

# Final Project

*Victor*

*October 21, 2017*

## Memo on final graphic

### Executive summary

The terrorist attack of 9/11 sounds like the epic act of terrorism and often makes one feel like America is the country with the most acts of terrorism. Given this data, I was curious to explore acts of terrorism in the US but it made sense in light of the aforementioned rationale to have a perspective of global terrorism and picture terrorism in America within the context of similar acts in other parts of the world.

Contrary to what might appear intuitive, America is a pretty safe environment comparative to other regions. It however that most vulnerable terrorism in the world are Iraq, Pakistan, Afghanistan, India, Philippines, Thailand, Somalia, Russia, North Africa, etc.

Tracking terrorism in the US shows a sharp decrease between pre and post 9/11 numbers. One trend that however remains a steadily increasing means of terrorism is facility/infrastructure attack. Bombing/Explosion terrorism comes second to facility/infrastructure attack. It appears that given increased security measures by the US government after 9/11, facility/infrastructure attack serves as an easy alternative to explosive attacks by terrorists.

A further look at these two forms of attack revealed that bombing/explosion attacks mostly occurred in May while facility/infrastructure attack often occurred on a Monday.

What is (are) your main question(s)? What is your story? What does the final graphic show?

Going into this analysis, I was driven by curiosity on answers to four basic questions 1. Is the United States the country with most terrorism? 2. What has been the trend in terrorism in America, pre and post 9/11? 3. What is the distribution of terrorism across states; in essence where are the most vulnerable spots? 4. Which months and days are terrorists likely to attack?

These are answers that Americans seek as they reflect how safe they are. The story behind this is just about breaking down terrorism in the US and informing people about the change that has occurred since 9/11 and also possibly about the new form that terrorism might have taken. Choosing career, retirement and vacation can all be informed by a knowledge that answers to these questions might reveal.

My final graphic is titled "breaking down terrorism in the United States". It summarizes my findings by showing a global map highlighting terrorism across the globe. An arrow directs the reader to a line graph that shows the trend of the top four acts of terrorism in the US, pre and post 9/11. A more detailed map of the US is also presented to indicate vulnerable states based on terrorism recorded in these states post 9/11. Finally there are two bar charts showing the most likely day of the week and month of attack respectively for each of the top two forms of terrorism in the US.

### Data background

The data was obtained from the National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland. It was structured into attacks recorded at a particular geographic coordinate in a city/town of a state, country, region with columns on the attack type, target, summary of incident, etc. There were a lot of missing data as well. All together it was a pretty good data set.

### Data cleaning

To eliminate any bias (in the sense that most acts of terrorism in the 70's and 80's were in pursuit of civil rights, independence and other human rights activities) the data for plotting terrorism on a global scale

was cleaned to reflect attacks that have occurred since 2000 in which case the terrorist incidents have less likelihood of being for any of the reasons above.

Also, in many instances, some data had to be eliminated because some values were missing. For instance in creating the data for plotting likely day or month for attack, data that contained zero (0) for day or month was removed. In a peculiar case of plotting the US map, Hawaii and Alaska were purposefully taken off to reflect mainland USA. The rest involved sorting by according to dates and other categories to capture the relevance of the questions.

Individual figures

Figure 1

This figure is called the Global map. It shows terrorism across the globe. First of all I imported the data into R, filtered the data for years 2000 and beyond and got a count of how many times terrorism is recorded in each country. I then loaded the world map, joined it to the framed data and plotted it after sorting out Antarctica.

Following the CRAP principles, I chose a contrast that brought out the disparity between countries at a glance. To make the map truthful, I had to regroup the data by attacks below, within and above certain threshold in order to adjust the color scale to reflect the differences among countries effectively

Figure 2

For this map (a line graph showing trend for various forms of terrorism in the US, pre and post 9/11), I first did a cumulative sum of attacks by forms of terrorism in the US, realized that it will be helpful to plot just the top four forms of terrorism to avoid the graph being too busy (i.e. to ensure meaningful proximity between lines). I chose bright colors for beauty and visibility. I also set the x axis label to show beginning year, 2002 and ending year so that my audience can picture the change in trend, pre and post 9/11. This made the line graph more meaningful.

Figure 3

Creating the chart showing attacks in US followed the same approach as I did for the global map except for the nuances of a US shape to create the map instead and removing Alaska and Hawaii to get the map to reflect mainland USA.

Figure 4

My final graphs show which day and month it's likely to have any of the first two popular kind of attacks. The process was keenly embedded on the CRAP principles to ensure the result was truthful. As such, the data frame that was created eliminated all months and days that were entered as zero. To ensure alignment, the y axis labels were reversed and the legend was also deleted since it was the same as the y axis labels. Very beautiful colors were used to maintain contrast and repetition for that of the months as well.

The most important uncertainty I hope the chart communicates to my audience is that except for May being the most likely month for explosive attacks, there is no predictability as to which day or month an attack does happen for these two forms of terrorism.

Final figure

To tell the story effectively, I chose to start with the global terrorism map, have an arrow point from North America to the map showing the trend of terrorism pre and post 9/11 and another arrow pointing from the trend in the post 9/11 to show how that translates to the distribution of the terrorism observed in the US post 9/11. It was apparent that California be called out since it was the most vulnerable state. Finally, I have at the bottom of the figure, the two bar charts showing the likely day and month for attacks.

I chose to title the final output "breaking down terrorism in the US" because that is the story it tells to my audience. The colors, fonts and even the arrows were chosen to add beauty to the figure but more importantly to maintain simplicity, communicate truth and present uncertainty as well.

# Breaking Down Terrorism in the US

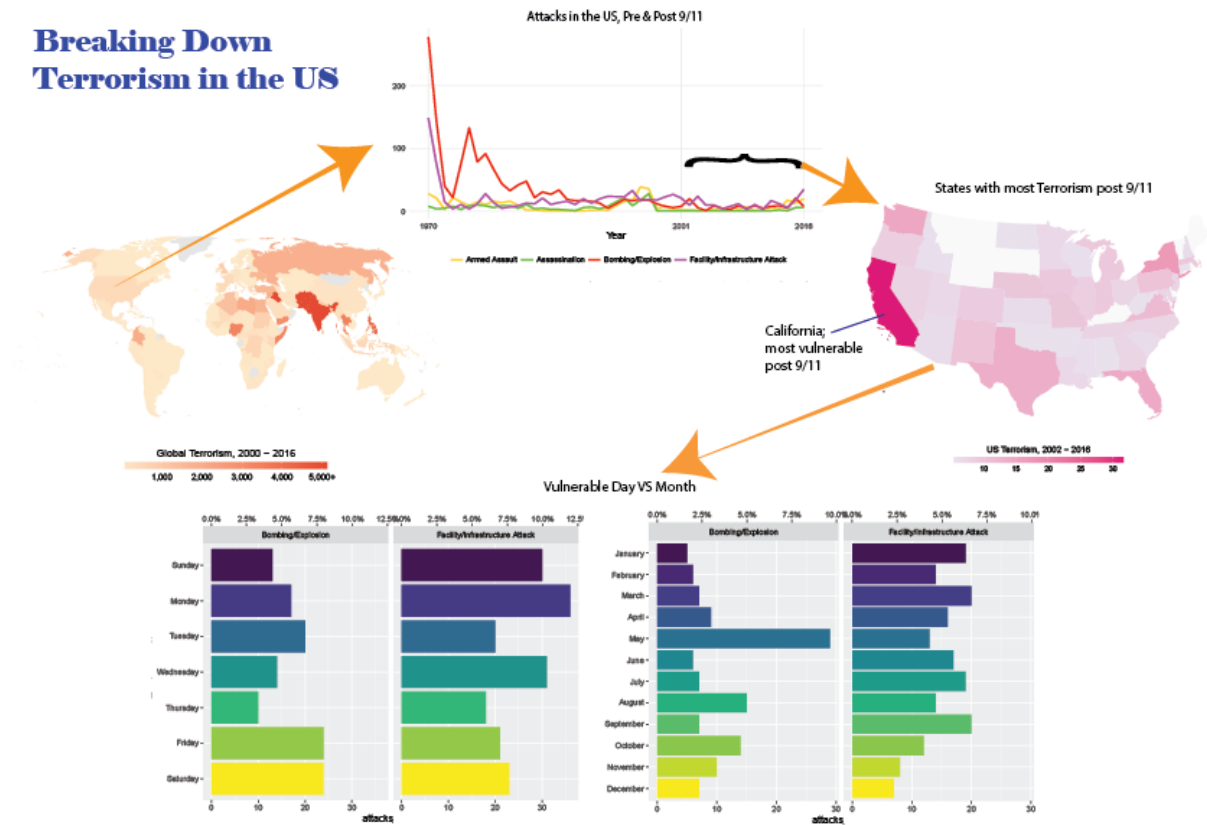


Figure 1: Caption for your final figure

```

library(tidyverse)
library(countrycode)
library(lubridate)
library(scales)
library(sf)
library(gapminder)
library(forcats)
library(ggrepel)
windowsFonts(`Roboto Condensed` = windowsFont("Roboto Condensed"))

```

Chart 1: Map showing global terrorism from 2000 to 2016

```
RawData <- read_csv("data/globalterrorismdb_0617dist.csv")
```

```

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   eventid = col_double(),
##   iyear = col_integer(),
##   imonth = col_integer(),
##   iday = col_integer(),
##   extended = col_integer(),
##   country = col_integer(),
##   region = col_integer(),
##   latitude = col_double(),
##   longitude = col_double(),
##   specificity = col_integer(),
##   vicinity = col_integer(),
##   crit1 = col_integer(),
##   crit2 = col_integer(),
##   crit3 = col_integer(),
##   doubtterr = col_integer(),
##   alternative = col_integer(),
##   multiple = col_integer(),
##   success = col_integer(),
##   suicide = col_integer(),
##   attacktype1 = col_integer()
##   # ... with 44 more columns
## )

## See spec(...) for full column specifications.

## Warning in rbind(names(probs), probs_f): number of columns of result is not
## a multiple of vector length (arg 1)

## Warning: 246 parsing failures.
## row # A tibble: 5 x 5 col      row      col      expected actual expected  <int>    <chr>
## ... ..
## See problems(...) for more details.

global_attack <- RawData %>%
  filter(iyear >= 2000) %>%
  group_by(country_txt) %>%
  summarize(attacks = n()) %>%
  mutate(iso_a3 = countrycode(country_txt, "country.name", "iso3c")) %>%
  filter(!is.na(iso_a3)) %>%

```

```

mutate(truncated = ifelse(attacks>5000, 5000, attacks))

## Warning in countrycode(country_txt, "country.name", "iso3c"): Some values were not matched unambiguou
world_shapes <- st_read("data/ne_50m_admin_0_countries.shp",
  stringsAsFactors = FALSE)

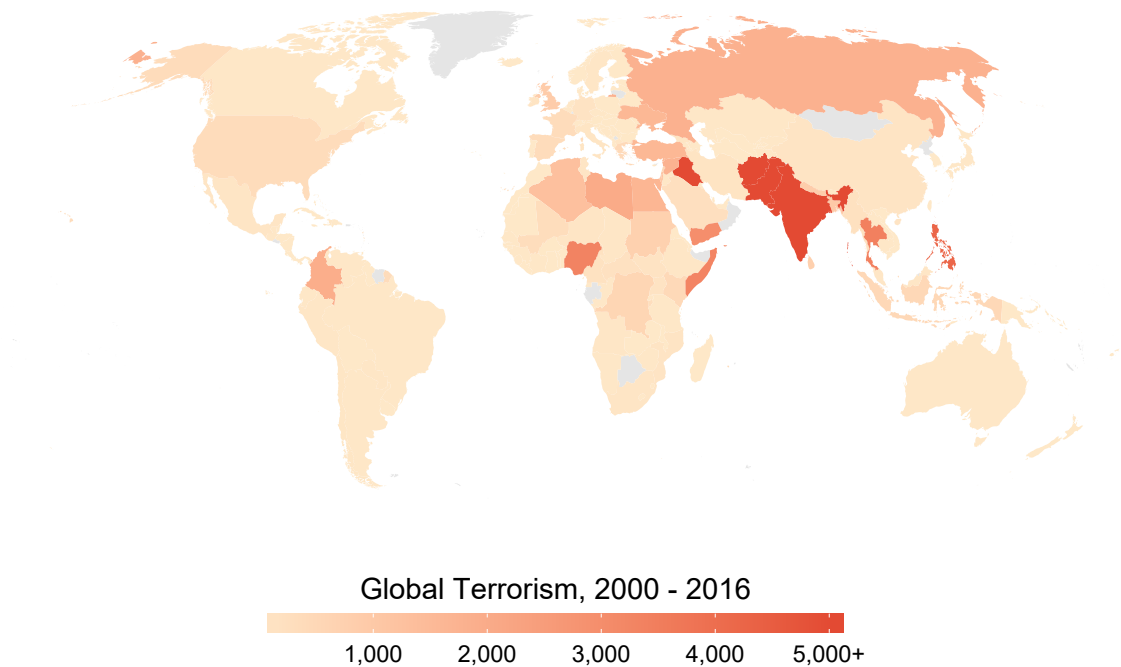
## Reading layer `ne_50m_admin_0_countries' from data source `C:\Users\ODAVIC\Desktop\Semester 3\Data V
## Simple feature collection with 241 features and 63 fields
## geometry type: MULTIPOLYGON
## dimension: XY
## bbox: xmin: -180 ymin: -89.99893 xmax: 180 ymax: 83.59961
## epsg (SRID): 4326
## proj4string: +proj=longlat +datum=WGS84 +no_defs
attack_map <- world_shapes %>%
  left_join(global_attack, by = "iso_a3") %>%
  filter(iso_a3 != "ATA")

Global_map <- ggplot(attack_map) +
  geom_sf(aes(fill = truncated), color = NA) +
  coord_sf(crs = st_crs(54030)) +
  scale_fill_gradient(low = "#fee8c8", high = "#e34a33", na.value = "grey90",
    labels = c("0", "1,000", "2,000", "3,000", "4,000", "5,000+")) +
  guides(fill = guide_colorbar(title.position = "top",
    title.hjust = "0.5",
    title = "Global Terrorism, 2000 - 2016",
    barwidth = 15, barheight = 0.5)) +

  theme_void() +
  theme(legend.position = "bottom")

Global_map

```



```
#ggsave(Global_map, filename = "Output/Global_map.pdf", width = 7, height = 5)
```

Chart 2: Evolution chart showing trend of attack in the United States

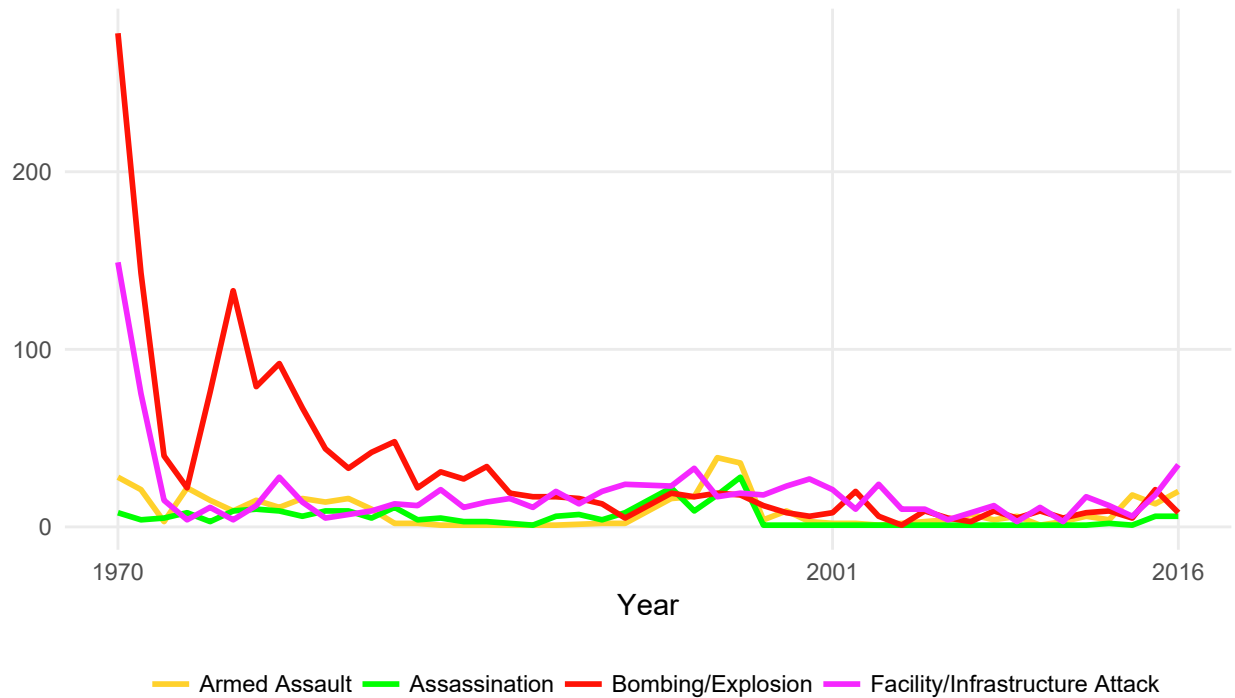
```
attack_evolution <- RawData %>%
  filter(region_txt == "North America", attacktype1_txt %in% c("Bombing/Explosion", "Facility/Infrastruc
  #filter(iyear >= 2002) %>%
  group_by(iyear,attacktype1_txt) %>%
  summarize(attacks = n())
```

```
Evolution <- ggplot(attack_evolution, aes(x = iyear, y = attacks, color = attacktype1_txt))+
  geom_line(size = 1) +
  scale_y_continuous() +
  scale_x_continuous(breaks = c(1970, 2001,2016)) +
  scale_color_manual(values = c("#FFD12C", "green", "#FF1306", "#F527FF"))+
  labs(x = "Year", y = NULL, title = "Terrorism in US, Pre & Post 9/11 ", subtitle = "What is the most
  theme_minimal() +
  theme(legend.position = "bottom",
        legend.title = element_blank(),
        panel.grid.minor = element_blank(),
        plot.caption = element_text(size = 8),
        text = element_text(),
        plot.title = element_text(),
        plot.subtitle = element_text())
```

Evolution

## Terrorism in US, Pre & Post 9/11

What is the most likely form of terrorism today?



Source: National Consortium for the Study of Terrorism and Responses to Terrorism (START), University of Maryland

```
#ggsave(Evolution, filename = "Output/Evolution.pdf", width = 7, height = 5)
```

Chart 3: Taking a closer look at vulnerable spots in the US after 9/11

```
US_attack <- RawData %>%
  filter(iyear >= 2002, country_txt == "United States") %>%
  group_by(provstate, country_txt) %>%
  summarize(attacks = n()) %>%
  mutate(iso_a3 = countrycode(country_txt, "country.name", "iso3c")) %>%
  filter(!is.na(iso_a3)) %>%
  rename(NAME = provstate)

US_shapes <- st_read("data/cb_2016_us_state_20m.shp",
  stringsAsFactors = FALSE) %>%
  filter(!(NAME %in% c("Alaska", "Hawaii", "Puerto Rico"))) )

## Reading layer `cb_2016_us_state_20m' from data source `C:\Users\ODAVIC\Desktop\Semester 3\Data Viz\b
## Simple feature collection with 52 features and 9 fields
## geometry type: MULTIPOLYGON
## dimension: XY
## bbox: xmin: -179.1743 ymin: 17.91377 xmax: 179.7739 ymax: 71.35256
## epsg (SRID): 4269
## proj4string: +proj=longlat +datum=NAD83 +no_defs

attack_map <- US_shapes %>%
  left_join(US_attack, by = "NAME")
```

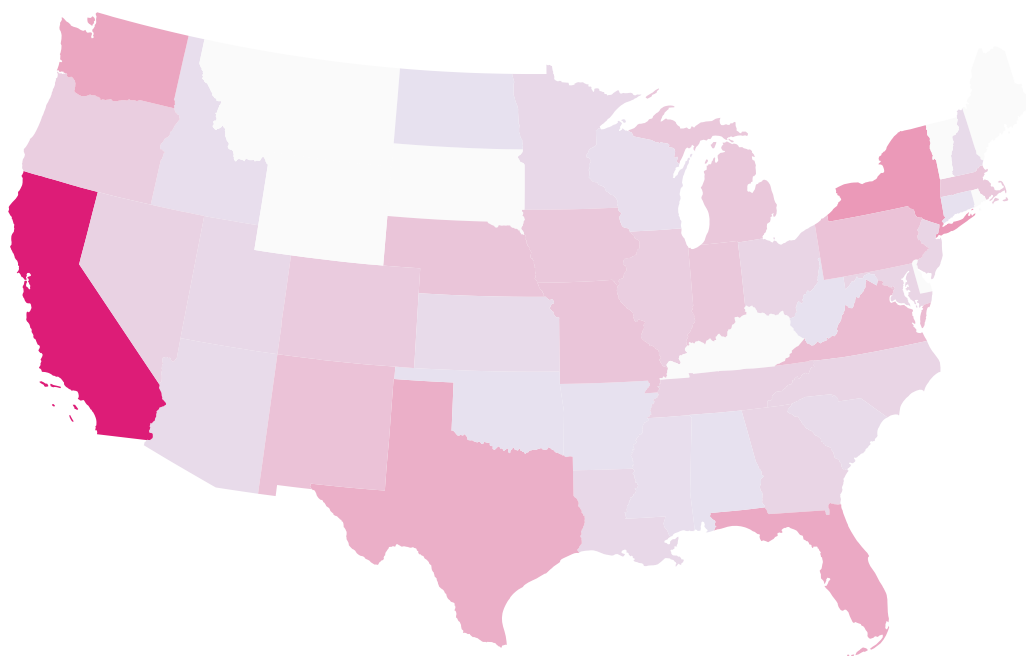
```

US_map <- ggplot(attack_map) +
  geom_sf(aes(fill = attacks), color = NA) +
  coord_sf(crs = st_crs(102003)) +
  scale_fill_gradient(low = "#e7e1ef", high = "#dd1c77", na.value = "grey98",
    labels = c("5", "10", "15", "20", "25", "30")) +
  guides(fill = guide_colorbar(title.position = "top",
    title.hjust = "0.5",
    title = "US Terrorism, 2002 - 2016",
    barwidth = 15, barheight = 0.5)) +

  theme_void() +
  theme(legend.position = "bottom")

```

US\_map



US Terrorism, 2002 - 2016



```

#ggsave(US_map, filename = "output/US_map.pdf", width = 7, height = 5)

```

Chart 4: Bar Chart showing the likely weekday for types of terrorist attack

```

attack_day <- RawData %>%
  filter(iyear >= 2002, iday != 0,
    attacktype1_txt %in% c("Bombing/Explosion",
      "Facility/Infrastructure Attack")) %>%
  mutate(cleandate = ymd(paste(iyear, imonth, iday, sep = "-"))) %>%
  mutate(weekday = wday(cleandate, label = TRUE, abbr = FALSE)) %>%
  group_by(region_txt, attacktype1_txt, weekday) %>%
  summarise(attacks_pday = n()) %>%
  filter(region_txt == "North America") %>%

```



```

ungroup() %>%
mutate(percent = attacks_pday / sum(attacks_pday))

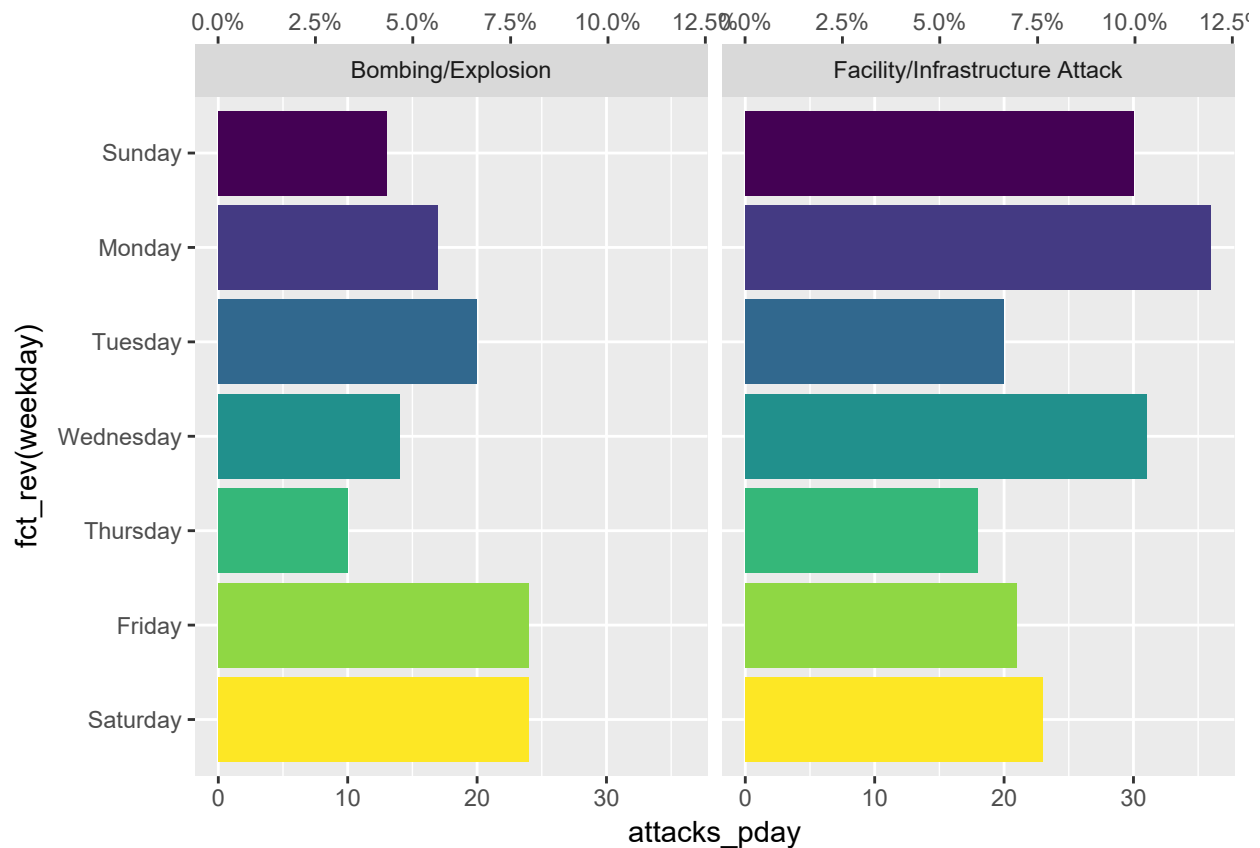
```

```

Day <- ggplot(attack_day, aes(x = fct_rev(weekday), y = attacks_pday, fill = weekday)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_y_continuous(sec.axis = sec_axis(~ . / sum(attack_day$attacks_pday), labels=scales::percent)) +
  facet_wrap(~attacktype1_txt) +
  coord_flip()+
  guides(fill = FALSE)

```

Day



```

#ggsave(Day, filename = "output/Day.pdf", width = 7, height = 5)

```

Chart 5: Chart showing terrorism per month

```

attack_month <- RawData %>%
  filter(iyear >= 2002, imonth != 0,
  attacktype1_txt %in% c("Bombing/Explosion",
                        "Facility/Infrastructure Attack")) %>%
  mutate(cleandate = ymd(paste(iyear,imonth,iday,sep = "-"))) %>%
  mutate(Month = month(cleandate, label = TRUE, abbr = FALSE))%>%
  group_by(region_txt, attacktype1_txt, Month) %>%
  summarise(attacks_pday = n()) %>%
  filter(region_txt == "North America", !is.na(Month)) %>%
  ungroup() %>%
  mutate(percent = attacks_pday / sum(attacks_pday))

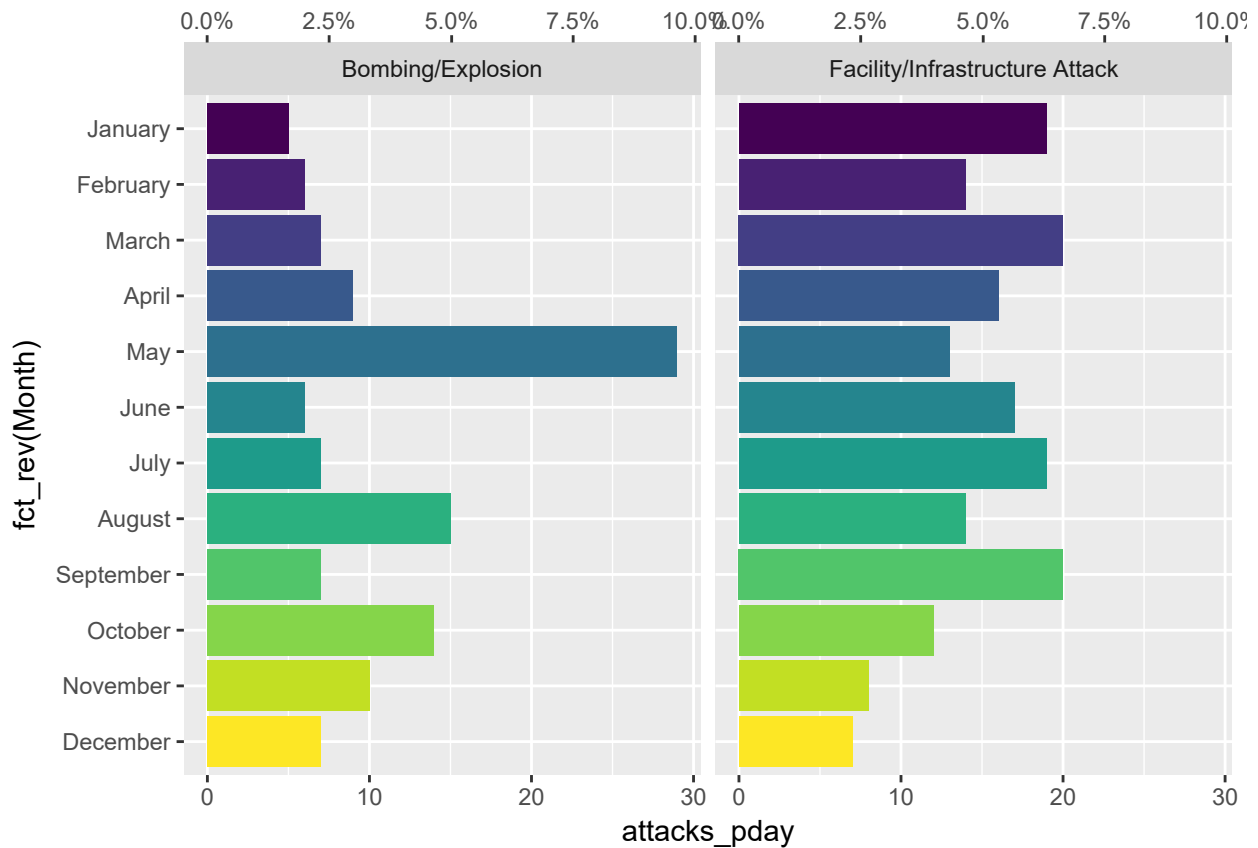
```

```
## Warning: 73 failed to parse.
```

```
#mutate(highlight = ifelse(Month == "May", TRUE, FALSE))
```

```
Monthly <- ggplot(attack_month, aes(x = fct_rev(Month), y = attacks_pday, fill = Month)) +  
  geom_bar(stat = "identity", position = "dodge") +  
  scale_y_continuous(sec.axis = sec_axis(~ . / sum(attack_month$attacks_pday), labels=scales::percent))  
  #scale_fill_manual(values = c())+  
  facet_wrap(~attacktype1_txt)+  
  coord_flip()+  
  guides(fill = FALSE)
```

Monthly



```
#ggsave(Monthly, filename = "output/Monthly.pdf", width = 7, height = 5)
```