

ols

dain

2021 8 16

```
#필요한 라이브러리
library(data.table)
library(carData)
library(car)
library(corrplot)
```

```
## corrplot 0.90 loaded
```

```
library(magrittr)
library(ggplot2)
library(lmtest)
```

```
## 필요한 패키지를 로딩중입니다: zoo
```

```
##
## 다음의 패키지를 부착합니다: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##   as.Date, as.Date.numeric
```

```
library(AER)
```

```
## 필요한 패키지를 로딩중입니다: sandwich
```

```
## 필요한 패키지를 로딩중입니다: survival
```

```
library(MASS)
library(dplyr)
```

```
##
## 다음의 패키지를 부착합니다: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
##
##   select
```

```
## The following object is masked from 'package:car':
##
##   recode
```

```
## The following objects are masked from 'package:data.table':  
##  
##   between, first, last
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(sp)  
library(sf)
```

```
## Linking to GEOS 3.9.0, GDAL 3.2.1, PROJ 7.2.1
```

```
library(tmap)  
library(RANN)  
library(spData)  
library(spdep)  
library(rgdal)
```

```
## rgdal: version: 1.5-23, (SVN revision 1121)  
## Geospatial Data Abstraction Library extensions to R successfully loaded  
## Loaded GDAL runtime: GDAL 3.2.1, released 2020/12/29  
## Path to GDAL shared files: C:/Users/user/Documents/R/win-library/4.1/rgdal/gdal  
## GDAL binary built with GEOS: TRUE  
## Loaded PROJ runtime: Rel. 7.2.1, January 1st, 2021, [PJ_VERSION: 721]  
## Path to PROJ shared files: C:/Users/user/Documents/R/win-library/4.1/rgdal/proj  
## PROJ CDN enabled: FALSE  
## Linking to sp version:1.4-5  
## To mute warnings of possible GDAL/OSR exportToProj4() degradation,  
## use options("rgdal_show_exportToProj4_warnings"="none") before loading rgdal.  
## Overwritten PROJ_LIB was C:/Users/user/Documents/R/win-library/4.1/rgdal/proj
```

```
library(RColorBrewer)  
library(leaflet)  
library(maptools)
```

```
## Checking rgeos availability: TRUE
```

```
library(gvlma)  
library(tidyr)
```

```
##  
## 다음의 패키지를 부착합니다: 'tidyr'
```

```
## The following object is masked from 'package:magrittr':  
##  
##     extract
```

```
library(leaps)  
library(broom)  
library(rsq)  
library(gridExtra)
```

```
##  
## 다음의 패키지를 부착합니다: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':  
##  
##     combine
```

```
library(spgwr)
```

```
## NOTE: This package does not constitute approval of GWR  
## as a method of spatial analysis; see example(gwr)
```

```
library(spatialreg)
```

```
## 필요한 패키지를 로딩중입니다: Matrix
```

```
##  
## 다음의 패키지를 부착합니다: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':  
##  
##     expand, pack, unpack
```

```
##  
## 다음의 패키지를 부착합니다: 'spatialreg'
```

```
## The following objects are masked from 'package:spdep':  
##  
##     as.spam.listw, as_dgRMatrix_listw, as_dsCMatrix_I,  
##     as_dsCMatrix_IrW, as_dsTMatrix_listw, can.be.simmed, cheb_setup,  
##     create_WX, do_ldet, eigen_pre_setup, eigen_setup, eigenw,  
##     errorsarlm, get.ClusterOption, get.coresOption, get.mcOption,  
##     get.VerboseOption, get.ZeroPolicyOption, GMargminImage, GMerrorsar,  
##     griffith_sone, gstsIs, Hausman.test, impacts, intImpacts,  
##     Jacobian_W, jacobianSetup, l_max, lagmess, lagsarlm, lextB,  
##     lextS, lextW, lmSLX, LU_prepermutate_setup, LU_setup,  
##     Matrix_J_setup, Matrix_setup, mcdet_setup, MCMCsamp, ME, mom_calc,  
##     mom_calc_int2, moments_setup, powerWeights, sacsarlm,
```

```
## SE_classic_setup, SE_interp_setup, SE_whichMin_setup,
## set.ClusterOption, set.coresOption, set.mcOption,
## set.VerboseOption, set.ZeroPolicyOption, similar.listw, spam_setup,
## spam_update_setup, SpatialFiltering, spautolm, spBreg_err,
## spBreg_lag, spBreg_sac, stsls, subgraph_eigenw, trW
```

```
library(geojsonio)
```

```
## Registered S3 method overwritten by 'geojsonsf':
## method from
## print.geojson geojson
```

```
##
## 다음의 패키지를 부착합니다: 'geojsonio'
```

```
## The following object is masked from 'package:base':
##
## pretty
```

```
gpclibPermit
```

```
## function ()
## {
##   if ("gpclib" %in% .packages(all.available = TRUE))
##     assign("gpclib", TRUE, envir = .MAPTOOLS_CACHE)
##   if (gpclibPermitStatus())
##     warning("support for gpclib will be withdrawn from maptools at the next major rele
ase")
##   gpclibPermitStatus()
## }
## <bytecode: 0x000000002a54b1f8>
## <environment: namespace:maptools>
```

```
library(maptools)
gpclibPermit
```

```
## function ()
## {
##   if ("gpclib" %in% .packages(all.available = TRUE))
##     assign("gpclib", TRUE, envir = .MAPTOOLS_CACHE)
##   if (gpclibPermitStatus())
##     warning("support for gpclib will be withdrawn from maptools at the next major rele
ase")
##   gpclibPermitStatus()
## }
## <bytecode: 0x000000002a54b1f8>
## <environment: namespace:maptools>
```

```
final_data <- st_read('D:/data/seoul_exam_re.geojson')
```

```
## Reading layer `seoul_exam_re' from data source `D:\data\seoul_exam_re.geojson' using driver `GeoJSON'
## Simple feature collection with 19153 features and 18 fields
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: 126.7645 ymin: 37.4283 xmax: 127.1838 ymax: 37.70146
## Geodetic CRS: WGS 84
```

```
colnames(final_data)
```

```
## [1] "TOT_REG_CD"      "ADM_NM"          "ADM_CD"          "면적"
## [5] "X"              "Y"              "공공기관"        "보육시설"
## [9] "사회복지시"     "지하철.1km이내." "지하철거리점수"  "승하차평균"
## [13] "구별총생활"     "구별총생활인구"  "상가수"          "어린이집개수"
## [17] "버스정류장수"    "버스승하차"      "geometry"
```

```
head(final_data)
```

```
## Simple feature collection with 6 features and 18 fields
## Geometry type: MULTIPOLYGON
## Dimension: XY
## Bounding box: xmin: 126.9611 ymin: 37.573 xmax: 126.9746 ymax: 37.57954
## Geodetic CRS: WGS 84
##      TOT_REG_CD ADM_NM ADM_CD 면적      X      Y 공공기관 보육시설
## 1 1101053010006 사직동 1101053 18424 126.9703 37.57830      8      2
## 2 1101053010001 사직동 1101053 18821 126.9652 37.57415      6      4
## 3 1101053010003 사직동 1101053 218981 126.9655 37.57651     10      6
## 4 1101053010002 사직동 1101053 37082 126.9690 37.57709      7      2
## 5 1101053010005 사직동 1101053 70996 126.9728 37.57803     11      2
## 6 1101053010004 사직동 1101053 44920 126.9717 37.57664     10      2
## 사회복지시 지하철.1km이내. 지하철거리점수 승하차평균 구별총생활
## 1      0      2      27      52315      367654
## 2      0      4      24      0      277143
## 3      0      4      26      34690      1245899
## 4      0      3      27      52315      1373417
## 5      0      3      29      52315      1038484
## 6      0      4      29      52315      1515584
## 구별총생활인구 상가수 어린이집개수 버스정류장수 버스승하차
## 1 11859.80645      29      8      2 109.516129
## 2 8940.096774      7      6      3 196.5591398
## 3 40190.29032     37     10     12 497.1962366
## 4 44303.77419     74      7      2 428.1774194
## 5 33499.48387    122     11      6 2136.021505
## 6 48889.80645    113     10      5 3152.083871
##
##      geometry
## 1 MULTIPOLYGON (((126.9703 37...
## 2 MULTIPOLYGON (((126.9661 37...
## 3 MULTIPOLYGON (((126.9665 37...
## 4 MULTIPOLYGON (((126.9688 37...
## 5 MULTIPOLYGON (((126.974 37...
## 6 MULTIPOLYGON (((126.9714 37...
```

```
str(final_data)
```

```
## Classes 'sf' and 'data.frame': 19153 obs. of 19 variables:
## $ TOT_REG_CD : chr "1101053010006" "1101053010001" "1101053010003" "1101053010002"
## ...
## $ ADM_NM : chr "사직동" "사직동" "사직동" "사직동" ...
## $ ADM_CD : chr "1101053" "1101053" "1101053" "1101053" ...
## $ 면적 : int 18424 18821 218981 37082 70996 44920 83191 513924 37110 40788 ...
## $ X : num 127 127 127 127 127 ...
## $ Y : num 37.6 37.6 37.6 37.6 37.6 ...
## $ 공공기관 : chr "8" "6" "10" "7" ...
## $ 보육시설 : chr "2" "4" "6" "2" ...
## $ 사회복지시 : chr "0" "0" "0" "0" ...
## $ 지하철.1km이내.: chr "2" "4" "4" "3" ...
## $ 지하철거리점수 : chr "27" "24" "26" "27" ...
## $ 승하차평균 : chr "52315" "0" "34690" "52315" ...
## $ 구별총생활 : chr "367654" "277143" "1245899" "1373417" ...
## $ 구별총생활인구 : chr "11859.80645" "8940.096774" "40190.29032" "44303.77419" ...
## $ 상가수 : chr "29" "7" "37" "74" ...
## $ 어린이집개수 : chr "8" "6" "10" "7" ...
## $ 버스정류장수 : chr "2" "3" "12" "2" ...
## $ 버스승하차 : chr "109.516129" "196.5591398" "497.1962366" "428.1774194" ...
## $ geometry : sfc_MULTIPOLYGON of length 19153; first list element: List of 1
## ..$ :List of 1
## .. ..$ : num [1:136, 1:2] 127 127 127 127 127 ...
## ..- attr(*, "class")= chr [1:3] "XY" "MULTIPOLYGON" "sfg"
## - attr(*, "sf_column")= chr "geometry"
## - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",...: NA NA NA NA NA NA NA NA NA NA NA ...
## ..- attr(*, "names")= chr [1:18] "TOT_REG_CD" "ADM_NM" "ADM_CD" "면적" ...
```

*#chr*이라서 바꾼거임

```
final_data$공공기관 <- as.numeric(final_data$공공기관)
final_data$보육시설 <- as.numeric(final_data$보육시설)
final_data$사회복지시 <- as.numeric(final_data$사회복지시)
final_data$버스정류장 <- as.numeric(final_data$버스정류장)
final_data$버스승하차 <- as.numeric(final_data$버스승하차)
final_data$지하철.1km이내. <- as.numeric(final_data$지하철.1km이내.)
final_data$지하철거리점수 <- as.numeric(final_data$지하철거리점수)
```

```
## Warning: 강제형변환에 의해 생성된 NA 입니다
```

```
final_data$승하차평균 <- as.numeric(final_data$승하차평균)
final_data$구별총생활 <- as.numeric(final_data$구별총생활)
final_data$구별총생활인구 <- as.numeric(final_data$구별총생활인구)
final_data$상가수 <- as.numeric(final_data$상가수)
final_data$어린이집개수 <- as.numeric(final_data$어린이집개수)
```

```
str(final_data)
```

```
## Classes 'sf' and 'data.frame': 19153 obs. of 20 variables:
## $ TOT_REG_CD : chr "1101053010006" "1101053010001" "1101053010003" "1101053010002"
## ...
```

```
## $ ADM_NM      : chr  "사직동" "사직동" "사직동" "사직동" ...
## $ ADM_CD      : chr  "1101053" "1101053" "1101053" "1101053" ...
## $ 면적        : int   18424 18821 218981 37082 70996 44920 83191 513924 37110 40788 ...
## $ X           : num   127 127 127 127 127 ...
## $ Y           : num   37.6 37.6 37.6 37.6 37.6 ...
## $ 공공기관    : num    8 6 10 7 11 10 9 28 7 9 ...
## $ 보육시설    : num    2 4 6 2 2 2 4 7 7 2 ...
## $ 사회복지시  : num    0 0 0 0 0 0 0 1 1 0 ...
## $ 지하철.1km이내.: num    2 4 4 3 3 4 6 8 4 6 ...
## $ 지하철거리점수 : num   27 24 26 27 29 29 26 29 24 28 ...
## $ 승하차평균   : num  52315 0 34690 52315 52315 ...
## $ 구별총생활   : num  367654 277143 1245899 1373417 1038484 ...
## $ 구별총생활인구 : num  11860 8940 40190 44304 33499 ...
## $ 상가수       : num    29 7 37 74 122 113 53 424 6 118 ...
## $ 어린이집개수 : num    8 6 10 7 11 10 9 28 7 9 ...
## $ 버스정류장수 : chr   "2" "3" "12" "2" ...
## $ 버스승하차   : num   110 197 497 428 2136 ...
## $ geometry     :sfc_MULTIPOLYGON of length 19153; first list element: List of 1
## ..$ :List of 1
## .. ..$ : num [1:136, 1:2] 127 127 127 127 127 ...
## ..- attr(*, "class")= chr [1:3] "XY" "MULTIPOLYGON" "sfg"
## $ 버스정류장   : num    2 3 12 2 6 5 0 22 1 0 ...
## - attr(*, "sf_column")= chr "geometry"
## - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",...: NA NA NA NA NA NA NA NA NA NA NA ...
## ..- attr(*, "names")= chr [1:19] "TOT_REG_CD" "ADM_NM" "ADM_CD" "면적" ...
```

#변수별 na확인하는 함수

```
apply(final_data,2,function(x) sum(is.na(x)))
```

	TOT_REG_CD	ADM_NM	ADM_CD	면적	X
	0	0	0	0	0
	Y	공공기관	보육시설	사회복지시	지하철.1km이내.
	0	0	0	0	0
	지하철거리점수	승하차평균	구별총생활	구별총생활인구	상가수
	10	0	0	0	10
	어린이집개수	버스정류장수	버스승하차	geometry	버스정류장
	415	0	0	0	0

```
final_data$상가수[is.na(final_data$상가수)] <- 0
final_data$지하철거리점수[is.na(final_data$지하철거리점수)] <- 0 #이래도 되나
final_data$어린이집개수[is.na(final_data$어린이집개수)] <- 0
sum(is.na(final_data))
```

```
## [1] 0
```

#결측치 분포

```
library(mice)
```

```
##
## 다음의 패키지를 부착합니다: 'mice'
```

```
## The following object is masked from 'package:stats':  
##  
## filter
```

```
## The following objects are masked from 'package:base':  
##  
## cbind, rbind
```

```
library(VIM)
```

```
## 필요한 패키지를 로딩중입니다: colorspace
```

```
## 필요한 패키지를 로딩중입니다: grid
```

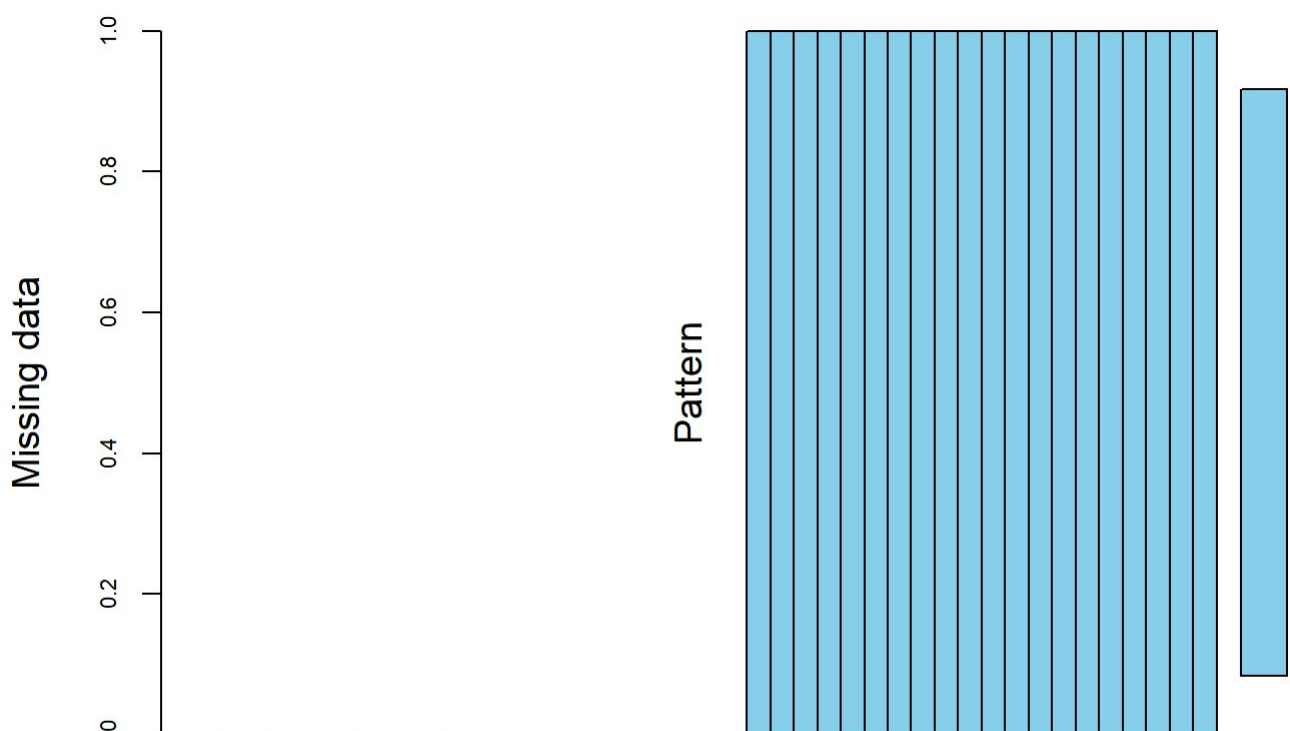
```
## VIM is ready to use.
```

```
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
```

```
##  
## 다음의 패키지를 부착합니다: 'VIM'
```

```
## The following object is masked from 'package:datasets':  
##  
## sleep
```

```
aggr(final_data, cex.axis=.7, gap=3, ylab=c("Missing data", "Pattern"))
```



0

TOT_REG_CD
ADM_NM
ADM_CD
면적
X
Y
공공기관
보육시설
사회복지시
지하철.1km이내.
지하철거리점수
승하차평균
구별총생활
구별총생활인구
상가수
어린이집개수
버스정류장수
버스승하차
geometry
버스정류장

TOT_REG_CD
ADM_NM
ADM_CD
면적
X
Y
공공기관
보육시설
사회복지시
지하철.1km이내.
지하철거리점수
승하차평균
구별총생활
구별총생활인구
상가수
어린이집개수
버스정류장수
버스승하차
geometry
버스정류장

```
str(final_data)
```

```
## Classes 'sf' and 'data.frame':  19153 obs. of  20 variables:
## $ TOT_REG_CD      : chr  "1101053010006" "1101053010001" "1101053010003" "1101053010002"
## ...
## $ ADM_NM          : chr  "사직동" "사직동" "사직동" "사직동" ...
## $ ADM_CD          : chr  "1101053" "1101053" "1101053" "1101053" ...
## $ 면적            : int  18424 18821 218981 37082 70996 44920 83191 513924 37110 40788 ...
## $ X               : num  127 127 127 127 127 ...
## $ Y               : num  37.6 37.6 37.6 37.6 37.6 ...
## $ 공공기관        : num  8 6 10 7 11 10 9 28 7 9 ...
## $ 보육시설        : num  2 4 6 2 2 2 4 7 7 2 ...
## $ 사회복지시      : num  0 0 0 0 0 0 0 1 1 0 ...
## $ 지하철.1km이내.: num  2 4 4 3 3 4 6 8 4 6 ...
## $ 지하철거리점수 : num  27 24 26 27 29 29 26 29 24 28 ...
## $ 승하차평균      : num  52315 0 34690 52315 52315 ...
## $ 구별총생활      : num  367654 277143 1245899 1373417 1038484 ...
## $ 구별총생활인구 : num  11860 8940 40190 44304 33499 ...
## $ 상가수          : num  29 7 37 74 122 113 53 424 6 118 ...
## $ 어린이집개수    : num  8 6 10 7 11 10 9 28 7 9 ...
## $ 버스정류장수    : chr  "2" "3" "12" "2" ...
## $ 버스승하차      : num  110 197 497 428 2136 ...
## $ geometry        :sfc_MULTIPOLYGON of length 19153; first list element: List of 1
## ..$ :List of 1
## .. ..$ : num [1:136, 1:2] 127 127 127 127 127 ...
## ..- attr(*, "class")= chr [1:3] "XY" "MULTIPOLYGON" "sfg"
## $ 버스정류장      : num  2 3 12 2 6 5 0 22 1 0 ...
## - attr(*, "sf_column")= chr "geometry"
## - attr(*, "agr")= Factor w/ 3 levels "constant","aggregate",...: NA NA NA NA NA NA NA NA NA NA NA ...
## ..- attr(*, "names")= chr [1:19] "TOT_REG_CD" "ADM_NM" "ADM_CD" "면적" ...
```

```
final_data_1 <- final_data %>% select (TOT_REG_CD, 버스정류장, 공공기관, 보육시설, 사회복지시, 어린이집개수, 버스승하차, 지하철.1km이내., 승하차평균, 지하철거리점수, 상가수, 구별총생활인구)
```

```
final_data_1$시설수 <- final_data_1$공공기관 + final_data_1$보육시설 + final_data_1$사회복지시+
final_data_1$어린이집개수
head(final_data_1$시설수)
```

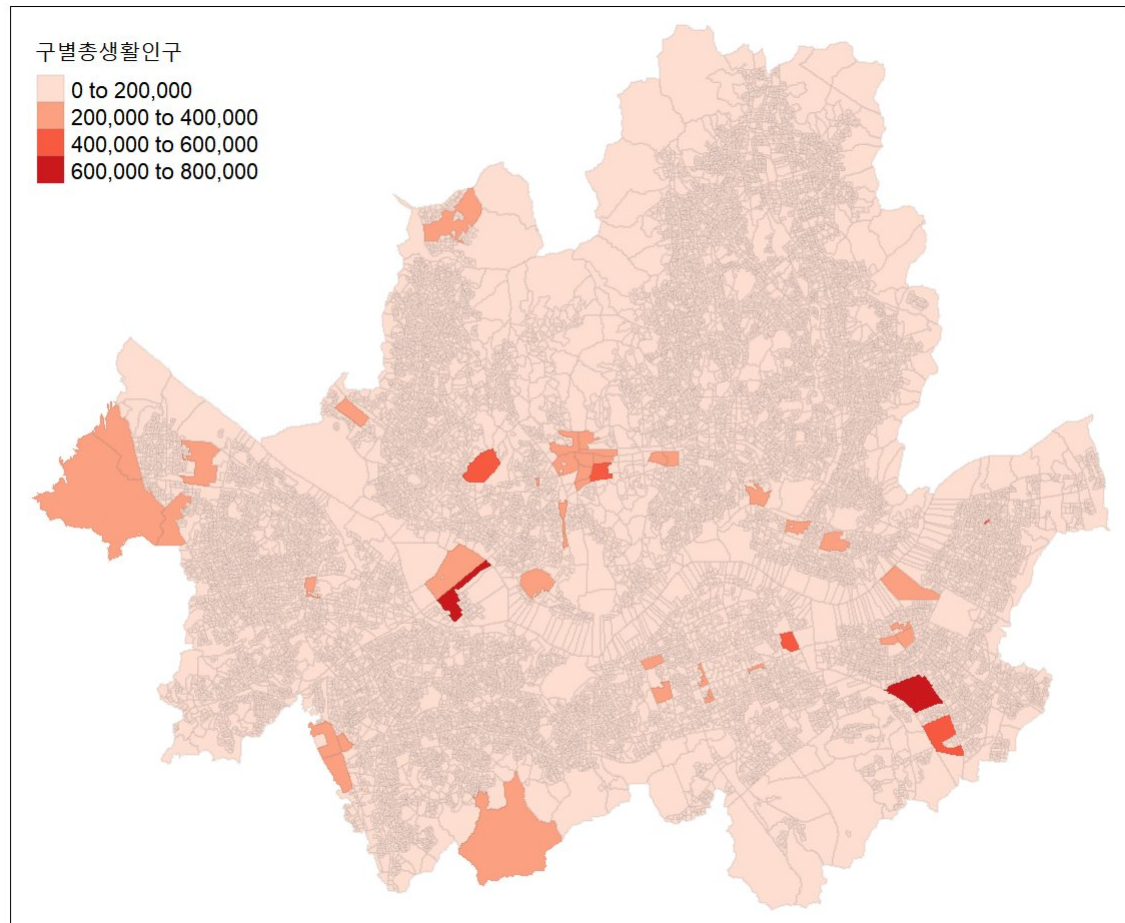
```
## [1] 18 16 26 16 24 22
```

```
options(repr.plot.width = 10, repr.plot.height = 10)
tmap_mode("plot")
```

```
## tmap mode set to plotting
```

```
tmap_options(check.and.fix = TRUE)
tm_shape(final_data_1) +tm_fill('구별총생활인구', style='pretty', palette='Reds') +tm_borders
(alpha=.1)
```

```
## Warning: The shape final_data_1 is invalid. See sf::st_is_valid
```



```
##공간데이터로 바꾸기
sp_data <- as(final_data_1, 'Spatial')
sp_data
```

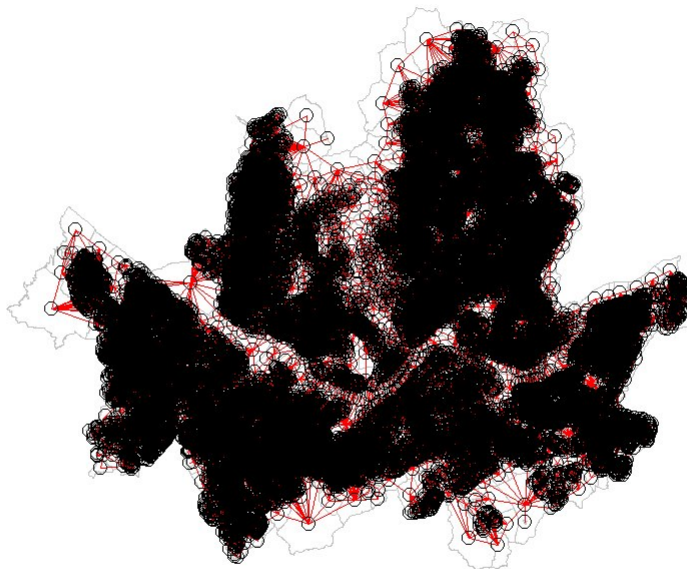
```
## class      : SpatialPolygonsDataFrame
## features   : 19153
## extent    : 126.7645, 127.1838, 37.4283, 37.70146 (xmin, xmax, ymin, ymax)
## crs       : +proj=longlat +datum=WGS84 +no_defs
## variables  : 13
## names      : TOT_REG_CD, 버스정류장, 공공기관, 보육시설, 사회복지시, 어린이집개수, 버스승
하차, 지하철.1km이내., 승하차평균, 지하철거리점수, 상가수, 구별총생활인구, 시설수
## min values : 1101053010001, 0, 0, 0, 0, 0,
0, 0, 0, 0, 18.09677419, 0
## max values : 1125074033301, 42, 40, 15, 34, 40, 145
11.12903, 12, 218065, 29, 2073, 761567.2258, 95
```

```
##이웃(neighbours)정의
w <- poly2nb(sp_data)
```

w

```
## Neighbour list object:  
## Number of regions: 19153  
## Number of nonzero links: 116200  
## Percentage nonzero weights: 0.03167616  
## Average number of links: 6.066935
```

```
# 각자별 이웃관계 시각화  
plot(sp_data, col='white', border='grey', lwd=0.1)  
xy <- coordinates(sp_data)  
  
plot(w, xy, col='red', lwd=0.1, add=TRUE)
```



```
##가중치행렬 생성  
ww <- nb2listw(w, style='W') ## row-standardised matrix  
ww
```

```
## Characteristics of weights list object:  
## Neighbour list object:  
## Number of regions: 19153  
## Number of nonzero links: 116200  
## Percentage nonzero weights: 0.03167616  
## Average number of links: 6.066935  
##  
## Weights style: W
```

```
## Weights constants summary:
##      n      nn      S0      S1      S2
## W 19153 366837409 19153 6758.993 81960.13
```

#####moran's i 검정 (ols가 아닌, 공간모형을 사용해야 하는 이유) 종속변수인 생활의 전역적 공간적 자기 상관성이 있는지 판단하기 위해 Moran's I 검정

```
results <- moran.mc(final_data_1$구별총생활인구, ww, nsim=99) # nsim : # of permutations want
to run
results
```

```
##
## Monte-Carlo simulation of Moran I
##
## data: final_data_1$구별총생활인구
## weights: ww
## number of simulations + 1: 100
##
## statistic = 0.27435, observed rank = 100, p-value = 0.01
## alternative hypothesis: greater
```

Moran's I지수가 0.27435로 유의하게 나타나는 것을 볼 수 있음 귀무가설 : 아무런 패턴이 존재하지 않는다. 귀무가설 기각 => 자기상관성이 강하게 존재한다는 것을 의미

#####LISA분석 - 왜 공간적 모델 중에서도 국지적 모델(GWR)을 써야 하는지에 대한 이유

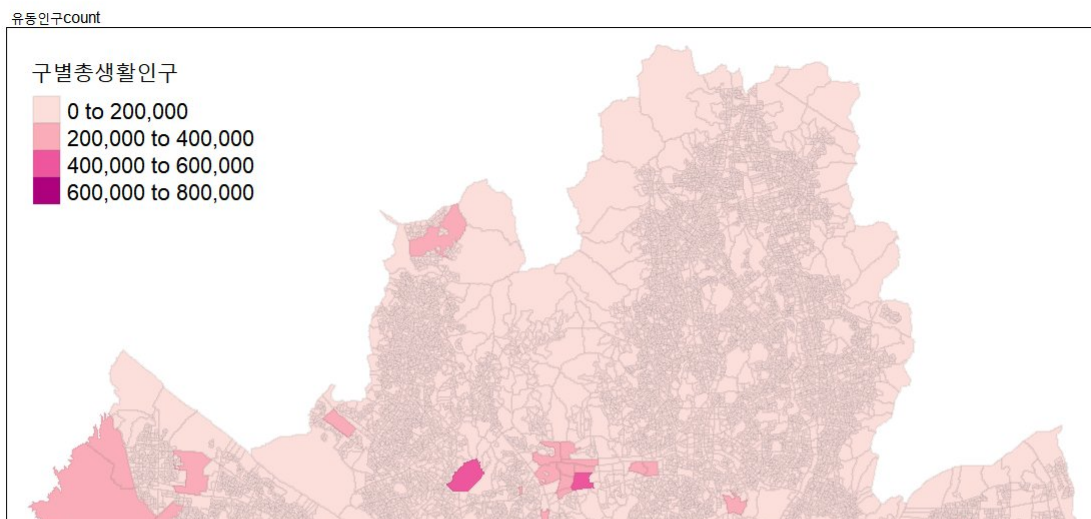
Moran's I는 전역적인 값이며, 보다 국지적으로 어떤 지역들이 서로 유사한 값들을 가지며 클러스터하고 있는지 분석하기 위해 국지적 자기상관성 LISA 분석을 실시

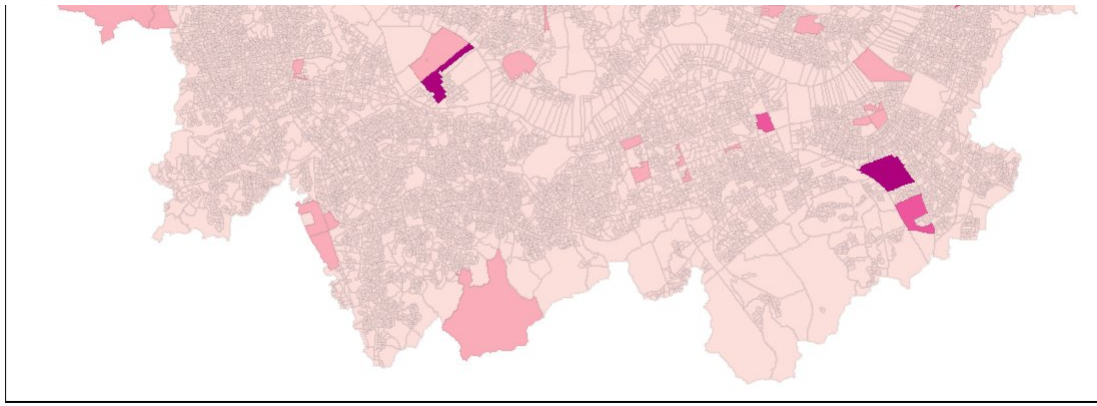
```
tmap_mode('plot')
```

```
## tmap mode set to plotting
```

```
tm_shape(final_data_1) + tm_fill('구별총생활인구', style='pretty', palette='RdPu') +tm_borders
(alpha=.1) + tm_layout(main.title='유동인구count', main.title.size=0.5)
```

```
## Warning: The shape final_data_1 is invalid. See sf::st_is_valid
```





```
local_m <- localmoran(final_data_1$구별총생활인구, ww)
summary(local_m)
```

```
##           Ii           E.Ii           Var.Ii           Z.Ii
## Min.      : -6.64135   Min.      :-5.221e-05   Min.      :0.01971   Min.      : -11.5731
## 1st Qu.: -0.01341   1st Qu.: -5.221e-05   1st Qu.:0.14110   1st Qu.: -0.0317
## Median :  0.02381   Median : -5.221e-05   Median :0.16463   Median :  0.0578
## Mean      :  0.27435   Mean      : -5.221e-05   Mean      :0.18910   Mean      :  0.8044
## 3rd Qu.:  0.09387   3rd Qu.: -5.221e-05   3rd Qu.:0.19757   3rd Qu.:  0.2225
## Max.      :194.47860   Max.      : -5.221e-05   Max.      :0.98804   Max.      :479.3108
## Pr(z > 0)
## Min.      :0.0000
## 1st Qu.:0.4120
## Median :0.4769
## Mean      :0.4551
## 3rd Qu.:0.5126
## Max.      :1.0000
```

```
final_data_1$s_구별총생활인구 <- scale(final_data_1$구별총생활인구) %>% as.vector()
```

```
# spatial lag 변수 생성
final_data_1$lag_s_구별총생활인구 <- lag.listw(ww, final_data_1$구별총생활인구)
```

```
summary(final_data_1$s_구별총생활인구)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
## -0.61462 -0.36462 -0.22248  0.00000  0.03295 33.65883
```

```
summary(final_data_1$lag_s_구별총생활인구)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##   484.6   9471.5 13262.7 18301.4 20150.2 381802.1
```

```
final_data_1 <- st_as_sf(final_data_1) %>%
  mutate(quad_sig = ifelse(final_data_1$s_구별총생활인구 > 0 &
                           final_data_1$lag_s_구별총생활인구 > 0 &
                           local_m[,5] <= 0.05,
                           "high-high",
```



```

ifelse(final_data_1$s_구별총생활인구 <= 0 &
      final_data_1$lag_s_구별총생활인구 <= 0 &
      local_m[,5] <= 0.05,
      "low-low",
      ifelse(final_data_1$s_구별총생활인구 > 0 &
            final_data_1$lag_s_구별총생활인구 <= 0 &
            local_m[,5] <= 0.05,
            "high-low",
            ifelse(final_data_1$s_구별총생활인구 <= 0 &
                  final_data_1$lag_s_구별총생활인구 > 0 &
                  local_m[,5] <= 0.05,
                  "low-high",
                  "non-significant")))))

```

```
table(final_data_1$quad_sig)
```

```
##
##      high-high non-significant
##           962          18191

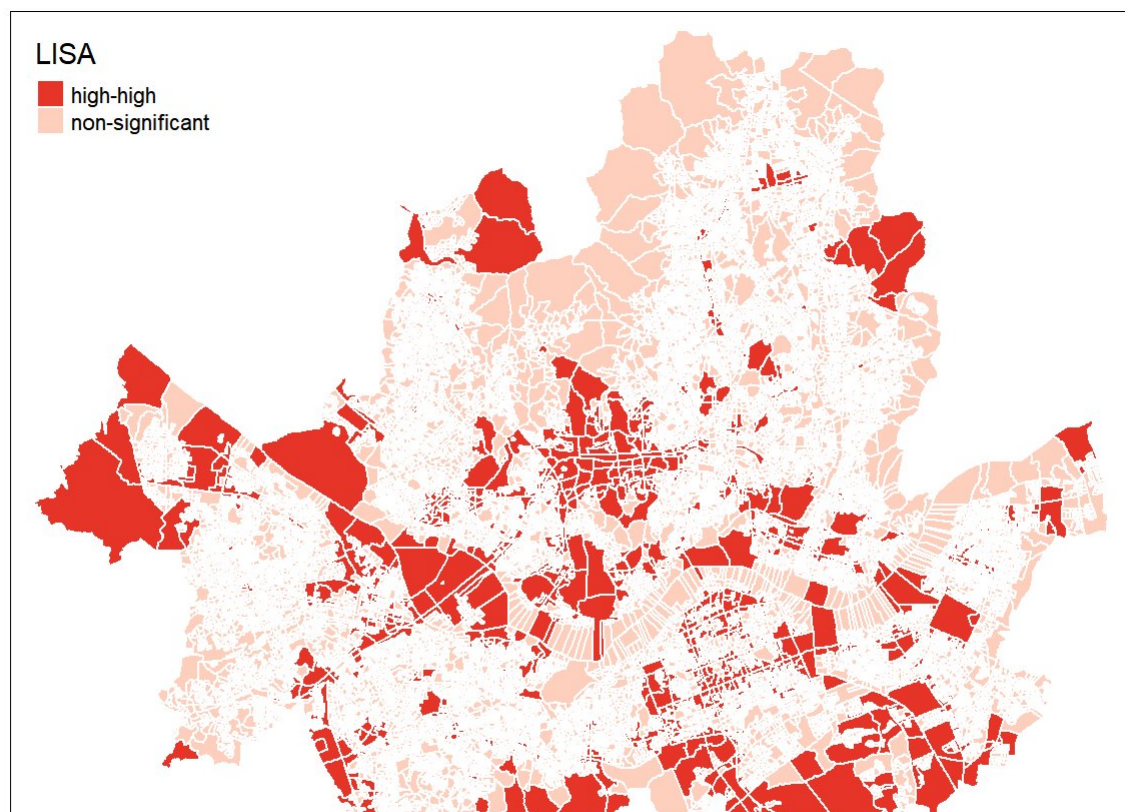
```

```
nrow(local_m[local_m[,5] <= 0.05,])
```

```
## [1] 962
```

```
qtm(final_data_1, fill='quad_sig', fill.title='LISA',
     fill.palette = "-Reds", border='white')
```

```
## Warning: The shape final_data_1 is invalid. See sf::st_is_valid
```





Moran's I 는 전역적인 값이며, 보다 국지적으로 어떤 지역들이 서로 유사한 값들을 갖고 있는지 확인하기 위해 국지적 자기상관성 LISA분석 실시

전역적 : 공간적의존성 국지적 : 공간적의존성 뿐만 아니라 공간적 이질성도 고려

공간의존성 : 공간상의 한 지점에서 측정된 현상이 다른 지점에서의 측정값과 서로 관련이 있을 때 나타난

공간이질성 : 각 지점이나 지역에 내재된 고유성에서 발생하게 되는 자료의 지리적 변이나 통계적 특성의 차이를 말한다

공간가중행렬을 토대로 Moran's I 값을 분석 -> 이를 통해 공간적 상관성 여부를 판단 Moran's I 값은 1에 가까울 수록 공간적 상관성이 큼 0.27435으로 어느 수준의 공간적 자기 상관성이 존재함 교통사고가 발생한 곳에 더욱 사고가 많이 발생하며 공간적 군집을 이루는 것 확인가능 => 이는 공간적자기상관성을 고려하는 공간회귀모형 이 필요하다고 판단가능

으아아아아ㅏㅇ 이제 써도 됨 =====>인데 (전역적 공간회귀모형 vs 국지적 공간회귀모형) gwr은 이제 국 지적 공간회귀모형임....) + 국지적은 : 공간적의존성 뿐만 아니라 공간적 이질성도 고려 가능

#ols회귀분석

```
colnames(final_data_1)
```

```
## [1] "TOT_REG_CD"      "버스정류장"      "공공기관"
## [4] "보육시설"        "사회복지시"     "어린이집개수"
## [7] "버스승하차"       "지하철.1km이내." "승하차평균"
## [10] "지하철거리점수"  "상가수"         "구별총생활인구"
## [13] "geometry"        "시설수"         "s_구별총생활인구"
## [16] "lag_s_구별총생활인구" "quad_sig"
```

```
## 회귀분석에 쓸 컬럼만 데이터프레임으로 변환
```

```
dat <- data.frame(final_data_1) %>% dplyr::select (버스정류장, 시설수, 버스승하차, 지하철.1km이
내., 승하차평균, 지하철거리점수, 상가수, 구별총생활인구)
colnames(dat)
```

```
## [1] "버스정류장"      "시설수"         "버스승하차"      "지하철.1km이내."
## [5] "승하차평균"     "지하철거리점수" "상가수"         "구별총생활인구"
```

```
## OLS식 구하기
```

```
lm_ols1 <- lm(구별총생활인구 ~ . , data=dat)
summary(lm_ols1)
```

```
##
## Call:
## lm(formula = 구별총생활인구 ~ . , data = dat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -281032  -4957    -713    3287  534905
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.112e+03  7.561e+02   5.438 5.46e-08 ***
## 버스정류장     2.198e+03  4.795e+01  45.852 < 2e-16 ***
## 시설수        7.310e+01  1.395e+01   5.240 1.62e-07 ***
## 버스승하차     1.056e+00  1.317e-01   8.018 1.14e-15 ***
## 지하철.1km이내. 1.472e+03  9.064e+01  16.246 < 2e-16 ***
## 승하차평균     5.588e-02  3.853e-03  14.501 < 2e-16 ***
## 지하철거리점수 -2.071e+02  3.541e+01  -5.848 5.06e-09 ***
## 상가수         2.599e+02  2.535e+00  102.534 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14520 on 19145 degrees of freedom
## Multiple R-squared:  0.5733, Adjusted R-squared:  0.5731
## F-statistic: 3674 on 7 and 19145 DF, p-value: < 2.2e-16
```

R-squared는 결정계수를 의미하며, 회귀모형의 설명력을 표현하는 용도로 사용 P-value와 같이 0과 1 사이의 값으로 나타나는데, 0에 가까울수록 설명력이 낮고, 1에 가까울수록 높다고 해석. 학계에서는 0.6이상, 마케팅조사 실무에서는 0.3이상이면 의미가 있다고 해석 이 수치가 의미하는 것은 원인변수가 결과변수를 설명하는 정도로서 R Square가 0.5725라는 것은 원인변수들의 분산이 결과변수의 분산을 57%정도 설명한다는 뜻

회귀분석(Regression Analysis)을 실시했을 때 이 R Squarer값이 0.4이상 이 되지 않으면 나머지 지표들은 볼 필요도 없이 의미가 없는 것이므로 반드시 이 지표를 먼저 봐서 0.4이상인지를 확인하여야 함

(현재 우리가 분석한 결과는 0.3077이니 유의미하다 볼 수 있음) # distance 안하고 지하철거리점수로 한거임 F 검정의 p-value가 0.05 미만일 때, 귀무가설 H0 는 기각되고, 'Y는 X에 대한 회귀식으로 설명 가능하다'는 대립가설 H1이 채택되며, 최종적으로 추정된 회귀식은 적합하다는 결론이 도출

```
##다중공선성 확인
vif(lm_ols1)
```

```
##      버스정류장      시설수      버스승하차  지하철.1km이내.      승하차평균
##      1.352317      1.093086      1.172320      1.570643      1.266167
##  지하철거리점수      상가수
##      1.762353      1.325126
```

vif가 10이 넘어가면, 심각한 다중공선성이 있다고 판단합니다.

가정 진단

ols의 기본 가정 4가지 (선형성, 오차항 독립성, 등분산성, 오차항 정규성)

오차항의 모든 분산은 모든 관찰치에서 일정할 것 -> 이문제를 해결 못하면 이분산성 문제(Heteroscedasticity)가 발생

이를 브루쉬 파간 검정 (Breusch-Pagan test, BP 검정) 을 통해 알 수 있다.

- p-value $\geq 0.05 \Rightarrow$ 등분산성 Homoskedasticity
- p-value $< 0.05 \Rightarrow$ 이분산성 Heteroskedasticity

```
bptest(lm_ols1) ##등분산이라는 귀무가설을 기각하므로, 이분산성이 있다고 판단
```

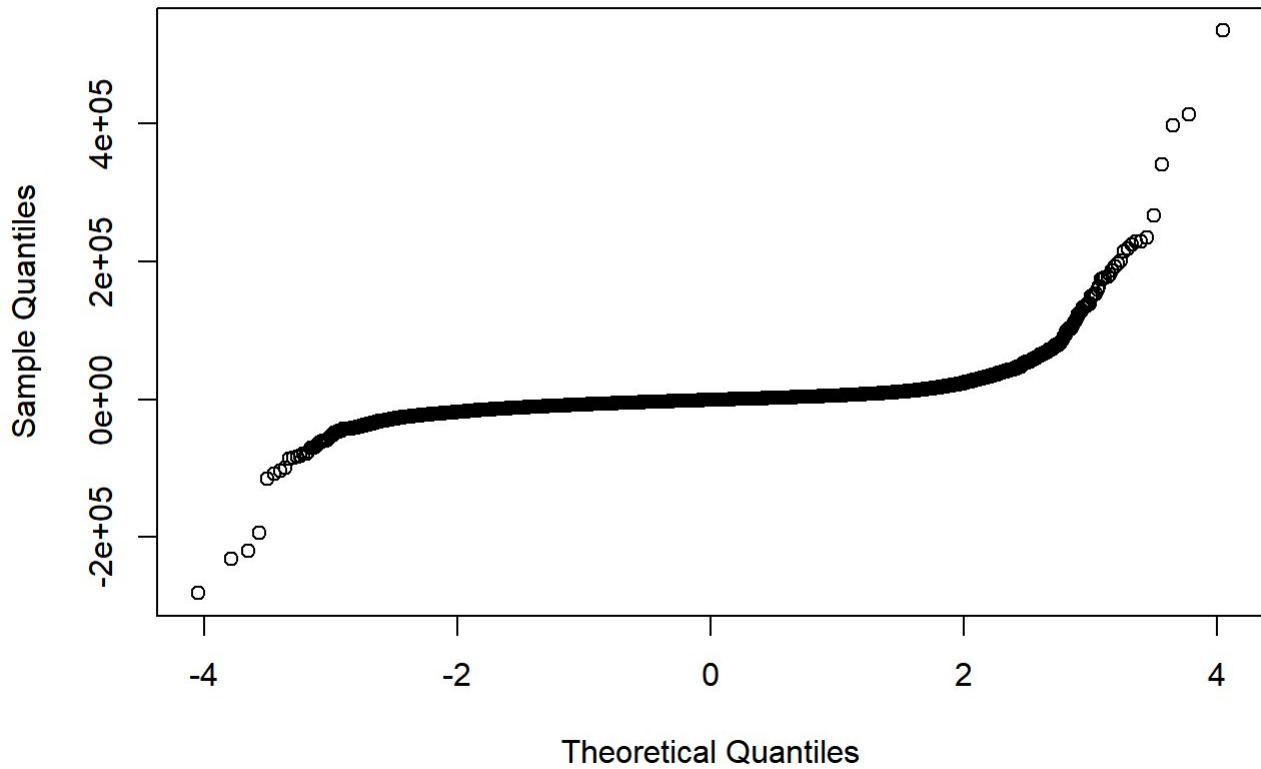
```
##
```



```
## studentized Breusch-Pagan test
##
## data: lm_ols1
## BP = 2822.4, df = 7, p-value < 2.2e-16
```

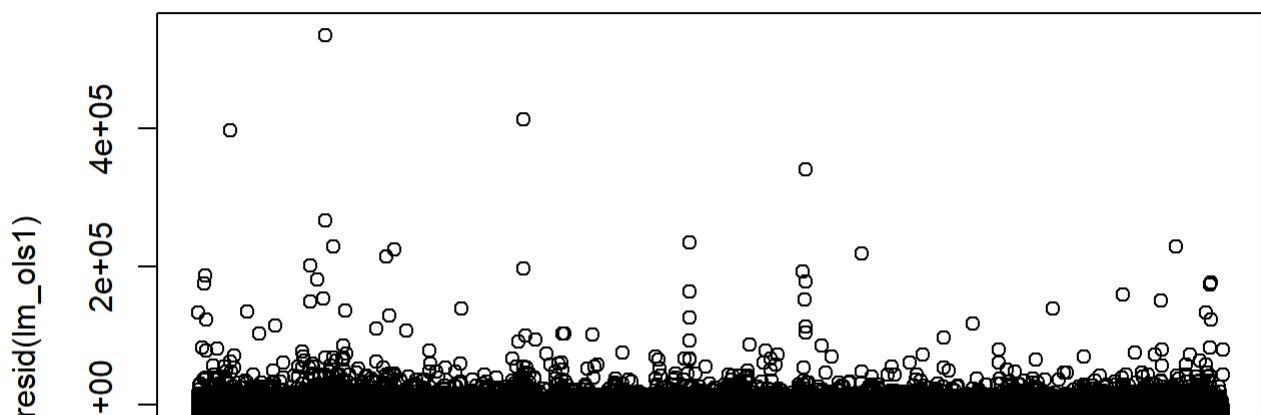
```
##정규성
qqnorm(lm_ols1$residuals)
```

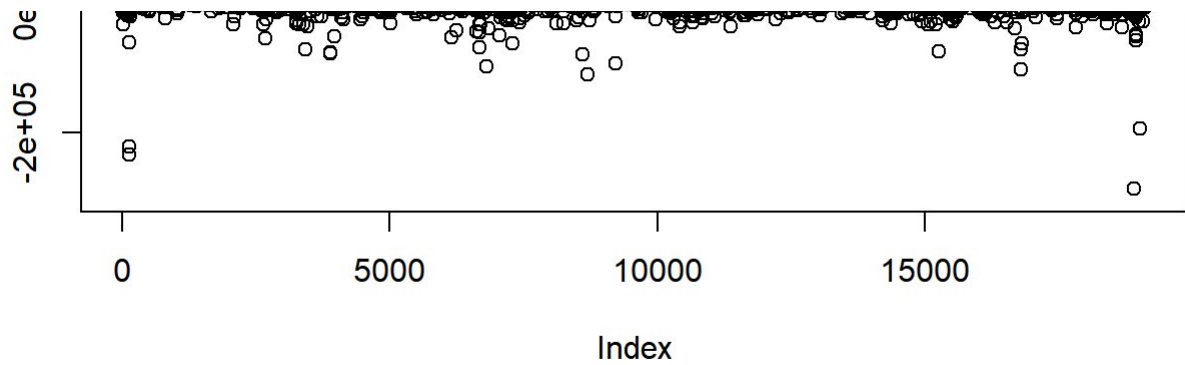
Normal Q-Q Plot



=> 정규성 가정 위배

```
##독립성
plot(resid(lm_ols1))
```





=> 등분산성(오차항의 모든 분산은 모든 관찰치에서 일정할 것) 가정 위배

이분산성 문제를 가지고 있다는 것으로 해석할 수 있으며 즉, 위 모델이 분석에 적합하지 않다

```
dwtest(lm_ols1)
```

```
##
## Durbin-Watson test
##
## data: lm_ols1
## DW = 1.845, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is greater than 0
```

=> 독립성 가정 위배 =====> 오차항 사이 상관관계 있음

```
#공간데이터로 변환
final_sp <- as(final_data, "Spatial")
```

```
##이웃 정의
w <- poly2nb(final_sp)
w
```

```
## Neighbour list object:
## Number of regions: 19153
## Number of nonzero links: 116200
## Percentage nonzero weights: 0.03167616
## Average number of links: 6.066935
```

```
##공간가중치행렬 만들기
wm <- nb2mat(w, style='W')
rwm <- mat2listw(wm, style='W')
```

```
options("scipen" = 100)
lm.morantest(lm_ols1, rwm, alternative="two.sided")
```

```
##
## Global Moran I for regression residuals
##
## data:
## model: lm(formula = 구별총생활인구 ~ ., data = dat)
```

```
## weights: rwm
##
## Moran I statistic standard deviate = 42.886, p-value <
## 0.00000000000000022
## alternative hypothesis: two.sided
## sample estimates:
## Observed Moran I      Expectation      Variance
##      0.18369524681    -0.00026815095    0.00001840059
```

일반회귀모형 잔차의 분포가 공간적 자기상관성을 띄는지 검정 Moran's i 지수가 0.17이며 유의수준0.05수준에서 기각

=> 그러므로 GWR 시작