tp1

2022-11-24

R Markdown

R Markdown permet de créer des documents dynamiques, qui incluent du texte mis en forme, des équations, et du code R.

Cet outil est très utile pour écrire des rapports techniques. Il permet, dans un seul document, d'exposer le contexte du problème, la méthode de résolution, et les résultats de l'analyse.

De nombreuses ressources sont disponibles en ligne. Voir par exemple l'introduction officielle à R markdown, ainsi que la fiche synthétique.

Installation

Pour utiliser R markdown, il suffit d'installer la librairie associée, en tapant la commande suivante dans la console:

```
install.packages("rmarkdown")
```

Création d'un document

Une fois l'installation effectuée, le plus simple est de créer un nouveau document en utilisant l'interface de RStudio.

Sélectionner le menu:

File > New File > R markdown...

Dans l'onglet Document (sélectionné par défaut), vous pouvez saisir le titre et la, le ou les auteur · e · s du document.

Trois formats sont proposée:

- HTML: permet une mise en forme dynamique, utile pour les sites internet.
- PDF: permet une mise en forme fixe, utile pour un rendu "officiel" ou papier.
- Word: permet de générer des documents éditables, qui peuvent ensuite être partagés avec des collaborateurs ou collaboratrices.

Pour cette séance, on choisi le format HTML, le plus simple en terme de mise en page.

Premier document

Une fois le document créé, RStudio propose un contenu "didactique" par défaut.

Vous pouvez compiler ce document en cliquant sur la commande Knit en haut à gauche, à côté d'une pelotte de laine (en anglais, "to knit" signifie "tricoter").

Si tout fonctionne bien, RStudio va générer un document HTML, qu'il ouvre dans une nouvelle fenêtre.

Vous pouvez étudier ce premier document, et voir comment la source (le .Rmd) influe sur la sortie (le .html).

Code R et graphiques

Supposons que l'on mène l'expérience suivante.
Une urne contient 300 boules, 100 rouges, 100 bleues et 100 vertes.
On tire une boule au hasard, on note sa couleur, et on la met de côté.
On reproduit cette expérience 60 fois.

Avec R, on peut simuler cette expérience aléatoire de la façon suivante.

```
# set.seed(412)                          ## Reproductibilité
urne <- c(rep("rouge", 100),           ## 100 boules rouges
          rep("bleue", 100),           ## 100 boules bleues
          rep("verte", 100))           ## 100 boules vertes
n_exp <- 60                            ## Nombre de fois où je fais l'expérience
echantillon <- sample(urne,            ## échantillonne les boules
                      n_exp,             ## n_exp fois
                      replace = FALSE) ## sans remise
```

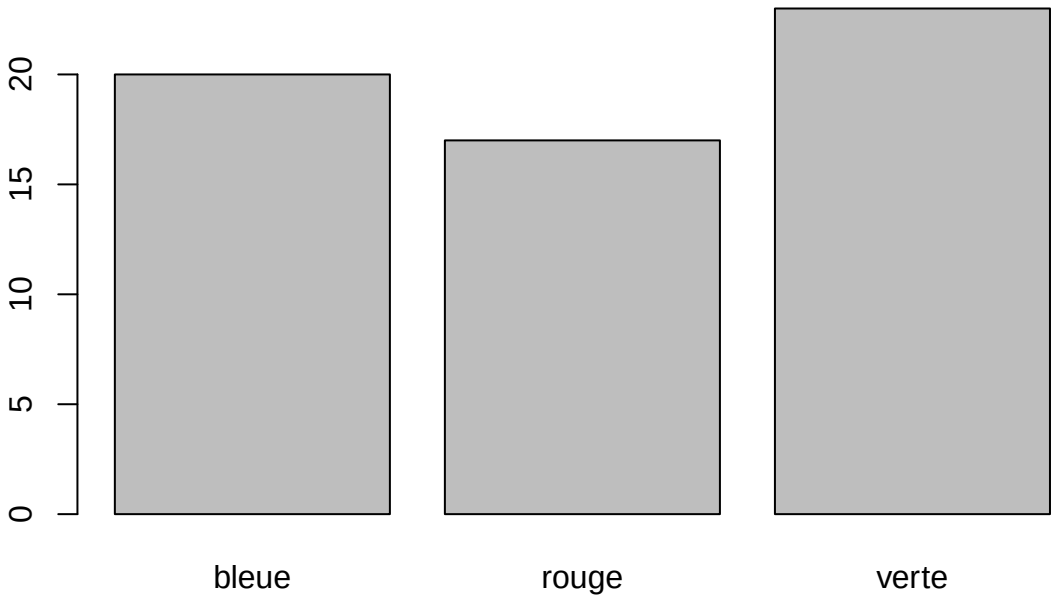
On peut ensuite afficher un résumé de l'experience (nombre de fois où chaque boule a été tirée).

```
res <- table(echantillon)              ## Résumé de l'échantillon
res
```

```
## echantillon
## bleue rouge verte
##    20    17    23
```

Et tracer ce résultat.

```
barplot(res)                           ## Diagramme en bar
```



Il est aussi possible d'utiliser les résultats dans le texte.

Par exemple, sur les 60 expériences, on a tiré ici 20 fois la boule bleue.

Exercice

> **À rendre**: votre rapport *compilé* sous format `html`. Rédigez un document `R markdown` qui réponde au problème exposé ci-dessous. Le document doit être auto-suffisant, et exposer clairement

le problème, les analyses, et les conclusions.

Il contiendra des sections, des mots mis en valeurs en **gras** et en *italique*, et du `code` dans le texte.

Il contiendra également des blocs de code `R`, qui serviront à répondre au problème.

Il pourra contenir des équations Latex, et des listes, numérotées ou non.

On pêche des poissons dans le Lez.

On suppose qu'il y a en tout 10 000 poissons dans le Lez, dont 2 000 rouges, 3 000 verts et 5 000 bleus (le Lez est très pollué). On suppose cependant que l'on n'a pas accès à cette information (on ne sait pas combien il y a de poissons en tout, ni combien de chaque couleur).

On se pose la question : "Quelle est la proportion de poissons rouges, verts et bleus dans le Lez ?"

On pêche 100 poissons, que l'on garde pour les exposer en aquarium, et l'on note leur couleur.

Décrivez l'expérience statistique : question posée, individus, population, échantillon, taille, type de variable mesurée. - **Quelle est la proportion de poissons rouges, verts et bleus dans le Lez ?** - Individus : les poissons du Lez. - Population : 10 000. - Echantillon : 100. - Type de varaibles : qualitatives (couleur des poissons). La variable a-t-elle bien des modalités incompatible, exhaustives et sans ambiguité ? Oui, car les poissons ne sont pas de couleurs mélangées. Simulez les données correspondant à cette expérience, en utilisant la fonction `sample`.

```
couleurs = c(rep('bleu', 5000), rep('rouge', 2000), rep('vert', 3000))
poissons = sample(couleurs, size=100, replace=TRUE)
```

Calculer la fréquence empirique de chaque couleur de poissons en utilisant la fonction `table`.

```
res_poissons <- table(poissons)
print(res_poissons)
```

```
## poissons
##  bleu rouge  vert
##    56    18    26
```

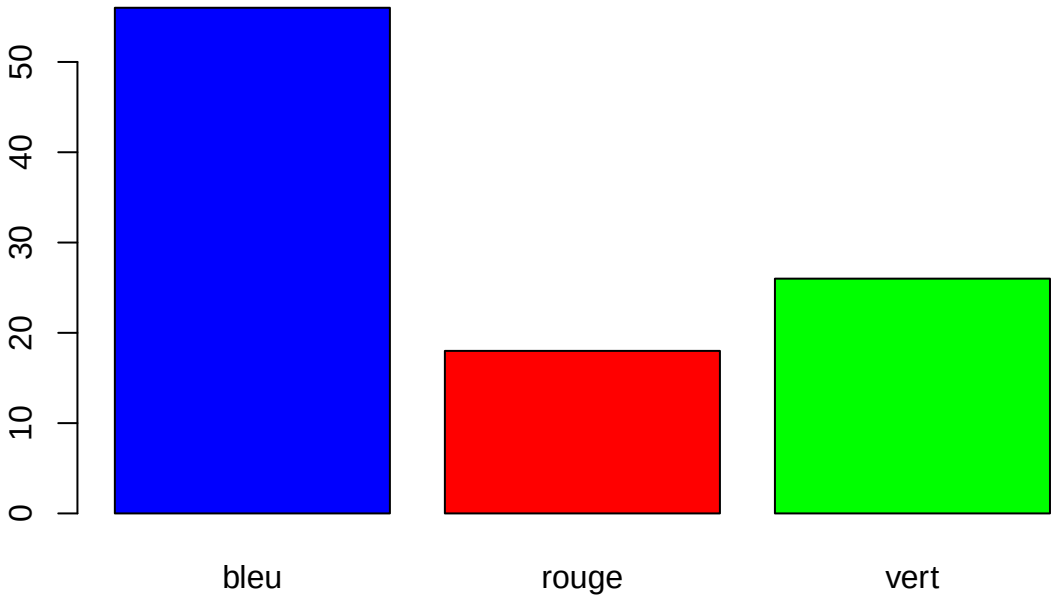
Quelle est la proportion estimée de poissons rouges, verts et bleus ?

Proportion estimée : 50 bleus | 20 rouges | 30 verts. > Ce résultat vous surprend-il ?

Non.

Tracer ces fréquences empiriques en utilisant la fonction `barplot`.

```
barplot(res_poissons, col=c("blue", "red", "green"))                    ## Diagramme en bar
```



Que se passe-t-il lorsque vous compilez votre document plusieurs fois de suite ? Les résultats numériques changent-ils ?

Les résultats ne changent pas car la fonction `set.seed` est déjà utilisée dans l'exemple plus haut. Cependant, lorsque la fonction est commentée, *les résultats changent à chaque éxecution.*

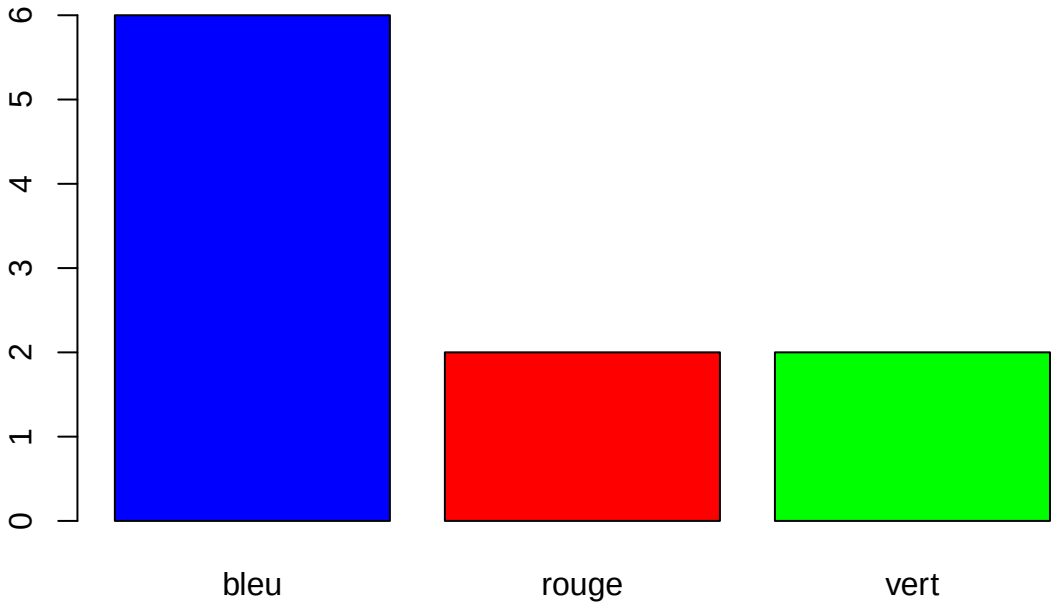
Est-ce normal ?

C'est normal, car sample effectue un tirage aléatoire.

Utilisez la fonction `set.seed` pour produire un document reproductible.

On suppose que, par manque de budget, on ne peut pêcher que 10 poissons. Quels résultats obtenez-vous ? L'estimation des proportions est-elle bonne dans ce cas ?

```
poissons10 = sample(couleurs, size=10, replace=TRUE)
res_poissons10 <- table(poissons10)
barplot(res_poissons10, col=c("blue", "red", "green"))
```



On suppose maintenant que, grâce à une collecte de fonds en ligne, on peut pêcher 1 000 poissons. Est-ce que cela améliore l'estimation ?

Cela **améliore bien l'estimaiton**.

```
poissons10000 = sample(couleurs, size=10000, replace=TRUE)
res_poissons10000 <- table(poissons10000)
barplot(res_poissons10000, col=c("blue", "red", "green") )
```