a) What is Data Analytics?

Data analytics (DA) is the process of examining data sets to draw conclusions about the information they contain. It helps to make more-informed business decisions and by scientists and researchers to validate or reject scientific models, theories and hypotheses.

We use various algorithm or manual methods to analyze data and present them visually for better understanding. For instance, investing many data sets to derive relationship between variables.

b) What is the difference between Data Analytics and Data Science?

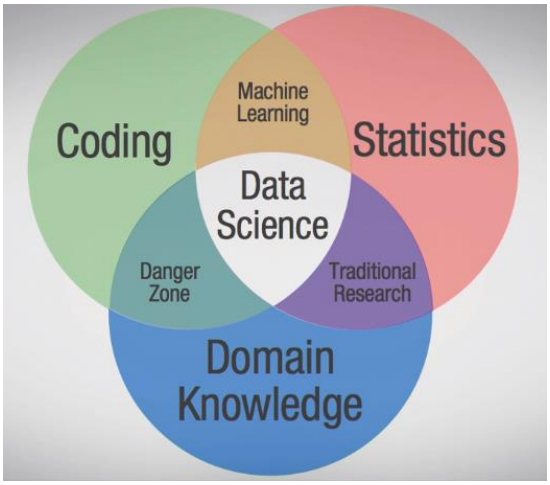
Data Science is the combination of statistics, mathematics, programming, problem-solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing and aligning the data.

Think of data science is the house that hold the tools and methods, data analytics is a specific room in that house. It is related and similar to data science, but more specific and concentrated. Data analytics is generally more focused than data science because instead of just looking for connections between data, data analysts have a specific goal in minding that they are sorting through data to look for ways to support. Data analytics is often automated to provide insights in certain areas.

Normally a data scientist is expected to formulate the questions that will help a business and then proceed in solving them, while a data analyst is given questions by the business team and pursues a solution with that guidance.

Both roles are expected to write queries, work with engineering teams to source the right data, perform data munging (getting data into the right format, convenient for analysis/interpretation) and derive information from data. However, in most cases a data analyst is not expected to build statistical models or be hands-on in machine learning and advanced programming. Instead, a data analyst typically works on simpler structured SQL or similar databases or with other BI tools/packages.

The data scientist role also calls for strong data visualization skills and the ability to convert data into a business story. A data analyst is normally not expected to transform data and analysis into a business scenario and roadmap. Following is the Venn-diagram for data science.



c) What is the utility of Data Analytics in Governance, why is it needed in the current times?

Implementing data analytics in the government projects delivers clear benefits. It helps to make prudent policy and reduce risk. We can use analytics to automate assessments on forecasting processes, and with banks to create their own recommendation engines.

Analytics helps to understand and learn from what has happened in the past. As data becomes more important to the government, you may move from 'descriptive' analytics to 'prescriptive', i.e. where should the government policy go next? Some of the benefits are as follows-

1. Raw data comes at us with overwhelming velocity and volume. But we are fundamentally visual creatures, so graphical representations will always prove more insightful to us than columns and rows of numbers. That's why data visualization is so important; it allows anyone from within a department to quickly grasp difficult concepts and identify new patterns within data without the need for complex analysis
2. **Government Administration**: Back Office (Finance, Human Resources, Procurement)
   Data analytics helps agencies identify departments that are not in normal range, areas where retirement will impact operation, types of skills that will be needed, query the data to find encumbrances, where spending at current rates will exceed budget and why.
3. **Health & Human Services:** Predictive data analytics uses all data available to identify most at-risk children, those least likely to be re-unified and the best services for each child.
4. Fraud Detection/Prevention Fraud, waste and abuse is something all government agencies keep a watchful eye on. But the use of predictive data analytics can help agencies stay ahead of fraudsters and allow machine learning to identify schemes and the big business behind fraud.
5. Airports/ Railways – enhance customer experience and reduce costs, predictive data analytics helps identify terminals not in use to save on maintenance and heating/AC costs, best locations for retail, where long lines are anticipated to assign staff, route expansions to maximize airport revenue.

I will be using Terrorism data to show how analytics can help government to make robust policy to fight against it.

d) Define a use-case according to your own understanding where Data Analytics can be used in Governance? (Build a model using R/Python)

In the current era, the severe problem which not only India but entire world is facing, is Terrorism. It is threat to mankind and very difficult to predict future attack- how, when, and where, they are going to happen next due to unavailability of real data. Therefore, counter terrorism is an act which prevent such attacks and saves mankind. Nowadays, every country is focusing on building strong and intelligent system which could accurately predict and enable them to draw conclusion to take required safety measures for providing safe and peaceful life to humans.

In this report, I will provide descriptive and inferential statistics of Terror attacks happened in India using GTD data provided by the university of Maryland. Data can be found online *http://www.start.umd.edu/.* Technology used for the below analysis is R using various libraries such ggplot, random forest, tidyverse, data table etc. Please find R code and data file here.


TA_D.R


Terror_India.csv

Descriptive Analysis

Please note missing and unknown data are purposely included for descriptive analysis to show the crisis of real data.

Date	provstate	attacktype1_txt	targtype1_txt	gname	nkill
13-Jan-16	Odisha	Bombing/Explosion	Unknown	Unknown	0
14-Jan-16	Delhi	Facility/Infrastructure Attack	Airports & Aircraft	Hindu Illaignar Sena	0
14-Jan-16	Odisha	Armed Assault	Private Citizens & Property	Maoists	1
14-Jan-16	Bihar	Armed Assault	Violent Political Party	Unknown	1
15-Jan-16	Haryana	Bombing/Explosion	Transportation	Students Islamic Movement of India (SIMI)	0
15-Jan-16	Meghalaya	Hostage Taking (Kidnapping)	Private Citizens & Property	Achik National Liberation Army (ANLA)	0
15-Jan-16	Chhattisgarh	Armed Assault	Police	Maoists	4
15-Jan-16	Jammu and Kashmir	Bombing/Explosion	Unknown	Unknown	0
15-Jan-16	Odisha	Hostage Taking (Kidnapping)	Government (General)	Communist Party of India - Maoist (CPI-Maoist)	1
16-Jan-16	Kerala	Bombing/Explosion	Police	Democratic Youth Federation of India (DYFI)	0
16-Jan-16	Kerala	Bombing/Explosion	Police	Democratic Youth Federation of India (DYFI)	0
16-Jan-16	Bihar	Bombing/Explosion	Private Citizens & Property	Unknown	0
17-Jan-16	Andhra Pradesh	Bombing/Explosion	Religious Figures/Institutions	Unknown	0
17-Jan-16	Chhattisgarh	Bombing/Explosion	Military	Unknown	0
18-Jan-16	Chhattisgarh	Bombing/Explosion	Military	Maoists	1
18-Jan-16	Punjab	Armed Assault	Terrorists/Non-State Militia	Unknown	0
19-Jan-16	Assam	Bombing/Explosion	Private Citizens & Property	Unknown	1

Figure 1

Figure 1 is showing sample data of 10 rows in html formatted table

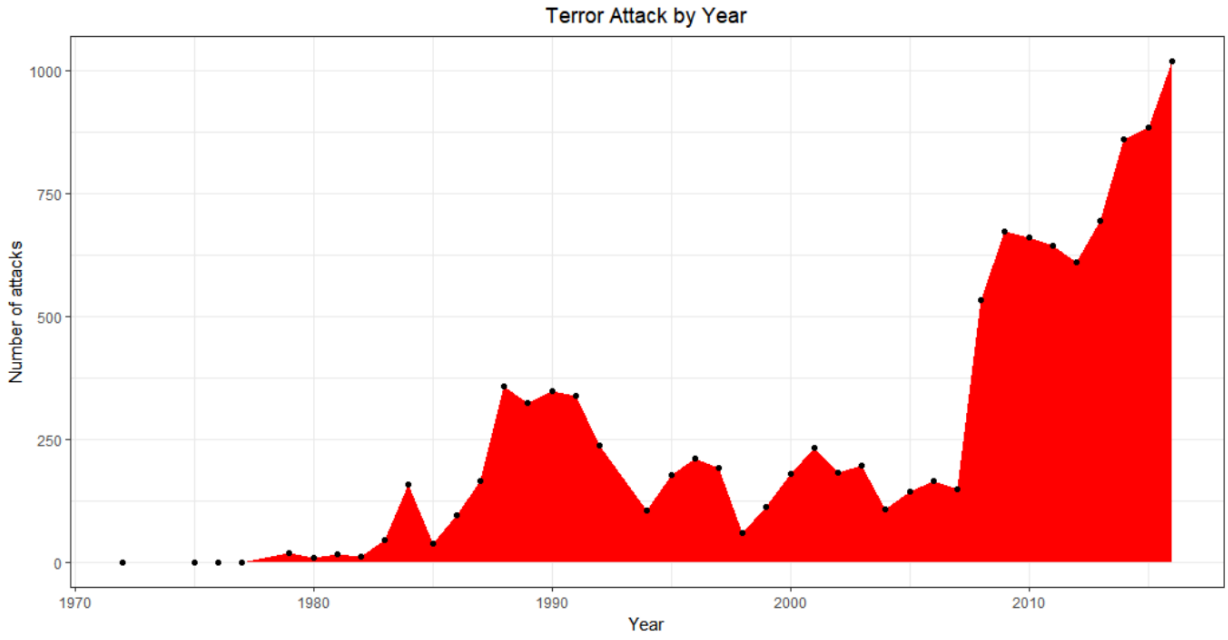


Figure 2

As we can see from Figure 2, number of terror attacks are increasing by year. After 2010, attacks have increased tremendously.

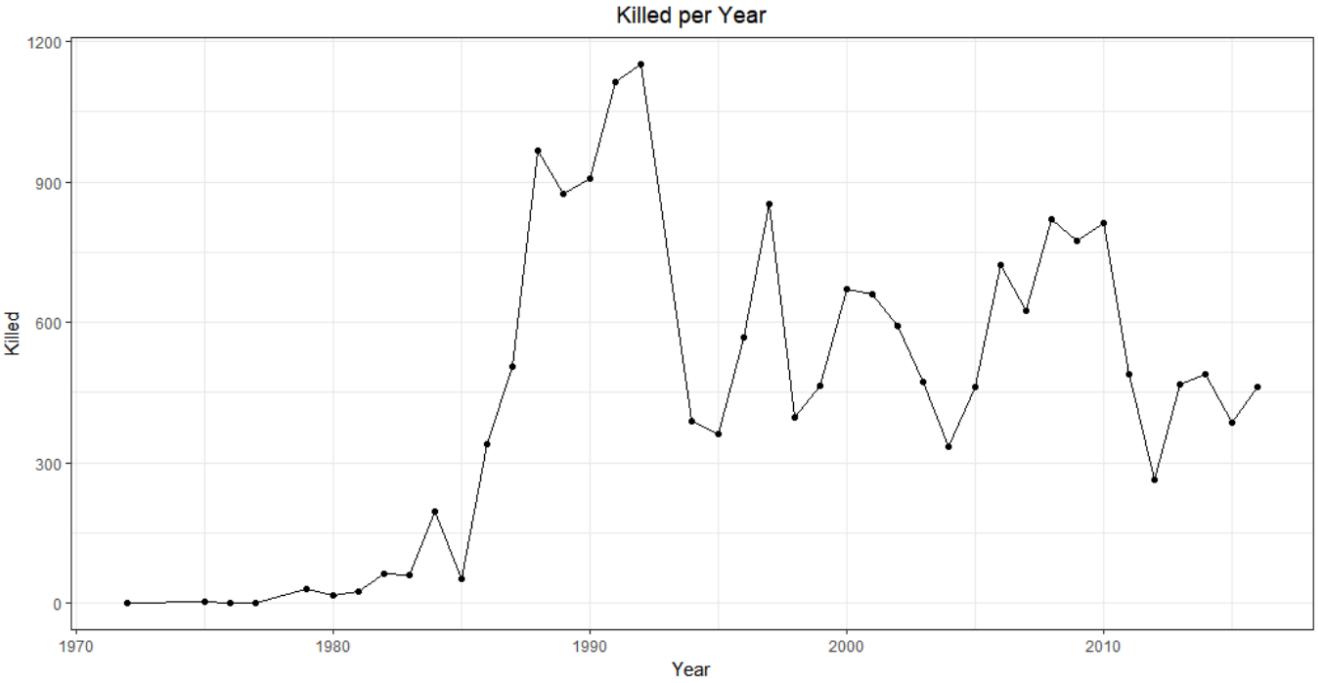


Figure 3

Figure 3, line chart depicts the number of killed per year due to terror attacks. Number of casualties have augmented in 1985 -1992

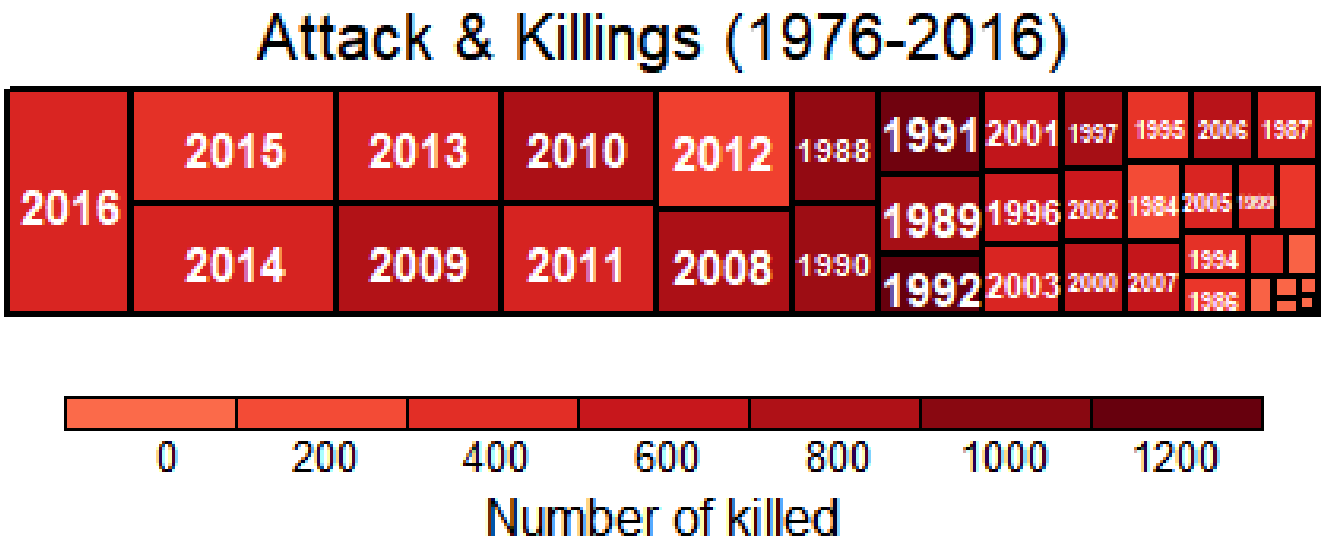


Figure 4

The above Heat map describe the number of attacks and casualties by year. Here size of rectangle box is used for number of attacks and color is used for casualties. It clearly says, in the period of 1991-1992, the casualties are more than 2400 combined while number of attack was relatively low.

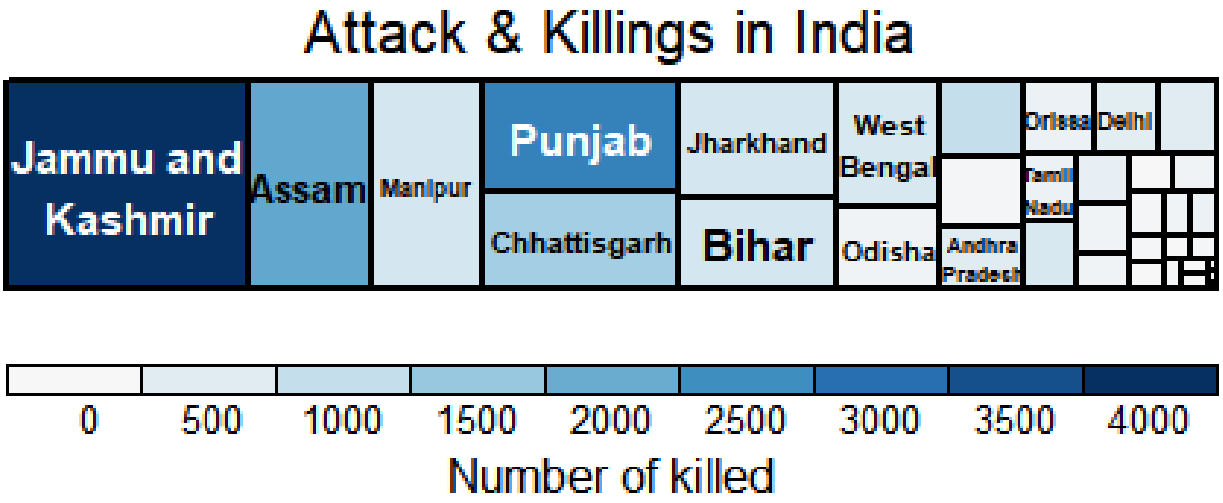


Figure 5

The above heat chart shows the attacks and casualties data by state.

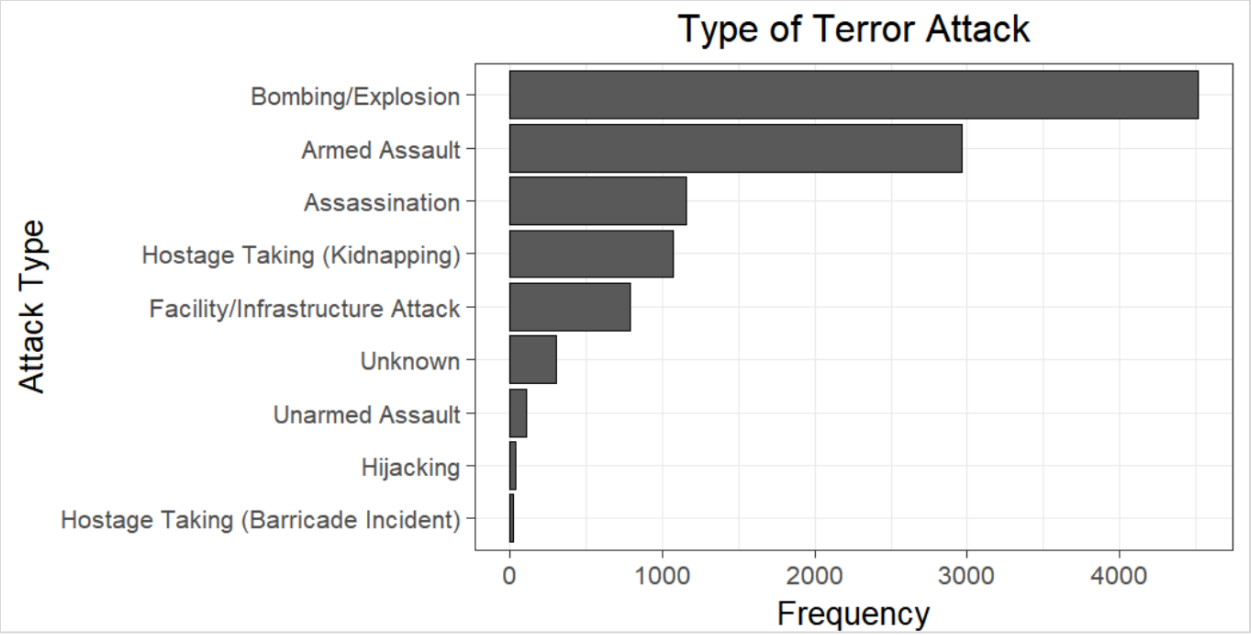


Figure 6

The bar chart shows the type of terror attack happened in India since 1976.

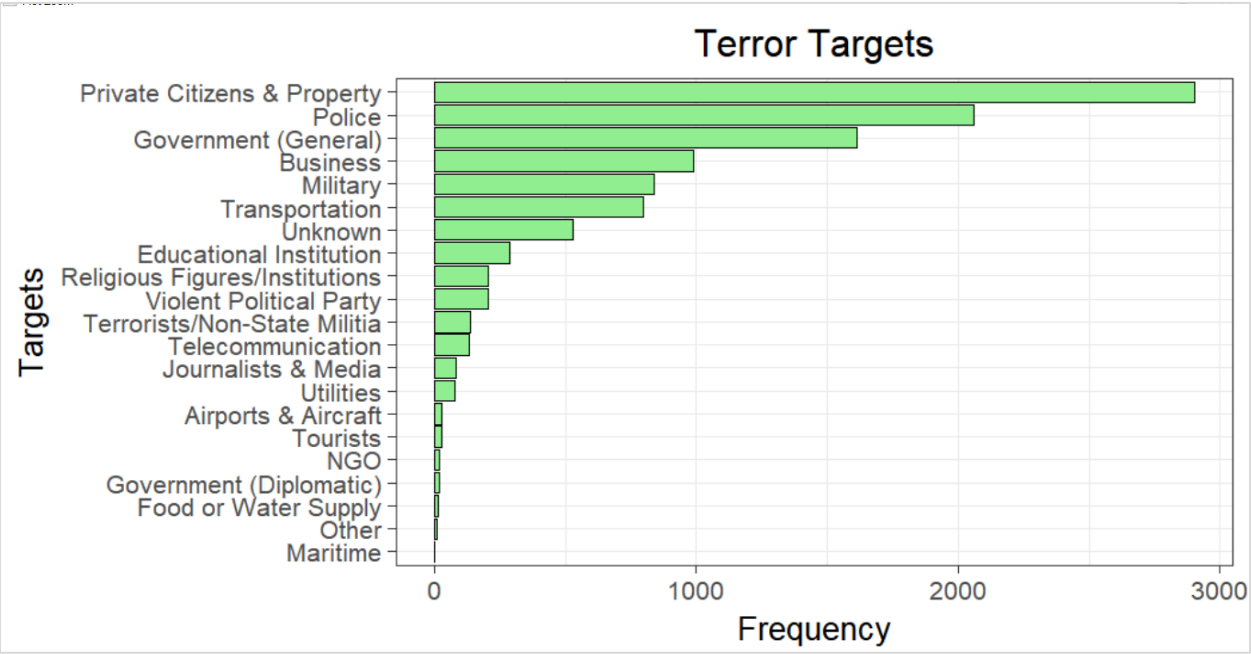


Figure 7

Fig.7 shows the targets chosen for the attack by terrorist group since 1976. Mostly, Terrorist attack on citizen, police and government.

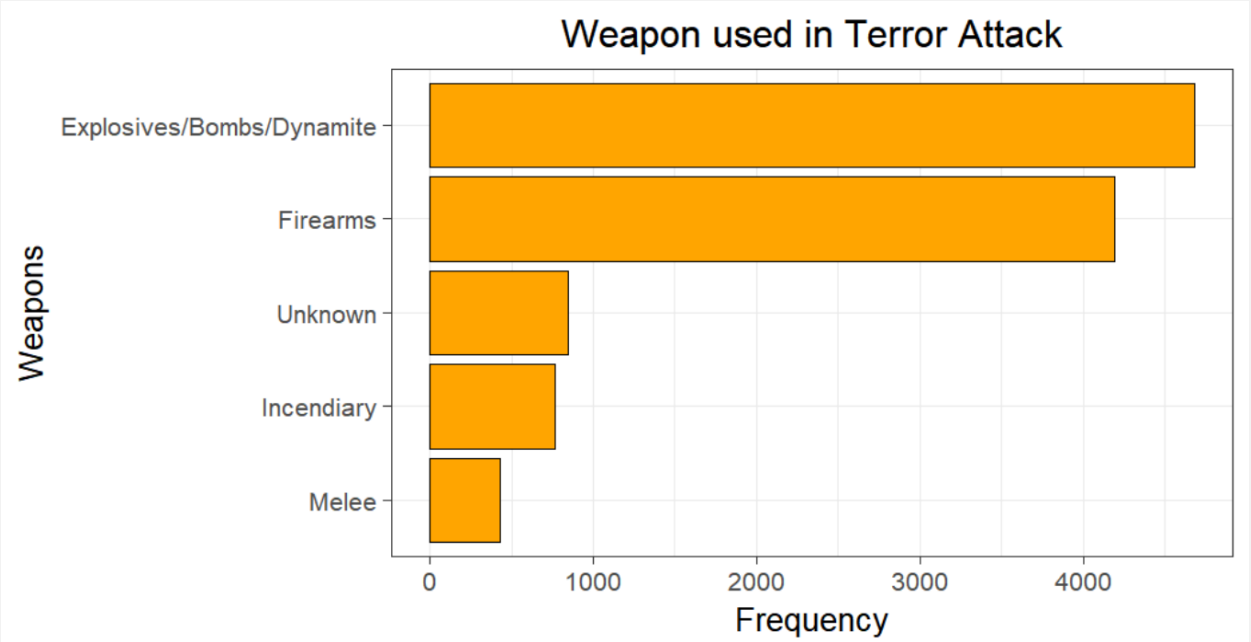


Figure 8

Fig 8 shows the weapons used for attack by terrorist. It is clearly evident that explosives, bombs and dynamite are being used the most since 1976.

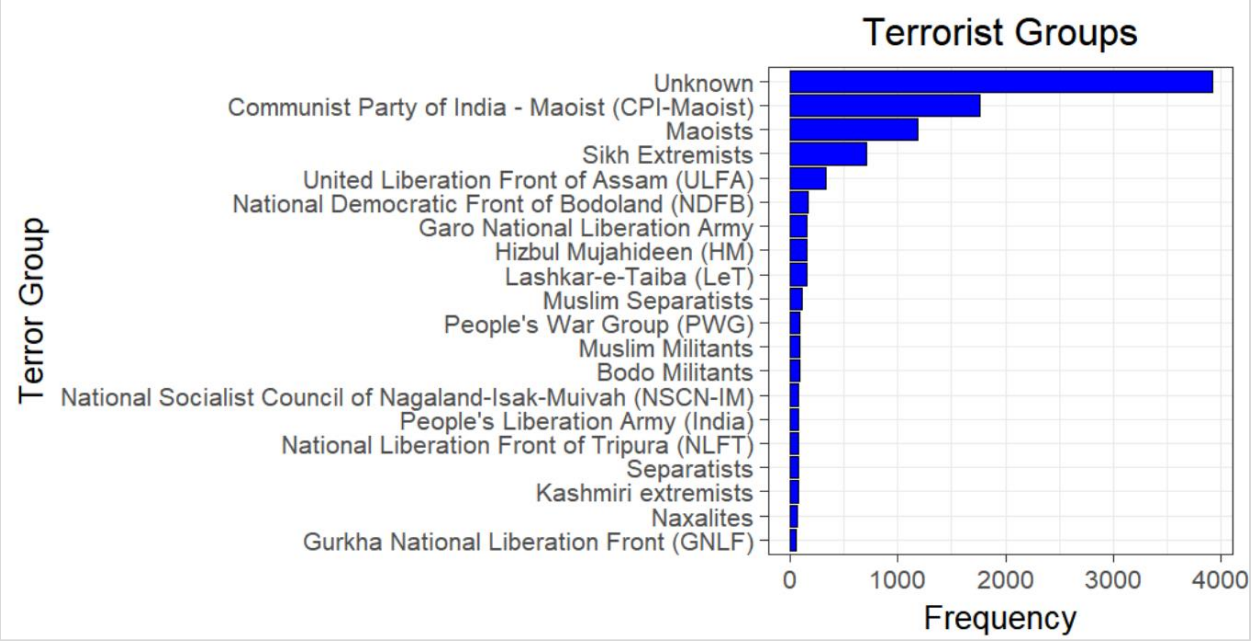


Figure 9

Fig 9, shows the terrorist group- Maoist who is mostly active through out 1976

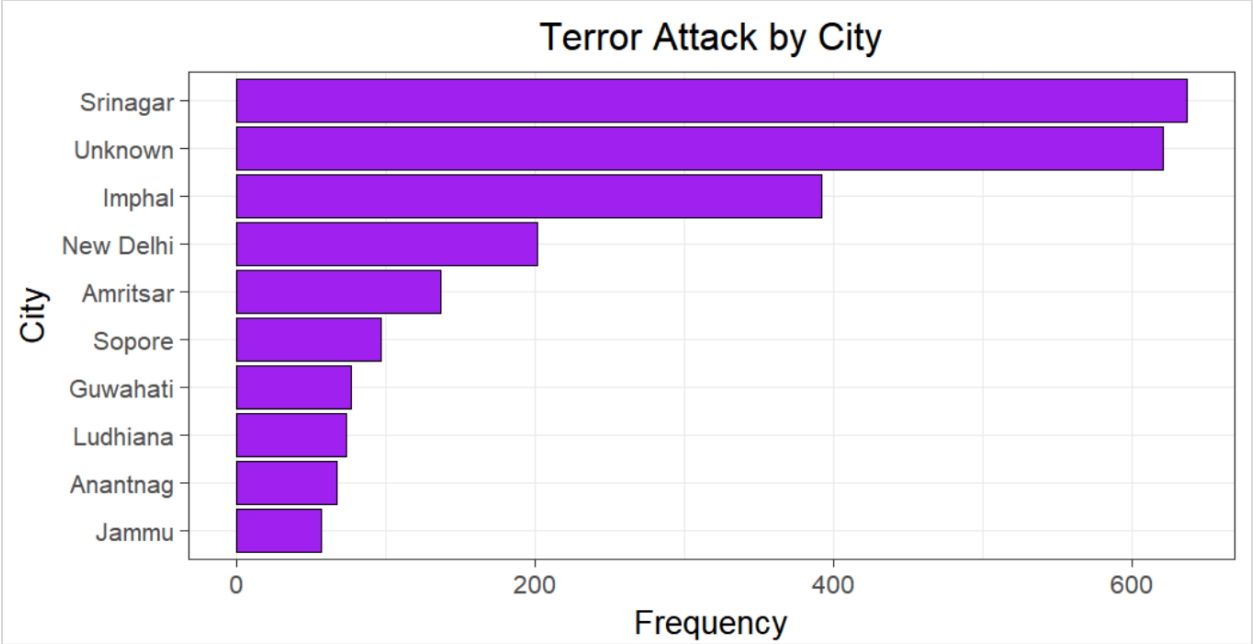
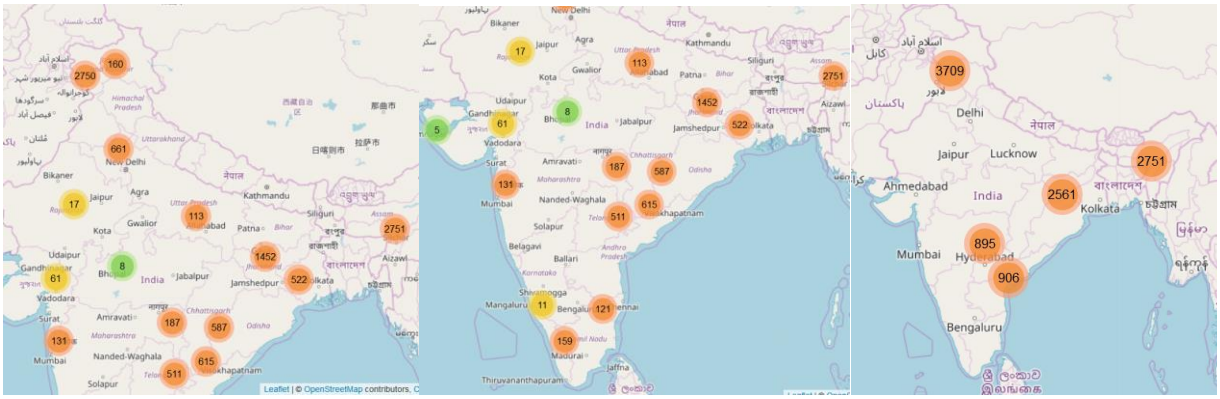


Figure 10

Fig 10 shows the top cities which has high number of attacks.

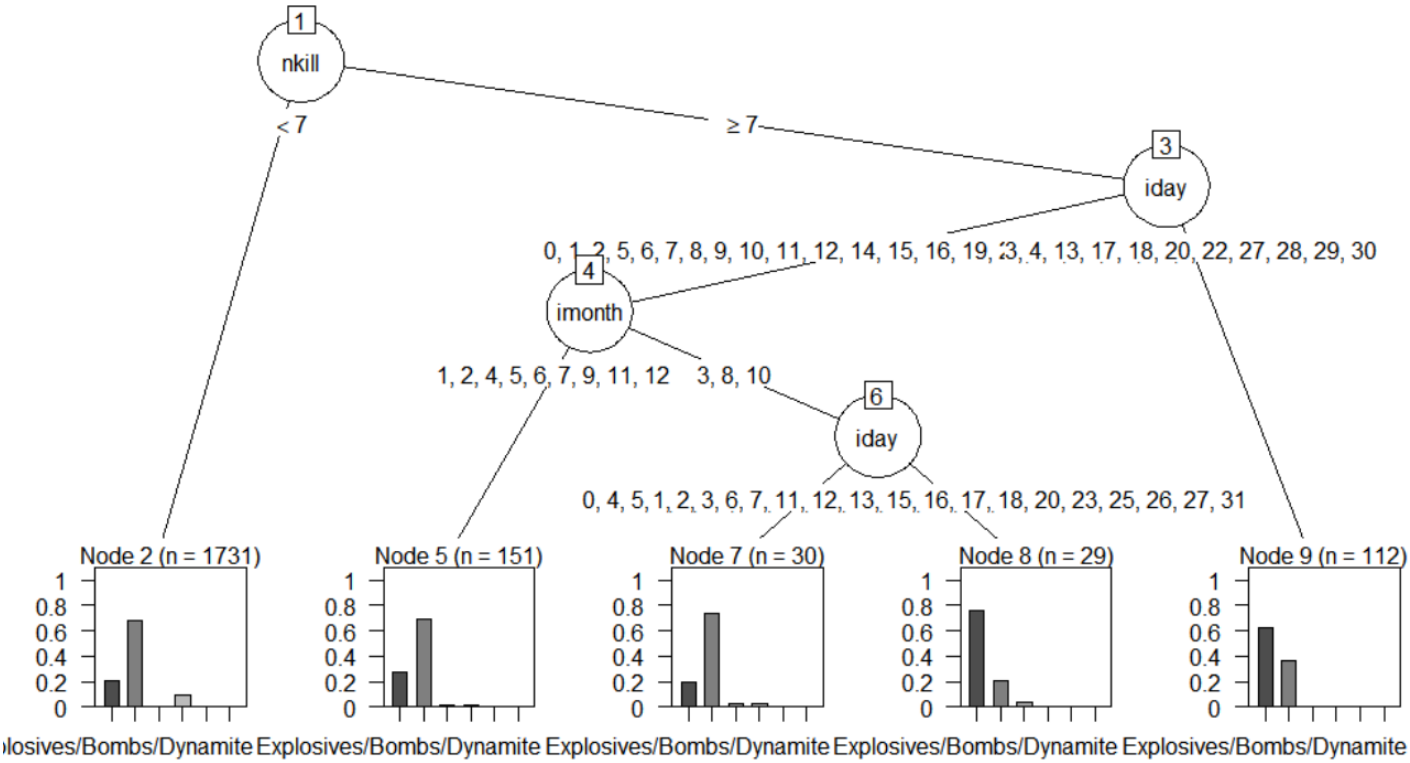


The above leaflet graph shows the attack per city. If we scroll the map its depicts the exact number of attacks.

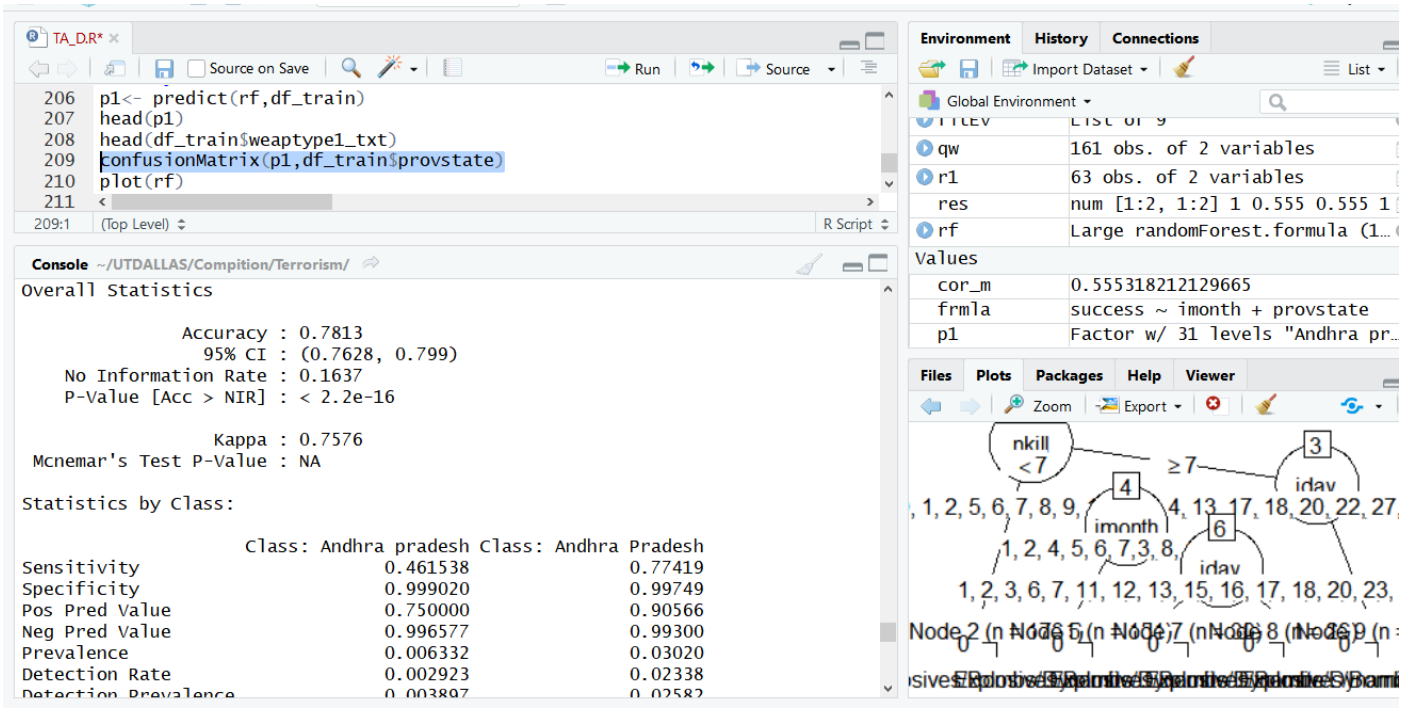
Prediction part

1. Prediction of weapon used

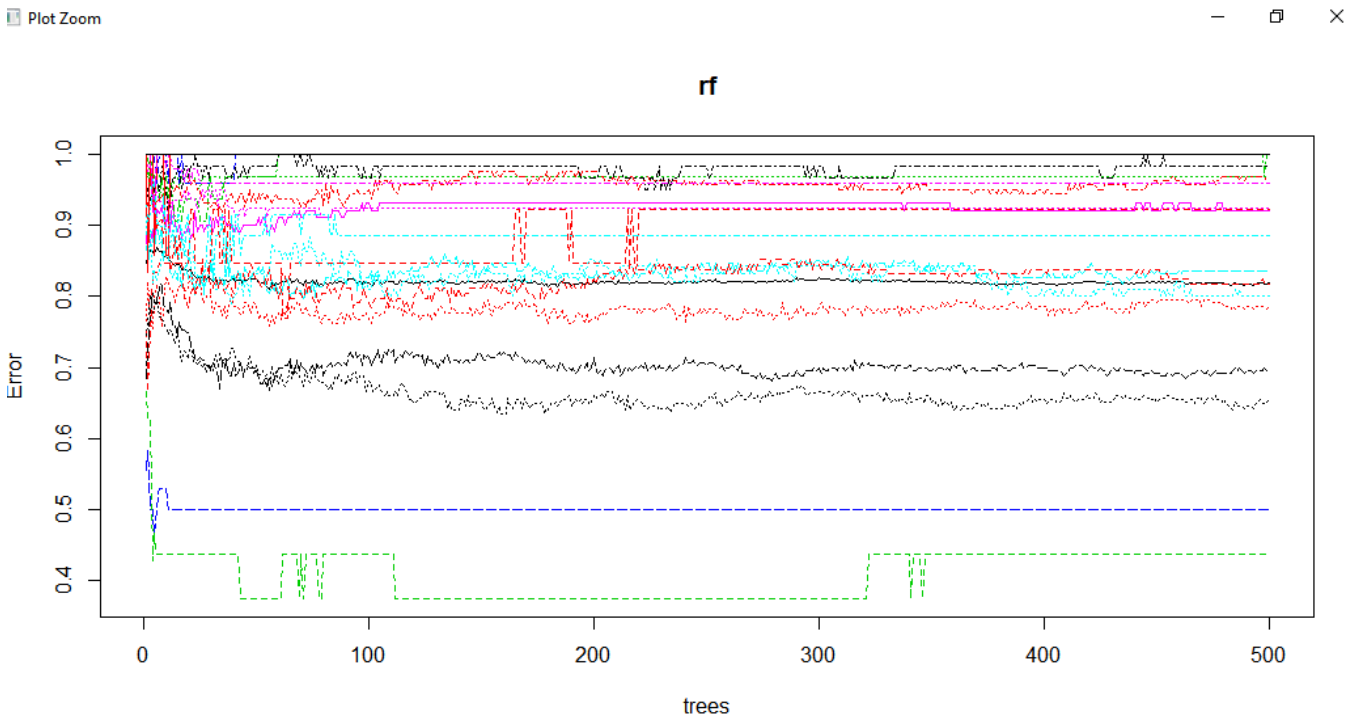
I am using decision tree from evtree package to predict weapon type. The terminal nodes contain box plot which tells the probability of which type of weapon has the chance to be used in the attack.



2. Prediction of state: - Below is the random forest algorithm result which shows the confusion matrix. I used random forest algorithm to predict the state where the next attack is going to happen.I divided the data in to two part train and test.



Below graph shows the error rate for random forest algorithm



Number of casualties' prediction: - I used linear regression to predict number of casualties using nwounds variable. R square is bad due to absence of other dependent variable

Interpretation

For every 4 wound persons is associated with the death of 1 person.

```
Residuals:
    Min      1Q  Median      3Q     Max
-23.695  -1.853  -1.853   0.424  79.147

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.853379   0.124398   22.94   <2e-16 ***
nwound      0.246135   0.009953   24.73   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.345 on 2051 degrees of freedom
Multiple R-squared:  0.2297,    Adjusted R-squared:  0.2293
```