

# Bilan\_Jour1

Cédric HASSEN-KHODJA

## 1. Importation et préparation de données:

Importez le jeu de données "mtcars" dans R.

Supprimez les colonnes "vs" et "am".

Renommez la colonne "wt" en "weight".

Ajoutez une nouvelle colonne "disp\_per\_cyl" et calculez ce ratio.

## 2. Manipulation de données:

Importez le jeu de données "diamonds" à partir du package "ggplot2".

Utilisez la fonction de votre choix pour extraire les diamants de plus de 1 carat.

Créez une nouvelle colonne appelée "price\_per\_carat" qui calcule le prix par carat de chaque diamant.

Utilisez la fonction "aggregate()" pour trouver le prix moyen par carat pour chaque type de coupe de diamant.

## 3. Fusion de données:

Importez les fichiers "sales.csv" et "products.csv" dans R.

Fusionnez les deux jeux de données en utilisant la fonction "merge" et la colonne "Product et Price" comme clé de fusion.

Ajoutez une nouvelle colonne appelée "Revenu" qui calcule le chiffre d'affaires total de chaque vente en multipliant la quantité vendue par le prix unitaire.

## 4. Nettoyage de données:

Importez le fichier "housing.csv" dans R.

Inspectez le fichier pour vérifier le type de données et le nombre de valeurs manquantes dans chaque colonne.

Identifiez les observations qui ont des valeurs manquantes.

Supprimez les observations avec des valeurs manquantes.

## 5. Sélection de données:

Importez le jeu de données "iris".

Extraire les observations avec une longueur de pétale supérieure à 4,5 cm.

Sélectionnez les colonnes "Species", "Petal.Length" et "Petal.Width".

Triez les observations par ordre croissant de longueur de pétale.

## 6. Création de fonctions:

Créez une fonction qui prend en entrée une liste de nombres et renvoie la somme de ces nombres. Testez votre fonction avec quelques exemples.

## 7. Manipulation de chaînes de caractères:

Importez le jeu de données "mtcars".

Extraire les noms de voitures qui contiennent la lettre "C" dans leur nom.

## 8. Simulation de données:

Créez une simulation pour simuler 1000 lancers de deux dés équilibrés.

Calculez la somme des deux dés pour chaque lancer et affichez un histogramme de la distribution des sommes.

## 9. Manipulation de données avec dplyr:

Importez le jeu de données "mtcars".

Groupez les données par le nombre de cylindres et calculez la moyenne de la puissance et du poids pour chaque groupe.

## 10. Manipulation de données avec tidyr:

Importez le jeu de données "billboard" à partir de la bibliothèque "tidyr".

Manipulez les données de telle sorte que chaque ligne contienne une chanson avec son titre, son artiste, sa semaine de classement et sa position dans les charts.

## 11. Tidyverse - Groupement et agrégation de données:

Importez le jeu de données "diamonds" à partir du package "ggplot2".

Regroupez les données par coupe de diamant et couleur puis calculez le prix moyen, la médiane et l'écart-type pour chaque groupe.

## 12. Tidyverse - Filtrage et tri des données:

Importez le jeu de données "mtcars".

Extraire les voitures qui ont une consommation de carburant supérieure à 20 miles par gallon. Ensuite, triez les voitures par cylindrée décroissante.

## 13. Exportation de données:

Exportez le résultat précédent dans un fichier CSV nommé "mtcars\_data.csv" sur votre ordinateur.

## 14. Filtre de données:

Sélectionnez toutes les observations dans le jeu de données "diamonds" où la qualité de la coupe est "Premium".

Sélectionnez toutes les observations dans le jeu de données "mpg" où la consommation de carburant (mpg) est supérieure à 25.

## 15. Sélection de colonnes:

Sélectionnez les colonnes "mpg" et "hp" dans le jeu de données "mtcars".

## 16. Ajout de colonnes:

Ajoutez une colonne "price\_per\_carat" dans le jeu de données "diamonds" qui calcule le prix par carat pour chaque diamant.

Ajoutez une colonne "cyl\_density" dans le jeu de données "mtcars" qui calcule la densité de cylindres pour chaque voiture (nombre de cylindres par unité de poids).

## 17. Tri des données:

Triez les observations dans le jeu de données "iris" selon la longueur de pétale décroissante (colonne "petal\_length").

Triez les observations dans le jeu de données "diamonds" selon le prix croissant.

## 18. Agrégation de données avec des fonctions personnalisées:

Créez une fonction qui calcule le pourcentage de valeurs manquantes dans une colonne et utilisez-la pour calculer le pourcentage de valeurs manquantes pour chaque colonne dans le jeu de données "titanic".

## 19. Résumé de données:

Calculez la moyenne et l'écart-type de la longueur de sépale pour chaque espèce dans le jeu de données "iris".

Calculez le nombre moyen de passagers pour chaque classe de cabine dans le jeu de données "titanic".