

資料

## 「地域における自殺の基礎資料」を用いた 自治体別のコロプレス図（色分け統計地図）の作成

新海浩之<sup>1</sup>

### 【要旨】

都道府県及び市町村は当該地域内の自殺対策に当たって、自殺総合対策大綱及び地域の実情等を勘案して自殺対策計画を策定することとされている（自殺対策基本法第13条）が、各自治体において科学的根拠に基づく自殺対策を発展させるためには、既存の統計資料を最大限に活用することが第一歩である。現在、警察によって収集された自殺者の統計が厚生労働省において「地域における自殺の基礎資料」として公表されているが、その内容は非常に豊富であるものの、数値の羅列であり、やや無味乾燥で一見して理解するのが難しい。

そこで、本稿においては、オープンソースのソフトウェア及び地図情報を用いて同基礎資料所収の統計に基づく市町村単位で色分けした統計地図（コロプレス図）として作成する手順を示す。

コロプレス図は、数値を表等に配置した統計資料に比較すると精緻さでは劣るものの、視覚化に優れているため、職員及び住民等に対して訴求力の高い資料となり、自殺対策の発展に資する可能性が期待される。

キーワード：自殺、視覚化、コロプレス図、R

### 1. はじめに

自殺対策基本法第13条においては、都道府県及び市町村は自殺総合対策大綱及び地域の実情等を勘案して、自殺対策計画を策定することとされている。各自治体において科学的根拠に基づく自殺対策を発展させるために、統計資料を活用することが第一歩であることは言うまでもない。

いのち支える自殺対策推進センターは毎年、各自治体向けにその自治体における自殺の資料を「地域自殺実態プロファイル」として提供しており、これにより地域における詳細な自殺の実態を把握できる。この地域自殺実態プロファイルの活用場面のひとつ

として、自殺対策の推進のための大前提として、地域の関係者が当該地域の自殺の実態に関する認識を共有することがあげられる<sup>6)</sup>。

地域自殺実態プロファイルは、各自治体について年代・性別・職業等の組み合わせで（例えば、「40～59歳での男性の無職で独り暮らし」など）での自殺者の動向を知ることが可能とするものであるが、自治体ごとに作成されているため、都道府県としては所属自治体間の差異の全体像が把握しにくいという問題点もある。

一方、厚生労働省が発表している「地域における自殺の基礎資料」<sup>8)</sup>には男女別、年齢層別、職業別等

<sup>1</sup> 神奈川大学法学部

の自殺者数が詳細に掲載されており、自治体間の比較も可能だが、スプレッドシート上に数値が並べられているものであり、そのままでは傾向を掴むことは難しい。無機質な数値の羅列から意味やパターンを見出し、問題性を直感的に理解するためには、データをグラフやチャート等によって可視化することが必要であると言われている<sup>13)</sup>。

そこで、本資料においてはデータ可視化の一つの方法として、オープンソースのソフトウェア及び地図情報データを用いて都道府県・市町村ごとの自殺データを地図上に色分けするコロプレス図を作成する手順を示す。

## 2. コロプレス図とは

コロプレス図とは、コロプレスマップとも呼ばれ、1次元のデータを地図上で異なる色や模様によって分類し、対応する領域（今回の場合は行政区画が対象）を塗り分けるものである。コロプレス図は、数値を表等に配置した統計資料に比較すると精緻さでは劣るものの、視覚化に優れているため、訴求力の高い資料となることが期待できるが、以下のような効用と限界が挙げられる。

### 2.1 コロプレス図の効用

コロプレス図の効用としては、以下の4点が挙げられる。

- データの視覚的な比較が容易：コロプレス図は地理的な領域ごとに色やパターンなどを用いてデータを表現するものであるため、異なる領域間での数値的な差を色の濃淡や明暗などによって比較するにあたって直感的な理解が可能となる。
- 分布のパターンの把握：地理的に隣接する領域同士が似たような色や濃度を持つ場合、その領域間で共通の傾向やパターンが存在する可能性が考えられ、これによって領域ごとの分布のパターンを把握しやすくなる。
- 効果的なコミュニケーション手段：視覚的な要素が強調されているため、一般の人々や非専門家にも情報をわかりやすく伝えることが可能となる。政策決定者や一般の人々に対してデータを分かりやすく伝える手段として活用できる。
- 領域間の相対的な重要性の把握：特定の領域が他の領域と比べて高いまたは低い値を持っている場合、その地域の相対的な重要性を容易に把握することが可能となる。したがって、政策立案や戦略の決定において役立つ。

### 2.2 コロプレス図の限界

コロプレス図の限界としては、以下の3点が挙げられる。

- 領域のサイズ効果：領域全体が同じ色で表示されることから、領域のサイズが大きい場合、データが一様に分散していない限り、領域の実際の特徴を正確に反映しない可能性もある。
- データの不均一性：領域ごとの人口密度や地域の大きさによって、データの解釈が歪められる可能性がある。人口密度が高い領域と低い領域を同じように比較することは、誤った結論につながる可能性がある。
- 色彩やパターンの認知の問題：可視化技術にあっては、使用する色の選択によって問題の緊急度や深刻度に関する認知に差が生じる可能性があると言われている。また、不用意にカラフルなものを作成すると却って誤った印象を与えることになる可能性がある<sup>11)</sup>。

## 3. 作成にあたって必要なデータとソフトウェア

コロプレス図の作成にあたっては、地図上に表示するデータ（本稿の場合自殺データ）、地図データ、

それらのデータを処理するためのソフトウェアが必要となる。

### 3.1 地図上に表示するデータ

地域の自殺を示すデータには様々なものがあるが<sup>2</sup>、本稿では厚生労働省が発表している「地域における自殺の基礎資料」に所収の自殺統計を用いる。自殺統計とは、都道府県警察で検視・死体調査により遺体の死因を自殺と判断したケースについて、案件毎に自殺統計原票を作成してデータ化したものを警察庁において取りまとめているものである。この情報の提供を受けて厚生労働省において資料化したものが、「地域における自殺の基礎資料」である<sup>8)</sup>。

### 3.2 地図データ（シェープファイル）

コロプレス図の作成には、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地理情報システム（GIS：Geographic Information System）が必要である。身近な例を挙げると、スマートフォン等で地図ソフトを使ったり、自動車の位置情報を示すための基盤である。本稿の目的のためには、パソコンで読み取ることのできる、都道府県および市町村の境界を示した白地図のようなデータが必要となる。地理情報システム（GIS）に利用するそのようなデータには様々な形式がありうるが、今回は「シェープファイル」を用いる。シェープファイルとは、地理空間情報システム（GIS）で使用されるファイル形式であり、地図上の対象物の位置や形状、属性情報を点（ポイント）・線（ライン）・面（ポリゴン）で表現するベクターデータを格納しているものであり、最新のものに比較して情報量に制約があるもの

の、仕様が公開されているため、汎用性が高いことが特徴である。

### 3.3 データ処理のためのソフトウェア

データに加えて、それを処理して地図上に作画するソフトウェアが必要となるが、本稿では無償のオープンソースの統計ソフトである、「R」を用いる<sup>3)</sup>。Rは無償でありながら、世界中のユーザーが開発したパッケージを用いることで、高価な市販ソフトウェアと同等若しくはそれを上回る統計計算が可能以上に非常に優れたグラフィックエンジンを備えているため、今回の目的に適している。

なお、一般によく用いられている Excel でも、新しいバージョンであれば「塗り分け地図」という機能を用いてコロプレス図を作成することは可能ではあるが、標準機能では都道府県単位の地図しか作成することはできない。<sup>3 4</sup>

## 4. 作業の概要

作業の概要を示すと以下のとおりとなる。

- Rの入手
- 必要なデータ（自殺関連データ、シェープファイル）の入手
- Rのパッケージの準備・データの読み込み・加工
- 自殺関連データとシェープファイルの合体
- コロプレス図の作成

本稿では、都道府県別の自殺者数、自殺死亡率に関するコロプレス図を作成する手順を先に示す。次いで市区町村別のコロプレス図を作成するが、都道府県別の図を作成する手順と重なる部分は省略する。

<sup>2</sup> 自殺統計以外では人口動態統計などが考えられる。

<sup>3</sup> Excelでも、先ほどのシェープファイルと合わせて用いると、市町村単位のコロプレス図を作成することは可能である。しかし、その場合本稿で紹介するRで行う方法に比べて格段に煩雑な前処理が必要となる。

<sup>4</sup> そのほかに利用可能なソフトウェアとしては、地図作成に特化した無償のソフトウェアである、QGIS (<https://qgis.org/ja/site/about/index.html#>) を用いる方法や、有償のソフトウェアでは、ArcGIS (<https://www.esri.com/products/arcgis/>) などが挙げられる。

## 4.1 Rの入手

Rのインストールにあたっては、使用しているOS (Windows, MacOS, Linux) に従ったRの最新版をダウンロードしてインストールする。また、RStudioというIDE (統合型開発環境) も同時にダウンロードしておくにより使いやすくなる。特に、RStudioで用いることのできる *RMarkdown* という記法により、Rの操作に必要なコマンド文の記述とそれに伴う解説や文章を同時に作成することができ、再現性の高い開発が可能となるため、利用を推奨する<sup>9)</sup>。本稿の執筆も基本的に *RMarkdown* を用いている。

## 4.2 R操作の基本

Rの使用に関する基本的な部分のみを紹介する。Rは、最近の一般のソフトウェアのようにマウスなどでダイアログボックスをクリックして操作を行うGUI (Graphical User Interface) のソフトウェアではなく、コマンド文を入力することによって動作させるCUI (Character User Interface) ソフトウェアで

```
> a <- 1+2 # 1+2 の計算結果を a というオブジェクトに代入
> a # a を入力すると
[1] 3 # a に入っている計算結果が表示される。
> b <- 4+5 # 4+5 の計算結果を b というオブジェクトに代入
> b
[1] 9 # b に入っている計算結果が表示される
> c <- a*b # a*b の計算結果を c というオブジェクトに代入 (乗算、除算の表現はエクセル等と同様)
> c # c と入力すると
[1] 27 # c に入っている計算結果が表示される
> a <- ("hello world") # a というオブジェクトに "hello world" という文章を代入
> print(a) # print(オブジェクト名) という関数で () 内のオブジェクトを表示させる
[1] "hello world" # 表示がされた
```

Rの使用方法の詳細については紙幅の関係で割愛するが、多くの優れた概説書があり<sup>2)</sup>、また、インターネット上にも多くの解説が掲載されているので、そちらを参照いただきたい。

あるため、最初はとっつきづらいものであるが、他のソフトに比較してコマンドが人間の言語 (英語) に近いものであるため、以下の例を見れば自ずと内容は理解できると考える。

なお、コマンド文の後ろの#に続けて任意の文字を入力することができる。本稿でもこれ以降、#の後に適宜各コマンドの説明を付すことにする。

Rにおいてはデータや一時的なベクトル、変数等は全てオブジェクトというものに格納して処理する。そのためには、オブジェクトの名前に対して左矢印<-を付して対象を代入すると、命令を書き込む。代入にあたっては、パッケージ等で定められた関数(命令)を用いることもある。オブジェクトの名称は任意であるが、Rの他のコマンドと重複しないようなものを選ぶ必要がある。

作成したオブジェクト同士で演算することも可能である。

ごく簡単な例を示す。

## 4.3 データの入手

### 4.3.1 自殺関連データ

最初に自殺関連のデータについて、必要な情報をダウンロードする。

今回は、厚生労働省の HP「地域における自殺の基礎資料」で 2022 年の確定値データを用いる<sup>8)</sup>。

上記サイトをみると、令和 4 年の各月別のデータの下に「令和 4 年確定値その 1」という項目と「令和 4 年確定値その 2」という項目が存在する。このうち「令和 4 年確定値その 1」は自殺日を基準とするデータであり、「令和 4 年確定値その 2」は発見

日を基準とするデータとなっている。必要なものを選定すればいいわけであるが、ここでは自殺日基準のデータを用いる。

「令和 4 年確定値その 1」のハイパーリンクからダウンロードされたファイルは `R4KAKUTEI-CHIKI01` というフォルダに圧縮されており、以下のファイルが収められている。

```
2022-0-2022CDR-22-00013_A1-4 表(全国・自殺日).xls
2022-0-2022CDR-22-00013_A5 表(県・自殺日・住居地).xls
2022-0-2022CDR-22-00013_A6 表(県・自殺日・発見地).xls
2022-0-2022CDR-22-00013_A7 表(市町村・自殺日・住居地)clon.xls
2022-0-2022CDR-22-00013_A8 表(市町村・自殺日・発見地)clon.xls
R4「地域における自殺の基礎資料」の利用にあたって.pdf
```

必要なものを選択するが、本稿では自殺日・住居地のデータを用いるため、以下の二つを使用する。

```
2022-0-2022CDR-22-00013_A5 表(県・自殺日・住居地).xls
2022-0-2022CDR-22-00013_A7 表(市町村・自殺日・住居地)clon.xls
```

ただし、このままではファイル名が長すぎ、また、R においては「-」や「()」があると処理ができないためファイル名を以下のように変更する。

```
2022_A5.xls
2022_A7.xls
```

このファイルを R の作業ディレクトリ下の `/Data` ディレクトリに保存しておく。

#### 4.3.2 シェープファイル

都道府県単位のシェープファイルは後述する R のパッケージに付属しているものを用いるため、個別にダウンロードする必要はない。

市区町村単位のシェープファイルは、国土交通省国土地理院が公表している国土数値情報ダウンロードサイトから入手する<sup>10)</sup>。国土数値情報は 5 分野に分かれているが、今回はこの中から「2.政策地域」の行政地域データを利用する。利用するデータに合わせて、全国、地方単位、県単位のデータがそれぞれ

入手可能であるが、本稿では全国版のデータを用いた方法を示す。

ダウンロードサイトから、「行政地域>行政区域(ポリゴン)」を選択すると、全国、地方、都道府県単位でファイルをダウンロードすることができる(ダウンロードの際にアンケートに回答する必要がある)。なお、全国版のファイルは 424MB とやや大きいので注意が必要である。

ダウンロードされたものは、`N03-20230101_GML` というフォルダになっており、その中で、`N03-23_230101.shp` がシェープファイルであるが、後述する R での読み込みでは他のファイルも必要となるため、フォルダの中に収められているファイルを R の作業ディレクトリ下の `/Data` ディレクトリに保存しておく。

#### 4.4 パッケージの導入

R では世界各国の開発研究者が無償で多くのパッケージ(追加機能のようなもの)を開発しており、それらを導入することで、より簡易にデータの処理

ができる。本稿ではパッケージについて言及するときは{パッケージ名}と表記する。

今回用いるのは以下のパッケージである。

- {tidyverse} R へのデータの読み込み、読み込んだデータの抽出、変形、加工、可視化といった、データ分析における基本的な作業を効率的に行うためのさまざまな機能を提供しているほか、ggplot という非常に優れた作画機能を持っている<sup>4)</sup>。
- {readxl} エクセル形式でのデータの読み込みに関して、シートや範囲を指定するために用いる。
- {NipponMap} 都道府県単位の簡易なシェープファイルが格納されている。
- {sf} シェープファイルを R に読み込ませるために用いる。

使用するパッケージはあらかじめ R にインストールしなければならないため、R 上で、`install.packages("パッケージ名")`としてインストールする。パッケージ名の周りのダブルクォテーションは必須である。

パッケージはインストールしただけでは使用することができないため、必要な時にライブラリ (R 中のフォルダのようなもの) に呼びだしておかなければならない。パッケージを呼び出すコマンドは、`library(パッケージ名)`である (このときはダブルクォテーションは不要)。

今回の場合だと以下のとおりとなる。

```
all_sui <- read_excel("Data/2022_A5.xls", range = "総数 (秘)!B8:F56") #自殺日・住居地のデータの読み込み
# 総数 (秘)シートを対象にする。
# 都道府県名とコード、自殺者数・自殺死亡率以外の数値は不要なので、B~F列しかいらぬ。
# 4行目以下しか必要ではない
male_sui <- read_excel("Data/2022_A5.xls", range = "男 (秘)!B8:F56")
female_sui <- read_excel("Data/2022_A5.xls", range = "女 (秘)!B8:F56") #元データのシートのフォントに注意。
```

```
library(tidyverse)
library(readxl)
library(forcats)
library(NipponMap)
library(sf)
```

まず、都道府県単位での地図の作成の手順を示す。

#### 4.5 データの読み込み (都道府県単位の方法)

まず、{readxl}パッケージの `read_excel` 関数を用いて、先ほど/Data フォルダに格納した自殺データのうち、都道府県単位のデータである、2022\_A5.xlsを読み込む。

ただし、2022\_A5.xlsには「総数 (秘)」、「男 (秘)」、「女 (秘)」という複数のシートが含まれており、R のオブジェクトとしては、複数のシートは取り扱えないため、それぞれを別のオブジェクトに代入し、後で結合する。

また、2022\_A5.xlsにはさまざまなデータが含まれているが、本稿では解説のために、「自殺者数」及び「自殺死亡率」の数値しか用いないため、当該部分のみを読み込んでいます。2022\_A5.xlsの他の数値を使用する際は列名指定の部分(「総数 (秘)!B8:F56」等としている部分。指定の仕方は Excel の方法と同じ。)を変更すれば良い。<sup>5)</sup>

<sup>5)</sup> なお、政府発行の統計データでは、タイトル行やフォーマットが年によって変更されることがあり、それによってデータの開始位置がずれることがあるため、注意が必要である。

続いて、都道府県単位のシェープファイルを読み込む。これは先ほど導入した *NipponMap* に収められているので、以下のコマンドで読み込むことができる。

```
map <- read_sf(system.file("shapes/jpn.shp", package = "NipponMap"))
# NipponMap 都道府県のシェープファイルをも map オブジェクトに代入する。
```

#### 4.6 データの加工（都道府県単位の方法）

読み込んだデータは総数、男性、女性と別々のオブジェクトに分かれているため、今後コロプレス図作成のために一つにまとめる必要がある。また、必要があれば、新しい変数を作成することもある。このように、データを変換することを「加工」と言ったり、「前処理」と言ったりする。Excel などでも可能であるが、一度変更してしまったものは元に戻せないため、間違いがあった際に遡って探り当てるのは困難である。以下のようにコマンド文によって作成しておけば、オリジナルのデータには手を加えていないので、何度でも異なる加工を繰り返すことが

できる。間違いがあった際はその間違いの部分だけを修正すれば良く、効率的である。これにより、再現性が高い処理が実現できる。

加工の手順は以下の各コマンドの説明を見ていただきたいが、`%>%` という見慣れないものが登場する。これは上述の *tidyverse* パッケージの導入で用いることのできる「パイプ関数」と言われるもので、関数の左側にあるオブジェクトに対して右側の処理を行うというものであり、`%>%` を連続して用いて一連の処理を連続して記載することを可能にしている。パイプ関数の使用により、右矢印 `->` を使用して、コマンド文末においたオブジェクトに対する代入の処理が可能となる。

```
all_sui %>% filter(!is.na(自殺死亡率)) -> all_sui #自殺死亡率がブランクであるものを取り除く
male_sui %>% filter(!is.na(自殺死亡率)) -> male_sui
female_sui %>% filter(!is.na(自殺死亡率)) -> female_sui
all_sui %>%
  mutate(男性自殺者数=male_sui$自殺者数) %>% # mutate 関数により、male_sui にある「自殺者数」というコラムを「男性自殺者」として all_sui に加える)
  mutate(男性自殺死亡率=male_sui$自殺死亡率)%>%
  mutate(女性自殺者数=female_sui$自殺者数) %>%
  mutate(女性自殺死亡率 = female_sui$自殺死亡率) -> all_sui # ここまでの一連の変形を all_sui に上書き
all_sui$都道府県コード <- formatC(all_sui$都道府県コード, width=2, flag = "0")
#all_sui の都道府県コードが一桁である場合に頭に 0 をつけた。後のシェープファイルとの結合のために必要。
rm(male_sui)
rm(female_sui) # all_sui に情報が集約されたため、不要となった male_sui 及び female_sui を削除
```

#### 4.7 データの結合

都道府県単位の男女別自殺者数等を読み込んだオブジェクト *all\_sui* とシェープファイルを読み込んだオブジェクト *map* を結合する。結合する場合、二

つのオブジェクトの間で共通のデータを「鍵」にしなければいけないが、ここでは、*map* オブジェクトの「*jiscode*」と *all\_sui* の「都道府県コード」が共通であるため、これを鍵として結合する。

```
map_all_sui <- left_join(map, all_sui, by = c("jiscode"="都道府県コード"))
```

#map と all\_sui のデータを jiscode と都道府県コードを鍵として結合する。

#### 4.8 各都道府県の自殺死亡率の視覚化（コロプレス図）

データの準備ができたので、都道府県の自殺死亡率を地図上に描画するためのコマンドを示す。描画

にあたっては使用するのは `{tidyverse}` パッケージに収められている極めて優れた作画機能を持つ、`ggplot` を使用する<sup>1)</sup>。

```
map_all_sui %>%
```

```
ggplot(aes(fill = 自殺死亡率)) +
```

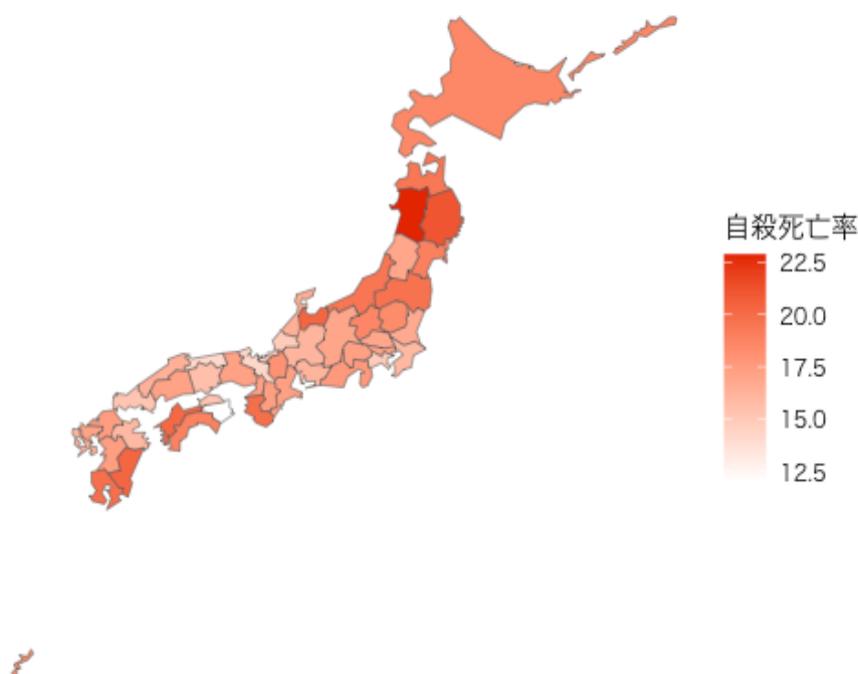
```
  geom_sf() +
```

```
  scale_fill_gradient(low = "white", high = "#E22400") +
```

```
  annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料", hjust=1, vjust=-.4, size=3, family = "HiraKakuPro-W3") +
```

```
  theme_void(base_family = "HiraKakuPro-W3")
```

```
ggsave("全国自殺死亡率 map2022.jpg", width=20, height =15, units="cm") #作図したコロプレス図をファイルに保存
```



出典：厚生労働省地域における自殺の基礎資料

図1 都道府県別の自殺死亡率（男女計，2022）

上述のとおり、作成にあたっては色の選択が重要であるが、ここでは `scale_fill_gradient` 関数によって、低い値を白で、高い値を鮮やかな赤 `#E22400` に徐々に変化する色調を用いた。従って、自殺死亡率が高くなると赤が濃くなっていくような配色になっている。

詳しいコマンドの説明は省略するが、コマンド内の「自殺死亡率」という部分を「自殺死亡者数」と変更すれば、死亡者数を基準にしたグラフを作成することができる。

最後の `ggsave` から始まるコマンドによって、生成したコロプレス図を画像ファイル (`jpg`) として出力している。この `jpg` ファイルは他のプレゼンテーションソフトなどで利用できる。

```
map_all_sui %>%
  ggplot(aes(fill = 男性自殺死亡率)) +
    geom_sf() +
    scale_fill_gradient(low = "white", high = "#E22400") +
    annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料", hjust=1, vjust = -.4, size=2, family =
"HiraKakuPro-W3") +
    theme_void(base_family = "HiraKakuPro-W3")
```

男性だけに限ると、秋田県と宮崎県は比較的高い状態のままであるが、富山県は目立たなくなったように見える。

#### 4.10 各都道府県の自殺死亡率（女性）

続いて、「男性自殺死亡率」の部分を変更すると女性の自殺死亡率の分布のコロプレス図が作成できる。

図を見ると、秋田県、岩手県、富山県で自殺死亡率が相対的に高く、京都府と徳島県で低いように見える。

#### 4.9 各都道府県の自殺死亡率（男性）

このようなコマンドによる作図の利点は、データオブジェクトに入っている情報であれば、コマンドの一部を書き換えることで簡単に別の図を作成できるところにある。例えば、上記のコマンドの「自殺死亡率」を「男性自殺死亡率」とすれば各都道府県の男性自殺死亡率のコロプレス図を作成することができる。<sup>6</sup>

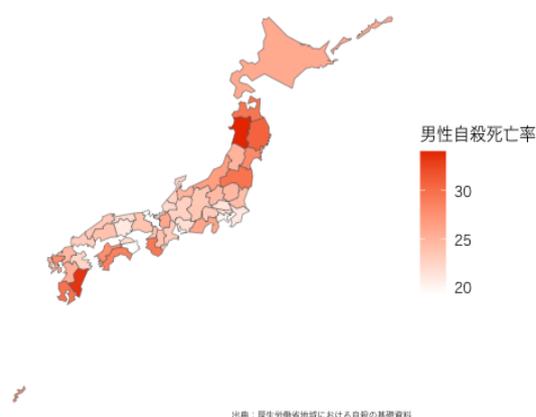


図2 都道府県別の自殺死亡率（男性, 2022）

```
map_all_sui %>%
  ggplot(aes(fill = 女性自殺死亡率)) +
```

<sup>6</sup> コマンドの書き換えで別の図を作成する場合、画像ファイルの作成コマンドでも同様にファイル名を変更しておかないと元の図が新しいもので上書きされてしまうので、注意が必要である。

```
map_all_sui %>%
ggplot(aes(fill = 女性自殺死亡率)) +
  geom_sf() +
  scale_fill_gradient(low = "white", high = "#E22400") +
  annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料", hjust=1, vjust=-.4, size=3, family =
"HiraKakuPro-W3") +
  theme_void(base_family = "HiraKakuPro-W3")
```

女性だけに限ると、富山県と愛媛県の自殺死亡率が比較的高いようにみる。また、千葉県に関してやや目立つようである。

様々な種類の図を簡単に作成することができ、訴求力の高い資料を手早く作成することができる。

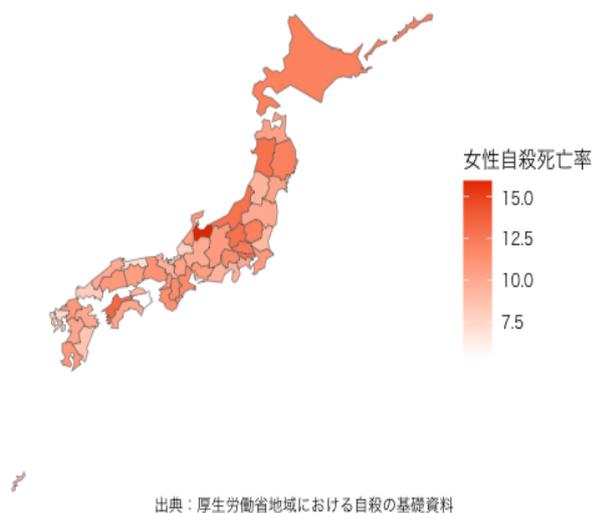


図3 都道府県別の自殺死亡率（女性, 2022）

このように、一度コロプレス図を作成するコマンド文を作成すると、その後は該当箇所を変更するだけで、男女計の自殺死亡率、男性のみ、女性のみ等、

## 5. 市区町村データの利用

地域における自殺者の分布についてより詳しく見るためには、都道府県における所属市区町村の状況を見る必要がある。

そこで、次に、市区町村データを用いたコロプレス図を作成する手順を示す。R やそれに伴うパッケージの導入などについては都道府県別の図を作成したものと共通であるため説明は省略し、関連のデータの読み込み、加工、結合および作成についてのみ解説する。

### 5.1 データの読み込み（市区町村単位の方法）

読み込むデータは `2022_A7.xls` であるが、手順は基本的に都道府県単位の場合と同様である。ここでも「自殺者数」及び「自殺死亡率」の数値のみを読み込んでいる。

```
mun_sui <- read_excel("Data/2022_A7.xls", range = "総数 (秘)!B8:G1667")
mun_sui %>% filter (市区町村コード!="NA") -> mun_sui
mun_male <- read_excel("Data/2022_A7.xls", range = "男 (秘)!B8:G1667")
mun_male %>% filter (市区町村コード!="NA") %>%
  mutate(男性自殺者数=自殺者数) %>%
  mutate(男性自殺死亡率=自殺死亡率) %>%
  select( 市区町村コード, 男性自殺者数, 男性自殺死亡率) -> mun_male
```

```
mun_female <- read_excel("Data/2022_A7.xls", range = "女 (秘)!B8:G1667")
mun_female %>% filter (市区町村コード!="NA") %>%
  mutate(女性自殺者数=自殺者数) %>%
  mutate(女性自殺死亡率 = 自殺死亡率)%>%
  select( 市区町村コード, 女性自殺者数, 女性自殺死亡率) -> mun_female
```

続いて、市町村単位のシェープファイルを読み込む。これは先ほど/Data ディレクトリに保存した `N03-20230101_GML` 関連のファイルを読み込む。

コマンド上は `.shp` ファイルだけを読み込んでいるように見えるが、他のファイルも使われている。

```
mun_map <- read_sf("Data/N03-23_230101.shp", options = "ENCODING=CP932", stringsAsFactors=FALSE)
```

## 5.2 データの加工（市区町村単位の方法）

都道府県のデータの読み込みの際と同様、総数、男性及び女性について別々のオブジェクトに読み込んだため、それらを一つにまとめる必要がある。今

回は非常に大きなデータとなるため、`{tidyverse}` パッケージの関数による簡易な方法を示した。

後でシェープファイルとの結合のために、市区町村コードを統一化するための加工を行なっている。

```
mun_sui %>% left_join(mun_male, by="市区町村コード") %>%
  left_join(mun_female, by="市区町村コード") ->mun_sui
rm(mun_male)
rm(mun_female) # 不要になった `mun_male` 及び `mun_female` を削除
mun_sui$市区町村コード <- as.integer(mun_sui$市区町村コード) # 市区町村コードを整数化する
mun_sui$市区町村コード <- formatC(mun_sui$市区町村コード, width=6, flag="0") # 市区町村コードが五桁のものもあるため頭に0をつけて、六桁に揃えた
mun_sui$市区町村コード <- str_sub(string = mun_sui$市区町村コード, start=1, end=5) # 市区町村コードを五桁に揃える (一番後ろは不要)
```

シェープファイルを代入した `mun_map` オブジェクトと `mun_sui` オブジェクトを共通の市区町村

コードを鍵にして結合し、`mun_all_map` オブジェクトを作成する。

```
mun_all_map <- left_join (mun_map, mun_sui, by =c("N03_007" = "市区町村コード")) # mun_map の "N03_007" と mun_sui の "市区町村コード" が共通しているので、それを鍵に結合する。
```

## 5.3 市区町村単位の自殺者数の視覚化（コロプレス図：神奈川県の場合）

ここまでで作成した `mun_all_sui` オブジェクトには日本の全ての市区町村の位置情報と自殺関連のデータが含まれているため、そのままコロプレス図を作成した場合、日本地図の上に全自治体の自殺者数が色分けされて作成されることとなる。しかし、そ

れでは色分けが細かくなりすぎてかえって可視化の意味がなくなってしまう。ある程度地域を区切って（例えば、都道府県ごとに）コロプレス図を作成する方が現実的である。

`mun_all_sui` オブジェクトには、`N03_001` 列に都道府県名が入っているため、`{tidyverse}` パッケージの `filter` 関数を用いて適切な都道府県のデータ行を抽出する。ここでも元のオブジェクトの形を崩さず

に該当部分だけを抽出することができるため、同一のオブジェクトから様々な異なる抽出ができ、非常に効率的である。

ここでは、神奈川県に該当するデータを抽出してその所属市区町村の自殺者数のコロプレス図を作成する手順を示す。

```
mun_all_map %>% filter(N03_001 == "神奈川県") %>% # `mun_all_map` から神奈川県に該当するものだけを抽出
  ggplot(aes(fill = 自殺者数)) + #自殺者数について作成
    geom_sf() +
    scale_fill_gradient(low = "white", high = "#E22400") +
    annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料¥n なお、グレーは当該年に自殺者がなかったこと (NA) を示す", hjust=1, vjust = -.4, size=2, family = "HiraKakuPro-W3") +
    theme_void(base_family = "HiraKakuPro-W3")
```

ここでは、横須賀市と藤沢市の自殺者数が他の自治体に比較して多かったことがわかる。

なお、市区町村においては、該当年に自殺者がいない場合もあるため、そのような地域は該当値なし (NA) として、グレーで塗りつぶしている。2022 年においては、山北町と清川村が NA であった。

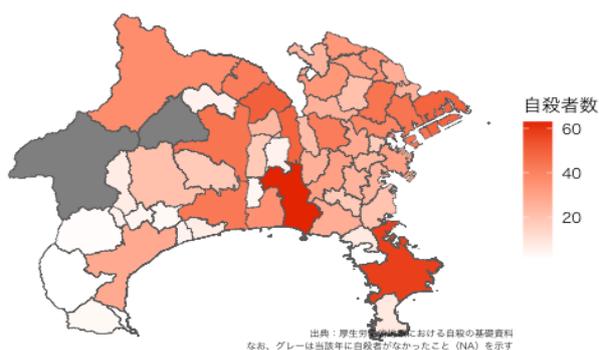


図 4 神奈川県の自殺者数 (男女計, 2022)

### 5.3.1 神奈川県の自殺死亡率

図 4 では、神奈川県の各自治体における自殺者数を見たが、藤沢市と横須賀市は人口が多いために自殺者が多かったという可能性もあるため、人口に対する自殺者の割合を示す自殺死亡率で見た場合には異なる様相を示す可能性もある。そこで、同じ神奈川県での市区町村の地図で自殺死亡率を示すと以下のようなになる。

```
mun_all_map %>% filter(N03_001 == "神奈川県") %>%
  ggplot(aes(fill = 自殺死亡率)) +
    geom_sf() +
    scale_fill_gradient(low = "white", high = "#E22400") +
    annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料¥n なお、グレーは当該年に自殺者がなかったこと (NA) を示す", hjust=1, vjust = -.4, size=2, family = "HiraKakuPro-W3") +
    theme_void(base_family = "HiraKakuPro-W3")
```

すると、自殺者数で表示した際は突出して見えた横須賀市と藤沢市は目立たなくなり、代わりに松田町が濃くなっている。これは、例えば 2022 年の自殺者数は藤沢市で 63 人、松田町では 7 人であったが、同地域の人口の差から、自殺死亡率で見るとそれぞれ 14.2 及び 65.0 となることから生じる現象である。

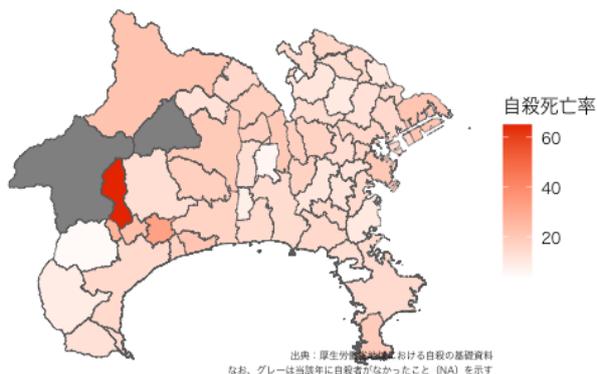


図 5 神奈川県の上殺死亡率（男女計、2022）

#### 5.4 政令指定都市のコロプレス図

今回作成したコロプレス図では、政令指定都市を含む都道府県の場合、政令指定都市内の区と一般の市区町村が同列に描画されている。 `mun_all_map` をみると `N03_003` 列に都市名があるため、「横浜市」だけを抽出したコロプレス図を作成することもできる。ここでは、女性の自殺死亡率をマッピングする。

また、色の効果の違いを見るために女性自殺死亡率が高くなると白から濃い青 "#000088" へのグラデーションで表現されるように作成した。

```
mun_all_map %>% filter(N03_003 == "横浜市") %>% # 横浜市のみを抽出した。
  ggplot(aes(fill = 女性自殺死亡率)) +
  geom_sf() +
  scale_fill_gradient(low = "white", high = "#000088") +
# ここで、最高値を濃い青 (#000088) になるよう変更している。
  annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料¥n
  なお、グレーは当該年に自殺者がな
  かったこと (NA) を示す", hjust=1, vjust=-.4, size=2, family = "HiraKakuPro-W3") +
  theme_void(base_family = "HiraKakuPro-W3")
```

ここでは、西区をはじめとして緑区、栄区が相対的に高く、一方で、鶴見区、保土ヶ谷区が低いことがわかる。しかし、色の効果としては青よりも赤の方が訴求力が高いように感じられる。

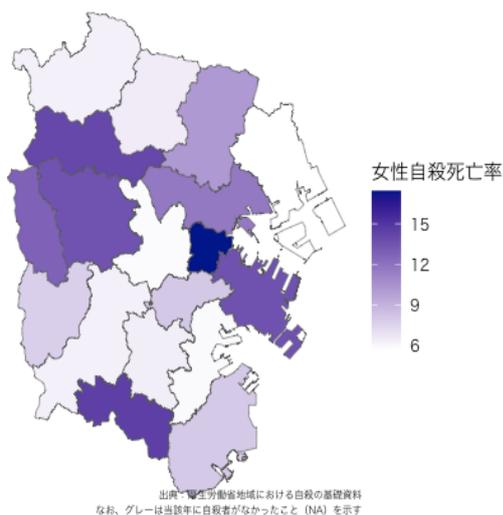


図6 横浜市の自殺死亡率（女性, 2022）

## 6. コロプレス図作成時の注意：島嶼部などを含む場合

コロプレス図では、データ上に存在する市区町村データを位置座標に従って機械的に描画するため、東京都や鹿児島県、沖縄県等、県域に離島が含まれる場合には、島嶼部を除く手立てをとらないと地図の領域が著しく広がる結果、市町村の領域は極めて小さく表現されてしまい、かえって視認性が悪い図になってしまう。

ここでは、東京都の例を示す。

### 6.1 島嶼部を含む東京都の自殺者数のコロプレス図

島嶼部を除かずに東京都の自殺者数をコロプレス図に作成してみる。

```
mun_all_map %>% filter(N03_001 == "東京都") %>% # `mun_all_map`から東京都のデータを抽出
  ggplot(aes(fill = 自殺者数)) +
  geom_sf() +
  scale_fill_gradient(low = "white", high = "#E22400") +
  annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料。なお、グレーは当該年に自殺者がなかったこと（NA）を示す", hjust=1, vjust=-.4, size=2, family = "HiraKakuPro-W3") +
  theme_void(base_family = "HiraKakuPro-W3")
```

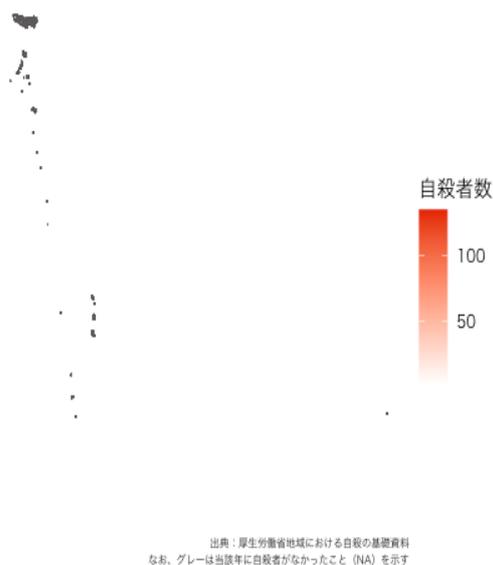


図7 東京都の自殺者数（男女計, 2022）

東京都は行政的には南は小笠原列島まで含むため、領域が広くなりすぎ、可視化の意味をなさない図となってしまった。

### 6.2 島嶼部を除いた東京都の自殺者数のコロプレス図

東京都の島嶼部は、大島町、利島村、新島村、神津島村、三宅村、八丈村等が知られているが、GISデータでは、鳥島、ベヨネース列岩、須美寿島等という一般にはあまり知られていない地域も網羅されているため、これらを漏れなく特定して *islands* というオブジェクトに代入（これは文字ベクトルと言われるものである。）し、元のデータからその部分を取り除いた地図を作成する。

```
islands <-c("大島町", "利島村", "新島村", "神津島村", "三宅村", "御蔵島村",
           "八丈町", "青ヶ島村", "小笠原村", "鳥島", "ペヨネース列島",
           "須美寿島", "中央防波堤外側廃棄物処理場（中潮橋南側）",
           "孀婦岩", "所属未定地", "荒川河口部") # 島嶼部の名称を `islands` というオブジェクトに指定

mun_all_map %>% filter(N03_001 == "東京都", !N03_004 %in% islands) %>% # 都道府県が「東京都」であるが、市区町村名
に `islands` に入っているものが該当すると取り除くという趣旨

ggplot(aes(fill = 自殺者数)) +
geom_sf() +
scale_fill_gradient(low = "white", high = "#E22400") +
annotate("text", x=Inf, y=-Inf, label="出典：厚生労働省地域における自殺の基礎資料Yn なお、グレーは当該年に自殺者がな
かったこと (NA) を示す", hjust=1, vjust = -.4, size=2, family = "HiraKakuPro-W3") +
theme_void(base_family = "HiraKakuPro-W3")
```

作業の結果、視認性の高いコロプレス図が作成された。特別区では足立区、大田区及び世田谷区の自殺者が多く、多摩地域では八王子市の自殺者数が多い状況が見て取れる。

示すコロプレス図の作成の方法をみてきた。データの視覚化にはほかにもさまざまな方法があり、Rがその用途に用いることができることはもちろん、Rは統計ソフトとして最新の統計分析を行うこともできるものであることから、今後の分析に役立てていただきたい。

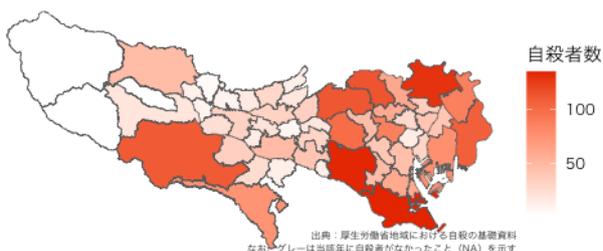


図8 東京都（島嶼部を除く）の自殺者数（男女計, 2022）

また、本稿で紹介したコロプレス図では、自殺者が多い（色の濃い）市区町村があっても、その地域の名称やそこにおける自殺者数といった情報をすぐに把握できず、他の地図情報を参照する必要がある。その問題を克服するためには、動的なコロプレス図を作成する必要があるが、Rで`leaflet`というパッケージを用いてインターネット上の地図にhtml方式で動的な地図をプロットし、マウスオーバーによって具体的な市区町村名や自殺者数等の統計数値を表示させ、拡大や移動も可能な視覚化も可能であるが、紙幅の関係で方法の共有は別の機会に譲る。

## 7. 終わりに：コロプレス図の活用とその先

本稿ではデータ視覚化の一つの方法として、Rの優れたグラフィック処理能力を用いて都道府県別及び市区町村別に自殺関連の情報を地図上色分けして

コロプレス図は自治体相互の自殺の状況を概観するために有用な可視化技術と言え、情報を可視化することによって地域住民や政策担当者に地域の自殺の概要を把握し、現状について考える機会を持って

もらう等、地域における自殺対策の推進に役立つものと思われる。

本稿では 2022 年単年の情報のみに基づく図を作成したが、実務上は時系列を追って同一のコロプレス図を作成することで自殺関連の情報の変遷について市区町村ごとの比較を行うことも効果的であろう。また、住居地と発見地の情報を並べること、発生時間別の情報を使用することなどにより、自殺の実態に対する多面的な理解が可能になると考えられる。さらに、自殺単体の情報に限らず、生活保護加入世帯数、職業の分布情報等関連の情報との組み合わせ等によってより多角的な自殺やその他の地域の問題の分布を空間地理的な観点から分析する例もある<sup>5, 12)</sup>。

一方でコロプレス図には 2.2 で述べた限界もある。コロプレス図は単位とする領域（本稿の例では自治体）を同一のデータで塗りつぶすことから、どのような情報を用いるにせよ、問題があたかも当該自治体に均一に存在しているかのような印象を与えてしまう可能性がある。これは市区町村単位での集計情報を行政区域に作図するというコロプレス図の特質上避けることのできない課題である。

現在、GIS 技術の発展には目覚ましいものがあり、街区レベルや大字・町丁目レベルといったより細かい地図情報を用いたり、三次元的な情報をプロットしたり、交通網上に人流データを加えたりすることも可能となっている。個別の自殺情報やその経緯を地理的にピンポイントで示すことも技術的には可能であり、現実につきまといや子どもへの不審な声かけ事案などを町丁目レベルで注意喚起する事件事故発生マップのような利用例もある<sup>7)</sup>。ただし、そのような利用のためには「地域における自殺の基礎資料」のような集計データではなく、個人レベルの粒度の情報の利用が必要となる。これは前述した様々な情報の組み合わせの際にも必要となる。自殺に関する情報は極めて機微な個人情報であることに配慮することが大前提ではあるものの、今後の自殺対策

のさらなる発展を目指す場合、将来的には検討の必要性もあるのではないだろうか。

## 文献

- 1) Chang, Winston (2013) 『R グラフィックス クックブック : ggplot2 によるグラフ作成のレシピ集』 オライリー・ジャパン, オーム社 (発売).
- 2) Lander, Jared P. (2018) 『みんなの R』 マイナビ出版.
- 3) R Core Team (2023) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.
- 4) Wickham, H, Averick M, Bryan J, Chang W, et. al. (2019) Welcome to the tidyverse. Journal of Open Source Software, , 4(43), .
- 5) 岡 檀・久保田 貴文・椿 広計, 他 (2017) 「和歌山県の地理的特性と自殺率の関係: 地理情報システムによる可視化と地域差を表す指標の検討。」『日本公衆衛生雑誌,』 64: 36-41.
- 6) 金子 善博・藤田 幸司 (2018) 「地域自殺実態プロファイル」と「地域自殺対策政策パッケージ」. 『法律のひろば,』 71: 14-21.
- 7) 警視庁 事件事故発生マップ. [https://www.keishicho.metro.tokyo.lg.jp/jiken\\_jiko/hassei/map\\_annai.html](https://www.keishicho.metro.tokyo.lg.jp/jiken_jiko/hassei/map_annai.html).
- 8) 厚生労働省 地域における自殺の基礎資料. <https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000140901.html>.
- 9) 高橋 康介 (2018) 『再現可能性のすゝめ: RStudio によるデータ解析とレポート作成.』 共立出版.
- 10) 国土地理院 国土数値情報ダウンロードサイト. <https://nlftp.mlit.go.jp/ksj/index.html>.
- 11) 山根 万由子・雨宮 護・白川 真裕, 他 (2020) 「犯罪発生マップにおける地図表現

- の実態と閲覧者の認知への影響.』『都市計画論文集』, 55: 385-392.
- 12) 入江 安子・南 由貴代・上浦 千明, 他 (2015) 「地域診断における GIS の活用.」『奈良県立医科大学医学部看護学科紀要』, 11: 3-13.
- 13) 鈴木 雅彦・鈴木 嘉右 (2015) 「データ可視化の必要性と意義 : データビジュアライゼーションとは(<特集>情報をわかりやすくするデザイン)」。『情報の科学と技術』, 65: 470-475.

Technical Note

## Choropleth Maps (color-coded statistical map) Using "Basic Data on Suicide in the Community."

Hiroyuki Shinkai

### 【Abstract】

Prefectural and municipal governments are required to formulate a basic plan for suicide countermeasures in accordance with the actual situation (Article 13 of the Basic Law on Suicide Countermeasures). To develop evidence-based suicide countermeasures in each municipality, the first step is to make maximum use of existing statistical data. Currently, statistics on suicides collected by the police are published by the Ministry of Health, Labor and Welfare as "Basic Data on Suicide in Local Communities". While they are extremely rich in content, they are a list of numerical values, somewhat tasteless and difficult to understand at first glance.

This paper presents methods for creating color-coded statistical maps (choropleth maps) for each municipality based on the figures contained in the above-mentioned "Basic Data", using open-source software and map information.

Although the choropleth maps are less precise than statistical data in which numerical values are arranged in tables, they are superior in visualization and are expected to be highly appealing to staff and residents, thereby contributing to the development of suicide countermeasures.

**Keywords:** suicide, visualization, choropleth map, R statistic