Laboratorium 14 – Analiza i bazy danych

Wprowadzenie do R-clustering.

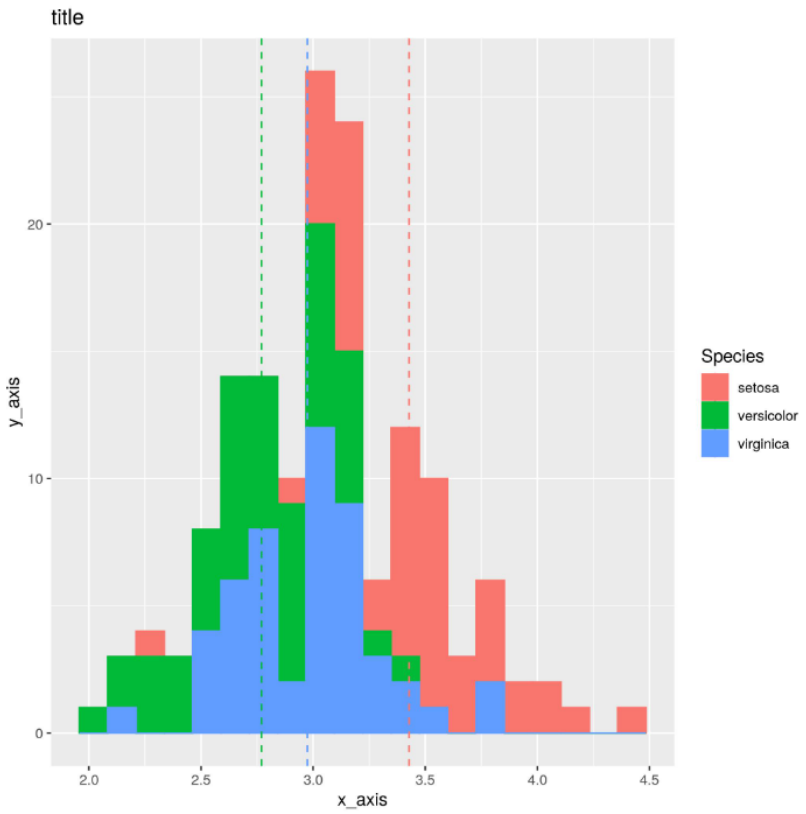
Jakub Sacha, gr.1

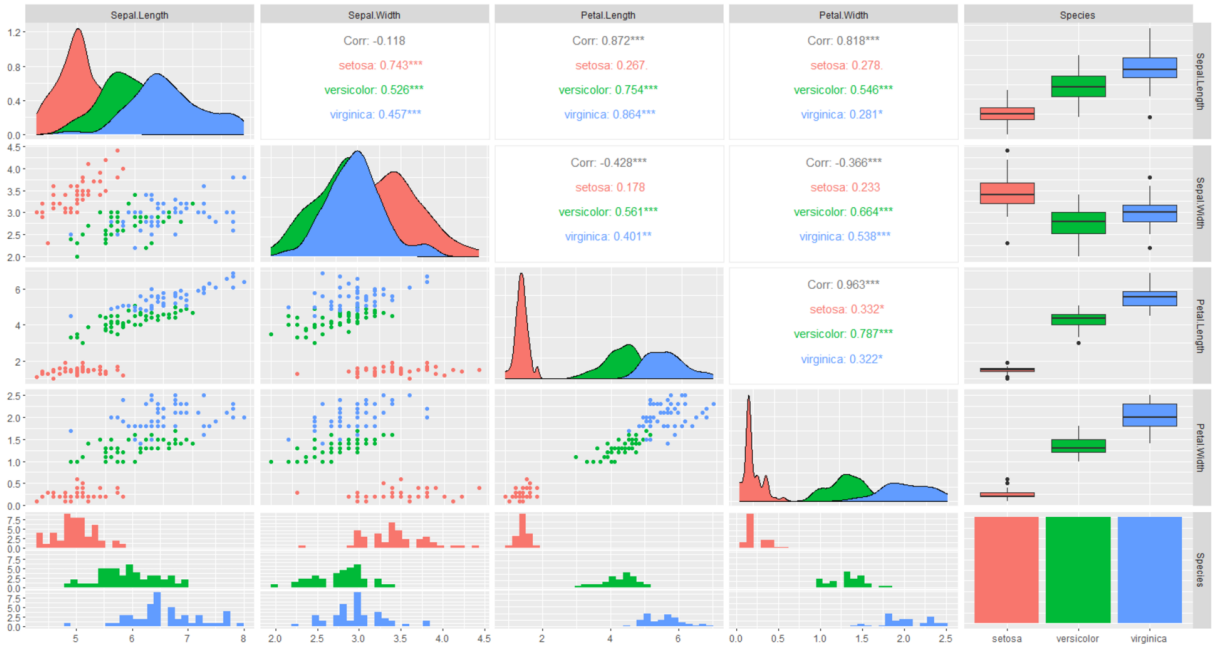
Programy realizujące zadania 1, 2 i 3 oraz ich wyniki.

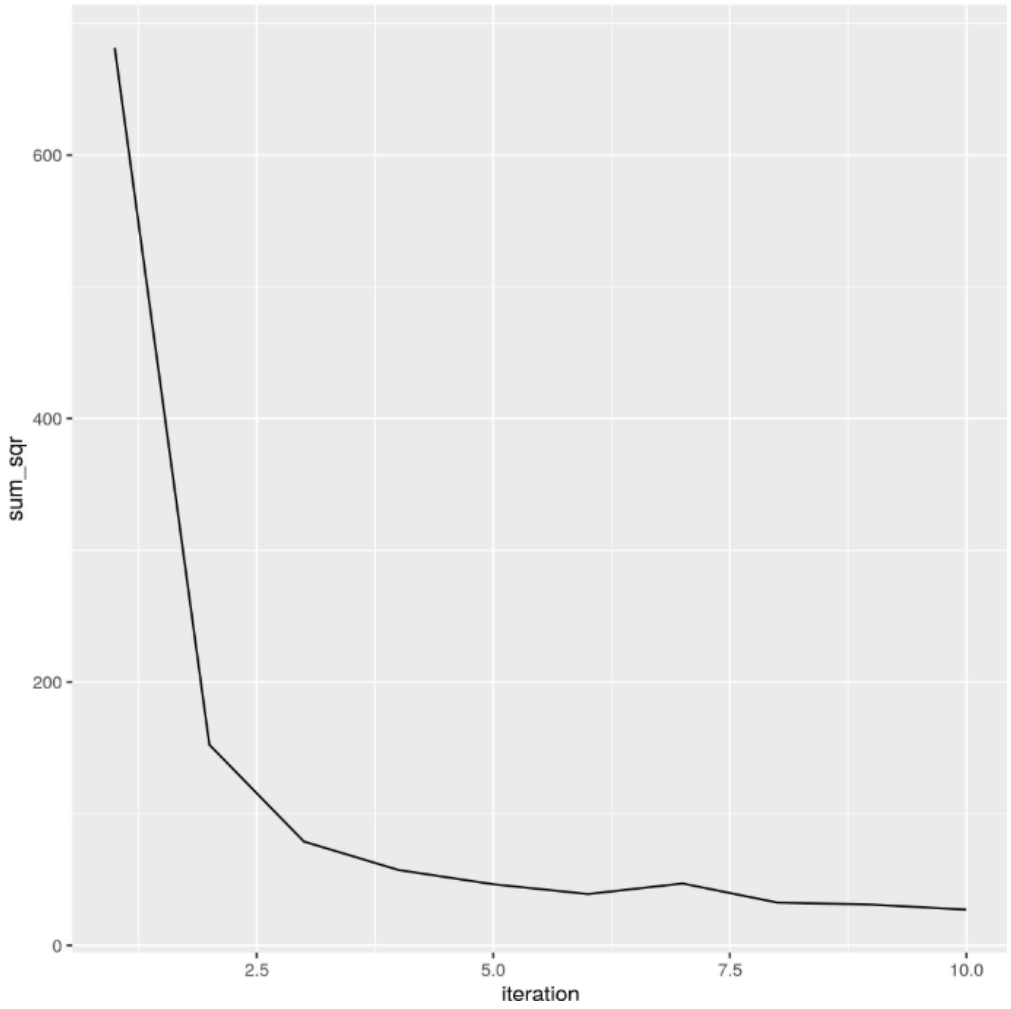
```r
library(magrittr)
library(ggplot2)
library(datasets)
library(GGally)

### Zadanie 1 ###
list_data <- c(1,2,3,4,5,6,7,8,9,10)

list_data <- sqrt(sum(sin(log2(list_data))))
print (paste('Lista po wykonaniu odpowiednich obliczeń',list_data))
data(iris)
print(head(iris))

agg <- iris %>%
    aggregate(. ~Species, .,mean)
print(agg)


### Zadanie 2 ###
a <- ggplot(iris, aes(x = Sepal.Width)) +
    geom_histogram(aes(fill=Species, color=Species), bins=20) +
    geom_vline(data=agg, aes(xintercept=Sepal.Width, color=Species), linetype="dashed") +
labs(x='x_axis', y='y_axis', title='title')
ggsave("/home/rplot.jpg", plot = a)
a <- ggpairs(data = iris, aes(color = Species))
ggsave("/home/rplot2.jpg", plot = a)
```

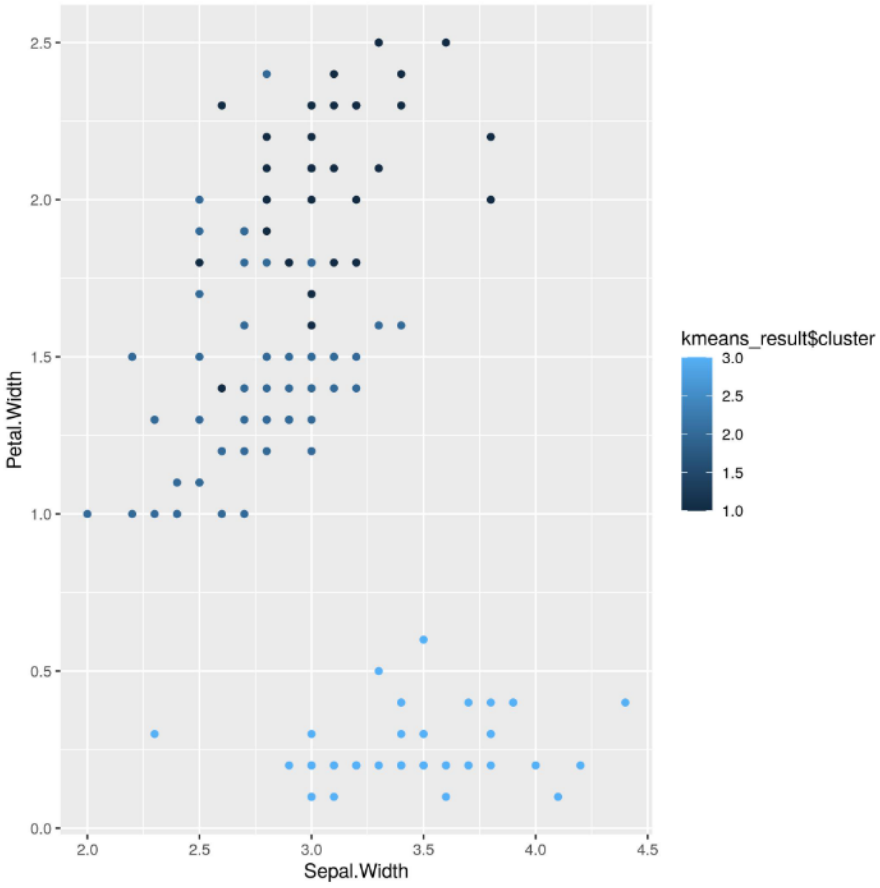
```r

### Zadanie 3 ###
x <- iris[, 1:4]
y <- iris[,5]

sum_sqr <-c()

for (i in 1:10){
    kmeans_result <- kmeans(x, i)
    sum_sqr <- append(sum_sqr, kmeans_result$tot.withinss)
}
a <- ggplot(data.frame(iteration = 1:length(sum_sqr), value = sum_sqr), aes(x = iteration, y = sum_sqr)) +
geom_line()
ggsave("/home/rplot3.jpg", plot = a)

kmeans_result <- kmeans(x, 3)
a <- ggplot(iris, aes(x = Sepal.Width, y = Petal.Width, color= kmeans_result$cluster)) + geom_point()
ggsave("/home/rplot4.jpg", plot = a)

a <- ggplot(iris, aes(x = Sepal.Width, y = Petal.Width, color = Species)) + geom_point()
ggsave("/home/rplot5.jpg", plot = a)
```

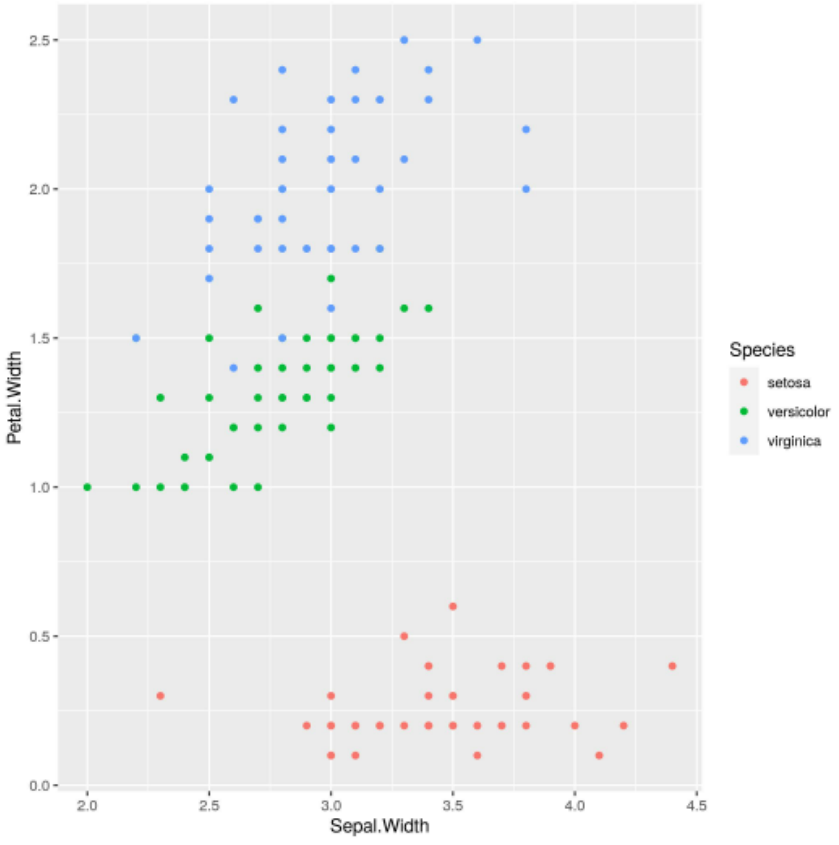
```
> source("/home/docker/lab14.r", encoding = "UTF-8")
[1] "Lista po wykonaniu odpowiednich obliczeń 2.06673513765335"
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
     Species Sepal.Length Sepal.Width Petal.Length Petal.Width
1     setosa        5.006       3.428        1.462       0.246
2 versicolor        5.936       2.770        4.260       1.326
3  virginica        6.588       2.974        5.552       2.026
Saving 7 x 7 in image
Saving 7 x 7 in image
 plot: [5,1] [=======================================>--------] 84% est: 0s `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
 plot: [5,2] [========================================>-------] 88% est: 0s `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
 plot: [5,3] [=========================================>----] 92% est: 0s `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
 plot: [5,4] [==========================================>---] 96% est: 0s `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
Saving 7 x 7 in image
Saving 7 x 7 in image
Saving 7 x 7 in image
```




Sepal.Length
Sepal.Width
Petal.Length
Petal.Width
Species
Corr: -0.118
setosa: 0.743***
versicolor: 0.526***
virginica: 0.457***
Corr: 0.872***
setosa: 0.267.
versicolor: 0.754***
virginica: 0.864***
Corr: 0.818***
setosa: 0.278.
versicolor: 0.546***
virginica: 0.281*
Corr: -0.428***
setosa: 0.178
versicolor: 0.561***
virginica: 0.401**
Corr: -0.366***
setosa: 0.233
versicolor: 0.664***
virginica: 0.538***
Corr: 0.963***
setosa: 0.332*
versicolor: 0.787***
virginica: 0.322*
setosa
versicolor
virginica


sum_sqr
iteration
600
400
200
0
2.5
5.0
7.5
10.0


Petal.Width
Sepal.Width
kmeans_result$cluster
3.0
2.5
2.0
1.5
1.0
0.0
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
4.0
4.5


Petal.Width
Sepal.Width
Species
setosa
versicolor
virginica
0.0
0.5
1.0
1.5
2.0
2.5
2.0
2.5
3.0
3.5
4.0
4.5