Contents

1 Abstract

Terrorism. The word carries such ghastly connotations that we immediately start thinking of beards, assault rifles, screams, pain and death. But what do we know about terrorism really? We do know of the atrocities committed and the global grief that ensues but do we know why they do what they do? Imagine there was a way to understand/analyze the various motives behind a terrorist attack with the help of real, tangible data. Imagine a world where you could predict an imminent threat to a city or a nation. Imagine being able to foresee the number of casualties of an attack. This project is a stepping stone to that shore.

Our process of arriving at this topic involved looking into the most prevalent annoyances of the world. What do we need solved the most at the moment, was what we were asking ourselves. We started looking at everything from why our heaters don't work often to why world peace doesn't exist. Terrorism and the implications of it stood out the most. Numerous research papers and countless hours of philosophical rumination after, we realized there's still so much we don't know about the subject. Why do the countries that get attacked get attacked? Can we ever predict the reason behind the thought that leads to terrorists doing what they do? Though the papers helped us grasp the basic ideologies/motivations behind a terrorist attack, we wanted to draw solid conclusions from real world data sets and that's when we decided we had to do this.

2 Overview

2.1 ***Objectives***

- Can parameters like GDP, freedom, the scale of corruption and previous attack parameters of a nation help predict future terrorist attack zones?
- Can the number of casualties be predicted from our dataset-exclusive features?
- How prone is a happy country to terrorist attacks vs an unhappy or war-torn country?
- Where, geographically speaking, are terrorists most active?

2.2 ***Research findings***

- The GTD is the most comprehensive unclassified database on terrorist event in the world.
- It includes more than 27,000 bombings, 12,000 assassinations and 2,900 kidnappings since 1970.
- For each incident, information is available on the location and date of the corresponding incident, summing up to 128 variables.

2.3 ***Methodology***

- Forward selection, backward selection, Mallow's Cp statistics for feature selection.
- LDA/logistic regression/RandomForest for predicting future attacks.
- Linear regression/logistic regression/RandomForest for predicting casualties.

- Repeated Cross Validation for re-sampling.
- Please note that we weren't able to use KNN regression because of the presence of a lot of missing values.

3 Data sources

Our data is a combination of two individual data sets called 'Global Terrorism Database' and 'World Happiness Report' that can be found on Kaggle.

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

Variables: >100 variables on attack locations, perpetrators, attacks types, weapons used, etc..

The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 155 countries by their happiness levels, was released at the United Nations at an event celebrating International Day of Happiness on March 20th. The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

Variables: Happiness score estimates, economic production, social support, life expectancy, freedom, absence of corruption, and generosity of most of the nations of the world.

4 Data Processing

The first major issue we had to deal with with respect to the data was the presence of a a lot of missing values. We tried to impute the missing data with the median/mode method but that lead to highly inaccurate values that completely threw off our analysis/predictions. Right as we were looking into imputational methods involving predictions, we realized we cannot, rather should not impute some of the said values as it just wouldn't be right on our part. One cannot just decide if a country is happy or not or predict the location of a country's previous attack. That would change things and we unanimously decided not to do it.

Next, all there was left to do was merge the data sets, remove the redundant parameters (redundant in the sense that we thought the parameters wouldn't be useful for our purposes) and rename some of them for easy interpretation. 78 variables including approxdate, extended, resolution, region, specificity and vicinity were removed.

5 Data Analysis

5.1 ***Visualization***

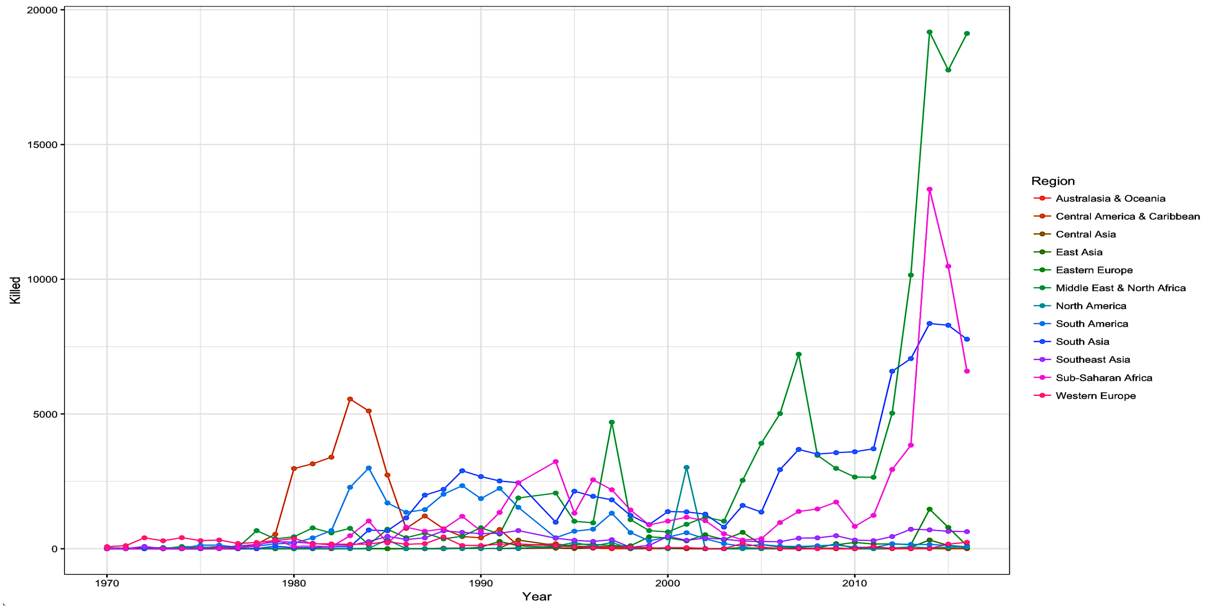


Figure 1: Killings per year per country

As we can see from Figure 1, Central America had the most killings in the 1980s. This was probably during the Central American crisis which began in the late 1970s. The Middle East & North Africa, South asia and Sub-Saharan Africa had mojor spikes in the number of killings after the year 2000. These correlate to the several attacks in the period including the battle of Iraq, battle of Baghdad, the conflicts in Lebanon, Iran and Syria and the wars in Kenya, Ethiopia and Sudan.

Next, we can see on Figure2 that the most common type of weapon used was bombs/explosives. It only makes sense for them to be the most common method because the terrorists would want to cause maximum death and destruction which is only possible with bombs. We also have to note that firearms are the next most used weapon. Even this makes sense because whats the next best thing after bombs? Guns. It is also important to note that the mode of many of the attacks is unknown.

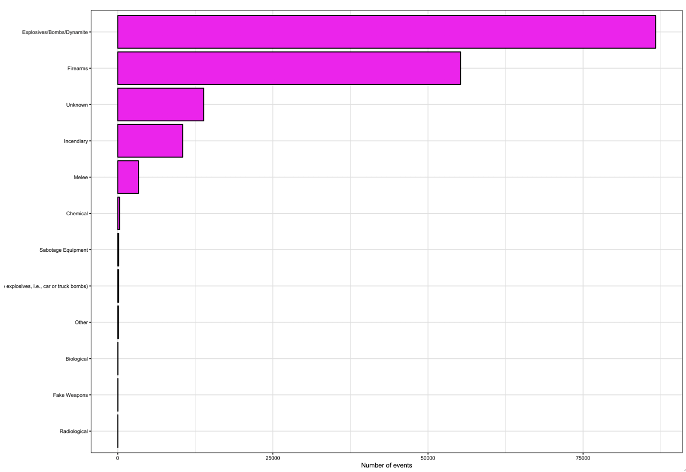


Figure 2: Common weapons used for the attacks

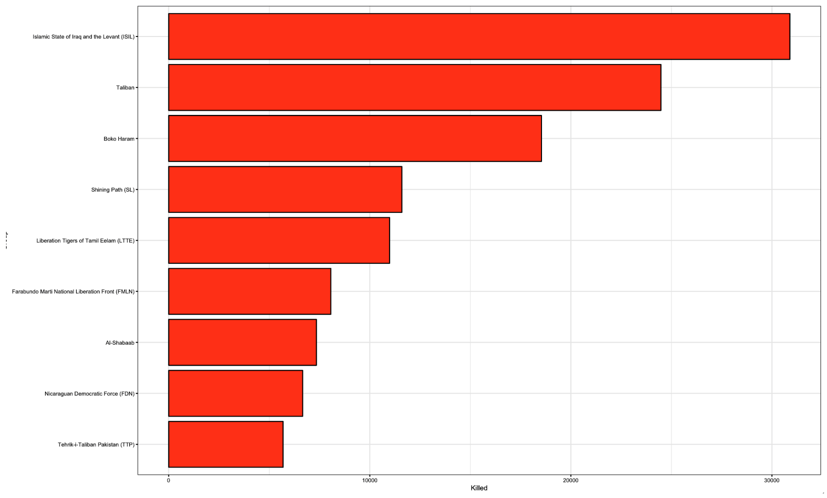


Figure 3: The number of killings per group

As you may have guessed, Figure 3 on the left here tells us that the most active terrorist organization is the Islamic State of Iraq and the Levant (ISIL). Second comes the seasoned one: The Taliban. We were surprised not to find Al-Qaeda but not so much seeing Boko Haram and Liberation Tigers of Tamil Eelam (LTTE) on the 3rd and 5th position respectively. A compelling point to take away here is the scale of death caused by a relatively less known group called the Shining path (SL).

Let us first notice the most prevalent and important words of the word cloud on the right here (Figure 4): Islamic, majority, minority, elections, Shiite, Iraq, sectarian, communities and retaliation. This word cloud tells more about the attacks and the basic thinking behind them than you may realize. The existence of 'Shiite' and 'Islamic' tell us many of the attacks were a direct result of clashes between Sunnis and Shiites, which are the major branches of Islam in the Middle East. 'Majority' and 'Minority' tell us how disputes between the majority and minority classes of religions might have lead to some of the attacks. We can also conclude that some of the attacks are a form of retaliation and some due to sectarianism within the communities of the world. Finally, we can also notice that 'Iraq' pops up a lot as well as a testament to all the horrors taking place in the country.

Figure 4: Word Cloud of the attack summaries

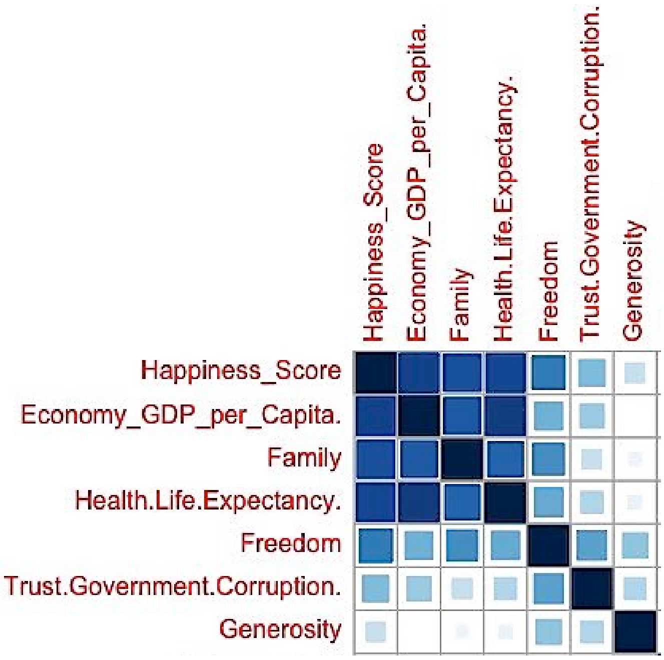


Figure 5: Correlation matrix of happiness score with its factors

Coming to the most important part, the Happiness score, let's see how correlated the happiness score of a country is with the factors deciding it from Figure 5. On careful observation, we can see that the happiness score is majorly related to the Gross Domestic Product (GDP), family presence and the life expectancy of a nation (albeit family being slightly less correlated). Apparently, the GDP decides how happy a country is more than the freedom one enjoys in the country. The level of corruption and generosity index have lesser effect on the happiness score. It is also interesting to note that the first four parameters: Happiness score, GDP, family and life expectancy are highly correlated with each other.

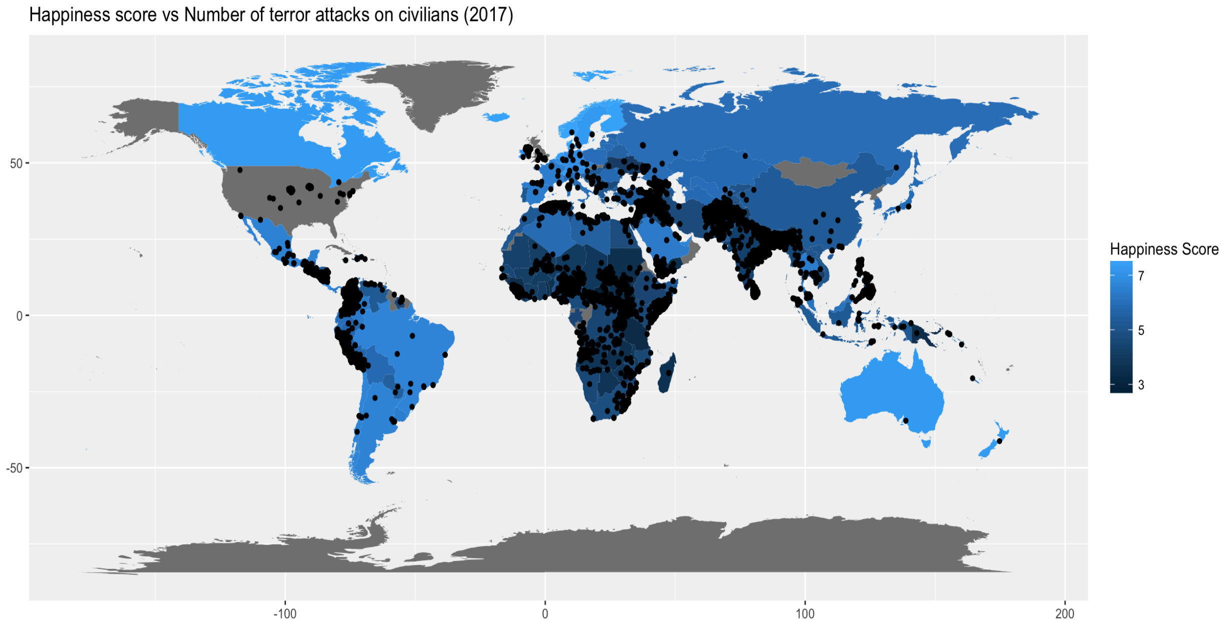


Figure 6: Happiness score vs. the number of attacks

Figure 6 above is the most important visualization of them all. This shows us how the happiness score of a country weighs against its probability of getting attacked. The darker shades of blue (and more towards black) are the most unhappy countries and the lighter shades of blue are the happy ones.

5.2 ***Feature selection***

Remember how we discarded some of the variables we thought weren't useful for our analysis? We employed the following methods to further weed out the insignificant ones:

- Forward selection
- Backward selection
- Mallow's Cp statistic

5.2.1 Forward Selection

Forward selection is a type of regression where we build regression model from a set of candidate predictor variables by entering predictors based on p values, in a stepwise manner until there is no variable left to enter any more. To put it simply, we go about adding variables to the regression with the least p-value each time. The model should include all the candidate predictor variables.

Candidate parameters:

1. attackmode
2. targettype
3. targsubtype
4. weaptype
5. property_value
6. Economy_GDP_per_Capita
7. Family
8. Health.Life.Expectancy
9. Freedom
10. Generosity
11. Trust.Government.Corruption

Parameters removed:

1. property_value
2. weap_type

Final model output:

Model Summary

R	0.949	RMSE	0.728
R-Squared	0.900	Coef. Var	19.200
Adj. R-Squared	0.900	MSE	0.530
Pred R-Squared	0.900	MAE	0.559

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	792663.065	9	88073.674	166188.709	0.0000
Residual	87969.956	165993	0.530		
Total	880633.020	166002			

Parameter Estimates

Model	Beta	Std.Error	Std.Beta	t	Sig	lower	upper
(Intercept)	1.912	0.010		197.149	0.000	1.893	1.931
Trust.Government.Corruption.	-1.156	0.004	-0.428	-277.478	0.000	-1.164	-1.148
Family	2.280	0.010	0.405	238.190	0.000	2.262	2.299
Generosity	1.746	0.016	0.102	107.573	0.000	1.714	1.777
Health.Life.Expectancy.	1.819	0.016	0.164	110.893	0.000	1.786	1.851
Economy_GDP_per_Capita.	-0.707	0.010	-0.099	-69.092	0.000	-0.727	-0.687
Freedom	0.746	0.015	0.048	48.723	0.000	0.716	0.775
attackmode	-0.011	0.001	-0.009	-12.012	0.000	-0.013	-0.010
number_kill	-0.001	0.000	-0.007	-8.746	0.000	-0.002	-0.001
targettype	0.002	0.000	0.005	6.824	0.000	0.001	0.002

Elimination Summary

	Removed
1	property_value
2	weaptype

5.2.2 Backward Selection

Backward selection is a type of regression where we build regression model from a set of candidate predictor variables by removing predictors based on p values, in a stepwise manner until there is no variable left to remove any more. This is essentially the opposite of Forward Selection. In other words, we begin the regression with the full feature set and we keep removing variables that lead to high p-values. The model should include all the candidate predictor variables.

Please note that the candidate parameters are the same as forward selection. Following are the results of the backward selection method:

We are eliminating variables based on p value...

Parameters removed:

1. property_value

No more variables satisfy the condition of p value = 0.3

Final model output:

Model Summary

R	0.838	RMSE	0.541
R-Squared	0.702	Coef. Var	10.370
Adj. R-Squared	0.702	MSE	0.293
Pred R-Squared	0.702	MAE	0.445

RMSE: Root Mean Square Error
MSE: Mean Square Error
MAE: Mean Absolute Error

ANOVA

	Sum of Squares	DF	Mean Square	F	Sig.
Regression	109176.118	11	9925.102	33888.312	0.0000
Residual	46391.398	158399	0.293		
Total	155567.516	158410			

Parameter Estimates

Model	Beta	Std.Error	Std.Beta	t	Sig	lower	upper
(Intercept)	2.147	0.008		261.231	0.000	2.131	2.163
attackmode	-0.008	0.001	-0.015	-8.048	0.000	-0.009	-0.006
targettype	0.013	0.001	0.087	9.802	0.000	0.011	0.016
targsubtype	-0.002	0.000	-0.076	-8.510	0.000	-0.003	-0.002
weaptype	-0.003	0.001	-0.006	-3.239	0.001	-0.004	-0.001
number_kill	-0.002	0.000	-0.021	-15.310	0.000	-0.002	-0.002
Economy_GDP_per_Capita.	-0.123	0.009	-0.041	-13.787	0.000	-0.141	-0.106
Family	1.498	0.008	0.458	176.924	0.000	1.482	1.515
Health.Life.Expectancy.	1.772	0.013	0.362	138.317	0.000	1.746	1.797
Freedom	1.138	0.013	0.166	87.930	0.000	1.112	1.163
Generosity	0.255	0.011	0.033	22.549	0.000	0.233	0.277
Trust.Government.Corruption.	2.272	0.025	0.140	91.800	0.000	2.224	2.321

Elimination Summary

	Removed	R-Square	R-Square	C(p)	AIC	RMSE
1	property_value	0.7018	0.7018	16566.2007	255035.7436	0.5412

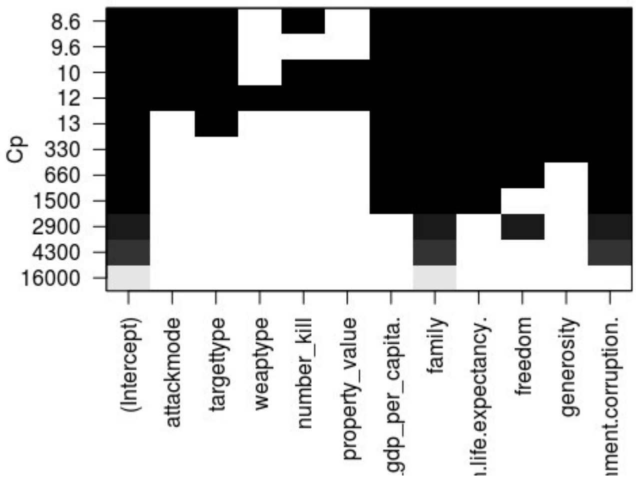
5.2.3 Mallows's Cp

Mallows' Cp is a statistic that can help you choose between competing multiple regression models. Mallows' Cp compares the full model to models with the best subsets of predictors. It helps you strike an important balance with the number of predictors in the model. A model with too many predictors can be relatively imprecise while a model with too few predictors can produce biased estimates. Remember that Mallows' Cp to compare regression models is only valid when you start with the same complete set of predictors.

A Mallows' Cp value that is close to the number of predictors plus the constant indicates that the model produces relatively precise and unbiased estimates.

A Mallows' Cp value that is greater than the number of predictors plus the constant indicates that the model is biased and does not fit the data well.

We have used the *regsubsets* function in R to find the best model using the 'exhaustive' method and plotting the output with the 'Cp' statistic. The plot is as follows:



After going through the results of the three methods, we can see that the weapon type and the property value of the damage previously caused are irrelevant in predicting the future attacks.

6 Model Training

We have two things to predict in our project:

- Predict the occurrence of future terrorist attack on a country given it's present condition of well being in terms of factors like corruption, health etc.
- Prediction of the number of casualties given the weapon types used, etc.

6.1 *Predicting future attacks*

Since a column related to this is not available in our xls, we add a column for future attack susing the baseline that the number of attacks have been less in a happy country. So if the happiness rank of a country is more than 100 there are higher chances of future attacks than on countries with rank lower than 100. We then factorize the response for the purpose of our regression as shown in the code below:

```
terrorism_FA$FutureAttack <- (terrorism_FA$Happiness_Rank > 100)
terrorism_FA$FutureAttack = factor(terrorism_FA$FutureAttack)
```

The data is then partitioned into 80-20 train-test sets and the three types of regressions are performed on the train data.

```
lm1_cv_terr <-train(as.numeric(FutureAttack)~attackmode+Economy_GDP_per_Capita.+...,
                        data = terrorism_FA, method = "lda",
```

```
                              trControl = ctrlTerr)

glm1_cv_terr <-train(FutureAttack~attackmode+Economy_GDP_per_Capita.++...,
                              data = terrorism_FA, method = "glm",
                              trControl = ctrlTerr)

RF_Terr_fit <-randomForest(FutureAttack ~ attackmode+Economy_GDP_per_Capita.+...,
                                        data=trainingTerr,
                                        importance=TRUE,
                                        ntree=200)
```

We found 200 to be the best value of entry for RandomForest in our particular case because we have a good amount of data and moe the value of entry, better the results. Please not that the parameters in trainControl() specify the details for training the regression model on the global terrorism data with happiness score. For the purpose of resampling, we choose to use "repeated cv" which is cross validation repeated **n** number of times.

```
ctrlTerr <- trainControl(method = "repeatedcv",  repeats = 5)
```

Cross Validation helps us obtain the test error rate of a statistical learning method/model of regression. Repeating it 'n' number of times makes it similar to k-fold cross validation. Since the validation is repeated, the final test error rate is the average of the test error rate of each of the 5 repeats. Also training it on 80% of the observations gives it the advantage of LOOCV. Thus the results of this training followed by validation give us results with low bias and low variance as well.

6.2 ***Predicting casualties***

Unlike the previous prediction this column is available in our xls. Also, we don't need to factorize since the values in casualties are quantitative.

So we directly start with data partitioning with 80% as train and 20% as the validation/testing set and perform our regressions on the train set.

```
#Linear
lm1_cv <- train(number_kill~success+multiple_attack+attackmode+targettype+...,
                              data = trainingTerr_C,
                              method = "lm",
                              trControl = ctrlTerr_C,na.action = na.pass

#Logistic
glm1_cv_C <- train(number_kill~success+multiple_attack+attackmode+targettype+...,
                              data = trainingTerr_C,
                              method = "glm",
                              trControl = ctrlTerr_C, na.action = na.pass)

#RandomForest
RF_Terr_fit_C <- randomForest(number_kill~success+multiple_attack+attackmode+...,
                              data=trainingTerr_C,
                              importance=TRUE,
                              ntree=50)
```

We decided 50 to be the best value of entry in this case because of the lack of data and a higher entry value would have given us incorrect/improper results. Also, note that the parameters in trainControl() specify the details for training the regression model on the global terrorism data with happiness score. For the purpose of resampling, we chose to use “repeated cv” which is cross validation repeated n number of times

```
ctrlTerr_C <- trainControl(method = "repeatedcv", repeats = 5)
```

Cross Validation helps to give the test error rate of a statistical learning method or the model of regression in simpler terms. Repeating it for a number of times makes it similar to k-fold cross validation. Since the validation is repeated, the final test error rate is the average of test error rate of each of the 5 repeats. Also training it on a major 80% observation gives it the advantage of LOOCV. Thus the results of this training followed by validation shall give us results with low bias and low variance as well. The results of training on each sample are shown below:

	RMSE	Rsquared	MAE	Resample
1	8.236655	0.0050552687	2.926947	Fold01 Rep1
2	7.993493	0.0056661129	2.935250	Fold02 Rep1
3	16.120262	0.0009100217	3.102228	Fold03 Rep1
4	15.435170	0.0014960939	3.079869	Fold04 Rep1
5	8.077991	0.0051264995	2.874326	Fold05 Rep1
6	8.467224	0.0050040012	2.933742	Fold06 Rep1
7	14.195815	0.0014211088	3.040424	Fold07 Rep1
8	8.945177	0.0037126035	2.861024	Fold08 Rep1
9	10.666022	0.0034478544	2.970937	Fold09 Rep1
10	12.652409	0.0024798042	2.982636	Fold10 Rep1

7 Model Validation

Validation for predictions from regression models is of utmost importance. It is well possible to develop a model with apparently provides adequate predictions, but fails when validated on new observations.

7.1 *Validation for future attacks*

Since we have a goal to predict future attacks, validation of the new observation needs to taken care of. We have used confusion matrix and AUCROC curve to validate the data using the fitted model.

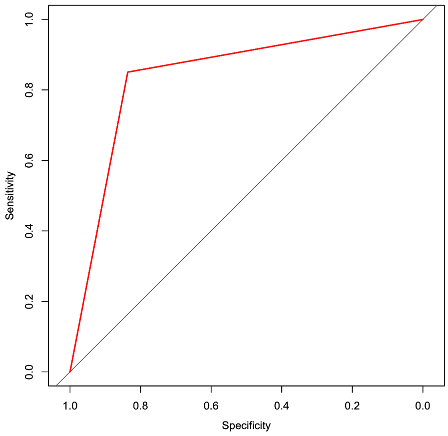
- **Confusion matrix**: A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix.
- **ROC curve:** It stands for receiver operating characteristiccurve which is a graphical plot that illustrates thetrue positive rate(TPR) against thefalse positive rate(FPR) at various threshold settings.

7.1.1 Logistic Regression

The confusion matrix for the validation gives the below matrix:

Prediction	**TRUE**	**FALSE**
FALSE	15975	2172
TRUE	3153	12218

Only 2172 samples re predicted to be wrong as through he training mode the possibility of the attack is low but after the validation it depicts a high possibility. In the similar manner, 3172 samples are predicted the low possibility of attack whereas the training data shows a high possibility. The total accuracy obtained by confusion matrix tends to be **84.11**%.



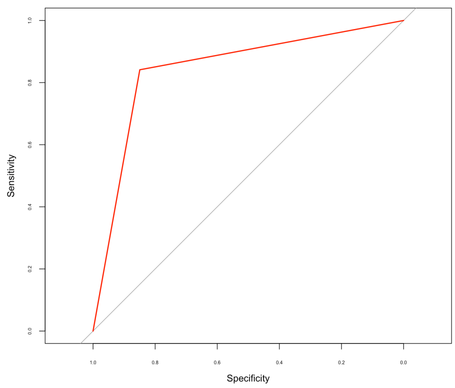
The area under AUC ROC curve is **0.8421**.

7.1.2 LDA

The confusion matrix for the validation gives the below matrix:

Prediction	**TRUE**	**FALSE**
FALSE	16248	2284
TRUE	2880	12106

Only 2284 samples re predicted to be wrong as through the training mode the possibility of the attack is low but after the validation it depicts a high possibility. In the similar manner, 2880 samples are predicted the low possibility of attack whereas the training data shows a high possibility. The total accuracy obtained by confusion matrix tends to be **84.59**%.It is slightly higher than the logistic regression.



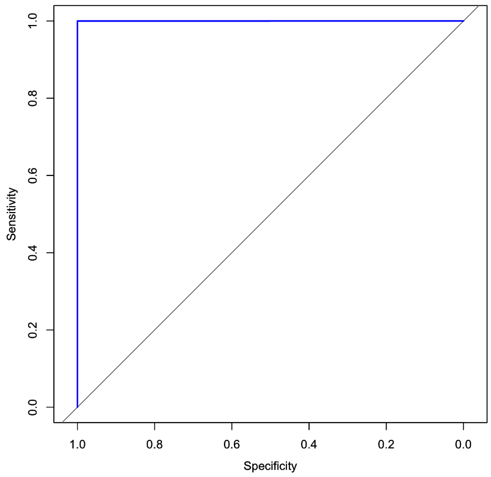
The area under AUC ROC curve is **0.8454**.

7.1.3 RandomForest

The confusion matrix for the validation gives the below matrix:

Prediction	**TRUE**	**FALSE**
FALSE	19128	0
TRUE	0	14390

It is the scenario for a perfect classification which is highly likely to be true. The total accuracy obtained by confusion matrix tends to be 100%. Hence we ignored the random forest results in our case.



The area under AUC ROC curve is **1**.

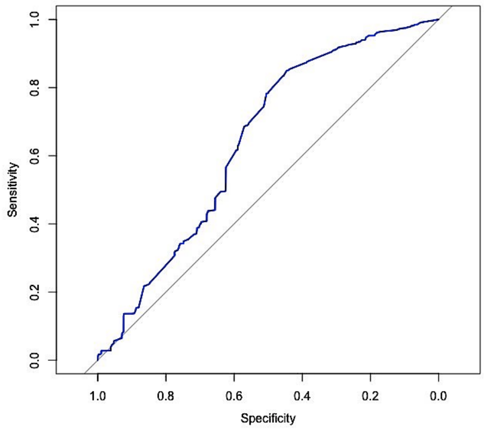
After all the observations, we conclude that the **LDA** performed better for our data with an accuracy of 84.59% for prediction the future attacks.

7.2 *Validating the casualties*

For validating the casualties we are using the Root Mean Square Error and AUC ROC curve techniques.

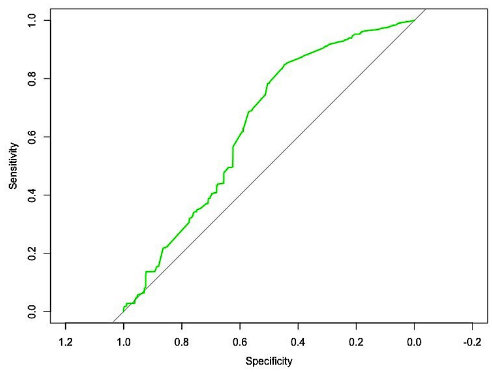
7.2.1 Logistic Regression

Logistic Regression: The RMSE value obtained is **8.02** whereas the AUC ROC curve is **0.6588**



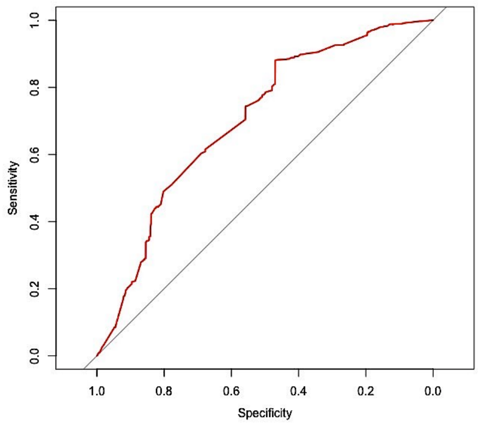
7.2.2 Linear Regression

Logistic Regression: The RMSE value obtained is **8.02** whereas the AUC ROC curve is **0.6588**



7.2.3 RandomForest

Logistic Regression: The RMSE value obtained is **7.94** whereas the AUC ROC curve is **0.71**



After all the observations, we conclude that **Random forest** is the best model since it gives low RMSE value as compared to others but the AUC ROC curve has a high value of 0.71 in comparison to other models.

8 Conclusion

Based on our analysis we can say that LDA and Logistic regression performed well with an accuracy of about 85% without any bias. RandomForest gave RoC and accuracy of 100% which too good to believe as this may be a result of overfitting of data. Given the present condition of a country in terms of factors like Economy_GDP_Per_Capita, Family.Health.Life.Expectancy, Freedom, Generosity, Trust.Government.Corruption, we were able to predict the occurrence of a future attack. Thus if the factors change leading to an increase in the chances of future attack, the data can be used as an alarm for the government.

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE  TRUE
     FALSE 16248  2284
     TRUE   2880 12106

               Accuracy : 0.8459
                 95% CI : (0.842, 0.8498)
    No Information Rate : 0.5707
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6872
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8494
            Specificity : 0.8413
         Pos Pred Value : 0.8768
         Neg Pred Value : 0.8078
             Prevalence : 0.5707
         Detection Rate : 0.4848
   Detection Prevalence : 0.5529
      Balanced Accuracy : 0.8454

       'Positive' Class : FALSE
```

(a) LDA

```
Confusion Matrix and Statistics

          Reference
Prediction FALSE  TRUE
     FALSE 15975  2172
     TRUE   3153 12218

               Accuracy : 0.8411
                 95% CI : (0.8372, 0.845)
    No Information Rate : 0.5707
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6785
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.8352
            Specificity : 0.8491
         Pos Pred Value : 0.8803
         Neg Pred Value : 0.7949
             Prevalence : 0.5707
         Detection Rate : 0.4766
   Detection Prevalence : 0.5414
      Balanced Accuracy : 0.8421

       'Positive' Class : FALSE
```

(b) Logistic Regression

Figure 7: Summaries for LDA and lm

Also, based on our analysis we can say that RandomForest model with RMSE of 7.94 performed well although we will need more data on casualties in order to make better predictions. Whereas Linear and logistic regression gave RMSE of 8.02 which is good but not better than RandomForest. Depending on the details of an attack the government of a country which is prone to terrorist attacks can make preparations to attend the casualties with the basic survival needs. However due to unavailability of casualties in existing file, there is a high bias with an accuracy of roughly 70%.

9 Source Code

The following libraries have been made use of for this project. Please click on them to know more about their documentation and dependencies.

leaps, randomForest, Metrics, rsq, rworldmap, ggplot2, leaflet, treemap, corrplot, tm, corpus, tidytext, tidyr, wordcloud, knitr, kableExtra, formattable, dplyr, olsrr, caret, pROC.

10 References

1. Borooah, Vani K. 2009. "Terrorist Incidents in India, 1998-2004: A Quantitative Analysis of Fatality Rates." Terrorism & Political Violence 21:476-498

2. Choi, Seung-Whan. 2010. "Fighting Terrorism through the Rule of Law?" Journal of Conflict Resolution 54:940-966.

3. Enders, Walter, Todd Sandler, and Khrsrav Gaibulloev. 2011. "Domestic Versus Transnational Terrorism: Data, Decomposition, and Dynamics." Journal of Peace Research 48:319-337.

4. Gaibulloev, Khusrav, and Todd Sandler. 2011. "The Adverse Effect of Transnational and Domestic Terrorism on Growth in Africa." Journal of Peace Research 48:355-371.

5. Godwin, Alex, Remco Change,Robert Kosara, and William Ribarsky. 2008. "Visual Analysis of Entity Relationships in Global Terrorism Database" Proceedings of SPIE 6983, 69830G; doi:10.1117/12.778084

6. Kis-Katos, Krisztina, Helge Liebert, and Gunther G. Schulze. Forthcoming. "On the Origin of Domestic and International Terrorism." European Journal of Political Economy

7. LaFree, Gary. 2009. Criminology's third war: Special issue on terrorism and responses to terrorism. Criminology and Public Policy 8:431-444

8. LaFree, Gary and Laura Dugan. 2007. "Introducing the Global Terrorism Database." Terrorism and Political Violence 19:181-204.

9. LaFree, Gary and Laura Dugan. 2009. "Tracking Global Terrorism, 1970-2004." In To Protect and to Serve: Police and Policing in an Age of Terrorism, David Weisburd, Thomas Feucht, Idit Hakimi, Lois Mock and Simon Perry (eds.). New York: Springer.

10. LaFree, Gary, Laura Duga, Min Xie, and Piyusah Singh. Forthcoming. "Spatial and Temporal Patterns of Terrorist Attacks by ETA 1970-2007." Journal of Quantitative Criminology.