# Data

The data was collected from a variety of online sources (should add appendix at end of report) and was processed into a grid of identically sized square cells for the city. Each was one of three types of spatial data and the type dictated how the data was processed. *Area level* data is where a region on the map is assigned a value for the variable. For example, a census tract may have an mean income or it may have instances of victim-based crime. After processing, all data becomes area level, for spatial data analysis to proceed. When area level data is processed into the grid, each grid value becomes the weighted average – by area – of the regions contained within the grid cell. [picture for clarity?] The second type of data is *point pattern*. Point pattern data is where discrete occurrences of the variable of interest are located on a map. For example, this may be the locations of schools within the city or instances of crime. Point pattern data is processed in one of two ways, depending on the variable. The first determines the distance to the nearest occurrence from the center of the grid cell. The other counts the number of occurrences within the grid cell. For example, it is more relevant to the analysis to know the (Euclidean) distance of the cell to the nearest hospital rather than the number of hospitals within the cell. Or the number of victim-based crime incidents, rather than the distance to the nearest one. The final type of data is *geostatistical* data or raster. This data is processed from discrete locations which have a recorded value, e.g. rainfall or temperature gauges across the city. The values are processed by using the geospatial points within the grid, and calculating the cell’s mean, median, max, and min for each of the variables. Finally, for each of the covariates processed, a spatially lagged value is assigned to the cell. The spatially lagged value is simply the average of the surrounding (queen’s move) cells, and is analogous to the time lagging variables in time-series analysis. A function to process a range of data into a grid was developed in R [insert all references to R libraries used here]. The function has the ability to change the size of the cell grid, so that any conclusions reached can be tested for spatial discretization size sensitivity.