Implementation of LISA - Variation of Income across different LGAs in Australia

May 9, 2018

Why Spatial Econometrics

Everything is related to everything else, but near things are more related than distant things
Tobler's (1979) first law of geagraphy

- Spatial Dependence : Existemce of functional relationship between what happens at one point in space and what happens elsewhere
- Spatial Heterogeneity: lack of uniformity of effect of space
  Example: Urban places have unequal population and income levels

Why Spatial Econometrics

Neighbors in Space

Expressing a way the structure of spatial dependence is to be incorporated in a model

- Lagged variable in time series analysis is unambiguous
- Spatial Lag?
  - Neighborhood and nearest neighbor $\{$j | P([x_i])$\neq$ P[x_i|x_j] and d_ij $< \epsilon\}$

Why Spatial Econometrics

Spatial Contiguity Matrices

- Binary Contiguity Matrix
- Rook Contiguity
- Bishop Contiguity
- Queen Contiguity
- Distance Based Weights
- K-nearest neighbor weights

From where did we get the Data

Income Data Source

- Australian Bureau Of Statistics (http://stat.data.abs.gov.au)
- Economy -> Finance -> Household Income -> Census 2016, Total Household Income (Weekly) by Household Composition (LGA)
- Customise layout():
  Page - Census Year, Household Composition
  Row - State, Region, Geography Level
  Column - Total Household Income Weekly

From where did we get the Data

Shapefile Data Source

- Australian Bureau Of Statistics (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.003July%202016?OpenDocument)
- Name of the file (Zip) to download: Local Government Areas ASGS Ed 2016 Digital Boundaries in ESRI Shapefile Format

How does the Data look

```
## # A tibble: 121,880 x 6
##    Weekly_Household_Income Annual_Household~ State  LGA_
##    <chr>                   <chr>              <chr>    <
##  1 $1,000-$1,249           $52,000-$64,999    New S~     1
##  2 $1,000-$1,249           $52,000-$64,999    New S~     1
##  3 $1,000-$1,249           $52,000-$64,999    New S~     1
##  4 $1,000-$1,249           $52,000-$64,999    New S~     1
##  5 $1,000-$1,249           $52,000-$64,999    New S~     1
##  6 $1,000-$1,249           $52,000-$64,999    New S~     1
##  7 $1,000-$1,249           $52,000-$64,999    New S~     1
##  8 $1,000-$1,249           $52,000-$64,999    New S~     1
##  9 $1,000-$1,249           $52,000-$64,999    New S~     1
## 10 $1,000-$1,249           $52,000-$64,999    New S~     1
## # ... with 121,870 more rows
```

Analysis

- Number of people for 21 income categories provided for each regions (Histogram for different LGAs with bins as the income categories)
- A density plot created using the historgram
- Compute p= 0.01, 0.02, ..., 0.99 for each LGA

Analysis

Table2

```
## # A tibble: 1,662 x 3
## # Groups:   LGA_2016 [?]
##    LGA_2016 Weekly_Household_Income Number_People
##       <int> <chr>                           <int>
##  1    10050 High                            13715
##  2    10050 Low                              6075
##  3    10050 Medium                          15489
##  4    10130 High                             7178
##  5    10130 Low                              3138
##  6    10130 Medium                           8104
##  7    10250 High                            11057
##  8    10250 Low                              4869
##  9    10250 Medium                          13679
## 10    10300 High                              518
## # ... with 1,652 more rows
```

Analysis

Obtain one representative income for each LGA

Weekly household median incomes assumed for each class:

- Low: 250A$
- Medium: 1000A$
- High: 5000A$

Reading LGA Data

- Reading OGR vector maps into Spatial objects
- The Geospatial Data Abstraction Library (GDAL) is a computer software library for reading and writing raster and vector geospatial data formats
- The rgdal package has been around for more than a decade and provides bindings to the incredible Geospatial Data Abstraction Library (GDAL) for reading, writing and converting between spatial formats

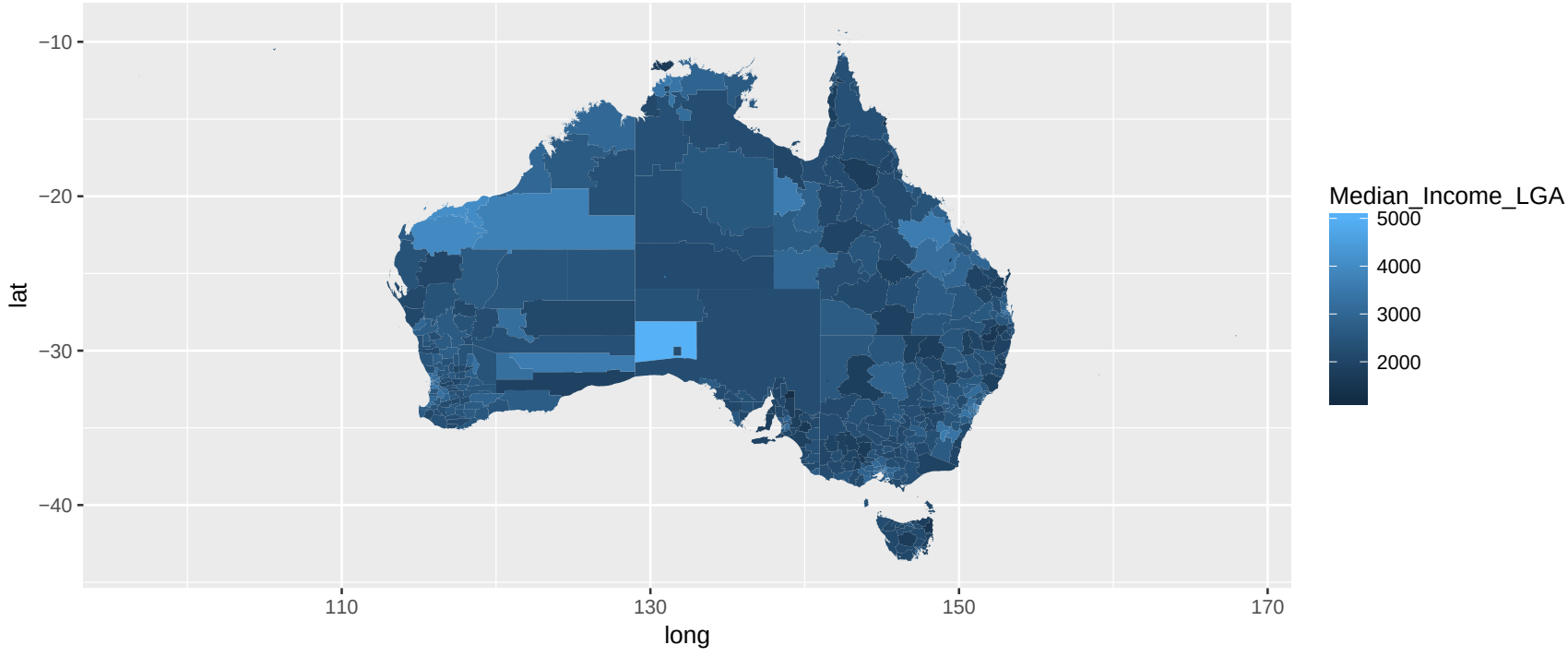
```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/sgup0008/Documents/ABS_Income/ABS_Income
## with 563 features
## It has 5 fields

## Warning in rgdal::readOGR("../LGA_shapefiles", layer = "
## Dropping null geometries: 131, 132, 213, 214, 293, 294,
## 536, 537, 556, 557, 559, 560, 562, 563

## Regions defined for each Polygons
```

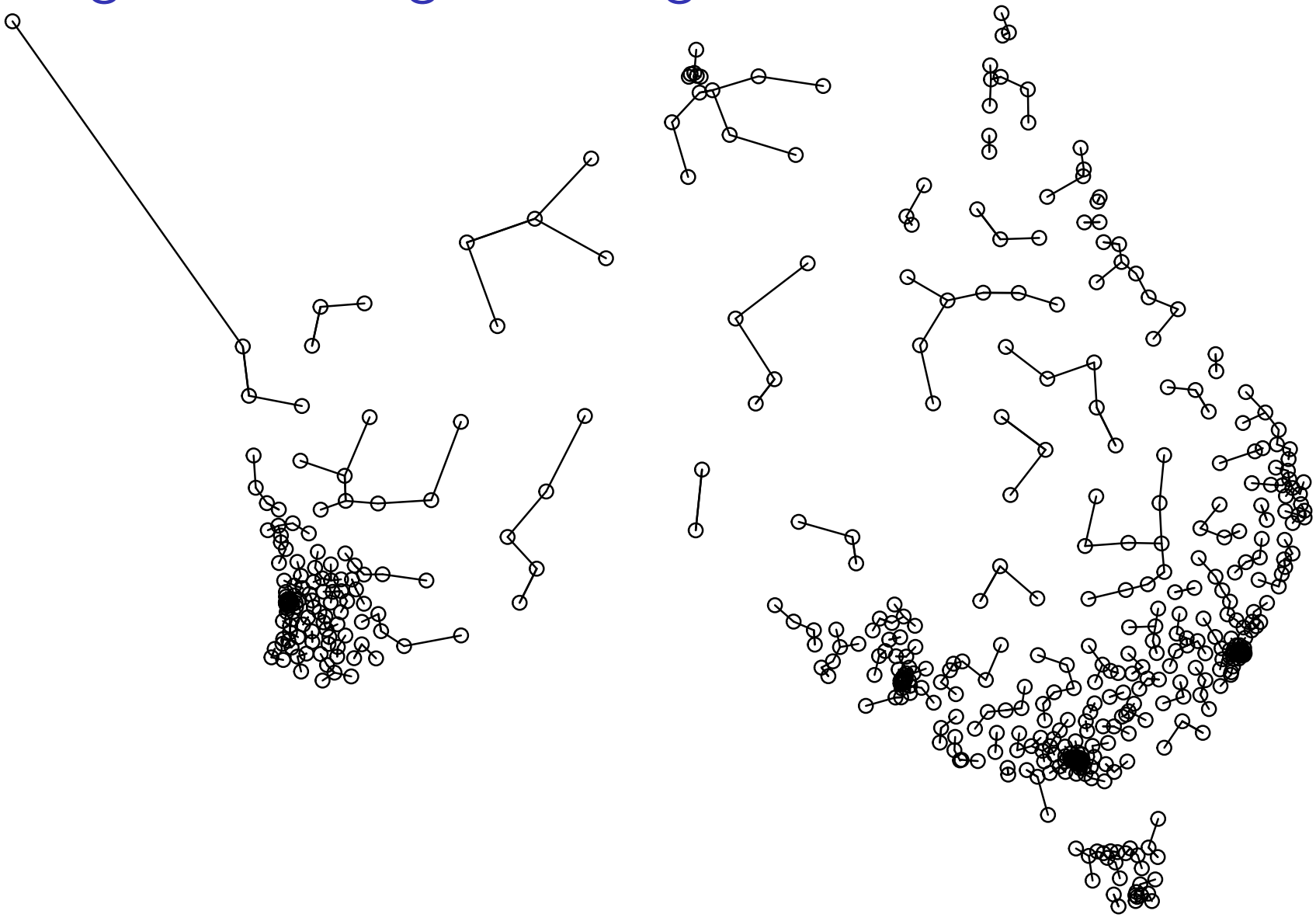
HeatMap - Maps with Structural Breaks

```
## Joining, by = "id"
```



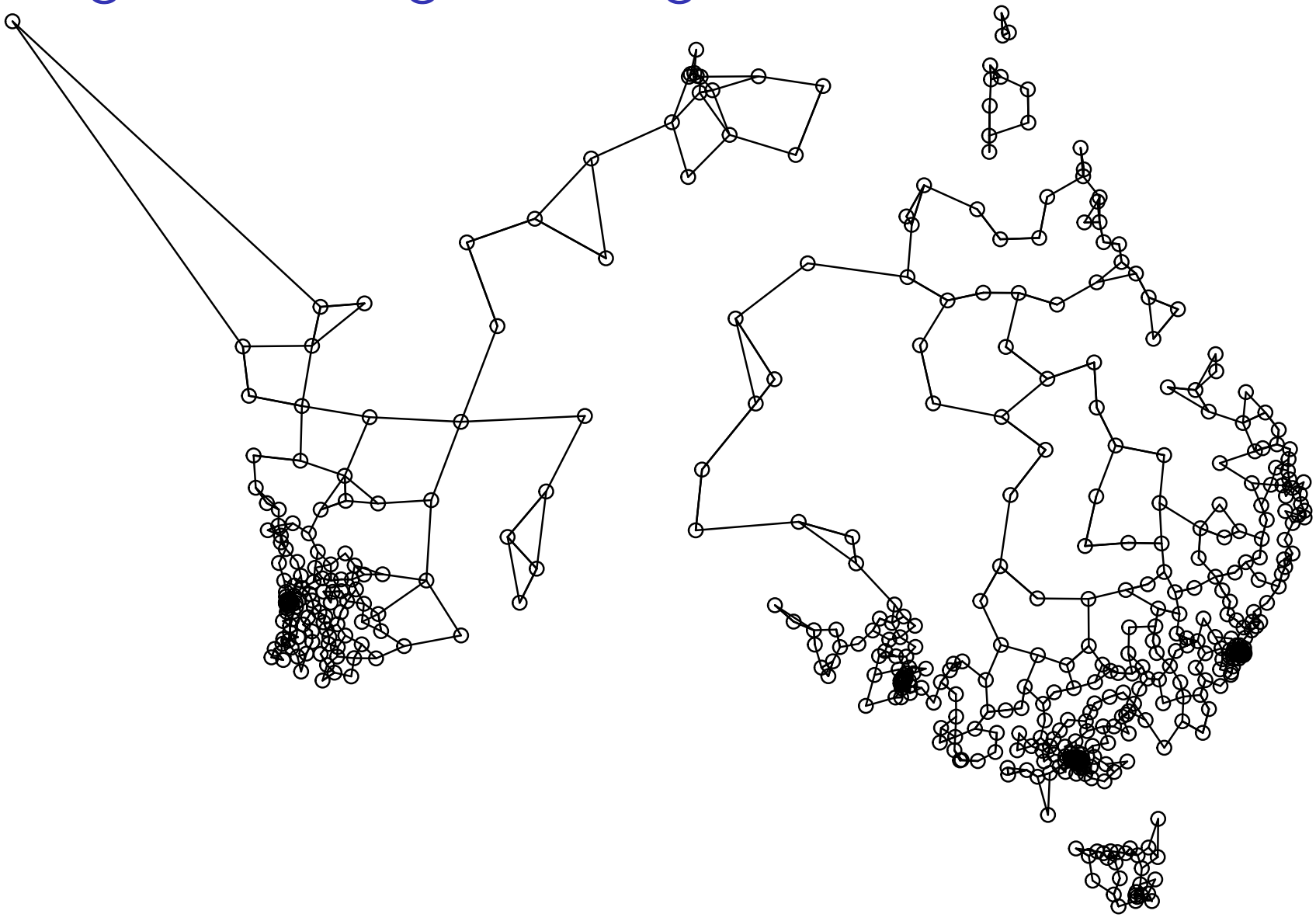

```
## [1] 1090
```

Plotting nearest neighbors neighbours K=1



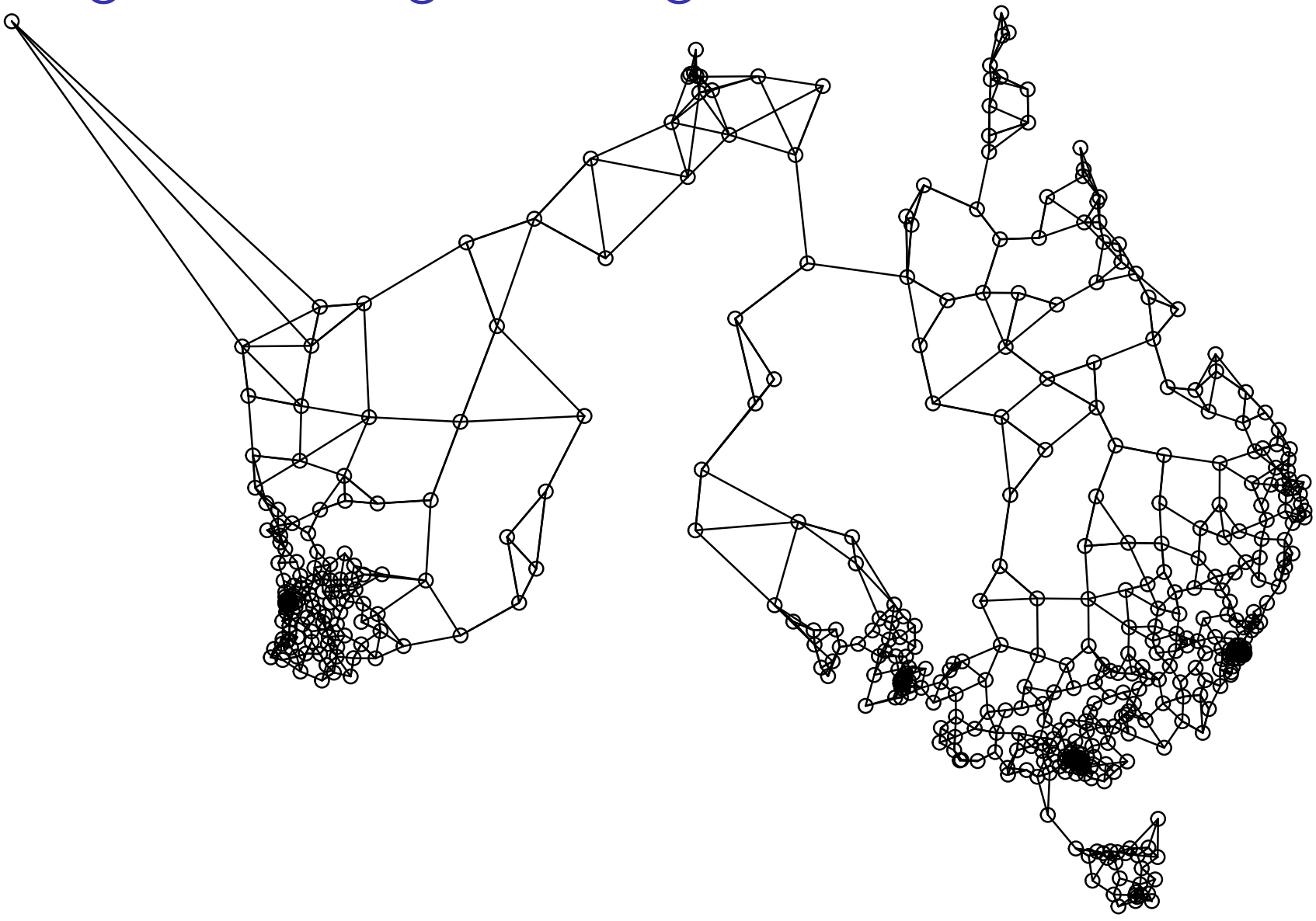
```
## [1] 1612.478
```

Plotting nearest neighbors neighbours K=2



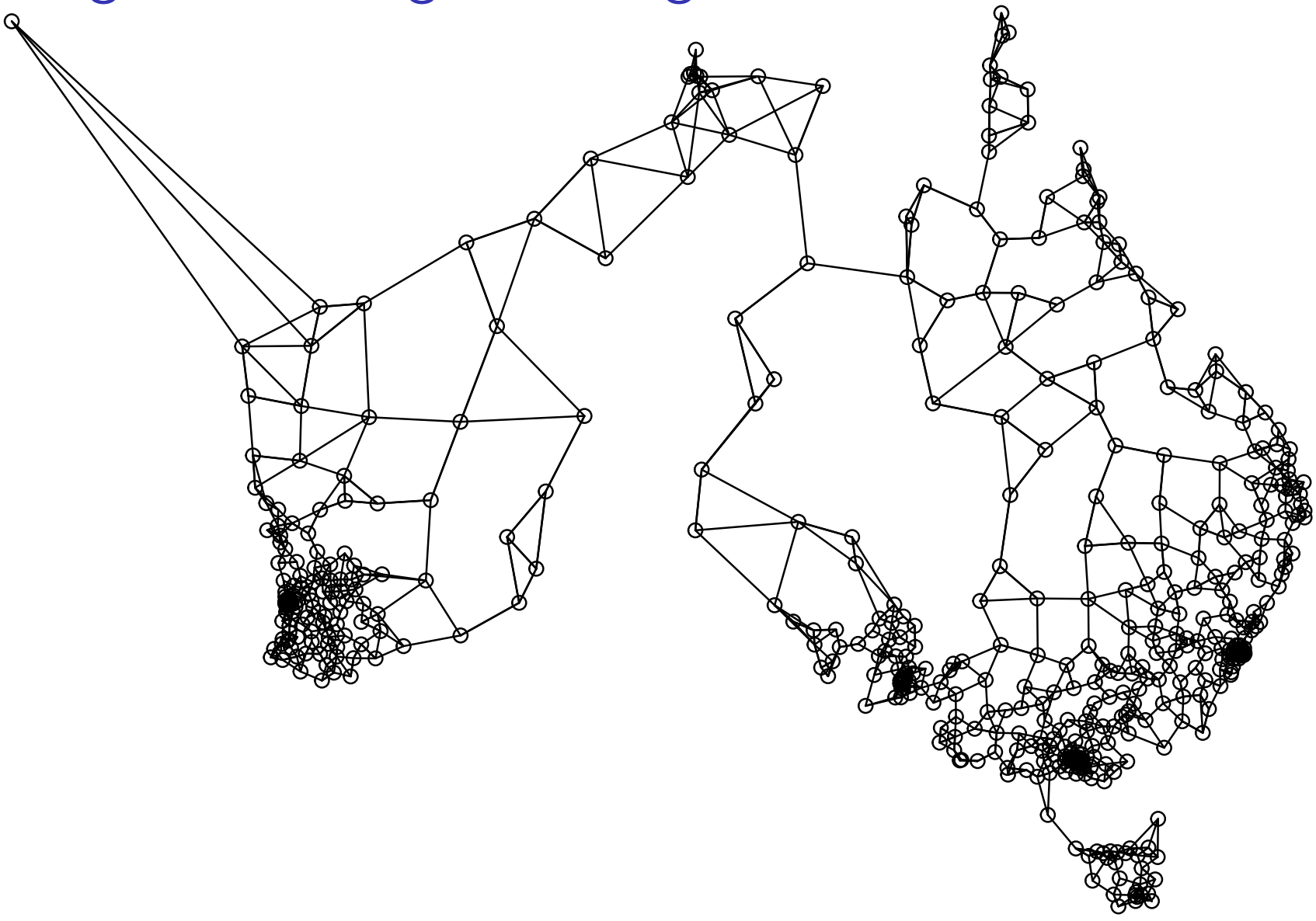
```
## [1] 1690.147
```

Plotting nearest neighbors neighbours K=3



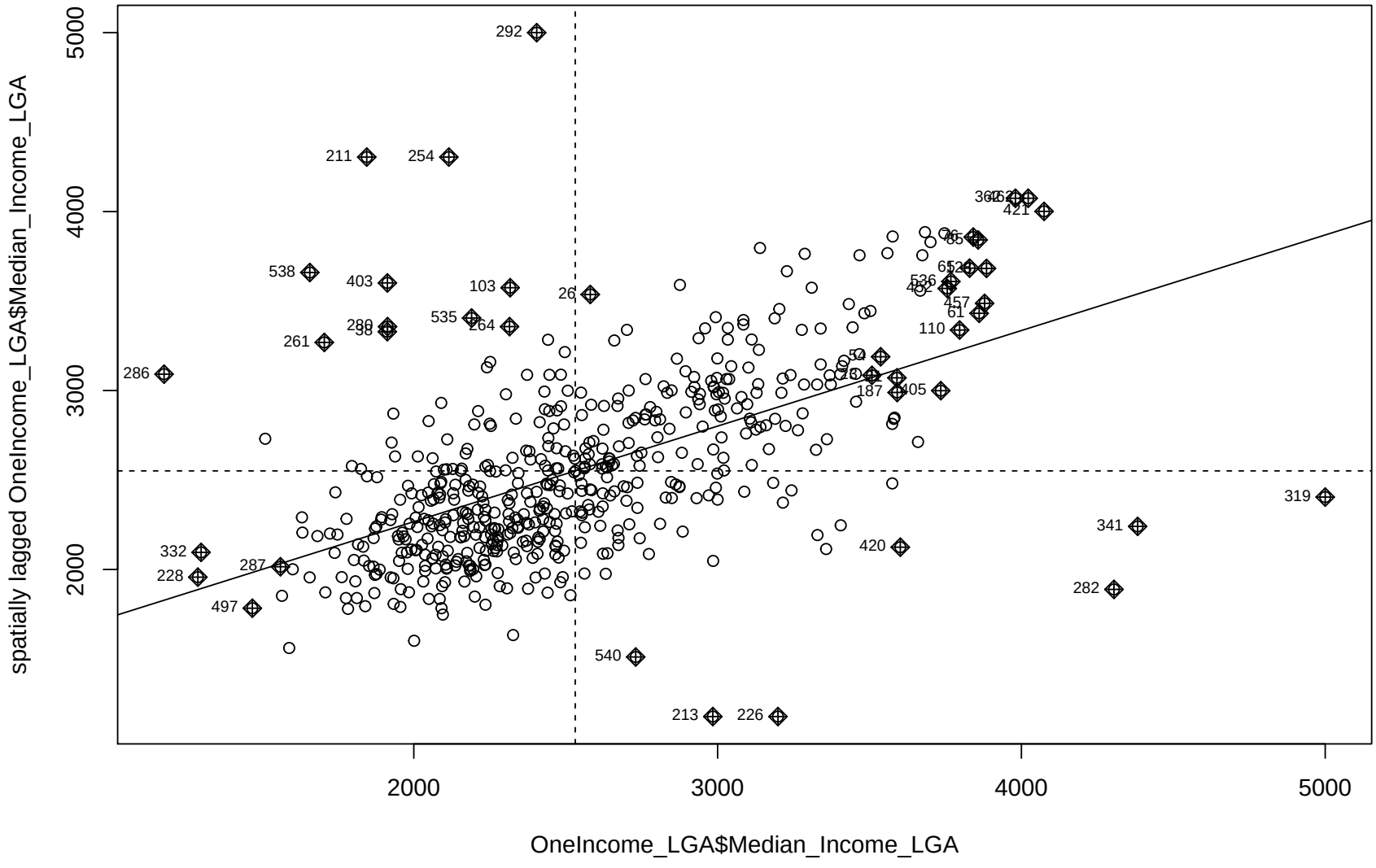
```
## [1] 1778.608
```

Plotting nearest neighbors neighbours K=4



```
## [1] 1792.845
```

Moran Scatterplot



Local Moran

```
##        Ii                E.Ii               Var.Ii       
##  Min.   :-4.079221   Min.   :-0.001838   Min.   :0.3305  
##  1st Qu.:-0.000144   1st Qu.:-0.001838   1st Qu.:0.4966  
##  Median : 0.258115   Median :-0.001838   Median :0.9951  
##  Mean   : 0.534091   Mean   :-0.001838   Mean   :0.7399  
##  3rd Qu.: 0.739304   3rd Qu.:-0.001838   3rd Qu.:0.9951  
##  Max.   : 7.714157   Max.   :-0.001838   Max.   :0.9951  
##       Z.Ii             Pr(z > 0)     
##  Min.   :-6.642709   Min.   :0.0000  
##  1st Qu.: 0.001698   1st Qu.:0.1726  
##  Median : 0.299093   Median :0.3824  
##  Mean   : 0.646667   Mean   :0.3524  
##  3rd Qu.: 0.944061   3rd Qu.:0.4993  
##  Max.   :10.794443   Max.   :1.0000
```

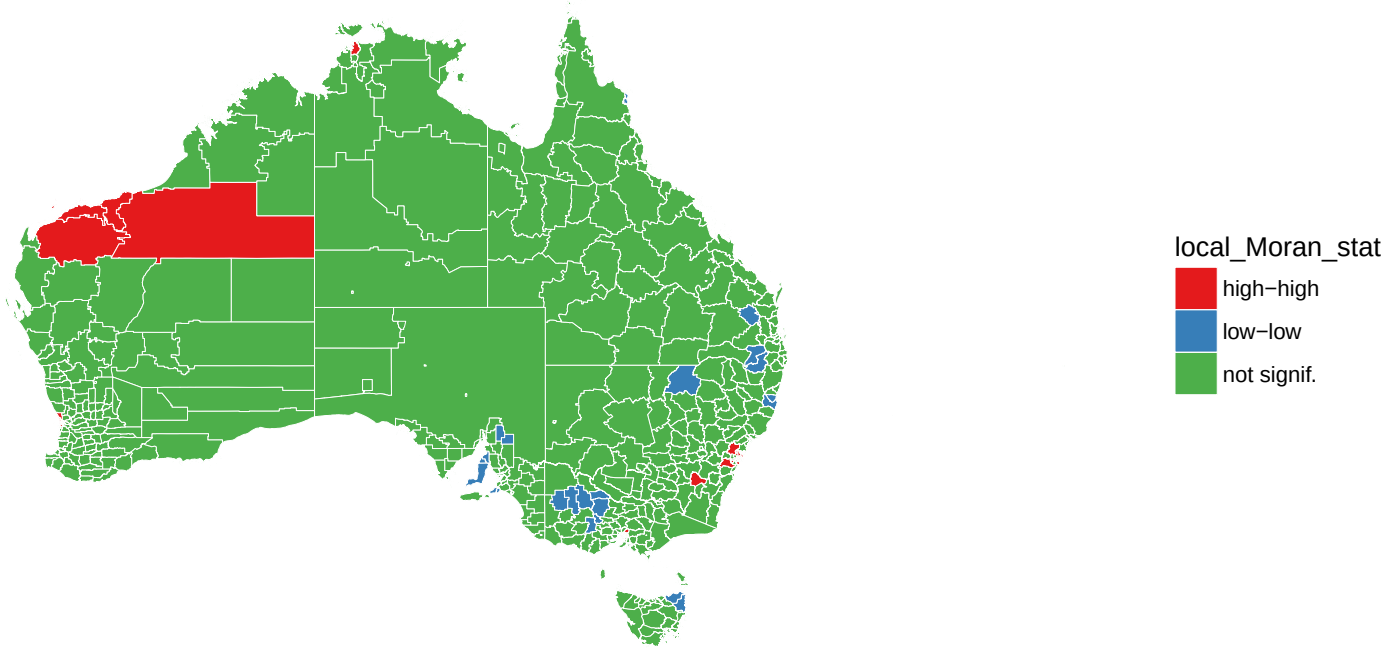
Local Significance Plot using Moran's Statistic

```
## # A tibble: 3 x 2
##   local_Moran_stat     n
##   <chr>            <int>
## 1 high-high           51
## 2 low-low             25
## 3 not signif.        469
```

Local Significance Plot using Moran's Statistic

```
## Joining, by = c("LGA_2016", "Median_Income_LGA", "Median
```

```
## Joining, by = "id"
```



Local Significance Plot using Gi star Statistic

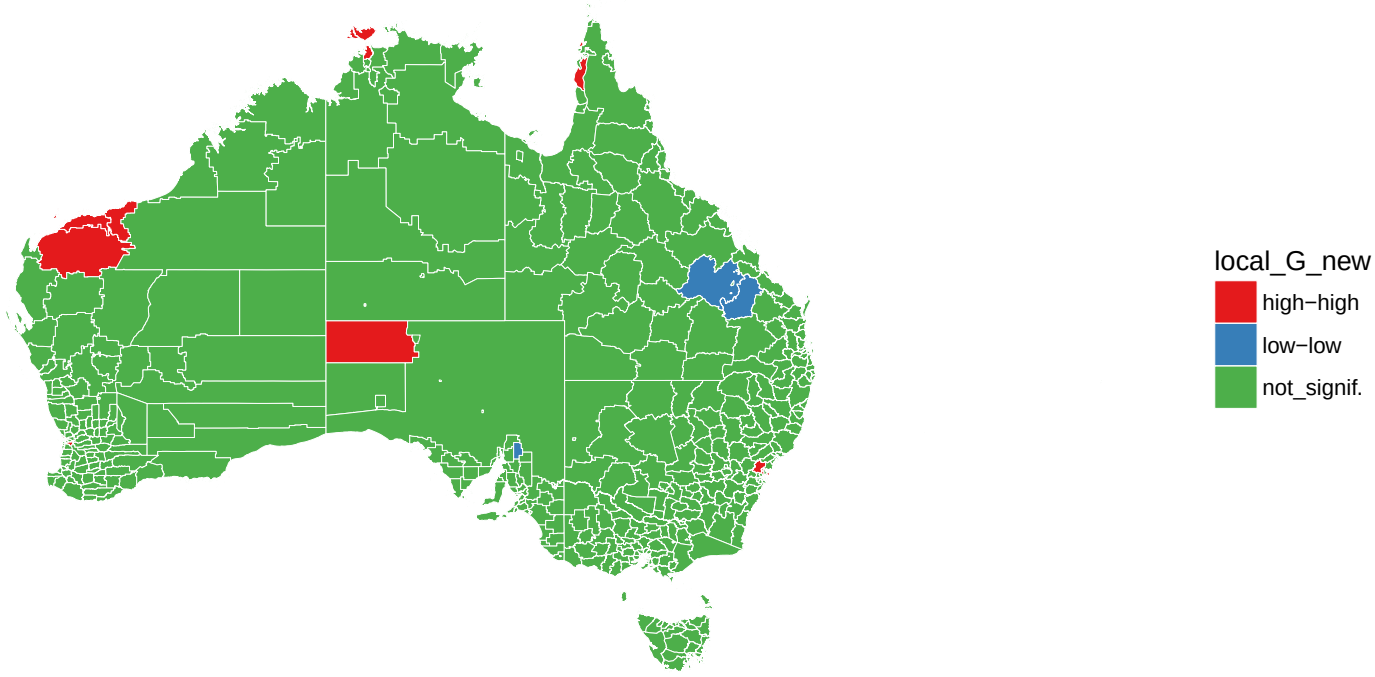
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -2.49074 -0.78212 -0.16746  0.04306  0.79739  4.51540
```

```
## # A tibble: 3 x 2
##   local_G_new     n
##   <fct>       <int>
## 1 high-high      35
## 2 low-low         4
## 3 not_signif.   506
```

Local Significance Plot using Gi star Statistic

```
## Joining, by = c("LGA_2016", "Median_Income_LGA", "Median
```

```
## Joining, by = "id"
```



Regions spatially significant in local Moran and not in local Gi star

x
Bayside (C)
Boroondara (C)
Camden (A)
Canada Bay (A)
Cockburn (C)
East Fremantle (T)
East Pilbara (S)
Georges River (A)
Hornsby (A)
Inner West (A)
Joondalup (C)
Manningham (C)
Melville (C)
Nillumbik (S)
Subiaco (C)

Regions spatially significant in local Moran and not in local Gi star

x
Anangu Pitjantjatjara (AC)
Armadale (C)
Aurukun (S)
Mapoon (S)
Tiwi Islands (R)