CRIME TYPE PREDICTION THROUGH DEMOGRAPHIC AND CRIME DATA IN THE CITY OF LOS ANGELES

TEAM KRISPY CRIME DONUTS

ANALYTICS PROBLEM AND RESEARCH QUESTION

- Literature review
- Past theory, methodology, data
- What`s the impact on policy? Why is this topic important?

PAST THEORETICAL AND EMPIRICAL MODELS

- Modeling approach
- Endogenous/exogenous variables to be explained
- How are variables measured
- hypotheses

RESEARCH DESIGN

1. Why
2. Theoretical description, r/ps between variables
3. Data source
4. Variable operationalization

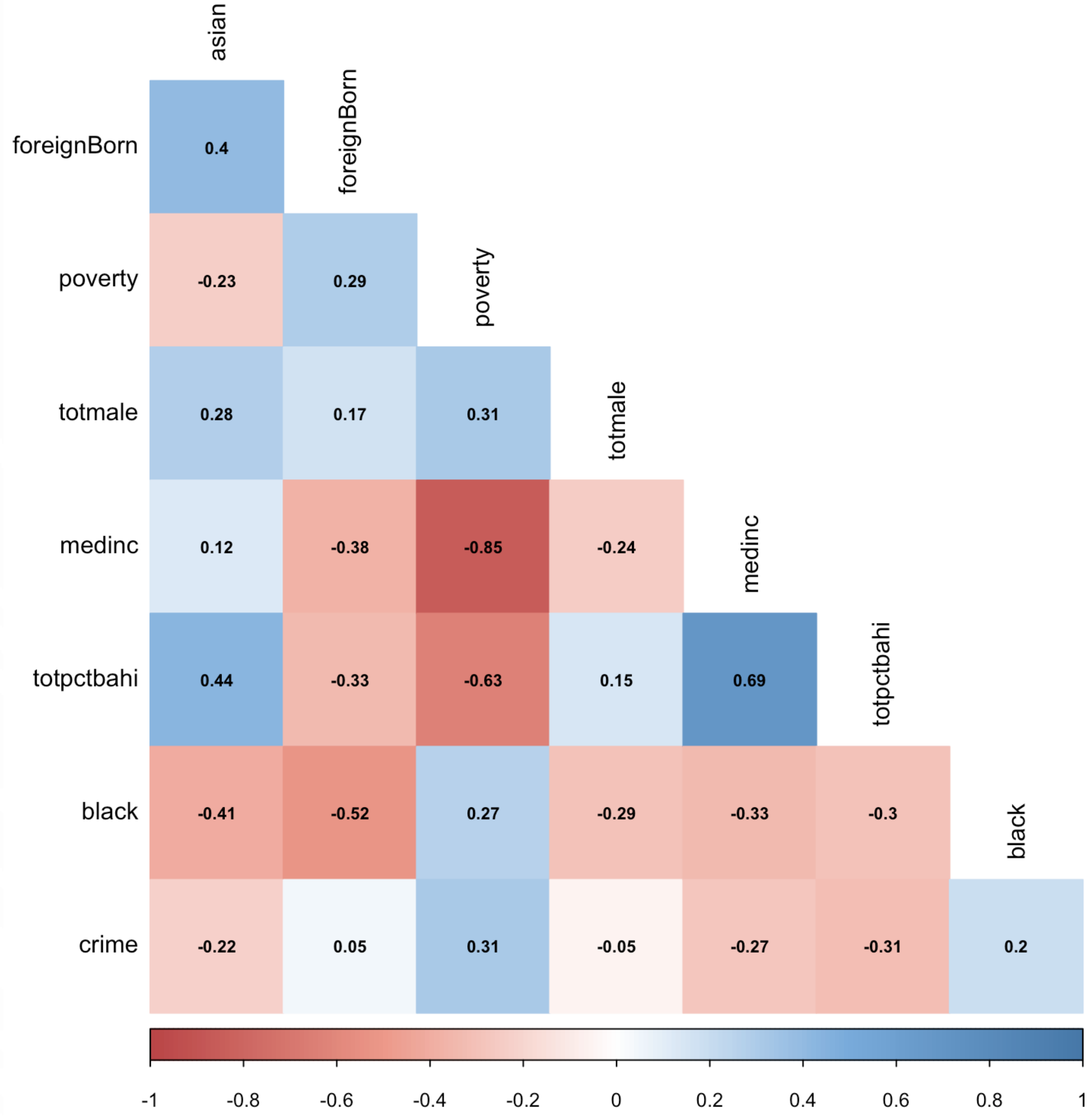
HIGH LEVEL ARCHITECTURE SLIDE- FLOW CHARTS

MEASUREMENT LIMITATIONS, SHORTCOMINGS

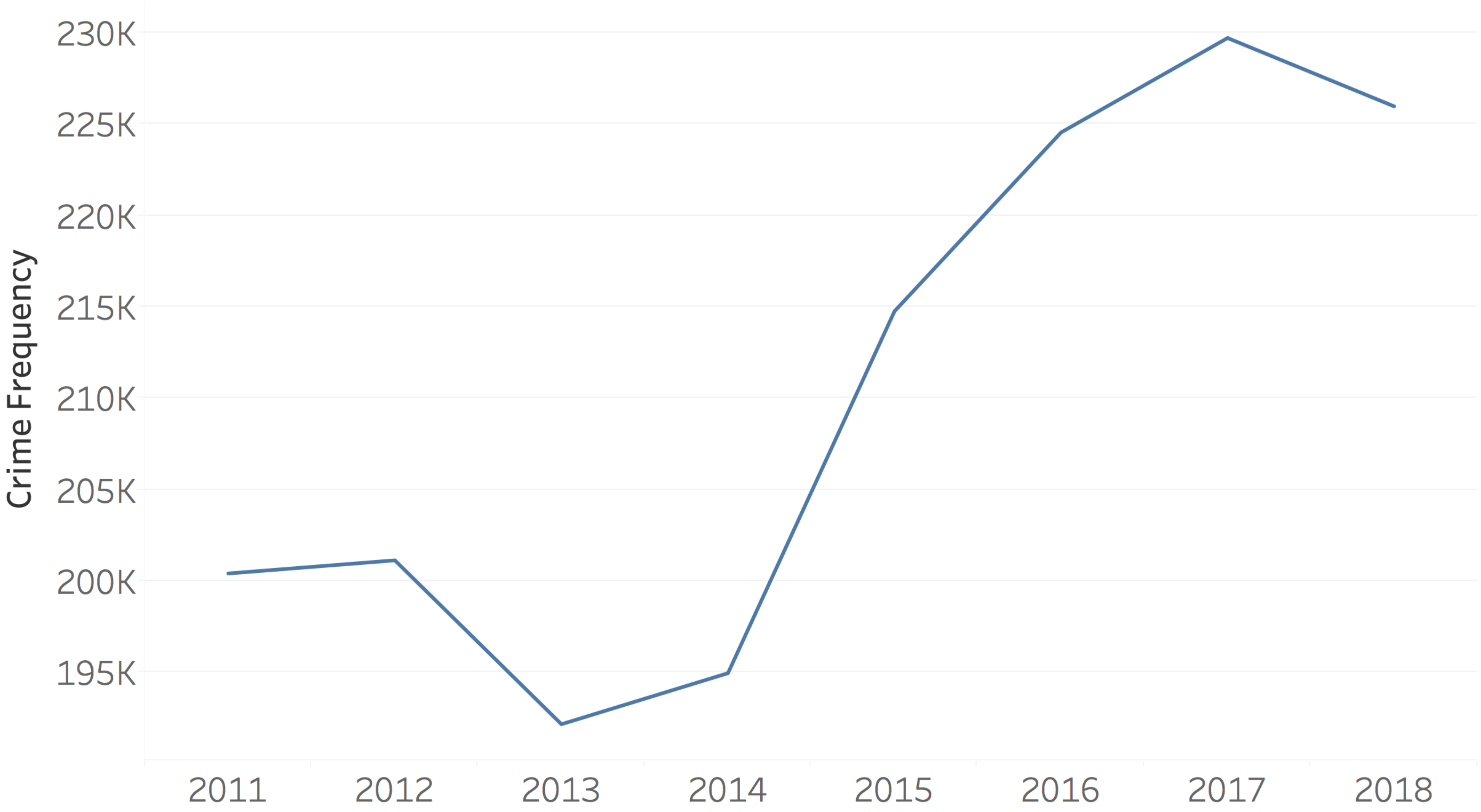
- Problems with data
- Describe what you did to data
- [Mark visual]

FULL MODEL DESCRIPTION: DV AND IV'S

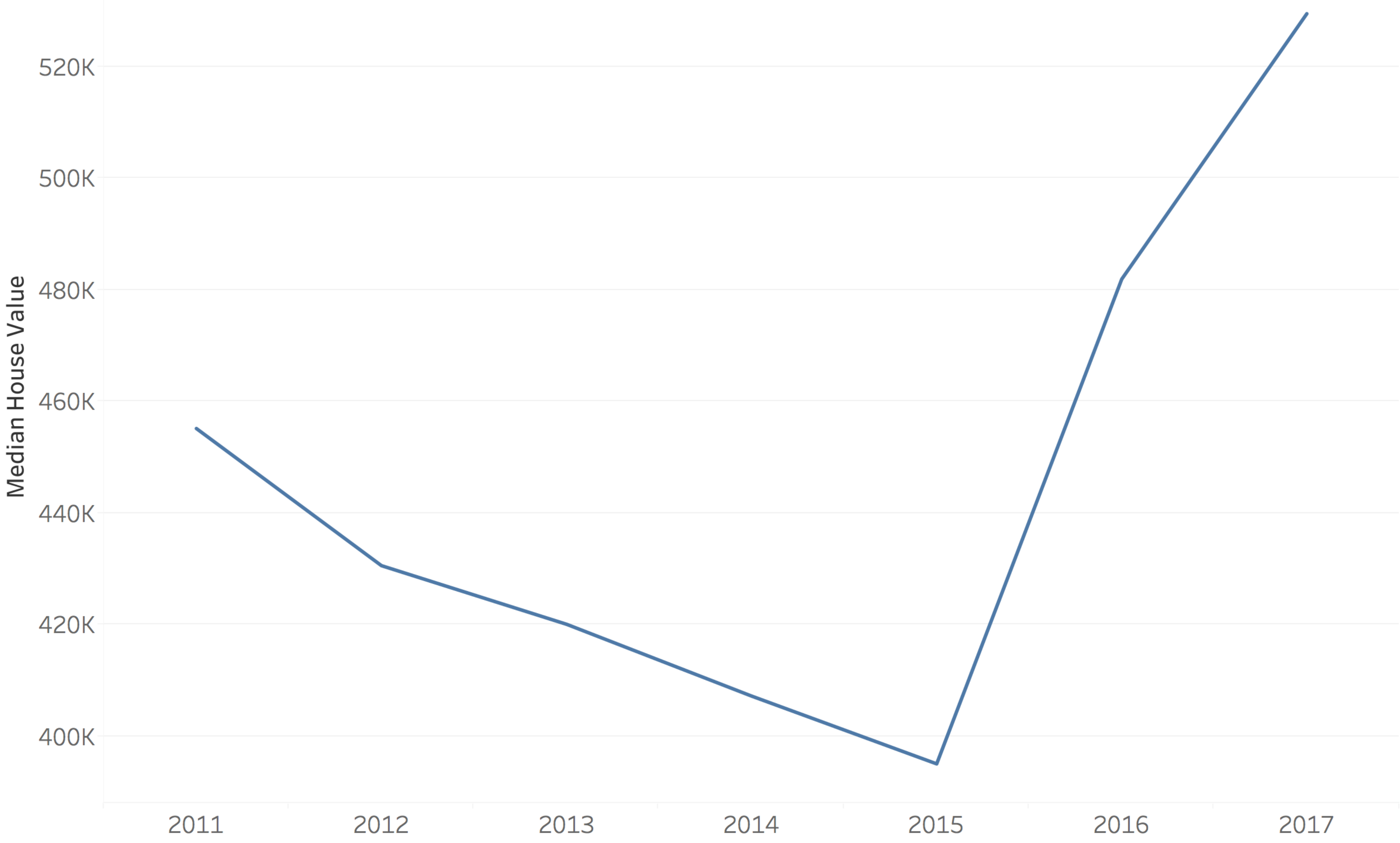
EDA – CORRELATION PLOT- TO BE FIXED



CRIME FREQUENCY OVER TIME

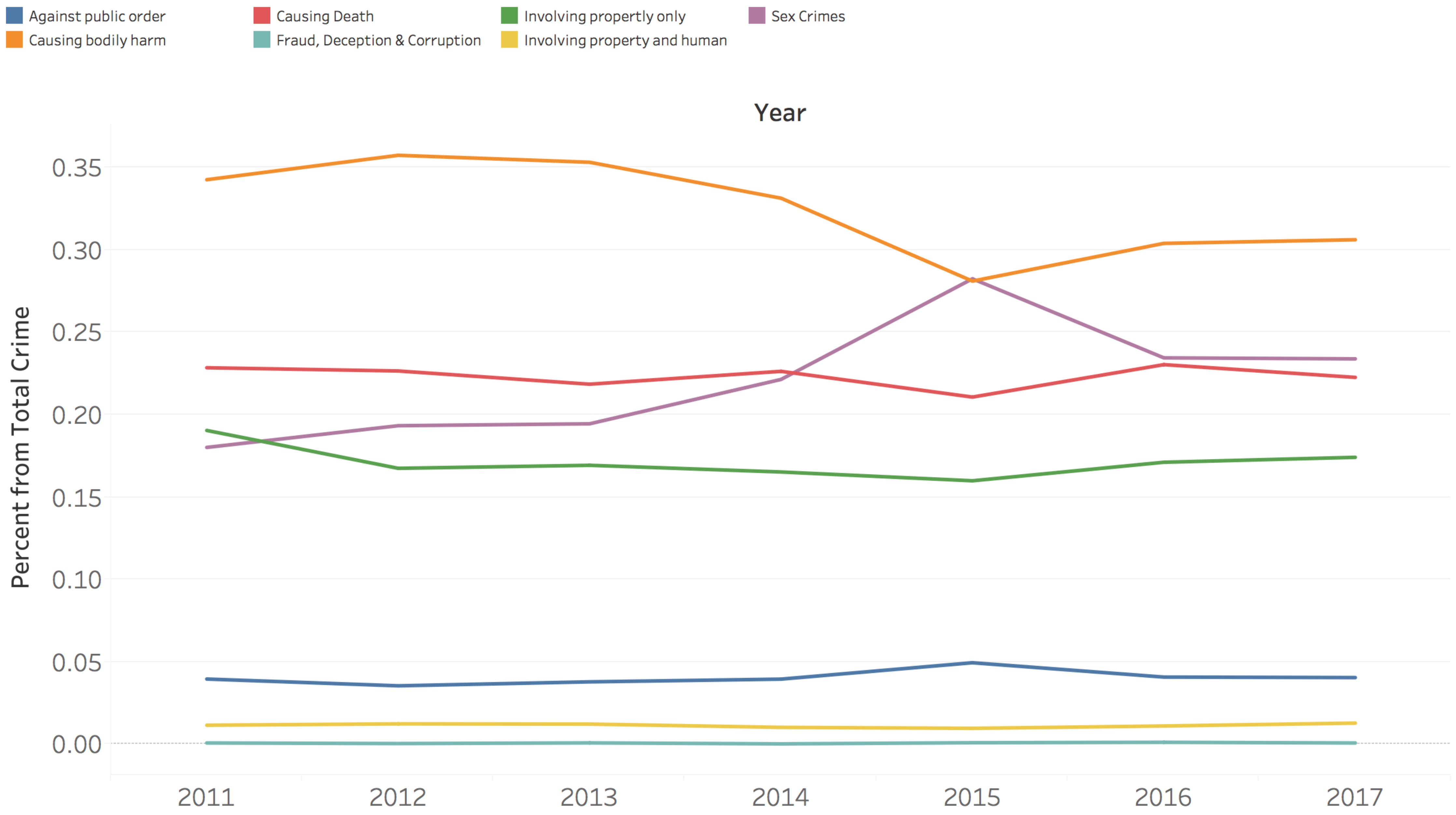


MEDIAN HOUSE VALUE OVER TIME



CRIME FREQUENCY VS MEDIAN HOUSE VALUE

CRIME BY CATEGORY OVER TIME



4-FOLD CROSS VALIDATION

- Accuracy for RF 4, 47.66% Accuracy

MODEL ESTIMATION, BASED ON ACTUAL DATA

Variable	OLS	RF	K-means	Logit
V1				
Vn				
	R^2, Adj R^2, Pseudo …			
Model Evaluation				
Accuracy				
Precision				
Recall				
AUC				

MODEL ESTIMATION, BASED ON PCA DATA

Variable	OLS	RF	K-means	Logit
V1				
Vn				
	R^2, Adj R^2, Pseudo …			
Model Evaluation				
Accuracy				
Precision				
Recall				
AUC				

FINDINGS SUMMARIZED IN 2 TABLES

POLICY IMPLICATIONS

CONCLUSIONS

RESEARCH QUESTION

WHAT IS THE OCCURRENCE OF CRIME BY TYPE BASED ON THE DEMOGRAPHIC AND CRIME DATA?

METHODOLOGY

- Random forest
- Logit
- K-Means
- PCA

Training set 2011-2016

Testing 2017

DEPENDENT VARIABLE

DV- Crime Category

Against public order

Causing bodily harm

Causing death

Fraud, Deception and corruption

Involving property only

Involving property and human

Sex Crimes

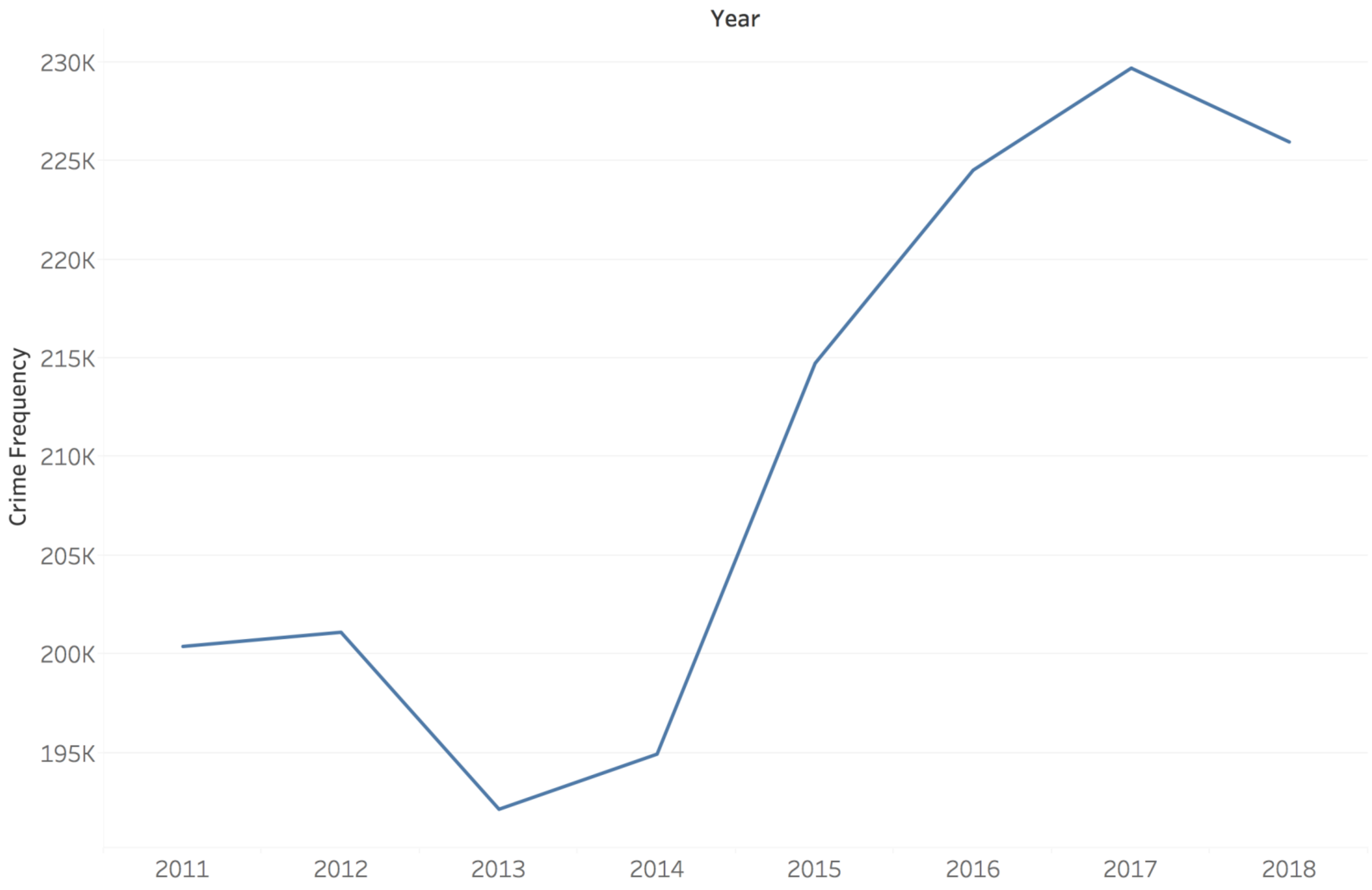
Other

VARIABLES

IV`s

Weapon type	Poverty
Victim age	Ethnicity
Day, month	Median Income
Victim ethnicity	Median Home Value
Area Name	Migration, foreign born
Population	Median Age
Unemployment	Total Population
Ethnicity	Educ, Eng prof

CRIME FREQUENCY OVER TIME



PEARSON CORRELATION MATRIX

	Victim Age	Victim Sex	Weapon Type	Ethnic African American	Highschool	Poverty
Victim Age	1	0.03332009	-0.0289069	0.01804324	0.00363339	-0.0248123
Victim Sex	0.03332009	1	-0.3328478	-0.0692281	-0.0417218	-0.0293806
Weapon.Type	-0.0289069	-0.3328478	1	-0.0451047	-0.013855	-0.024921
Ethnic African American	0.01804324	-0.0692281	-0.0451047	1	0.34112	0.26803514
Highschool	0.00363339	-0.0417218	-0.013855	0.34112	1	0.18425026
Poverty	-0.0248123	-0.0293806	-0.024921	0.26803514	0.18425026	1

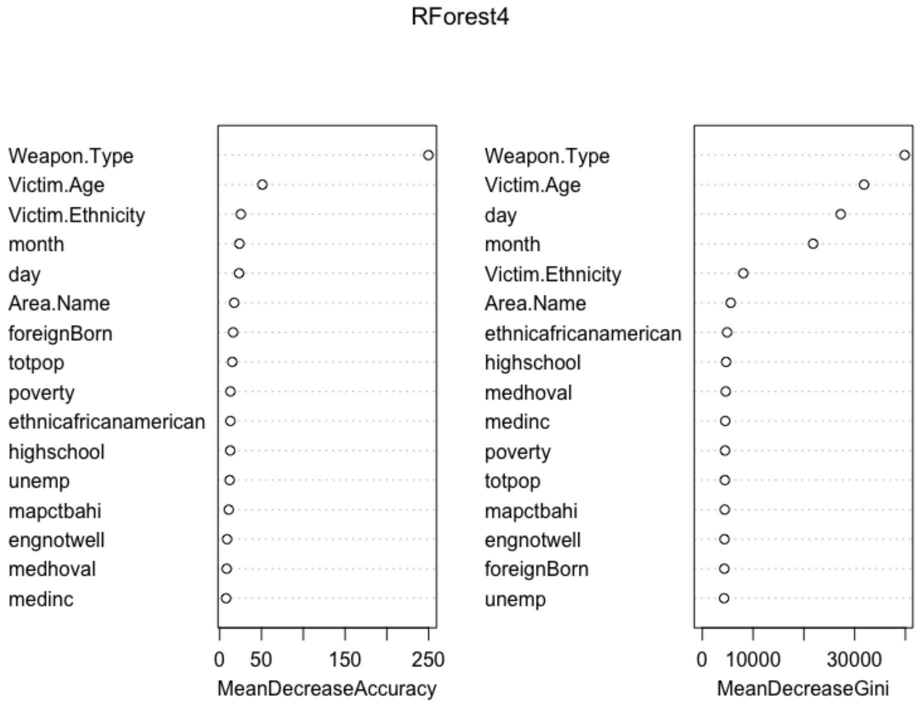
RANDOM FOREST MODEL

Crime Category ~ Area Name, Victim Age, Weapon Type, Month, Day, Victim Ethnicity, Poverty, Med income, Med Home Value, Unemployment, Total Population, Ethnic African American, English not Well, Foreign Born, High School, Male % from Total with Bachelors

Accuracy, training - 78.43%, AUC - 62.23%

Accuracy, testing - 46.15%

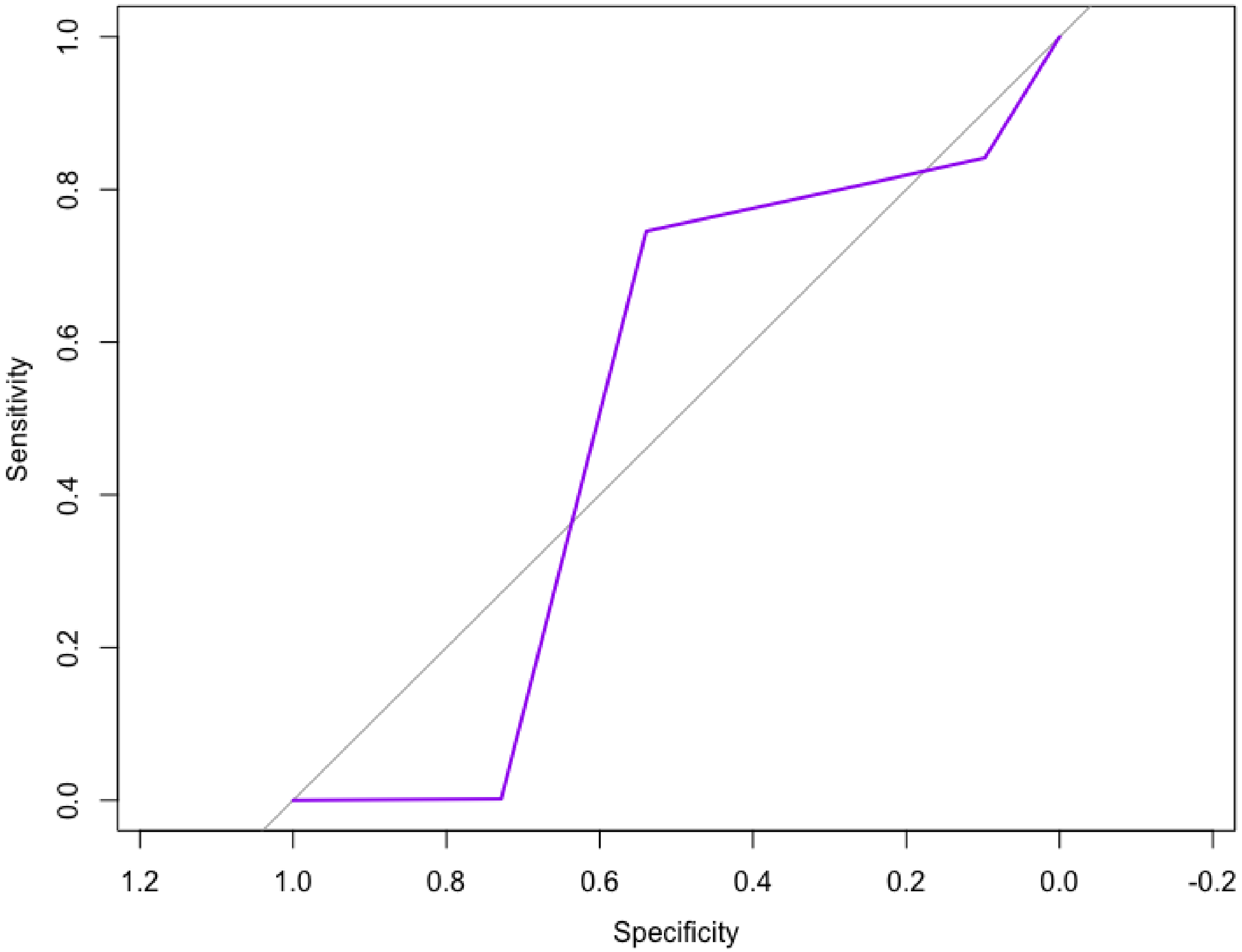
VARIABLE IMPORTANCE PLOT



RANDOM FOREST MODEL, TRAINING DATA

Crime Category	Sensitivity	Specificity
Against Public Order	0.53404	0.99853
Causing bodily harm	0.8500	0.8520
Causing death	0.9001	0.9503
Fraud, Deception, Corruption	0.4748201	0.9999770
Involving property only	0.6660	0.9776
Involving property and human	0.261131	0.999890
Sex crimes	0.7381	0.9260
Other	0.494827	0.999908

TRADE-OFF BETWEEN SENSITIVITY AND SPECIFICITY, RANDOM FOREST ON TRAINING DATA



RANDOM FOREST, TESTING DATA

Crime Category	**Sensitivity**	**Specificity**
Against Public Order	0.027695	0.994153
Causing bodily harm	0.4924	0.7324
Causing death	0.6860	0.8137
Fraud, Deception, Corruption	0.000e+00	1.000e+00
Involving property only	0.25432	0.92628
Involving property and human	0.0000000	0.9998241
Sex crimes	0.4820	0.8157
Other	1.464e-03	9.999e-01

LOGIT MODEL

Crime Category ~ Area Name, Victim Age, Weapon Type, Month, Day, Victim Ethnicity, Poverty, Med income, Med Home Value, Unemployment, Total Population, Ethnic African American, English not Well, Foreign Born, High School, Male % from Total with Bachelors

- Accuracy -0.4662
- AUC- 0.5201

LOGIT MODEL, TRAINING DATA

Accuracy – 46.62%

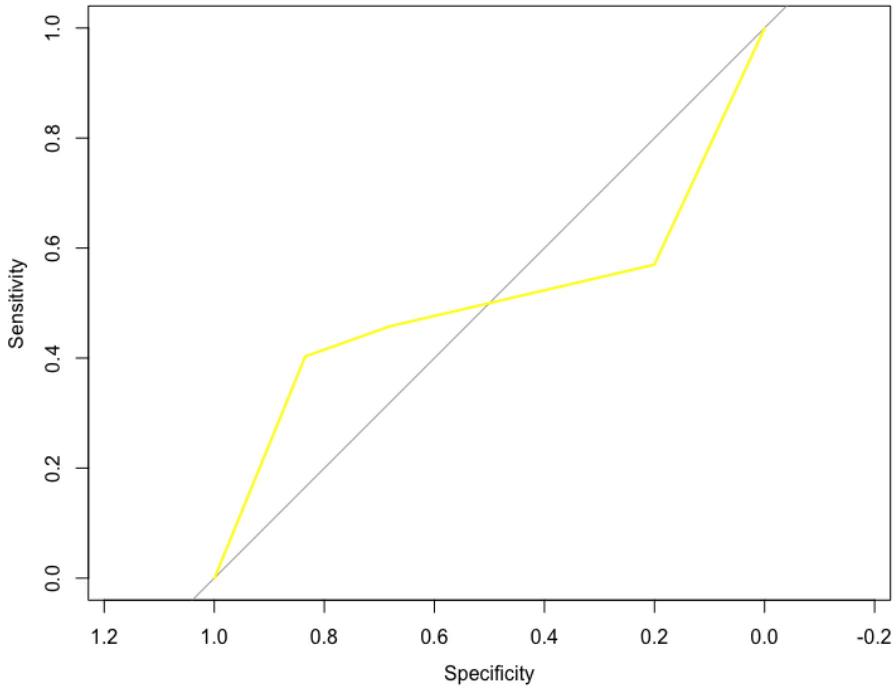
Crime Category	**Sensitivity**	**Specificity**
Against Public Order	7.274e-05	1.000e+00
Causing bodily harm	0.4299	**0.7768**
Causing death	**0.7217**	**0.8405**
Fraud, Deception, Corruption	0.0000000	1.0000000
Involving property only	0.25851	**0.93105**
Involving property and human	0.0000000	1.00000
Sex crimes	0.5491	**0.7411**
Other	0.000e+00	1.000e+00

LOGIT MODEL, TESTING DATA

Accuracy - 0.4351

Crime Category	**Sensitivity**	**Specificity**
Against Public Order	0.00000	1.000e+00
Causing bodily harm	0.3251	**0.8072**
Causing death	**0.7038**	**0.8121**
Fraud, Deception, Corruption	0.0000000	1.0000000
Involving property only	0.26751	**0.91821**
Involving property and human	0.0000000	1.00000
Sex crimes	0.5661	**0.7163**
Other	0.000e+00	0.9998787

TRADE-OFF BETWEEN SENSITIVITY AND SPECIFICITY, LOGIT TRAINING



LOGIT MODEL, TRAINING DATA

Accuracy – 46.62%

Crime Category	**Sensitivity**	**Specificity**
Against Public Order	7.274e-05	1.000e+00
Causing bodily harm	0.4299	0.7768
Causing death	0.7217	0.8405
Fraud, Deception, Corruption	0.0000000	1.0000000
Involving property only	0.25851	0.93105
Involving property and human	0.0000000	1.00000
Sex crimes	0.5491	0.7411
Other	0.000e+00	1.000e+00