

# 可视化课程作业

司徒雪颖  
中央财经大学

## 一、可视化优秀例子



图表 1 佳木斯 8 小时人口迁移图

图 1 来源：百度 8 小时迁移图 <http://qianxi.baidu.com>

数据类型：分类数据

图类型：地图，关系图

信息：快速看出某 8 小时内佳木斯人口迁移的最多的城市和省份

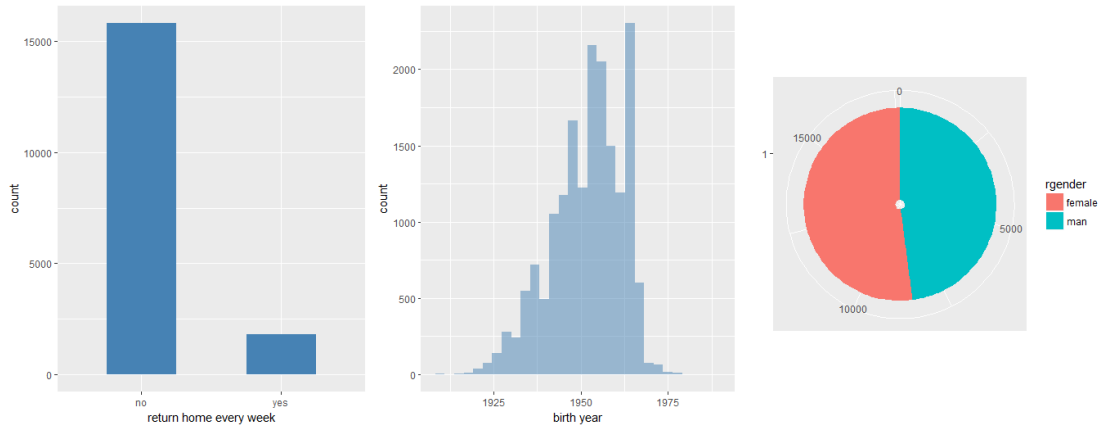
## 二、R 基础绘图和 ggplot2 包绘图对比

```
library(foreign)
library(RColorBrewer)
library(ggplot2)
mydata = read.dta("demographic_background.dta", convert.factors=F)
mydata2 = data.frame(mydata$ba005, mydata$rgender, mydata$ba002_1) #选
列
colnames(mydata2) = c("ba005", "rgender", "ba002_1")
head(mydata2)
mydata2 = na.omit(mydata2)
mydata2$rgender[mydata2$rgender=="0"] = NA
```

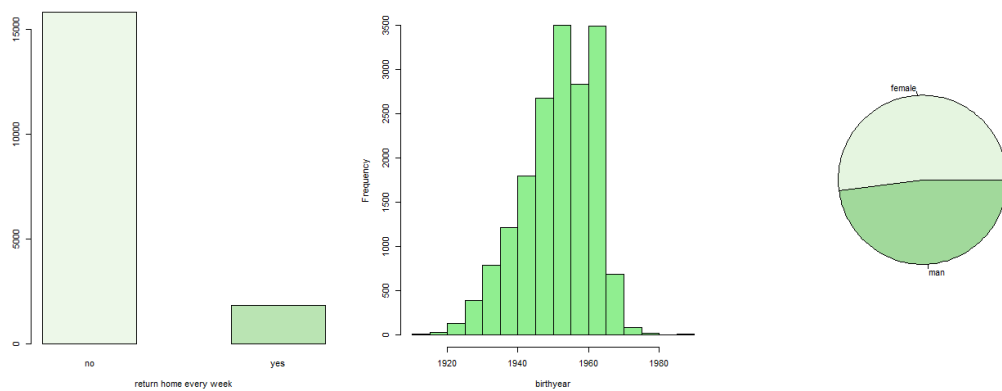
```

mydata2$rgender[mydata2$rgender==1] = "man"
mydata2$rgender[mydata2$rgender==2] = "female"
mydata2$ba005[mydata2$ba005==1] = "yes"
mydata2$ba005[mydata2$ba005==2] = "no"
#ggplot2
library("grid")
grid.newpage()
pushViewport(viewport(layout = grid.layout(1,3)))
vplayout <- function(x, y) {
  viewport(layout.pos.row = x, layout.pos.col = y)
}
#条形图
p = ggplot(mydata2, mapping = aes(ba005, fill = I("steelblue")))
p = p+ geom_bar(width = 0.5) + xlab("return home every week")
print(p, vp = vplayout(1,1))
#直方图
p = ggplot(mydata2, aes(ba002_1))
p = p+geom_histogram(bins = 30, aes(y = ..count..), alpha = 0.5, fill =
I("steelblue"))
p = p+xlab("birth year")
print(p, vp = vplayout(1,2))
#饼图
p = ggplot(na.omit(mydata2), aes(x = factor(1), fill = rgender)) +
geom_bar()
p = p + coord_polar(theta = 'y') +ylab("")+xlab("")
print(p, vp = vplayout(1,3))
#R 基础绘图
par(mfrow = c(1,3))
barplot(table(mydata2$ba005), col = brewer.pal(5, "Greens"), space =
1, xlab = "return home every week")
hist(mydata2$ba002_1, col = "light green", xlab = "birthyear", main =
"")
pie(table(mydata2$rgender), col = brewer.pal(2, "Greens"))

```



图表 2 ggplot2 绘图案例



图表 3 R 基础绘图案例

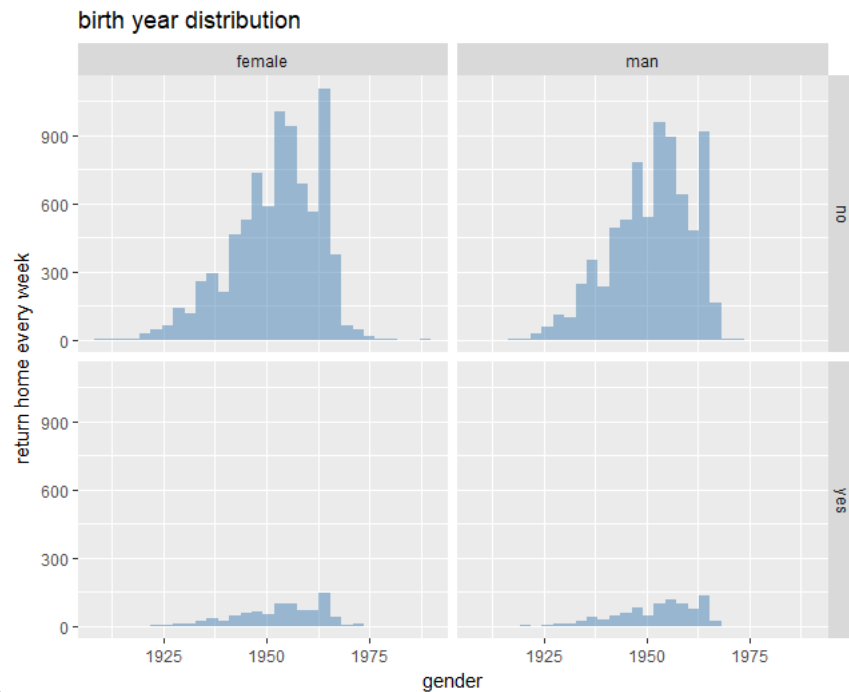
图 2 为 ggplot2 绘图，图 3 为 R 基础绘图，图 2 图 3 为相同变量不同绘制方法的对比。绘制条形图的变量为 ba003（是否周末回家），绘制直方图的变量为 ba002\_1（出生年份），绘制饼图的变量为 rgender（性别）。

ggplot2 与 R 基础绘图函数的区别在于：

1. 基础图形没有背景，ggplot2 绘图默认灰色栅栏背景。
2. 基础图形采用的是“纸-笔”的模型，只能在图形的最顶端进行画图，不能修改或删除已有的内容。ggplot2 可以让用户使轻松修改图形类型、数据等。
3. 基础绘图展示信息单一，而 ggplot2 可以通过添加分类变量，绘制分面图展示变量间的关系

```
#分面
p = ggplot(na.omit(mydata2), aes(ba002_1))
```

```
p = p+geom_histogram(bins = 30,aes(y = ..count..),alpha = 0.5,fill =
I("steelblue"))
p+facet_grid(ba005~rgender)+ylab("return home every
week")+xlab("gender")+labs(title = "birth year distribution")
```

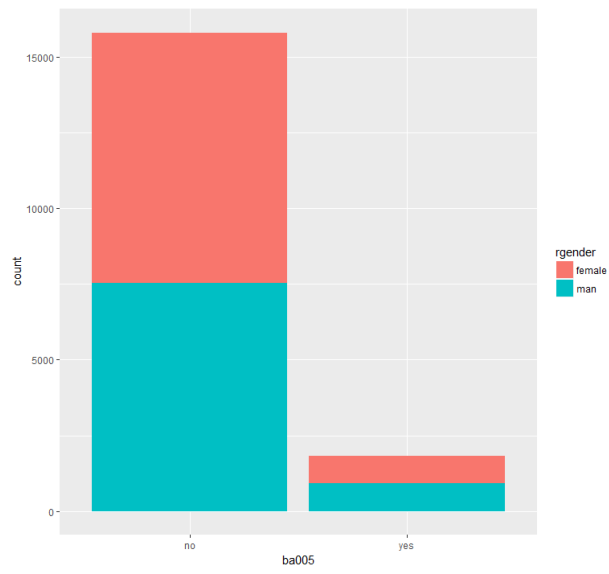


图表 4 ggplot2 分面绘图

图 4 为 ggplot2 绘制的分面图，主要对象为 ba002\_1（出生年份），横向分类为 rgender（性别），纵向分类为 ba003（是否周末回家）。从图 4 可以看出，选 yes（周末回家）的人很少，但在性别、出生年份上无差异。

### 三、考察两两变量的关系

这个数据集是问卷数据集，变量多是定类或定序变量，考察两两变量的关系，多用堆叠柱状图反映，如图 5 为 ba003 与 rgender 的堆积柱状图，没有反映出两者有明显联系。



图表 5 性别与是否回家的堆积柱状图