

Visualizing and Exploratory Data Analysis for Small Area Suicide Data

Takafumi Kubota^{*†}

Abstract—This study visualized small area suicide data especially municipality divided cross tabulation data in Japan. The data are categorized by several variables such as year, sex, age groups and so on. In this study exploratory data analysis were also performed. An implementation for this study is application of R package *shiny* and applied software of *Exploratory* that is powered by R. Tokushima prefecture was selected as the example of the data. From the results it was found Tokushima city had the highest number of suicide, and in south area there were municipalities which had relatively low number of suicide. The number of suicide of age group fifty to sixty-nine years old was the highest in age group and in the group male was slightly higher than female. In month trend, March and September were higher and male was more than double of female in each month, and the number of suicide of male in Tokushima city and age group fifty to sixty-nine years old are higher than other municipalities, sex and age groups.

Keyword: *visualizing; small area data; shiny; exploratory.*

1 Introduction

Suicide is one of the most serious issues that should be solved in Japan. Historical figures show that the number of suicide was stable at around 25,000 until 1997, but it had rapidly increased over 30,000 in 1998, and had been at a high level more than 10 years. Recently the figure was gradually decreasing, but it remains high level among other developed countries, even though fund for the Urgent Enhancement of Local Suicide Prevention Measures was launched in 2009.

In order to understand the situation by areal statistics of suicide in Japan, Fujita, et al. (2009) collected a broad number of statistics. The data divided in terms of area, year, age group and sex. Since there is no variable for reason relation, it is difficult to grasp the reason why people committed suicide.

On the other hand, the Cabinet Office (2016) has started detail data of several kinds of suicide in Japan by aggregating suicide data from National Police Agency. Kubota and Tsubaki (2014) introduced visualization ways by web application for reason-specified suicide data in Japan, and Kubota, et al. (2015) introduced visualization ways for panel data separated by prefectures in Japan by R package *micromap*.

However, there is necessary for personnel of suicide in each prefecture or municipality to apply smaller area such as municipality to find out regional specific policy making. "Basic area statistics of suicide by prefecture" of the Cabinet Office's data are divided in several kinds of regions not only prefectures but also municipalities.

Therefore, an objective of this study is to visualize municipality divided small area data of suicide. In this study Tokushima prefecture was selected because there are both high and low suicide risk areas in the prefecture.

2 Data

The data "Basic area statistics of suicide by prefecture" which are divided by municipalities have several variables such as year, sex, age group, housemate, occupation, reason, place, way, date, time periods, and attempted. Table 1 shows an example of the data. The data were used to visualize the number of suicide in each variable in shiny application. For drawing map shape file of GADM database of Global Administrative

* School of Management and Information, Tama University, 4-1-1 Hijirigaoka, Tama, Tokyo, 206-0022, Japan, E-mail: kubota@tama.ac.jp, Tel: +81-42-337-7164

Areas (Hijmans, 2016) in Japan was used.

To repeat count number times of cross tabulation cell, raw data is restructured from cross tabulation (table 2).

Table 1: Example of "Basic area statistics of suicide by prefecture"

City code	City name	Number of suicide	Suicide rate	Age group			...	Occupation			...	
				Under 20	20-29	.		Family	Employer	Unemployment		
										Student		...
...												
36202	Tokushima	48	18.64	0	8		11	30	3			
...												

Table 2: Data example for *Exploratory*

City code	Age group	Sex	Year	Month
36201	2.y30_49	1. male	2012	01
36202	2.y30_49	1. male	2012	01
36202	2.y30_49	1. male	2012	01
36204	1.y0_29	1. male	2012	01
...

3 Visualizing and exploratory data analysis

3.1 Shiny

An application for visualization of suicide data was implemented by an R package *shiny* (RStudio, Inc., 2016). Packages of *spdep* and *maptools* were also used for the application to read shape file and to draw choropleth map. In the application user can select municipality, year, sex, type, age and so on, then choropleth map is drawn (Figure 1) and cross tabulation of the municipality is calculated and shown in "Table" tab (Figure 2).

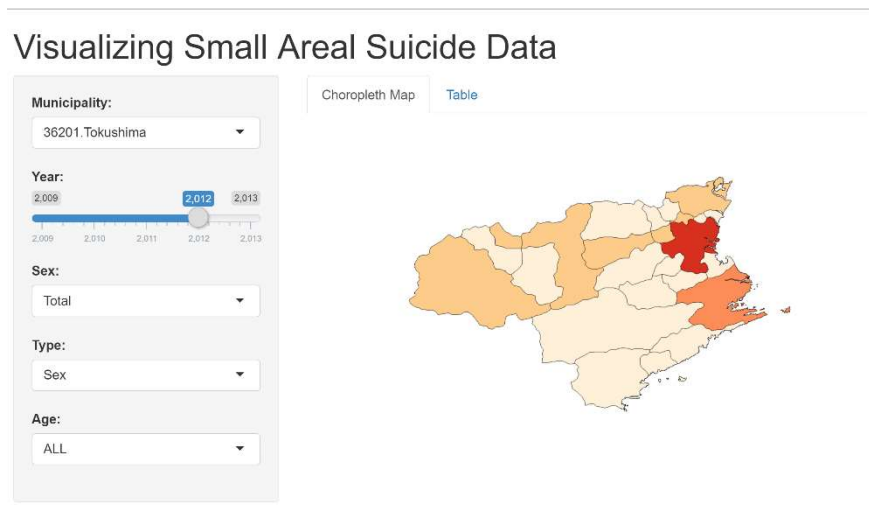


Figure 1: Choropleth map of area suicide data of Tokushima prefecture by municipalities in 2012.

Visualizing Small Areal Suicide Data

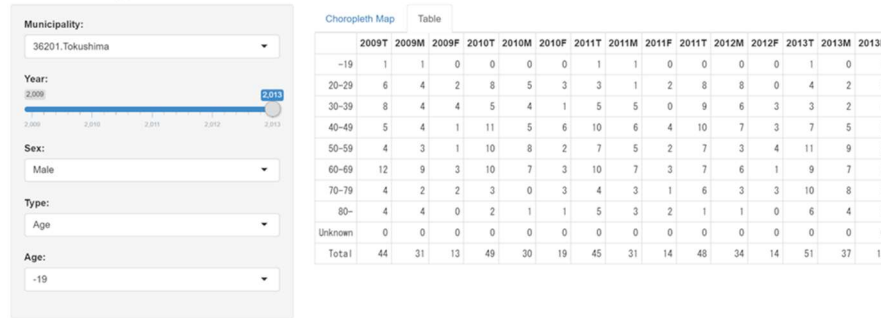


Figure 2: Cross tabulation of suicide data at Tokushima city (municipality) of Tokushima prefecture in 2012.

3.2 Exploratory

An exploratory data analysis was applied by *Exploratory* (Exploratory, Inc., 2016). *Exploratory* is one desktop app for data wrangling, visualization, and advanced analytics, and it is powered by R. Figure 3 shows stacked bar chart of the number of suicide in categories of age groups and sex (top left), line chart of monthly suicide number in classified by sex (top right), stacked bar chart of the number of suicide in categories of municipalities and sex (bottom left), and heat map of the number of suicide in categorized of municipalities and age groups (bottom right).

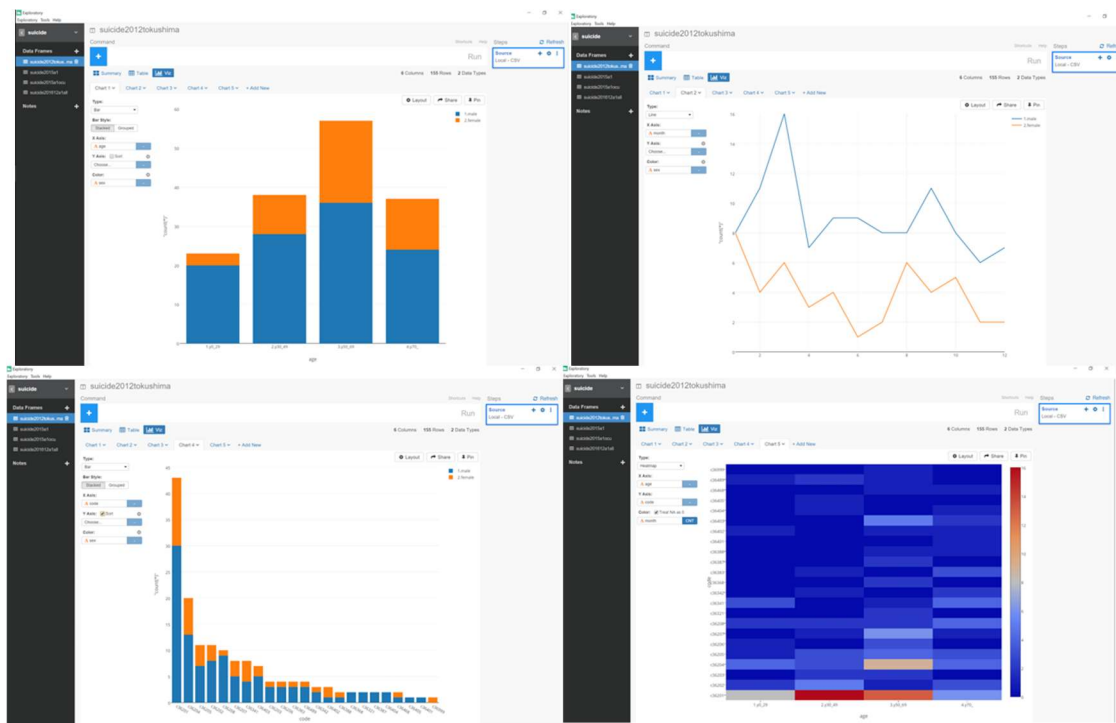


Figure 3: Hard copy of the results of *Exploratory*; Stacked bar chart of age groups and sex (top left), line chart of month and sex (top right), stacked bar chart of municipalities and sex (bottom left), and heat map of municipalities and age groups (bottom right).

4 Discussions and summary

The application for visualization of small area especially municipality divided suicide data was implemented. From the result of choropleth map, it was found Tokushima city had the highest number of suicide, and in south area there were municipalities which had relatively low number of suicide for total number of suicide in Tokushima prefecture in 2012. The application also provided the selected cross tabulation to user. This is municipality version of cross tabulation for suicide data, although prefecture version of the data is available in the Cabinet Office (2016).

From the result of exploratory data analysis in Figure 3, the number of suicide of age group fifty to sixty-nine years old was the highest in age group and male was slightly higher than female. In month trend, March and September were higher and in the group male was more than double of female in each month. Both months are the end of fiscal year, therefore to find out the relations between these facts will be future studies.

From the result of area cross tabulation, the number of suicide of male in Tokushima city and age group fifty to sixty-nine years old are higher than other municipalities, sex and age groups. The fact will be applied to regional specific policy making.

Suicide risk is not only calculated by the number of suicide, but also it should be calculated by suicide rate or the other factors, therefore, for the future studies, the other data such as population should be connected to the suicide data and also exploratory data analysis should be performed.

Acknowledgment

This work was partly supported by KAKENHI (26330052; Grant-in-Aid for Scientific Research(C)).

References

- [1] Cabinet Office (2016), Statistics of Suicide. Suicide Person Based on Suicidal Statistics (in Japanese), URL: <http://www8.cao.go.jp/jisatsutaisaku/toukei/>
- [2] Exploratory, Inc. (2016), Exploratory, URL: <https://exploratory.io>
- [3] Fujita, T. (2009), Statistics of Community for the Death from Suicide. National Institute of Mental Health, National Center of Neurology and Psychiatry, Japan.
- [4] Hijmans, R. (2016), GADM database of Global Administrative Areas, URL: <http://www.gadm.org>
- [5] Kubota, T. and Tsubaki, H. (2014), Visualization for reason-specified suicide data in Japan, The 21st International Conference on Computational Statistics (COMPSTAT2014).
- [6] Kubota, T., Iizuka, M. and Tsubaki, H. (2015), Visualization for spatial and paneled data of reason-specified suicide data by prefecture in Japan, the 60th International Statistical Institute World Statistics Congress.
- [7] Ministry of Land, Infrastructure, Transport and Tourism (2014), National Land Numerical Information download service, National Land Information Division, National Spatial Planning and Regional Policy Bureau, MLIT of Japan, URL: <http://nlftp.mlit.go.jp/ksj-e/index.html>
- [8] RStudio, Inc. (2016), Shiny by RStudio A web application framework for R, URL: <http://shiny.rstudio.com/>
- [9] Statistics Bureau (2010), National Population Census., URL: <http://www.stat.go.jp/data/kokusei/2010/index.htm>