

Aide pour le rapport du projet

author 1

author 2

2025-07-23

Table of contents

1	Consignes de rédaction	1
2	Quelques éléments de Quarto	2
3	Etude du jeu de données Iris	2
3.1	Récupération des données	3
3.2	Statistiques descriptives	3
3.2.1	La variable <i>Species</i>	3
3.2.2	Les 4 variables quantitatives	5
	References	6

Ce document a pour but de vous donner des consignes pour la rédaction du rapport et vous donner des éléments de démarrage pour la rédaction d'un rapport en Quarto.

1 Consignes de rédaction

Votre rapport doit synthétiser votre travail d'étude du jeu de données abordé durant le projet. Il doit comprendre :

- une organisation par sections, sous-sections, ... une introduction et une conclusion
- pour chaque question: expliquez vos choix de méthodes, mettez-les en application et commentez les résultats
- Toute figure doit avoir une légende et doit être commentée
- Même remarque pour les tableaux de résultats
-

2 Quelques éléments de Quarto

Des éléments de rédaction d'un document Quarto sont donnés dans la partie 4 des tutoriels R ([ici](#)) et dans le TP1.

Nous rappelons ici quelques points :

- Vous devez créer un documents Quarto au format de sortie PDF
- Vous pouvez organiser votre document en sections, sous-sections, ... grâce à `#`, `##`, ...
- Vous pouvez dans l'en-tête de votre document Quarto
 - préciser le titre du document, les auteurs, la date, ...
 - ajouter une table des matières avec l'option `toc`
 - ajouter des macro Latex dans `header-includes`
 - ajouter une bibliographie avec `bibliography`:
 - ...
- Vous pouvez mettre du code R dans des chunks R et jouer sur les options comme
 - `#| echo: false` pour ne pas afficher le code dans le rapport
 - `#| eval: false` pour ne pas l'évaluer dans le rapport
 - `#| fig-height: , #| fig-width: , ...` pour maîtriser la taille des figures
 - `#| fig-cap:` pour mettre une légende aux figures
 - `#| message: false` pour ne pas afficher les messages de R
 - `#| warning: false` pour ne pas afficher les warnings de R
 - ...

Vous pouvez appliquer des options permanentes aux chunks en les ajoutant dans l'en-tête dans la partie `execute`:

- Vous pouvez mettre des formules mathématiques Latex entre `$...$`. On peut aussi utiliser `\begin{equation} \end{equation}`, ...
- Références croisées : Pour citer une Figure, une table, ... dans le texte, il faut préciser l'option `#| label:` et dans le texte utiliser `@ ...` avec `fig-` pour une figure, `tbl-` pour une table. Un exemple pour chaque est donné dans la suite de ce document. Surtout, ne pas écrire des phrases du type “d'après la figure ci-dessous” !

3 Etude du jeu de données Iris

On va ici utiliser le célèbre jeu de données des Iris pour illustrer quelques points de rédaction en Quarto. Vous êtes donc invités à parcourir en même temps le .pdf et le .qmd pour comprendre les points de syntaxe.

3.1 Récupération des données

Les données Iris ont été collectées par Edgar Anderson (Anderson 1935). Ce sont les mesures en centimètres des variables suivantes : longueur du sépale (Sepal.Length), largeur du sépale (Sepal.Width), longueur du pétale (Petal.Length) et largeur du pétale (Petal.Width) pour trois espèces d'iris : *Iris setosa*, *I. versicolor* et *I. virginica*. Les données sont disponibles de base sous R et on les récupère donc avec la fonction `data(iris)`. Les premières lignes du jeu de données sont affichés dans la Table 1.

Table 1: Les premières lignes du jeu de données iris

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa

On retrouve bien que les données sont composées de 150 individus, de 4 variables quantitatives et d'une variable qualitative *Species*. Dans la suite, nous notons Y la variable *Species* et \mathbf{X} la matrice composée des 4 autres variables

$$\mathbf{X} = (X_{ij})_{i \in \{1, \dots, 150\}, j \in \{1, \dots, 4\}}.$$

3.2 Statistiques descriptives

Nous faisons ici quelques statistiques descriptives pour prendre en main les données.

3.2.1 La variable *Species*

Nous commençons par la variable *Species* (vecteur Y) qui est une variable qualitative. La Figure 1 nous permet de contrôler que nous avons bien 50 individus par espèce.

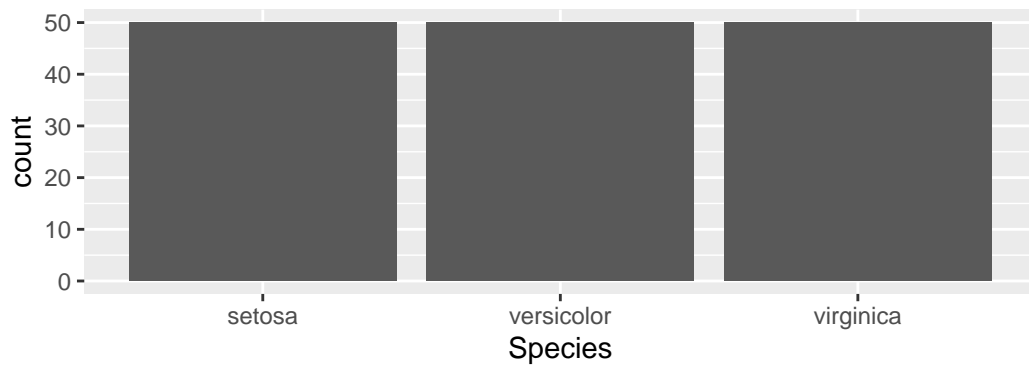


Figure 1: Barplot de la variable Species

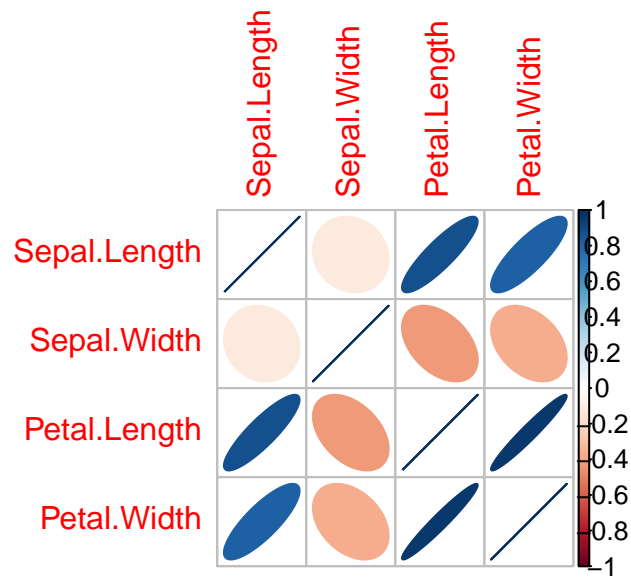


Figure 2: Matrice des corrélations entre les 4 variables quantitatives

3.2.2 Les 4 variables quantitatives

Nous nous intéressons maintenant aux 4 variables quantitatives (matrice X). La Figure 2 montre les corrélations entre les 4 variables. On peut remarquer que la largeur et la longueur des pétales sont fortement corrélées positivement, ce n'est pas le cas pour les sépales.

A l'aide de la librairie **FactoMineR**, on met en place une ACP. On représente en Figure 3 les individus projetés sur le premier plan factoriel, la couleur correspondant à l'espèce ainsi que les corrélations des variables quantitatives initiales avec les deux premières composantes principales.

```
respca=PCA(iris,quali.sup=5,graph=F)
fviz_pca_ind(respca,label="none",habillage=5)+
  theme(legend.position="bottom")
fviz_pca_var(respca)
```

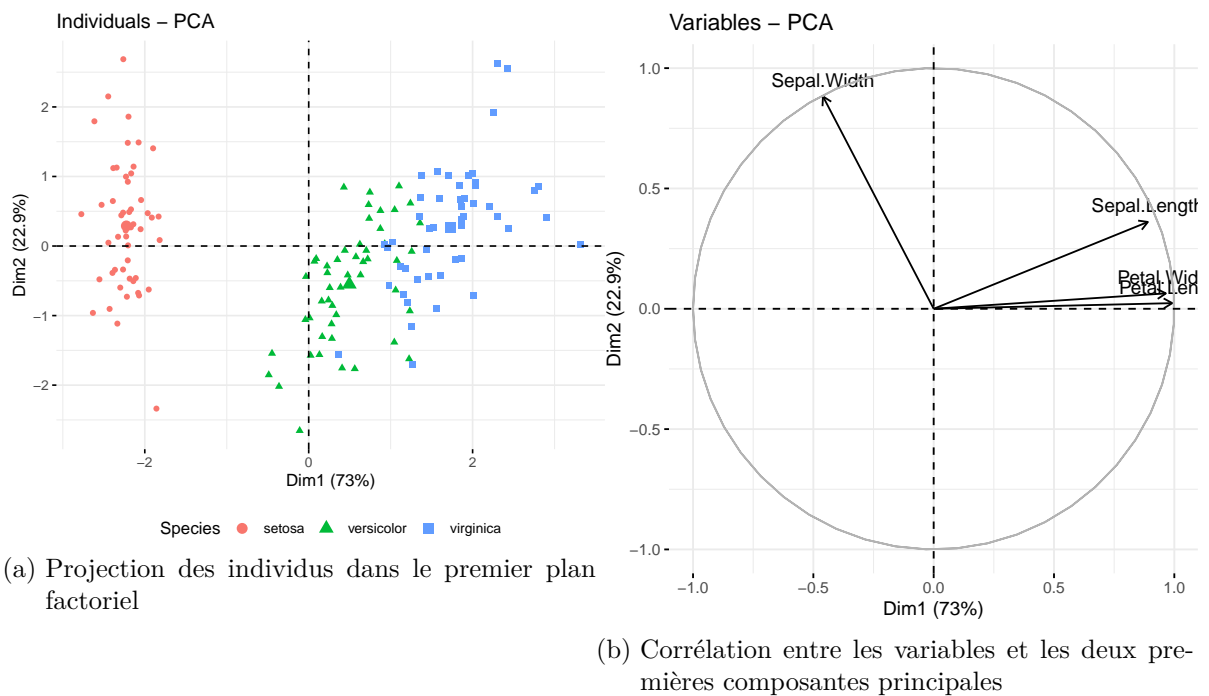


Figure 3: Résultats de l'ACP

On peut remarquer que les *Setosa* se distinguent des deux autres espèces principalement par la largeur de leur sépales. On peut appuyer ce point à l'aide de la Figure 4.

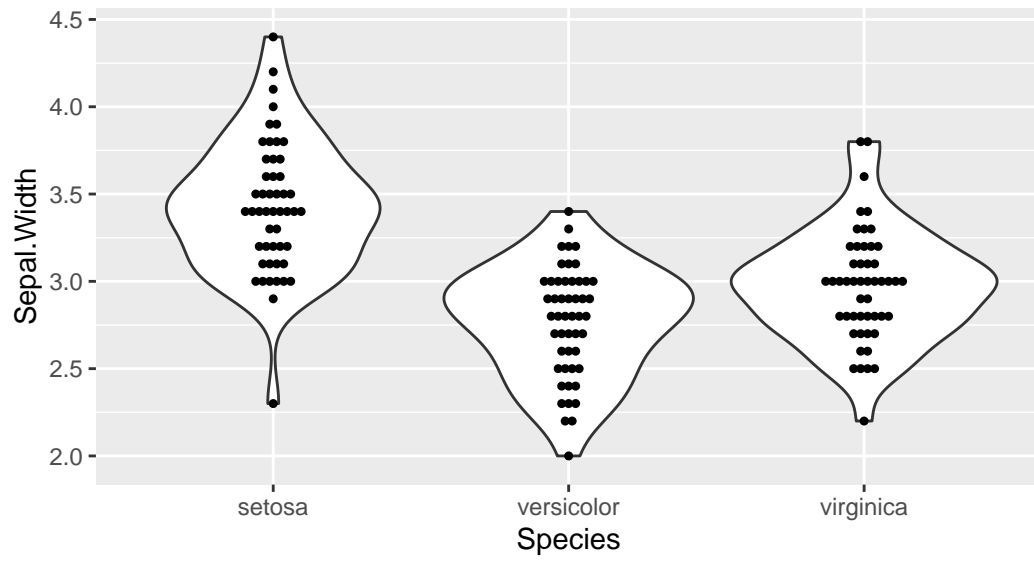


Figure 4: Violin plots de la largeur des sépales pour chaque espèce

References

Anderson, Edgar. 1935. "The Irises of the Gaspé Peninsula." *Bull. Am. Iris Soc.* 59: 2–5.