HOME VALUE PATROL

Team 3

Akashdeep Jaswal
Aparna Ramasamy
Atinderpal Singh
Varuna Damodaran

HOME VALUE PATROL

Agenda

- ★ Introduction & Context
- ★ What? - Research Question
- ★ Data Description
- ★ Exploratory Data Analysis
- ★ Data Cleaning and Curation
- ★ Regression Analysis
- ★ Conclusion

WHY ?
this project

FBI

8,277,829 property crime offenses in the nation in 2014.

$14.3 billion estimated losses caused by property crimes in 2014.

Source: https://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2014/crime-in-the-u.s.-2014/offenses-known-to-law-enforcement/property-crime

WHAT? are we doing

RESEARCH QUESTION

Does crime in a neighborhood affect the home values over time?

HOME VALUE PATROL

EMERGENCY DIAL 911

DATA description

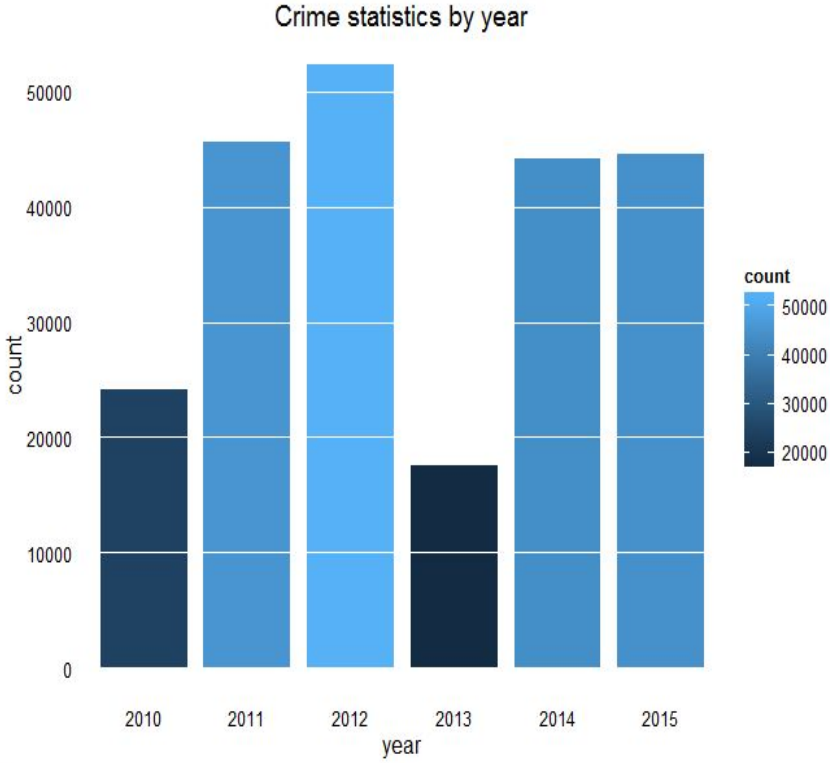
Seattle 911 Response

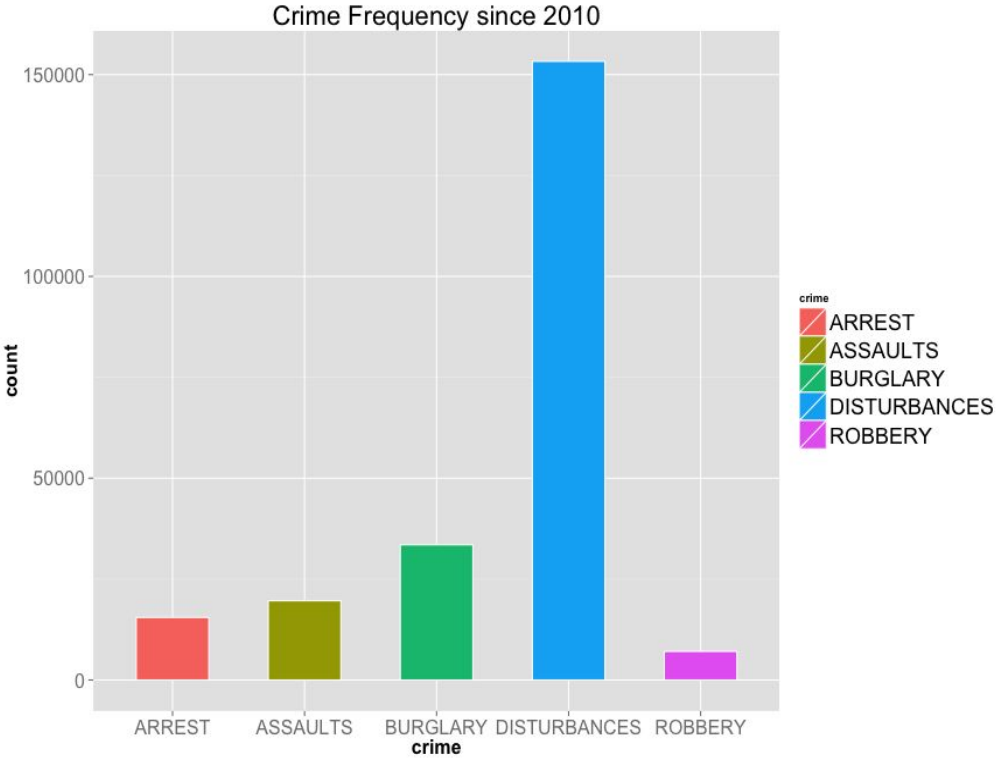
- ★ 1.14 million records
- ★ Timeline 2010-15
- ★ 19 variables
- ★ 46 crime types

Zillow Data Research

- ★ Time-series data
- ★ Dimensions: 5692 x 88
- ★ Year, Month, ZHVI, Neighborhood, Region

EXPLORATORY Data Analysis

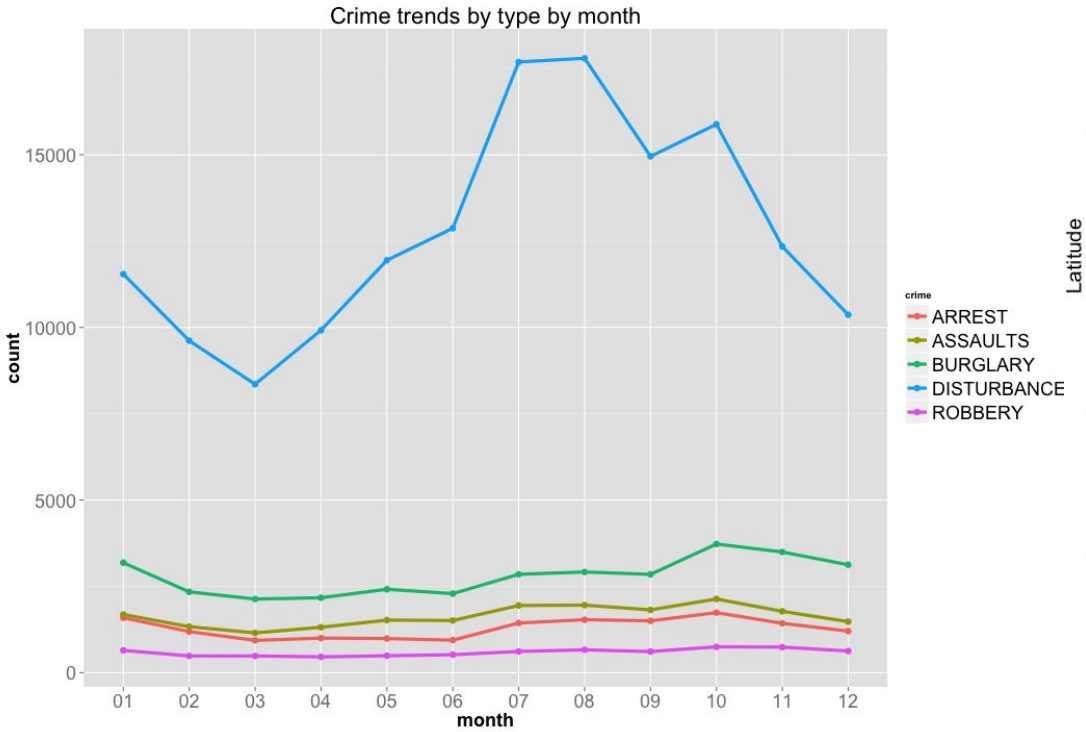


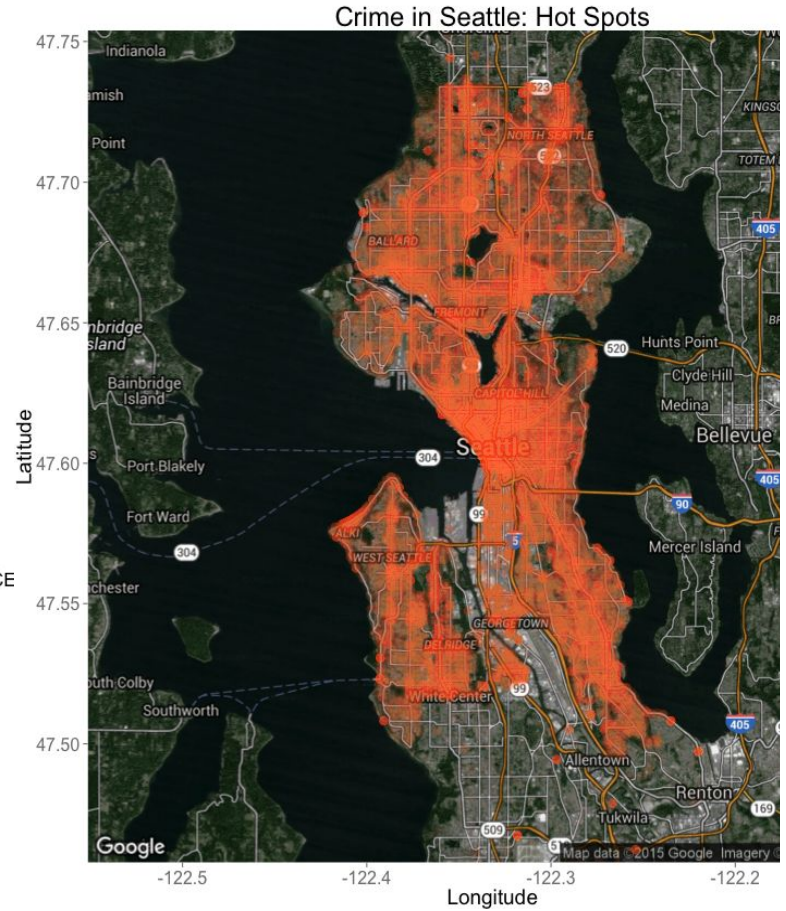


HOME VALUE PATROL

EMERGENCY DIAL 911

EXPLORATORY Data Analysis



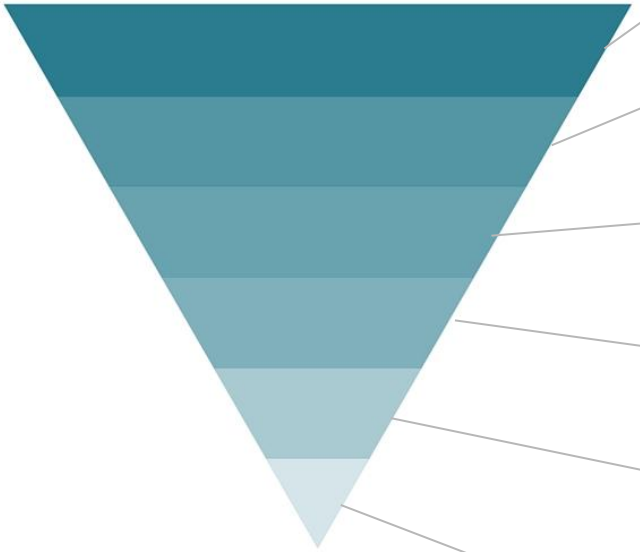


HOME VALUE PATROL

EMERGENCY DIAL 911

DATA!

Cleaning & Curation Methods



Getting rid of unwanted records and columns

Reverse Geocoding Latitudes & Longitudes ~ High Scale Conversion using Shapefiles in R

Data Transformation ~ Tidy Data (Method applied Hadley Wickham - Journal of Statistical Software, vol. 59, 2014.)

Data Aggregation by Year, Month & Neighborhood

Merging the clean datasets by Aggregated Matching

Ready for Regression Analysis

HOME VALUE PATROL

EMERGENCY DIAL 911

DATA!

Cleaning - R Packages used

gpclib - General Polygon Clipping Library for R

rgeos - Interface to Geometry Engine - Open Source (GEOS)

maptools - Reading / Handling Spatial Objects

rgdal - Bindings for the Geospatial Data Abstraction Library

spatialEco - Spatial Analysis and Modelling

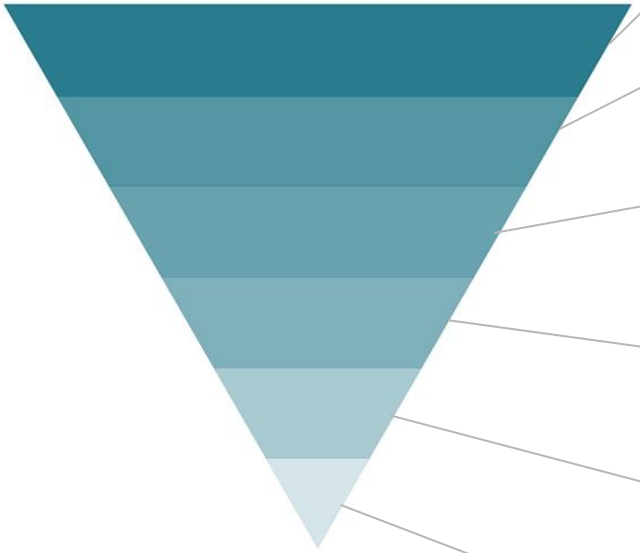
tidyr | ggplot2 | dplyr

Over 300 lines of R code written for the project:

https://github.com/akashjaswal/home-value-patrol/blob/master/project_code.R

DATA!

Cleaning & Curation numbers



1.14 Million records narrowed down to 229K values

Reverse geocoded 229K different Latitude-Longitude combinations and mapped them to 82 neighborhood zones

Zillow data transformed from dim [5692 x 88] to [20592 x 4] dataframes (Reduce & Transpose)

229K values aggregated to 5904 rows (5 Years, 60 Months, 82 Neighborhoods, 5 Crime types)

Final Merged Dataset to append the Zillow Home Value to each of the corresponding aggregated 911 value

Ready for Regression Analysis

HOME VALUE PATROL

EMERGENCY DIAL 911

Data Analysis

Phew, Finally

Crimes aggregated by month, year and neighborhood with corresponding value of house during the same period and neighborhood

project_code.R* | aggregate.911.neighbor... | zillow.911.merged

Filter

	Neighborhood	Year	Month	Assault.Count	Burglary.Count	Arrest.Count	Disturbance.Count	Robbery.Count	home.value
4088	South Beacon Hill	2011	05	14	7	9	28	7	251500
2812	Minor	2014	04	6	8	9	63	1	445100
372	Bitter Lake	2012	03	5	8	9	41	2	253900
1399	First Hill	2011	07	13	8	9	78	4	216800
1740	Greenwood	2011	03	2	9	9	32	0	320500
3063	North College Park	2012	03	3	9	9	31	1	298900
216	Beacon Hill	2010	09	2	9	9	35	2	311200

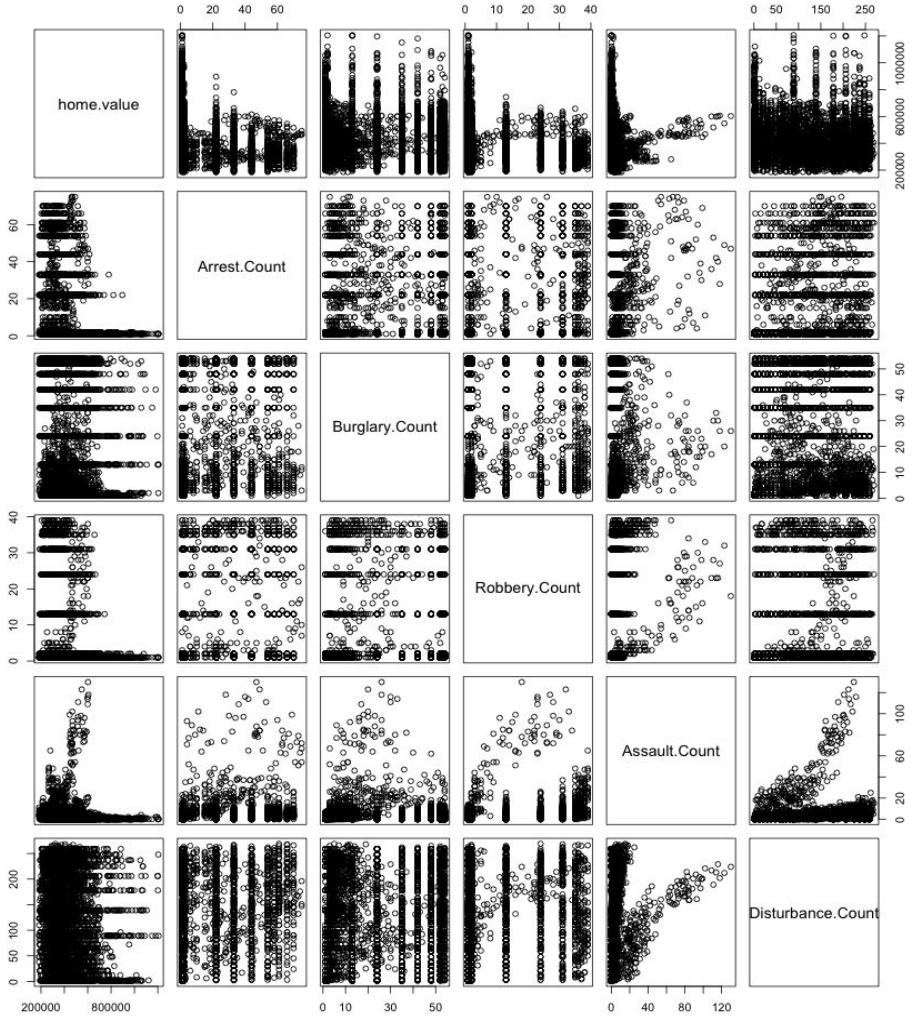
WHAT IS OUR MODEL?

Predicting home values as a function of property crimes

Data Analysis

Collinearity

- ✔ Linear Relationship between data
- ✔ Nearly normally-distributed residuals
- ✔ Constant variability
- ✔ Independent observations



Data Analysis

Multiple Regression

Multiple R-squared: 0.1106

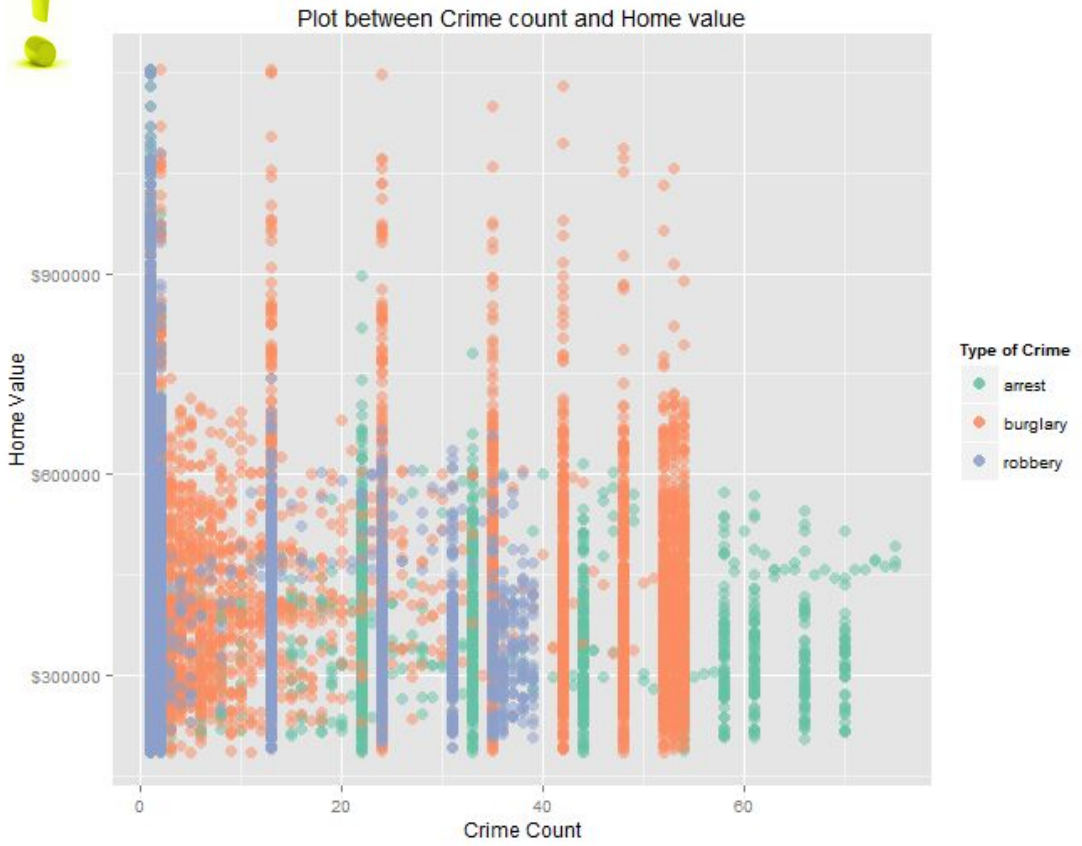
home.value = β_0 + β_1* **Assault.Count** + β_2* **Arrest.Count** + β_3* **Burglary.Count** + β_4* **Disturbance.Count** + β_5* **Robbery.Count**

	Estimate	Pr (>|t|)	95% Confidence Interval
(Intercept)	469028.28	< 2e-16	461599.3, 476457.2
Assault.Count	1401.19	5.18e-09	931.7, 1870.6
Arrest.Count	-2278.03	< 2e-16	-2531.6, -2024.4
Burglary.Count	-114.53	0.2973	-329.9, 100.87
Disturbance.Count	56.38	0.0273	6.3, 106.4
Robbery.Count	-2413.51	< 2e-16	-2876.40, -1950.6

Data Analysis

Model Interpretation

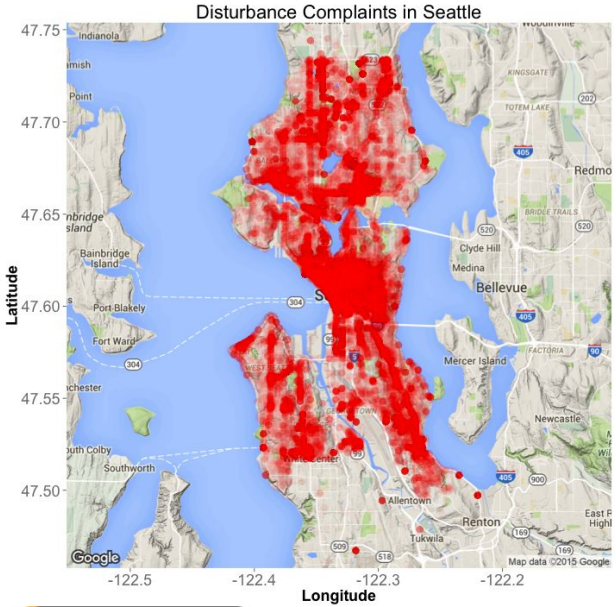
Assault Count	▲ $ 1401.19
Arrest Count	▼ $ 2278.03
Burglary Count	▼ $ 114.53
Disturbance Count	▲ ▼ $ 56.38
Robbery Count	$ 2413.51



Data Analysis

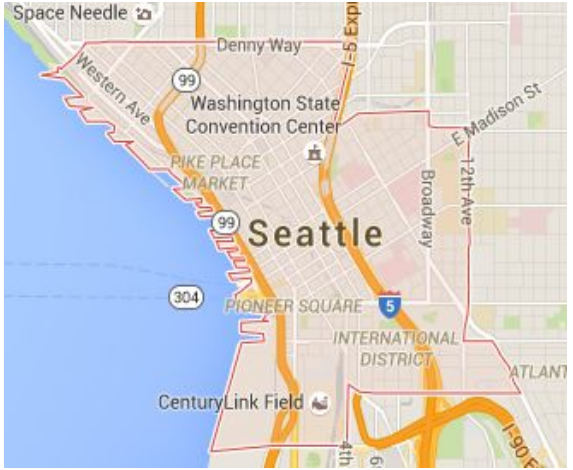
Special Inferences

Disturbance complaints have **less or no correlation** with time or neighborhood ~ evenly distributed



Seattlers like to **Party**!

Since 2010, 4471 of total 17523 **assault** cases reported in Downtown Seattle



You are less likely to be arrested if you stay in an **expensive** house!

Burglary - The victim **does not** have to be present - Makes this less egregious and hence it affects the home value with lesser magnitude.

Robbery - The victim **must be** present at the scene - More egregious crime, has higher negative impact on the value of the homes.

HOME VALUE PATROL

EMERGENCY DIAL 911

Project Scope !

Limitations and Key Learnings

Other Factors include:

- Distance from Workplace, UW and SeaTac Airport.
- Access to Transportation & Departmental Stores.
- History of neighborhood and past data.

Key Learnings

- ✔ Data Cleaning
- ✔ R Programming
- ✔ Open-ended questions
- ✔ Data Science

Thank you!

HOME VALUE PATROL