

空間情報を用いた社会・経済分析(第3回)

統計数理研究所 村上大輔

dmuraka@ism.ac.jp

担当回（前半）

内容

- 第2回(4/21 月)：空間データの処理・地図化
- 第3回(4/28 月)：探索的空間データ解析
- 第4回(5/8, 木)：空間計量経済モデルと応用



各回で統計ソフトウェアRを用いた実

スライド・Rコード(slide_lec3.ppt, code_lec3.R)

https://github.com/dmuraka/HIAS_class

- 質問等は村上(dmuraka@ism.ac.jp)までご連絡ください

先週に作成した 住宅地公示地価データ

列名	説明
price	住宅地公示地価(円/m ²)
rice, agri, ..., out	1km ² メッシュ内の土地利用面積(m ²)
st_dist	最寄駅距離(m)
tk_dist	東京駅までの距離(m)

```
> price[1:10,]  
Simple feature collection with 2129 features and 18 fields  
Geometry type: POINT  
Dimension: XY  
Bounding box: xmin: 139.1364 ymin: 27.0949 xmax: 142.2034 ymax: 35.83208  
Geodetic CRS: JGD2011  
First 10 features:  
      price      status      mesh rice agri forest wild building  road  rail  
1 3960000      住宅 53394529    0    0 209134    0  575119 62740    0  
2 2530000      住宅 53394519    0    0  94120    0  564720 188240    0  
3 4830000      住宅 53394528    0    0  31370    0  784257 125481 20914  
4 1970000      住宅 53394539    0    0  62734    0  690072  62734 52278  
5 3680000      住宅 53394539    0    0  62734    0  690072  62734 52278  
6 2270000      住宅 53394519    0    0  94120    0  564720 188240    0  
7 3740000      住宅 53394529    0    0 209134    0  575119  62740    0  
11 3790000 住宅,店舗,事務所 53394631    0    0    0    0  763251 135922 125466  
12 3200000 住宅,店舗,事務所 53394641    0    0    0    0  773627 188180  62727  
13 1580000      住宅,事務所 53394641    0    0    0    0  773627 188180  62727  
      other  river beach ocean golf out      geometry      st_dist      tk_dist  
1 62740 135937    0    0    0    0 POINT (139.7448 35.69014) 599.62116 2331.644 [m]  
2 83662 115036    0    0    0    0 POINT (139.7375 35.6812) 333.27568 2762.764 [m]  
3 62741  20914    0    0    0    0 POINT (139.7329 35.68814) 348.31769 3277.265 [m]  
4 73189 104556    0    0    0    0 POINT (139.7465 35.69863) 303.54010 2761.092 [m]  
5 73189 104556    0    0    0    0 POINT (139.7462 35.69608) 453.15420 2589.836 [m]  
6 83662 115036    0    0    0    0 POINT (139.7408 35.68065) 285.30953 2468.656 [m]  
7 62740 135937    0    0    0    0 POINT (139.7408 35.68827) 332.41378 2594.560 [m]  
11    0  20911    0    0    0    0 POINT (139.7749 35.69587)  42.66372 1752.159 [m]  
12 10454  10454    0    0    0    0 POINT (139.771 35.70278)  72.51797 2424.979 [m]  
13 10454  10454    0    0    0    0 POINT (139.769 35.70212) 260.74456 2339.240 [m]
```

線形回帰モデル(さわりだけ)

説明変数(x_1, \dots, x_K)が被説明変数(y ; 住宅地価)に及ぼす影響の強さを推定

$$y = \beta_0 + \sum_{k=1}^K x_k \beta_k + \varepsilon$$

説明変数: x_k (例: 最寄駅距離、東京駅までの距離)

回帰係数: β_k (k番目の説明変数からの影響の強さを表す)

住宅地価の場合

最寄駅までの距離 (st_dist): x_1

β_1

東京駅までの距離 (tk_dist): x_2

β_2

残差(説明されない要因): ε

駅距離が1単位増えたら
地価はいくつ増えるか

- $\beta_1 > 0$: 駅距離が増えるほど高騰
- $\beta_1 < 0$: 駅距離が増えるほど低下

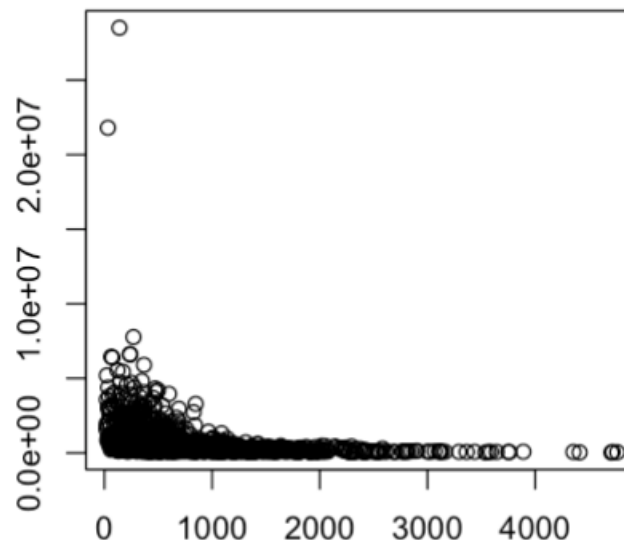
住宅地価 (price): y

被説明変数と説明変数の関係の確認

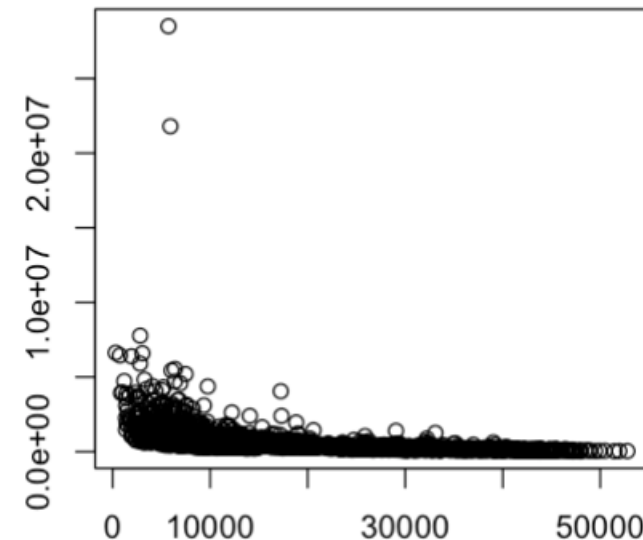
- 片対数: y だけ対数
- 両対数: y も x も対数

→ `st_dist` と `tk_dist` に関しては、
今回は両対数が良さそう

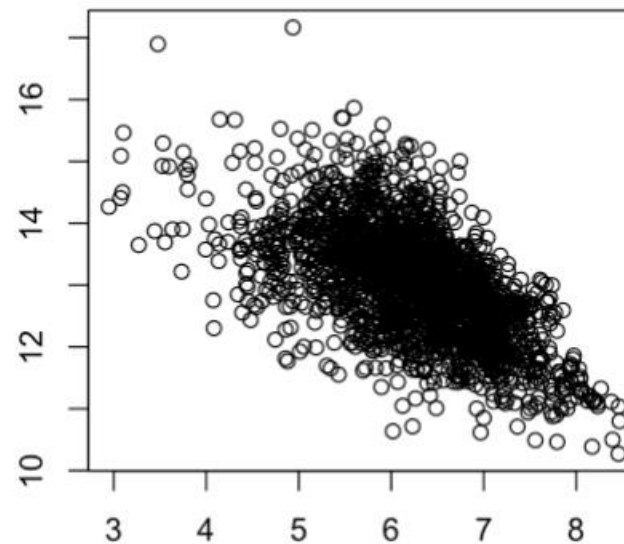
> `plot(st_dist, price)`



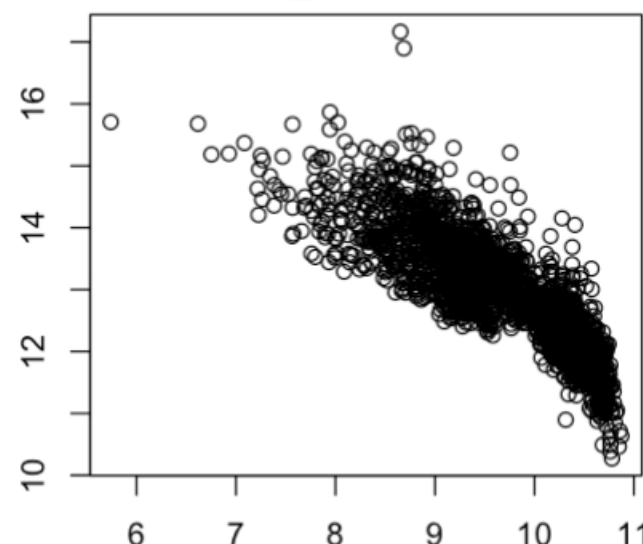
> `plot(tk_dist, price)`



> `plot(log(st_dist), log(price))`



> `plot(log(tk_dist), log(price))`



線形回帰モデルの推定結果

住宅地価
(対数)

最寄駅距離
(対数)

東京駅距離
(対数)

1km²メッシュ内の各土地利用の面積(m²)

データ名

```
> mod <- lm(log(price) ~ log(st_dist) + log(tk_dist) + rice + agri + forest + wild + building + road + rail + river, data = dprice3)
> summary(mod)
```

回帰係数

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.282e+01	2.131e-01	107.100	< 2e-16	***
log(st_dist)	-2.863e-01	1.282e-02	-22.332	< 2e-16	***
log(tk_dist)	-8.208e-01	1.601e-02	-51.258	< 2e-16	***
rice	-1.542e-06	1.173e-06	-1.315	0.18874	
agri	-1.303e-06	2.328e-07	-5.600	2.43e-08	***
forest	-1.046e-06	1.771e-07	-5.907	4.06e-09	***
wild	-3.198e-06	2.045e-06	-1.564	0.11803	
building	-1.903e-09	1.355e-07	-0.014	0.98879	
road	-7.711e-07	2.856e-07	-2.700	0.00699	**
rail	2.186e-07	3.145e-07	0.695	0.48716	
river	-1.113e-06	1.696e-07	-6.562	6.65e-11	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4045 on 2100 degrees of freedom
Multiple R-squared: 0.7956, Adjusted R-squared: 0.7946
F-statistic: 817.2 on 10 and 2100 DF, p-value: < 2.2e-16

有意性(効果の有無)

*** : 0.1%水準で有意
** : 1%水準で有意
* : 5%水準で有意
 : 0.1%水準で有意

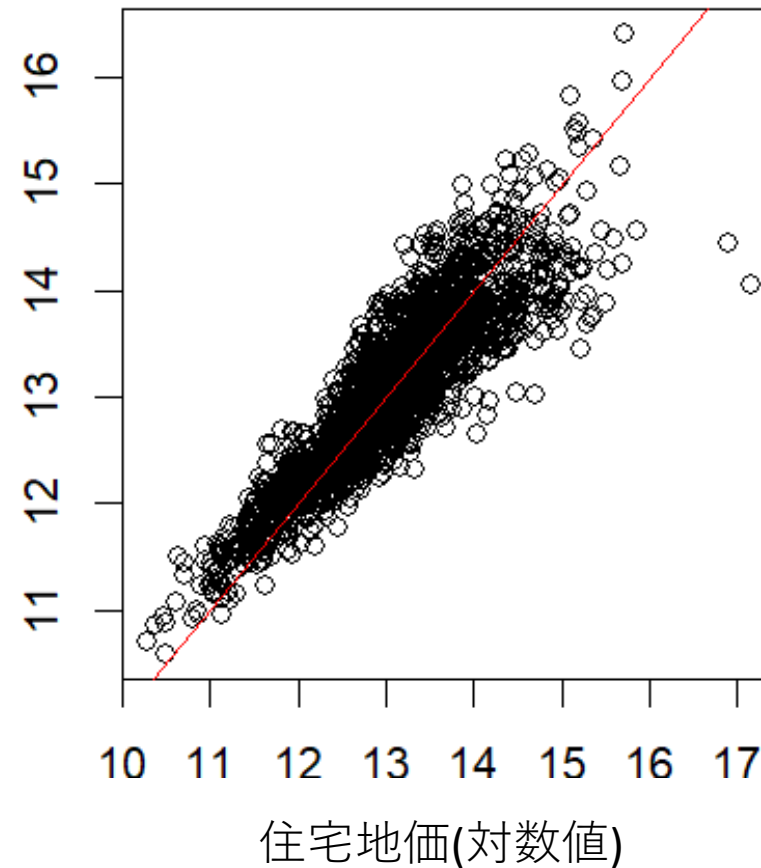
自由度調整済み 決定係数

地価の変動の
79.46%が説明された

実際の住宅地価と予測値の比較

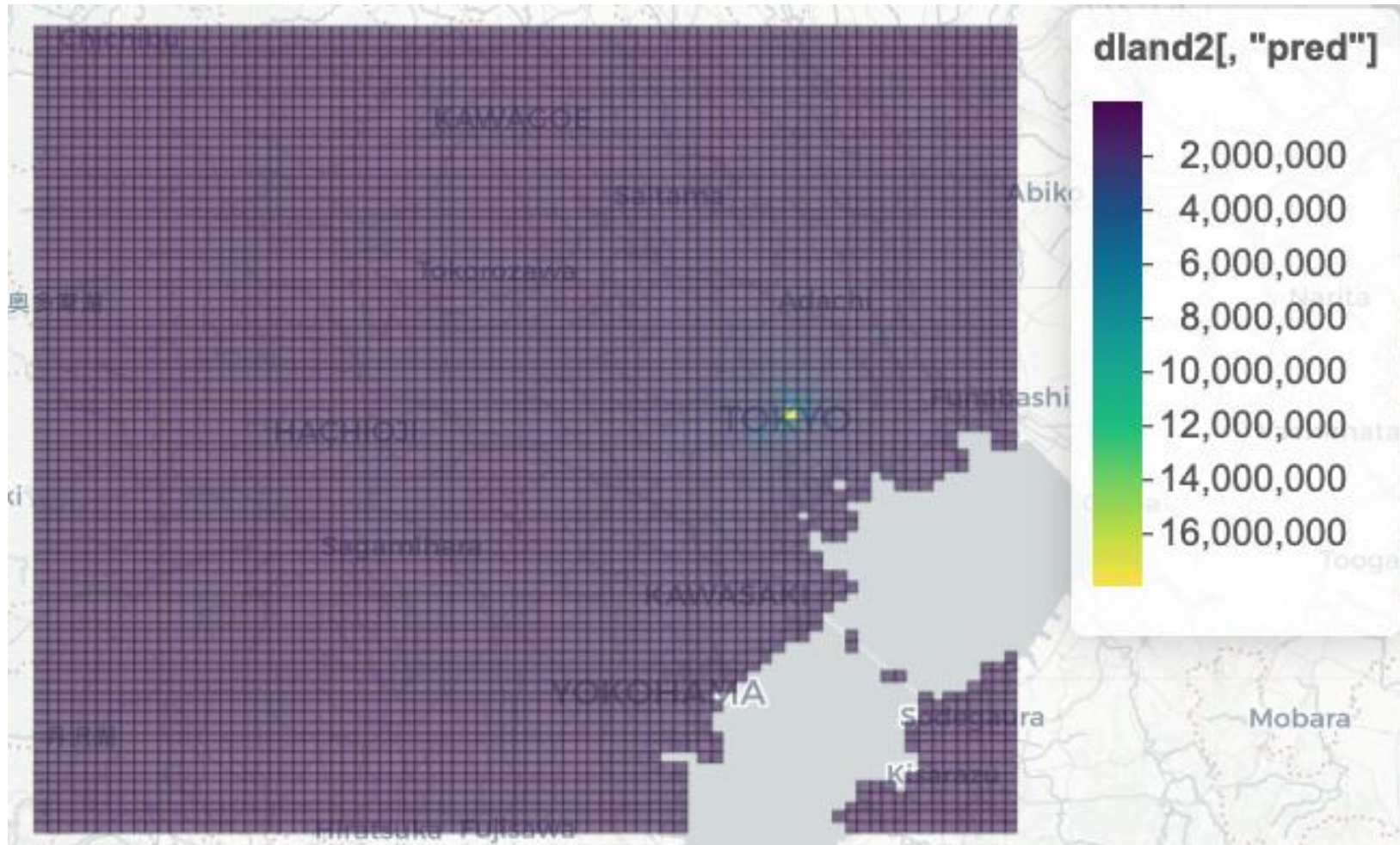
```
> log_pred <- predict(mod)
> plot( log(dprice3$price), log_pred )
> abline(0,1,col="red")
```

予測値(対数値)



地価の予測

```
> log_pred0 <- predict(mod,newdata=dland2) # 土地利用3次メッシュ毎の住宅地価(対数値)の予測値  
> dland2$pred <- exp(log_pred0)           # 住宅地価(実数値)の予測値  
> mapview(dland2[,“pred”])                # プロット
```



カラーパレットの作成 (RColorBrewerパッケージ)

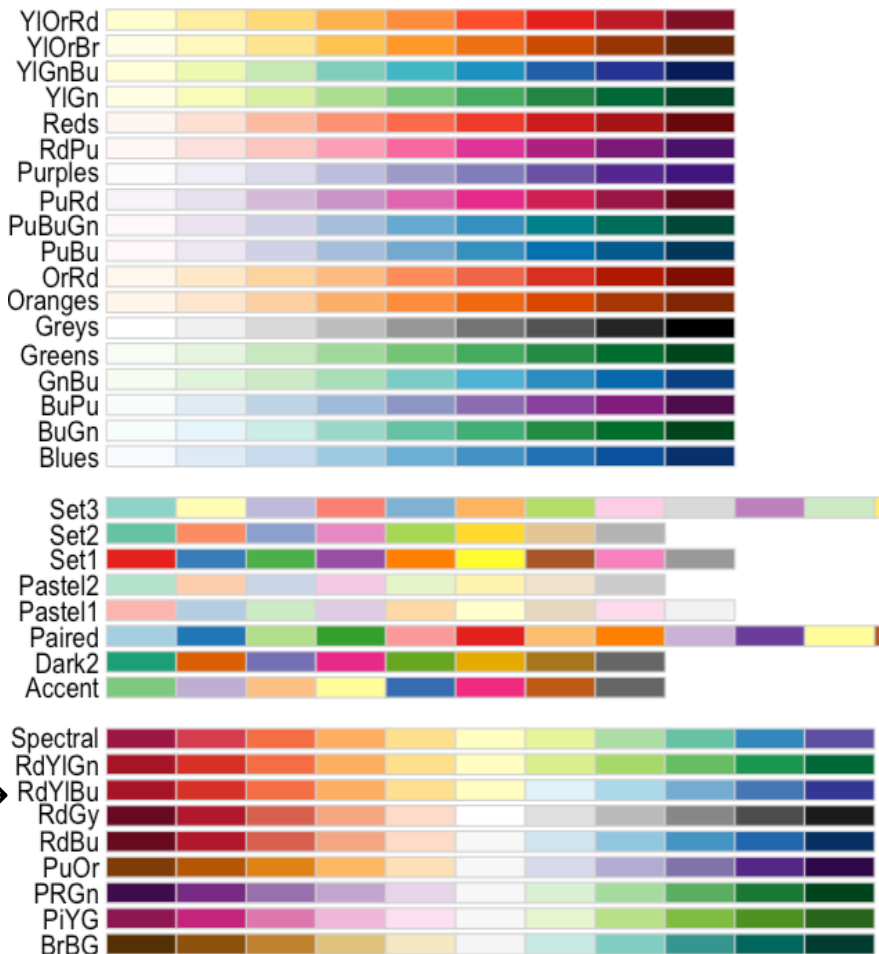
カラーパレットの生成

```
> breaks <-c(0,100000,200000,300000,400000,  
             500000,600000,800000,1200000,  
             2000000, max(dland2$pred)) # 色の区切り位置  
> pal0 <- brewer.pal(length(breaks)-1, "RdYlGn") # パレット生成  
> pal <- rev( pal0 ) # 色を逆順に(赤=大きい)
```

RdYlGn(Red-Yellow-Greenの略)→

カラーパレットの確認

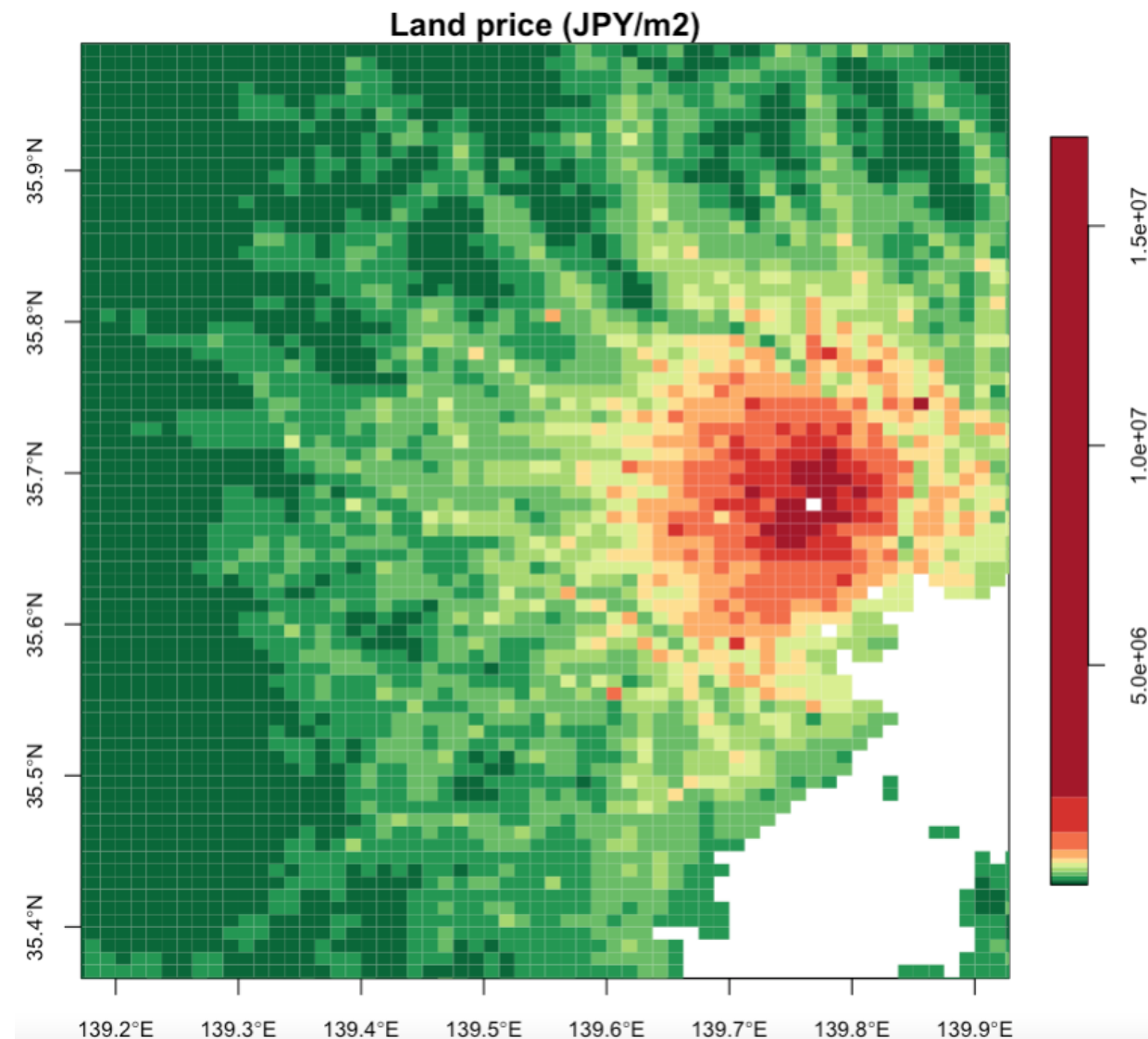
> display.brewer.all()



plot関数を用いた地図化

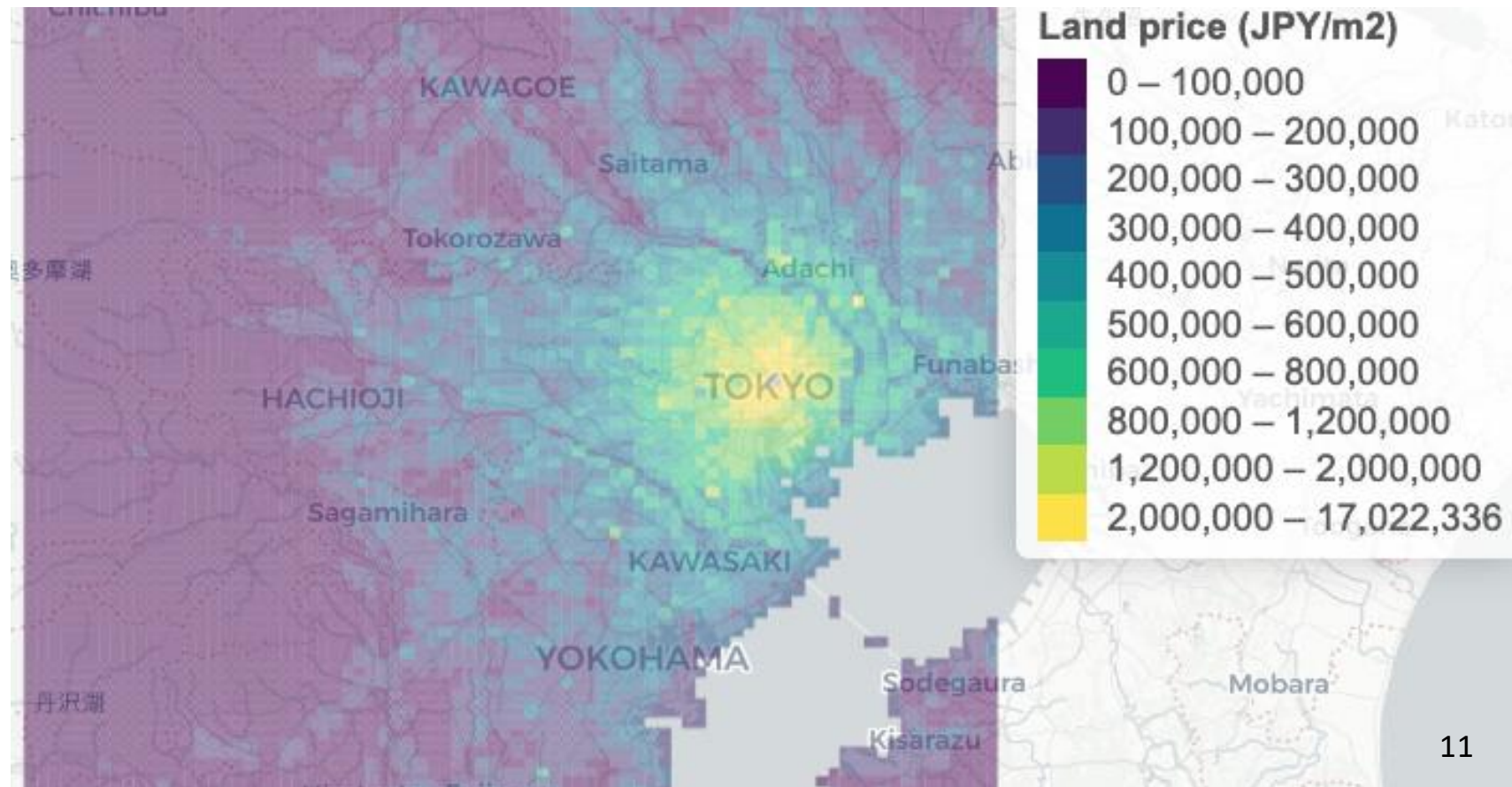
- レポートなどの地図作成に便利

```
plot(dland2["pred"], # 住宅地価をプロット
     axes = TRUE,    # XY座標値を表示
     cex.axis = 0.8, # XY座標値のフォントサイズ
     pal = pal,       # カラーパレット
     breaks = breaks, # 色の区切り位置
     pch = 20,        # シンボル(例：20は円、15は四角)
     cex = 0.8,       # シンボルのサイズ(デフォルトは1)
     key.pos = 4,      # 凡例の位置 (4は図の右)
     key.length = 0.8, # 判例の長さ (1=図の幅)
     xlim = c(139.2, 139.9), ylim = c(35.5, 35.85), # 表示範囲
     border=NA,        # メッシュの枠線を消す
     main = "Land price (JPY/m2)") # タイトル
```



```
> mapview(dland2[, "pred"], # 住宅地価をプロット
  cex=4, # プロットするシンボルのサイズ
  lwd=0, # シンボルの枠線の太さ (0は枠線なし)
  at = breaks, # 色の区切り位置
  alpha.regions=0.8, # 透過度 (1は透過なし)
  legend=TRUE, # TRUEの場合、凡例を表示
  layer.name="Land price (JPY/m2)" # タイトル
```

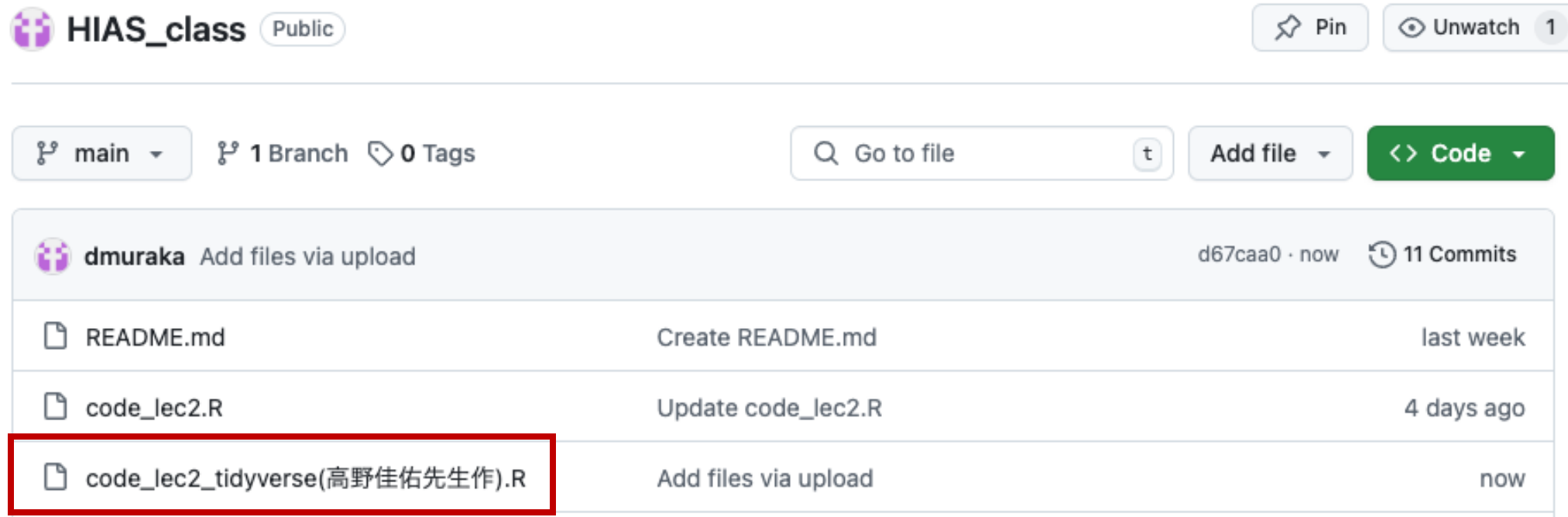
mapviewを用いた地図化



余談：tidyverseを用いた実装

- 高野啓介先生がtidyverseパッケージを用いてナウい感じに再実装してくれました

https://github.com/dmuraka/HIAS_class



HIAS_class Public

Pin Unwatch 1

main 1 Branch 0 Tags

Go to file t Add file <> Code

dmuraka Add files via upload d67caa0 · now 11 Commits

README.md	Create README.md	last week
code_lec2.R	Update code_lec2.R	4 days ago
code_lec2_tidyverse(高野佳佑先生作).R	Add files via upload	now

今後の予定等

内容

- 第2回(4/21 月) : 空間データの処理・地図化
- **第3回(4/28 月) : 探索的空間データ解析**
- 第4回(5/8, 木) : 空間計量経済モデルと応用

↑

各回で統計ソフトウェアRを用いた実例を紹介

スライド・Rコード置き場(slide_lec3.ppt, code_lec3.R)

https://github.com/dmuraka/HIAS_class

- 質問等は村上(dmuraka@ism.ac.jp)までご連絡ください

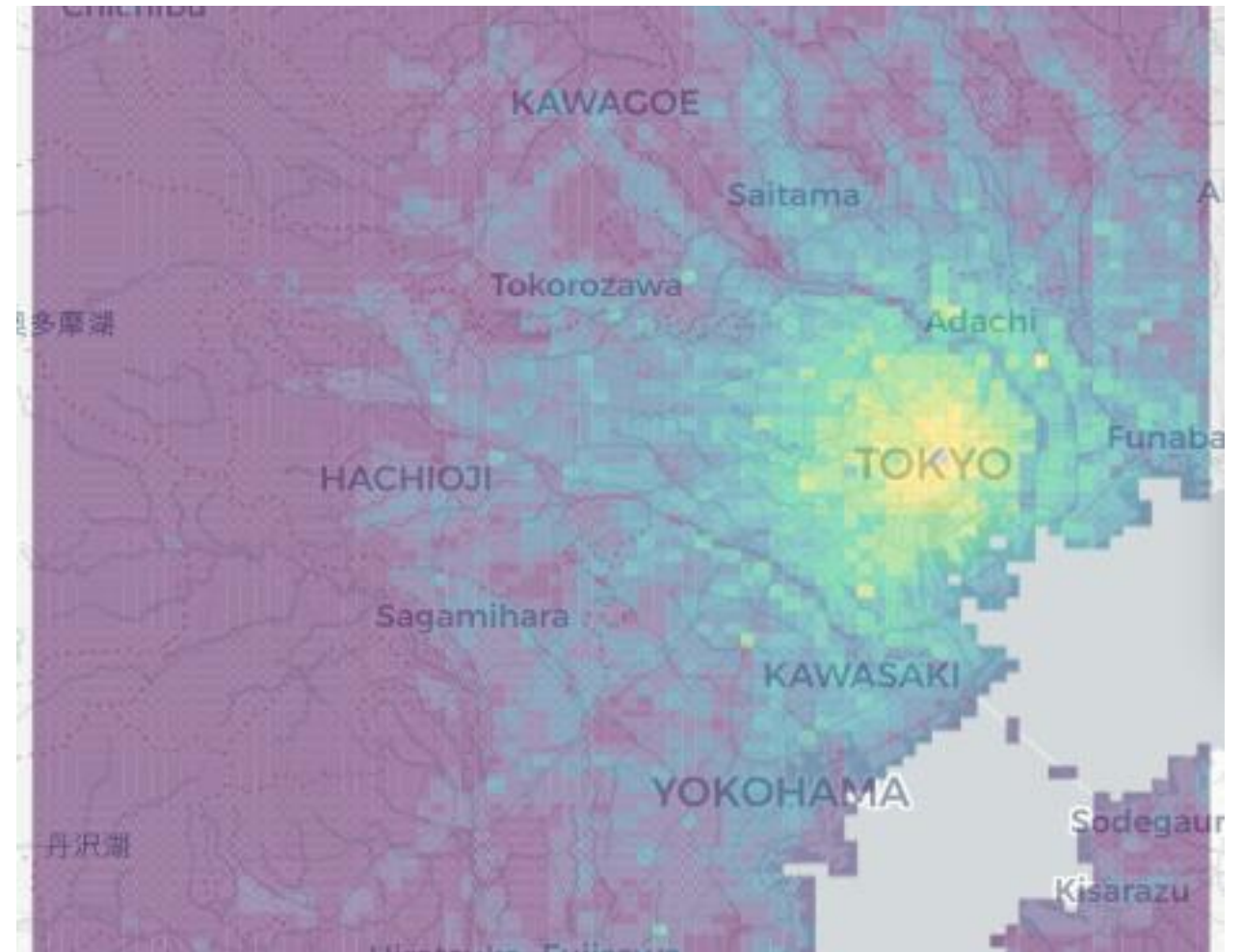
空間相関(spatial autocorrelation/spatial dependence)

近所と強く相関するという

空間データの一般的性質

≡ クラスター(or チェッカーボード)

パターンを持つ



R. A. Fisher (1935)

“After choosing the *area* we usually have no guidance beyond the widely verifiable fact that patches in *close proximity* are commonly *more alike*, as judged by the yield of crops, than those which are *far apart*.”

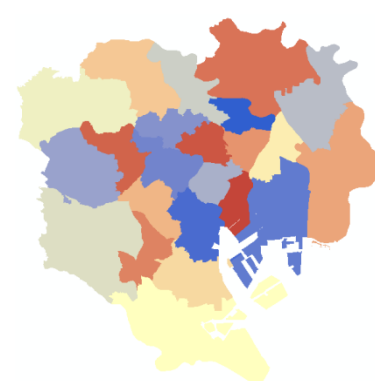
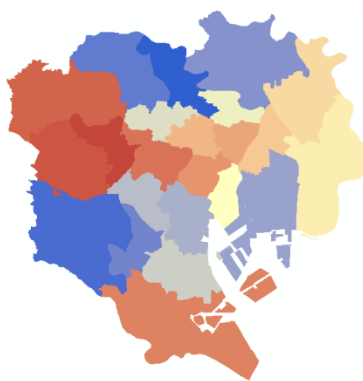
空間相関 = 自分と近隣の相関

正の空間相関

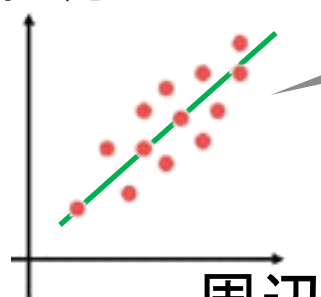
負の空間相関



値
高
低



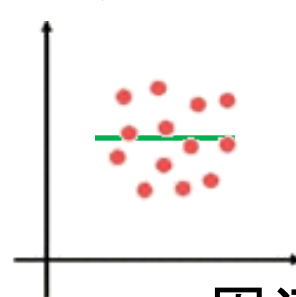
自分



周辺

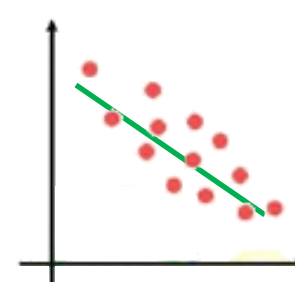
自分が高いと
周辺も低い

自分



周辺

自分



周辺

なぜ**正**の空間相関が生じるか

- 潜在的要因

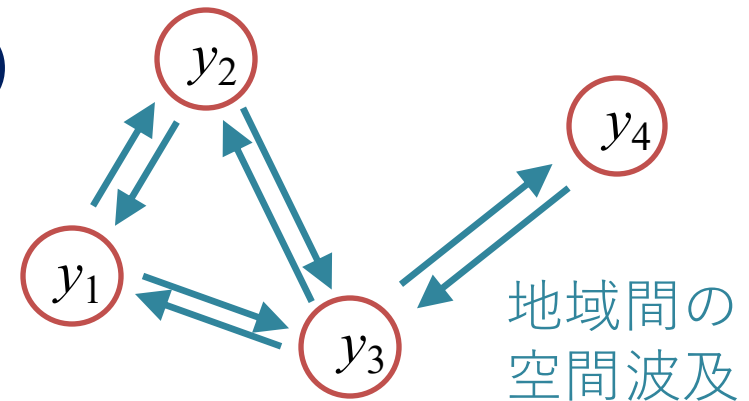
(大気, 水, 施設立地, ...)

例えば地価
の場合



- 空間的な相互作用 (人々, 経済,...)

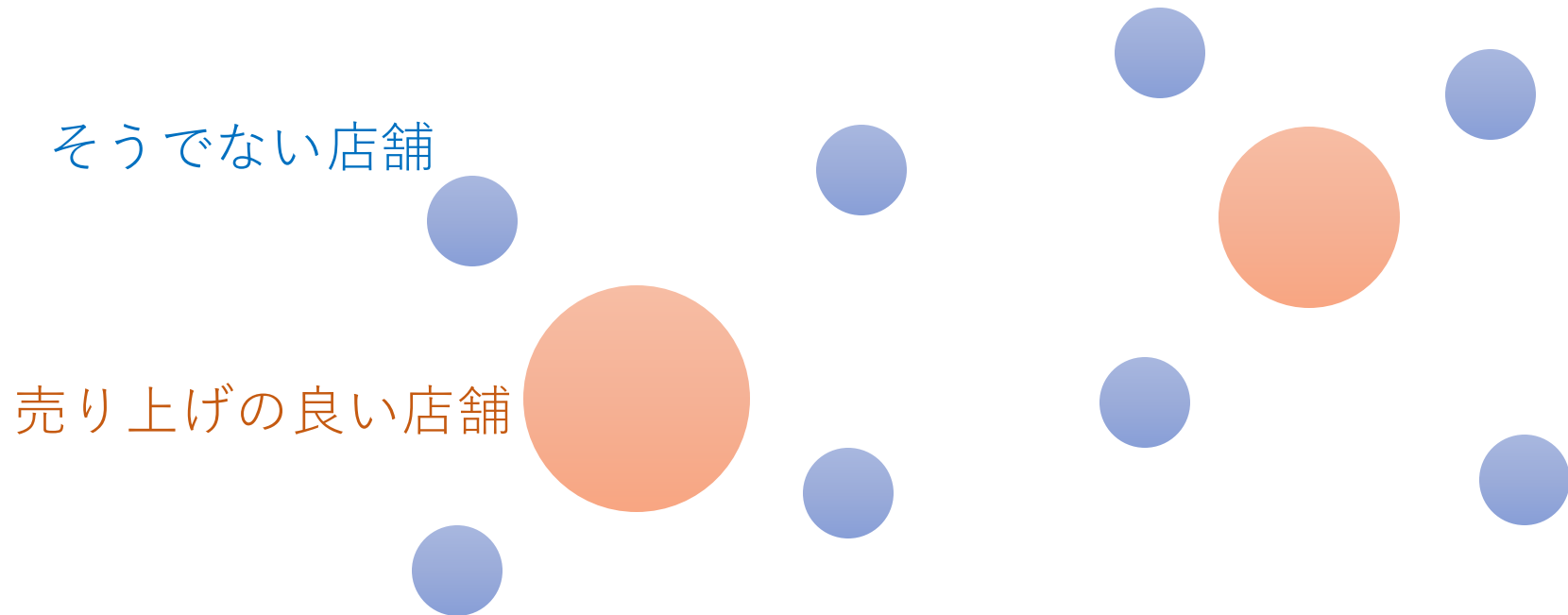
- 人や物が活発に交流した結果
地域間が強固に結びつく
- 隣接自治体が税率を上げたのを見て、
自分も税率を上げる



なぜ**負**の空間相関が生じるか

- **空間競争 (Spatial competition)**

- 例：店舗の売り上げ(大型店の近くの中規模店など)
森林の成長

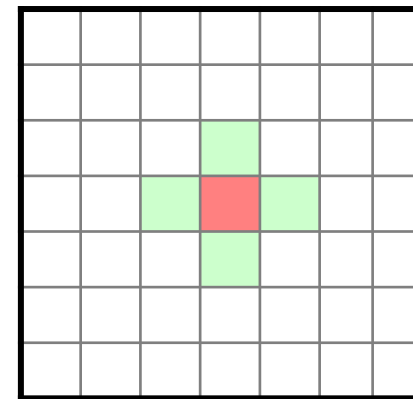


モランI統計量 ≡ 自分と近所との相関係数

近隣は
平均以上？

自分は
平均以上？

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_j - \bar{y})(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2}$$

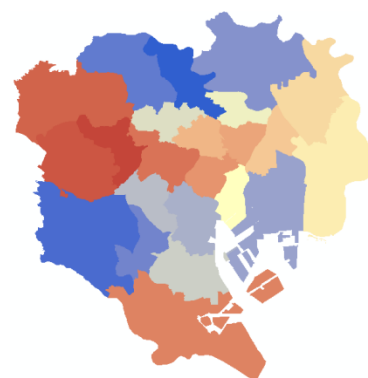


y_i : 地域*i*の観測値 \bar{y} : 標本平均 n : 標本数 w_{ij} : 地域*j*の空間重み (例: 近隣1それ以外0)

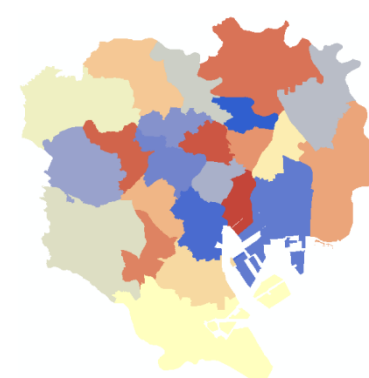
データ
の値
大
小



$I=0.78$



$I \approx 0.00$



$I = -0.28$

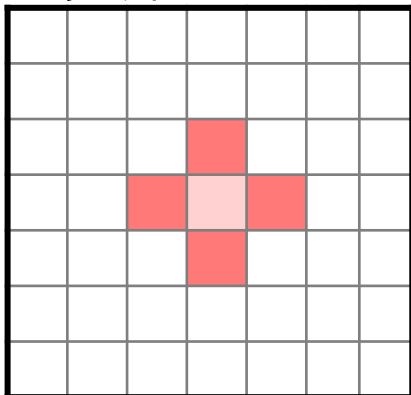
空間重み(w_{ij})の与え方：1または0

隣接に基づく方法

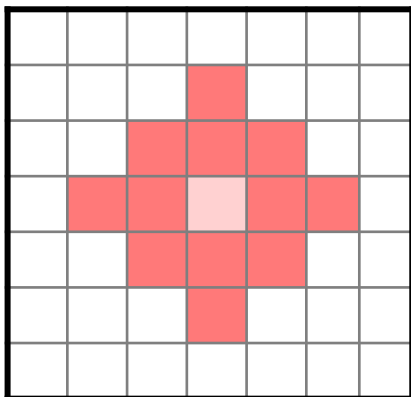
ルーク型

境界を共有

1次

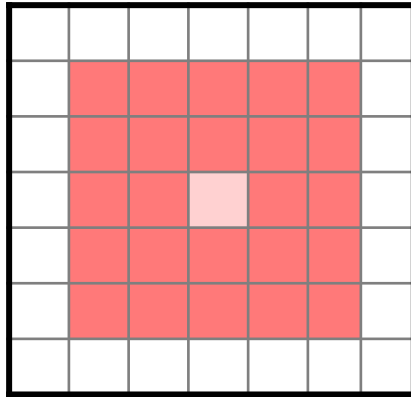
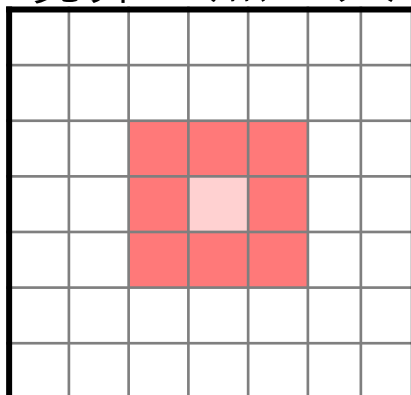


2次



クイーン型

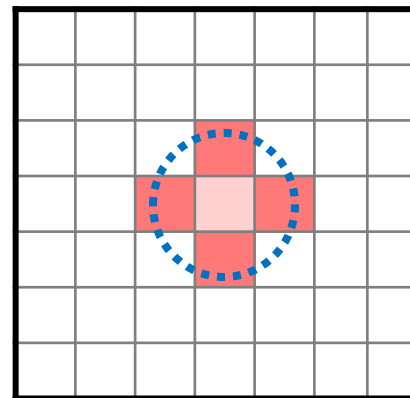
境界or点を共有



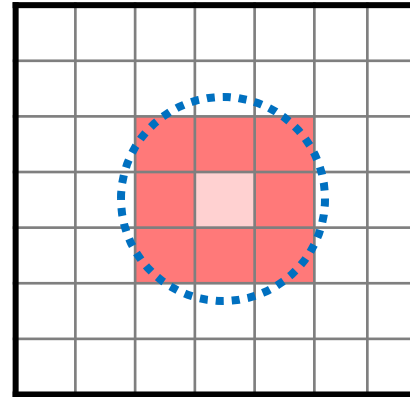
距離に基づく方法

最近隣 k 地点

4近傍

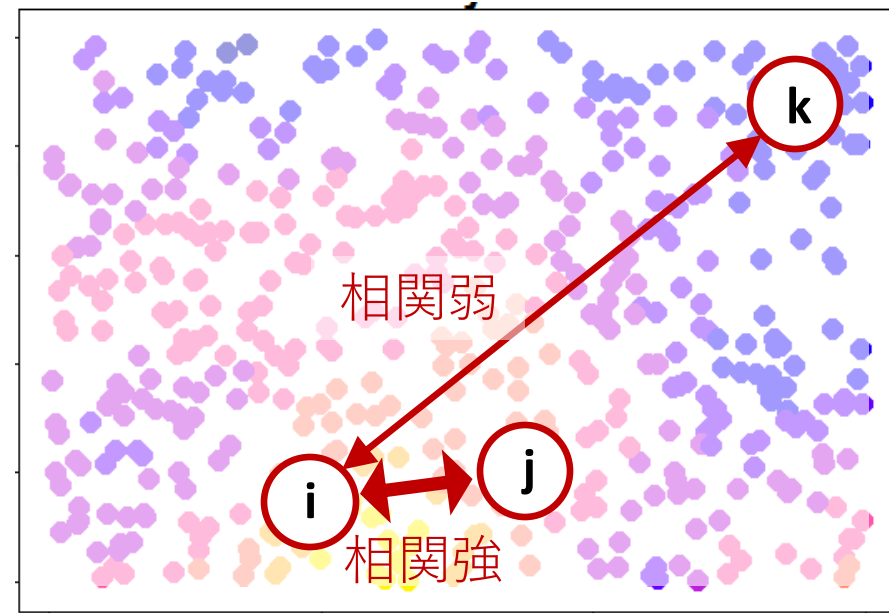
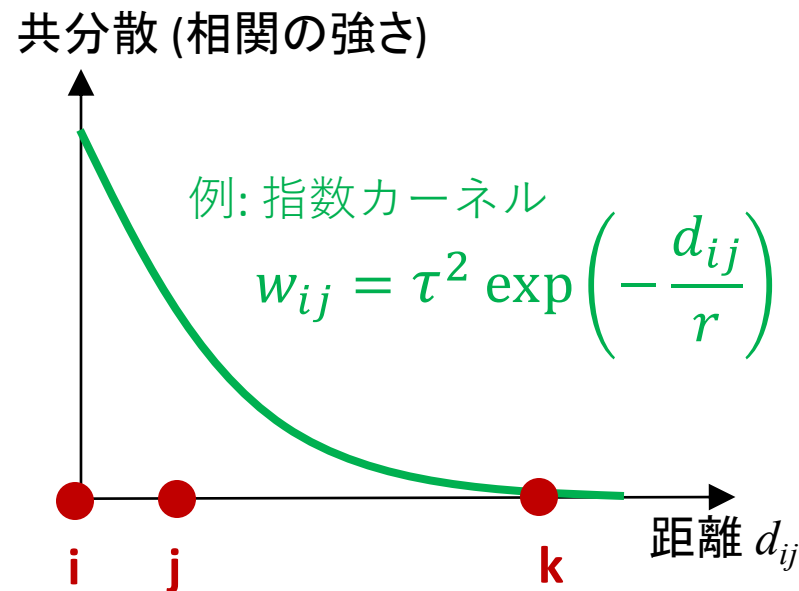


8近傍



空間重み(w_{ij})の与え方：距離の減衰関数/カーネル

- ゾーン*i*からの距離が離れるにつれて空間重みは減衰



モランI統計量の目安(Griffith, 2003)

- おおむね-1～1の間

- **正** : **正** の空間相関

- **負** : **負** の空間相関

- 目安 (Griffith, 2003)

- $I = 0.90 \sim 1.00$: Marked
 - $I = 0.75 \sim 0.90$: Strong
 - $I = 0.50 \sim 0.75$: Moderate
 - $I = 0.25 \sim 0.50$: Weak
 - $I = 0.00 \sim 0.25$: Slight

都道府県別人口

```
> shp      <- system.file("shapes/jpn.shp", package = "NipponMap")[1]
> pref0    <- st_read(shp) :シェープファイルの読み込み(通常はここから)
> pref     <- pref0[pref0$name != "Okinawa",] :沖縄は除外
> st_crs(pref)<-4326      :座標参照系(crs)の指定: 世界測地系(WGS84)+地理座標系
```

```
> pref
```

Simple feature collection with 46 features and 5 fields

Geometry type: MULTIPOLYGON

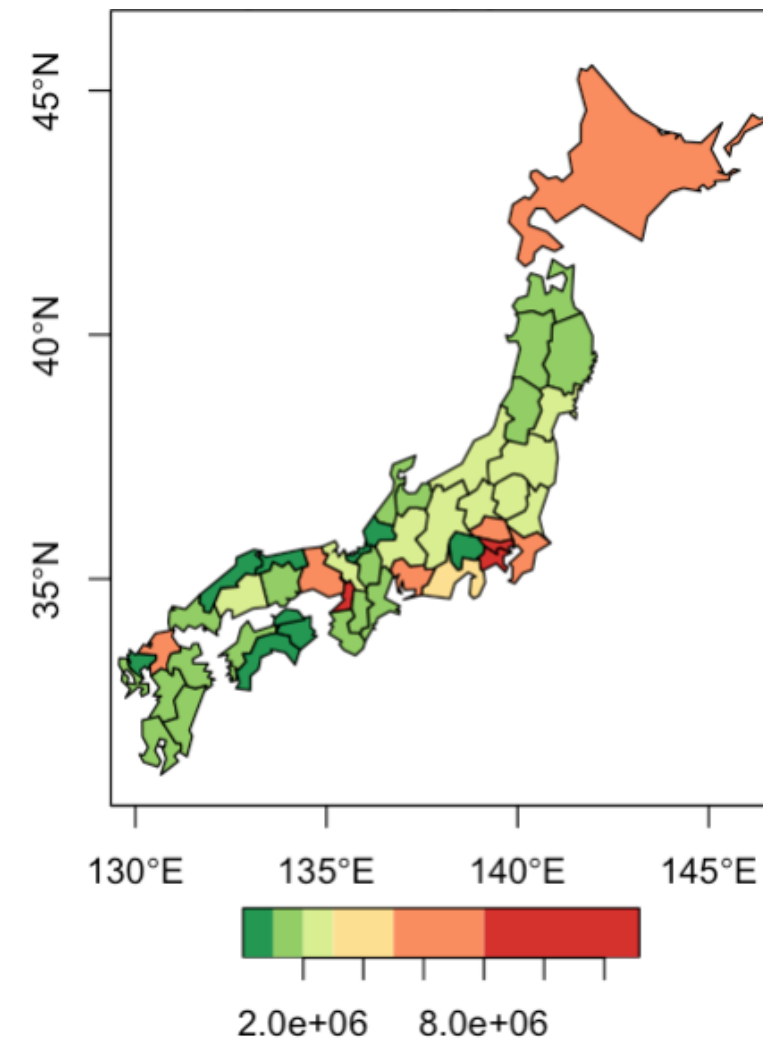
Dimension: XY

Bounding box: xmin: 129.5697 ymin: 30.9898 xmax: 148.8678 ymax: 45.5331

Geodetic CRS: WGS 84

First 10 features:

	SP_ID	jiscode	name	population	region	geometry
1	1	01	Hokkaido	5506419	Hokkaido	MULTIPOLYGON (((139.7707 42...
2	2	02	Aomori	1373339	Tohoku	MULTIPOLYGON (((140.8727 40...
3	3	03	Iwate	1330147	Tohoku	MULTIPOLYGON (((140.7862 39...
4	4	04	Miyagi	2348165	Tohoku	MULTIPOLYGON (((140.2802 38...



モランI統計量の評価

```
> coords <- st_coordinates(st_centroid(pref)) :各都道府県の重心点の位置座標
> knn <- knearneigh(coords,4) :最近隣4都道府県の探索 (knn形式)
> nb <- knn2nb(knn) :最近隣4ゾーンの一覧(nb形式)
> w <- nb2listw(nb) :空間重み (listw形式): 最近隣4ゾーンは1、それ以外は0

> pop <- pref$population :都道府県別人口
> moran <- moran.test(pop, listw=w) :モランI統計量
> moran
```

Moran I test under randomisation

data: pop
weights: w

Moran I statistic standard deviate = 3.8875, p-value = 5.064e-05
alternative hypothesis: greater
sample estimates:

Moran I statistic
0.320743957

Expectation	Variance
-0.022222222	0.007783232

帰無仮説の下での
期待値と分散

正の空間相関があると言って良い？

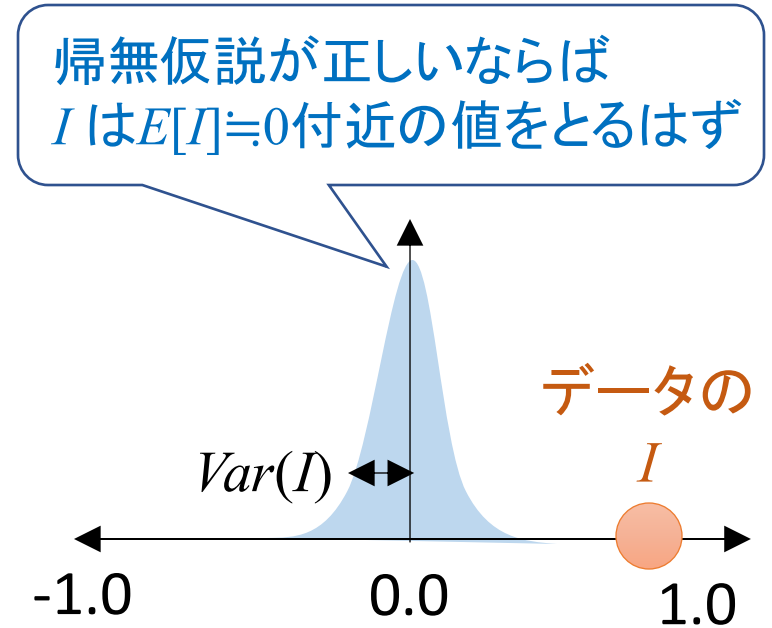
補足：仮説検定

1: 仮説を決める

- 帰無仮説： $I = 0$ (空間相関なし)
- 対立仮説： $I \neq 0$

2: 帰無仮説の下での I の分布を求める

- 帰無仮説が正しい場合に I がとりうる値の範囲を求める $\rightarrow E[I], Var[I]$



3: データからモラン I 統計量を求める

- データから得られた I が上記範囲の外側
 \rightarrow 帰無仮説は棄却 $\rightarrow I \neq 0$ を採択

都道府県別人口密度の場合

```
> moran <- moran.test(pop, listw=w)
> moran
```

Moran I test under randomisation

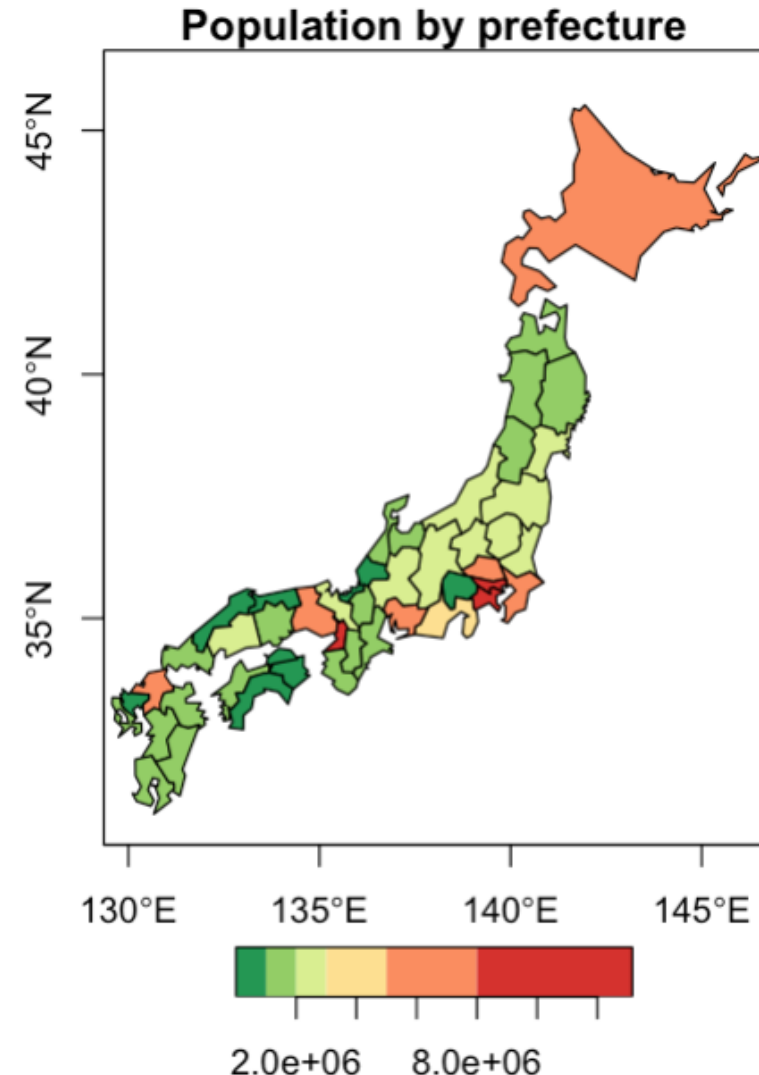
data: pop
weights: w

Moran I statistic standard deviate = 3.8875, p-value = 5.064e-05

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.320743957	-0.022222222	0.007783232



帰無仮説(空間相関なし)が正しい場合に今回の推定値を得る確率(p値)は0.00005064

- p値が0.10以下：10%水準で統計的に有意 ←より有意になりやすい
- p値が0.05以下：5%水準で統計的に有意 ←最も標準的
- p値が0.01以下：1%水準で統計的に有意 ←より有意になりにくい

大域と局所の空間相関

- Global indicator of spatial association (GISA)

- 全域についての検定
- モランI統計量, ギアリーC統計量,...

分解



モランIでは
地域差がみれない。。

- Local indicator of spatial association (LISA)

- 地域（地点）毎の検定
- ローカルモランI統計量, ローカルギアリーC統計量,...

モランI統計量の分解

$$I \rightarrow \frac{1}{n} \sum_{i=1}^n I_i$$

整理

ローカルモランI統計量

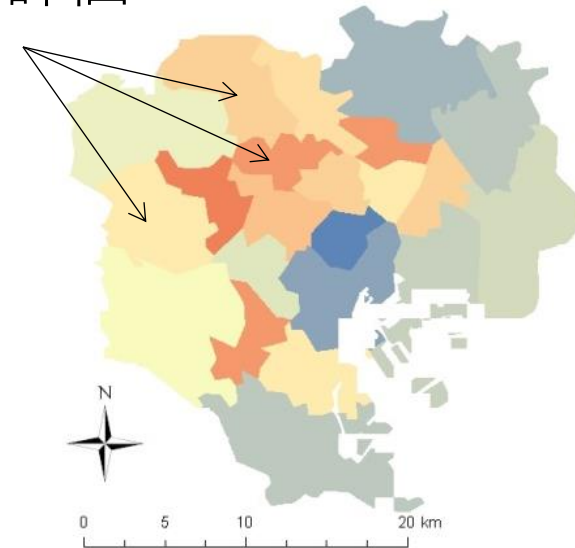
• 自分と**周辺**の相関を評価

$$I_i = \frac{1}{m} (y_i - \bar{y}) \sum_{j=1}^J w_{ij} (y_j - \bar{y})$$

自分

周辺

各ゾーン周辺の
空間相関を評価



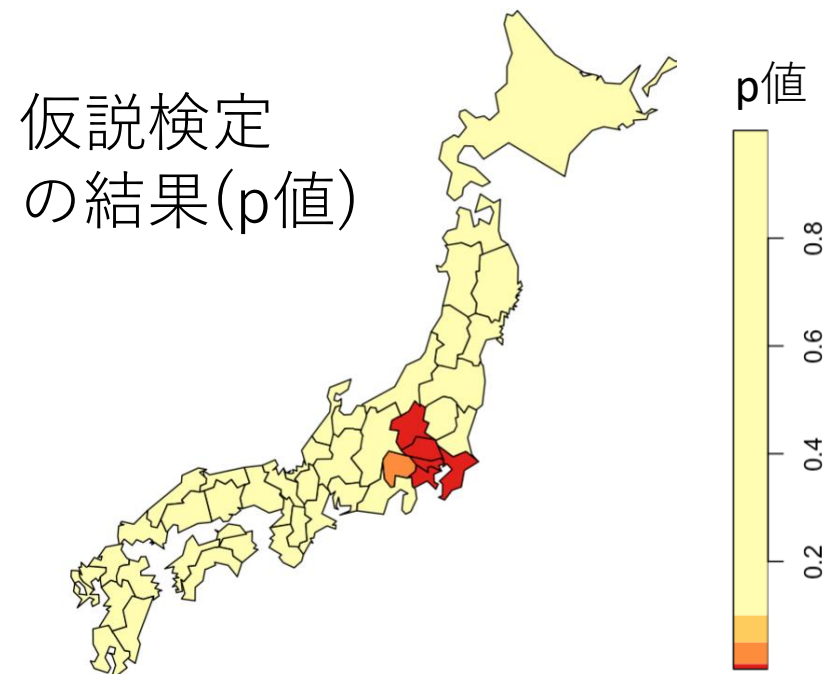
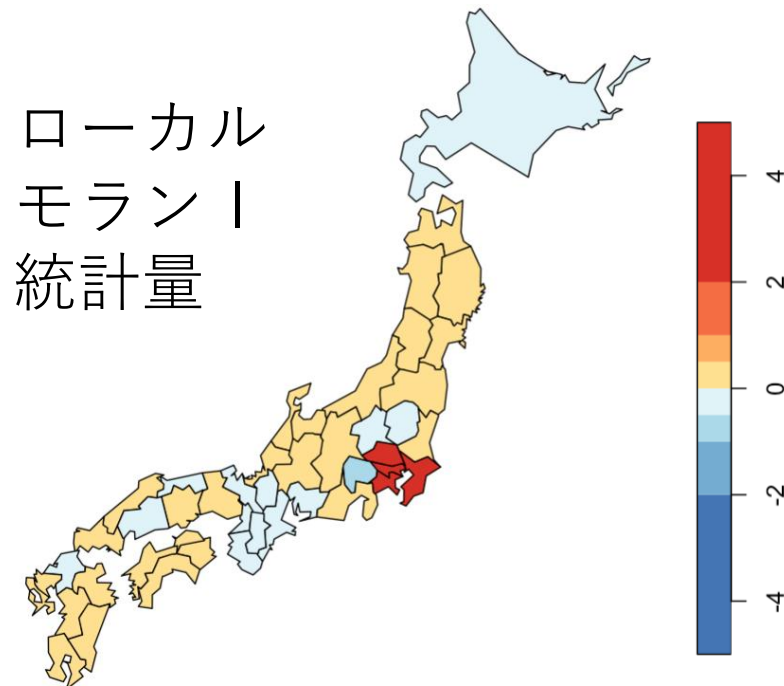
$$m = \frac{1}{n-1} \sum_{i=1, i \neq j}^I (y_i - \bar{y})^2 - \bar{y}^2$$

ローカルモランI統計量の評価

```
> lmoran <- localmoran(pop, listw=w)
> lmoran[1:3,]
```

	Ii	E.Ii	Var.Ii	Z.Ii	Pr(z != E(Ii))
1	-0.4690880	-0.023537521	0.24628916	-0.8977890	0.3692980
2	0.2450530	-0.005917060	0.06303149	0.9996390	0.3174853
3	0.2505732	-0.006293181	0.06701276	0.9922671	0.3210673

li : ローカルモランI
E.li, Var.li : 帰無仮説の下でのliの期待値と分散
Z.li : z値 = $(li - E.li) / \sqrt{Var.li}$ 。liを基準化したもの。Z.li $\geq |1.96|$ の場合に5%水準で有意
Pr(z!=E(li)) : p値 = 帰無仮説



モラン散布図

$$I_i = \frac{1}{m} (y_i - \bar{y}) \sum_{j=1}^J w_{ij} (y_j - \bar{y})$$

自分 周辺

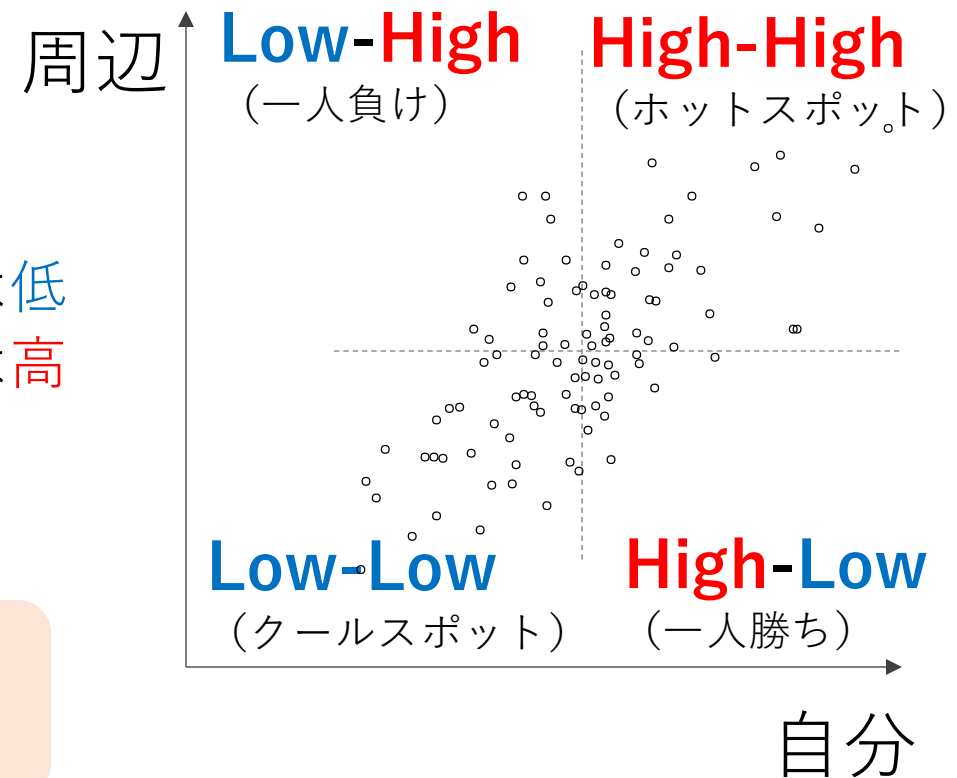
I_i が大きい2ケース

- **High-High** : 自分+周辺が高
- **Low-Low** : 自分+周辺が低

I_i が小さい2ケース

- **High-Low** : 自分は高, 周辺は低
- **Low-High** : 自分は低, 周辺は高

地域が分類できる



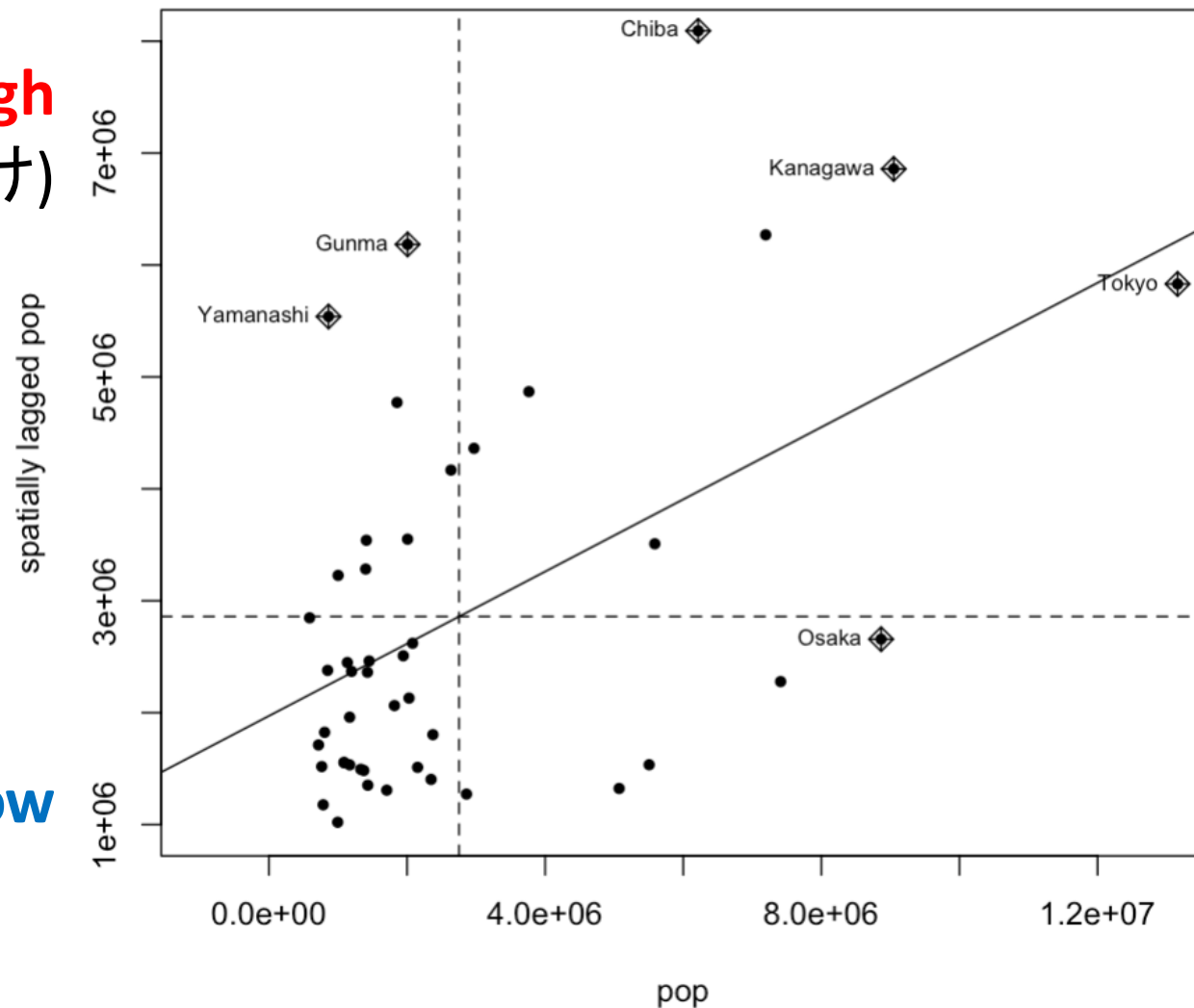
モラン散布図

都道府県名 図形の形(●)

```
> moran.plot(pop, listw=w, labels=pref$name, pch=20,  
+            xlim=c(-1000000,13000000),ylim=c(1000000,8000000))
```

Low-High
(一人負け)

Low-Low



High-High

High-Low
(一人勝ち)

モラン散布図に基づく分類図

```
> moran_cl<- attr(lmoran,"quadr")$mean :分類結果
> pref$moran_cl<-moran_cl :分類結果を都道府県別の新たな列(moran_cl)に追加
> levels(moran_cl) :分類の確認
[1] "Low-Low" "High-Low" "Low-High" "High-High"
```

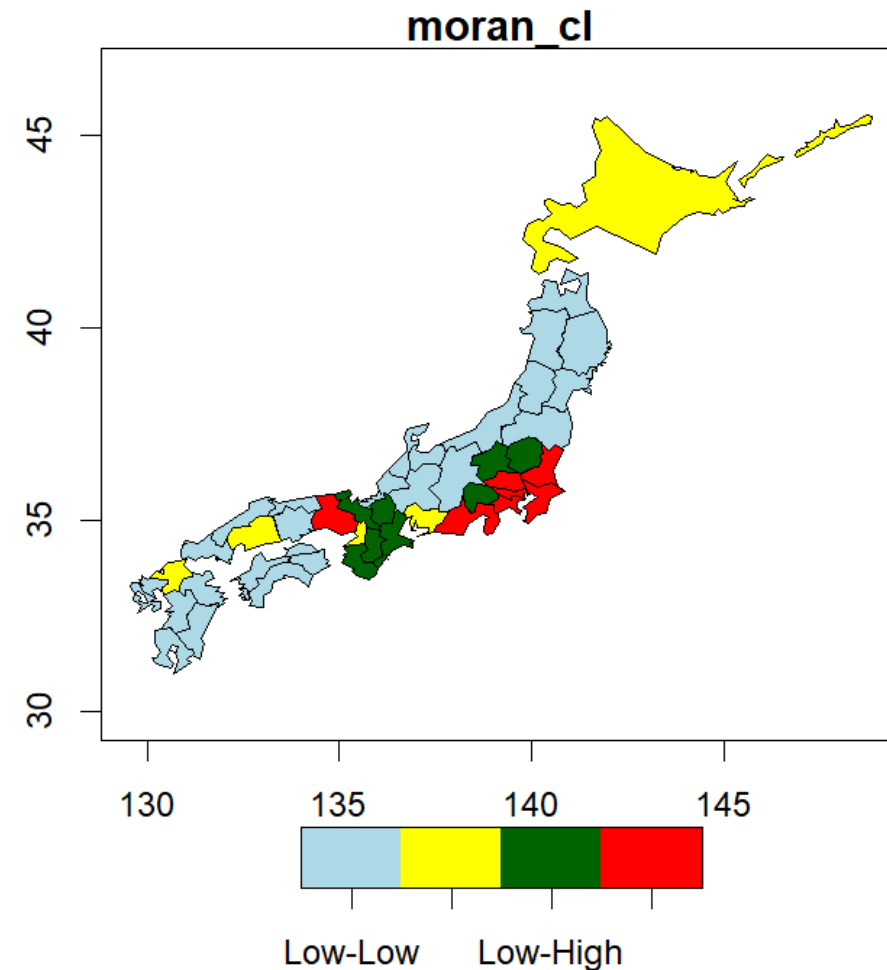
```
> plot(pref[, "moran_cl"],
+       axes=TRUE, :位置座標を表示
+       pal = c("light blue", :Low-Low:薄い青
+              "yellow", :High-Low:黄
+              "dark green", :Low-High:濃い緑
+              "red"), :High-High: 赤
+       key.pos=1, :凡例の位置
+       key.length=0.5) :凡例の長さ
```

High-High: 関東、兵庫

High-Low : 北海道, 愛知, 大阪, 広島, 福岡

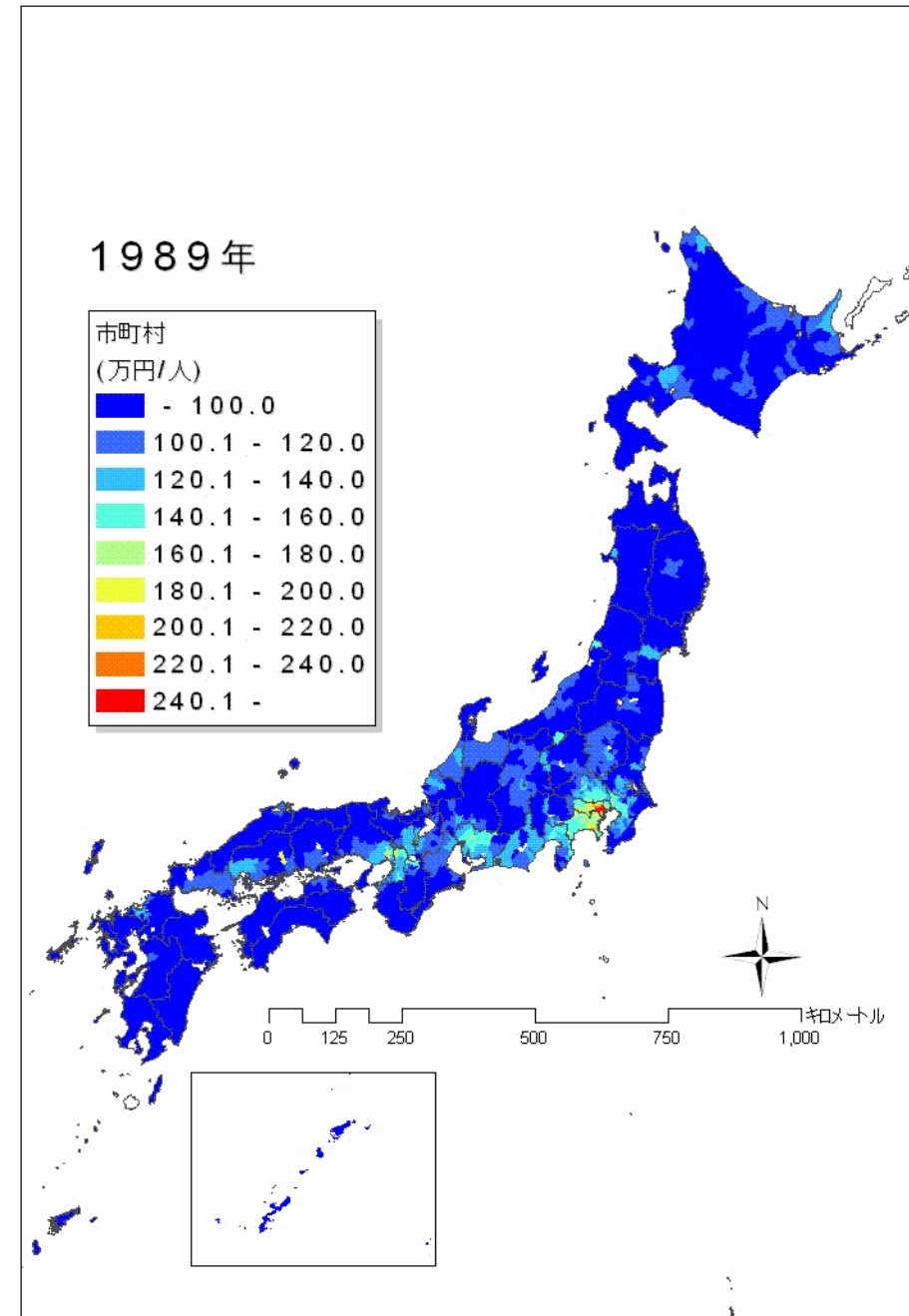
Low-High : 栃木, 群馬, 京都, 滋賀, 三重, 和歌山

Low-Low : その他



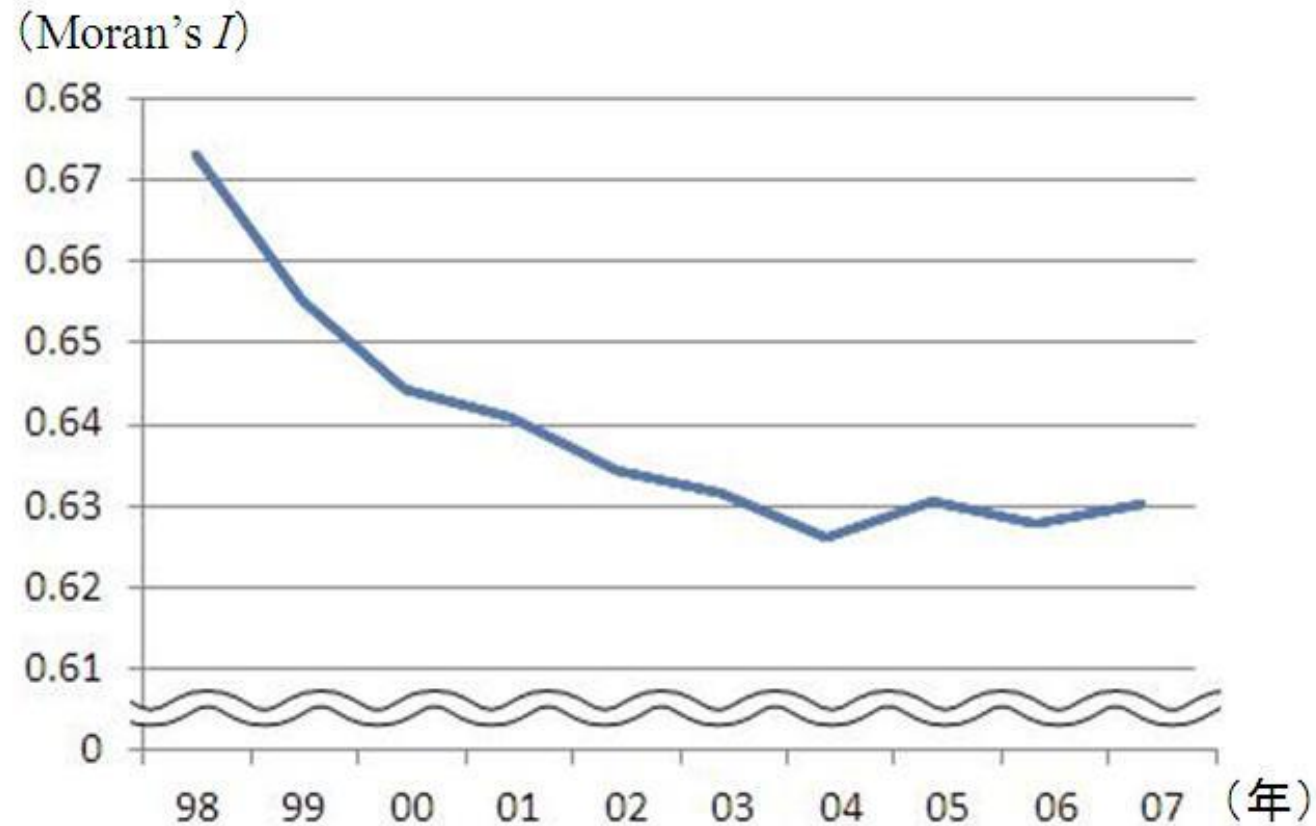
我が国の 所得格差分析 (1998-2007) への応用

Tamesue, K., Tsutsumi, M., & Yamagata, Y.
(2013). Income disparity and correlation in
Japan. *Review of Urban & Regional
Development Studies*, 25(1), 2-15.

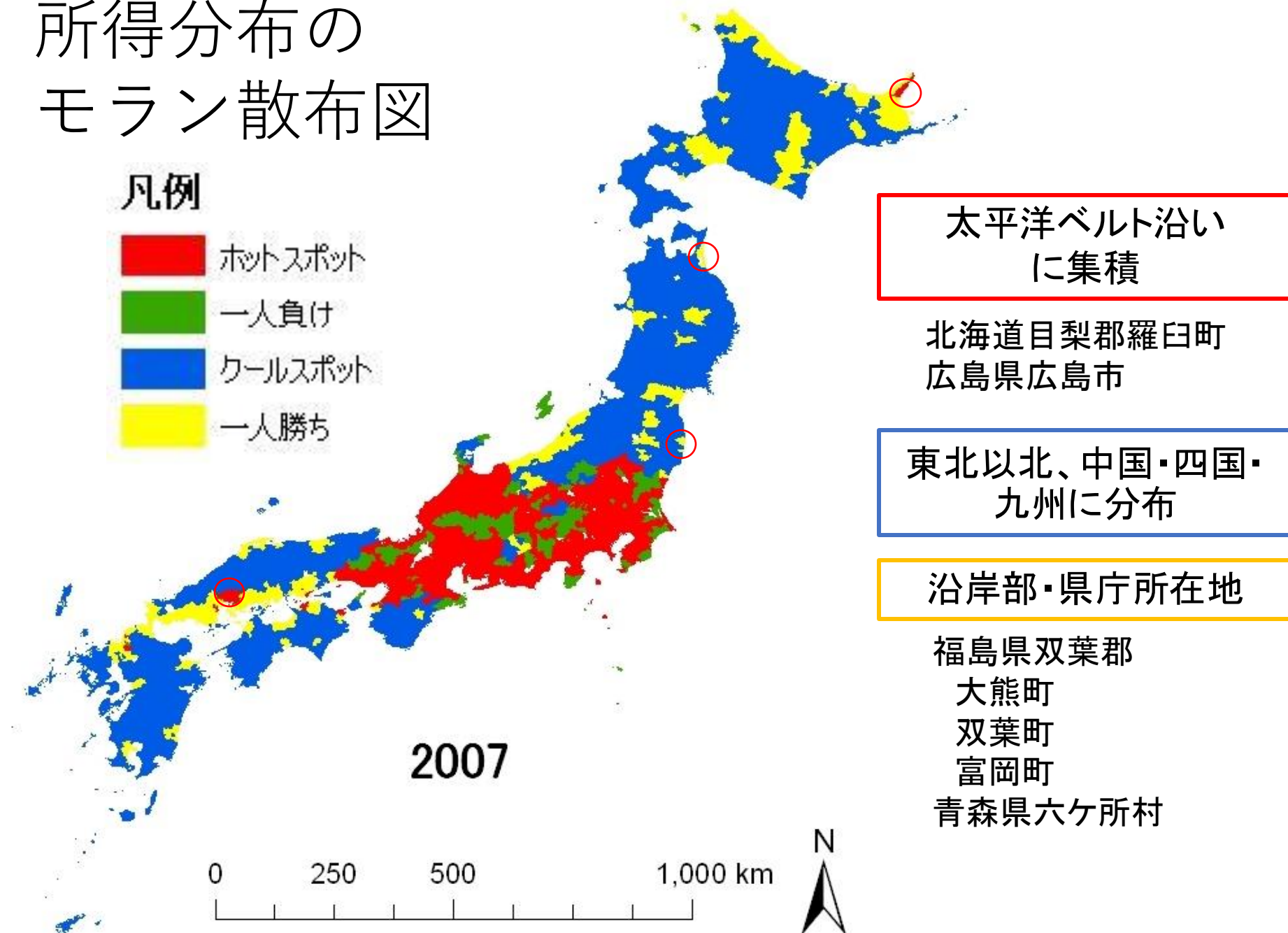


モランI統計量の変化

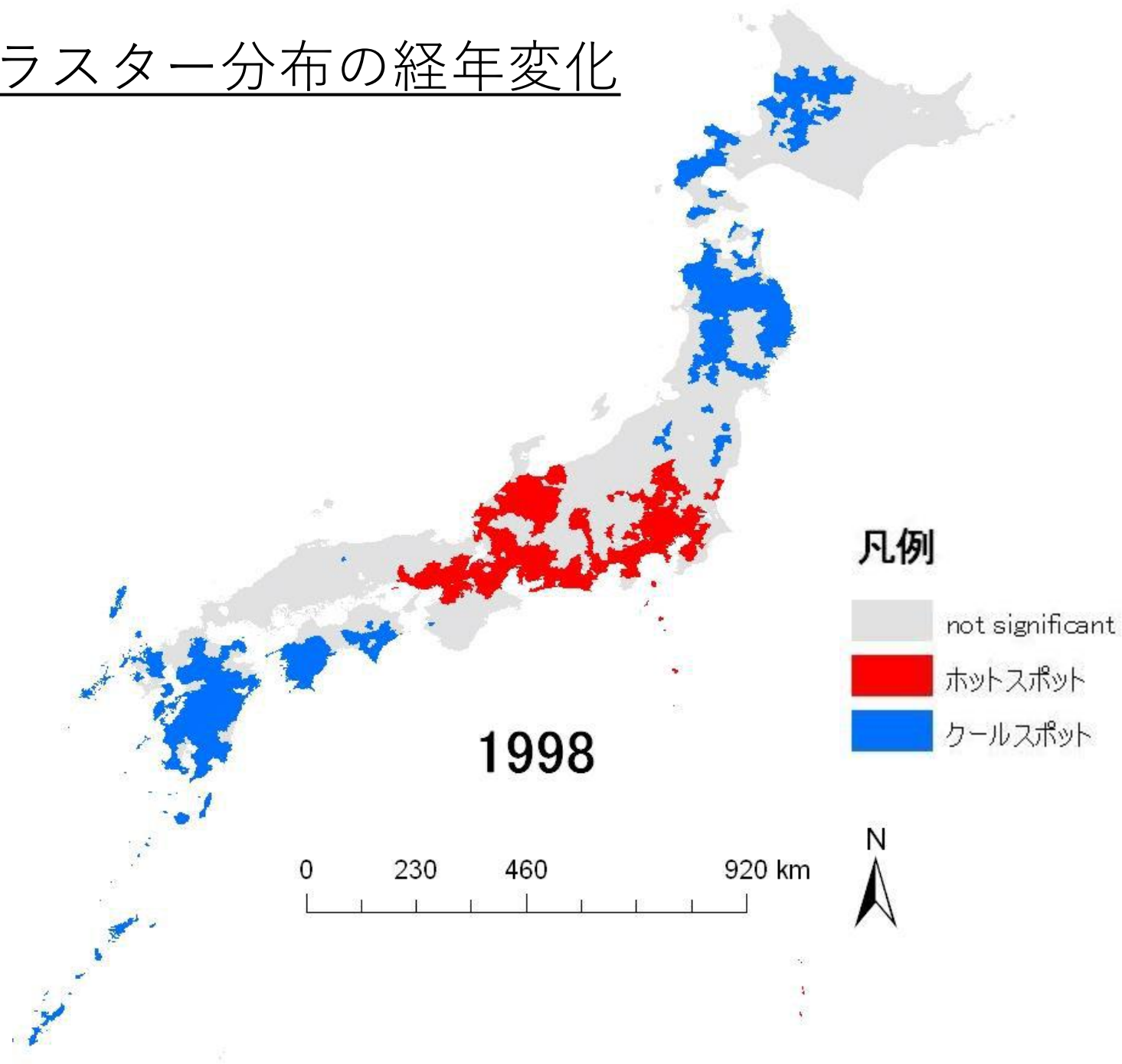
- Tamesue et al. (2013)
 - 市区町村別の一人当たり所得
 - 年々減衰。隣接市区町村との相関が弱まる傾向



所得分布の モラン散布図



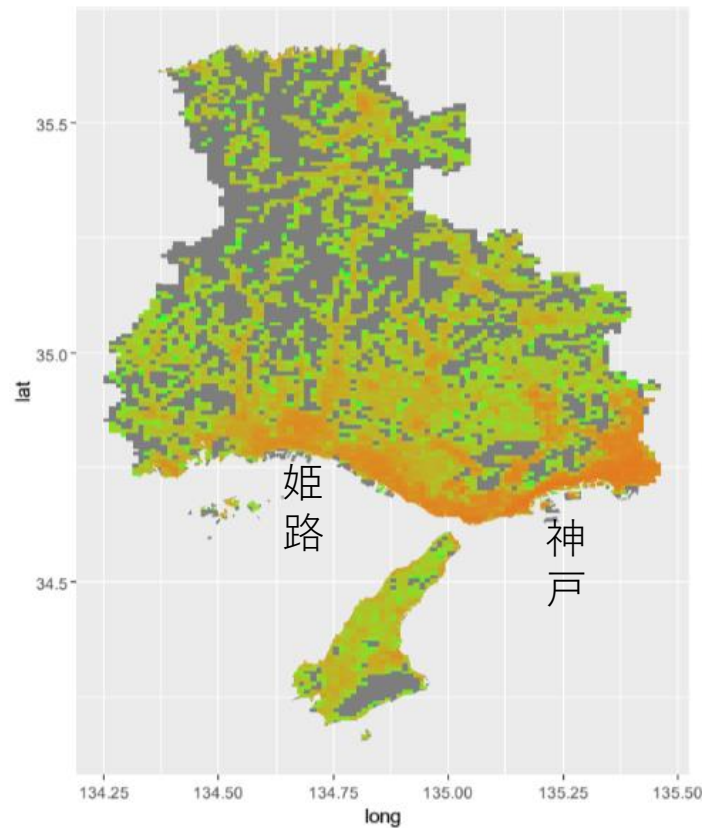
有意クラスター分布の経年変化



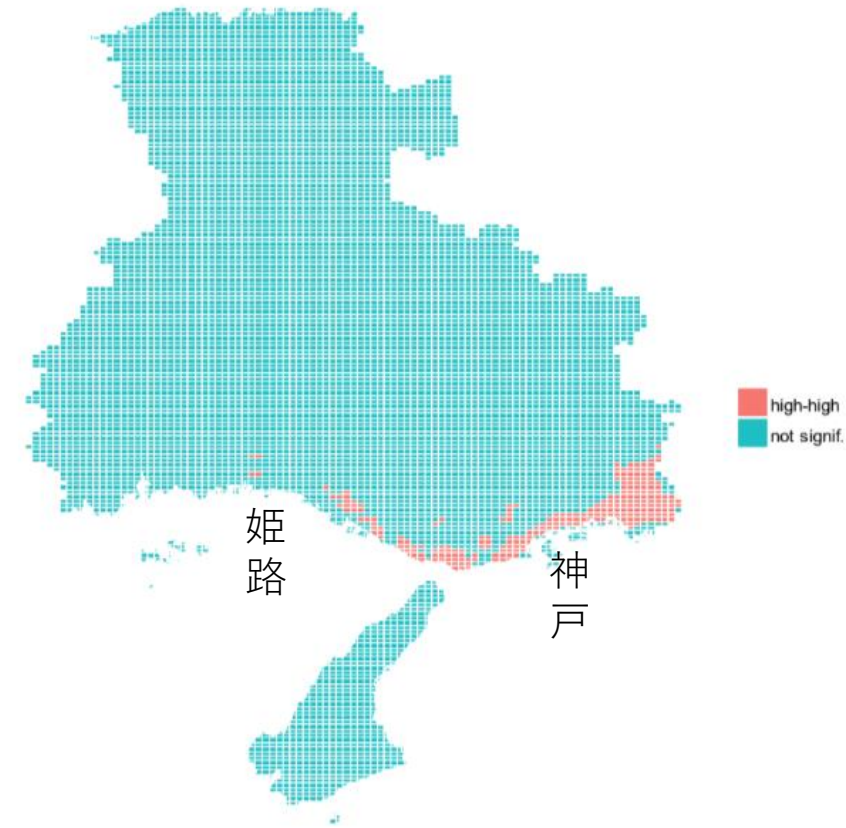
モランI統計量の特徴

瀬谷創准教授(神戸大)
よりお借りしました

- 平均との比較に基づくため、局所的な傾向に対する感度が低い



人口密度(兵庫県)



モラン散布図

ギアリーC統計量

自分と近隣の差

$$C = \frac{n-1}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (y_i - y_j)^2}{\sum_i (y_i - \bar{y})^2}$$

y_i : i 番目の標本

\bar{y} : 標本平均

n : 標本数

w_{ij} : 空間重み (例: 近隣は1それ以外0)

正の空間相関

$C < 1$



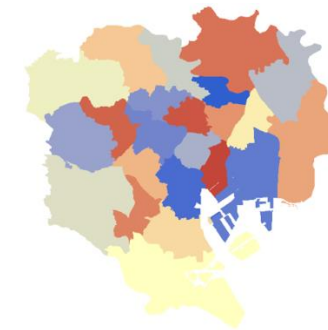
空間相関なし

$C \approx 1$



負の空間相関

$C > 1$



ローカル・ギアリーC統計量

$$C_i = \sum_j w_{ij} (y_i - y_j)^2$$

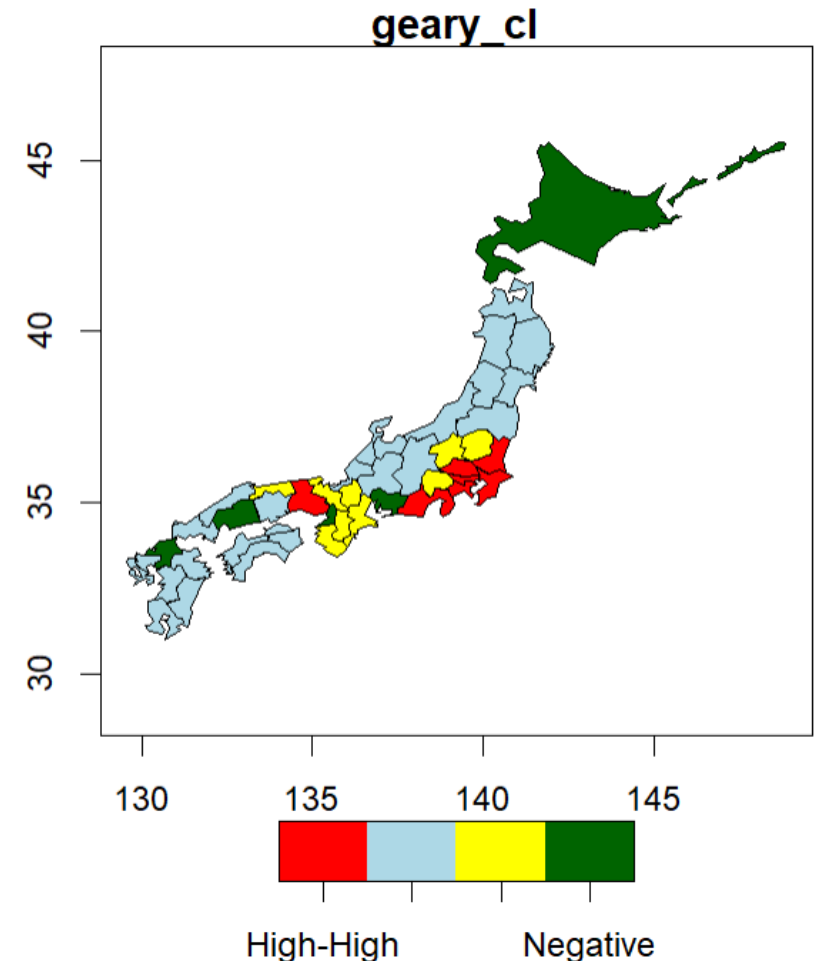
ローカル・ギアリーC統計量に基づく分類図

```
> lgeary0 <- localC_perm(pop, listw=w) : ローカルGeary's C
> geary_cl <- attr(lgeary0,"cluster")   : 分類結果
> pref$geary_cl <- geary_cl             : 分類結果を都道府県別の新たな列(geary_cl)に追加
> levels(geary_cl)                     : 分類の確認
[1] "High-High"      "Low-Low"        "Other Positive" "Negative"

> plot(pref[, "geary_cl"],
+       axes=TRUE,
+       pal = c("red",           : High-High : 赤
+               "light blue",   : Low-Low  : 薄青
+               "yellow",       : Other.Pos.: 黄色
+               "dark green"), : Negative : 濃緑
+       key.pos=1,
+       key.length=0.5)
```

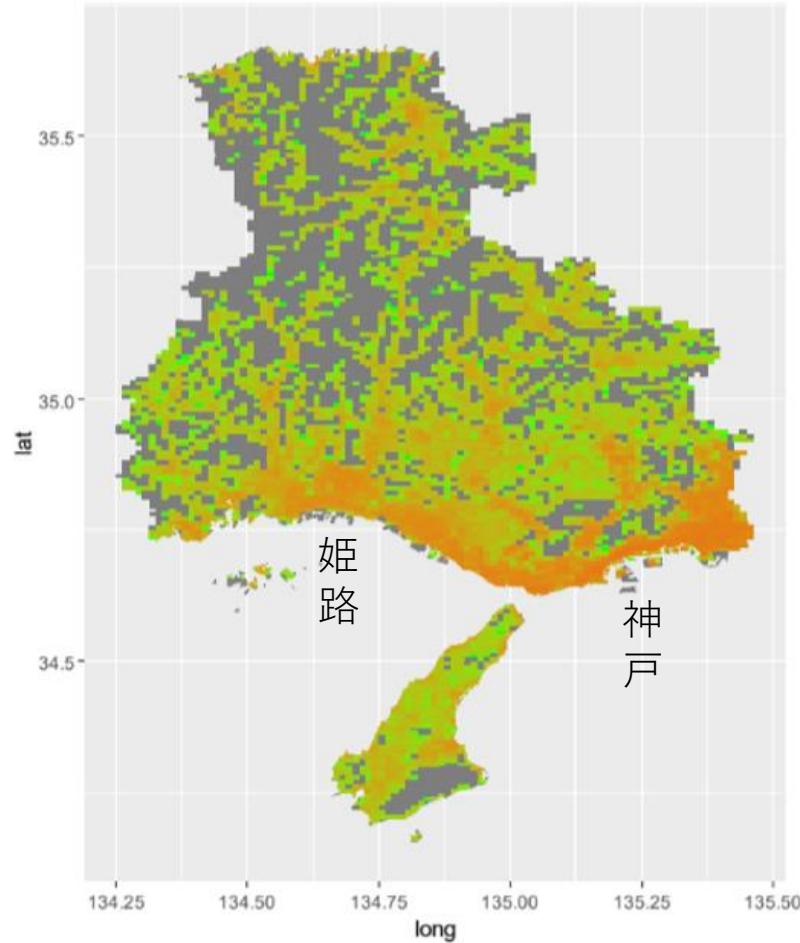
ギアリーCに基づく4分類

High-High	: 自分も隣も高い	} $C < 1$
Low-Low	: 自分も隣も低い	
Other positive:	隣接と類似(上記以外)	
Negative	: 隣接と逆の傾向	$C > 1$

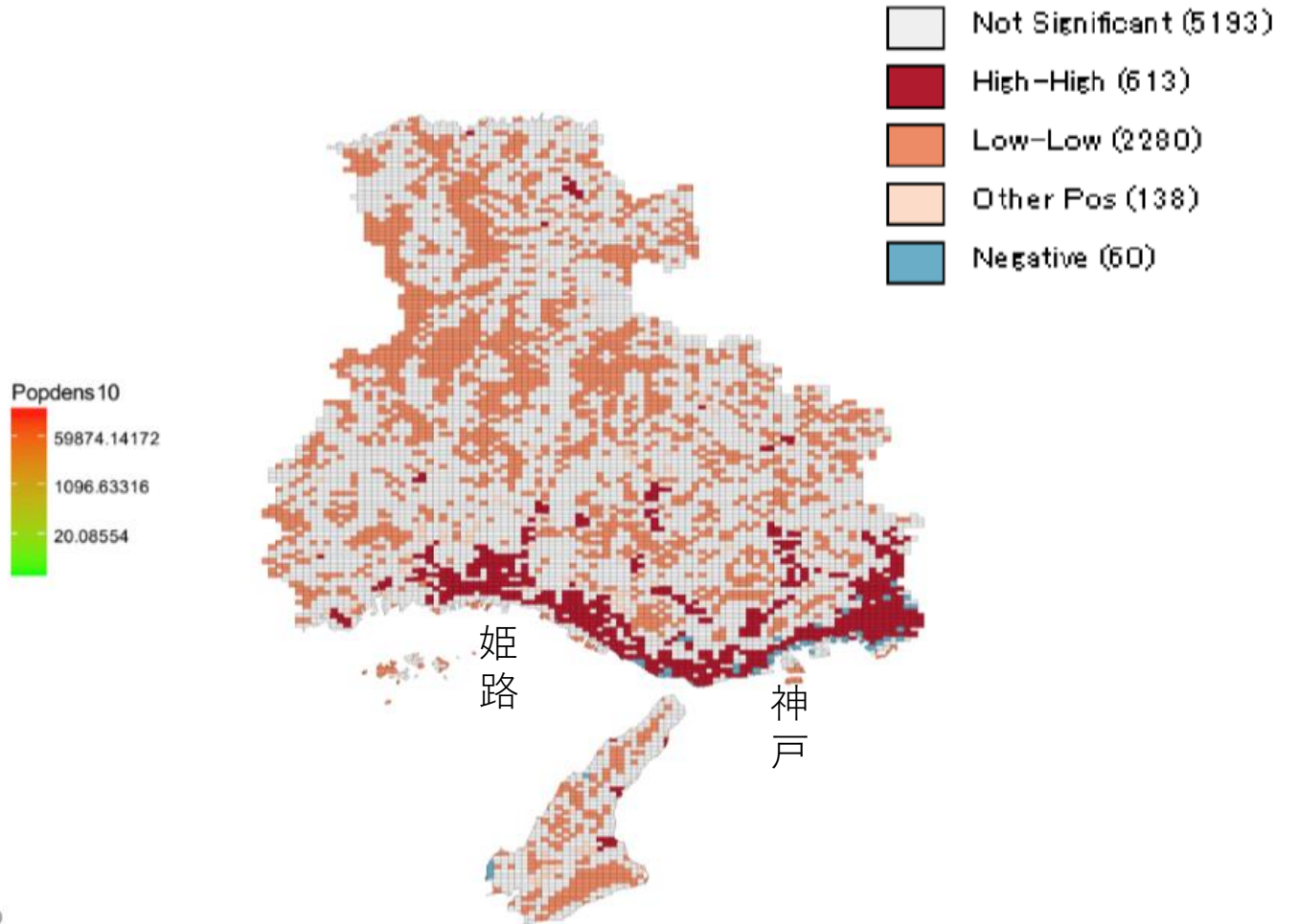


ギアリCの方が局所的傾向が捉えられる

※ただし、個人的経験としては解釈しにくい結果が得れる場合もあり、モランIの方が解釈性に優れる印象



人口密度(兵庫県)



ローカルギアリーCに基づく分類

まとめ

- 空間データの基本的な処理・作図方法を紹介
- **spdep**による探索的な空間分析の方法を説明
 - 大域空間統計量
 - モランI, ギアリーC, ...
 - 局所空間統計量
 - ローカルモランI, ローカルギアリーC, Getis/Ord G*, LOSH統計量, ...
 - データの特性を把握するために広く用いられている。