

Exercise Set 2-2

1) Here is code to do the whole procedure. You'll see that petal length and width are strongly associated and that all three species can be reasonably clearly distinguished on the basis of petal length and width.

```
#Compute the sample mean and median
mean(iris$Petal.Length)
median(iris$Petal.Length)

#Histograms provide a visual summary of the distribution
hist(iris$Petal.Length, xlab = "Petal Length")
mean(iris$Petal.Length[iris$Species=="setosa"])

#We'll skip the conditional.mean function and use tapply()
#instead.
tapply(iris$Petal.Length, iris$Species, mean)
tapply(iris$Petal.Width, iris$Species, mean)

#A basic boxplot
boxplot(iris$Petal.Length ~ iris$Species)
title(xlab = "Species", ylab = "Petal Length") #Add axis labels

#A scatterplot
plot(iris$Petal.Length, iris$Petal.Width, pch =
as.numeric(iris$Species),
      xlab = "Petal Length", ylab = "Petal Width")
#And add a legend
legend("topleft", pch = c(1,2,3), legend = c("setosa",
"versicolor", "virginica"))
```

2) Code is in the question. The `gpairs()` plot includes scatterplots for every possible pair of variables in the `iris` dataset.