

## STI Cases in the Chicago area, a study from 00-14

This data had many challenges. First and foremost it is a small data set with only 14 years of data collected from 88 key community areas in Chicago. Unable to visualize all the data and looking for the most complete sets to compare I settled on the areas which had over 80,000 people per community area as they were likely to have the best insight. Into how the population as a whole might be predicted. I chose not to compare the lesser populated areas as many areas had missing data and therefore a complete analysis could not be completed. This dataset(s) was obtained from:

[https://www.healthdata.gov/search/type/dataset?query=STI&sort\\_by=changed&sort\\_order=DESC](https://www.healthdata.gov/search/type/dataset?query=STI&sort_by=changed&sort_order=DESC) and was cleaned and wrangled using a multitude of techniques.

First the data was wrangled via Excel, as it is not as foreign to me, however after the visualization module I then went back and cleaned it using Python techniques with pandas and matplotlib.

Represented below are the basic trends seen in Lake View, Near North Side, West Town, and Austin, for both Males and Females in both Chlamydia and Gonorrhea studies.

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.

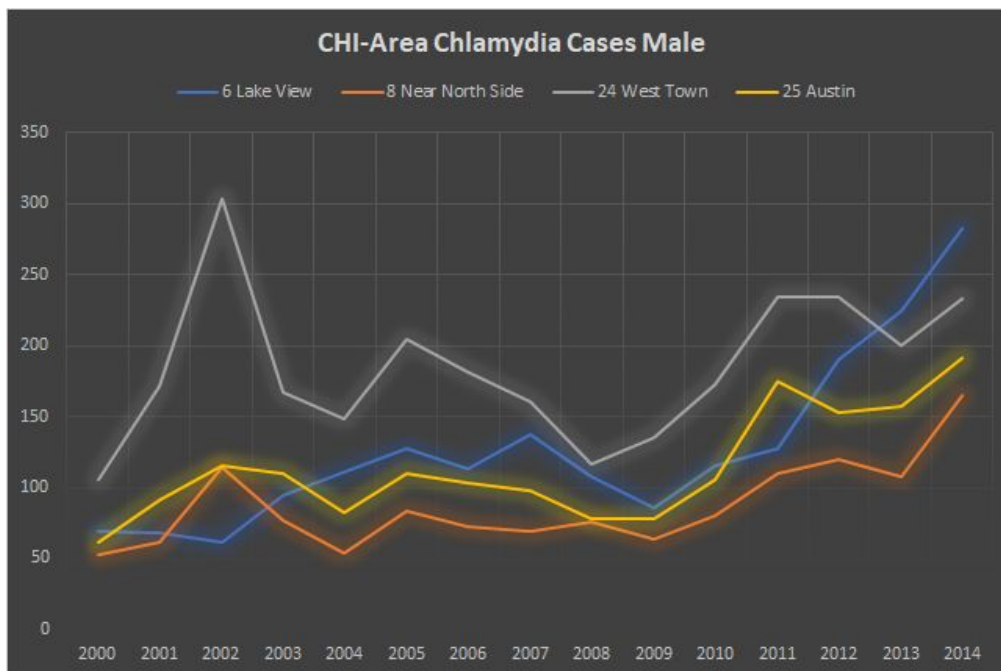


Figure 1. Spike in 2002 for West Town

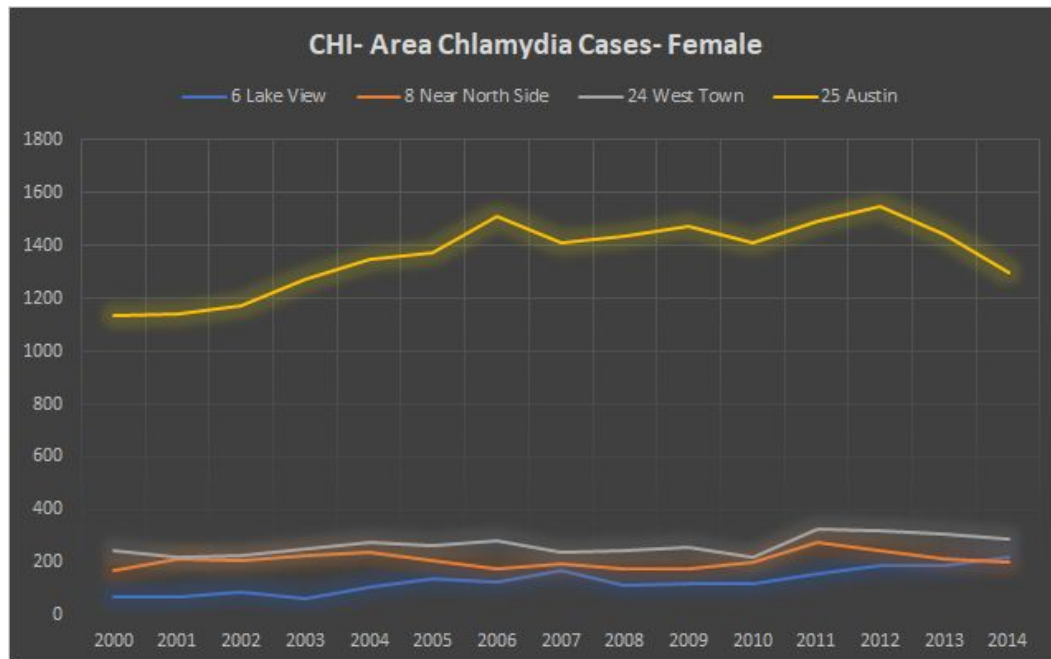


Figure 2. Note Austin's very high numbers compared to the other areas

#### Chlamydia Means

	F	M
Lake View	130.6667	127.8
Near North Side	209	87.06667
West Town	265.6	184.6667
Austin	1364.6	113.8667

Table. 1.

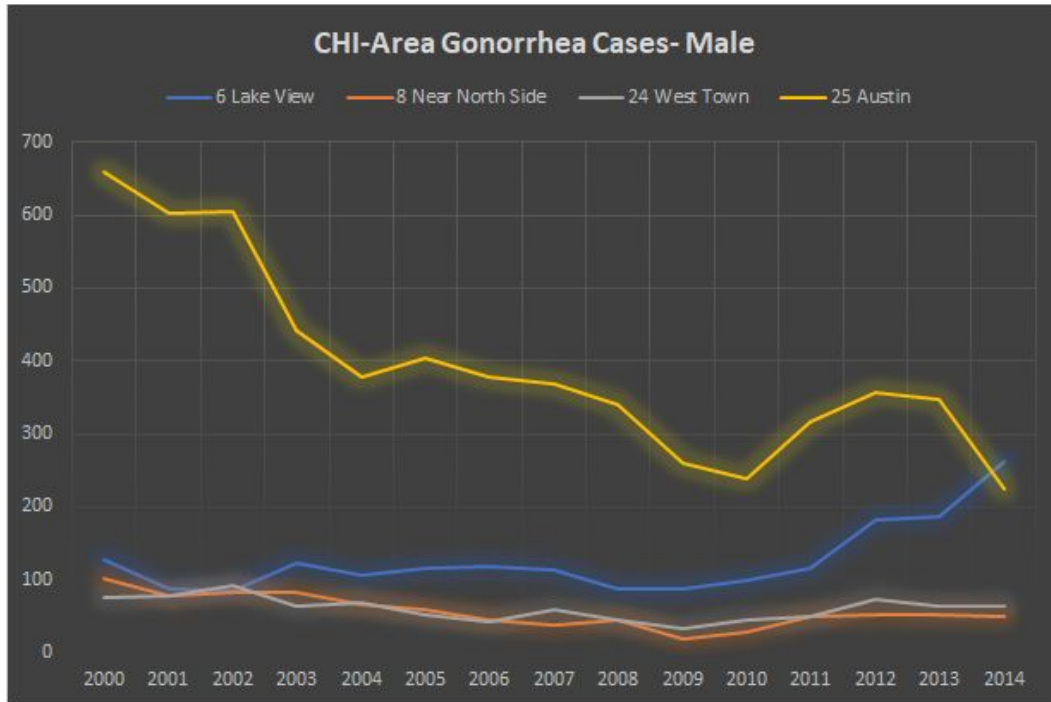


Figure 3. Austin's rate comes down within the other ranges

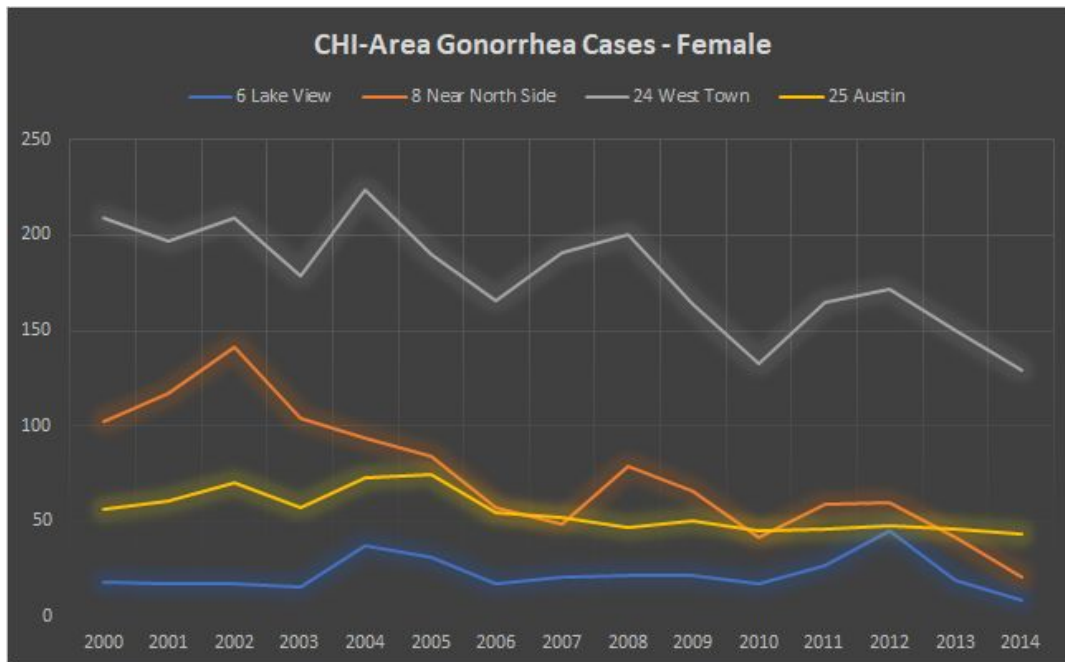


Figure 4.

Gonorrhea			Gonorrhea-BS	
	F	M	F	M
Lake View	22.33333	126.4667	23.06667	132.3333
Near North Side	74.46667	56.8	82.73333	57.13333
West Town	178.5333	60.33333	162.9333	64.53333
Austin	54.93333	395.0667	53.2	406.4667

Table 2. Original data and resampled bootstrap data- 1 sample

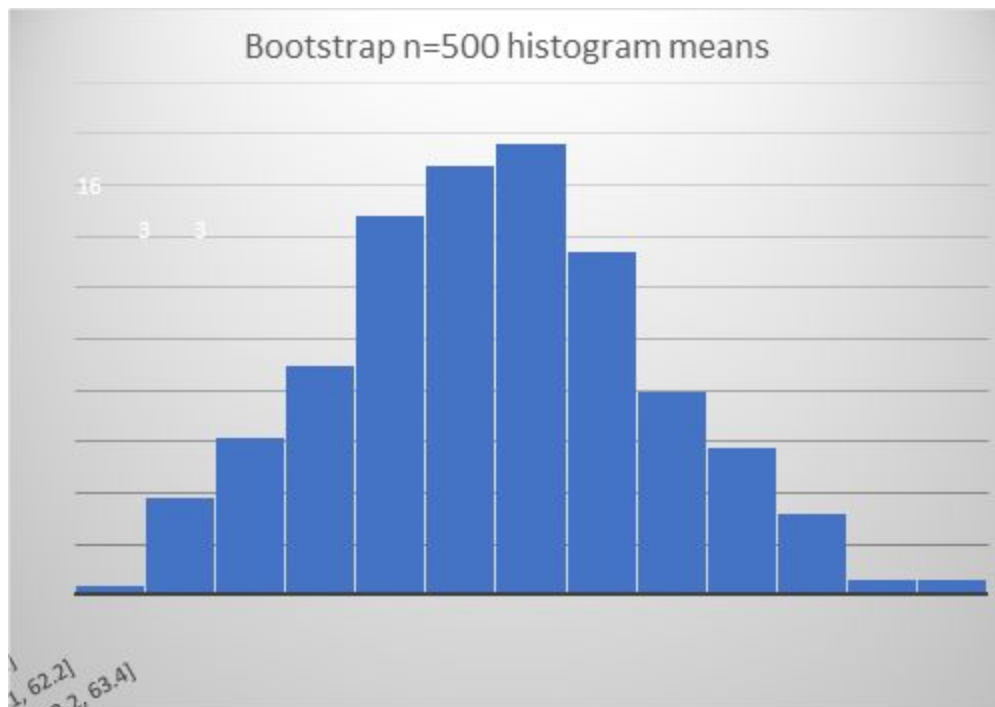


Figure 5. Sample of Austin data gon-Female n=500 The distribution is normal.

As we can see this data does not have any outliers and is clearly within the possibility of reason.

mean	54.95422488
confid interval	1.96

stdev	2.705598805
n	501
margin of error	0.236919391
lower bound	54.71730549
upper bound	55.19114427

Table 3. Stats of the n=501 sample Gon-Female

Looking at the original data for Gon-Female for Austin and West Town (the two largest in population), running a regression we find the p-value of 0.002173 which is < 5%, so in this case we can throw out the null hypothesis that the areas are representative of the whole population. We can see just from this data that even the two largest in populations are statistically different from one another and thus each community area must be treated differently and likely has other contributing factors.