Project: Data Science

Sean E. Curl

April 4, 2018

## # INTRODUCTION

## Information on more than 170,000 Terrorist Attacks

## The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2016 (with annual updates planned for the future). The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 170,000 cases. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland.

The goal of this project is to explore the aspects and variation of terrorist events around the world from 1970 to 2016. The exploration, we hope, yields some interesting insights into the realm of terrorism such as: Who are the most active terrorist group within a specific region? What is the most devastating type of terrorist attack? When, where, and how many are terrorist attack have occurred?

We also wanted to build an easy to use and visualize Shiny application in R. The Shiny app will have many graphs and plots—some of which will be interactive. We want to provide the user will the choice to select from various data fields, so that they could output plots and visualization which might be of interest to them personally. Ultimately, we want the application to be focused around a user discovery experience.

## # LOAD:

## We wanted to provide a quick overview of the data types of the fields we SELECTED within the Global Terrorism Database. As you can see, the majority of data types were converted by R to be integers, factors, or doubles. The conversion are exactly what we wanted; it’s now easier to any of these variables for visualization and exploration.

# DATA DICTIONARY: http://start.umd.edu/gtd/downloads/Codebook.pdf

### Load and Clean

dataset <- read.csv("C:/Users/PC - Main/Desktop/Data Science/globalterrorismdb\_0617dist.csv")  
  
terrorism <- dataset %>%  
 select("Year" = iyear,   
 "Month" = imonth,   
 "Day" = iday,   
 "Country" = country\_txt,   
 "Region" = region\_txt,   
 "AttackType" = attacktype1\_txt,   
 "Target" = target1,   
 "Killed" = nkill,   
 "Wounded" = nwound,   
 "Summary" = summary,   
 "Group" = gname,   
 "Target\_Type" = targtype1\_txt,   
 "Weapon\_type" = weapsubtype1\_txt,   
 "Motive" = motive,  
 "City" = city,  
 "lat" = latitude,  
 "lon" = longitude,  
 "City" = city) %>%  
   
 mutate(Killed = as.numeric(Killed),  
 Wounded = as.numeric(Wounded), lat = as.numeric(lat), long = as.numeric(lon)) %>%  
 mutate(Casualties = Killed + Wounded) %>%  
 glimpse()

## Observations: 170,350  
## Variables: 19  
## $ Year <int> 1970, 1970, 1970, 1970, 1970, 1970, 1970, 1970, 19...  
## $ Month <int> 7, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...  
## $ Day <int> 2, 0, 0, 0, 0, 1, 2, 2, 2, 3, 1, 6, 8, 9, 9, 10, 1...  
## $ Country <fct> Dominican Republic, Mexico, Philippines, Greece, J...  
## $ Region <fct> Central America & Caribbean, North America, Southe...  
## $ AttackType <fct> Assassination, Hostage Taking (Kidnapping), Assass...  
## $ Target <fct> Julio Guzman, Nadine Chaval, daughter, Employee, U...  
## $ Killed <dbl> 1, 0, 1, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA,...  
## $ Wounded <dbl> 0, 0, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA,...  
## $ Summary <fct> , , , , , 1/1/1970: Unknown African American assai...  
## $ Group <fct> MANO-D, 23rd of September Communist League, Unknow...  
## $ Target\_Type <fct> Private Citizens & Property, Government (Diplomati...  
## $ Weapon\_type <fct> , , , Unknown Explosive Type, , Unknown Gun Type, ...  
## $ Motive <fct> , , , , , To protest the Cairo Illinois Police Dep...  
## $ City <fct> Santo Domingo, Mexico city, Unknown, Athens, Fukou...  
## $ lat <dbl> 18.45679, 19.43261, 15.47860, 37.98377, 33.58041, ...  
## $ lon <dbl> -69.95116, -99.13321, 120.59974, 23.72816, 130.396...  
## $ long <dbl> -69.95116, -99.13321, 120.59974, 23.72816, 130.396...  
## $ Casualties <dbl> 1, 0, 1, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA,...

terrorism %>%   
 map\_dbl(~sum(is.na(.)))

## Year Month Day Country Region AttackType   
## 0 0 0 0 0 0   
## Target Killed Wounded Summary Group Target\_Type   
## 1 9682 15325 0 0 0   
## Weapon\_type Motive City lat lon long   
## 0 0 0 4606 4606 4606   
## Casualties   
## 15826

## # MISSING VALUES

## Let’s see how many missing values there are:

## We can see there’s a lot of different missing values: 1 Target; 9682 Killed, 15325 Wounded; 4606 Lat; 4606 Lon; and 15826 casualties. However, these NA’s are not really a concern for us given that our goal is simply to delve into exploratory analysis of the data. Additionally, it’s important to note that there’s nearly 170,000 records! Therefore, we can afford to have missing values, especially given that the NA fields like Killed, wounded, or casualties might be empty for a reason (e.g. there was no one killed!).

Quickly quoting the data dictionary: “Where there is evidence of fatalities, but a figure is not reported or it is too vague to be of use, this field remains blank. If information is missing regarding the number of victims killed in an attack, but perpetrator fatalities are known, this value will reflect only the number of perpetrators who died as a result of the incident. Likewise, if information on the number of perpetrators killed in an attack is missing, but victim fatalities are known, this field will only report the number of victims killed in the incident.”

## High Impact Areas

country\_highest\_attacks <- terrorism %>%  
 group\_by(Country) %>%  
 summarise(n = n()) %>%  
 arrange(desc(n)) %>%  
 slice(1)  
  
region\_highest\_attacks <- terrorism %>%  
 group\_by(Region) %>%  
 summarise(n = n()) %>%  
 arrange(desc(n)) %>%  
 slice(1)  
  
highest\_killed <- terrorism %>%  
 group\_by(Year, Country, Killed) %>%  
 summarise(max(Killed)) %>%  
 arrange(desc(Killed)) %>%  
 head(1)  
  
paste("Country with Highest Terrorist Attacks:", country\_highest\_attacks$Country)

## [1] "Country with Highest Terrorist Attacks: Iraq"

paste("Region with Highest Terrorist Attacks:", region\_highest\_attacks$Region)

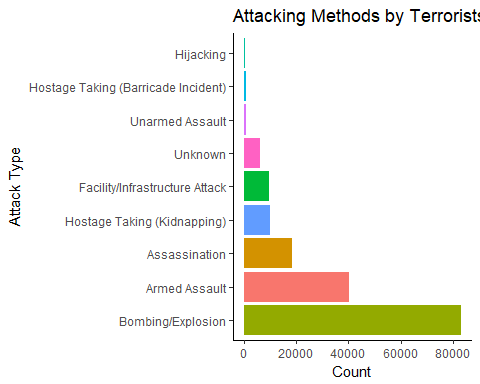
## [1] "Region with Highest Terrorist Attacks: Middle East & North Africa"

paste("Maximum peopled killed in an attack are:", "In", highest\_killed$Year, highest\_killed$`max(Killed)`, "peopled died in",highest\_killed$Country)

## [1] "Maximum peopled killed in an attack are: In 2014 1500 peopled died in Iraq"

## Popular Attack Types

attacking\_methods <- terrorism %>%  
 group\_by(AttackType) %>%  
 summarise(n = n()) %>%  
 arrange(n) %>%  
 ggplot(mapping = aes(reorder(AttackType, -n), n, fill = AttackType)) +   
 geom\_bar(stat = "identity", show.legend = FALSE) +  
 theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust=1)) +  
 ylab("Count") +  
 xlab("Attack Type") +   
 ggtitle("Attacking Methods by Terrorists") +  
 theme\_classic() +  
 coord\_flip()  
  
attacking\_methods



Top Three Attack Types:

Bombing/Explosion: ~82,000

Can include either high or low explosives (including a dirty bomb) but does not include a nuclear explosive device that releases energy from fission and/or fusion, or an incendiary device where decomposition takes place at a much slower rate. If an attack involves certain classes of explosive devices along with firearms, incendiaries, or sharp objects, then the attack is coded as an armed assault only.

Armed Assault: ~40,000

An attack whose primary objective is to cause physical harm or death directly to human beings by use of a firearm, incendiary, or sharp instrument (knife, etc.). Not to include attacks involving the use of fists, rocks, sticks, or other handheld (less-than-lethal) weapons. Also includes attacks involving certain classes of explosive devices in addition to firearms, incendiaries, or sharp instruments. The explosive device subcategories that are included in this classification are grenades, projectiles, and unknown or other explosive devices that are thrown.

Assassination: ~20,000

An act whose primary objective is to kill one or more specific, prominent individuals. Usually carried out on persons of some note, such as high-ranking military officers, government officials, celebrities, etc. Not to include attacks on non-specific members of a targeted group. The killing of a police officer would be an armed assault unless there is reason to believe the attackers singled out a particularly prominent officer for assassination.

## Global Terror Maps

Geo.Country = geocode(unique(terrorism$Country))

Geo.CountryCountry) Geo.Country <- Geo.Country %>% select(Country, lon, lat) %>% arrange(Country)

graph\_data <- merge(Geo.Country, terrorism, by = “Country”) %>% group\_by(Country, lon, lat, Casualties) %>% summarise(total = sum(Casualties))

mp <- NULL mapWorld <- borders(“world”, colour=“gray50”, fill=“gray50”) mp <- ggplot() + mapWorld + geom\_point(data = graph\_data, mapping = aes(x=lon, y=lat, group = Country, size = total, color = “#000099”), show.legend = FALSE) + theme\_classic() + theme(axis.title = element\_blank(), axis.text = element\_blank(), axis.ticks = element\_blank(), axis.line = element\_blank()) + ggtitle(“Casualties by Country since 1970”)

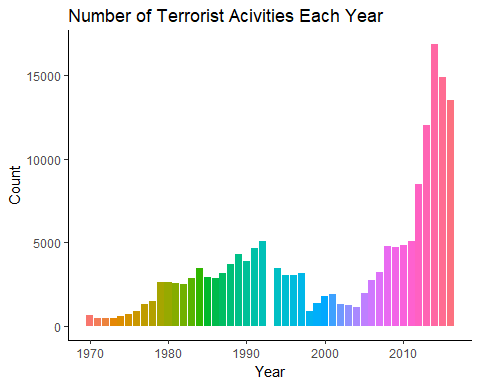
mp

test <- merge(Geo.Country, terrorism, by = “Country”) %>% group\_by(Group) %>% summarise(n = n()) %>% arrange(desc(n)) %>% slice(1:20)

graph\_data1 <- merge(Geo.Country, terrorism, by = “Country”) %>% filter(Group == test$Group) %>% group\_by(Country, lon, lat, Group) %>% summarise(n = n()) %>% arrange(desc(n))

lp <- ggplot() + mapWorld + geom\_point(data = graph\_data1, mapping = aes(x=lon, y=lat, size = n, color = Group)) + theme\_classic() + theme(axis.title = element\_blank(), axis.text = element\_blank(), axis.ticks = element\_blank(), axis.line = element\_blank()) + ggtitle(“Events by the Top 20 Terrorist Organizations since 1970”) lp

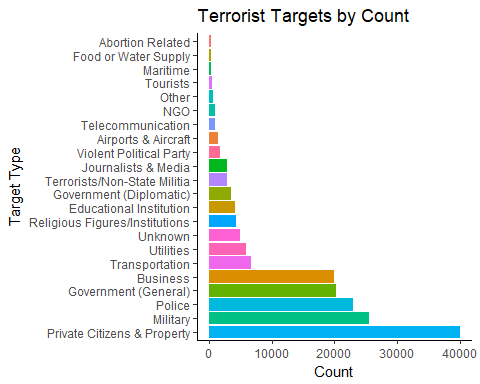
# 1.2 [6]  
count <- terrorism %>%  
 group\_by(Year) %>%  
 summarise(n = n())  
  
ggplot(count,   
 aes(x = Year,   
 y = n, fill = factor(Year))) +  
 geom\_bar(stat = "identity", show.legend = FALSE) +  
 xlab("Year") +  
 ylab("Count") +  
 ggtitle("Number of Terrorist Acivities Each Year") +  
 theme\_classic()



Top Three Years by Count:

[NEED SUM(MAX(YEAR)) COUNT]

# 1.2 [11]  
count <- terrorism %>%  
 group\_by(Target\_Type) %>%  
 summarise(n = n())  
  
ggplot(count,   
 aes(reorder(x = Target\_Type, -n),   
 y = n, fill = factor(Target\_Type))) +  
 geom\_bar(stat = "identity", show.legend = FALSE) +  
 xlab("Target Type") +  
 ylab("Count") +  
 ggtitle("Terrorist Targets by Count") +  
 theme(axis.text.x=element\_text(angle = 90, hjust =1)) +  
 theme\_classic() +  
 coord\_flip()



Top Four Targets by Count:

Private Citizens and Property: ~40,000

This value includes attacks on individuals, the public in general or attacks in public areas including markets, commercial streets, busy intersections and pedestrian malls.

Also includes ambiguous cases where the target/victim was a named individual, or where the target/victim of an attack could be identified by name, age, occupation, gender or nationality. This value also includes ceremonial events, such as weddings and funerals. The GTD contains a number of attacks against students. If these attacks are not expressly against a school, university or other educational institution or are not carried out in an educational setting, these attacks are coded using this value.

Also, includes incidents involving political supporters as private citizens and property, provided that these supporters are not part of a government-sponsored event.

Finally, this value includes police informers. Does not include attacks causing civilian casualties in businesses such as restaurants, cafes or movie theaters (these categories are coded as “Business”).

Military: ~26,000

Includes attacks against military units, patrols, barracks, convoys, jeeps, and aircraft. Also includes attacks on recruiting sites, and soldiers engaged in internal policing functions such as at checkpoints and in anti-narcotics activities. This category also includes peacekeeping units that conduct military operations (e.g., AMISOM)

Excludes attacks against non-state militias and guerrillas, these types of attacks are coded as “Terrorist/Non-state Militias”.

Police: ~24,000

This value includes attacks on members of the police force or police installations; this includes police boxes, patrols headquarters, academies, cars, checkpoints, etc. Includes attacks against jails or prison facilities, or jail or prison staff or guards.

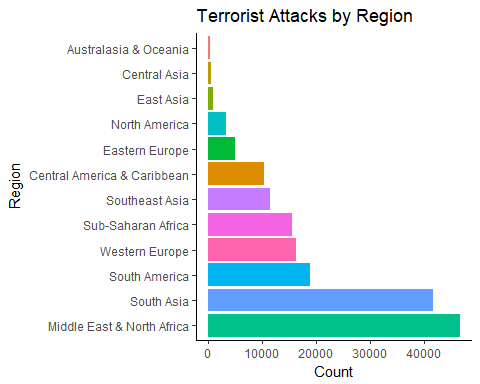
Government (General) / Business: ~22,000

Any attack on a government building; government member, former members, including members of political parties in official capacities, their convoys, or events sponsored by political parties; political movements; or a government sponsored institution where the attack is expressly carried out to harm the government.

This value includes attacks on judges, public attorneys (e.g., prosecutors), courts and court systems, politicians, royalty, head of state, government employees (unless police or military), election-related attacks, or intelligence agencies and spies.

This value does not include attacks on political candidates for office or members of political parties that do not hold an elected office (these attacks are captured in “Private Citizens and Property”).

# 1.4 [11]  
count <- terrorism %>%  
 group\_by(Region) %>%  
 summarise(n = n())  
  
ggplot(count,   
 aes(reorder(x = Region, -n),   
 y = n, fill = factor(Region))) +  
 geom\_bar(stat = "identity", show.legend = FALSE) +  
 xlab("Region") +  
 ylab("Count") +  
 ggtitle("Terrorist Attacks by Region") +  
 theme(axis.text.x=element\_text(angle = 90, hjust =1)) +  
 theme\_classic() +  
 coord\_flip()



Top Four Terrorist Attacks by Region:

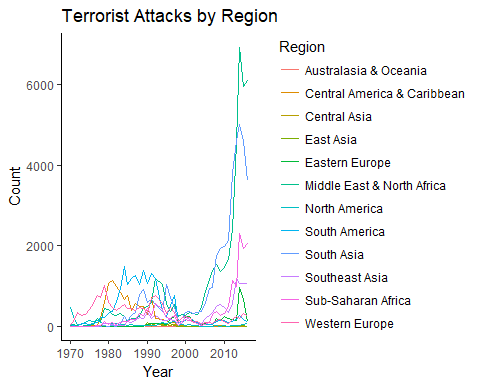
Middle East & North Africa: ~44,000

South Asia: ~42,000

South America: ~19,000

Western Europe: ~16,000

# 1.4 [12]  
count <- terrorism %>%  
 group\_by(Region, Year) %>%  
 summarise(n = n())  
  
ggplot(count,   
 aes(x = Year,   
 y = n, colour = Region)) +  
 geom\_line(stat = "identity") +  
 xlab("Year") +  
 ylab("Count") +  
 ggtitle("Terrorist Attacks by Region") +  
 theme(axis.text.x=element\_text(angle = 90, hjust =1)) +  
 theme\_classic()



Terrorist Attacks by Region & Year:

Middle East & North Africa: ~44,000

Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Libya, Morocco, North Yemen, Qatar, Saudi Arabia, South Yemen, Syria, Tunisia, Turkey, United Arab Emirates, West Bank and Gaza Strip, Western Sahara, Yemen

South Asia: ~42,000

Afghanistan, Bangladesh, Bhutan, India, Maldives, Mauritius, Nepal, Pakistan, Sri Lanka

South America: ~19,000

Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Falkland Islands, French Guiana, Guyana, Paraguay, Peru, Suriname, Uruguay, Venezuela

Western Europe: ~16,000

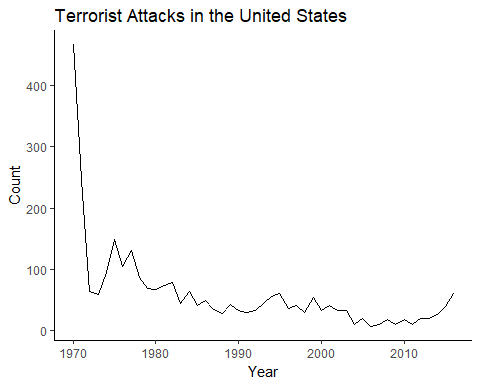
Andorra, Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Gibraltar, Greece, Iceland, Ireland, Italy, Luxembourg, Malta, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, Vatican City, West Germany (FRG)

# 1.5 [18]

count <- terrorism %>% filter(Group == test$Group & Group != “Unknown”) %>% group\_by(Group, Year) %>% summarise(n = n())

ggplot(count, aes(x = Year, y = n, colour = Group)) + geom\_line(stat = “identity”) + xlab(“Year”) + ylab(“Count”) + ggtitle(“Terrorist Attacks by Group”) + theme(axis.text.x=element\_text(angle = 90, hjust =1)) + theme\_classic()

count <- terrorism %>%  
 filter(Country == "United States") %>%  
 group\_by(Year) %>%  
 summarise(n = n())  
  
ggplot(count,   
 aes(x = Year,   
 y = n)) +  
 geom\_line(stat = "identity") +  
 xlab("Year") +  
 ylab("Count") +  
 ggtitle("Terrorist Attacks in the United States") +  
 theme(axis.text.x=element\_text(angle = 90, hjust =1)) +  
 theme\_classic()



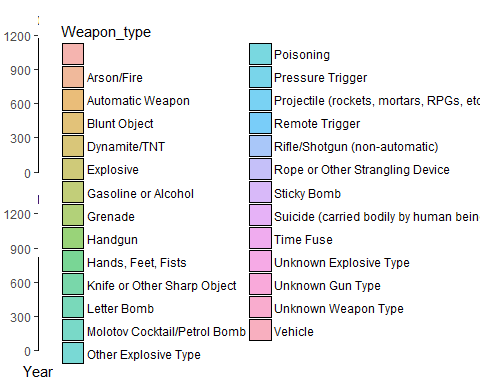
NEED A SUM COUNT SHOW TOP YEARS

NEED A SUM COUNT BY STATE

highest\_killed\_USA <- terrorism %>%  
 filter(Country == "United States") %>%  
 group\_by(Year, City, Casualties) %>%  
 summarise(max(Casualties)) %>%  
 arrange(desc(Casualties)) %>%  
 head(10)  
  
data.frame(highest\_killed\_USA)

## Year City Casualties max.Casualties.  
## 1 2001 New York City 8749 8749  
## 2 2001 New York City 8747 8747  
## 3 1995 Oklahoma City 818 818  
## 4 1984 The Dalles 751 751  
## 5 2001 Arlington 295 295  
## 6 2013 West 166 166  
## 7 2013 Boston 134 134  
## 8 2013 Boston 133 133  
## 9 1996 Atlanta 111 111  
## 10 2016 Orlando 103 103

terrorism %>%  
 filter(Country == "United States") %>%  
 mutate(bKilled = as.factor(ifelse(Killed > 1, 1, 0))) %>%  
 na.omit() %>%  
 ggplot(aes(Year, fill = Weapon\_type)) +  
 geom\_density(alpha = 0.5) +  
 scale\_x\_log10() +  
 facet\_wrap(~ bKilled, ncol = 1)



NEED SUM COUNT BY STATE

NEED SUM COUNT BY YEAR