My Example Document

Brandon Kopp

7/13/2019

Fisher's Iris Data

by Brandon Kopp

The famous **Fisher 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*.

`iris` is a data frame with 150 cases (rows) and 5 variables (columns) named:

1. Sepal.Length
2. Sepal.Width
3. Petal.Length
4. Petal.Width
5. Species

For more information about this data set, see Fisher's Iris Data Set on Wikipedia.

Load Libraries

Load the `dplyr` and `ggplot2` packages for later use.

```
library(dplyr)
library(ggplot2)
```

Data Transformation

Calculate new variables that represent the area of the Sepals and Petals (Length X Width).

```
iris_df <- iris %>%
  mutate(sepal_area = Sepal.Length * Sepal.Width,
         petal_area = Petal.Length * Petal.Width)
```

Data Table

Display the data using a table.

```
knitr::kable(head(iris_df))
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species	sepal_area	petal_area
5.1	3.5	1.4	0.2	setosa	17.85	0.28
4.9	3.0	1.4	0.2	setosa	14.70	0.28
4.7	3.2	1.3	0.2	setosa	15.04	0.26
4.6	3.1	1.5	0.2	setosa	14.26	0.30
5.0	3.6	1.4	0.2	setosa	18.00	0.28
5.4	3.9	1.7	0.4	setosa	21.06	0.68

Plot

Create a scatterplot of the Sepal Area and Petal Area variables calculated above.

```
ggplot(iris_df, aes(sepal_area, petal_area)) +
  geom_point(aes(color=Species), alpha=0.6, size=3) +
  labs(x="Sepal Area (in sq cm)", y="Petal Area (in sq cm)") +
  theme_bw()
```