# Journalist Deaths Across the World

**Karen Sanchez Trejo**
Indiana University Bloomington
School of Informatics and Computing
karsanc@iu.edu
(residential section student)

**Arnav Arnav**
Indiana University Bloomington
School of Informatics and Computing
aarnav@iu.edu
(online section student)

### Abstract

Journalism is one of the most important occupation in a free world. Hard working journalists spend enormous amounts of their time and energy to dig into the situations and provide us with information that most people would miss otherwise. Good journalism keeps the organizations in power in check and empower the citizens in a democracy with knowledge about the world. The job however comes with its risks. Over the years journalists have had to face various perilous situations and many of them have been fatal. In this project, we aim to visualize the information about the journalists who gave their life for the job and attempt to understand any reasons that gravity of the situation and aim find out which topics and geographical regions have proven to be more fatal than others.

**Keywords :** Journalist deaths, Time series, Data maps, Network graphs

## Introduction

There have been numerous attacks against journalists. One of the most noticeable in the recent past has been the attack on the office of the French satirical newspaper Charlie Hebdo in Paris in 2015. In the tragic event two armed men forced their way into the office and went on to shoot journalists in the head (Wikipedia contributors 2018a). The reporters in the office have been vocal in asking for protection ever since (Marlowe 2018). The newspaper employs a private security firm for their offices but reporters believe that spending huge amounts on security of press in a democratic country should not be necessary in the first place.

There are organizations working to protect journalists such as the Reporters Without Borders, and the Council of Europe. The Reporters without Borders (RSF) is a non profit organization based in Paris, advocating freedom of press and information. The RSF has been actively, since its foundation in 1981 and to protect journalists from attacks, reduce censorship of the press and aid persecuted journalists and provides for the safety of war correspondents (Wikipedia contributors 2018e). The RSF has been campaigning for the creation of a United Nations body under the Secretary General for the protection of Journalists (Reporters Without Borders 2018b).

Another organization is the Committee to Protect Journalists (CPJ), an American non-profit organization that was founded in 1981 in response to the harassment of the Paraguayan journalist Alcibiades Gonzlez Delvalle. The committee has been advocating press freedom ever since and holds gatherings to honor the journalists who face attacks, intimidation and even prison in order to report the world events. The CPJ maintains a running list of the journalists killed across the globe, and publishes a magazine and surveys that detail the accounts of such attacks and other information about the freedom of the press (Wikipedia contributors 2018b) (Committee to Protect Journalists 2018c).

In this project, we aim to visualize the deaths of journalists across the world from 1992 to March 2016, and hope to shine some light on these deaths. The data was provided by the CPJ on Kaggle (Committee to Protect Journalists 2016). A good visualization would help understand the importance of the issue and aid in further analysis of the problem to and help minimize the dangers faced by media professionals and help freedom of press.

## Background and Related Work

Protection of media personnel and staff and freedom of the speech and press are important issues and various important works have preceded this one. Here we list some important works and visualizations related to the same topic as this project.

A 2014 article by the Columbia Journal Review (Langley 2014) addresses violations of free expression across Europe with the help of a data map of the continent in Figure 1. The article emphasizes the importance of such visualizations in bringing these attacks into common knowledge and highlights the seriousness of such attacks. The visualization here can be confusing for various reasons. Firstly, the dots on the map are too big, and countries with multiple dots, (specially eastern European regions in the map) can be hard to distinguish from each other. This can be improved by using a color suitable color map and representing the data on the map as a heat map or a density map, with the help of colors instead of dots.

The Reporters Without Borders in figure 2 visualize the freedom index of different countries in the world, on a world map. The countries are ranked based on their freedom scores and users can learn more about the specific countries. The in-

formation about the countries includes the general situation and conditions of the media and the press and lists recent relevant news articles. The visualization can be improved by adding other information about other variables such as economic conditions in the countries, and it would help in comparison to provide context of freedom indices of the countries in the previous and following years.

Seelke in (Seelke 2018), in a Congressional Research memorandum, addresses the issue of violence and threats against journalists and media personnel in Mexico and performs detailed comparisons with other Latin American countries. The work includes two important visualizations. Figure 3 compares homicides of Journalists and media workers in Mexico as reported in three different databases from the year 2007 to 2017 using a bar graph. Figure 4 compares the difference perpetrators of violence against journalists in 2017 in Mexico using a well labeled pie chart in 2017. In the first visualization in figure 3, although we see a trend, it might be difficult to interpret since it is a bar chart. A better way would be to augment the bar chart with a a line chart comparing the trend obtained from the three datasets. The pie chart in figure 4, is not a good representation of quantities, since using areas (or sectors) to represent a scalar quantities can be misleading. A better visualization would be to augment the chart with a bar graph comparing those values.

The CPJ website (Committee to Protect Journalists 2018c) displays visualizations of the dataset that we analyzed in this project. It includes a data map in figure 5, where counts of deaths are represented with circles, which area is proportional to the count. When we represent data with area the viewer could be mislead, people underestimate larger quantities because they do not look large enough. Another problem of this visualization is that in some cases it is hard to distinguish to what country each circle belongs. Although, this problem is solved adding labels that appear when one clicks on the circle. We can also find a histogram that shows the number of deaths by year in figure 6, which could be better visualized as a time series.

The works stated in this section establish a strong foundation for the project. In this project we aim to provide helpful visualization to help understand the problem better and help further the cause.

## Objectives

The objective of this project is to visualize the information about the journalists who gave their life for the job through exploratory analysis, data maps and network diagrams. Creating interactive visualization, improving upon previous work for the same problem and discovering interesting insights in the data.

## Dataset

The dataset provided by the CPJ on Kaggle (Committee to Protect Journalists 2016) contains information of journalist deaths from 1992 to March 2016 in 105 countries. CPJ includes in its definition of journalists the staff, freelancers, bloggers, stringers and journalists citizen who report news in the public domain regardless of the medium, which could be printed, photographic, radio, television or electronic. The dataset includes 18 variables: Motive of murder, date of death, name of the journalist, sex, country where they were killed, organization they were working on, nationality, medium (TV, radio, etc.), job (reporter, camera operator, editor, etc), coverage (crime, politics, war, etc.), freelance (yes or no), local/foreign (origin of the journalist), source of fire, type of death, impunity for murder (monitoring conducted by CPJ about the legal process for confirmed murder cases), taken captive (yes or no), threatened (yes or no) and tortured (yes or no).

## Techniques

- **Data Maps:** Geographical Information Systems (GIS) are applications that capture, analyze and visualize data that is linked to geographical locations. Mapping data to geographical locations properly can help users to analyze spatial information (wiki.gis contributors 2018). It is helpful in such applications to visualize the data on a geographical map on a particular location of the world. Since the data we have from CPJ (Committee to Protect Journalists 2016) has countries of the deaths of journalists, we can visualize the data with data maps.

- **Network Graphs:** Representing data as graphs can allow us to visualize and discover relationships that might be missed otherwise. A graph is defined in graph theory as a set of vertices and edges. each vertex represents a data point and connections or relations between the vertices are represented by edges (Wikipedia contributors 2018c). Networkx (Networkx Contributors 2018) is a python library that allows creation, manipulation and visualization on complex graphs and networks. The visualization library Plotly (Plotly contributors 2018) now also supports network graph visualizations, and can visualize networks created using Networkx.

- **Time Series:** A time series describes how quantitative variable behaves through time. The samples should be taken at successive equally spaced points in time. Time series plots and analysis are important to understand trends in variables and how they interact with each other.

Altair (Altair 2018) and plotly (Plotly contributors 2018) are commonly used libraries that make creating interactive visualizations easier. Altair is a declarative statistical visualization library that provides a python wrapper to create Vega visualizations, while plotly provides an interface to create and host interactive visualizations in various programming languages including python.

## Process

The data had to be cleaned and preprocessed before generating visualizations.

The variable date of death was converted to datetime data type. We created a Pandas period object for the month and the year, when needed for time series visualizations and grouped the data accordingly to get counts for the respective visualizations. In order to create a data map in Altair, we

needed 3 digit ISO 3316 country codes to identify the country to be visualized. Since we did not have this information in the dataset, we merged the table available on wikipedia (Wikipedia contributors 2018d).

Then, we created visualization to explore the data, such as bar plots and a time series with Altair and Plotly libraries. We focused on making them interactive.

We created a color map that shows the number of deaths in each country with the library Altair. We used viridis as the color map, since it has most of the desirable properties for visualizations, such as being color-blind and print friendly. We carried out numerous attempts to add an slider to the map that would change the year that is visualized. Nevertheless, we did not succeed and there were no examples that use sliders with maps in altair. So instead we used HTML to create the slider.

Finally, we visualized the variables coverage, medium and jobs. Since these variables contain more than one value, for example a person can cover war and politics, we used graphical networks to visualize them.

## Results and Discussion

Since we wanted to make interactive visualizations, we created most of our visualizations in Altair (Altair 2018) and Plotly (Plotly contributors 2018). The interactive visualizations can be seen on the web page (here), and the code can be found in the github repository for the project (here).

### Exploratory Analysis

In order to get a better understanding of the data and to find the best visualization of the data, we started with some exploratory analysis.

An interactive time series plot of the number of deaths in each month, as shown in figure 7 can tell us that there is no specific trend across all countries. We can however see certain peaks that correspond to specific events.

A bar plot showing the number of deaths (figure 8) in all countries across all the years in the dataset (1992 -2016) shows that there are a few countries that are clearly more dangerous, with Iraq having the highest number of deaths followed by Philippines. It should be noted that the dataset contains death in Syria only after 2010. A time series graph of 5 countries with the most number of deaths (figure 10) reveals this fact and shows an increasing trend in the number of deaths in Pakistan and Mexico. The number of deaths in Iraq go down in 2007.

Next we draw staked bar charts that show us the number of male and female media workers in the data (figure 9), the number of freelance media workers against the number of media workers associated with specific news organizations (figure 11), and the impunity of murder (figure 14). The last chart is particularly interesting and along with a bar chart of counts of each type of death (figure 13) tells us that most of the deaths in the data were confirmed murders and these murders were not investigated or only partially investigated. We see from figure 12 that most of the people killed were local to the country of death.

### Data Maps

We generated a data map using Altair. We chose equirectangular projection for our data maps.

Figure 15 shows the deaths in all the countries across all the years in our dataset. As can be seen from the bar plots, we see that Iraq has the highest number of deaths.

We created such maps for all the years in the dataset and constructed an interactive visualization that allows the viewers to select the year with the help of a slider. Moving from 1992 to 2016, one can clearly see how the scenario changes in different countries, and we were able to find some interesting years.

Figure 16 shows a map of 1995 where Algeria has the highest number of deaths. These are related to the Algerian civil war following a change in the political scenario in the country. 2009 shows a high number of deaths in Philippines, as can be seen from figure 17, which can be attributed to the Philippines massacre of 2009. In figure 18 we can see an increase in the number of deaths in Syria in 2010, as the number of deaths in Iraq go down.

### Network Diagrams

To visualize the different multi-valued variables, we plotted three network diagrams, using the kamada kawai force directed layout in 3 dimensions using plotly, which allows viewers to zoom in on nodes, and rotate the visualization to see other nodes that might be hidden.

From the network diagram for coverage shown in figure 19, we can see that most of the journalists in the dataset were killed while reporting on topics of war, and corruption.

Another network diagram for specific media jobs reveals that most media workers in the dataset worked as print reporters, camera operators, or broadcast personnel. Another network diagram reveals that most journalists worked in print media, television or radio, while there are very few online journalists in the data.

## Future Work

- The project can be extended by getting more information about the deaths of various journalists from twitter and news articles to better understand the scenario and use this information to perform detailed analysis on the data.

- The data maps can be augmented by adding annotations about the political situations in some of the volatile countries across various years, to visualize the data along with a story.

- Deeper network analysis can be performed on the data collected from the news articles.

## Acknowledgements

We would like to thank Kaggle for their data platform that makes accessing and downloading datasets very easy for data scientists.

## References

[Altair 2018] Altair. 2018. Altair: Declarative visualization in python. Altair docs.

[Committee to Protect Journalists 2016] Committee to Protect Journalists. 2016. Journalists killed worldwide since 1992. Kaggle.

[Committee to Protect Journalists 2018a] Committee to Protect Journalists. 2018a. 1321 journalists killed. Committee to Protect Journalists Website.

[Committee to Protect Journalists 2018b] Committee to Protect Journalists. 2018b. 42 journalists killed. Committee to Protect Journalists Website.

[Committee to Protect Journalists 2018c] Committee to Protect Journalists. 2018c. Committee to protect journalists. CPJ website.

[Langley 2014] Langley, A. 2014. Maps reveal violence against journalists. Columbia Journal Review.

[Marlowe 2018] Marlowe, L. 2018. Three years after charlie hebdo, magazines staff live in fear. the Irish Times Website.

[Networkx Contributors 2018] Networkx Contributors. 2018. Software for complex networks. Networkx website.

[Plotly contributors 2018] Plotly contributors. 2018. Network graphs in python. Plotly docs wensite.

[Reporters Without Borders 2018a] Reporters Without Borders. 2018a. 2018 world press freedom index. Reporters Without Borders website.

[Reporters Without Borders 2018b] Reporters Without Borders. 2018b. Protecting journalists. Reporters Without Borders Website.

[Seelke 2018] Seelke, C. R. 2018. Violence against journalists and media workers in mexico and u. s. policy.

[wiki.gis contributors 2018] wiki.gis contributors. 2018. Geographic information system. wiki.GIS.com website.

[Wikipedia contributors 2018a] Wikipedia contributors. 2018a. Charlie hebdo shooting — Wikipedia, the free encyclopedia. [Online; accessed 16-September-2018].

[Wikipedia contributors 2018b] Wikipedia contributors. 2018b. Committee to protect journalists — Wikipedia, the free encyclopedia. [Online; accessed 16-September-2018].

[Wikipedia contributors 2018c] Wikipedia contributors. 2018c. Graph theory — Wikipedia, the free encyclopedia. [Online; accessed 17-September-2018].

[Wikipedia contributors 2018d] Wikipedia contributors. 2018d. List of iso 3166 country codes — Wikipedia, the free encyclopedia. [Online; accessed 13-December-2018].

[Wikipedia contributors 2018e] Wikipedia contributors. 2018e. Reporters without borders — Wikipedia, the free encyclopedia. [Online; accessed 16-September-2018].

Figure 1: Violations of Free Expressions in Europe: 2014 (Langley 2014)



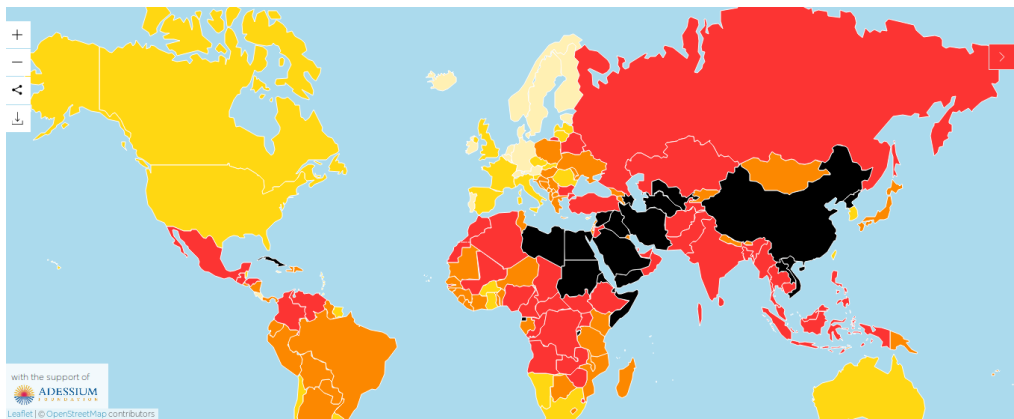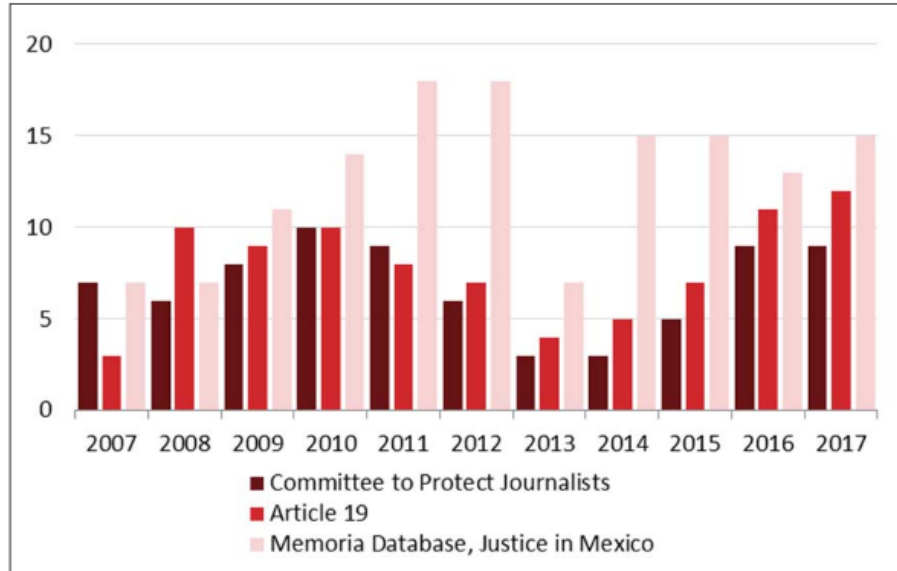Figure 2: Freedom index across the world by RSF (Reporters Without Borders 2018a)

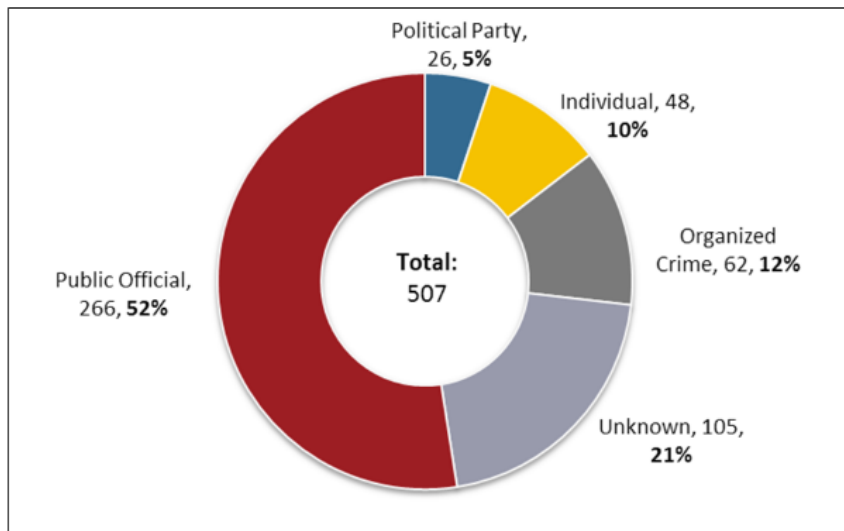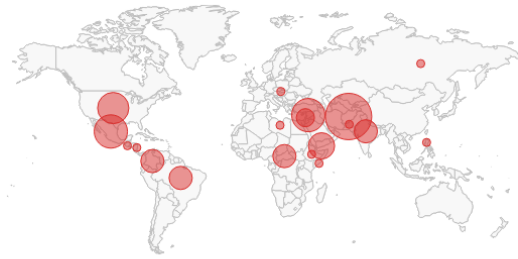Figure 3: Homicides of Journalists ans Media Workers in Mexico: 2007-2017 (Seelke 2018)



Figure 4: Perpetrators of Violence against Journalists in Mexico: 2017 (Seelke 2018)
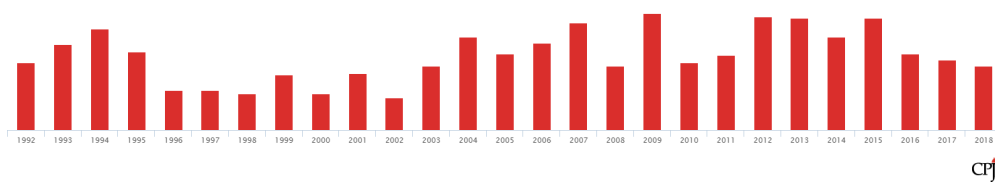
## 62 Journalists Killed

in 2018



Figure 5: Map of journalists killed in 2018 (Committee to Protect Journalists 2018b)

## 1321 Journalists Killed

between 1992 and 2018 / Motive Confirmed



Figure 6: Histogram of journalists killed in 2018 (Committee to Protect Journalists 2018a)
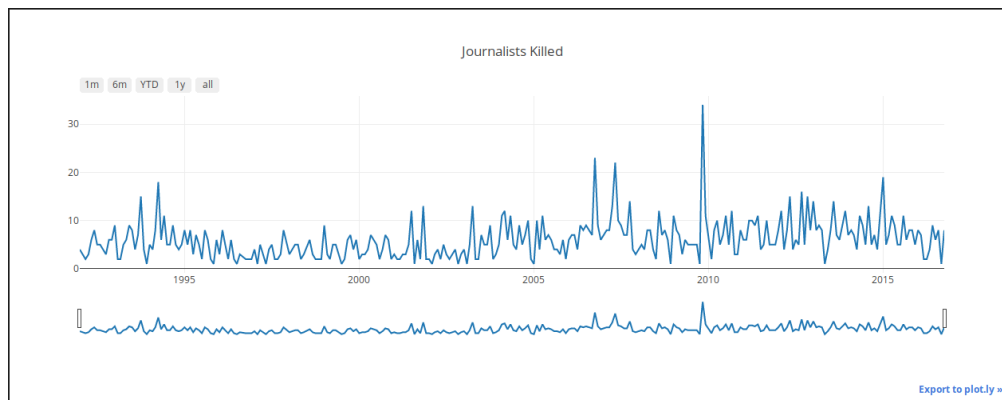


Figure 7: An interactive time series visualization of the number of deaths per month
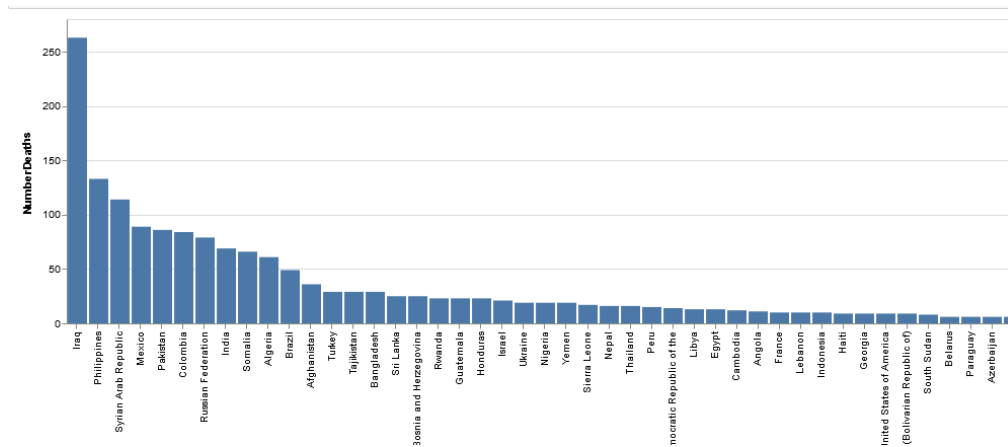
Figure 8: Number of deaths in each country
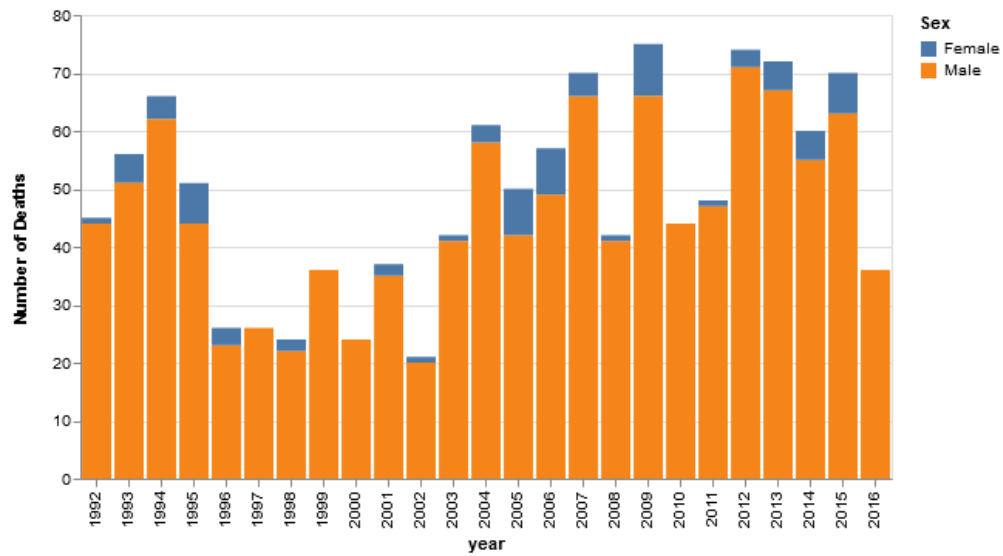


Figure 9: Number of males and females in the dataset over the years
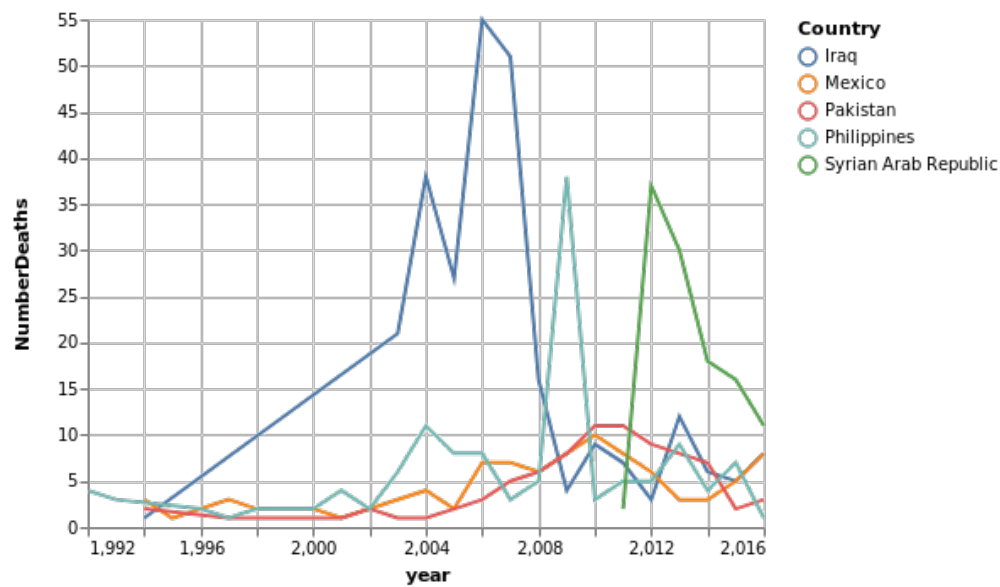
Figure 10: A time series of the 5 countries with the most deaths
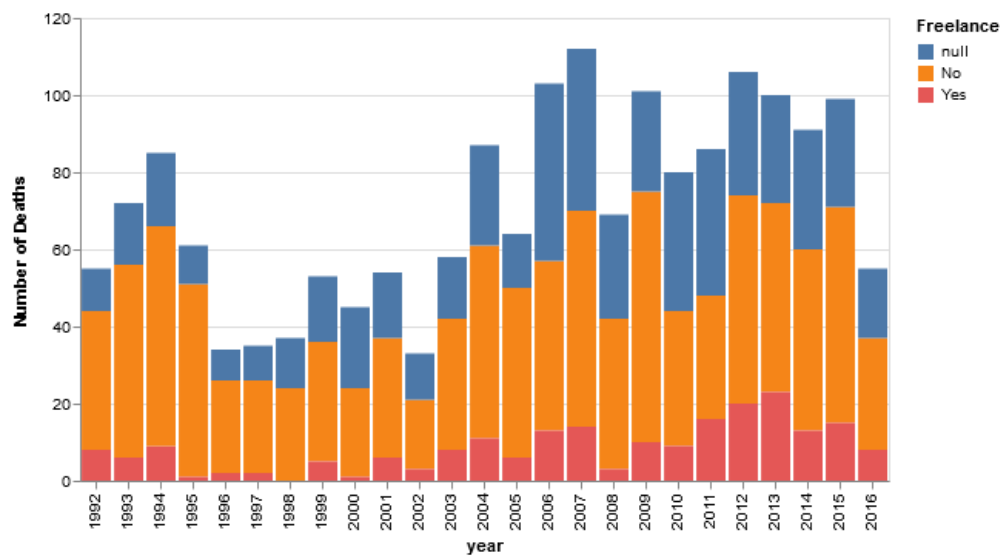


Figure 11: Number of freelance journalists over the years
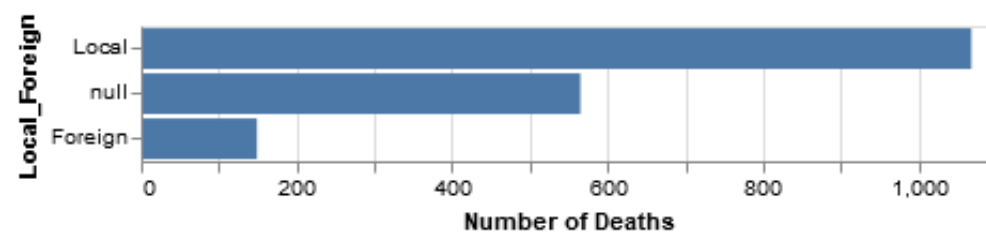


Figure 12: Most media workers were locals to the country they were killed in
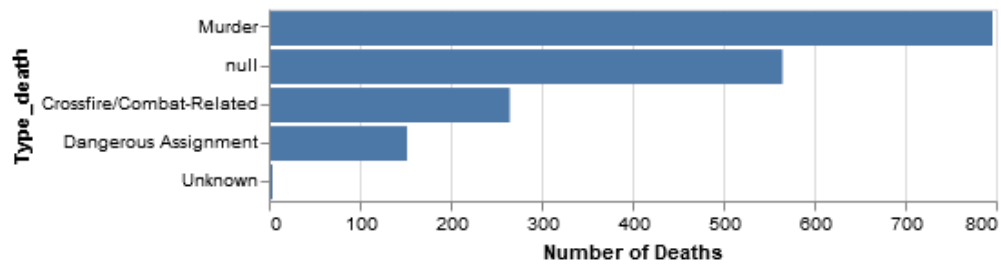
Figure 13: Counts for the type of death of the media worker
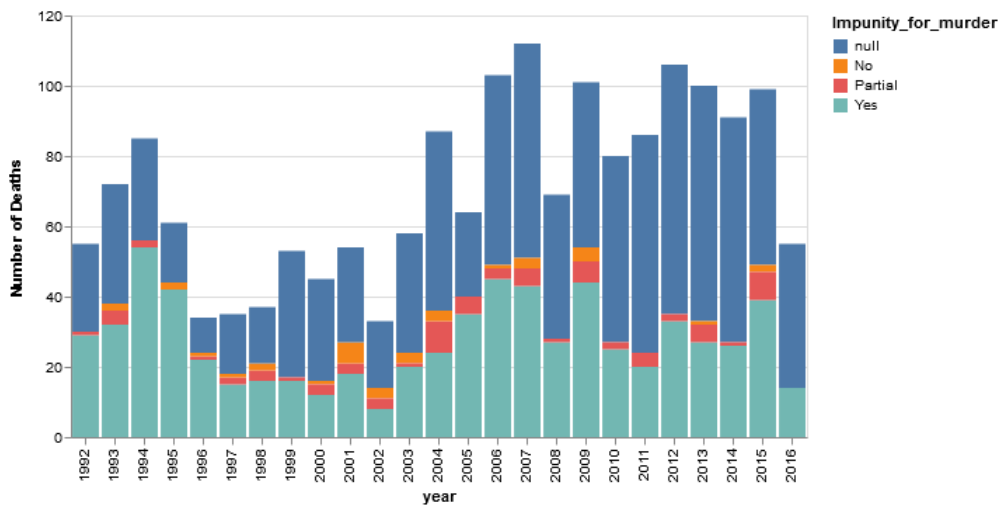


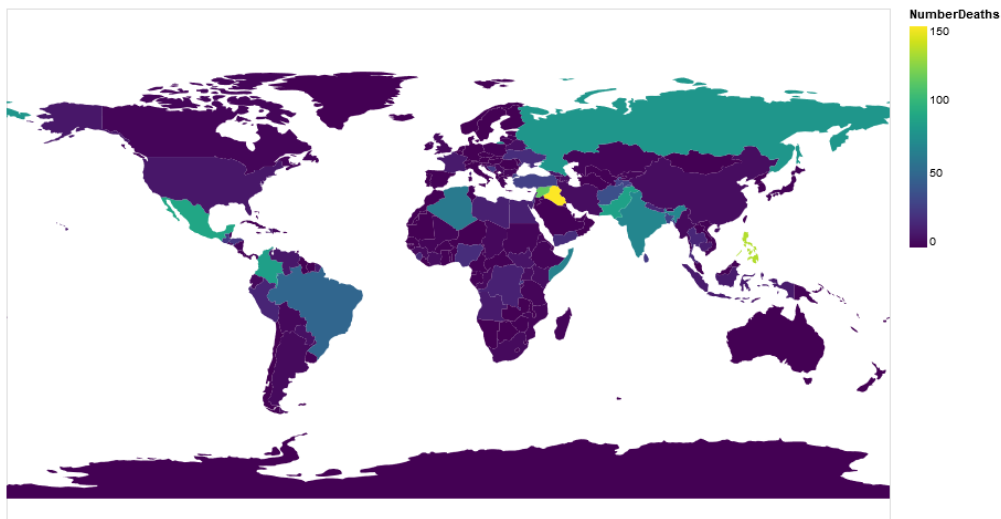Figure 14: Impunity of death: Most cases were not investigated, or only partially investigated



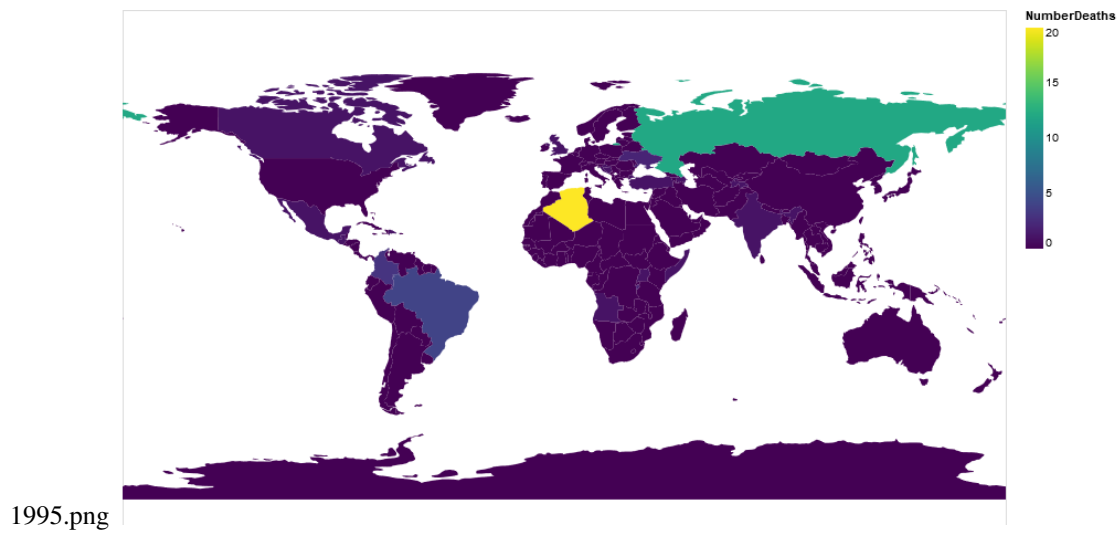Figure 15: The number of deaths in each country from 1992 to 2016

1995.png

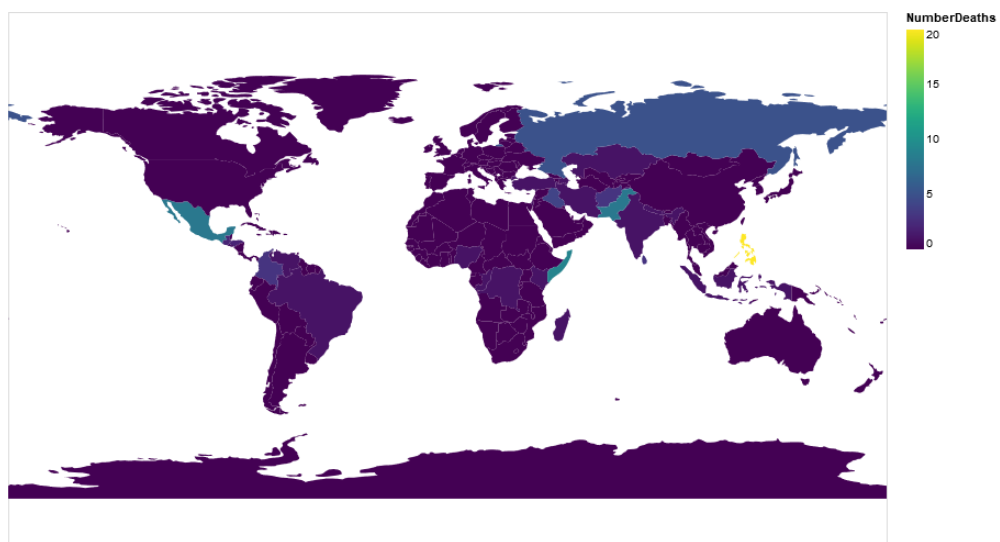Figure 16: 1995: A high number of deaths in Algeria due to the Algrian Civil War



Figure 17: 2009: A high number of deaths in Philippines due to the Philippines massacre
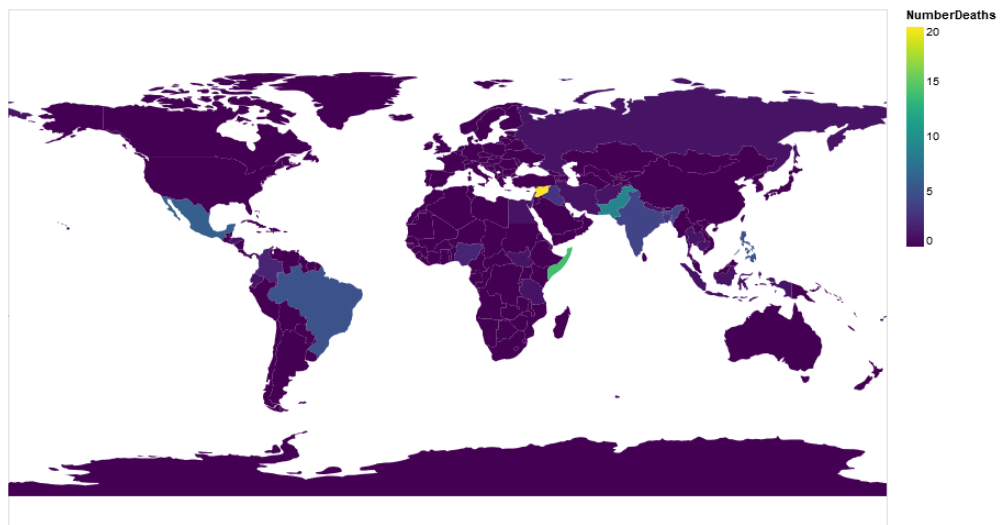
Figure 18: 2012: An increase in the number of deaths in Syria due to the Syrian Crisis
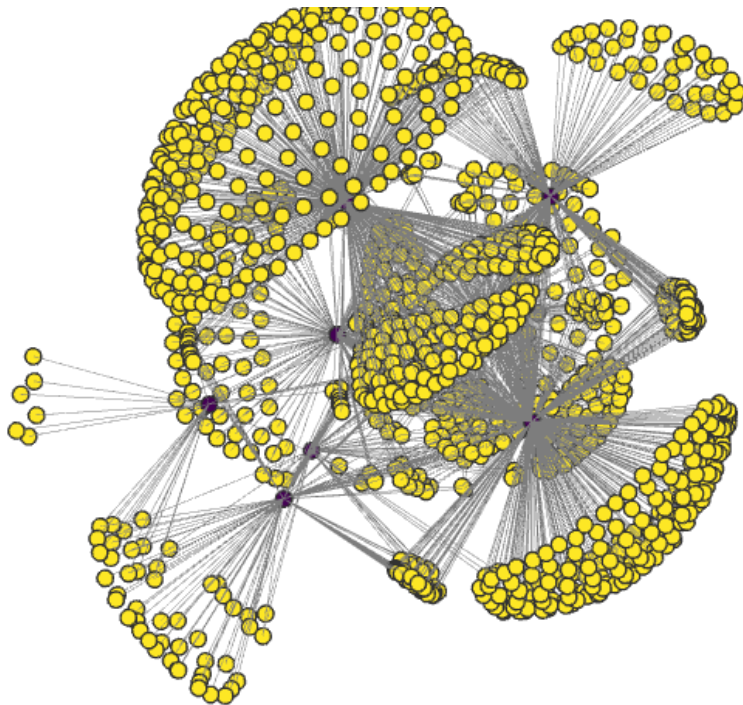


Figure 19: Network diagram of journalists in yellow and the topic of coverage in purple