

Master 1 – BDIA

Analyse et visualisation des données

Sujet TP1

Exercice 1 : Créez votre premier vecteur

1. Créer un vecteur contenant les nombres pairs de 2 à 10.
2. Calculer la somme et la moyenne.
3. Afficher le vecteur, la somme et la moyenne.

Indications

En R, la fonction `cat()` (abréviation de "concatenate and print") est utilisée pour afficher du texte ou des valeurs à l'écran. Contrairement à `print()`, elle ne retourne pas de guillemets autour des chaînes de caractères et permet de concaténer plusieurs éléments pour les afficher dans une seule ligne.

Saut de ligne : `"\n"` ajoute un saut de ligne après l'affichage pour une meilleure lisibilité.

Exercice 2 : Températures

Voici les températures Fahrenheit relevées sur une semaine, du lundi au dimanche :
53.6, 59.0, 50.0, 57.2, 64.4, 68.0, 60.8.

Travail à réaliser :

1. Convertir les températures Fahrenheit (F) en Celsius (C) à l'aide de la formule :

$$T_C = (T_F - 32) \times \frac{5}{9} \quad (1)$$

2. Calculer la température moyenne de la semaine en Celsius.
3. Afficher les jours avec des températures supérieures à 15°C.
4. Identifier la température minimale et maximale en Celsius.
5. Augmenter toutes les températures de 2°C (scénario d'un réchauffement).
6. Afficher l'histogramme des températures avec les jours de la semaine en abscisse.

On peut créer autant d'objet qu'il y a de jour de la semaine et leur affecter les températures. Cependant on préférera créer un vecteur avec les différentes températures pour ensuite réaliser les calculs sur les données.

Exercice 3 : Gestion des valeurs manquantes

Il peut arriver que certaines valeurs soient non renseignées. En R, ces valeurs s'écrivent NA (pour not available). À partir du vecteur suivant :

```
1 conjoint1 <- c(1200, 1180, 1750, NA)
```

1. Afficher le minimum et le maximum en ignorant les valeurs manquantes.
2. Afficher la moyenne avec deux décimales.

Indications

na.rm = TRUE : est utilisé dans min(), max(), et mean() pour ignorer les valeurs NA lors des calculs.

Arrondir la moyenne à deux décimales : round(moyenne, 2)

Affiche exactement deux décimales, même si la moyenne est un nombre entier :

format(..., nsmall = 2) :

Exercice 4 : Analyse climatique

Les températures et précipitations mensuelles à Lyon en 2023 :

```
1 temperature <- c(3.4, 4.8, 8.4, 11.4, 15.8, 19.4, 22.2, 21.6, 17.6,
  13.4, 7.6, 4.4)
2 precipitations <- c(47.2, 44.1, 50.4, 74.9, 90.8, 75.6, 63.7, 62, 87.5,
  98.6, 81.9, 55.2)
```

1. Calculer la température moyenne annuelle.
2. Calculer la somme totale des précipitations.
3. Visualiser les données avec ggplot2 en choisissant un graphique adapté à ces données.

Indications

Un data.frame est une structure de données utilisée pour stocker des données tabulaires, similaires à une table ou une feuille Excel. Chaque colonne d'un data.frame correspond à une variable, et chaque ligne correspond à une observation ou un enregistrement.

Un facteur (factor) est une structure de données utilisée pour représenter des variables catégoriques, comme les jours de la semaine, les types de fruits, ou les réponses à un questionnaire ("Oui", "Non"). Par défaut, les données sont triées par ordre alphabétique (exemple : "Avr", "Déc", "Fév" pour les mois). Pour forcer l'ordre chronologique des mois de l'année faire : factor(mois, levels = mois).

Exercice 5 : Calcul de revenus par ménage

On dispose des revenus des conjoints et du nombre de personnes dans chaque ménage :

```
1 conjoint1 <- c(1200, 1180, 1750, 2100)
2 conjoint2 <- c(1450, 1870, 1690, 0)
3 nb_personnes <- c(4, 2, 3, 2)
```

1. Calculer le revenu total pour chaque ménage.
2. Diviser ce revenu par le nombre de personnes.

Exercice 6 : Jeu de données mtcars

R propose de nombreux jeux de données (dataset) accessibles avec la commande `data()`. Le jeu de données `mtcars` contient des informations sur les performances de 32 modèles de voitures américaines avec 11 variables numériques comme le montre le tableau 1. Le gallon est une unité de mesure de volume utilisée principalement dans les pays anglo-saxons, il correspond à 3,78 litres. Le mile équivaut à 1,6 kilomètre. La colonne `mpg` signifie mile par gallon et correspond à la consommation de carburant des véhicules.

Variable	Description	Type
<code>mpg</code>	Miles per gallon (miles par gallon)	Numérique
<code>cyl</code>	Nombre de cylindres	Numérique
<code>disp</code>	Cylindrée (en pouces cubes)	Numérique
<code>hp</code>	Puissance (en chevaux-vapeur)	Numérique
<code>drat</code>	Ratio de l'essieu arrière	Numérique
<code>wt</code>	Poids de la voiture (en milliers de livres)	Numérique
<code>qsec</code>	Temps pour parcourir 1/4 de mile (secondes)	Numérique
<code>vs</code>	Type de moteur (0 = en V, 1 = en ligne)	Numérique
<code>am</code>	Type de transmission (0 auto., 1 manuelle)	Numérique
<code>gear</code>	Nombre de vitesses	Numérique
<code>carb</code>	Nombre de carburateurs	Numérique

Table 1: Description des variables du dataset `mtcars`

On souhaite étudier les données de `mtcars` et pour cela, faire :

1. Charger les données.
2. Connaître la structure des données.
3. Afficher les premières lignes.
4. Ajouter une colonne nommée `consommation` et convertir les données `mpg` pour obtenir une consommation en Litre pour 100 kilomètres.
5. Afficher un résumé statistiques des données.
6. Afficher les voitures ayant une (`consommation`) > 20 litres.

7. Afficher uniquement les colonnes `consommation`, `hp` et `gear`.
8. Calculer et afficher la moyenne de la consommation pour les voitures avec transmission manuelle.
9. Créer un graphique "nuage de points" pour illustrer la relation entre le poids (`wt`) des véhicules et la `consommation`.

Conversion miles/gallon en kilomètres/litre

Pour convertir une consommation exprimée en miles par gallon (**mpg**) en litres par 100 kilomètres (**L/100 km**), voici les étapes détaillées :

$$\begin{aligned} 1 \text{ mile} &= 1.60934 \text{ kilomètres}, \\ 1 \text{ gallon (US)} &= 3.78541 \text{ litres}. \end{aligned}$$

Pour convertir **mpg** en **km/L**, on multiplie **mpg** par le ratio :

$$\text{km/L} = \text{mpg} \times \frac{1.60934}{3.78541}.$$

Conversion de kilomètres/litre en litres/100 kilomètres

Pour obtenir une consommation en **L/100 km**, on prend l'inverse de **km/L** et on multiplie par 100 :

$$\text{L/100 km} = \frac{100}{\text{km/L}}.$$

En remplaçant **km/L** par sa formule en fonction de **mpg** :

$$\text{L/100 km} = \frac{100}{\text{mpg} \times \frac{1.60934}{3.78541}}.$$

Simplification

En simplifiant :

$$\text{L/100 km} = \frac{100 \times 3.78541}{1.60934} \times \frac{1}{\text{mpg}}.$$

La constante est alors calculée comme suit :

$$\frac{100 \times 3.78541}{1.60934} = \frac{378.541}{1.60934} \approx 235.214583.$$

Ainsi, la formule finale pour convertir **mpg** en **L/100 km** est donnée par :

$$\text{L/100 km} = \frac{235.214583}{\text{mpg}}.$$