

# iris\_report

Dehlia,Elaf,Pierre

2025-06-05

## Introduction

Ce projet explore le célèbre dataset Iris, afin de révéler les relations entre les mesures des fleurs et leurs espèces.

L'objectif est de comprendre les données, détecter des tendances et préparer une base solide pour une classification future.

## Exploration des données

Le dataset contient 150 observations réparties en 3 espèces : *setosa*, *versicolor*, *virginica*.

Les variables mesurées sont la longueur et la largeur des sépales et des pétales.

Aucune valeur manquante n'a été détectée.

## Statistiques descriptives

Moyennes et écart-types des mesures par espèce :

```
library(dplyr)

##
## Attachement du package : 'dplyr'

## Les objets suivants sont masqués depuis 'package:stats':
##
##      filter, lag

## Les objets suivants sont masqués depuis 'package:base':
##
##      intersect, setdiff, setequal, union

iris %>%
  group_by(Species) %>%
  summarise(
    moy_long_sepale = mean(Sepal.Length),
    ecart_type_petale = sd(Petal.Width)
  )
```

```
## # A tibble: 3 x 3
##   Species    moy_long_sepale ecart_type_petale
##   <fct>          <dbl>          <dbl>
## 1 setosa          5.01            0.105
## 2 versicolor      5.94            0.198
## 3 virginica       6.59            0.275
```

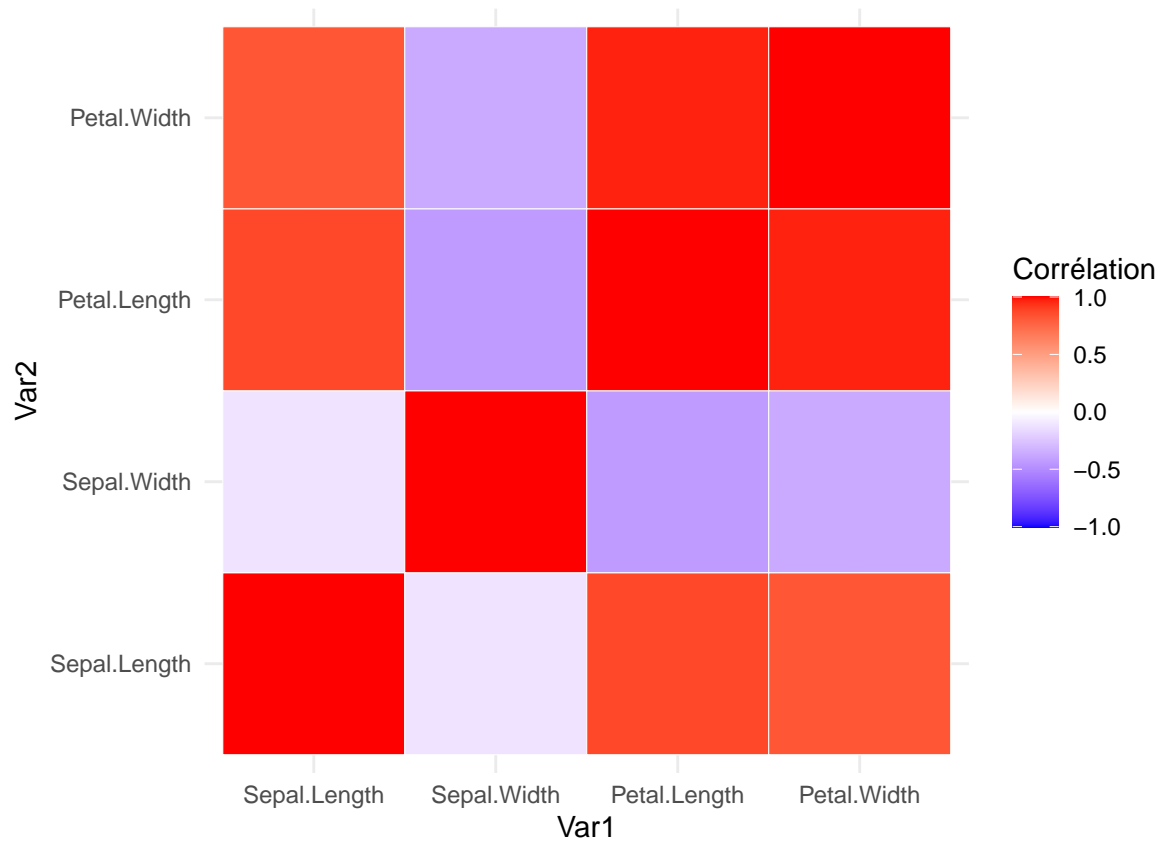
#Matrice de corrélation entre les variables continues :

```
cor_matrix <- cor(iris %>% select(-Species))
cor_matrix
```

```
##           Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length    1.0000000  -0.1175698    0.8717538    0.8179411
## Sepal.Width     -0.1175698    1.0000000   -0.4284401   -0.3661259
## Petal.Length     0.8717538  -0.4284401    1.0000000    0.9628654
## Petal.Width      0.8179411  -0.3661259    0.9628654    1.0000000
```

La heatmap ci-dessous illustre ces corrélations. On note une forte corrélation positive entre la longueur et la largeur des pétales.

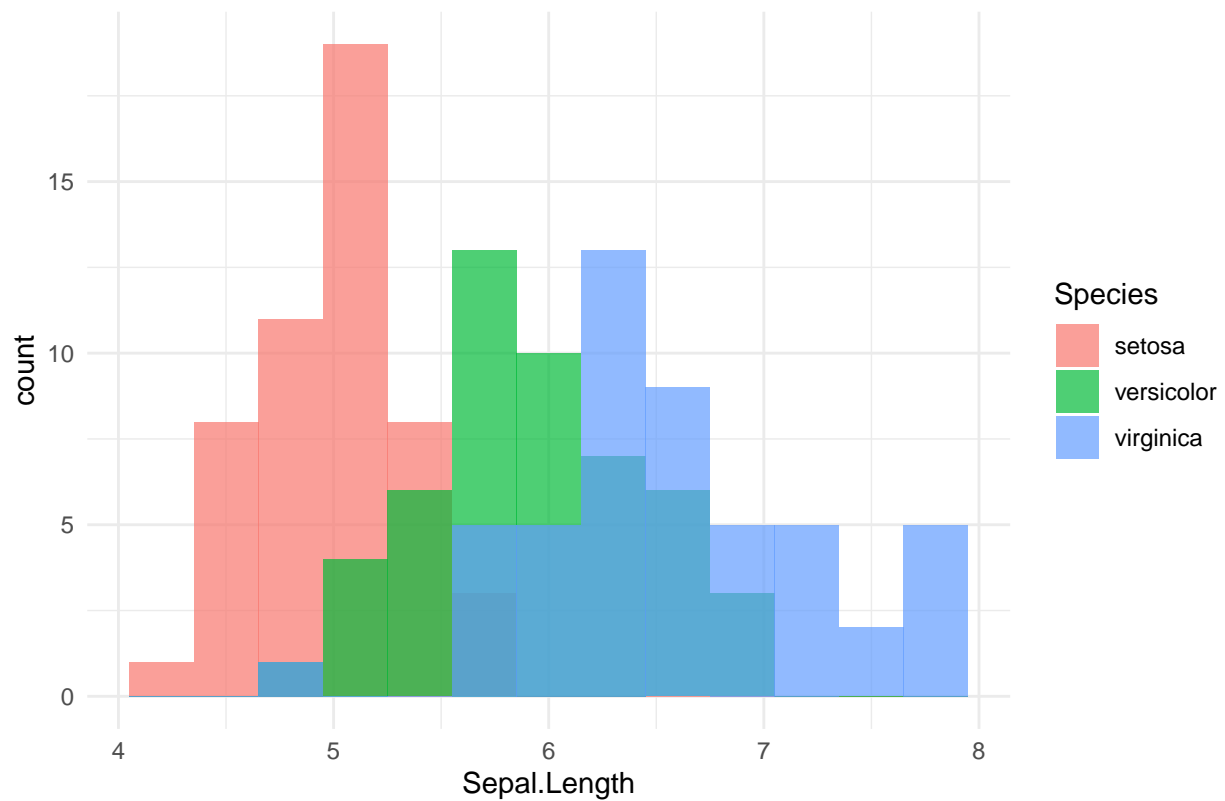
```
library(ggplot2)
library(reshape2)
melted_cor <- melt(cor_matrix)
ggplot(melted_cor, aes(Var1, Var2, fill = value)) +
  geom_tile(color = "white") +
  scale_fill_gradient2(
    low = "blue", high = "red", mid = "white",
    midpoint = 0, limit = c(-1,1), space = "Lab",
    name = "Corrélation"
  ) +
  theme_minimal() +
  coord_fixed()
```



## Visualisations Histogramme des longueurs de sépales selon les espèces :

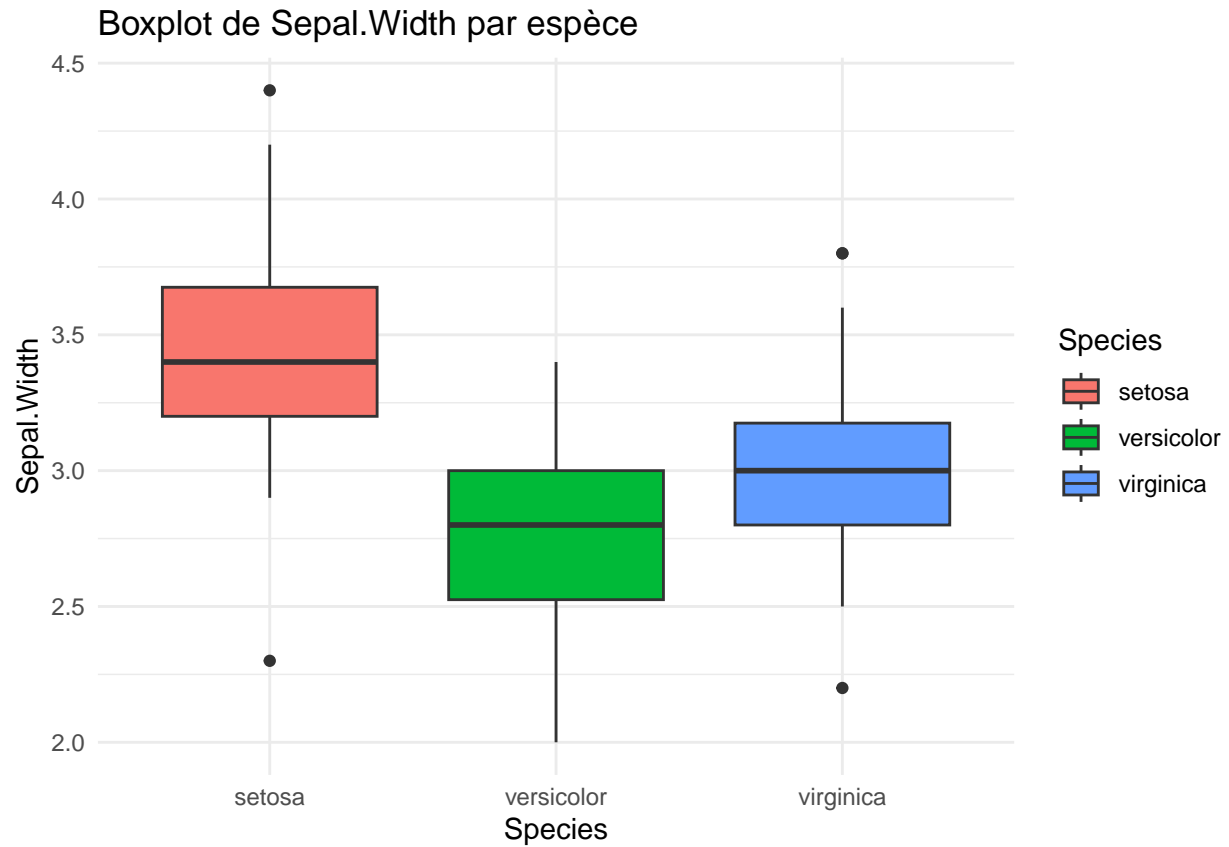
```
ggplot(iris, aes(x = Sepal.Length, fill = Species)) +
  geom_histogram(binwidth = 0.3, alpha = 0.7, position = "identity") +
  labs(title = "Distribution des longueurs de sépales") +
  theme_minimal()
```

Distribution des longueurs de sépales



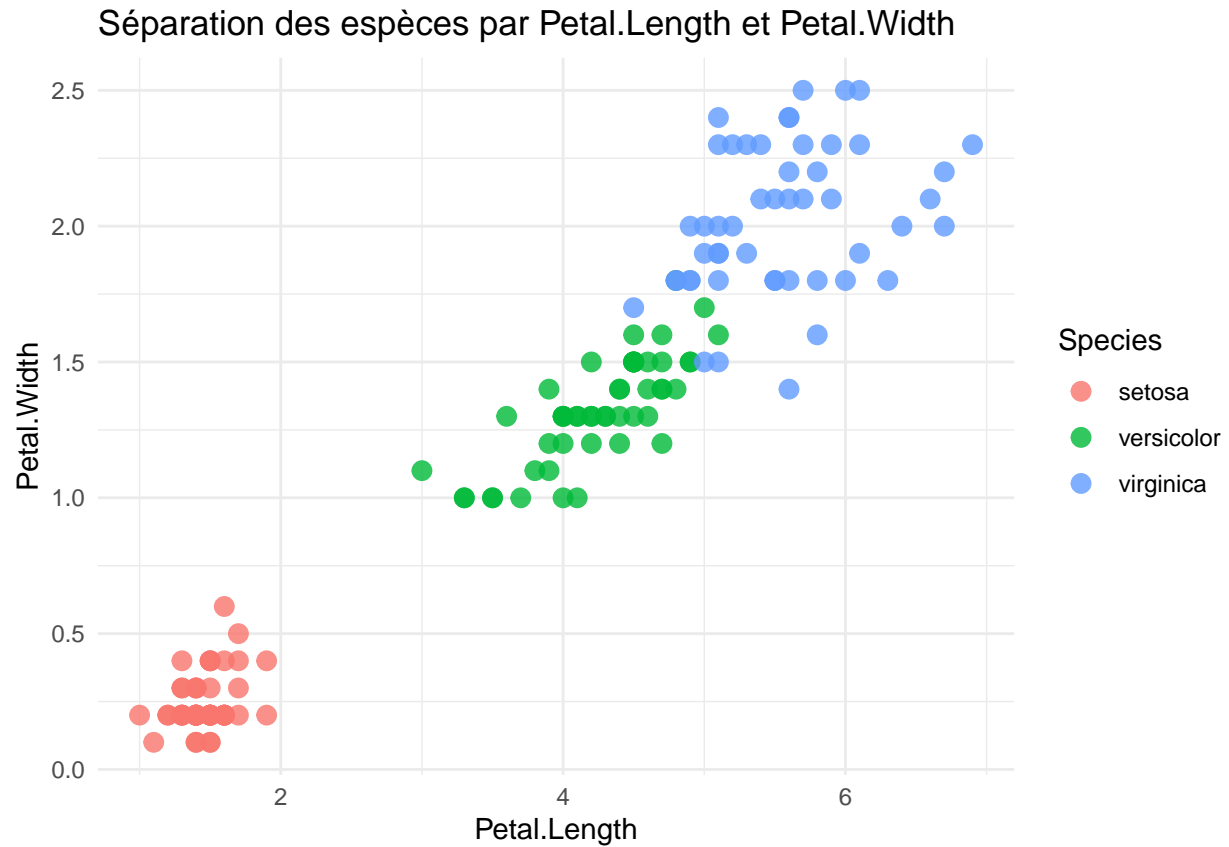
Boxplot des largeurs de sépales par espèce (outliers visibles) :

```
ggplot(iris, aes(x = Species, y = Sepal.Width, fill = Species)) +  
  geom_boxplot() +  
  labs(title = "Boxplot de Sepal.Width par espèce") +  
  theme_minimal()
```



Scatter plot séparant bien les espèces selon longueur et largeur des pétales :

```
ggplot(iris, aes(x = Petal.Length, y = Petal.Width, color = Species)) +  
  geom_point(size = 3, alpha = 0.8) +  
  labs(title = "Séparation des espèces par Petal.Length et Petal.Width") +  
  theme_minimal()
```



#### ##Conclusion

Les variables Petal.Length et Petal.Width sont fortement corrélées et permettent de distinguer clairement les espèces. Setosa est bien séparée des autres, tandis que versicolor et virginica se chevauchent un peu plus. Quelques valeurs aberrantes ont été détectées dans Sepal.Width.

Ces résultats posent une base solide pour une classification automatique des espèces.