SUICIDE BOMBING & DRONE STRIKES

PAKISTAN'S STORY

FIT5147 R Project Report

APRIL 28, 2018

MONASH UNIVERSITY
Syed Ali Alim Rizvi - 28984773

Table of Contents

Table of Figures

INTRODUCTION

Terrorism has plagued major countries for over a decade and many countries have started experiencing acts of terrorism now more than ever. Terrorism is the general term given to the intentional act of violence to cause fear in the hearts of people for a specific gain, whether it be political, religious, financial etc.

The September 11 attacks in New York City and Washington DC gave rise to the United States 'War on Terror'. This term was originally against Al Qaeda and the US-led mission in Afghanistan. However due to the rejection of surrendering the Arab Mujahidin by the Taliban leader after 9/11, President Bush of the United States declared no difference between Al Qaeda and Taliban. Due to the high impact of this mission in Afghanistan, many Taliban and Al Qaeda members alike started crossing into Pakistan's northern regions through the border. This caused high pressure by the US on Pakistan's government to intervene and retaliate. The United States from the year 2004 started its target on militant groups in the two northern regions of Pakistan, Khyber Pakhtunkhwa (KPK) and the Federally Administered Tribal Areas (FATA).

Pakistan has suffered in both micro as well as macro levels due to the impact of terrorism and drone strikes conducted. The terrorist attacks in Pakistan are mainly done through suicide attacks. Hence this report focuses only on attacks done by terrorist groups by suicide bombing. The report will use publicly available data to identify the following:

- Is there a relation between the suicide bombing attacks and the number of drone attacks in Pakistan? What is this relation? What are the casualties?
- Does the lack of safety or security concerns affect the emigration of citizens out of Pakistan? Do citizens consider emigration for survival?
- Is there any trend that can be identified in the suicide bombing attacks?

As for any data science project, the task of exploring and visualising data starts with data pre-processing (Wrangling). This is done to make sure there are no inconsistencies in the data and the data is cleaned and in the format that is required to do the analysis.

DATA WRANGLING/CHECKING

The data for Suicide bombing attacks[i] as well as the drone attacks[ii] were obtained from the Kaggle website. The data for emigration[iii] counts by occupation, province, and skill type were obtained from Pakistan's Bureau of Emigration & Overseas Employment. The data for the latitudes and longitudes for the emigration data is taken from the dataset publishing language website.[iv]

Almost all the wrangling is done using R except for two instances, extracting missing Hijri Dates from suicide data, and minor corrections and data transposing for Emigration Data; both of which were done in Excel.

Suicide Attack Data

Most of the wrangling and data checking was performed on this dataset. It had plenty of missing values as well as included data that was quite inconsistent.

Importing the data: The data included Islamic Dates for suicide attacks which included characters that were not common for UTF-8, hence, the data was imported using file encoding of 'Latin-1'. The structure for the imported data is shown in Figure 1.

Cleaning Gregorian Dates: The date attribute had to be converted as the Date type in R. Hence the format was check so that it could be converted correctly. The dates were in different formats. **1.** the dates were in long format but there were written in two different ways, 'Friday-April 1-2011' and 'Tuesday-Oct-25-2016', with a space in the middle or a hyphen. **2.** the Months were either in the abbreviated forms such as 'Feb' or full form such as 'February'.

```
'data.frame':   496 obs. of  26 variables:
 $ S.                     : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Date                   : Factor w/ 455 levels "Friday-April 1-2011",..: 267 151 440 41 38 291 192 61 145 318 ...
 $ Islamic.Date           : Factor w/ 454 levels "05 Zeqad 1437 A.H.",..: 252 27 256 356 367 334 427 133 48 98 ...
 $ Blast.Day.Type         : Factor w/ 4 levels "","Holiday","Weekend",..: 2 4 4 4 4 2 4 4 4 4 ...
 $ Holiday.Type           : Factor w/ 16 levels "","Ashura","Ashura Holiday",..: 16 1 1 1 1 5 1 1 1 1 ...
 $ Time                   : Factor w/ 217 levels "","10:00:00 AM",..: 209 209 146 24 209 60 1 39 196 209 ...
 $ City                   : Factor w/ 93 levels "Attock ","ATTOCK",..: 25 30 31 30 67 69 69 31 31 58 ...
 $ Latitude               : num  33.7 25 25 25 30.2 ...
 $ Longitude              : Factor w/ 101 levels "","62.35","66.4447",..: 87 9 9 9 17 86 86 9 9 31 ...
 $ Province               : Factor w/ 9 levels "AJK","Balochistan",..: 4 9 9 9 3 8 8 9 9 6 ...
 $ Location               : Factor w/ 488 levels "","1)First blast lower mal gate 2)Ghaznavi street urdu bazar 3) Bhati Chowk
",..: 117 324 344 471 159 183 481 194 160 76 ...
 $ Location.Category      : Factor w/ 26 levels ""," ","Airport",..: 9 20 16 9 23 19 23 23 23 18 ...
 $ Location.Sensitivity   : Factor w/ 5 levels "","High","low",..: 2 4 5 2 5 4 5 5 5 2 ...
 $ Open.Closed.Space      : Factor w/ 7 levels "","closed","Closed",..: 3 3 3 3 3 5 3 3 3 5 ...
 $ Influencing.Event.Event: Factor w/ 175 levels ""," ","A Compound of an anti-Taliban militia commander Mullah Nabi Hanafi was
targetted by militants",..: 1 1 1 1 48 125 90 62 1 1 ...
 $ Target.Type            : Factor w/ 25 levels "","advocates (lawyers)",..: 11 16 11 11 23 17 23 23 23 17 ...
 $ Targeted.Sect.if.any   : Factor w/ 9 levels "","Ahmedi","Christian",..: 5 5 3 3 7 5 6 6 7 5 ...
 $ Killed.Min             : int  14 NA 13 NA 44 16 NA 14 16 NA ...
 $ Killed.Max             : int  15 3 15 12 47 18 1 15 18 2 ...
 $ Injured.Min            : int  NA NA 20 NA NA NA 3 96 NA NA ...
 $ Injured.Max            : Factor w/ 92 levels "","0","1","10",..: 77 43 56 68 79 67 55 32 50 30 ...
 $ No..of.Suicide.Blasts  : Factor w/ 6 levels "","1","2","3",..: 3 2 2 2 2 3 2 2 2 2 ...
 $ Explosive.Weight..max. : Factor w/ 117 levels "","100","1000kg",..: 1 1 57 NA NA 64 NA 59 NA 1 ...
 $ Hospital.Names         : Factor w/ 241 levels "","1.Agency Headquarters \nhospital ",..: 1 1 54 NA 23 37 NA 114 123 1 ...
 $ Temperature.C.         : num  15.8 23.8 31.5 31.4 33.1 ...
 $ Temperature.F.         : num  60.5 74.8 88.6 88.6 91.6 ...
```

Figure 1: Structure of Suicide Attack Data; before cleaning

A function was created and used to replace all the hashes at incorrect places with spaces. Then, the format of the months was corrected by bringing all the months into a consistent format. Thirdly, one date was misspelt. This was identified as the date format function could not identify it as a correct date; this date was then correctly spelt and replaced. Lastly, the dates once in a consistent format were converted as a date type.

Four extra columns were added to the data frame to split the date into its components, Weekday, day, month, and year. Part of the code used to clean Gregorian dates is shown in Figure 2: Gregorian Date Cleaning Code Parts.

Cleaning/Imputing Islamic Dates:

Due to the lack of packages in R as well as an exact function to convert Gregorian dates to Hijri dates, the missing values were imputed using Excel's Date formatting to Hijri Calendar. Since Excel would revert the date formatting back to Gregorian even after saving the file the Hijri date was converted to TEXT format using =TEXT() function with the format "[$-,106]B2d/mm/yyyy;@".

After imputation, the next task was to extract the date part from Islamic dates that were in two different formats, Excel and the original, as well as the fact that the original dates had multiple different spellings for different months.

A regex was created to extract the date parts and create a data frame of given dates with different formats and then a dictionary was used to clean the month names and bring them to a correct format. Once the data frame

Date	Islamic.Date	dirty_isl_month	isl_day	isl_month	isl_year_AH
2016-08-08	05 Zeqad 1437 A.H.	Zeqad	05	dhu al-qadah	1437
2017-04-05	08 Rajab 1438 A.H	Rajab	08	rajab	1438
2017-02-07	09 Jamadi-ul-Awal 1438 A.H.	Jamadi-ul-Awal	09	jumada al-ula	1438
2014-11-02	10/01/1436	01	10	muharram	1436
2015-10-23	10/01/1437	01	10	muharram	1437
2012-03-03	10/04/1433	04	10	rabi ath-thani	1433

Figure 3: Islamic Dates Cleaned Output

had clean values it was merged with the original data. Parts of the codes used to get this done is shown in Figure 4: Islamic Date Cleaning Code Parts as well as a segment of the cleaned output in Figure 3: Islamic Dates Cleaned Output

Cleaning Provinces: Most of the values were in the correct format except for a few spelling errors which were fixed easily.

Cleaning Blast Day Type: This attribute showed whether the day of the blast was a working day, holiday, or a weekend. The only issue in this was that there were several missing values hence imputation was needed. To impute this the gr_weekday attribute, that was created from the corrected Gregorian date, was used to identify whether the day was a Saturday or a Sunday or a weekday. Using this the missing values of the blast day type were imputed.

Cleaning Location Category: This attribute showed the type of location that the suicide bombing attack occurred; these included areas like Airport, Residence, Market, Education, etc. Other than the spelling errors and data synonyms (Foreign, Foreigner) there were missing values in the data set as well. To correctly impute this data, a location detail attribute was used which provided the detail of the location. An example of the location detail is "Entrance of a Shiite neighbourhood in southwestern Pakistan". Hence by using certain keywords in the detail of the location the type of the location could be imputed. For example, in the detail location given above the word 'neighbourhood' can be used to identify that the location of the target was of type residence.

To create a list of such keywords the details of the Location data were skimmed through and certain keywords were picked. The code for imputing the Location type by extracting information from Location details is given below in Figure 5: Code to Extract Location Type from Location Details while examples of the values in the attribute location details are shown in Figure 6: Location Details

```
#check the dirty dates and the format and track indexes
hash_count <- str_count(sa$Date, "-")
space_count <- str_count(sa$Date, " ")
#check unique
cat('hash:', unique(hash_count), ' | space:', unique(space_count))

#function to clean date
gr_date_clean <- function(x) {
    ind_sec = str_locate_all(x, '-')[[1]][2]
    str_sub(x,ind_sec, ind_sec) <- " "
    return (x) }

#create cleaned dataframe
temp <- t(data.frame(lapply(sa$Date[which(hash_count == 3)], gr_date_clean)))
row.names(temp) <- NULL
#replace unclean values with cleaned values
sa$Date[which(hash_count == 3)] <- temp

#fixing month format
sa$Date <- unlist(sapply(sa$Date, \
                         function(x) str_replace_all(x, '-Jan ', '-January ')))
...
sa$Date <- unlist(sapply(sa$Date, \
                         function(x) str_replace_all(x, '-Dec ', '-December ')))

#fixing date
sa[which(sa$Islamic.Date == '13/11/1436'),'Date'] <- 'Thursday-August 27-2015'

#converting to type Date
a <- sa %>% filter(Islamic.Date == '13/11/1436') %>% select(Date) %>% as.character

#converting Date type to Date
sa$Date <- as.Date(sa$Date, format = '%A-%B %d-%Y')

#adding columns by splitting date
sa$gr_week_day <- format(sa$Date, '%a')
...
sa$gr_year <- format(sa$Date, '%Y')
```

Figure 2: Gregorian Date Cleaning Code Parts

```
#using regex to extract date parts
pattern <- '([:digit:]{1,})[ |/]([:print:]*)[ |/]([:digit:]{1,})([:print:]*)'
islamic_date_parts <- str_match(sa$Islamic.Date, pattern)

#create dataframe from extracted parts
isl_df <- data.frame('isl_date'=islamic_date_parts[,1],\
                     'isl_day'=islamic_date_parts[,2], \
                     'dirty_isl_month' = islamic_date_parts[,3], \
                     'isl_year_AH'=islamic_date_parts[,4])

#convert all to character
isl_df$isl_date <- as.character(isl_df$isl_date)
isl_df$isl_day <- as.character(isl_df$isl_day)
isl_df$dirty_isl_month <- as.character(isl_df$dirty_isl_month)
isl_df$isl_year_AH <- as.character(isl_df$isl_year_AH)

#dictionary for cleaning islamic months
islamic_months <- list(
    'MuHarram' = 'muharram',
    '01' = 'muharram',
    ...
    '12' = 'dhu al-hijjah'
)

# fix islamic date using the dictionary created
isl_df$isl_month <- unlist(lapply(isl_df$dirty_isl_month, \
                                  function(x) islamic_months[[x]]))

#merge dataframes
sa2 <- merge(sa, isl_df, by.x='Islamic.Date', by.y='isl_date', all.x=TRUE)

#remove duplicated and remove incorrect indices
sa2 <- sa2[!duplicated(sa2),]
rownames(sa2) <- NULL
```

Figure 4: Islamic Date Cleaning Code Parts

Cleaning Open/Closed (Environment Type): This attribute has 2 values, either Open or Closed. Open indicates that the environment in which the suicide blast occurred was open, such as a ground, while Closed indicates a closed environment, such as a building. Other than some issues

Location
Bedian Road
Attack on Muharram Procession in Jacobabad near Lashari Muhalla
Imambargah
At PoliticalParty MP's office on College road Taunsa DGK.
Entrance of a Shiite neighborhood in southwestern Pakistan
Jinnah Airport-Airport Road Karachi

Figure 6: Location Details

```
sa2$Location.Category[sa2$Location.Category == '' & \
                      str_detect(tolower(sa2$Location), paste(c('college', 'school', 'children'), \
                                                               collapse = '|'))] <- 'Educational'
sa2$Location.Category[sa2$Location.Category == '' & \
                      str_detect(tolower(sa2$Location), paste(c('imambargah', 'church', 'muharram', 'shiite'), c\
                                                               ollapse = '|'))] <- 'Religious'
...
sa2$Location.Category[sa2$Location.Category == '' & \
                      str_detect(tolower(sa2$Location), paste(c('road', 'vehicle', 'bus', 'chowk'), \
                                                               collapse = '|'))] <- 'Mobile'
```

Figure 5: Code to Extract Location Type from Location Details

with whitespace and lower and upper case of the values the main issue was that of missing values. Here the missing values were imputed using the Location Detail attribute as well. Here the approach was to identify areas that were closed using keywords such as buildings, office, facility, school etc and then the remaining areas would be labelled as open.

Cleaning Target Type: This attribute indicates the type of the targets of the suicide bomb blasts (Civilian, Military, Foreigners). This attribute had missing values as well as a few misspellings and synonym data. While the misspellings and data synonyms were identified and cleaned, the missing Target type was identified using the Location Category that was cleaned above. For example, if the Location category was either one of the Residential, Educational, Market or Mobile then the Target Type of the suicide blast was imputed as Civilian. The Code for cleaning this is shown in Figure 7: Code Part for Imputing Target Type

```
sa2[sa2$Target.Type == '' & sa2$Location.Category %in% c('Residence', 'Educational', 'Market', 'Mobile'), \
    'Target.Type'] <- 'Civilian'
```

Figure 7: Code Part for Imputing Target Type

Cleaning Target Sect if Any: Target Sect provides values for what religious sect of the population was targeted if indeed one was targeted. Hence this provides a filter on the Target Type Religious. Since we have many sects in Pakistan such as Shiites, Sunni, Ahmadis, Christians etc, the missing data for this attribute was imputed using the Location details attribute provided. Picking up keywords such as church, Shiite, Imambargah etc the Target Sect was filled.

Cleaning Time: This attribute indicated the time of the attack. This attribute was also very dirty. Some examples of the values included 7:00 am, 7:00:00 am, After Noon, before Maghrib prayer etc. A few more examples are provided in Figure 8: Dirty Time Examples Since the exact time from values like near Friday Prayers cannot be identified, the time attribute was broken into 4 different parts of the day. These included:

- Morning: 5 am to 12 pm
- Afternoon: 12 pm to 5 pm
- Evening: 5 pm to 9 pm
- Night: 9 pm to 5 am

11:35:00 PM
11.35 am
11:40:00 AM
:
After Friday prayers
After Midnight
After noon
around 4:30 PM
around 5pm and second attack after 15min/5:16:00 PM and second attack after 25 min
between 7:00-7:40 AM

Figure 8: Dirty Time Examples

To do this, regexes were created to either identify words such as Friday Prayers or actual time formats provided. Different patterns of writing time were considered as the times were all written in different formats with some being mentioned in the middle of complete sentences as shown in Figure 8: Dirty Time Examples.

Part of the code segment used to extract the converted time formats from the dirty times is shown in Figure 9: Code Part to Clean Time

The drawback in cleaning this attribute is that part of the accuracy is lost. For example, even the values for which the exact time of the blast was provided, they were converted into either Morning, Afternoon, Evening, or Night, which resulted in a vaguer timing.

Another drawback is that for the values that are missing there is no way to impute the values. Hence only the values of time that were provided were cleaned and no imputation for the missing values was done.

```
#cleaning time
sa2$Time[which(str_detect(tolower(sa2$Time), \
                    paste(c('morning', '([56789]|10|11)([:.]\\d\\d(:(\\d\\d)?)?)? ?am'), \
                        collapse = '|')))] <- 'Morning'
sa2$Time[which(str_detect(tolower(sa2$Time), \
                    paste(c('[after]? ?noon', 'friday prayer', 'jumma', '([1234]|12)([:.]\\d\\d(:(\\d\\d)?)?)? ?pm'), \
                        collapse = '|')))] <- 'Afternoon'
sa2$Time[which(str_detect(tolower(sa2$Time), \
                    paste(c('evening', 'aftar', 'maghrib', '[5678]([:.]\\d\\d(:(\\d\\d)?)?)? ?pm'), \
                        collapse = '|')))] <- 'Evening'
...
```

Figure 9: Code Part to Clean Time

Removing Unwanted Columns:

A few attributes that either was not used for the data analysis that needed to be done or those that required a greater deal of data cleaning were excluded from the final dataset.

One of the attributes, Holiday Type, would have been quite helpful in seeing trends in the suicide attack data, it showed the type of holiday it was on the day of the suicide blast. However, most of the data for this attribute was missing and from the data that did exist most of it was the weekend. Since this attribute would not have added any value for the analysis, it was removed from the data.

The names of the cities were also mentioned in the data, but the names of the cities were quite dirty. Many different spellings for different cities and hence required a lot of cleaning. Since the data for latitudes and longitudes of the location of the attacks was provided with the attribute for city names was discarded; mainly due to the limitation of time.

The number of suicide blasts, Explosive weights, and influencing event attributes although interesting, had to be excluded due to the dirty values and the restriction in the time available to wrangle the data. Location sensitivity was another variable that explained how sensitive the location was but since there was a lack of metadata and no explanation was given as to what was considered sensitive and why this attribute had to be discarded.

Other attributes such as injured min, injured max, killed min, Temperature, Hospital names etc were discarded from the final dataset as these attributes didn't seem to

```
'data.frame':   496 obs. of  17 variables:
 $ Date                   : Date, format: "2016-08-08" "2017-04-05" ...
 $ sa_Blast.Day.Type      : chr  "Working Day" "Working Day" "Working Day" "Holiday" ...
 $ sa_Time                : chr  "" "Morning" "Morning" "" ...
 $ sa_City                : chr  "quetta" "lahore" "bannu" "wagah" ...
 $ sa_Latitude            : num  30.2 31.5 32.9 31.6 28.3 ...
 $ sa_Longitude           : num  67 74.4 70.6 74.6 68.7 ...
 $ sa_Province            : chr  "Balochistan" "Punjab" "KPK" "Punjab" ...
 $ sa_Location.Category   : chr  "Government" "Mobile" "Police" "Military" ...
 $ sa_Open.Closed.Space   : chr  "Closed" "Open" "Open" "Open" ...
 $ sa_Target.Type         : chr  "Judges & Lawyers" "Military" "Police" "Civilian" ...
 $ sa_Targeted.Sect.if.any : chr  "None" "None" "None" "None" ...
 $ sa_Killed.Max          : int  70 7 0 55 23 1 4 62 95 4 ...
 $ sa_No..of.Suicide.Blasts: num  NA NA 1 NA NA 1 NA NA 2 1 ...
 $ sa_isl_day             : chr  "05" "08" "09" "10" ...
 $ sa_isl_year_AH         : chr  "1437" "1438" "1438" "1436" ...
 $ sa_isl_month           : chr  "dhu al-qadah" "rajab" "jumada al-ula" "muharram" ...
 $ sa                     : num  1 1 1 1 1 1 1 1 1 1 ...
```

Figure 10: Final Suicide Data

be linked to the scope of the report and hence we not spend time on for wrangling due to the time limitation. The structure of the final suicide data after cleaning the required columns is shown in Figure 10: Final Suicide Data

Drone Strike Data

All the wrangling for this data set was done in R. Unlike the suicide attacks data set, the data presented here was almost in a clean format.

Importing Data: This data was also imported using Latin-1 file encoding as UTF-8 was unable to do so due to the presence of certain characters. The structure of the dataset after the import is shown in Figure 11: Structure of Drone Attack Data; Before Cleaning

```
'data.frame':   405 obs. of  26 variables:
 $ S.                   : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Date                 : Factor w/ 344 levels "","Friday, April 16, 2010",..: 34 174 193 20 23 101 4 263 43 67 ...
 $ Islamic.Date         : Factor w/ 344 levels "","10/01/1431",..: 16 243 248 308 62 333 4 275 175 192 ...
 $ Time                 : Factor w/ 77 levels "","1:00","10:00",..: 44 50 1 1 52 1 1 5 1 1 ...
 $ Location             : Factor w/ 304 levels "","Acha Dara area ",..: 218 192 106 230 36 34 229 179 45 122 ...
 $ City                 : Factor w/ 16 levels "","Bajaur  Agency",..: 14 11 11 11 2 2 11 11 11 3 ...
 $ Province             : Factor w/ 4 levels "","Balochistan",..: 3 3 3 3 3 3 3 3 3 4 ...
 $ No.of.Strike         : int  1 1 1 1 1 1 1 3 1 1 ...
 $ Al.Qaeda             : int  NA 1 1 NA NA NA NA NA NA NA ...
 $ Taliban              : int  1 NA NA NA NA NA NA NA NA NA ...
 $ Civilians.Min        : int  0 0 0 NA 0 0 0 20 5 NA ...
 $ Civilians.Max        : int  4 1 1 NA 18 82 2 32 8 NA ...
 $ Foreigners.Min       : int  NA NA NA NA NA NA NA NA NA NA ...
 $ Foreigners.Max       : int  NA NA 3 NA NA NA 2 NA 2 NA ...
 $ Total.Died.Min       : int  NA NA NA NA NA NA NA 20 5 NA ...
 $ Total.Died.Mix       : int  5 2 5 3 18 82 4 32 10 NA ...
 $ Injured.Min          : int  NA NA NA NA NA NA NA NA 6 NA ...
 $ Injured.Max          : int  NA NA 2 2 2 NA 2 15 12 1 ...
 $ Women.Children       : Factor w/ 4 levels "","0","N","Y": 3 3 1 1 4 4 1 4 1 1 ...
 $ Special.Mention..Site.: Factor w/ 309 levels "","2 Drones fired at a compound ",..: 41 77 92 1 268 91 31 90 144 1 ...
 $ Comments             : Factor w/ 173 levels "","10 Missiles were fired on 2 compounds",..: 172 31 108 1 92 3 96 96 67 1 ...
 $ References           : Factor w/ 374 levels "","AFP 15-Jun-2011/PBC-AFP-15-Jun-2011   Dawn News 15-Jun-2011/PBC-DWN-15-Jun-
2011  Daily Jang 15-Jun-2011/PBC-DJN-15-Jun-2011",..: 29 343 45 360 155 30 362 220 221 36 ...
 $ Longitude            : num  69.9 70.1 70.1 70.1 71.5 ...
 $ Latitude             : num  33 33 33 33 34.7 ...
 $ Temperature.C.       : num  28.475 11.475 7.08 0.535 10.025 ...
 $ Temperature.F.       : num  83.3 52.7 44.7 33 50 ...
```

Figure 11: Structure of Drone Attack Data; Before Cleaning

Cleaning Date: The data format in the dataset was consistent. The date attribute was converted from Factor to the Date format in R using the format it was provided in.

After conversion, the date was broken in separate parts Weekday, Day, Month, Year, as done before with suicide attack data.

Cleaning Province: The data for the province was already clean and the values did not need to be cleaned further or wrangled.

Cleaning Civilians Max and Total Died: The values for the total civilian casualties (Civilians.Max) included negative values shown in Figure 12, which is not possible. The summary of the values is given in Figure 15: Summary of Civilian Casualties After a bit of data checking it was obvious that Total = Civilians + Foreigner + Taliban + Al Qaeda casualties.

Al.Qaeda	Taliban	Civilians.Min	Civilians.Max	Foreigners.Min	Foreigners.Max	Total.Died.Min	Total.Died.Max	Injured.Min	Injured.Max	Women.Children
NA	6	0	-6	NA	NA	NA	5	NA	3	Y
NA	5	0	-5	NA	NA	NA	9	NA	NA	N
NA	4	0	-4	NA	NA	NA	6	NA	NA	N
NA	5	-4	-1	NA	NA	4	39	NA	NA	N
49	662	1304	2542	68	372	1943	3658	402	1329	

Figure 12: Part of Data related to Casualties

ıaeda	Taliban	Civilians.Min	Civilians.Max	Foreigners.Min	Foreigners.Max	Total.Died.Min	Total.Died.Max	Injured.Min	Injured.Max	Women.Children	new.tot.died
NA	NA	NA	NA	NA	NA	NA	3	NA	2		0
1	NA	5	7	5	7	12	13	NA	2	Y	15
NA	NA	7	8	NA	6	8	13	NA	3		14
NA	NA	3	5	NA	4	5	7	3	6		9
NA	12	0	6	NA	NA	7	12	NA	2		18
7	1	8	13	NA	NA	13	16	NA	NA		21

Figure 13: Creating new Total Casualties column with correct values

```
#adding new column for corrected new total died
ta <- ta %>% rowwise() %>% mutate('new.tot.died' = sum(Taliban, Al.Qaeda, Civilians.Max, Foreigners.Max, na.rm=TRUE))
ta %>% filter(Total.Died.Max != new.tot.died) %>% select(c(filt, 'new.tot.died'))

#fixing total died and changing its value
ta[,'Total.Died.Max'] <- ta %>% select(Total.Died.Max, new.tot.died) %>% \
    rowwise() %>% transmute('Total.Died.Max' = \
                            ifelse((is.na(Total.Died.Max))|(Total.Died.Max>new.tot.died), Total.Died.Max, new.tot.died))

#removing the additional column
ta <- ta %>% select(-new.tot.died)

#fixing negative civilians died
ta[neg_civ_died,'Civilians.Max'] <- ta[neg_civ_died,] %>% \
    select(filt) %>% rowwise() %>% transmute('Civilians.Max' = \
                                  sum(Total.Died.Max, -Foreigners.Max, -Al.Qaeda, -Taliban, na.rm=TRUE))

#replacing the old data frame with the cleaned one
da <- ta
```

Figure 14: Fixing Total Casualties and Civilian Casualties

Another issue was that in certain incidents the

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  -6.00    3.00    5.00   14.16    8.50 2542.00      46
```

Figure 15: Summary of Civilian Casualties

total casualties were less than the overall sum of the other casualties. This meant that there needed to be cleaning for the total casualties and then using this data correcting the values civilians. A new data frame was created where a new column was created where the values of total died were from the equation mentioned above. An example of this with total casualties as well as the correct column can be seen in Figure 13: Creating new Total Casualties column with correct values

Then only the values of the original total casualties that were less than the new total calculated casualties were replaced, after which the new column was removed.

The negative civilians were fixed using the new corrected Total casualties and the original data set was replaced with the corrected one.

The code for cleaning both the total died, and civilian casualties are shown in Figure 14: Fixing Total Casualties and Civilian Casualties

Cleaning Women and Children: This was a Boolean attribute indicating whether there were any woman and children in the casualties of the drone strike. It contained either missing values or 'Y' indicating a Yes or a 'N' indicating that there were no women or children casualties. There were minor spelling issues in the data that were corrected easily.

Figure 16: Incorrect Latitudes and Longitudes

Cleaning Latitudes and Longitudes: After the first visualisation of the latitudes and longitudes of the drone attacks, shown in Figure 16: Incorrect Latitudes and Longitudes, it was apparent that some of the values were incorrect, as the plotted points were falling outside of Pakistan. After a bit of viewing the data, it was identified that some values of the latitude and longitudes were incorrectly swapped with each other. Hence to fix it the values for those incorrect rows were swapped. The code to do is shown in Figure 17: Code Incorrect Lat and Long.

The structure of the final data after cleaning and removing the unwanted columns is shown in Figure 18: Final Corrected Drone Strike Data

Merging the Suicide Attacks and Drone Strikes Data

An outer join was done on both the data frames based on the date attribute. The date parts that were created for each dataset were removed so that they could be created for the final merged data frame. Before merging the two data frame's attributes were prefixed with either 'sa_' indicating that the attributes are for suicide attack data set or 'da_' indicating that the attributes are for the drone attacks datasets. The code for merging the data frames in shown in Figure 19: Merging Suicide Attacks and Drone Strikes Data and the final structure for the merged data frame is shown in Figure 20.

Emigration Data

The emigration data was imported as four different files, The counts per year of Emigration of citizens out of Pakistan by;

1. The country they were going to
2. Occupation of emigrants
3. Province
4. Occupation Category (not used in the exploration part of the report as the data didn't seem to be that meaningful)

The data was already clean, but a bit of formatting was done to make it compatible with other data. The data for Country was transposed to get years as rows and country counts as columns. This wrangling was done in Excel when the data was initially downloaded.

In the data set for the count of emigration by country, the country names were changed to match the names of the countries provided in the data set obtained for the Latitudes and Longitudes of the countries.

Latitudes and Longitudes Data

This dataset was imported to complement the emigration data by country. This was done so that we could visualise the countries based on the latitudes and longitudes in the data exploration segment of the report. The values were just copied and pasted from the website into an excel sheet and no specific file was imported.

Writing the Cleaned Datasets

The cleaned data sets were written back to be used for the data exploration part of the report. The code for doing this is provided in Figure 21.

```
col <- c('Longitude','Latitude')
da[which(da$Latitude > 36 & da$Longitude < 60), col] <- da \
    %>% filter(Latitude > 36 & Longitude < 60) %>% select(Latitude, Longitude)
```

to Fix

Figure 17: Code to Fix Incorrect Lat and Long

```
'data.frame':	404 obs. of  14 variables:
 $ Date             : Date, format: "2004-06-18" "2005-05-08" ...
 $ da_Time          : chr  "22:00" "23:30" "" "" ...
 $ da_City          : chr  "South Waziristan" "North Waziristan" "North Waziristan" "North Waziristan" ...
 $ da_Province      : chr  "FATA" "FATA" "FATA" "FATA" ...
 $ da_No.of.Strike  : int  1 1 1 1 1 1 1 3 1 1 ...
 $ da_Al.Qaeda      : int  NA 1 1 NA NA NA NA NA NA NA ...
 $ da_Taliban       : int  1 NA NA NA NA NA NA NA NA NA ...
 $ da_Civilians.Max : num  4 1 1 NA 18 82 2 32 8 NA ...
 $ da_Foreigners.Max: int  NA NA 3 NA NA NA 2 NA 2 NA ...
 $ da_Total.Died.Max: num  5 2 5 3 18 82 4 32 10 NA ...
 $ da_Women.Children: chr  "N" "N" "" "" ...
 $ da_Longitude     : num  69.9 70.1 70.1 70.1 71.5 ...
 $ da_Latitude      : num  33 33 33 33 34.7 ...
 $ da               : num  1 1 1 1 1 1 1 1 1 1 ...
```

Figure 18: Final Corrected Drone Strike Data

```
#renaming Suicide Attacks Attributes
names(sa_final)[2:length(names(sa_final))] <- paste('sa', names(sa_final), sep='_')[2:length(names(sa_final))]
#renaming Drone Strikes Atrributes
names(da)[2:length(names(da))] <- paste('da', names(da), sep='_')[2:length(names(da))]

#merging both the data frame with outer join
df <- merge(sa_final, da, by='Date', all=TRUE)

#ensuring no duplicates
df <- df[!duplicated(df),]
#fixing row indexes
rownames(df) <- NULL

#adding date parts into the data frame by splitting date
df$gr_week_day <- format(df$Date, '%a')
df$gr_day <- format(df$Date, '%d')
df$gr_month <- format(df$Date, '%m')
df$gr_year <- format(df$Date, '%Y')
```

Figure 19: Merging Suicide Attacks and Drone Strikes Data

```
'data.frame':	847 obs. of  34 variables:
 $ Date                  : Date, format: "1995-11-19" "2000-11-06" ...
 $ sa_Blast.Day.Type     : chr  "Holiday" "Working Day" "Working Day" "Working Day" ...
 $ sa_Time               : chr  "" "" "Morning" "Morning" ...
 $ sa_City               : chr  "islamabad" "karachi" "karachi" "karachi" ...
 $ sa_Latitude           : num  33.7 25 25 25 30.2 ...
 $ sa_Longitude          : num  73.1 67 67 67 67 ...
 $ sa_Province           : chr  "Capital" "Sindh" "Sindh" "Sindh" ...
 $ sa_Location.Category  : chr  "Foreign" "Office Building" "Hotel" "Foreign" ...
 $ sa_Open.Closed.Space  : chr  "Closed" "Closed" "Closed" "Closed" ...
 $ sa_Target.Type        : chr  "Foreigner" "Media" "Foreigner" "Foreigner" ...
 $ sa_Targeted.Sect.if.any : chr  "None" "None" "Christian" "Christian" ...
 $ sa_Killed.Max         : int  15 3 15 12 47 18 1 15 18 2 ...
 $ sa_No..of.Suicide.Blasts: num  2 1 1 1 1 2 1 1 1 1 ...
 $ sa_isl_day            : chr  "25" "10" "25" "3" ...
 $ sa_isl_year_AH        : chr  "1416" "1421" "1423" "1423" ...
 $ sa_isl_month          : chr  "jumada al-akhira" "shaban" "safar" "rabi ath-thani" ...
 $ sa                    : num  1 1 1 1 1 1 1 1 1 1 ...
 $ da_Time               : chr  NA NA NA NA ...
 $ da_City               : chr  NA NA NA NA ...
 $ da_Province           : chr  NA NA NA NA ...
 $ da_No.of.Strike       : int  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Al.Qaeda           : int  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Taliban            : int  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Civilians.Max      : num  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Foreigners.Max     : int  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Total.Died.Max     : num  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Women.Children     : chr  NA NA NA NA ...
 $ da_Longitude          : num  NA NA NA NA NA NA NA NA NA NA ...
 $ da_Latitude           : num  NA NA NA NA NA NA NA NA NA NA ...
 $ da                    : num  NA NA NA NA NA NA NA NA NA NA ...
 $ gr_week_day           : chr  "Sun" "Mon" "Wed" "Fri" ...
 $ gr_day                : chr  "19" "06" "08" "14" ...
 $ gr_month              : chr  "11" "11" "05" "06" ...
 $ gr_year               : chr  "1995" "2000" "2002" "2002" ...
```

Figure 20: Structure of Merged Data from Suicide and Drone Attacks

```
# writing data
write.csv(df, file='r_project_cleaned.csv') #merged data for suicide and Drone Attacks
write.csv(em_oc, file='em_oc.csv') #ocupation data
write.csv(em_co, file='em_co.csv') #country data
write.csv(em_pr, file='em_pr.csv') #province data
write.csv(em_ca, file='em_ca.csv') #ocupation category data
write.csv(em_co_lat_long, file='em_co_lat_long.csv') #lats and longs of emigration country
```

Figure 21: Code for writing csv files of Cleaned Data

DATA EXPLORATION

Question 1: Is there a relation between the suicide bombing attacks and the number of drone attacks in Pakistan? What is this relation? What are the casualties?

Once the data had been cleaned, it was ready for visualization and exploration. The first task was to view the number of suicide attacks and the drone attacks over the years. It's apparent, in Figure 23, that the suicide attacks start to increase after the year 2001 and reach their peak at around 2009 after which there is a decline. We can also see another increase in the number of suicide bombings after the year 2015. This increase could be due to the entry of Taliban and Al Qaeda groups into Pakistan after the inception of the US-led War in Afghanistan. But the sudden rise and fall needed an explanation.

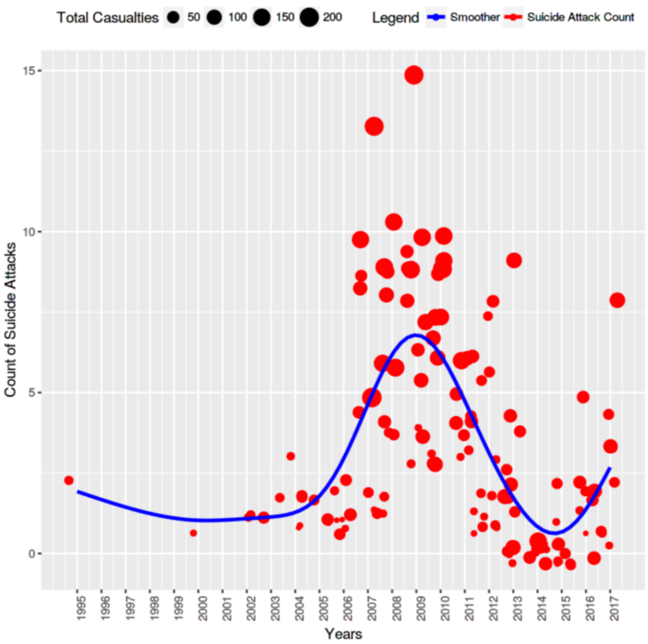

Figure 23: Suicide Attacks by Month

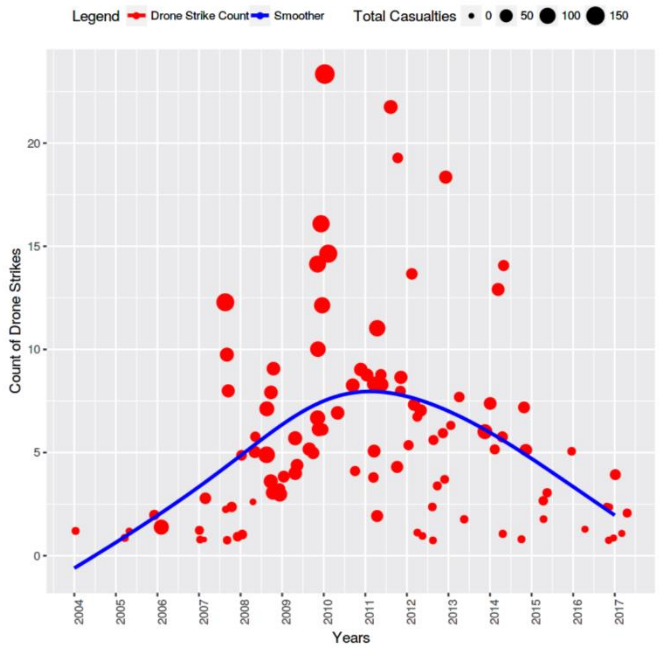

Figure 22: Drone Strikes by Month

After viewing drone attacks, in Figure 22, we can see that they start to occur around the year 2004 and experienced a steady increase till the year 2011 after which there was a steady decline till the year 2017. The rate of increase is almost the same as the rate of decrease over the years. To get a better understanding of the data both the count of drone strikes as well as the number of suicide bombings, the data for both were plotted in one graph to see the general trend between the two.

When viewed together, as seen in Figure 24, we can see that the suicide bombings start to increase after the September 11th attack in New York City and Washington DC. The drone strikes started in 2004. The interesting take from this is that the sudden rise in the count of suicide bombing attacks is during the exact time there is a rise in the drone attacks. We can also see that the count of drone strikes per month increased by a very large margin during Obamas Presidency.

We can see in Figure 22 and Figure 23 that not only did the count of attacks increased for both the suicide as well as drone attacks, the number of casualties also rose quite high during these months. But what can we tell about these casualties? Who are they? For this the type of casualties can we view. Let's start off with the drone strikes.

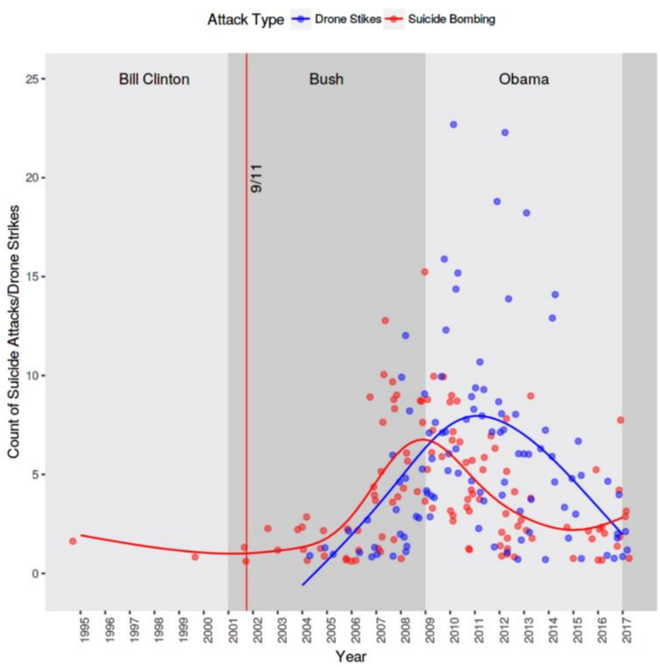

Figure 24: Drone Attacks vs Suicide Bombings by Months

As seen in Figure 25: Drone Strike Casualties by Year, the number of casualties peaked during the year 2010 and each year, leading up to the year 2013, the total casualties of civilians were the most compared to any other. After the year 2013, we can see that the number of casualties for Taliban was the highest. An aggregate of the total deaths caused by drone strikes can we viewed in Figure 26: Drone Strike Casualties by Type. We can see that over the years the highest number of casualties caused by drone strikes conducted by the United States have been Pakistani Citizens and not the Al Qaeda or the Taliban which according to the US was their mission.

Unlike the drone strikes conducted by the US, that were restricted only to FATA and KPK, suicide bombings could be conducted in any location. Hence to understand the effects of the casualties, the plot for the total casualties by type is created with the count of attacks. In Figure 27 we can see that the top 5 targeted types are Military, Police, Civilians, Government Officials, and Religious, in the given order. Although the count of attacks on civilians is 3rd highest but the highest casualties by a large amount are in civilians. Figure 28 also compliments this, as we can see the

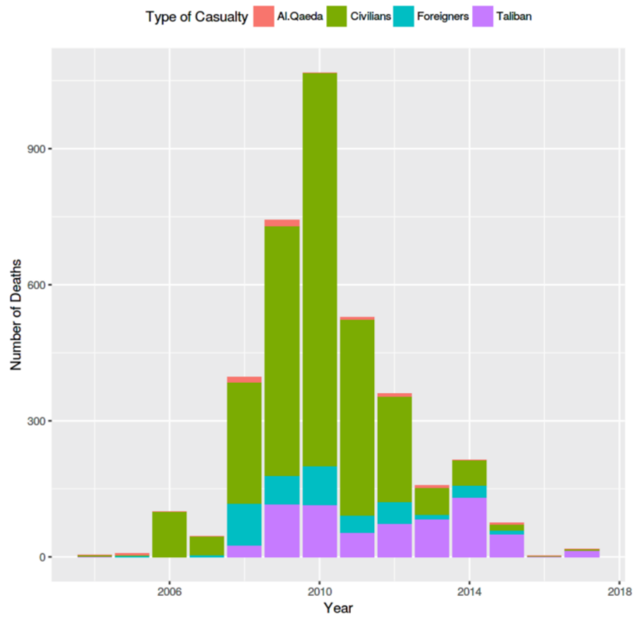

Figure 25: Drone Strike Casualties by Year

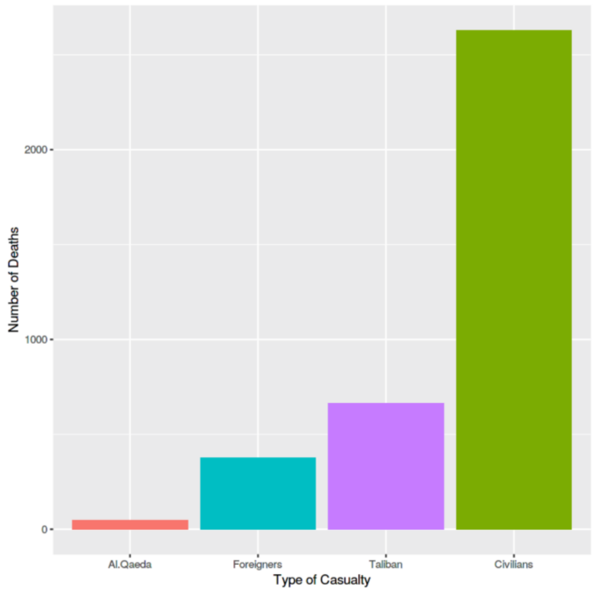

Figure 26: Drone Strike Casualties by Type

count of attacks on Police, Military, Government Officials, and Civilians all increased during the years of the drone strikes. One theory could be that the suicide bombing is focused on Military, Government, Police is because they might be retaliating against the drone attacks being done. This theory also adds v value to the fact that the high casualties in civilians could be because, firstly civilians are easy targets, secondly, targeting civilians puts very high pressure on the government. To check this, we can view the target location type by years as shown in Figure 30, in which we can see that the count of attacks done on the locations related to police, government officials, military as well as civilian related locations such as mobile, markets etc all shot up in the year 2007. This is the exact year during which the intensity of drone attacks also increased drastically as seen in Figure 22.

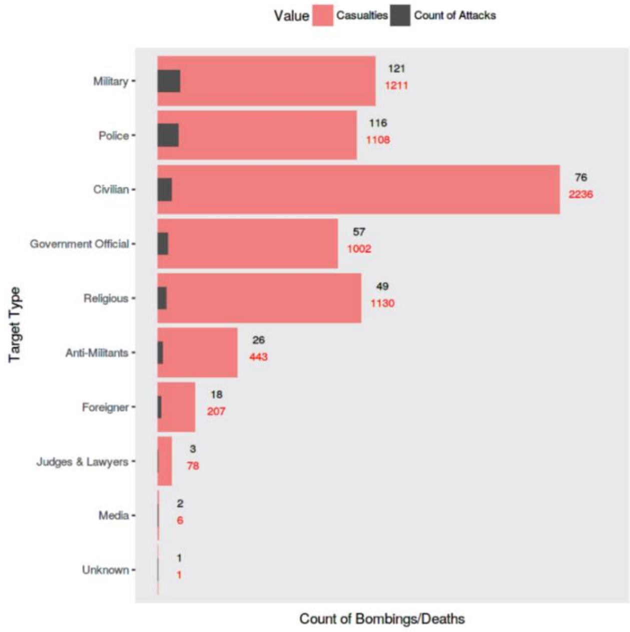

Figure 27: Casualties of Suicide Attacks and Count of Attacks

Now that there is some idea as to what the casualty types were for both it was time to view the data province wise to check whether there was anything interesting in that. In Figure 29: Suicide Bombings & Drone Strikes Location Wise, we can see that the drone strikes were only focused in FATA and KPK (we already knew that) and that the suicide bombings are not region specific. We couldn't find anything to take from this plot, even if there is something there, it cannot be view directly.

Although there is no pattern available on the macro level if we try to check the data only provinces we might be able to capture something. In Figure

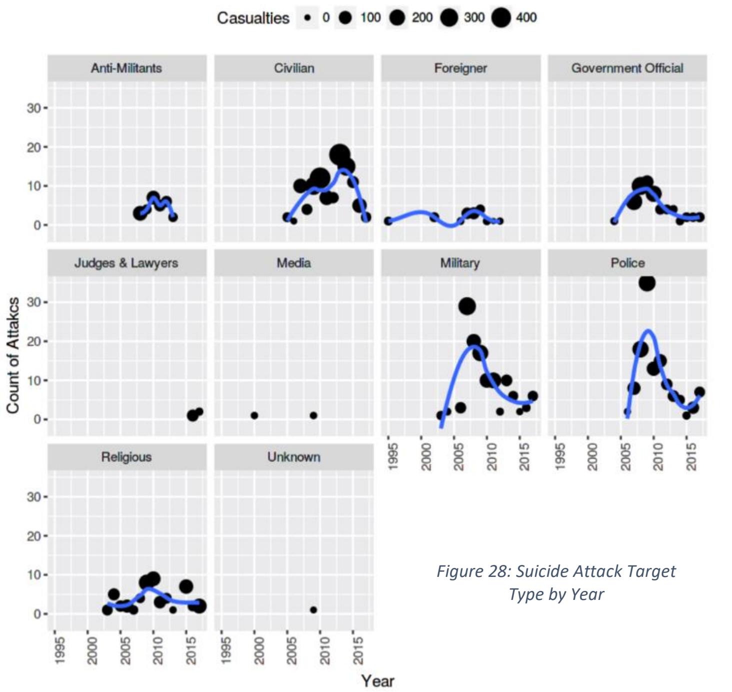

Figure 28: Suicide Attack Target Type by Year

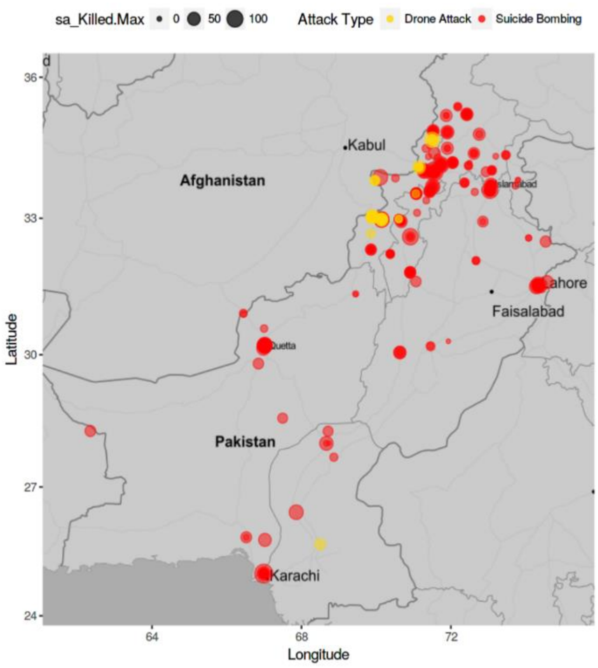

Figure 29: Suicide Bombings & Drone Strikes Location Wise

Figure 30: Suicide attack Target Location Type by Year

33.1, we can see that that almost all the drone strikes were in FATA while only a handful in KPK. An interesting find is that the highest number of suicide attacks are in KPK and decrease as the distance from KPK decreases. If we try to go deeper and view the count of attacks not only by Provinces but also of Target Types with Provinces, we get the plot shown in Figure 33.2. We can see that some of the Target Types are mainly focused in northern locations, but the facet figures are somewhat similar. Hence a bar chart of the same data is shown in Figure 33.3. Through this, it is very easy to see that Military, Government officials, Anti-Militants and Civilian Target Types are most common in and near KPK, where the highest counts are seen in KPK. Target type of Police is interesting as well as in other provinces the attacks focused on police are somewhat similar except for KPK where the count of attacks skyrockets. Other Target Types such as Religious, Judges and Lawyers, and foreigners are not only low in numbers compared to others but are also similar throughout the regions.

This is a very beneficial find as most drone attacks occur in FATA which is next to KPK. FATA is Federally Administered Tribal Areas and mostly a war zone, where nearly 2 million people were dispersed as the schools, hospitals, housed were all destroyed. Hence it somewhat makes sense that in terms of retaliation the next best option for the terrorist groups was KPK which is right next to FATA. The retaliation attacks seem to be region specific but other attacks such as religious, Judges & Lawyers or Foreigners seem to be unaffected by region.

Question 2: Does the lack of safety or security concerns affect the emigration of citizens out of Pakistan? Do citizens consider emigration for survival?

After the above exploration, it was apparent that all areas of Pakistan were being affected one way or the other with the impact of Suicide bombings as well as indirectly with Drone Strikes. But to understand this effect better, I decided to use data of people emigrating out of Pakistan and into other countries. Emigration counts say a lot about one country; people emigrate for employability purposes, health, better living standards, and most importantly security.

When the count of emigration out of Pakistan is plotted over the years, as in Figure 35, we can see that there is a sudden increase in the count right after the incident of 9/11. This seems to fit our general view that as the security situation in the country worsens more and more people tend to leave the country to settle elsewhere. But since as explored before, most of the attacks are focused on certain regions, viewing the emigration data by the province would tend to make more sense. Hence in Figure 34, we can see that the main driving force for the rise in total emigration is the count of Skilled and Un-Skilled Labour after 9/11. As seen in Figure 33.1: Drone Strike & Suicide Bombing by Province, we know that the most affected areas by suicide attacks were KPK and Punjab. KPK is mostly rural and has unskilled labour, while Punjab has a mixture of skilled

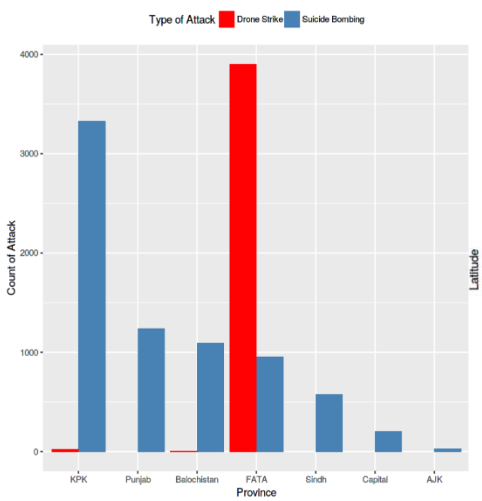

Figure 33.1: Drone Strike & Suicide Bombing by Province

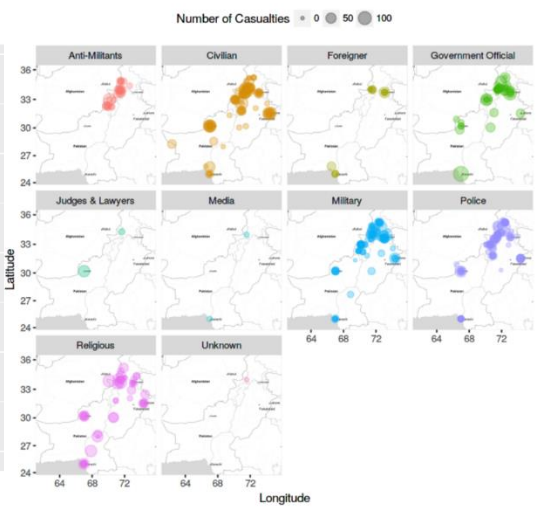

Figure 33.2: Suicide Bombing by Target Type and Location

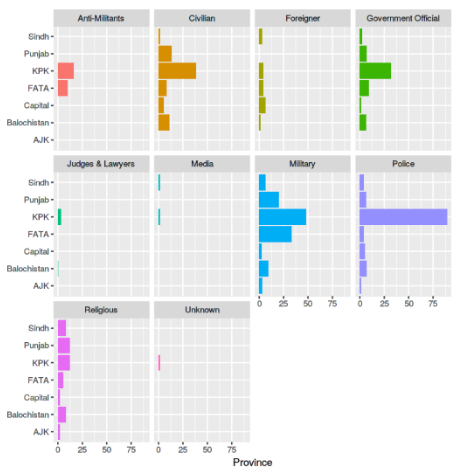

Figure 33.3: Suicide Bombing by Target Type & Province

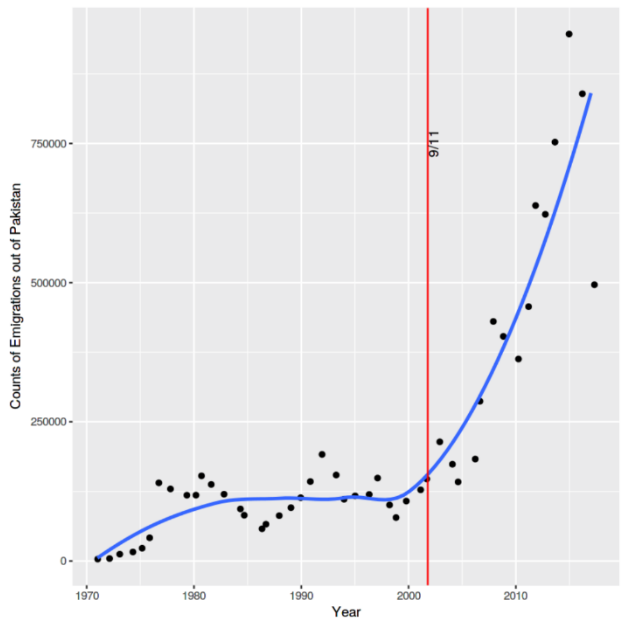

Figure 35: Emigration Counts Out of Pakistan by Year

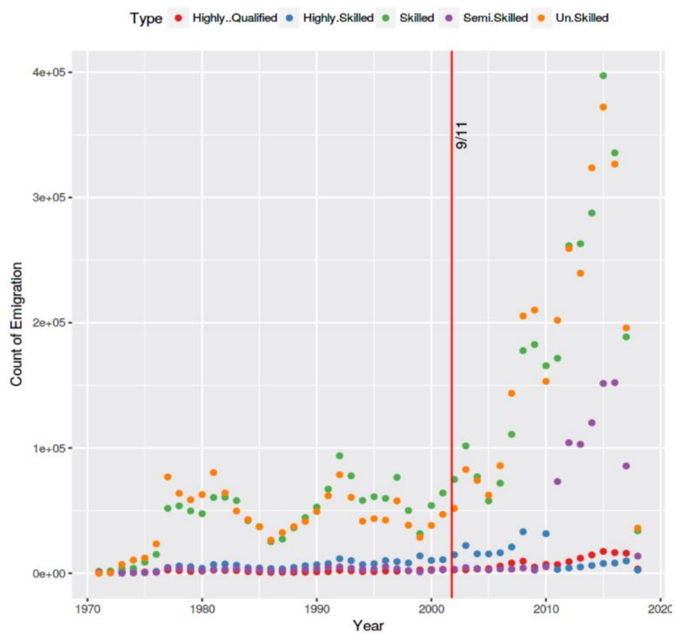

Figure 34: Emigration Counts out of Pakistan by Year & Skill Type

labours. To understand and get an idea of which province had the most people emigrating out of, the emigration data was plotted over years by province.

As per our guess, in Figure 36: Emigration Counts out of Pakistan by Year & Province we can see that the emigrants are mainly from the provinces of Punjab and KPK. This tends to prove the point that due to the security issues surrounding the areas of most impact i.e. KPK and Punjab, people are trying to move out of the country and relocate for their own safety.

Figure 37: Emigration out of Pakistan Flow Map shows the flow of people migrating out of Pakistan and shows where they are relocating. The highest count of emigrations was for the countries Saudi Arabia, United Arab Emirate, Oman, Qatar and Bahrain, represented by the red points. The colour of the lines did not affect the flow map as the countries are very close to each other and the colours do not show well.

Figure 37: Emigration out of Pakistan Flow Map

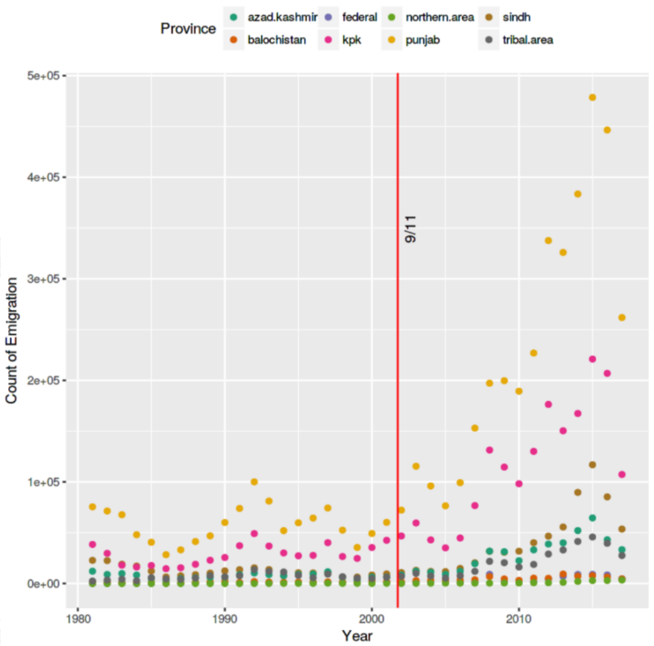

Figure 36: Emigration Counts out of Pakistan by Year & Province

Question 3: Is there any trend that can be identified in the suicide bombing attacks?

We have already identified partly that the suicide attacks were mainly focused on Government Officials, Military, Anti Militants, Police and Civilians, but the general similarities on how the suicide attacks occurred have not yet been explored. Hence to explore this

In Figure 39: Suicide Bombing Attacks by Gregorian Months, we can see that the count of casualties by suicide bombing attacks is higher in the months of February and March and then drops down until September when the counts rise till the end of December. February and March have major events due to Kashmir day (February) and Pakistan day (March). Both these days are important in Pakistan's history and events are conducted by both civilians as well as governmental events. December is important in terms of two major events being the Birthday of Quaid-e-Azam (founder of Pakistan-25th Dec) and new year's. 25th December includes events hosted again by both civilians as well as government while new year's is mainly celebrated by civilians. The high count in September to November could be due to multiple reasons and I couldn't seem to identify as to why it could be happening. One theory might be that there might be more government-related events happening during these months and since a lot of the suicide attacks are government targeted, this might be increasing the numbers.

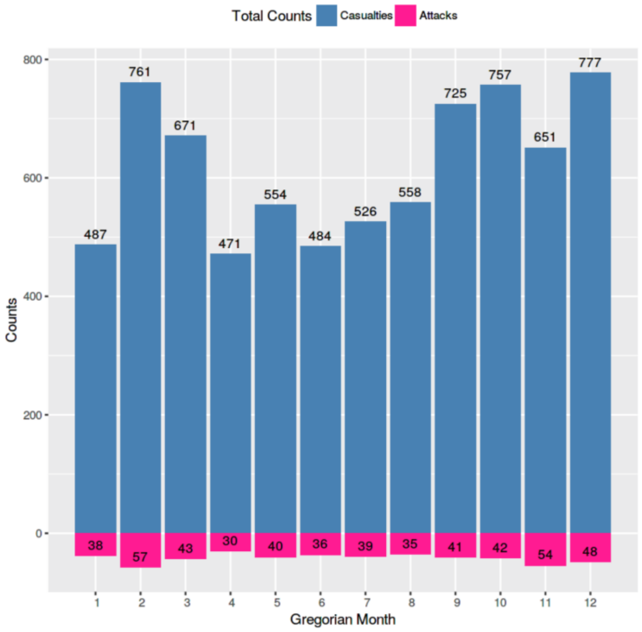

Figure 39: Suicide Bombing Attacks by Gregorian Months

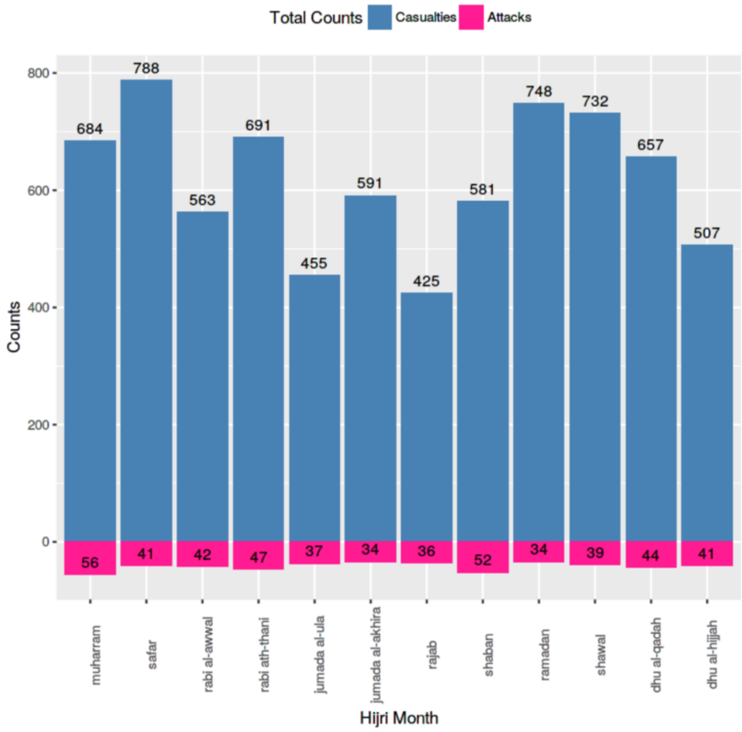

Figure 38: Suicide Bombing Attacks by Islamic Months

Since Pakistan is a Muslim state with most of its population being Muslim viewing the count of casualties and suicide bombing attacks in different Islamic months might unveil some trends, and hence it was plotted. We can see in Figure 38, that the number of casualties is very high in Ramadan and Shawal and the highest in Safar. Ramadan is the month where Muslims fast and hence there are iftars of Muslim gatherings where Muslims can be a high target. In Shawal there is Eid ul-Fitr, it is the Eid of celebration after Ramadan ends and hence there are multiple events of happiness during this month and hence if targeted there might be high casualties. Although during the month of Dhu al-Hijjah there is another Eid for Muslims, this Eid is not considered as happening as the other one and hence there are fewer gatherings and we can see fewer casualties. Why the month of Safar is the highest in terms of casualties seems a bit confusing.

In terms of the number of attacks, the highest number of attacks are in the months of Muharram and Shaban. Muharram is a very important and sad month for Muslims but the 'Shia' sect of Muslims tend to have ongoing activities in this month and are very highly vulnerable. Despite the high number of attacks, the security for all the events, especially in Muharram, is very high and hence there are fewer chances of successful attacks and are mostly countered.

Since a solid trend could not be identified for different months the suicide attack data was plotted against weekdays. In Figure 41: Suicide Attacks by Week Days, we can see that the count of attacks and casualties is higher in Monday, Thursday, and Friday, with Friday being the Highest. The main reasoning for this could be that in Pakistan Saturday and Sunday are weekends and holidays for work, school etc. Hence most of the activity is done on Thursday and Friday, and on Monday the activity is high as most of the work resumes are the weekend. One reason for the Friday count

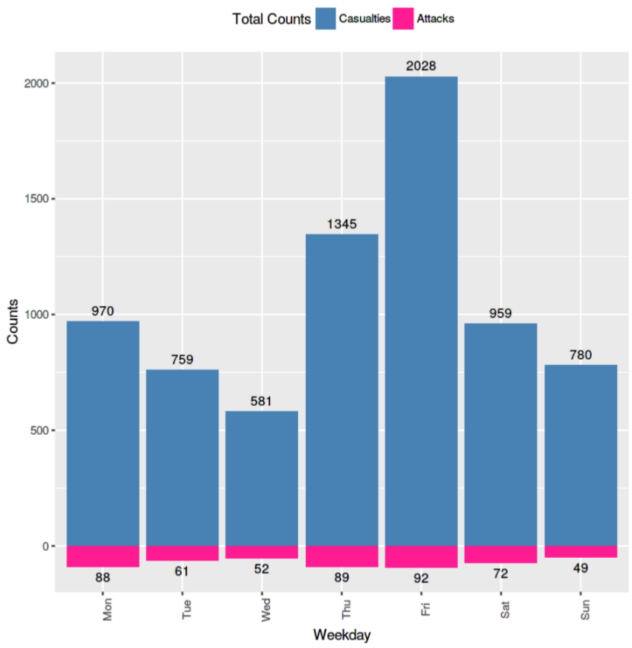

Figure 41: Suicide Attacks by Week Days

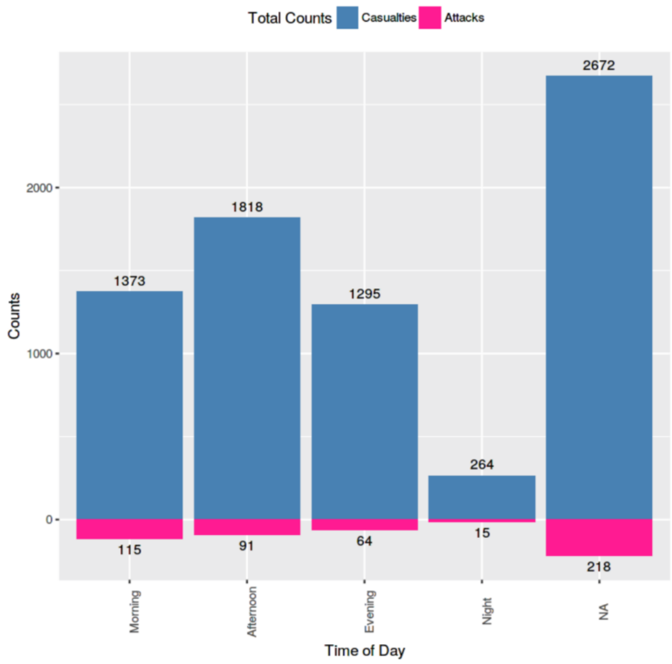

Figure 41: Suicide Attacks by Day Time

being the highest could be that there is Friday prayer for Muslims and they are vulnerable targets are it is conducted in a crowd rather than at home with crowds sometimes so large that entire roads are covered for Friday prayers.

To check this theory the attacks were checked with a time of day as well. In Figure 41: Suicide Attacks by Day Time, we can see that the most attacks are done during the afternoon and the least are done during the night. This could be because afternoons are the most active in Pakistan and Nights unlike other countries are the least active since Pakistan is a Muslim State and there is very little nightlife.

Another characteristic that could lead to some sort of a pattern identification was the environment type for the suicide bombings that occur. Environment type basically is either open (markets, parks etc) or closed (buildings, houses etc). Figure 43 and Figure 44, show that there are more suicide blasts in open areas compared to closed. This was expected as open areas are easier to target compared to closed areas due to the presence of security. To identify better on what target locations types were targeted in what ways, Figure 42 was created, where we can see that Military, Police, and Markets are mainly targeted in open environments. This could be mainly true for getting through Military, Police and Mall

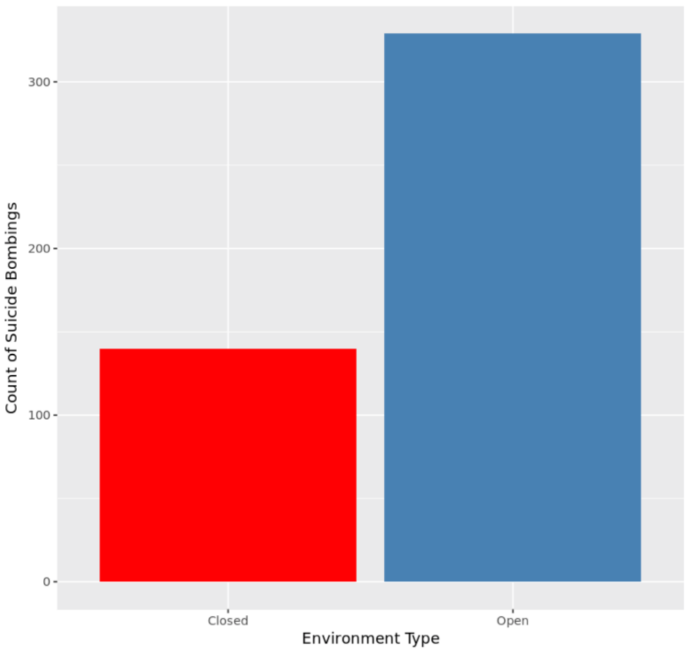

Figure 44: Suicide Bombings by Environment

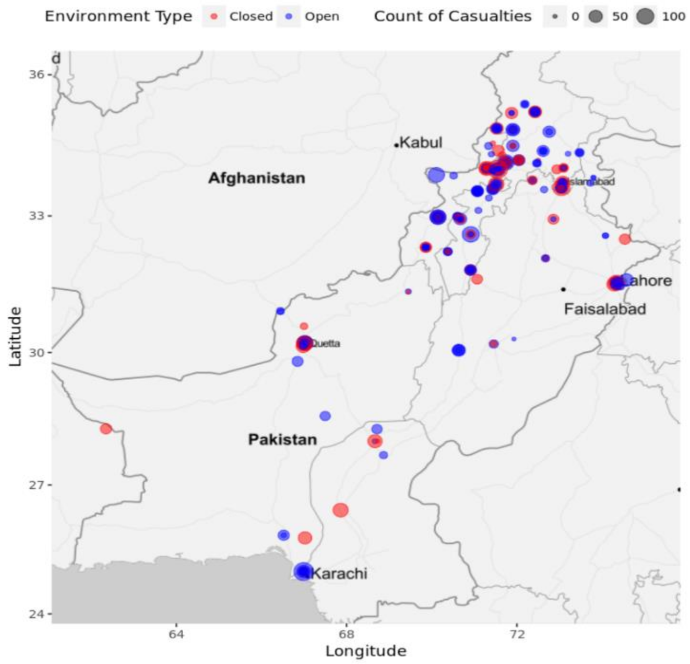

Figure 43: Suicide Bombings by Environment

security in Pakistan is very difficult. Another interesting find is that Religious locations are generally targeted in closed environments. This somewhat makes sense as mostly Religious activities mainly occur in closed areas such as Mosques, Churches etc.

CONCLUSION

After doing the exploration we can see that there seems to be a clear correlation between the number of suicide bombing attacks and the start of the drone strikes in the region of FATA and KPK. As the drone strikes started occurring the number of suicide attacks skyrocketed. We can also see that after a certain point where the number of drone strikes peaked higher than the count of suicide attacks during Obama's tenure in the white house, 2009, the count of suicide attacks started to decline. This could be taken as an indication that the drone strikes were successful, although we see a rise in suicide attacks again after 2015, which might indicate that the terrorists might just have gone into hiding for a while.

After further exploration, we could identify that the rise in suicide attacks was mainly driven by the increased attacks on Military, Police, Anti-Militants, Government Officials and Civilians which the attacks on religious and foreigners did not contribute much to the total rise in numbers. Usually, this is not the case with general Islamic terrorist groups as they tend to focus their attacks on foreigners (mostly anyone not from their religion), or Religious Islamic sects such as against Shia, Sunni, Ahmadis etc. Through the exploration, it is a clear indication that the rise in attacks was due to the retaliation for the drone strikes getting conducted against the terrorists. Police, Military etc. are a clear indication of retaliation but there are high civilian casualties as targeting civilians is the best way to put pressure on any government.

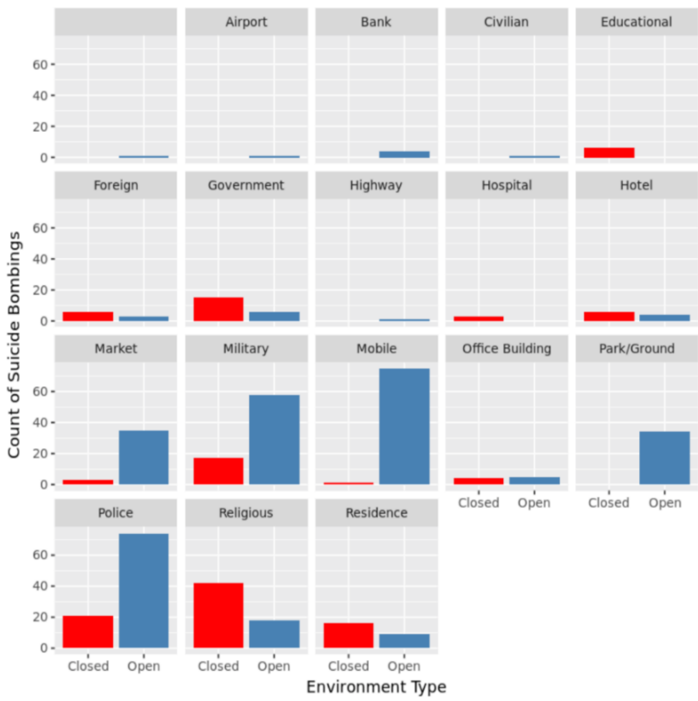

Figure 42: Suicide Bombing Environment Type by Location Category

Besides the Target type, we have seen that in suicide bombings the targeted group was Pakistani citizens, whether they be civilians or related to the Police, Military etc. With the drone strikes, we can see that most of the casualties were civilians, which is shocking as the main purpose of these were to target terrorists (Al Qaeda and Taliban).

We also saw that the impact of the suicide bombings, as well as the drone strikes, were quite high on the Pakistani citizens (as expected). Firstly, although most of the drone strikes were focused in FATA the maximum number of suicide attacks were in KPK and there decreased as the distance increased from FATA. There are a few points to take from this. First, we know already that the impact of the clash has resulted in the destruction of almost all schools, hospitals, houses etc in FATA, this led to most of its population had to relocate in surrounding provinces. Hence, to target civilians or any other target group, the terrorists had to move to their next closest potentially high populated areas i.e. KPK and Punjab. AJK, although being very close to FATA, has very high security due to the current issue over Kashmir between Pakistan and India and hence the terrorist groups, as seen in the plots, were unable to create havoc here. Secondly, we can see a huge rise in people emigrating out of Pakistan after the start of this clash. The counts see a steep rise after 9/11, mostly in provinces highly affected by the clash mainly KPK contributing to the emigration of Unskilled labour, and Punjab contributing to the skilled labour. This is due to the fact that there was a high intake of such skill levels by Saudi Arabia and UAE.

In terms of the similarities between the attacks, we can see that most of the attacks are targeted during the winter season and in terms of Hijri dates, the attacks are focused more in the months of Safar, Ramadan and Shawal. This could mainly be due to the type of holidays and type of events that occur during these times, mainly both for civilians and government events alike. However, we can see a much-focused trend when we view the attacks over different times of the day and weekdays. Most of the attacks are focused around Thursday, Friday and Monday i.e. days surrounding the weekend when the activity is highest in Pakistan. In terms of time we see a high count of attacks and casualties in the Afternoon as mainly most attacks occur on Fridays and the afternoon, the time during the compulsory Friday prayer is the most suitable for the terrorists to attack. In terms of the environment type, we can see that there is not much a trend.

REFLECTION

Although the report might have been able to find good (subjective) findings from the data, there are some steps that could not be taken due to time, data acquisition issues or other restrictions, that if they were taken they might have improved the quality of the findings.

Firstly, more statistical tests might have been taken in order to calculate the strengths of the correlation between many variables, such as the actual correlation between the suicide bombing attack numbers and the number of drone strikes. This would not only compliment our findings from visualizations but would also have identified any incorrect correlations that might have been thought correct visually but might have been coincidences.

Secondly, different locations of Pakistan have different populations. If the visualizations and relations were created on normalized data of for example emigrations or number of casualties, the visualization might have been considered more accurate and representative of the actual situation. That said normalised data is always a good choice is considered better for any visualization but due to time constraints, this could not be done in this report.

Thirdly, in terms of identifying trends and similarities between suicide bombing attacks, more data could have been used to go deeper into the analysis and identify trends. An example of this could be using data of political events getting conducted to identify why certain Gregorian months had more count of suicide attacks than others.

Lastly and most importantly, an important finding for the project was that keeping aside the argument of whether the drone attacks were successful in combating the suicide bombings (terrorist activities), the main effect of this was on Pakistan's public. There are many arguments against the use of drone strikes but using this exploration we can see that firstly; the strikes not only caused a lot of civilian casualties themselves but also aggravated the suicide attacks on Pakistani Population. Secondly, it created a war zone in FATA, resulting in families relocating to other areas by leaving behind their lives, but also the loss in terms of destruction caused. Thirdly, the security situation resulted in a lot of the skilled and unskilled labour of Pakistan emigrating out of the country. The labour of the country is the driving force behind it and loss of it can lead to many negative economic effects on any country. In a nutshell, taking the step of drone strikes should have been taken by considering all the outcomes it might have on the country. Maybe, if another method of tackling the issue of terrorism, the negative effects might have been less, or maybe this was the only way; this debate is for another time.

REFERENCES

Data sets

[i] https://www.kaggle.com/zusmani/pakistandroneattacks
[ii] https://www.kaggle.com/zusmani/pakistansuicideattacks
[iii] http://beoe.gov.pk/reports-and-statistics
[iv] https://developers.google.com/public-data/docs/canonical/countries_csv

Research paper

http://watson.brown.edu/costsofwar/files/cow/imce/papers/2015/The%20Impact%20of%20Drone%20Strikes%20in%20Pakistan.pdf
Zeeshan-ul-hassan Usmani and Hira Bashir, "The Impact of Drone Strikes in Pakistan", Cost of War Project, Brown University, December 16, 2014