

文章编号: 2095-2163(2019)05-0347-04

中图分类号: TP302.4

文献标志码: A

R语言在经济学GIS绘图中的应用

姚清仿

(贵州民族大学人文科技学院, 贵阳 550025)

摘要: 随着不同学科的交叉融合发展,基于地理信息系统(GIS)的绘图在经济学中发挥着越来越重要的作用。而具有强大可视化绘图功能的R语言,通过编写新代码或调整已有代码可轻松实现科研中数据呈现与图形绘制的要求。文章以贵州省2017年9个地级行政区的GDP增长率为例,使用3.5.1版本的R语言、R编辑器RStudio并载入相关的R-package绘制出有指北针、比例尺、不同类型图例的地图并提供相应的脚本与说明,为经济学研究中直观地表述研究结果提供了良好的范例,以方便从事经济学研究的人员直接更改使用。

关键词: GIS; R语言; 经济学

Application of R Language in economics GIS mapping

YAO Qingfang

(College of Humanities and Sciences, Guizhou Minzu University, Guiyang 550025, China)

[Abstract] With the development of interdisciplinary integration, mapping based on Geographic Information System (GIS) plays an increasingly important role in economics. R language, which has powerful visual drawing function, can easily meet the requirements of data presentation and graphics drawing in scientific research by writing new codes or adjusting existing codes. Taking the GDP growth rate of nine prefecture-level administrative regions in Guizhou Province in 2017 as an example, this paper draws maps of different types of legends and provides corresponding scripts and explanations by using R language of version 3.5.1, RStudio editor and loading relevant R-package, which provides an intuitive description of the research results in economic research. Thus, people engaged in economic research can implement direct use from this example.

[Key words] GIS; R language; economics

0 引言

随着不同学科的交叉融合发展,基于地理信息系统(GIS)的绘图在经济学中发挥着越来越重要的作用。R语言同时具有统计分析、数据处理和可视化绘图三大功能^[1]。是一个自由、免费、源代码开放的软件,是一种适合推广应用于制图的优秀工具^[2]。此外,R语言与其它数绘图工具不同,其在具体操作时可轻松地编写代码或者调整其它用户的代码,适应特定图形绘制要求^[3]。虽然目前已有许多介绍R与GIS绘图的资料^[4-6],但还没有关于R语言就地图的指北针、比例尺、图例的连续或者离散进行整合绘制的资料。此外,R语言版本和R的扩展包间存在兼容问题,导致图形绘制过程中一些函数或者功能的无法使用。并且即使是相同功能的R-package也需要不同参数进行调用^[7],这成为经济学研究者的困难之处。基于此,本文以贵州省2017年9个地级行政区的GDP增长率为例,使用3.5.1版本的R语言、R编辑器RStudio并载入相关的R-package包,绘制出有指北针、比例尺、图例的地图并

提供相应的脚本与说明,旨在为经济学研究中直观地表述研究结果提供了良好的范例,具有一定的借鉴意义。

1 材料与方法

1.1 R语言与RStudio软件的下载

3.5.1版本X64位的R语言安装程序包从<https://cran.r-project.org/bin/windows/base/old/3.5.1/>下载。由于RStudio集成了丰富的开发界面,提供了方便的函数名识别及搜索功能,极大地方便了R编程计算^[8]。RStudio安装程序包可从<https://www.rstudio.com/products/rstudio/download/>下载。

1.2 R-package的下载及其使用

R语言绘图功能的实现很大程度上依赖于不同的R-package包,其中包含了许多用户可以直接调用的函数^[9]。本文需要下载的R-package软件包内容见表1。

作者简介: 姚清仿(1992-),男,硕士,主要研究方向:农业经济与贸易。

收稿日期: 2019-08-12

表1 GIS图绘制所需要的R-package

Tab. 1 The R-package for GIS plot

R-package	作用特点
Maptools, sp	处理空间的对象或数据,配合 ggsn 包的使用
ggplot2	创建图形的可视化
ggsn	比例尺、指北针的绘制
viridis	离散图例的绘制

R语言使用时 R-package 只需要安装一次,但每次使用均需要重新加载^[10]。下载和加载 R-package 包的代码分别为:

```
Install, packages("maptools") # 下载 maptools 包
library(maptools) # 加载 maptools 包
```

1.3 数据的读取

省、市、县的数据来自于国家基础地理信息中心,也可以来自非官方的全球行政区域数据库 GADM2。当数据存储格式为 .shp 时,使用 rgdal 包中的函数 readOGR(file.choose()) 读取。其中函数:

file.choose() 可以实现手动点击数据文件,从而将数据输入 R 语言运行的环境中。其执行语句如下:

```
library(rgdal) # 加载 rgdal 的包。
```

```
CHN3 = readOGR(file.choose()) # 读取 .shp 格式的数据。
```

当数据的存储格式为 .rds 时,使用 base 包中的函数 readRDS(file.choose()), base 属于 R 语言中基础包,其中的函数如 file.choose()、readOGR() 无需加载即可调用。其执行语句如下:

```
CHN3 = readOGR(file.choose()) # 读取 .rds 格式的数据。
```

贵州省 2017 年 9 个地级行政区的 GDP 增长率的数据来自于贵州省统计局的《贵州统计年鉴 2018》。通常需要将其于城市的 ID 对应赋值之后,使用函数 read.csv() 将 GDP 增长率的数据读入到 R 语言环境中。其执行语句如下:

```
guizhou = subset(CHN3, CHN3@data$NAME_1 == "Guizhou") # 获取贵州省的地理数据。
```

Guizhou = data.frame(row.names(guizhou@data), guizhou@data) # 将贵州省地级市的编号整合为数据框。由于读入的地理信息的数据是 SpatialPolygonsDataFrame 类,需要通过 @ 取得其中的数据。

```
CityID = Guizhou[, c("row.names.Guizhou.data.", "NL_NAME_2")] # 获取城市编号和名称对应的数据。
```

```
read.csv(file.choose()) # 读取存储格式为 .csv 并含有城市编号的 GDP 增长率的数据。
```

1.4 绘图主要步骤及其代码实现

在读取数据的基础之上,需要将数据加工处理为绘图的数据,即根据城市的编号将地理数据和 GDP 增长率的数据的整合。首先,将 SpatialPolygonsDataFrame 类的地理数据转换为数据框类型。其次,将地理的信息数据和贵州省 9 个地级行政区的 GDP 增长率的数据根据城市编码合并。再次,通过计算每个城市中心的地理坐标,可在每个地级行政区添加对应的区域名称。最后,调用函数 ggplot() 即可绘制相应的 GIS 图。具体代码如下:

```
gzmap = fortify(guizhou) # 将地理信息的数据转换为数据框的格式。
```

```
GZmap = merge(gzmap, GDP, by.x = "id", by.y = "ID", all.x = TRUE) # 将 GDP 增长率和数据框类型的地理数据按照城市编码 ID 合并。
```

```
pos = data.frame(row.names(coordinates(guizhou)), coordinates(guizhou), guizhou@data$NL_NAME_2) # row.names(coordinates(guizhou)) 是各个行政区的编码, coordinates(guizhou) 是计算各个行政区的地理中心坐标。 guizhou@data$NL_NAME_2 为各个行政区名称, 将三者整合为数据框。
```

```
names(pos) = c("ID", "Lon", "Lat", "cityname") # 将整合数据的各列重新命名, 分别表示城市编码、城市经度、城市纬度、城市名称。
```

2 图例结果说明

2.1 图例连续的 GIS 图

为了更加直观的展示 2017 年贵州省 9 个地级行政区的 GDP 增长率,通过执行相应的 R 语言代码绘制了其对应的地理信息图。具体代码如下:

```
library(maptools)
```

```
library(sp)
```

```
library(ggplot2)
```

```
library(ggsn)
```

```
library(viridis)
```

```
GDP = read.csv(file.choose())
```

```
names(GDP)[1] = "ID" # 将读入的 GDP 增长率的数据的第一列变量命名为 ID。
```

```
CHN3 <- readRDS(file.choose())
```

```
guizhou = subset(CHN3, CHN3@data$NAME_1 == "Guizhou")
```

```
Guizhou = data.frame(row.names(guizhou@
```

```

data), guizhou@data)
CityID = Guizhou[, c("row.names.guizhou", "NL_NAME_2")]
pos = data.frame(row.names(coordinates(guizhou)), coordinates(guizhou), guizhou@data$NL_NAME_2)
names(pos) = c("ID", "Lon", "Lat", "cityname")
gzmap = fortify(guizhou)
GZmap = merge(gzmap, GDP, by.x = "id", by.y = "ID", all.x = TRUE)
ggplot(GZmap, aes(long, lat)) + #ggplot 进行图形的绘制, 将数据可视化。
scale_fill_gradient(low = "green", high = "red", name = "GDP 增长率(%)") + #绘制连续型的图例, 指定其颜色的变化范围
geom_polygon(aes(group = group, fill = GDPincrease), colour = "black") + #设置区域的填充数据
coord_equal() + #指定图形的横坐标和纵坐标的显示比例
north(GZmap, location = "topleft", symbol = 9) + #绘制指北针
scalebar(GZmap, location = "bottomright", dist = 100, height = 0.025, dist_unit = "km", st.size = 2.5, st.dist = 0.04, transform = TRUE, model = "GRS67") + #绘制比例尺
geom_text(aes(x = pos$Lon, y = pos$Lat, label = cityname), data = pos) +
theme(
  panel.grid = element_blank(), #设置网格线
  panel.background = element_blank(), #设置图形的背景
  axis.text = element_blank(), #设置坐标轴的信息
  axis.ticks = element_blank(), #设置坐标轴的刻度
  axis.title = element_blank(), #设置坐标轴的标题
  legend.position = "right" #指定图例的位置。
)

```

在以上代码中, 指北针函数的参数 `symbol` 表示

指北针的不同类型, 可以使用函数 `northSymbols()` 查看并选择相应的类型。比例尺的函数中, `dist` 用于设置图上 1 单位代表的实际距离的数值。`st.size` 表示的是数值的字体大小, `st.dist` 表示的是数值与比例图标的距离, `Model` 表示比例尺图标的显示类型, 执行函数 `help(scalebar)` 便可查找相应的 `Model` 类型。

执行以上代码, 即可得到含有贵州省 9 个地级行政区的 GDP 增长率的 GIS 图。具体效果如图 1 所示。结果显示, 黔东南苗族侗族自治州的 GDP 增长率小于 7%, 显示为白色。其中颜色最深的是黔西南布依苗族自治州, 其 GDP 增长率组最高约为 12.5%。其它的行政区的颜色越深, GDP 增长率相应的就越高。

2.2 图例离散的 GIS 图

在经济学研究中, 有时需要将数据进行分类展示。为此, 需要使用图例离散函数 `scale_fill_viridis()`, 该函数中的参数 `direction` 表示离散分类的排序方式。`Option` 表示其中颜色的显示方式, 有 A-E5 种选择方式。`alpha` 设置颜色的透明度, `begin`、`end` 分别设置颜色的色调。具体代码如下:

```

ggplot(GZmap, aes(long, lat)) +
scale_fill_viridis(discrete = TRUE, labels = c("[5.2, 11.4]", "(11.4, 12.1]", "(12.1, 12.5]"), name = "GDP 增长率(%)", alpha = 0.4, begin = 0.15, end = 0.3, direction = -1, option = "E") + #设置离散图例。
geom_polygon(aes(group = group, fill = cut_number(GDPincrease, 3)), colour = "black") +
coord_equal() +
north(GZmap, location = "topleft", symbol = 9) +
scalebar(GZmap, location = "bottomright", dist = 100, height = 0.025, dist_unit = "km", st.size = 2.5, st.dist = 0.04, transform = TRUE, model = "GRS67") +
geom_text(aes(x = pos$Lon, y = pos$Lat, label = cityname), data = pos) +
theme(
  panel.grid = element_blank(),
  panel.background = element_blank(),
  axis.text = element_blank(),

```

(下转第 353 页)