1. Data Sources and Data Pre-Processing:

1.1 Community Areas:

Source: City of Chicago Open Data Portal: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6

The 77 current community areas of Chicago were represented in a polygon vector format. These community area boundaries provided a spatial basis for the other data layers.

1.2 Population Data:

Source: http://www.actforchildren.org/wp-content/uploads/2018/01/Census-Data-by-Chicago-Community-Area-2017.pdf

The Population data at the community area level was difficult to find as population is recorded at the census block level. Since census blocks overlap in certain areas, spatially joining this data is not reliable and its computationally inefficient. A PDF file posted by Illinois Act for Children contained the required 2010 census data summarized by community area. A simple PDF scraper written in Python3 was used to quickly obtain the community area population data. The community area populations were then joined to the community areas layer by attribute.

1.3 Average Life Expectancy Data:

Source: City of Chicago Open Data Portal: https://data.cityofchicago.org/Health-Human-Services/Public-Health-Statistics-Life-Expectancy-By-Commun/qjr3-bm53

The Average Life Expectancy Data calculated from the 2010 U.S. Census was obtained from a table. The Life Expectancy's by community area values were then joined to the community areas layer by attribute.

1.4 Crime Data:

Source: City of Chicago Open Data Portal: https://data.cityofchicago.org/Public-Safety/Crimes-2018/3i3m-jwuy

The 2018 Crime Incidents data were represented as points derived from a CSV file. The City of Chicago claims that these points are recorded at the block level and locations are approximate for privacy reasons. The 260,000+ crime incident points for 2018 were then spatially joined to the community area boundaries layer. The sum of all criminal incidents was recorded for each community area in Chicago. The crime counts per community area were divided by population, and then multiplied by 1000 to create a crime per 1000 people statistic.

1.5 Hardship Index Data:

Source: City of Chicago Open Data Portal: https://data.cityofchicago.org/Health-Human-Services/Census-Data-Selected-socioeconomic-indicators-in-C/kn9c-c2s2

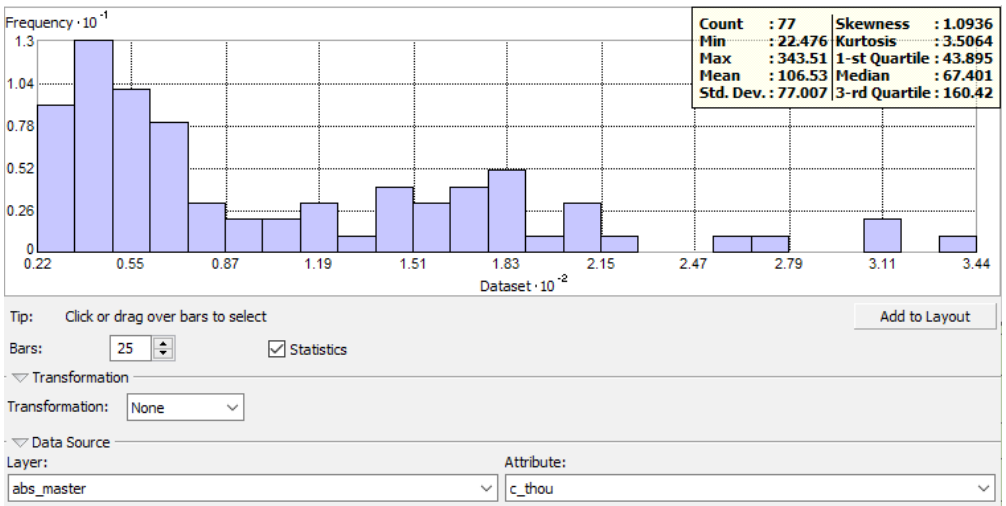
The Hardship Index Data was derived from a table containing the various socioeconomic indicators per community area in Chicago. These socioeconomic indicators were calculated from the 2010 U.S. Census. The Hardship Index for each community area was joined to the community areas layer by attribute. This Index is calculated on a 1 – 100 scale by the U.S. Census Bureau by incorporating per capita income, percentage of housing crowded, percentage of households below poverty line, percentage of individuals age 16 or over that are unemployed, percentage of individuals age 25 or over without a high school diploma and percentage of individuals under the age of 18 and over the age of 64. These indices gives each community area a unique ranking between 1 – 100. Since there are only 77 community areas the values are distributed evenly with some gaps. The distinct values make this statistic excellent for comparing community areas however, the U.S. Census Bureau claims that this data cannot be compared to other regions as it normalized for Chicago.

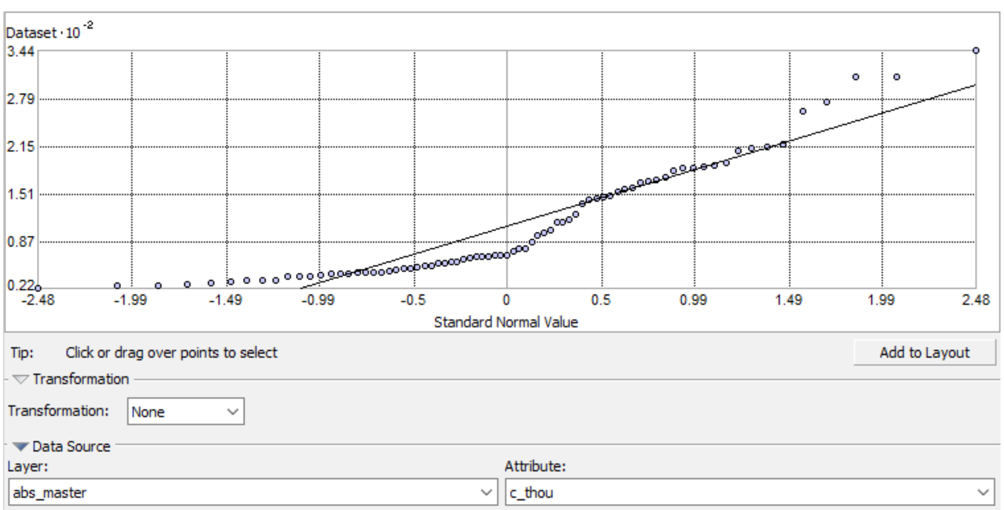
2. Attribute Analysis

2.1 Crime per 1000 People:

Before classifying into 3 classes using Natural Breaks the Crime per 1000 People statistic was presented in ratio format as the difference between community areas was exact and there is no possibility of negative values. After classification the Crime per 1000 People Statistic was presented in interval format where the class boundaries act as the intervals.

The following histogram and QQ-Plot show the distribution of the data:



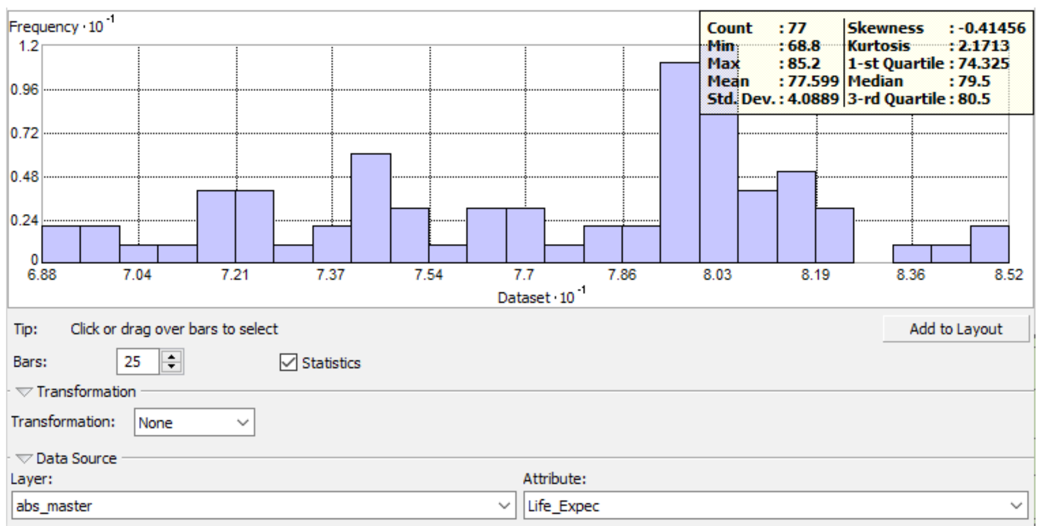


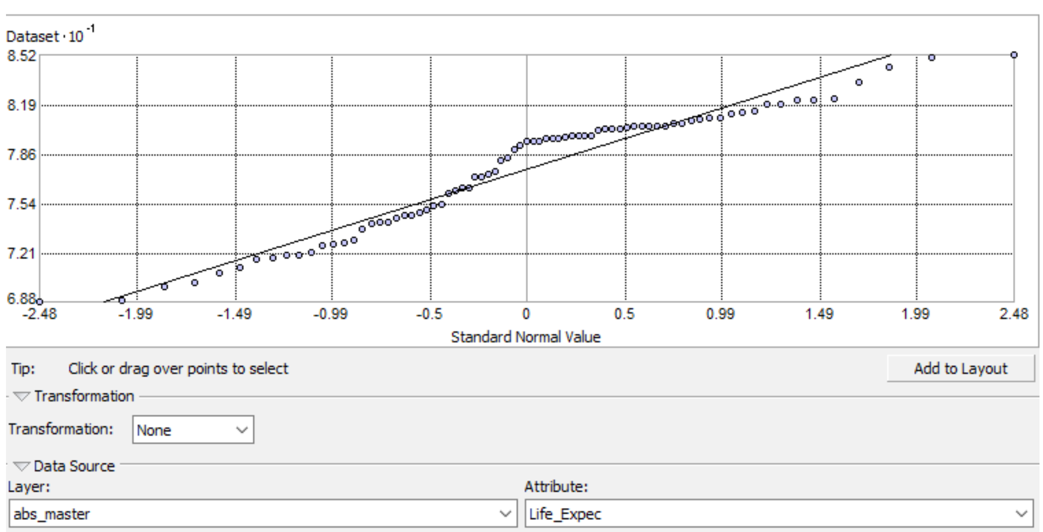
The data does not follow a Gaussian curve as there is a non linear pattern shown in the QQ-Plot. In addition, the histogram shows a positive skew, as a majority of the values are located in the lower to middle regions.

2.2 Average Life Expectancy:

Before classifying into 3 classes using Natural Breaks the Average Life Expectancy statistic was presented in ratio format as the difference between community areas was exact and there is no possibility of negative values. After classification the Average Life Expectancy statistic was presented in interval format where the class boundaries act as the intervals.

The following histogram and QQ-Plot show the distribution of the Average Life Expectancy data:



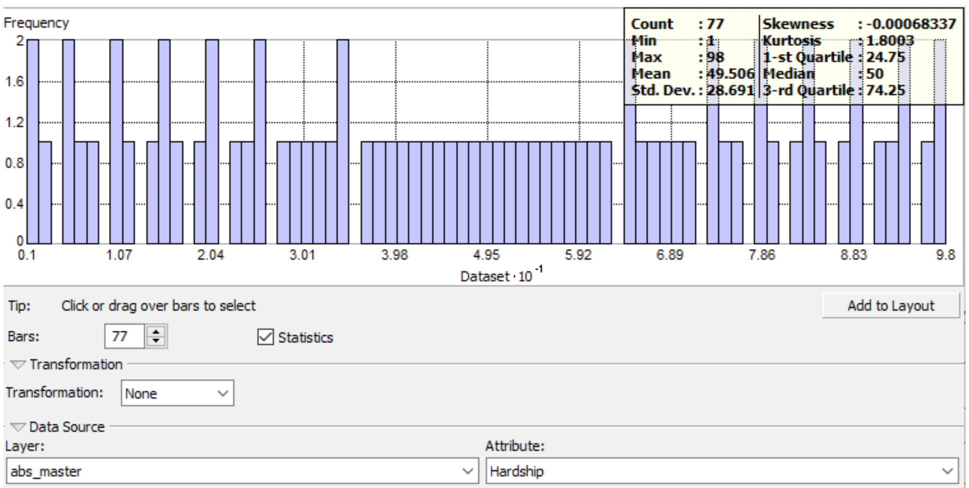


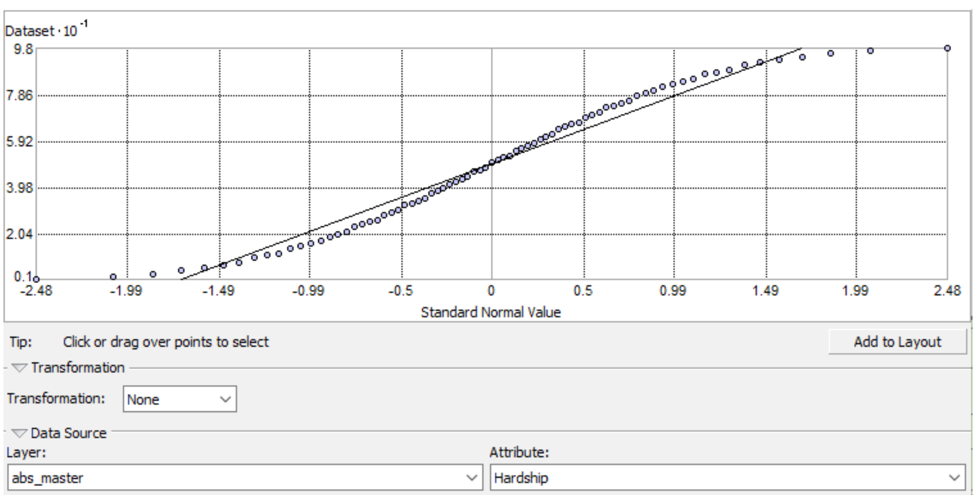
The data does not follow a Gaussian curve as there is a non linear pattern shown in the QQ-Plot. In addition, the histogram shows a negative skew, as a majority of the values are located in the positive standard deviation regions of the distribution.

2.3 Hardship Index

The Hardship index was obtained and presented in ordinal format as each community area has a unique index corresponding to a ranking. This ranking is determined by the six socioeconomic indicators outlined in Section 1.5. The extrusion level is proportional to this ranking and allows the reader to compare community areas to determine which one has a greater or lesser Hardship Index ranking.

The following histogram and QQ-Plot show the distribution of the data:





The distributions are as expected for distinct ordinal data. The gaps in the histogram are the result of the 1 - 100 scale and there only being 77 distinct community areas.

3. GVF Results, Classification and Justification

3.1 GVF Results:

The following table shows the GVF values and class bounds for the classification methods tested on the map:

Classification	Statistic	Class 1 Upper Bound	Class 2 Upper Bound	Class 3 Upper Bound	GVF Value
Natural Breaks	Crime per 1000 People	100.59	216.98	343.51	0.9
Equal Interval	Crime per 1000 People	129.49	236.5	343.51	0.9
Natural Breaks	Average Life Expectancy	72.9	77.5	85.19	0.88
Equal Interval	Average Life Expectancy	74.27	79.73	85.19	0.84

3.2 Classification and Justification:

Due to the nature of the “3D” choropleth, 3 classes was deemed as the maximum as the number of possible colours or classes is 2 to the power of n, where n is an integer representing the number of classes. 4 or more classes would have resulted in many possible values which would have difficult to interpret. The natural breaks method was chosen due to the non-Gaussian distributions in both the crime and life expectancy data sets. In addition, it resulted in a higher GVF value for the Average Life Expectancy statistic.

The Hardship Index variable was not classified as the extrusion was proportional to the level of hardship in each community area. The magnitude of the extrusion was determined visually however, this variable was intended to be used for pair-wise comparisons between community areas and not as a means to quantify the hardship index directly.

4. Symbol Parameters

4.1 Background:

A “background surface” was created by extruding a 20 percent grey polygon file of the community areas by a factor of the hardship index of each polygon feature divided by 2000. This level of extrusion was deemed visibly suitable as it showed adequate variation between community areas. This layer provides an opaque surface that supports the overlaying of two transparent layers to make a “3D” choropleth.

4.2 Crime per 1000 People:

The Crime per 1000 people layer was given a red colour palette progressing in a linear fashion through the red colours. The first class was given an HSV value of (0,33,100), the second was given a value of (0, 66, 100) and the third class was given a value of (0,100,100). This layer was made 10% transparent and then extruded by a factor of the hardship index of each polygon feature divided by 2000.

4.3 Average Life Expectancy:

The Average Life Expectancy layer was given a blue colour palette progressing in a linear fashion through the blue colours. The first class was given an HSV value of (60,33,100), the second was given a value of (60, 66, 100) and the third class was given a value of (60,100,100). This layer was made 20% transparent and then extruded by a factor of the hardship index of each polygon feature divided by 2000. This layer was made more transparent then the red layer due to the visually dominating effect of blue colours. This was to make each colour appear more “equally weighted”.

4.4 Legend:

The legend of the map is one of its key features. The x-axis represents the 3 Crime per 1000 people classes, changing in intensity from left to right, and the y-axis represents the 3 Life Expectancy classes changing in intensity from bottom to top. Finally, the z-axis shows the scale of the 3D extrusion using the Hardship Index variable.

5. Map Message

The aim of this map was to show the spatial distribution of hardship in a large US City such as Chicago. It was hypothesized that the inner city areas would see much higher rates of hardship and crime, and a lower average life expectancy. The map clearly shows this pattern in the two high intensity areas in the North-East and Central regions of the map. These areas display a purple colour and a large extrusion, showing high crime, hard living conditions and low life expectancy. The message of this map is complex and it is not meant to try and show causation between crime and life expectancy or crime and hardship or other relationships. It shows that getting out of difficult living conditions is an uphill battle. There are numerous factors that influence hardship and quality of life, far too many to be considered on any one map however, maps such as this one provide a means to display a spatial pattern. This map shows that even in a relatively small geographic area, various socioeconomic factors can introduce large variation in the attributes of particular areas.

Another characteristic of this map that was crucial in maintaining readability and understanding was the decision to not use any symbols. This “3D” choropleth incorporates 3 variables by manipulating the geography of the landscape, creating a beautiful surface that is unobstructed by “loud” symbols. This was an attempt to let the data “speak for itself” and the map presents the reader with a lot of data (over 300,000 unique entries) in a manageable form factor. In conclusion, this map was an attempt to integrate all the skills I have learned in this class as well as other skills I have acquired along the way into something meaningful and important for people to read.