Sumi Choudhury HW #2

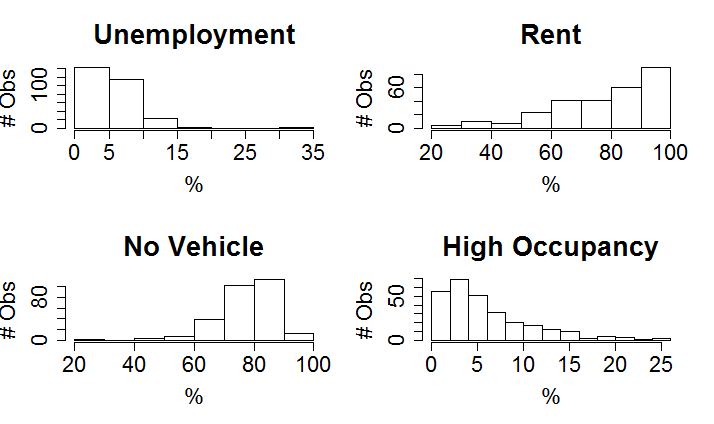
SDGB 7844 October 19, 2016

Solutions HW 2

1) A census tract is an area roughly equivalent to a neighborhood established by the Bureau of Census for analyzing populations. It generally encompasses a population between 1,200 to 8,000 people with an optimum size of 4,000 people. It is considered by the Bureau of Census as “relatively permanent”, but it can change over time. In New York County, there are **288** census tracts; this is shown by the number of observations in R Studio in the workspace.

2) One advantage of computing estimates after combining 5 years of data is reliability. Multi-year estimates are given for areas with populations of less than 65,000. This type of estimate is beneficial in providing a better overall picture of the population characteristics, in case a point-in-time estimate might reflect drastic anomalies due to environmental, economic or other influential factors. One disadvantage of computing estimates after combining 5 years of data is less current data. If we want to look at the most recent data, multi-year estimates provide less current information because they are based on the data sample from the most previous year and the multiple years previous to it. For areas with minimal changes taking place, the less current data will not have significant influence on the estimates.

3 Histograms for each variable:



Unemployment is positively skewed with a mean higher than the median. Rent is negatively skewed with a mean less than the median. No Vehicle is also negatively skewed and the mean is less than the median. High Occupancy is positively skewed with a mean higher than the median.

Summary Statistics:

Mean Median Standard Deviation Max. Min.

Unemployment 5.839716 5.0 3.592804 33.3 0.0

Rent 78.102509 81.6 18.215975 100.0 21.8

No Vehicle 77.580645 78.7 9.547186 96.0 21.9

High Occupancy 5.884588 4.4 4.935244 25.1 0.0

4) Number of missing observations for each variable:

Unemployment Rent No Vehicle High Occupancy

6 9 9 9

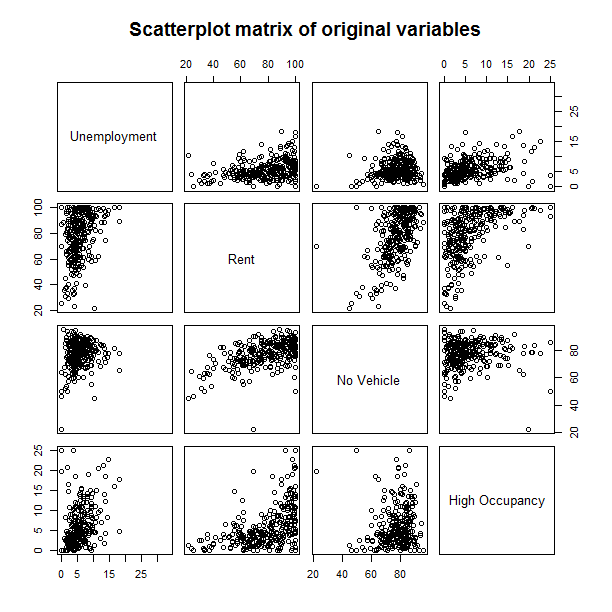
Census tracts with incomplete data: 9

Percentage of census tracts with missing data: 3.125%

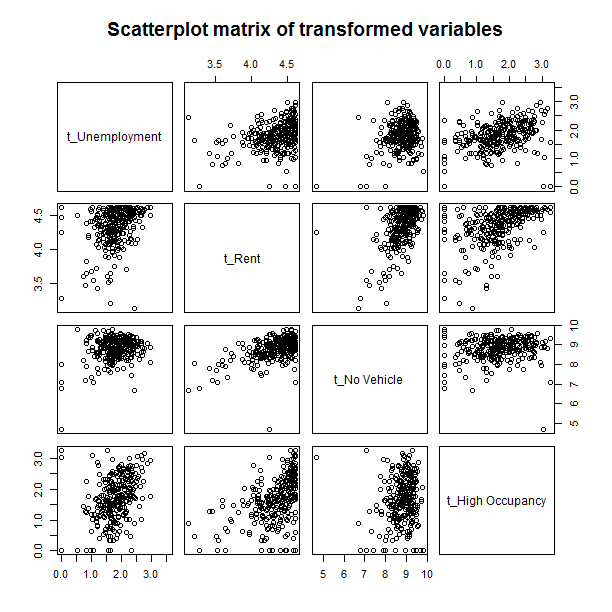
[1] 0.03125

The missing observations are not a problem for our analysis due to the relatively small percentage of incomplete census tracts. We can exclude the incomplete records from our analysis when, for example, we are calculating the summary statistics. We should keep our incomplete data because we will need them to include within the boundaries of the map.

5) The scatterplot matrix shows that there is no distinct linear relationship between the variables.



The scatterplot matrix shows that there is no distinct linear relationship between the transformed variables.



Correlation matrix of the transformed variables:

t\_Unemployment t\_Rent t\_No Vehicle t\_High Occupancy

t\_Unemployment 1.0000000 0.3235757 0.2492961 0.3666114

t\_Rent 0.3235757 1.0000000 0.5467847 0.4792976

t\_No Vehicle 0.2492961 0.5467847 1.0000000 0.1371552

t\_High Occupancy 0.3666114 0.4792976 0.1371552 1.0000000

There is some relatively higher correlation between Rent and No Vehicle but this could be due to the many people living in the metropolitan area choosing not to own cars. There are no negative correlations, which is a good sign since negative correlations would cancel the variables out. Also, we would prefer **not** to have high correlations, as in constructing an index we would like for each variable to capture a different component of deprivation.

6) Census tract that is most deprived: census tract 261, Townsend index level 5.235171

GEO.id2 GEO.display.label Townsend Index Rank

263 36061026100 Census Tract 261, New York County, New York 5.235171 1

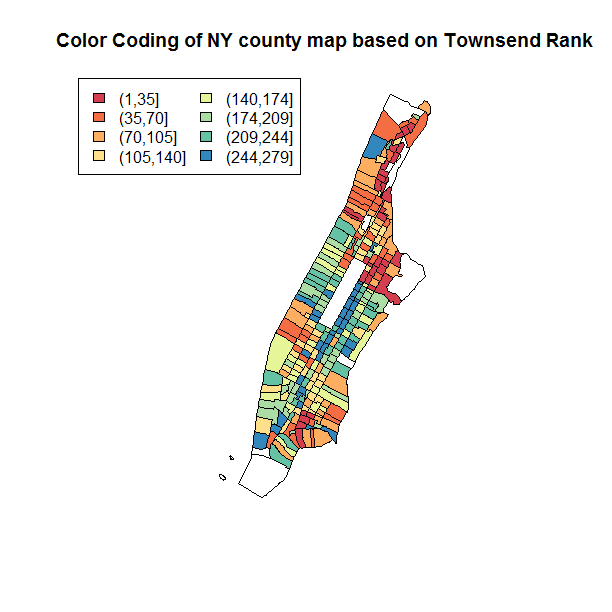
Census tract that is least deprived: census tract 112.02, Townsend index level -13.06891

GEO.id2 GEO.display.label Townsend Index Rank

112 36061011202 Census Tract 112.02, New York County, New York -13.06891 279

7) Calculating margins of error is a critical component of determining the precision of the indices. We may have over or underestimated some of the indices, since the variables from which they are based upon are themselves estimates. Without being able to assess the confidence level that a given sample estimate is within an acceptable distance from the population parameter, we risk also not being able to assess correctly about which areas are the most deprived and require additional funding.

8)



Based on the color-coded map, the least deprived regions are in the upper-east side, which is to be expected as it is known to be an affluent area. Farther north are the most deprived areas which are nearby Harlem. On the lower right side is the East Village which is more deprived than the West Village. The large rectangle in the middle of the map is Central Park. In general, the upper west side is less deprived than north of Central Park and the lower-east side.

9) Lowenstein is in census tract 145:

GEO.id GEO.id2 GEO.display.label

146 1400000US36061014500 36061014500 Census Tract 145, New York County, New York

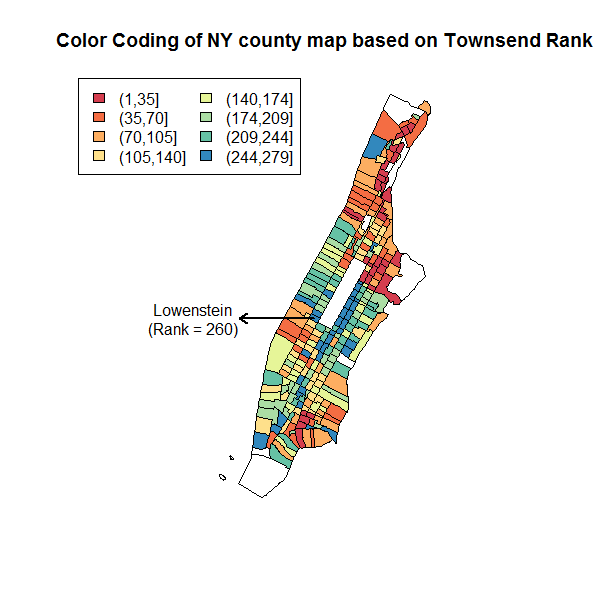
Unemployment Rent No Vehicle High Occupancy t\_Unemployment t\_Rent t\_No Vehicle

146 2.5 66.3 72.2 0.5 1.252763 4.20916 8.497058

t\_High Occupancy Townsend Index Rank

146 0.4054651 -3.773305 260

The deprivation level rank is 260.



10) Since New York County is an urban county, it would not make sense to compute the Townsend index values for all census tracts within New York State combined. This is because certain critical variables that we have used to compute our deprivation indices in this analysis would be impacted in a way which would make this study less meaningful. For example, in rural counties, many more people would own cars because public transportation is far less widely available. Also, it is likely that more people in rural areas would own a home, however home ownership in rural counties is not always a good measure of a deprivation variable. Owning a home in a rural area is likely much more affordable than owning an apartment in a metropolitan area. Thus, ownership of a vehicle or a home in a rural area cannot be utilized to construct meaningful indices when evaluating against indices in the urban areas.