introduction to ✴ spatial data

introduction

Spatial data encompasses **physical space** including geography and geometry: data obtained from cellular phones, connected devices, biomedical data, oil exploration, real estate prediction, agriculture, internet, natural gas networks, transportation networks, and many other **physical data networks**.

It is important to initially **map** spatial data for understanding of the orientation and **density** for creating predictions of future spatial data characteristics

initializing the spatial data analysis process

Spatial data has unique characteristics. For example, the illustration (right) maps Forced Entry Burglaries in New Haven CT, Sept 2014. **Point Data** numerically represents the location of each burglary; but often **Raster Data** (a count of point data within each cell).

Point Data is not always uniformly distributed (if one area has a higher frequency density than another or if the data is not captured consistently in all circumstance)

A traditional option would be to compute the physical distance (**Euclidean distance**) between the points. However, this is not a best practice with spatial data. An example for reason is if in the example of home burglaries, if the homes were separated by a fence, the **Euclidean distance** would not take into account the barrier between data points.
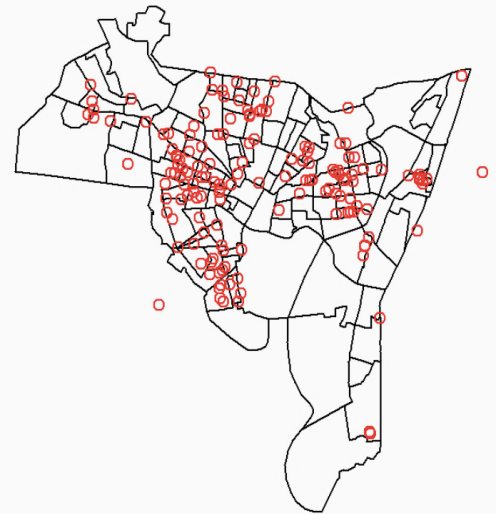
Alternative measures of defining **distance** between data points could take into account the characteristics of the population. For example, driving distance could account for barriers between points. Additionally, a mixture of distance definitions could be applied based on local circumstance.

A **Distance Matrix** is often required to plot the distance between all points. This is not always applicable for problems that do not have a finite number of entities to compute distances on.

	Crime 1	Crime 2	Crime 3	Crime 4	Crime 5	...	
Crime 1	0	3.329	1.309	...	...		
Crime 2	2.463	0	2.235	...	...		
Crime 3	1.125	2.376	0				
Crime 4	:	:	:				
Crime 5							
:							

Spatial Data Warning Considerations

- ¨ coordinate systems should not be confused with each other (state plane and gps coordinates)
- ¨ Euclidean distances along the Earth should not be manually computed as the Earth is curved
- ¨ Various Visualizations will aid in sanity checking the analysis methods

Other types of Spatial Data include **Space-Time Data** that comprise occurrence times of the various Spatial Events. Additionally, **Labeled Spatial Data** could include Binary Classification of an observed classification; **Real-Valued Labels** are also equally probable.