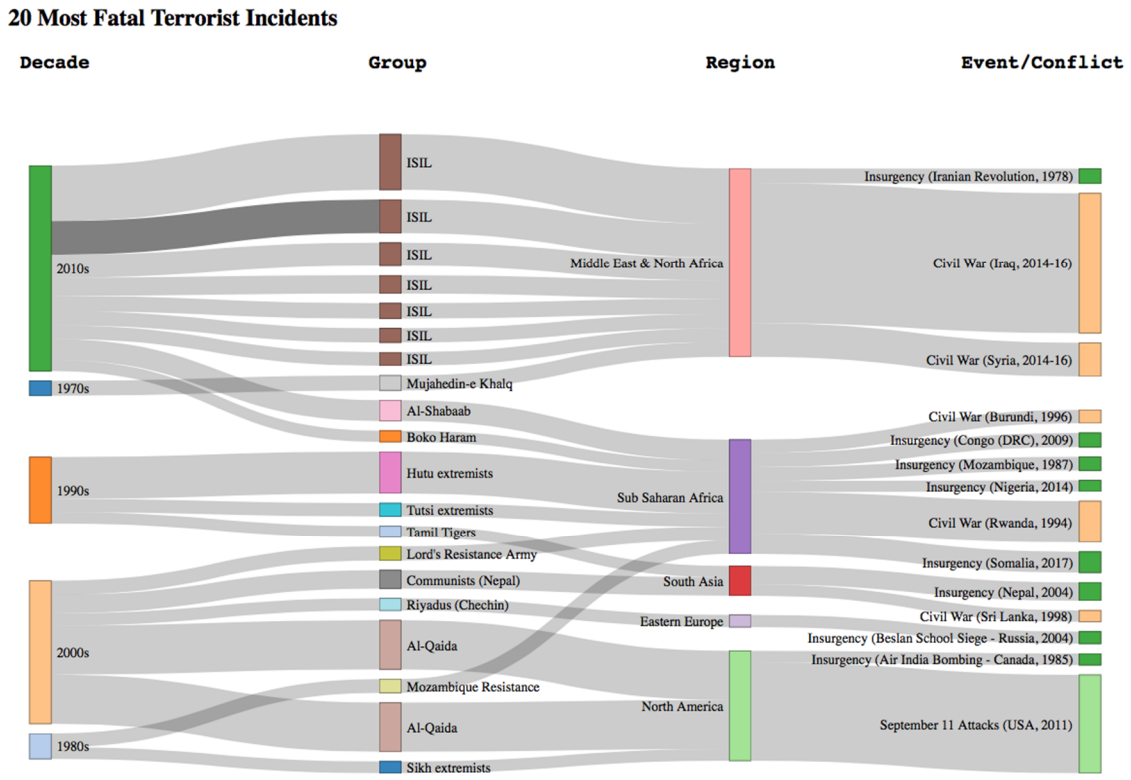
In the Sankey diagram, each key feature is represented by a separate axis: time (decade), actor (the terrorist organisation), region and event/conflict type. The width of edges conveys the number of fatalities. Sankey diagrams are interactive, highlighting linked edges and nodes when one hovers over the relevant portion of the diagram.

3.2 Implementation

NetworkD3 in R was used to implement the Sankey diagram. Nodes, edges and edge widths are detailed in a JSON file. Each item within each feature axis is represented by an individual node, with edges connecting items in axes and their width representing the number of fatalities. The layers are created through the ordering of the edges between the nodes. Network D3 optimises the ordering of the nodes in each axis to minimise edge crossing through specification of the 'number of iterations' of the underlying optimisation algorithm (1000 iterations used here).

The data was pre-processed (e.g. creation of new data features like decade) and analysed (grouped by the relevant data feature and ordered by the number of fatalities) using R. A JSON file was created to represent each node and edge based on this analysis. The Sankey diagram was constructed using NetworkD3's sankeyNetwork in

NetworkD3 library. The resulting diagram was saved in HTML format (with some formatting using HTML widgets) to allow for an interactive graphic.



3.3 Evaluation

3.3.1 Insights Obtained

Presentation of the 20 most deadly events using a Sankey diagram highlighted:

- Terrorist events are getting deadlier, with the current decade (2010s) seeing the deadliest event, the highest number of deadly events, and the highest total fatalities from such events. This was followed by the preceding decade (2000s).
- The recent rise in highly fatal events has been driven by the Islamic State terrorist group (ISIL), impacting the Middle East through the Iraqi and Syrian Civil Wars.
- By contrast, the drivers on the second most impacted region, Sub-Saharan Africa, have been more diverse by time, actor, and event/country.
- South Asia has a small number of very deadly events despite having a high number of overall fatalities in the GTD database.
- North America features prominently due to the September 11 attacks, despite having a (relatively) low number of fatalities overall in the GTD database.
- 19 of the 20 deadliest events are rooted in regional/local conflict (civil war or insurgency), the one exception being the September 11 attacks.

3.3.2 Results

The visualisation was evaluated using group self-evaluation based on Nielsen's heuristics. For this purpose, group members not involved in creating the Sankey

diagram acted as 'experts' and noted the strengths and rated the weaknesses of the visualisation. The results are tabled below.

Strengths

Description of Strengths	Heuristics Related
Ability to select edges, makes it more clear	User control and freedom. Flexibility and efficiency of use.
Hovering of edges makes it easy to navigate	User control and freedom.

Weaknesses / Problems

Description of Weakness	Heuristics Related	Severity Rating				
		Josh	Ignacio	Michael	Ereina	Chau
Cluttered with information	Aesthetic and minimalist design.	0	0	0	0	0
You can't link the event to the group directly	Match between system and real world.	2	2	2	2	2
You need to get used to the visual for a while	Flexibility and efficiency of use. Help and documentation.	1	1	1	0	1
The decade is not ordered chronologically	Match between system and real world.	1	1	2	1	1

3.3.3 Discussion

Most weaknesses identified were considered non-fatal and largely minimal/cosmetic. The inability to track the event directly to the group involved was noted. Curing this deficiency would involve creating new edges which would clutter the visualisation. Most flows are readily interpretable as is.

There is a trade-off between the amount of information conveyed via text and the potential to clutter the visualisation. The graphic does have a lot of text, although most 'experts' did not consider this a problem. The purpose of the visualisation was to help obtain insight, so the trade-off landed with more information. The balance could be different if the purpose was different.

Sankey diagrams order nodes to avoid crossings between flows (edges), which in itself is a challenging combinatorial problem (Zarate, Bodic, Dwyer, Gange & Stuckey, 2018). This provides a visually appealing design. One side-effect is that nodes on each axis can be re-ordered from original inputs. This only really affects the time (decade) axis. The trade-off between an expected ordering and minimising edge crossing and maximising visual impact fell to the visual impact. Key insights remain unaltered by the final node ordering here.

Sankey diagrams (including this one) score well on aesthetic graph drawing metrics as they have no edge bends or node overlaps, are quite symmetric, and nodes and edges are generally uniformly spaced (with variation in edge size here corresponding to the number of fatalities, which is an important effect to convey). Overall, this provides for a clean and clear visualisation.

4. Task 3: Show the relationship of the origin of the most active terrorist groups to their distribution of attack

4.1 Design

4.1.1 Analysis

The analysis approach for task 4 involved narrowing down terrorist groups of interest to a threshold of the number of attacks above 2000 which resulted in 11 different terrorist groups. As stated in the GTD documentation, there exist generic terrorist group values such as ‘Separatists’ which were filtered due to their problematic ambiguities. Since the origins of terrorist groups are not included in the GTD, research was conducted to gather accurate origin information.

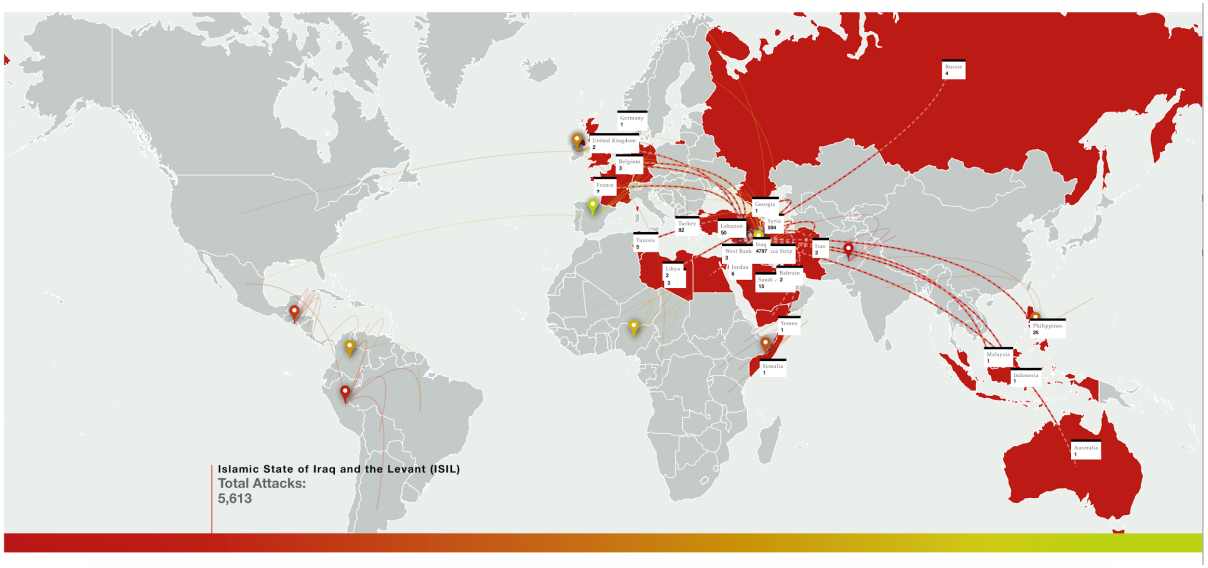
4.1.2 Visualisation

We aimed to visualise the distribution of attacks per terrorist group geographically using a world map. Source nodes (the pinpoints) are situated at the country of origin of a given terrorist group and edges with dynamic motion reflect the direction of attacks from the source node to a destination (the target regions of attack). The map is interactive, such that when a source node is clicked, countries affected by a certain terrorist group are highlighted to clearly show the regional distribution of attacks.

4.2 Implementation

For Task 4, we used D3.js as it enabled us to overlay node and edge objects across a geographical map. It also provided us with the flexibility to incorporate interaction, namely selection of terrorist groups to highlight affected countries as well as animation of edges to display directions. Firstly, the raw data is preprocessed using Microsoft Excel to count the number of attacks by each terrorist group and apply our predefined threshold. Following this, we retain attack target information for each attack and categorise them by country. Next, we conduct research on the origin of each terrorist

group and create a directed edge from the origin to each target country. The data is then converted to JSON file and imported to d3.js. The colour of the origins and the corresponding countries of attack varies from red to green, indicating the decrease in the number of attacks and is presented in the bar legend below the graph. The source code is adopted from WIRED (2017).



4.3 Evaluation

4.3.1 Results

Presentation of Task 4 using our interactive map uncovered a few interesting insights:

- There exists homogeneous intra-regional activity for some groups: FMLN, FARC and SL in Central and South America, Boko Haram and Al-Shabaab in Africa.
- Other groups show dominating intra-regional activity with some inter-regional edge cases: ETA (dominant in Europe, with few attacks in Mexico), IRA (dominant in Europe with few attacks in the US and Zaire), NPA (dominant in the Philippines, with few attacks in India and Pakistan) and PKK (dominant in Turkey and some parts in Europe with few attacks in Iraq, Syria and Iran).
- Taliban, having the highest number of total attacks (7,478) of the 11 groups, showed a relatively small geographical spread as these attacks concentrated in Afghanistan (the origin), as well as Turkmenistan and Pakistan, which border Afghanistan.
- ISIL exhibited the most distribution of attacks across the globe from their source origin in the Middle East. Their attacks have affected Australia, parts of South East Asia, Europe and North Africa.

Nielsen's Heuristics was used to evaluate the visualisation. Observations of strengths and weaknesses are shown below:

Strengths

Description of Strengths	Heuristics Related
Very clear to see the origin and the destination	Match between system and real world.
Interactive and hovering will give more information	User control and freedom. Flexibility and efficiency of use.
The colour scale and legend help users to understand the different number of attacks	Visibility of system status.

Weaknesses / Problems

Description of Weakness	Heuristics Related	Severity Rating				
		Josh	Ignacio	Michael	Tim	Chau
The font size is too small	Aesthetic and minimalist design. Visibility of system status.	2	2	3	2	2
Without clicking, the edges are difficult to interpret	Flexibility and efficiency of use.	2	2	1	2	1
The size of the highlighted country is disproportionate to the number of attacks	Match between system and real world.	0	0	0	0	0

4.3.2 Discussion

Most weaknesses have been evaluated as largely minor on the usability of the graph overall. The small font size was regarded as the attribute that may negatively affect usability the most. The use of small font size was considered a trade-off to minimise visual obstructions of country labels hiding country visibility (especially in regions such as Europe where countries have a relatively small size) and/or the overlapping of the labels themselves which would decrease readability. The natural next step would be to explore the implementation of a zooming function to improve visibility and allow the user the freedom to centre on selected locations. This would allow us to incorporate and display more textual information by country, such as dates of attacks and number of fatalities.

5. Task 4: Analyse the pattern of terrorist attacks on each region based on dates, periods and link to public holidays

5.1 Design

5.1.1 Analysis

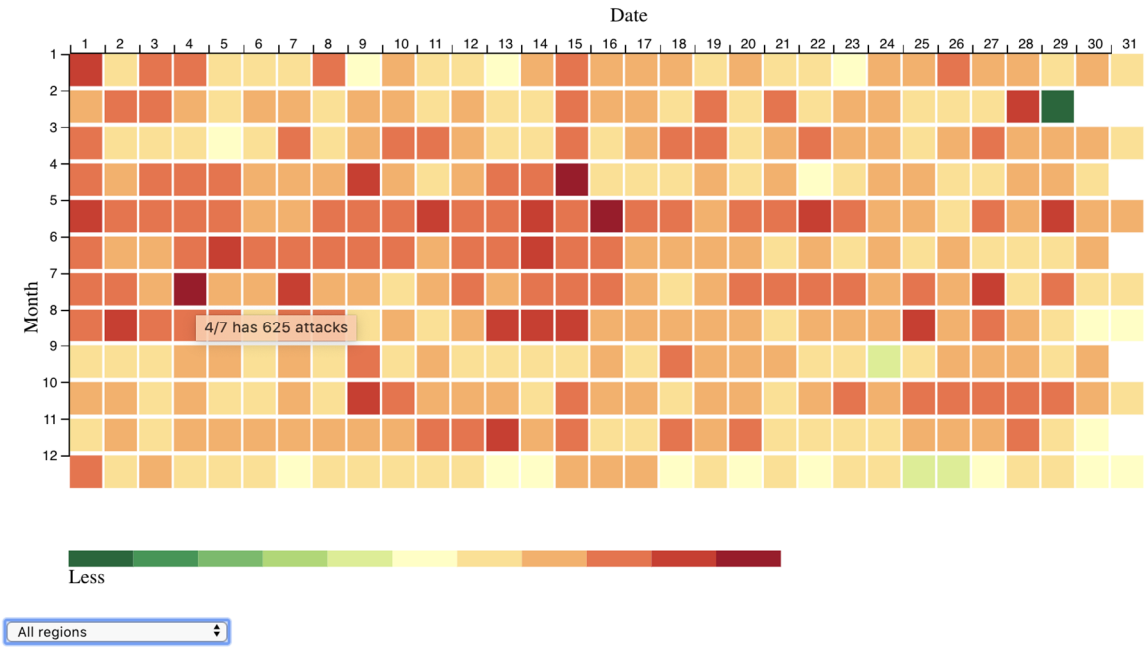
The analysis approach for Task 4 is to identify any patterns in the number of attacks in terms of days and months. We suspect that terrorists tend to attack during public holidays such as New Year's Eve, due to the large number of people gathering at public places, so more victims would be impacted. We also produce the same graph for each region because countries in the same region may share the same religions and public holidays. Using heatmap as described below, we could identify which days, months or periods of the year having the most or least number of attacks and analyze the relationship between the outliers (if any) to holiday seasons.

5.1.2 Visualisation

For Task 4, we use a heatmap calendar view to present the data. The value of each day of the year is quantized into a diverging colour scale, ranging from green to red indicating the increase in values. The vertical axis presents the months (1 - 12) and the horizontal axis presents the days (1 - 31). We apply the same concept to produce a graph for each region. Users are able to select the region and have more information about the number of attacks by hovering to any day on the graph.

5.2 Implementation

There are several tools to implement the Calendar Heatmap for Task 4 such as D3.js, Tableau, Excel PivotTable, etc. In this assignment, we use D3.js as it provides more interactive functions. Firstly, the raw data is preprocessed using Microsoft Excel to count the number of attacks and the number of victims for each day of the year from 1970 to 2017 and for each region. Next, it is converted to a CSV file and loaded to D3.js to visualise the values.



5.3 Evaluation

5.3.1 Results

From the *All regions* graph we can notice two significant dates, which are 1 January (New Year Day) having a very high number of attacks and 25, 26 December (Christmas Day) having a very low number of attacks. Furthermore, terrorists tend to attack in the middle of the year, from April to August.

Initially, we also notice 4 July having a very high number of attacks and suspect it is due to the USA's Independence Day. However, after re-checking with *North America* region graph, the number of attacks in North America is only 21 compared to 625 overall. Thus, we reject our hypothesis in this case. Therefore, for every suspected day, further investigation must be done in order to confirm/reject the hypothesis.

Nielsen's Heuristics was used to evaluate the visualisation. Observations of strengths and weaknesses are shown below:

Strengths

Description of Strengths	Heuristics Related
Colour scales help to identify the patterns in the calendar	Aesthetic and minimalist design. Match between system and real world.
Interactive and hovering show more information	User control and freedom. Flexibility and efficiency of use.
Users can select the region	User control and freedom. Flexibility and efficiency of use.

Weaknesses / Problems

Description of Weakness	Heuristics Related	Severity Rating				
		Josh	Ignacio	Michael	Tim	Ereina
There is no graph at first glance. Users have to select a random region to show the information	Flexibility and efficiency of use. User control and freedom.	2	1	2	2	2
The legend should have more information such as the range of the number of attacks for each colour	Visibility of system status.	1	1	1	1	1

5.3.2 Discussion

The graph can be improved by showing more information like common public holidays and religious dates, as well as when we hover to a specific day, the years having the highest number of attacks and locations can be shown on the screen. One major problem we found out is that most religious dates do not follow Gregorian calendar such as Ramadan fasting period. Going forward, we can produce the same maps for each year from 1970 to 2017, so that we can better see the trend or pattern changes, as well as mitigate the difficulty in identifying and keeping track of the public holidays which do not follow the Gregorian calendar.

6. Conclusion

This project allowed us to explore different aspects of the terrorist activities in the GTD, focusing on geographical trends, groups, number of fatalities, and incident dates. We presented four tasks related to this analysis and implemented various tools to visualise the results. Our visualisations focused mainly on the historically most important groups. We developed several visualisations to explore and comprehend these aspects and find additional interesting patterns like the distribution of attacks through the calendar year.

Based on the implementation, we found interesting patterns related to the cluster and trends of group attacks on the geographical map over time. Using a Sankey diagram, we discovered interesting insights from 20 most deadly terrorist attacks. Furthermore, another interactive map showed us the spread and mobility of the groups to other countries and regions. Finally, the calendar heat map presents the patterns on the number of attacks on some holidays by region.

We performed the evaluation of the four tasks using Heuristic Evaluation based on criteria and severity rating set by Nielsen (1995a, 1995b). To summarise, most of our visualisations are interactive and hence become the strengths of our approaches. We found out that most problems or weaknesses in our visualisation are trivial. However, there are some issues that might need to be addressed for future works.

7. References

Global Terrorism Database. (2018). Retrieved from https://www.start.umd.edu/gtd/

Nielsen, J. (1995a). 10 Heuristics for User Interface Design. Retrieved from https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J. (1995b). Severity Ratings for Usability Problems. Retrieved from https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/

WIRED. (2017). How Facebook Helps to Reveal the Fate of Missing Syrian Refugees. Retrieved from https://www.wired.com/2017/04/locating-missing-refugees-social-media/

Zarate, D., Bodic, P., Dwyer, T., Gange, G., & Stuckey, P. (2018). Optimal Sankey Diagrams Via Integer Programming. *2018 IEEE Pacific Visualization Symposium (Pacificvis)*, 135-139. doi: 10.1109/pacificvis.2018.00025.

Appendix 1 - Group Meeting Minutes

Meeting Week 6

06 SEPTEMBER 2018 / 05:30 PM / NEW LAW BUILDING

Attendees

Ereina Gomez, Tim Nallaiah, Bao Chau Ngo, Michael Kam, Ignacio Colino

Agenda

1. Introduce team members
2. Select the data set for the group project

Discussion

- Discuss the potential analysis and visualisation of each data set
- Discuss the positives and negatives of each problem
- Decided on Global Terrorism Data (GTD) from Kaggle (Problem No. 8)

Action Items

1. Analyse and explore the dataset for discussion on tasks. Assigned to: everyone

Meeting Week 7

13 SEPTEMBER 2018 / 08:00 PM / NEW LAW BUILDING

Attendees

Ereina Gomez, Tim Nallaiah, Bao Chau Ngo, Michael Kam, Ignacio Colino, Josh Pierre

Agenda

1. Allocate work for the initial report
2. Discuss the findings from data exploration
3. Discuss potential tasks

Discussion

- Relevant tasks for the GTD:
    1. Show the terrorist activities trend over time in regions or countries.
    2. Measure the impact of the terrorist group on different attributes by the countries.
    3. Show the terrorist group change trend over time and the interaction between the groups.
    4. Analyse of the attacks based on the month/day/public holiday on that country.
- Possible visualisation approaches: parallel coordinates for showing the relationship of different attributes, animation of the terrorist activities on the map over time, stream graph, temporal map, and the terrorist group similarity network graph drawing.
- Possible implementation tools to be used: Tableau, NetworkX, D3, yEd, Python, etc.

Action Items

1. Create Introduction and Planning part for the Initial Report. Assigned to: Josh. Deadline: 20 September at 12 am.
2. Create Task 1 (Design and Approach, Implementation, Evaluation). Assigned to: Ignacio. Deadline: 20 September at 12 am.
3. Create Task 2 (Design and Approach, Implementation, Evaluation). Assigned to: Michael. Deadline: 20 September at 12 am.
4. Create Task 3 (Design and Approach, Implementation, Evaluation). Assigned to: Tim. Deadline: 20 September at 12 am.
5. Create Task 4 (Design and Approach, Implementation, Evaluation). Assigned to: Chau and Ereina. Deadline: 20 September at 12 am.

Meeting Week 8

21 SEPTEMBER 2018 / 09:30 PM / CONFERENCE CALL

Attendees

Ereina Gomez, Tim Nallaiah, Bao Chau Ngo, Michael Kam, Ignacio Colino, Josh Pierre

Agenda

1. Discuss the findings from dataset and visualisation
2. Discuss the next steps

Discussion

- Need to explore the data so we can validate the tasks.
- Need to find a connection between the tasks so it creates a theme or story.
- Some possible attributes of the dataset that can be used in the analysis: number of killed, number of wounded, group name, target type, weapon type, attack type

Action Items

1. Analyse and start creating visualisation form the data. Assigned to: everyone. Update deadline: 26 September.

Meeting on Mid Semester Break

26 SEPTEMBER 2018 / 06:00 PM / Abercrombie Learning Hub

Attendees

Ereina Gomez, Tim Nallaiah, Bao Chau Ngo, Michael Kam, Ignacio Colino, Josh Pierre

Agenda

1. Discuss the findings from dataset and visualisation
2. Discuss the next steps

Discussion

- Need to explore the data so we can validate the tasks.
- Need to find a connection between the tasks so it creates a theme or story.

- Some possible attributes of the dataset that can be used in the analysis: number of killed, number of wounded, group name, target type, weapon type, attack type

Action Items

1. Analyse and start creating visualisation form the data. Assigned to everyone. Update deadline: 26 September.

Meeting Week 9

4 OCTOBER 2018 / 08:10 PM / New Law Building

Attendees

Ereina Gomez, Bao Chau Ngo, Michael Kam, Ignacio Colino, Josh Pierre

Agenda

1. Update on the progress for analysis and visualisation
2. Discuss the findings related to the tasks

Discussion

- Chau showed her visualisation for in the form of heatmap on the calendar date and month to display the pattern of the attacks. She and Ereina also worked on an interactive map using d3.js for showing the origin of the groups and the spread to other countries. Need to add another task for this.
- Ignacio gave an update on the animation is still on progress for showing the trend of terrorist activities on the map using Python.
- Josh showed bar charts for the number of attacks and groups according to the number of attacks using Power BI.
- Tim had created a Sankey and chord diagram for the group names using R.
- Michael explained how he was experimenting with Parallel Coordinates and radar charts. However, the task and visualisation do not give new insight. Removal of the task.

Action Items

1. Create the presentation slides. Assigned to everyone. Deadline: 10 October.
2. Continue work on the visualisation up to the point it can be included in the presentation. Assigned to everyone. Deadline: 10 October.

Meeting Week 11

18 OCTOBER 2018 / 05:15 PM / Carslaw Learning Hub

Attendees

Ereina Gomez, Bao Chau Ngo, Michael Kam, Ignacio Colino, Josh Pierre, Tim Nallaiah

Agenda

1. Presentation rehearsals
2. Final report discussion
3. Evaluation

Discussion

- Rehearsed the presentation for a few times and made some adjustment from the team's suggestions.
- Need to finalise the visualisation by next week.
- Explored the task evaluation methods for next week.

Action Items

1. Finalising the visualisation and start writing the report. Assigned to: everyone. Update deadline: 25 October.

Meeting Week 12

25 OCTOBER 2018 / 06:00 PM / Carslaw Learning Hub

Attendees

Ereina Gomez, Bao Chau Ngo, Michael Kam, Ignacio Colino, Tim Nallaiah

Agenda

1. Discuss the final report allocation and structure
2. Discuss the tasks evaluation

Discussion

- Michael showed the interactive map for Task 1 as another approach than the animation.
- Discussed the different evaluation methods from the lecture. Decided on the Heuristic Evaluation.
- Performed Heuristic Evaluation on each visualisation based on strengths and weaknesses.
- Agreed on the format of the final report, dividing by tasks

Action Items

1. Write the Introduction part. Assigned to: Josh. Deadline: 1 November.
2. Write the analysis, visualisation, implementation, and evaluation for Task 1. Assigned to: Ignacio and Michael. Deadline: 1 November.
3. Write the analysis, visualisation, implementation, and evaluation for Task 2. Assigned to: Tim. Deadline: 1 November.
4. Write the analysis, visualisation, implementation, and evaluation for Task 3 and 4. Assigned to: Chau and Ereina. Deadline: 1 November.
5. Compile the minutes of meetings. Assigned to: Michael. Deadline: 1 November.

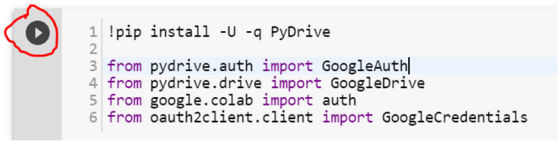
Appendix 2 - Instructions to View the Visualisations

Task 1

Task 1 was developed using the Google Colaboratory environment which allows to collaboratively and simultaneously develop python code. There is no need to download or install specific packages locally. However, the Python notebook, the animation video, legends for the animated map, and the interactive HTML file are also attached in the folder for "Task 1" in this assignment to be reviewed.

To get both the animation map and interactive map follow the next steps:

1. Open this link
2. Click on each of the chunks of code and click on the play button (following the sequence of chunks)



3. A pop-up to download the visualizations will appear
4. Download the visualizations
5. Open the animated map with a video player
6. Open the interactive map (HTML file) with a web browser (e.g.: Chrome)
    a. You can zoom in by scrolling or with the + and - buttons
    b. The decades can be filtered in and out with the checkboxes on the top right menu
    c. Information for each node can be explored by clicking on the different nodes

Task 2

Task 2 is implemented using R programming to produce the interactive HTML file. The code in the form of R Markdown file is included in the "Task 2" folder. Furthermore, the JSON file named "top20_v4.json" is there to support the visualisation. To view the

visualisation, you can open the “Task 2 Sankey_4axis.html” file using a web browser. There is no need to define the python web server.

Task 3 and Task 4

To open the HTML file for Task 3 and Task 4, please follow the instruction:

1. Open a Command Prompt or Terminal
2. Go to the directory where you have the “Task 3” or “Task 4” folder.
3. Run **python -V** and note the first digit of the version number returned.
4. Depending on the version of Python you have on your machine, you can:
    a. For version 2.xx: Run **python -m SimpleHTTPServer 8888**
    b. For version 3.xx: Run **python -m http.server 8888**
5. You can then open a web browser and enter the address in **localhost:8888**
6. Navigate and open the file “Task 3 terrorist_groups.html” to view the visualisation for Task 3 or “Task 4 calendar_heatmap(region).html” to view the visualisation for Task 4.
7. For Task 4’s heat map, you need to change the region first to see the visualisation.