UNITED STATES MILITARY ACADEMY

FINAL PROJECT REPORT

MA478: GENERALIZED LINEAR MODELS

SECTION H2

COL NICHOLAS CLARK

By

CADET ISABELLA PALCHAK '24, CO D1

WEST POINT, NEW YORK

07 MAY 2024

IP I CERTIFY THAT I HAVE COMPLETELY DOCUMENTED ALL SOURCES THAT I USED TO COMPLETE THIS ASSIGNMENT AND THAT I ACKNOWLEDGED ALL ASSISTANCE I RECEIVED IN THE COMPLETION OF THIS ASSIGNMENT.

_______ I CERTIFY THAT I DID NOT USE ANY SOURCES OR RECEIVE ANY ASSISTANCE REQUIRING DOCUMENTATION WHILE COMPLETING THIS ASSIGNMENT.

SIGNATURE: Isabella Palchak

Final Project Report

CDT Isabella Palchak

May 7, 2024

Abstract

We were tasked with building a statistical model to capture the impact of potential socio-economic and environmental covariates on the number of burglaries in Chicago. We investigated the effects of 'population,' a measurement of 'unemployment,' a measurement of 'wealth,' count of 'young males,' and average 'monthly precipitation' on the number of burglaries by Chicago census block. Our exploration resulted in the construction of two Poisson mixed effects models and one Poisson random effects model. Comparison of the models indicate that our Poisson random effects model best captures the impacts of the socio-economic and environmental covariates on the number of burglaries by census block in Chicago. Our results suggests that the Poisson Random Effects model was most effective. Additionally, we found a decreasing trend in burglaries across time and that wealth has a negative impact on burglary rates.

1 Key Words

Socio-economic variables: We measure this through the following covariates: 'population,' a measurement of 'wealth,' a measurement of 'unemployment,' and the count of 'young males' by Chicago census block.

Fixed Effect: An estimate that attempts to capture the effect of a covariate on our variable of interest.

Random Effect: A random variable that captures unique attributes of a covariate. To describe a random effect, we do not report an estimate, rather, we report variance and standard deviation to capture its distribution.

2 Introduction

We were tasked with building a statistical model to capture the effects of socio-economic and environmental impacts on the number of burglaries in Chicago. This project is motivated by Chicago's crime history. The city often makes national headlines for the murders, assaults, burglaries, and other violent crimes that occur within its boundaries.[1] However, not all areas of Chicago are equally dangerous. According to *Crain's Chicago Business*, the North Side of Chicago is considered relatively safe with average crime rates. Whereas the West Side is thought to be one of the most violent neighborhoods in the world.[2] These polarizing burrows of Chicago have very distinct socio-economic characteristics. For example, the West Side contains many poverty-stricken and underdeveloped communities. Whereas the North Side sports more average statistics across the board. Knowing that Chicago crime rates differ by location, we aimed to model the burglary counts in Chicago by census block. We also know that police presence is associated with decreased crime rates.[3] So, our overarching goal was to develop a statistical model that could provide insight into how the police forces could better deploy its resources to reduce its overall crime rates.

[1]Ruteki, Jared. "Chicago's Murder Count Makes National Headlines, But Accurately Talking About Crime Data Requires Nuance." *WTTW Chicago.* August 31, 2013. Accessed April 2024. https://news.wttw.com/2023/08/31/chicago-s-murder-count-makes-national-headlines-accurately-talking-about-crime-data

[2]Hendershot, Steve. "The Inequality of Saftey." *Crain's Chicago Business.* October 24, 2022. https://www.chicagobusiness .com/rains-forum-safer-chicago/chicago-violence-problem-debate-safety-inequality

[3]Ibid.

3 Literature Review

Other Applied Statistics and Data Science researchers have addressed similar investigations of Chicago's crime data. We will briefly explore three peer-reviewed journal articles and explain how my investigation was similar or different. Each article has many contributors, so we will reference the relevant articles by their sources: Missouri State University, University of Baltimore, and the University of Arizona. First, we will explain the results of spatial explorations, then the results of temporal explorations, then briefly explain how this literature is relevant to our project.

An exploration done by Jun Luo at Missouri State University, "Multi-spatiotemporal patterns of residential burglary crimes in Chicago: 2006-2016," concludes the burglary rates in Chicago are not static. Rather, Luo found that the rates burglary crimes depend on both location and time. He utilized geospatial information systems programs to conduct his research and produce heatmaps that he used to visualize how the crime hotspots change. Lou also found that the hotspots on his maps depended on the type of spatial clustering he used.[4] For example, crime hotspots vary depending on whether the data was clustered police district, or police and community patrolling zones. The outcome of this research demonstrates that crime rates depend on location. Lou's research is not unique, other researchers have taken similar approaches in exploring crime data in Chicago.

A team of researchers from the University of Baltimore produced a research article, "Crime Analysis in Chicago City," that demonstrated similar conclusions. Like Lou, these researchers clustered the crime data to identify crime hotspots. However, these researchers utilized different techniques to conduct the analysis. These researchers used K-means clustering and Spatial Clustering to investigate crime hotspots. K-means clustering is a method grouping data observations that have similar characteristics into clusters. They also used a computer program, called SatScan, that clusters inputted data by using Poisson modeling and then overlays the results onto a map for explainability.[5] So the researchers from both Missouri State University and the University of Baltimore identified hotspots in Chicago crimes and recommend further explorations to identify characteristics of can explain why crime rates depend on location.

In addition to spatial modeling, researchers have investigated Chicago crime data across time. Lou, from Missouri State University, and researchers at the University of Arizona make conclusions that Chicago crime rates differ temporally. Lou explored Chicago crime data by hour, day, and month while simultaneously considering the location clusters previously discussed. Similarly, researchers from the University of Arizona also explored the Chicago crime data across years and found that crime rates depend on both holidays and weekdays.[6] Both research groups conclude that time affects crime rates in Chicago, and certain times, such as weekdays or holidays, seem to have significantly different crime counts.

Similar to our project, the researchers from the University of Arizona investigated the impact of environmental factors on Chicago crime rates. They explored how weather impacts crime rates and found that aggressive crime rates depend on the temperature. They discovered that temperature impacts the location of crime, generally determining whether the crimes occur inside or outside. Additionally, our exploration will investigate the impact of location and time on the crime rates, just as the researchers from Missouri State University and the Universtiy of Baltimore researched. However, we differ from this literature because we will investigate clusters of census block groups and the total number of burglaries by month. I will explain the methodology in greater detail below.

[4]Luo, Jun. "Multi-spatiotemporal patterns of residential burglary crimes in Chicago: 2006-2016." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4 (2017)*: 193. https://doi.org/10.5194/isprs-annals-IV-4-W2-193-2017.

[5]Alqahtani A, Garima A, and Alaiad A, "Crime Analysis in Chicago City," *2019 10th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2019, pp. 166-172, doi: 10.1109/IACS.2019.8809142.

[6]Towers S, Chen S, Malik A, Ebert D (2018) "Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective." *PLoS ONE 13(10)*: e0205151. https://doi.org/10.1371/journal.pone.0205151

4 Methodology

We were tasked to build and develop a statistical model that captures the effects of potential socio-economic and environmental covariates on the count of Chicago burglaries. Utilizing our provided data, our problem statement was to develop and choose a model that best captures the effects of 'population,' 'unemployment,' 'wealth,' count of 'young males,' and 'monthly precipitation' on the number of 'burglaries' by census block in Chicago.

The data that we used for this project was retrieved from GitHub. We were provided with a spreadsheet containing the Chicago burglary data for 552 different census block groups over the years 2010 to 2015. This data was count-data where each row represented a census block, meaning that there were 552 rows of observations. The columns represented every month from 2010 to 2015. So, an individual cell in the spreadsheet contained the count of burglaries within a specific census block during a particular month and year.

Additionally, we had separate spreadsheets containing data for the following covariates: a measure of 'unemployment', a measure of 'wealth', total 'population', and a count of the number of 'young males' in each census block. These four spreadsheets contained 552 observations and one column containing the information described above. However, the methods used to measure 'unemployment' and 'wealth' were not specified. These factors were investigated to explore the potential effects of socio-economic factors on the count of burglaries. Additionally, we incorporated the average 'monthly precipitation' retrieved from the National Weather Service. The precipitation data was included to investigate potential effects of environmental factors on the rate of burglaries. Precipitation only measures rainfall, meaning that it does not account for snow in the winter months.

We utilized the following flowchart to complete our project:

1. *Compile the dataframe:* condense the 'crime', 'population', 'uemployment', 'young males', 'wealth', and 'precipitation' spreadsheets into a single dataframe.
2. *Explore the data:* briefly explore the distributions of the covariates and variable of interest.
3. *Transform the data:* pivot our data from wide format to long format. Aggregate the count of burglaries over the years and condense it into a single column for each census block. Meaning that there will be 12 rows for each census block, where each row represents the count of burglaries between 2010-2015 in that month.
4. *Explore data again:* Investigate the covariates again. Identify the spike in burglary counts during the warmer months, and the decreasing trend over all the years.
5. *Build models:*
   (a) Poisson Mixed Effects 1 (the effect of each variable is the same for each block)
   (b) Random Effects Model
   (c) Mixed Effects 2 (the effect of each variable differs by block)
6. *Explain the models:* Interpret our results and expxlain our models in the context of the problem. Use AIC to choose the best model.

5 Data Preparation

The first step in our exploration was to condense the multiple sources of data into a single dataset. After merging the data together in Excel, we aggregated the burglaries across time. This was one of the most challenging aspects of this whole project. After out initial exploration of the data, we determined that we would to build three Poisson models. However, the data we previously worked with in class was in a different format than the data we were provided with for this project. The challenge was recognizing that our data was not formatted properly to code the models that we built on paper. In order to code our models in R-Studio we needed to transform our data into a different format. The challenge was visualizing how we needed to transform the data in order to code the models we built.

Once we recognized that we needed to aggregate the count of burglaries over time, then the process itself was quite simple. However, it was very tedious because we made our data transformations in Excel. Our final data frame consisted of 6,624 rows where each of the 552 census block had twelve observations–the count of total burglaries for each month between 2010-2015. Additionally, each block had its associated static measurements of 'unemployment,' 'population', 'wealth,' and 'young males' across the five years. The measurement of 'precipitation' was static across location but varied across the months and years.

6 Data Exploration

Our initial data exploration revealed an overall decreasing trend in the rate of burglaries across the five years. This trend can be visualized by the line chart in *Figure 1* in *Appendix A*. This visual aligns with previous research indicating that Chicago crime rates have been on a steady decline. A Yale University researcher, Dr. Andrew V. Papachristos, examined crime trends and visualized a continual decrease in crime rates between 1965 to 2013.[7] So, our data aligns with findings from past research. In addition to the general decline across the years, there appears to be a cyclical spike and drop-off in the yearly crime rates, again see *Figure 1* in *Appendix A*. This figure indicates that there are higher crime rates in the summer months and lower crime rates in the winter months. Finally, *Figure 1* is colored by precipitation. Where the highest precipitation is seen with the rise in crime and lower precipitation is seen with the low rates in crime. This visual was created prior to aggregating burglaries across time.

We also created *Figure 2* in *Appendix A*, by plotting the observations of 'wealth' against the total 'population'. The method by which 'wealth' was measured is unknown. However, this visual gives us insight because it shows that wealth increases as population increases. Meaning that wealth appears to be an aggregated measurement for all of the individuals within a census block. *Figure 2* indicates that 'wealth' and 'population' are highly correlated. To account for this colinearity, we created a new variable where we divided 'wealth' by the total 'population' in each block. This created a new measurement of standardized wealth. This allows us to account for both wealth and population without over estimating the effects of either covariate.

We chose to investigate all of the provided potential covariates, however we only presented the best versions of the three models with the lowest AIC. So the models in the following section do not account for the count of 'young males' or the measurement of 'unemployment' because they did not significantly improve the model's AIC. We recognize that adding complexity to models will increase the AIC but we chose the simpler models because the negligible change in AIC was not worth the added complexity.

[7] Papachristos, Dr. Andrew. "48 Years of Crime in Chicago: A Descriptive Analysis of Serious Crime Trends from 1965 to 2013." *Yale University: Institution for Social and Policy Studies.* https://doi.org/10.1371/journal.pone.0205151

7 Build Models

We built three Poisson models to account for the socio-economic and environmental factors that we expect to impact the number of burglaries in Chicago. The models were all Poisson models because our data was count data, which indicated that a Poisson distribution would be appropriate. The models are shown below.

Model 1. The first model we built was a Poisson Mixed Effects model. This model included one fixed effect for 'wealth,' one fixed effect for 'precipitation,' an offset to account for changes in 'population,' and one random effect to account for the unique aspects of each census block. In Model 1, $i = census\ block$ and $j = month$ where x_{1i} : $wealth$, x_{2j} : $weather$, and p_i : $population$. Additionally, $Y_{ij} \sim Po(\lambda_{ij})$ and $\gamma_i \sim N(0, \sigma^2_\gamma)$ The model and its estimates are shown below:

$$log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1i} + log(p_i) + \gamma_i + \beta_2 x_{2j}$$

Model 1			
Fixed Effects	Estimate	Standard Error	P-value
Intercept	-5.196476	0.032555	$< 2e-16$
Wealth	-0.226501	0.022912	$< 2e-16$
Precipitation	0.025157	0.003174	$2.26e-15$
Random Effects	Variance	Standard Deviation	
γ_i	0.273	0.5225	

Table 1: This table contains the values for the fixed and random effects in Model 1

Model 2. The second model we produced was a Poisson Random Effects model. This model included a fixed effect for 'wealth,' a random effect for block during each month, a random effect for block, a random effect for month, and an offset to account for the 'population.' In Model 2, $i = census\ block$ and $j = month$ where x_{1i} : $wealth$, ν_j : $unique\ aspects\ of\ month\ j$, ϵ_{ij} : $unique\ aspects\ of\ each\ block\ i\ during\ month\ j$, γ_j : $unique\ aspects\ of\ block\ i$, and p_i : $population$. Additionally, $Y_{ij} \sim Po(\lambda_{ij})$, $\nu_i \sim N(0, \sigma^2_\nu)$, $\epsilon_i \sim N(0, \sigma^2_\epsilon)$ and $\gamma_i \sim N(0, \sigma^2_\gamma)$. Model 2 and its estimates are shown below:

$$log(\lambda_{ij}) = \beta_0 + \beta_1 x_{1i} + \nu_j + \epsilon_{ij} + \gamma_i + log(p_i)$$

Model 2			
Fixed Effects	Estimate	Standard Error	P-value
Intercept	-5.196476	0.032555	$< 2e-16$
Wealth	-0.226501	0.022912	$< 2e-16$
Random Effects	Variance	Standard Deviation	
ν_i	0.03251	0.1803	
ϵ_{ij}	0.02964	0.1722	
γ_i	0.27089	0.5205	

Table 2: This table contains the fixed and random effects for Model 2

Model 3. The third model we built was another Poisson Mixed Effects model. This model contained a random effect for block, a fixed effect for block, and a random effect to capture the unique aspect of 'wealth' on block. In contrast to Model 1, this model assumes that 'wealth' impacts different blocks differently. In Model 3, $i = census\ block$ and $j = month$ where x_{1i} : $wealth$, ν_{0i} : $unique\ aspects\ of\ block$, ν_{1i} : $unique\ aspects\ of\ wealth\ given\ block\ i$, and p_i : $population$. Additionally, $Y_{ij} \sim Po(\lambda_{ij})$, $\nu_{0i} \sim N(0, \sigma^2_{\nu_{0i}})$, and $\nu_{1i} \sim N(0, \sigma^2_{\nu_{1i}})$. The model and its estimates are shown below:

$$log(\lambda_i) = \beta_0 + (\beta_1 + x_{1i})x_i + log(p_i)$$

Model 3			
Fixed Effects	Estimate	Standard Error	P-value
Intercept	-5.02861	0.02223	$< 2e-16$
Wealth	-0.20995	0.02776	$3.97e-14$
Random Effects	Variance	Standard Deviation	
ν_{0i}	0.20498	0.4527	
ν_{1i}	0.07362	0.2713	

Table 3: This table contains the fixed and random effects for Model 3

8 Select Models

After we built our three models, we had to select the best one. Model 1, Model 2, and Model 3 were all nested models. Meaning that the most complex model contains the simplest model inside its structure. We used AIC to compare the models. The AIC for Model 1, Model 2, and Model 3 are 34815.7, 33465.6, and 34965.4 respectively. The AIC was the lowest for model 2, which is why we determined that model 2 was the best model. Model 2 best captured the effects of socio-economic and environmental factors given to us. We are aware that AIC increases with complexity. So some of the increase in AIC is due to the added complexity of larger models. However, we still chose model 2 as our best model because AIC is our best method of comparing effectiveness of the models we produced.

9 Discussion

Analyzing our chosen model, Model 2, we determine that 'wealth' has a negative effect on the burglary rate in a given Chicago census block. Some limitations to our work include a failure to account for the effect that both the number of 'young males' and the measurement of 'unemployment' had on the burglary rates in Chicago. We had originally intended to include 'young males' and 'unemployment' in our models. However, incorporating these covariates did not decrease the AIC significantly enough for us to include them in our findings.

For practical applications, this information can be used to determine better methods of allocating law enforcement within their patrolling zones. For example, there may be more efficient ways to distribute police forces in the winter because we know that crime rates tend to drop in the winter. Additionally, if crime rates spike in the summer, Chicago can adjust their law enforcement methods by increasing the number of patrols. If we can predict higher burglary rates in specific census blocks, then Chicago may take narrow its focus of crime prevention. This allows policy makers or law enforcement to determine how to more effectively allocate resources such as police patrols.

For future work, we recommend investigating why they burglary rates increase in the summer. We wonder if it is too cold to commit crimes in the winter? Or is it because tourism increases in the summer? Or because homeowners are gone on vacation?

Ethical issues that may arise when dealing with this form of data include generalization of results to a population that it cannot not be generalized to. For example, suppose we stated that there are higher burglary rates amongst the least wealthy blocks. The intention of our research is to provide findings that allow policy makers to make informed decisions that lead to decreased crime rates in Chicago. But however, if we explain our results poorly the legislatures my misinterpret our findings. Say someone generalizes our conclusions to say that poor people commit the most burglaries suggest that law enforcement must target or arrest more of the poorer community. That would be an unethical misinterpretation of our data. It is the responsibility of the researcher to make findings understandable to the average reader. Researchers are ethically required to fully explain and interpret their results in order to prevent unethical adaptions of statistics or conclusions.

10 Appendix A

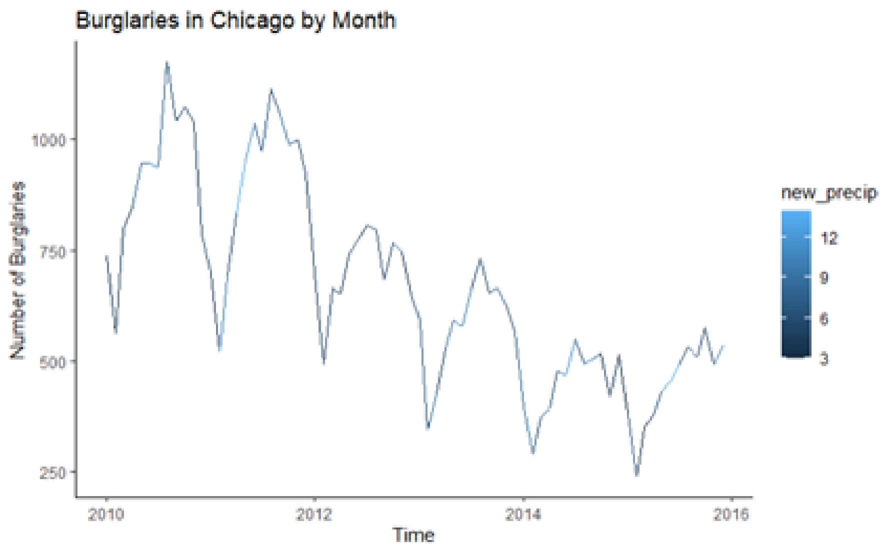


Figure 1: Plot of Burglaries between 2010-2015, colored by Average Monthly Precipitation.

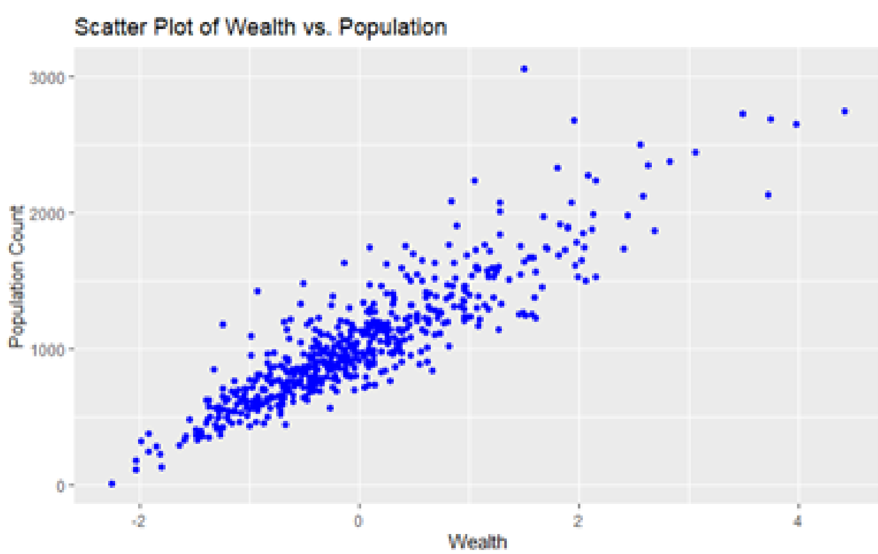


Figure 2: Scatter Plot that shows the colinearity between the covariates 'Wealth' and 'Population.'

Works Cited

Alqahtani A, Garima A, and Alaiad A, "Crime Analysis in Chicago City," *2019 10th International Conference on Information and Communication Systems (ICICS)*, Irbid, Jordan, 2019. West Point, NY. Accessed April 2024. pp. 166-172, doi: 10.1109/IACS.2019.8809142.

Agresti, Alan. *Foundations of Linear and Generalized Linear Models.* First Edition. (Wiley). I referenced the Agresti textbook throughout the semester for the quizzes. West Point, NY. Accessed April 2024.

"Crime Report: Most Dangerous Parts of Chicago." *Cove*. December 23, 2023. West Point, NY. Accessed April 2024.https://www.covesmart.com/blog/most-dangerous-parts-of-chicago-how-to-stay-safe-in-a-red-zone/#:~:text=Rape%2C%20robbery%2C%20and%20assault%20are%20all%20more%20common%20than%20average,a%20victim%20of%20property%20crime

Faraway, Julian. *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. Second Edition. I referenced the Faraway textbook throughout the semester for the quizzes. West Point, NY. Accessed April 2024.

Hendershot, Steve. "The Inequalities of safety: All big cities have a violence problem. Chicago's is different." *Crain's Chicago Business*. October 24, 2022. West Point, NY. Accessed April 2024. https://www.chicagobusiness .com/rains-forum-safer-chicago/chicago-violence-problem-debate-safety-inequality

Luo, Jun. "Multi-spatiotemporal patterns of residential burglary crimes in Chicago: 2006-2016." I*SPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4* (2017): 193. West Point, NY. Accessed April 2024. https://doi.org/10.5194/isprs-annals-IV-4-W2-193-2017.

Papachristos, Dr. Andrew. "48 Years of Crime in Chicago: A Descriptive Analysis of Serious Crime Trends from 1965 to 2013." *Yale University: Institution for Social and Policy Studies*. West Point, NY. Accessed April 2024.

Ruteki, Jared. "Chicago's Murder Count Makes National Headlines, But Accurately Talking About Crime Data Requires Nuance." *WTTW Chicago*. August 31, 2013. Accessed April 2024. https://news.wttw.com/2023/08/31/chicago-s-murder-count-makes-national-headlines-accurately-talking-about-crime-data

Towers S, Chen S, Malik A, Ebert D (2018) "Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective." *PLoS ONE* 13(10): e0205151. West Point, NY. Accessed April 2024. https://doi.org/10.1371/journal.pone.0205151

R Notebook

```r
library(faraway)
library(tidyverse)
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3

## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(knitr)
library(ggplot2)
library(nnet)
library(broom)
library(ggrepel)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.3.3

## Loading required package: Matrix

## Warning: package 'Matrix' was built under R version 4.3.3

##
## Attaching package: 'Matrix'
##
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
clean_crime <- read_csv(file= paste("clean_crime.csv", sep = ""))
```

```
## Rows: 6624 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (10): block, month_burg, wealth, ym, unemp, pop, precip, ym_pop, wealth_...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

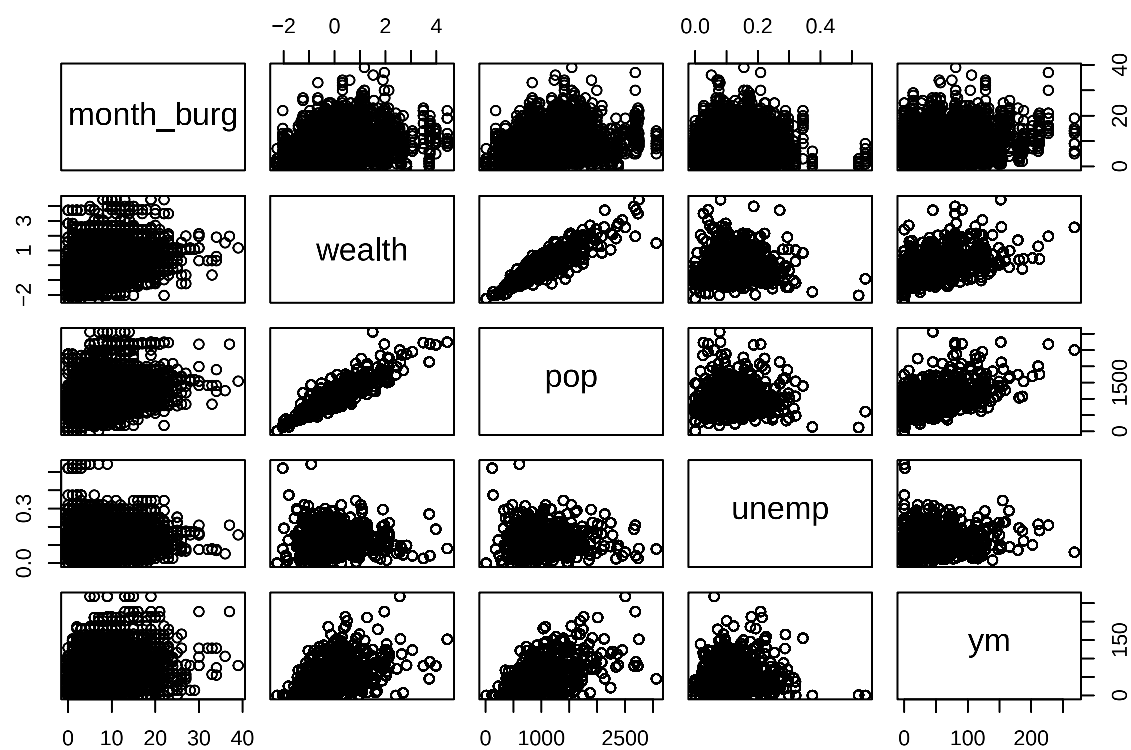
```r
crime <- read_csv(file= paste("crime.csv", sep = ""))
```

```
## Rows: 552 Columns: 72
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (72): 201001, 201002, 201003, 201004, 201005, 201006, 201007, 201008, 20...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
percip <- read_csv(file= paste("percip.csv", sep = ""))
```

```
## Rows: 72 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (2): month_year, m_precip
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
pairs(month_burg ~ wealth + pop+ unemp + ym , data = clean_crime)
```



```r
# mod1 <- multinom(total_crime ~ unemp + pop + ym + wealth, combined)

# Poisson Regression Model

# mod1 <- glm(total_crime ~ pop + unemp + ym +wealth, data = combined, family = poisson)
# summary(mod1)
```

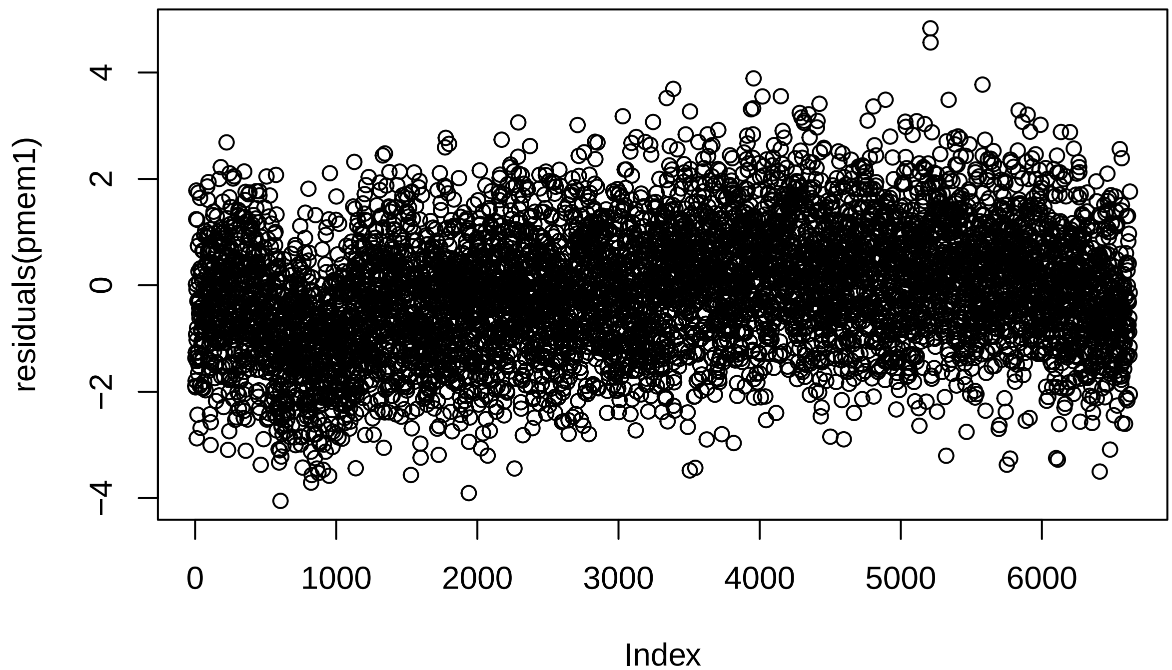
```r
# working <- clean_crime %>%
#   group_by(block)%>%
#   summarize(count = n())
#
# view(working)

pmem1 <- glmer(month_burg ~ wealth +  offset(log(pop)) + (1|block) + precip, family = poisson, data = cl

summary(pmem1)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: month_burg ~ wealth + offset(log(pop)) + (1 | block) + precip
##    Data: clean_crime
##
##      AIC      BIC   logLik deviance df.resid
##  34815.7  34842.9 -17403.9  34807.7     6620
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2253 -0.8359 -0.1096  0.6828  5.7788
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  block  (Intercept) 0.273    0.5225
## Number of obs: 6624, groups:  block, 552
##
## Fixed effects:
##              Estimate Std. Error  z value Pr(>|z|)
## (Intercept) -5.196476   0.032555 -159.619  < 2e-16 ***
## wealth      -0.226501   0.022912   -9.886  < 2e-16 ***
## precip       0.025157   0.003174    7.926 2.26e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr) wealth
## wealth -0.006
## precip -0.712  0.000
```

```r
plot(residuals(pmem1))
```



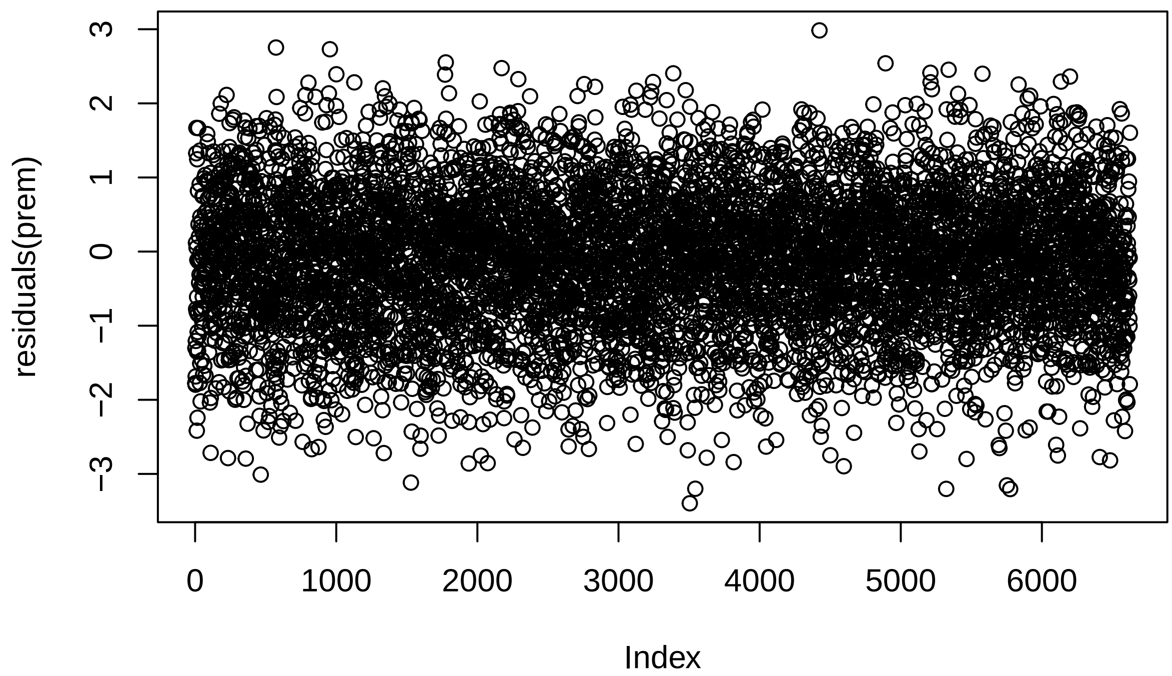

```r
prem <- glmer(month_burg ~ wealth + (1|month) + (1|block:month) + (1|block) + offset(log(pop)), family =

summary(prem)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula: month_burg ~ wealth + (1 | month) + (1 | block:month) + (1 |
##     block) + offset(log(pop))
##    Data: clean_crime
##
##      AIC      BIC   logLik deviance df.resid
##  33465.6  33499.5 -16727.8  33455.6     6619
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5196 -0.6625 -0.0721  0.5131  3.9678
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  block:month (Intercept) 0.02964  0.1722
##  block       (Intercept) 0.27089  0.5205
##  month       (Intercept) 0.03251  0.1803
## Number of obs: 6624, groups:  block:month, 6624; block, 552; month, 12
```

```
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.04201    0.05687 -88.651   <2e-16 ***
## wealth      -0.22631    0.02292  -9.873   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## wealth -0.004
```

```r
plot(residuals(prem))
```

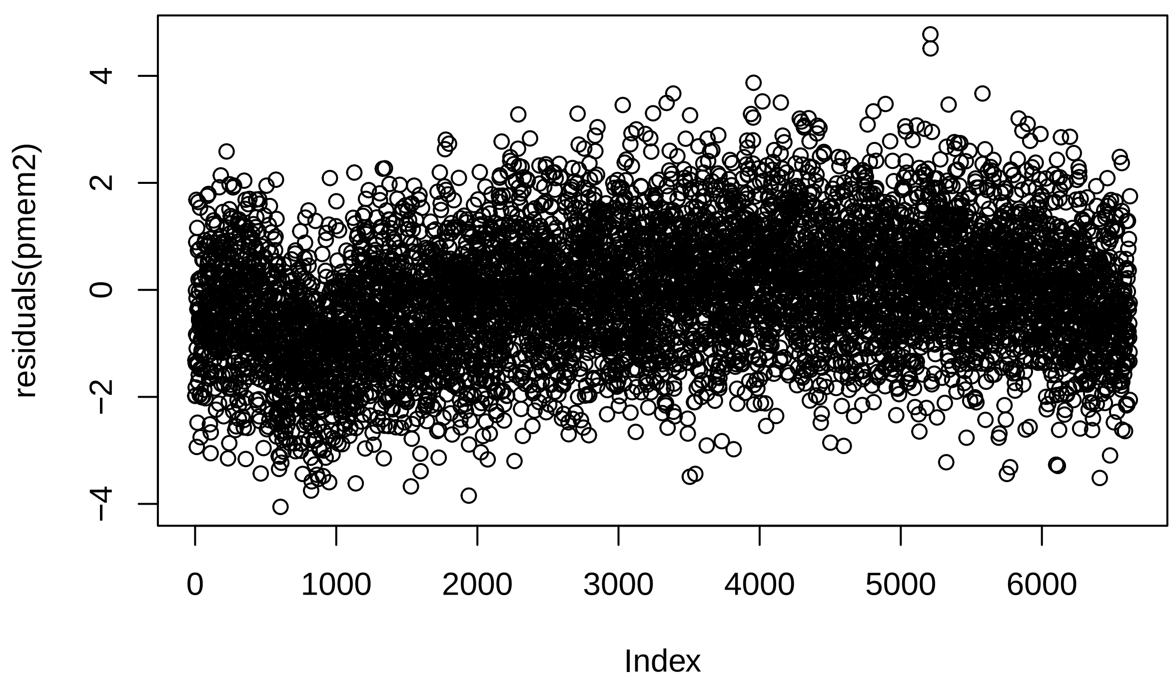



```r
pmem2 <- glmer(month_burg ~ (1|block) + (0+wealth|block)+wealth + offset(log(pop)), family = poisson, da
# COL Clark also mentioned being able to do this: - can choose whichever is easier to describe???
#pmem2 <- glmer(month_burg ~ (wealth|block)+wealth + offset(log(pop)), family = poisson, data = clean_c
summary(pmem2)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: poisson  ( log )
## Formula:
## month_burg ~ (1 | block) + (0 + wealth | block) + wealth + offset(log(pop))
##    Data: clean_crime
```

```
##
##      AIC      BIC   logLik deviance df.resid
##  34848.1  34875.3 -17420.0  34840.1     6620
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2275 -0.8410 -0.1026  0.6965  5.6981
##
## Random effects:
##  Groups  Name        Variance Std.Dev.
##  block   (Intercept) 0.20498  0.4527
##  block.1 wealth      0.07362  0.2713
## Number of obs: 6624, groups:  block, 552
##
## Fixed effects:
##             Estimate Std. Error  z value Pr(>|z|)
## (Intercept) -5.02861    0.02223 -226.188  < 2e-16 ***
## wealth      -0.20995    0.02776   -7.562 3.97e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##        (Intr)
## wealth 0.065
```

```r
plot(residuals(pmem2))
```



```r
crime_long <- crime %>%
  ## select(- "...1") %>%
  # subset(df, select = -"...1") %>%
  summarize(total_crime = rowSums(crime))
```

```
## Warning: Returning more (or less) than 1 row per `summarise()` group was deprecated in
## dplyr 1.1.0.
## i Please use `reframe()` instead.
## i When switching from `summarise()` to `reframe()`, remember that `reframe()`
##   always returns an ungrouped data frame and adjust accordingly.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
#view(crime)
crime_plot <- ggplot(crime_long, aes(x = month, y = total_count, group = year)) +
  geom_line() +
  facet_wrap(~ year) + # Facet wrap by year
  theme_minimal() +
  labs(x = "Month", y = "Total Count of Burglaries", title = "Monthly Crime Count by Year") #+
  #theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1)) # Rotate x-axis labels if needed
```

```r
# crime <- crime
# #view(crime)
# # population by census block group
# pop <- pop
# # centered and scaled average family income by census block group (2015 dollars)
# wealth <- wealth
# names(wealth)[names(wealth) == "...1"] <- "stand_wealth"
# df <- data.frame(wealth = wealth)
# names(df)[1] <- "wealth"
# view(wealth)
# # number of young males by census block group (15-20 yr olds)
# ym <- ym
# # percentage unemployed by census block group
# unemp <- unemp
```

```r
library(tidyverse)
library(lubridate)


crime_data <- crime


long_crime_data <- crime_data %>%
  pivot_longer(
    cols = everything(),
    names_to = "month_year",
    values_to = "burglaries"
  )
```

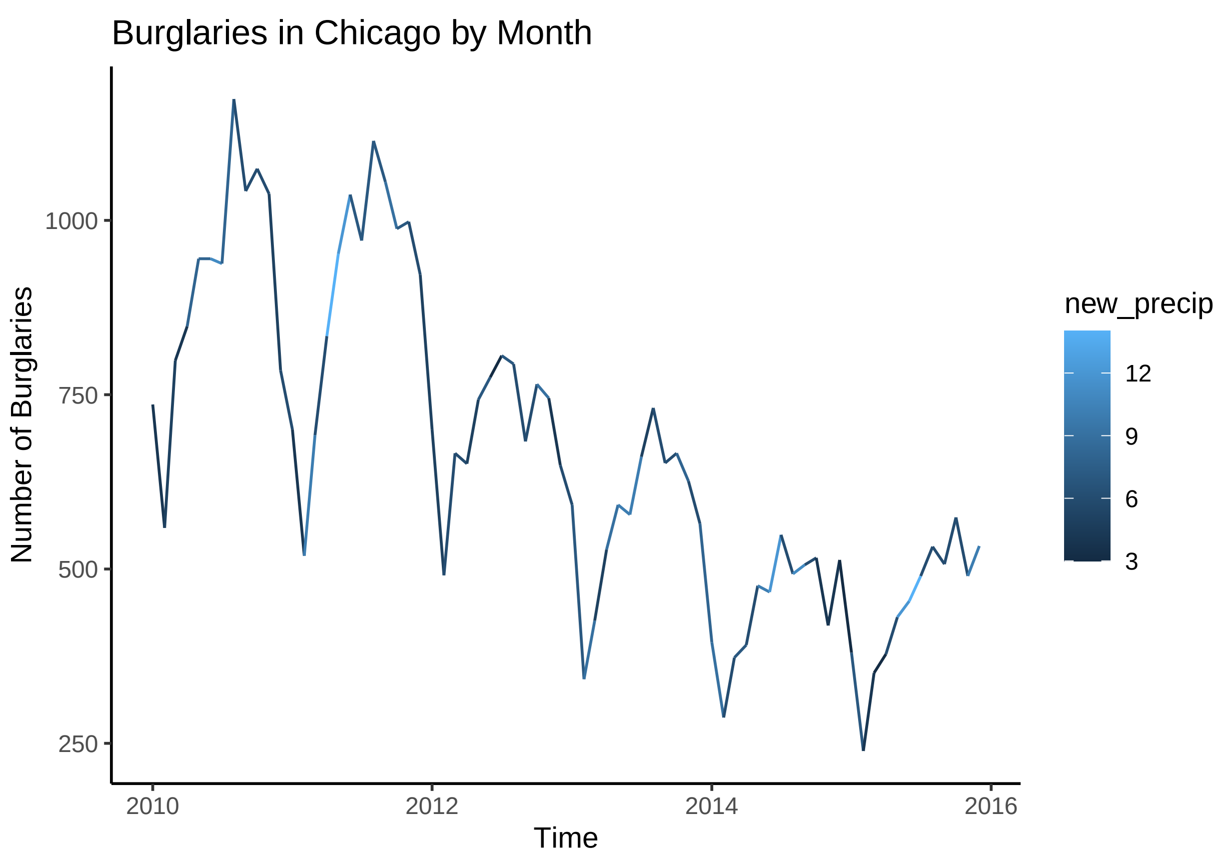
```r
works <- long_crime_data %>%
  group_by(month_year) %>%
  summarize(burglaries = sum(burglaries))

percip$month_year <- as.character(percip$month_year)
works <- left_join(works,percip, by = "month_year")
works$month_year <- ym(works$month_year)
works$month_year <- as.Date(works$month_year)

works <- works %>%
  mutate(new_precip = percip$m_precip)
# percip$month_year <- ym(percip$month_year)
# percip$month_year <- as.Date(percip$month_year)




ggplot(works, aes(x = month_year, y = burglaries, color = new_precip)) +
  geom_line()+
  labs(title = "Burglaries in Chicago by Month",
       x = "Time",
       y = "Number of Burglaries") +
  theme_classic()
```



```r
percip$month_year <- as.character(percip$month_year)
works$month_year <- ym(works$month_year)
```

```
## Warning: All formats failed to parse. No formats found.
```

```r
works$month_year <- as.Date(works$month_year)
percip$month_year <- ym(percip$month_year)
percip$month_year <- as.Date(percip$month_year)
works <- left_join(works,percip, by = "month_year")
```