

R 软件绘图中中文渲染解决方案

2021 年 3 月 7 日

1 引言

R 通常情况下既是指 R 语言也是指 R Project，后者是由社区发起的一个项目，提供了 R 语言的解释器的实现以及源码。在 R 语言中，向量、矩阵、高维数组和表格型数据 (`data.frame`) 的处理变得非常方便和容易，并且活跃的开源社区也为 R 实现了很多使用的软件包 (`package`)，从而 R 常常被用在数据分析和数据挖掘任务中。它的一个常用的功能是作图，也就是数据可视化，在这其中，有时候我们希望在图表上附上中文的注解、图示、文本段落和标签，以此来增加图表的易读性，但是默认情况下，不使用三方软件包，只用 R 自带的绘图函数，在 Mac 系统上进行绘图时，会遇到中文字体显示不出的问题。为此，社区提供了许多解决方案，我们将要介绍其中的两个，分别是 `ragg` 和 `showtext`。

在 R 中，图形输出是通过设备 (`device`) 进行的，所谓的设备主要包括一系列图形函数的实现，例如画路径的函数，画多边形的函数，栅格化的函数，以及渲染字体的函数 (`text()`)，中文显示问题其实跟 R 自带的图形设备有关，默认只支持几种很少的字体，所以当需要渲染中文字体（或者一些 CJK 字体）时，就经常会出现问题。这些软件包的原理，主要是替换 `device` 中负责渲染字体的那些函数，当然了，所谓的设备并不是指物理设备，而是指一系列的 R 需要显示图形时所调用的函数。

2 问题复现

操作系统是 macOS Sierra 10.12.6，R 的版本是 3.6.3，`ggplot2` 的版本是 3.3.2，代码：

```
1 library(ggplot2);
2
3 t = "中文内容";
4
5 p = ggplot() +
6   geom_text(
7     aes(x = 0, y = 0, label = t),
8     family = "Noto Serif SC",
9     size = 24
10  );
11
12 pdf("example-0.pdf", 7, 4);
13 print(p);
14 dev.off();
```

会报错：

```
1 Error in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
2 invalid font type
3 Calls: print ... drawDetails -> drawDetails.text -> grid.Call.graphics
```

```

4 In addition: Warning messages:
5 1: In grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
6 font family 'Noto Serif SC' not found in PostScript font database
7 2: In grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
8 font family 'Noto Serif SC' not found in PostScript font database
9 3: In grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
10 font family 'Noto Serif SC' not found in PostScript font database
11 4: In grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
12 font family 'Noto Serif SC' not found in PostScript font database
13 5: In grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
14 font family 'Noto Serif SC' not found in PostScript font database
15 Execution halted

```

但是我能确定这个字体是安装了的。现在如果去掉字体设置：

```

1 library(ggplot2);
2
3 t = "中文内容";
4
5 p = ggplot() +
6   geom_text(aes(x = 0, y = 0, label = t));
7
8 pdf("example-0.pdf", 7, 4);
9 print(p);
10 dev.off();

```

会发现中文无法正常显示 (图 2-1)



图 2-1: 中文无法正常显示

包括将 pdf 改为 png，将 example-0.pdf 改为 example-0.png 也无济于事。

3 ragg 解决方案

首先我们安装 ragg，如果安装了则直接载入：

```
install.packages("ragg");
```

以下是代码：

```

1 library(ggplot2);
2 library(ragg);
3
4 content = "中文内容";
5 p = ggplot() +
6   geom_text(

```

```

7     aes(x = 0, y = 2, label = content),
8     family = "Noto Serif SC",
9     size = 18
10 );
11
12
13 agg_png("noto_1.png");
14 print(p);
15 dev.off();

```

效果如图



图 3-1: 使用 ragg 包渲染包含中文的图形

我们又加载了 `ragg` 包，并且将设备 `pdf` 换成设备 `agg_png`，并且在 `geom_text` 函数中指定了中文字体，由此我们看出 `ragg` 包的使用方式主要是在设备名前面加上 `agg_`，目前我们发现 `ragg` 还不支持 `pdf`，所以我们使用了 PNG 作为输出格式，图 3-2列出了 `ragg` 包提供的可支持自定义字体的图形设备：

```

> library("ragg");
> agg_
agg_capture      agg_png      agg_supertransparent
agg_jpeg         agg_ppm      agg_tiff

```

图 3-2: ragg 包提供的图形 device

事实上，`ragg` 通过 `systemfonts` 包将系统安装的字体载入，这里 Noto Serif SC 是我自行安装的一个字体。所以说，`ragg` 实质上是通过实现支持系统安装字体的图形设备 `agg_*` 来解决这个中文字体渲染问题。理论上，系统安装了的字体，都可以通过 `ragg` 包在 R 中使用。

4 showtext 解决方案

`showtext` 出现得更早，并且相比 `ragg`，它还支持pdf设备输出，首先安装 `showtext` 包：

```
install.packages("showtext");
```

然后下面是一个实例：

```

1 library(ggplot2);
2 library(showtext);
3
4 path_to_font = "../...路径.../NotoSerifSC-Regular.otf";
5 font_add("nsscr", path_to_font);
6
7 t = "中文内容";
8

```

```

9 p = ggplot() +
10 geom_text(
11   aes(x = 0, y = 0, label = t),
12   family = "nsscr",
13   size = 24
14 );
15
16 pdf("showtext.pdf");
17 showtext_begin();
18 print(p);
19 showtext_end();
20 dev.off();

```



图 4-1: 使用 showtext 使 R 支持中文渲染

这里我们需要知道字体的路径，然后通过 `font_add` 函数加载，而 `nsscr` 则是指定给这个字体的一个别名，或者我们也可以通过 `font_add_google` 函数加载谷歌字体。 `showtext` 的工作方式主要是体现在 `showtext_begin` 和 `showtext_end` 这两个函数上面，要注意的是，`showtext_begin` 和 `showtext_end` 执行的时候要求已经有图形设备打开，所以

```
pdf("showtext.pdf");
```

是放在

```
showtext_begin();
```

之前，并且

```
dev.off();
```

是放在

```
showtext_end();
```

之后。实际上还有更加简单的使用方式：

```

1 library(ggplot2);
2 library(showtext);
3
4 showtext_auto();
5
6 path_to_font = "/path/to/xxx.otf";
7 font_add("nsscr", path_to_font);
8
9 t = "中文内容";
10
11 p = ggplot() +
12 geom_text(

```

```
13     aes(x = 0, y = 0, label = t),
14     family = "nsscr",
15     size = 24
16 );
17
18 pdf("showtext1.pdf");
19 print(p);
20 dev.off();
```

即在代码的开头就放置 `showtext_auto`，让系统来判断什么时候需要 `showtext_begin()` 什么时候需要 `showtext_end()`。 `showtext` 所做的呢，就是从字体文件中提取出字形 (glyphs)，然后利用这些字形数据，自己实现渲染字体的函数，然后当 `showtext_begin` 执行时，R 自带图形设备函数（例如 `pdf`）所依赖的底层字体渲染函数会被替换成 `showtext` 的实现，而 `showtext_end` 则是将替换恢复。

5 总结

`ragg` 包通过实现自己的支持系统字体的图形设备 `agg_*` 使得系统字体在 R 中可用，但是不支持 PDF 输出；而 `showtext` 则可以读入字形文件，自己实现渲染逻辑，并且将 R 自带图形设备的渲染逻辑引导至自己实现的函数，实现自定义字体的支持，同时相比 `ragg` 它还支持 PDF 输出。