Data Visualisation
DALT7016

Nikolaos Alexoudis
19178012

MsC In Data Analytics

WORD COUNT: 1360

Data Base

The dataset I chose about the assessment in the Data Visualisation module is The Lost Journalists Dataset. This dataset consists of 1879 reported missing cases of journalists around the world from 1992 to 2017. This dataset has as its columns reported details considering:

- The date of the incident.
- The name of the victim.
- Sexual orientation of the victim (If confirmed).
- Country that the incident took place.
- Organisation.
- Nationality.
- Medium of reach.
- Job
- Topic Coverage.
- Indigenousy.
- Type of group that was associated with the incident.
- Type of death.
- Impunity of the propagating member (Yes, No and Partial) .
- Captivity of the victim (True or False) .
- Reported threats against the Victim (True or False) .
- Reported Torture conducted against the victim (True or False) .

Plus additional information that was retrieved from the plotly country codes library, namely each country's three letter code. Every other data we used is a product of the original data. This data is widely available on Kaggle and is free with no copyright violations. There are no ethical violations concerning the display of loss, since each case is not referred to by name. This database is very informative and has a multitude of variables that could be displayed with different visualisations.

Brief Formulation

The purpose of this data visualisation project is to create an interactive map that shows which regimes around the world are or were the least friendly on free speech. The visualisation aims to show both qualitative and aggregate data. This aims to clarify the reasons between the majority of the cases and to be correlated with historical events to shed some light upon those crimes.

This map will be used in a BBC associates meeting. Prevalent topics that will be discussed are the fair pay rate among journalists that cover stories in the most dangerous countries. Those visualisations need to be interactive since there is much information on display through three distinct dimensions (latitude, longitude and time). The user should be able to traverse through the map and get some essential information on the majority of the

cases just by mousing over a country at a specific time. The information then would be correlated with news of that year and then set the tone for future predictions on coverage of similar events.

The project's place in the purpose map will be Exploratory and Feeling. Exploratory because the data will be used by the user to get their own insights in conjunction with other data. The engagement with the material will definitely try to ease comprehension but the emotive repercussions are difficult to ignore. We have to take into account that the crowd this visualisation is presented are journalists themselves.

Working with Data

As soon as we download our data we can see some problems concerning the column and row names. Namely some column names contain spaces, some are quite complicated and long and others do not really make any sense. There are eight columns containing names from x1 to x8 and their variables are mostly NA's. We will simply ignore those. Then we change the names of some columns i.e. 'Impunity (for Murder)' to 'Impunity'.

Next we will change the values of all our binary variables to 0 and 1 in case we need to use them later on. Now the date format is really detailed and would not work with our representation because the time axis would have so many ticks. In order to avoid this we will transform the date column to contain only the years. Some data is being lost because the dates are descriptive i.e. 'Somewhere between May 3rd and May 9th', but that is ok since the loss is not bigger than 100 entries. Now the years go from 1992 to 2008.

The problem we face now is that the country names that will need to be mapped on the plotly object might be different than the ones plotly understands. Luckily there is a library online with the country names and each ISO-3 code associated with them. We now only need to left-join the two tables, but first we must make sure that each country name matches in the two tables. Most names match but some entries have alternative names i.e. 'UK' instead of 'United Kingdom'. We manually change those. Now some other names may be controversial fr example 'Israel and the Occupied Palestinian Territory' has to be changed to just 'Israel'. We only do this for the purposes of this investigation without trying to choose a side in the conflict. It is also evident that there were some border changes in the time span from 1992 to 2008. The Yugoslav war as a result had the creation of multiple states in the Balkans. To avoid the problem of mapping a country that does not exist anymore, we will replace its name with the name of its spiritual successor. So Yugoslavia will be renamed to just Serbia. This convention is taken since other Yugoslav states (that are not Serbia) are mentioned at the same time as Yugoslavia.

For the next part we will take some aggregate data from the already existing dataframe to create our hover object. We will take the total kill count per place and year, the most common reason of death and the most covered topic. Then we will try to map all this information with plotly. The resulting map should look like this:

Journalist Deaths Demo
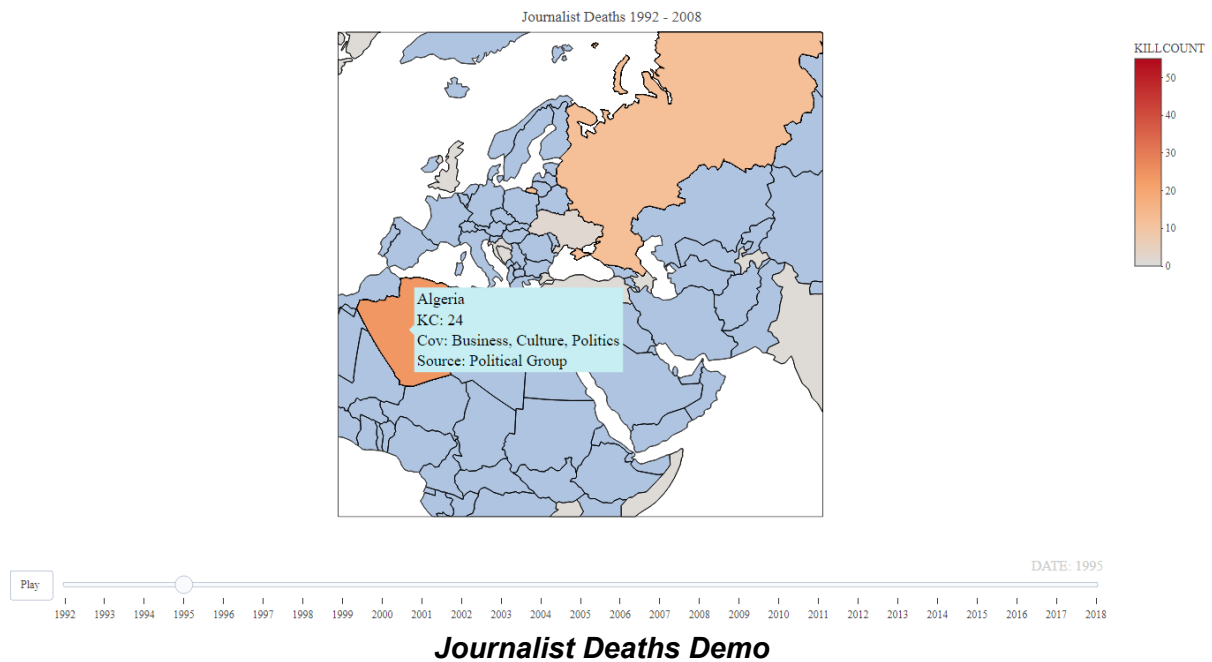
Note that the projection we see is a zoom in window of the globe. We can zoom in and out at will. The colour we used for the land is light blue (#afc4e0). The palette for the kill counts of each country is 'RdOrYl' meaning red, orange and yellow. We chose these colours so they would have the most contrast between them. The hover is in turquoise (c7eef2) not for any particular reason other than aesthetics. The font that is being used is 'DM Sans' a thicker version of 'Gil Sans' that the BBC actually uses. On the hover we see the country name, the kill count for that year, the most common coverage story and source of fire. The time bar is traversable and the play button allows for a mini animation of the evolution of journalist deaths to take place.

Second visualisation

The second visualisation is a concatenation of pie charts and a bar plot of aggregate data. The data that is displayed deals with the most deaths per country, the gender that was attacked the most, the medium that was attacked the most and the ratio of the type of death in each case. In those mini visualisations we used some of the data we manipulated before in binary form and an aggregate table we created. The colour palette here is 'PuOr' (purple and orange) clearly for aesthetics and for it to be colour blind friendly. This chart could be used in a briefing environment where it would be displayed in a board to give a more holistic insight of the findings of the first visualisation, along with the display of further information.
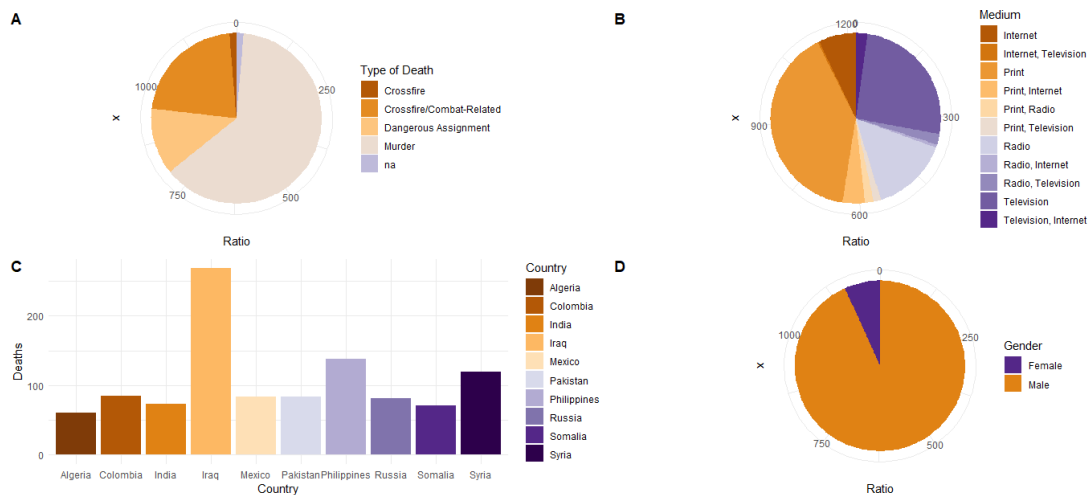
Aggregate Bar Charts and Bar Plots

Conclusion & Reflections

We can gather a lot of information from the two above visualisations. For example the second visualisation tells us that the most hostile country for journalists in the world is Iraq by a far margin this could be due to the Iraq war. The most attacked gender was males, possibly because they tend to be assigned to more hostile coverage missions. The most common way to go for a journalist is planned murder implying intention in the carrying out of the killing. Now the mediums in the medium pie chart might be difficult to recognise individually but that is exactly the point. Most journalists go between many different mediums so the lines actually get blurry in real life as well. We only need to notice that the spectrum goes like that: Dark Orange- Internet, Light Orange- Print, Light Purple- Radio and Deep Purple- Television. First visualisation can give a rough estimation of the hotpoints a journalist should avoid and the type of reasons associated with this decision.

The sources of the dataset are:
- BBC: *The lost journalists of 2016*
-NRT: *AT LEAST 122 MEDIA PROFESSIONALS KILLED GLOBALLY IN 2016*
-All Africa: *Africa: Journalist Killings Ease From Record Highs As Murders Down, Combat Deaths Up*
-Plotly Library: *Plotly Countries and Codes*