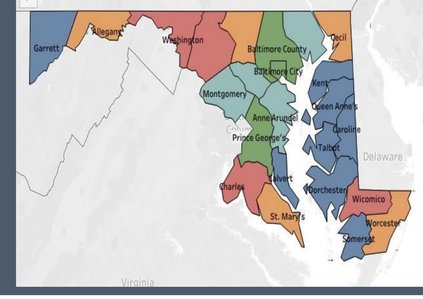
Maryland Crime Rate Trend Analysis

IS 733 Data Mining | Poster Presentation
Akshay Negi, Nishigandha Karle, Ninad Sawant, Prerana Prabhakar,
Vamshi Krishna Yenmangandla,
University of Maryland,Baltimore County



OVERVIEW

We utilize data mining methodology to perform an exploratory data analysis on Maryland's crime rate, which has a higher violent crime rate (4.7) than the national average (3.7). The goal is to identify 1032 rows and classify crime patterns, crime based on locations and figure out which of Maryland's counties falls into the highest crime rate categories. The first thing that matters most is the safety. Given that Maryland is known for its violence, predicting crime will help in demonstrating which county has the highest risk factor. Our motive is to raise consciousness among people about the factors that cause unsafe behaviour. Analysis of crime can be provided to Police department which will help them increase the security or avoid crimes from happening. By understanding the patterns of criminal behaviour, many criminal investigations can be solved.

DATA PREPROCESSING

Data Cleaning: The dataset was available in a comma-separated values (CSV) file, contained over 1032 records ranging from 1975 to present.The raw data have been transformed in a useful and efficient format by doing Data cleaning using python. All the blank rows were taken care of and made data error free.

Data Re-scaling: We need to re-scale the YEAR variable, before building a linear mixed-effects regression model. Regression models gives best results when the intercept is near zero but YEAR starts at 1975, in the raw data. The model will fail to converge, if we use YEAR without re-scaling. We'll create a new variable, YEAR_R that starts at zero rather than 1975.

```
# A tibble: 6 x 4
  JURISDICTION     YEAR POPULATION crime_rate
  <chr>           <dbl>      <dbl>      <dbl>
1 Allegany County  1975      79655       178.
2 Allegany County  1976      83923       104.
3 Allegany County  1977      82102       155.
4 Allegany County  1978      79966       128.
5 Allegany County  1979      79721       138
6 Allegany County  1980      80461       148.
>
```

```
> head(crime_use)
# A tibble: 6 x 5
  JURISDICTION     YEAR POPULATION crime_rate YEAR_R
  <chr>           <dbl>      <dbl>      <dbl>  <dbl>
1 Allegany County  1975      79655       178.      0
2 Allegany County  1976      83923       104.      1
3 Allegany County  1977      82102       155.      2
4 Allegany County  1978      79966       128.      3
5 Allegany County  1979      79721       138       4
6 Allegany County  1980      80461       148.      5
>
```

METHODS

A scatter plot was plotted to check the correlation between the crime rate and Population. This scatter plot was for the entire state of Maryland. It showed that less the population more the crime rate, but it should always be inverse so plotted a scatter plot considering individual county. The first plot was for Baltimore city, it showed a positive correlation between population and crime rate but checked for Howard County then there was a negative correlation. To understand the crime trend in every county, LMER (linear mixed-effect regression) model is used. LMER deals with fixed and random effects of variables in the form of the equation of a line. Considering the data, the equation is crime_rate ~ Year_R + (Year_R | Jurisdiction). Here crime_rate is the dependent variable to be predicted and Year_R is the intercept and Jurisdiction is the grouping variable. So, the dependent variable crime_rate is predicted by an intercept and a random error term for the intercept.

Table 1: LMER generated values for the county to plot the map

RESULTS

The LMER model in (**Figure 1**) show various trends in the state of Maryland. As the name suggests a linear trend but the map depicts that there is no linear trend in the state. The crime rate changes for every county. The scale on the right hand side shows the percentage increase in the crime rate on the color band from red to the highest to sky blue to the lowest. Seeing at **(Figure 4)**, the percentage change between the crime rate for all counties. For example the county which depicted negative change is indicated as light red in the map which tells that the rate is reduced for that county by that percentage. **(Figure 4)** also tells us that Baltimore city has the highest crime rate for the span of 40 years but when we compare the crime trend the city is showing a decrease in the crime. The **(Table 1)** is showing the various calculations done to calculate the slope of every county to be plotted on the map. The values are generated by using LMER and then another formula is applied to calculate the county slope and similarly US map is used for depicting this trend.**(Figure 3)** shows the values of LMER predicted using scatter plot. **(Figure 2)** is showing the various crimes happening in Maryland according to the year and their trends like for which year it is increasing and similarly for which year it is decreasing.

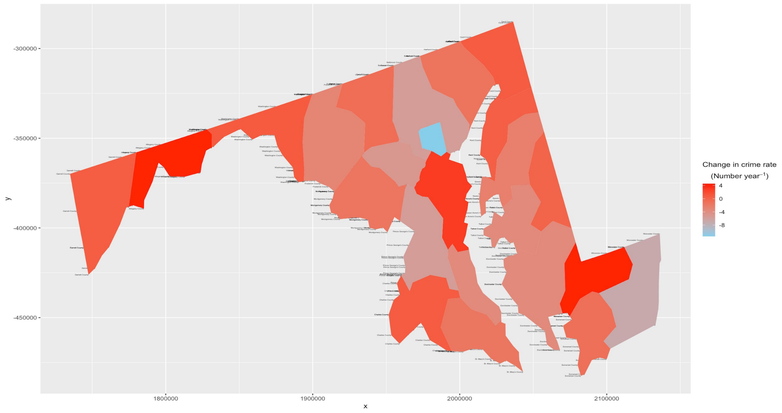


Figure 1: Trends in the Maryland crime data using LMER(Linear Mixed Effect Regression)

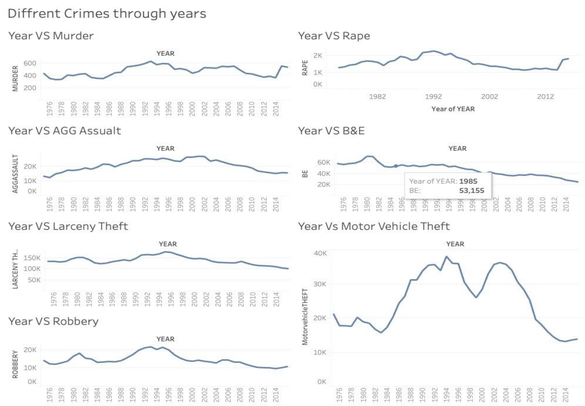


Figure 2: Comparing different crimes against year.

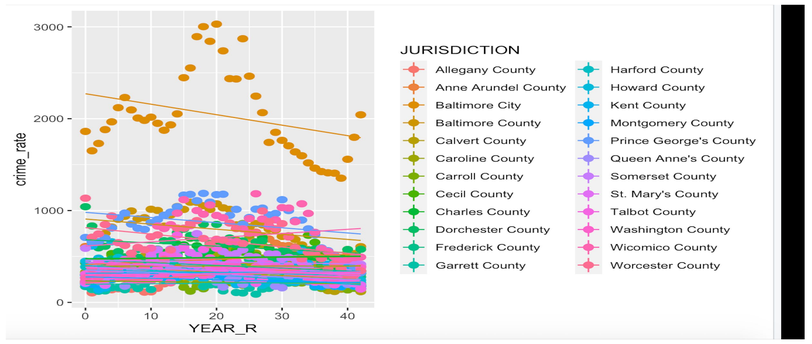


Figure 3: Scatter plot for the LMER values. Both fixed and random effect.

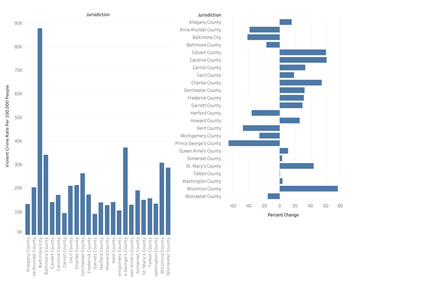


Figure 4: Percent change in crime rate for counties in the state of Maryland

Figure 5: Population vs the crime total

CONCLUSION

Crime rates are varying across the entire world, so it is very difficult to detect . By visualising the trends in the crime rate, the policy makers can decide on different types of plans based on the trends or the happenings of crime in that county. The data which we are dealing with has been collected from **Maryland Statistical Analysis Center** and **Open Maryland** Site . The data is collected for 40 year time span and this data is enough to visualise and detect the trends to help government to increase the safety measures in that particular county.As the concern was to show trends, regression model was used. Post research Linear Mixed Effect Regression(LMER) was used to show the trend. To cross verify the result of the LMER , we plotted the same trend using Tableau to show the percent change in the crime and the result matches with the LMER output. After analysing the data there were few questions like whether the population influences crime rate, this question can be answered through Tableau Visualisation as shown in **(Figure 5)** and it can be found that it has a negative correlation and thus population does not matter for crime to take place .Hence we conclude that there is no linear trend in state of Maryland, and it differs from county to county.

REFERENCES

Fitting Linear Mixed-Effects Models Using lme4.
Retrieved from https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf

A beginners guide for Lmer.
Retrieved from https://rstudio-pubs-static.s3.amazonaws.com/63556_e35cc7e2dfb54a5bb551f3fa4b3ec4ae.html

Statistics with R.
Retrieved from http://sia.webpopix.org/lme.html

US map using R.
Retrieved from https://cran.r-project.org/web/packages/usmap/vignettes/mapping.html