FINAL PROJECT

BDA 504 Data Visualization

24 May 2018

This document was prepared for final project of BDA 504 course, MEF University.

Türkan Sevgili
sevgilit@mef.edu.tr

Table of Contents

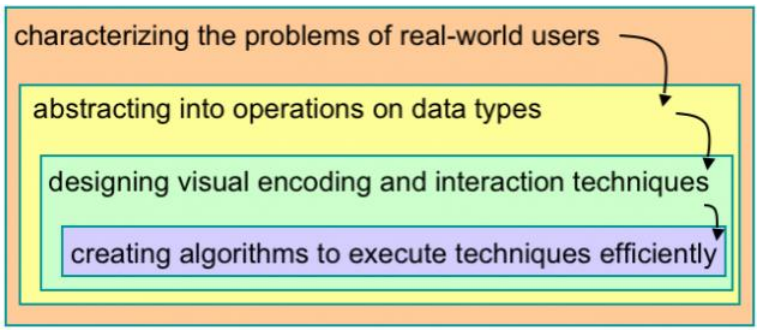
A. Introduction

In this report, a selected data was visualized as the final project of lesson BDA504. An interactive visualization was prepared. Firstly, I want to start with most commonly preferred definition of data visualization.

Data visualizations are dynamic expression of data that can be filtered, aggregated and further manipulated "on the fly" to produce an enhanced analytic view of the data.

Ben Fry describes the process of creating a data visualization in a very accessible way. In his *Phd thesis on "Computational Information Design"* he outlines the following process: [1]

After visualization research, Tamara Munzner has defined another workflow for creating visualizations: *A Nested Model for Visualization Design and Validation.* [2]



I have to take into consideration above two processes while preparing visualization. I tried to choose a tidy format dataset that explained briefly in *this document*. [3] Because tidy datasets are easy to manipulate, model and visualize. Structure of my dataset like that; each variable is a column, each observation is a row. Sincerely my dataset is in a tidy format, so I have not to need to realize all above processes.

Dataset Description, Content and Data fields

"Terrorism_Turkey" dataset which I have been used in project is part of The Global Terrorism Database ***(GTD)*** [4] that includes information on terrorist events around the World (205 countries, more than 33000 cities, from 1970 through 2016 on at least 45 variables for each case.

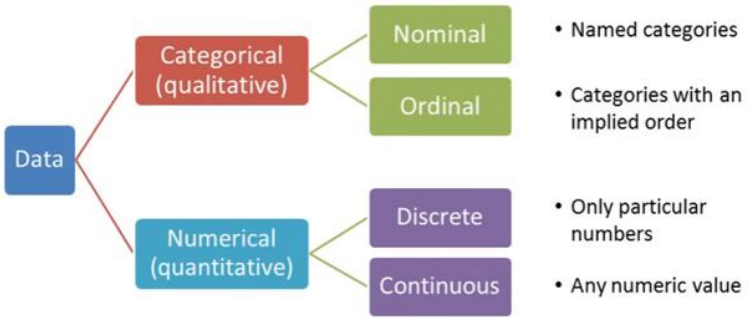
Geography: Turkey
Time period: 1970-2016, except 1993
Unit of analysis: Attack
Variables: 18 variables on location, tactics, perpetrators, targets, and groups
Sources: Unclassified media articles

The data set consists of 4106 observations and 18 variables. Necessary information about mostly used fields can be found below;

- eventid: Incidents from the GTD follow a 12-digit Event ID system.
- iyear: This field contains the year in which the incident occurred.
- imonth: This field contains the number of the month in which the incident occurred.
- iday: This field contains the numeric day of the month on which the incident occurred.
- country: This field identifies the country code country or location where the incident occurred.
- city: This field contains the name of the city, village, or town in which the incident occurred.
- latitude: This field records the latitude (based on WGS1984 standards) of the city in which the event occurred.
- longitude: This field records the longitude of the city in which the event occurred.
- success: (1 = "Yes" The incident was successful, 0 = "No" The incident was not successful.)
- attacktype: This field captures the general method of attack and often reflects the broad class of tactics used.
- targetype: The target/victim type field captures the general type of target/victim.
- gname : This field contains the name of the group that carried out the attack.
- weapontype : Weapons used in an attack are recorded for each case.
- nkill : This field stores the number of total confirmed fatalities for the incident.
- nwound : This field records the number of confirmed non-fatal injuries to both perpetrators and victims.

Type of Data

Basic overview of the types of data is shown below;



Based on above figure, I determined my dataset variables' data types. You can show in below table;

Variables	Data Types			
	Qualitative		Quantitative	
	Categorical/Nominal	Ordinal	Discrete	Continuous
eventid	✓			
iyear			✓	
imonth			✓	
iday			✓	
country_txt	✓			
city	✓			
latitude				✓
longitude				✓
success	✓ (also binary)			
attacktype	✓			
targetype	✓			
gname	✓			
weapontype	✓			
nkill			✓	
nwound			✓	

Objectives

In order to measure the understanding and influence of the user/reader, I think that my purpose and message must be clear. So I determined some objectives below for interactive visualization.

- The purpose of the analysis and visualization is to understand the Turkey's Terrorism Statistics on the basis of the cities and years.
- A clear understanding of the data to see if we can locate useful insights about where and when terror attacks occur in Turkey.
- To create an opinion on the relatively risky and secure/safe cities and years.

Data Structure

Spatial structure: Map

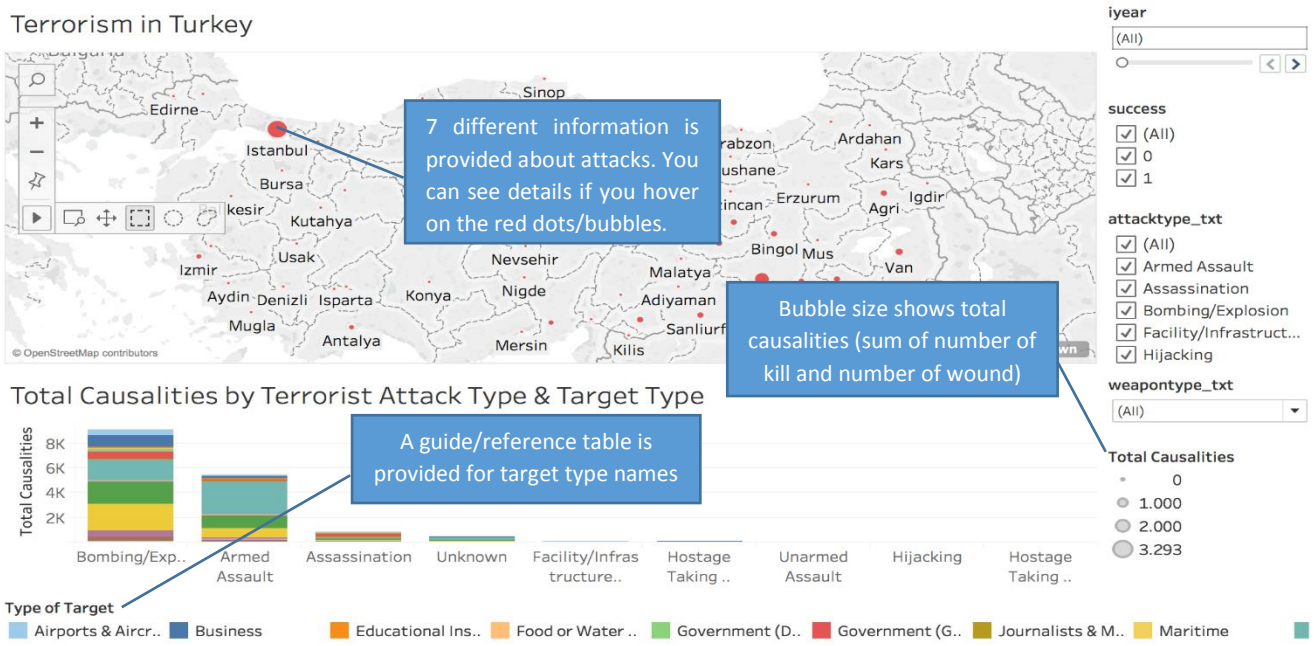
In order to displays phenomena such as social, political, economic or cultural issues with the purpose of revealing patterns and frequencies in the geography where they occur.

- Thematic map: representations of attribute data (quantitative or qualitative) on a base map
- Dot maps:
  Circle is most common shape
  Some maps use the visual variable of shape to differentiate categories
  Color hue is also used
  Effective is portraying relative densities and bad at displaying absolute quantities
  Require equal-area map projection

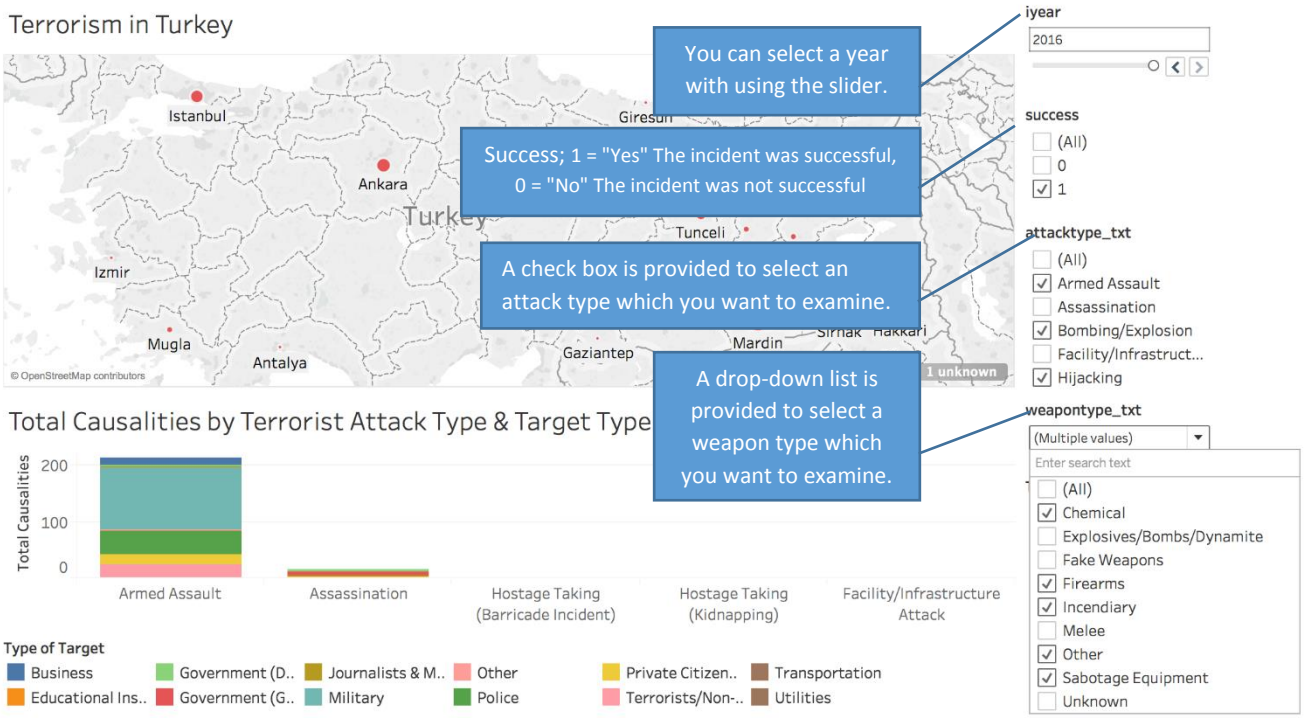
Terrorism_Turkey dataset consists temporal references; years from 1970 till 2016, also consists geographical information (longitude/latitude and city names). I can say that according to above information about data structures, my data structure can appropriate "Spatial Time Series". Also it can suitable both thematic and dot maps. In general all two structure is part of spatial structure.

B. Visualization

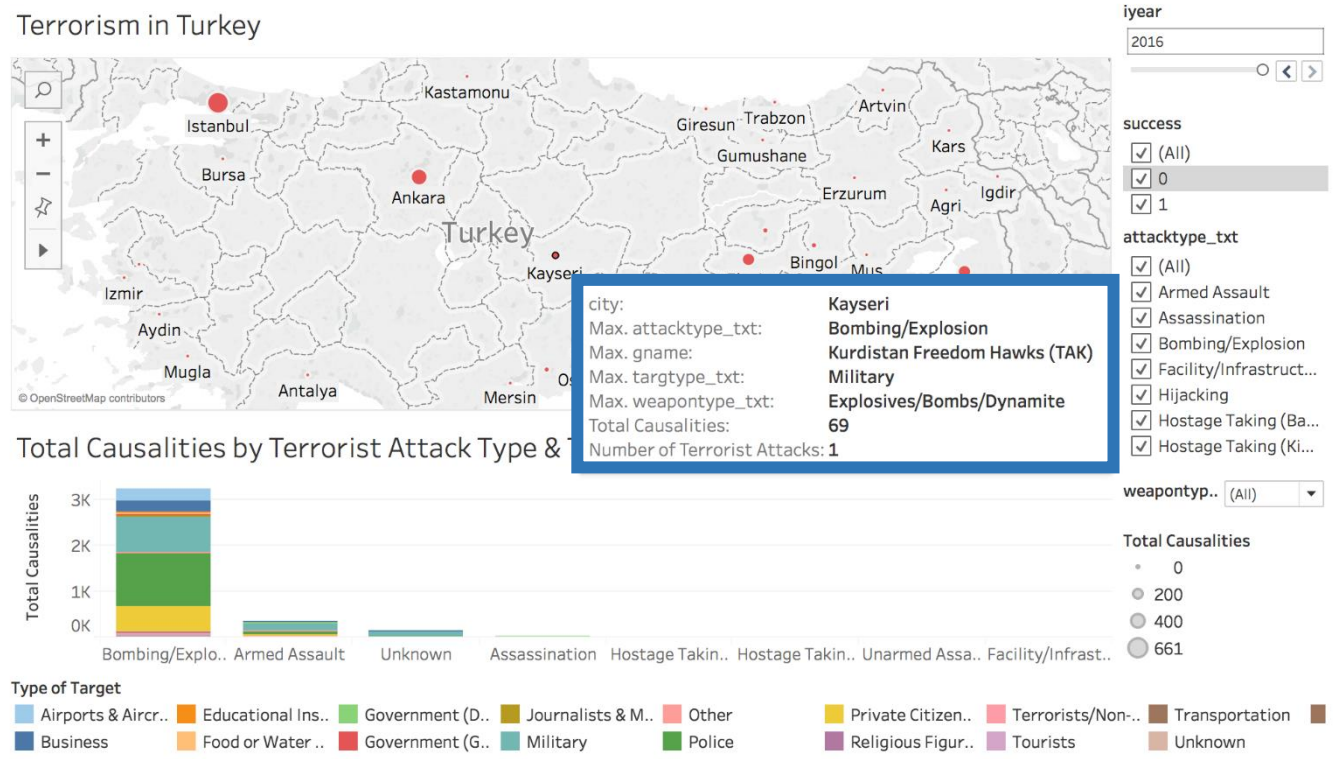
Interactive visualization shows information about over 4K terrorist attacks of various types in Turkey in the time range between 1970 and 2016 on map. Drilling into data with filters, you can have a pattern about the cities that are more frequently victims of terrorist attacks. You can see below the screenshots of the interactive visualization and some quick explanations;



This view is filtered to show a map for the cities which are success terrorist events with selected attack and weapon type in 2016 and details about the terrorist attack. Also stacked bar chart (bottom of visualization) shows with same filters total causalities by attack and target type.



Clicking on a city will filter the map to show incidents specific to that city. To view details on incident, hover on the red dots. 7 different information are shown in detail. They are; city where terrorist attack occur, which type of attack is most used, which terrorist group perform the most attack, who is target the most by terrorist groups, which type of weapon is most used. Additionally number of terrorist attacks and total causalities are shown as measure.



Visualization Technique

Interactive and exploratory data visualizations are allowing users/readers to visualize original data sets in order to draw their own conclusions. So I desired to prepare an interactive visualization.

In my dataset, each city has a latitude and longitude, so the cities can be mapped as a two-dimensional plot, with the minimum and maximum values for the latitude and longitude used for the start and end of the scale in each dimension.

Because of dataset consists geographical information I think map visualization is the best visualization technique for me. Also global mapping models are universally accepted as the key to efficiently organizing spatial datasets. *Link* [5]

I decided an interactive geographical map for my visualization technique from article you sent us to read that help us our project *Information Visualization Techniques Usage Model.* [6]

Table 1. Information Visualization Technique Guide

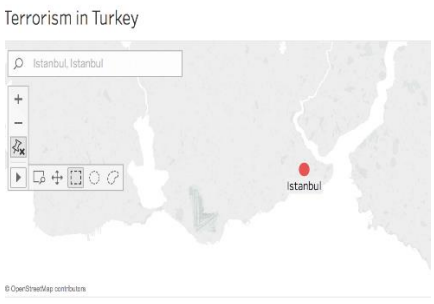
Domain of Usage	Nature of the Data	Function of IV tool	Interaction Techniques	Visualization Techniques
Geographical Information Science	Multidimensional	To explore geographic health data in order to gain insight into the complex and interdependent factors that cause epidemics.	Brushing, Selection, classification, and colour scheme	Scatter plot, bivariate map, time series plot, and parallel coordinate plot.
		For spatiotemporal crime analysis. Used to explore geographical data on criminal activity for crime detection and prevention purposes	Linear and Composite animation functions, temporal legend, zooming	Scatter plots, interactive geographical map
	Geospatial data	To convey research information in an interactive manner to wide audience of different background.	Explore, Selection, View	Interactive geographical map,

Interaction Technique

Interaction provides the user control or explore the data. Also it might cover things like selecting a subset of the data or changing the viewpoint by requiring the data is designed differently.

“Panning and Zooming” helps the user to change from smaller view to a more detailed view of the data (enlarge any section of a graph). I think that more the user zooms into the visualization, the higher the data resolution, so I chose panning and zooming that is also most used/appropriate technique in map visualization.

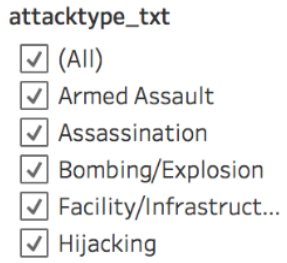
There is a “Search” button which provides convenient to find a city more detailed. For example; if you want to see only terrorist attacks which occurs in İstanbul, you can search and easily see in larger view.



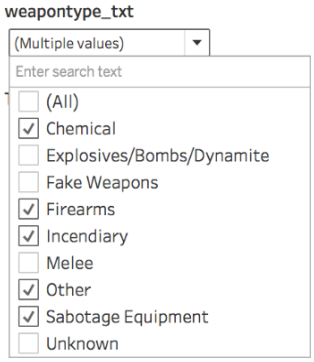
Also I preferred to use slider, drop-down list and check box to user for make some interaction with filtering. If needed, one can change the filter selections on the right to view information about any time period, success information and attack or weapon type.

Slider is suitable for year selection. You can drag the range slider to visualize terrorist events that were occur in 2014.

Success has two options which means 1 = "Yes" The incident was successful, 0 = "No" The incident was not successful, so I chose tick box. Tick box fields are boolean values that represent either true or false.



I preferred a check box for attack type. A check box filter is used to filter to combinations of values in the data column. Only rows containing values with marked check boxes remain in the visualization.



A typical use for checkboxes is to allow the user to show or hide elements on the data visualization by clicking on the corresponding legend item checkbox. I preferred a drop down list for weapon type that used in terror attacks. Selecting all check boxes means no filtering is made, and clearing all check boxes means everything is filtered out.

A legend automatically provides a key to the colors used by data visualization controls such as charts and maps. I can guide the users/readers by including the legend (maybe I call reference table) with target type names and bubble sizes on right and bottom of the visualization. Color legends are very useful for being able to tell exactly what a color references inside of a chart but it takes up a lot of real estate on a dashboard that could be used for other purposes. In my current example, I have a dashboard with two charts, one of them -in bar chart- have a set of color legends.

If users/readers want, they can access -with hover on the dot- a tooltip on each city that shows below information;

city:
Max. attacktype_txt:
Max. gname:
Max. targtype_txt:
Max. weapontype_txt:
Total Causalities:
Number of Terrorist Attacks:

Visual Mappings (Encodings)

Following tables shows and summarizes briefly that which visual encoding can be used which data types in general;

	Categorical	Ordered	Quantitative
Size		✓	✓
Hue	✓		
Saturation		✓	✓
Orientation	✓	✓	✓
Shape	✓		
Texture	✓		

Ranking of Applicability of Properties for Different Data Types
(Mackinlay 88, Not Empirically Verified)

QUANTITATIVE	ORDINAL	NOMINAL
Position	Position	Position
Length	Density	Color Hue
Angle	Color Saturation	Texture
Slope	Color Hue	Connection
Area	Texture	Containment
Volume	Connection	Density
Density	Containment	Color Saturation
Color Saturation	Length	Shape
Color Hue	Angle	Length

Color hue→ I have had an idea of coding the bubbles in colors and labeling them, and that could add some valuable data to the visualization and I applied. I preferred red color to show terror attacks because in general perception red color symbolizes negativity, so I think it is suitable that red dots refers fatality. Color-saturation might be used but saturation is generally suitable for ordered variables which my dataset do not have. In stacked bar chart, I also used color hue to symbolize target types. There are more subcategory for target so I chose this color palette because all target types represents different target.

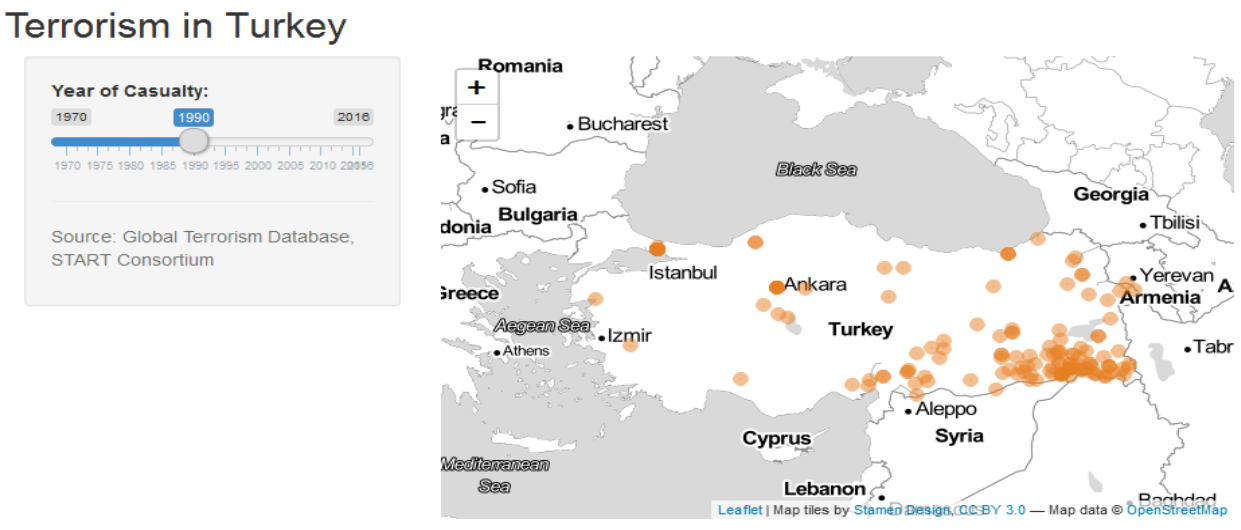
Shape & Size→ I prefer bubble which is most common shape in map visualization. Bubble sizes are shown total causalities - sum of number of killed and number of wounded people- during a terrorist attack. Size encodings generally uses for quantitative variables so I prefer size to show total causalities.

There is no any metaphor used in the visualization.

Visualization Tool

First I tried to visualize my data, Turkey Terrorism, with R Shiny package. I wanted to study with R to provide user an interaction and elasticity with adding some drill-downs. But I am still a beginner in R and I could not give more information. Then I decided to visualize with Tableau. It is true tool for me. Tableau provides hidden opportunities with interactive visualizations with only drag-drop in a short time. I can easily drill down and filter my data on the fly in dashboard tab.

You can access interactive R visualization trial study with following link; https://sevgilit.shinyapps.io/GTD_Shiny/. Also you can reach *twb* file for Tableau interactive visualization in my Github page (additionally it will sent you in attachment)



C. Conclusion/ Lessons learned

In the simplest way, I can say, this project and course had added value to me about data visualization. There are many ways to visualize data, new tools and chart types develop constantly. I learned with this project that I must focus on the principle that a visualization should clarify and summarize the key message rather than confusing and overloading the user/reader with unnecessary information.

D. References

[1] http://benfry.com/phd/dissertation-110323c.pdf
[2] http://www.cs.ubc.ca/labs/imager/tr/2009/NestedModel/NestedModel.pdf
[3] http://vita.had. co.nz/papers/tidy-data.pdf
[4]http://start.umd.edu/gtd/about/
[5] https://journals.lib.unb.ca/index.php/ihr/article/viewFile/23318/27093
[6] http://www.icoci.cms.net.my/proceedings/2013/PDF/PID45.pdf

E. Appendix

R codes are provided below for R Shiny visualization that one used in this project.

```
# This is a Shiny web application. You can run the application by clicking
# the 'Run App' button above.
# Find out more about building applications with Shiny here:
#    http://shiny.rstudio.com/

library(shiny)
library(leaflet)

# Define UI
ui <- fluidPage(
  # Give the page a title
  titlePanel("Terrorism in Turkey"),
  # Generate a row with a sidebar
  sidebarPanel(
    sliderInput(inputId = "year", label = "Year of Casualty:",
                value = 2010, min = 1970, max = 2016, sep=""),
    hr(),
    helpText("Source: Global Terrorism Database, START Consortium")
  ),
  # Show a plot of the generated distribution
  mainPanel(
    leafletOutput("map")
  )
)
# Define server logic

server <- shinyServer(function(input, output) {
  p = readRDS("rds/gtd_data.rds")

  output$map = renderLeaflet({
    leaflet() %>%
      addProviderTiles("Stamen.TonerLite") %>%
      setView(lng = 34.85427, lat = 39.91987, zoom = 5)
  })
  observe({
    leafletProxy("map") %>% clearShapes()%>%
      addCircles(data = p[grepl(input$year, p$year),],color = '#e67e22',weight = 10)
  })
})
# Run the application
shinyApp(ui, server)
```