**HW 2: [Minxia Ji]-[1:15]**

Question 1

A census tract is a geographic region defined for the purpose of taking a [census](https://en.wikipedia.org/wiki/Census" \o "Census).Usually these coincide with the [limits](https://en.wikipedia.org/wiki/Border" \o "Border) of [cities](https://en.wikipedia.org/wiki/Cities" \o "Cities), [towns](https://en.wikipedia.org/wiki/Town" \o "Town) or other administrative areas and several tracts commonly exist within a county.

Census tracts represent the smallest territorial unit for which population data are available in many countries.In the United States, census tracts are subdivided into [block groups](https://en.wikipedia.org/wiki/Census_block_group" \o "Census block group) and [census blocks](https://en.wikipedia.org/wiki/Census_block" \o "Census block).

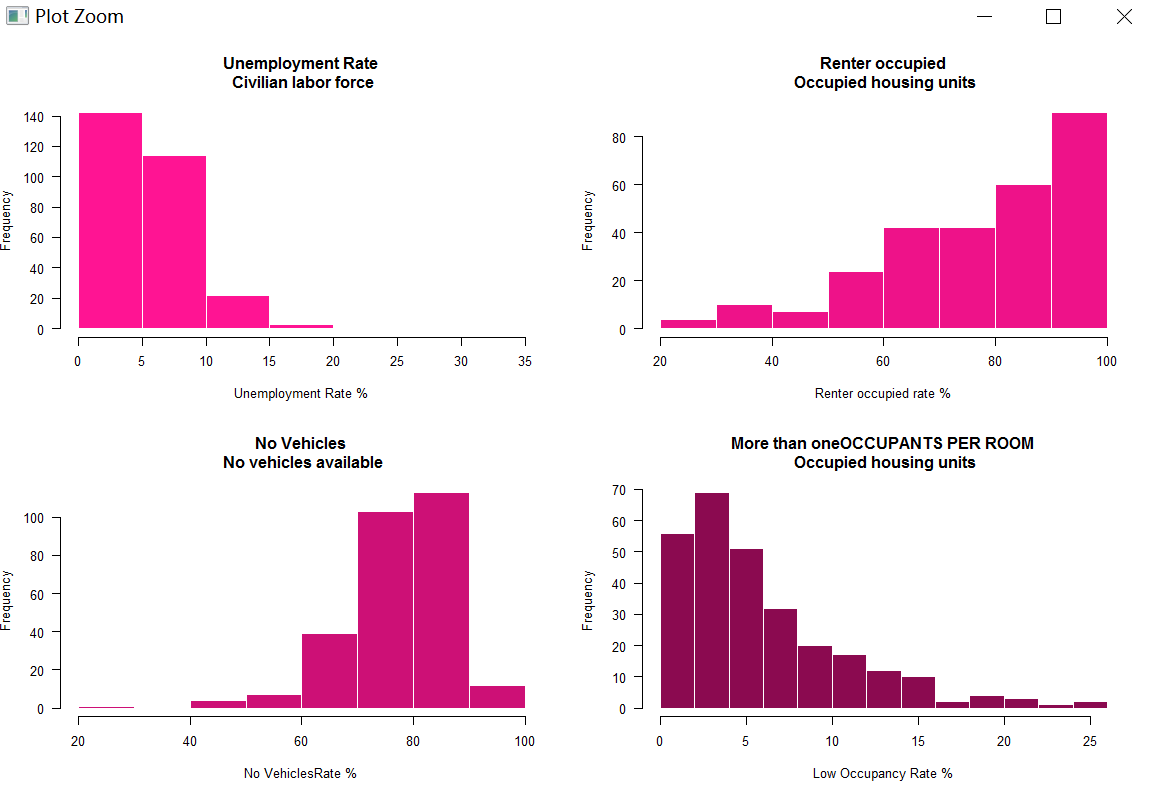
There are 288 census tracts in New York County.

Question 2

The advantage is that with 5 years data, we can get a comprehensive understanding of deprivation.In such a short period like one year, it is possible that extreme events might influence rent price or other issues, then influence much on deprivation rate of whole the year.Thus, the data in such a short period may not sufficient to illustrate deprivation rate.

The disadvantage is that New York, where economy grows in a dramatic speed, changes so rapidly from day by day, let alone year by year.Using 5 years data may not appropriate to represent the current deprivation rate in New York.

Question 3



The unemployment chart is ***right skewed.***When the unemployment rate going up, the frequency is going down.The histogram is like a slope.In most of census tracts, the unemployment rate is in 0%~ 10%.In some census tracts, the unemployment rate is in 10%~15%.Only in a few census tracts that the unemployment rate is in 15%~20% and 30%~35%.

The renter chart is ***left skewed***.We can see the bars tend to go up.It shows that in most areas we have high rent ration.It means that in New York County, houses are mostly rented out.

The No Vehicles chart is also ***left skewed***.The bars are going up from 20%~90%.In 90%~100%, the frequency is going down.That means in most areas, the ratio of households who have no cars is between 60%~90%.

The low occupancy rate chart is ***right skewed***.We can tell that the bars tend to going down when the percentage becomes higher.That means in most areas of New York County, most people live in overcrowded houses where more than one person live in a room.

Statistics

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | “Unemployment” | “housingTenure” | “noVehicles” | “lowOccupancy” |
| mean | 5.839716 | 78.102509 | 77.580645 | 94.115412 |
| median | 5.0 | 81.6 | 78.7 | 95.6 |
| Standard deviation | 3.592804 | 18.215975 | 9.547186 | 4.935244 |
| maximum | 33.3 | 100.0 | 96.0 | 100.0 |
| minimum | 0.0 | 21.8 | 21.9 | 74.9 |

Question 4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | unemployment | housingTenure | noVehicles | lowOccupancy |
| How many missing | 6 | 9 | 9 | 9 |

In data DP03 where we get unemployment data, 2.076125% percentage of census tracts do not have complete data.

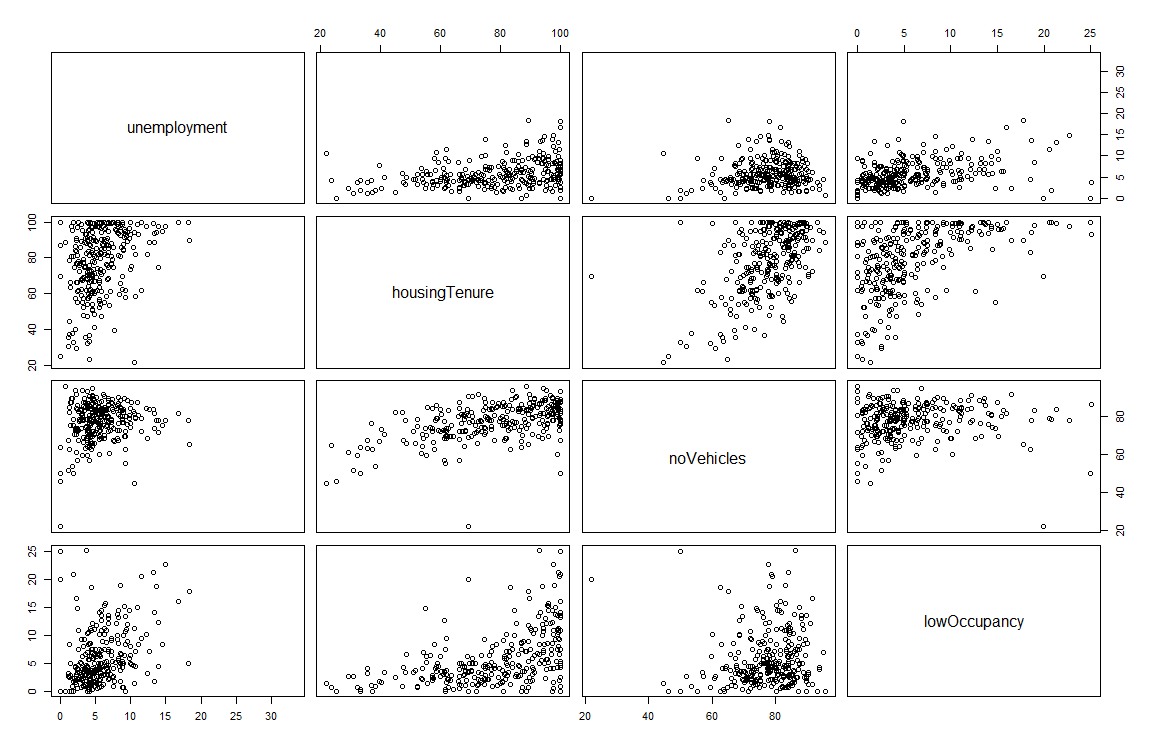
In data DP04, where we get housingTenure, noVehicles and lowOccupancy datas, 3.114187% percentage of census tracts do not have complete data.

This is not a problem of the analysis.Because the rates are far less than 10%, it is manageable.

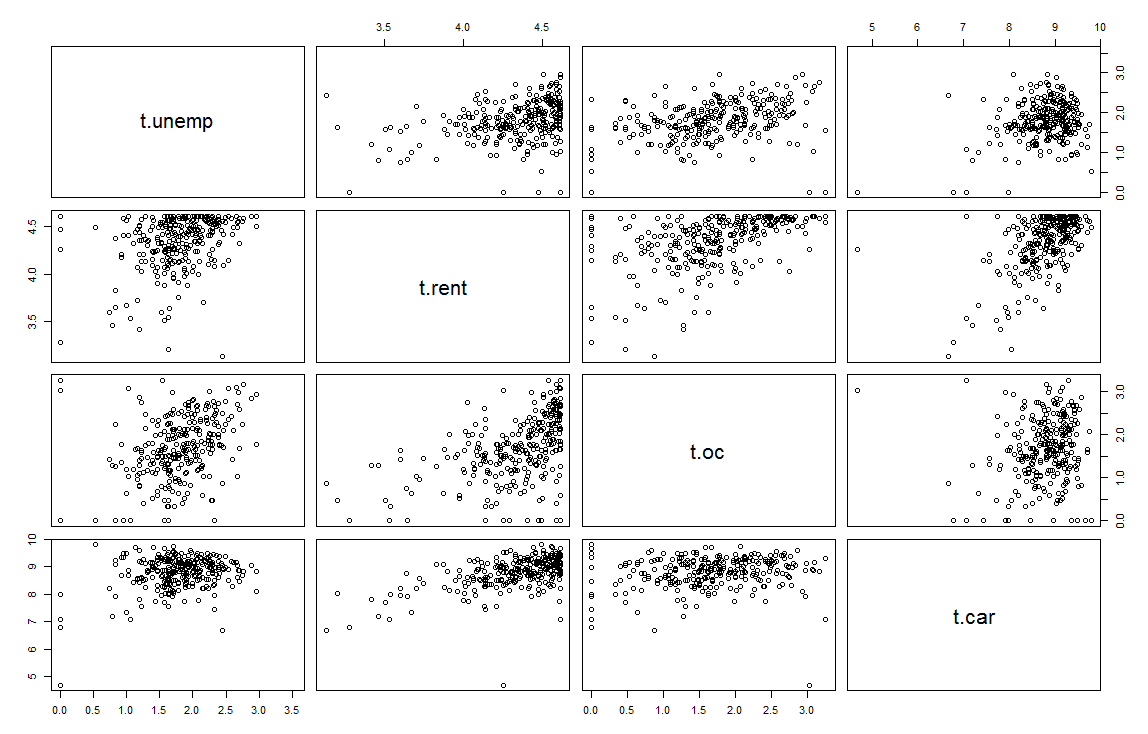
Question 5

#There’s **no linear** relation between these variables.

#There’s **no linear** relation between these transformed variables



*Before transformation*

**

*After transformation*

t.unemp t.rent t.oc t.car

t.unemp 1.0000000 0.3235757 0.3666114 0.2492961

t.rent 0.3235757 1.0000000 0.4792976 0.5467847

t.oc 0.3666114 0.4792976 1.0000000 0.1371552

t.car 0.2492961 0.5467847 0.1371552 1.0000000

We can conclude that unemployment has weak correlation with other variables; housing tenure shows slightly correlation with no vehicles and low occupancy.In addition, no vehicles has little correlation with low occupancy.

Question 6

Most deprived:

Census Tract 261, New York County, New York: 5.235171

Least deprived:

Census Tract 112.02, New York County, New York: -13.06891

Question 7

The margin of error means random sampling error in the data.

It implicates that the results after analyzing samples might not accurately represent the actual situation.The larger the margin of error, the less confidence we would have in terms of the accuracy of analysis results.

Question 8

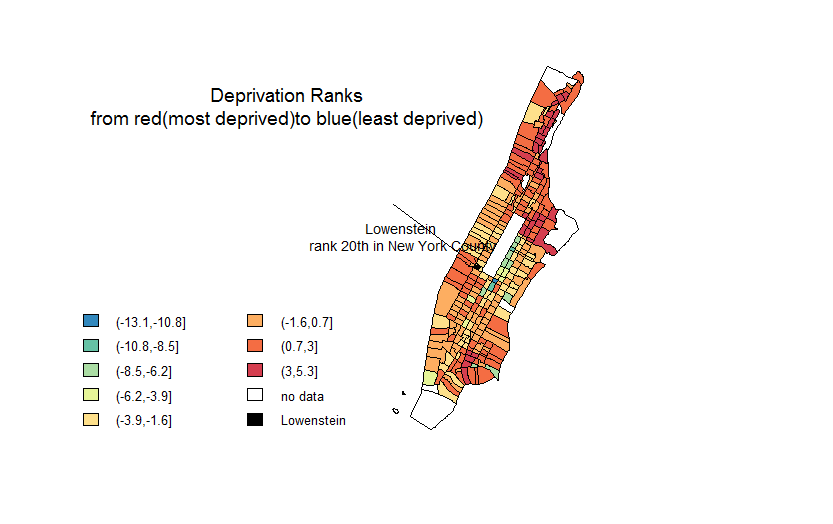
#See map in Question 9

The map shows that the Upper Town is **more deprived**.especially the Upper East side. The deprivation level in Midtown is lower than Upper Town and Lower Manhattan has nonuniform distribution of deprivation.The **least deprived area** shows around the central park, especially in Upper east side.

The large rectangle in the middle of map represents **Central Park**.

Question 9

Census Tract 145, New York County, New York(GEO.id:36061014500) is where Lowenstein locates. It ranks 20th in New York County regarding the deprivation level.



Question 10

It doesn’t make sense to compute the Townsend index values for all census tracts within New York state combined.

Because in rural areas, housing price is much cheaper.Most people own houses rather than own an apartment or rent an apartment.Seldom do people rent houses.Households usually own one home or even more homes.In addition, the traffic in NY County is not so good.People would like to choose public transportation in NY rather than buy a car.But in other areas, people are much more likely to buy a car.So these variables in townsend index are not sufficient to know the real deprivation rate regarding to all census tracts within New York state combined.