

連動する統計グラフィックスによる 多変量地理情報データの視覚化

小林 郁典¹・山本 由和¹・中野 純司^{2,3}・Jung Jin Lee⁴

(受付 2006 年 8 月 2 日; 改訂 2007 年 3 月 2 日)

要 旨

近年の情報技術の発達により、地理情報データを誰もが比較的容易に入手することができるようになってきている。地理情報データは地球表面の位置に関する情報をもつデータであるため、その視覚化には地図を利用したグラフィックスが有用である。しかしながら、これまで提案されてきた地図を利用したグラフィックスでは一度に表示できる変数の数に制限があり、多変量データをうまく視覚化することができないという問題がある。

そこで、本論文では、多変量地理情報データを視覚化するひとつの方法として、条件付きコロプレスマップと平行座標プロットなどの連動が、多変量地理情報データの探索的な解析を支援する道具として有用であることを示す。

ここで提案する統計グラフィックスは、われわれが開発を進めている統計解析システム Jasp で実装された。本論文では、その機能と使い方を実際のデータを解析しながら説明する。

キーワード：コロプレスマップ, 条件付きコロプレスマップ, 平行座標プロット, Jasp.

1. はじめに

地理情報データは、居住地域、住所、事故現場など、地球表面の位置に関する情報をもつ。近年、このようなデータは、行政や企業のマーケティング部門での空間分析や地域間比較などのために盛んに利用されている。また、そのような解析を円滑に行うための解析ツールの開発が進み、地理情報データを利用した有益な研究成果が報告されるようになってきている(古田 他, 2005)。

地理情報データの解析を行う場合、位置関係を考慮しながら解析を進めたり、位置関係がわかるように解析結果を表示したりすることが要求される。このためには、地図を基本としたグラフィックスが有用である。例えば、地図上の各領域内に、その面積で量を表したドットを描画した地図(ドットマップ)や、円グラフを描画した地図(円グラフマップ)が有名である(Chang, 2006)。

これらのグラフィックスでは、領域別の予測結果や分類結果などの 1 つの変数を表示するこ

¹ 徳島文理大学 工学部：〒769-2193 香川県さぬき市志度 1314-1

² 統計数理研究所：〒106-8569 東京都港区南麻布 4-6-7

³ 総合研究大学院大学 複合科学研究科統計科学専攻：〒106-8569 東京都港区南麻布 4-6-7

⁴ Department of Statistics, Soong Sil University, 1-1 Sangdo-5 Dong, Dongjak-Gu, Seoul 156-743, Korea

とは有効であるが、表示する変数が多くなると視覚的な理解が困難になる。つまり、変数間の関係をこれらのグラフィックスから推測したり、データを見ながら外れ値の疑いのある観測値を抽出したりすることがむずかしい。もちろん、位置関係を考慮しなくてよければ散布図や平行座標プロットなどの多変量解析用の汎用的な統計グラフィックスを利用してデータを観察することはできる。しかしながら、位置関係を考慮しなければ十分な分析結果が得られないことは明らかである。

そこで、本研究では、領域の位置関係を示す地図を基本としたグラフィックスと多変量のデータを視覚化することができる統計グラフィックスが連動する解析環境を構築した。特に、地図を基本としたグラフィックスとしては領域で観測されたデータをカテゴリごとにその領域の色によって表現する地形図であるコロプレスマップ(choropleth map)とそれに条件を付加した条件付きコロプレスマップ(conditioned choropleth map)を、統計グラフィックスとしては多変量データを一度に視覚化することのできる平行座標プロットを利用する。

われわれは、多変量地理情報データ用のグラフィックスを統計解析システム Jasp (Nakano et al., 2000) 上で実装した。Jasp は、Java 言語を開発言語としたオープンソースの汎用統計解析システムであり、基本的な統計関数、代表的なデータマイニング手法、グラフ機能をもち、さらに、ネットワーク環境への対応や利用者のレベルに応じた拡張性を持っている。

本論文の2章では、コロプレスマップ、条件付きコロプレスマップ、平行座標プロットについて個別に説明し、3章では2章で説明したグラフィックスを連動させることによる利点を解析事例によって示す。4章では、本研究と関連する研究活動について言及する。最後の5章はまとめである。

2. 地理情報データ解析のためのグラフィックス

本章では、多変量地理情報データを解析するために利用するグラフィックスについて説明する。

2.1 コロプレスマップ

コロプレスマップとは、1次元のデータを異なる色や模様などによって分類し、それによって対応する領域(行政区域や統計区など)を塗り分けた地図のことである。コロプレスマップを描くためには、領域の境界情報とそれらの塗り方を与える必要がある。

Jasp でコロプレスマップを表示するためには、関数

```
CMap(mapObject, dataObject, vName, col)
```

を利用する。ここで、第1引数の mapObject は地図の境界線を描くために所定の形式の情報を持ったデータオブジェクト、第2引数の dataObject は色を塗るための情報を持ったデータオブジェクト、第3引数の vName は dataObject の中の色指定に利用するカテゴリ変数名、第4引数の col は塗り分ける色情報の配列を指定する。なお、dataObject の中には mapObject が持つ地域コードと同じものを保持する必要がある。また、vName と col は省略することができる。vName が省略された場合には、cData の1列目の変数が vName として設定され、col が省略された場合には、Jasp が自動的に塗る色を決める。

CMap により描かれた地形図上では、マウスクリックにより特定の領域を、マウスドラッグによりドラッグ内の複数の領域を選択状態にすることができる。そして、この選択情報をデータオブジェクトとして抽出して、2次的に利用することが可能である。

ここで CMap の利用例を示す。解析するのは、総務省統計局のサイト「統計でみる都道府県・市区町村(社会・人口統計体系)」(<http://www.stat.go.jp/data/ssds/index.htm>) で公開され

ているデータのうち、徳島県(50市町村別)に関するデータ(変数72個)と、65歳以上の高齢者の1,000人あたりの自殺者数(1997年からの7年間の平均値)である。Jaspに組み込まれているクラスタリング(K-Means法)手法をこれに適用し、その結果をコロプレスマップで表示させることを想定する。なお、クラスタ数として3と4を設定した。クラスタ数が3の場合の作業をJasp言語で記述すると次のようになる。

```
tokushima = read("ssds_tokushima.jdt")
dataObject = tokushima[,"id"]
mapObject = read("mapTokushima.jdt")
km3 = wekaKMeansModel(dataObject, 3)
col3 = clusterData(km3, dataObject)
kMeansResult3 = CMap(mapObject, col3)
kMeansResult3
```

1行目で先ほどのデータを読み込み、次の行で読み込んだデータから地域番号を削除している。そして、3行目で徳島県の地図を描くための境界値のデータを読み込み、4行目でK-Means法によるクラスタリングを行っている。変数col3には、クラスタリング結果の色情報が代入され、最後の2行でクラスタリング結果によって色分けされたコロプレスマップを作成し表示する。クラスタ数を4にするには、プログラム中の4行目以降の3を4に置き換えたもので実現できる。これらを実行したJaspのGUI Windowが図1である。

図1の左側の3つのグループに分類した方は、水色の領域が比較的人口が密集している場所(市部やベッドタウンエリア)を、緑色の領域が水色に次いで人口が密集している場所(平地にある田園地帯)を、青色の領域が人があまり住んでいない場所(過疎部・山間部)を表している。図1の右側の4つのグループに分類した方は、3つに分けた場合の青色の領域をさらに2つのグループに分けたことにほぼなっている。4つに分けた方の青色は、過疎部でも比較的人が住んでいる場所である。また、黄色の領域は、四国山地に位置する可住領域の少ない山間部である。

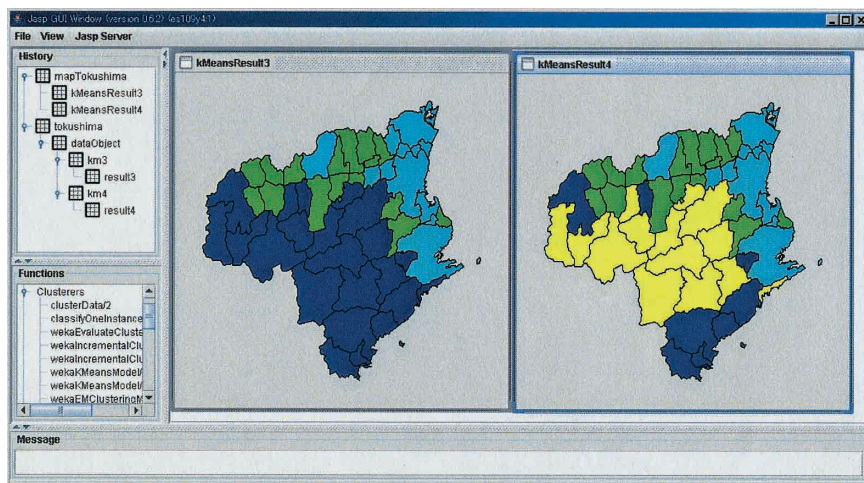


図1. クラスタリング結果で色分けをした徳島県のコロプレスマップ。

クラスタリング結果の妥当性は、解析者が関連する地理的な知識にもとづいて判断しなければならないので、このように、地図上でクラスタリング結果を表示し、そして、見比べることができれば、解析者にとっては判断がしやすくなる。また、Jaspには線形回帰モデル、C4.5、ニューラルネットワークモデルをデータに適用する関数が備わっている(小林 他, 2006)ので、それらを使った解析の結果をコロプレスマップ上に表現するようなことも可能である。

2.2 条件付きコロプレスマップ

一般に多変量データにおいては変数間にある何らかの関係を見たいことが多い。このようなデータでは、ある条件により観測値をいくつかのグループに分けてから解析をすると関係がわかりやすくなることがある。多変量地理情報データに関しても同様であるので、その地理的特性を視覚化するために、われわれは条件付きのコロプレスマップを利用することにした。

条件付きコロプレスマップは、2つの変数の区間を条件として、その条件にあてはまる領域だけに色をつけたコロプレスマップを行列形式に並べて描いたものである。これは、コロプレスマップを描くために採用した変数(目的変数)に対して条件として考える2つの変数(条件変数)がどう関係しているのかを観察したり、目的変数と条件変数に関するモデルを導出するとき利用される(Carr et al., 2000)。条件付きコロプレスマップを使用した解析では、目的変数や条件変数、表示するコロプレスマップの数などをマウス操作などによって対話的に操作できることが重要である。

Jaspで条件付きコロプレスマップを表示するためには、関数

```
CCMaps(mapObject, dataObject)
```

を利用する。ここで、第1引数のmapObjectは地図の境界線を描くための情報を持ったデータオブジェクト、第2引数は目的変数と2つの変数が含まれるデータオブジェクトである。これらはCMap関数の引数と同じものである。

図2は、徳島県の50市町村別のデータを条件付きコロプレスマップで解析している例である。

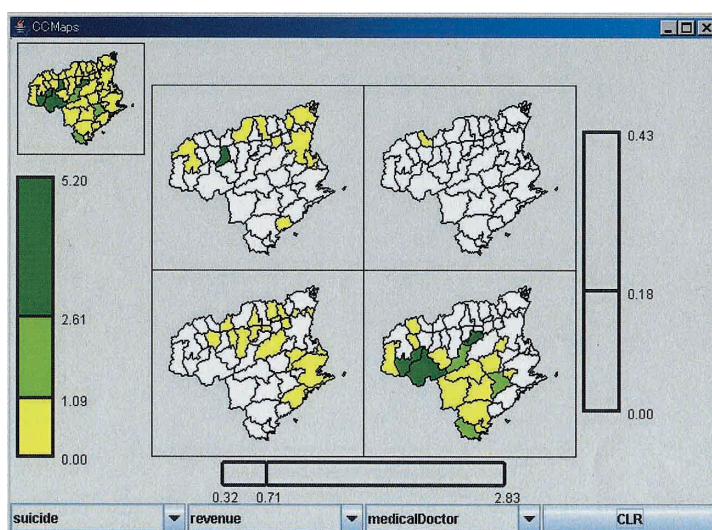


図2. 条件付きコロプレスマップの表示例.

図2の左上の小さな地図が、目的変数を利用して描いた一般のコロプレスマップである。どのように色分けしているのかは、図の左端にあるスライダバーから判断できる。この図の場合、目的変数のためのスライダバーは3つに区分され、それぞれの境界値はスライダバーの横に表示されている数値から読み取ることができる。

図2の中央にある2行2列の格子状に配置されているものが、2つの条件変数によって区分されたコロプレスマップである。2つの条件変数は、下部の横方向と右端の縦方向のスライダバーにそれぞれ割り当てられており、どのような条件が設定されているかはスライダバーの境界値から読み取ることができる。この例では、条件変数は共に2つの階級に分割されている。条件付きコロプレスマップにあるこれら3つのスライダバーでは、マウス操作によって境界値を変更したり、分割数を増減させたりすることができる。

図2の下端にある3つのコンボボックスは、左から目的変数、横方向の条件変数、縦方向の条件変数を指定するためのものである。この例では、目的変数として65歳以上の高齢者の地域人口1,000人あたりの自殺率(変数名は `suicide`、以下これを“高齢者自殺率”とよぶ)が、横方向の条件変数として地域人口あたりの歳入総額(変数名は `revenue`、単位は万円/人、以下これを“歳入総額”とよぶ)が、縦方向の条件変数として地域人口百人あたりの医師の数(変数名は `medicalDoctor`、以下これを“医師数”とよぶ)がそれぞれ設定されている。

まず、図2中の左上のコロプレスマップから、高齢者自殺率の高い地区は山間部に多く見受けられることがわかる。特に、濃い緑色の地区はとりわけ交通事情の悪い山間部にある。次に、格子状に並んでいる徳島県の地図を観察すると、条件として設定した歳入総額と医師数についていくつかの知見を得ることができる。歳入総額が少ない地区は、徳島県の市部やベッドタウンエリアがほとんどであり、多い地区は、過疎部・山間部がほとんどである。また、医師数が多い地区は、市部や県立病院などの総合病院が存在するところやその近辺である。高齢者自殺率が高い地区は、歳入総額が比較的多く、そして医師数が比較的少ないところであることがわかる。

Jasp で実装した条件付きコロプレスマップでは、次に示す方法で領域を選択状態(強調表示状態)にすることができる。(1)地形図上の特定領域をマウスクリックすると、その領域が選択状態になる。(2)3つのスライダバー上の区間をクリックすることで、対応する領域が選択状態になる。ただし、条件変数のための2つのスライダバーでは、対応する行か列(横方向のスライダバーの場合は列、縦方向のスライダバーの場合は行)の中に含まれるすべての色つきの領域が選択状態になる。(3)格子状に表示されている各枠の中の領域が表示されていない場所をクリックすると、その枠内の中の色つきのすべての領域が選択状態になる。

選択状態になった領域は、関数

```
getSelectedRegions(ccmObject)
```

を利用することで、その地区コードの配列を抽出することができる。ここで、第1引数 `ccmObject` は、CCMaps 関数によって生成されるオブジェクトである。この関数は、CMap 関数で生成されるコロプレスマップに対しても適用することができる。この関数を利用すれば、選択状態であった地区のデータだけを抽出し、回帰分析などを適用することができる。

2.3 平行座標プロット

平行座標プロット(Parallel Coordinate Plot, PCP)は、変数ごとの座標を平行に並べ、直行座標では不可能な4次元以上のデータを同時に表示することができるグラフィックスである。観測値の値は、平行座標軸上の点を通る折れ線で表現される(Inselberg, 1985)。平行座標プロットの重要な利点は、ひとつの観測値の多数の変数における値や他の観測値とそれらの相対位置

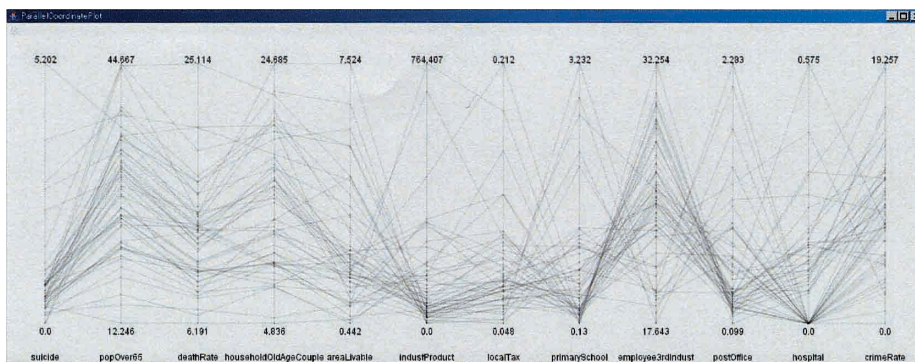


図 3. 平行座標プロットの表示例.

を同時に眺められることである.

Jasp で平行座標プロットを表示させる場合には、関数

```
PCP(dataObject)
```

を利用する. ここで, dataObject は, Jasp が管理する一般的なデータオブジェクトである.

この関数を利用した平行座標プロットの例を図 3 に示す. この図も, 前例と同様に徳島県のデータを表示している. 図中には折れ線が 50 本あり, それぞれが 50 市町村のどれかを表している. また, 変数としては, 左の縦軸から前述の高齢者自殺率, 65 歳以上の人口比率, 死亡率, 65 歳以上の高齢者夫婦だけの世帯, 一人あたりの可住地域面積, 一人あたりの製造品出荷額, 一人あたりの地方税収入額, 住民 1,000 人あたりの小学校数, 第 3 次産業就業者比率, 住民 1,000 人あたりの郵便局数, 住民 1,000 人あたりの一般病院数, 住民 1,000 人あたりの刑法犯発生件数の計 12 変数である. 徳島県のデータには 73 個の変数があり, 平行座標プロットを利用すればすべての変数を一度に表示させることもできるが, ここでは見やすいように 12 変数に限定した. 従って, 図中には 12 本の縦軸が表示されている.

縦軸の上端にはその変数の最大値が, また, 下端にはその変数の最小値が表示される. そして, その最小値の下には変数名が表示される. 平行座標プロットでは, マウス操作によって折れ線を選択状態にしたり, 変数の表示位置を入れ替えたり, 変数の軸の上下を逆転したりすることができる.

3. グラフィックス間の連動

多変量地理情報データを解析するためには, 上述したコロプレスマップ, 条件付きコロプレスマップ, 平行座標プロット, 及びその他の統計グラフィックス(ヒストグラム, ボックスプロット, 散布図行列など)が連動しなければならない.

すなわち, 地図上の領域と各観測値は連携しており, どれかのグラフィックス上で観測値がマウス操作により選択状態になると, それに対応する別のグラフィックスの領域あるいは観測値が自動的に強調表示される. この機能により, ささまざまなグラフィックスを介して地理情報データの特徴を観察したり, 知見を得たりすることができる.

ここで, 前述と同じ徳島県のデータを用いて, 連動する統計グラフィックスによる解析例を示す.

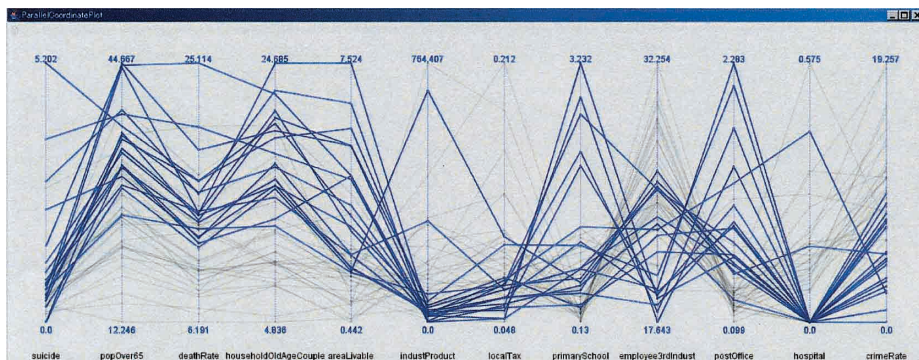


図 4. 複数の観測値が強調表示状態になっている平行座標プロット。

まず、図 2 の横方向の条件変数としては歳入総額が選択されているが、これは、高齢者の自殺が、住む地域の経済的な状況と関係があるという考えに基づいて選択した。また、縦方向の条件としては医師数が選択されているが、これは、高齢者の自殺が住む地域の医療機関の数と関係があるという考えに基づいて選択した。

図 2 の格子状の徳島県地図を見ると、高齢者自殺率の高い地区、つまり深い緑色で塗られたところは、右下の地図に集まっていることがわかる。この地図は、“住民ひとりあたりの歳入総額が 0.71~2.83 万円、医師数が 0.18 以下”に該当する行政区域である。これが徳島県の高齢者自殺率の高い地区に関するひとつの特徴である。

条件付きコロプレスマップは、目的変数と 2 つの条件変数の計 3 つの変数についての情報を示すが、これ以外の変数に関しての情報は見ることができない。そこで、この行政区域の他の変数との関係を観察するために、マウスによって条件付きコロプレスマップの右下の地図の背景をクリックする。そうすると、図 3 の平行座標プロットは、図 4 のように対応する行政区域（この場合 17 個）の折れ線が強調表示される。

強調された折れ線が縦軸上のどこに位置しているのかを観察することで、選択状態にある行政区域と他の変数との関係を知ることができる。例えば、製造品出荷額を表す変数 `industProduct`（左から 6 つ目の縦軸）の軸上では、強調された 17 本の多くの折れ線が最小値の近辺に集まっていることがわかる。これは、該当する地域の多くが、ほとんど工業化されていないことを表している。つまり、歳入総額が比較的多くて医師数が比較的少ない地域は、製造品出荷額は低く、そして、高齢者自殺率も高い傾向にある、ということが言える。

選択されている地域に見受けられるこの他の傾向としては、65 歳以上の人口比率が高い (`popOver65`)、小学校数が多い (`primarySchool`)、郵便局の数が多い (`postOffice`)、ほとんどの地区に一般病院がない (`hospital`)、地方税収入額が少ないので経済的に潤っていない (`localTax`)、第 3 次産業就業者比率が低い傾向にある (`employee3rdIndust`)、治安が比較的良い (`crimeRate`)、のようなことが挙げられる。

次に、高齢者の自殺と関係が見受けられた変数のうち、小学校数に注目してみる。図 5 は、平行座標プロットのマウス操作による選択機能を利用して、小学校数の多い折れ線（上位 5 つ）を選択状態にしたものである。この図から、小学校数の多い地域は、(1) 高齢者自殺率が高い (`suicide`)、(2) 65 歳以上の人口比率が高い (`popOver65`)、(3) 製造品出荷額と地方税収入額が低い (`industProduct` と `localTax`)、(4) 一般病院がない (`hospital`)、ということがわかる。

ただし、平行座標プロットでは小学校の多い地域が徳島県のどのあたりのかがわからない。

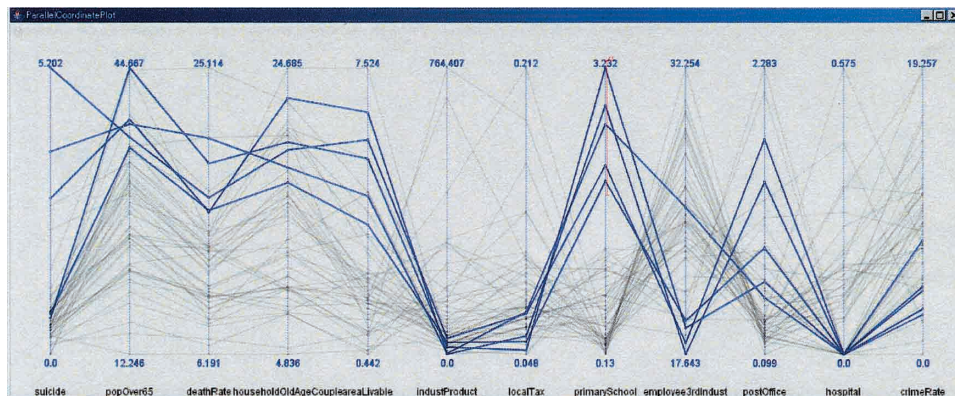


図 5. 小学校数の多い地域(上位 5 つ) を選択した状態の平行座標プロット.

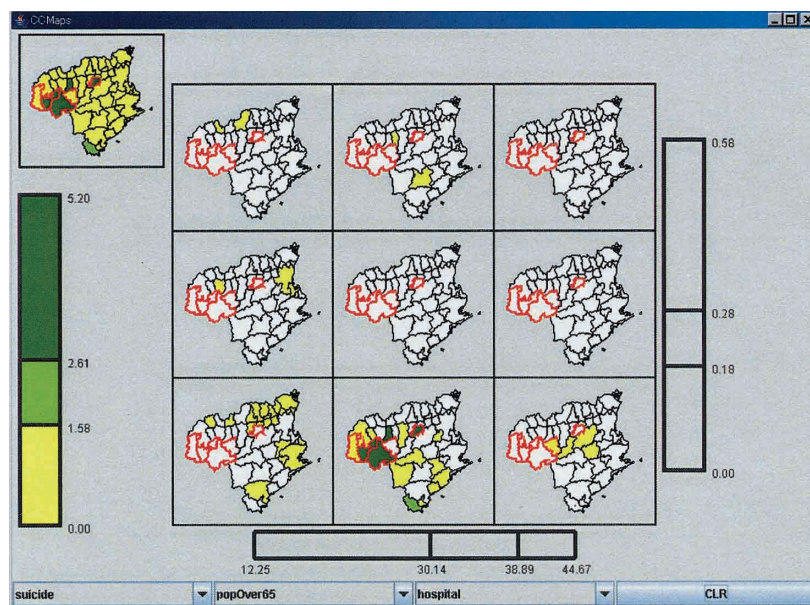


図 6. 図 5 に対応した条件付きコロプレスマップ.

そこで、これを観察するために、コロプレスマップか条件付きコロプレスマップを参照する。図 6 は、図 5 に対応した条件付きコロプレスマップである。図 6 中の境界が赤くなっている地域が、小学校数の多い地区(上位 5 つ)である。これから、強調表示されている地域は、徳島県の山間部に位置していることがわかる。また、高齢者自殺率の高い地域と重なっていることも見てとれる。

小学校数の多い地区は、高齢者が多く、病院がないということなので、図 6 のように、横方向の条件変数として 65 歳以上の人口比率 (popOver65)、縦方向の条件変数として一般病院数

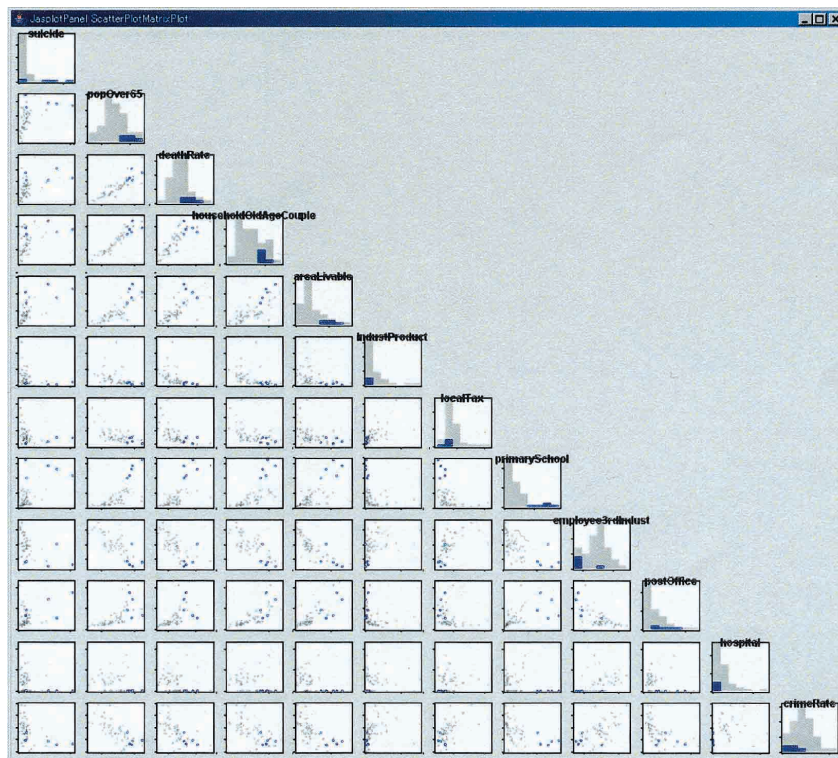


図 7. 図 5 に対応した散布図行列.

(hospital)をそれぞれ設定して、高齢者自殺率との関係を観察する。このとき、2つの条件変数は共に3つの階級にわけている。

図6は、スライダーを調整して、高齢者自殺率の高い地区を3行2列目の地図中に集めた状況を示している。このことから、高齢者自殺率が高い地区は、65歳以上の人口比率が30.14~38.89%の間(最も高い地域ではない)、一般病院数が0.18以下(一般病院が少ない地域)、小学校数が多い(上位5つ)、ということがわかる。

このようにJaspに追加した解析環境は、条件付きコロプレスマップで条件により対象の領域を限定し、その限定された領域の他の変数との特徴や傾向を平行座標プロットで観察したり、逆に、平行座標プロットから着目する変数とデータ全体の傾向を観察しながら、影響のありそうな変数を選択(最大2つ)し、条件付きコロプレスマップでこれらの変数との傾向を観察したりすることが探索的に行いやすい。

さらに、Jaspでは図7に示すような散布図行列を表示させ、各変数のヒストグラムと2変数間の関係(散布図)を観察することもできる。図7は、図5に連動しており、対応した小学校数の多い地区が強調表示されている。小学校数の多い地区は図中で青く強調表示されているので、その地区が他の変数とどのように関係しているのかがわかる。また強調表示を考慮せず、2変数間に線形関係があるかどうかなどを観察することができる。ただし、散布図行列は、あまり多くの変数を用いると表示が小さくなり正しく読み取ることが困難になる。

4. 関連する研究活動

近年、地理情報データを管理し、視覚化する道具として地理情報システム (Geographic Information System, GIS) が注目を集めている。GIS は、位置情報を有するデータを効率的に蓄積、検索、変換して、空間解析や地図出力、さらに意思決定支援を行うコンピュータ・システムである (村山, 2005)。普及している GIS としては、ESRI 社の ArcView が有名である。GIS の定義を鑑みれば、本研究で提案した環境は GIS の実現として位置づけられるであろう。現時点では本研究で述べたような出力地図と統計グラフィックスが連携して操作できるような対話的なインタフェースをもつ GIS は見受けられないが、いずれ普及してくるものと思われる。

一方で、汎用的な統計解析システムにおいても地図に基づいたグラフィックスを利用して地理情報データを表示するための機能の充実が図られている。たとえば、世界的に広く利用されている統計解析システム R (Ihaka and Gentleman, 1996) にも GIS を実現できるようにいくつかの機能の実装が試みられている (Bivand, 1999)。R には多くのグラフィックス機能が実装されているが、それらによって表示されたグラフィックスに対して対話的な操作を行うことは容易ではない。現在、この欠点を補うために、iPlots (<http://www.iplots.org/>) という対話的なグラフ操作ができるような機能の実装が進められている。

条件をつけて統計グラフィックスを描くというアイデアは Cleveland (1993) により提案された。このアイデアは、統計解析システム S (Chambers and Hastie, 1992) の Trellis Graphics として実現され、その有効性が示された。そして、現在では、R の `coplot` コマンド (Murrell, 2005) として広く使われている。われわれのシステムでは、これらの有用性を踏襲しながら、さらに、条件の基準を対話的に変更することを可能にし、条件によって分けた観測値を次の解析に連続的に利用することができるようにしている。

平行座標プロットと散布図やコロプレスマップを連動させた多変量地理情報データの視覚化手法は Edsall (2003) によって報告されている。これは、マウス操作により平行座標プロット上で観測値を表す折れ線を選択状態にすると、連結された散布図やコロプレスマップ上の対応する点や領域が強調表示されるというものである。これにより解析者は、データの位置関係を観察しながら多変量の地理情報データの特徴を把握することができる。しかしながら、Edsall の方法は、平行座標プロット上のマウス操作とその効果については言及しているが、散布図やコロプレスマップ上での対話的な利用については踏み込んだ議論がされていない。条件付きコロプレスマップの利用や観測値の強調表示ができることが Edsall とわれわれの提案の違いである。地理情報データの解析では、平行座標空間でのデータの操作だけでなく、地図上でのデータの操作も解析上必要なことは明らかであろう。また、この効果については3章で言及している。

5. おわりに

本研究では、多変量の地理情報データを視覚化するひとつの方法として、各領域の位置関係を観察するために目的変数に対して2つの変数による条件付けを行える条件付きコロプレスマップと、多変量データをひとつの図上に表示することができる平行座標プロットを双方向に連携することを提案した。これにより、解析の初期段階でデータの全体像を眺めたり、偏りや外れ値の有無を観察することが地図を基本としたグラフィックスだけを利用するよりも容易にできることを示した。

これまで地理情報データを解析する場合、各領域の位置関係が重要な要因となるため、ドットマップやコロプレスマップのような地図を基本としたグラフィックスを利用することが好まれてきた。しかしながら、地図の場合には、表現できる情報は条件付きコロプレスマップを利用した場合でもせいぜい3変数までである。それ以上の多変量データの場合には、利用者が

なんらかの方法で変数の選択を行うことが必要となるが、変数が増えてくればその作業は大きな負担となる。そのために、われわれは、平行座標プロットのように多変量を同時に表示できる統計グラフィックスを、地図を基本としたグラフィックスと密接に連携して利用することにした。

ここで紹介した解析環境は汎用統計解析システム Jasp に実装したので、他の統計手法やグラフと同時に利用することができる。Jasp はわれわれのサイト (<http://jasp.ism.ac.jp/>) において公開されている。

謝 辞

本論文で使用した徳島県の高齢者の自殺に関するデータは徳島県警察本部捜査一課からのものである。また、査読者ならびに編集委員の先生方には有益なご助言と修正のご指摘を賜りました。記して深い感謝の意を表します。

参 考 文 献

- Bivand, R. S. (1999). Integrating GRASS 5.0 and R: GIS and modern statistics for data analysis, *Proceedings 7th Scandinavian Research Conference on Geographical Information Science*, 111–127, Aalborg, Denmark.
- Carr, D. B., Wallin, J. F. and Carr, D. A. (2000). Two new templates for epidemiology applications: Linked micromap plots and conditioned choropleth maps, *Statistics in Medicine*, **19**, 2521–2538.
- Chambers, J. and Hastie, T. (eds.) (1992). *Statistical Models in S*, Chapman Hall, New York.
- Chang, K. (2006). *Introduction to Geographic Information Systems 3rd Edition*, McGraw-Hill, New York.
- Cleveland, W. S. (1993). *Visualizing Data*, Hobart Press, Summit, New Jersey.
- Edsall, R. M. (2003). The parallel coordinate plot in action: Design and use for geographics visualization, *Computational Statistics and Data Analysis*, **43**, 605–619.
- 古田 均, 吉川 眞, 田中成典, 北川悦司 編(2005). 『基礎からわかる GIS』, 森北出版, 東京.
- Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics*, **5**, 299–314.
- Inselberg, A. (1985). The plane with parallel coordinates, *Visual Computer*, **1**, 69–91.
- 小林郁典, 中野純司, 山本由和, 藤原丈史 (2006). Java による統計システムへのデータマイニング機能の実装, *計算機統計学*, **18**(1), 15–25.
- 村山祐司 編(2005). 『地理情報システム』, 朝倉出版, 東京.
- Murrell, P. (2005). *R Graphics*, Chapman & Hall/CRC, Boca Raton.
- Nakano, J., Fujiwara, T., Yamamoto, Y. and Kobayashi, I. (2000). A statistical package based on Pnuts, *COMPSTAT2000 Proceedings in Computational Statistics* (eds. J. G. Bethlehem and P. G. M. van der Heijden), 361–366, Physica-Verlag, Heidelberg.

Multivariate Geographical Data Visualization Using Linked Statistical Graphics

Ikunori Kobayashi¹, Yoshikazu Yamamoto¹, Junji Nakano² and Jung Jin Lee³

¹Faculty of Engineering, Tokushima Bunri University

²The Institute of Statistical Mathematics

³Department of Statistics, Soong Sil University

Multivariate geographical data have become widely available and are being used for various purposes because of the development of information technologies. As geographical data include locations of regions, map-based graphics such as dot maps and choropleth maps are suitable for expressing them. Map-based graphics, however, are not suitable for expressing high dimensional data.

This paper proposes the use of linked statistical graphics, especially conditioned choropleth maps (CCmaps) and parallel coordinate plots (PCPs), for visualizing and analyzing multivariate geographical data.

The CCmaps is an extension of a choropleth map and arranges several choropleth maps according to conditions defined by two other variables. The PCP is appropriate for describing multivariate data in a single graph, but are not good at showing location information.

We illustrate that linked views and interactive operations on these graphics are effective for geographical data analysis. The proposed functions are implemented in a statistical analysis system Jasp.