

Exercices - Semaine 1

Raphaël Nedellec

Préparation

1. Vérifiez que les options globales de RStudio sont conformes aux options recommandées dans le cours.
2. Créez un projet RStudio intitulé `cours_r_semaine_1`. Ce projet sera créé vide. Les options utilisation de `renv` et de `git` peuvent être ignorées.

Exercice

Pour cet exercice, nous allons télécharger un jeu de données issu de <https://data.gouv.fr> sur les infrastructures olympiques des JO 2024 à Paris. Le jeu de données est [disponible ici](#). L'objectif de cet exercice est d'utiliser uniquement des fonctions de R base. N'utilisez pas de librairie tierce sauf mention contraire.

1. Créez un script `.R` intitulé `exercice.R`.
2. Lisez le fichier téléchargé (ou directement depuis le site, cf aide plus bas) avec la fonction `read.table`. Le résultat sera affecté à l'objet de nom `data_ex`. Le jeu de données contient 4 colonnes. Quels sont les noms et la nature des colonnes ? Combien de lignes contient la `data.frame` ?
3. Combien y a t'il de sites olympiques ?
4. Combien y a t'il de sites paralympiques ?
5. Quels sont les sites qui accueillent plusieurs disciplines sportives ?
6. Quels sont les disciplines para-olympiques accueillies dans ces sites franciliens ?
7. Quel(s) site(s) accueille(nt) le plus de disciplines différentes ?
8. Quel discipline aura lieu sur le plus grand nombre de sites ? Quels sont ces sites ?
9. A vol d'oiseau, quels sont les sites les deux sites les plus proches ?
10. Quels sont les deux sites les plus éloignés ?
11. Vous êtes un heureux millionnaire et souhaitez investir dans un bien immobilier. Vous décidez d'acheter un appartement situé au barycentre de l'ensemble des sites olympiques. Où se situe cet appartement ?

Aide pour l'exercice

- Utilisez la fonction `?<code>` pour déterminer les bons paramètres de lecture. Il est aussi possible de lire directement le jeu de données depuis l'url suivant <https://data.iledefrance.fr/api/explore/v2.1/catalog/datasets/paris-2024-sites-olympiques-et-paralympiques-franciliens/exports/csv> en utilisant la fonction `read.table` et en le passant en paramètre.
- Les fonctions `grep`, `grepl`, `strsplit`, `regmatches` pourront vous être utiles pour identifier tout ou partie d'une chaîne de caractères, ou les manipuler. N'hésitez pas à regarder la documentation pour trouver la fonction appropriée.
- La fonction `distHaversine` est une fonction du package `geosphere` pour calculer des distances sur une sphère entre deux points géolocalisés.
- Vous pouvez installer le package `tidygeocoder` et utiliser la fonction `reverse_geocode` pour identifier le nom d'une adresse à partir d'une localisation.