Résumé Exposé R

Ousseynou GUEYE, Marie Agathe SECK, Dieynaba KA

2023-07-02

# Janitor

C’est un package qui permet de faire une analyse pour s’assurer qu’on a une bonne base. Il permet aussi de faire toute sorte d’opération sur les tableaux.

*Les fonctions essentielles de JANITOR* Il y a des fonctions qui permettent le nettoyage des données et des fonctions d’exploration de données. La fonction clean-names () : Qui permet de changer le nom des variables dans une base afin de rendre les variables beaucoup plus parlante. La fonction remove-empty () : Elle permet de supprimer les valeurs manquantes dans une base de données. La fonction remove-constant () : Elle permet de supprimer une colonne ou il y a des valeurs constantes (par exemple une colonne qui qui contient des valeurs identiques). La fonction get-dupes () : Elle examine les observations qui se répètent, elle permet de supprimer les doublons. La fonction –tabyl () : Pour faire sortir des combinaisons de tableaux. Cependant, elle n’est pas une fonction du package JANITOR

*Les limites du package :* JANITOR ne fait pas de tableaux, il faut faire appel à d’autres fonctions d’autres packages pour faire le format des tableaux.

# GTSUMMARY

Il permet de créer des tableaux statistiques descriptifs, on peut combiner plusieurs tableaux avec ce package.

*Les fonctions essentielles :* tbl-summary () : pour créer des tableaux flextable () : Pour exporter des tableaux (Word, PDF etc.) digits () : permet d’ajuster le nombre de chiffres après la virgule. tbl-cross () : Permet de créer un tableau de contingence en indiquant ce qu’on doit mettre en ligne et en colonne. tbl-split() : Permet de diviser le tableau et sortir une partie. tbl-stack () : Fusionne des tableaux par lignes tbl-merge () : Fusionne des tableaux par colonnes tbl-strata() : permet de faire une comparaison d’un groupe dans un autre groupe

# Rmarkdown

Rmarkdown est une bibliothèque sur R qui nous permet de produire des rapports, texte, article… L’extension rmarkdown permet de générer des documents de manière dynamique en mélangeant texte mis en forme et résultats produits par du code R. Les documents générés peuvent être au format HTML, PDF, Word, et bien d’autres. C’est donc un outil très pratique pour l’exportation, la communication et la diffusion de résultats d’analyse.

# Quarto

Dans R, le package “quarto” est une bibliothèque qui,comme pour R Markdown, permet de créer des documents reproductibles et interactifs mélangeant du code R, du texte, des images et des visualisations.A l’image de R Markdown, Il a ete conçu pour faciliter la création de rapports, de présentations et de livres.En comparaison avec R Markdown ,Quarto vient apporter plus de possibilités en termes de production de document comme: -creer des tableaux plus jolis que ceux proposés par R Markdown; -Personnalisation avancée : “quarto” offre une plus grande flexibilité de personnalisation par rapport à R Markdown. Vous pouvez utiliser des thèmes personnalisés, des styles CSS, des modèles HTML et d’autres outils de personnalisation pour ajuster l’apparence et le style de vos documents “quarto” selon vos besoins et vos préférences. -Prendre en charge plusieurs langages de programmation : Alors que R Markdown est principalement axé sur R, “quarto” offre la possibilité d’utiliser plusieurs langages de programmation, tels que R, Python et Julia, dans un même document. Cela permet d’intégrer facilement du code provenant de différentes sources et d’exploiter les fonctionnalités spécifiques de chaque langage.

Notons que l’une des limites de Quarto est ,q’un livre en quarto book ne propose pas pe sortie en word . la sortie est soit en format pdf ou HTML. Conseil: Même s’il est vrai que Quarto apporte également beaucoup de falicité quant à la rédaction de documents, Le choix entre Quarto et R Markdown dépendra des besoins spécifiques du projet, des langages de programmation utilisés et des fonctionnalités requises pour la génération de documents .

## R vers Excel

Le statisticien est souvent appelé utilisé à la fois R et excel pour aboutir aux résultats issus des traitements de données avant d’en sortir les rapports. Pour faire preuve d’éfficience, R nous offre cette facilté transcrire des résultats R vers Excel plus rapidement à l’aide d’un certain nombre de packages. Ainsi le passage de R vers Excel et vice versa est assuré par plusieurs packages dont on peut en citer:

-readxl:Ce package permet d’importer des données à partir de fichiers Excel

-writexl:Ce package permet d’exporter des données depuis R vers des fichiers Excel (.xlsx) sans nécessiter Microsoft Excel lui-même. Il offre une interface simple pour créer des fichiers Excel avec différentes feuilles de calcul et formats.

-openxlsx:Ce package permet de lire, d’écrire et de manipuler des fichiers Excel (.xlsx). Il offre des fonctionnalités avancées telles que la création de feuilles de calcul, l’ajout de graphiques, la gestion des styles et des formats de cellules, etc.

-xlsx:Ce package fournit des fonctions pour lire, écrire et manipuler des fichiers Excel (.xlsx) en R. Il permet également de créer des feuilles de calcul, de gérer les formats de cellules et de fusionner des cellules.

-r2excel:Le package “r2excel” permet de générer des fichiers Excel à partir de données et de résultats d’analyses réalisées dans R. Il fournit des fonctionnalités pour créer des feuilles de calcul, ajouter des données, formater les cellules, créer des graphiques et exporter les résultats dans des fichiers Excel.

En résumé, à travers les packages ci dessus, il est plus facilite pour l’analyste de données ou le statisticien de générer ses statistiques dans le format Excel à partir des résultats d’analyse obtenus sous R, avec possibilité de les présenter de manière conviviale,et d’appliquer des formats personnalisés .

## Réticulate

Dans la même optique que R ver Excel, le package Réticulate vient répondre à un besoin de communiquer avec deux langage : R et python.En d’autres termes,le package Réticulate a ete concu pour établir une communication trés fluide entre les R users et les Python users. Notons également que tout comme Reticulate,rPython,PythonInR, Jupyter Notebooks offrent aussi cette avantage d’utiliser du python dans R. De manière concret, Reticulate assure ce trait d’union entre R et Python en faisant principalement:

* Appeler de Python à partir de R de diverses manières, R Markdown, référencement de scripts Python, importation de modules Python et utilisation interactive de Python dans une session R;
* Convertir les objets R en Python et vice versa (par exemple, entre des data frame R et Pandas, ou entre des matrices R et des tableaux NumPy) Pour utiliser le package reticulate,il préférable s’assurer avoir réaliser un certain nombre d’instructions apres l’installation du dit package, notemment:
* Installer python dans sa machine (une version >= 2.7 est recommendée) Si Python n’est pas installé, vous pouvez toujors le faire directement en ligne, avec annaconda ou directement sur R avec Reticulate en utilisant la commande :
* Configurer l’environnement Python (ie spécifier la version de Python à utiliser)

Après les configuration nécessaires, le package Reticulate peut etre utilisé à des fins suivant: Excécuter un script Python dans R:

* Installer et importer des modules Python On peut utiliser la fonction import() pour importer n’importe quel module Python et l’appeler à partir de R après l’avoir installé avec py\_install
* Appeler des fonction Python depuis R.
* Ecrire et exécuter du code Python soit avec R Markdown, soit avec la fonction “py\_run\_string” dans reticulate, soit également d’acceder à une console par exemple en utilisant dans une console de R la fonction “repl\_python”.
* Accéder à des objets python dans R et vice versa
* Conversion des objets de python vers R et vice versa: notons ici qu’il faut connaitre les équivalents des objets R en python

Il est toutefois important de noter que ce jeu de vas et viens entre les deux lagages recquiere une bonne maitrise des synthaxe de R et de python.

# Text mining avec R

Text mining avec R est le traitement statistique de texte en utilisant certains packages disponibles dans R. Si nous avons une base de donnees, nous pouvons faire le prétraitement de texte, faire apparaître la fréquence des mots ou des ngrammes. Nous pouvons aussi déterminer la matrice terme des mots… Les packages utilisés sont entre autres *tidytext*, *dpyr* et *tm*.

Nous pouvons supprimer les nombres, convertir le texte en minuscule, supprimer une liste de mots non désirés, rechercher le mot le plus fréquent…

Ces packages aussi peuvent nous permettre de déterminer l’histogramme des mots utilisés dans un texte ainsi que le nuage des mots avec le package *tm*.

# Calcul parallèle avec R

Le traitement des données peut parfois présenter des limites, lorsque la quantité de données est grande, telles que des coûts en temps de calcul élevés, des problèmes de mémoire et des performances inférieures.

Pour optimiser le temps d’exécution d’un programme sur R, différentes stratégies peuvent être utilisées, telles que l’utilisation de fonctions optimisées, l’exploitation des calculs vectoriels et matriciels, et la minimisation des allocations mémoire inutiles. Cependant, lorsque ces méthodes ne suffisent pas, la programmation parallèle offre une solution appropriée. L’objectif de la programmation parallèle est d’accélérer les calculs informatiques en exploitant simultanément plusieurs unités de calcul.

Le principe du calcul parallèle est de subdiviser un programme initial en n blocs de programmes exécutés simultanément et retournant ainsi le résultat recherché.

Les packages comme *Parallel* inclus dans R, *doParallel*, *rmr2* sont souvent utilisés.

Nous pouvons à l’aide des fonctions comme *Parallel::Detectcores()* pour déterminer le nombre de coeur de la machine, *rmr2::mapreduce()* qui est une fonction clée qui fournit une interface pour exécuter des calculs distribués.

# Cartographie avec R

La cartographie des données géographiques avec R comprend plusieurs étapes essentielles. Tout d’abord, on utilise des packages tels que “sf” ou “sp” pour importer les données géographiques dans R, qu’elles soient au format shapefile, GeoJSON ou via des API en ligne. Ensuite, il est nécessaire de préparer les données en effectuant des opérations telles que la projection des coordonnées, la fusion des sources de données ou la manipulation des attributs.

Une fois les données préparées, on peut créer une carte de base en utilisant des packages tels que “ggplot2” et en spécifiant les coordonnées géographiques, les fonds de carte, les axes et les légendes. Les données préparées peuvent ensuite être ajoutées à la carte, en incorporant des points, des lignes ou des polygones correspondant aux données et en utilisant les attributs pour définir les couleurs, les tailles ou les symboles des éléments.

Pour personnaliser la carte, on peut utiliser les fonctionnalités de personnalisation de R pour ajouter des titres, des étiquettes, des échelles, des légendes ou d’autres éléments de design, et appliquer des techniques de mise en forme graphique. Enfin, une fois la carte créée, elle peut être exportée dans différents formats (PNG, JPEG, PDF) et intégrée dans des rapports ou des applications interactives.

La cartographie avec R offre de nombreuses possibilités pour visualiser et analyser efficacement des données géographiques dans des domaines variés tels que l’analyse de données, la géographie, l’épidémiologie, l’environnement et l’urbanisme.

# RShiny

RShiny est une bibliothèque permettant de développer des applications web interactives avec R. Les applications Shiny se composent d’un fichier UI pour définir l’apparence de l’interface utilisateur et d’un fichier Server contenant le code R pour le traitement des données.

On peut creer une interface en utilisant différents éléments tels que les boutons et les graphiques interactifs. Il y a aussi la programmation réactive qui permet de lier les entrées de l’utilisateur aux sorties correspondantes.

On peut aussi personnaliser l’apparence de l’application les options de déploiement. Bien que RShiny ait des limitations, telles que la dépendance à R et la limitation de personnalisation de l’interface utilisateur, il reste une solution populaire pour le développement d’applications web interactives avec R. Il est important de prendre en compte ces limitations lors de la planification et de l’adaptation de l’utilisation de RShiny selon les besoins du projet