



Présentation du logiciel R

Réunion du Comité Technique Systèmes
d'Information

03 Décembre 2015



Les besoins

- Gestion des données au sein d'un système d'information (SI)
- Nécessité de croiser les données avec celles d'autres organismes
- Valorisation des bases de données existantes
- Solutions techniques favorisant le travail en équipe (cf recherche reproductible) et facilitant la publication des résultats.
- Le tout avec des fichiers de plus en plus importants.



Les solutions courantes actuelles

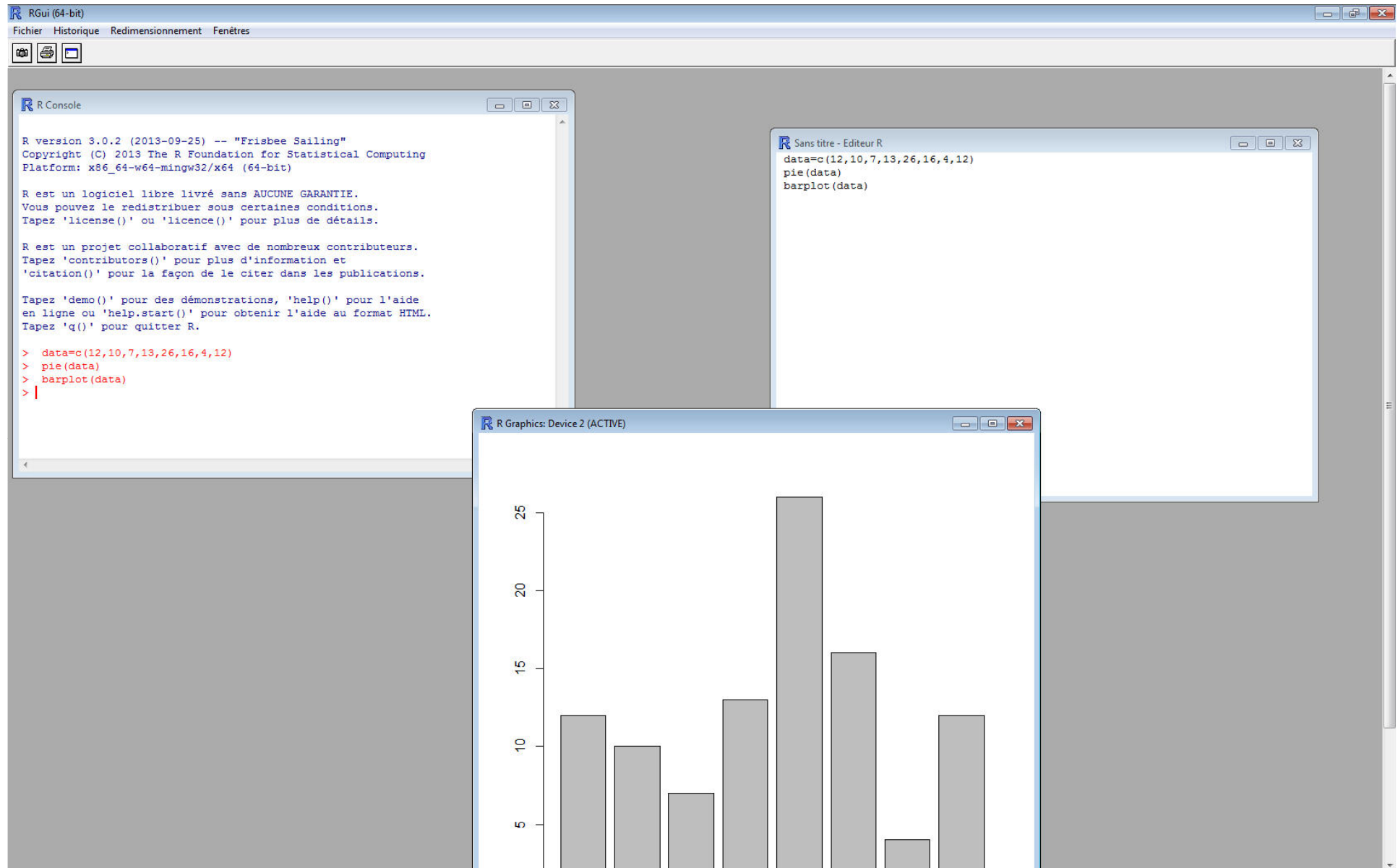
- Tableur
- Access
- SAS
- BD PostgreSQL (PostGIS)
- les produits Arcgis
- Python



Le logiciel R

- Logiciel et langage de programmation d'analyse statistique
- Existe depuis 1993
- licence libre
- logiciel disponible sous différents systèmes d'exploitation (Windows, Linux, Mac)

Le logiciel R





Les avantages

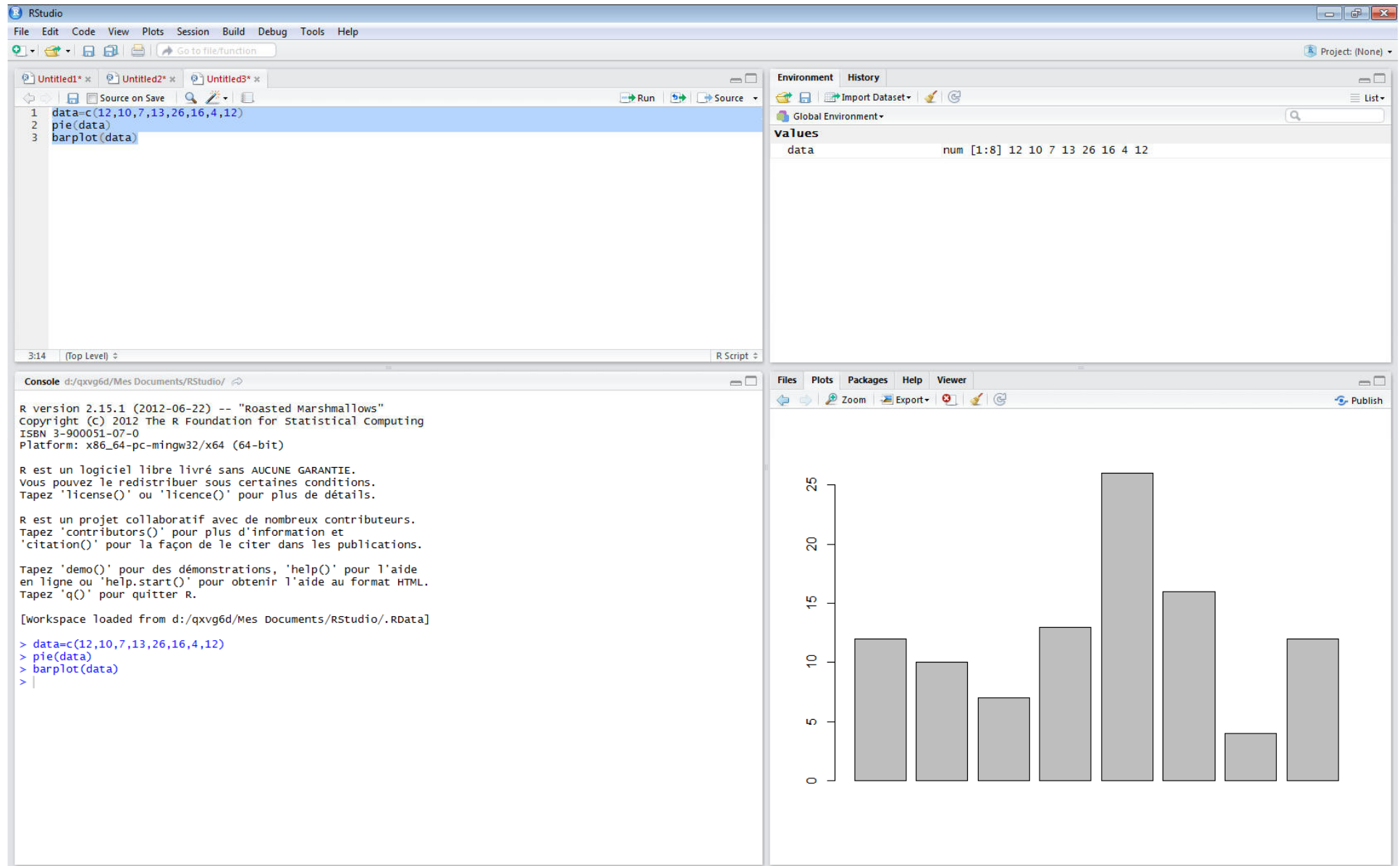
- Logiciel très puissant, extensions (7.400)
- Excellentes capacités graphiques
- Écosystème R : R est capable de travailler avec de nombreux autres formats et logiciels (LateX, PDF...)
- Possibilité d'extraire des données web.
- Un développement très actif lié à une forte communauté d'utilisateurs et l'implication croissante de sociétés commerciales.
- Usage cartographique ([Lien](#))



Les inconvénients

- Logiciel en anglais et nécessité de programmation = phase formation
- Interface graphique très sommaire, mais il en existe d'autres à disposition dont RStudio
- Données traitées dans la mémoire vive (avantage et inconvénient)

R Studio





Usage sommaire

- Gestion des données tabulaires
 - Importation et visualisation des données
 - Mise en forme des tables
 - Gestion des valeurs manquantes
 - Jointures et agrégations
 - Exportation des données mises en forme
- Accessible au débutant très rapidement à partir de quelques fonctions seulement.



Usage avancé :

- Travail collaboratif et publications
 - Tout le processus de traitement des données dans R est enregistré dans un script (programme)
 - Un document (html, pdf, docx, ods) peut être généré directement dans le script
 - En actualisant les données de départ et relançant le script le document est actualisé (tables, indicateurs, graphes, hors commentaires).
 - La diffusion des données et du script commenté est alors assimilable en partie à de la « recherche reproductible ».
- Assez facile pour des documents de travail, compliqué pour des rapports publiables. Pour des infographies élaborées, le recours à la PAO reste d'actualité.



Usage avancé (2)

- Présentation



Utilisations

- A l'INSEE
 - Analyse de données (du tri à plat aux analyses factorielles en passant par la réalisation des graphiques de travail en particulier les boîtes à moustache)
 - Développement d'outil « cliquable » (distancier, appli-carto)
- A la DREAL
 - Utilisation expérimentale en complément à SAS.



Liens

- R :
 - Le site du projet : <http://cran.rproject.org/>
 - Les packages classés par thème : <https://cran.r-project.org/web/views/>
- Rstudio : <https://www.rstudio.com/>
- Autres :
 - Forum français des utilisateurs de R : <http://forums.cirad.fr/logiciel-R/>
 - Un site sur le langage R : <http://www.duclert.org/Aide-memoire-R/Le-langage/Introduction.php>
 - Un forum en anglais des utilisateurs de R : <http://www.r-bloggers.com/>
 - R et espace est un livre complet qui présente l'utilisation du logiciel pour l'analyse de données simples géo-référencées et leur analyse : <http://framabook.org/r-et-espace/>
 - Formation en ligne ([Lien](#))



Merci de votre attention