Stats 340 Project Final Report

Names: Bhavishya Hasija, Brenden Paddock, Esmma Almousa, Michael Kornely, Neel Chaudhari

Group Name: hammerhead

NetIDs: bhasija, brpaddock, ealmousa, mkornely, nchaudhari

Abstract:

Our analysis focuses on two datasets. The first dataset, called “Inmates Under Custody: Beginning 2008” incorporated personal factors of each inmate convicted in the state of New York from 2008-2021. Our second dataset, called county_complete, consisted of demographic data of each county in the United States. The two datasets were filtered and joined together by only New York counties. We performed PCA, hot encoding, and clustering on this data in order for us to answer our statistical question: **Are personal factors (including age, gender, and race/ethnicity) predictors for the level of facility security (medium facility security, or maximum facility security) placed on inmates who were convicted of the same crime? Furthermore, are predictors using demographic data from the county of conviction, statistically significant in predicting the security level?** Our first analysis found that age was a highly significant factor in determining the security level for inmates who committed the same crimes. One surprising finding was that gender is not a significant predictor of security level for crimes like 2nd-degree murder and burglary. The second analysis found that the education and unemployment rates were good predictors of the security levels for those inmates. We were surprised to see that non-English speaking households were a significant predictor for this model.

Discussion on Statistical Question:

When interpreting the definition of Medium Security Level Facilities, these facilities are standard facilities used to house the majority of criminals. Although routines are still strictly enforced by guards and workers, prisoners still typically encounter typical interactions with other inmates and the facility is enclosed by a razor-wire fence. Maximum Facility Security levels, however, are specifically designed for inmates who pose a greater threat to society due to their violent behaviors and actions. These inmates require a much more strict level of supervision. These facilities have multiple fences around their building or even a sturdy wall. This is an important factor in prison facilities not only for the sake of society since heavier restrictions are placed on inmates who are believed to have a higher chance of attempting to escape but also for the protection of guards and other inmates in the facility.

Based on this information, we were interested in discovering whether or not personal factors about the convicted individual could potentially be a predictor for the security level that they are sentenced with. When we first discovered our dataset, we were most intrigued by this variable and wanted to learn more about its importance and influence on the criminal justice system. Since individuals placed under maximum security are typically labeled as more likely to “pose a greater threat,” we wanted to investigate whether or not there were any potential predictors of this determination such as age, race, or gender. Additionally, our dataset incorporated the county of the indictment based on the inmate’s actions. We discovered another dataset with county demographics that we wanted to merge with our original dataset(Inmates Under Custody: Beginning 2008) and find whether or not demographic data was statistically significant in predicting the facility security level for a convicted criminal. This part of the question was important to us because we wanted to take a look at the county population that was being affected by the crime and whether or not their society’s demographics were relevant to the security level decision.

Link to where the New York Crime dataset can be found:
https://catalog.data.gov/dataset/inmates-under-custody-beginning-2008
(We downloaded and used the CSV file)

This dataset, called “Inmates Under Custody: Beginning 2008” can be found on data.gov which is an open-source of U.S. government data. Esmma initially found this dataset. This set captures information about inmates in New York Prison Facilities from March 31, 2008, until August 6th, 2021. The following data were collected on each inmate in this dataset: year, latest admission type, county of the indictment, gender, most serious crime, current age, housing facility, facility security, and race/ethnicity. We are interested specifically in using the following variables to answer our question:

Variables Used:

Snapshot Year: Year represented by the snapshot population. Data represents inmates under custody on March 31, the end of the Fiscal Year.
Latest Admission Type: New Court Commitment = new admission to prison on current offense; Returned Parole Violator (RPV) = return to prison for a technical violation of parole conditions; Other = This category represents returns to prison for something other than a new court commitment or a parole violation. It consists primarily of returns from other agencies and returns from court-ordered discharge.
County of Indictment: County of indictment associated with the most serious conviction offense
Gender: Inmate gender
Most Serious Crime: Most serious conviction offense of inmate, as defined by the offense with the longest maximum sentence.
Housing Facility: Facility in which inmate was housed as of the filing date
Security Level: Security Level associated with inmate housing facility as of the file date
Race/Ethnicity: Race/Ethnicity as reported by the inmate, adjusted for country of birth and parental country of birth to determine Hispanic ethnicity. Inmates with Hispanic ethnicity will be reported as Hispanic, regardless of reported race.

We will primarily be looking at the Most Serious Crime, Age, Race/Ethnicity, and County of Indictment variables. We categorized age into groups (17-26, 27-36, 37-46, etc). Since the data dates back to 2008, it is possible things have changed in the incarceration process in New York. To account for this, we are using data from certain years/time periods. There are also over 528 crimes. We will use the top 5-10 committed crimes in that period since we have much more data on those crimes than some of the more specific ones for our specific procedures. The larger amounts of data will allow us to build a stronger model, with hopefully, less variance.

This dataset interested us because it contained many personal factors about each inmate in New York prison facilities that would allow us to test whether or not these factors have an influence on crimes committed. The reader should care about this project because there have been drastic increases in incarceration in the United States and has been an arising issue for the past few years. New York has one of the highest percentages of its population in prison. Analyzing data on the inmates in prison institutions may help us find some answers on why and how this is happening and how we can “target” specific groups that have higher rates of a type of crime or recidivism so we can prevent it from happening in the first place. Given that New York is a diverse and populous state, we are given a dataset that could be more generalized to the United States as a whole.

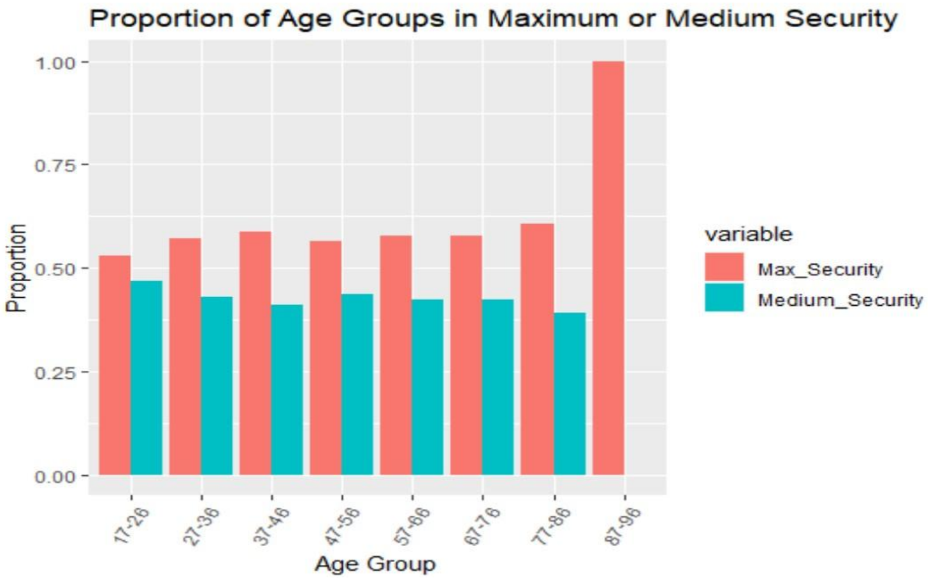
Link to where the demographic dataset can be found:
https://www.openintro.org/data/index.php?data=county_complete (CSV provided in the course files)

Variables Used from Counties Dataset:

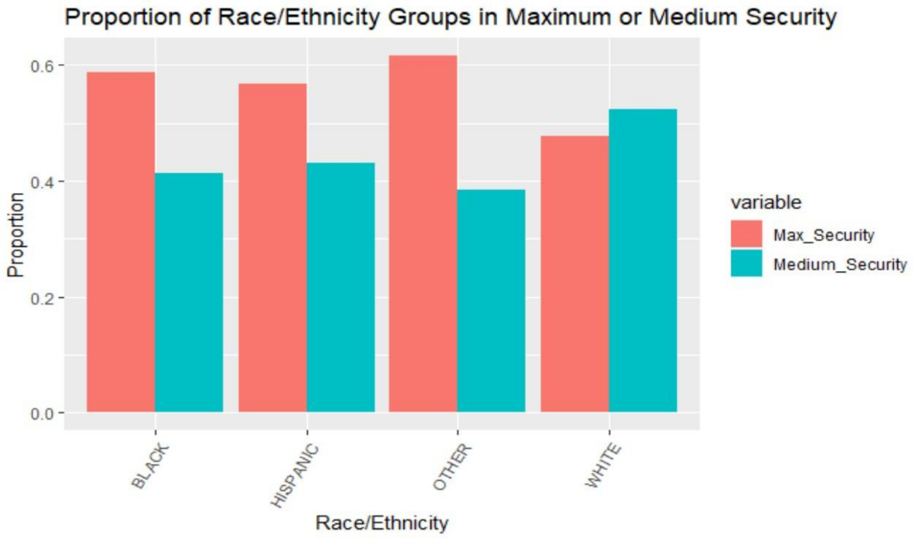
Some_college_2016: Percent of population with some college education (2012-2016)
Hs_grad_2016:Percent of population 25 and older that is a high school graduate(2012-2016)
Bachelors_2016:Percent of population that earned a bachelor's degree (2012-2016).
Poverty_2016: Percent of population below poverty level (2012-2016).
Unemployed_2016: Number of civilians unemployed in 2016.
Bachelors_2019: Percent of population 25 and older that earned a Bachelor's degree or higher (2015-2019).
Households_speak_limited_english_2019:Percent of limited English-speaking households (2015-2019).
Households_speak_other_2019: Percent of households speaking non European or Asian/Pacific Island language (2015-2019).
Households_speak_spanish_2019: Percent of households speaking Spanish (2015-2019).
Housing_mobile_homes_2019: Percent of housing units in mobile homes and other types of units (2015-2019).
Hs_grad_2019: Percent of population 25 and older that is a high school graduate (2015-2019).
Mean_household_income_2019: Mean household income (2019 dollars, 2015-2019).
Median_indviidual_income_2019: Median individual income (2019).
Poverty_2019: Percent of population below the poverty level (2015-2019).
Uninsured_2019: Percent of population below the poverty level (2015-2019).
Some_college_2017: Percent of population with some college education (2017).
Hs_grad_2017: Percent of population that is a high school graduate (2017)
Bachelors_2017: Percent of population that earned a bachelor's degree (2017)
Poverty_2017: Percent of population below poverty level (2017)
Median_household_income_2017: Median household income (2017)
Uninsured_2017: Percent of population who are uninsured (2017)
Speak_english_only_2017: Percent of population that speaks English only (2017)

We used the county_complete dataset which contains a large collection of various statistics for each county taken from a variety of surveys, censuses, and studies across several years. The dataset contains a variety of variables taken from several years which we will incorporate into our logistic models. The variables we will be including for our models is:

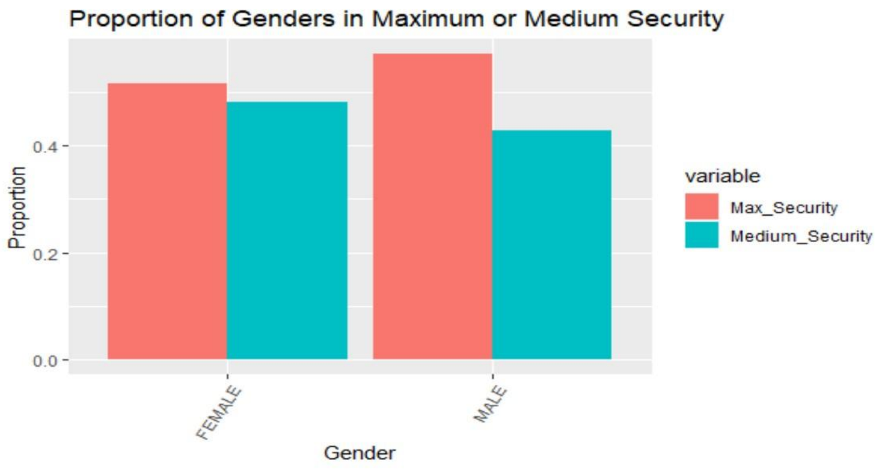
Since the New York dataset provides only a few variables regarding the individual, we decided that we could merge the two datasets by the county in order to get some more demographics about the community. For example, their education level, the poverty level, languages spoken, income, and more. This could help us reduce the variance of our models to help us predict why some criminals go into specific security prisons.



The plot represents the proportions of each age group placed in Maximum or Medium Security for their crime. The youngest age group (17-26 years old) had the lowest proportion with approximately 53% of those in this group placed under Maximum Security, while the eldest age group (87-96 years old) were all placed under Maximum Security. This made us question whether or not elderly individuals were automatically sent to Maximum Facility Securities due to their old age which is linked to other factors (such as poor health or a slow decline in life quality due to old age).



The plot represents the proportions of each Race/Ethnicity placed in Maximum or Medium Security for their crime. White individuals had the lowest proportion placed under Maximum Security at approximately 48%. The category “other” had the highest proportion with approximately 62% placed under Maximum Security. Black individuals had the second-highest proportion placed under Maximum Security at approximately 59%, and Hispanic individuals had the third-highest with about 57% placed under Maximum Security. This plot made us question why the proportion of White individuals placed under Maximum Security was so much lower and whether or not race is a predictor for the level of facility security a person is placed under for their crime.



The plot represents the proportions of each gender placed in Maximum or Medium Security for their crime. Approximately 57% of males were placed under Maximum Security while approximately 51% of females were placed under Maximum Security. Since males had a higher proportion in Maximum Security, we questioned whether or not gender was a potential predictor for the facility security level of inmates convicted of a crime.

Statistical Procedures

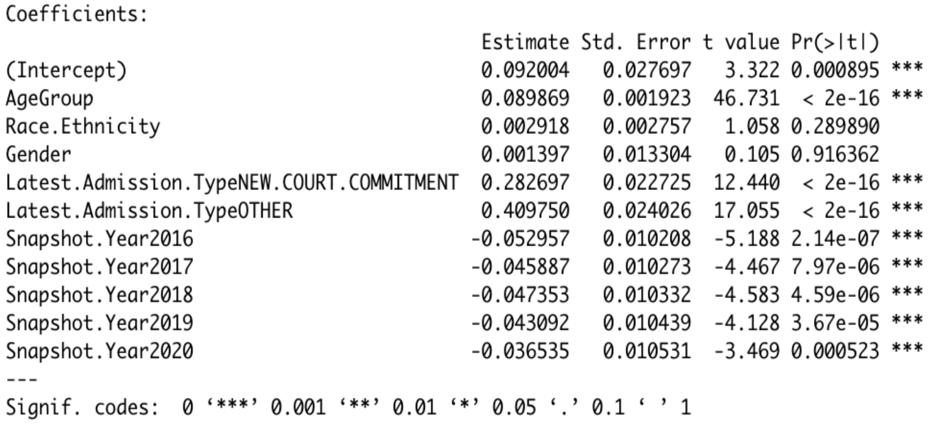
We decided to proceed in two different ways in the statistical procedure when making our logistic models.

First, using solely the data provided by the New York Inmates data, we created 5 models, filtering the data by the specific crime type in the 2016-2021 time period. The crimes we will be focusing on (in order of most committed):

- 2nd-degree murder
- 1st-degree robbery
- 2nd-degree possession of weaponry
- 1st-degree manslaughter
- 2nd-degree burglary

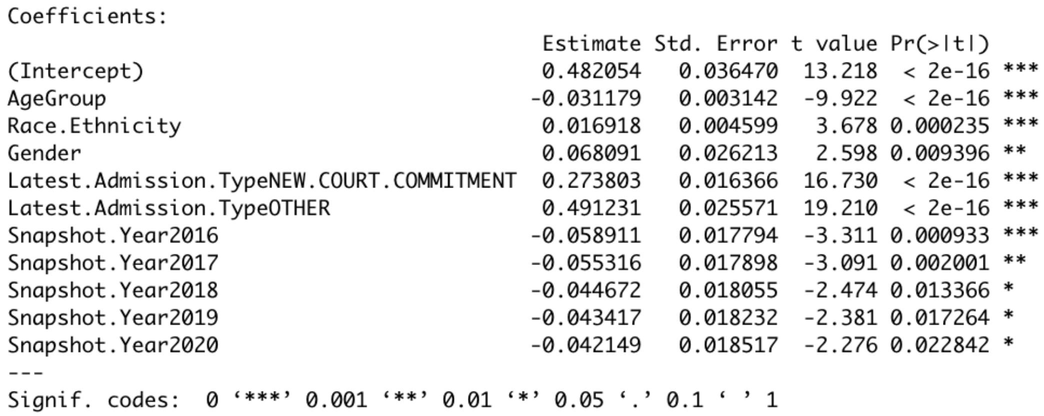
Here are the models:

- 2nd-degree murder



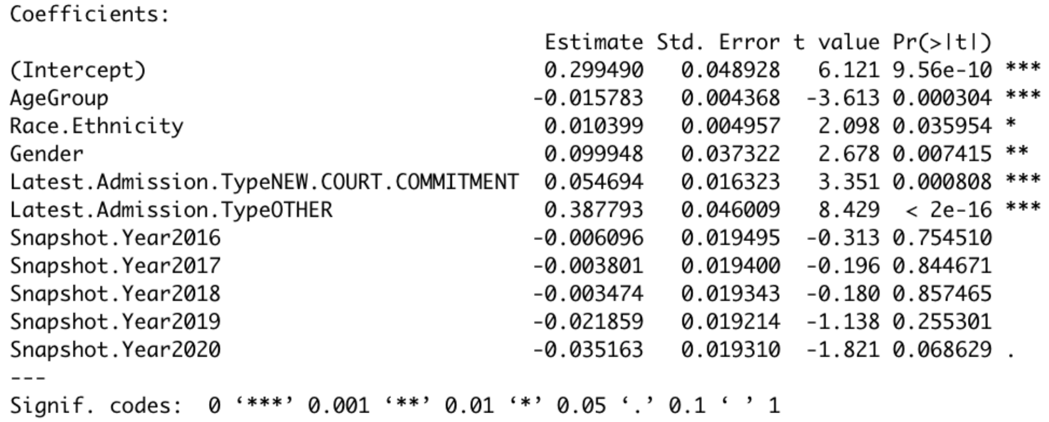
```
Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)    
(Intercept)                         0.092004   0.027697   3.322 0.000895 ***
AgeGroup                            0.089869   0.001923  46.731  < 2e-16 ***
Race.Ethnicity                      0.002918   0.002757   1.058 0.289890    
Gender                              0.001397   0.013304   0.105 0.916362    
Latest.Admission.TypeNEW.COURT.COMMITMENT  0.282697   0.022725  12.440  < 2e-16 ***
Latest.Admission.TypeOTHER          0.409750   0.024026  17.055  < 2e-16 ***
Snapshot.Year2016                  -0.052957   0.010208  -5.188 2.14e-07 ***
Snapshot.Year2017                  -0.045887   0.010273  -4.467 7.97e-06 ***
Snapshot.Year2018                  -0.047353   0.010332  -4.583 4.59e-06 ***
Snapshot.Year2019                  -0.043092   0.010439  -4.128 3.67e-05 ***
Snapshot.Year2020                  -0.036535   0.010531  -3.469 0.000523 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 1st-degree robbery



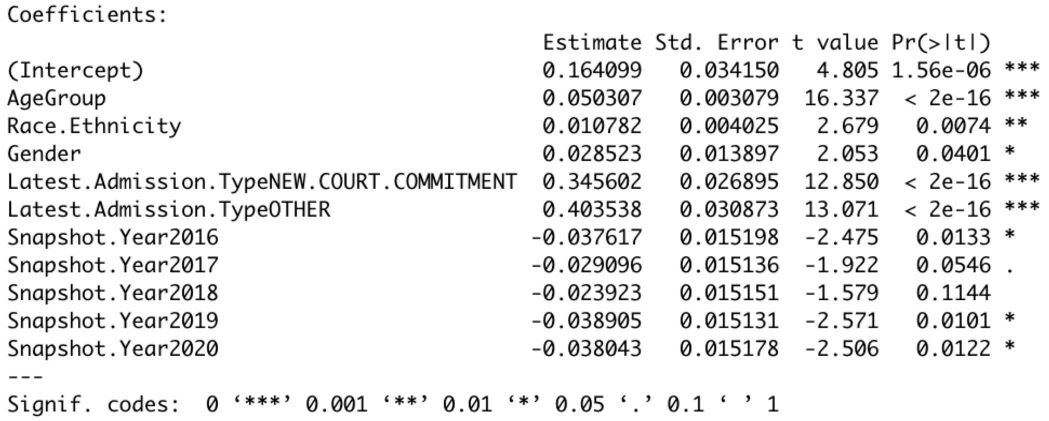
```
Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                           0.482054   0.036470  13.218  < 2e-16 ***
AgeGroup                             -0.031179   0.003142  -9.922  < 2e-16 ***
Race.Ethnicity                        0.016918   0.004599   3.678 0.000235 ***
Gender                                0.068091   0.026213   2.598 0.009396 ** 
Latest.Admission.TypeNEW.COURT.COMMITMENT  0.273803   0.016366  16.730  < 2e-16 ***
Latest.Admission.TypeOTHER            0.491231   0.025571  19.210  < 2e-16 ***
Snapshot.Year2016                    -0.058911   0.017794  -3.311 0.000933 ***
Snapshot.Year2017                    -0.055316   0.017898  -3.091 0.002001 ** 
Snapshot.Year2018                    -0.044672   0.018055  -2.474 0.013366 *  
Snapshot.Year2019                    -0.043417   0.018232  -2.381 0.017264 *  
Snapshot.Year2020                    -0.042149   0.018517  -2.276 0.022842 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 2nd-degree possession of weaponry



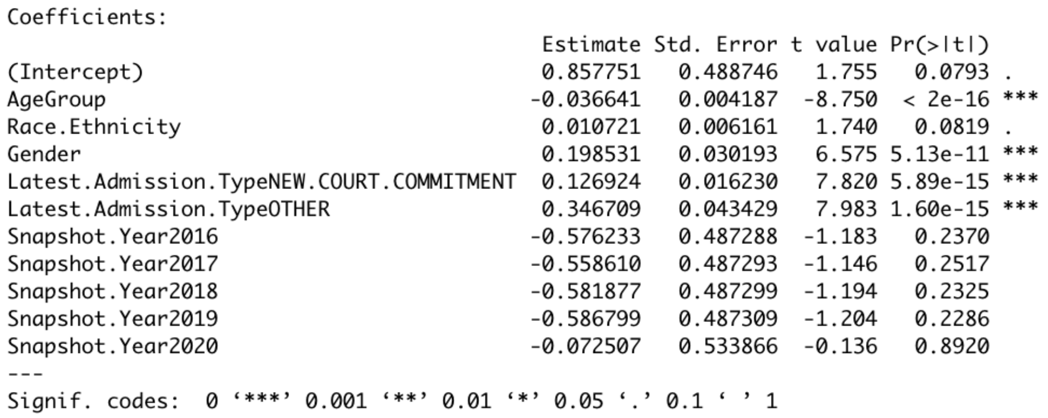
```
Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                           0.299490   0.048928   6.121 9.56e-10 ***
AgeGroup                             -0.015783   0.004368  -3.613 0.000304 ***
Race.Ethnicity                        0.010399   0.004957   2.098 0.035954 *  
Gender                                0.099948   0.037322   2.678 0.007415 ** 
Latest.Admission.TypeNEW.COURT.COMMITMENT  0.054694   0.016323   3.351 0.000808 ***
Latest.Admission.TypeOTHER            0.387793   0.046009   8.429  < 2e-16 ***
Snapshot.Year2016                    -0.006096   0.019495  -0.313 0.754510    
Snapshot.Year2017                    -0.003801   0.019400  -0.196 0.844671    
Snapshot.Year2018                    -0.003474   0.019343  -0.180 0.857465    
Snapshot.Year2019                    -0.021859   0.019214  -1.138 0.255301    
Snapshot.Year2020                    -0.035163   0.019310  -1.821 0.068629 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 1st-degree manslaughter



```
Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                           0.164099   0.034150   4.805 1.56e-06 ***
AgeGroup                              0.050307   0.003079  16.337  < 2e-16 ***
Race.Ethnicity                        0.010782   0.004025   2.679   0.0074 ** 
Gender                                0.028523   0.013897   2.053   0.0401 *  
Latest.Admission.TypeNEW.COURT.COMMITMENT  0.345602   0.026895  12.850  < 2e-16 ***
Latest.Admission.TypeOTHER            0.403538   0.030873  13.071  < 2e-16 ***
Snapshot.Year2016                    -0.037617   0.015198  -2.475   0.0133 *  
Snapshot.Year2017                    -0.029096   0.015136  -1.922   0.0546 .  
Snapshot.Year2018                    -0.023923   0.015151  -1.579   0.1144    
Snapshot.Year2019                    -0.038905   0.015131  -2.571   0.0101 *  
Snapshot.Year2020                    -0.038043   0.015178  -2.506   0.0122 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 2nd-degree burglary



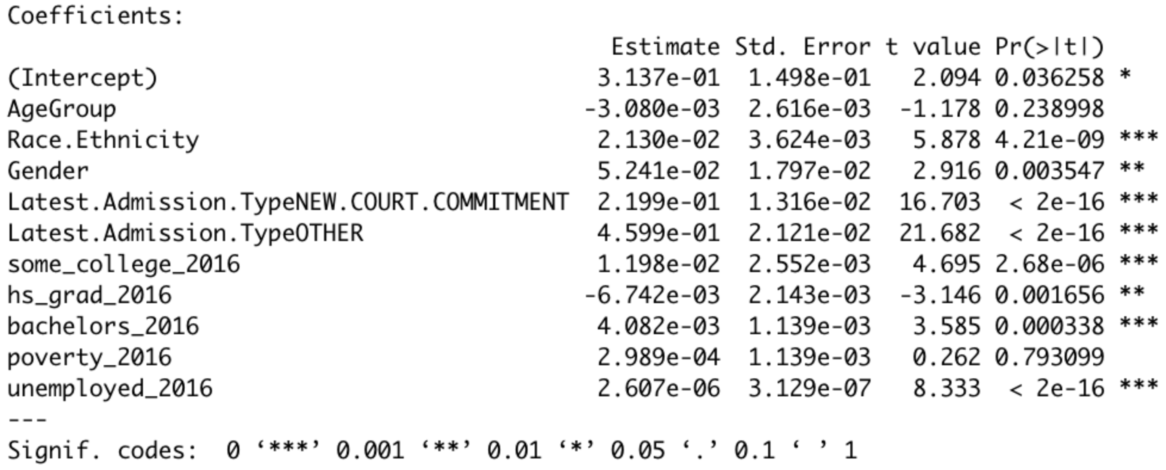
```
Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)    
(Intercept)                              0.857751   0.488746   1.755   0.0793 .  
AgeGroup                                -0.036641   0.004187  -8.750  < 2e-16 ***
Race.Ethnicity                           0.010721   0.006161   1.740   0.0819 .  
Gender                                   0.198531   0.030193   6.575 5.13e-11 ***
Latest.Admission.TypeNEW.COURT.COMMITMENT  0.126924   0.016230   7.820 5.89e-15 ***
Latest.Admission.TypeOTHER               0.346709   0.043429   7.983 1.60e-15 ***
Snapshot.Year2016                       -0.576233   0.487288  -1.183   0.2370    
Snapshot.Year2017                       -0.558610   0.487293  -1.146   0.2517    
Snapshot.Year2018                       -0.581877   0.487299  -1.194   0.2325    
Snapshot.Year2019                       -0.586799   0.487309  -1.204   0.2286    
Snapshot.Year2020                       -0.072507   0.533866  -0.136   0.8920    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis of statistically significant values while determining the level of security:

- 2nd-degree murder - All factors were statistically significant except for Race and Gender
- 1st-degree robbery -All factors were statistically significant
- 2nd-degree possession of weaponry - All factors were statistically significant except all the Snapshot.Year features of 2016 - 2020
- 1st-degree manslaughter - All factors were statistically significant except Snapshot.Year2017 and 2018
- 2nd-degree burglary - For this case only AgeGroup, Gender, Latest.Admission, and Latest.Admission.TypeNEW.COURT.COMMITMENT appear to be statistically significant

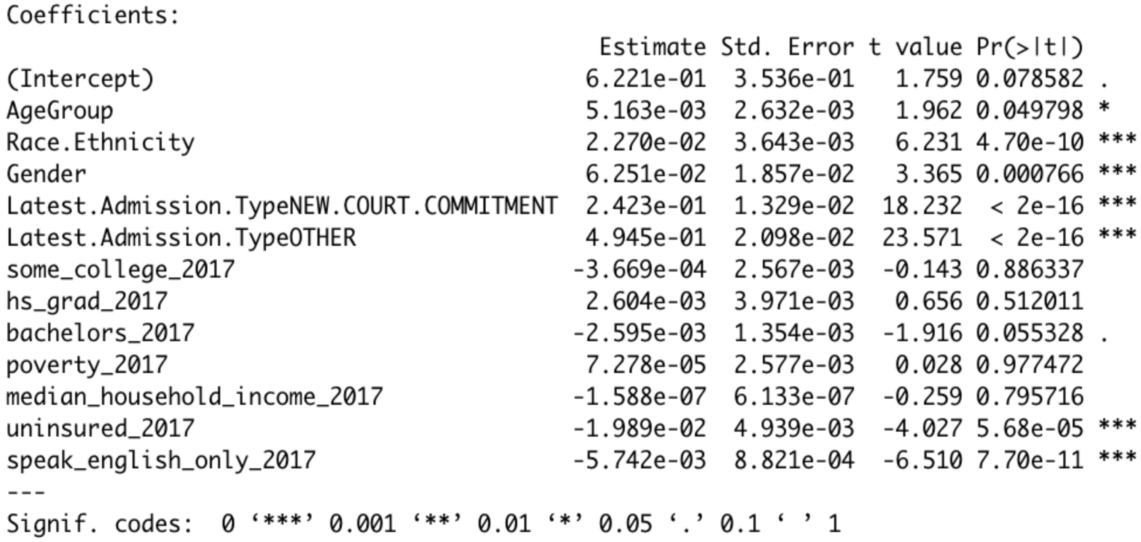
Our Second Statistical Procedure was looking at the county demographics that affected the security level for the top 10 committed crimes in New York in the years 2016, 2017, and 2019. Here is our procedure:

- 2016:



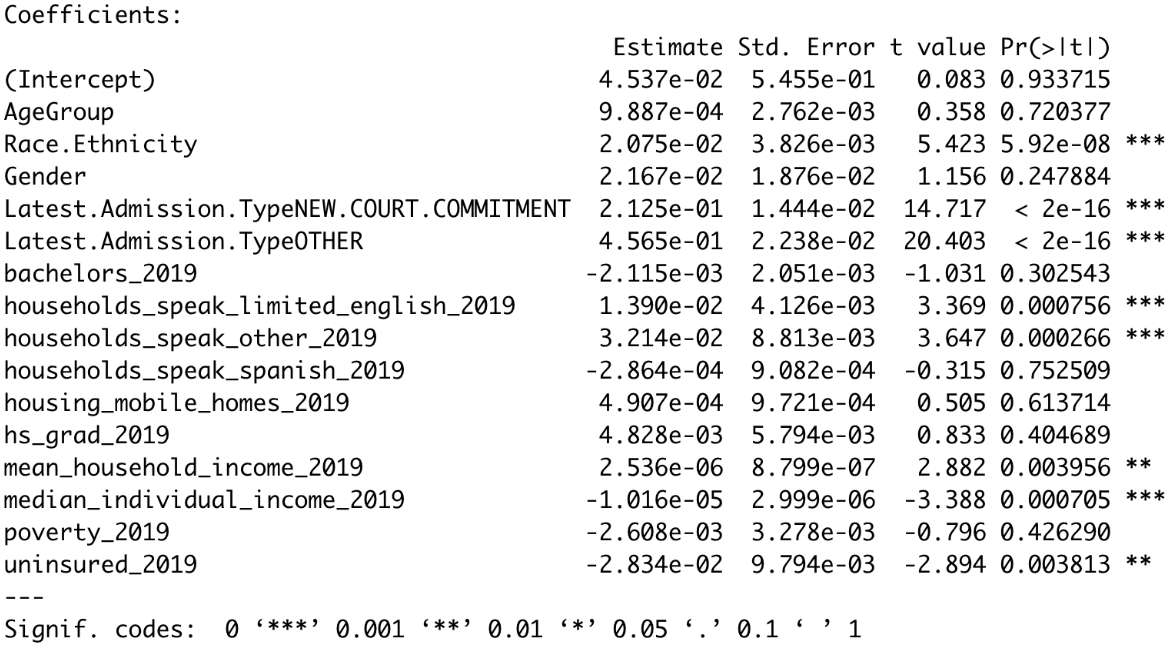
```
Coefficients:
                                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                               3.137e-01  1.498e-01   2.094 0.036258 *  
AgeGroup                                 -3.080e-03  2.616e-03  -1.178 0.238998    
Race.Ethnicity                            2.130e-02  3.624e-03   5.878 4.21e-09 ***
Gender                                    5.241e-02  1.797e-02   2.916 0.003547 ** 
Latest.Admission.TypeNEW.COURT.COMMITMENT  2.199e-01  1.316e-02  16.703  < 2e-16 ***
Latest.Admission.TypeOTHER                4.599e-01  2.121e-02  21.682  < 2e-16 ***
some_college_2016                         1.198e-02  2.552e-03   4.695 2.68e-06 ***
hs_grad_2016                             -6.742e-03  2.143e-03  -3.146 0.001656 ** 
bachelors_2016                            4.082e-03  1.139e-03   3.585 0.000338 ***
poverty_2016                              2.989e-04  1.139e-03   0.262 0.793099    
unemployed_2016                           2.607e-06  3.129e-07   8.333  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 2017:



```
Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)                          6.221e-01  3.536e-01   1.759 0.078582 .  
AgeGroup                             5.163e-03  2.632e-03   1.962 0.049798 *  
Race.Ethnicity                       2.270e-02  3.643e-03   6.231 4.70e-10 ***
Gender                               6.251e-02  1.857e-02   3.365 0.000766 ***
Latest.Admission.TypeNEW.COURT.COMMITMENT  2.423e-01  1.329e-02  18.232  < 2e-16 ***
Latest.Admission.TypeOTHER           4.945e-01  2.098e-02  23.571  < 2e-16 ***
some_college_2017                   -3.669e-04  2.567e-03  -0.143 0.886337    
hs_grad_2017                         2.604e-03  3.971e-03   0.656 0.512011    
bachelors_2017                      -2.595e-03  1.354e-03  -1.916 0.055328 .  
poverty_2017                         7.278e-05  2.577e-03   0.028 0.977472    
median_household_income_2017        -1.588e-07  6.133e-07  -0.259 0.795716    
uninsured_2017                      -1.989e-02  4.939e-03  -4.027 5.68e-05 ***
speak_english_only_2017             -5.742e-03  8.821e-04  -6.510 7.70e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

- 2019:



```
Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)                              4.537e-02  5.455e-01   0.083 0.933715    
AgeGroup                                 9.887e-04  2.762e-03   0.358 0.720377    
Race.Ethnicity                           2.075e-02  3.826e-03   5.423 5.92e-08 ***
Gender                                   2.167e-02  1.876e-02   1.156 0.247884    
Latest.Admission.TypeNEW.COURT.COMMITMENT  2.125e-01  1.444e-02  14.717  < 2e-16 ***
Latest.Admission.TypeOTHER               4.565e-01  2.238e-02  20.403  < 2e-16 ***
bachelors_2019                          -2.115e-03  2.051e-03  -1.031 0.302543    
households_speak_limited_english_2019    1.390e-02  4.126e-03   3.369 0.000756 ***
households_speak_other_2019              3.214e-02  8.813e-03   3.647 0.000266 ***
households_speak_spanish_2019           -2.864e-04  9.082e-04  -0.315 0.752509    
housing_mobile_homes_2019                4.907e-04  9.721e-04   0.505 0.613714    
hs_grad_2019                             4.828e-03  5.794e-03   0.833 0.404689    
mean_household_income_2019               2.536e-06  8.799e-07   2.882 0.003956 ** 
median_individual_income_2019           -1.016e-05  2.999e-06  -3.388 0.000705 ***
poverty_2019                            -2.608e-03  3.278e-03  -0.796 0.426290    
uninsured_2019                          -2.834e-02  9.794e-03  -2.894 0.003813 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Analysis of statistically significant values:

- The county demographics that were statistically significant in security facility levels in 2016 were: some_college_2016, hs_grad_2016, bachelors_2016, and unemployed_2016 at the alpha level of .05.
- The county demographics that were statistically significant in security facility levels in 2017 were: uninsured_2017, and speak_english_only_2017.
- The county demographics that were statistically significant in security facility levels in 2019 were: households_speak_limited_english_2019, households_speak_other_2019, mean_household_income_2019, median_household_income 2019, and uninsured_2019.

Why Logistic Regression for Both Procedures?

We think logistic regression was the best statistical model for both our procedures since we are working on a binary categorical variable, plus it gives us the statistical significance of each predictor and variable, making it easy for readers and us to see what variables could be predictors for the level of security, for each crime like in our first procedure, or each year for our second procedure. Logistic regression also works great while classifying with many factors, which is what we needed. We tried doing some other tests that failed which will be explained below.

Challenges

- Finding useful and relevant predictors has been difficult. We are assuming the predictors we chose for our logistic regression affect the security level the most.
- Finding and using the correct statistical procedures has also been a challenge. We believe using logistic regression is the best procedure, but there could be a better one
- Accounting for changes in trends in the New York Prison system over the years as well as choosing crimes. The large quantity of data would take a lot of time to analyze even if we were working on this full time.
- Preparing the data. Several of the variables had to be one-hot encoded.
- Fitting the data without using family=binomial in our method calls
- Finding demographic data was difficult, and it really restricted us from our analysis.

Dead Ends

- Tried to implement a dendrogram to show hierarchical clustering but ran into a vector memory exhausted error. Getting rid of data would not be ideal as the rest of the inferences were made based on the entire dataset.
- We also tried performing PCA and then clustering but that did not produce desirable results. This was because of the fact that our data was too specific. Along with that, the use of OneHotEncoding made creating plots difficult on the transformed categorical data. We believe we did not have enough variability or the necessary columns in our data to perform clustering and classification.
- We figured out some possible solutions to this problem. Maybe if we were focusing on the population of people inside the prisons(highly variable), and then performed an analysis on that, we could have had better results with clustering.
- If we were able to find more demographic data, it would have been easier to analyze

Brief Summary of Findings and Potential Future Research Questions and Procedures

In our first statistical procedure, we learned that in the top five crimes committed in New York through 2016-2021, your age was a significant factor amongst your inmates who committed the same crime as you in determining security level. Surpsingly, if you were convicted of second-degree murder or burglary, your race did not play many factors into your security level as the other three crimes committed. Gender was only a significant player with 1st-degree robbery, 2nd-degree possession of weaponry, 1st-degree manslaughter, and 2nd-degree burglary. Admission type was statistically significant in all five of these crimes. Finally, when the crime was committed played an important part in most of these crimes. It was interesting to see how committing 1st-degree murder in 2017 and 2018 affected your security level less compared to your other inmates who committed it in 2016, 2019, 2020, and 2021. New admission types and other types of admission were significant for every crime as well. Overall, all security levels for committing the same types of crimes are dependent on the crime itself..

In our second statistical procedure, it was found that in 2016, education and unemployment rates were statistically significant to the security facility levels placed on inmates. This may be because counties with high education and employment rates are less likely to have high crime rates, therefore the areas with less graduation rates and higher unemployment may lead to stricter facility security due to more activity in serious crime in the area. In 2017, it was found that the variables uninsured_2017 and speak_english_only_2017 were statistically significant variables in predicting the facility security level for inmates. In 2019, it was found that households_speak_limited_english_2019, households_speak_other_2019, mean_household_income_2019, median_household_income 2019, and uninsured_2019 were statistically significant demographic variables in the level of facility security. This may be because counties with higher mean and median household incomes and lower uninsured rates are less likely to experience high rates of crime in their county. It was interesting to see that households that spoke limited English or other languages were statistically significant variables in this test.

- Due to a large amount of hot encoding on the original crimes dataset, it would make sense to make a decision tree as many of the variables could be modified to have a binary outcome. (i.e yes or no is a member of a specific race)
    - With that, using Random Forests to find the best fit, attempting to reduce the variance as much as possible.
    - Due to just being introduced to this procedure, we did not feel confident enough to include this in our report.
- In the future, an interesting question to test would be: Does the number of female or male maximum security prison institutions in a county have an effect on the number of females or males sentenced to Maximum Security Facility prisons in New York? For this question, we would still need to add data on how many female/male prison populations are in each county and whether or not this has an effect on the number that are sentenced to each type of facility.
- Another interesting question can be predicting security level, given all the relevant features. This would require a lot more data and features to be even considered a viable application. Another aspect to be considered is the moral viability of this model but it can possibly be the norm of the future.