

Agence Nationale de la Statistique et de la démographie (ANSD)

Ecole nationale de la statistique et de l'analyse économique

Pierre Ndiaye (ENSAE)

Dakar, Juin 2023

Résumé- Synthèse succ_Exposé_2023

Groupe: 2

Pathé diagne-Famara Sadio-Fallou Badji

Elèves ingénieurs Statisticiens Économistes en première année ISE.

pathediagne25@gmail.com

Professeur:

Mr Hema Aboubacar.

OBJECTIFS GENERALES

L'objectif général de ce document est de fournir un résumé synthétique complet et synthétique de l'ensemble des exposés présentés, mettant en évidence les principaux points discutés, les concepts clés abordés et les conclusions importantes tirées. Cela permettra à tout un chacun de bénéficier d'une vision d'ensemble de tous les sujets traités lors de l'exposé et de comprendre les principaux enseignements et messages transmis par les différents exposants.

Nous avons suivi avec beaucoup d'intérêt le troisième exposé qui a porté sur rmarkdown.

Dans cette présentation, les présentateurs ont exploré les fonctionnalités et l'importance de R Markdown, un langage de balisage puissant pour la génération de documents dynamiques.

Leurs développements et précisions nous ont permis de mieux comprendre l'importance de R Markdown dans la création de rapports interactifs, de présentations et d'autres types de documents.

Cela nous a permis d'apprécier son utilité dans le domaine de l'analyse de données.

Ainsi, les points ci-dessous ont été développés tout au long de cet exposé.

I. Fonctionnalités de R Markdown.

S'agissant des fonctionnalités de rmarkdown les points importants ont été relevés.

- La syntaxe de base de R Markdown permet de combiner facilement du texte et du code dans un même document, facilitant ainsi la création de documents bien structurés.
- L'intégration de code R dans les documents R Markdown offre la possibilité d'exécuter des analyses, de générer des graphiques et de présenter les résultats de manière interactive.
- Avec R Markdown, les graphiques et les résultats peuvent être affichés de manière dynamique, permettant une exploration interactive des données.
- R Markdown prend en charge plusieurs formats de sortie tels que HTML, PDF et Word, offrant ainsi une flexibilité pour générer des documents dans le format souhaité.

II. Gestion des titres et sous titres

Nous avons appris comment gérer les titres et sous-titres en utilisant la syntaxe Markdown.

- ✓ Pour spécifier le titre principal de votre document, on utilise le symbole "#" suivi du texte du titre. Par exemple, "# Titre Principal".
- ✓ Sous-titres thématiques : On utilise le symbole "#" suivi du texte du sous-titre pour créer des sous-titres thématiques.
- ✓ Il est possible d'utiliser plusieurs niveaux de sous-titres en ajoutant des "#" supplémentaires. Par exemple, "## Sous-titre 1", "### Sous-titre 2".

III. Utilisation et manipulation avancée de R Markdown .

- En utilisant R Markdown, il est possible d'inclure des images, des tableaux et des formules mathématiques pour enrichir les documents.
- La gestion des références bibliographiques dans R Markdown facilite l'intégration des sources et des citations dans les documents académiques.
- L'automatisation de la génération de rapports avec R Markdown permet de gagner du temps en générant automatiquement les documents à partir des données et du code.

Résumé-synthèse-exposés

IV. Avantages de l'utilisation de R Markdown

A travers l'exposer nous avons noté que l'utilisation de rmarkdown procure plusieurs avantages notamment:

- La reproductibilité des résultats qui est l'un des principaux avantages de l'utilisation de R Markdown, permettant de partager et de reproduire facilement les analyses.
- Grâce à R Markdown, le partage du code et des résultats devient simple, la collaboration et la transparence dans les projets d'analyse de données devient plus facile.
- L'automatisation de la génération de rapports avec R Markdown permet d'économiser du temps et de mettre en oeuvre des rapports d'analyse très présentables.

V. Exemples d'utilisation de R Markdown

- Exemples concrets de création de rapports interactifs avec R Markdown, mettant en évidence la flexibilité et l'interactivité offertes par cette approche.
- Nous avons vu comment R Markdown peut être utilisé pour générer des présentations Beamer, permettant une intégration harmonieuse du contenu et des résultats de l'analyse.
- Les exposants ont également examiné des exemples de production de documents scientifiques avec R Markdown, montrant comment il peut faciliter la création de documents bien structurés et professionnels.

En Conclusion, R Markdown offre des fonctionnalités puissantes pour la génération de documents dynamiques et reproductibles, combinant du texte, du code et des résultats dans un seul document.

R Quarto est un package R puissant et polyvalent pour la création de rapports dynamiques et interactifs. L'exposé sur R Quarto a mis en lumière ses fonctionnalités essentielles et son importance dans la génération de rapports.

I. Fonctionnalités de R Quarto :

Syntaxe claire et simple : R Quarto utilise une syntaxe intuitive qui permet de combiner facilement du texte et du code pour créer des rapports bien structurés.

Intégration de code R : R Quarto permet d'inclure du code R dans les rapports, ce qui facilite l'analyse des données, la création de graphiques et la génération de résultats interactifs.

II. Personnalisation avancée

R Quarto offre une grande flexibilité en termes de personnalisation des rapports, avec des options pour modifier l'apparence, le style et le formatage du contenu.

Prise en charge de multiples formats de sortie : R Quarto prend en charge plusieurs formats de sortie tels que HTML, PDF et Word, offrant ainsi une adaptabilité pour générer des rapports dans le format souhaité.

II. Gestion des titres et sous-titres :

R Quarto utilise une syntaxe similaire à R Markdown pour gérer les titres et les sous-titres. Les symboles "#" sont utilisés pour spécifier les niveaux de titres, permettant ainsi une hiérarchisation claire du contenu. En utilisant cette syntaxe, il est facile de créer des titres principaux, des sous-titres thématiques et des sous-sous-titres pour organiser efficacement le rapport.

III. Utilisation et manipulation avancée de R Quarto :

L'exposant montre que R Quarto offre des fonctionnalités avancées pour enrichir les rapports, y compris l'inclusion d'images, de tableaux et de formules mathématiques.

La gestion des références bibliographiques dans R Quarto facilite l'intégration des sources et des citations dans les documents académiques.

L'automatisation de la génération de rapports avec R Quarto permet de gagner du temps en générant automatiquement les rapports à partir des données et du code.

IV. Avantages de l'utilisation de R Quarto :

La principale force de R Quarto réside dans sa capacité à créer des rapports interactifs et dynamiques, permettant aux utilisateurs d'explorer les données et d'interagir avec les résultats.

R Quarto favorise la reproductibilité des résultats, facilitant le partage et la reproduction des analyses.

Grâce à R Quarto, le partage du code, des résultats et des visualisations devient simple, favorisant la collaboration et la transparence dans les projets d'analyse de données.

En résumé, l'exposé sur R Quarto a souligné son utilité dans la création de rapports dynamiques et interactifs. Ses fonctionnalités, la gestion des titres et sous-titres, ainsi que l'utilisation avancée avec inclusion d'images et de références bibliographiques ont été discutées. Les avantages de R Quarto incluent la reproductibilité des résultats, le partage facile du code et des résultats, ainsi que la facilité d'automatisation de la génération de rapports.

Le textemining

Le textemining permet de classer automatiquement des documents, d'obtenir un aperçu du contenu d'un document sans le lire, d'alimenter de manière automatique des bases de données, de faire de la veille sur des corpus documentaires et d'enrichir l'index d'un moteur de recherche pour améliorer la consultation des documents.

❖ Packages nécessaires

Les packages nécessaires pour faire le prétraitement sont :

- Library (tidytext)
- Library (dplyr)
- Library (tm)

tidytext et dplyr

➤ ###installation et téléchargement###

```
install.package("dplyr") ; install.packages("tidytext")
```

```
Library(dplyr) ; Library(tidytext)
```

🚦 Les fonctions nécessaires :

tidytext:: unnest_tokens(tbl,output,input,token,format,to_lower) : permet de séparer les mots d'un texte en lignes distinctes dans un dataframe ou tibble

tidytext:: stop_words() : permet d'arrêter une liste de mots figurant dans un texte

dplyr:: anti_join() : permet de supprimer les lignes d'un dataframe qui correspondent à celles d'un autre dataframe

dplyr:: filter() : permet de filtrer les lignes d'un dataframe selon des conditions spécifiées

dplyr:: count() : compte le nombre d'occurrences de chaque valeur unique dans une colonne spécifiée d'un dataframe.

dplyr:: mutate() : ajoute de nouvelles colonnes ou modifie les colonnes existantes dans un dataframe

dplyr:: tibble() : crée un nouveau tibble à partir des arguments fournis.

tm : c'est le package textmining le plus populaire de R

➤ ###installation et téléchargement###

```
install.package("tmr")
```

Library(tm)

Les fonctions nécessaires :

tm:: tm_map(txtel, removePunctuation) : permet de supprimer les ponctuations

```
### txtel <- removePunctuation(txtel)
```

tm:: tm_map(txtel, removeNumbers) : permet de supprimer les mots d'une liste

```
### txtel <- removeNumbers (txtel)
```

tm:: tm_map(txtel, removeWords, stopwords("english")) : permet de supprimer les mots dans un texte

```
### txtel <- removeWords(txtel, c("for", "his", "a", "of", "is", "hes", "IN", "I", "but", "i"))
```

tm:: tm_map(txtel, stripWhitespace) :

tm:: tm_map(txtel, tm_reduce) : tm_reduce est une fonction qui combine plusieurs transformations (mappages) en une seule. Elle prend en entrée un corpus et une liste de fonctions de transformation `tm` et renvoie une seule fonction de transformation `tm` obtenue en pliant les fonctions de la liste de droite à gauche.

tm:: VCorpus(VectorSource()) : permet de faire des transformations dans un format corpus

```
### txtel <- Corpus(VectorSource(txtel))
```

le package wordcloud :

est un générateur de nuages de mots disponible en R pour nous aider à analyser les textes et à visualiser rapidement les mots-clés sous forme de nuage de mots

➤ ###installation et téléchargement###

```
install.packages("wordcloud")
```

```
Library(wordcloud)
```

Les fonctions nécessaires :

wordcloud(): pour créer un nuage de mots à partir d'un vecteur de mots et de leurs fréquences

devtools::install_github("ifellows/wordcloud"): pour installer le package à partir de GitHub

scale: un paramètre dans la fonction `wordcloud()` permet de spécifier la plage de taille des mots

Le package gtsummary

Le package gtsummary fournit un moyen élégant et flexible de créer des tableaux analytiques et

de résumé prêts à être publiés en utilisant le langage de programmation R. Le package

gtsummary résume les ensembles de données, les modèles de régression et plus encore, en utilisant des valeurs par défaut sensées avec des capacités hautement personnalisables.

➤ ###installation et téléchargement###

```
Install.packages("gtsummary")
```

```
Library (gtsummary)
```

Les fonctions nécessaires :

- `tbl_summary()`: Résumez les data frames ou les tibbles pour présenter des statistiques descriptives, comparer les données démographiques de groupe
- `add_p()` ; `add_n()`, `bold_labels()`, etc.: Personnalisez les tableaux ``gtsummary`` en utilisant une liste croissante de fonctions de formatage/style: tout, des statistiques et des tests à utiliser au nombre de décimales à arrondir, en passant par les étiquettes en gras, les catégories d'indentation et plus encore
- On peut également utiliser des thèmes pour personnaliser l'apparence de vos tableaux avec des fonctions telles que ``set_gtsummary_theme()``, ``theme_gtsummary_journal()``
- `tbl_cross` : est une fonction qui crée un tableau croisé de statistiques récapitulatives.
- `tbl_continuous` : est une fonction qui résume une variable continue
- `tbl_uvregression()` : est utile pour réaliser plusieurs régressions univariées. Il faut lui passer un tableau ne contenant que la variable à expliquer et les variables explicatives. La variable à expliquer sera indiquée avec `y`. L'argument `method` indique la fonction à utiliser pour le calcul des modèles univariés.

Le package Janitor

L'étude du package 'janitor' est importante car il fournit des fonctions simples pour examiner et nettoyer les données sales. Il a été conçu pour les utilisateurs débutants et intermédiaires de R et est optimisé pour la convivialité. Les utilisateurs avancés de R peuvent déjà effectuer toutes les tâches couvertes ici, mais avec janitor, ils peuvent le faire plus rapidement et économiser leur réflexion pour les choses amusantes.

Parmi les fonctions principales de Janitor, on peut citer:

Formater parfaitement les noms de colonnes de data.frame;

Créer et formater des tableaux de fréquence d'une, deux ou trois variables

Pensez à une table améliorée (); et fournir d'autres outils pour nettoyer et examiner les data.frames.

➤ `###installation et téléchargement###`

```
install.packages("janitor")
```

```
library(janitor)
```

🔧 Les fonctions nécessaires :

- `clean_names()`: Cette fonction est utilisée pour changer et nettoyer les noms des colonnes dans les data frames.
- `remove_empty()`: peut être utilisée pour supprimer les lignes et les colonnes vides dans un jeu de données
- `get_dupes()`: Cette fonction récupère les doublons dans le jeu de données afin qu'ils puissent être examinés lors des opérations de nettoyage des données.

- `tabyl()`: Cette fonction est une version améliorée de la fonction `table()` qui permet de créer et de formater des tableaux de fréquence d'une, deux ou trois variables.

R vers EXCEL

R est un logiciel libre qui permet d'effectuer des graphiques et calculs mathématiques. Il est très utile pour les analyses statistiques, la mise en forme de graphiques, la création de plans d'expérience, le traitement de signal, l'utilisation comme un outil cartographique ou encore l'analyse d'images. D'un autre côté, Excel est un tableur qui permet de consolider des données issues de différentes sources au sein d'un même rapport³.

La combinaison des deux logiciels peut donc être très utile dans la réalisation d'un projet en permettant une analyse approfondie des données avec R et une consolidation et présentation claire des résultats avec Excel.

Les packages nécessaires :

Pour combiner R et Excel, on peut utiliser les packages suivants :

1. **"readxl"** : Ce package permet de lire les données à partir de fichiers Excel (.xls et .xlsx) dans R. Il fournit des fonctions simples pour extraire des feuilles de calcul, des plages de cellules spécifiques et d'autres informations à partir de fichiers Excel.

➤ `### installation et téléchargement###`

```
library(readxl)
```

2. **"writexl"** : Ce package permet d'écrire des données depuis R vers des fichiers Excel (.xlsx) sans avoir besoin d'Excel lui-même. Il offre des fonctionnalités pour créer de nouvelles feuilles de calcul, écrire des données dans des plages de cellules spécifiques et gérer les formats de données.

➤ `### installation et téléchargement###`

```
library("writexl")
```

3. **"openxlsx"**: Ce package fournit des fonctionnalités avancées pour lire, écrire et manipuler des fichiers Excel. Il permet de lire et d'écrire des fichiers Excel dans différents formats (xls, xlsx) et offre des options pour gérer les styles, les formules, les images et autres objets présents dans les fichiers Excel.

➤ `### installation et téléchargement###`

```
library("openxlsx")
```

4. **"xlsx"** : Ce package permet de lire et d'écrire des fichiers Excel en utilisant la bibliothèque Apache POI Java. Il prend en charge les formats de fichier .xlsx et .xls et fournit des fonctions pour extraire des données, créer de nouvelles feuilles de calcul, écrire des données dans des plages de cellules spécifiques,

➤ ### installation et téléchargement###

Library("xlsx")

5. **"XLConnect"** : Ce package fournit une interface pour lire, écrire et manipuler des fichiers Excel en utilisant la bibliothèque Java Apache POI. Il permet d'accéder aux données, aux feuilles de calcul et aux cellules dans les fichiers Excel, ainsi que d'effectuer

➤ ### installation et téléchargement###

Library("XLConnect")

6. **'r2excel'**:le package r2excel contient des fonctions pour rapidement et facilement importer, écrire et personnaliser des fichiers Excel

➤ ### installation et téléchargement###

install.packages("devtools")

devtools::install_github("kassambara/r2excel")

library(r2excel)

Les fonctions nécessaires :

- xlsx.addHeader pour ajouter des titres
- xlsx.addPlot pour ajouter des graphiques
- xlsx.addParagraph pour ajouter des paragraphes de textes
- xlsx.addTable pour ajouter des data.frames.
- xlsx.addLineBreak pour ajouter un saut de ligne
- xlsx.addHyperlink pour ajouter un hyper lien
- xlsx.writeFile : pour écrire une table de données dans un fichier Excel
- xlsx.writeMultipleData : pour exporter rapidement plusieurs tables de données dans le même classeur Excel.
- xlsx.openFile : pour ouvrir et voir un fichier Excel

Résolution de système d'équations non linéaires avec R

Résoudre un système d'équations non linéaires dans un langage de programmation comme R peut être important pour trouver des solutions à des problèmes complexes. Un système d'équations non linéaires est un système dans lequel au moins l'une des équations n'est pas linéaire. La solution d'un tel système peut être trouvée en utilisant des méthodes numériques, qui peuvent être implémentées dans un langage de programmation comme R. Cela peut être utile pour résoudre des problèmes dans divers domaines tels que l'ingénierie, la physique, l'économie, etc.

Les packages nécessaires :

Il y'a des packages sur R qui permettent de faire la résolution de système d'équation non linéaires. Parmi ces packages, il y'a certains qui permettent de faire la résolution directe et d'autres la

résolution indirecte.

➤ **Le package `nlesqlv`** : contient une fonction '**`nlesqlv()`**' permettant de résoudre des systèmes d'équation en utilisant la méthode de Newton et la méthode de Broyden

➤ **Le package `rootsolve`** : est une bibliothèque de fonctions disponibles dans le logiciel R qui fournit des outils pour la résolution numérique de systèmes d'équations non linéaires.

Dans ce package, il y a la fonction `multiroot()` qui utilise la méthode de newton pour résoudre des systèmes d'équations non linéaires.

➤ **Le package `stats`** : est un package de base de R contenant une fonction '**`optim()`**' permettant de faire la résolution des systèmes non linéaires. La fonction '**`optim()`**' en R est une fonction d'optimisation à usage général basée sur les algorithmes et de gradient conjugué.

➤ **Le package '`pracma`'** aussi, propose des méthodes numériques pour résoudre des équations non linéaires. Par exemple, les fonctions `fsolve()` et `broyden()` en utilisant la méthode de Newton ou la méthode de Broyden, respectivement.

REMARQUE :

Pour chaque fonction, on doit donner la valeur initiale et la fonction ou les systèmes d'équation et d'autres paramètres comme l'erreur ; c'est-à-dire la différence entre la solution cherchée et le nombre maximal d'itération.

R shiny

Rshiny est un package qui facilite la création d'application web interactives directement à partir de R et l'analyse des données. Les applications shiny sont contenues dans un seul script appelé `app.R` qui est dans un répertoire. On exécute ces applications avec `runApp`. Les entrées inputs et les sorties outputs forment les composants interactifs de de l'application. Les composantes d'appR sont l'UI, server et le shinyApp.

- L'UI : c'est ce qu'on voit. Il contrôle la mise en page et l'apparence de l'application.
- Le server : c'est le cerveau. Il contient les instructions dont l'ordinateur a besoin pour créer une application.
- La shinyApp crée des objets d'application à partir de Shiny à partir d'une paire UI/server explicite.

La barre latérale Layout a une fonction dédié `sidebarLayout()`. C'est une mise en page qui utilisée dans une mise en page fluide. La fonction `fluidRow()` permet de créer des lignes. Pour créer des colonnes, on utilise `column()` à l'intérieur de `fluidRow`.

La fonction `fluidPage` redessine dynamiquement les composants pour utiliser la totalité de l'espace disponible. Elle construit un layout de type grid constitué de lignes, elles-mêmes constituées de colonnes. Les lignes assurent que les éléments qu'elles contiennent apparaissent alignés horizontalement. Les colonnes définissent l'espace horizontal alloué à chaque élément. Cet espace est découpé en douze unités.

Les panels sont utilisés pour regrouper des éléments dans `fluidPage`. On peut ajouter un titre avec la fonction `titlePanel`. La partie principale de l'affichage est gérée dans la fonction `mainPanel`.

La fonction `renderPlot` permet de retourner un graphique `ggplot2` qui est lié à un composant de l'UI référencé par son `outputId` et la fonction `renderTable`, un tableau.

Il est possible de créer une liste de sélection qui peut être utilisée pour choisir un ou plusieurs éléments dans une liste de valeurs avec la fonction `selectinput`.

La fonction `updateselectinput` permet au server de réécrire le contenu d'un `selectInput` en utilisant les informations des objets `input` et `output`.

Cartographie sur R

Le logiciel R peut être utilisé pour la création de cartes simples permettant la représentation d'effectifs, de proportions ou de variables qualitatives pour les zones géographiques. Le stockage des données utilisées repose sur la classe d'objets `SpatialPolygonDataFrame`.

Pour représenter des données géographiques, il y a deux modes fondamentaux : le mode vecteur et le mode raster.

Un raster est une image constituée d'une matrice de pixels. Chaque pixel ayant ses propres valeurs. La résolution d'un raster dépend du nombre de pixels qu'il contient. Plus il y a de pixels dans une image, meilleur est sa résolution et sa qualité visuelle. Les rasters permettent de représenter des données continues et des données discrètes. Au format raster, les données continues peuvent être des données d'altitude. Pour importer des données raster sur R, on peut utiliser la fonction `raster()` issue du package `raster`.

Un vecteur représente un objet. Un vecteur est calculé, il demande donc quelques traitements de la part de l'ordinateur pour être affiché, mais peu de mémoire pour être stocké. Pour le format `shapefile`, les données spatiales sont stockées dans plusieurs fichiers qui portent le même nom avec des extensions différentes. Le fichier qui porte l'extension `.shp` contient toute l'information liée à la géométrie des unités spatiales. Il doit être accompagné de deux autres fichiers portant l'extension `.dbf` (dBase File) et `.shx` (Shape Index File). Le format GSON a l'avantage de contenir l'information dans un seul fichier.

La fonction `read.csv2()` permet d'importer des données classiques, des données non spatiales sur R.

La fonction `coordinates()` permet de passer d'un objet de classe `dataframe` à un objet de classe `spatial`.

La fonction `st_as_sf()` permet de transformer un objet de classe `dataframe` en objet de classe `shapefile`.

La fonction `readORG()` issu du package `rgdal` permet d'importer les types de données spatiales.

La fonction `st_read()` permet de lire de nombreux types de fichiers spatiaux. Et la fonction `st_write()` permet d'exporter de nombreux types de fichiers.

Pour représenter des cartes de base, on utilise la fonction `plot()` qui, appliquée à un objet `spatial` va seulement présenter la géométrie de l'objet. On peut ensuite utiliser les fonctions graphiques de base (`title()`, `legend()`, etc.) pour orner le graphique. Pour la norme `sf`, le principe est différent, une carte par variable sera automatiquement représentée. Pour ne sélectionner que la géométrie, il aurait fallu faire appel à la fonction `st_geometry()`.

Pour fusionner des objets spatiaux, on utilise la fonction `unionSpatialPolygons()` qui est dans le package `maptools`. L'objet créé, de classe `SpatialPolygons` ne contient que la géométrie des observations et pas de `dataframe`. On associe les géométries au jeu de données en utilisant la fonction `SpatialPolygonsDataFrame()`.

Pour pouvoir ajouter une unité spatiale à un objet `spatial`, il faut que les deux objets aient les mêmes attributs (variables) et on applique la fonction `spRbind` pour le collage.

Concernant l'affichage des étiquettes, on utilise `mf_label()` du package `mapsf`.

Pour afficher plusieurs cartes sur la même figure, on utilise l'argument `mfrow` de la fonction `par()`.

Calculs en parallèles

Le calcul en parallèle permet d'effectuer plus rapidement un calcul informatique en exploitant simultanément plusieurs unités de calcul. R exécute le programme plus rapidement de manière synchrone sur des bases de données volumineuses en utilisant simultanément les cœurs de l'ordinateur. Pour importer une grande base sur R sans pour autant la télécharger, on fait appel à la fonction `fs()` du package `fs`.

Pour pouvoir faire du calcul en parallèle, il est nécessaire d'importer le package `parallel` qui se base sur `apply`. La fonction `detectcores()` permet de déterminer le nombre de cœurs de l'ordinateur et la

fonction `stopcluster()` permet de stopper et libérer les différents workers.
Les packages `doParallel` et `foreach` permettent d'effectuer des boucles parallèles en R.
Il faut toutefois savoir qu'il est important de vider la mémoire de la machine avant de commencer le travail. On peut le faire avec la fonction `rm(list=ls())`.

R vers python reticulate

Nous avons assisté avec beaucoup d'intérêt à la présentation du groupe 9 qui portait sur la transition de R vers Python en utilisant le package Reticulate. Voici un résumé synthétique des points saillants de l'exposé

Introduction à Reticulate :

Les présentateurs ont commencé par expliquer ce qu'est Reticulate, un package de R permettant d'utiliser des fonctions et des scripts R dans un environnement Python. Cela permet aux utilisateurs de tirer parti des fonctionnalités de R tout en bénéficiant de l'écosystème Python. A travers l'exposé nous sommes capable de réaliser les manipulations en amont, nécessaires pour pouvoir installer le package reticulate, les extensions python nécessaires, etc...

Avantages de la transition : Les présentateurs ont souligné les avantages de passer de R à Python en utilisant Reticulate. Cela inclut la possibilité de combiner les capacités de R et de Python dans un même projet, d'accéder à une plus grande variété de packages et de bénéficier d'une plus grande flexibilité dans le développement et le déploiement de projets.

Syntaxe et utilisation de Reticulate : Les présentateurs ont présenté la syntaxe de base de Reticulate et ont montré des exemples concrets d'utilisation. Ils ont souligné la simplicité de l'intégration de scripts R dans du code Python existant et ont expliqué comment utiliser des fonctions R dans Python.

Conversion de données : Un aspect important de la transition de R à Python est la conversion des données entre les deux langages. Les présentateurs ont discuté des différentes méthodes de conversion de données, y compris l'utilisation de bibliothèques telles que pandas pour manipuler les données dans Python.

Cas d'utilisation : Les présentateurs ont partagé des exemples de cas d'utilisation où la combinaison de R et de Python avec Reticulate a été bénéfique. Cela inclut l'analyse de données, la visualisation, l'apprentissage automatique et d'autres domaines de l'analyse quantitative.

En conclusion, cet exposé sur la transition de R vers Python avec Reticulate nous a permis de comprendre les avantages et les possibilités offertes par l'utilisation de ce package. Nous avons appris comment intégrer facilement des scripts R dans du code Python, convertir des données entre les deux langages et exploiter les fonctionnalités de chaque langage dans nos projets d'analyse de données.

CONCLUSIONS ET REMARQUES

Nous avons assisté aux différents exposés que nous jugeons très intéressants sur différents sujets, notamment R Markdown et le package Reticulate, R vers Excel, RCartographie.....pour l'interopérabilité entre R et divers outils de traitement, d'analyse et d'exploration des données. Ces exposés ont mis en avant les fonctionnalités et l'importance de ces outils dans le domaine de l'analyse de données, en permettant la création de documents dynamiques et interactifs, ainsi que leur utilisation conjointe avec le langage R.

Enfin, ces exposés ont fourni des informations précieuses sur des outils puissants pour l'analyse de données, en mettant en avant leur utilité et leurs avantages pour les utilisateurs. Ils nous ont également permis en tant que participants de mieux comprendre comment ces outils peuvent être utilisés dans nos propres projets d'analyse de données et être plus efficient, plus performants.
