**Analysis on Immigrants**

**from Travel Ban Countries**

*By Josh, Pony, So, and Thu Ha*

**Introduction & Background**

According to the Council on Foreign Relations, “Immigrants comprise about 14 percent of the U.S. population: [more than forty-three million](https://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states) out of a total of about 323 million people”. These immigrants come in search of refuge and better opportunities that are often lacking in their respective countries. However, factors such as hard work, strong values and education that have been the motifs for achieving success have failed to offer immigrants such as *Valentino Achak Deng* in *What is the What: An Autobiography of* Valentino Achak Deng, success in America. Reflecting on his life, Valentino “ wonders whether leaving a war-torn continent and moving to one where they are not accepting of refugees from a war-torn continent really made things any different for him; to his mind, he has only exchanged one form of persecution for another one. ”

In addition, with the new administration, negative attitudes towards immigrants especially from muslim majority countries have arisen and aided in the implementation of new immigration policies that ban people from these countries to ensure “the safety and security of the American people.” Our study analyzes whether immigration from the 8 travel ban countries (Egypt, Iran, Libya, Sudan, Yemen, Iraq, Somalia, and Syria) impacts the safety of the American people, as well as analyzing the immigrant’s life in the United States.

The main purpose of the study is to analyze the negative stereotypes behind the immigrants from 8 travel ban countries in the United States and the condition of living in the US. First, we will make clusters of U.S states based on important features from the exploratory data analysis (Christan percent per state, Red State Voting per state, and etc. ), assign a cluster number to individuals based on where they resided, and compare immigrants by clusters to analyze characteristics of individuals from different clusters. Then, we will address negative stereotypes associated with crime and violence against immigrants, and examine whether the empirical data support these stereotypes. Finally, we analyze the level of success that immigrants attain, based on different demographic factors. We will compare immigrants and people in the US by different demographic factors. In summary, the paper is composed of three parts, which highlight different aspects of social justice issues related to immigrants from 8 travel ban countries:

1. Cluster Analysis of the characteristics of individuals based their assigned cluster number. We can investigate discriminant variables among different clusters of states, which might indicate preferences of accepting certain groups of immigrants.
2. Longitudinal Model to investigate the relationship between violent crime and immigrant arrival from the travel ban countries. This will provide us empirical evidence to debunk negative stereotypes and questionable claims against immigrants.
3. Item Response Theory (IRT) to create a success score that shows the interaction between the individuals and the items. This will help us to assess whether immigrants get to achieve the same level of success as non immigrants.

**Method/Analysis**

**(1) Features of individuals from different states by cluster analysis**

*Overview of Cluster Analysis*

In the first part of our analysis, we use unsupervised learning tools to make clusters of states in the US based on important features from our EDA and analyze individuals by clusters. This allows us to discover measurement features, without predicting a response variable. In order to investigate features of each state and immigrants living within each state, two different datasets were used: State level data that has four variables and individual level data that has 17 variables[[1]](#footnote-0). First, K-nearest neighbor clustering was used to make clusters of states. After classifying each state, every individual was assigned to a cluster number based on where they resided in the US. Principal Component Analysis (PCA) was used to measure weights of variables of individual data and see correlations of variables. Characteristics of individuals from different clusters were analyzed by comparing variables from individual data with accounting weights of variables from the PCA.

*K-nearest neighbor classification*

From the state level data, we used K-nearest neighbor classification in order to make cluster groups. Based on the sum of square of the model, three was the optimal number of cluster to make. 25 different random starting assignments were used.

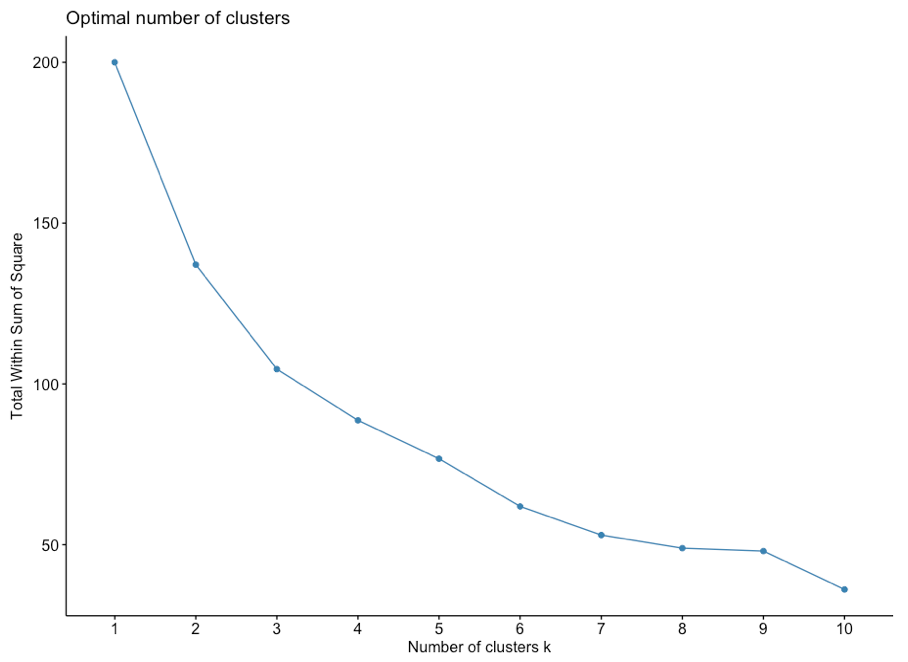


Figure 1: Choosing the optimal number of cluster based on Sum of Square

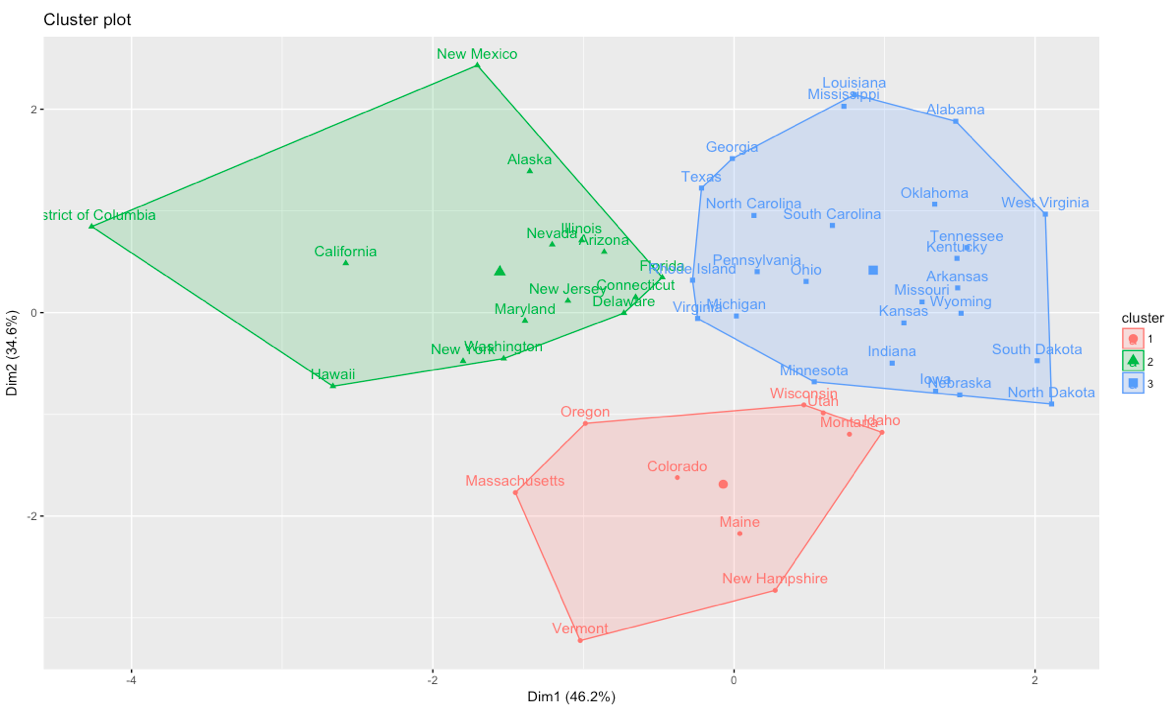


Figure 2: Three clusters by k-nearest neighbor classification

We can see that several of states are grouped by geographical location. And some of common features by group were examined and identified as below:

***Group 1: Predominantly white, liberal, rich states (mostly New England states)***

***Group 2: Diverse population, blue states, and less rich states***

***Group 3: (Mostly) poor and predominantly white states (mostly Southern states)***

*Principal Components Analysis(PCA)*

Principal Component Analysis (PCA) was used to measure weights of variables of individual data and see correlations of variables. Since there are 17 variables in individual data, it is crucial to use PCA for reducing dimensions of data. PCA analyzes effects of each of 17 variables of individual data. These variables were scaled before conducting PCA.

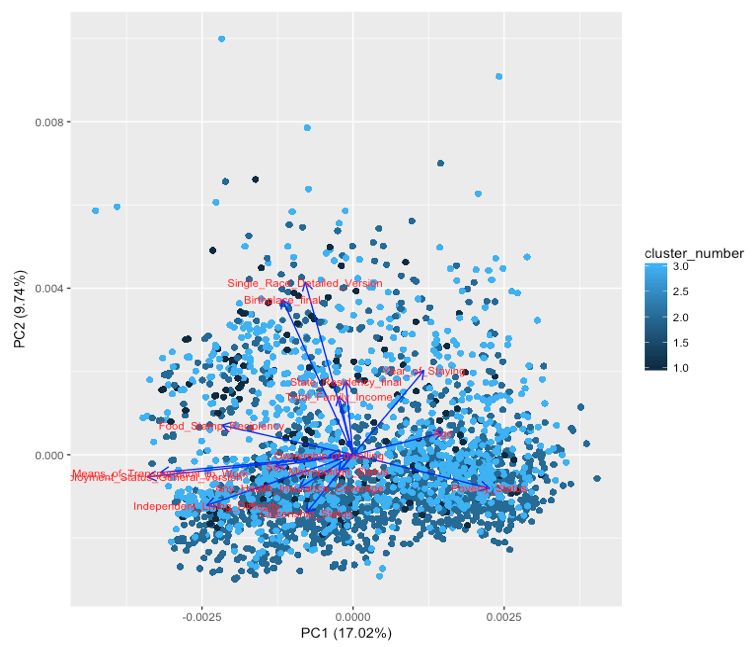


Figure 3: Principal Components Analysis (PCA) plot

Race, Birthplace, Independent living difficulty, Poverty, Food Stamp Recipients, Means of Transportation, Employment Status have larger weights than Total family income, Age, Sex, State Residency, Citizenship Status, and Health Insurance Status. Also, Year of Staying, Age, and Poverty Status are less correlated to other variables and are likely to have negative correlation with other variables. The correlation can be verified by the correlogram, and the trend follows as PCA plot implies about variables.

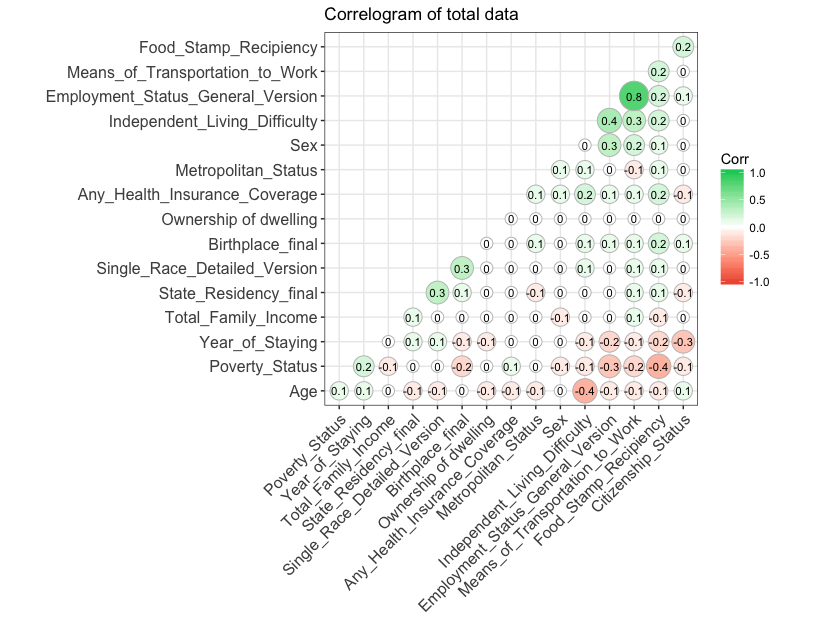


Figure 4: Correlation graph between 17 variables

Like Age , Poverty, and Year of Staying have negative correlations with many of the other variables, which is consistent with the PCA plot. The highest correlation is between Mean of Transportation and Employment status (0.8), which means that an individual is more likely to have a car if you are employed. However, since most of the variables are categorical, we might not capture accurate relationships between variables.

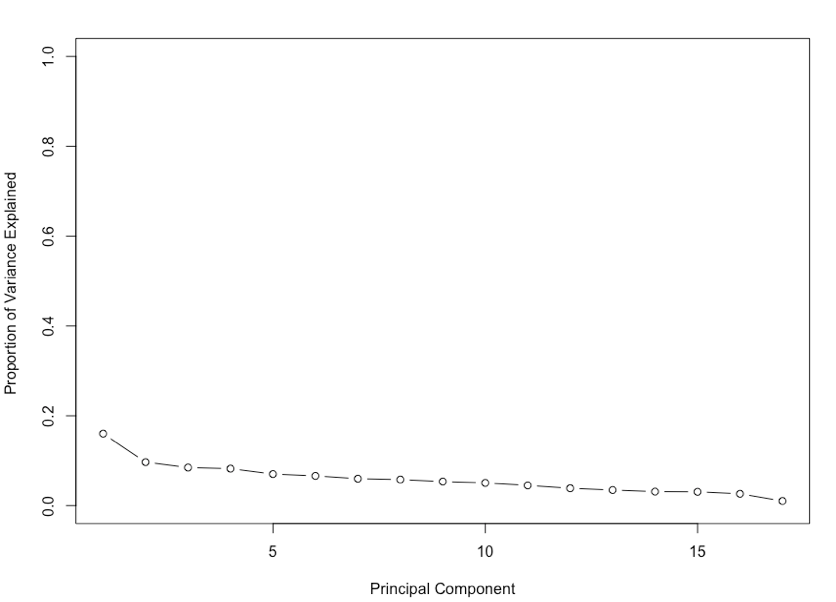


Figure 5: Proportion of variance explained by each variable

Each variable has a small proportion of variance explained (less than 20%), which implies that there is no heavily weighted variable within these variables.

*Cluster Comparison*

17 variables from the individual dataset were analyzed and compared by three cluster groups that were divided by the K-nearest neighbor classification. More weighted variables from PCA , such as Birthplace, Employment Status, Independent living difficulty, and means of transportation, and more informational variables, such as Total Family Income and Year of Staying, were closely examined.

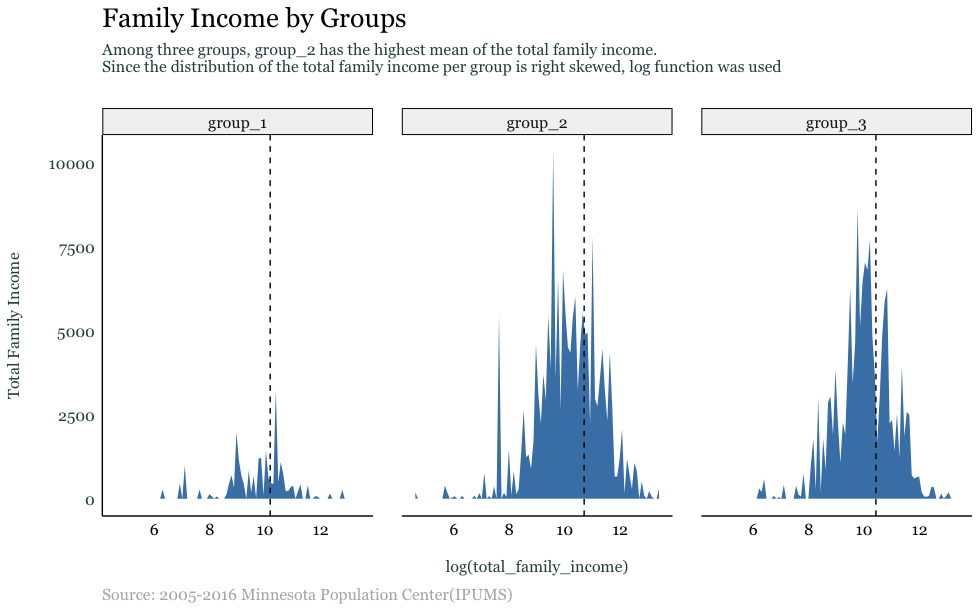


Figure 6:Total family income by cluster

Because the distribution of total family income by groups is right skewed, we used the log function was used to visualize data. Group 2 has the highest total family income among three groups (with log function: group 1=10.17, group 2=10.70, group 3=10.41, and all groups=10.43 | without log function: group 1=$26,226, group 2=$44,175, group 3=$33,411, and all groups= $34,604 )

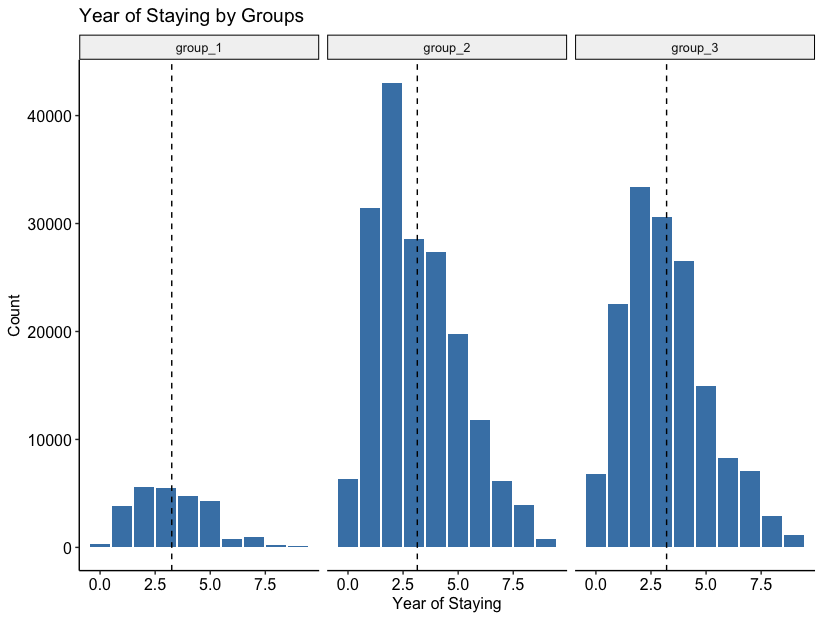


Figure 7:Year of staying in the US by cluster

Group 2 has the highest mean of year of staying, but the mean of year of staying among three groups are very similar (group 1=3.25 years, group 2=3.15 years, group 3=3.21 years). The mean of year of staying for all groups is 3.20 years.

*Ratio comparison*

Since there is a large difference of the number of sample for each group, categorical variables below were compared by ratio, which was calculated by the number of people belongs to a certain variable/total number of people in a group. A lighter color indicates a higher ratio.

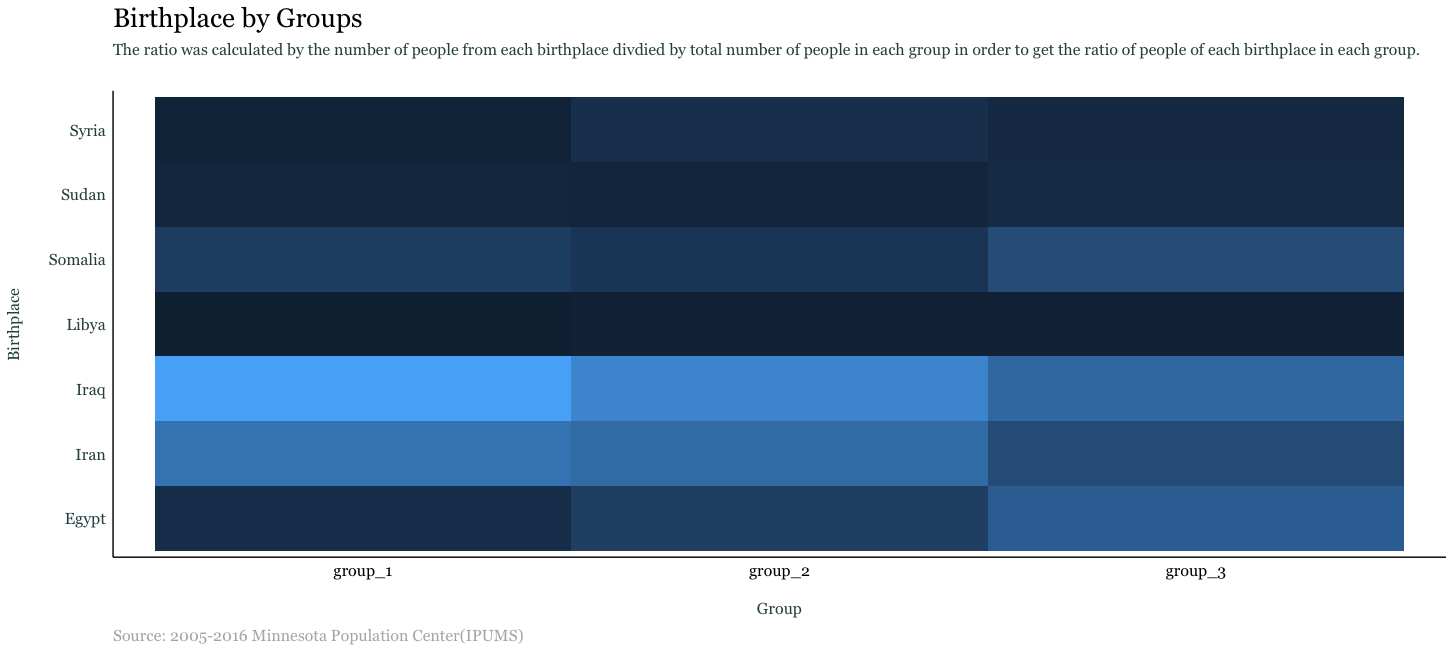


Figure 8: Birthplace of individuals by cluster

The plot indicates that the large proportion of people from group 1 is from Iraq (0.443) and Iran (0.316). Also, the large proportion of people from group 2 is from Iraq (0.370) and Iran (0.292). On the other hand, the large proportion of people from group 3 is from Iraq (0.279) and Egypt (0.245). For all of groups, the largest proportion of people is from Iraq. And group 3 has more even ratio than group 1 and group 2. Group 1 has the largest proportion of people from Iraq.

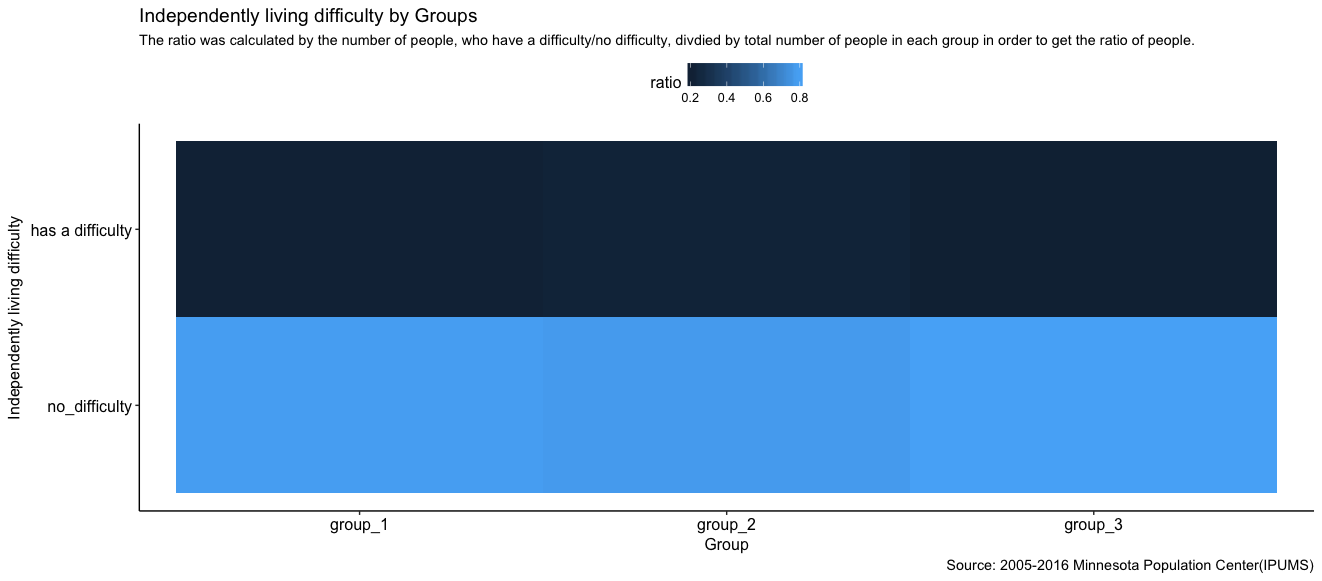


Figure 9: Independently living difficulty by cluster

The plot indicates that most people from all groups have no difficulty living independently (group 1: 0.796, group 2: 0.786, group 3: 0.806)

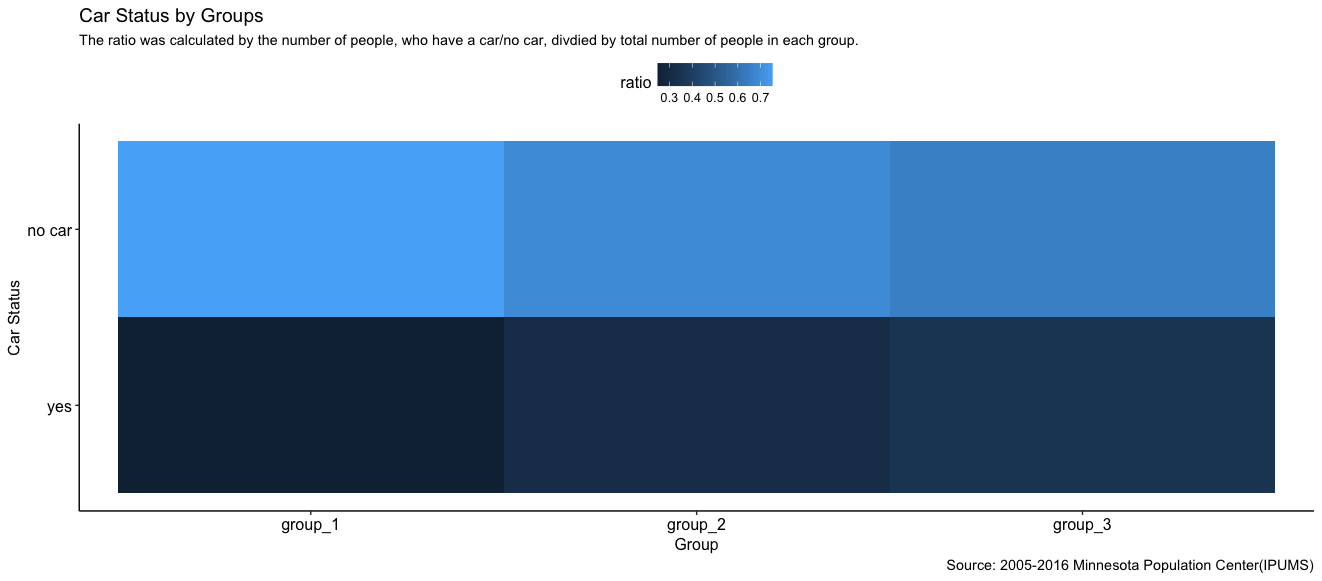


Figure 10: Possession of car by cluster

The plot indicates that a large proportion of people from group 1 have no car (0.753). However, group 2 and group 3 have similar ratios of people who have a car, compared to people who do not have a car (group 2 no car: 0.686 and group 3 no car: 0.650)

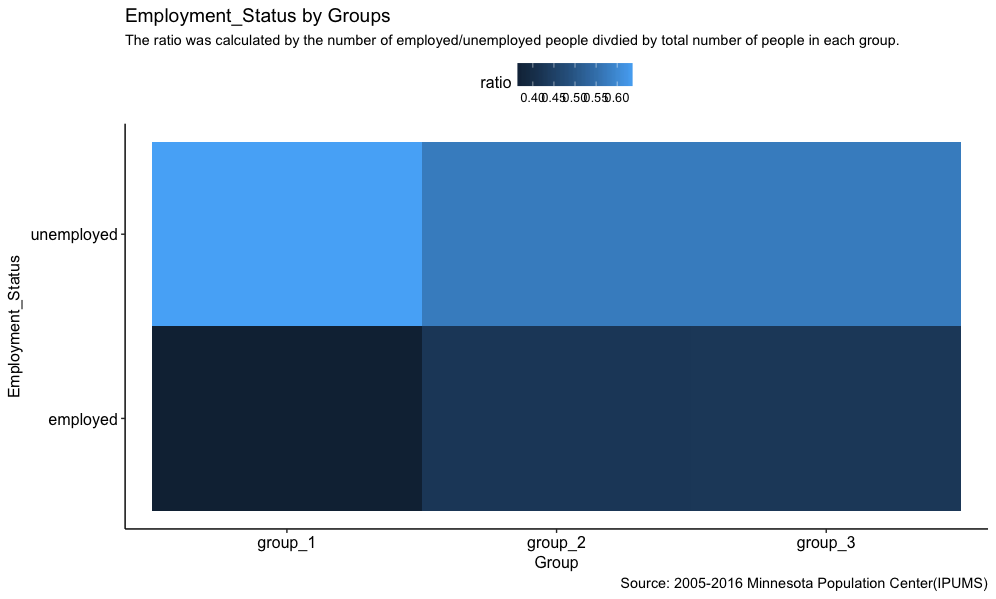


Figure 11:Employment status by cluster

The plot indicates that the large proportion of people from group 1 is unemployed (0.638). However, group 2 and group 3 have similar unemployment ratio (group 2: 0.574, group 3: 0.573).

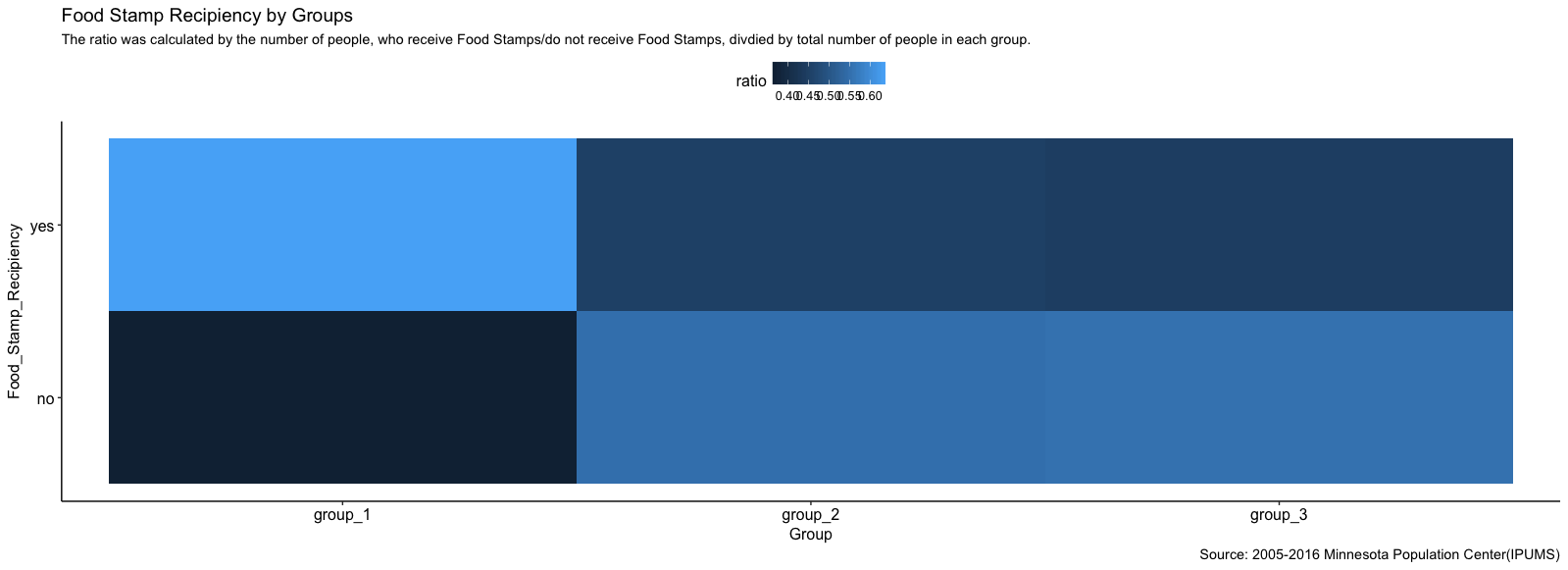


Figure 11: Food Stamps recipients by cluster

The plot indicates that a large proportion of people from group 1 receive Food Stamps (0.631). However, group 2 and group 3 have less number of people who receive Food Stamps than people who receive Food Stamps (group 2 Yes: 0.452 and group 3 Yes: 0.447)

*Inference*

We can conclude that individuals from group 1 are less privileged than individuals from other two groups. Compared to group 2 and 3, group 1 has higher ratio of people who are unemployed, do not have a car, receive Food Stamps, and have the lowest mean of family income of all three groups. For the demography of groups, the highest proportion of people in group 1 is from Iraq (0.443), Iran (0.316), and Somalia (0.129). Group 2 and Group 3 tend to have similar trend except group 3 has more diverse birthplaces.

Results can be interpreted that immigrants living predominantly white, liberal, and rich states tend to be underprivileged. Because immigrants are under-privileged in more privileged states, the disparity of living condition can cause the polarization of immigrant group and non-immigrant group. This polarization issue might be related to negative stereotypes that immigrants are poor, which can lead to a difficult transition for immigrants to adapt their life in the US.

**(2) Relationship between violent crime and immigrant by multilevel modeling**

*Data organization*

In order to investigate the relationship between violent crime rate and immigrant arrival from 2005 to 2014, state-level data was collected from various government websites. For part 2, we used data for 10 travel ban countries (Egypt, Iran, Mali, Libya, Sudan, Yemen, Iraq, Somalia, South Sudan and Syria) instead of 8 countries. District of Columbia was removed from all the datasets because of the unusually high crime rate.

With the longitudinal structure of the dataset, we have measurements of violent crime rate, immigrant arrival, population, etc. at different time points from 2005 to 2014 for each of the 50 states. Thus, we will address our research question at two levels (see Appendix, Data table for longitudinal model):

* Level 1: within-state covariates that change over time (main variables: immigrant arrival, violent crime rate, firearms recovered, firearm regulations)
* Level 2: state-specific variables that do not change over time (main variables: percent population living in urban area)

*Exploratory Data Analysis*

In Figure 13, we see that violent crime rates for 50 states decrease over time. However, there are differences among states in starting points in 2005 and slopes - changes in violent crime rate over the 10 year period, which suggests that multilevel model with state as level 2 is appropriate. Figure 14 shows that immigrant arrivals from the 10 travel ban countries stay the same or slightly increase over time, with California accepting the highest number of immigrants.

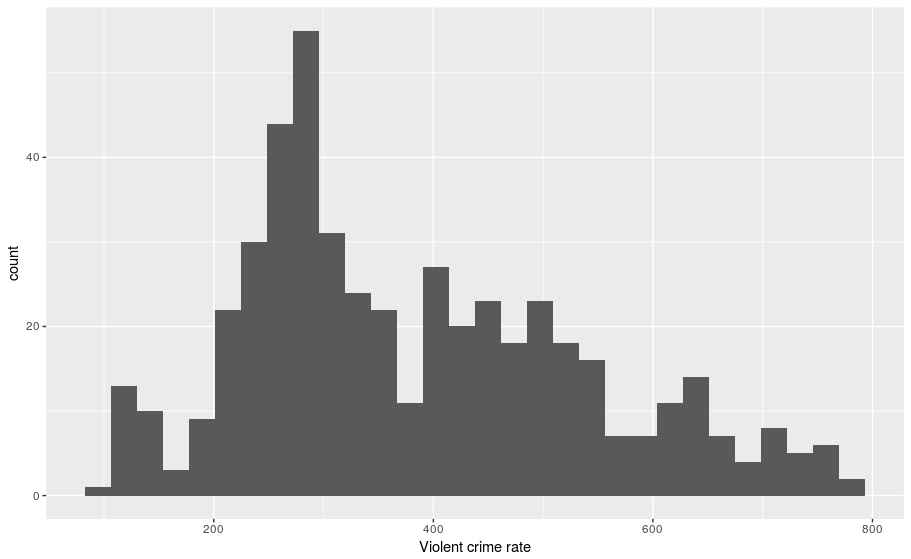
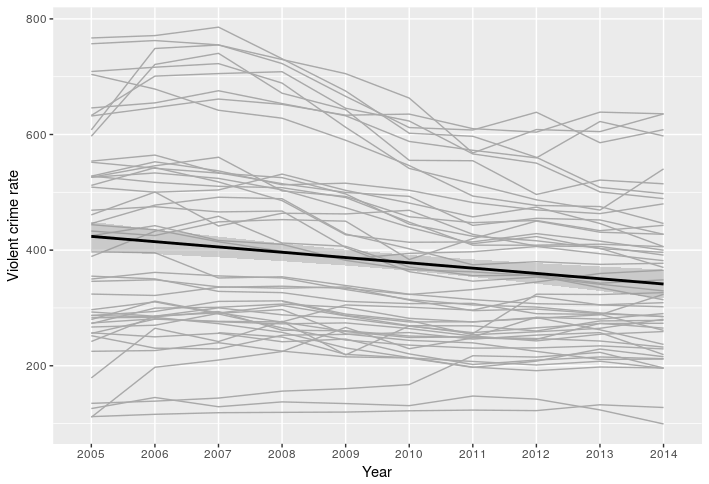
****

Figure 12: Histogram of violent crime rate Figure 13: Violent crime rate of 50 states from 2005 to 2014

While violent crime rate seems to be positively correlated with immigrant arrival, the positive trend disappears when individual states are examined (Figure 15). Most of the states have negative or no relationship between immigrant arrival and violent crime rate with a few exceptions (Delaware, Nevada, North Dakota and Tennessee) (Figure 16).

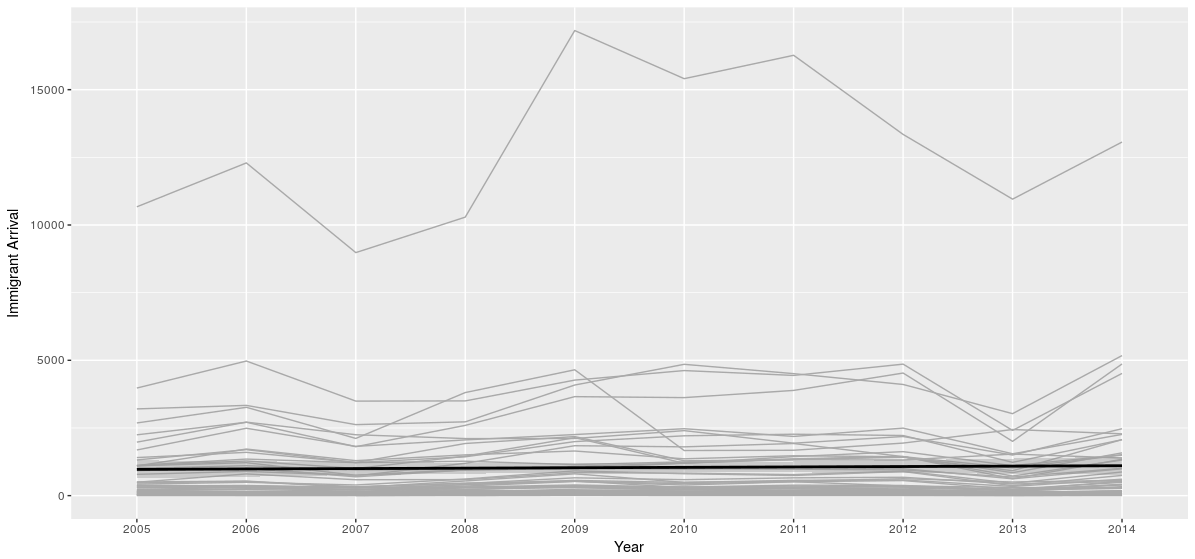
****

Figure 14: Immigrant arrival in 50 states from 2005 to 2014

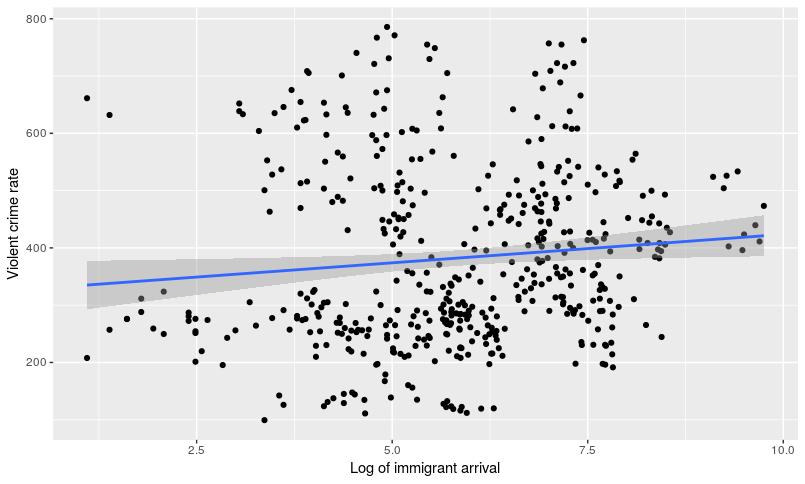


Figure 15: Violent crime rate by immigrant arrival (all 50 states and 10 years)



Figure 16: Violent crime rate by immigrant arrival for individual states

Among other level 1 covariates:

* Total number of firearm regulations has a slightly positive correlation with violent crime rate (Figure 17). When we look at individual states, however, most of them do not have total firearm regulations changed across years and thus it is hard to detect any linear trends.
* The ranges of violent crime rate across three legislation compositions are roughly equal, with Republican having slightly higher median rate (Figure 18).
* There seems to be a positive correlation between the number of firearms recovered and violent crime rate (Figure 19). Even when we investigate individual states, we can still see positive trends in most of the states.

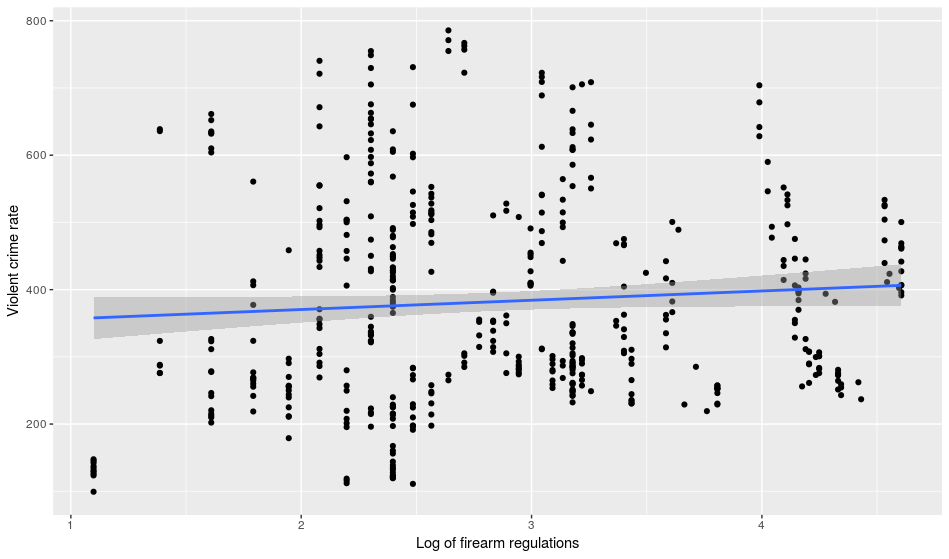


Figure 17: Violent crime rate by number of firearm regulations

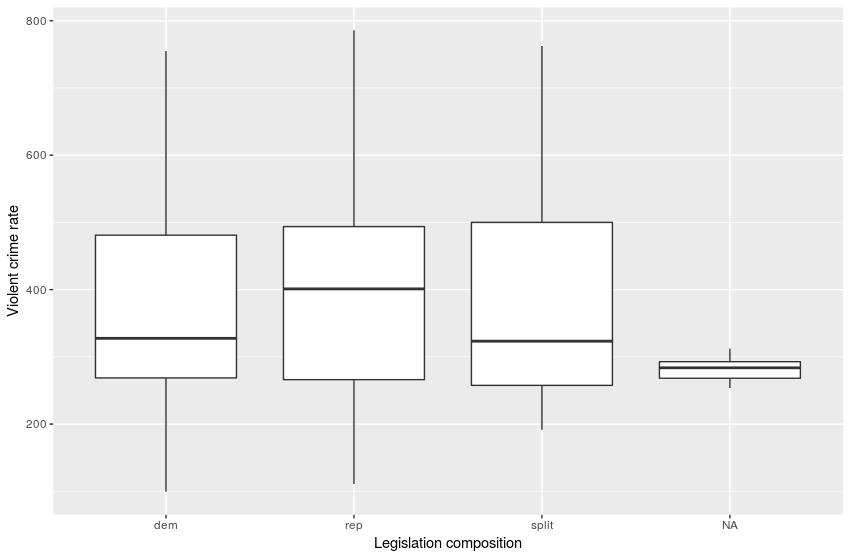


Figure 18: Violent crime rate by legislation composition

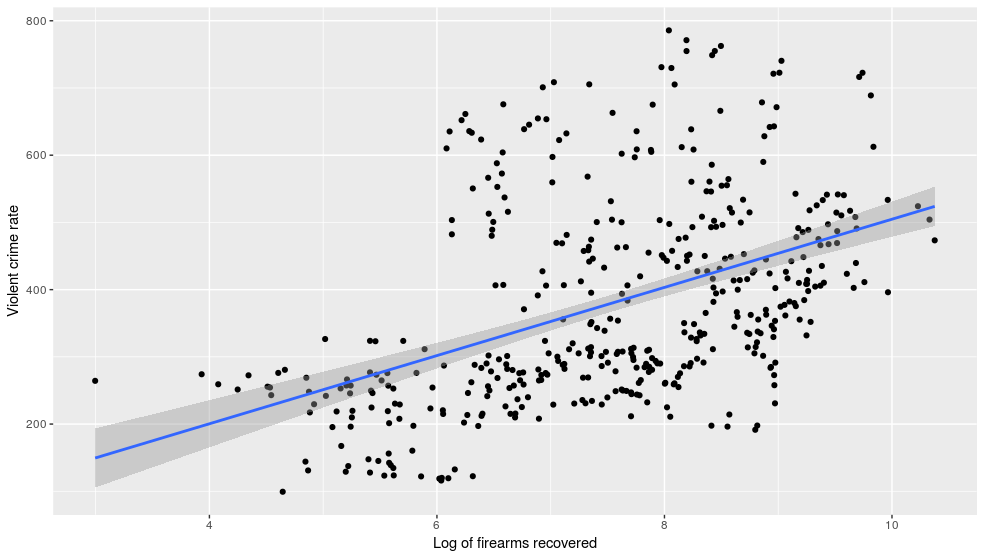


Figure 19: Violent crime rate by firearms recovered

Next, we will examine the relationships between violent crime rate and level 2 predictors. While violent crime rate increases with higher percent population living in urban area, it does not vary across percentage of urban area. Violent crime rate differs across three state clusters, with group 2 having the highest crime rate. However, since this cluster variable was constructed based on measurements of recent years (percent vote for Trump in 2016), we think it is not appropriate to include this in our model as a level 2 covariate.

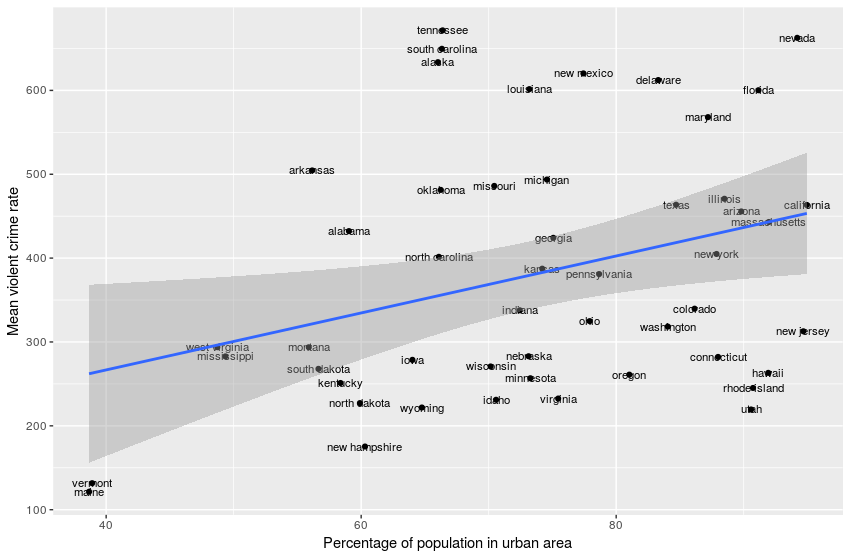
****

Figure 20: Mean violent crime rate by percent population living in urban area

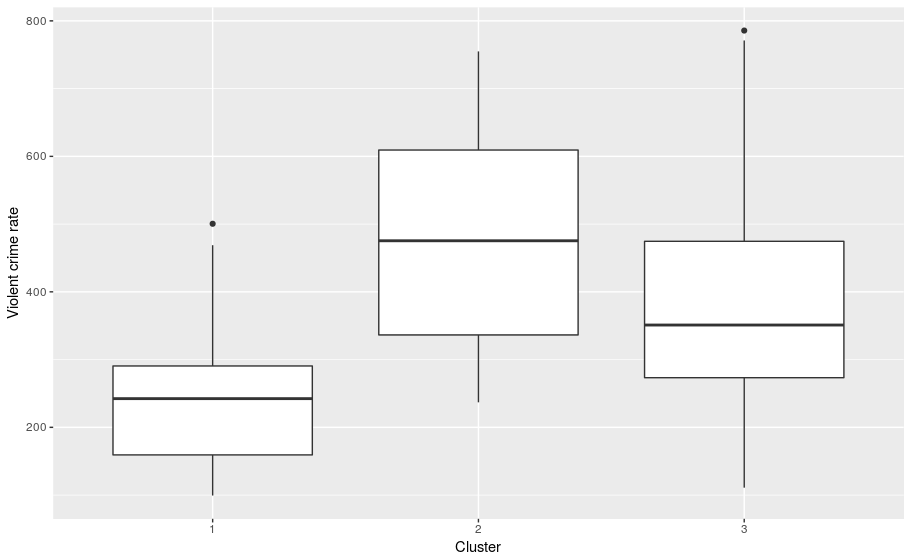
****

Figure 21: Mean violent crime rate by state cluster

Based on the results of the exploratory data analysis, the best predictors for our longitudinal model to explain violent crime rate are immigrant arrival, firearms recovered and percent population living in urban area. But first, let’s explore some simpler models that look at trends over time.

*Model 1: Unconditional growth model*

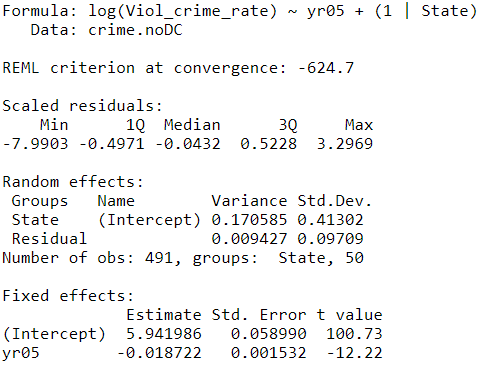
* Level One:

Yij = ai + bi\*Year05ij + ϵij

* Level Two:

ai = α0 +ui

bi = β0 + vi



Looking at the unconditional growth model with year from 2005 as the sole predictor, we can see that violent crime rate decreases over time (t=-12.22), and most of the total variation in violent crime rate is attributable to difference among states (AIC=-616.7, BIC=-599.9).

*Model 2: Immigrant arrival as the sole level 1 predictor*

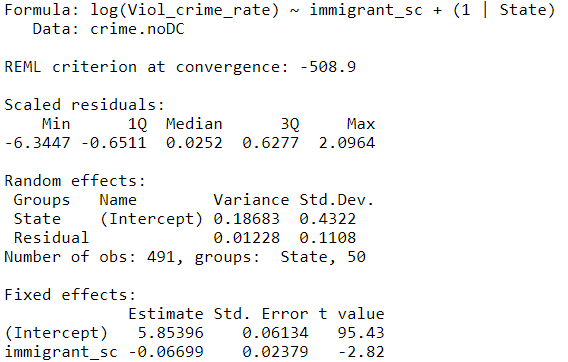
* Level One:

Yij = ai + bi\*Immigrant\_arrivalij + ϵij

* Level Two:

ai = α0 + ui

bi = β0 + vi



This multilevel model with immigrant arrival as the sole predictor shows that as immigrant arrival increases, violent crime decreases (t=-2.82), and a large portion of the total variability in violent crime rate is attributable to difference among states. However, AIC (-500.9) and BIC (-484.1) of model 2 are higher than those of the unconditional growth model.

Likelihood ratio test, AIC and BIC are used to compare models with different subset of predictors and choose the one that best explains violent crime rate. This final model includes immigrant arrival, year from 2005 and firearm recovered as level 1 predictors, and percent population living in urban area as a level 2 predictor (AIC=-701.2, BIC=-664.4). Even when we control for other covariates, immigrant arrival still has a negative relationship with violent crime rate, but now with an insignificant coefficient t=-1.32.

*Final model:*

* Level One:

Yij = ai + bi\*Immigrant\_arrivalij + ci\*Year05ij + di\*Firearmij+ ϵij

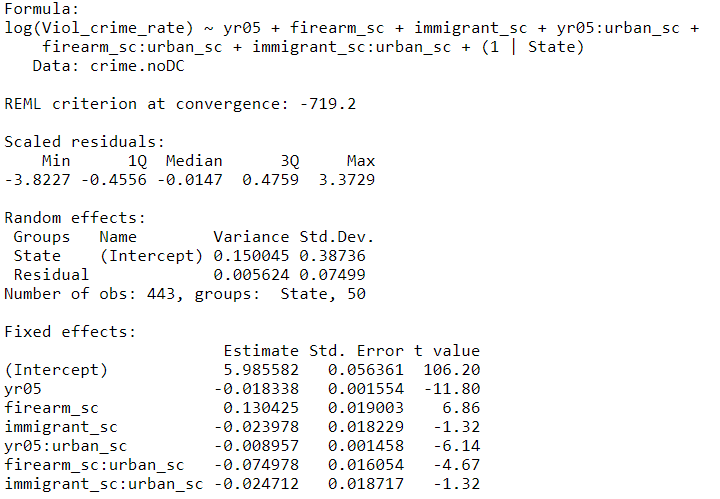
* Level Two:

ai = α0 + α1\*UrbanPop + ui

bi = β0 + β1\*UrbanPop

ci = γ0 + γ1\*UrbanPop

di = δ0 + δ1\*UrbanPop



Even though state cluster is not included in our final model, it is interesting to explore how state cluster is related to violent crime rate and immigrant arrival. We fit a linear trend to the data from each of the 50 states using year from 2005 as the time variable. Figure 22a shows that group 1 (predominantly white, liberal, rich states) starts out with the lowest violent crime rate in 2005, while group 2 (diverse population, blue and less rich states) starts out with the highest rate. However, violent crime rate in group 2 also decreases fastest over the course of 10 years, while group 1’s rate decreases slowest (Figure 22b). Figure 23 shows that group 2 has the widest range of immigrant acceptance over that 10 year period compared to group 1 and 3. Thus, it indicates that while states within cluster that has higher violent crime rates do not always accept more immigrants.

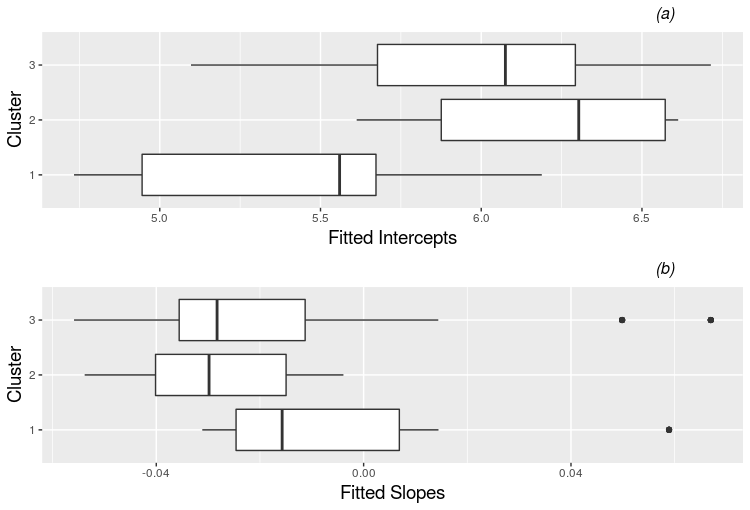


Figure 22: (a) Fitted intercepts and (b) slopes of fitted regression lines by state cluster

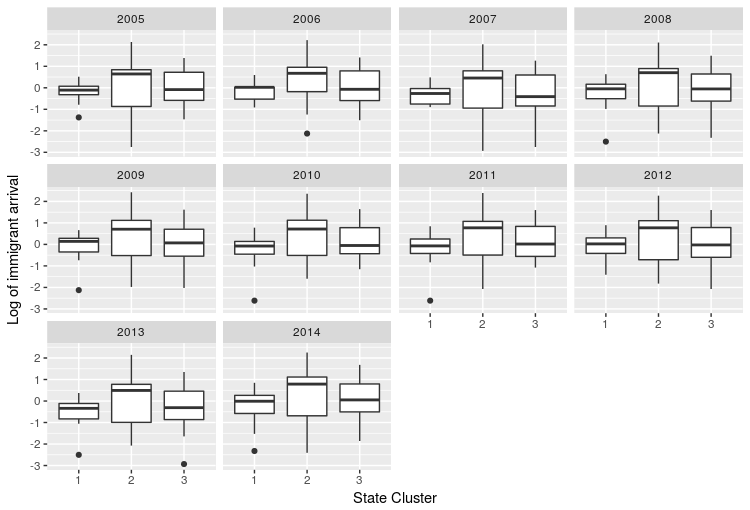


Figure 23: Log of immigrant arrival by state cluster from 2005 to 2014

*Terrorism incidents*

In addition to violent crime rate, we also investigate terrorism incidents in the US because terrorism is one of the stereotypes associated with immigrants from majority-Muslim countries. As we can see from Figure 22 and 23 below, there are not many terrorism incidents from 2005 - 2016. The number of incidents caused by Muslim associated terrorist groups or Iraqi extremists does not seem to correlate with immigrant arrival.

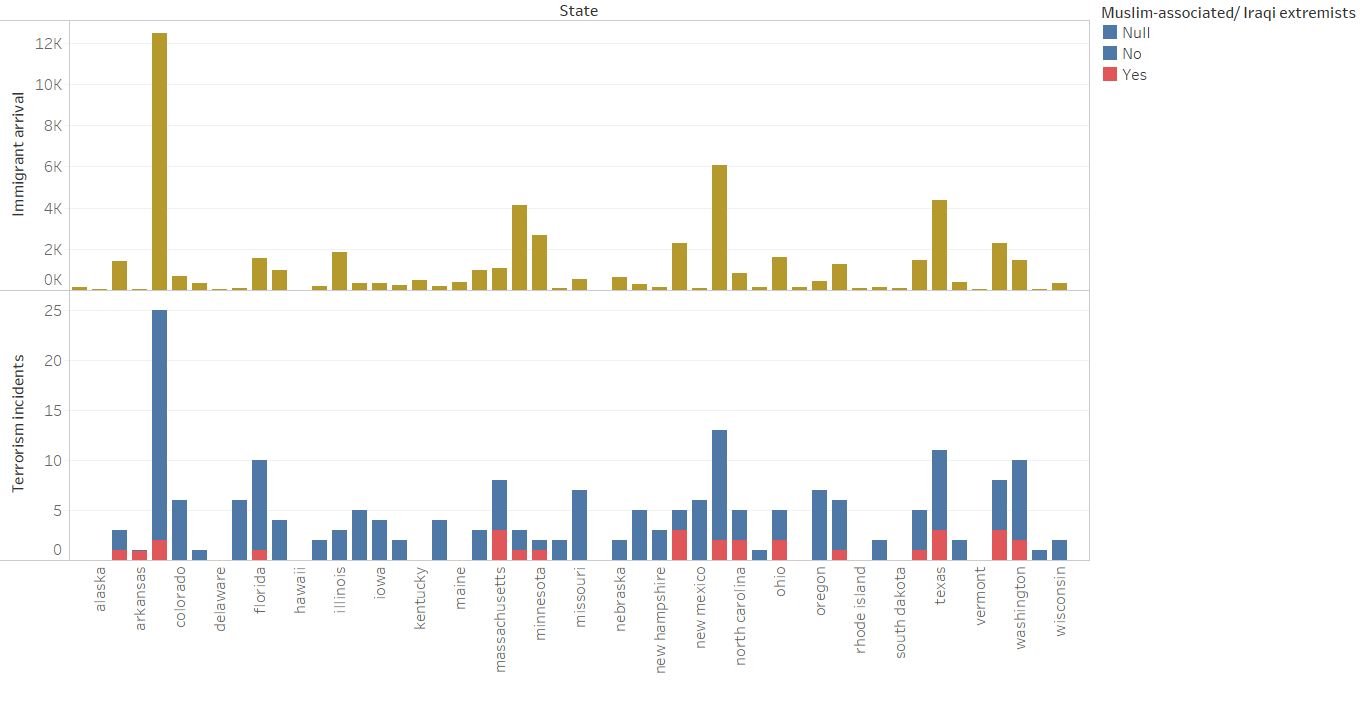


Figure 24: Count of terrorism incidents and immigrant arrival by states

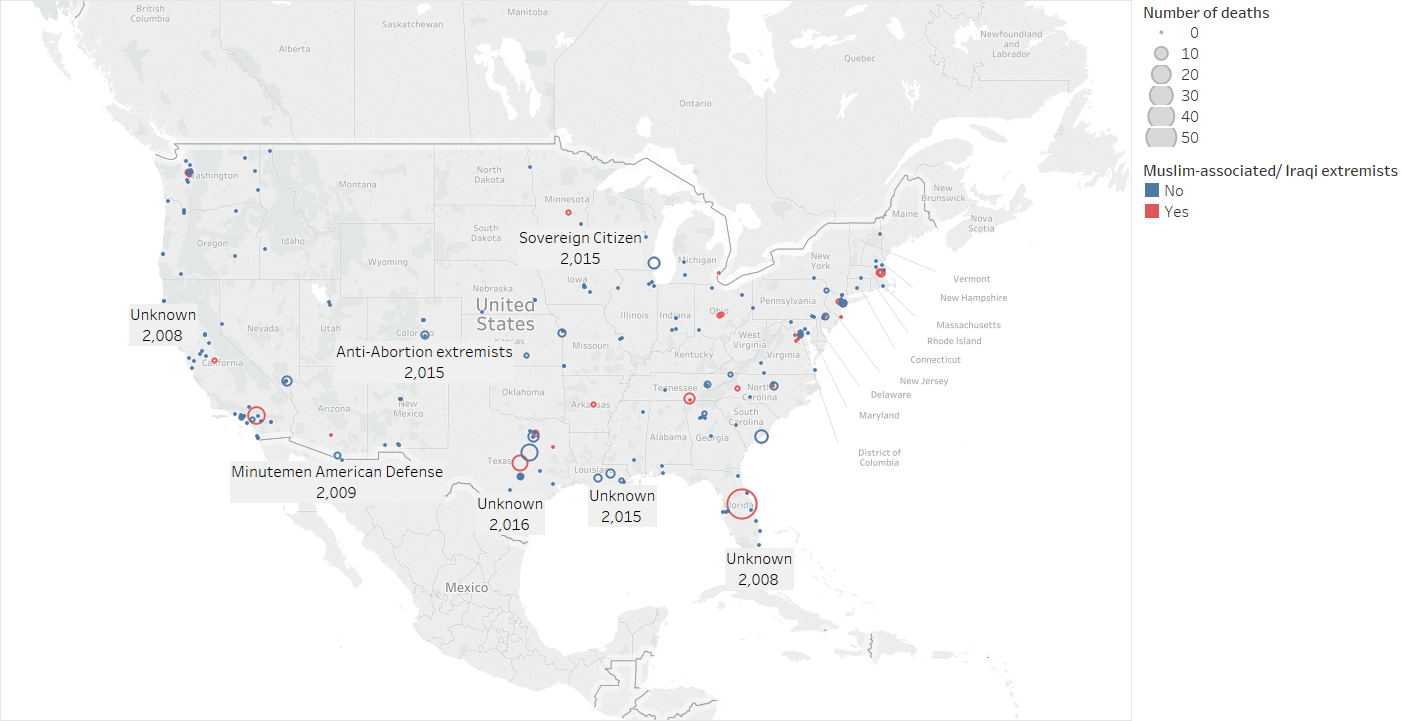


Figure 25: Terrorism incidents on US map

*Inference*

Our analysis suggests that immigrant arrival from 10 travel ban countries is not associated with violent crime rate (insignificant negative coefficient) and there are no correlation between immigrant arrival and terrorism incidents.

**(3) Measuring success by item response theory**

In order to analyze success of individuals, we used itemized responses to rank the success of individuals. Into our itemized response, we used the binary variables: Employment, Poverty, Foodstamp, Ownhome, and Health. We then created the itemized response plot and used each individual’s success score (z1) as the response variable for our linear model. We used the best subset selection in Rstudio, to maximize the R-squared value with success score as our predictor.

*Item Response*

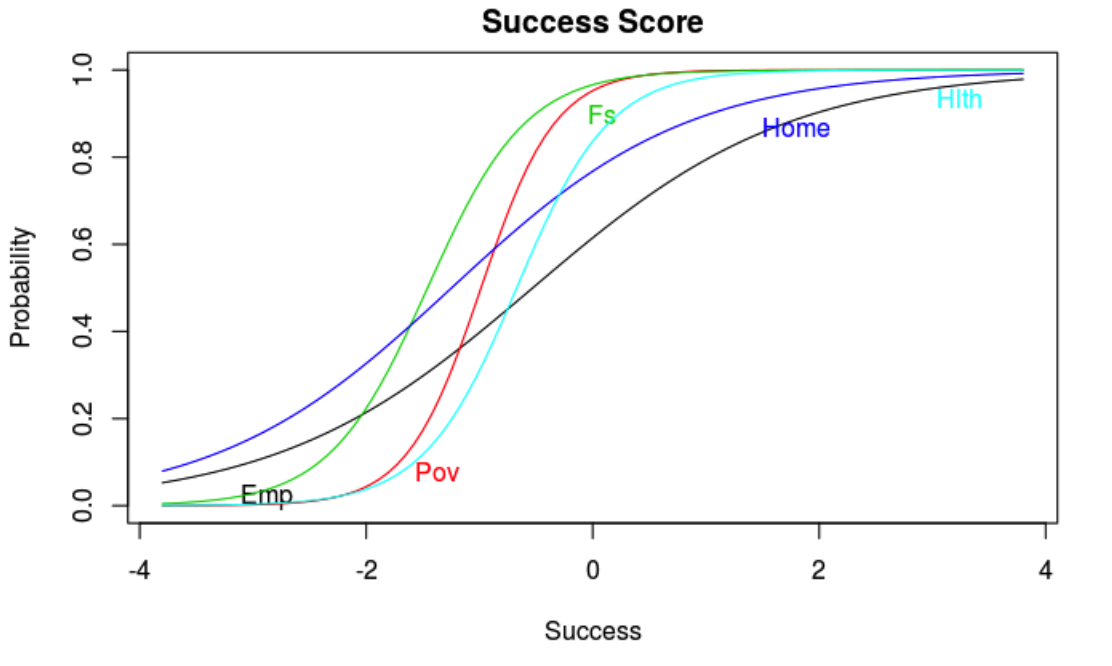


Figure 26: Item response for success score for both immigrants and non-immigrants

**Overall Success Item Response Parameters**

|  |  |  |
| --- | --- | --- |
| Variable | Difficulty | Discrimination |
| Emp | -0.5293679 | 0.8833654 |
| Pov | -0.9856305 | 3.0443455 |
| Fs | -1.4579614 | 2.2981252 |
| Home | -1.2457414 | 0.9595585 |
| Hlth | -0.6710364 | 2.4315518 |

Figure 1 shows the success score based on item response. The highest discrimination parameter is poverty (3.044), meaning that the probability of a person living above the poverty level increases most rapidly as the success score increases. So then out of our 5 variables, poverty affects the success score the most. Employment has the lowest discrimination parameter (0.88336), meaning thatemployment affects success score the least.

For example, when the probability of having health insurance is 0.5, the success score is -0.67, after controlling for the other variables. The difficulty scores indicate that owning a home (-1.25) and not depending on assistance to purchase food (-1.45796), is hardest to attain, while being employed is easiest (-.52). The difficulty score indicates the success score when the probability a given variable is 0.5, after accounting for other variables.

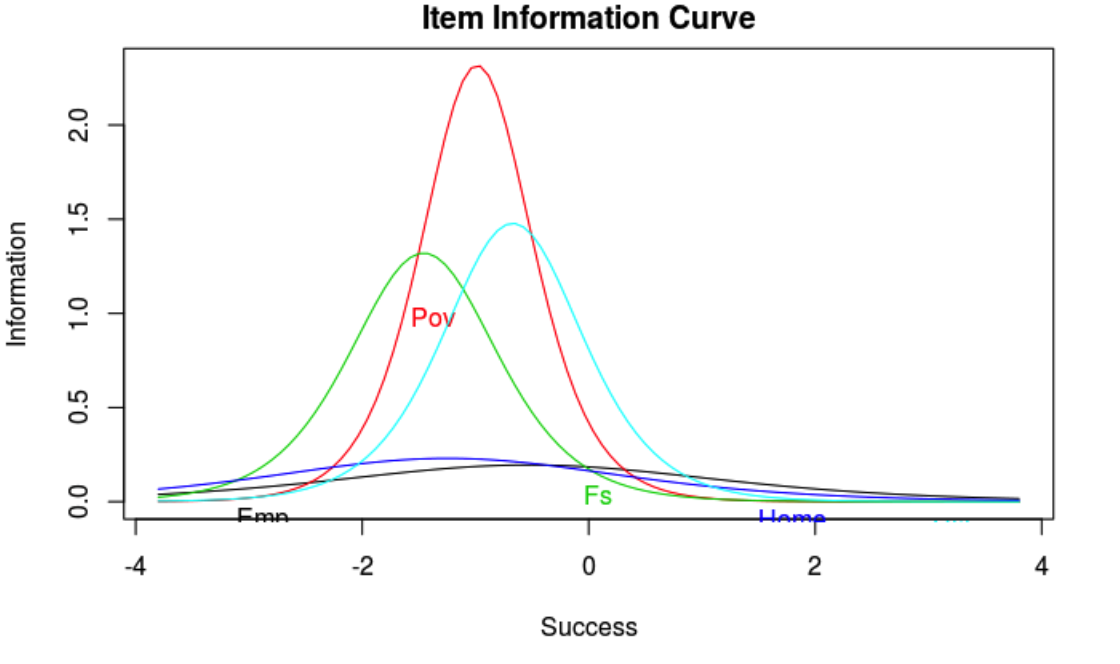


Figure 27: Item Information curve for success score for both immigrants and non-immigrants

From the item information curve (Figure 2), we see that poverty gives the most information about success score (approximately 2.3), with food stamps and healthcare giving between 1 and 1.5 on the information curve. Health insurance and employment status give us the least information about ability. This is consistent with our discrimination parameters.

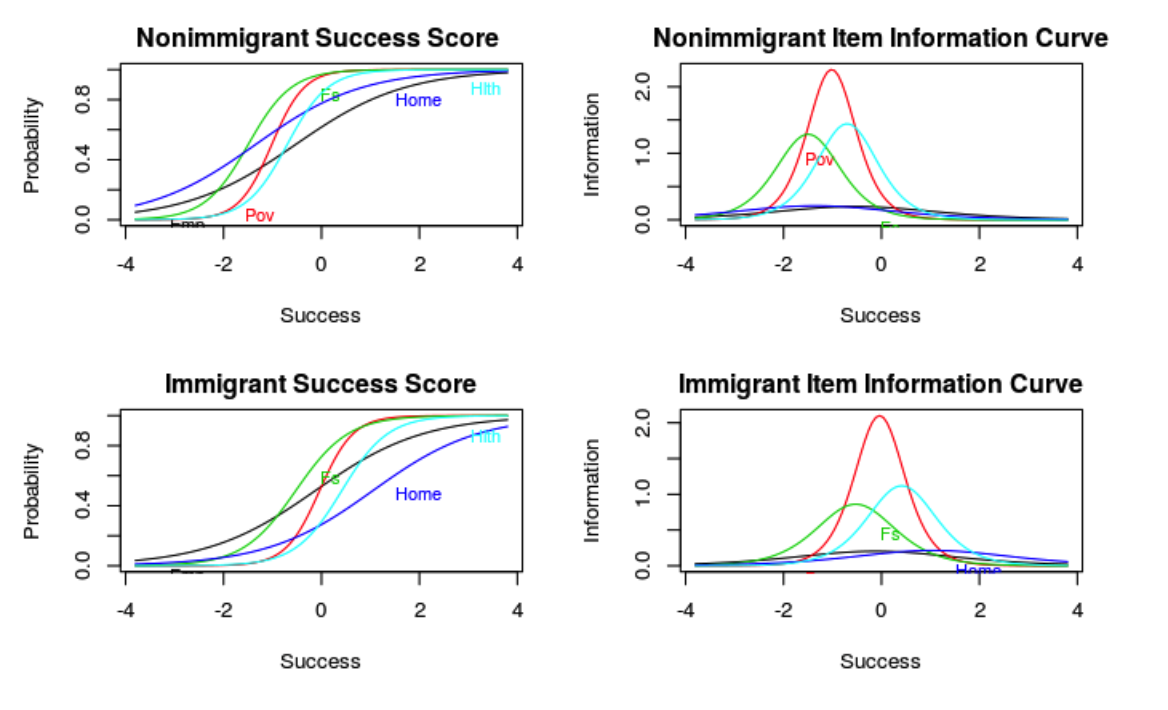


Figure 28: Success curve and item information curve for immigrants and nonimmigrants individually

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nonimmigrant Success Item Response Parameters**   |  |  |  | | --- | --- | --- | | Variable | Difficulty | Discrimination | | Emp | -0.534169 | 0.8936458 | | Pov | -1.0171565 | 3.0058008 | | Fs | -1.4982122 | 2.2678067 | | Home | -1.3511159 | 0.9175723 | | Hlth | -0.7015069 | 2.4023518 | | **Immigrant Success Item Response Parameters**   |  |  |  | | --- | --- | --- | | Variable | Difficulty | Discrimination | | Emp | -0.1272168 | 0.9019129 | | Pov | -0.0395891 | 2.8985569 | | Fs | -0.526123 | 1.8563316 | | Home | 1.0479563 | 0.9257858 | | Hlth | 0.41413216 | 2.1158999 | |

In the nonimmigrant success item response parameters, we look specifically at attaining success in the nonimmigrant population (Figure 3). The highest discrimination parameter is again poverty (3.0058), meaning that the probability of a person being above the poverty level increases most rapidly as the success score increases. Also again, employment has the lowest discrimination parameter (0.8936), meaning that employment affects success score the least. It is interesting to see that owning a home does not have as much of an impact in the nonimmigrant population as it does in the overall population (.917 vs .96). Food stamps has the lowest difficulty score, meaning that the the when the probability of having a food stamp is .5, the success score is at -1.4982, holding other variables constant. This is rather low compared to employment, which shows that when the probability of being employed is .5, the success score is at -.534, holding other variables constant. The item information curve gives a similar story.

In the immigrant success item response, we look solely at success in the immigrant population. Similar to the overall and the nonimmigrant population, the highest discrimination parameter is poverty (2.89). Also, employment has the lowest discrimination parameter (0..9019), meaning that employment affects success score the least. It is interesting to see that, although the discrimination score for poverty decreased from the nonimmigrant score (2.89 vs 3.005), the discrimination score for employment (.901) and owning a home (.2957) increased. This means that, out of the immigrants, employment and owning a home matter more for success, than they would for nonimmigrants. Food stamps has the lowest difficulty score away from, meaning that the the when the probability of having a food stamp is .5, the success score is at -.526, holding other variables constant. When an immigrant owns a home, they have a very high difficulty score (1.0479), meaning that when the probability of owning a home is .5, the success score is 1.05 after holding other variables constant.

Overall, it appears as though when a person owns food stamps, it is very difficult to attain success in all groups. Also, a person being above the poverty level seems to have the most impact on having a higher success score. It is interesting to see that the discrimination parameter of employment and owning a home are higher for success in the immigrant population, than it is in the nonimmigrant population; although poverty is larger in success for the nonimmigrant population than it is in the immigrant population.

*Success mini-EDA*

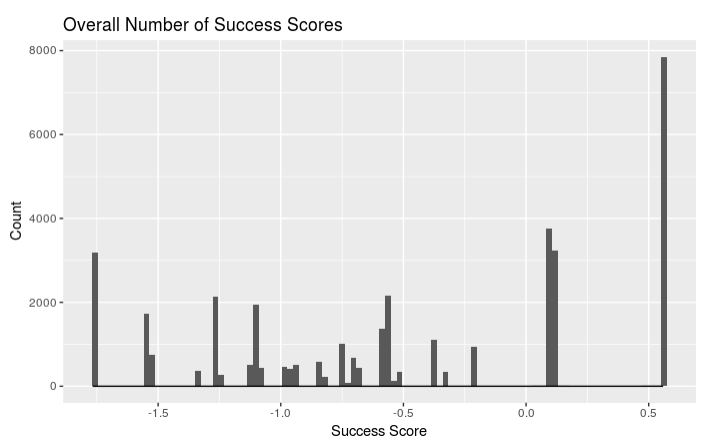
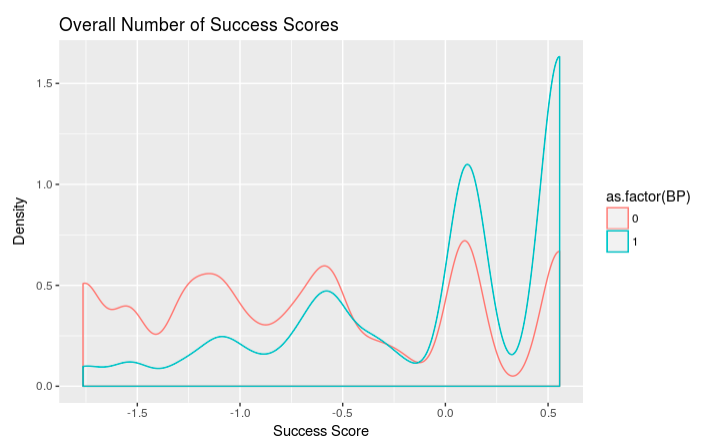
**

Figure 29: Frequency histogram of success scores Figure 30: Density plot of success scores, differentiated by immigrants and nonimmigrants

From the overall success score histogram, we can see that there is a large number of people who have success scores of 0.55. However, when we further look at the density plot, we can see that a large proportion of the people who have success scores over 0.5, are nonimmigrants. From the density plot, we find that a small area of the nonimmigrant population is below a success score of 0, and a large amount are between 0 and 0.55. However, from the immigrant population, we find that there is a larger area of that population that is less than 0, and not many who have a success score of over 0.

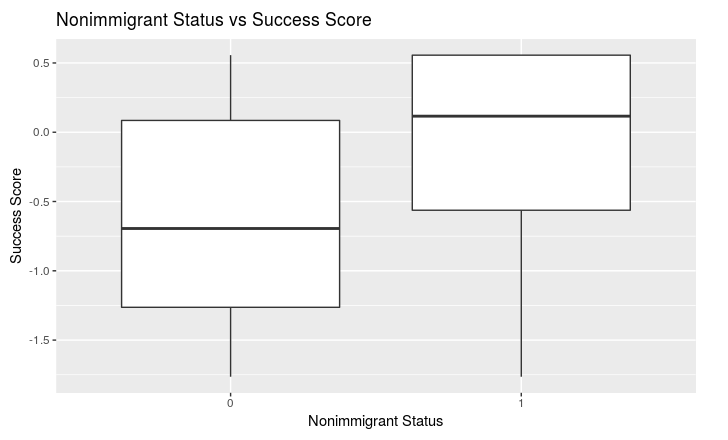


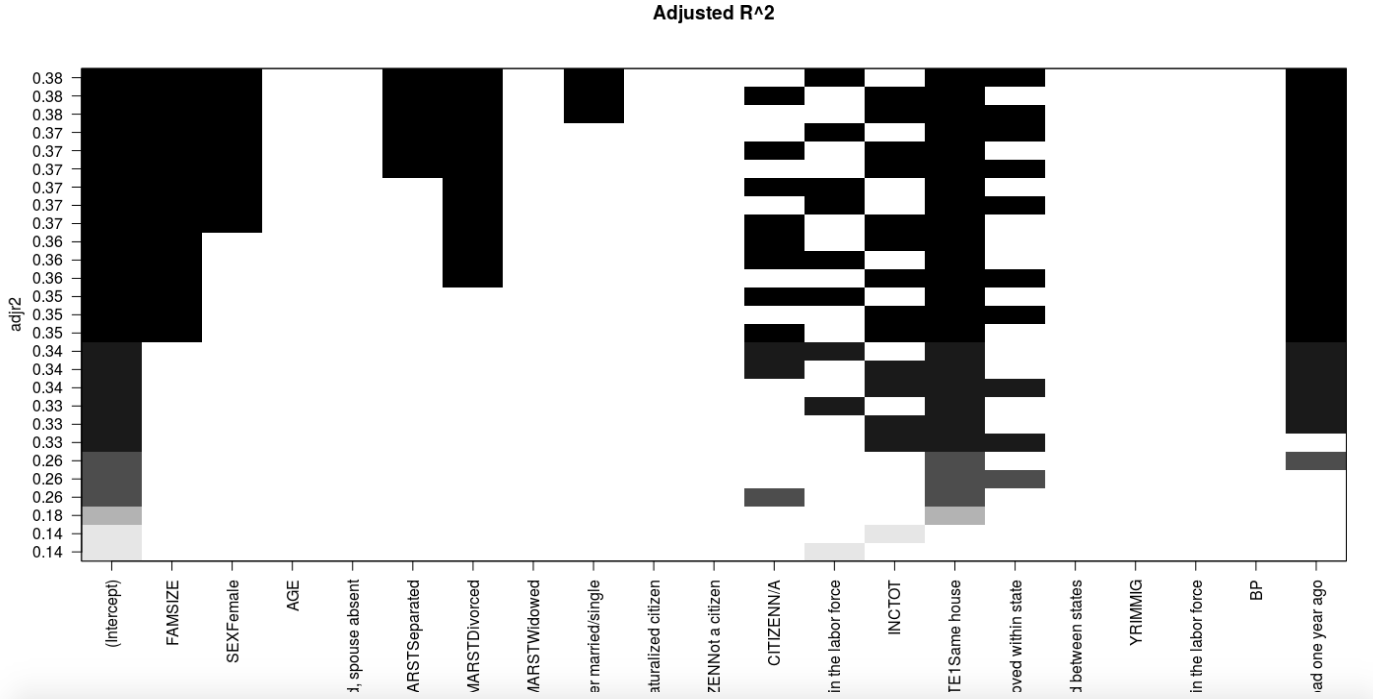
Figure 31: Boxplot showing the differences in success scores for immigrants and nonimmigrants

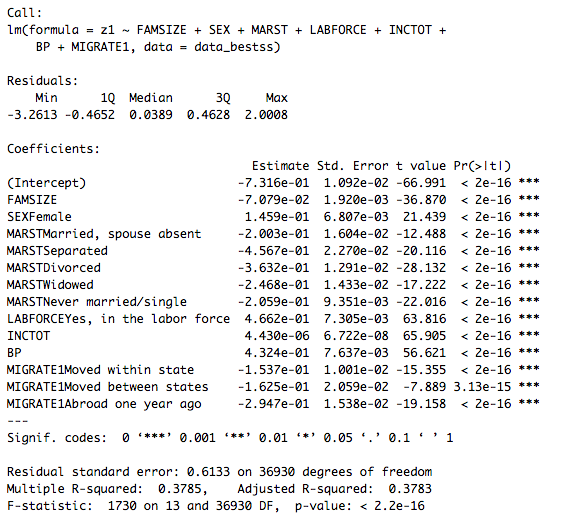
|  |  |  |
| --- | --- | --- |
| Group | Success Mean | Success Median |
| Overall | -0.4773753 | -0.5628491 |
| Immigrant | -0.6343786 | -0.694841 |
| Nonimmigrant | 0.05434582 | 0.1158217 |

When comparing the mean success scores for immigrants and nonimmigrants, we can see that immigrants have lower mean success scores (-0.634) than nonimmigrants (0.054). Furthermore, from the boxplot we can see that the nonimmigrants have clearly higher success scores (median: 0.1158 vs -0.6948). Whereas the immigrant population is normally distributed, the nonimmigrant population is skewed left, showing that the majority of nonimmigrants are more successful than immigrants. We can even see that the top 75% of the nonimmigrant population is still doing better than the bottom 50% of the immigrant population.

*Best Subsets & Model Selection*

In order to select our model, we used a best subset analysis to maximize our R-squared value. We chose 10 variables that we thought appropriate, to include in our best subset analysis: Family size, Sex, Age, Marital Status, Citizenship, Immigration Year, Labor force, total income, non immigration/immigration indicator, and 1 year migration status. From the best subsets plot, we can see that in order to maximize an adjusted R-squared of approximately 0.38, we should include: Family size, Sex, Marital Status, Immigration Year, Insurance coverage, labor force, and 1 year migration status. Though our indicator variable for nonimmigrant or immigrant is not part of the best-subset, it is a variable we wanted to look at so we added it to the model. We also added total income, because we felt like it is a good variable that we would be able to compare between the two different groups, and a good measure of success. Though our final model has an adjusted R-squared of 0.3785 instead of 0.38, we felt as though it is okay to lose some of the variability, if we can also account for income and immigrants/nonimmigrants.

*Final Model*



|  |  |  |  |
| --- | --- | --- | --- |
| GVIF | Df | GVIF^(1/(2\*Df)) | |
| FAMSIZE | 1.212395 | 1 | 1.101088 |
| SEX | 1.136873 | 1 | 1.066242 |
| MARST | 1.2989 | 5 | 1.026497 |
| LABFORCE | 1.236581 | 1 | 1.112016 |
| INCTOT | 1.179323 | 1 | 1.085966 |
| BP | 1.130599 | 1 | 1.063296 |
| MIGRATE1 | 1.058363 | 3 | 1.009499 |

Our final model includes family size, sex, marital status, health insurance coverage, labor force, total income, immigrant/non immigrant indicator, and migration status. When comparing this model to a smaller model with only the immigrant/non-immigrant indicator, we found that the larger model was more significant (F = 1330.8, p = 2.2e-16). We also compared our model to a model without 1 year migration status, and also found that the larger model was more significant (F = 195.83, p = 2.2e-16). Also, when looking at the variance inflation factors,, we observe that all of the factors are less than 1.3, which indicates low collinearity between all of our variables.

From our summary statistics, we see that if a person is single, male, married, does not have health insurance, is not in the labor force, is not an immigrant, and did not move within the last year, then the estimated success score is -0.73. For every one person increases in family size the estimated success score decreases by 0.07. When we look specifically at the immigrant/non-immigrant indicator (BP), we find that the estimated success score (z1) increases by .43 points if the person is a non-immigrant, after accounting for all other variables (p = 2e-16). Another interesting notion is that if a person is in the labor force the estimated success score increases by 0.466, after accounting for other variables. This is the highest increase for all of the variables in our model.

**Conclusion and limitation**

Our longitudinal model suggests that there is no significant relationship between violent crime rate and immigrant arrival from the travel ban countries from 2005 to 2014. This invalidates the negative stereotypes that drive the Trump Administration to ban people from Muslim-majority countries to ensure “the safety and security of the American people”. In fact, cluster analysis infers that immigrants living predominantly white, liberal, and rich states tend to be underprivileged. In these states, we see lower total family income and a higher ratio of immigrants who are unemployed, do not own a car, and receive Food Stamps compared to those living in other states. This is because immigrants within more privileged states are often under-privileged making their living conditions more polarized. This likely prompts negative stereotypes that immigrants are poor, which then makes their integration into the US even harder. In fact, results from the Item Response Theory authenticate this further by showing that immigrants obtain significantly less success than non-immigrants. This is because their definition of success is achieved through owning a home instead of having health insurance as it is the case for nonimmigrants. This leads to our conclusion that there is an economic gap between immigrants and nonimmigrants which could be attributed to the disparity in the available opportunities for both groups. Further research should be done to create policy changes that provide economic equity and prevent polarization of immigrants and nonimmigrants.

Our research project faces several limitations. It is often hard to evaluate results from unsupervised learning methods because there is no such prediction model to validate results, which are subject to different interpretations. In this way, although there is no such true answer to compare, the results can be used to check works before modeling. Because of time limit, we could not include and control for more state-level variables in the longitudinal model that explains violent crime rate. We were not able to model terrorism incident data because it is too sparse. For the analysis that uses item response theory, we assume independence between all variables. However, it is difficult to assume true independence between all factors (e.g. poverty and food stamps).

**Appendix**

*Data table for the cluster analysis*

*State Level Data*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Quantitative or Categorical** | **Meaning** | **Possible Responses** |
| White\_population | Quantitative | The percent of white population in a state | 0-100% |
| Redstate\_vote | Quantitative | The percent of vote for trump in election 2016 in a state | 0-100% |
| Christian\_Population | Quantitative | The percent of christian population in a state | 0-100% |
| Unemployment\_rate | Quantitative | The unemployment rate from 2016 to 2017 for each state | 0-100% |

*Individual Level Data*

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Name** | **Quantitative or Categorical** | **Meaning** | **Possible Responses** |
| Metropolitan status | Categorical | Whether the household was located within a metropolitan area | 1 (metra area), 2 (non-metra) |
| Ownership of dwelling | Categorical | Whether the housing unit was rented or owned by its inhabitants. Housing units acquired with a mortgage or other lending arrangement(s) are classified as "owned," even if repayment was not yet completed. | 1(owned),  2 (not owned) |
| Food\_stamp\_recipiency | Categorical | Whether anyone in the household received Food Stamps (now called the Supplemental Nutrition Assistance Program, or SNAP) at any time in the past 12 months. | 1(no),  2(yes) |
| Sex | Categorical | Whether the person was male or female. | 1(male), 2(female) |
| Age | Quantitative | The person's age in years as of the last birthday | 0-100 |
| Citizenship\_Status | Categorical | The citizenship status of respondents, distinguishing between naturalized citizens and non­citizens | 1(foreign born), 2(citizen), 3(not citizen) |
| Race | Categorical | The race of respondents | 1(White), 2(Black),3 (American Indian/Alaska Native),4 (Asian/pacific islander),5 (other race, non-Hispanic) |
| Any\_health\_insurance\_converage | Categorical | Whether persons had any health insurance coverage at the time of interview | 1(no health insurance coverage),  2 (with health insurance coverage) |
| Employment\_general\_version | Categorical | Whether the respondent was a part of the labor force ­­ working or seeking work ­­ and, if so, whether the person was currently unemployed | 1(employed), 2(not\_employed) |
| Independent\_Living\_Difficulty | Categorical | Whether the respondent has any physical, mental, or emotional condition lasting six months or more that makes it difficult or impossible to perform basic activities outside the home alone. | 1(has no difficulty), 2(has a difficulty) |
| Poverty | Quantitative | Each family's total income for the previous year as a percentage of the poverty thresholds established by the Social Security Administration in 1964 and subsequently revised in 1980, adjusted for inflation | 0-501 |
| Means of transportation | Categorical | Whether the respondent has a car or not | 1(car),  2(no car) |
| Total\_Family\_Income | Quantitative | The total pre­tax money income earned by one's family (as defined by FAMUNIT) from all sources for the previous year. | 0-651000 |
| Birthplace\_final | Categorical | Where the person was born | 1(Egypt), 2(Iran), 3(Iraq), 4(Libya), 5(Somalia), 6(Sudan), 7(Syria) |
| Cluster\_number | Categorical | Which cluster the person belongs | 1-3 |
| State\_Residency\_final | Categorical | The U.S. state where the respondent was living 1 year ago. | 1-51 states (Alphabetically named) |

*Data table for Longitudinal Model*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable Name** | **Type** | **Definition** | **Values** | **Level** |
| Immigrant arrival | Quantitative  (main explanatory variable) | Number of immigrant arrival from 10 travel ban countries in a state for one year | 3 - 17184 immigrants | 1 |
| Violent crime rate | Quantitative  (response variable) | Rate of violent crime per 100,000 population | 99.3 - 785.7 | 1 |
| Violent crime | Quantitative | Total number of violent crime | 622-194483 crimes | 1 |
| Population | Quantitative | Number of residents | 522830 - 38802500 residents | 1 |
| Firearms recovered | Quantitative | Total number of firearms recovered | 20-32069  firearms | 1 |
| Law total | Quantitative | Total number of firearm regulations | 3-100 laws | 1 |
| Legislation composition | Categorical | Partisan composition of state legislatures | dem=democrat  rep=republican  split | 1 |
| Poppct\_urban | Quantitative | Percentage of the state’s population lives in urban areas | 38.66% - 94.95% | 2 |
| Areapct\_urban | Quantitative | Percentage of urban areas, defined as either Urbanized Areas of 50,000 or more people or Urban Clusters of at least 2,500 and less than 50,000 people. | 0.05% - 39.7% | 2 |
| Cluster | Categorical | Which cluster the state belongs to based on White population, Christian population, unemployment rate and percent of vote for Trump | 1, 2, 3 | 2 |

*Data table for Item Response Theory*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable Name** | **Meaning** | **Variable Type** | **Values** | **Used for/in:** |
| z1 | Success Score | Response | Continuous from-1.7644 to 0.5563 | Response Variable |
| Employment | Employment Status | Binary | 1= Employed  0= Not-employed | IRT |
| Poverty | Poverty Status | Binary | 1= Above Poverty  0=Below Poverty | IRT |
| Foodstamp | Receiving food stamps | Binary | 1= Not receiving food stamps  0=Receiving food stamps | IRT |
| Ownhome | Home ownership | Binary | 1= Owning a home  0=Not owning a home | IRT |
| Health | Have health insurance | Binary | 1= Have health insurance  0=Don't have health insurance | IRT |
| FAMSIZE | Family Size | Explanatory | Discrete from 1 to 19 | Final Model |
| SEX | Sex | Explanatory | Male or Female | Final Model |
| AGE | Age | Explanatory | Discrete from 25 to 95 | Best Subset Analysis |
| MARST | Marital Status | Explanatory | Married, spouse present;  Married, spouse absent;  Separated;  Divorced;  Widowed;  Never married/single | Final Model |
| CITIZEN | Citizenship status | Explanatory | Born abroad of American Parents;  Naturalized citizen;  Not a citizen;  N/A | Best Subset Analysis |
| YRIMMIG | Immigration Year | Explanatory | Immigration Year from 2001-2016 | Best Subset Analysis |
| LABFORCE | Labor Force status | Explanatory | No, not in the labor force;  Yes, in the labor force | Final Model |
| INCTOT | Total income | Explanatory | Total income from 0 to 888000 | Final Model |
| BP | Immigrant/nonimmigrant indicator | Explanatory | 1=Nonimmigrant;  0=Immigrant | Final Model |
| MIGRATE1 | 1 Year Migration Status | Explanatory | Same house;  Moved within state;  Moved between states;  Abroad one year | Final Model |

1. Detailed information about variables is in Appendix. [↑](#footnote-ref-0)