

第5章 R 绘图基础

1. 使用 mtcars 数据集，绘制散点图矩阵，并观察哪些变量对之间具有分布规律性。

【参考答案】通过 pairs() 函数创建 mtcars 数据集所有变量对的散点图矩阵。

`pairs(mtcars, main="mtcars数据集散点矩阵图")`

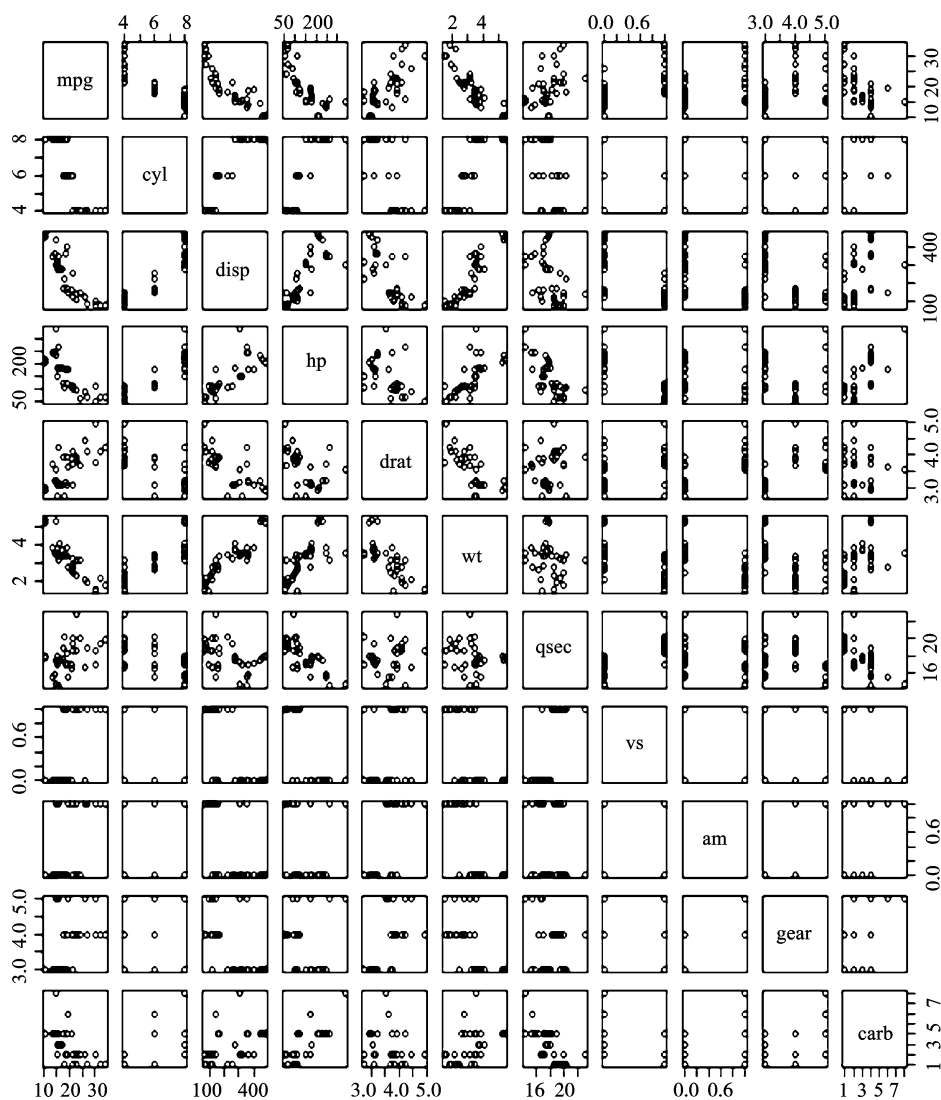


图 5-1 mtcars 数据集散点矩阵图

以 mtcars 数据集中 mpg、disp、drat、wt 四属性为例做变量对分布规律性分析。由散点图(图 5-1)可知 mpg、disp、drat、wt 四个变量，相关度最高的是 disp 和 wt 属性，表明两者关联性大，disp 的波动受 wt 影响最大；相关度最低的是 mpg 和 drat 属性，表明两者关联性相对其他属性较小，mpg 的波动受 drat 影响相比其他属性最小。

其他属性分布规律分析与此类似。

2. 使用 `mtcars` 数据集，分别以传输类型(`am`)和气缸数(`cyl`)分组，绘制耗油(`mpg`)、马力(`hp`)、重量(`wt`)等变量的箱线图，并判断是否存在显著差异。

【参考答案】

(1) `mpg`(耗油)

```
set.seed(3)
# 通过set.seed()来固定每次sample随机给出的三种颜色
col <- sample(colors(),3)
# boxplot(mpg ~ cyl + am,data=mtcars,at = c(1:3,5:7),col=col)
boxplot(mpg ~ cyl + am,data=mtcars,at = c(1:3,5:7),
        ylab="mpg(耗油)",col=col,main="耗油箱线图")
```

结果分析：通过耗油箱线图(图 5-2)可知，`mpg`(耗油)分别在 `cyl`(气缸数)和 `am`(传输类型)条件下，数据都无异常值。`mpg` 在 `cyl` 中的分布相对均匀，而在 `am` 中主要集中在第一组，同时在 `am` 中耗油整体比在 `cyl` 中高。

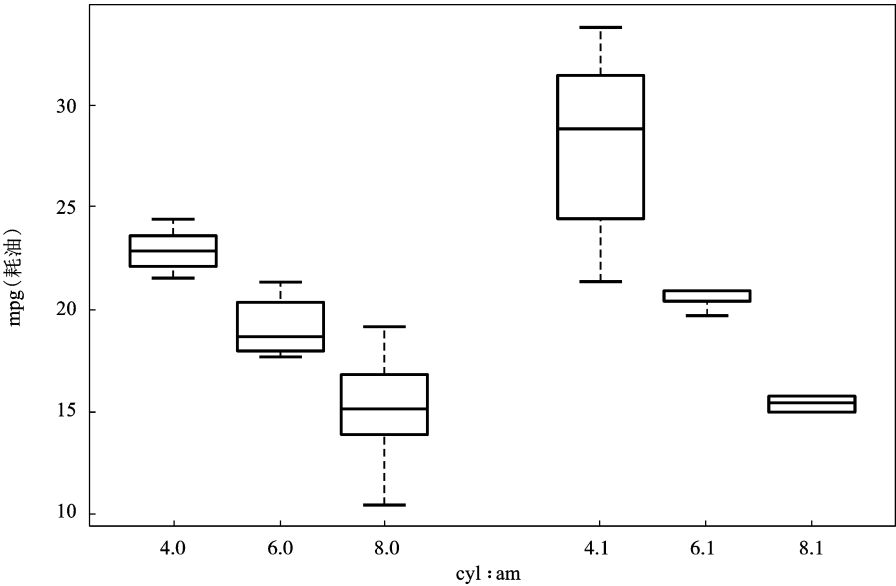


图 5-2 耗油箱线图

(2) `hp`(马力)

```
> set.seed(2)
> col<-sample(colors(),3)
> boxplot(hp ~ cyl + am,data=mtcars,at = c(1:3,5:7),
+        ylab="hp(马力)",col=col,main="马力箱线图")
```

结果分析：通过马力箱线图(图 5-3)可知，`hp`(马力)分别在 `cyl`(气缸数)和 `am`(传输类型)条件下，数据都无异常值。`hp` 随着 `cyl` 和 `am` 的增大，`hp` 也增加，但 `am` 对 `hp` 影响比 `cyl` 大。

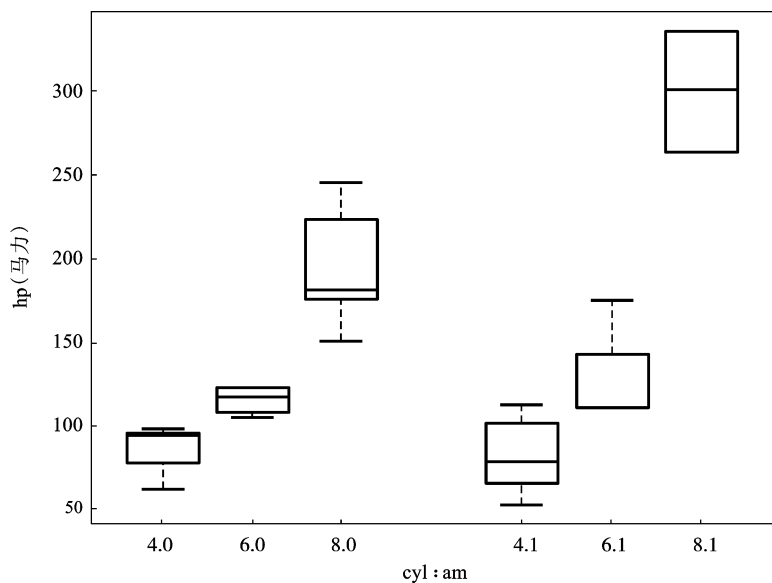


图 5-3 马力箱线图

(3) wt(重量)

```
> set.seed(1)
> col<-sample(colors(),3)
> boxplot(wt ~ cyl + am,data=mtcars,at = c(1:3,5:7),
+         ylab="wt(重量)",col=col,main="重量箱线图")
```

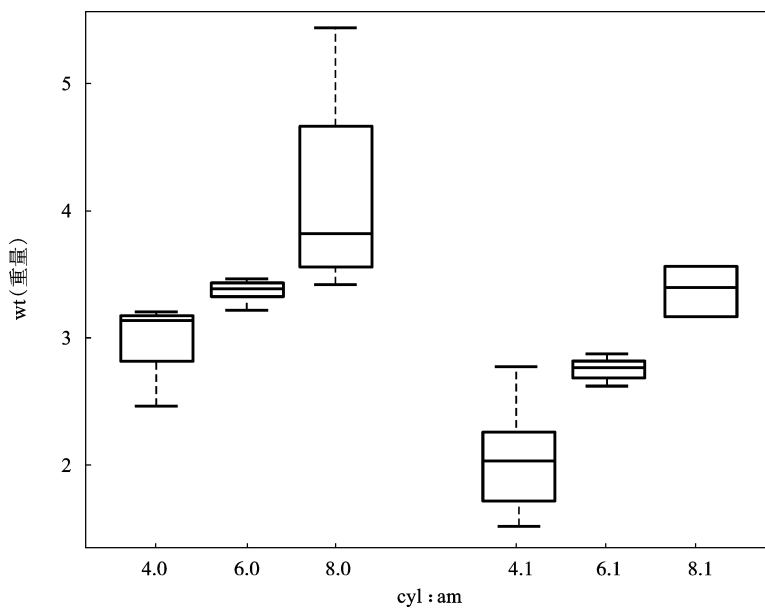


图 5-4 重量箱线图

结果分析：通过重量箱线图(图 5-4)可知，wt(重量)分别在 cyl(气缸数)和 am(传输类型)条件下，数据都无异常值。随着 cyl 和 am 的增加，wt 也增加，但 am 对 wt 影响比 cyl 小。

其余 mtcars 数据变量判断差异方法与上面类似。

3. 使用 mtcars 数据集，利用 ggplot2 绘图工具绘制马力 (hp) 和油耗 (mpg) 散点图，分别使用传输类型 (am) 和气缸数 (cyl) 确定图例形状，并尝试不同涂色方法。

【参考答案】

(1)使用传输类型 (am) 作为图例形状绘图

```
> library(ggplot2)
> ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(am)))+geom_point()
```

①调用 ggplot2 包中的 ggplot 函数绘图；上述代码中，参数指定使用 mtcars 数据集作为绘图数据来源；参数 aes 中设置横纵坐标对应的向量；以 am 作为绘图的形状；geom_point() 函数用于绘制点。执行效果如图 5-5 所示。

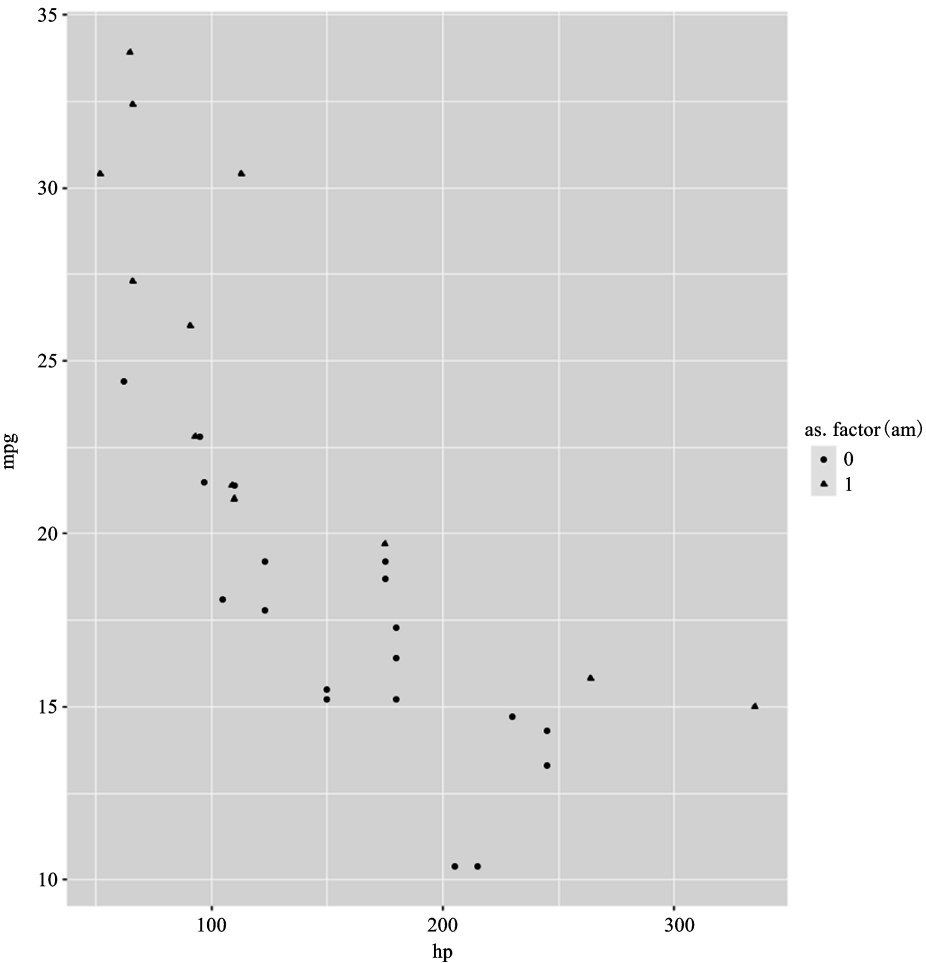


图 5-5 执行效果图

②在此基础上添加颜色。

```
ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(am), colour=hp))+geom_point()
```

执行效果如图 5-6 所示。

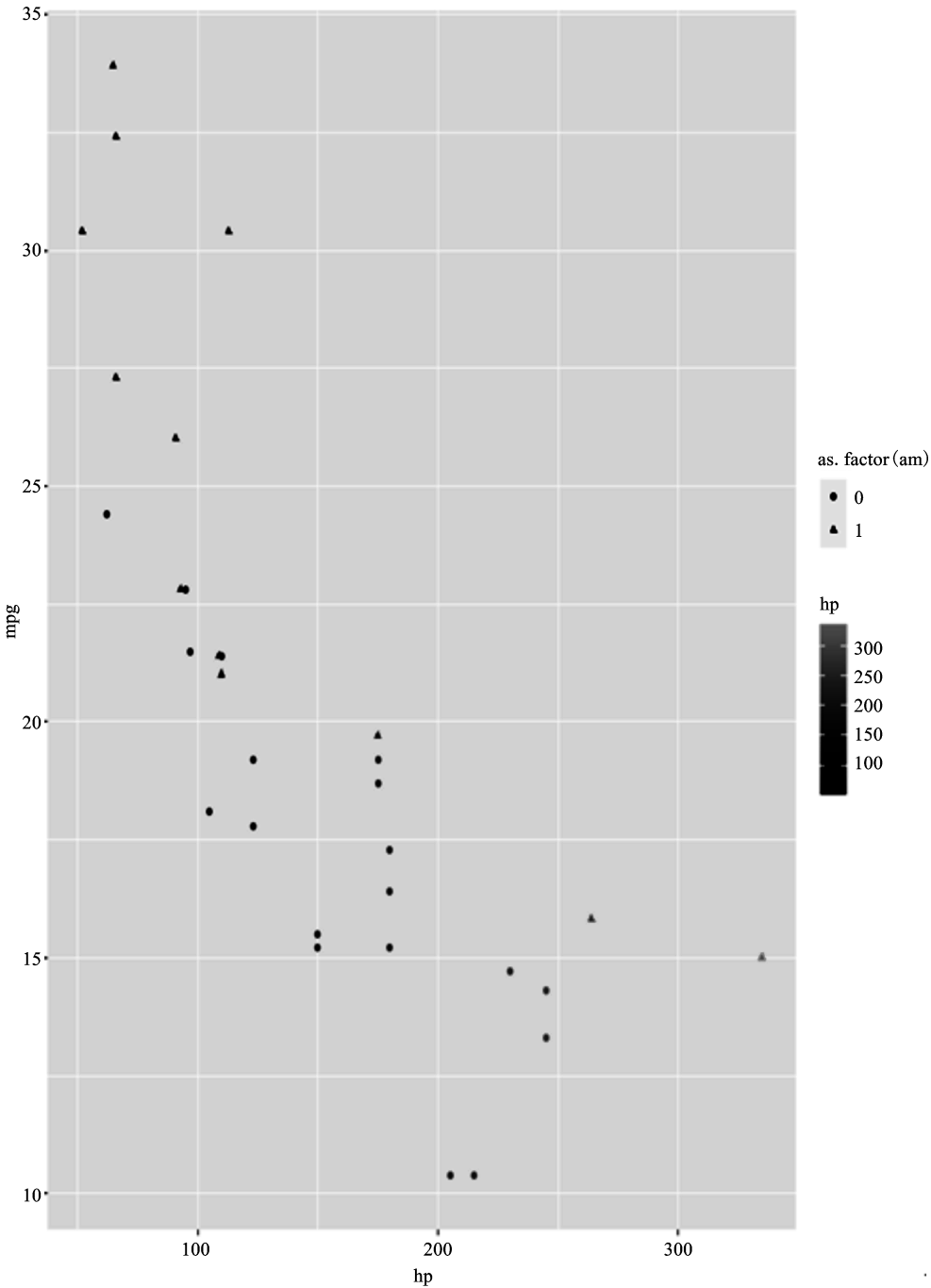


图 5-6

③通过转换因子(factor)涂色。

```
ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(am), colour=factor(hp)))+geom_point()
```

执行效果如图 5-7 所示。

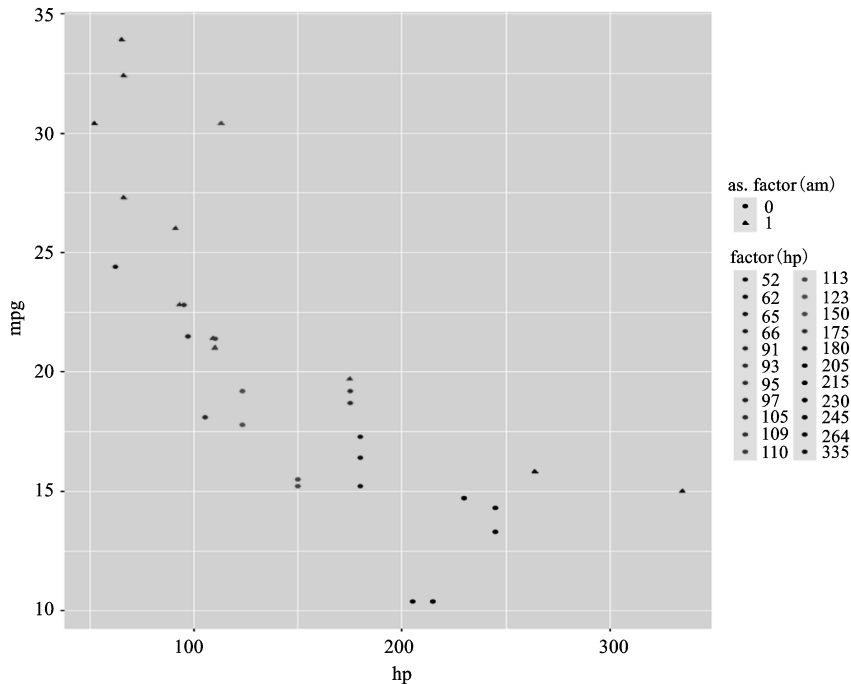


图 5-7

```
ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(am), colour=factor(mpg)))+geom_point()
```

执行效果如图 5-8 所示。

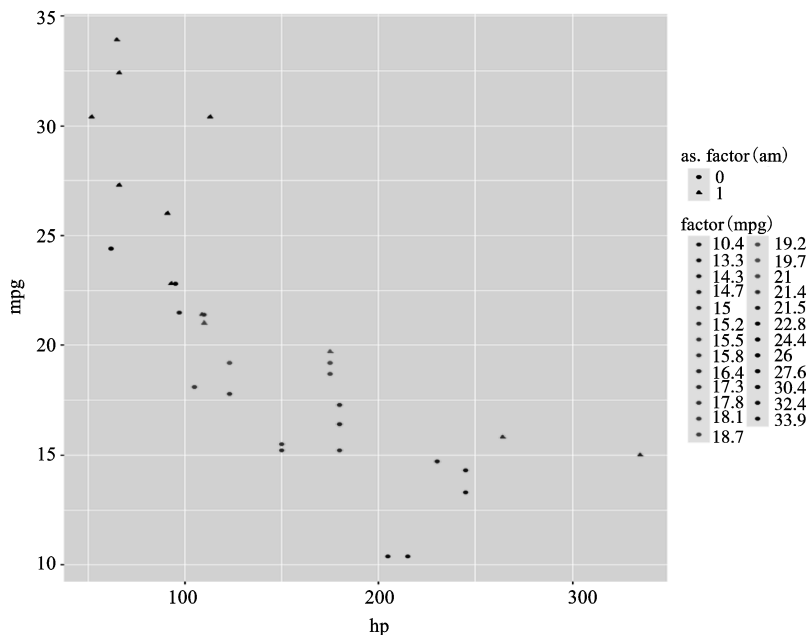


图 5-8

(2)使用传输类型(cyl)作为图例形状绘图

```
ggplot(mtcars,aes(x=hp,y=mpg,shape=as.factor(cyl)))+geom_point()
```

①调用 ggplot2 包中的 ggplot 函数绘图；上述代码中，参数指定使用 mtcars 数据库作为绘图数据来源；参数 aes 中设置横纵坐标对应的向量；以 cyl 作为绘图的形状；geom_point() 函数用于绘制点。执行效果如图 5-9 所示。

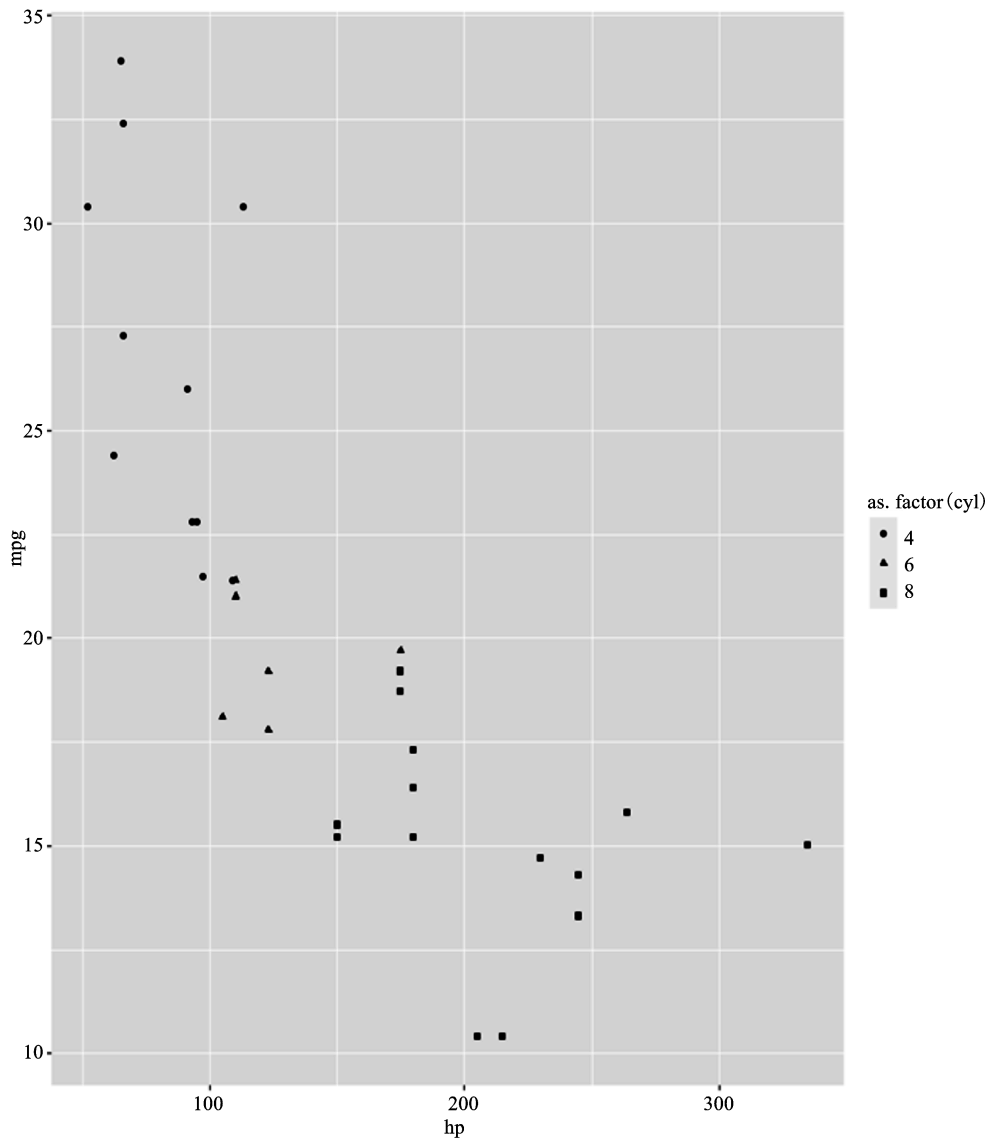


图 5-9

②在此基础上添加颜色。

```
ggplot(mtcars,aes(x=hp,y=mpg,shape=as.factor(cyl),colour=hp))+geom_point()
```

执行效果如图 5-10 所示。

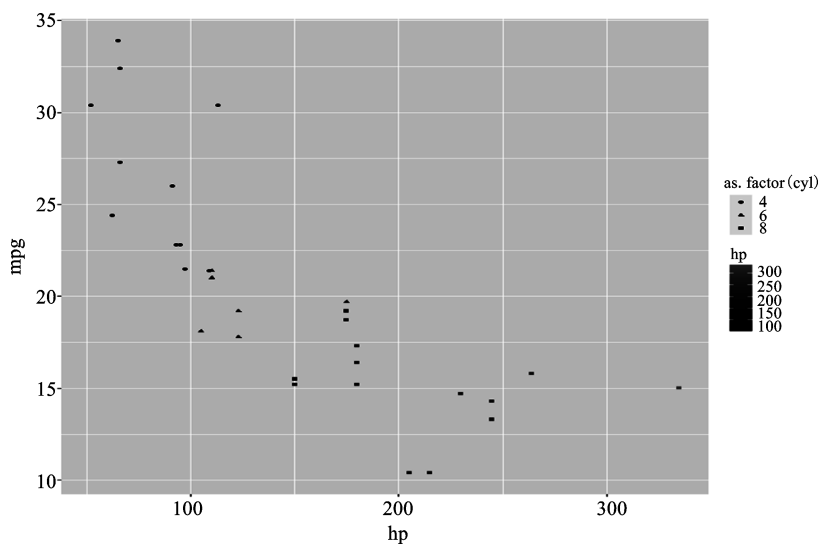


图 5-10

③通过转换因子(factor)涂色。

```
ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(cyl), colour=factor(mpg)))+geom_point()
```

执行效果如图 5-11 所示。

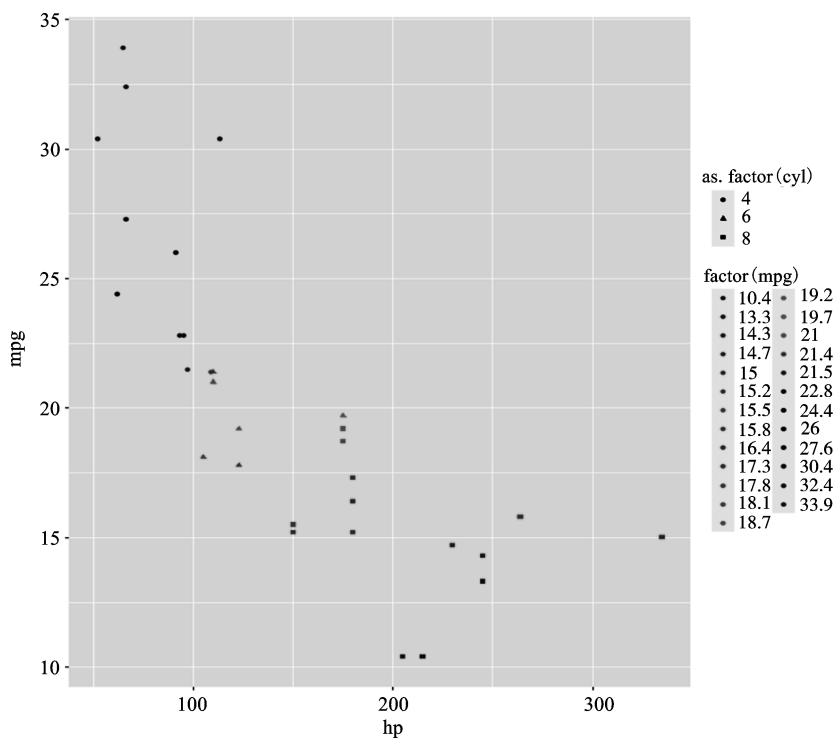


图 5-11

`ggplot(mtcars, aes(x=hp, y=mpg, shape=as.factor(cyl), colour=factor(hp)))+geom_point()`

执行效果如图 5-12。

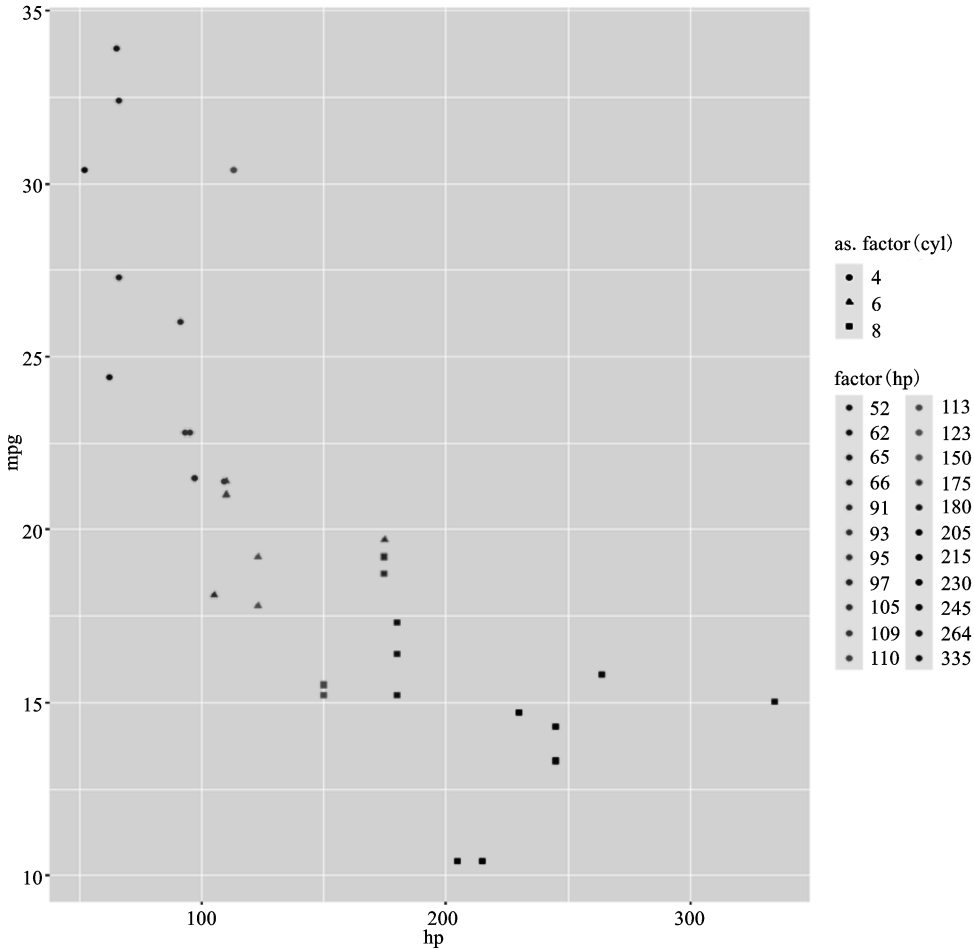


图 5-12

此外，还可以将 geom point 中加入 size 参数调节散点的大小。