Praktikum 1 – Data Analitik  
**HUBUNGAN ANTARA SEBUAH VARIABEL KATEGORIK DENGAN SEBUAH VARIABEL NUMERIK**

Dalam praktikum ini akan dipelajari dan dipraktekkan bagaimana mengetahui adanya hubungan antar variabel dalam sebuah data multivariat atau data yang berisi lebih dari dua variabel. Variabel sendiri dapat diartikan sebagai sebuah konsep yang memiliki variasi nilai. Hubungan (relationship) antar variable dapat terjadi antara sebuah variabel kategorik dengan sebuah variabel kontinu, antar dua variabel kategorik serta antar dua variabel kontinu. Data latihan yang dipergunakan adalah data mtcars, airquality dan iris yang merupakan data latihan yang tersimpan di R.

|  |
| --- |
| > library(RMySQL)  > con = dbConnect(MySQL(), user = 'root', password = '', dbname =  + 'db\_iris', host = 'localhost')  > myQuery <- "select \* from iris;"  > df <- dbGetQuery(con, myQuery)  > View(df) |

**Hubungan antara sebuah variabel kategorik dengan sebuah variabel kontinu**

> by(df$Petal\_width, df$Species\_name, mean)

df$Species\_name: Setosa

[1] 0.246

---------------------------------------------------------------------------------------

df$Species\_name: Verginica

[1] 2.026

---------------------------------------------------------------------------------------

df$Species\_name: Versicolor

[1] 1.326

> by(df$Petal\_width, df$Species\_name, sd)

df$Species\_name: Setosa

[1] 0.1053856

---------------------------------------------------------------------------------------

df$Species\_name: Verginica

[1] 0.2746501

---------------------------------------------------------------------------------------

df$Species\_name: Versicolor

[1] 0.1977527

> by(df$Petal\_width, df$Species\_name, summary)

df$Species\_name: Setosa

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.100 0.200 0.200 0.246 0.300 0.600

---------------------------------------------------------------------------------------

df$Species\_name: Verginica

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.400 1.800 2.000 2.026 2.300 2.500

---------------------------------------------------------------------------------------

df$Species\_name: Versicolor

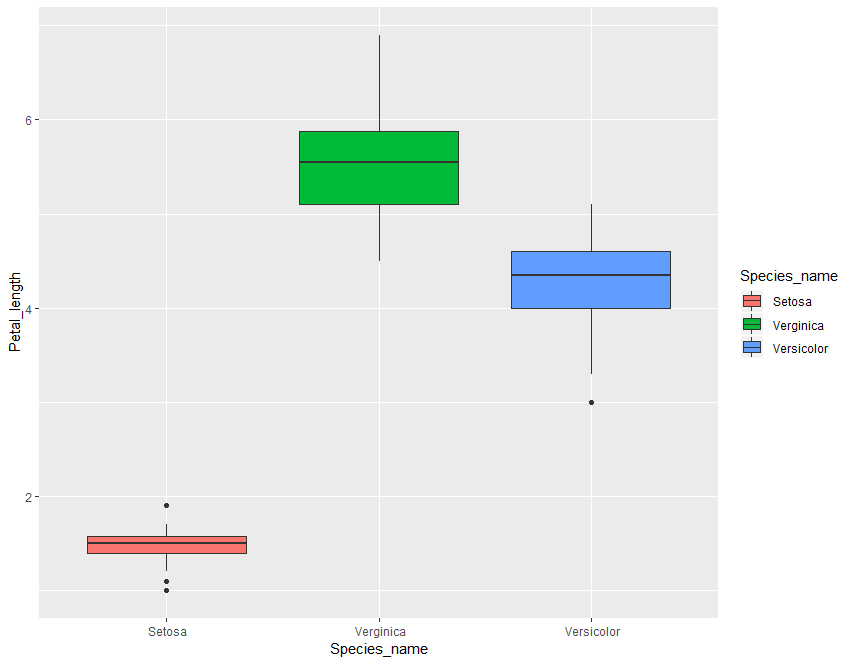
Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.200 1.300 1.326 1.500 1.800

**Metode visualisasi**

> library(ggplot2)

> qplot(Species\_name,Petal\_length, data=df, geom="boxplot", fill=Species\_name)



**Tugas 1 :**

Data: House Price Data

Data menunjukkan contoh harga rumah (dan karakteristik rumah yang terkait)  
untuk area metropolitan AS. Data antara lain menunjukkan identitas rumah, harga jual   
(dalam US $), luas (dalam kaki persegi), jumlah kamar tidur dan kamar mandi, jenis dinding serta lokasi rumah.  
Tujuan utama analisis dalam tugas 1 ini melihat adanya hubungan antara sebuah variabel numerik dengan sebuah variabel kategorik antara lain jenis dinding serta lokasi rumah. Misalnya apakah bahan dinding  
(yaitu, batu bata vs nonbrick) memiliki dampak signifikan pada harga? Atau, apakah penting  
lingkungan mana rumah itu berada? Jawaban atas pertanyaan-pertanyaan ini bisa membantu  
calon pembeli memutuskan berapa besar tawaran untuk rumah.

Gunakan analisis statistiska deskriptif baik yang bersifat numerik (rerata, simpangan baku, dll) serta beberapa metode grafikal (boxplot, density plot, histogram, dll). Berikan penjelasan untuk setiap output yang dihasilkan serta tulis R code yang digunakan. Tampilan grafik yang berwarna tentunya lebih baik.

Catatan : Kalimat yang dipergunakan untuk membuat penjelasan setiap output antar praktikan tidak boleh sama