

# What leads to Crime in US

Presented by Kimia Nafar

# The Idea

- Are there any common factors in crime?
- Does age, median income, or population correlate to crime rates in a state or a county?
- Does the type of crime matter?

- Find datasets on:
  - Population in US (per state and county)
  - Mean and Median household income
  - Age
  - Financial standing of each location
  - Crime rates and types in each area
- Look for anything interesting

# The Plan

# Getting Data

- **Crime data:**  
<https://crime-data-explorer.fr.cloud.gov>
- **City and area income and population:**  
<https://www.kaggle.com>
- **State income, and poverty rates:**  
<https://www.deptofnumbers.com>
- **Age distribution in states:**  
<https://worldpopulationreview.com>

# The Data

- County vs State data
  - More data points
  - Running models
- Data per city or zip code area
  - Grouping the data to get county and state information
- State Income, Age, Poverty rates
  - Some data not available on a smaller scale for counties
  - New datasets, cross checked with county data to ensure accuracy

# Cleaning The Data

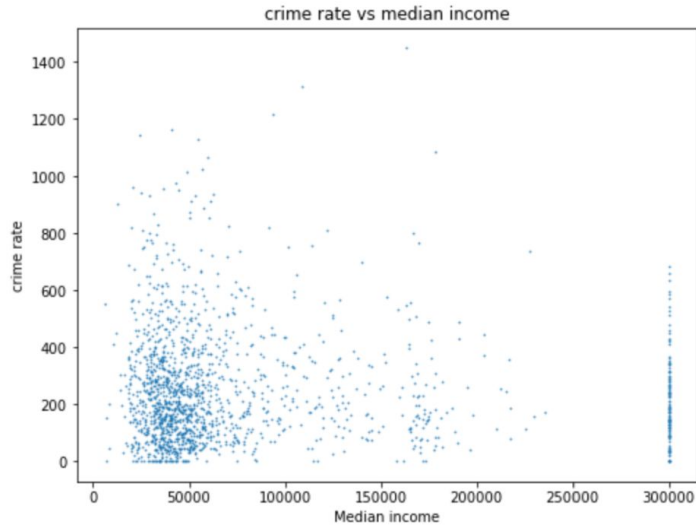
Noticed unusual zeros in the smaller scale census

```
income[income.Mean == 0]
```

	State_Code	State_Name	State_ab	County	Mean	Median	Stdev	sum_w	county_name
id									
1011500	1	Alabama	AL	St. Clair	0	0	0	0.0	St. Clair, AL
1021496	1	Alabama	AL	Autauga	0	0	0	0.0	Autauga, AL
1023736	1	Alabama	AL	Autauga	0	0	0	0.0	Autauga, AL
102806	1	Alabama	AL	Autauga	0	0	0	0.0	Autauga, AL
201540	2	Alaska	AK	Ketchikan Gateway Borough	0	0	0	0.0	Ketchikan Gateway Borough, AK
...	...	...	...	...	...	...	...	...	...
7202276	72	Puerto Rico	PR	Adjuntas Municipio	0	0	0	0.0	Adjuntas Municipio, PR
72023176	72	Puerto Rico	PR	Adjuntas Municipio	0	0	0	0.0	Adjuntas Municipio, PR
7202656	72	Puerto Rico	PR	Adjuntas Municipio	0	0	0	0.0	Adjuntas Municipio, PR
7202916	72	Puerto Rico	PR	Adjuntas Municipio	0	0	0	0.0	Adjuntas Municipio, PR
7202986	72	Puerto Rico	PR	Adjuntas Municipio	0	0	0	0.0	Adjuntas Municipio, PR

315 rows × 9 columns

# Cleaning The Data



When plotting the data noticed unrealistic median income for counties

# Inaccuracies in data

```
county_income[county_income['Median'] > 150000]  
county_income.to_csv()
```

	Mean	Median	Stdev	sum_w
county_name				
Aleutians West Census Area, AK	67102.000000	300000.0	41868.0	11.799612
Amador, CA	52845.666667	216879.0	33820.0	1806.751295
Austin, TX	120476.000000	300000.0	64727.0	78.759689
Banks, GA	38679.000000	166327.5	22154.5	133.763538
Barnes, ND	63411.000000	300000.0	38685.0	7.999840
...	...	...	...	...
Winona, MN	69253.500000	180100.0	49228.5	337.314335
Worcester, MD	74971.000000	300000.0	46013.0	59.998000
Wyandot, OH	50341.000000	300000.0	15855.0	36.331755
Yakima, WA	57904.000000	176645.0	39109.0	699.529935
Yazoo, MS	28333.000000	300000.0	3696.0	6.000000

235 rows × 4 columns

## Real Median Household Income for Austin Texas

Show dollars as: [Nominal](#) [Real](#)

	2019	1 Year Change	3 Year Change
US	\$65,712	+4.21%	+7.05%
Texas	\$64,034	+3.74%	+6.26%
Austin	\$80,954	+3.36%	+7.02%



# Final columns

## County

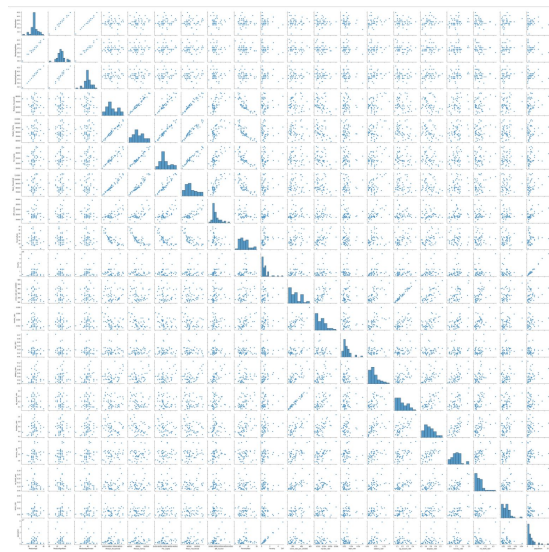
```
['county_name',  
 'Mean',  
 'Median',  
 'Stdev',  
 'crime_rate_per_100000',  
 'population',  
 'murder_rate',  
 'rape_rate',  
 'robbery_rate',  
 'ag_assault_rate',  
 'burglary_rate',  
 'larceny_rate',  
 'mv_theft_rate',  
 'arson_rate',  
 'State_ab',  
 'diff_income']
```

## State

```
['State',  
 'MedianAge',  
 'MedianAgeMale',  
 'MedianAgeFemale',  
 'Median_Household',  
 'Median_Family',  
 'Per_Capita',  
 'Mean_Household',  
 'diff_income',  
 'PovertyRate',  
 'Poverty',  
 'State_ab',  
 'crime_rate_per_100000',  
 'murder_rate',  
 'rape_rate',  
 'robbery_rate',  
 'ag_assault_rate',  
 'burglary_rate',  
 'larceny_rate',  
 'mv_theft_rate',  
 'arson_rate',  
 'population']
```

# Finding Something Interesting

- Correlation graphs
- Pairplot
  - Did not go well



# Heatmap - county data

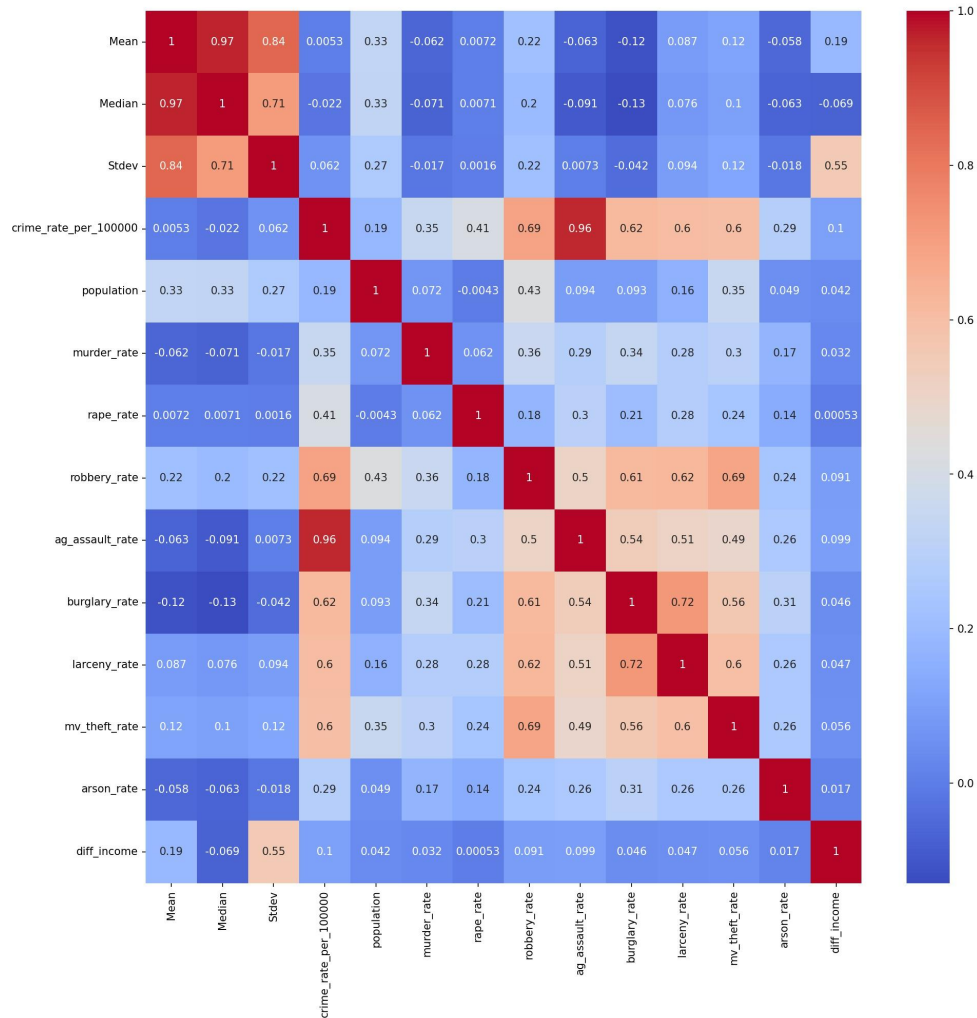
Much less correlation than expected

The ones that made more sense:

- Crime rate vs robbery rate
- Crime rate vs aggravated assault rate

Noteworthy:

- Population vs robbery rate
- Population vs motor vehicle theft rate



# Heatmap - state data

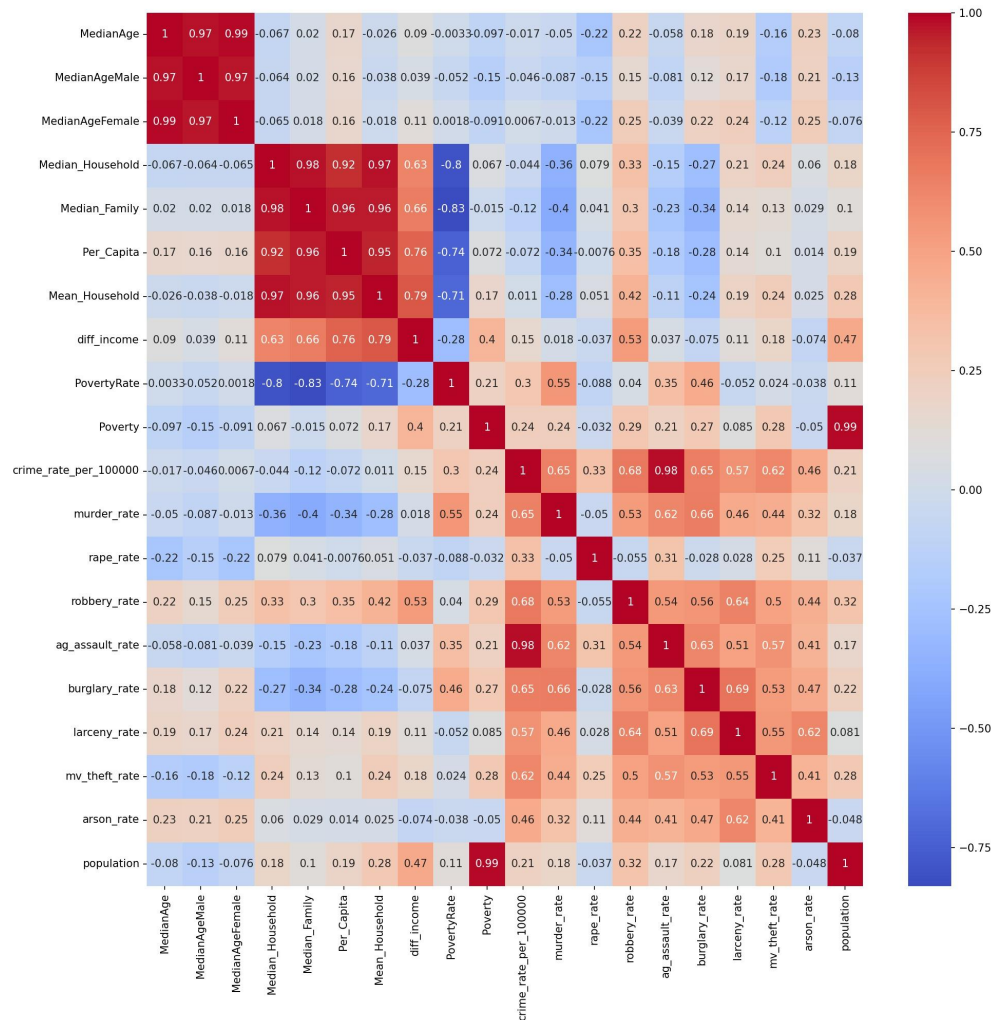
More correlation than before

Recurring observations:

- Crime rate vs robbery rate
- Crime rate vs aggravated assault rate

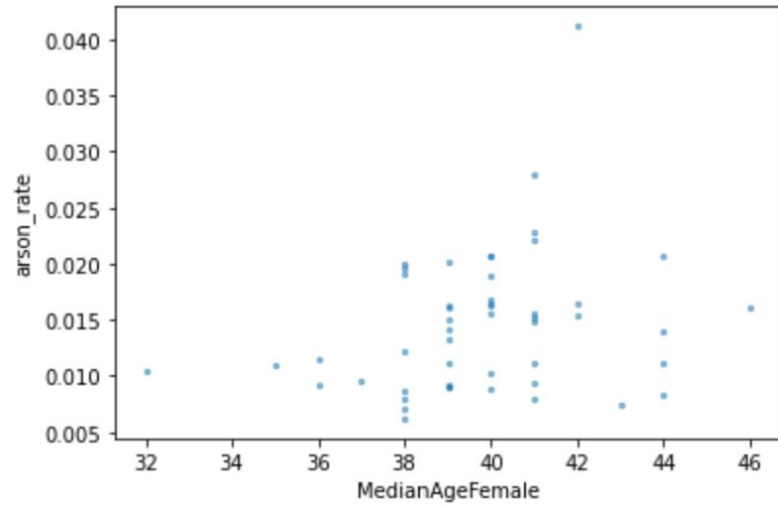
Noteworthy:

- Poverty rate vs burglary rate
- Poverty rate vs murder rate
- Diff income vs robbery rate
- Median female age vs arson



# Finding Something Interesting

- Does population affect robbery and vehicle theft?
- Does poverty rate affect murder and robbery?
- What is the deal with median female age and arson rate?

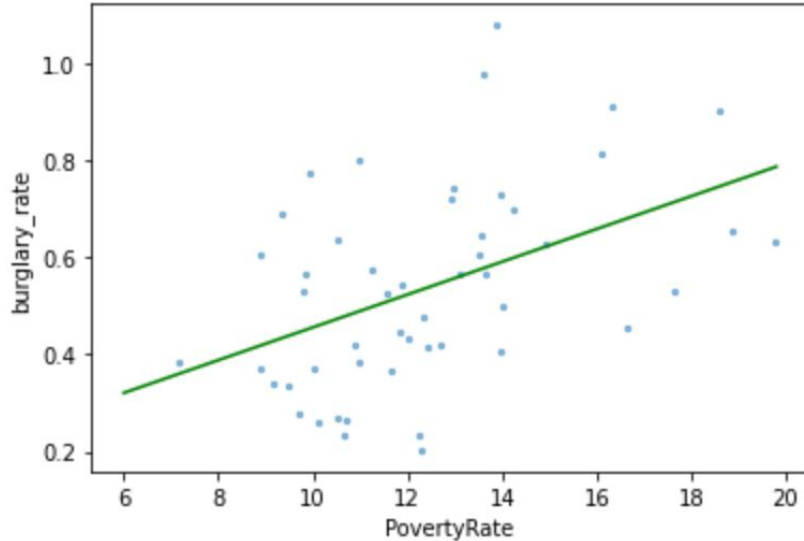


# Fact checking

**Table 1**  
**Gender of Serial Arsonists**  
**(N = 83)**

Gender	Number	Percent
Male	78	94.0
Female	5	6.0

# Poverty rate vs burglary rate



Linear regression:

Slope: 0.03384576204093982

Intercept: 0.11774672877163694

Pvalue: 0.0008858167379710507

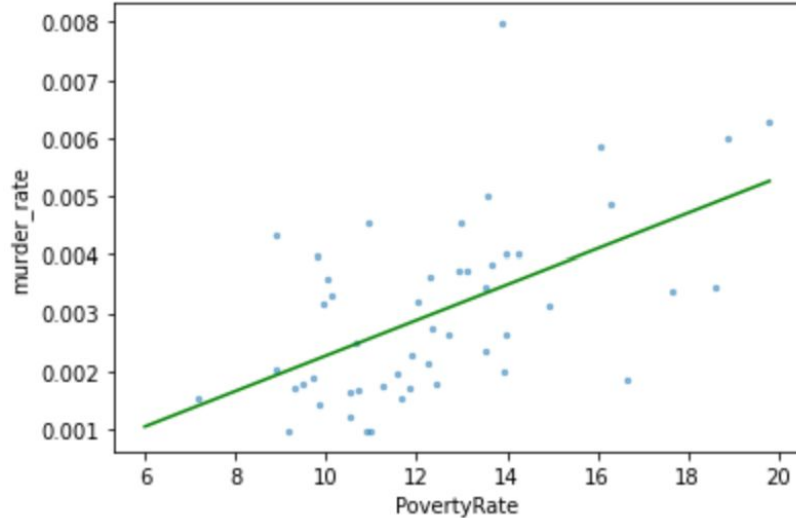
Null hypothesis: the slope of the line is zero

z\_score() standardization:

Slope: 0.4555694921041263

Intercept: 1.2251915489166151e-16

# Poverty rate vs murder rate



Linear regression:

Slope: 0.0003055289890320273

Intercept: -0.0007890128042497718

Pvalue: 3.077942719604505e-05

Small slope:

Murder rates compared to burglary rates are much smaller

z\_score() standardization:

Slope: 0.5533257324719432

Intercept: -1.806506299038438e-16



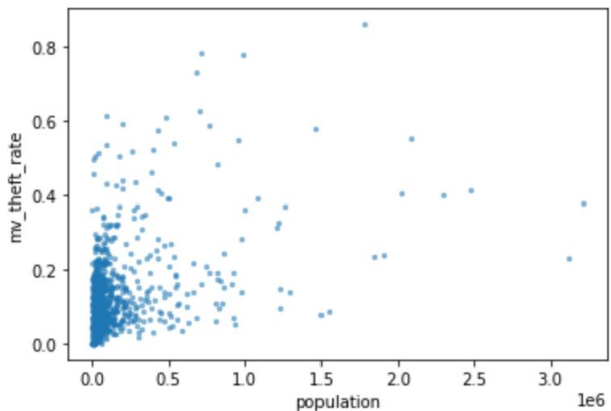
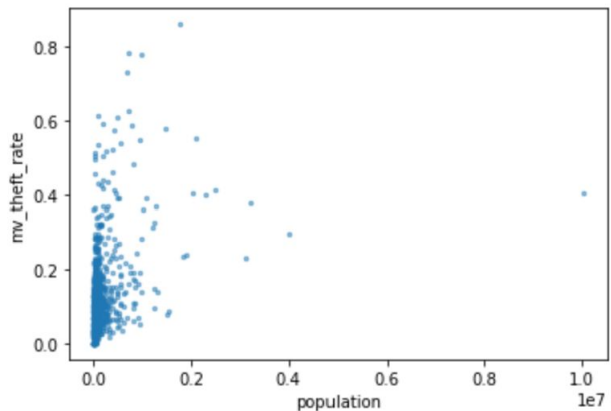
# K-Nearest Neighbors Regressor

```
from sklearn.neighbors import KNeighborsRegressor
```

A k-nearest neighbours regressor finds the k nearest training points and use their values to make a prediction. Usually take the mean of the k points, but could be weighted mean, median, etc.

K-nearest neighbors algorithm can be used both in classification and regression

## Population



## Training the model

```
model score for training data= 0.5188760475969103  
model score for test data= 0.2233782141798717
```

```
r2 score = 0.2233782141798717
```

## Normal Linear Regression

```
Slope = 5.239621817556642e-08  
Intercept = 0.030329669590205947  
Pvalue = 6.8152198100808e-55
```

# Results

The results were less not as strong as originally expected.

- The correlation between poverty rate and murder and burglary was visible and I could access it

However,

- The relation between population and robbery was not strong enough for the model to be successful

# Questions?

(I may not be able to answer them)

CREDITS: This presentation template was created  
by **Slidesgo**, including icons by **Flaticon**, and  
infographics & images by **Freepik**