



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

TERRORISMO EN EL MUNDO

VISUALIZACIÓN

MEMORIA TRABAJO ACADÉMICO



Grado en Ciencia de Datos

Álvaro Mazcuñán Herreros
Miquel Marín Colomé

ÍNDICE

1. INTRODUCCIÓN	2
2. FASES DE VISUALIZACIÓN	2
2.1. ADQUISICIÓN	2
2.2. FILTRADO.....	2
2.3. FORMATEO	3
2.4. MINADO	3
2.5. REPRESENTACIÓN	4
2.6. INTERACCIÓN.....	7
2.7. REFINADO	8
3. CONCLUSIÓN	8

ÍNDICE DE ILUSTRACIONES

Ilustración 1 - Mapa ataques terroristas	4
Ilustración 2 - Choropleth ataques terroristas	4
Ilustración 3 - Secuencia de colores (paletas)	5
Ilustración 4 - Gráfico barras ataques populares	5
Ilustración 5 - Gráfico barras impacto económico ataques terroristas.....	6
Ilustración 6 - Gráfico de líneas fallecidos y heridos en España.....	6
Ilustración 7 - Gráfico de barras fallecidos y heridos por tipo de ataque en Nigeria	7
Ilustración 8 - Dashboard completo información por País	7

1. INTRODUCCIÓN

Año tras año suceden guerras, conflictos y revueltas llevándose decenas de vidas por delante, destrozando ciudades, regiones y países. También dejan un gran rebaño de heridos, saturando hospitales y centros médicos de zonas conflictivas.

Cuando sucede un ataque y es contado en la televisión, al espectador se le pueden ocurrir preguntas como: “¿Dónde se concentran más los ataques terroristas?”, “¿Qué países son los más afectados?”, “¿Qué tipos de ataque son los más populares?”. Estas preguntas no tienen una respuesta instantánea, pero gracias a este proyecto es posible resolver estas cuestiones.

2. FASES DE VISUALIZACIÓN

Una vez introducido el objetivo de este trabajo, se puede pasar ya a la parte importante de éste, la descripción de cada una de las fases de visualización.

2.1. ADQUISICIÓN

La base de datos escogida es de la conocida web *Kaggle* (<https://www.kaggle.com/START-UMD/gtd>). Dicha base, almacena información sobre ataques terroristas desde el año 1970 hasta el 2017. La obtención de dichos datos fue sencilla. Se descargó en formato csv. Junto con la base de datos, también se obtuvo un PDF que explicaba qué tipo de información guardaba cada variable (metadatos).

2.2. FILTRADO

Dado el gran volumen de datos que almacena la tabla “*terrorism*”, se requería reducir ese peso para poder agilizar los cálculos necesarios para las siguientes fases. Por ello, se eliminaron todos los datos menos los de interés.

Esta fase se realizó en *Python*, ya que es una potente herramienta para tareas de filtrado y formateo de datos. Antes de realizar cualquier modificación, la base de datos contaba con 135 columnas para 176330 ataques. Era completamente necesario reducir las dimensiones de ésta.

Por ello, se seleccionaron únicamente las variables que serían útiles para los análisis venideros. Estas variables en cuestión son:

“*year*”: almacena el año del ataque. Toma valores desde 1970 hasta 2017.

“*month*”: almacena el mes del ataque.

“*day*”: guarda el día del ataque.

“*crit1*”, “*crit2*”, “*crit3*”: almacenan la información de por qué se realizó el ataque.

“*country_txt*”: en qué país ocurrió el ataque.

“*region_txt*”: región donde sucedió el ataque.

“*provstate*”: provincia/estado donde se produjo el ataque.

“*city*”: ciudad del ataque.

“*latitude*” y “*longitude*”: punto geográfico exacto donde sucedió el ataque.

“*attacktype1*”: tipo de ataque. Variable cuantitativa categórica que indica los tipos de ataque terrorista como bombardeo, asesinato, rapto...

“*suicide*”: Variable cuantitativa categórica que dice si el ataque fue suicida o no.

“*weaptype1*”: Variable cuantitativa categórica que informa del tipo de arma utilizada en el ataque.

“*nkill*” y “*nwound*”: número de muertos y heridos en un ataque, respectivamente.

“*success*”: Variable cuantitativa categórica. ¿Tuvo éxito el ataque o no?

“*propextend*”: Variable cuantitativa categórica que informa del impacto económico del ataque. Almacena si el ataque provocó un daño inferior a 1M de dólares, inferior a 1 billón de dólares o superior a esta cantidad.

Algunas de estas variables se utilizaron únicamente para fines exploratorios y no entraron en el *dashboard* final. Es el caso de “*crit1*”, “*crit2*”, “*crit3*”, “*region_txt*” y “*success*”, entre otras.

2.3. FORMATEO

Muchas de estas variables venían en el formato equivocado. Todas las que eran cuantitativas categóricas tomaban valores *float*, teniendo el carácter decimal sin aportar información alguna. Por ello, se transformaron a cadena de caracteres. Además, algunas cadenas como el año, mes y día eran enteros y necesitaban cambiar el tipo. Así, todas las variables pasan a ser *string*, a excepción de “*nkill*” y “*nwound*” que son enteros.

Al tener muchas observaciones, es normal encontrar datos faltantes. Se realizó una prueba de eliminar todos estos valores, pero la base de datos pasaba de 170000 ataques a apenas 40000. Como se perdía mucha información, se dejaron estos valores faltantes. Como excepción, en las dos variables numéricas mencionadas anteriormente, en vez de dejar estos valores faltantes, se imputaron estos huecos por la media de cada columna.

Todas estas modificaciones están recogidas en el archivo “*Limpieza_Y_Recodificación.ipynb*”, donde se explica paso a paso el proceso de formateo de estas variables.

Además, para representar los países en uno de los gráficos, éste requería que los territorios del mundo tuvieran un código exacto para poder pintarlos sobre el mapa en cuestión. Debido a esto, se necesitó un cambio en el nombre de los mismos. Para realizar esta modificación, se echó mano de la librería *countrycode* que transformaba el nombre del país al código correspondiente. Por ejemplo, se pasaba de “Spain” a “ESP”.

2.4. MINADO

El proyecto ya pasa a R. En este lenguaje se preparó el código para realizar cada gráfica. Se implementaron numerosas recodificaciones, filtros y obtención de estadísticos para la correcta implementación de los elementos gráficos. Por ejemplo, para la creación de muchas de las gráficas se necesitaba cierta información concreta para un país en un periodo de tiempo concreto. Para ello, el uso de sumatorios ha sido clave para lograr el dato que se requería.

Al principio, se tenía todo en el mismo documento pero el tiempo de ejecución superaba los 3 minutos cada vez que se quería obtener el *dashboard* final. Debido a esto, todos los cálculos de minado de datos se separaron del código en *Rmarkdown* y se ejecutó una vez en otro documento guardando así en un *csv* la salida de los algoritmos creados para simplificar la información propios de este paso. Al guardar estas tablas en un archivo aparte, se logró reducir el tiempo de ejecución de 3 minutos a apenas 10 segundos. Así que no aparecerá nada de minado de datos en el código final, pero sí se ha tenido en cuenta esta capa de visualización de información.

Como observación, se ve con claridad que, en la última pestaña del *dashboard*, se ha implementado la capa de minado, ya que se realizan sumatorios tanto del número de fallecidos como de heridos en el país seleccionado.

2.5. REPRESENTACIÓN

Previamente a la realización del *dashboard* en cuestión, se ha ejecutado cada una de las gráficas contenidas en este de forma separada. La primera de ellas, y la más importante, es la visualización de un mapa del mundo en el cual se grafican unos puntos encima de éste. Dichos puntos tratan todos y cada uno de los ataques terroristas que sucedieron entre el año 1970 y el 2017.

Se piensa que al realizar esta gráfica se puede obtener desde el principio una visualización muy general del tema que se está tratando y, de esta forma, al visualizar el resto de elementos gráficos del *dashboard*, se conocerá a la perfección la naturaleza de los datos utilizados.

En un principio se quisieron pintar dichos puntos en color rojo. No obstante, en la siguiente fase (Refinado) se explicarán las razones por los cuales se ha cambiado finalmente a un color amarillo.

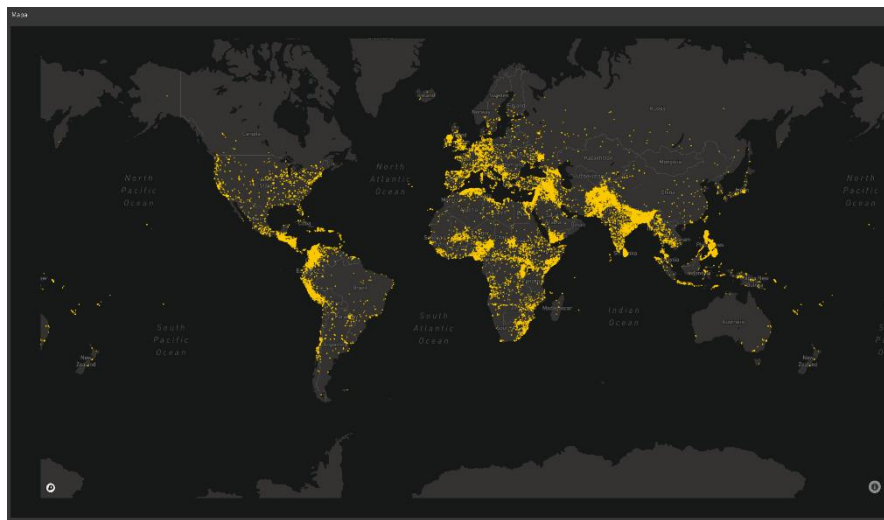


Ilustración 1 - Mapa ataques terroristas

A continuación, una vez se ha obtenido una visión general de los ataques mediante el mapamundi, se pasa ahora a la realización de otro mapa, en este caso un *Choropleth*, en el que los países se somborean de distintos colores, frecuentemente de la misma gama cromática. Mediante dicha representación gráfica se pretende mostrar el número de fallecidos totales por país.

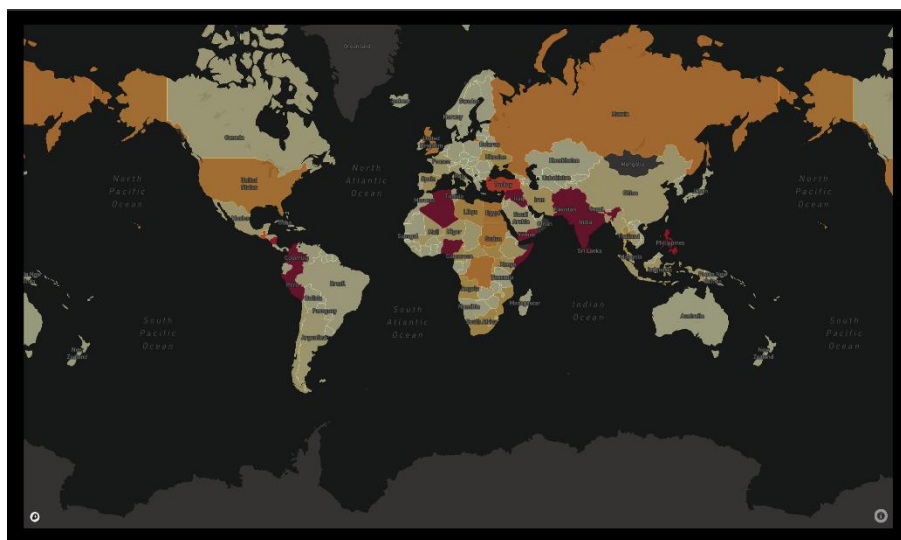


Ilustración 2 - Choropleth ataques terroristas

La secuencia de colores que se ha escogido para representar dicha información es la “YlOrRd” que contiene desde un amarillo muy claro a un rojo oscuro. Se ha elegido dicha paleta ya que se están tratando datos acerca de terrorismo y conviene utilizar colores “cálidos” para representar dicha información.

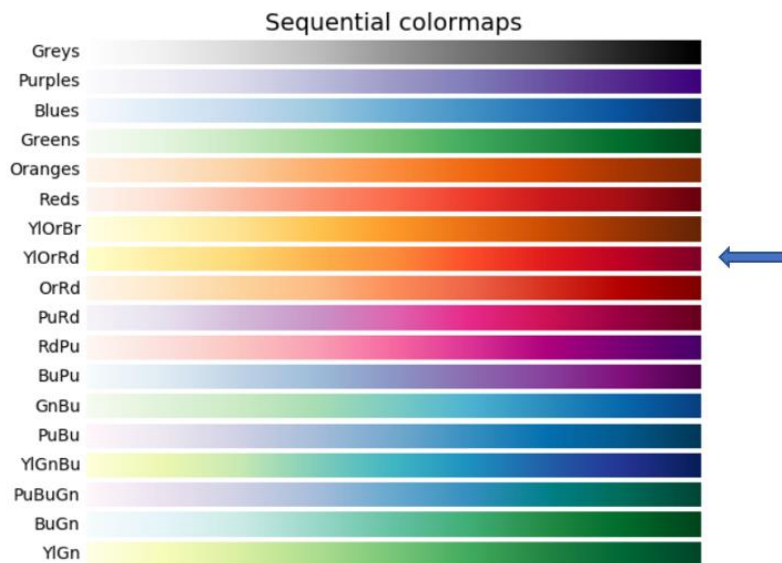


Ilustración 3 - Secuencia de colores

Ahora se pasaría a graficar información más detallada. Para empezar, se vuelve a implementar el mismo mapa que al principio pero con un matiz que se explicará en la fase de interacción. Junto a este mapa, aparecen dos gráficos más.

El primero de ellos es un gráfico de barras que contiene información de los ataques más populares que se produjeron entre los años que se disponen en la base de datos. Para representar dicha información se utilizó la variable “*attacktype1*”. En cada una de las barras, se muestra el conteo de ataques realizados de ese tipo. Posteriormente, en las secciones de Refinado e Interacción se explicaran cada una de las mejoras que se han realizado en esta representación.

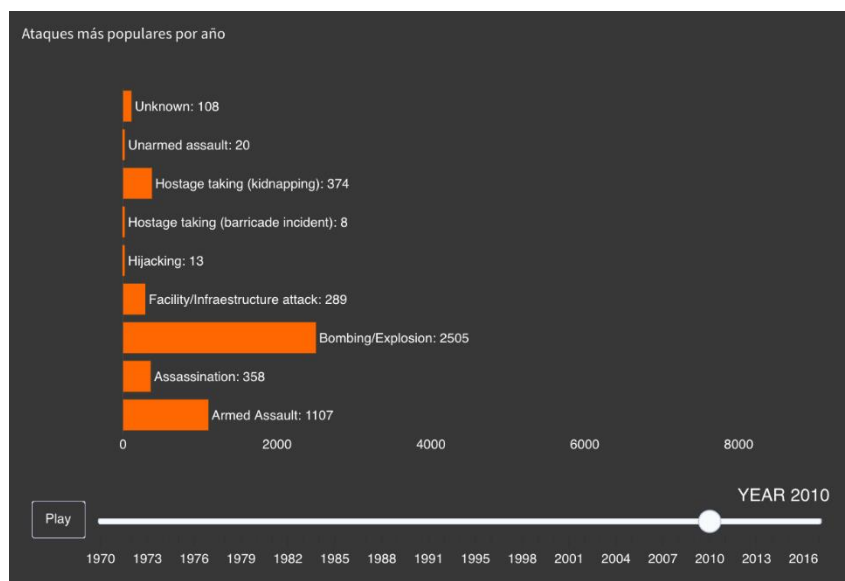


Ilustración 4 - Gráfico barras ataques populares

¹ https://matplotlib.org/3.1.1/gallery/color/colormap_reference.html

Seguidamente, junto con el gráfico que se acaba de presentar, se adjunta otro gráfico de barras. En este caso se ha utilizado la variable “*propextend*”. Dicha variable, tal y como se ha explicado en el apartado de la fase de Filtrado, explica el impacto económico que ha tenido ese ataque en concreto. De forma similar al anterior gráfico, se han contabilizado el número de ataques por cada uno de los impactos económicos que han supuesto.

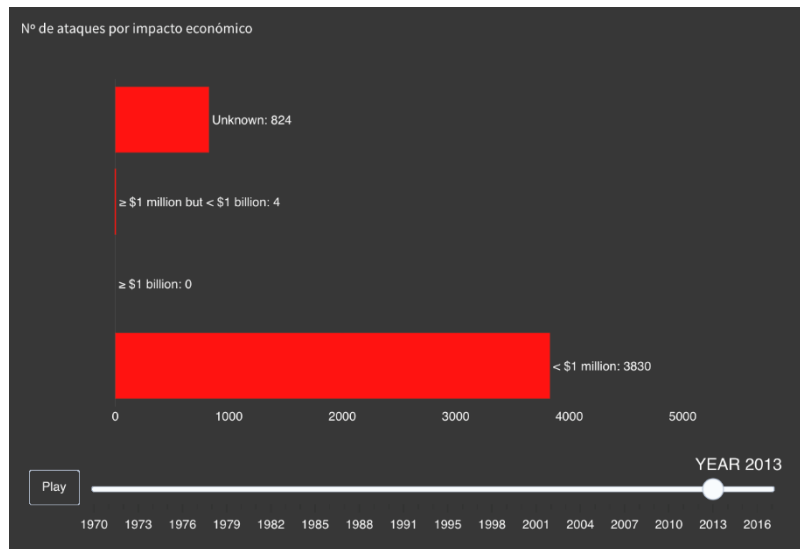


Ilustración 5 - Gráfico barras impacto económico ataques

Finalmente, en la última pestaña del *dashboard*, se describe información aún más detallada, en este caso, se muestran los datos por un país en concreto. Por una parte, se presenta el mapa con los ataques pintados en la parte superior y, en la parte inferior, se realizan tres tipos de gráficos tanto para representar fallecidos como heridos de los ataques terroristas.

- El primero de estos es un elemento gráfico conocido como *ValueBox* que, como el nombre indica, se trata de una caja que indica un valor en concreto. Para este caso, se utilizarán dos de estos gráficos en los que se mostrará el número de fallecidos y el número de heridos totales para un país en concreto.
- El segundo tipo es un gráfico de líneas que indica la evolución de los fallecidos/heridos de ese país en concreto. Para ello se ha obtenido la suma de dichas variables para cada año.
- Y el tercero es otro gráfico de barras que está relacionado con el que se ha comentado anteriormente. No obstante, en este caso, en lugar de contabilizar el número de ataques realizados de cada tipo, se obtiene la suma de fallecidos/heridos que han provocado cada uno de estos para cada año.

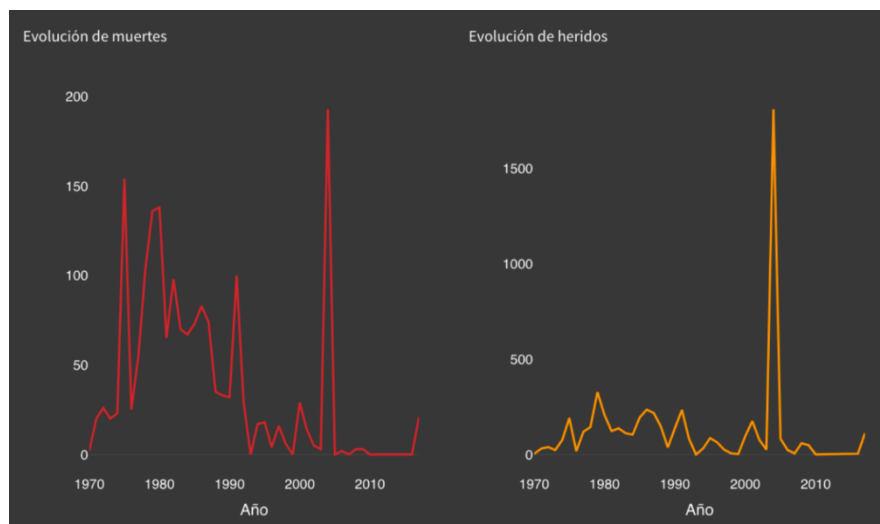


Ilustración 6 - Gráfico de líneas fallecidos y heridos en España

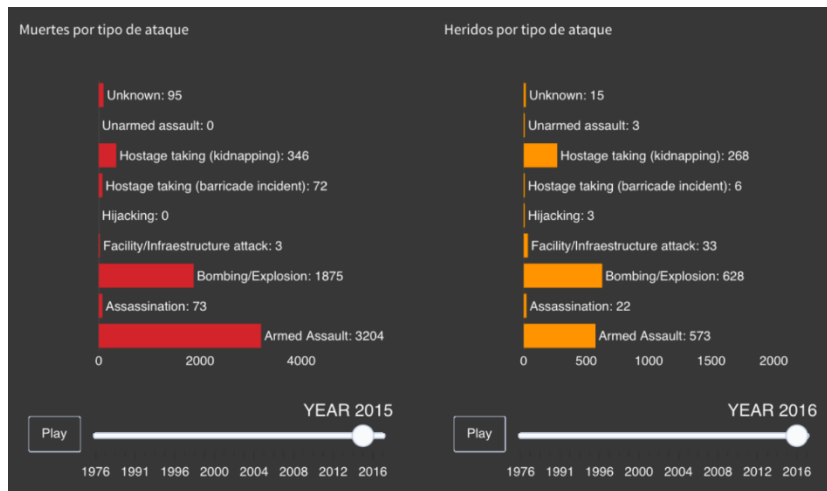


Ilustración 7 - Gráfico de barras fallecidos y heridos por tipo de ataque en Nigeria

2.6. INTERACCIÓN

Para hacer que las gráficas fueran mucho más dinámicas y que el usuario pudiese investigar y observar la información que deseara, se implementaron numerosos botones y seleccionadores.

Primeramente, en la pestaña en la que se trata información de manera más detallada, este inciso consiste en mostrar la información según el tipo de ataque. A la izquierda de los gráficos, aparece un desplegable para seleccionar el tipo de ataque. Al seleccionar, por ejemplo, “Asesinato”, el mapa pasará a mostrar los ataques que hayan sido asesinatos. En esta misma pestaña, se implementa un botón en las dos gráficas de abajo para inicializar la interacción a lo largo de los años y visualizar de manera mucho más divertida esta información.

En el caso de la última pestaña, el procedimiento es muy parecido pero varía en el seleccionador, que ahora será por país y no por tipo de ataque. Además, este cambio afecta a todos los elementos gráficos presentes en dicha pestaña. El botón tiene el mismo uso que en la anterior explicación.

Es oportuno comentar que el seleccionador utilizado no es el que ofrece *Plotly*, sino es el que ofrece la librería *Shiny*. Para poder utilizarlo, se emplea la función *selectInput* para seleccionar y, posteriormente, se engloba el código de *Plotly* bajo la función *renderPlotly*. Para que el valor seleccionado haga efecto sobre el gráfico, el elemento en cuestión se invocará con *input\$Tipo* (siendo *Tipo* el nombre que se le da al seleccionador).

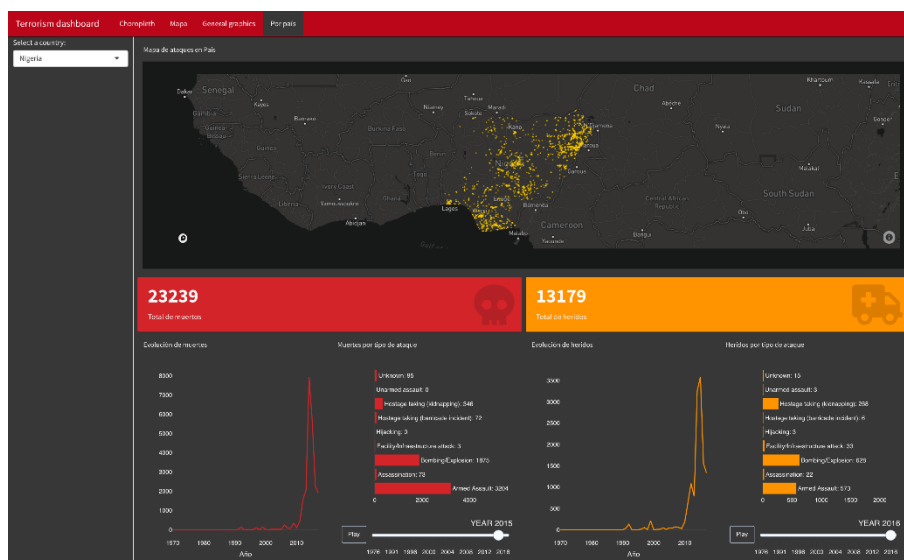


Ilustración 8 - Dashboard completo información por País

2.7. REFINADO

Finalmente, una vez se han obtenido todos y cada uno de los gráficos y añadida la interacción entre estos, se pasaría ya a la última fase de visualización, la de refinado, que consistiría en mejorar cada uno de las representaciones planteadas desde un principio.

En la fase de representación, el primer gráfico introducido ha sido el mapa con los puntos (en referencia a los ataques) pintados sobre éste. En un principio se habían coloreado en rojo pero, se terminó cambiando a un amarillo. Esta decisión viene dada durante la realización de la última pestaña del *dashboard*, en la cual se representan los gráficos de fallecidos y heridos junto con el mapa. Si se dejara el color rojo sobre los puntos, cualquiera que visualizase el *dashboard* lo podría asociar a fallecidos y no a un ataque en general.

En cuanto a las tonalidades en la pestaña de las gráficas detalladas, dado que el mapa se visualiza en color amarillo, se quiso dar a entender que lo que se representaba en los dos gráficos de la parte inferior es otro tipo de información diferente a la del mapa pero, en cierta medida, relacionada. Por ello, se ha decidido emplear colores análogos al amarillo del mapa para los otros gráficos, tomando valores anaranjados.

Finalmente tocaría hablar acerca de los colores de las gráficas de la última pestaña, referentes a los heridos y fallecidos en cada uno de los ataques. Tal y como se ha comentado anteriormente, el mapa (que se encuentra en la parte superior de la pestaña) estará pintado con los puntos en amarillo. Por lo tanto, se ha decidido escoger unos colores análogos que representen información muy similar. Por consiguiente, se elige un color rojo no muy saturado para representar la información de fallecidos y un naranja no muy oscuro para los heridos.

Cabe añadir que en los “*ValueBox*” mencionados anteriormente, se añaden dos iconos para aclarar aún más qué representa cada uno de estos gráficos. Para los fallecidos se incluye una calavera y, para los heridos, una ambulancia.

3. CONCLUSIÓN

En referencia a las preguntas de la introducción, este proyecto acerca al usuario a posibles respuestas.

Para la pregunta “*¿Dónde se concentran más los ataques terroristas?*” La respuesta está clara, a simple vista, India y Bangladesh se llevan el gato al agua. Con tan solo observar el mapa global de los ataques, casi ni se aprecia el área de dichos países de la cantidad de ataques que han recibido.

“*¿Qué pasa con los países más afectados?*” Observando el *Choropleth* se aprecia que los que están pintados en un color más oscuro son los más afectados. De estos, Irak es el que más muertes ha registrado (se puede observar concretamente el valor posicionándose sobre dicho país → 79565)

Y para la pregunta “*¿Qué tipos de ataque son los más populares?*” Los asaltos armados y los bombardeos son los más comunes, sobre todo en Oriente Medio y Asia.