

# DATA ANALYSIS WITH R

## A LA DECOUVERTE DE R

**Arsène W. KIEMA** *Professionnal Data Scientist*

arsenekiema@statplus.net / (00226) 64 66 62 63

9/5/24



# Qu'est-ce que R?

R est un langage de programmation et un environnement logiciel dédié aux statistiques et à la science des données. Il est utilisé dans divers domaines notamment pour l'analyse statistique, la visualisation des données, le calcul scientifique, l'Intelligence Artificielle, etc.

# Origine et évolution

## Création

R a été créé par Ross Ihaka et Robert Gentleman au début des années 1990 à l'Université d'Auckland, en Nouvelle-Zélande. Leur objectif était de développer un outil gratuit et open-source pour l'analyse statistique.

# Origine et évolution

## Création

R a été créé par Ross Ihaka et Robert Gentleman au début des années 1990 à l'Université d'Auckland, en Nouvelle-Zélande. Leur objectif était de développer un outil gratuit et open-source pour l'analyse statistique.

## Nom

Le nom "R" a été choisi en partie en référence aux prénoms des créateurs (Ross et Robert) et aussi en hommage au langage S, développé chez Bell Labs.

# Origine et évolution

## Création

R a été créé par Ross Ihaka et Robert Gentleman au début des années 1990 à l'Université d'Auckland, en Nouvelle-Zélande. Leur objectif était de développer un outil gratuit et open-source pour l'analyse statistique.

## Nom

Le nom "R" a été choisi en partie en référence aux prénoms des créateurs (Ross et Robert) et aussi en hommage au langage S, développé chez Bell Labs.

## Évolution

Depuis sa création, R a connu une évolution rapide grâce à une large communauté de développeurs et d'utilisateurs. De nombreuses versions ont été publiées, chacune apportant des améliorations et de nouvelles fonctionnalités.

# Origine et évolution

## Popularité

R est devenu très populaire dans les milieux académiques, les institutions gouvernementales, les entreprises et les startups technologiques en raison de sa puissance, sa flexibilité et sa capacité à gérer de grands ensembles de données.

# Caractéristiques clés

## Syntaxe conviviale

R est connu pour sa syntaxe intuitive et lisible, qui facilite la manipulation et l'analyse de données complexes.

## Analyse statistique

R offre une vaste gamme de fonctions statistiques, des méthodes de base aux analyses avancées, ce qui en fait un choix privilégié pour les statisticiens.

# Caractéristiques clés

## Extensions et packages

La communauté R a développé des milliers de packages qui ajoutent des fonctionnalités spécifiques, de l'apprentissage automatique à la visualisation en passant par la manipulation de données. CRAN (Comprehensive R Archive Network) est la principale source de packages R.

## Graphiques et visualisation

R excelle dans la création de graphiques de haute qualité, permettant de visualiser efficacement les données et les résultats d'analyse.

## Intégration de données

R peut se connecter à diverses sources de données, qu'il s'agisse de fichiers plats, de bases de données relationnelles ou même de flux en direct.



# Quelques avantages de R

## Open Source

R est un logiciel open source, ce qui signifie qu'il est librement disponible et peut être utilisé par quiconque sans frais. Cela a contribué à la croissance rapide de la communauté R et à la création de nombreux packages et ressources.

## Vaste Écosystème de Packages

R possède un écosystème riche en packages qui fournissent une multitude de fonctionnalités spécialisées. Qu'il s'agisse de statistiques, d'apprentissage automatique, de visualisation, de manipulation de données ou d'autres domaines, il y a probablement un package R pour répondre à vos besoins.

# Quelques avantages de R

## Puissant pour l'Analyse Statistique

R est conçu par et pour des statisticiens, ce qui se reflète dans ses capacités avancées d'analyse statistique. Des tests d'hypothèses classiques aux méthodes plus complexes comme les modèles linéaires généralisés, R offre une gamme complète d'outils pour les analyses statistiques.

## Visualisation de Données

R excelle dans la création de graphiques et de visualisations de données. Le package 'ggplot2' en particulier permet de créer des graphiques esthétiquement plaisants et informatifs avec une syntaxe relativement simple.

# Quelques avantages de R

## Communauté Active

R a une communauté en pleine croissance composée de chercheurs, de statisticiens, de scientifiques des données et d'autres professionnels. Cette communauté active contribue à la création de packages, au partage de connaissances et à la résolution de problèmes.

## Intégration de Données

R peut se connecter à diverses sources de données, notamment des fichiers plats, des bases de données relationnelles, des données en streaming et des services web. Cela facilite la manipulation et l'analyse de données provenant de sources variées.

# Quelques avantages de R

## Personnalisation et Répétabilité

Les scripts R sont réutilisables et permettent une personnalisation avancée. Cela facilite la répétabilité des analyses et l'automatisation des tâches.

# Quelques avantages de R

## Personnalisation et Répétabilité

Les scripts R sont réutilisables et permettent une personnalisation avancée. Cela facilite la répétabilité des analyses et l'automatisation des tâches.

## Documentation Abondante

Avec une abondance de ressources en ligne, de tutoriels et de documentation, les nouveaux utilisateurs peuvent rapidement apprendre R et commencer à l'utiliser efficacement.

# Quelques avantages de R

## Personnalisation et Répétabilité

Les scripts R sont réutilisables et permettent une personnalisation avancée. Cela facilite la répétabilité des analyses et l'automatisation des tâches.

## Documentation Abondante

Avec une abondance de ressources en ligne, de tutoriels et de documentation, les nouveaux utilisateurs peuvent rapidement apprendre R et commencer à l'utiliser efficacement.

## Multiplateforme

R fonctionne sur différentes plateformes, y compris Windows, macOS et Linux, ce qui le rend accessible à un large public.

# Domaines d'application de R

## Statistiques

Analyse statistique descriptive et inférentielle, modélisation statistique, tests d'hypothèses.

# Domaines d'application de R

## Statistiques

Analyse statistique descriptive et inférentielle, modélisation statistique, tests d'hypothèses.

## Science des données

Préparation des données, exploration des données, visualisation des données, apprentissage automatique.



# Domaines d'application de R

## Statistiques

Analyse statistique descriptive et inférentielle, modélisation statistique, tests d'hypothèses.

## Science des données

Préparation des données, exploration des données, visualisation des données, apprentissage automatique.

## Bioinformatique

Analyse des données génomiques, protéomiques, et transcriptomiques, visualisation de séquences biologiques.

# Domaines d'application de R

## Finance

Analyse des marchés financiers, modélisation financière, gestion des risques, prévisions économiques.

# Domaines d'application de R

## Finance

Analyse des marchés financiers, modélisation financière, gestion des risques, prévisions économiques.

## Géosciences

Analyse spatiale, géostatistique, cartographie.

# Domaines d'application de R

## Finance

Analyse des marchés financiers, modélisation financière, gestion des risques, prévisions économiques.

## Géosciences

Analyse spatiale, géostatistique, cartographie.

## Marketing

Analyse des comportements des consommateurs, segmentation de marché, modèles de prédiction des ventes.

## Enseignement et recherche académique

Utilisation dans les cours de statistiques, de science des données, et dans la recherche académique.

# Téléchargement et installation de R

## Téléchargement de R

- ➊ Rendez-vous sur le site officiel du projet R pour le calcul statistique à l'adresse <https://www.r-project.org/>.
- ➋ Cliquez sur le lien "download R" dans la section « Getting Started »
- ➌ Choisissez un site miroir CRAN proche de votre emplacement dans la liste des sites miroirs.
- ➍ Sélectionnez votre système d'exploitation (Windows) et cliquez sur l'option "base".
- ➎ Téléchargez la dernière version de R pour Windows en cliquant sur le lien fourni.

# Téléchargement et installation de R

## Installation de R

- 1 Localisez le fichier téléchargé (par exemple, "R-x.x.x-win.exe") et double-cliquez dessus pour lancer l'installateur.
- 2 Suivez les instructions de l'assistant d'installation
- 3 Les paramètres par défaut devraient être suffisants pour la plupart des utilisateurs, vous pouvez donc généralement continuer en cliquant sur "Next" ou "Install" jusqu'à ce que l'installation soit terminée.

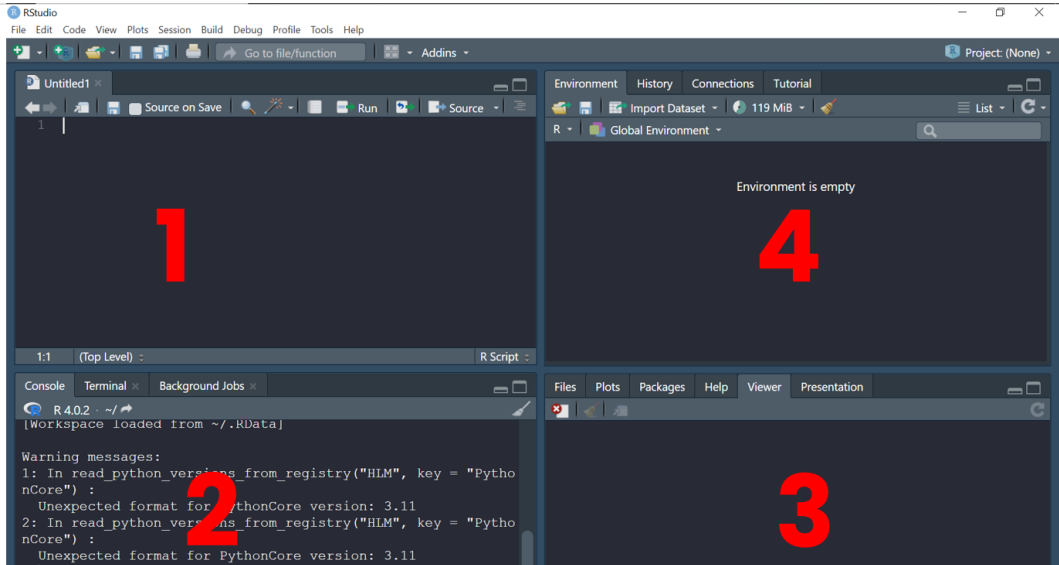


# Qu'est-ce que RStudio ?

RStudio est un environnement de développement intégré (IDE) spécialement conçu pour le langage de programmation R. C'est l'un des outils les plus populaires parmi les utilisateurs de R pour développer, tester, déboguer, analyser et visualiser des données à l'aide du langage R. RStudio vise à rendre le processus de travail avec R plus efficace et convivial en fournissant une interface utilisateur intuitive et une gamme de fonctionnalités utiles.



# Qu'est-ce que RStudio



# Quelques points clés de RStudio

## Interface Conviviale

RStudio offre une interface utilisateur ergonomique qui intègre différents panneaux pour éditer le code, afficher les résultats, explorer les variables, visualiser les graphiques et accéder aux fichiers et aux packages.

## Éditeur de Code

L'éditeur de code intégré à RStudio propose des fonctionnalités telles que la coloration syntaxique, l'autocomplétion, la mise en forme du code, la numérotation des lignes et des outils de débogage.

# Quelques points clés de RStudio

## Console R

RStudio dispose d'une console R intégrée qui permet d'exécuter des commandes R en temps réel et d'afficher les résultats directement dans l'interface. Cela permet de tester rapidement des commandes et de voir les sorties sans avoir à quitter l'IDE.

## Environnement de Travail

RStudio fournit un environnement de travail organisé avec des onglets pour les scripts, les graphiques, la console, l'historique des commandes et plus encore, facilitant la gestion de multiples aspects de votre projet.

# Téléchargement et installation de RStudio

## Téléchargement de RStudio

- ➊ Rendez-vous sur le site officiel de RStudio à l'adresse <https://www.rstudio.com/>.
- ➋ Cliquez sur le menu déroulant "Products" dans la barre de navigation supérieure et sélectionnez "RStudio IDE".
- ➌ Faites défiler jusqu'au niveau du bouton "Download RStudio".
- ➍ Cliquez ensuite sur "download rstudio desktop for windows" pour débuter le téléchargement.

## Installation de RStudio

- ➊ Localisez le fichier téléchargé (par exemple, "RStudio-x.x.x.exe") et double-cliquez dessus pour lancer l'installateur.
- ➋ Suivez les instructions de l'assistant d'installation.

# Téléchargement et installation de RStudio

Une fois le processus d'installation terminé, vous devriez avoir à la fois R et RStudio installés sur votre ordinateur Windows. Vous pouvez lancer RStudio et commencer à l'utiliser pour vos tâches d'analyse statistique et de traitement des données.

