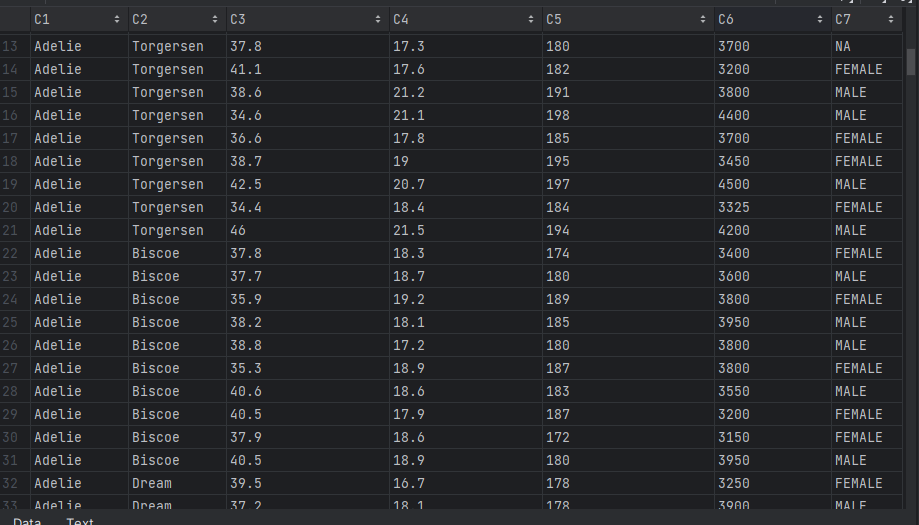
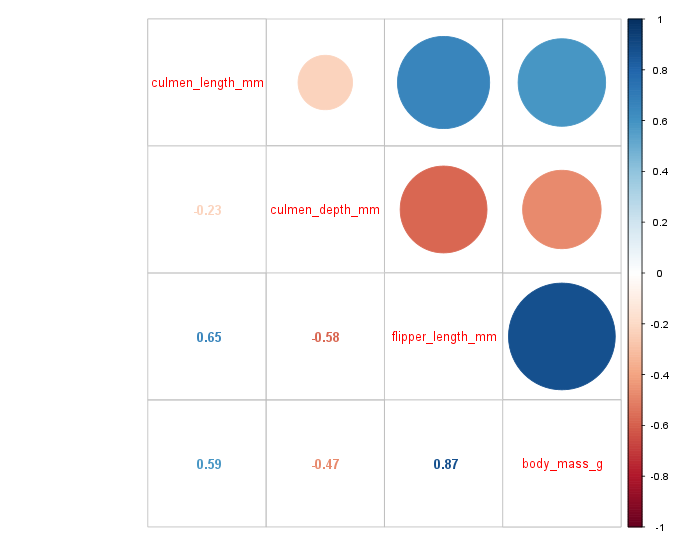
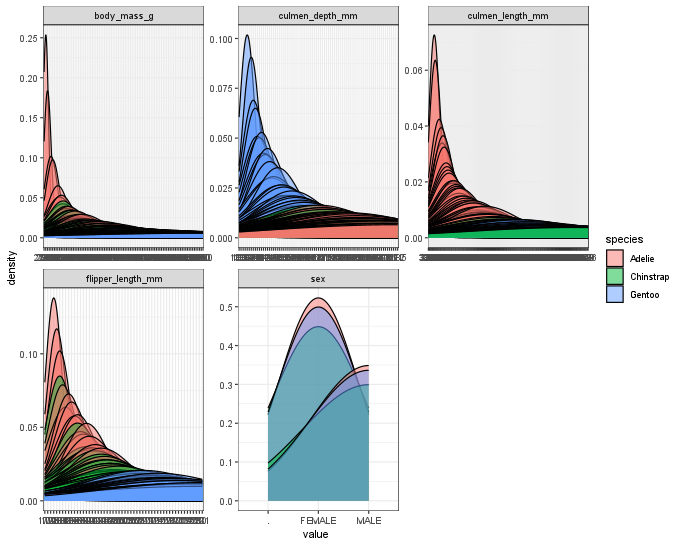
基于所给数据集”penguins\_size.csv”，可以进行回归分析研究企鹅大小体重在不同群系、种群间的差异，使用决策树分析研究企鹅特征分类。



首先使用Logistc法进行回归分析：



获取种群密度图

Logistic回归物种

Call:

glm(formula = species ~ culmen\_length\_mm + culmen\_depth\_mm +

body\_mass\_g, family = "binomial", data = peng\_spec\_train)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -36.436194 34.676402 -1.051 0.293

culmen\_length\_mm 3.731275 2.412647 1.547 0.122

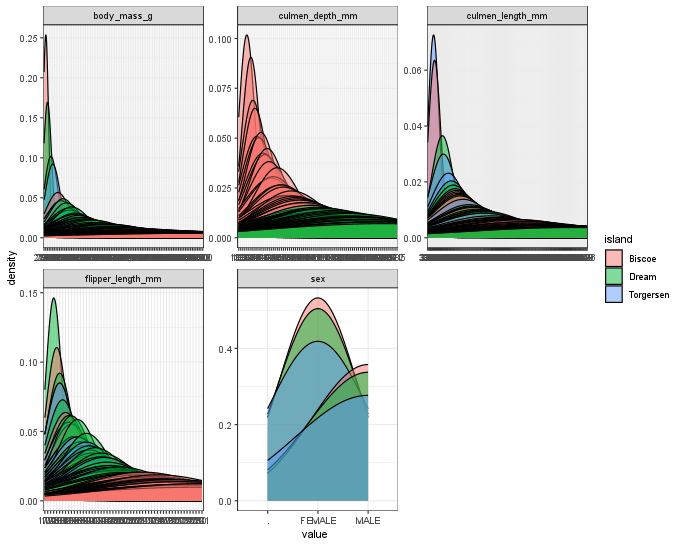
culmen\_depth\_mm -6.131967 4.058482 -1.511 0.131

body\_mass\_g -0.004344 0.003171 -1.370 0.171

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 330.5988 on 240 degrees of freedom

获取岛屿密度图

Logistic回归岛屿

Call:

glm(formula = island ~ culmen\_length\_mm + culmen\_depth\_mm + body\_mass\_g,

family = "binomial", data = peng\_isl\_train)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -10.089019 3.381820 -2.983 0.00285 \*\*

culmen\_length\_mm 0.151013 0.051489 2.933 0.00336 \*\*

culmen\_depth\_mm 0.813049 0.148003 5.493 3.94e-08 \*\*\*

body\_mass\_g -0.002577 0.000426 -6.049 1.46e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 333.99 on 240 degrees of freedom

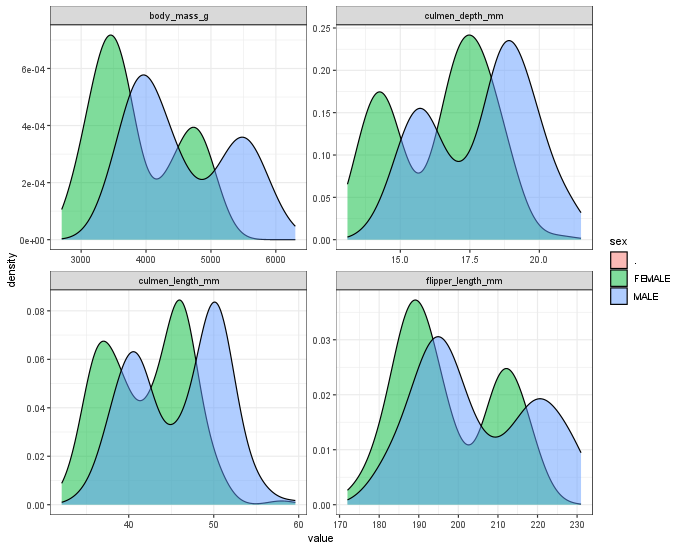
Residual deviance: 159.69 on 237 degrees of freedom

(因为不存在，1个观察量被删除了)

AIC: 167.69

Number of Fisher Scoring iterations: 6

获取性别密度图



Logistc回归性别

Call:

glm(formula = sex ~ culmen\_length\_mm + culmen\_depth\_mm + flipper\_length\_mm +

body\_mass\_g, family = "binomial", data = peng\_sex\_train)

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 51.687957 9.431095 5.481 4.24e-08 \*\*\*

culmen\_length\_mm -0.194732 0.065659 -2.966 0.00302 \*\*

culmen\_depth\_mm -1.968162 0.292037 -6.739 1.59e-11 \*\*\*

flipper\_length\_mm 0.072714 0.044516 1.633 0.10237

body\_mass\_g -0.005726 0.001041 -5.503 3.73e-08 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 324.38 on 233 degrees of freedom

Residual deviance: 113.54 on 229 degrees of freedom

(因为不存在，1个观察量被删除了)

AIC: 123.54

Number of Fisher Scoring iterations: 7