Migration Patterns in the United States and Their Effect on Terror

Alan H. Blair

George Mason University  
4400 University Drive

Fairfax, VA 22030  
(703) 966-5355

alan.homer.blair@gmail.com

Anya A. Mityushina

George Mason University  
4400 University Drive

Fairfax, VA 22030  
(512) 965 - 6418

mityushina.a@gmail.com

**ABSTRACT**

This paper evaluates the frequency and outcomes of terrorist acts on United States soil by foreign-born immigrants and U.S. born citizens. The research team uses a machine learning and data mining approach to identifying key attributes of the subpopulations that are terrorist prone. To acquire the necessary data, the research team started with STARTs dataset on US 2000-2014 terrorism incidents. [1] The GMU team performed extensive data wrangling to construct a comprehensive terrorism dataset that: (a) identified the individuals involved and their immigration status, (b) included FBI data on planned and attempted terror acts, and (c) extended incident cover to include 2015. In addition, the team actively searched diverse data sources (mainly via Google & Wikipedia) to fill any holes in the START coverage. This allowed the construction of statistical models that identify the key attributes. This new constructed dataset also provides a new possibility to examine the problem of terrorism with a systematic approach based on a feature vector and some target variables (i.e., Number of terror incidents, Number Killed, Number Wounded). The team finds that from 2000-2015 the Foreign Born population did not commit any acts of terror. The only exception is immigrants who are born orprediction with Boosting. The research is set to provide initial guidance in terrorism and foreign policy.

**Figure 1: The Changing Face of America, 1965 - 2065**

**CCS Concepts**

• Information systems~Data mining • Information systems~Data dictionaries

• Computing methodologies~Unsupervised learning.

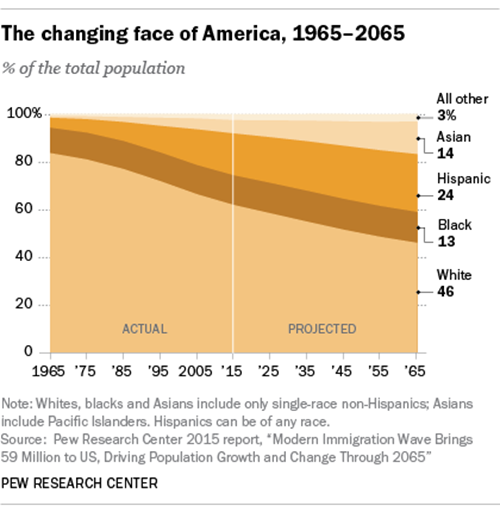
**Keywords**

Machine learning; Unsupervised learning; Terrorism; Migration; United States; Data Mining; Immigration Policy; Foreign Policy, K-Nearest-Neighbor, Multiple Linear Regression, Random Trees Ensemble.

# INTRODUCTION

Migration in the United States is often perceived to correlate with increasing rates of terrorism and crime as well as economically displacing the native-born population.

The 1965 change in immigration law set the stage for a massive change in the number, racial make-up and socio-economic status of the US population. This is outlined in Figure One. From 1965 the US (Non-Hispanic) White population changed from 85% of the population to 62% in 2015. Before 1965 the US had a quota system that had an explicit preference for immigrants from northwestern Europe. The aim was to keep the demographic make-up of the US unchanged. Immigration law then strongly restricted immigration from Asia, Africa and parts of the Caribbean. The shift away from ethnic selection in U.S. immigration policy was primarily a response to foreign policy pressures emanating from the growing number of independent Asian, African, and Latin American countries that sought to delegitimize racism. President Johnson signed this law thinking it would not have a massive impact on the native-born population. [2] Figure One below indicates a different narrative.



Most recently, this topic is being debated in the 2016 U.S. presidential race. The vast differences in foreign policy positions among presidential candidates are important to note, as United States immigration policy heavily relies on Executive power. For example, the Republican candidate, Donald Trump, has made removal of illegal immigrants and prevention of future illegal immigration as well as the deportation of criminal aliens key positions in his campaign. [3] He also supports a temporary ban on Muslim immigration because of the threat of Islamic Terrorism. [4] The Democratic candidate, Hilary Clinton, essentially believes in open borders, and that all illegal aliens should become citizens, while only violent criminal aliens should be deported. She also adds that any family member who shows up at a U.S. border can join family members already in the U.S. [5], and that there should be no ban on Muslim immigration. [6]. Most Western countries have had high and rising levels of immigration even though a majority of their populations consistently want immigration to stabilize or decrease. That public sentiment has been apparent since the early 1960s. [7] This paper is a data driven contribution to this ongoing debate.

Today, most research on migration in the United States suggests that outcomes are positive and the legal immigrant population actually lowers the crime in the area where they live and work. This correlates with our original hypothesis that the U.S. born populations are more likely to commit acts of terror than their immigrant counterparts. Primarily, we hope to deploy a machine learning algorithm that will help us uncover patterns in the terrorism of native and immigrant populations from 2000 to 2015.

Over the next 45 years’ current projections indicate the U.S. population will grow by 98.1 million. As shown in Figure 2, the population is expected to increase by 2.1 million people per year, the Foreign born will grow by 78 million and become almost 19% of the population. If the children of these immigrants are counted they will comprise around 40% of the US population; this makes the debate on immigration even more important.

**Figure 2: US Population By Nativity 2014 to 2016**



In all cases we are interested if there is a difference between the occurrence of terror incidents and results between the foreign born and the U.S. native-born populations. Usually, researchers are unable to measure these systemic relations between a feature vector and its target variable. They base their conclusion on sample, which are validated using hypothesis tests. In this case the team constructed a data set of all terror incidents in the US from 2000-through-2015. Since, this research is based on a population rather than a sample some hypothesis tests are less applicable. Our approach will utilize Data Mining and Machine Learning, which is the construction of a model that represents the systematic information that a feature vector provides about its target variable. Data mining is primarily concerned with looking for unsuspected systematic information. In practice researches do want to test the specific hypotheses that data mining algorithms generate

# METHODOLOGY

## Data:

*2.1.1 Original Data Set*

For this study, the team began with utilizing the START database, which is currently the most comprehensive unclassified data base on terrorist events in the world. This is the National Consortium for the Study of Terrorism and Responses to Terrorism (START). Since the research focused on United States, only those records were extracted for study. The Global Terrorism Database (GTD) deploys a Data Collection Methodology for collecting new records and updating existing records. The records go back to the early 1970s. Due to time limits and the fact that online information on individuals involved in terror incidents prior to 2000 is limited the team focused on US terror incidents from 2000 to the present. Thes initial dataset had around 304 records of incidents in the United States from 2000-2014.

*2.1.2 Enriching the dataset with individual records*

While the START dataset provided a good understanding of incidents in the United States, it did not provide us the information on individuals who committed these terror acts. A given act could be performed by one individual or many individuals. One person could perform multiple terrorist acts. Basically, the relationship between a Terror Incident and an Individual was many-to-many. In addition, it was missing data after 2014. Upon inspection, the team realized the START dataset did not include FBI data on planned and attempted terrorist acts. With this in mind, the team performed a data wrangling stage where we identified new incidents for 2015 and 2016. We also cross-checked the START dataset with datasets constructed by a terrorist researcher in North Carolina, terror incident lists from Council on American–Islamic Relations (CAIR), lists from Jewish organizations, Wikipedia, and news reports. A new incident record was created only if one did not already exist in CAIR. Each was analyzed against existing dataset, and a new record was only created if one did not exist. Collection of new data followed the GTD definition of a terrorist attack. Currently, GTD defines a terrorist attack as the threatened or actual use of illegal force and violence by a non‐state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation.

.The team enriched the START dataset with individuals involved in each incident. Each incident was enriched with a terrorist’s name, this sometimes created multiple rows for each incident. We also expanded data collection to include name, age, sex race, religion, ethnicity, education marital status, number of children, migration status, country of origin and data source(s) (e.g., URL links). We independently identified the motive for the act as described by the perpetrators. START also tracks terror group names and motivations at the incident level. The data dictionary used is outlined in Appendix A.

The research team constructed a database of terror incidents that occurred on US soil from 2000-to the present.. In order to compare Foreign Born to Native-Born terrorists the team collected basic demographic information about the individuals involved in these attacks. The 18 variables collected are outlined in the following Feature vector. This data would let the team determine the involvement of Native-Born and Foreign born individuals in these terrorist attacks. Since, the attacks number less than 200 from 2000-to-2015, the dataset is descriptive and not a sample. The motive behind the attack was almost always based on the statements of the attacker(s). The team found that most attacks were: (1) Islamic-Supremacy, (2) Eco-terrorism, (3) White Supremacy, (4) Anti-Government, and (5) Anti-Abortion, (6) Revenge (e.g. stage attack to get back at an ex-spouse), and (7) Other. These motivations are useful for visualization, but because of leakage were not a good predictor (i.e., attack had to occur before the motive was known). This dataset would allow the research team to test if foreign-born individuals were more involved in terror acts than native born and what if anything was the role of religion.

**Table 2: Response Variables (Attack Outcomes-3 Attributes)**

The goal of the overall dataset was to look at the terrorism rate per immigrant and country of origin. Specifically, the research is focused on the immigrant vs. foreign-born rates of:

* Planned Terrorism Acts on US soil
* Attempted Terrorism Acts on US soil
* Success of overall incident (excludes planned or attempted)
  + Terrorism Deaths on US soil
  + .Terrorism wounded persons on US soil
  + Property Damage

The enriched dataset contained 581 total records for individuals and incidents in the United States.

*2.1.3 Data schema*

The chosen schema for the new constructed dataset is as described in the table below.

**Table 1: Feature Vecter (26 attributes & Primary Keys**

|  |  |
| --- | --- |
| **Incident** | **Terrorist** |
| eventId  eventId2 | age  DoB |
| Summary  Data Source | Terrorist (name) |
| iyear  imonth  iday  Approxdate | sex  Num\_of Children  conversion  Year of Conversion  religion |
| city  provstate  State\_num | race  ethnicity |
|  | education |
| State\_Population | Immigrant Status Year of US Entry Country\_of\_Origin |
| Motive | motivation |

# The response variables are: (#1) Planned Attack, (#2) Attempted Attack, and #3) Successful Attack, each is mutually exclusive. A Planned Attack, cannot be in the same category as an Attempted Attack (e.g., failed car bombs in Times Square). A successful attack is where there is some combination of property damage, victims wounded, and victims killed

|  |
| --- |
| **Response Variables** |
| 1. Planned Attack - categorical/dichotomous (Y/N) |
| 1. Attempted Attack –categorical/dichotomous (Y/N) |
| 1. Success – categorical/dichotomous (Y/N)  * property damage * number wounded * number killed |

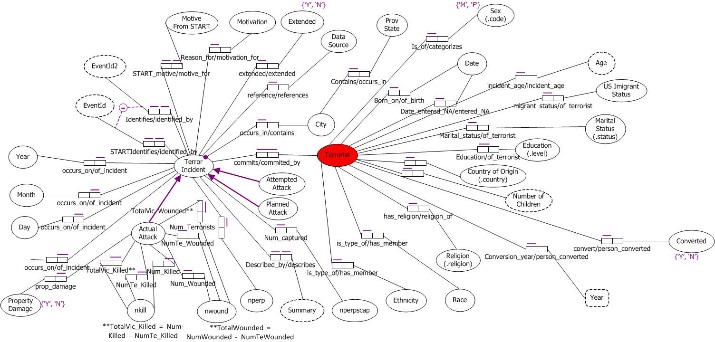
.

The response variables are mainly categorical (dichotomous or binary Y/N values for Planned, Attempted, and Successful). Two Success sub-elements are Ratio (Killed and Wounded). The induction of models these cases involve Multi Linear Regression, Tree-Induction, and K Nearest Neighbor Clustering. The models are designed to answer questions on Foreign-born vs. Native-born terrorism

*2.1.4. Data Model*

A major innovation in this analysis was to take the START database and link to each Terror Incident the individual(s) who committed these acts and what were their major demographic characteristics. It was surprising difficult to collect this data. As far as the team can tell no other research group in the US took this step. The Object Role (data) Model (ORM) of our feature set and target data elements is outlined below. Each incident can have one or more associated terrorists who commit the act. The act and individual properties are on different conceptual levels. A lack of a common database resulted in considerable reworking during the data collection process*.*

**Figure 3: ORM Features & Response Attr. Data Model**

**

*2.1.5. Data Cleaning*

Data cleaning was a necessary component to prepare the data for visualization. The data was cleaned for missing variables, and creating dummy variables for the categorical values. Missing variables were either treated with a statistical variable incretion (mode, median, or mean), or they were treated with deletion. Categorical data was treated with dummy variables by assigning each a numerical category. These rules are outlined in Table 3 below.

**Table 3: Major Data Cleaning Rules**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Treatment** | **Rows Affected** |
| Sex | Mode | 142 |
| Age | Mode - 21 | 123 |
| Marital Status | Mode - Single | 273 |
| Race\_Census | Mode-White | 165 |
| Motive\_2 | Delete Record | 31 |
| Immigrant Status | Delete Record | 123 |
| Planned Attack | Mode | NONE |
| Attempted Attack | Mode -0 | 1 |
| Property | Mode -0 | 1 |
| Num Killed | Mode | NONE |
| Num\_Wounded | Mode | NONE |
| success | Mode | NONE |

Categorical variables were also given transformed by creating a Category Score. The variables were assigned a number starting with 1. Ex. Male =1, Female =2.

*2.2.1 Tools*

Data Mining: The team utilized Microsoft Office Excel 2016 for initial data where housing, with the plan to move the dataset at a later time to a SQL DBMS. Data Mining and Machine Learning was conducted utilizing the Student Version of Frontline Solver – XLMiner add in for Excel.

Data Visualization**:** The team utilized Tableau 9.0.3– Student version, M.S. Excel 2016 with add-ins for Network Visualizations, and XL Miner Charts.

**2.3. Data Mining**

The team is using a classification mode to help determine the class probability estimation of terror incidents. For example, we sought to answer the questions of what is the probability that an immigrant will be involved in a terror attack (planned, attempted, or successful) vs a Native born individual. The team has the data to model systemic information between the Feature Vector and Categorical Response Variables for Foreign vs. Native-born groups. The first phase of the mode is to identify what drives terrorism. The second phase would be to gather more vector points, and test the model against general populations. This paper only covers the first phase.

*2.3.1 Feature Selection*

In order to help us identify the makeup of our data, we conducted a discovery process – feature selection. This process is conducted by evaluating all feature vectors towards one response variable. The resulting outcome is a gain ration, mutual information, and feature importance chi-square test. For our example we conducted this process 6 times for each of our response variables. The results were ordered by top 5 features, which were later used to help build our models.

*2.3.2 Model Selection*

Our data contained both categorical (transformed and numeric), as well as one continuous variable. Due to the nature of our data, we experimented with K-Nearest Neighbor, Multiple Linear Regression, and Regression Tree. For each model 60% of the dataset was randomly selected for training and 40% was utilized for testing the model. Where available, the data was also normalized to help with performance considering the 2001 terrorist activity in United States.

*2.3.2.1 K-Nearest Neighbor*

In the k-nearest-neighbor prediction method, the training data set is used to predict the value of a variable of interest for each member of a "target" data set. [8] The structure of the data generally consists of a variable of interest ("number killed," for example), and a number of additional predictor variables (age, religion, country of origin, etc.).

For this model we selected Sex, Age, Marital Status, Country of Origin, Race, Religion, and Immigration Status in hopes to predict the target variables for number killed, planned attack, attempted attack, wounded, property damage, and overall success. For each simulation we chose to normalized the data, and selected *k* =20. The model output provided results for the best *k*.

*2.3.2.2. Multiple Linear Regression*

Linear regression is performed on a dataset either to predict the response variable based on the predictor variable, or to study the relationship between the response variable and predictor variables. For example, using linear regression, the crime rate of a state can be explained as a function of demographic factors such as population, education, male to female ratio etc.

This procedure performs linear regression on a selected dataset that fits a linear model of the form

**Y= b0 + b1X1 + b2X2+ .... + bkXk+ e**

where Y is the dependent variable (response), X1, X2,.. .,Xk are the independent variables (predictors) and e is the random error. b0 , b1, b2, .... bk are known as the regression coefficients, which are estimated from the data. The multiple linear regression algorithm in XLMiner chooses regression coefficients to minimize the difference between the predicted and actual values. [9]

For this model we selected Sex, Age, Marital Status, Country of Origin, Race, Religion, and Immigration Status in hopes to predict the output variables for number killed, planned attack, attempted attack, wounded, property damage, and overall success.

The weight variable that we tested was sex, age, country of origin, and immigration status.

*2.3.2.3 Regression Tree Prediction Method*

As with all regression techniques, XLMiner assumes the existence of a single output (response) variable and one or more input (predictor) variables. The output variable is numerical. The general regression tree building methodology allows input variables to be a mixture of continuous and categorical variables. A decision tree is generated where each decision node in the tree contains a test on some input variable's value. The terminal nodes of the tree contain the predicted output variable values.

A Regression tree may be considered as a **variant of decision trees**, designed to **approximate real-valued functions** instead of being used for classification method [9] .

50 weak learners were utilized for this particular training. We selected Sex, Age, Marital Status, Country of Origin, Race, Religion, and Immigration Status in hopes to predict the target variables for number killed, planned attack, attempted attack, wounded, property damage, and overall success.

# RESULTS

## Model Performance

Three different models ran for each response variable against the predictor variable. Total 18 models were reviewed for prediction output of the model against the test dataset. Motive and state of incident were left off from model variables, as this might cause a potential ecological fallacy or

### K-Nearest Neighbor

The data suggests that our strongest predictors are **age** and **country of origin** for planned, actual, and success (where success is greater than 0 only if the attack resulted in death, property damage, or wounded individual).

#### Planned Attack

The model was almost 4 times better at prediction than a random assignment. Our best k=18, and the lift charts indicate a good performance. The data suggests that our strongest predictors are **age** and **country of origin**.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 5.99047619 | 0.148132158 | 2.03338E-18 |
| Validation | 23.04368659 | 0.355828136 | 0.005104203 |

#### Attempted Attack

The model for training was 7 times better at predicting an attempted attack than a random assignment, with the best k =20. The training data preformed at 3 times better than a random assignment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 8.522142857 | 0.176682 | 0.006227 |
| Validation | 21.36569177 | 0.342628 | 0.027124 |

#### Success of Attack

The model for training was 3 times better at predicting an attempted attack than a random assignment, with the best k =20. The training data preformed at 2.5 times better than a random assignment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 22.63929 | 0.287972 | -0.0348 |
| Validation | 37.49659 | 0.4539 | -0.07926 |

*3.1.2. Multiple Linear Regression*

White utilizing the backwards elimination for variable selection, we acquired Age and Country of Origin when the weight was Sex. Sex and Country of origin were the best predictors when the weight was Age. The model best preformed with more than 2 variables, with some expanding up to 6 coefficients.

*3.1.2.1 Planned Attack*

The training data preformed at 2.5 times better than a random assignment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 1647.655986 | 2.456699 | 1.075504912 |
| Validation | 26.88451253 | 0.38434 | -0.03578053 |

*Results shown for Model with Age as weight.*

*3.1.2.2 Attempted Attack*

This model preformed best with 5 coefficients: (Age, CountryOrigin, Marital Status, Race, Immigrant Status\_ord). The predictive outcome of the model was about 2 times better than a random variable.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 37.40303708 | 0.3701452 | -0.002618 |
| Validation | 20.76806364 | 0.3378021 | -0.02013041 |

*3.1.2.3 Successful Attack*

This model preformed best with 5 coefficients: (Age, CountryOrigin, Marital Status, Race, Immigrant Status\_ord). The predictive outcome of the model was about 2 times better than a random variable.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 58.30553196 | 0.46214 | 0.012072621 |
| Validation | 36.89351634 | 0.4502351 | -0.00724999 |

*3.1.3. Regression Tree Prediction Method*

The overall data tended to split best on Age, Race, and Country of Origin.

*3.1.3.1 Planned Attack*

This model usually preforms 3 time better than a random assignment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 8.29057408 | 0.17426532 | -3.73124E-17 |
| Validation | 24.3414344 | 0.36571047 | 0.004074401 |

*3.1.3.2 Attempted Attack*

This model usually preforms 4 time better than a random assignment.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 10.2930873 | 0.19417419 | 1.12852E-17 |
| Validation | 26.8943239 | 0.38440995 | -0.022296445 |

*3.1.3.3 Successful Attack*

This model usually preforms 3 time better than a random assignment.

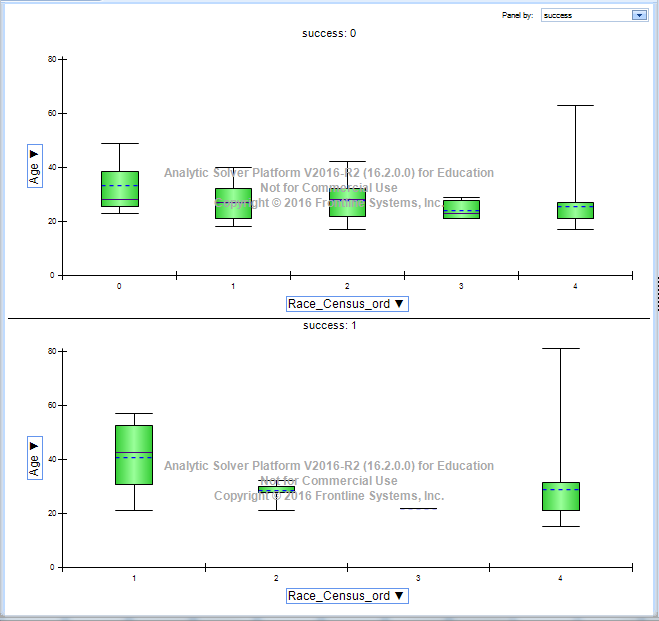
|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total sum of squared errors** | **RMS Error** | **Average Error** |
| Training | 23.2984931 | 0.29213433 | -6.42547E-17 |
| Validation | 39.6721912 | 0.46688229 | -0.013134701 |

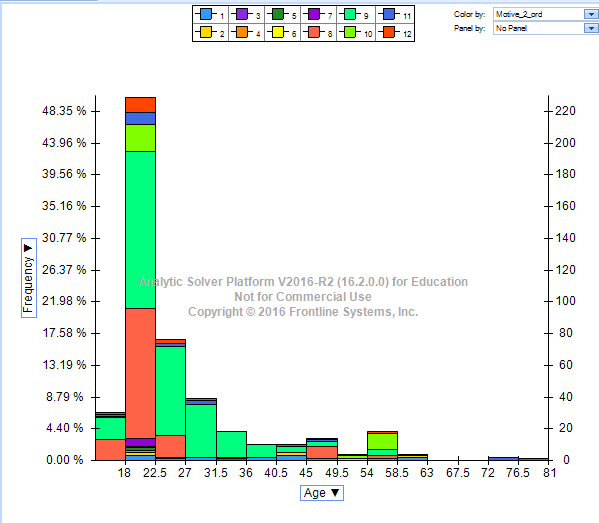
# CONCLUSION

Data Mining and Machine Learning provided us with the information that most terrorism acts of success, planned, or attempted, can be split off if we pay attention to age, religion, country of origin, race, and sex of the individual. This is an interesting finding that majority of the terror deaths are by young males that are Islamic and categorized as white by the U.S. Census 2010.

In Figure 4, we can see that the age is still relatively consistent by race. When we look at the age of those who are actually successful, the race plays a bigger role. Here category 0 and 4 are more present. Category 0 designates individuals who identify with 2+ races, and 4 designates individuals who are classified as white under the 2010 Census classification.

Figure 4: Relationships of Age and Race on the success (1) or Not (0) of a terrorist attack, visualized by boxplots.

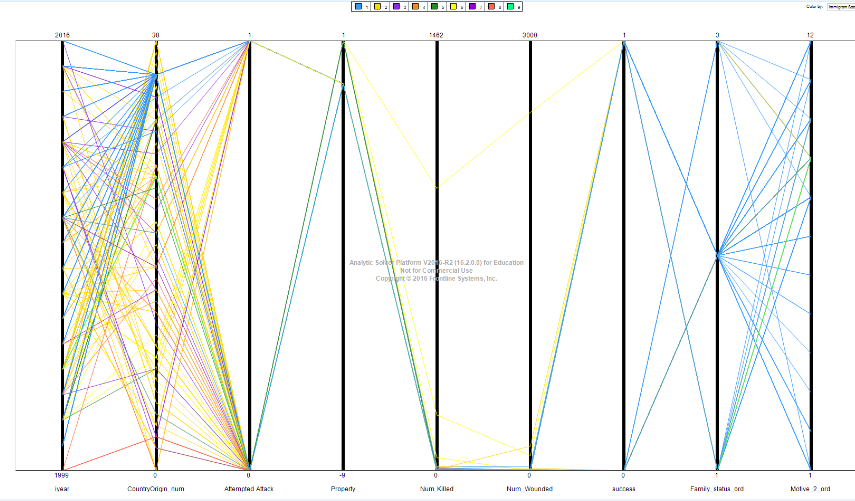
.



When looking at age in comparison with motivation, Table 5, we see that age is consistently skewed to the left with the most occurring variable at 21 years of age .

We further look at country of origin, and success of various attacks. **Figure 6,** describes a story between the categories for country of origin, year, attacks types, motive, and more. From there we can view that the split on country of origin is significant and apparent. This is also supported in our models’ findings.

Table 6: Multiple variable network diagram.

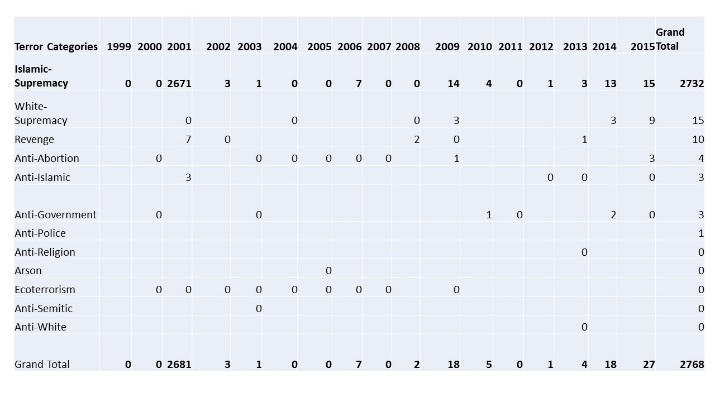


Through analysis of models and exploratory data, we observe that most terrorist attacks in the United States do not kill or wound people. In addition, it should be noted that terrorist attacks attract significant attention, but (excluding 9/11) they result in at most tens of deaths/year. In comparison, United States has around 15,000 murders each year. Terror deaths are thus around .0006% of murder deaths in a given year.

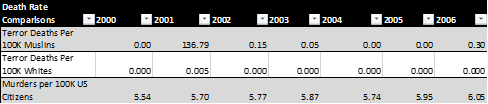
Our second major finding is that Foreign-Born Immigrants who are not Islamic do not commit terror acts in the United States. AT query of our descriptive statistics shows that from 2000 - 2015 non-Islamic immigrants do not commit acts of terrorism. [10]. However, Table 4 below shows Islamic terror kills many more Americans than native-born non-Islamic groups. For example, even excluding 9/11 Islamic terror killed 61 people since 2000; this compares to 32 terrorist deaths from all other motivations.

**Table 5: Death Rate Comparisons 2000 – 2015**

**Table 4: Response Variables (Attack Outcomes-3 Attributes)**



The team’s third major finding is that compared to some naïve predictions, the models do split on identification of Islamic religion. Those immigrants who identify with Islam (and increasingly native-born Islamists) do carry a weight in the terrorism cases. On average they are 100X more likely to commit an act of terrorism than a native non-Hispanic white individual. This 100X greater propensity to terror deaths provides some rational for Trump’s supports of a temporary ban on Muslim immigration. The following Table 5 provides the relevant data. This Terror death data is normalized relative to the 3.2 million Muslims estimated to live in the US [11] and the 198.1 million non-Hispanic whites.



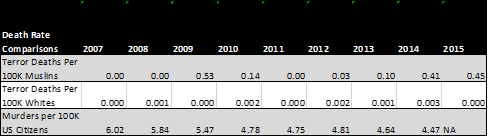


Table 5 is the start of an expected values analysis. It outlines the normalized probability of terror deaths. The team’s major finding is that Foreign-Born Immigrants who are not Islamic pose no Terror problem to the US. A query of our descriptive statistics shows that from 2000 - 2015 non-Islamic immigrants do not commit acts of Terror. Over Entire 2000-2015 Period each 100K Muslin Immigrants Results in 83 more US deaths. This is mainly because 9/11. Muslim immigration to the US averages around 100Kpersons per year. Muslim immigrant cause terror deaths per capita 100X more than non-Hispanic whites. Immigrant Hispanics, excepting Muslim Hispanics, do not commit acts of terror.

In our future research the team hopes to review actual crime statistics and thus provide a more complete story of immigration impacts, especially for illegal aliens. Moreover, we hope to gather even more extensive data on immigrant impacts on US economic growth and welfare expenditures, as well as immigrant victimization. Ideally the team can start to identify the predictors for immigrants (or native-born) who commit terror acts, crimes or fail economically and those who do not. The team wants to also look for lags. For example, the Islamic terrorists captured or killed in 2015 generally arrived in the US during the 1990s. Ideally the team will expand the terror dataset to include 1990 to 2000.

# ACKNOWLEDGMENTS

Thank you to the START, Frontline Analytics for providing the necessary tools to construct the dataset, and to GMU.

# REFERENCES

1. LaFree G., Director, START National Consortium for the Study of Terrorism and Responses to Terrorism. START: A Center of Excellence of the U.S. Department of Homeland Security, University of Maryland, College Park, MD 20742, USA 301.405.6600 https://www.start.umd.edu/gtd/
2. Walsh K., 2015 (October 2) 50 Years Ago, Immigration Changed in America: America's current racial makeup is rooted in a decades-old legislative overhaul, US News & World Report, see: http://www.usnews.com/news/articles/2015/10/02/50-years-ago-immigration-changed-in-america
3. IMMIGRATION REFORM THAT WILL MAKE AMERICA GREAT AGAIN, Donald Trump for President https://www.donaldjtrump.com/positions/immigration-reform
4. Bergman R, Donald Trump's Call to Ban Muslim Immigrants, Atlantic (Dec. 7, 2015) http://www.theatlantic.com/politics/archive/2015/12/donald-trumps-call-to-ban-muslim-immigrants/419298/
5. America needs comprehensive immigration reform with a pathway to citizenship, Hilary for America https://www.hillaryclinton.com/issues/immigration-reform/
6. Tani M, Hillary Clinton responded to a Muslim veteran's question on Islamophobia with a scathing criticism of Donald Trump <http://www.businessinsider.com/hillary-clinton-donald-trump-muslim-veteran-cnn-town-hall-2016-1>
7. Hochschild JL, Mollenkopf J. The Complexities of Immigration: Why Western Countries Struggle with Immigration Politics and Policies. In: Delivering Citizenship. edited by Bertelsmann Stiftung, European Policy Centre, Migration Policy Institute. Berlin, Germany: Verlag Bertelsmann Stiftung ; 2009. Retrieved from http://scholar.harvard.edu/jlhochschild/publications/complexities-immigration-why-western-countries-struggle-immigration-politi
8. Series. ACM, New York, NY, 19-33. DOI= <http://doi.acm.org/10.1145/90417.90738>.
9. Robert Shapiro and Jiwon Vellucci . 2010. The Impact of Immigration and Immigration Reform on the Wages of American Workers. (May 2010). Retrieved April 30, 2016 from <http://ndn.org/sites/default/files/paper/impact_of_immigration_on_us_wages_-_shapiro-vellucci_-_final-_may_24_2010_0.pdf>
10. National Consortium for the Study of Terrorism and Responses to Terrorism (START). (2015). Global Terrorism Database [Data file]. Retrieved from <http://www.start.umd.edu/gtd>
11. Besheer Mohamed, Jan. 6, 2016 A new estimate of the U.S. Muslim population. BESHEER MOHAMED Retrieved from http://www.pewresearch.org/fact-tank/2016/01/06/a-new-estimate-of-the-u-s-muslim-population

# APPENDIX: Categorical Variables

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Sex | 1 | Male |  | Country of Origin | 0 | Afganistan |
|  | 2 | Female |  |  | 2 | Albania |
|  |  |  |  |  | 3 | Algeria |
| Marital Status |  |  |  |  | 4 | Bosnia |
|  | 2 | Single |  |  | 5 | Caribbean |
|  | 1 | Married |  |  | 5 | Carribean |
|  | 3 | Divorced |  |  | 6 | Chechnya |
| Converted\_flag |  |  |  |  | 7 | Cuba |
|  | 3 | Don't know |  |  | 8 | Dominican Republic |
|  | 1 | No |  |  | 9 | Egypt |
|  | 2 | Yes |  |  | 10 | England |
|  |  |  |  |  | 11 | Ethiopia |
| Race\_Census | 0 | 2+ races |  |  | 12 | France |
|  | 1 | Asian |  |  | 13 | Guinea |
|  | 2 | Black |  |  | 14 | Haiti |
|  | 3 | Hispanic or Latino |  |  | 15 | India |
|  | 4 | White |  |  | 16 | Iran |
|  |  |  |  |  | 16 | Iraq |
| Motive |  |  |  |  | 17 | Jamaica |
|  | 1 | Anti-Government |  |  | 18 | Jordan |
|  | 2 | Anti-Islamic |  |  | 19 | Kazahstan |
|  | 3 | Anti-Police |  |  | 20 | Korea |
|  | 4 | Anti-Religion |  |  | 21 | Kosovo |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 5 | Anti-Semitic |  |  |  | 22 | Lebannon |
|  | 6 | Anti-White |  |  |  | 23 | Mexico |
|  | 7 | Arson |  |  |  | 24 | Morocco |
|  | 8 | Ecoterrorism |  |  |  | 25 | Nigeria |
|  | 9 | Islamic-Supremacy |  |  |  | 26 | Pakistan |
|  | 10 | Revenge |  |  |  | 27 | Palestine |
|  | 11 | Islamic-Supremacy |  |  |  | 28 | Philippines |
|  | 12 | Anti-Abortion |  |  |  | 29 | Puerto Rico |
|  |  |  |  |  |  | 30 | Russia |
| Religion | 0 | Christian |  |  |  | 31 | Saudi Arabia |
|  | 1 | Catholic |  |  |  | 32 | Somalia |
|  | 2 | Islam |  |  |  | 33 | Uzbekistan |
|  | 3 | Jewish |  |  |  | 34 | United Arab Emirates |
|  | 4 | Mormon |  |  |  |  |  |
|  | 10 | Other |  |  |  | 35 | United States |
| Immigrant Status |  |  |  |  |  | 36 | Uyghur |
|  | 1 | Born US Citizen |  |  |  | 37 | Venezuela |
|  | 2 | Naturalized Citizen |  |  |  | 38 | Yemen |
|  | 3 | Legal Permanent Alien |  |  |  |  |  |
|  | 4 | Student Visa |  |  |  |  |  |
|  | 5 | Temporary Visa |  |  |  |  |  |
|  | 6 | Tourist Visa |  |  |  |  |  |
|  | 7 | Refugee |  |  |  |  |  |
|  | 8 | Illegal Immigrant |  |  |  |  |  |
|  | 9 | Asylum |  |  |  |  |  |

8.APPENDIX 2 –GRAPHS AND CHARTS

Terrorism dataset variable distribution:

