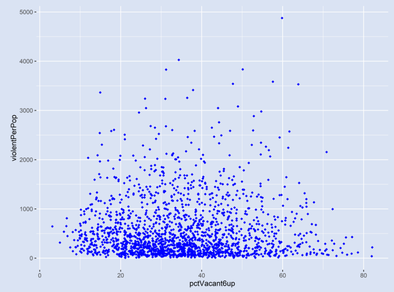
What Determines Crime Rates in the USA?

1710770, 1826598, 1831807, 1810820

Exploratory Data Analysis

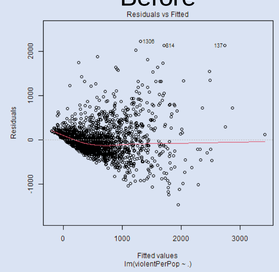
- Deleted Missing Values
- Merged Pacific Region into West
- Converted pctUrban into a factor variable with two categories
- Removed pctVacant6up and State from our models



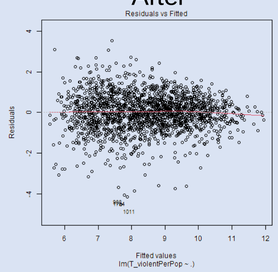
Transformations

- Transformed variables using BoxCox & Trial and Error
- This ensured modelling assumptions were fulfilled

Before



After



Influence Points

- Identified observations with high influence, using outlier and leverage plots
- These can affect the models - we removed the observations below

Group	Influence Points	Reason for Removal
Alaska	816, 1128, 1758	Not Representative
Zero values for ownHousQrange	40, 493, 520, 565, 1737	Invalid
High influence points	322, 1329	Very High Influence

Multicollinearity

- Identified several groups of highly correlated predictor variables

Least Significant	(1)	*pctKids2Par pctKidsBornNevrMarr*
	(2)	*pctCollGrad pctKids2Par pctKidsBornNevrMarr*
	(3)	*ownHousMed rentMed*
	(4)	*pctWdiv pctCollGrad popDensity*
	(5)	*pctLowEdu pctNotHSgrad popDensity*
	(6)	*medIncome pctEmploy pctHousOwnerOccup*
	(7)	*pctLowEdu pctNotHSgrad pctCollGrad pctWdiv*
Most Significant	(8)	*ownHousMed ownHousQrange*

Variable Selection

- Variable selection methods were compared by their predictive and explanatory power

	violentPerPop			nonViolPerPop		
Method	**p**	**R**	**RMSE**	**p**	**R**	**RMSE**
OLS	23	0.6624136	0.9003539	23	0.5551260	0.5470958
AIC	15	0.6648898	0.9014250	14	0.5587500	0.5481817
BIC	12	0.6642046	0.9039431	12	0.5539244	0.5518254
Ridge Regression	23	0.6677315	0.9141737	23	0.5599642	0.5578307
LASSO	12	0.6501593	0.9313836	9	0.5311418	0.5692164

Violent Model

$$\log_2(\text{violentPerPop}) = 8.9 - (2.0\times10^{-1})\text{NorthEast} + (4.6\times10^{-1})\text{West} + (3.7\times10^{-1})\text{South} + (3.8\times10^{-1})\text{pctUrban}(50,100] + (6.6\times10^{-3})\text{medIncome}^{0.5} - (1.8\times10^{-2})\text{pctWdiv} - (4.6\times10^{-6})\text{pctKids2Par}^{3} + (3.4\times10^{-1})\text{pctKidsBornNevrMarr}^{0.5} - (8.0\times10^{-21})\text{pctHousOccup}^{10} + (8.9\times10^{-2})\text{pctVacantBoarded}^{0.5} + (1.2\times10^{-1})\text{pctForeignBorn}^{0.5}$$

Non-Violent Model

$$\log_2(\text{nonViolPerPop}) = 13.1 - (9.5\times10^{-2})\text{MidWest} - (4.3\times10^{-1})\text{NorthEast} + (8.8\times10^{-2})\text{West} + (4.9\times10^{-2})\text{South} + (1.6\times10^{-1})\text{pctUrban}(50,100] - (2.5\times10^{-6})\text{pctKids2Par}^{3} - (1.2\times10^{-21})\text{pctHousOccup}^{10} - (3.2\times10^{-6})\text{pctHousOwnerOccup}^{2}$$

Below are the determinants with the most significant effect – However multicollinearity must be taken into account

Selected Variables	Violent Crime - Model Interpretation
NorthEast	If the region is NorthEast, violentPerPop decreases by ≈ 13% compared to Midwest
West	If the region is West, violentPerPop increases by ≈ 38% compared to Midwest
South	If region is South, violentPerPop increases by ≈ 29% compared to Midwest
pctUrban(50,100]	If pctUrban is between (50,100], violentPerPop increases by ≈ 30%
pctKids2Par	For every 1% increase in pctKids2Par, violentPerPop decreases by ≈ 3.5%
pctHousOccup	For every 1% increase in pctHousOccup, violentPerPop decreases by ≈ 3.1%

Selected Variables	Non-violent Crime - Model Interpretation
MidWest	If the region is MidWest, nonViolPerPop decreases by ≈ 6.4%
NorthEast	If the region is NorthEast, nonViolPerPop decreases by ≈ 26%
West	If the region is West, nonViolPerPop increases by ≈ 6.3%
South	If region is South, nonViolPerPop increases by ≈ 3.5%
pctUrban(50,100]	If pctUrban is between (50,100], nonViolPerPop increases by ≈ 12%
pctKids2Par	For every 1% increase in pctKids2Par, nonViolPerPop decreases by ≈ 2%

Are the causes of Violent and Non-Violent crime the same?

- The causes of violent and non-violent crime are very similar
- The key indicators are region, pctUrban, pctKids2Par and pctHousOccup. This is to be expected, as violentPerPop and nonViolPerPop are highly correlated
- There are however some important differences:
  - pctUrban is a stronger indicator of violentPerPop than it is of nonViolPerPop
  - Variables pctWdiv, pctVacantBoarded and pctForeignBorn are indicators of violentPerPop but none of them, or variables that they are colinear to, are indicators of nonViolPerPop

Are there areas of the US that do not conform to the general pattern?

- Florida had unusually high crime. It accounted for 1/5 of the high violent crime and almost 2/5 of the high non-violent crime representing 11% and 20% of its data points respectively.
- Massachusetts had unusually low crime. It accounted for roughly 1/5 of the violent and non-violent crime representing 7% and 10% of its data points respectively
- States RI, OH, IN and ME each had an observation that was an outlier with lower violentPerPop than predicted whilst CA had an observation with unusually high Violent and Non-Violent crime. These observations are spread across different states so there don`t appear to be any significantly not conformant areas.