

# Stataによる空間計量分析

## Mataを用いたプログラム開発を通じて

2017年9月16日（土）  
2017 Japanese Stata Users  
Group Meeting  
京都リサーチパーク

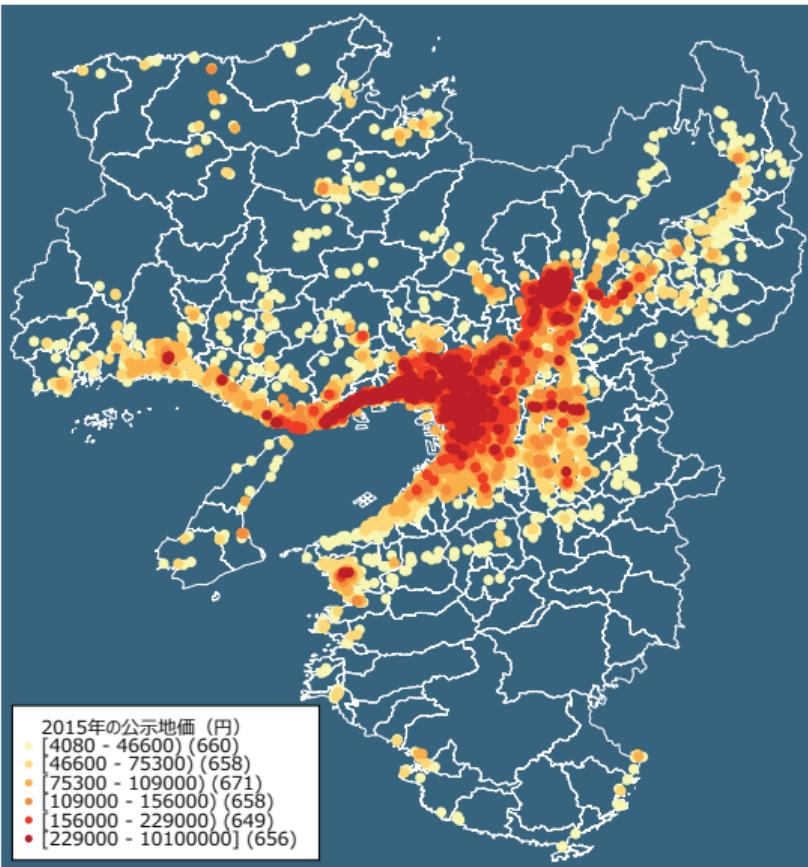
近藤恵介  
( 経済産業研究所 )



## ❖ 自己紹介 ❖

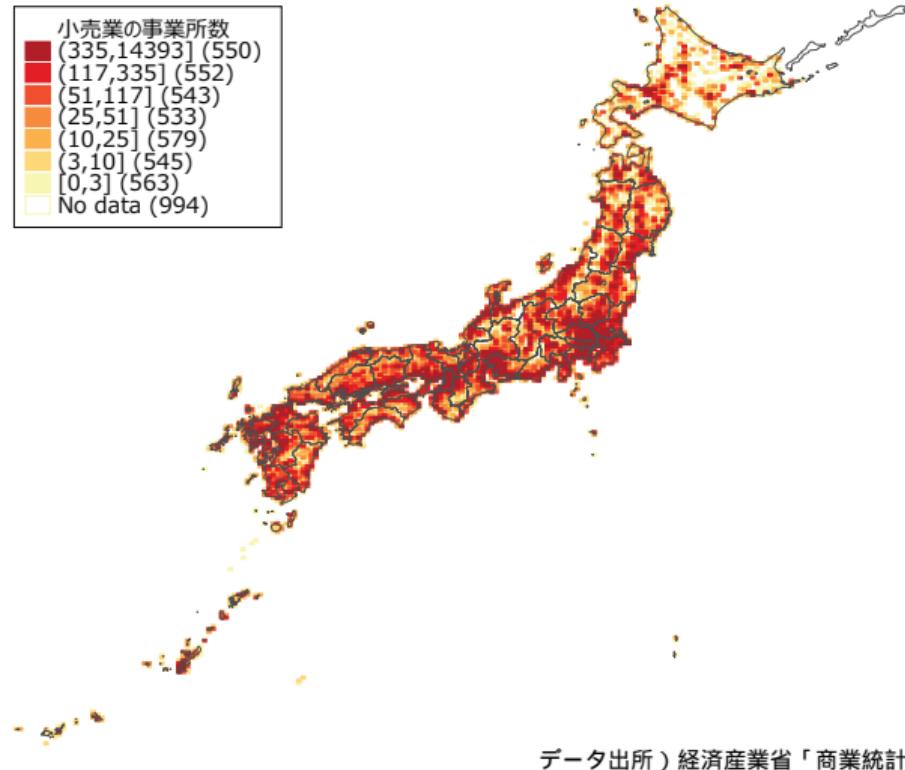
- ☞ 研究分野：空間経済学，集積の経済学
- ☞ これまでに開発したStata コマンド
  - ▶ getisord ( Hot Spot 分析を行うコマンド, *Stata Journal* )
  - ▶ spgen ( 空間ラグ変数を計算するコマンド, RePEc )
  - ▶ estquant ( Quantile Approach を行うコマンド, RePEc )

# はじめに



☞ Stata で地図表示が可能

小売業の事業所数	
(335,14393]	(550)
(117,335]	(552)
(51,117]	(543)
(25,51]	(533)
(10,25]	(579)
(3,10]	(545)
[0,3]	(563)
No data	(994)



## ◆ Stata 15 の新機能 ◆

### ☞ 空間計量経済学を公式にサポート

- ▶ 地図描画も公式コマンドとして採用されている
- ▶ **Sp コマンド**として各種コマンドが用意されている

# 1. 地理情報システム (GIS) と Stata

## ◆ 報告のアウトライン

1. 地理情報システム (GIS) と Stata
2. 空間計量分析とは？
3. プログラミング言語 Mata とは？
4. Stata による空間計量分析の例

# 1. 地理情報システム (GIS) と Stata

## ▶ 地理情報システム (Geographic Information System) とは

「地理的位置を手がかりに，位置に関する情報を持ったデータ（空間データ）を総合的に管理・加工し，視覚的に表示し，高度な分析や迅速な判断を可能にする技術である。」

出所 ) 国土地理院 (<http://www.gsi.go.jp/GIS/whatisgis.html>)

☞ 「視覚化」することで理解が深まる

# 1. 地理情報システム (GIS) と Stata

## ▶ Stata でも GIS が可能

Stata 14 以前は User-Written コマンドとして提供

- ▶ `shp2dta` ( 地図データ Shape 形式を Stata 形式に変換 )
- ▶ `spmap` ( データを地図上に表示 )

Stata 15 から正式にサポート

- ▶ `spshape2dta` ( 地図データ Shape 形式を Stata 形式に変換 )
- ▶ `grmap` ( データを地図上に表示 )

# 1. 地理情報システム (GIS) と Stata

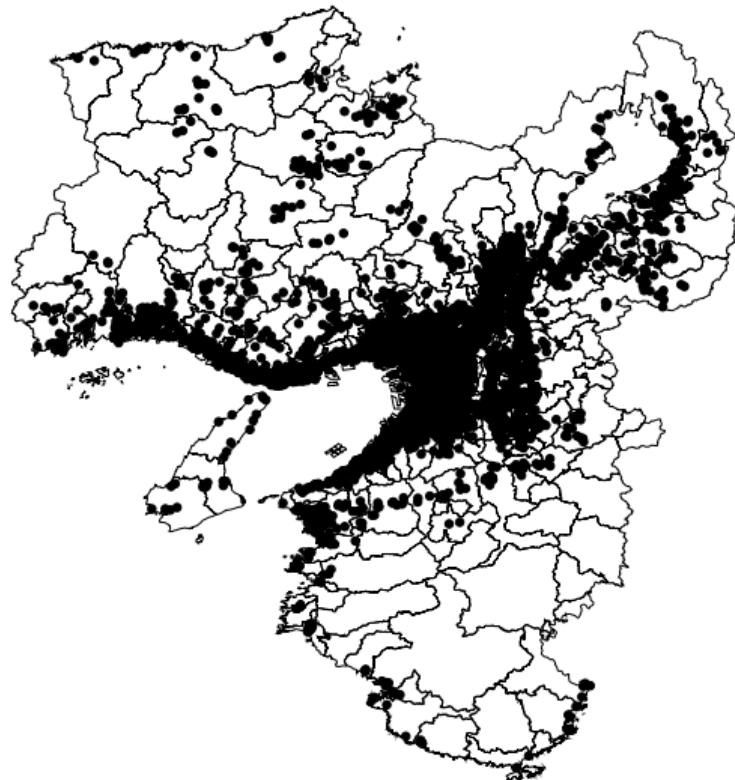


Fig: 関西圏の地価公示の調査地点

- ▶ ここでやっていること
  - 1. 地図 (Shape ファイル) の用意  
→ `shp2dta` で Stata 形式に変換
  - 2. 調査地点の緯度経度データを用意
  - 3. 地図を背景にして調査地点を表示  
→ `spmap` で地図表示が可能
- ☞ Stata 15 では , `spshape2dta`, `grmap`

▶ Code

データ出所 ) 国土数値情報、地価公示

# 1. 地理情報システム (GIS) と Stata

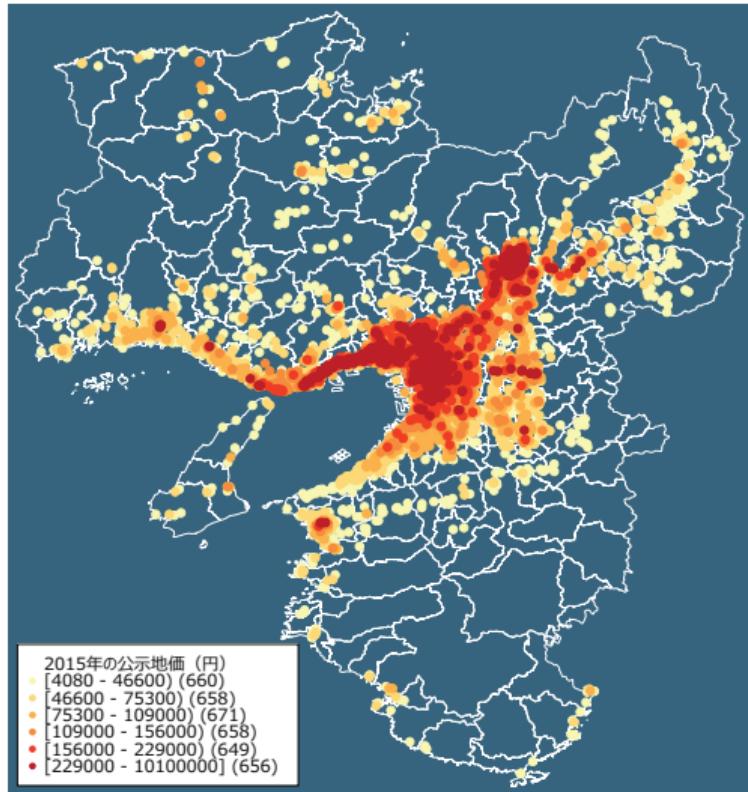


Fig: 関西圏の地価公示

さらに ,

- ▶ 地点上に公示地価の情報を付加
- ▶ 装飾の変更

▶ [Code](#)

データ出所 ) 国土数値情報、地価公示

# 1. 地理情報システム (GIS) と Stata

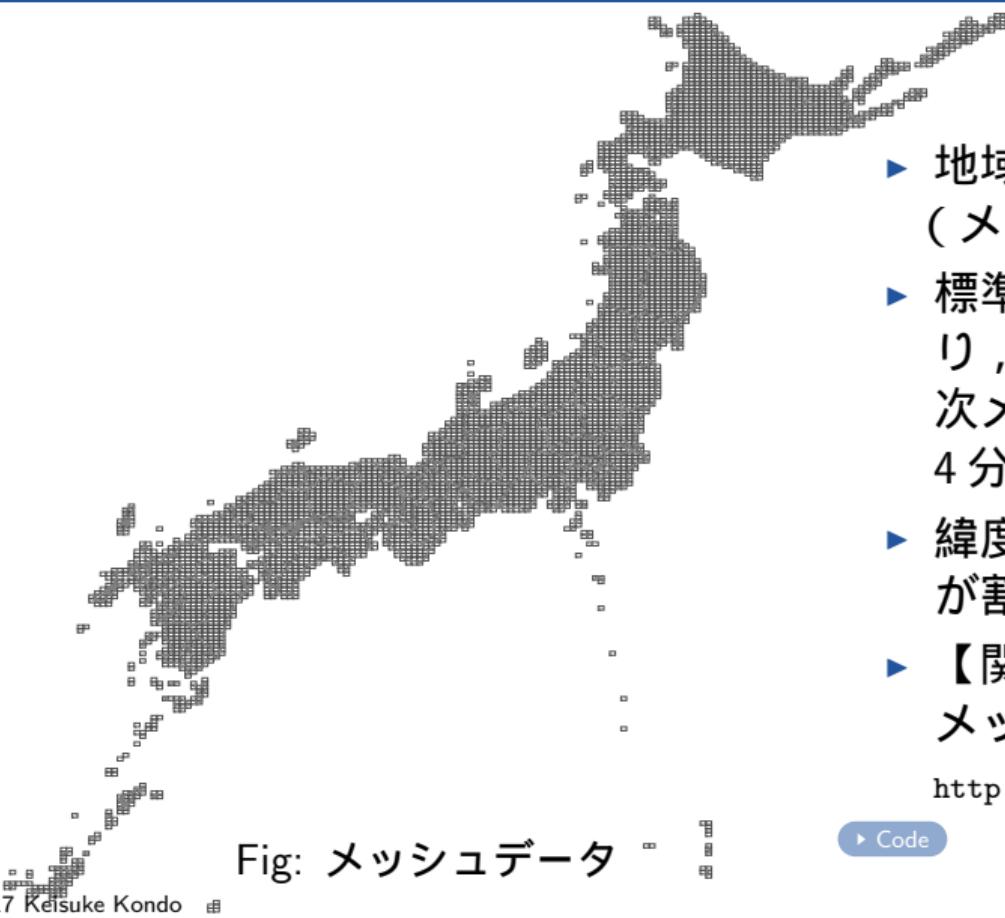


Fig: メッシュデータ

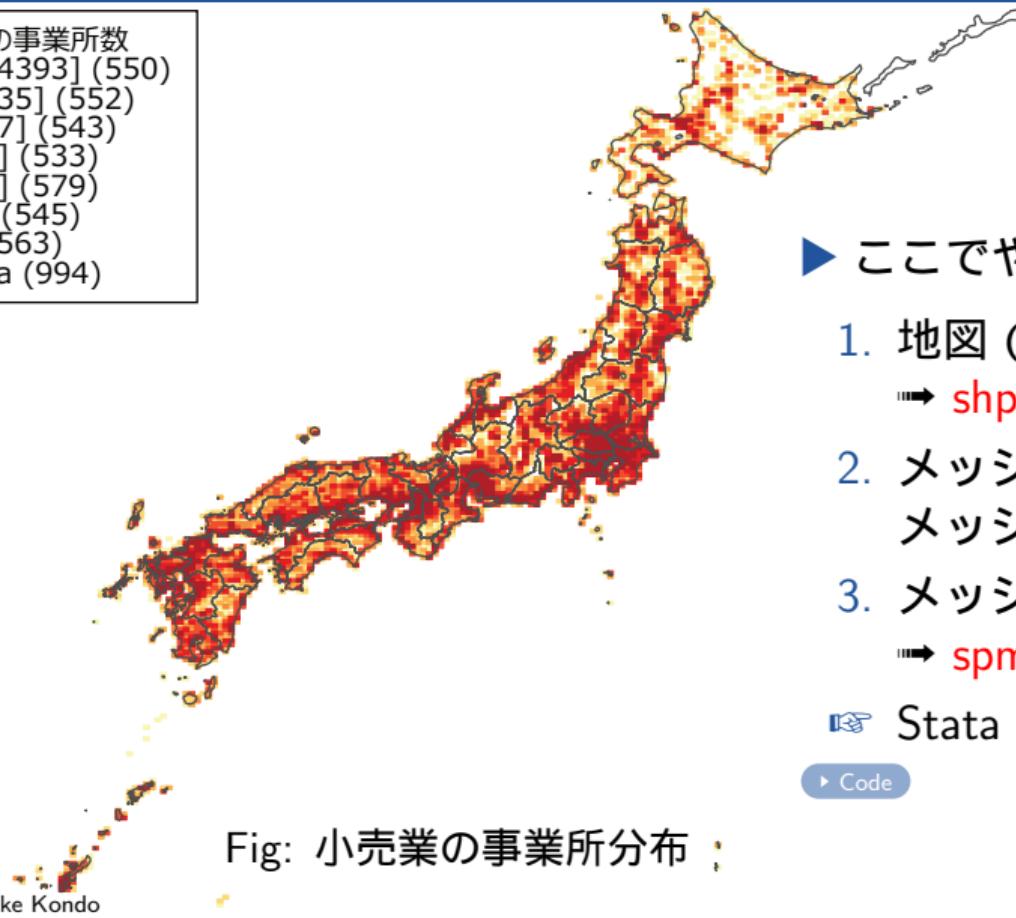
- ▶ 地域を緯度経度に基づいて網目（メッシュ）に区切った単位
- ▶ 標準地域メッシュが規定されており，1次メッシュ，2次メッシュ，3次メッシュ，2分の1地域メッシュ，4分の1地域メッシュ等がある
- ▶ 緯度経度に基づいてメッシュコードが割り振られている
- ▶ 【関連文献】総務省統計局「地域メッシュ統計について」

[http://www.stat.go.jp/data/mesh/m\\_tuite.htm](http://www.stat.go.jp/data/mesh/m_tuite.htm)

▶ Code

# 1. 地理情報システム (GIS) と Stata

小売業の事業所数	
(335,14393]	(550)
(117,335]	(552)
(51,117]	(543)
(25,51]	(533)
(10,25]	(579)
{3,10]	(545)
[0,3]	(563)
No data	(994)



▶ ここでやっていること

1. 地図 (Shape ファイル) の用意  
→ `shp2dta` で Stata 形式に変換
2. メッシュコードを用いて商業統計  
メッシュデータとマージ
3. メッシュデータを地図上に表示  
→ `spmap` で地図表示が可能  
☞ Stata 15 では , `spshape2dta`, `grmap`

▶ Code

Fig: 小売業の事業所分布 :

データ出所 ) 経済産業省「商業統計」

# 1. 地理情報システム (GIS) と Stata

## ▶ Stata の統計解析と GIS が同時に使用できる

### Stata 上でシームレスな作業が可能

- ▶ 空間データの視覚化だけでなく、統計解析を行うのが目的
- ▶ Stata の分析結果を地図上に表示できる  
(ただし、Shape ファイルが利用可能であることが前提)

### GIS 専門ソフトウェアのメリット

- ▶ 地図データの加工、強力な描画機能、より詳細な作業が可能

## 2. 空間計量分析とは？

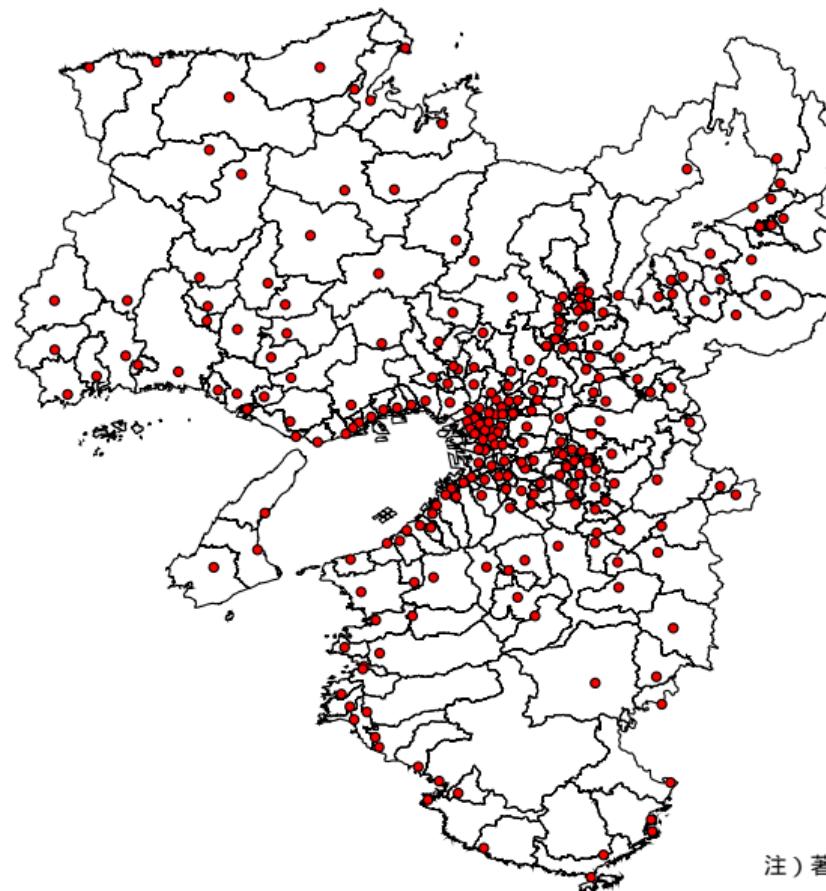
- ▶ 空間上で生じる相互作用を考慮した計量分析

一般的には  
「地理空間」  
(拡張可能)

波及効果  
シナジー効果  
ピア効果 etc.

## 2. 空間計量分析とは？

☞ 市区町村の例



注) 著者作成。上図では市区役所、町村役場が基準点。

## 2. 空間計量分析とは？

☞ 数学的に表すと、

$R$  は地域のインデックス

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1R} \\ d_{21} & d_{22} & \cdots & d_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ d_{R1} & d_{R2} & \cdots & d_{RR} \end{pmatrix}$$

☞ 各地域ペア  $(i, j)$  の地理的距離を 行列として表すことができる

## 2. 空間計量分析とは？

- ▶ **空間重み行列 (Spatial Weight Matrix) を導入**

分析に応じて作成方法が異なるので注意

## 2. 空間計量分析とは？

☞ 空間重み行列の一例

$$W = \begin{pmatrix} 0 & w_{12} & \cdots & w_{1R} \\ w_{21} & 0 & \cdots & w_{2R} \\ \vdots & \vdots & \ddots & \vdots \\ w_{R1} & w_{R2} & \cdots & 0 \end{pmatrix}$$

→ 総和が 1  
→ 総和が 1  
→ 総和が 1

距離の近い地域ほどより大きなウェイト (e.g., 距離の逆数)

☞ 空間計量経済学では行和を 1 に基準化することが多い

## 2. 空間計量分析とは？

- ☞ 行和が 1 になるように基準化するメリット ( $R = 3$  の場合)

$$\underbrace{Wx}_{\text{空間ラグ}} = \begin{matrix} \text{京都} \rightarrow \\ \text{大阪} \rightarrow \\ \text{兵庫} \rightarrow \end{matrix} \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

↓ 京都      ↓ 大阪      ↓ 兵庫

- ☞ 空間重み行列  $W$  ( $3 \times 3$  行列) と変数  $x$  ( $3 \times 1$  ベクトル) の積

## 2. 空間計量分析とは？

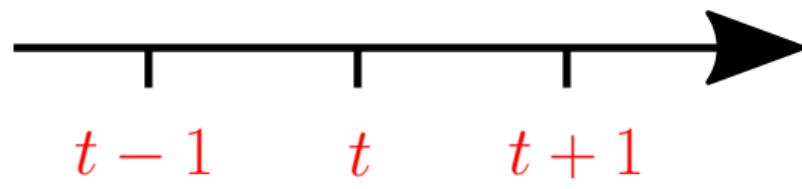
- 自身から見た周辺地域の加重平均として解釈が可能

$$\underbrace{Wx}_{\text{空間ラグ}} = \begin{pmatrix} w_{12}x_2 + w_{13}x_3 \\ \uparrow \text{大阪と兵庫の加重平均} \\ w_{21}x_1 + w_{23}x_3 \\ \uparrow \text{京都と兵庫の加重平均} \\ w_{31}x_1 + w_{32}x_2 \\ \uparrow \text{京都と大阪の加重平均} \end{pmatrix}$$

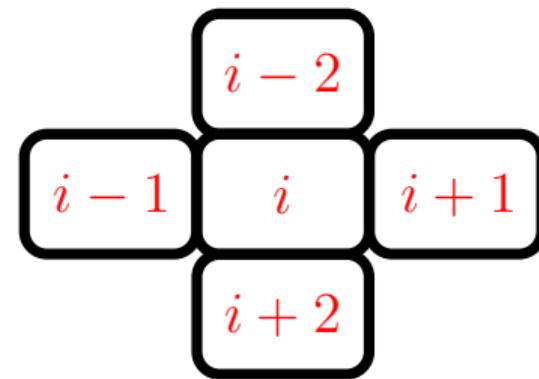
← 京都  
← 大阪  
← 兵庫

- 空間ラグ変数 (spatial lag)  $\xleftrightarrow{\text{比較}}$  時系列のラグ変数

## 2. 空間計量分析とは？



(a) 時系列のラグ



(b) 空間ラグ

Fig: 空間ラグのイメージ

注)著者作成

### 3. プログラミング言語 Mata とは？

- ▶ Mata を用いて **Stata** に実装させることが可能
  - ▶ 行列プログラミング言語
  - ▶ 閃いたアイデアを自由に Stata に取り込むことができる

### 3. プログラミング言語 Mata とは？

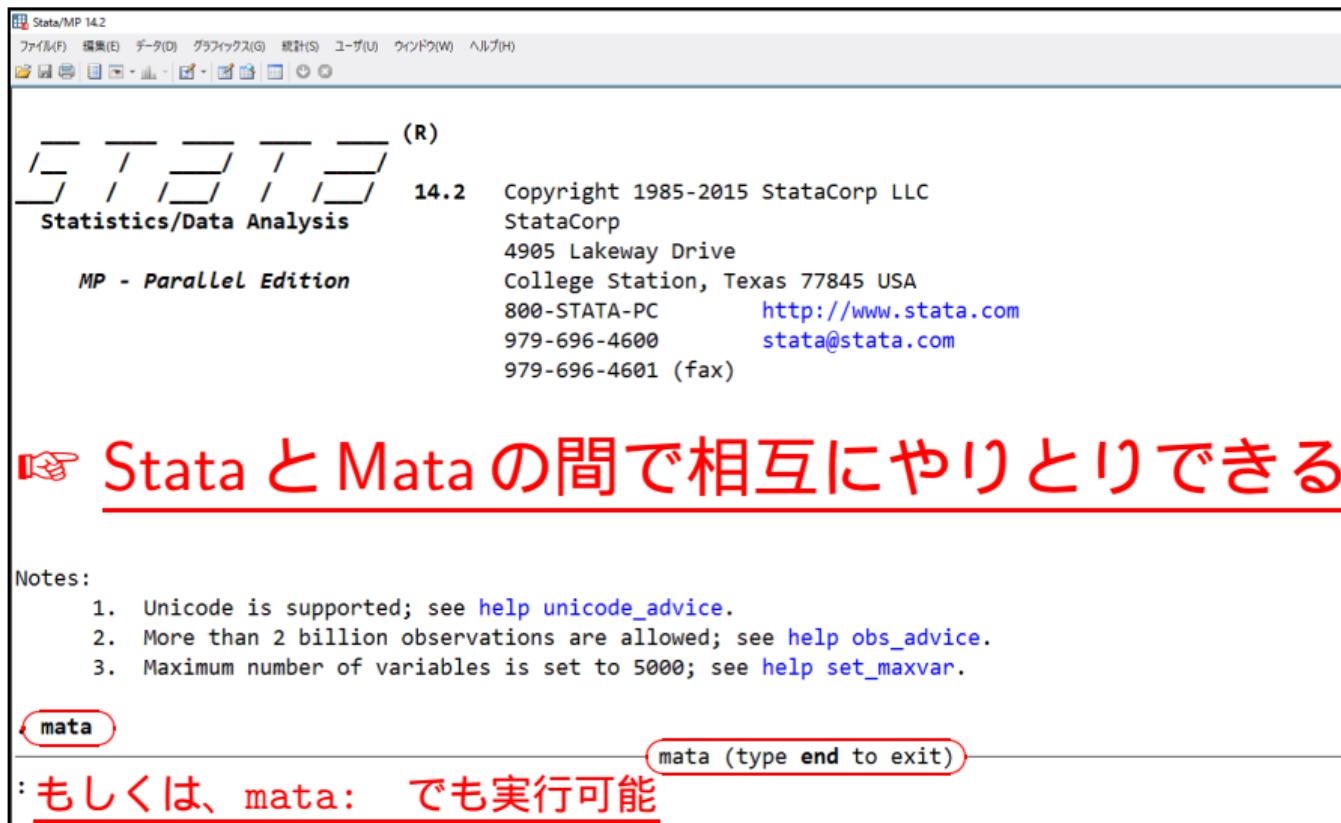


Fig: Stataにおいて Mata の起動

### 3. プログラミング言語 Mata とは？

Stata の緯度経度から空間ラグを計算

NEIG	CRIME	INC	HOVAL	x_cntrd	y_cntrd
1	18.801754	21.232	44.567001	-83.01151	40.01925
2	32.38776	4.477	33.200001	-83.00604	40.00999
3	38.425858	11.337	37.125	-83.00452	40.00074
4	.178269	8.438	75	-83.0207	39.99759
5	15.72598	19.531	80.467003	-82.99041	40.02917
6	30.626781	15.956	26.35	-82.98248	40.01301
7	50.73151	11.252	23.225	-82.98262	39.99761
8	26.066658	16.028999	28.75	-82.95131	40.00286
9	48.585487	9.873	18	-82.98406	39.98129
10	34.000835	13.598	96.400002	-82.92435	39.98895
11	36.868774	9.798	41.75	-82.91728	39.97852
12	20.048504	21.155001	47.733002	-82.91103	39.96805
13	19.145592	18.941999	40.299999	-82.90689	39.9549
14	18.905146	22.207001	42.099998	-82.89859	39.94416
15	27.822861	18.950001	42.5	-82.89921	39.93035
16	16.241299	29.833	61.950001	-82.91848	39.94498
17	.223797	31.07	81.266998	-82.93169	39.96703
18	30.515917	17.586	52.599998	-82.95589	39.98721
19	33.705048	11.709	30.450001	-82.95475	39.97309
20	40.969742	8.085	20.299999	-82.97211	39.97165
21	52.79443	10.822	34.099998	-82.95117	39.96388
22	41.968163	9.918	23.6	-82.94911	39.9572
23	39.175053	12.814	27	-82.95406	39.95043
24	53.710938	11.107	22.700001	-82.97378	39.94947
25	25.962263	16.961	33.5	-82.95514	39.94156
26	22.541491	18.796	35.799999	-82.96155	39.9305
27	26.645266	11.813	26.799999	-82.97194	39.93539
28	20.028489	14.135	27.722	-82.98660	39.93412

① Mata にデータを取り込む



```
. mata: lon, lat
```

	1	2
1	-83.01151	40.01925
2	-83.00604	40.00999
3	-83.00452	40.00074
4	-83.0207	39.99759
5	-82.99041	40.02917
6	-82.98248	40.01301
7	-82.98262	39.99761
8	-82.95131	40.00286
9	-82.98406	39.98129
10	-82.92435	39.98895

② Mata で距離行列を計算する



```
. mata: mD
```

[symmetric]

	1	2	3	4	5
1					
2	1.129271598	0			
3	2.140154396	1.035239354	0		
4	2.529783106	1.86082693	1.425270877	0	
5	2.111269638	2.513133044	3.378767049	4.356971208	0
6	2.573425758	2.039299664	2.323282429	3.685364889	1.917782061
7	3.443565965	2.426685956	1.902166219	3.251912647	3.566817672
8	5.452547152	4.739752065	4.549781144	5.954282595	4.435899773
9	4.822836279	3.698434588	2.778006809	3.614995999	5.34392125
10	8.167583088	7.356667267	6.970569962	8.284239871	7.194257774

## 4. Stataによる空間計量分析の例

- ▶ 空間ラグ変数を計算する **spgen** コマンドの開発

## 4. Stata による空間計量分析の例

### ▶ Mata で記述しているコアな部分

1. Stata データセットから空間重み行列  $W$  を作成する
2. 空間ラグ変数  $Wx$  を計算する
3. Stata データセットに計算結果を格納する

## 4. Stataによる空間計量分析の例

### ❖ コマンドの例

```
spgen var1, lat(y) lon(x) swm(pow 4) dist(.) dunit(km)
      緯度    経度    空間重み    距離の閾値    距離の単位
          行列の型
```



❖ var1 の空間ラグ変数をデータに保存する

## 4. Stataによる空間計量分析の例

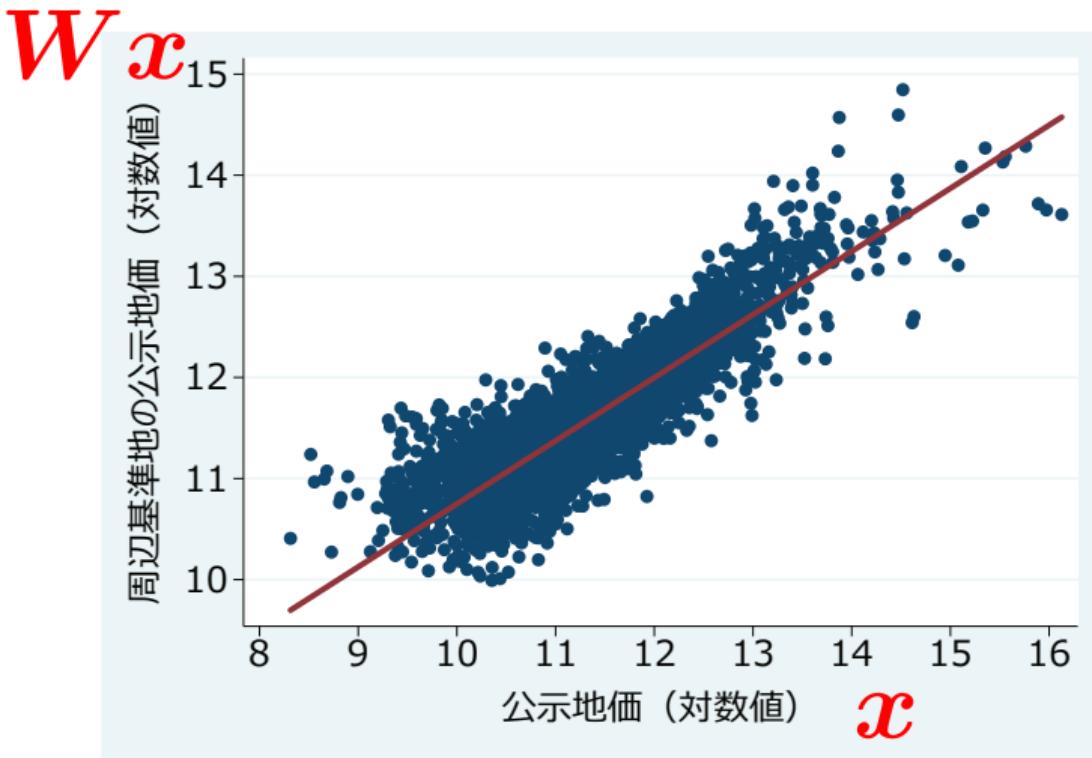
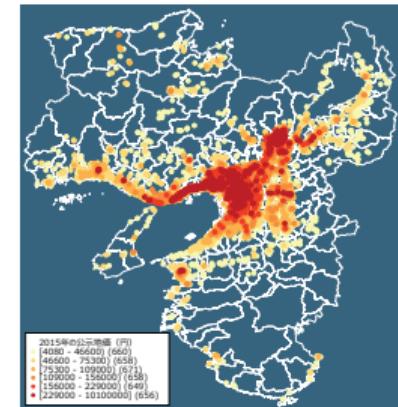


Fig: 空間相関



自身の周辺は同様の傾向を示す



global spatial autocorrelation  
(Moran's I)

注) 国土交通省「国土数値情報」の公示地価データもとに著者作成

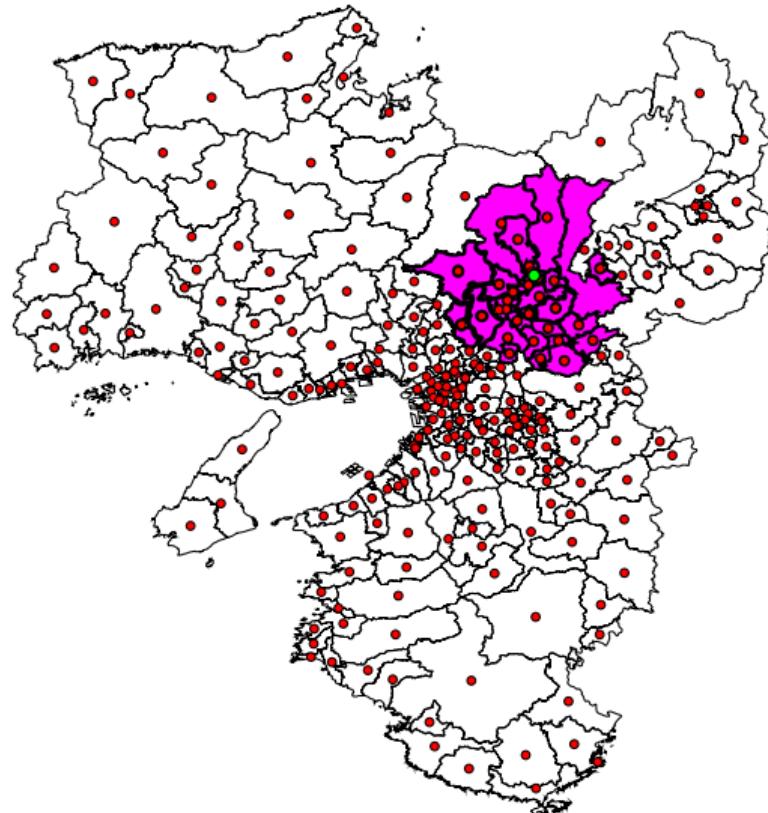
## 4. Stataによる空間計量分析の例

- ▶ ホットスポット分析を行う **getisord** コマンドの開発

## 4. Stataによる空間計量分析の例

- ▶ ホットスポット分析 (Getis-Ord  $G_i^*(d)$ ) のアイデア
  1. 自身とその周辺地域の局所和が全体の総和に占める割合
  2. 空間的にランダムに分布する場合を帰無仮説として検定

#### 4. Stata による空間計量分析の例



$$G_i^*(d) = \frac{\text{地域 } i \text{ とその周辺の局所和}}{\text{全体の総和}}$$

$\uparrow$   
 $d \text{ km} \text{ 内の局所的範囲}$

Fig: ホットスポット分析の直感的解釈

## 4. Stataによる空間計量分析の例

### ☞ コマンドの例 —

```
getisord var1, lat(y) lon(x) swm(bin) dist(30) dunit(km)
          緯度   経度   空間重み  距離の閾値  距離の単位
                           行列の型
```



☞ var1 の Getis-Ord  $G_i^*(d)$  統計量をデータに保存する

## 4. Stata による空間計量分析の例

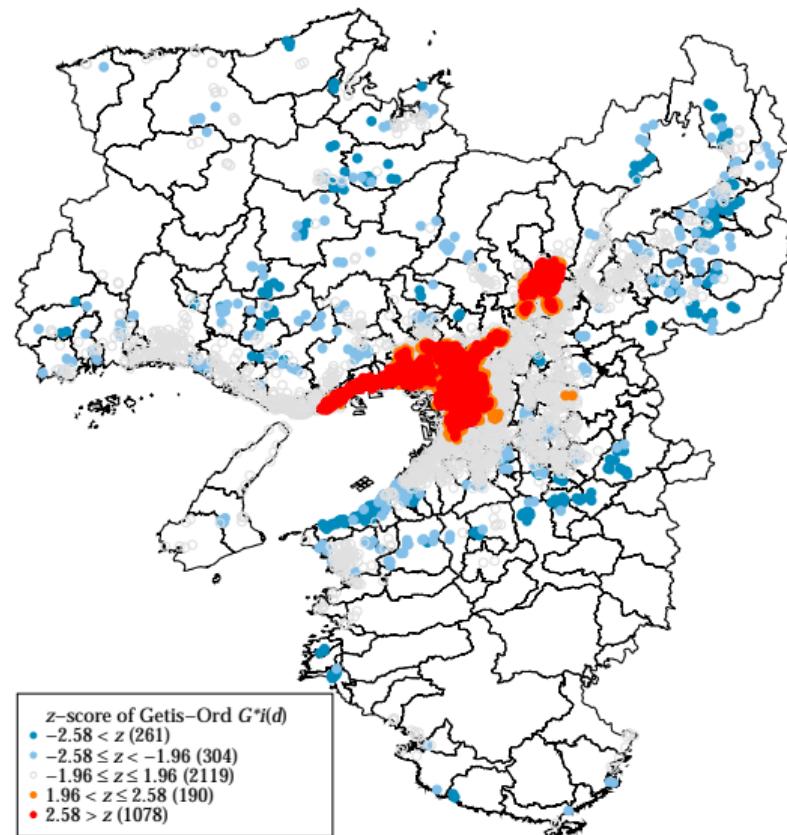


Fig: ホットスポット分析

注)著者作成

## 4. Stata による空間計量分析の例

### ▶ 空間計量経済学のコマンド

- ▶ Stata 15 から公式にサポート (e.g., spgenerate, spmatrix, spregress)
- ▶ Stata 14 以前は User-Written コマンド (e.g., spgen, spmat, spreg)

## 4. Stataによる空間計量分析の例

- ▶ 空間上で生じる相互作用を考慮した計量分析

$$y = X\beta + u$$
$$Wy \quad WX \quad Wu$$

- ☞ 3つのタイプの空間ラグ変数を回帰分析に取り込む

## 4. Stataによる空間計量分析の例

### ☞ 空間計量経済モデルの例 1

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

Stata 15における推定コマンドの一例（地図データがある場合）

- . spshape2dta "myfile.shp"
- . use "myfile.dta"
- . spset
- . spmatrix create contiguity W, normalize(row)
- . spregress y x1 x2, gs2sls dvarlag(W)

## 4. Stata による空間計量分析の例

### ☞ 空間計量経済モデルの例 2

$$y = X\beta + \mathbf{W}X\theta + u$$

Stata 15 における推定コマンドの一例（地図データがある場合）

- . spshape2dta "myfile.shp"
- . use "myfile.dta"
- . spset
- . spmatrix create contiguity W, normalize(row)
- . spregress y x1 x2, gs2sls ivarlag(W: x1 x2)

## 4. Stata による空間計量分析の例

### ☞ 空間計量経済モデルの例 3

$$y = X\beta + u, \quad u = \lambda \mathbf{W}u + \varepsilon$$

Stata 15 における推定コマンドの一例（地図データがある場合）

- . spshape2dta "myfile.shp"
- . use "myfile.dta"
- . spset
- . spmatrix create contiguity W, normalize(row)
- . spregress y x1 x2, gs2sls **errorlag(W)**

## 4. Stataによる空間計量分析の例

### ▶ Spコマンドの概要（地図データがある場合）

1. `spshape2dta` で Shape 形式 “myfile.shp” を変換すると，“myfile.dta”（データファイル）と “myfile\_shp.dta”（地図ファイル）の 2 つのファイルが作成される。
2. Stata で “myfile.dta”（データファイル）を読み込む。
3. `spset` でデータファイルと地図ファイルの共通 ID と地理情報を設定する。
4. `spmatrix` で地理情報を用いて空間重み行列を作成する（行和標準化）。
5. `spregress` で空間重み行列を取り込んだ空間計量モデルを推定する。  
★ `spregress` で ml（最尤法）と gs2sls（操作変数法）の 2 種類の推定方法が可能。  
★ `grmap` で地図上にデータを描画可能。  
★ `spivregress` で内生変数を含んだ空間計量モデルが推定可能。  
★ `spxtregress` でパネルデータを用いた空間計量モデルが推定可能。

## 4. Stata による空間計量分析の例

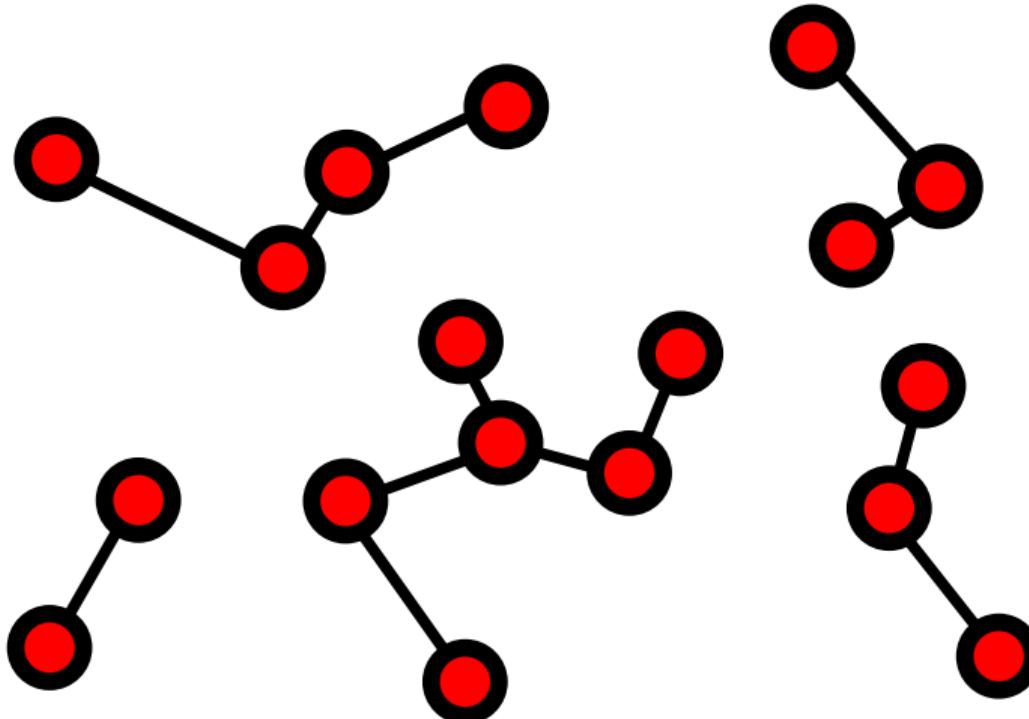
### ▶ **spgen** コマンドの利点 (公式コマンドとの比較)

- ▶ より直感的な操作が可能  
( 空間重み行列  $W$  と変数  $x$  の双方で順序を事前に揃える必要があった )  
( ただし Stata 15 の公式コマンドからは spset を導入し改善 )
- ▶ 空間重み行列の様々な形式に対応  
( 事前に用意された 2 種類の SWM 以外は Mata の操作が必要 )
- ▶ 空間重み行列のサイズが大きくても計算が可能  
( 行列サイズが大きくなると計算が困難になる )  
( Stata 15 の公式コマンドでは観測数に応じてスペック上の限界がある )

## 4. Stataによる空間計量分析の例

- ▶ 個人や企業の地理的ネットワークに拡張できる
  - ▶ 空間計量経済学では観測単位が地点や地域が対象
  - ▶ 個人や企業の地理的近接性はどのような影響があるのか
  - ▶ ジオコーディングにより地理座標を付与できるようになった

## 4. Stataによる空間計量分析の例



例えば、

- ▶ 近隣店舗との価格競争
- ▶ 近隣企業への知識波及

さらに、

- ▶ 「空間」は必ずしも「地理」だけではない
- ▶ ソーシャル・ネットワーク、企業間取引ネットワークとも方法論は密接に関連している

Fig: 企業の地理的ネットワークの例（最近傍でつながったネットワーク）

注)著者作成

- ▶ **Stata** における地図表示に関するコマンド
  - ▶ shp2dta/spshape2dta ( Shape 形式を Stata の DTA 形式に変換 )
  - ▶ spmap/grmap ( Stata で地図表示 . grmap は spmap から公式に採用 )
- ▶ **Stata** における空間計量分析に関するコマンド
  - ▶ spmat/spmatrix ( 空間重み行列を作成・格納 )
  - ▶ spreg/spregress ( 空間計量経済モデルの推定 )
  - ▶ spgen/spgenerate ( 空間ラグ変数の作成 )
  - ▶ geodist/spdistance ( 2 地点間の大圏距離を計算 )

右が Stata 15 で公式にサポートされているコマンド  
左が同等の機能を持つ User-Written コマンド

## ▶ GIS の関連情報

- ▶ 地理情報分析支援システム MANDARA  
(埼玉大学谷先生が開発されているフリー GIS ソフト)
- ▶ QGIS (フリーのオープンソース GIS ソフト)
- ▶ ArcGIS (ESRI 社が開発を行っている専門的な GIS ソフト)
- ▶ 国土交通省「国土数値情報」(地理データが利用可能)
- ▶ Center for Spatial Data Science, University of Chicago (地理データが利用可能)  
(URL: <http://spatial.uchicago.edu/>)

## ▶ Mata の関連情報

- ▶ Stata Blog: Programming (<http://blog.stata.com/category/programming/>)  
( Drukker 先生の “Programming an estimation command in Stata” シリーズの中で Mata の具体例について解説されている )
- ▶ Mata Reference Manual ( Stata 付属のマニュアル )

ご清聴ありがとうございました。  
どのようなコメントでも頂ければ幸いです。

Email: kondo-keisuke@rieti.go.jp

URL: <https://sites.google.com/site/keisukekondokk/>

# 【補足】1. 地理情報システム (GIS) と Stata

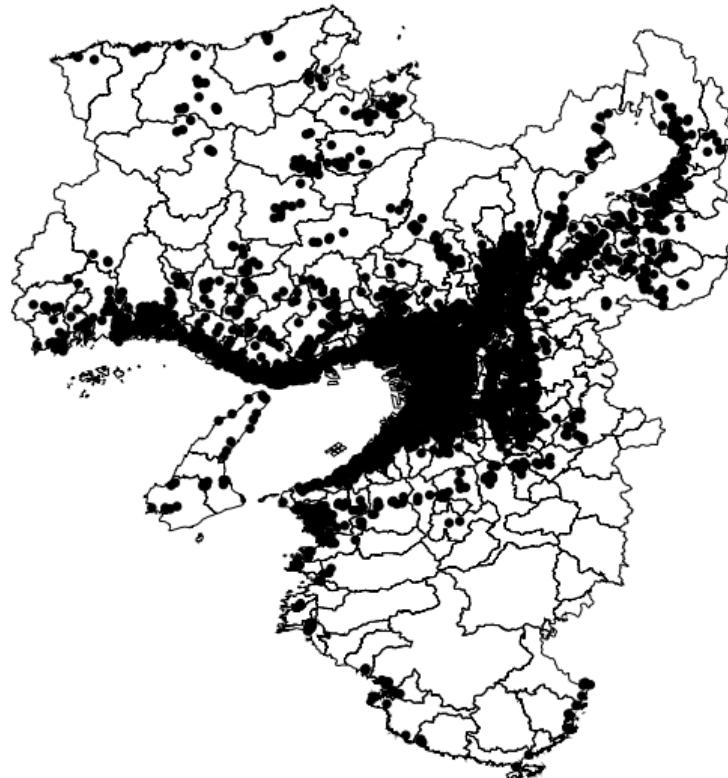


Fig: 関西圏の地価公示の調査地点

▶ Stata コード (Stata 14 以前は spmap )

```
grmap using "japan_muni_shp.dta" if /*  
*/ id_pref >= 25 & id_pref <= 30, /*  
*/ point(data("landprice_kansai-2015d.dta") /*  
*/ x(lon) y(lat) size(small)) /*  
*/ legend(size(*1.5))
```

◀ Return

- 1 id\_pref の指定で関西圏のみ表示
- 2 landprice\_kansai-2015d.dta は地価公示データ  
(緯度 lat と経度 lon を指定する)

データ出所) 国土数値情報、地価公示

# 【補足】1. 地理情報システム (GIS) と Stata

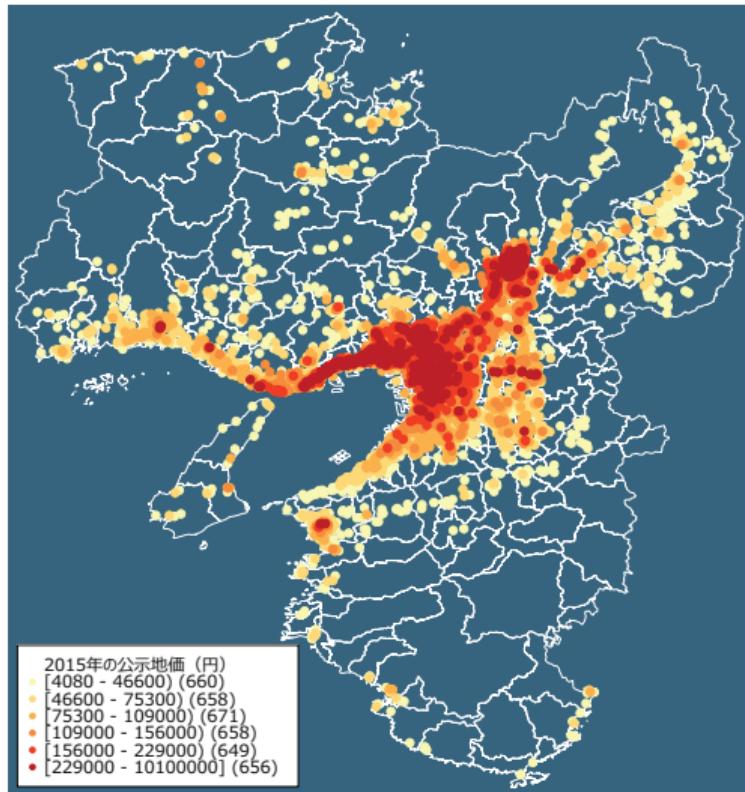


Fig: 関西圏の地価公示

▶ Stata コード (Stata 14 以前は spmap )

```
grmap using "japan_muni_shp.dta" if /*  
*/ id_pref >= 25 & id_pref <= 30, /*  
*/ ocolor(white) /*  
*/ point(data("landprice_kansai-2015d.dta") /*  
*/ x(lon) y(lat) by(ql_lp2015) fcolor(YlOrRd) /*  
*/ size(small) legenda(on) /*  
*/ legtitle("2015 年の公示地価 (円)") legc) /*  
*/ legend(size(*1.1) region(lcolor(black)) /*  
*/ fcolor(white)) position(8)) /*  
*/ plotregion(icolor(edkblue) color(edkblue)) /*  
*/ graphregion(icolor(edkblue) color(edkblue))
```

◀ Return

- 1 id\_pref の指定で関西圏のみ表示
- 2 landprice\_kansai-2015d.dta は地価公示データ  
(緯度 lat と経度 lon を指定する)
- 3 ql\_lp2015 は quantile 区分の変数 (事前に作成)

データ出所) 国土数値情報、地価公示

# 【補足】1. 地理情報システム (GIS) と Stata

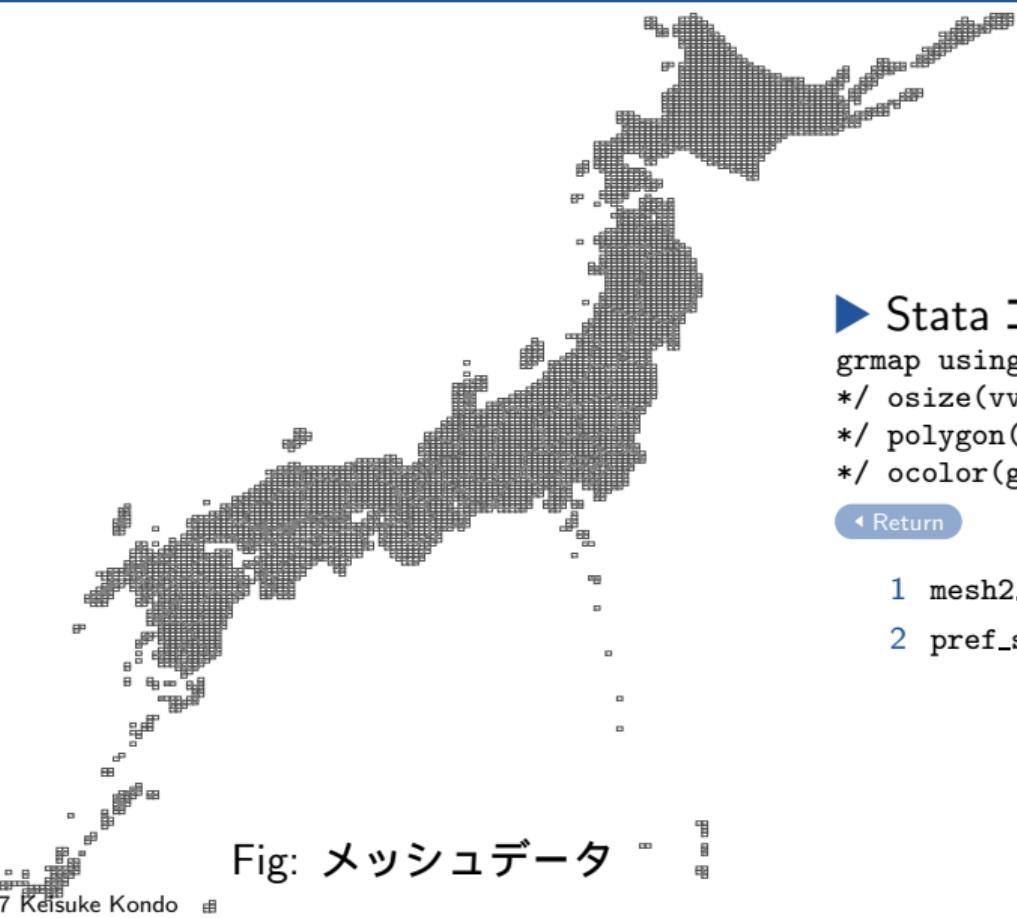


Fig: メッシュデータ

▶ Stata コード (Stata 14 以前は spmap )

```
grmap using "mesh2_shp.dta", /*  
/* osize(vvthin) ocolor(gs5) fcolor(None) /*  
/* polygon(data("pref_shp.dta") osize(vvthin) /*  
/* ocolor(gs8) fcolor(None))
```

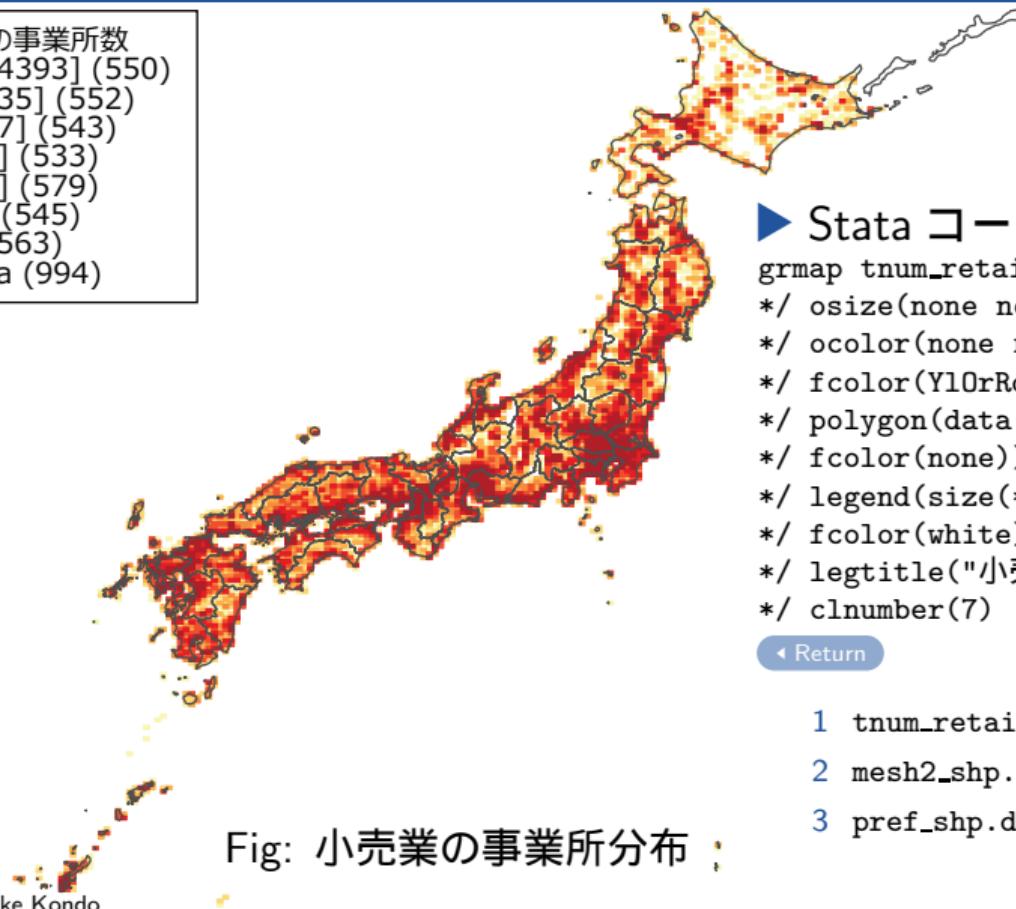
◀ Return

1 mesh2\_shp.dta はメッシュデータの地図ファイル

2 pref\_shp.dta は都道府県の地図ファイル

# 【補足】1. 地理情報システム (GIS) と Stata

小売業の事業所数	
(335,14393]	(550)
(117,335]	(552)
(51,117]	(543)
(25,51]	(533)
(10,25]	(579)
{3,10]	(545)
[0,3]	(563)
No data	(994)



▶ Stata コード (Stata 14 以前は spmap )

```
grmap tnum_retail using "mesh2_shp.dta", /*  
*/ osize(none none none none none none) /*  
*/ ocolor(none none none none none none) /*  
*/ fcolor(YlOrRd) ndocolor(none) ndfcolor(none) /*  
*/ polygon(data("pref_shp.dta") ocolor(gs5) /*  
*/ fcolor(none)) /*  
*/ legend(size(*1.4) region(lcolor(black) /*  
*/ fcolor(white)) position(11)) /*  
*/ legtitle("小売業の事業所数") legcount /*  
*/ clnumber(7)
```

[◀ Return](#)

- 1 tnum\_retail はメッシュ内の小売業の事業所数の変数
- 2 mesh2\_shp.dta はメッシュデータの地図ファイル
- 3 pref\_shp.dta は都道府県の地図ファイル

データ出所) 経済産業省「商業統計」