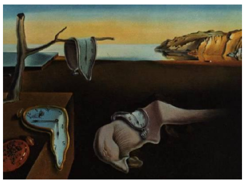


Journal of Statistical Software

November 2017, Volume 82, Issue 6. *doi: 10.18637/jss.v082.i06*

vdmR: Generating Web-Based Visual Data Mining Tools with R

Tomokazu Fujino
Fukuoka Women's University

Abstract

The **vdmR** package generates web-based visual data mining tools by adding interactive functions to **ggplot2** graphics. Brushing and linking between multiple plots is one of the main features of this package. Currently, scatter plots, histograms, parallel coordinate plots, and choropleth maps are supported in the **vdmR** package. In addition, identification on a plot is supported by linking the plot and the data table.

Keywords: visual data mining, SVG, interactive, choropleth map, R.

1. Motivation

Tukey (1977) stated that *"Exploratory data analysis is detective work – numerical detective work or counting detective work or graphical detective work."* Our interpretation is that graphical detective work (the process of finding knowledge from datasets by using statistical graphics) is a part of exploratory data analysis. We call this process "visual data mining". On the other hand, Wegman (2003) stated that visual data mining is an extension of exploratory data analysis and the difference between these two topics resides in the size and dimensionality of the datasets. Our **vdmR** package (Fujino 2017) takes its name from visual data mining, but currently it cannot deal with large scale datasets which have roughly 10 thousands or more records because of a limitation of the implementation of the package and the rendering capability of a web browser. We would like to improve the package capabilities in a future release.

Visual data mining plays an important role in the first step of exploratory data analysis. Visual data mining combines data mining and information visualization techniques, especially techniques using statistical graphics. An important feature of visual data mining is the interactivity of sophisticated statistical graphics. **GGobi** (Swayne, Temple Lang, Buja, and Cook 2003) is one of the successful implementations of visual data mining tools that include interactive and dynamic statistical graphics. When we think about interactive statistical

Feature/software	**vdmR**	**ggvis**	**GGobi**	**Mondrian**	**iplots**	**animint**
Labeling	✓		✓	✓		✓
Hovering	✓	✓	✓	✓		✓
Brusing	✓		✓	✓		
Painting (persistent brushing)	✓		✓	✓		
MLV	✓		✓	✓	✓	✓
Link with data table	✓		✓			
HTML output	✓	✓				✓
Link with choropleth maps	✓			✓		✓
R or standalone (SA)	R	R	SA	SA	R	R

Table 1: Comparison of statistical software and packages supporting interactive graphics.

graphics, we should refer to this software because it implements many kinds of interactive and dynamic features. If we restrict ourselves to only the interactive features, **Mondrian** (Theus 2002) is also a good model. The **vdmR** package was developed to generate visual data mining tools from R (R Core Team 2017), which include functionalities that are difficult to realize in **GGobi**. The key functionalities such as brushing, identification, and linking between basic statistical graphics are implemented in the visual data mining tool generated by the **vdmR** package.

Because **GGobi** was developed as standalone software, we need to install **GGobi** into the local PC environment to use its functionalities. Therefore, we cannot use **GGobi** in a web-based system. Moreover, plotting locations such as longitude and latitude of the observations in a dataset on a map and drawing choropleth maps are not included in **GGobi**, although such functionalities are useful for understanding the spatial trends in data. The **vdmR** package can handle these functionalities.

The **ggplot2** package (Wickham 2009) is widely used for creating statistical graphics in R. It is available from the Comprehensive R Archive Network (CRAN) and listed as a core package in the CRAN Task View on Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization (Lewin-Koh 2015). A lot of contributed packages depend on and import the **ggplot2** package for plotting. Two reasons that **ggplot2** is broadly used are as follows: it provides simple and beautiful plots even with simple commands and the grammar for generating **ggplot2** graphics is clear. The grammar describes the kinds of plots and variables that are mapped to the aesthetic attributes in a plot. The **vdmR** package first generates **ggplot2** graphics and then converts them into the Scalable Vector Graphics (SVG) format while maintaining structures such as points, lines, and polygons by using the **gridSVG** package (Murrell and Potter 2017).The **ggplot2** graphics are generated by using functions of the **grid** package, which are called for editing a structure of the graphics and adding interactive functions written in JavaScript.

The **ggvis** (Chang and Wickham 2016) package, which also provides statistical graphics with interactive functionalities, was developed as an extension of **ggplot2**. Although **ggvis** seems to follow the same concepts as **vdmR** for generating web-based statistical graphics, the grammatical concepts of **ggvis** are closer to **ggplot2** than to **vdmR**. For example, **ggvis** supports functions for adding new layers, which can be connected by pipe operators, whereas **vdmR** directly creates graphics with linked brushing (linking + brushing) and choropleth maps. The collection of these kinds of statistical graphics is referred to as “multiple linked views (MLV)”

Package	Function in **vdmR** package
ggplot2	Generating basic statistical plots
grid	Modifying **ggplot2** graphics
gridSVG (Murrell and Potter 2017)	Exporting **ggplot2** graphics to SVG
GGally (Schloerke *et al.* 2017)	Generating **ggplot2** based parallel coordinate plots
maptools (Bivand and Lewin-Koh 2017)	Reading spatial object from shape files
plyr (Wickham 2011)	Manipulating vectors for aesthetic mappings
dplyr (Wickham *et al.* 2017)	Manipulating data frame
broom (Robinson 2017)	Converting spatial object to data frame
rjson (Couture-Beil 2014)	Converting R object to JSON and embedding it in JavaScript
Rook (Horner 2014)	Launching local web server

Table 2: Package dependencies of **vdmR**.

(Buja, Cook, and Swayne 1996). Currently, **ggvis** does not support functions such as multiple linked views and linking with data tables. Table 1 shows a comparison of some of the standalone software packages and R packages that support interactive statistical graphics. In the table, the "labeling" and "hovering" means a function that shows attributes of the specified point by mouse operation. "Hovering" means displaying them transiently while the mouse cursor is on the specified point. "Labeling" means displaying them persistently when a mouse button is clicked on the specified point.

The dependencies of **vdmR** on other packages are listed in Table 2. The rest of this paper illustrates the details of the **vdmR** package.

2. Outline of usage

2.1. Installation and sample dataset

The **vdmR** package is available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=vdmR and can be installed from CRAN and loaded with the following commands.

```
R> install.packages("vdmR")
R> library("vdmR")
```

Several packages are required for **vdmR**, and all are installed at the same time if not installed previously. In the lines following the commands, **vdmR** provides a sample dataset for demonstration and illustration of the package.

```
R> data("vsfuk2012", package = "vdmR")
R> head(vsfuk2012[, 1:5])

  CityCode CityName Type FertilityRate MortalityRate
1    40133   Chuuou Ward           8.9           6.0
2    40132   Hakata Ward          10.4           6.9
```

```
3    40134   Minami Ward            9.9            7.3
4    40137   Sawara Ward            9.9            7.0
5    40131  Higashi Ward           10.5            7.1
6    40135    Nishi Ward           10.2            7.4
```

The `vsfuk2012` dataset gives the vital statistics for the Fukuoka prefecture of Japan from 2008 to 2012. The dataset consists of 72 rows and 17 columns. Each row indicates the result of one municipality of the Fukuoka prefecture. Each column shows one type of basic information of the municipalities such as code, name, or type (city, town, or village) or index such as population, fertility rate, or mortality rate. Information of all variables is given in the help document of the dataset.

2.2. Creating VDM tools

VDM tools from **vdmR** are created in the following two steps.

- Create each statistical plot in HTML files with SVG and JavaScript.
- Create the main window for launching each plot and open the window.

Here we show an example for generating two plots.

```
R> vscat(MortalityRate, FertilityRate, vsfuk2012, "scat01", "vsfuk2012")
R> vhist(MarriageRate, vsfuk2012, "hist01", "vsfuk2012")
```

Each command generates files in temporary directory for which the location can be changed by the argument of `path`. The files are named according to the following rule:

{*plot name*}.{*tag name*}.(`svg`.){*extension*}

All HTML files include the `<embed>` tag for the related SVG file, which contains graphical information, such as axes and sizes of the shapes in the plot. When opening an HTML file with a web browser, the plot appears. Note that the interactive functions are not enabled at this stage. To enable the interactive functions, the user needs to generate the "launcher" with the following command.

```
R> vlaunch(vsfuk2012, "main", "vsfuk2012")
```

If the web server for the R help is running, the command returns the following message.

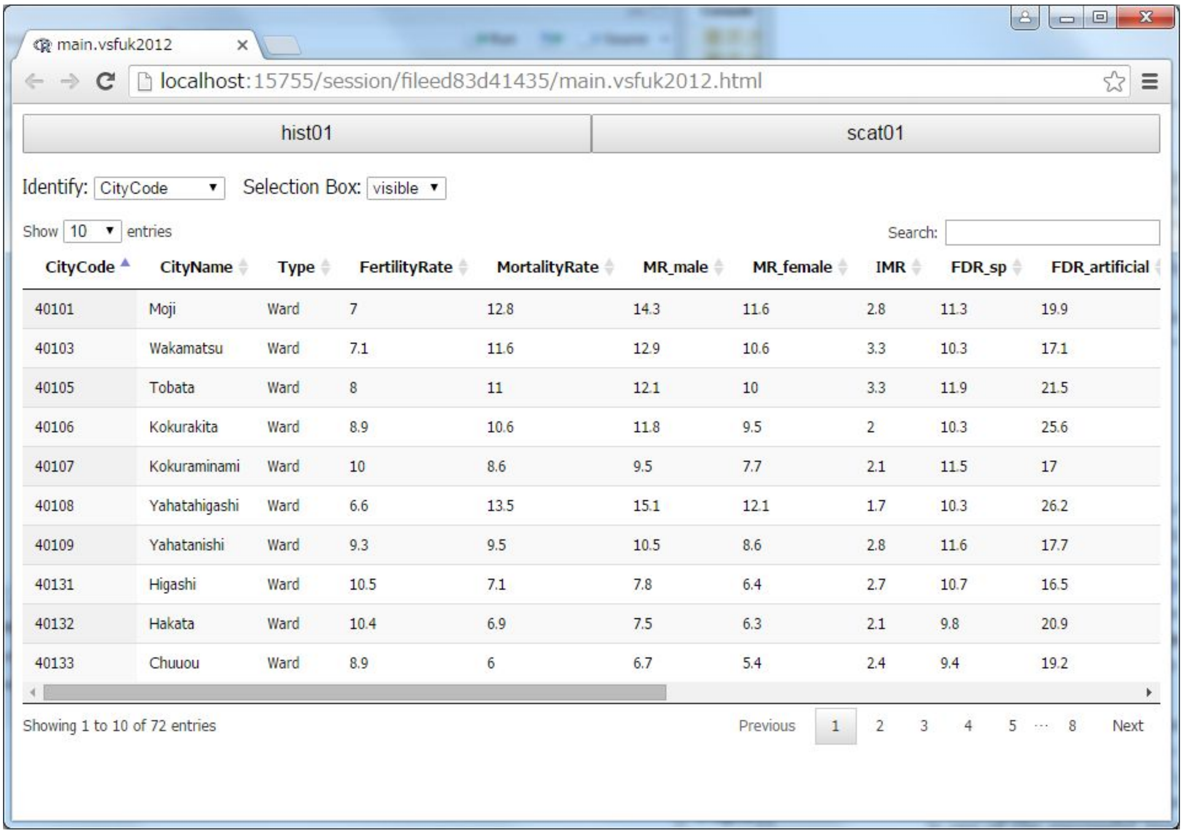
```
Error in startDynamicHelp(TRUE): server already running
```

In this case, the following command stops the server before the `vlaunch()` function.

```
R> tools::startDynamicHelp(FALSE)
```

The `vlaunch()` function generates an HTML file named in the following format.

{*launcher name*}.{*tag name*}.`html`



Identify: CityCode ▾ Selection Box: visible ▾

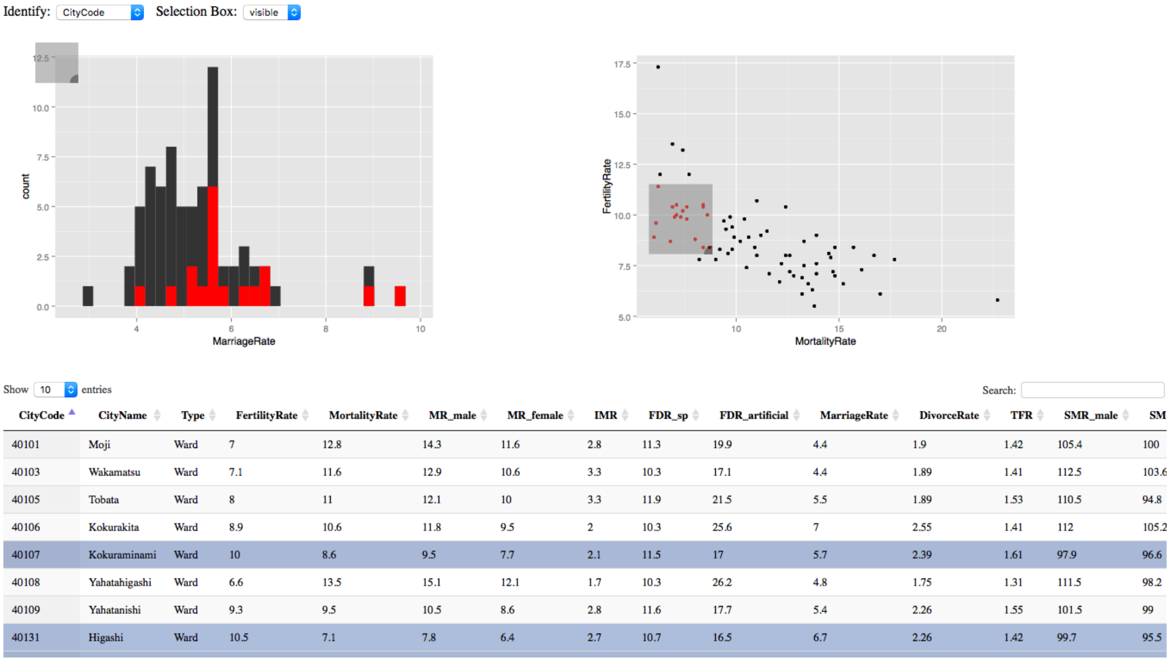
Show 10 ▾ entries

Search:

CityCode	CityName	Type	FertilityRate	MortalityRate	MR_male	MR_female	IMR	FDR_sp	FDR_artificial
40101	Moji	Ward	7	12.8	14.3	11.6	2.8	11.3	19.9
40103	Wakamatsu	Ward	7.1	11.6	12.9	10.6	3.3	10.3	17.1
40105	Tobata	Ward	8	11	12.1	10	3.3	11.9	21.5
40106	Kokurakita	Ward	8.9	10.6	11.8	9.5	2	10.3	25.6
40107	Kokuraminami	Ward	10	8.6	9.5	7.7	2.1	11.5	17
40108	Yahatahigashi	Ward	6.6	13.5	15.1	12.1	1.7	10.3	26.2
40109	Yahatanishi	Ward	9.3	9.5	10.5	8.6	2.8	11.6	17.7
40131	Higashi	Ward	10.5	7.1	7.8	6.4	2.7	10.7	16.5
40132	Hakata	Ward	10.4	6.9	7.5	6.3	2.1	9.8	20.9
40133	Chuuou	Ward	8.9	6	6.7	5.4	2.4	9.4	19.2

Showing 1 to 10 of 72 entries

Figure 1: Main window generated by the `vlaunch()` function.



Identify: CityCode Selection Box: visible

Show 10 entries

Search:

CityCode	CityName	Type	FertilityRate	MortalityRate	MR_male	MR_female	IMR	FDR_sp	FDR_artificial	MarriageRate	DivorceRate	TFR	SMR_male	SM
40101	Moji	Ward	7	12.8	14.3	11.6	2.8	11.3	19.9	4.4	1.9	1.42	105.4	100
40103	Wakamatsu	Ward	7.1	11.6	12.9	10.6	3.3	10.3	17.1	4.4	1.89	1.41	112.5	103.6
40105	Tobata	Ward	8	11	12.1	10	3.3	11.9	21.5	5.5	1.89	1.53	110.5	94.8
40106	Kokurakita	Ward	8.9	10.6	11.8	9.5	2	10.3	25.6	7	2.55	1.41	112	105.2
40107	Kokuraminami	Ward	10	8.6	9.5	7.7	2.1	11.5	17	5.7	2.39	1.61	97.9	96.6
40108	Yahatahigashi	Ward	6.6	13.5	15.1	12.1	1.7	10.3	26.2	4.8	1.75	1.31	111.5	98.2
40109	Yahatanishi	Ward	9.3	9.5	10.5	8.6	2.8	11.6	17.7	5.4	2.26	1.55	101.5	99
40131	Higashi	Ward	10.5	7.1	7.8	6.4	2.7	10.7	16.5	6.7	2.26	1.42	99.7	95.5

Figure 2: Inline frame version of a main window.

Some related files (`.js`, `.css`) are also generated. The above command then launches the default web browser and displays the HTML file (Figure 1). When `TRUE` is specified in the argument `iframe`, all of the plot windows are displayed in the main window by inline frame. (Figure 2).

```
R> vlaunch(vsfuk2012, "main", "vsfuk2012", iframe = TRUE)
```

The current VDM tool is supported in the latest versions of Google Chrome, Mozilla Firefox, and Safari, but not the Internet Explorer.

2.3. Manipulating VDM tools

When clicking one of the buttons located on the upper part of the main window, a corresponding plot window is opened (Figures 3, 4). The user can see the gray rectangle on the top left of the window. By moving this rectangle selection tool in the case of the scatter plot, the data points in the rectangle region are highlighted, and then a histogram of the selected data is drawn in the histogram window in highlighted color. In addition, the corresponding rows in the data table of the main window are highlighted (Figure 5).

The **vdmR** package supports persistent selection, which means that once objects such as points in the scatter plot are selected, the selected objects continue to be selected even if these objects are out of the region of the moved selection tool. In our VDM tools, double-clicking the selection tool enables the persistent selection mode. When double-clicking again, the tool returns to the normal (temporary) selection mode (Figure 6).

The selection tool is also resizable. When dragging the mouse to the bottom right of the selection tool, the region is resized. If a user wants to hide the selection tool, the second select form of the main window can be used to switch the visibility between “visible” and “hidden”.

When the mouse cursor is placed on a symbol, such as a point, line, or polygon corresponding to each data value, the value of the variable specified by the select form in the main window is displayed near the mouse cursor (hovering). Moreover, by clicking on the symbol, the display of the value remains (identifying).

A sortable table of the dataset is displayed in the main window. When clicking one of the variable labels, the rows are sorted based on descending or ascending order of the corresponding variable.

2.4. Interacting with R

When executing a copy command (pressing [Ctrl]+[c] in Windows) after selecting a subset of the dataset, the selected data are copied to the clipboard. Then, the data can be used for further analysis in R, as follows.

```
R> vsfuk2012.sub <- read.table("clipboard", header = TRUE)
```

MacOS users need to specify `pipe("pbpaste")` instead of `"clipboard"`. Currently, this feature is available only for Chrome and Firefox on Windows and MacOS, not on Linux and Safari on MacOS.

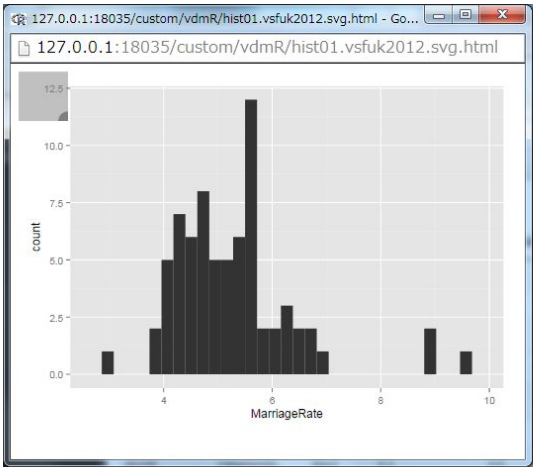

127.0.0.1:18035/custom/vdmR/hist01.vsfuk2012.svg.html - Go...
127.0.0.1:18035/custom/vdmR/hist01.vsfuk2012.svg.html
12.5
10.0
7.5
5.0
2.5
0.0
count
4
6
8
10
MarriageRate

Figure 3: Histogram window generated by `vhist()` function.

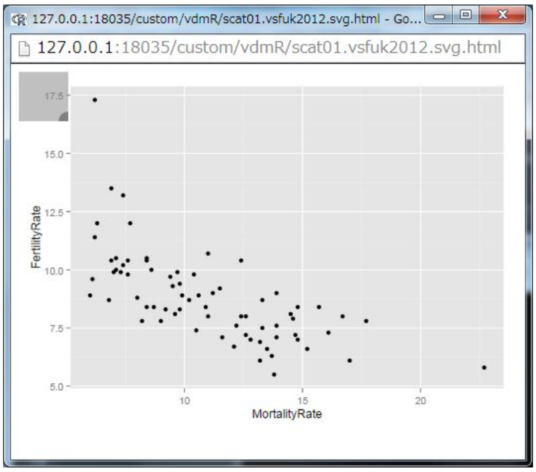
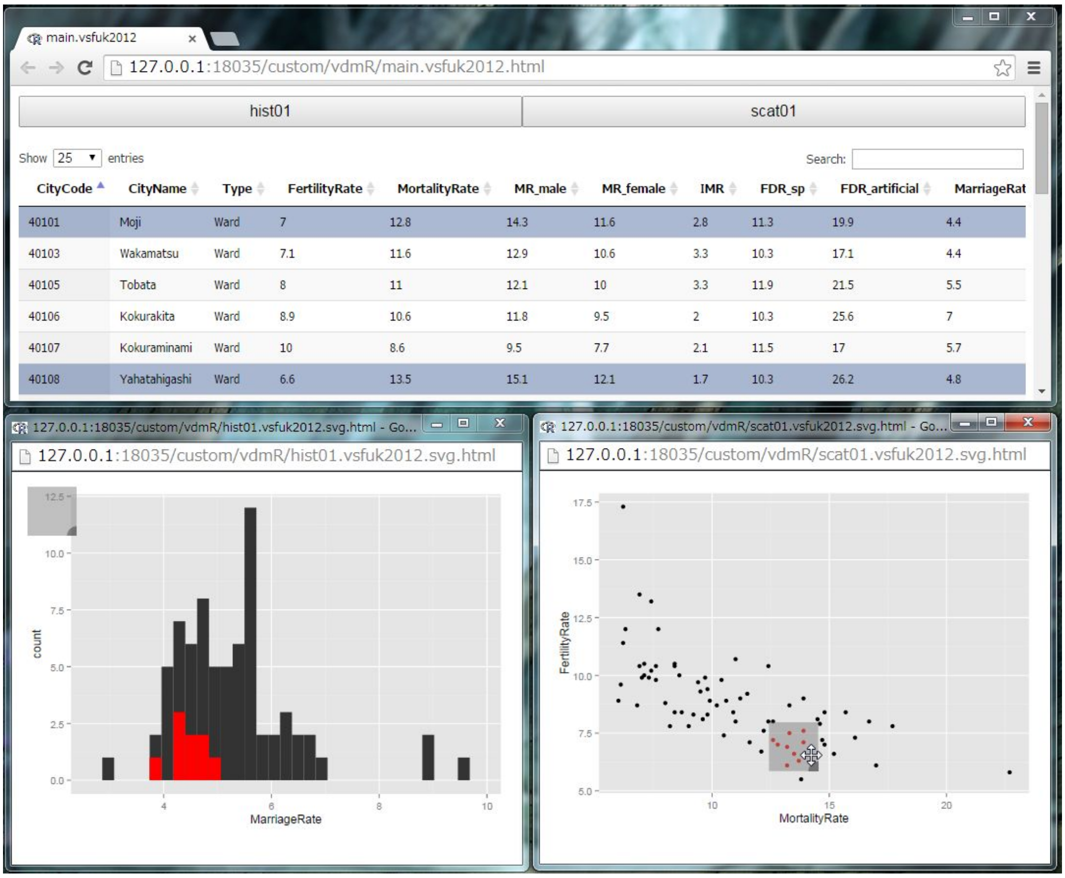

127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html - Go...
127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html
17.5
15.0
12.5
10.0
7.5
5.0
FertilityRate
10
15
20
MortalityRate

Figure 4: Scatter plot window generated by `vscat()` function.



CityCode	CityName	Type	FertilityRate	MortalityRate	MR_male	MR_female	IMR	FDR_sp	FDR_artificial	MarriageRat
40101	Moji	Ward	7	12.8	14.3	11.6	2.8	11.3	19.9	4.4
40103	Wakamatsu	Ward	7.1	11.6	12.9	10.6	3.3	10.3	17.1	4.4
40105	Tobata	Ward	8	11	12.1	10	3.3	11.9	21.5	5.5
40106	Kokurakita	Ward	8.9	10.6	11.8	9.5	2	10.3	25.6	7
40107	Kokuraminami	Ward	10	8.6	9.5	7.7	2.1	11.5	17	5.7
40108	Yahatahigashi	Ward	6.6	13.5	15.1	12.1	1.7	10.3	26.2	4.8

127.0.0.1:18035/custom/vdmR/hist01.vsfuk2012.svg.html - Go...
127.0.0.1:18035/custom/vdmR/hist01.vsfuk2012.svg.html
12.5
10.0
7.5
5.0
2.5
0.0
count
4
6
8
10
MarriageRate
127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html - Go...
127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html
17.5
15.0
12.5
10.0
7.5
5.0
FertilityRate
10
15
20
MortalityRate

Figure 5: Example of multiple linked views.

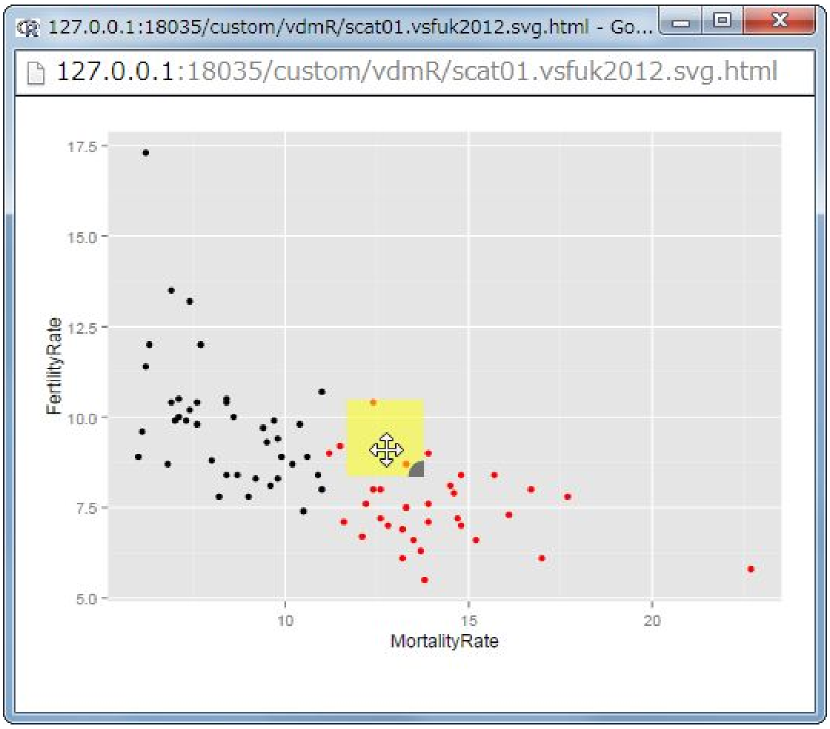

127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html - Go...
127.0.0.1:18035/custom/vdmR/scat01.vsfuk2012.svg.html
17.5
15.0
12.5
10.0
7.5
5.0
FertilityRate
10
15
20
MortalityRate

Figure 6: Example of persistent selection in a scatter plot.

3. Details of vdmR

3.1. Data table and launcher of other plots: `vlaunch()`

The `vlaunch()` function generates a main window that opens each pre-generated window, which includes a statistical plot with interactivity. First, by using regular expressions based on the `tag` argument, the function finds the pre-generated HTML and SVG files in the current working directory. In the example in Section 2.2, the `vlaunch()` function finds HTML files and an SVG file of the statistical plot. The name of the SVG file is in the format `*.vsfuk2012.svg.html`. Then, the function generates buttons for opening these files in the HTML file. All of the plot windows should be opened from the main window, because the multiple linked views (linked-brushing) can be implemented by using cross-document messaging through the main window. Cross-document messaging is one of the HTML5 technologies. However, cross-document messaging does not work correctly by opening HTML files directly in Chrome for security reasons. Thus, in the **vdmR** package, all graphics and data tables are displayed through the local web server, which is launched by the **Rook** package. The data table in the main window is implemented by **DataTables**, which is a plug-in for the **jQuery** JavaScript library. **DataTables** provides many useful functions for data tables in web pages, such as pagination, instant search, and multi-column ordering.

3.2. Scatter plot: `vscat()`

The `vscat()` function generates an interactive scatter plot in **ggplot2** style. The first two

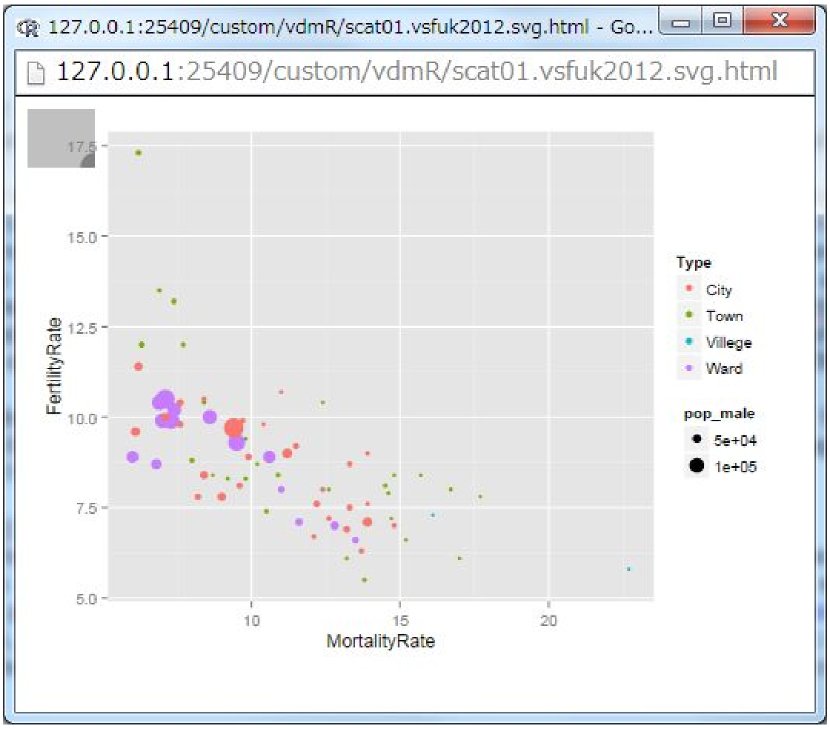

127.0.0.1:25409/custom/vdmR/scat01.vsfuk2012.svg.html - Go...
127.0.0.1:25409/custom/vdmR/scat01.vsfuk2012.svg.html
FertilityRate
17.5
15.0
12.5
10.0
7.5
5.0
10
15
20
MortalityRate
Type
City
Town
Villege
Ward
pop_male
5e+04
1e+05

Figure 7: Scatter plot with aesthetic mapping.

arguments `x, y` are the column names in the data frame given by the third argument of the function. The columns are mapped into the x-axis and the y-axis of the scatter plot, respectively. The dot argument (`...`) is passed to the `aes()` function of **ggplot2**. Thus, in the `vsfuk2012` dataset, each city's male population size corresponds to its point size on the scatter plot, and each city's type corresponds to its color on the plot by the following script:

```
R> vscat(MortalityRate, FertilityRate, vsfuk2012, "scat01", "vsfuk2012",
+     color = Type, size = pop_male)
R> vlaunch(vsfuk2012, "main", "vsfuk2012")
```

The output of this `vscat()` function (with aesthetic mapping) is shown in Figure 7. The user can set the color of all points to a specific color by setting the argument of the `I()` function to the character string of the color name or the color code, as in the following example. Note that it does not work without the `I()` function because the user selected color differs from the aesthetic mapping.

```
R> vscat(MortalityRate, FertilityRate, vsfuk2012, "scat01", "vsfuk2012",
+   color = I("darkgreen"), size = pop_male)
```

This grammar for the color setting is the same style used in the `qplot()` function of **ggplot2**.

3.3. Histogram: `vhist()`

The `vhist()` function generates an interactive histogram in **ggplot2** style. The arguments of the function are almost the same as those for the `vscat()` function except that the variable `y` does not exist. A user can add the following code to specify the color of the histogram:

```
R> vhist(MarriageRate, vsfuk2012, "hist01", "vsfuk2012",
+    fill = I("darkgreen"), color = I("black"))
R> vlaunch(vsfuk2012, "main", "vsfuk2012")
```

A histogram visualizes the frequency count table of the specific variable of the dataset. Thus, the graphic objects (bars of the histogram) do not have a one-to-one correspondence with a row of the dataset. This makes it more difficult to implement the linked-brushing function. We solved this by embedding the mapping information between each row of data and a histogram bar into the generated SVG files in JSON (JavaScript object notation) format by using the **rjson** package (Couture-Beil 2014).

3.4. Parallel coordinate plot: `vpcp()`

A parallel coordinate plot (Inselberg 1985; Wegman 1990) is well known as a powerful tool for visualizing multivariate data. The `vpcp()` function can generate a parallel coordinate plot with interactive facilities in **ggplot2** style. In **ggplot2**, the `ggpcp()` function can draw the parallel coordinate plot; however, the function is deprecated. Thus, we used the `ggparcoord()` function provided by the **GGally** package (Schloerke *et al.* 2017) in the `vpcp()` function, and so most of the arguments of the `vpcp()` function are the same as those of the `ggparcoord()` function. The example shown in Figure 8 is created with the following code.

```
R> vpcp(vsfuk2012, 4:17, "pcp1", "vsfuk2012", groupColumn = "Type",
+    scale = "uniminmax", missing = "min10")
R> vlaunch(vsfuk2012, "main", "vsfuk2012")
```

Note that the argument of `groupColumn` requires the character string of the column name. Therefore, double quotation marks are needed. Moreover, the argument of `alphaLines` can be used for an alpha-blending value of the lines in a pallalel coordinate plot, as follows:

```
R> vpcp(vsfuk2012, 4:17, "pcp1", "vsfuk2012", alphaLines = 0.1,
+    scale = "uniminmax", missing = "min10")
R> vlaunch(vsfuk2012, "main", "vsfuk2012")
```

The output is shown in Figure 9.

3.5. Choropleth map: `vcmap()`

The interactive version of a choropleth map, in which polygons of the areas are painted according to the value of the specific variable, is one of the central features of the **vdmR** package. By using a choropleth map in multiple linked views, it is easy to see the relationship between the spatial characteristics and the multivariate characteristics. The interactive linked micromap plot (Symanzik and Carr 2008) is one of the interactive applications of the choropleth map.

To generate the interactive choropleth map in **vdmR**, the user has to prepare the ESRI shapefile, including the region for the dataset. The shapefile has an attribute table with the common ID column of the data frame. The **vdmR** package provides a sample shapefile related to the data `vsfuk2012`. By using the **maptools** package (Bivand and Lewin-Koh 2017), it is possible to import the shapefile into the R environment as a '`SpatialPolygonsDataFrame`' object, as follows:

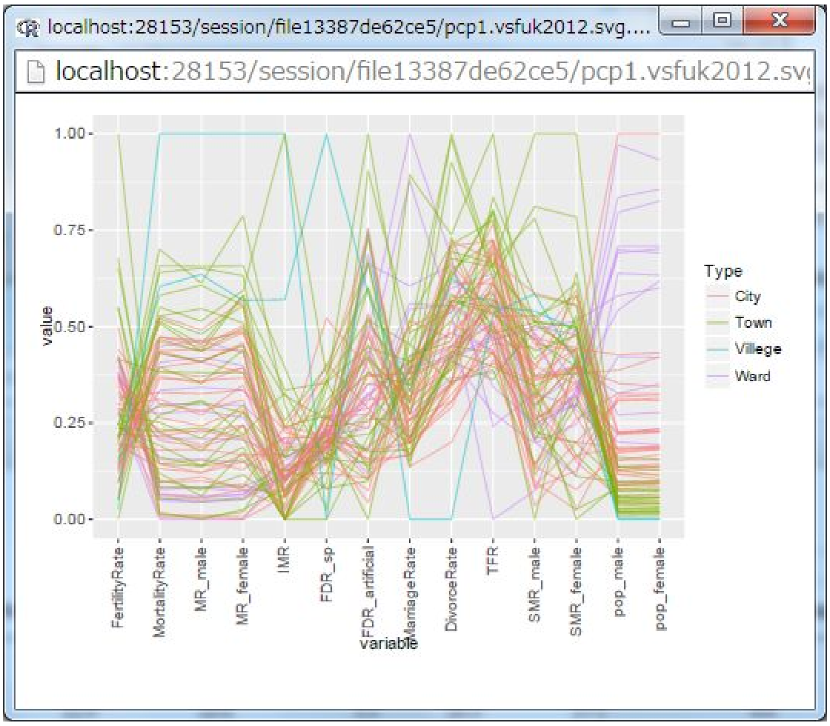


Figure 8: Parallel coordinate plot generated by `vpcp()` function.

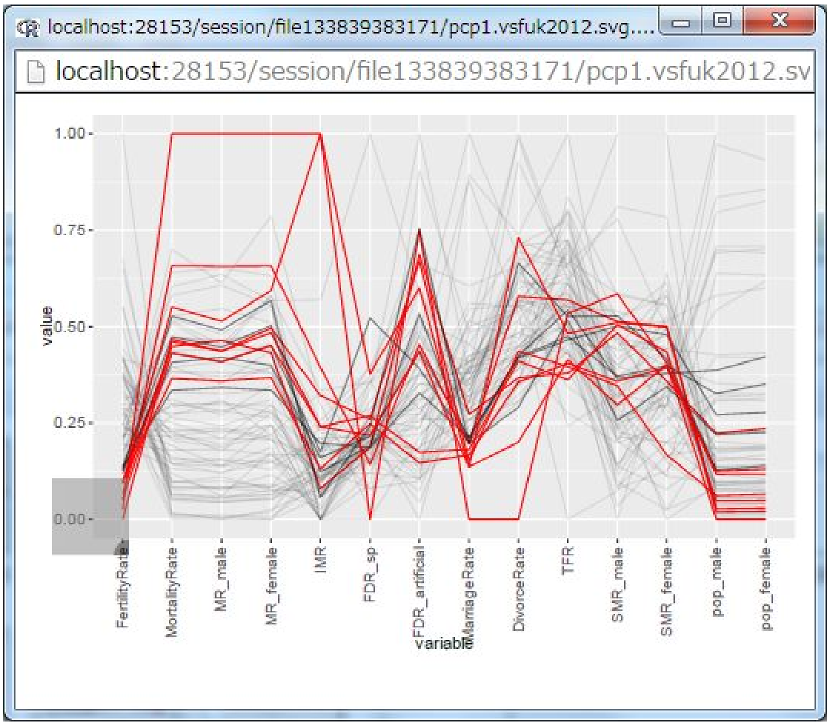


Figure 9: Alpha-blending and brushing in pallalel coordinate plot.

```
R> library("maptools")
R> shp.path <- file.path(system.file(package = "vdmR"),
+    "etc/shapes/fukuoka2012.shp")
R> vsfuk2012.spdf <- readShapeSpatial(shp.path, IDvar = "CityCode")
R> head(vsfuk2012.spdf@data)
```

```
      CityCode      CityName Type
40101    40101          Moji Ward
40103    40103     Wakamatsu Ward
40105    40105        Tobata Ward
40106    40106    Kokurakita Ward
40107    40107  Kokuraminami Ward
40108    40108 Yahatahigashi Ward
```

If the shapefile does not have the common ID column of the data frame, the user needs to edit the shapefile by using desktop GIS software, such as **QGIS** or **ArcGIS**, or else edit the data frame.

The `vcmap()` function provides the interactive choropleth map in the **vdmR** package. For example, the following code generates the interactive choropleth map of the fertility rate in the Fukuoka prefecture:

```
R> frcol <- scale_fill_gradient2(low = "blue", mid = "white", high = "red",
+    midpoint = median(vsfuk2012$FertilityRate))
R> vcmap(shp.path, vsfuk2012, "CityCode", "CityCode", "map1", "vsfuk2012",
+    fill = FertilityRate, ggscale = frcol)
```

Figure 10 shows the output of the choropleth map. The first and second arguments state the path to the shapefile and the data frame, respectively. The third and fourth arguments state the column names of the common ID for the attribute table and the data frame, respectively. The argument `fill` states the column name assigned to the color of polygons. If the user needs to use a specific color scale, the color scale generated by the `scale_fill_*()` function has to be passed to the argument `ggscale`.

In the `vcmap()` function, the shape file is imported from the specified directory as a spatial object, and then the function converts the object to a data frame by using the `tidy()` function defined in the **broom** package. By using this data frame, the `ggplot()` and the `geom_polygon()` functions draw a choropleth map.

The brushing operation is slightly different from that of other **vdmR** outputs. The select box does not appear on the choropleth map, so the user has to brush the polygons directly with the mouse pointer. The operation on the choropleth map is always a persistent selection. Double-clicking out of a region of the polygons resets all of the selections.

4. Application to web-based systems

Fujino (2007) proposed a new framework for WebGIS by using R and SVG. The **vdmR** package makes it easier to implement such a web-based system by including VDM tools. We developed the web interface of the **vdmR** package by **rApache** (Horner 2013) and the **brew**

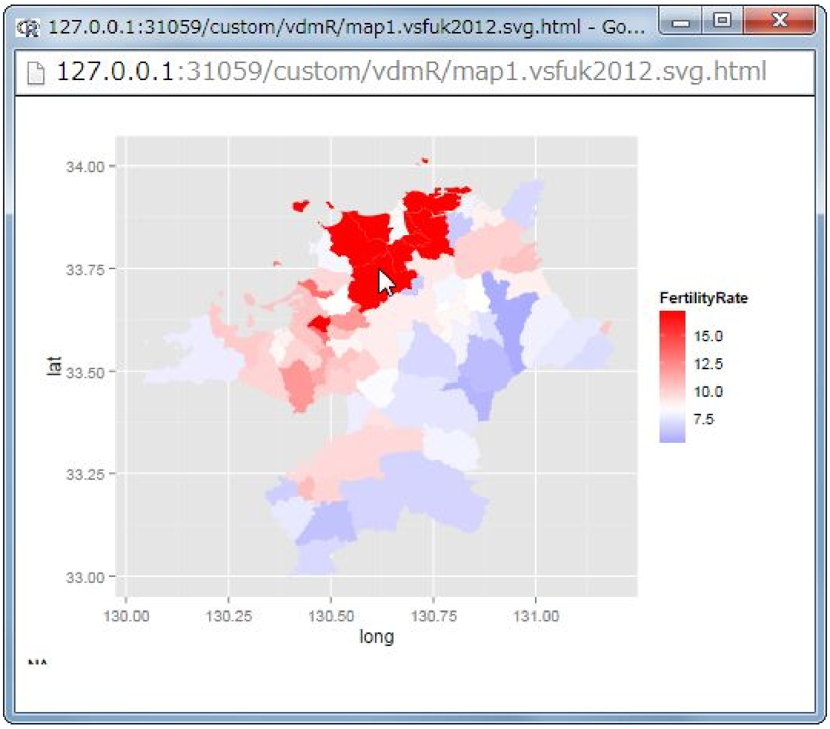

127.0.0.1:31059/custom/vdmR/map1.vsfuk2012.svg.html - Go...
127.0.0.1:31059/custom/vdmR/map1.vsfuk2012.svg.html
34.00
33.75
33.50
33.25
33.00
lat
130.00
130.25
130.50
130.75
131.00
long
FertilityRate
15.0
12.5
10.0
7.5

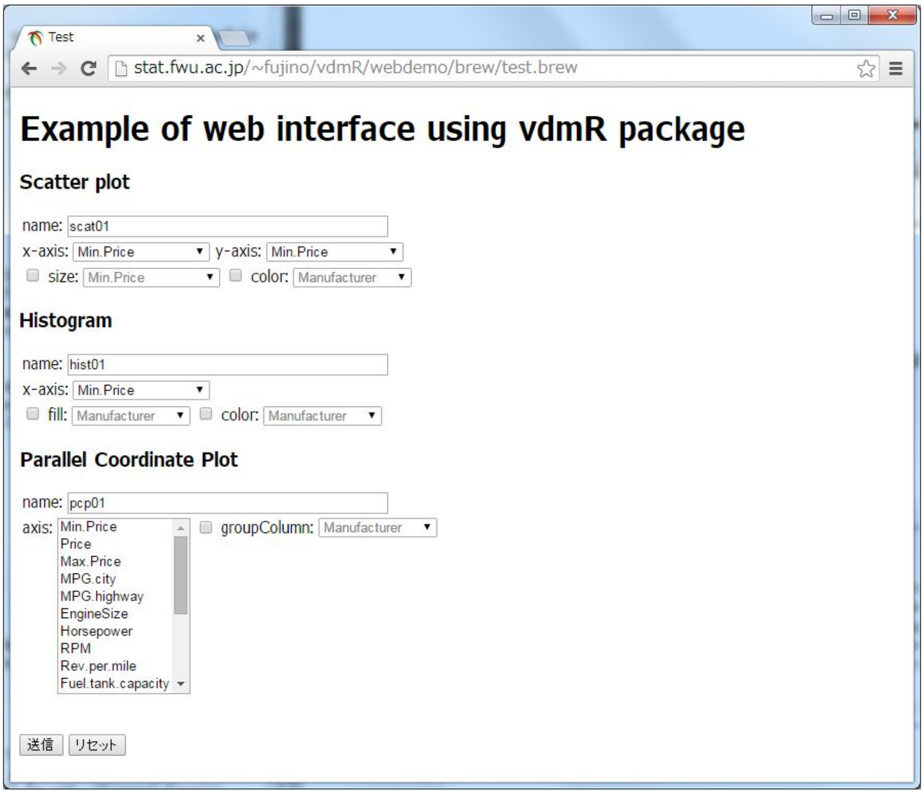
Figure 10: Choropleth map generated by `vcmap()` function.

package (Horner 2011) for a simple demonstration. This site is available on the following URL: `http://stat.fwu.ac.jp/~fujino/vdmR/webdemo/`. This site has three files (`index.html`, `test.brew`, `test.R`) and one temporary directory. First, a user chooses one of the datasets, which are familiar to R users, from the selection form (`index.html`). Or, the dataset can be extracted from the database for practical use. The `test.brew` file consists of a mixture of HTML and embedded R scripts. The **brew** package allows R to run these scripts and then overwrite them in the output of the results. The following scripts are part of `test.brew`.

```
x-axis: <select name="scx">
<%
dfm <- eval(parse(text=POST$dataset))
cn.num <- names(dfm)[sapply(dfm,class)%in%c("numeric","integer")]
cn.cat <- setdiff(names(dfm), cn.num)
%>
<%= paste0("<option value=\"",cn.num,"\">",cn.num,"</option>") %>
</select>
```

As shown in the above code, `test.brew` provides the form for setting parameters and variable names in the dataset, which is specified in the form of `index.html`. Then, `test.brew` sends them to the next R script (`test.R`), which performs **vdmR** functions (Figure 11). Finally, the functions of the **vdmR** package are executed and the main window of the VDM tools generated for the temporary directory on the server are shown on the web browser.

This framework could be broadly applicable to various kinds of online analysis systems, including business intelligence (BI) systems.



Test

stat.fwu.ac.jp/~fujino/vdmR/webdemo/brew/test.brew

Example of web interface using vdmR package

Scatter plot

name: scat01

x-axis: Min.Price ▾ y-axis: Min.Price ▾

□ size: Min.Price ▾ □ color: Manufacturer ▾

Histogram

name: hist01

x-axis: Min.Price ▾

□ fill: Manufacturer ▾ □ color: Manufacturer ▾

Parallel Coordinate Plot

name: pcp01

axis: Min.Price
Price
Max.Price
MPG.city
MPG.highway
EngineSize
Horsepower
RPM
Rev.per.mile
Fuel.tank.capacity

□ groupColumn: Manufacturer ▾

Figure 11: Demo: Web interface for **vdmR** package.

5. Conclusions and future work

The **vdmR** package provides a direct method for generating web-based and **ggplot2** style VDM tools from R. The VDM tools have functionalities such as linked-brushing between some kinds of statistical plots such as scatter plots, histograms, parallel coordinate plots, and choropleth maps. Aesthetic mappings such as color, size, and fill are also available. Currently, these plots and mappings are implemented in the **vdmR** package as **ggplot2** based functions. We would like to implement more functions in future releases, including facetting, multiple layers, and additional plots such as boxplots, time series plots, and mosaic plots. In addition, we need to increase the performance of the package.

VR-Gobi (Nelson, Cook, and Cruz-Neira 1999) proposed the concept of 3D VDM in virtual reality and developed a prototype of the system. We would like to develop more practical 3D VDM tools by using recent web technologies and extending the **vdmR** package. It would be possible to implement the web-based 3D VDM tools by using WebGL technology (Parisi 2012), which is supported by major web browsers by default. Fortunately, some tools exist, for example, **three.js**, which help to generate 3D graphics in the WebGL format.

References

Bivand R, Lewin-Koh N (2017). ***maptools**: Tools for Reading and Handling Spatial Objects.* R package version 0.9-2, URL `https://CRAN.R-project.org/package=maptools`.

Buja A, Cook D, Swayne DF (1996). "Interactive High-Dimensional Data Visualization." *Journal of Computational and Graphical Statistics*, **5**(1), 78–99. doi:10.1080/10618600.1996.10474696.

Chang W, Wickham H (2016). ***ggvis**: Interactive Grammar of Graphics.* R package version 0.4.3, URL https://CRAN.R-project.org/package=ggvis.

Couture-Beil A (2014). ***rjson**: JSON for R.* R package version 0.2.15, URL https://CRAN.R-project.org/package=rjson.

Fujino T (2007). "SVG+Ajax+R: A New Framework for WebGIS." *Computational Statistics*, **22**(4), 511–520. doi:10.1007/s00180-007-0058-8.

Fujino T (2017). ***vdmR**: Visual Data Mining Tools for R.* R package version 0.2.5, URL https://CRAN.R-project.org/package=vdmR.

Horner J (2011). ***brew**: Templating Framework for Report Generation.* R package version 1.0-6, URL https://CRAN.R-project.org/package=brew.

Horner J (2013). ***rApache**: Web Application Development with R and Apache.* URL http://www.rapache.net/.

Horner J (2014). ***Rook**: A Web Server Interface for R.* R package version 1.1-1, URL https://CRAN.R-project.org/package=Rook.

Inselberg A (1985). "The Plane with Parallel Coordinates." *The Visual Computer*, **1**(2), 69–91. doi:10.1007/bf01898350.

Lewin-Koh N (2015). *CRAN Task View: Graphic Displays & Dynamic Graphics & Graphic Devices & Visualization.* Version 2015-01-07, URL https://CRAN.R-project.org/view=Graphics.

Murrell P, Potter S (2017). ***gridSVG**: Export **grid** Graphics as SVG.* R package version 1.5-1, URL https://CRAN.R-project.org/package=gridSVG.

Nelson L, Cook D, Cruz-Neira C (1999). "**XGobi** vs the C2: Results of an Experiment Comparing Data Visualization in a 3-D Immersive Virtual Reality Environment with a 2-D Workstation Display." *Computational Statistics*, **14**(1), 39–51. doi:10.1007/pl00022704.

Parisi T (2012). *WebGL: Up and Running.* O'Reilly Media.

R Core Team (2017). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Robinson D (2017). ***broom**: Convert Statistical Analysis Objects into Tidy Data Frames.* R package version 0.4.2, URL https://CRAN.R-project.org/package=broom.

Schloerke B, Crowley J, Cook D, Briatte F, Marbach M, Thoen E, Elberg A, Larmarange J (2017). ***GGally**: Extension to **ggplot2**.* R package version 1.3.2, URL https://CRAN.R-project.org/package=GGally.

Swayne DF, Temple Lang D, Buja A, Cook D (2003). "**GGobi**: Evolving from **XGobi** into an Extensible Framework for Interactive Data Visualization." *Computational Statistics & Data Analysis*, **43**(4), 423–444. doi:10.1016/s0167-9473(02)00286-4.

Symanzik J, Carr DB (2008). "Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data." In *Handbook of Data Visualization*, Springer Handbooks Comp.Statistics, pp. 267–294. Springer-Verlag. doi:10.1007/978-3-540-33037-0_12.

Theus M (2002). "Interactive Data Visualization Using **Mondrian**." *Journal of Statistical Software*, **7**(11), 1–9. doi:10.18637/jss.v007.i11.

Tukey JW (1977). *EDA: Exploratory Data Analysis*. Addison-Wesley.

Wegman EJ (1990). "Hyperdimensional Data Analysis Using Parallel Coordinates." *Journal of the American Statistical Association*, **85**(411), 664–675. doi:10.2307/2290001.

Wegman EJ (2003). "Visual Data Mining." *Statistics in Medicine*, **22**(9), 1383–1397. doi:10.1002/sim.1502.

Wickham H (2009). ***ggplot2**: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wickham H (2011). "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software*, **40**(1), 1–29. doi:10.18637/jss.v040.i01.

Wickham H, François R, Henry L, Müller K (2017). ***dplyr**: A Grammar of Data Manipulation*. R package version 0.7.4, URL https://CRAN.R-project.org/package=dplyr.

Affiliation:

Tomokazu Fujino
Department of Environmental Science
International College of Arts and Sciences
Fukuoka Women's University
1-1-1 Kasumigaoka, Higashi-ku, Fukuoka, Japan
E-mail: fujino@fwu.ac.jp
URL: http://www.fwu.ac.jp/

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/

November 2017, Volume 82, Issue 6 *Submitted:* 2014-12-06
doi:10.18637/jss.v082.i06 *Accepted:* 2016-12-15