18/11/2018

Interactive Media Design

(CA1)

Mc Loughlin, Sean

x00130298

Contents

[CA Summary 2](#_Toc530321400)

[Background 3](#_Toc530321401)

[Dataset 4](#_Toc530321402)

[Seven Stages 5](#_Toc530321403)

[Acquire: 5](#_Toc530321404)

[Parse: 5](#_Toc530321405)

[Filter 7](#_Toc530321406)

[Mine 8](#_Toc530321407)

[Represent 8](#_Toc530321408)

[Refine 9](#_Toc530321409)

[Interact 10](#_Toc530321410)

[Problems and Solutions 11](#_Toc530321411)

[Conclusions 12](#_Toc530321412)

[Discoveries 12](#_Toc530321413)

[Further Analysis 12](#_Toc530321414)

# CA Summary

My overall goal for this CA is to create an interactive visualisation dashboard of a dataset which is user-friendly, easy to understand and use.

The dataset I’ve chosen to create this visualisation dashboard is based on terrorist attacks around the world. I’m focusing on the years 2010 to 2018. What I’m trying to achieve from studying this dataset is to understand it and being able to put some context on some certain trends that I may come across when doing the dashboard.

My other main objectives are to try and illustrate these trends and noticeable figures to the users so that they can fully get the context and information of this dataset purely by looking at the interactive visualisation dashboard that I provided to them.

# Background

The topic that I picked to do my dataset on is terrorist attacks worldwide. This will be focusing on terrorist attacks between 2010 and 2018. The reason I picked this data set is, as people, it is very often that we see the big terrorist attacks around the world on the news. Although this is terrible news, it seems that only the ‘big’ terrorist attacks are highlighted, it seems that the attacks that do be highlighted on the news or any media are mostly ‘western’ countries. One reason I chose to pick this dataset is to inform the user that the number of terrorist attacks that happen in the ‘western’ countries are miniscule compared to other regions of the world with barely any coverage from the media on it.

A reason why I choose to pick this dataset also is to learn about the severity of the number of attacks, number of people killed and wounded in other countries where I may not know about. This way I can learn about about the severity and frequency of attacks in different countries and then pass this information on to the user.

Before examining the dataset and getting my figures and visualizations together, one trend I’m expecting to be there is the increase of attacks in the Middle-east in about 2013-14 onwards. The reason I’m expecting there to be a huge ongoing increase in this period is due to the Islamic State (ISIS) involvement in terrorist attacks. One thing I will be looking for is the number of incidents that happened in the years before 2013-14 and during and after it.

# Dataset

The data set I’m using is called **Global Terrorism Database** and I got this dataset off **Kaggle**. The dataset was released in July 2018 and it covers more than 180,000 terrorist attacks worldwide from 1970-2018.

The following link is the website where I downloaded this dataset from:

<https://www.kaggle.com/START-UMD/gtd/version/3>

# Seven Stages

## Acquire:

I acquired my dataset from Kaggle. The name of the dataset is called Global Terrorism Database and it is up to date as of July 2018. The data in the dataset stretches back to 1970.

## Parse:

When I first opened the dataset, I saw that it needed a lot of cleaning up to do. I saw the names of the columns and the different categories, everything was very messy. One of the first thing I had to change with the data was to delete the unnecessary columns which I knew I wasn’t going to use. There were 135 columns originally and I narrowed it down to approximately 20 columns. I could afford to delete the majority of the columns in the dataset as there was unneeded data in them.

Once I narrowed all the columns down to what I thought was worth keeping, I decided to rename the columns to something that’s easier to read. The original names of the columns were named badly so I fixed this problem and gave these columns names that were understandable.

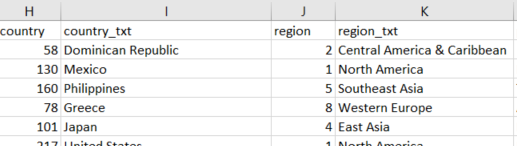


Figure 1: (Original Code) Example of bad column naming

The next thing I needed to do was to get a date format setup on the dataset. The original code that I had to work with had 3 separate columns, one for year, one for month and one for day. I wanted to put this into one column but couldn’t do it for the exact day. So, for the dates, I put them into quarterly dates (e.g. 1/1/2018, 1/4/2018, 1/7/2018 and 1/10/2018).

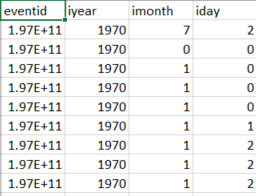
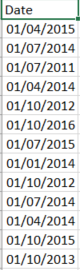
 

Figure 2: Date format on original code Figure 3: Updated date format

Another problem that arose from the date issue was that there were too many rows to manually sort out. Therefore, the rows from the years 1970 to 2009 had to be deleted due to the date not being correct.

The next thing I had an issue with was the blank spaces in the code. There was a lot of instances where there would be a city blank and therefore the country would be blank, however, the country region code would be there. Therefore, I needed to take a note down of all of the country codes and find the blank spaces where a country would be and then I would need to fill in the name of the country. Since the city column was blank I changed this to unknown.



Figure 4: Example of how the blank spaces for countries/cities looked

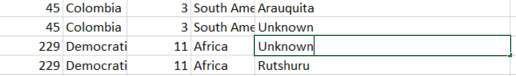


Figure 5: How I updated the blank countries/cities

## Filter

In total, there were more than 100 columns deleted from the dataset. I manually added one column to the code which was the ‘Date’ column. The majority of the rows that were deleted from the original code were useless. They mostly contained columns that were 95% blank due to some rows being more severe attacks than others. Some other columns were deleted due to a lack of context, some columns didn’t make sense and there was no true data sense linked with the rows.

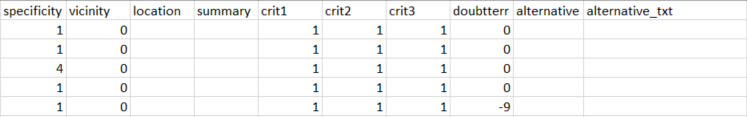


Figure 6: Example of some columns deleted due to no context

The following photo shows the lack of data in the rows where there wasn’t a huge attack recorded. There are barely any rows which contain the an ‘attacktype3’ information. These rows were deleted due to blank spaces and I felt it was unnecessary to keep 3 different types. I kept ‘attack type 1’ as this was the main form of attack and I can find it useful for my visualization.

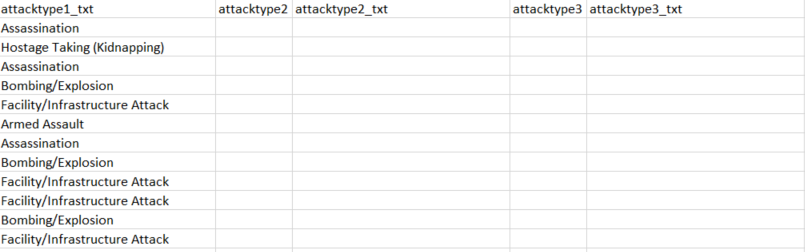


Figure 7: Unnecessary columns throughout the original dataset

## Mine

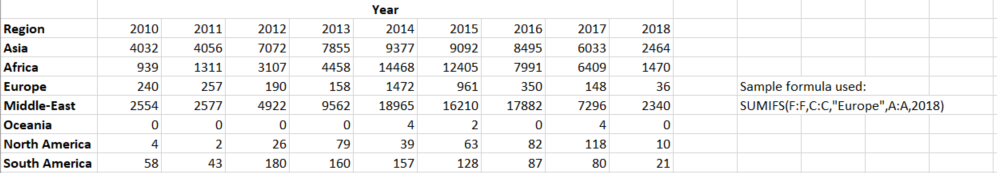
The following photo is a representation of the amount of people killed by terrorist attacks from 2010 to 2018. This is split up into each region. 

Figure 8: Mining the data

## Represent

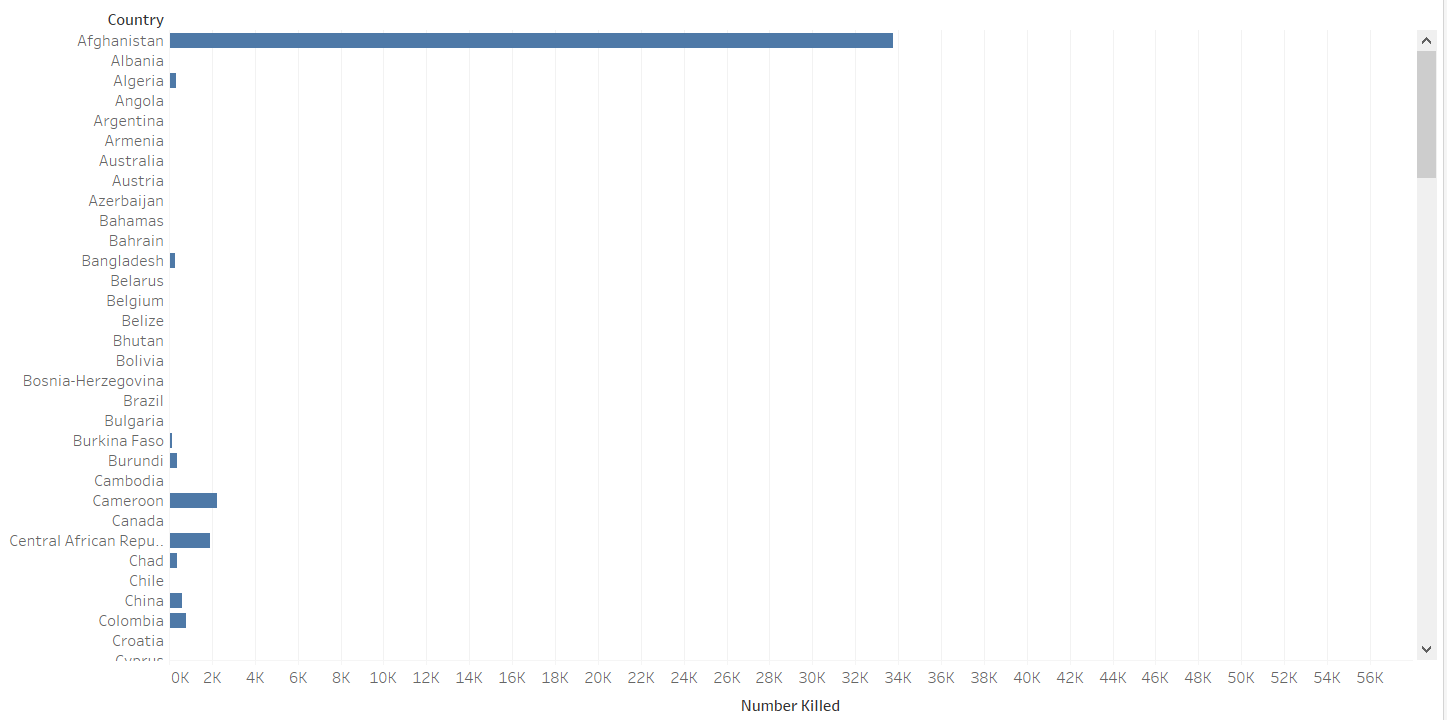
I wanted to do a breakdown of the number of attacks, people wounded, and people killed in each country. This can be represented in a bar graph, but the user would need to do a lot of scrolling in order to get information on all countries they may want to know about. Below shows an example of a bar graph of this representation. It is clear that this isn’t the right type of bar graph for this type of information due to too many countries and due to the huge numbers of Afghanistan which is approximately 34,000. The huge numbers of Afghanistan make it seem that the majority of the other countries are at 0. This data and graph will need to be refined to make this information clearer for the user.

Figure 9: Bar graph of all countries and people killed in them

## Refine

From the previous visualization, it was clear that the bar graph needed to be represented in a different type of way. The way I made this representation clearer was by making the data and countries into an interactive heatmap of the world. This heatmap of the world made it clear to the user as it shows every country in the dataset on one dashboard. The users are easily able to identify which country has the most deaths as the heatmap would indicate which is the country that has the most deaths. I also added a visualization on the sidebar of the dashboard for the user that tells them which countries are in the top 10 for total deaths. This allows the user to take in this information instantly without looking at the heatmap. For each country, I added a rollover which tells the user how many attacks have been involved with that country, how many people were wounded in the country and how many deaths. To make things easier for the user, I also added a highlight option. This would allow the user to put in the name of a country and this feature would highlight that country on the heatmap. This feature allows the user to get the information instantly on any country of their choice. This is also a useful feature as not every user would know the location of every country.

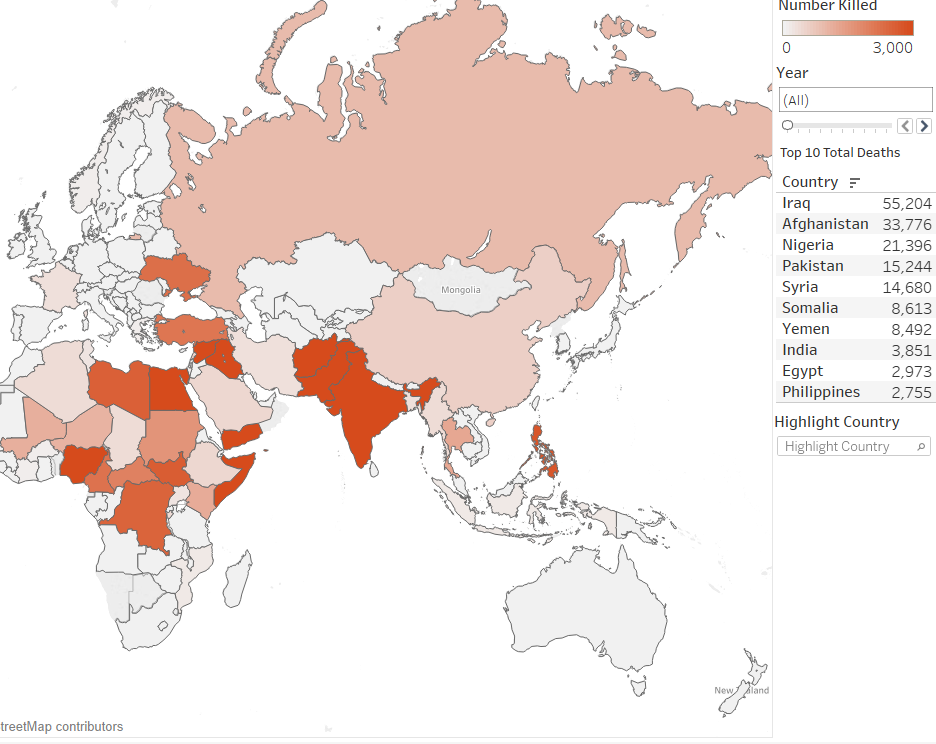


Figure 10: A refined representation of the list of countries

## Interact

The heatmap involves a lot of interaction from the user but one interaction that wasn’t mentioned in the refine phase was the interaction feature with the years. I needed to add an interaction for years because the dataset was from 2010 to 2018 and I wanted to add as much granularity for the user. This feature allows the users to either pick any year of their choice or to pick an accumulative sum over all the years available

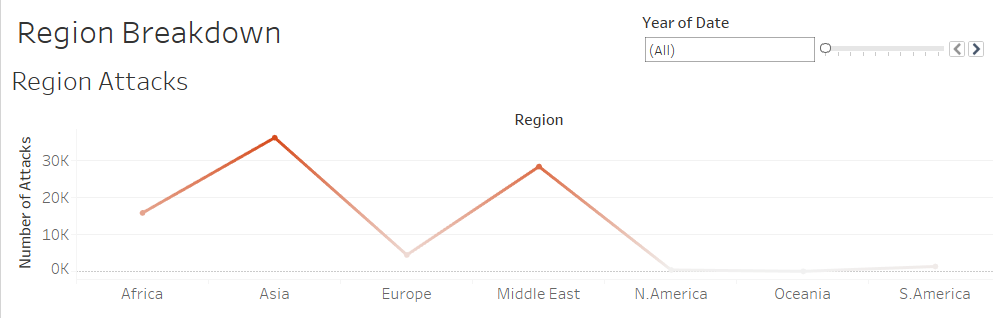


Figure 11: Visualisation with all years selected

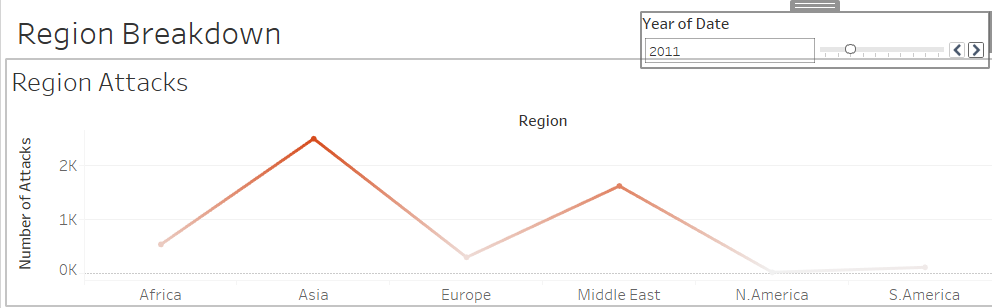


Figure 12: Visualisation with one year picked (2011)

# Problems and Solutions

There were some problems I came across when trying to create visualizations. Most of these problems came from the number of rows/columns were in the data due to the number of countries. This limited the number of visualizations I could create with countries due to this issue. I could solve this issue with number of deaths, wounded and attacks as I could integrate all of them into my heatmap of the world. I came across problems with this when I was trying to do a breakdown of each country in terms of the types of terrorist attacks that occurred in that country and the target of these attacks. I came across problems as there was too much data on the screen that made it borderline impossible for the user to understand.

Another problem that I came across was trying to get the colors on the heatmap to indicate how bad the situation was in that country. This problem happened because of the inflated number of deaths that Iraq has on the other countries. Iraq had 50,000+ deaths and no other country was even close to them. Therefore, when the heatmap was created, the only countries who had any type of color were the top 10 countries because they had a significant amount of deaths. Since all the other countries were grey, this gave the wrong visual impression to the user as it would indicate or make the user assume that there are very little deaths in a country which may have over 1,000 deaths. How I changed this is I altered the maximum number on the scale to have a more accurate visual representation of deaths in the countries with 1,000+ deaths. This addition was very much needed to get the point across to the user.

# Conclusions

## Discoveries

When I first started with this dataset I wanted to get a few years for the dataset to see the differences. Whether the numbers increased through the years or whether they decreased. I feel that 2010 to 2018 was a perfect number of years to choose from due to the recent troubles related to ISIS in 2013/2014. I wanted to see whether there was an increase in the attacks and find out more information about the extent and how much it increased.

In 2011 there were 2,590 bombing throughout the world. The number of bombing attacks in 2014 are almost quadruple the numbers of 2011. This is information that would be impossible to figure out from looking at the raw data.

When looking at deaths I was surprised to see how big the increase was in the middle-east. Comparing with 2010, where there were 2,554 deaths. This figure drastically increased in 2014 with the number of deaths going to 18,965 deaths, nearly 8 times the deaths compared to 2010. One other thing that surprised me was that the middle east didn’t have the biggest increase of deaths. The country with the biggest increase was Africa. In 2010 there were 939 deaths. This figure multiplies more than 14 times to 14,468 in 2014 and this was something I wasn’t expecting.

## Further Analysis

If I was to do some further analysis of this dataset. One of the first things that I would do would be to add some more granularity into the data by inserting the correct month instead of having each row based on the quarter it happened in. One other visualization I would like to do would be to have more visualization for the middle-east. Having the exact co-ordinates for each attack and being able to see the visual representation of the area of each attack on tableau. This would potentially show a trend for the attack target and why they may target that area (e.g. I would expect that there would be attacks on schools, hospitals, important roads out of the country etc.). This would be a lot of information for a user and it would be able to tell the story of that country.