R语言学习之道：罗马非一日建成

原创 2017-08-29 louwill [数据科学家养成记](https://mp.weixin.qq.com/s?__biz=MzI4ODY2NjYzMQ==&mid=2247483937&idx=1&sn=c6f7258e91cb2de1c0b62a6fdc44f358&chksm=ec3ba749db4c2e5f5205f5e136f41ea64813062509cf7746208d18b33ad8ea9e2bc8f0f0ed9e&scene=38&key=fef0cb6a4f422abce52a2649d3e31e44ff34df17435969ade8841aec07d3ea35b4655a3f25fd4b18eaa2f5879bf97a947bf9cf18ad35af2ea0eca801c20b795c7d69f043eca1c6ed408324c2ae88f394&ascene=7&uin=MjczMTU1MjUwMA%3D%3D&devicetype=Windows+10&version=62060038&pass_ticket=asf26keQy9psk7MnNEGtbSxrAuLfa0TjXGECGP%2BYQlvg7w8Vef4gSa3zI4X0i7GA&winzoom=1##)

https://mmbiz.qpic.cn/mmbiz_gif/4lN1XOZshfcRGz67CAts4Cia5Co0pqVrYkdRvLuJCKX6GURdrCeQicPa2JhIm3pxZr74DBleYvZ7jd4f4MUH7gXw/0?wx_fmt=gif&tp=webp&wxfrom=5&wx_lazy=1本文翻译和改编自**Sharp Sight Labs**

要想精通数据科学，大量的练习必不可少。说起来容易做起来难，实际上很多人并不知道该如何练习。

如何进行数据科学的训练是一个十分广泛的话题，这有点超出本文所要介绍的范围了，但在这里我还是想给大家介绍一些有效的建议：

在众多的数据科学技能之中，你一开始最应该掌握最重要的那些技术，然后通过写一些小的脚本来不断的练习直到你知道了这些技能的前因后果。在数据科学领域，知其所以然往往更重要。

稍微夸张一点说就是，你最终目标应该是闭着眼睛就可以复盘所有的代码。

在此信念下，我想给大家一些小的脚本案例供大家来练习。下面是一些关于绘制一些地图的R代码。（我希望大家能认真注意一下下面的代码有多简单，又有多少函数和命令是你实际需要知道的）

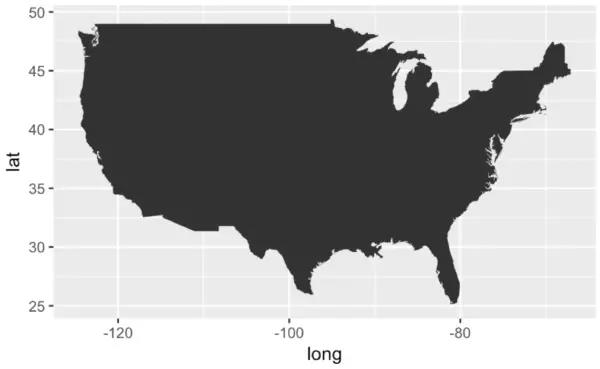
由易入手，逐个击破

先从一个简单的例子开始：创建一幅美国地图。

# MAP USA (STATES)  
map\_data("state") %>%

 ggplot(aes(x = long, y = lat, group = group)) +

   geom\_polygon()



三行代码（硬要算上library函数的话那就是四行）。

总之三行代码我们搞定了美国地图。

并且你仅需要改变5个字符（把usa改成state）就可以得到一幅显示各州的美国地图。

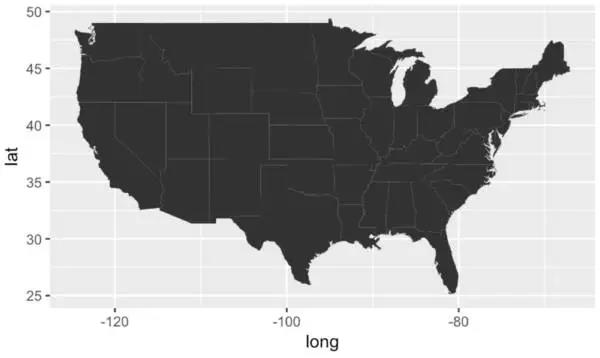
如果你再添加一条线，你可以使用filter函数对你的数据取个子集，然后展示一部分州的地图。

# MAP CALIFORNIA, NEVADA, OREGON, WASHINGTON  
# - to do this, we're using dplyr::filter()  
# ... otherwise, it's almost exactly the same as the previous code  
map\_data("state") %>%

 filter(region %in% c("california","nevada","oregon","washington")) %>%

 ggplot(aes(x = long, y = lat, group = group)) +

   geom\_polygon()



在这里我想强调的是如果你了解filter函数和管道操作符（%>%）的性能的话，你会意识到这有多么方便快捷。

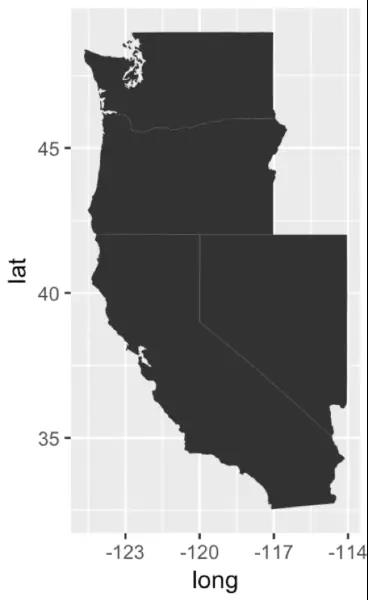
你要是再对这个图做一些变化的话，你也可以创建各个州下面的县的地图展示。

# MAP CALIFORNIA, NEVADA, OREGON, WASHINGTON  
# - to do this, we're using dplyr::filter()  
# ... otherwise, it's almost exactly the same as the previous code  
map\_data("state") %>%

 filter(region %in% c("california","nevada","oregon","washington")) %>%

 ggplot(aes(x = long, y = lat, group = group)) +

   geom\_polygon()

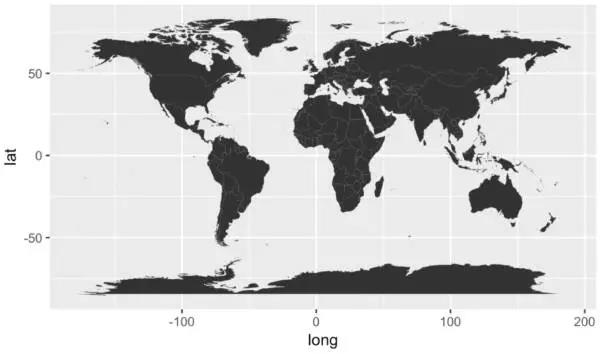


再修改一下代码世界地图也可以轻易画出来：

# MAP WORLD  
map\_data("world") %>%

 ggplot(aes(x = long, y = lat, group = group)) +

   geom\_polygon()



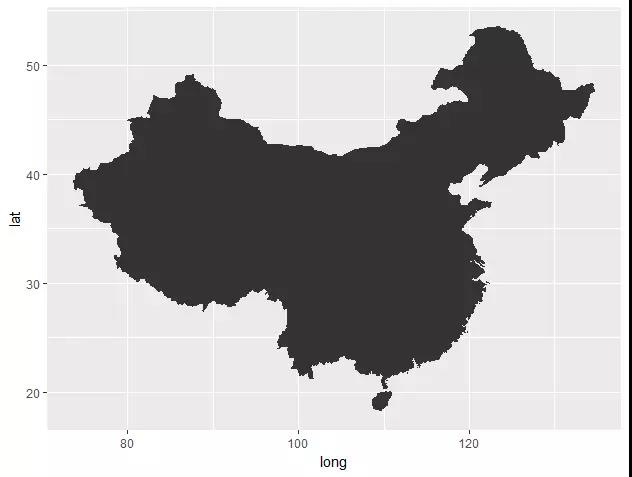
之后你可以选择绘制任意一个国家的地图，比如说中国：

# MAP China  
map\_data("world") %>%

 filter(region == 'China') %>%

 ggplot(aes(x = long, y = lat, group = group)) +

   geom\_polygon()



（改编者注：美帝国主义用心险恶，这里制成的中国地图没有台湾省，大家要注意）

我想再一次强调的是上面所有地图都只需要对我们一开始的那三行代码进行略微的修改就可以得到这些图形。在创建这样的基础地图的时候真的不需要我们用的多少复杂的方法，通常简单的两个R语言函数即可搞定。

由小见大，举一反三

也许你会问：为什么一定要练习这些基本的绘图函数？平时这些都太简单根本用不着。你应该认识到建成罗马城并非一日之功，学习R语言尤是如此。

此外，这些小的代码片段也是由一些基本函数构成的，你也可以把这些函数抽出来单独进行学习。所以精通一门语言的诀窍首先在于你要掌握一些基本的很微小的单个命令和函数。在你熟记这些语法和函数之后，这样的小脚本练习就让你有机会写一个稍微复杂一点的程序。不断的写这样的小片段代码，然后把它们组合起来你就可以完成更复杂的数据分析情形。多次练习之后，你就有信心直接上手一些小的数据分析项目了。

比如说，在经过上面很多次的绘制地图的小片段的代码练习后，你就可以绘制一些像下面这样复杂的地图：



如果你真想熟练掌握数据科学的技巧的话，在进行大的练习之前不断的找小的脚本片段进行练习。对这些练习代码反复拆分重组直到你可以不假思索的在你的电脑上敲出它们。你会回来感谢我的，祝好。

原文链接：

http://sharpsightlabs.com/blog/simple-practice-maps-tidyverse/