Student ID: 180008901

Interactive Visualisations Using Tableau Desktop

Introduction

This assignment is to design a visualisation that can be used to explore given data and be used to answers one open-end question.

The question that I selected is "**How diversely were journalist crisis distributed in reference to their area of coverage?**".

In order to visualise data, I started with pre-processing data. Because coverage attribute came in the form of multiple values separated by a comma so I implemented python script (preprocessing_v2.py) to separate it into multiple records with only one coverage per row. But, it will lead to duplicated data. To fix that, I added the `Id` attribute as a running index to keep the uniqueness of the record. However, after playing with the data, I found that some records are duplicated when it represents missing journalists so to get the accurate number of unique journalists, I, later, used `Full name` as a key instead.

Lastly, I filter out data from 2019 because I think that it does not reflect the whole year summary so it might distort the trend.

Description of the visualisation

Five attributes being used in the visualisation as shown in the table below.

Attribute	Attribute Type	Visual Variable	Interactive element
Coverage	Categorical	Shape	Click to filter
Country	Categorical	Position	Click to filter/Hyperlink
Status	Categorical	Colour Hue	Select
Number of records	Quantitative	Length/Colour Saturation	
Year	Quantitative	Position	

For categorical attributes, there are not many options for a categorical attribute. Because coverage will be used as the key variable so I decided to use it as the main navigation icon. I tried to use colour but due to its limited separability and discriminability -- it will be discussed in my sketches, I, later, changed to encode it with a big icon instead. Even shape is not the most effective identity channel but I made it sufficiently big to be popout. The country attribute is used as a map which is the most effective channel for the spatial attribute. Status is represented with colour hue because it has only 3 categories which

could be effectively encoded in term of discriminability. The colour palette was selected carefully to be related to death conveying mood and tone despite that it might be less discriminability.

For quantitative attributes, the number of records was encoded in two ways. It was encoded as colour saturation in the map to provide a brief overview which could be more effective to highlight serendipitous insight. It encourages suspicion which, next, I used the top five bar chart on the right side to reveal more insights. The number of records, here, is represented by length to provide accurate information. It is more effective for comparison due to its strong psychophysical power. In the bottom, the year attribute is utilised in another bar chart showing changes and trend line. I placed it in the bottom guiding the viewer to understand the changes as an overall summary.

Figure 1

To support data exploration, I added several interactive elements including coverage filter, country filter and a status selection box. The coverage filter is used as the big navigation bar to control the big picture of the visualisation. The top 5 countries chart, the map and trend chart will be changed according to those filters. The colour also changes indicating what journalist status that you are looking for as shown in figure 2. In addition, the viewer could traverse more to external resources by clicking a link in tooltip.

Figure 2 the visualisation result after selecting coverage=war and status=killed.

Lastly, the icons I used made by Freepik from www.flaticon.com (1). This visualisation was inspired by Sleeper R. 's work (2).

Insights from the visualisation

This visualisation leverages both spatial and chronological attribute to answer the question in many aspects. The design principle here is to continuously rise a new question throughout the visualisation.

Starting with how the crisis distributed across the working fields of those journalists. It can be seen by ordered topics from left to right. Top three of the problems were caused by politics, war and human right issue -- shown in figure 3.

Next to how it distributed. It apparently shows that this crisis occurs the most in the middle east countries -- Iraq, Turkey and Syria following with dictatorship countries -- the Philippines and Somalia.

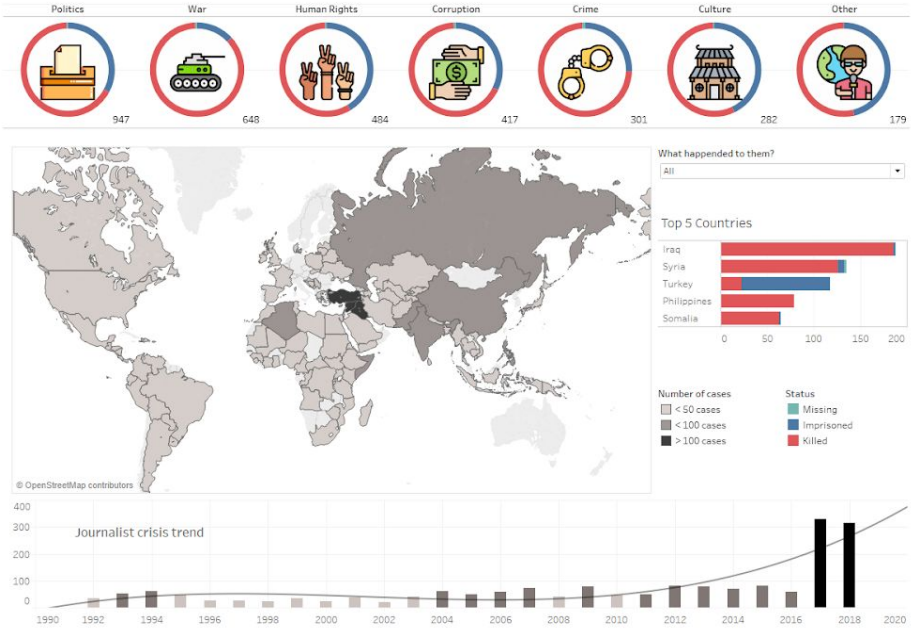

Figure 3

It leads to the next question; How sensitive are these topics? By combining those two filter, it shows that Politics is the most sensitive topic across the globe comparing with others as shown in figure 4. Different country has different painful topics.

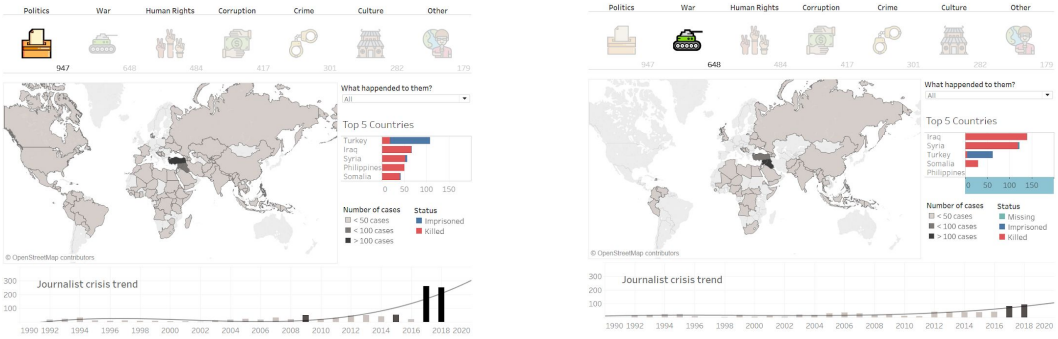

Figure 4

I surprisingly found that different countries treated journalist differently; Turkey and China tend to imprison the journalists but Iraq, Syria and the Philippines tend to kill them as shown in *Figure 5*.

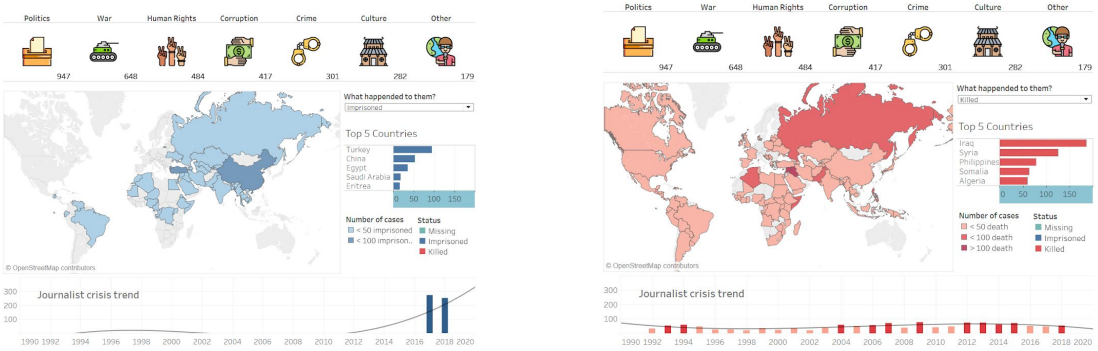

Figure 5

Lastly, the bottom, it summarises with the final question; how has it changed throughout history and what might happen in the future? It shows an upward trend in almost every context, which in 2017-2018, the cases increased by 470%. But only the killing surprisingly does not have this upward trend as shown in figure 5.

These are a few examples that I got from the visualisation but it could have much more than that.

Critical Discussion

This work has been designed to provide information and raise the new question. The visual variables and channels are selected without violating expressiveness and effectiveness rules. However, it is hard for precisely comparison between two categories due to data overwhelming -- too many information, does not highlight on a particular point, and change blindness -- people cannot capture all detail. Moreover, it also fails to consider the hierarchy attribute. The location could be allowed to drill down in a particular region. The timeline could allow to zoom investigating more in a certain period.

As far as the limitation is concerned, there are many questions that could be explored more for example what is the reasons behind the journalist deaths?,and how did they treat them?

Extension

I investigated more by examining search trends from Google Trends (3) and news headline dataset from Rohk's dataset in Kaggle (4). The dashboard was saved as 'P2-Extension.twbx'. I implemented as a proof of concept so the data has not been clean and it might need more work to do with context analysis.

As shown in figure 6, it reveals how people react to this crisis both from public reaction and though news media.

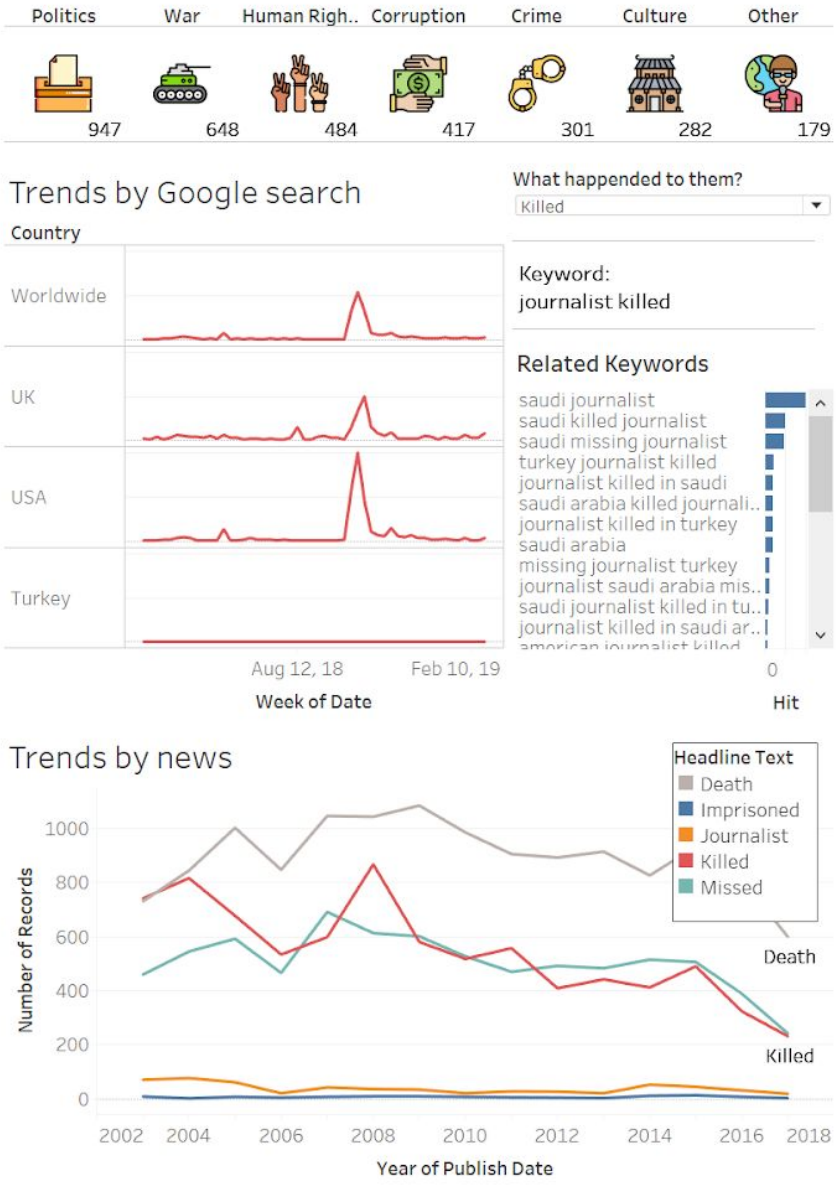

Figure 6

References

1. Free vector icons designed by Freepik [Internet]. Flaticon. [cited 2019 Feb 27]. Available from: https://www.flaticon.com/authors/freepik

2. Sleeper R. The 10 Highest Grossing Actors of All Time [Internet]. Tableau Public. 2019 [cited 2019 Feb 27]. Available from: https://public.tableau.com/en-us/s/gallery/10-highest-grossing-actors-all-time

3. Google Trends [Internet]. Google Trends. [cited 2019 Feb 28]. Available from: https://trends.google.com/trends

4. All the news [Internet]. [cited 2019 Feb 28]. Available from: https://www.kaggle.com/snapcrack/all-the-news