# Chicago-Community Area

## STI Rates and Stats

By Angela Baker

# Where did the data come from?

- These data were obtained from: https://www.healthdata.gov/search/type/dataset?query=STI&sort_by=changed&sort_order=DESC
- http://chicago-zone.blogspot.com/2014/03/chicago-zip-code-map-locate-chicago.html
- https://www.cityofchicago.org/dam/city/depts/cdph/policy_planning/PP_Web%20Health%20Care%20Facilities%20by%20Region.pdf
- Null: That the spread across the Chicago area communities can be represented as a whole and not by community area boundaries.
- Alternative: That the Chicago area communities differ in a statically significant manor.

# What was observed in these data?

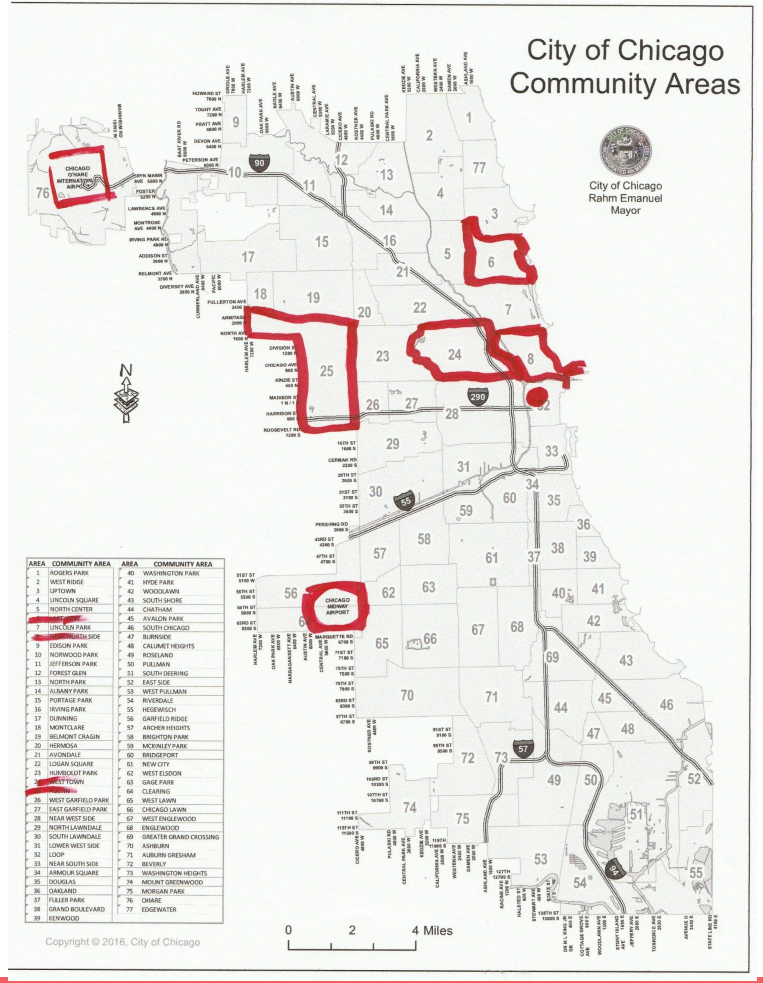
- Looking for patterns within the data
  - Is there a difference between males and females for cases?
  - Does any one community area have more or less cases based on size and access to free clinics?
  - Are there any patterns that can be predicted with Machine Learning?

- Using Python in conjunction with seaborn, matplotlib.pyplot, pandas and numpy

- Small data sets lead to difficult processing

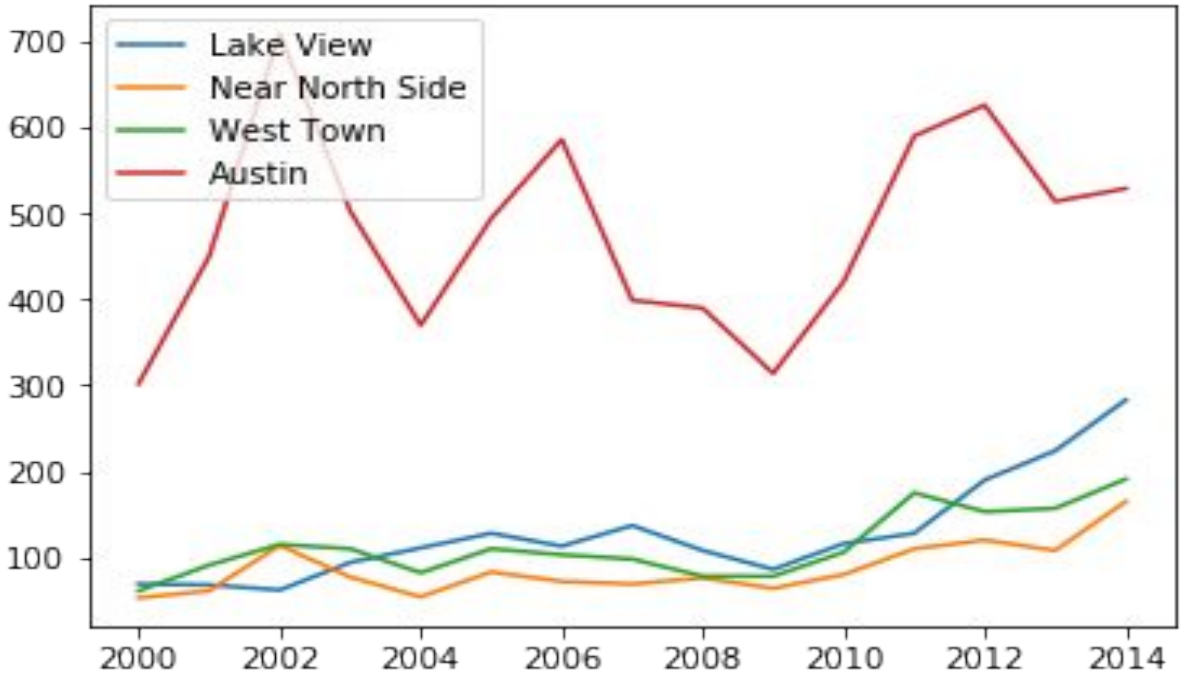
# Let's look at the highest population areas

| Area | Population_2010 | Clinics Available |
|---|---|---|
| Lake View (6) | 94,368 | 3 |
| Near North Side (8) | 80,484 | 2 |
| West Town (24) | 82,236 | 5 |
| Austin (25) | 98,514 | 7 |



# Chlamydia Cases for Males

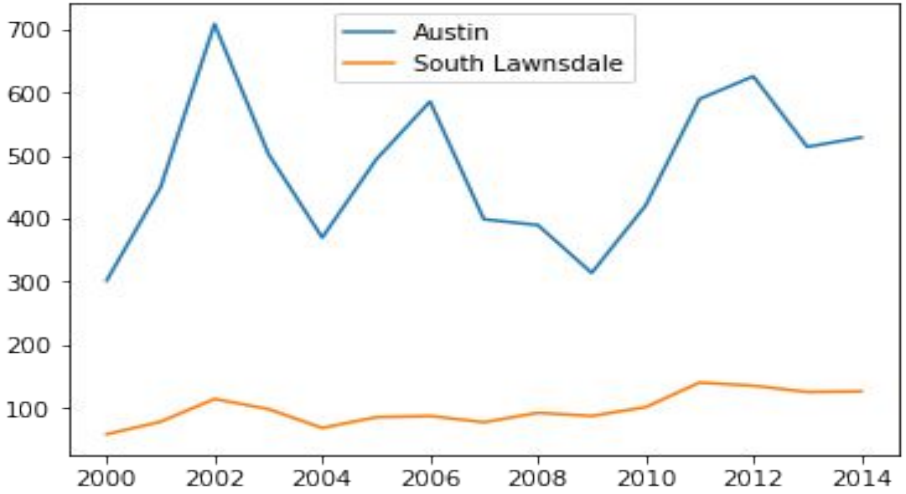
These are the 4 most populated community areas in the Chicago area and are not missing any data



# Taking a closer look part 1

Looking at Austin and South Lawndale who have roughly equal land mass away from the city center and equal access to free clinics.

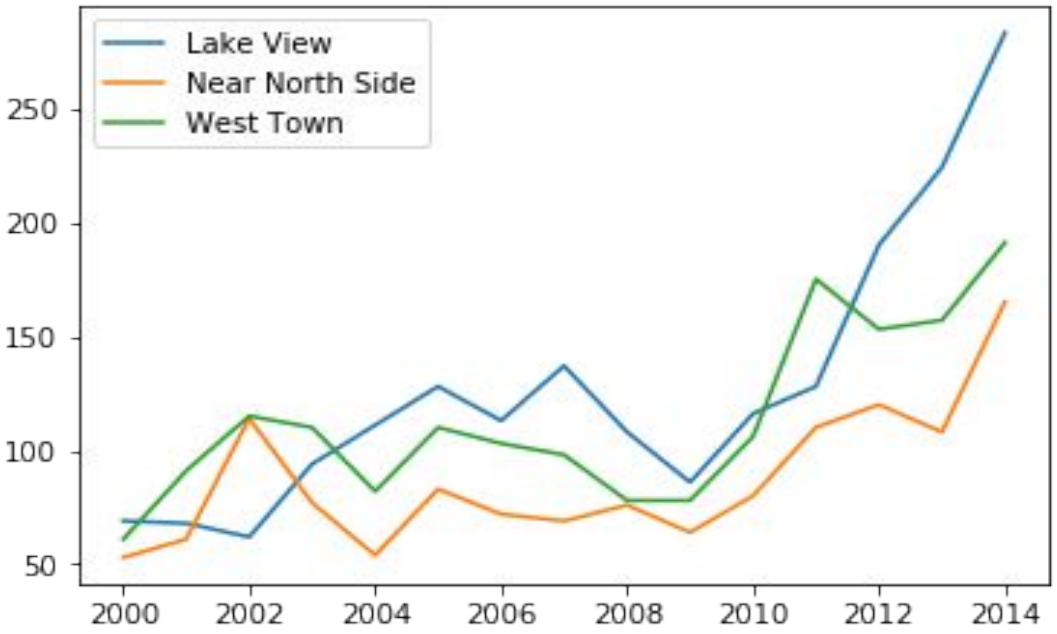
Austin still seems to be an outlier with almost 6x as many cases on some years.



# Closer look pt 2

Taking Austin out of the picture and only looking at the other populations over 80,000 we can see that there has been a drastic raise since 2010.

In almost every community area of Chicago there was a dip in population between 00-10 but there isn't any census data since then.



# Problems with the Data

- Small Data Set
  - Very Few complete columns and rows
  - Predictable trends are only visual
- No real categories
  - Making educated guesses about how trends will continue or drop is not statistically different
  - Also no real way to categorize new data
- Next time:
  - Find a bigger data set and let it do the taking instead of looking for a problem that data might be able to solve
  - Think about categories and predictive ability before diving into cleaning and wrangling for a data set that may not give any useful models

# Conclusion

While the Data taught allowed me to try many techniques and improve my research skills, it did not necessarily come out with any useful predictive models or categorical data. Noting that the data set was extremely small and therefore had very few points in which to create anything more than what can be seen visually.

While the sets were statistically different from one another and thus the Null Hypothesis : that the spread across the Chicago area communities can be represented as a whole and not by community area boundaries, is rejected and it would be useful to have these cases sorted by month or a longer data set could give more predictive power.