

# รายงาน

## เรื่อง Visualization of Iris Data Set

### โดย

07610477	นายศักดิ์ณรงค์	สมบัติเจริญ
620710405	นางสาวณัฐธิดา	ลาภธนชัย
620710407	นางสาวเพชรรัตน์	สุขอุบล
620710408	นางสาวสุธาทิพย์	แย้มกลิ่น

### เสนอ

อาจารย์ จิตดำรง ปรีชาสุข

รายงานนี้เป็นส่วนหนึ่งของรายวิชา 522151 Foundation of data science

ภาคเรียนที่ 2 ปีการศึกษา 2563

คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

# TERM PROJECT

## 522151-2561 FOUNDATION OF DATA SCIENCE

### VISUALIZATION PROJECT

MEMBERS :      07610477      นายศักดิ์ณรงค์      สมบัติเจริญ  
                         620710405      นางสาวณัฐธิดา      ลาภชนชัย  
                         620710407      นางสาวเพชรรัตน์      สุขอุบล  
                         620710408      นางสาวสุธาทิพย์      แย้มกลิ่น

DATASET :      Iris Data Set      [<https://archive.ics.uci.edu/ml/datasets/Iris>]



#### Iris Data Set

Download: [Data Folder](#) [Data Set Description](#)

Abstract: Famous database; from Fisher, 1936



Data Set Characteristics:	Multivariate	Number of Instances:	150	Area:	Life
Attribute Characteristics:	Real	Number of Attributes:	4	Date Donated	1988-07-01
Associated Tasks:	Classification	Missing Values?	No	Number of Web Hits:	3181677

#### Source:

Creator:

R. A. Fisher

Donor:

Michael Marshall (MARSHALL%PLU '@'io.arc.nasa.gov)

#### Data Set Information:

This is perhaps the best known database to be found in the pattern recognition literature. Fisher's paper is a classic in the field and is referenced frequently to this day. (See Duda & Hart, for example.) The data set contains 3 classes of 50 instances each, where each class refers to a type of iris plant. One class is linearly separable from the other 2, the latter are NOT linearly separable from each other.

Predicted attribute: class of iris plant.

This is an exceedingly simple domain.

This data differs from the data presented in Fisher's article (identified by Steve Chadwick, [spchadwick '@'espeedaz.net](mailto:spchadwick '@'espeedaz.net)). The 35th sample should be: 4.9,3.1,1.5,0.2,"Iris-setosa" where the error is in the fourth feature. The 38th sample: 4.9,3.6,1.4,0.1,"Iris-setosa" where the errors are in the second and third features.

#### Attribute Information:

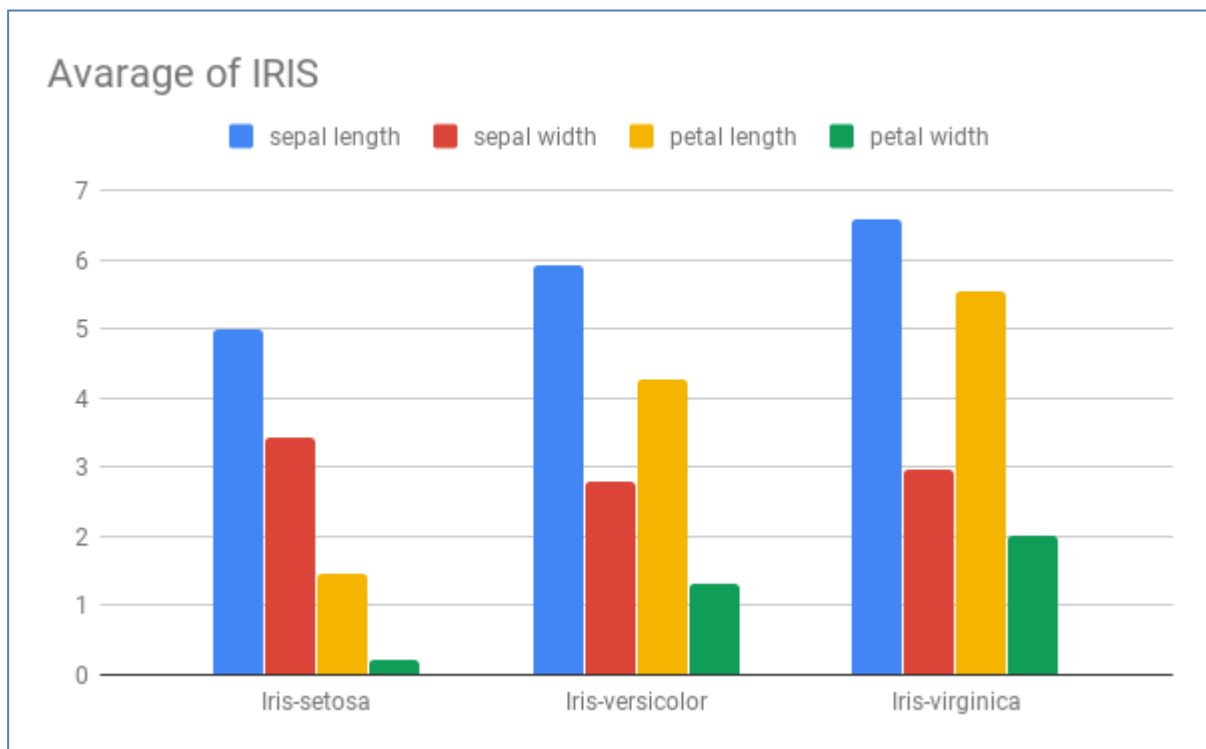
1. sepal length in cm  
2. sepal width in cm  
3. petal length in cm  
4. petal width in cm  
5. class:  
-- Iris Setosa  
-- Iris Versicolour  
-- Iris Virginica

### Dataset Name

- Attributes Number : 4
- Attributes Name : sepal length, sepal width, petal length, petal width
- Attributes Characteristics : Real
- Associated Tasks : Classification

- **Attributes Detail :**
  1. sepal length in cm
  2. sepal width in cm
  3. petal length in cm
  4. petal width in cm
  5. class:
    - Iris Setosa
    - Iris Versicolour
    - Iris Virginica
- **Class Name :** Iris Setosa (50), Iris Versicolour (50), Iris Virginica (50)
- **Number of instances :** 150

In Chart

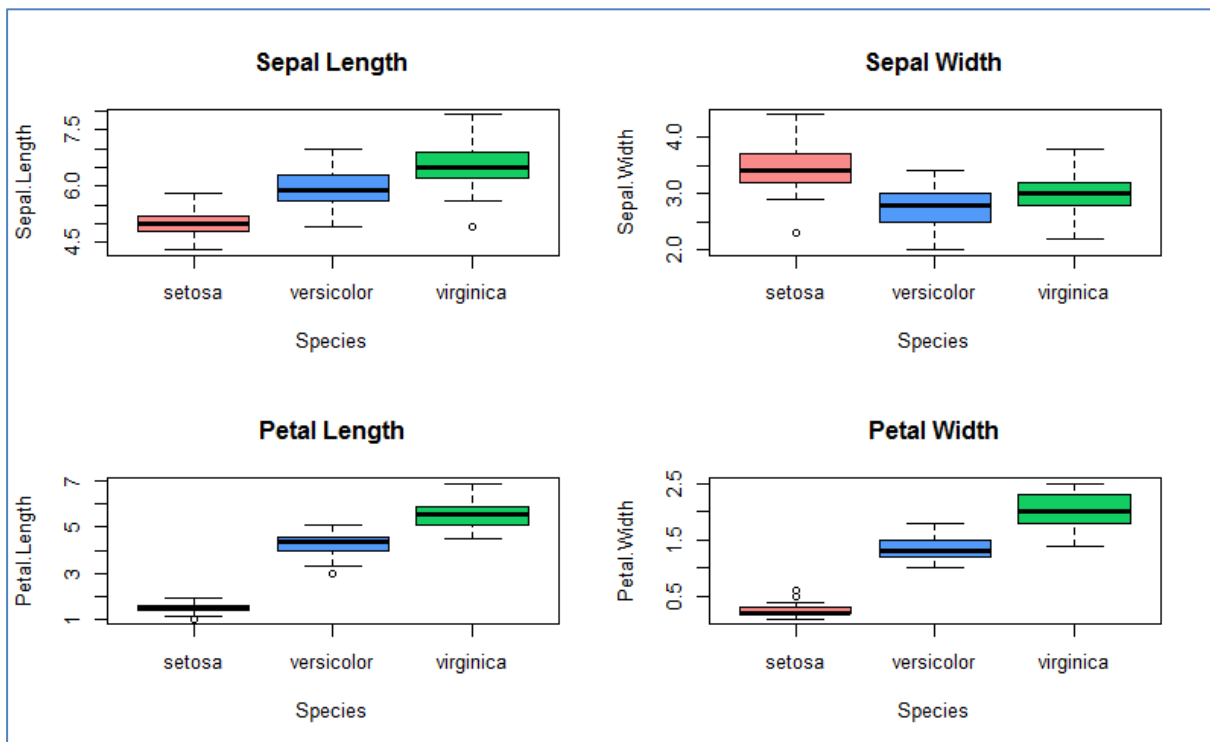


- **Type of Chart(1) :** Bar Chart / 6 มิติ
- **Point to Focus :** กราฟแท่งแสดงค่าเฉลี่ยของความยาวของกลีบเลี้ยง ความกว้างของกลีบเลี้ยง ความยาวของกลีบดอก และความกว้างของกลีบดอกของดอกไอริสแต่ละสายพันธุ์
- เพื่อการเปรียบเทียบของข้อมูล ซึ่งกราฟแท่งสามารถทำให้เข้าใจได้ง่ายขึ้น

```

> par(mfrow = c(2,2))
> boxplot (iris$Sepal.Length~iris$Species,col=c("#fa8a8a","#529afa","#11cd62"), ylab="Sepal.Length",
  xlab="Species", main="Sepal Length")
> boxplot (iris$Sepal.Width~iris$Species ,col=c("#fa8a8a","#529afa","#11cd62"), ylab="Sepal.Width" ,
  xlab="Species", main="Sepal Width" )
> boxplot (iris$Petal.Length~iris$Species,col=c("#fa8a8a","#529afa","#11cd62"), ylab="Petal.Length",
  xlab="Species", main="Petal Length")
> boxplot (iris$Petal.Width~iris$Species ,col=c("#fa8a8a","#529afa","#11cd62"), ylab="Petal.Width" ,
  xlab="Species", main="Petal Width" )

```

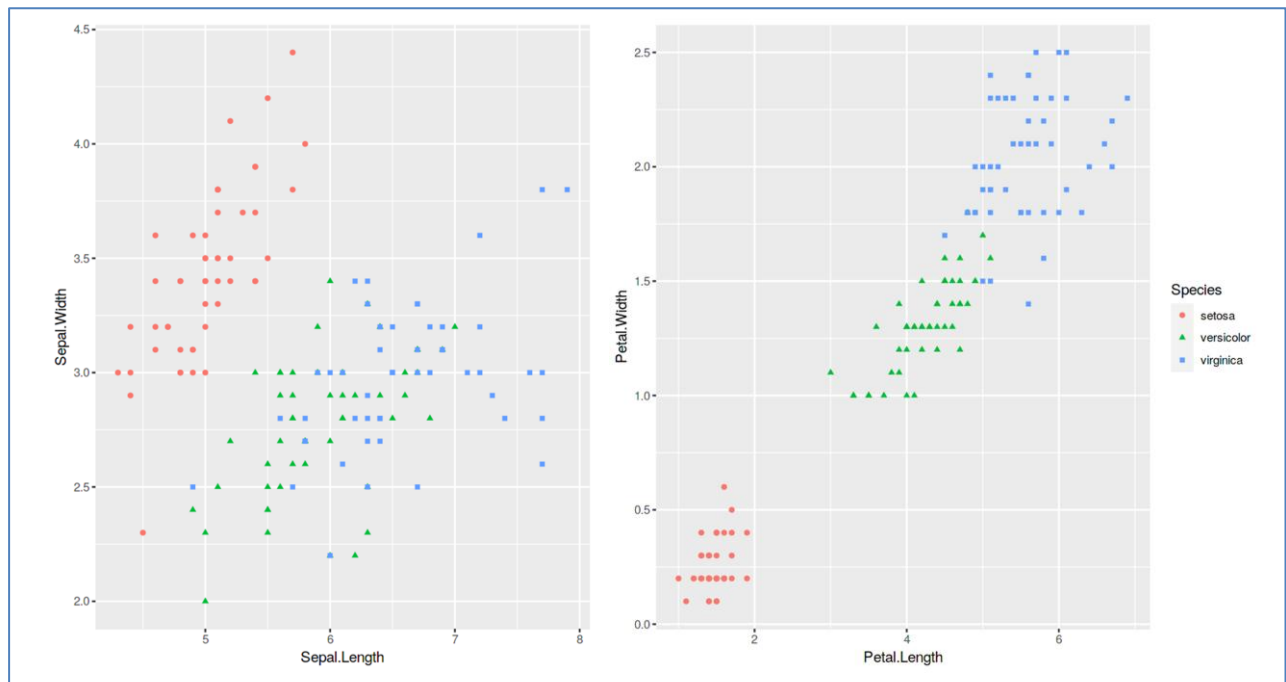


- Type of Chart(2) : Box Plot / 5 มิติ
- Point to Focus : กราฟ box plot แสดงค่ากลาง ค่าการกระจาย สัดส่วนข้อมูลที่มากหรือน้อยกว่าค่ากลาง ของความยาวของกลีบเลี้ยงและกลีบดอกของดอกไอริสแต่ละสายพันธุ์ และความกว้างของกลีบเลี้ยงและกลีบดอกของดอกไอริสแต่ละสายพันธุ์
- ดูข้อมูลที่อยู่ห่างจากกลุ่มมากๆ หรือค่านอกกลุ่มได้เพื่อเป็นการสำรวจข้อมูล

```

> install.packages(c("ggplot2", "dplyr"))
> library(ggplot2)
> library(dplyr)
> ggplot(data = iris, aes(x=Sepal.Length, y=Sepal.Width))+ geom_point(aes(color=Species, shape=Species))
> ggplot(data = iris, aes(x=Petal.Length, y= Petal.Width))+ geom_point(aes(color=Species, shape=Species))

```



- Type of Chart(3) : Scatter Plot / 5 มิติ
- Point to Focus : กราฟ Scatter plot แสดงความสัมพันธ์ของความกว้างและความยาวของกลีบเลี้ยงและกลีบดอกของดอกไอริสแต่ละสายพันธุ์
- แสดงการกระจายของข้อมูลและความสัมพันธ์ของ ความกว้างและความยาวของกลีบเลี้ยงและกลีบดอกของดอกไอริสแต่ละสายพันธุ์