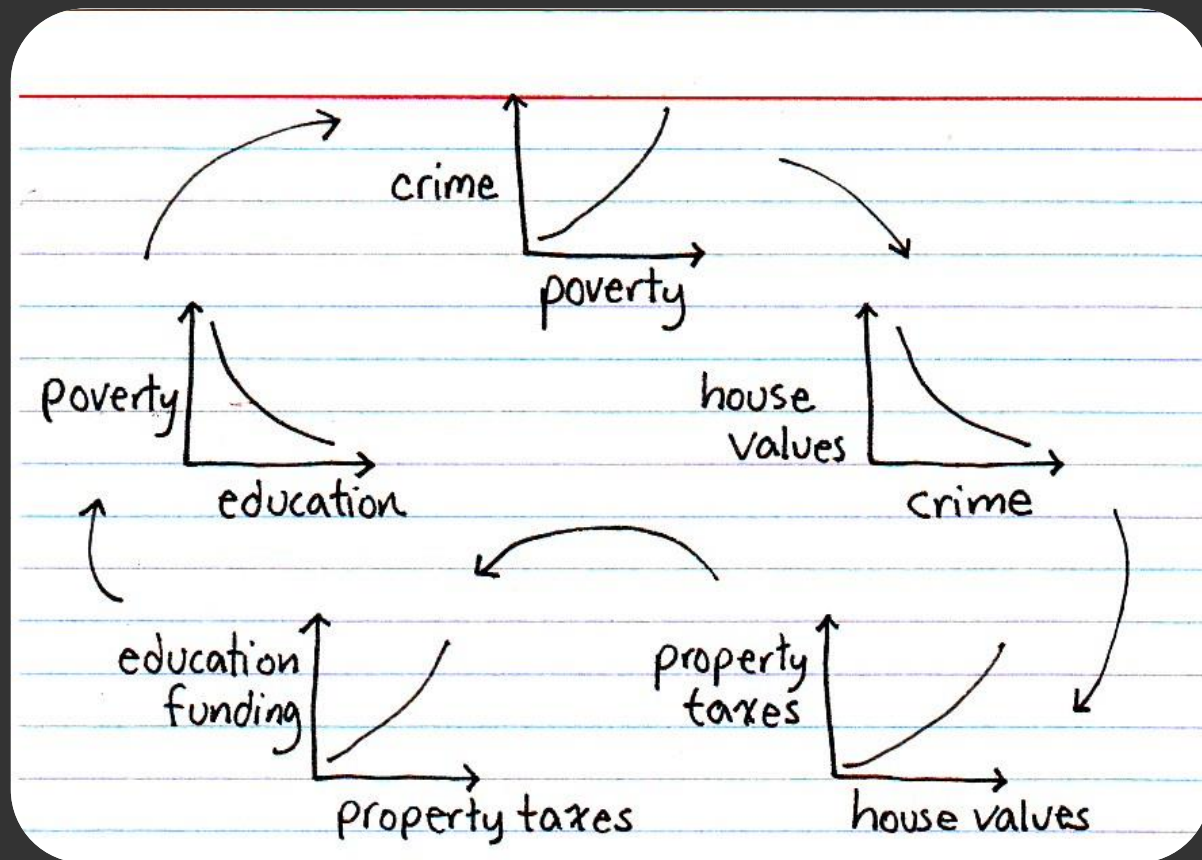




A black and white photograph featuring a horizontal strip of white police tape with the words "POLICE LINE DO NOT CROSS" printed in bold, black, sans-serif capital letters. The tape is positioned in the lower half of the frame. In the background, several bright, out-of-focus circular lights, likely from a police vehicle's emergency lights, are visible against a dark, blurred backdrop. The overall composition is centered and uses a shallow depth of field to emphasize the tape.

**POLICE LINE DO NOT CROSS**





# Question

What are the main factors correlated to violent crimes?

# contents

01	Data
02	Method
03	Results
04	Conclusion

>> 01 Used Data

# Communities and Crime Data

---

1827 Communities in the USA

Response Variable:

Violent crimes per 100k (1995 FBI UCR)

Explanatory variables:

122 socio-economic factors

(1990 US Census, 1990 US LEMAS )

---

# Violent Crimes

---

FBI says

1. Murder and non-negligent manslaughter
  2. Forcible rape
  3. Robbery
  4. Aggravated assault
-



>> 02 Method

Heatmap

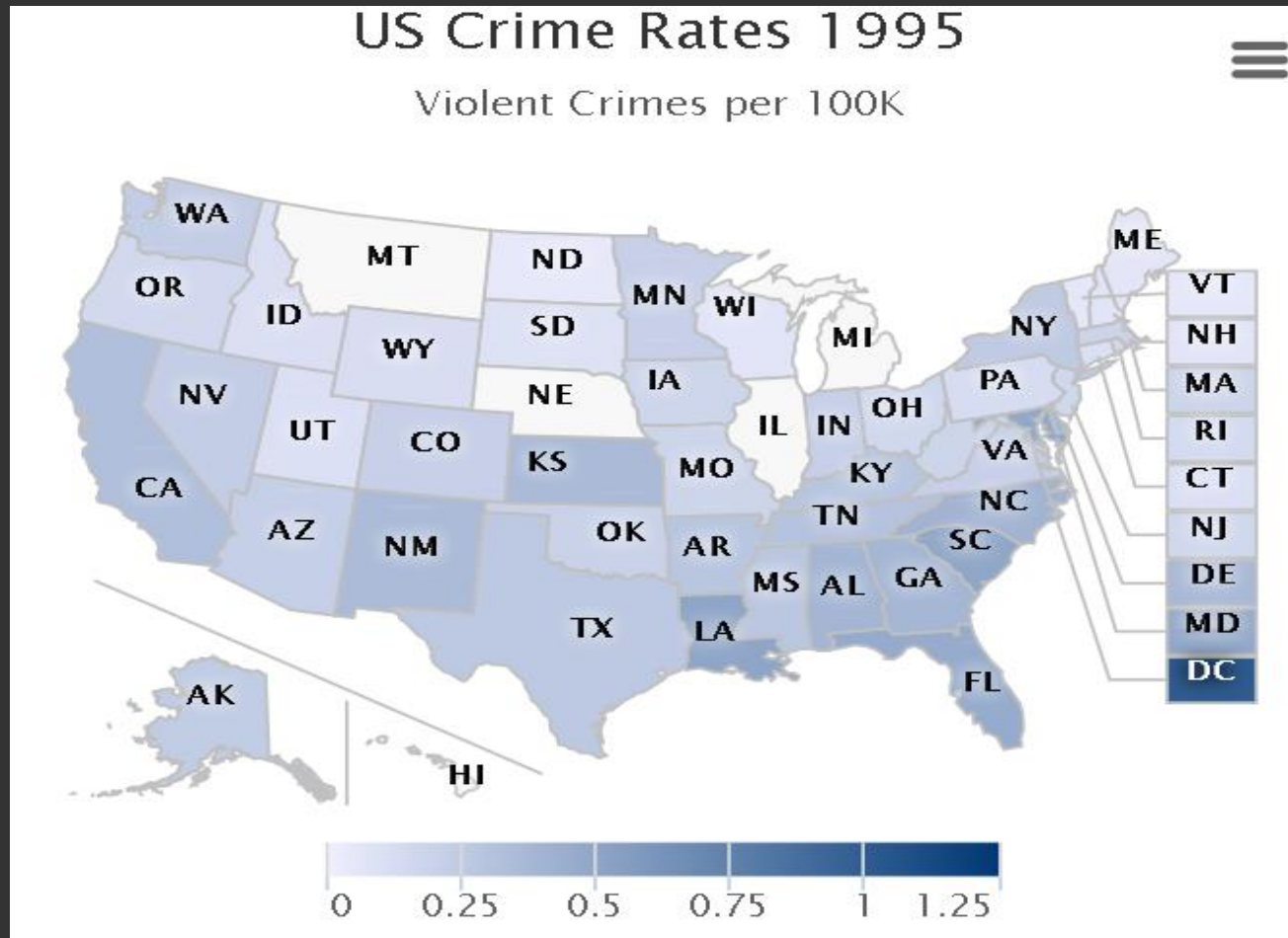
Ordinary  
Linear  
Regression

Logistic  
Regression



Analysis Process

## Graph1 – Violent Crimes By state on US Heatmap



<https://jsfiddle.net/6t6ubw1o/embedded/result/>

Heatmap

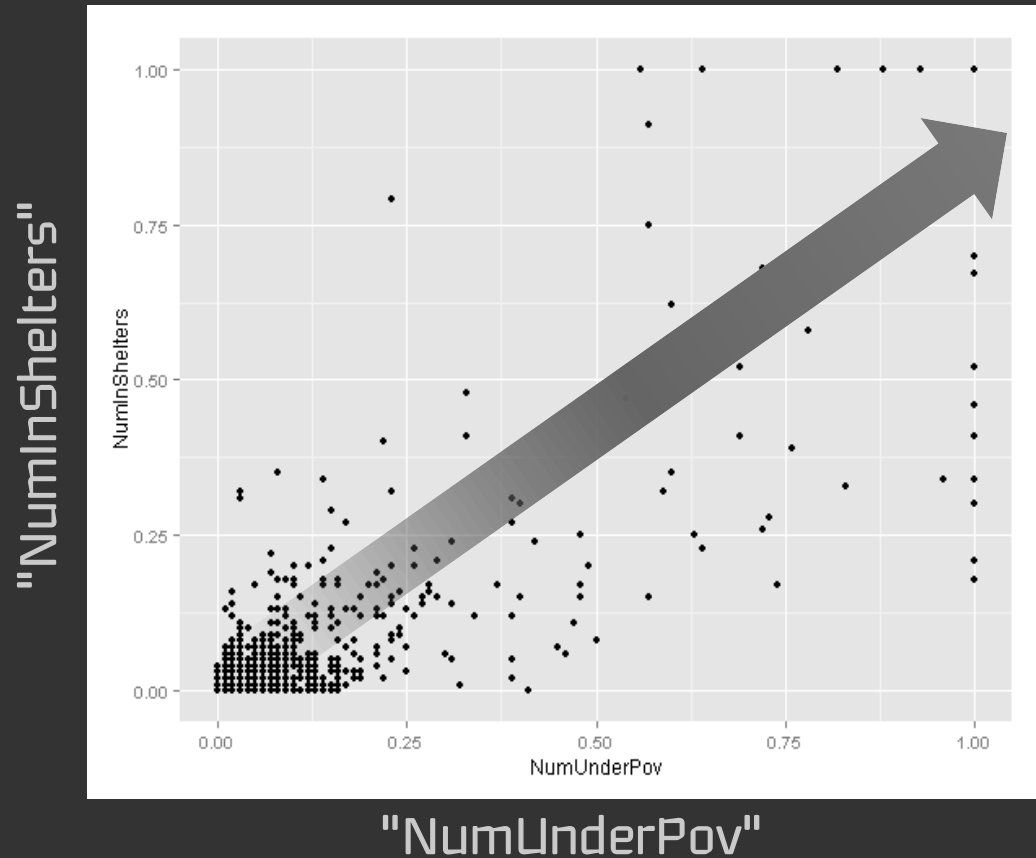
Ordinary  
Linear  
Regression

Logistic  
Regression



Analysis Process

## Correlation between two explanatory variables



Ex) Multicollinearity between  
"NumUnderPov"  
and  
"NumInShelters."

Heatmap

Ordinary  
Linear  
Regression

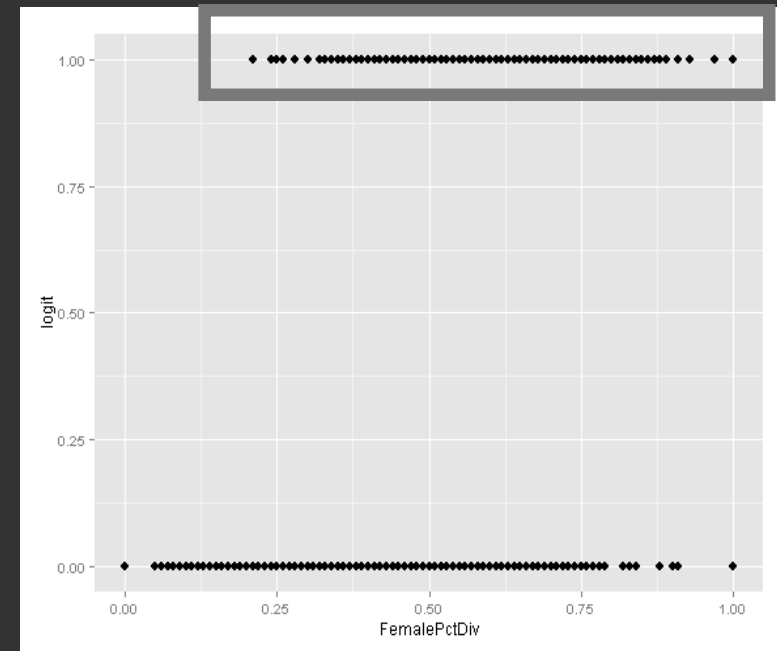
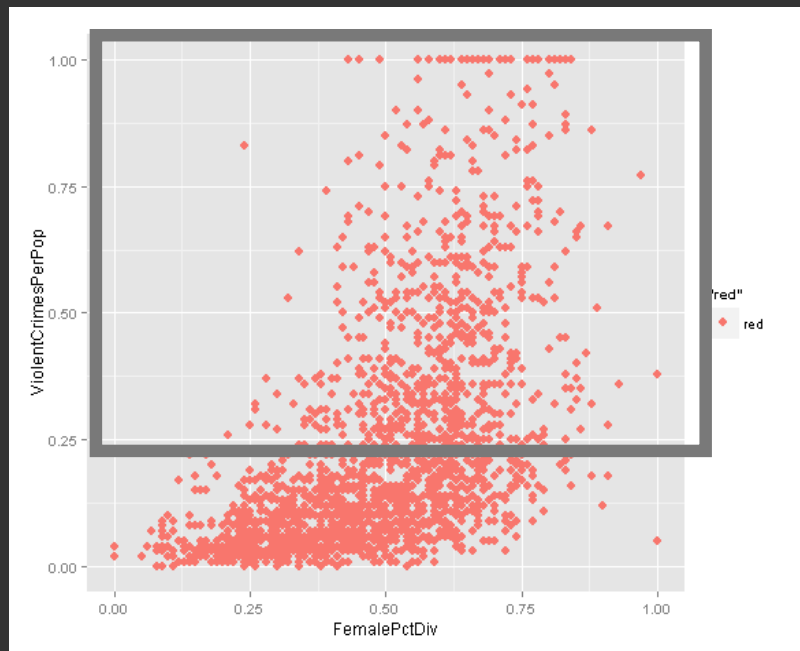
Logistic  
Regression



Analysis Process

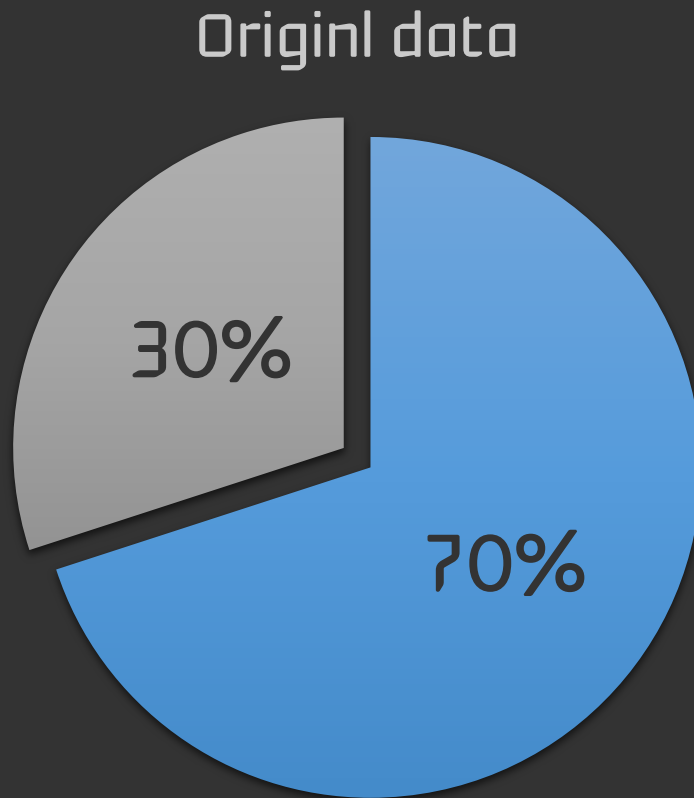


## Adding new response variable



`ViolentCrimesPerPop > 0.24` becomes "1" in logit variavle

## Training data // Test data



Training data : 70%  
\_to build formula

Test data : 30%  
\_to test formula

>> 03 Result

## Test data result

threshold	AUC	sensitivity	specificity	prop.correct
0.5	0.81514	0.7207207	0.9095745	0.8394649

The overall model's prediction power is about 83.9%

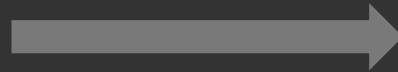
- Increase crime
- Decrease crime

Coefficients:	Estimate	Std.	Error	z-value
(Intercept)	7.8958	1.168	6.76	1.38E-11
NumUnderPov	10.1802	2.0442	4.98	6.36E-07
PctForeignBorn	3.6062	0.8416	4.285	1.83E-05
PctIlleg	4.2088	0.878	4.794	1.64E-06
pctWFarmSelf	1.5292	0.4924	3.105	0.0019
pctWInvInc	-6.2308	0.9681	-6.436	1.23E-10
PctKids2Par	-6.1354	1.2026	-5.102	3.36E-07
PctNotSpeakEnglWell	-3.0342	0.9082	-3.341	0.000835
PctPopUnderPov	-3.7541	0.8096	-4.637	3.53E-06
PctWorkMom	-2.2919	0.6168	-3.716	0.000202
PctHousOccup	-2.0396	0.4617	-4.417	9.99E-06
MedOwnCostPctIncNoMtg	-1.5791	0.4995	-3.161	0.001572

>> 04 Conclusion  
& furthermore

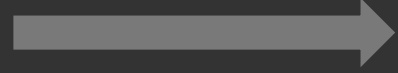


Economic status is  
Prominent factor



Defeat  
Poverty

Kids born to  
Single-parent family  
& never married



Caring  
Kids

Crime data  
in other countries and years

Compare common factors  
between different countries

Mash-up with other data sets that  
have different explanatory variables

QnA