**Practical No: 7**

**Aim: Logistics Regression**

> library(datasets)

> ir\_data<- iris

> head(ir\_data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 5.1 3.5 1.4 0.2 setosa

2 4.9 3.0 1.4 0.2 setosa

3 4.7 3.2 1.3 0.2 setosa

4 4.6 3.1 1.5 0.2 setosa

5 5.0 3.6 1.4 0.2 setosa

6 5.4 3.9 1.7 0.4 setosa

> str(ir\_data)

'data.frame': 150 obs. of 5 variables:

$ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...

$ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...

$ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...

$ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...

$ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

> levels(ir\_data$Species)

[1] "setosa" "versicolor" "virginica"

> sum(is.na(ir\_data))

[1] 0

> ir\_data<-ir\_data[1:100,]

> set.seed(100)

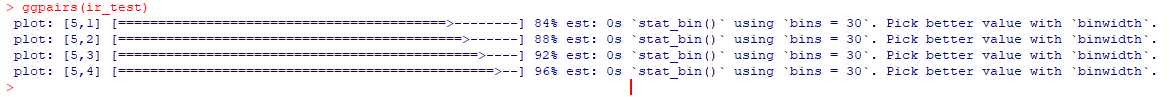
> samp<-sample(1:100,80)

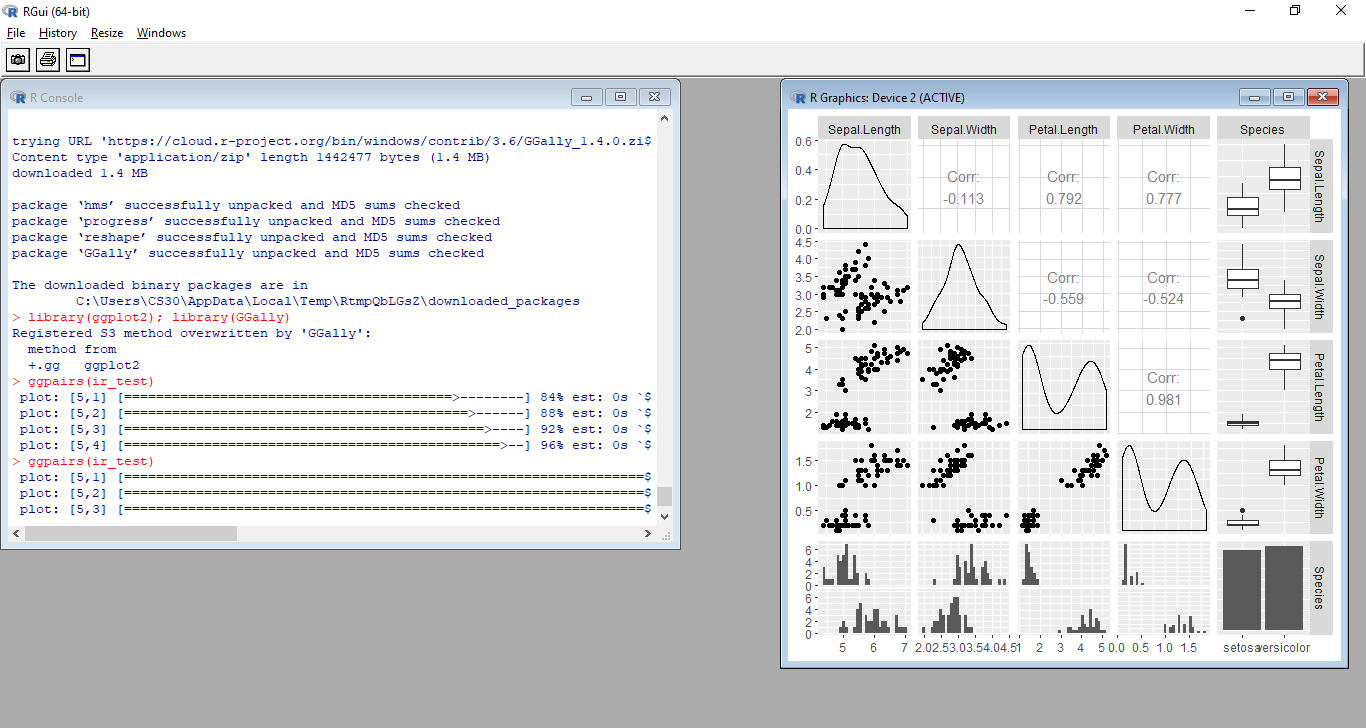
> ir\_test<-ir\_data[samp,]

> ir\_ctrl<-ir\_data[-samp,]

> library(ggplot2); library(GGally)

> ggpairs(ir\_test)





> y<-ir\_test$Species; x<-ir\_test$Sepal.Length

> glfit<-glm(y~x, family = 'binomial')

> summary(glfit)

Call:

glm(formula = y ~ x, family = "binomial")

Deviance Residuals:

Min 1Q Median 3Q Max

-1.94538 -0.50121 0.04079 0.45923 2.26238

Coefficients:

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

(Intercept) -25.386 5.517 -4.601 4.20e-06 \*\*\*

x 4.675 1.017 4.596 4.31e-06 \*\*\*

---

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 110.854 on 79 degrees of freedom

Residual deviance: 56.716 on 78 degrees of freedom

AIC: 60.716

Number of Fisher Scoring iterations: 6

> newdata<- data.frame(x=ir\_ctrl$Sepal.Length)

> predicted\_val<-predict(glfit, newdata, type="response")

> prediction<-data.frame(ir\_ctrl$Sepal.Length, ir\_ctrl$Species,predicted\_val)

> prediction

ir\_ctrl.Sepal.Length ir\_ctrl.Species predicted\_val

1 5.1 setosa 0.176005274

2 4.7 setosa 0.031871367

3 4.6 setosa 0.020210042

4 5.0 setosa 0.118037011

5 4.6 setosa 0.020210042

6 4.3 setosa 0.005048194

7 4.6 setosa 0.020210042

8 5.2 setosa 0.254235573

9 5.2 setosa 0.254235573

10 5.0 setosa 0.118037011

11 5.0 setosa 0.118037011

12 6.6 versicolor 0.995801728

13 5.2 versicolor 0.254235573

14 5.8 versicolor 0.849266756

15 6.2 versicolor 0.973373695

16 6.6 versicolor 0.995801728

17 5.5 versicolor 0.580872616

18 6.3 versicolor 0.983149322

19 5.7 versicolor 0.779260130

20 5.7 versicolor 0.779260130

> qplot(prediction[,1], round(prediction[,3]), col=prediction[,2], xlab = 'Sepal Length', ylab = 'Prediction using Logistic Reg.')

