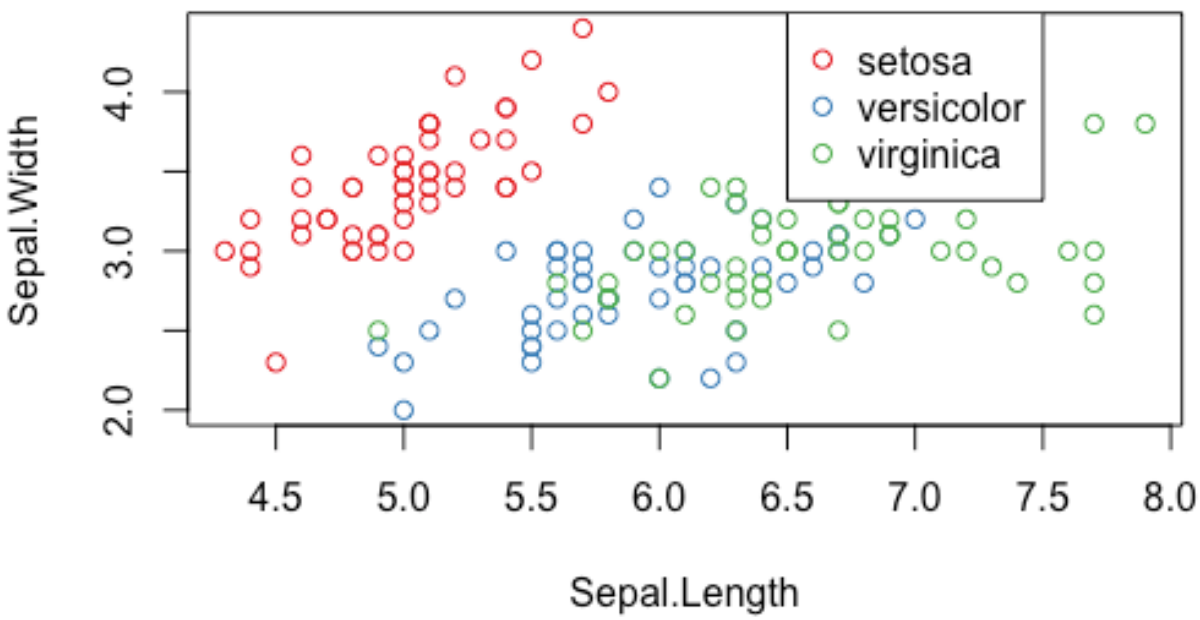
IRIS DATASET

1. Two classes in IRIS are more “similar” to each other. Find which ones using scatter plots. Lets say class 1 and class 2.

A.

```
data<- read.csv("train.csv")
IRIS1 <- data[1:86,2:6]
head(mydata)
plot(Sepal.Width ~ Sepal.Length, data=iris, col=brewer.pal(3, "Set1")[iris$Species])
legend(x=6.5, y=4.5, legend=levels(iris$Species), col=brewer.pal(3, "Set1"), pch=1)
```




From the above plot it is clear that setosa looks totally different and versicoloured and veronica are similar.

******************************************************************************

2. **Lets create a “meta-class” combining class 1 and class 2 (or whichever are the two most similar classes). Lets call it class 4.**

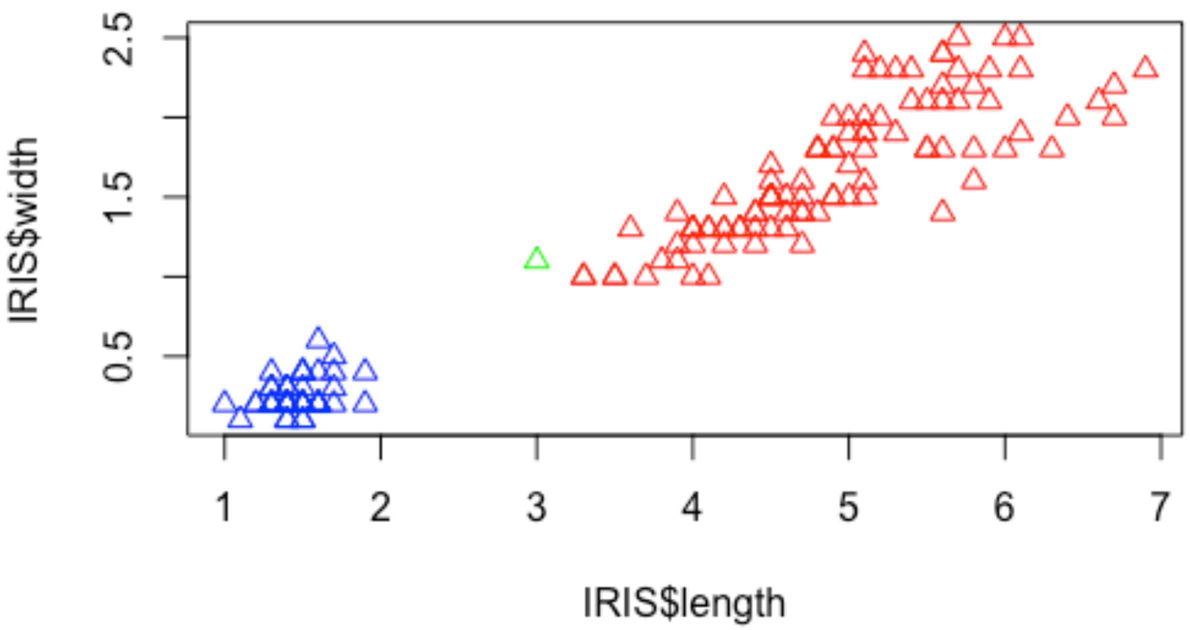
In the above graph I have decided to us both versicoloured and virginica in a single class and setosa in a different class. And using the below code I am performing the action, by creating a new column class which has value 1 for setosa and -1 for versicoloured and virginica.

```
library(MASS)
IRIS <- data.frame(length=IRIS1$Petal.Length,width=IRIS1$Petal.Width)
IRIS$class <- ifelse(IRIS1$Species == "setosa", +1, -1)
```

3. **Create the first Fisher projection by trying to discriminate class 3 (the different class) from class 4 (the meta-class).**
**Do this on training data only**

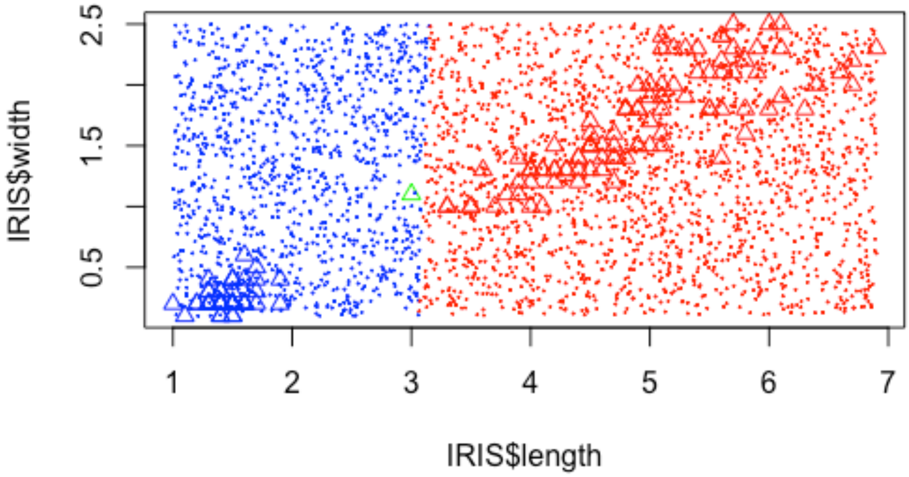
Using the below code I have created the LDA analysis and predicted the values and constructed the plot fo the predicted values.

```
LDA <- lda(class ~ length + width, data = IRIS)
IRIS$pred <- predict(LDA)$class
IRIS$colour <- 2+ifelse(IRIS$pred == IRIS$class, as.numeric(IRIS$class), 0)
color <- c("red" ,"green" , "blue" ,"red" ,"blue")
plot(IRIS$length,IRIS$width,col=color[IRIS$colour],pch=2,cex=1)
```




The above the plot clearly shows that blue is setosa and red could anything between viginica and versicoloured and green is the one predicted wrongly as setosa.
Below lets create few dummy values and create area for it.

```
rpt <- 3000
rpts <- function(v,rpt){runif(rpt,min=v[1],max=v[2])}
grid <- data.frame(apply(apply(IRIS[,1:2],2,range),2,rpts,pts))
grid$pred <- 3+as.numeric(predict(LDA,grid)$class)
points(grid$length,grid$width,col=color[grid$pred],cex=0.1) > sum(IRIS$class != IRIS$pred)
```

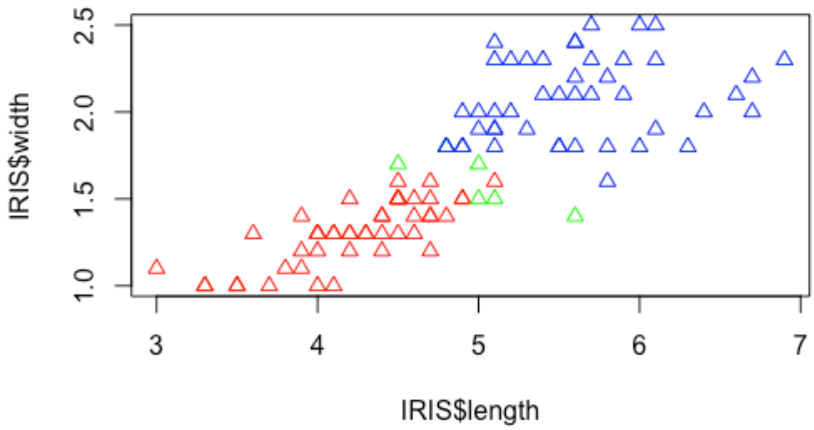



*****************************************************************************

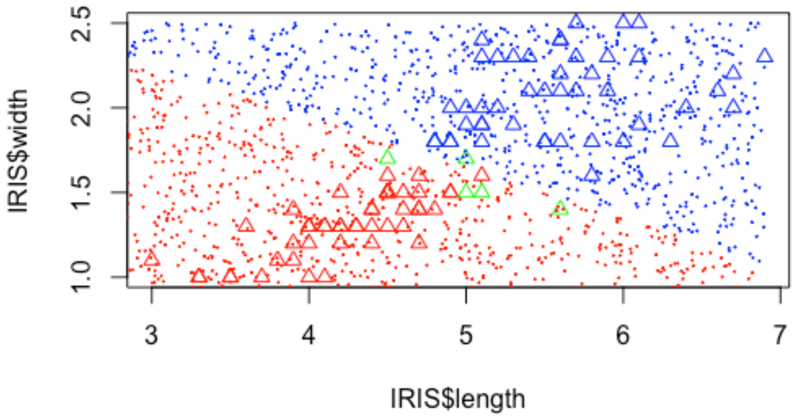
**4. Create the second Fisher projection by trying to discriminate class 1 from class 2 (the original two similar classes). Do this on training data only**

A. Lets now discriminate between versicoloured and Virginia and see what happens.
In the below we have excluded setosa from the dataset and performed LDA only on th petal length and width. In the below plot red is versicoloured and blue is Virginia and green are the versicoloured who are identified wrong.

```
IRIS1<-IRIS1[IRIS1$Species!="setosa",]
 IRIS <- data.frame(length=IRIS1$Petal.Length,width=IRIS1$Petal.Width)
 IRIS$class <- ifelse(IRIS1$Species == "virginica", +1, -1)
 LDA <- lda(class ~ length + width, data = IRIS)
 IRIS$pred <- predict(LDA)$class
 IRIS$colour <- 2+ifelse(IRIS$pred == IRIS$class, as.numeric(IRIS$class), 0)
 color <- c("red","green", "blue","red","blue")
 plot(IRIS$length,IRIS$width,col=color[IRIS$colour],pch=2,cex=1)
```



```
rep <- 3000
rpts <- function(v,pts){runif(pts,min=v[1],max=v[2])}
grid <- data.frame(apply(apply(X[,1:2],2,range),2,rpts,rep))
grid$pred <- 3+as.numeric(predict(LDA,grid)$class)
points(grid$length,grid$width,col=color[grid$pred],cex=0.1) > sum(X$class != X$pred)
```




********************************************************************************

**6.. Now project the entire data in these two projections and color code the class points. Dothisontestdataonly.**

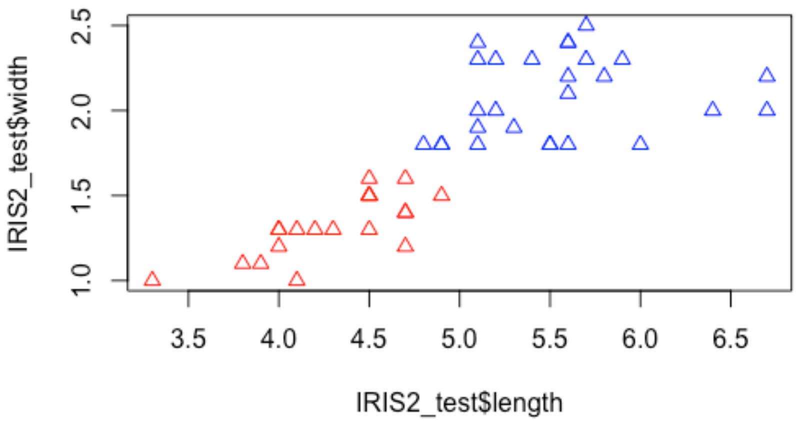
**A.**

```
test<- read.csv("test.csv")
dim(test)
head(test)
```

```
IRIS1_test <- data.frame(length=test$Petal.Length,width=test$Petal.Width)
IRIS1_test$class <- ifelse(test$Species == "setosa", +1, -1)
IRIS1_test$pred <- predict(LDA_2,IRIS1_test)$class
color <- c("red","green", "blue","red","blue")

IRIS2_test$colour <- 2+ifelse(IRIS2_test$pred == IRIS2_test$class, as.numeric(IRIS2_test$class), 0)
plot(IRIS2_test$length,IRIS2_test$width,col=color[IRIS2_test$colour],pch=2,cex=1)
```

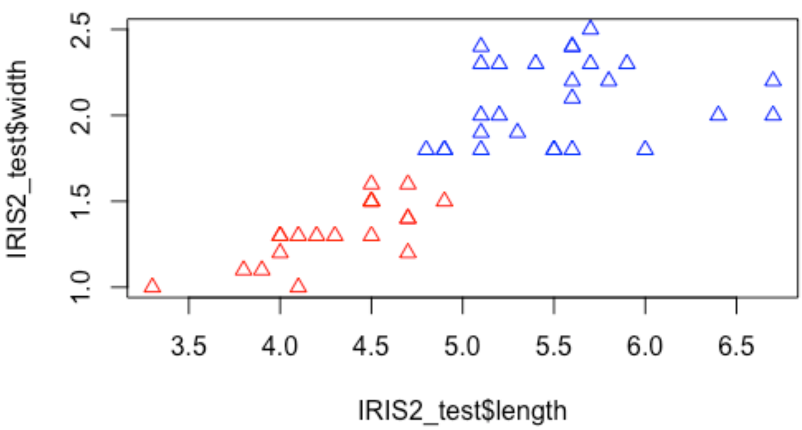
The above code works on the training data to distinct between setosa and other two class. The plot is below. The prediction is 100% accurate.




```
IRIS1_test<-test[test$Species!="setosa",]
IRIS2_test <-
data.frame(length=IRIS1_test$Petal.Length,width=IRIS1_test$Petal.Width)
IRIS2_test$class <- ifelse(IRIS1_test$Species == "virginica", +1, -1)
IRIS2_test$pred <- predict(LDA_2,IRIS2_test)$class
color <- c("red","green", "blue","red","blue")

IRIS2_test$colour <- 2+ifelse(IRIS2_test$pred == IRIS2_test$class, as.numeric(IRIS2_test$class), 0)
plot(IRIS2_test$length,IRIS2_test$width,col=color[IRIS2_test$colour],pch=2,cex=1)
```

Above is the code to distinct between Virginia and veriscolor on the test data and all the predicted values looks to be correct.




**********************************************************************************

**6. Comment on what you observed and did.**

**A.** We have used the LDA analysis to distinguish between different class , though there was no 100% accuracy in training data set we saw the 100% accuracy in prediction value. The other interesting point was we could distinguish between setosa and other two classes using just the petal length and width sepal details were not so much in use.

************************************************************************************

Acuracy Table

| Question | Data Set | Data type | Algorithm | Accuracy | |
|---|---|---|---|---|---|
| **P1** | **Iris** | Train | Fisher | 98% | Distinct  class sets and other two |
| **P1** | **Iris** | Train | Fisher | 88% | Between virgina and versacolor |
| **P1** | **Iris** | train | Fisher | 93% | Full data |
| **P1** | **Iris** | Test | Fisher | 100% | Full Data |