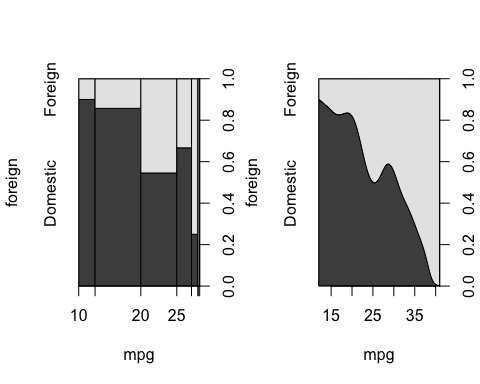
R语言-随堂实验作业16

### 请将附件中的auto.csv数据表文件读入R的工作空间中。

auto = read.csv("Datasets/auto.csv")

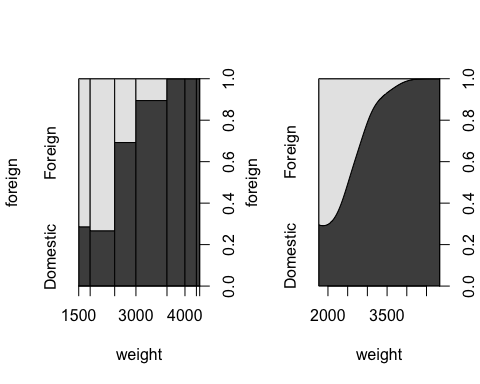
### 用合适的图形来可视化“foreign”和“mpg”两个变量之间的关系（以“foreign”为因变量，以“mpg”为自变量）。

layout(matrix(1:2, ncol = 2))  
plot(foreign ~ mpg, data = auto)  
cdplot(foreign ~ mpg, data = auto)



### 用合适的图形来可视化“foreign”和“weight”两个变量之间的关系（以“foreign”为因变量，以“weight”为自变量）。

layout(matrix(1:2, ncol = 2))  
plot(foreign ~ weight, data = auto)  
cdplot(foreign ~ weight, data = auto)



### 用逻辑回归模型评估“mpg”和“weight”变量是否值得被引入到一个预测模型中来预测一辆车是否为“Foreign”车（“foreign”变量），并解释显著的回归系数的估值的含义（基于使用“mpg”和“weight”两个自变量的模型来解释）。

auto=read.csv("Datasets/auto.csv",header=TRUE)  
str(auto)

## 'data.frame': 74 obs. of 4 variables:  
## $ make : Factor w/ 74 levels "AMC Concord",..: 1 2 3 7 8 9 10 11 12 13 ...  
## $ mpg : int 22 17 22 20 15 18 26 20 16 19 ...  
## $ weight : int 2930 3350 2640 3250 4080 3670 2230 3280 3880 3400 ...  
## $ foreign: Factor w/ 2 levels "Domestic","Foreign": 1 1 1 1 1 1 1 1 1 1 ...

auto\_logit <- glm(foreign ~ weight+mpg, data = auto, family = binomial())  
summary(auto\_logit)

##   
## Call:  
## glm(formula = foreign ~ weight + mpg, family = binomial(), data = auto)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.0436 -0.4285 -0.2207 0.5347 2.0679   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 13.708367 4.518709 3.034 0.002416 \*\*   
## weight -0.003907 0.001012 -3.862 0.000113 \*\*\*  
## mpg -0.168587 0.091917 -1.834 0.066637 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 90.066 on 73 degrees of freedom  
## Residual deviance: 54.350 on 71 degrees of freedom  
## AIC: 60.35  
##   
## Number of Fisher Scoring iterations: 6

##### 解答：逻辑回归模型为:.其中是“第个观测为Foreign车”的概率。是“第个观测为Foreign车”的优势的自然对数。（1）对应的p值等于0.000113，因此在的显著性水平上，我们推翻的原假设，因此我们得出weight变量值得被引入到模型中的结论。对应的p值等于0.066637，因此在的显著性水平上，我们不能推翻的原假设，因此我们得出mpg变量不值得被引入到模型中的结论。（2）基于如上使用“mpg”和“weight”两个自变量的模型，我们对显著的回归系数的解释如下（注：不显著的回归系数因为不能推翻其为0的原假设，所以可以不用进行解释了）：,即,所以，基于我们的数据，在其他变量的值不变的情况下，我们估计weight每增加1磅，“车为Foreign车”的优势变为原来的99.6%,这个结论在0.05的显著性水平上显著。