GEOG 5680 Example Markdown

Simon Brewer

May 13, 2015

Iris dataset: plots and correlations

Introduction

We are doing some exploratory statistics on the Fisher Iris dataset. Start by reading the file:

```
iris = read.csv("iris.csv")
```

Summary statistics

Calculate summary stats on all columns:

```
summary(iris)
```

```
##   Sepal.Length    Sepal.Width    Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species         Code
##  setosa     :50   Min.   :1
##  versicolor:50    1st Qu.:1
##  virginica :50    Median :2
##                   Mean   :2
##                   3rd Qu.:3
##                   Max.   :3
```

Make a plot

Just a scatterplot:

```
plot(iris$Sepal.Length, iris$Sepal.Width,
     col=iris$Species, pch=16)
```
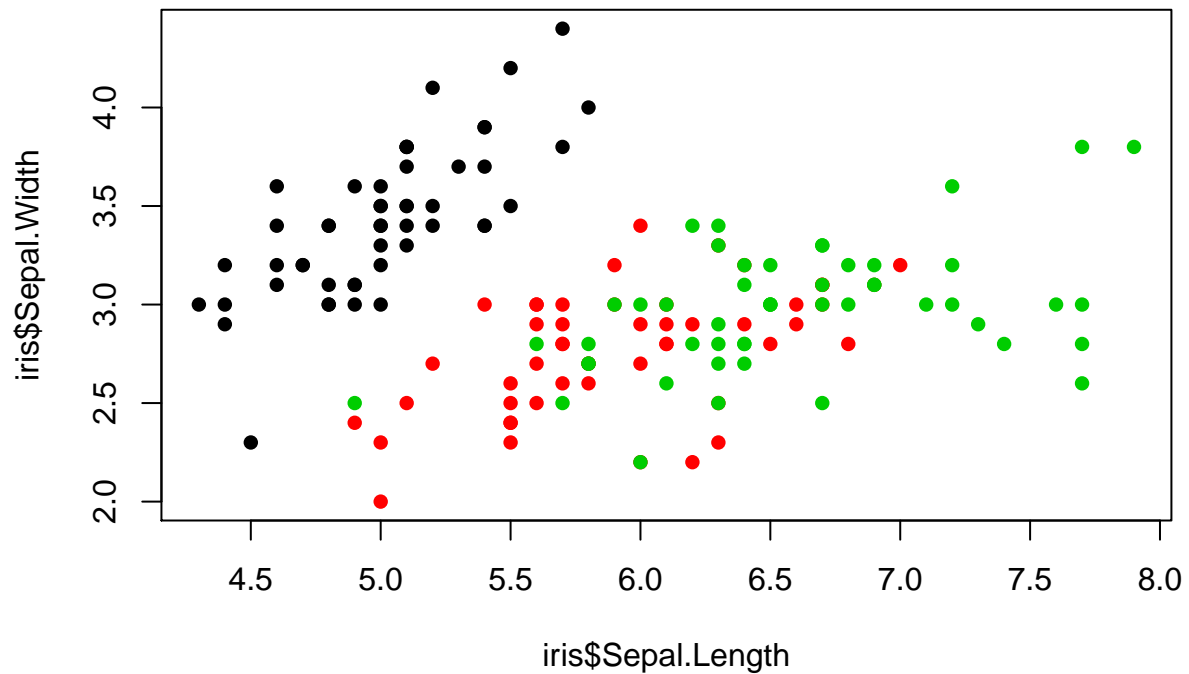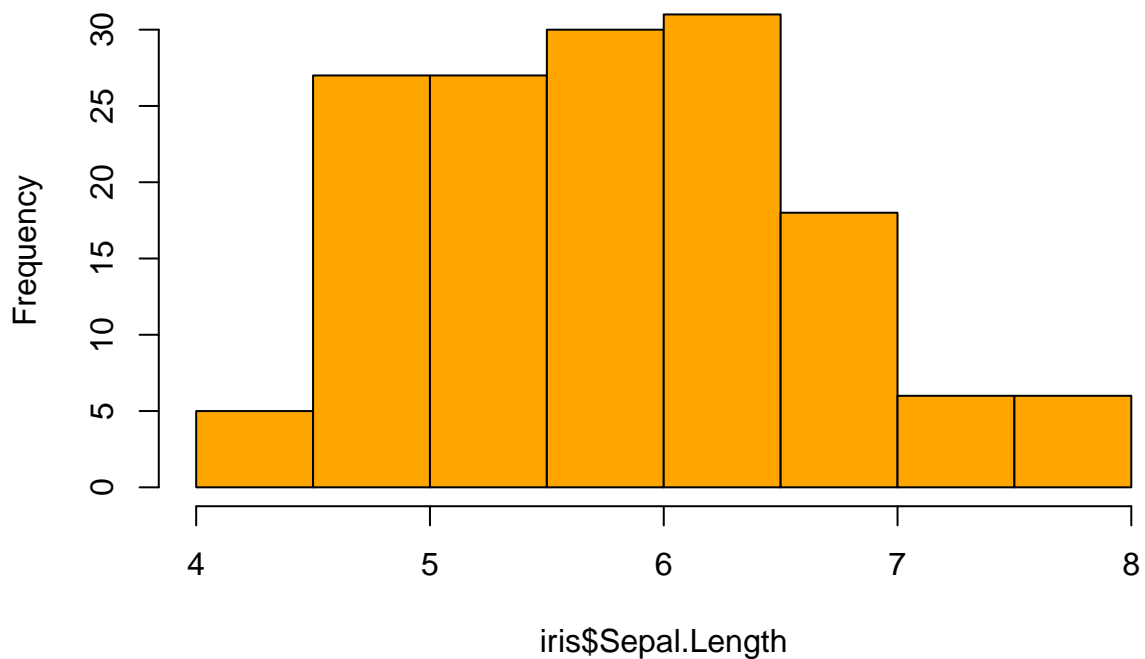
And a histogram:

```r
hist(iris$Sepal.Length, col="orange")
```

Histogram of iris$Sepal.Length

Correlations

Start this with a list

- Correlation only the variables
- These are columns 1 to 4

```r
cor(iris[,1:4])
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
## Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
## Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
## Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
```

Be aware of correlations!

Discussion

Well, we have found out lots of things about the Iris. Flowers with wide petals tend to have long petals. Who knew?