Homework 2

SDGB 7844, Prof. Nagaraja

Yuwen (Suyi) Wu, 330-530PM

10/17/2017

Question 1

Solution:

A census tract, census area, census district or meshblock is a geographic region defined for the purpose of taking a census

This is the Wikipedia,where we can search for the definition of Census Tract.

Census Tract 319, New York County, New York

Question 2

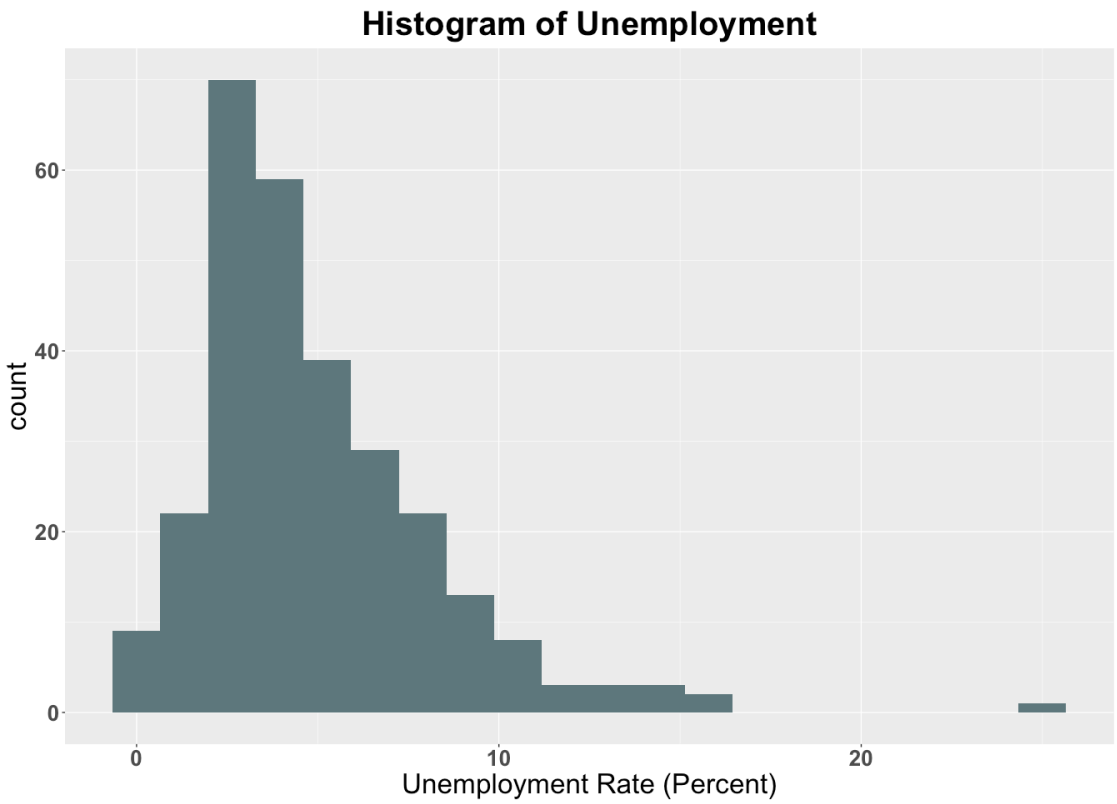
Solution:

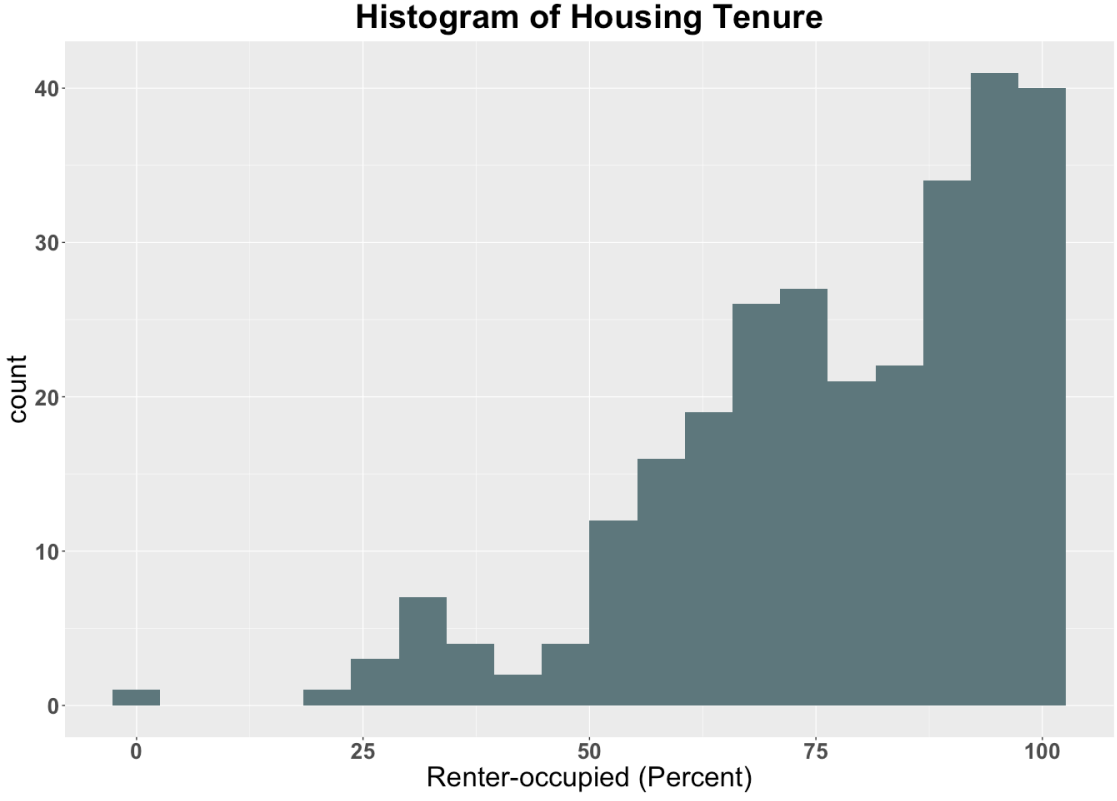
The advantage is that the larger number makes estimates more accurate, the disadvantage is that there are more outliers in larger number.

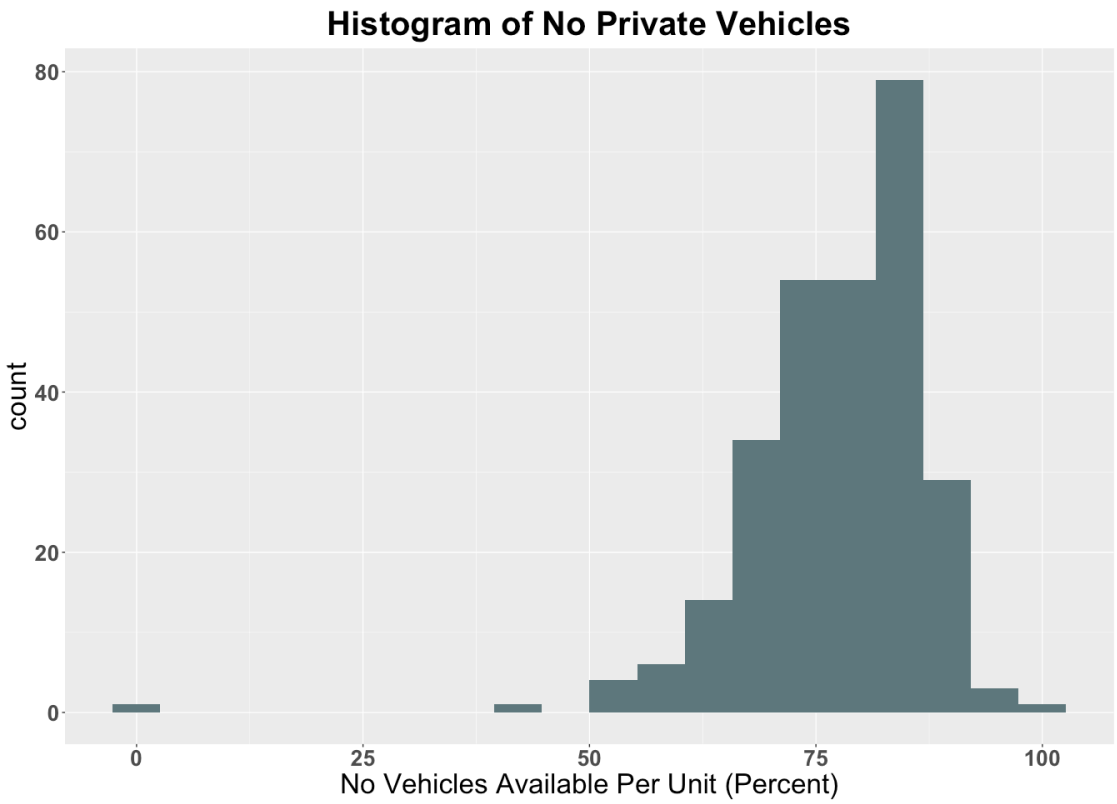
Question 3

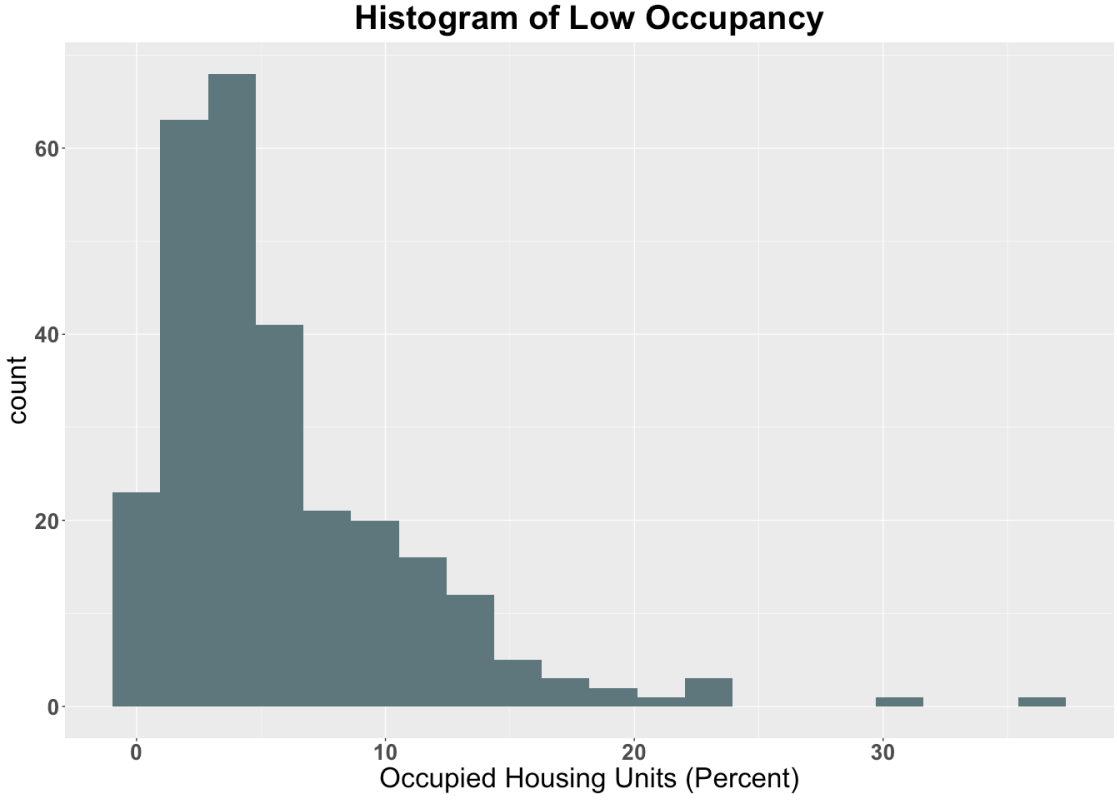
Solution:

```
##                    mean.z median.z      sd.z max.z min.z
## Unemployment     4.926855     4.00  3.250082  25.0     0
## Housing_Tenure  77.531429    80.75 19.173203 100.0     0
## No_Vehicles     77.413571    78.60  9.783584 100.0     0
## Low_Occupancy    5.883214     4.10  5.164891  36.4     0
```


Histogram of Unemployment
count
60
40
20
0
0
10
20
Unemployment Rate (Percent)


Histogram of Housing Tenure
count
40
30
20
10
0
0
25
50
75
100
Renter-occupied (Percent)


Histogram of No Private Vehicles
count
80
60
40
20
0
0
25
50
75
100
No Vehicles Available Per Unit (Percent)


Histogram of Low Occupancy
count
60
40
20
0
0
10
20
30
Occupied Housing Units (Percent)

the Unemployment rate Histogram is right (positive) skewness. the Housing Tenure Histogram is left skewness. the No Vehicles Histogram is left skewness. the low occupancy histogram is left skewness.

Question 4

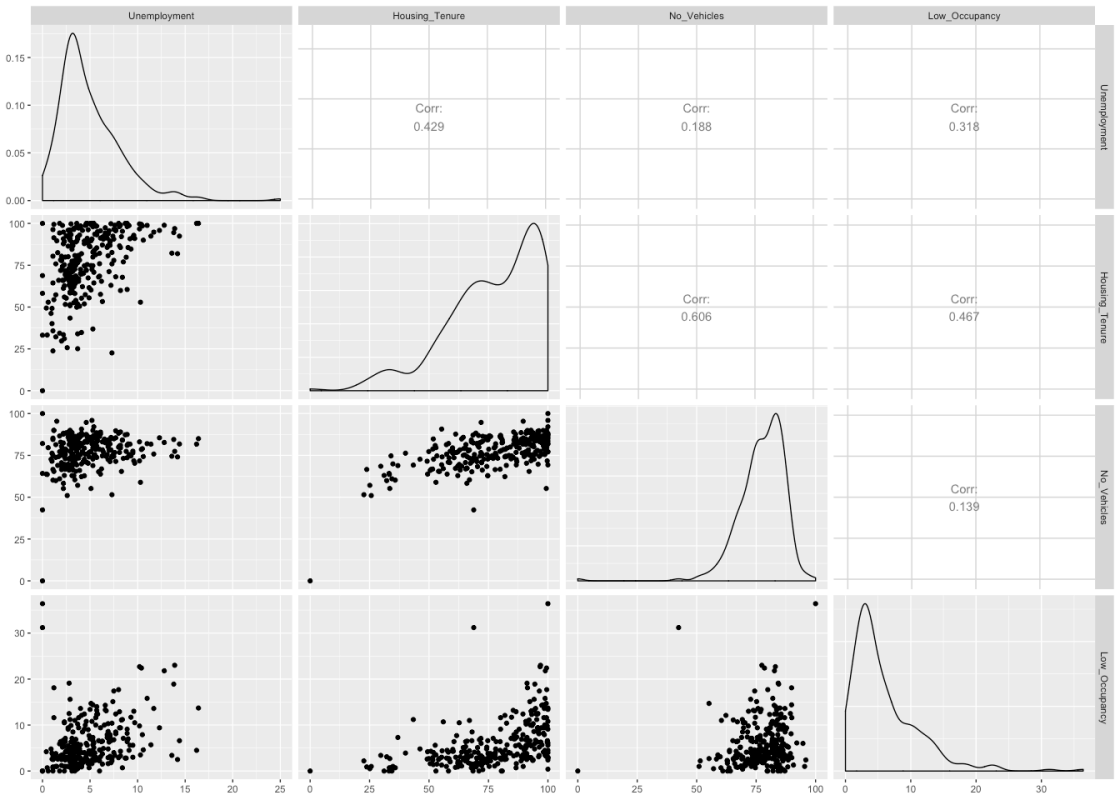
Solution:

```
sum(is.na(z$Unemployment))

## [1] 5

sum(is.na(z$Housing_Tenure))

## [1] 8

sum(is.na(z$No_Vehicles))

## [1] 8

sum(is.na(z$Low_Occupancy))

## [1] 8

1-nrow(z[complete.cases(z),] )/nrow(z)

## [1] 0.02777778
```

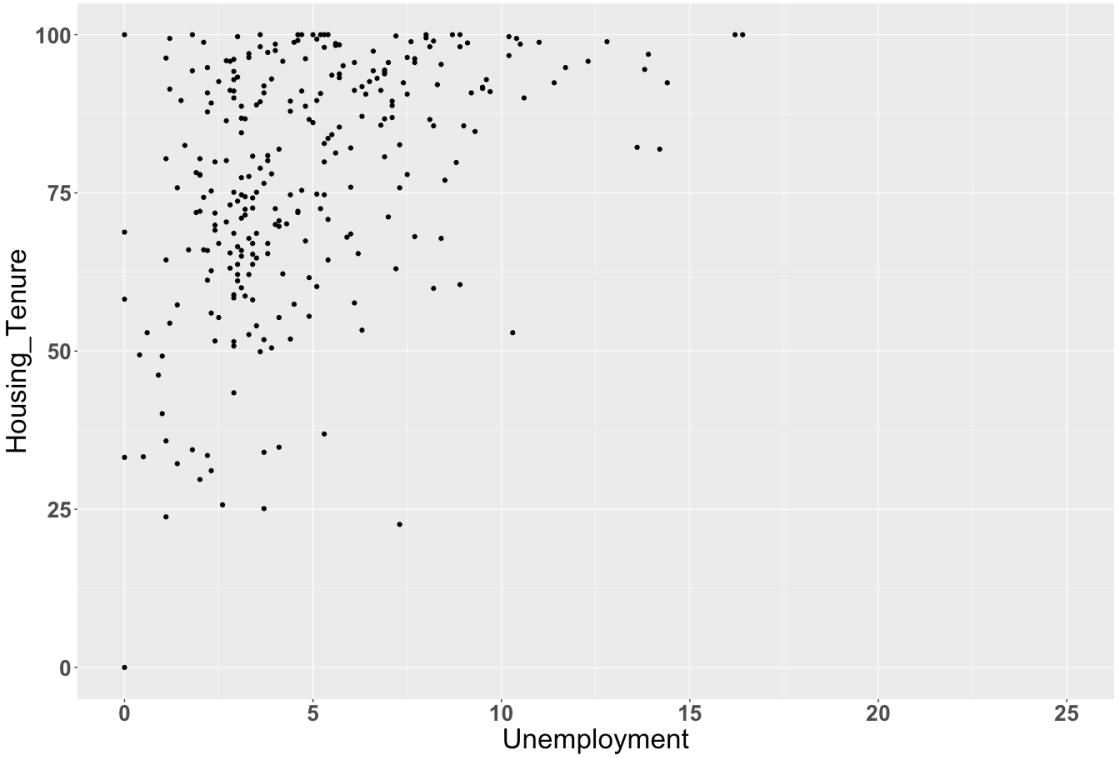
0.0277778 which is less than 10% so it is not the problem of analysis the missing reason is that the contract in business area of wall street or the contract in central park

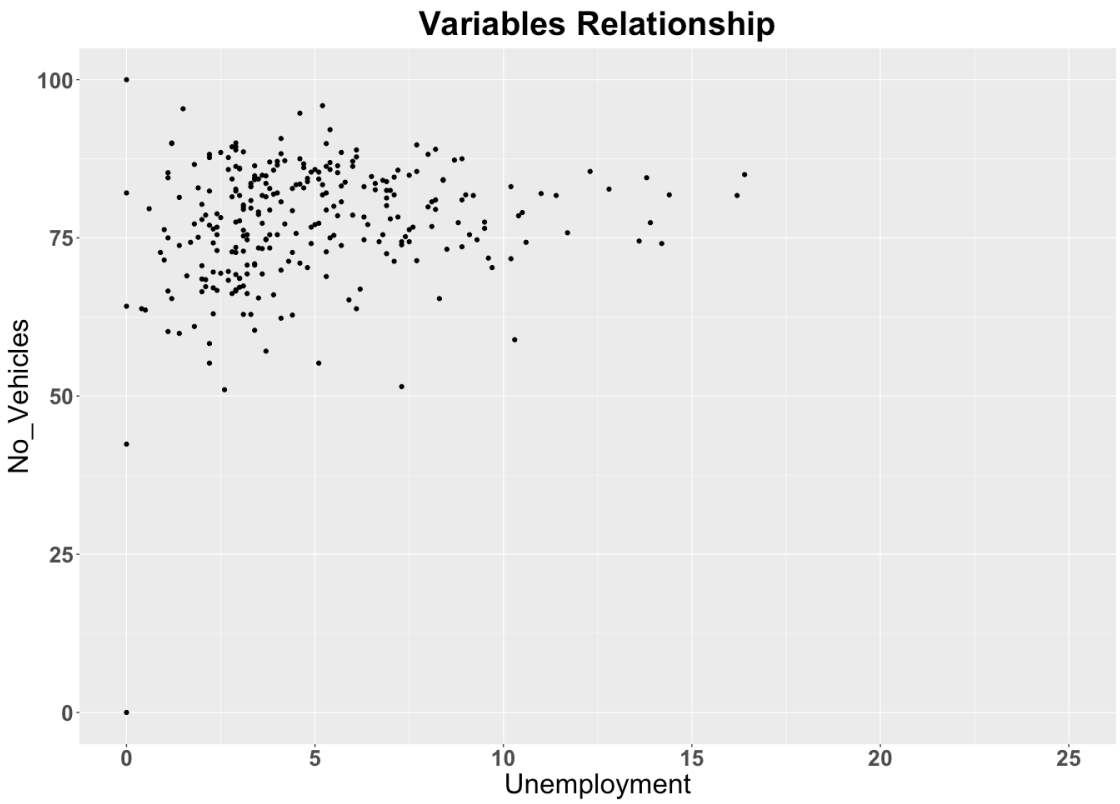
Question 5

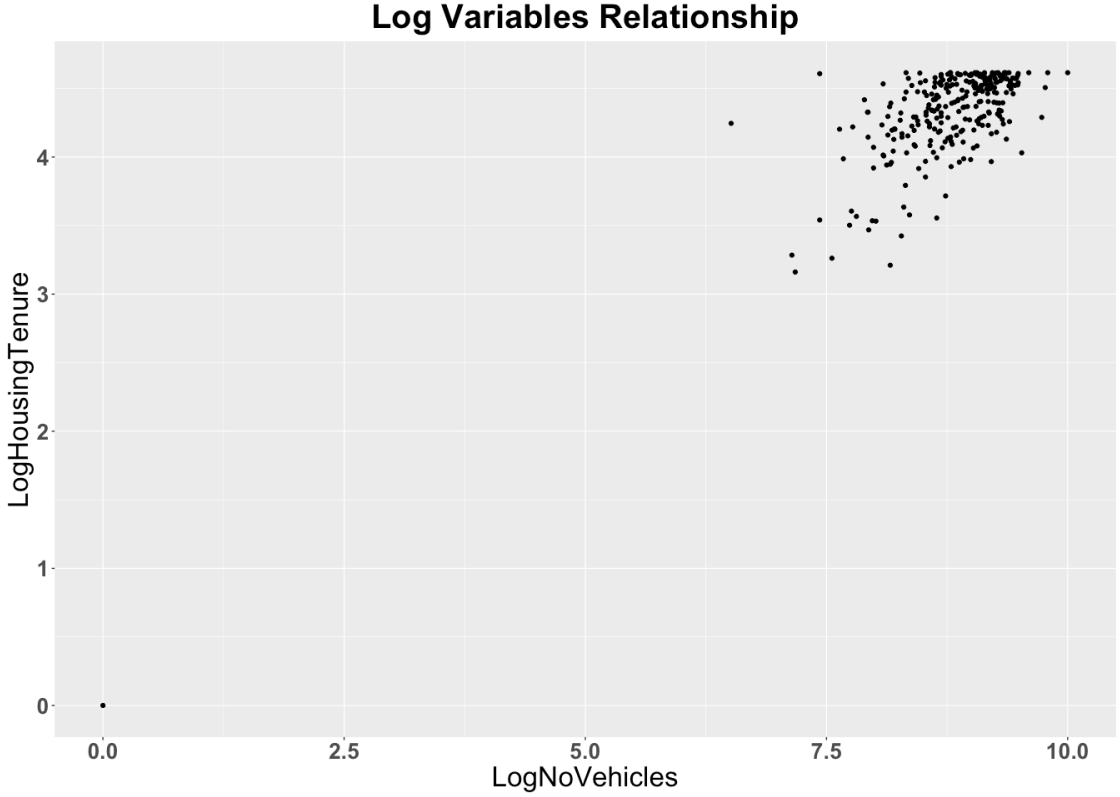
Solution:

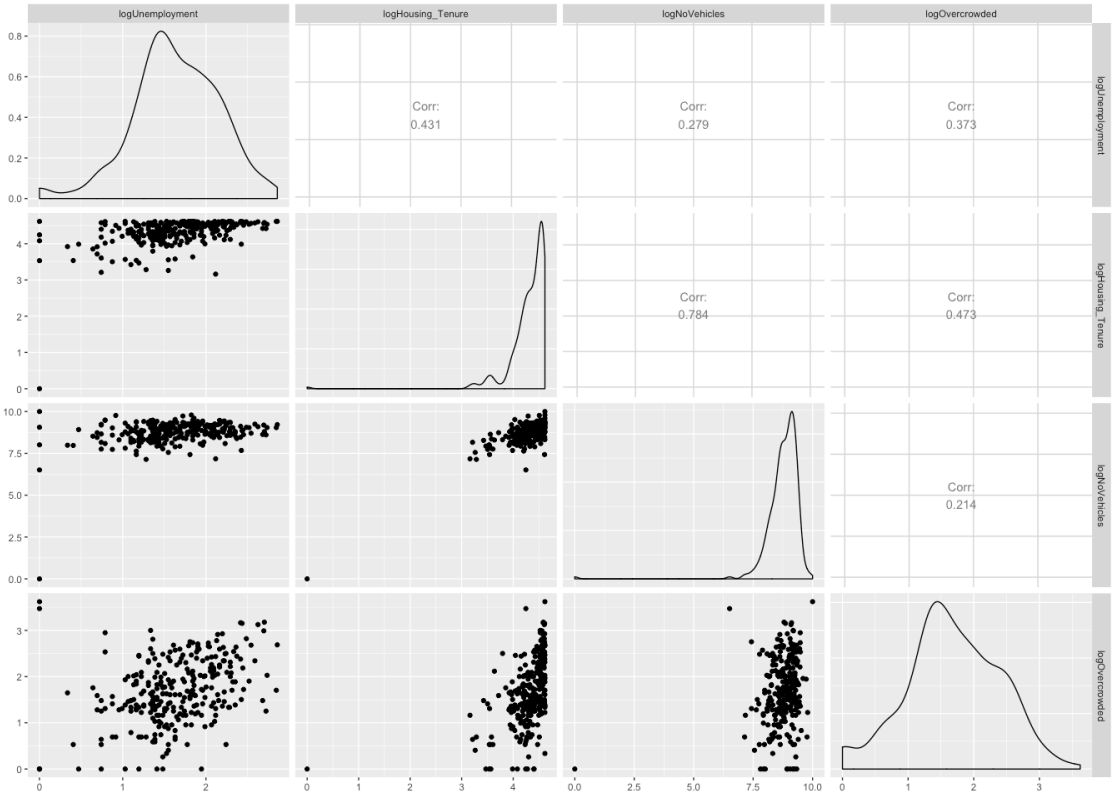


Variables Relationship




Variables Relationship
No_Vehicles
Unemployment
100
75
50
25
0
0
5
10
15
20
25


Log Variables Relationship
LogHousingTenure
LogNoVehicles
4
3
2
1
0
0.0
2.5
5.0
7.5
10.0




```
##   logUnemployment logHousing_Tenure logNoVehicles logOvercrowded
## 2       0.7884574          4.609162      9.486833       2.533697
## 3       2.1162555          4.341205      8.596511       1.308333
## 5       1.5686159          4.587006      8.689074       2.341806
## 6       1.5686159          4.395683      9.327379       2.533697
## 7       1.9021075          4.599152      9.121403       2.708050
## 8       1.1939225          4.502029      8.740709       1.335001
```

```
##     logUnemployment logHousing_Tenure logNoVehicles logOvercrowded
## 281        2.415914          4.612146      8.467585      2.3795461
## 282        2.128232          4.536891      8.671793      2.7788193
## 283        1.960095          4.070735      7.987490      1.7047481
## 284        2.230014          4.533674      8.087027      2.4932055
## 286        1.163151          4.562263      8.774964      2.0412203
## 287        1.458615          4.144721      7.930952      0.9932518
```

```
##                   logUnemployment logHousing_Tenure logNoVehicles
## logUnemployment         1.0000000         0.4311138     0.2788261
## logHousing_Tenure       0.4311138         1.0000000     0.7842310
## logNoVehicles           0.2788261         0.7842310     1.0000000
## logOvercrowded          0.3725944         0.4730620     0.2141658
##                   logOvercrowded
## logUnemployment        0.3725944
## logHousing_Tenure      0.4730620
## logNoVehicles          0.2141658
## logOvercrowded         1.0000000
```

ENTER YOUR ANSWER HERE.

It cannot be seen any relationship between raw data. However, after adjustment of data, from the correlation matrix, it shows these four variables are related.

Question 6

Solution:

```
##check data
##znew
mean.unemp<- mean(znew$logUnemployment,na.rm = TRUE)
mean.oc<- mean(znew$logOvercrowded,na.rm = TRUE)
mean.rent <- mean(znew$logHousing_Tenure,na.rm = TRUE)
mean.car <- mean(znew$logNoVehicles,na.rm = TRUE)
std.unemp <- sd(znew$logUnemployment,na.rm = TRUE)
std.oc <- sd(znew$logOvercrowded,na.rm = TRUE)
std.rent <- sd(znew$logHousing_Tenure,na.rm = TRUE)
std.car <- sd(znew$logNoVehicles,na.rm = TRUE)
standard.unemp <- (znew$logUnemployment-mean.unemp)/std.unemp
standard.oc<-(znew$logOvercrowded-mean.oc)/std.oc
standard.rent<-(znew$logHousing_Tenure-mean.rent)/std.rent
standard.car <- (znew$logNoVehicles-mean.car)/std.car
townsendIndex <-
data.frame(standard.unemp,standard.oc,standard.rent,standard.car)
head(townsendIndex)

##   standard.unemp standard.oc standard.rent standard.car
## 1             NA          NA            NA           NA
## 2     -1.5737101   1.1901476    0.75706986    0.9881078
## 3      0.8781135  -0.5281221    0.06591915   -0.2366199
## 4             NA          NA            NA           NA
## 5     -0.1331213   0.9210680    0.69992227   -0.1092906
## 6     -0.1331213   1.1901476    0.20643666    0.7687628

townsendIndex<-data.frame(znew$Geography,townsendIndex,
                          townsend = rowSums(townsendIndex))
head(townsendIndex)

##                                 znew.Geography standard.unemp standard.oc
## 1    Census Tract 1, New York County, New York             NA          NA
## 2 Census Tract 2.01, New York County, New York     -1.5737101   1.1901476
## 3 Census Tract 2.02, New York County, New York      0.8781135  -0.5281221
## 4    Census Tract 5, New York County, New York             NA          NA
## 5    Census Tract 6, New York County, New York     -0.1331213   0.9210680
## 6    Census Tract 7, New York County, New York     -0.1331213   1.1901476
##   standard.rent standard.car  townsend
## 1            NA           NA        NA
## 2    0.75706986    0.9881078 1.3616151
## 3    0.06591915   -0.2366199 0.1792907
## 4            NA           NA        NA
```

```
## 5    0.69992227   -0.1092906 1.3785783
## 6    0.20643666    0.7687628 2.0322258
```

```
tail(townsendIndex)
```

```
##                                    znew.Geography standard.unemp
## 283    Census Tract 307, New York County, New York      0.5897575
## 284    Census Tract 309, New York County, New York      1.0881731
## 285    Census Tract 311, New York County, New York             NA
## 286 Census Tract 317.03, New York County, New York     -0.8818262
## 287 Census Tract 317.04, New York County, New York     -0.3362416
## 288    Census Tract 319, New York County, New York             NA
##     standard.oc standard.rent standard.car   townsend
## 283  0.02775223    -0.6317118 -1.074389198 -1.0885913
## 284  1.13336852     0.5623614 -0.937466817  1.8464362
## 285          NA            NA           NA         NA
## 286  0.49957126     0.6361006  0.008860822  0.2627065
## 287 -0.96994526    -0.4408774 -1.152163179 -2.8992275
## 288          NA            NA           NA         NA
```

```
nrow(drop_na(townsendIndex))
```

```
## [1] 280
```

ENTER YOUR ANSWER HERE.

There are only 280 tracts could be compute with Townsend Index The row with NA value will be become NA in townsend column because the standardize process would make it become NA with NA minus a number

Question 7

Solution:

```
head(townsendIndex)
```

```
##                                 znew.Geography standard.unemp standard.oc
## 1    Census Tract 1, New York County, New York             NA          NA
## 2 Census Tract 2.01, New York County, New York     -1.5737101   1.1901476
## 3 Census Tract 2.02, New York County, New York      0.8781135  -0.5281221
## 4    Census Tract 5, New York County, New York             NA          NA
## 5    Census Tract 6, New York County, New York     -0.1331213   0.9210680
## 6    Census Tract 7, New York County, New York     -0.1331213   1.1901476
##   standard.rent standard.car  townsend
## 1            NA           NA        NA
## 2    0.75706986    0.9881078 1.3616151
## 3    0.06591915   -0.2366199 0.1792907
## 4            NA           NA        NA
## 5    0.69992227   -0.1092906 1.3785783
## 6    0.20643666    0.7687628 2.0322258
```

```
head(arrange(townsendIndex,townsend))
```

```
##                                    znew.Geography standard.unemp
## 1 Census Tract 217.03, New York County, New York     -3.0296230
## 2 Census Tract 112.02, New York County, New York     -3.0296230
## 3    Census Tract 122, New York County, New York     -2.2809181
## 4 Census Tract 160.01, New York County, New York     -0.6643359
## 5  Census Tract 14.01, New York County, New York     -1.6596109
## 6 Census Tract 150.01, New York County, New York     -1.1283963
##   standard.oc standard.rent standard.car   townsend
## 1   -2.362735    -11.131476  -12.0619864 -28.585820
## 2   -2.362735     -2.020704   -1.0399991  -8.453061
## 3   -1.618660     -2.013173   -1.0916247  -7.004376
## 4   -1.390768     -2.659248   -2.2382324  -6.952584
## 5   -1.462694     -2.849654   -0.8358709  -6.807829
## 6   -2.362735     -1.931753   -1.3182014  -6.741085
```

The Census Tract 217.03 is the least deprived Tract and the Census Tract 285 is the most deprived Tract. In the most deprived Tract, the unemployment rate is higher than almost 95% of the all tracts considering the standardized unemployment rate is 2.32. and the overcrowed rate is also higher than average. Therefore, I don't want live in such place.

Question 8

Solution:

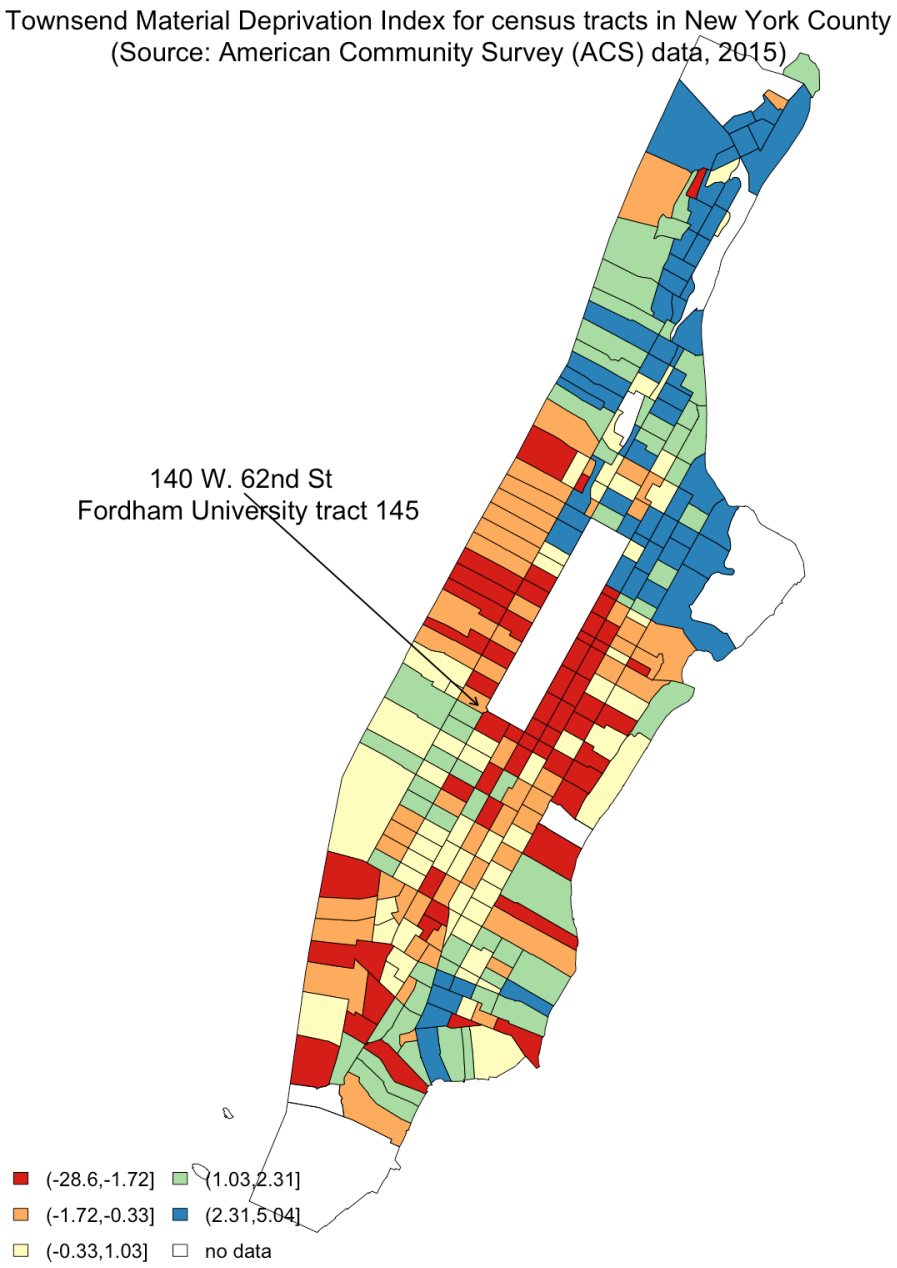
The margin of error is a statistic expressing the amount of random sampling error in a survey's results. We ignore these errors because there are small enough to have no influence on our data estimate.

Question 9

Solution:

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/wusuyi/Desktop/18Fall_Fordham/STAT METHODS &
COMI/Assignment/Homework2/tl_2015_36_tract", layer: "tl_2015_36_tract"
## with 4918 features
## It has 12 fields
## Integer64 fields read as strings:  ALAND AWATER

##   [1] 244 242 243 266 249 275 267 253 246 255 264 272 257 271 270 260 265
##  [18] 245 254 250 279 277 278 282 284 170  84 280 281 276 107 112 113 108
##  [35] 109 106 153 149 101  51  38 285 154 198 230 156 288 207 172 140 126
##  [52] 215  82  37  97  54 206 194 241 256  16 167 283 157  48 118 111 209
##  [69] 229 161  91 181 240  65 139 127 187  73  33  76  42  14  79  20 273
##  [86] 233 263   9 179 251 148  31 114  95  35 132 197 258 120 105  62  58
## [103] 196 186 212 269  89 164 214  81 142  71 110  50  24   7 262 222 286
## [120]  45 247 160 227 124 223  13 259 268   2 145  11 137   5 274  10   6
## [137]  12   3   8   1   4 261  86 252 239 220 287 248  22  26  25  41  21
## [154]  18  23  17  19  15  28  36  30  44  39  43  27  34  40  32  29  85
## [171]  69  46  60  52  63  53  55  56  57  67  66  64  68  72  49  70  47
```

```
## [188]  61  59  80  87  83  75  88  74  77  78  93  98  96 102 100  94  92
## [205]  90 103  99 104 115 116 123 121 122 125 117 119 134 128 135 129 131
## [222] 138 133 130 136 147 150 144 166 151 158 162 163 168 152 141 159 165
## [239] 146 155 143 182 188 177 184 180 183 173 189 175 178 174 185 169 192
## [256] 195 205 203 201 200 191 202 193 190 204 199 210 225 216 228 213 235
## [273] 232 221 219 208 231 211 224 226 218 217 236 238 234 237 176 171

## [1] 288   5

##         min       max
## x -74.04729 -73.90700
## y  40.67955  40.88221
```




```
##         min       max
## x -74.04729 -73.90700
## y  40.67955  40.88221

## quartz_off_screen
##                 2
```

As what I know , the upper east town would be so call “rich area”,which is same to my output graph. In the graph, the upper east town is red majority which means this area is less deprived area than other area in New York County. the Mid West Town is an average area which is also fit to my knowledge about where I live. The middle area of rectangualr is the central park which is no data.

Question 10

Solution:

```
## add a row of rank of townsend
townsendIndex <-
townsendIndex%>%arrange(desc(townsend))%>%mutate(no=rownames(townsendIndex))
head(townsendIndex) ## check the rank
```

```
##     znew.Geography standard.unemp standard.oc standard.rent standard.car
## 1 Census Tract 285       2.244945    1.406306     0.7724383   0.62042548
## 2 Census Tract 261       1.816915    2.021773     0.7441925   0.44766321
## 3 Census Tract 249       1.946096    1.831009     0.6280107   0.58306925
## 4 Census Tract 291       1.958531    2.093699     0.6920304   0.04017306
## 5 Census Tract 251       1.431442    2.076060     0.6867557   0.47787980
## 6 Census Tract 293       1.464125    2.058197     0.7570699   0.12586692
##   townsend city_group_equal color_city_equal no
## 1 5.044114      (2.31,5.04]          #2B83BA  1
## 2 5.030544      (2.31,5.04]          #2B83BA  2
## 3 4.988185      (2.31,5.04]          #2B83BA  3
## 4 4.784433      (2.31,5.04]          #2B83BA  4
## 5 4.672138      (2.31,5.04]          #2B83BA  5
## 6 4.405259      (2.31,5.04]          #2B83BA  6
```

```
filter(townsendIndex, znew.Geography == "Census Tract 145")
```

```
##     znew.Geography standard.unemp standard.oc standard.rent standard.car
## 1 Census Tract 145     -0.4697842  -0.9189486     -0.267013   0.06360417
##    townsend city_group_equal color_city_equal  no
## 1 -1.592142    (-1.72,-0.33]          #FDAE61 218
```

The census tract of 140W 62nd St. is 145 Tract This is the AmericanFactFinder,where we can search for the Census Tract of Fordham Univerisity. The rank of Cunsus Tract 145 is No. 218 from most deprived tract to lest deprived tract

Question 11

Solution:

We should not combine New York County to other county in New York States because the New York City is really different from other county considering its economy, location and business condition.