

Communicating Data with R Markdown Example

Daniel Fitton

23/04/2020

Edgar Anderson's Iris Data

This famous (Fisher's or Anderson's) iris data set gives the measurements in centimeters of the variables sepal length and width and petal length and width, respectively, for 50 flowers from each of 3 species of iris. The species are *Iris setosa*, *versicolor*, and *virginica*.

Here is a summary of the data:

```
## 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
## setosa :50
## versicolor:50
## virginica :50
##
##
##
```

Table

Here is an aggregated view of each species, calculated by taking the mean of each column. It also includes a new column which calculates the Total Length.

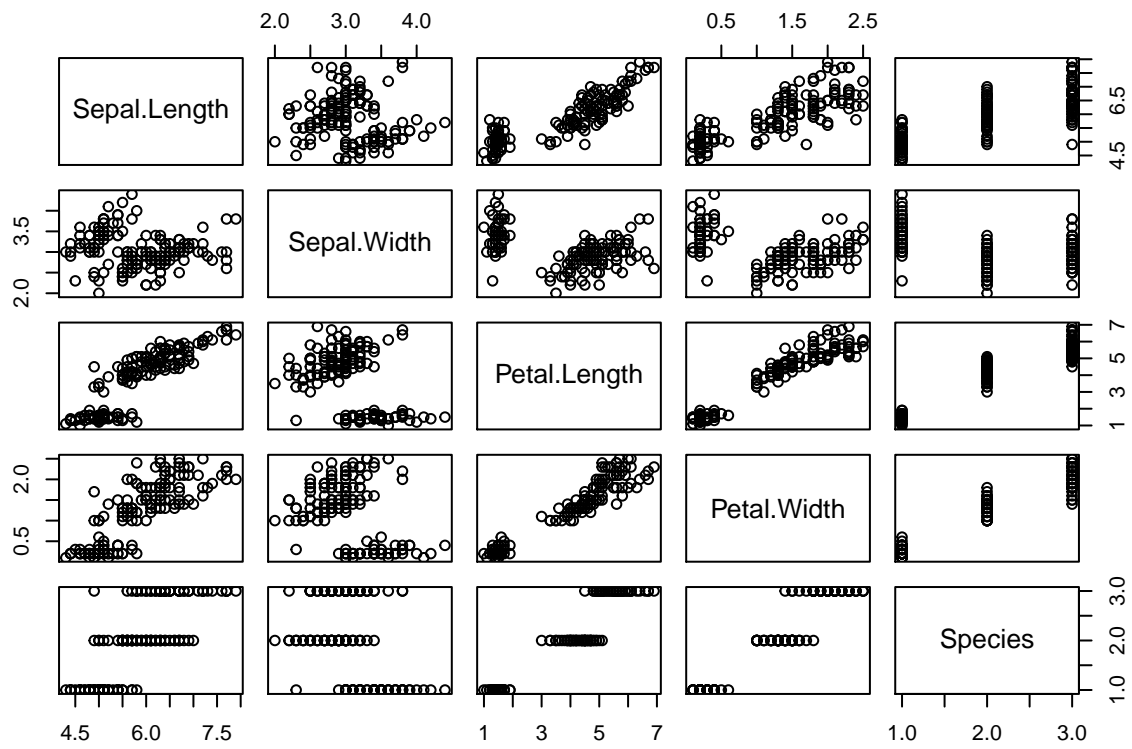
Species	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Total.Length
virginica	6.588	2.974	5.552	2.026	12.140
versicolor	5.936	2.770	4.260	1.326	10.196
setosa	5.006	3.428	1.462	0.246	6.468

Pair Plot

This visualisation shows how the distribution of data when comparing each variable in the dataset. There appears to be a positive correlation between the following:

- Sepal Length and Petal Length
- Sepal Length and Petal Width
- Petal Length and Petal Width

There appears to be clear groups, or clusters, in the dataset when comparing these variables as well. There doesn't, however, appear to be a clear correlation between Sepal Length and Sepal Width.



Petal Length vs. Petal Width

The clusters in the dataset appear to be caused by the type of iris species.

