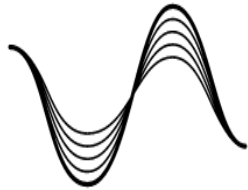# Using Open Data for Local Policy Making

Lessons from DSSG

*Jonathan Auerbach*
*New York City Council*

# Data Science for Social Good



Data Science for Social Good





Goal of Open Data is to "empower Chicago's residents by providing them with information necessary to participate in government in a meaningful manner" and "assist in identifying possible solutions to pressing governmental problems."
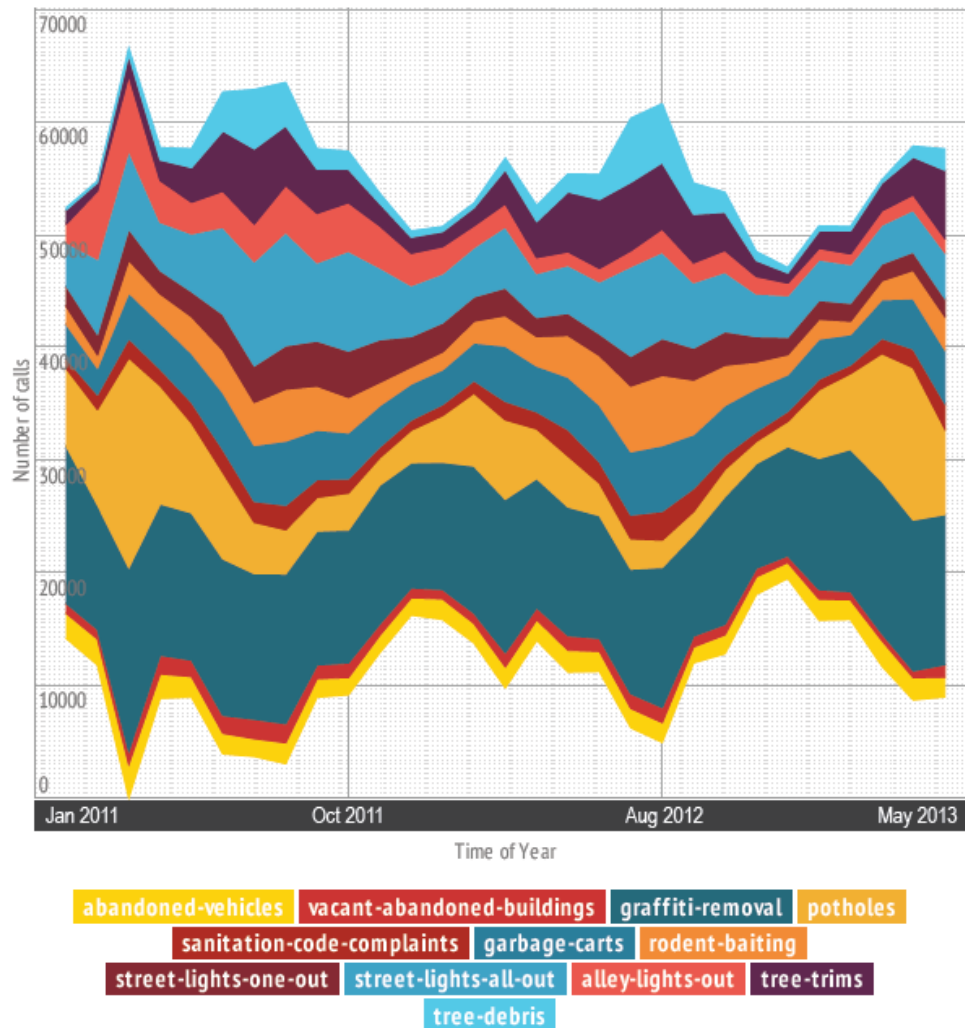
-Mayor Emanuel

# Chicago's Data Portal
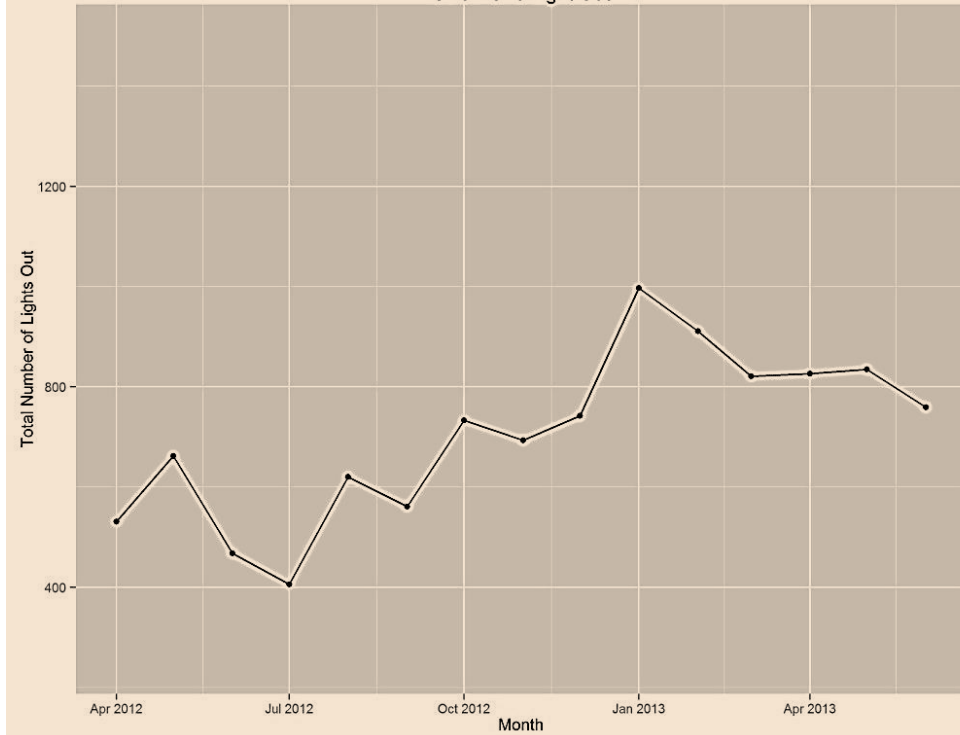
▸ Contains 200+ open datasets

▸ Open data is transactional data
   ▸ Transactional Data != Survey Data
   ▸ Open Portal Data is generally collected for accounting purposes
   ▸ We looked at Customer Service Requests (including 311 Calls) and Crime Data
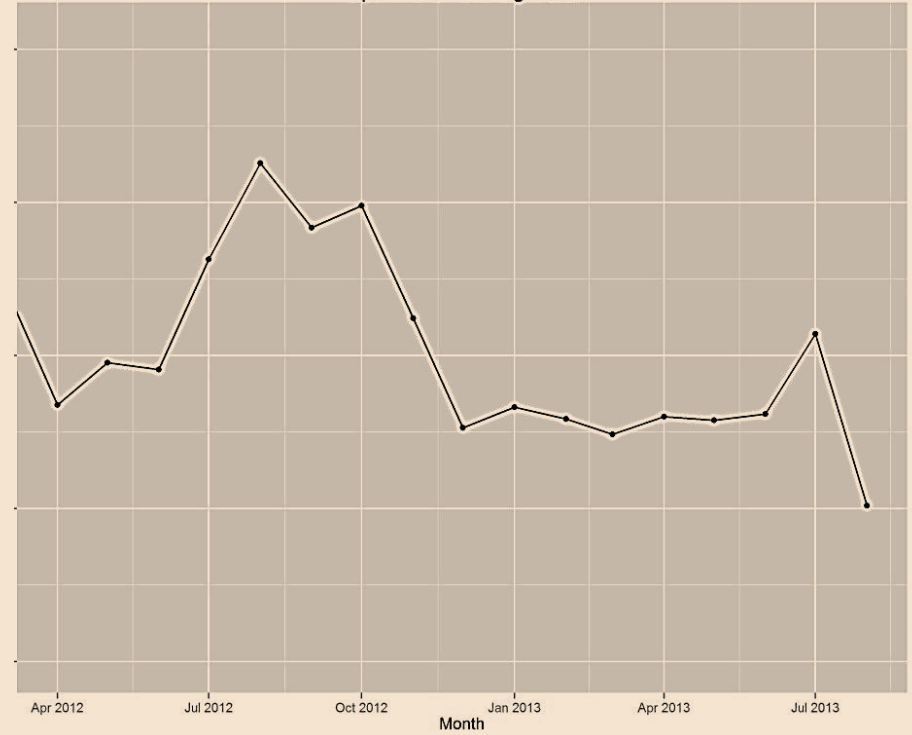
# Customer Service Requests (CSRs)

# Notable Limitations

# So Why Use Open Data in Policy?

- Policy making has been using data for a long time
  - Pension Financing (1920s)
  - Capital Financing (1950s)

- But policy decisions are rarely data driven
  - Parties make data-driven arguments but the advantage in decision making goes to those that have a privileged access to data

- We believed that (despite its limitations) open data can facilitate objective, data-driven decisions

# Empowering the Public

▸ Open data is used most effectively when policy makers use it to include the public more intelligently in decision making.

▸ Government should draw on the talents of the huge, "impartial" data using community.

    ▸ Build visualizations/tools

▸ But first the quality of the data has to be evaluated and documented.

    ▸ Lots of EDA

# Chicago Community Areas
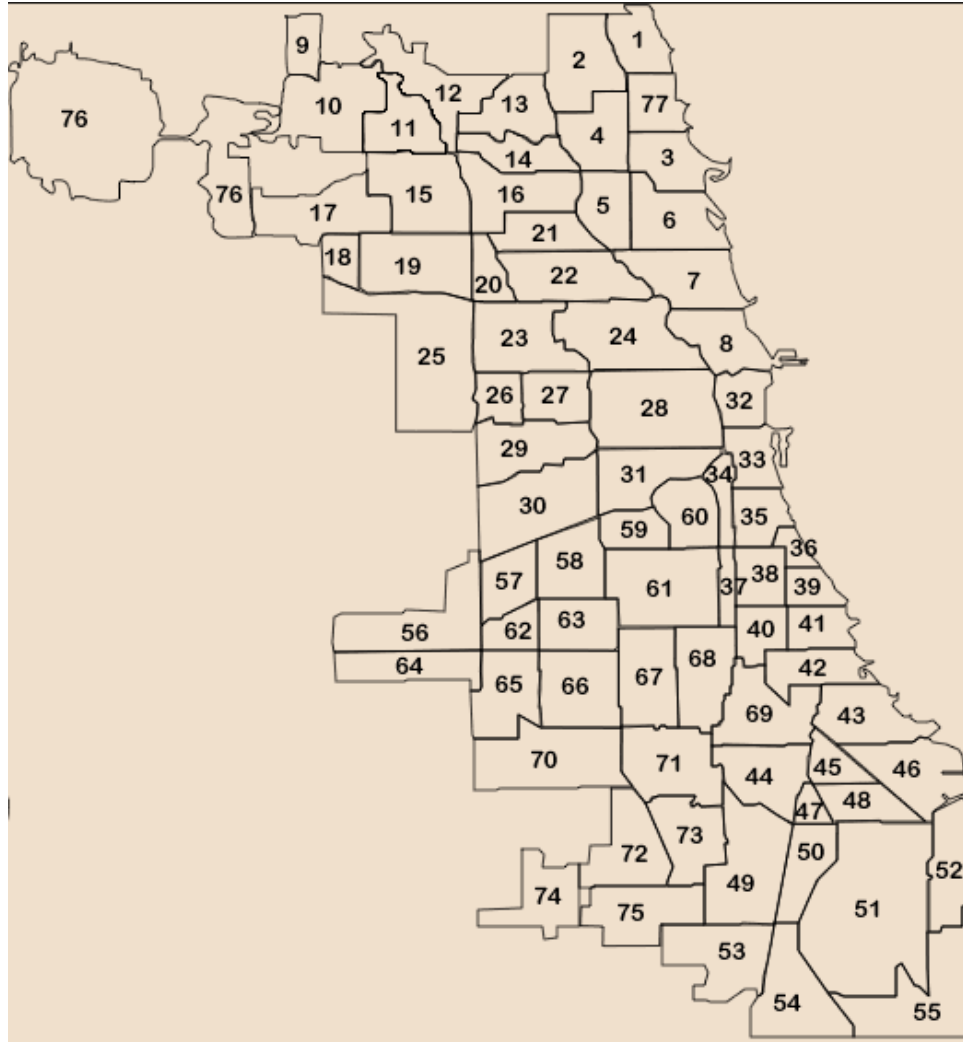
# Quiz: Match the Service Request to its Graph
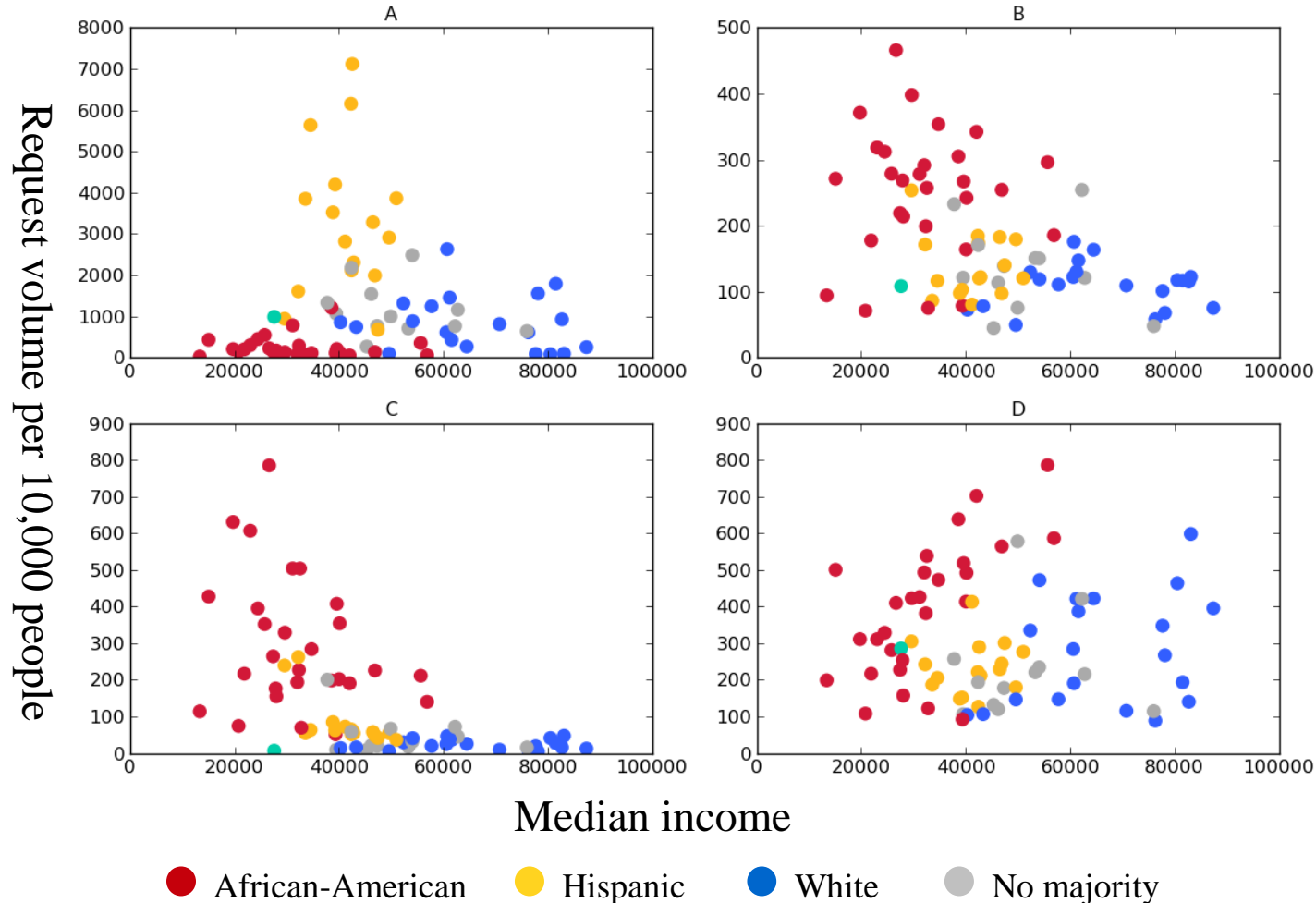


CSRs:

**Abandoned Building Notification**;

**Graffiti Removal Request**

**Sanitation Code Violation**

**Street Lights Out**

Legend:
- African-American (red)
- Hispanic (yellow)
- White (blue)
- No majority (grey)

X-axis: Median income
Y-axis: Request volume per 10,000 people

# Graffiti Removal Requests

# Census Tract Clusters



Cluster 2

Cluster 0
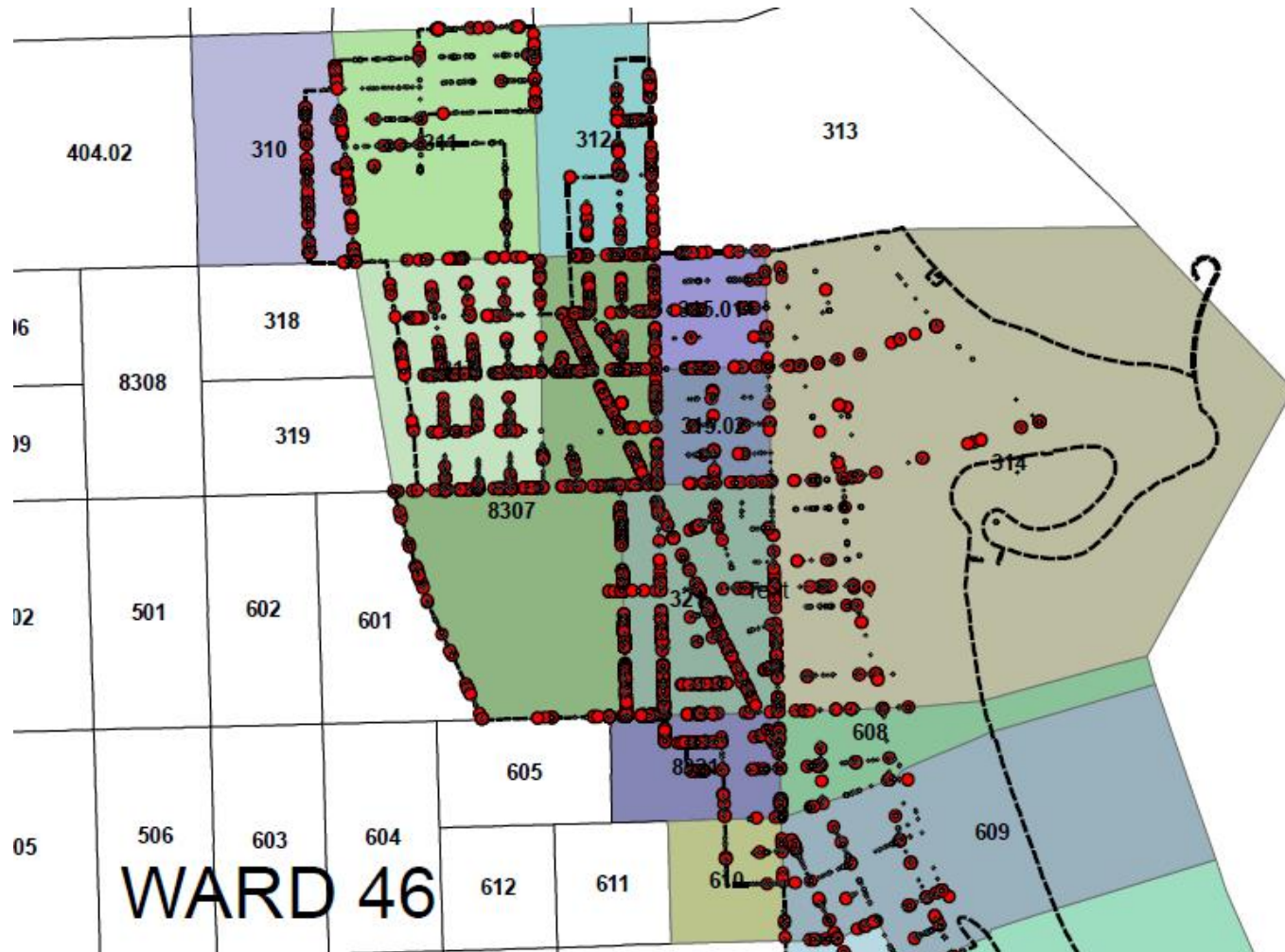
Cluster 1

Cluster 5

Cluster 3

Cluster 4
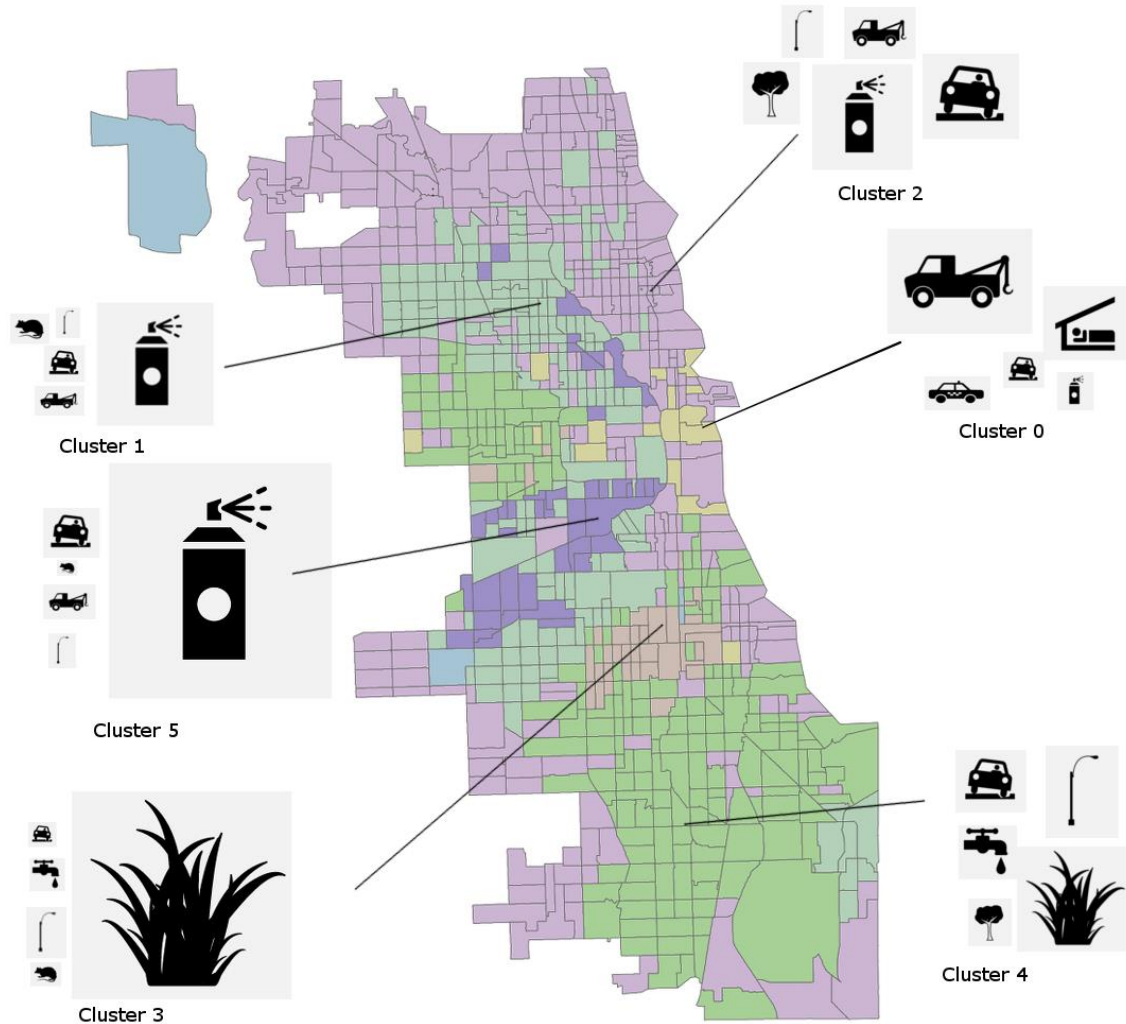
# Neighborhoods Defined by Services



Cluster 2

Cluster 1

Cluster 0

Cluster 5

Cluster 3

Cluster 4

⬤ White  ⬤ Hispanic  ⬤ African-American
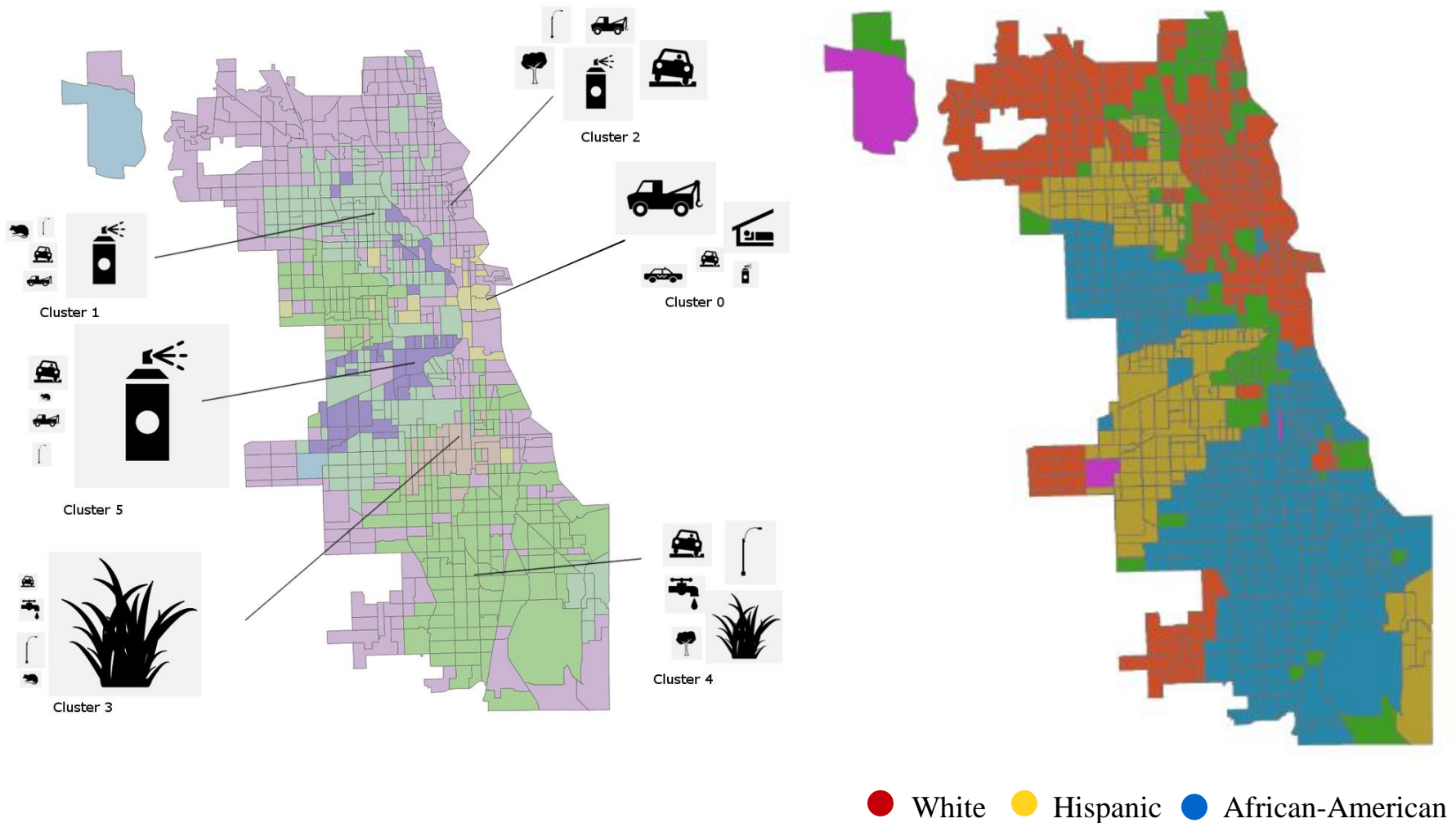
# Select Coefficients of Poisson Generalized Linear Models

| Service Request | Graffiti Removal | Potholes Filing |
|---|---|---|
| Intercept | 2.029 | 1.754 |
| Service Requests (Lag 1) | 0.008 | 0.009 |
| Service Requests (Lag 2) | 0.006 | 0.009 |
| Population (1000s) | 0.074 | 0.158 |
| Fraction 65+ | -3.356 | 1.445 |
| Fraction Black | -1.232 | 0.345 |
| Fraction Hispanic | 0.743 | 0.349 |
| Fraction below Poverty Line | -0.152 | -0.368 |
| Unemp Rate (%) | -0.002 | 0.005 |
| Median Income ($1000s) | -0.001 | 0.005 |

*** all coefficients significant at 1% level

# Do Light Outages Cause Crime?

- Looked at 27 models on the Community Area Level

  - 3 different types of "lights out"

  - 9 different crime types

  - Included various ACS 5-Year demographic and economic characteristics

| Table 2. Crimes by Type | |
|---|---|
| All Crimes (No Deceptive Practice) | 179514 |
| Narcotics | 36978 |
| Theft | 39308 |
| Battery | 33600 |
| Criminal Damage | 22452 |
| Motor Vehicle Theft | 19158 |
| Robbery | 13845 |
| Assault | 12324 |
| Burglary | 1420 |
| Homicide | 429 |
| Deceptive Practice | 3865 |

# Crime and Time until Completion
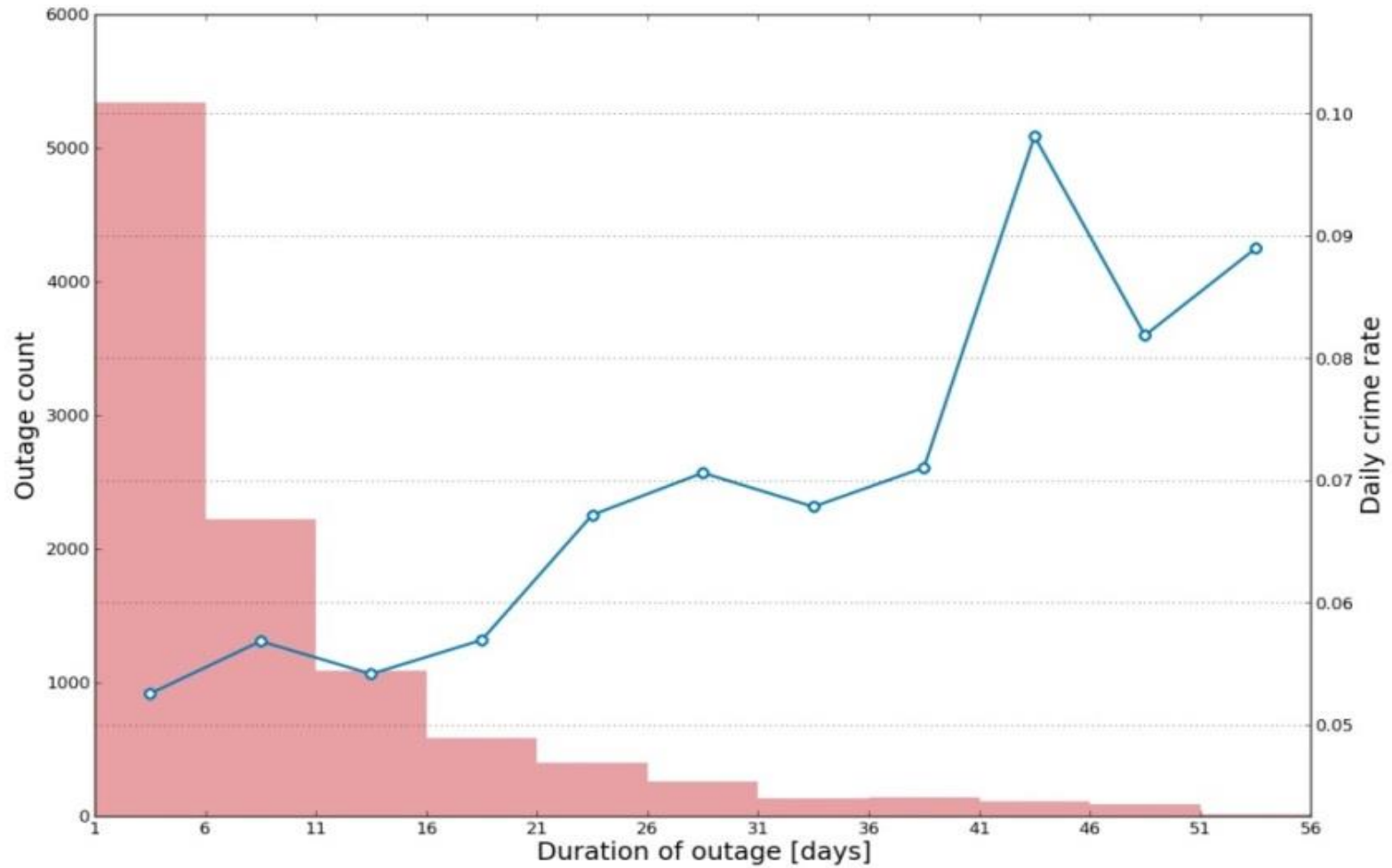
# Lights Out and Robberies

Table 6. Within Area % Difference in
Alley Robbery Rates by Community Area

| Community Area | | % Difference | | |
|---|---|---|---|---|
| Number | Name | % Diff | P Value | Significance |
| 7 | Lincoln Park | 341.9 | 0.001 | ** |
| 22 | Logan Square | 32.5 | 0.360 | |
| 23 | Humboldt Park | 52.1 | 0.094 | |
| 24 | West Town | 25.7 | 0.477 | |
| 25 | Austin | 33.6 | 0.108 | |
| 26 | West Garfield Park | 20.8 | 0.414 | |
| 27 | East Garfield Park | -42.7 | 0.241 | |
| 29 | North Lawndale | -6.6 | 0.816 | |
| 30 | South Lawndale | 49.0 | 0.210 | |
| 42 | Woodlawn | -51.0 | 0.178 | |
| 43 | South Shore | 1.2 | 0.955 | |
| 44 | Chatham | 53.7 | 0.022 | * |
| 46 | South Chicago | 4.3 | 0.879 | |
| 49 | Roseland | -24.3 | 0.275 | |
| 53 | West Pullman | 135.3 | 0.008 | ** |
| 66 | Chicago Lawn | -35.3 | 0.129 | |
| 67 | West Englewood | -31.4 | 0.126 | |
| 68 | Englewood | 127.7 | 0.001 | ** |
| 69 | Greater Grand Crossing | -20.3 | 0.425 | |
| 71 | Auburn Gresham | 52.1 | 0.030 | * |
| 73 | Washington Heights | 1.2 | 0.972 | |

**Significant at 1% level, *Significant at 5% level



**Percent Change**

| | |
|---|---|
| | NA |
| | -50 - 0 |
| | 0 - 50 |
| | 50 - 100 |
| | 100 - 200 |
| | 200 - 350 |