

Phoenix Crime Analysis

Lyttia Cancinos

Introduction

- Phoenix, AZ is the 5th largest city in the country with a population of over 1,600,000.
- Despite scorching summer temps and a chronically underfunded public education system, people are increasingly adopting Phoenix as their new home.
- In 2017, Phoenix showed the 2nd highest population growth among large cities with populations of 50,000 or more adding 24,036 people in 2017 (US Census Bureau).



The Problem

1. As of now, the Phoenix Police Department is dangerously understaffed.
 - a. Phoenix PD employs just over 3,000 officers, about 1,500 short of the national average, that are needed to police a city the size of Phoenix. (According to Ken Crane, president of the Phoenix Law Enforcement Association)
 - b. Although all hands are on deck and calls are prioritized, there has been increased call response wait times for many years.
2. Living in Phoenix, new residents quickly learn that near city center, crime has become practically unavoidable. Families than can afford it often move to the suburbs usually to provide a safer environment for their children.

Proposed Solutions

Goal # 1:

Strategically deploy officers and patrol areas of Phoenix by understanding crime trends:

- locate the areas of higher and lower crime rates
- determine when, where, and which types of crime occur most often

Goal # 2:

Examine the conditions in which these crimes occur in order to determine:

- contributing factors
- what doesn't seem to be making a difference
- use this information to create a plan to make the city safer and keep families from moving away

Crime Data Used

- City of Phoenix Open Data website found here: [Phoenix Crime Data](#).
 - Data is in the form of a csv file which is updated daily by 11am and includes incidents from November 1st, 2015 forward through 7 days prior to today's posting date.
- At the time of original download (7/2/2018 8:00pm), the data included 169,818 crimes and recorded data for 7 variables.
 - The file can be updated to include new crimes on a daily basis.

Crime Data Variables

1. Crime incident number
2. Earliest and latest date and time the crime occurred*
3. Uniform Crime Reporting (UCR) crime type:
 - a. Homicides, rapes, robberies, aggravated assaults, burglaries, thefts, motor vehicle thefts, arsons, and drug offenses*
4. Hundred block address (to preserve confidentiality)
5. Zip code*
6. Type of location/premise*

Property Data Used

- Median house price by zip code (as a measure of prosperity) will come from realtor.com residential listings database found here:

<https://www.realtor.com/research/data>
 - The data is “based on the most comprehensive and accurate database of MLS-listed for-sale homes in the industry”.
- The data includes 1,094,091 observations of 34 variables
 - I am only using data from Phoenix zip codes- 2,993 entries- and 2 variables.

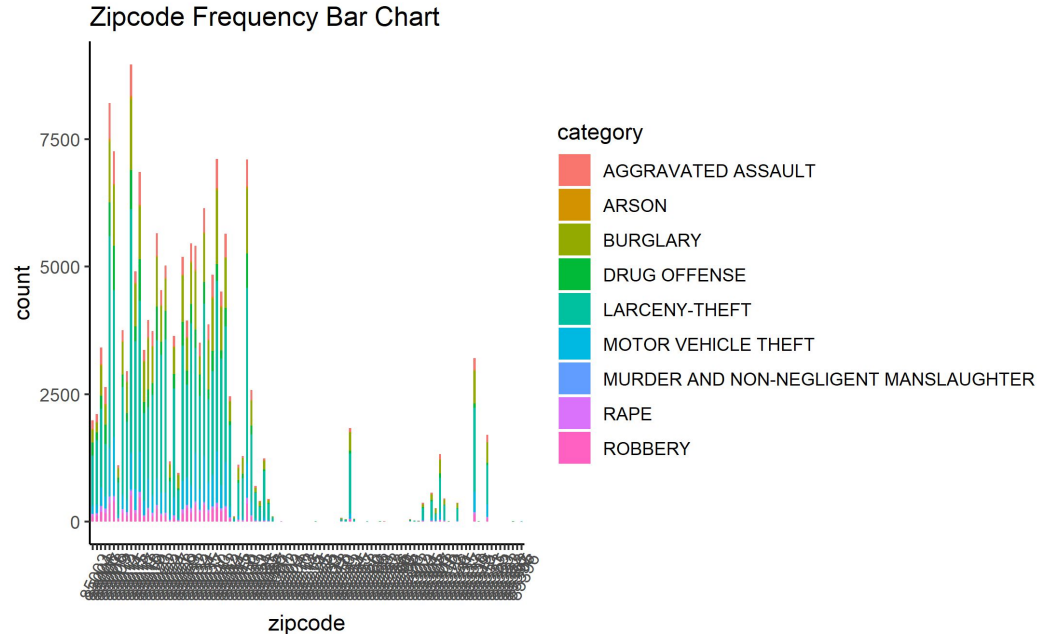
Property Data Variables

Data was provided by zip code from 5-1-12 to 5-1-2018

1. Median & Average list price*
2. Luxury list price
3. Median & Average days on market
4. Total active listings*
5. New listings
6. Price increases and price reductions

Data Preparation

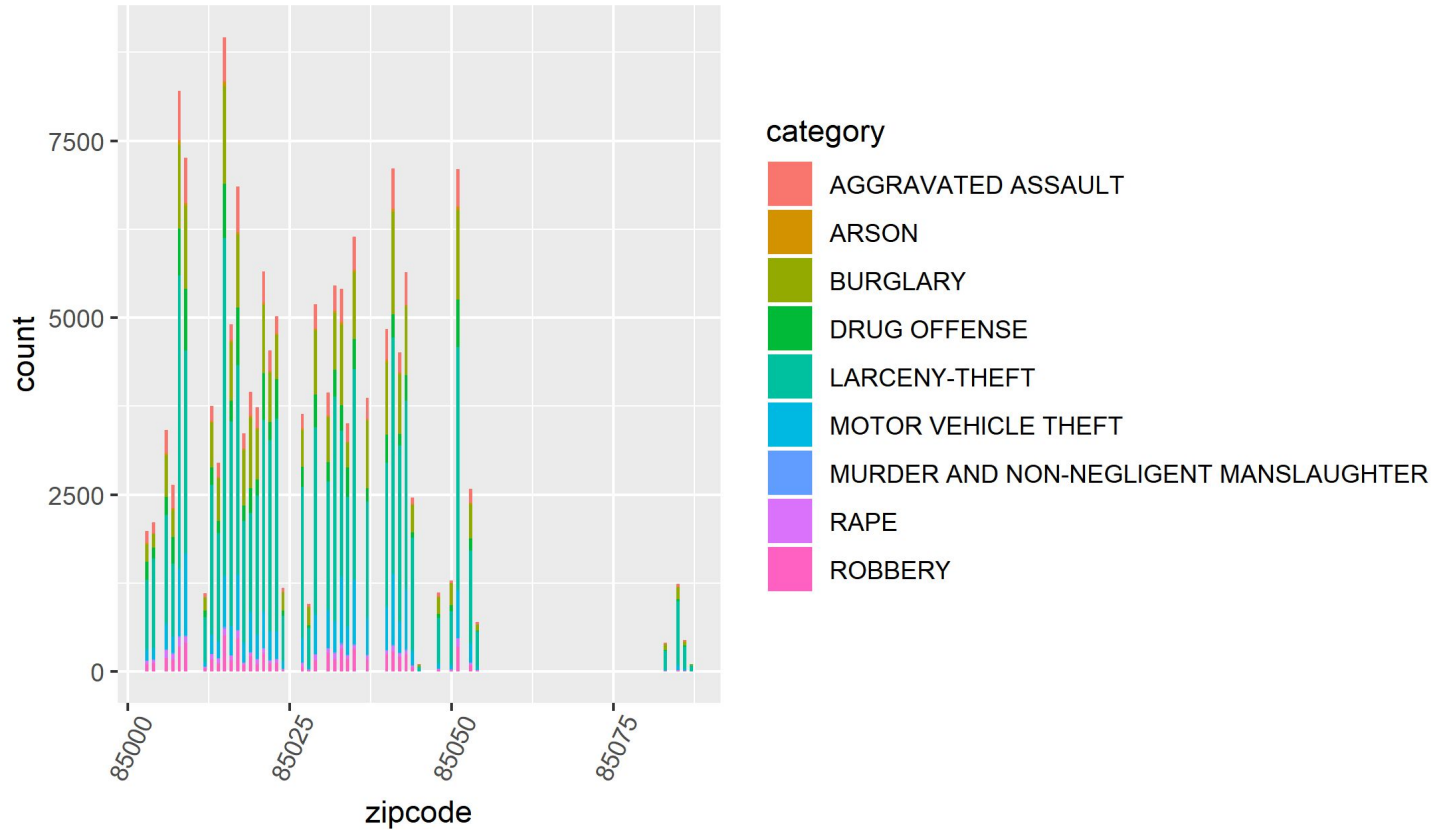
The data was wrangled using R scripts to clean, filter, and join the datasets as well as engineer other variables for use, and create some initial visualizations.



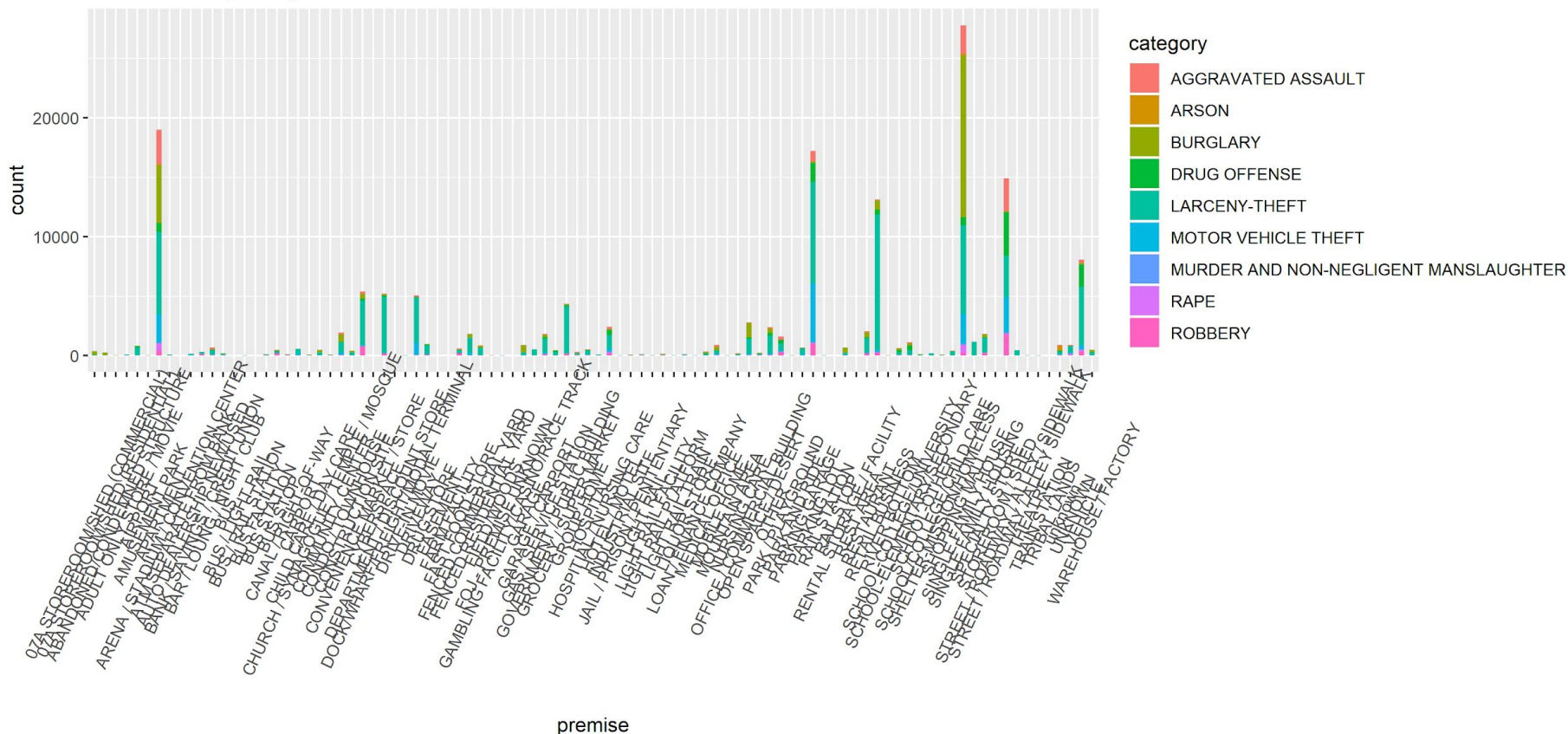
Zip-codes were filtered from the original data to more accurately represent a true crime count, since many of these zip-codes are located outside of official Phoenix boundaries, and likely have many additional crimes reported to their respective police departments.

Phoenix zip-codes start with 850. This left me with 42 zip-codes.

Zipcode Frequency Bar Chart



Premise Frequency Bar Chart



Original 94 premise types. Many are redundant and can be combined. Reduced to 53.

Premise Type distribution once combined.

Highest Crime Rates

1. Crime Type:
 - a. Non violent crimes occur more often than violent crimes, with burglary being the most common crime type
2. Zip-code:
 - a. 85015 has the highest total crime count with 8,965 crimes.
3. Premise Type:
 - a. Single family homes were the most likely premise types to be involved in crime
4. Housing Value:
 - a. Crimes more commonly occur in areas of median/mode housing value, rather than the lowest or upper ranges

Lowest Crime Rates

1. Crime Type:
 - a. violent types including
2. Zip-code:
 - a.
3. Premise Type:
 - a.
4. Housing Value:
 - a. upper range

Crime Trend by Property Value

Lower Housing

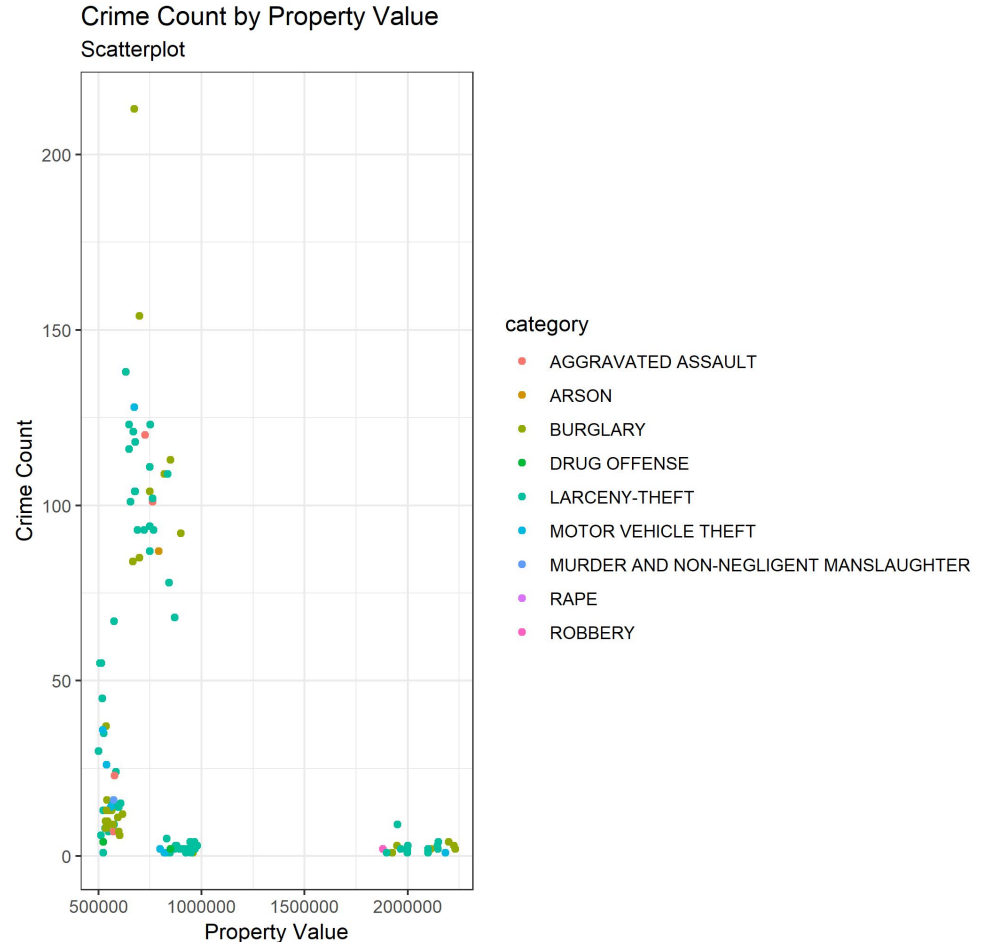
- _ crime counts are reported for housing in the _ property value range.

Middle Housing

- crime counts are reported for housing in the _ property value range.

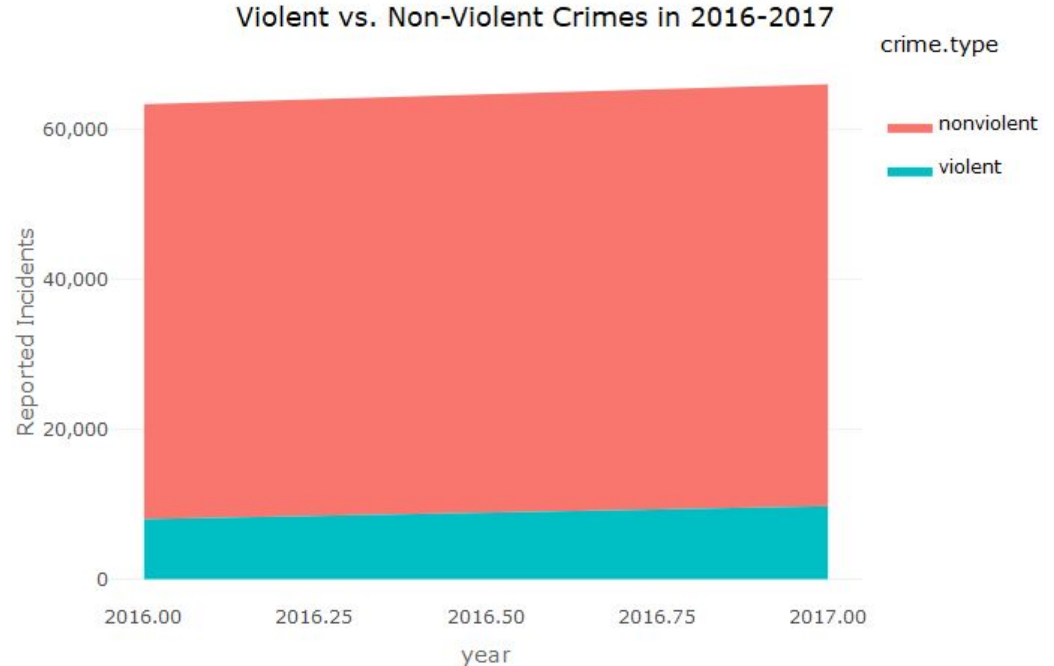
Upper Housing

- Much smaller crime counts are reported for housing in the 500,000- 2,000,000 property value range.

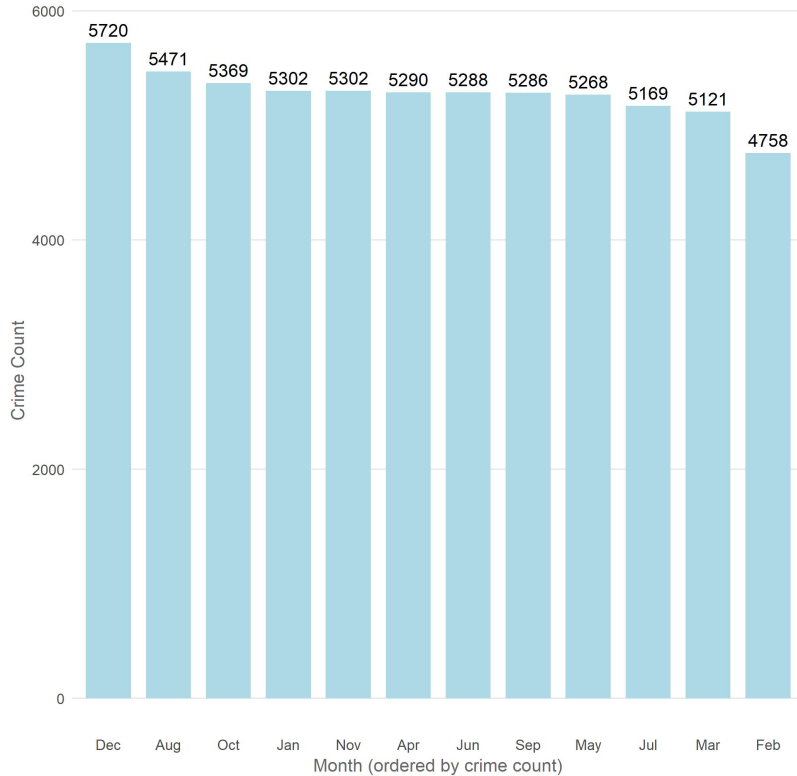


2016-2017 Violence Progression

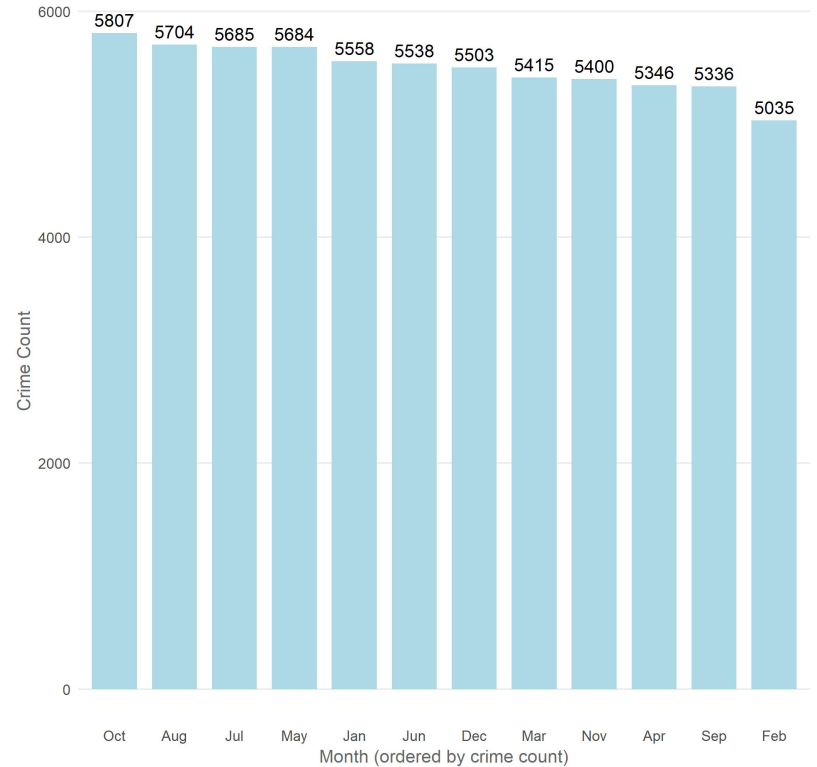
- The graph to the right shows a gradual increase in crimes overall from 2016 to 2017 with a much larger occurrence of non violent crime.
- Crime seemed pretty consistent throughout the year, averaging about 5500 crimes each month in 2016 and 2017.



Crimes reported to Phoenix PD each Month in 2016



Crimes reported to Phoenix PD each Month in 2017



Number of Crimes reported each month in 2016 and 2017 ordered by crime count.

These graphs show bottom limits of about 5,000 crimes to upper limits of 6,000 crimes reported to Phoenix PD.

Seasonal Trend

Is there a correlation between Crime and Seasons in Phoenix?

Trends by Day

Is there a correlation between Crime and Day of the Week/Month in Phoenix?

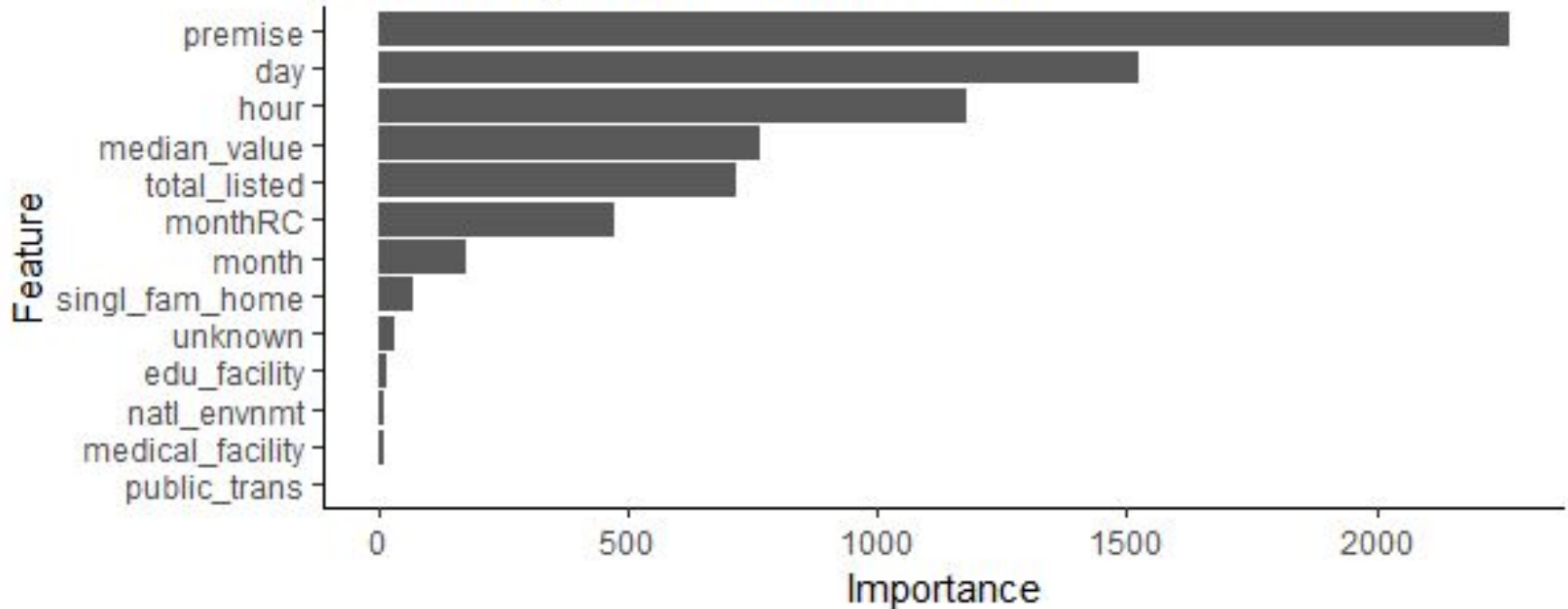
Methods Used

Since I wanted to fit a model that can find a pattern in categorical variables that may or may not be linear and may have a degree of multicollinearity, a decision tree algorithm works best.

The Random Forest algorithm is a predictive modeling tool that can find patterns in data by randomly creating numerous decision trees and determining the importance of each indicator to the outcome.

The algorithm produced by this process can then be used to predict whether a crime is likely to be violent or not, depending on the conditions presented.

Feature Importance: CrimeForest



This graph shows the degree to which each feature (or variable) makes a difference in the type of crime occurred. This translates to an importance rating for the model's prediction algorithm.

Model Evaluation

Confusion Matrix

Cross Validation

Proposed Use

This type of prediction could be used along with crime hotspot maps and other visualizations, to provide:

- decision support to police departments
- inform policy makers
- It could also be used in a program that assesses and assigns a community's crime rating to provide decision support to home buyers, and could be listed on websites that assist in home buying (like Zillow).

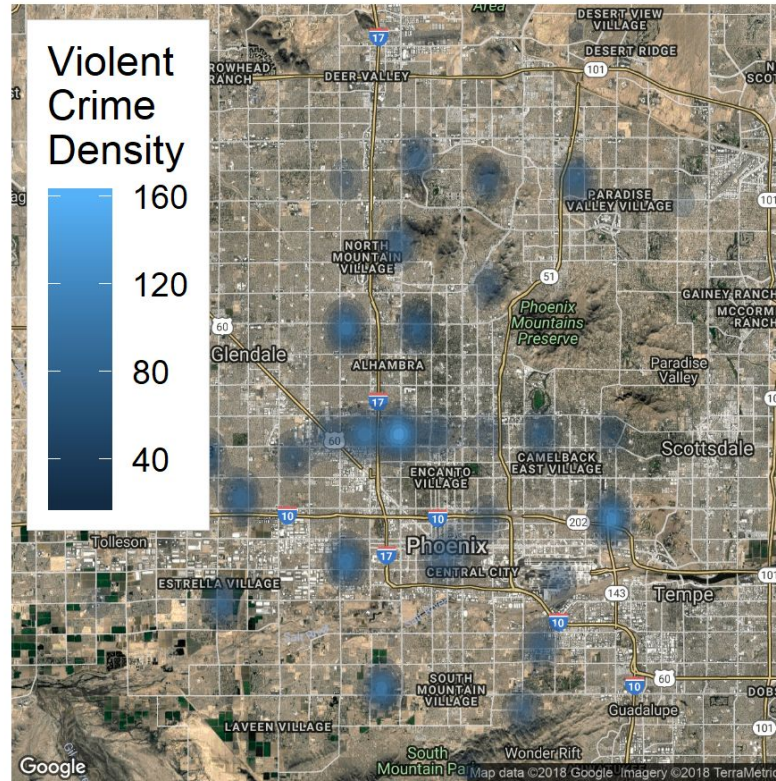
Advantages of Analyzing Crime

1. Phoenix law enforcement is extremely limited in officers and resources.
 - a. It is not possible to put equal resources into every investigation or to catch every criminal.
 - b. Random patrol is not the most effective way to use limited resources.
 - c. Analysis allows for strategic deployment of patrol officers.
2. Analysis can provide information that will support operations on a daily basis.
 - a. Creating and analyzing crime maps,
 - b. Provide up to date information on crime series, patterns, and trends,
 - c. Can link crimes to locations or suspects for investigators.

Limitations

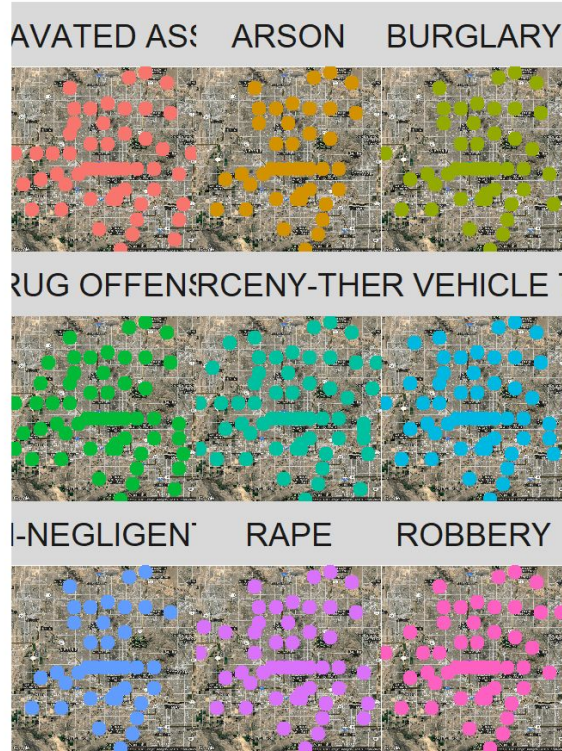
1. Open data = Confidentiality
 - a. Hop spot maps would be much more telling with precise location data for each crime
2. Free API requests = limited call capacity
 - a. Weather data was limited to 10 calls/minute, and full of error messages
 - b. R continuously went over call limits, even when limits were explicitly set
 - c. Too much data was missing. To figure this out, I need more time.
3. No data on criminals

Crime Density in Phoenix



This density map is the beginnings of a crime hot spot map, in need of more accurate lat/long coordinates to increase descriptive ability.

Crime in Phoenix



Crime Type geographic distribution: To be updated with alpha & jitter to show density

Proposed Next Steps & Further Research

I need more information to analyze before I can make more pointed predictions.

True crime analysis usually incorporates information about suspects and apprehended criminals as well as precise locations.

This information can help in further predicting repeat offenses and stopping crime before it happens.

While detailed data collection should be in place, it is understandably not open to the public.

Sources

Crime Data:

[[City of Phoenix Open](#)

[Data](#)](https://phoenixopendata.com/dataset/crime-data/resource/0ce3411a-2fc6-4302-a33f-167f68608a20?view_id=644b88ef-16b3-497d-9413-2ba3eedfd3c1)

Zip code lat/lon coordinates:

[[Free Zip Code Database](#)](http://federalgovernmentzipcodes.us/)

Property Data:

[[Realtor.com](#)](https://www.realtor.com/research/data)

Phoenix Area Base Map:

[[Google Maps Static API](#)](https://developers.google.com/maps/documentation/maps-static/intro)

GGMaps:

D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2.

The R Journal, 5(1), 144-161. URL

<http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>

Weather Data (weather conditions):

[[rwunderground API Rpackage](#)](https://github.com/ALShum/rwunderground)