

数据可视化

李伟

目录

包安装	2
第一个数据集	3
第一个图	4
课堂小练习	7
了解偏离点的汽车类型	8
分面技术	11
参考书目	12

包安装

首先, 需要安装相应的语言包 tidyverse: `install.packages(...)`.

```
install.packages("tidyverse")
```

其次, 我们加载所需的语言包。

```
library(tidyverse)
```

```
## -- Attaching packages ----- ti

## v ggplot2 3.2.1      v purrr   0.3.2
## v tibble  2.1.3      v dplyr   0.8.3
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## -- Conflicts ----- ti
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

注意, 每个包只需要安装一次, 但每次使用前都需要加载。

第一个数据集

第一个数据集来自于 `ggplot2` 包中的 `mpg`, 我们可以使用 “`?mpg`” 查看相应的帮助文档.

主要的参数如下:

`manufacturer`: 生产商;

`model`: 汽车型号;

`displ`: 发动机排量 (L);

`year`: 上市年份;

`cyl`: 汽缸数目;

`trans`: 变速器类型;

`drv`: 驱动方式;

`cty`: 在城市一加仑汽油相应的英里数;

`hwy`: 在高速路一加仑汽油相应的英里数;

`fl`: 燃油类型;

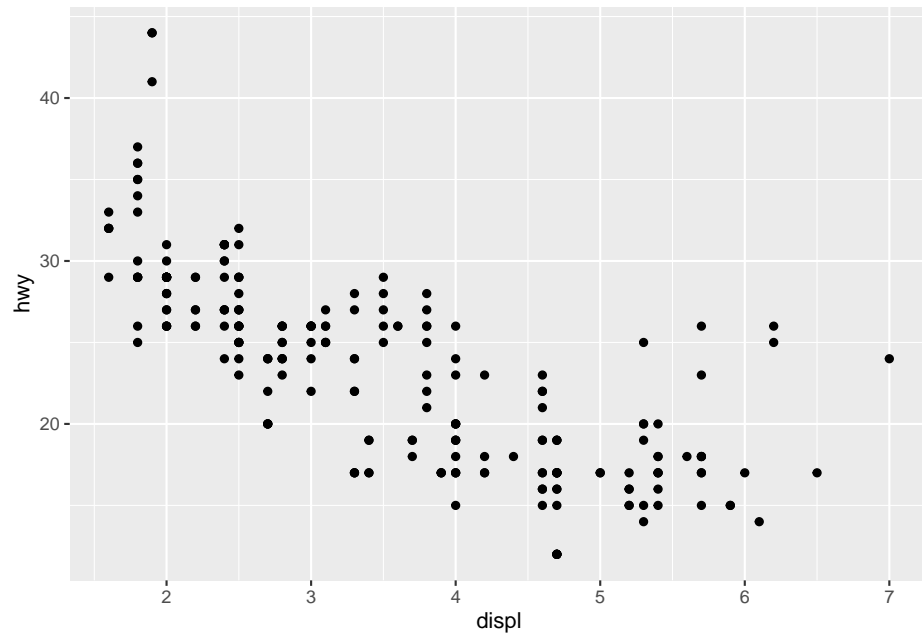
`class`: 汽车类型 (属于两座车, 紧凑型, SUV 等)

第一个图

现在，我们用 ggplot2 回答第一个问题：

是否发动机的排放量越大，其对应的耗油量越大？

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy))
```



ggplot2 包的作者给出了一个画图模板:

```
ggplot(data = <DATA>) +  
  <GEOM_FUNCTION>(mapping = aes(<MAPPINGS>))
```

课堂小练习

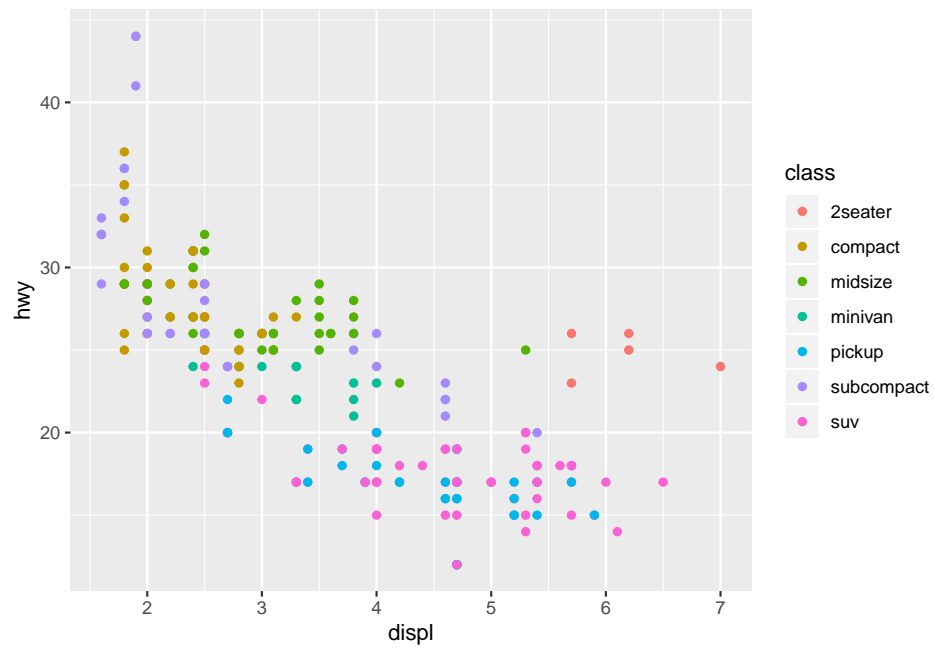
1. mpg 有多少行，有多少列？
2. 作出 hwy 与 cyl 之间的散点图。
3. 除了上述图，还可以做什么图形来回答刚刚的问题？

了解偏离点的汽车类型

在刚刚的图中，有几个点与大部分点的趋势是有偏离的。

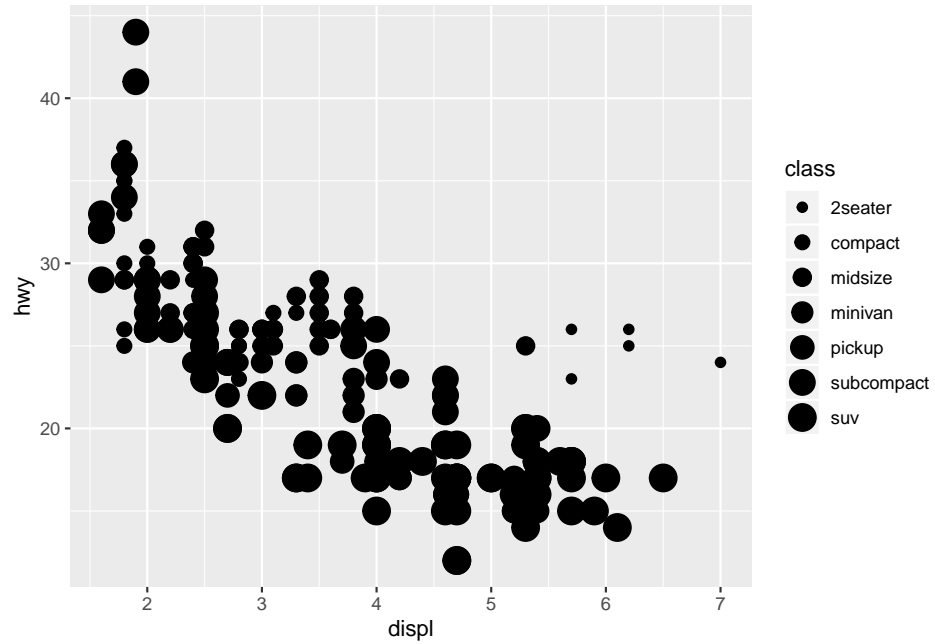
下面我们看一下是这些点代表的是什么类型的汽车。

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(displ, hwy, colour = class))
```




```
ggplot(data = mpg) +  
  geom_point(mapping = aes(displ, hwy, size = class))
```

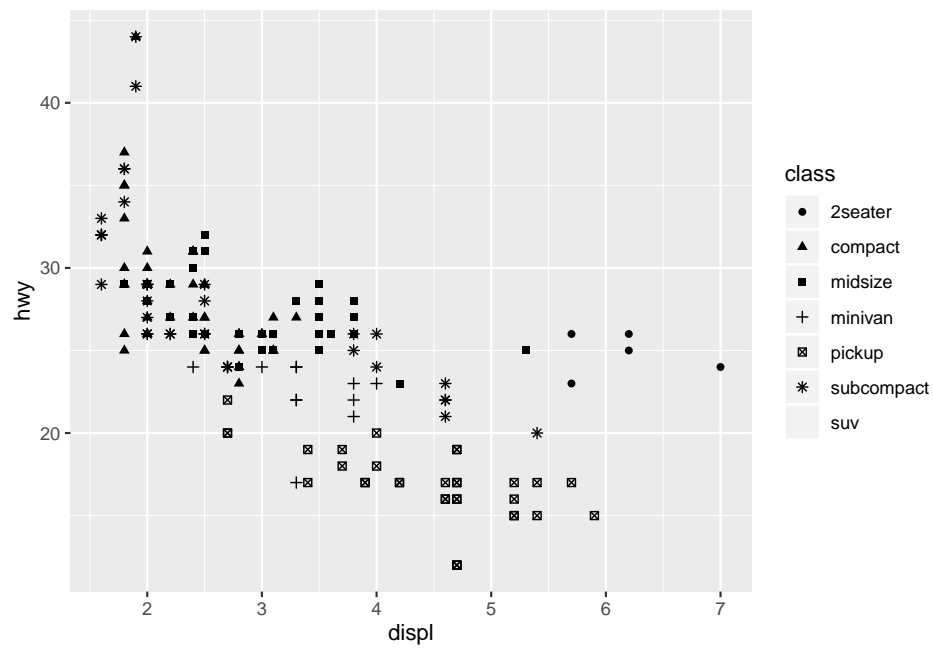
```
## Warning: Using size for a discrete variable is not advised.
```



```
ggplot(data = mpg) +  
  geom_point(mapping = aes(displ, hwy, shape = class))
```

```
## Warning: The shape palette can deal with a maximum of 6 discrete values  
## because more than 6 becomes difficult to discriminate; you have 7.  
## Consider specifying shapes manually if you must have them.
```

```
## Warning: Removed 62 rows containing missing values (geom_point).
```

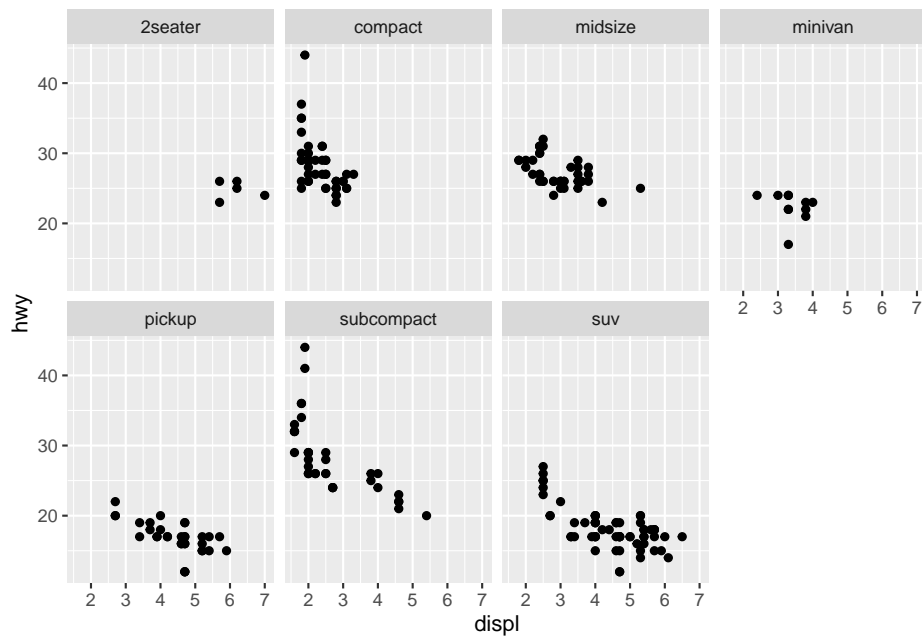


分面技术

如果我们单独看每一种类型的车辆，其排量与耗油量的关系，此时可以借助分面技术。

分面技术可以帮助我们针对不同的种类进行作图对比。

```
ggplot(data = mpg) +  
  geom_point(mapping = aes(x = displ, y = hwy)) +  
  facet_wrap(~ class, nrow = 2)
```



参考书目

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*.

Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science : Import, Tidy, Transform, Visualize, and Model Data*.