UNITED STATES MILITARY ACADEMY
WEST POINT®

MA478: Final Project

CDT Isabella Palchak
CDT Karly Parcell

25 April 2024

Problem Statement

Develop and choose a model that best captures the effects of…

- population
- unemployment
- wealth
- count of young males
- monthly precipitation

… on the number of burglaries by census block in Chicago.

Chicago Data

- Spatial data that consists of 552 census blocks
- *Population*: Total population in each census block
- *Unemployment*: measurement of unemployment in each block
- *Wealth*: measurement of wealth in each block
- *Young Males*: count of young males in each block
- *Precipitation*: Average monthly precipitation 2010-2015
- *Number of burglaries:* Count of burglaries by month & block

- Transformation: wide → long

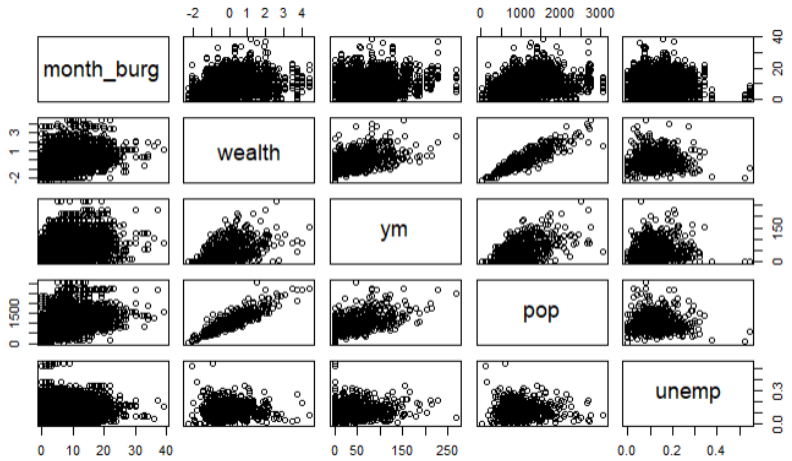
Research Question

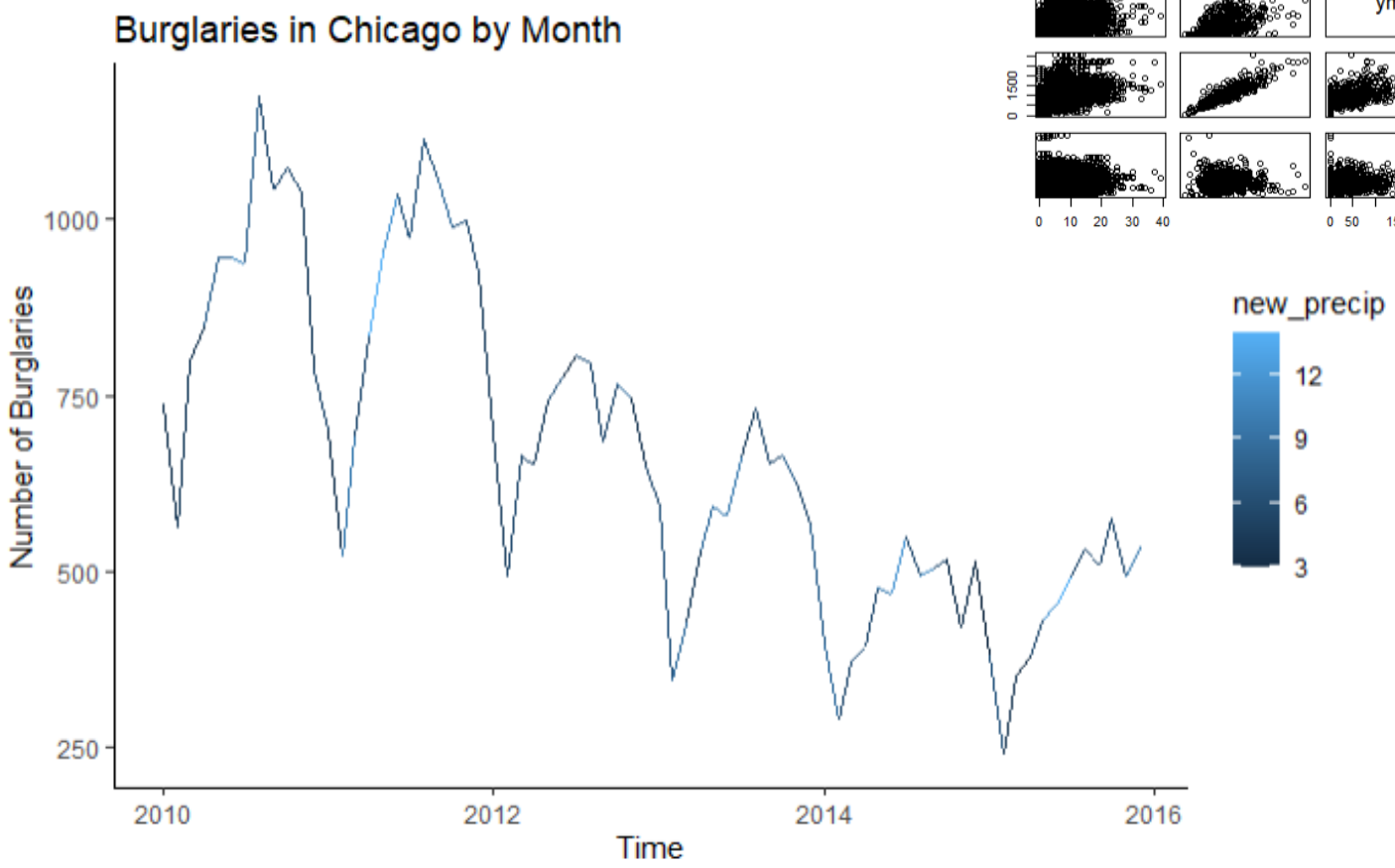
What are the effects of socio-economic and environmental factors on the number of burglaries by census block in Chicago?

- We suspect that high precipitation levels will be linked to fewer burglaries.
- We suspect that burglaries will be more prevalent in blocks with higher levels of wealth

Data Exploration

Pairs Plot





Methodologies

Clean data

- Condense the crime, pop, uemp, ym, wealth, precipitation
- Transform month_year data into one column
  (Twelve rows for each census block)

Explore data

- Investigate the covariates
- Identified spike in warmer months, decreasing burglary trend

Create models

- Poisson Mixed Effects #1 (effect is the same for each block)
- Random Effects Model
- Mixed Effects #2 (effect differs by block)

Explain models

- Context of the problem

Choose the best model

- AIC to choose the best model

Challenges

- Transforming the Data
  - Condense the month_year columns into a single column

- Filtering precipitation
  - Set threshold for precipitation $\geq$ 0.07
  - Considered “light rain”
  - Previously showed that all days had precipitation

Poisson Mixed Effects Model

$i = block$
$j = month$

$x_{1i} : wealth$
$x_{2j} : weather$
$p_i : population$

$Y_{ij} \sim Po(\lambda_{ij})$
$\gamma_i \sim N(0, \sigma_\gamma^2)$

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

	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

	Variance	Standard Dev.
γ_i	0.273	0.5225

MATHEMATICAL SCIENCES

Poisson Random Effects Model

$i = block$
$j = month$

$Y_{ij} \sim Po(\lambda_{ij})$

x_{1i}: $wealth$
p_i : $population$
υ_j : $unique\ aspects\ of\ month\ j$
ε_{ij} : $unique\ aspects\ of\ each\ block\ i\ during\ month\ j$
γ_i : $unique\ aspects\ of\ block\ i$

$$\log(\lambda_i) = \beta_0 + \beta_1 x_{1i} + \upsilon_j + \varepsilon_{ij} + \gamma_i + \log(p_i)$$

	Estimate	Standard Error	P-value
Intercept	-5.196476	0.032555	< 2e-16
Wealth	-0.226501	0.022912	< 2e-16

	Variance	Standard Dev.
υ_j	0.03251	0.1803
ε_{ij}	0.02964	0.1722
γ_i	0.27089	0.5205

MATHEMATICAL SCIENCES

Poisson Mixed Effects Model #2

$i = block$
$j = month$

$Y_i \sim Po(X_i)$
$\log(\lambda_i) = \eta_i$

$x_i : wealth$
$p_i : population$

$\upsilon_{0i} \sim N(0, \sigma^2_{\upsilon_{0i}})$
$\upsilon_{1i} \sim N(0, \sigma^2_{\upsilon_{1i}})$

$$\eta_i = \beta_0 + \upsilon_{0i} + (\beta_1 + \upsilon_{1i}) x_i + \log(p_i)$$

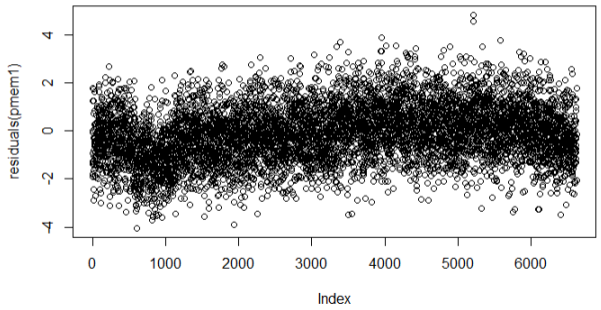
	Estimate	Standard Error	P-value
Intercept	-5.01756	0.02493	< 2e-16

	Variance	Standard Dev.
υ_{0i}	0.09318	0.3052
υ_{1i}	0.23437	0.4841

MATHEMATICAL SCIENCES

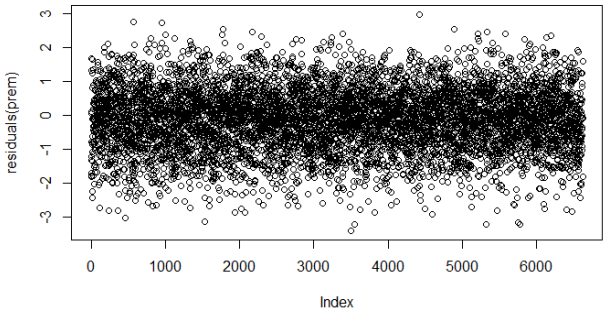
Model Selection

Model 1
(Poisson mixed-effects model #1)



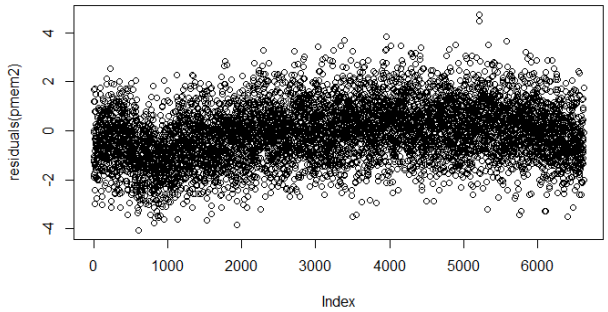
AIC : 34815.7

Model 2
(Poisson random effects model)



AIC : 33465.6

Model 3
(Poisson mixed-effects model #2)



AIC : 34965.4

MATHEMATICAL SCIENCES

Limitations, Assumptions, Future Work

- Failed to account for the count of young males
  - Tried to adjust the young male count for the total population
  - ym is likely a subset of the population
- Failed to account for unemployment
  - Negligible decrease in AIC was not worth the added complexity
- We do not know how wealth is measured

- Future work:
  - We want to investigate why burglary rates increase in the summer. Is it too cold? Too snowy? Are people not vacationing during the winter?

"So What?"

- If trends in burglary rates are identified and predictable, policy can be implemented to mitigate the issue.

- For example:
  - If crime rates spike in the summer or drastically drop when its raining, Chicago can adjust law enforcement accordingly.
  - If we can predict higher burglary rates in specific census blocks, then Chicago may take narrow its focus of crime prevention. More effective allocation of resources.

Works Cited

- **Break-in Increase in the Dry Season, Robbery Increase in the Rainy Season**
  - Afon, A. O., & Badiora, A. I. (2018). The Dynamics of Crime Opportunities: Evidences from Weather Conditions and Spatial Pattern of Residential Neighborhood in Ibadan, Nigeria. *Papers in Applied Geography*, 4(1), 1-20. https://doi.org/10.1080/23754931.2017.1346520

- **Possible Explanations for Increased Burglaries in the Summer**
  - https://policyadvice.net/insurance/insights/burglary-statistics/

- **What is Considered Precipitation**
  - https://windy.app/blog/how-do-we-measure-precipitation.html#:~:text=As%20a%20general%20rule%20in%20meteorology%20and%20weather,1%20square%20meter%20%2810.7%20sq%20foot%29%20of%20area.