

Workshop « R pour les archéologues ! »

Organisé par

Anaïs Vignoles

anaïs.l.vignoles@gmail.com

et

Aymeric Hermann

aymeric.hermann@cnr.fr

Un outil au service de la science reproductible

Le langage de programmation R permet de réaliser efficacement une grande variété de procédures, d'analyses et de visualisation de données. Que ce soit pour créer des documents, créer des graphiques, analyser des jeux de données semi-quantitatifs ou faire du SIG, R permet de faire à peu près tout ce que vous voulez. Il permet une meilleure reproductibilité en gardant une trace écrite (le fameux « code ») de toutes les étapes de votre travail d'analyse.

Les logiciels R et RStudio sont des logiciels puissants et gratuits développés en « source ouverte » (open source), et la création de *packages* (ensemble de fonctions qui améliorent ou étendent les fonctionnalités de base de R) est assurée par une grande communauté de programmeurs (et parfois collègues archéologues !) qui partagent le fruit de leur travail.

De façon générale, l'utilisation de R (ou de tout autre langage de programmation) dans la réalisation du travail de recherche permet de se questionner sur les logiques et la cohérence de nos pratiques en décomposant les choix que l'on opère dans le processus d'analyse des données, ainsi que dans certaines opérations de modification ou de réarrangement de la présentation de celles-ci dans le cadre de leur traitement et de leur publication.

L'univers de R peut être intimidant au premier abord, mais en réalité il s'agit avant tout d'un langage facile à appréhender et à utiliser. Le tout début peut être un peu décourageant car il faut se familiariser avec une toute nouvelle façon d'envisager ses données et les opérations que l'on souhaite faire ; mais la plupart des gens qui utilisent R s'accordent à dire que la courbe de progression est exponentielle et il est très facile de s'améliorer si l'on utilise cet outil régulièrement. En outre, il existe une quantité énorme de ressources en ligne et une communauté très dynamique sur les forums qui permettent de trouver rapidement la plupart des solutions aux problèmes courants. Et si vous aimez les casse-têtes et énigmes, travailler avec R peut rapidement se transformer en jeu d'enquête qui stimule la créativité dans la recherche de solutions à des problèmes !

Notre proposition de Workshop autour de R

L'objectif de ce petit Workshop de 4 séances de 2h est de vous introduire au langage de programmation R au travers de grandes notions ou familles d'opérations : explorer et arranger vos données ; visualiser des données quantitatives, qualitatives et/ou spatiales ; automatiser des opérations sur de grands jeux de données ; travailler à une meilleure reproductibilité et transparence de nos analyses. Les séances seront divisées en **1h consacrée à la découverte d'un ensemble de codes dans un « script »** et de différentes opérations prédéfinies, et **1h consacrée à une mise en situation** sur un jeu de données fourni par nos soins, avec de petits défis à relever (que vous pourrez continuer chez vous !)

Voici un programme préliminaire des choses que vous pourrez apprendre au cours de ces séances.

1) Se familiariser avec le langage de programmation R

Dans cette séance, vous apprendrez **les rudiments du langage R** : on y explorera un petit jeu de données (fictif) issu d'un tableau excel et l'on apprendra à créer un graphique simple et à l'enregistrer automatiquement.

2) Analyser et visualiser un jeu de données semi-quantitatif

Dans cette séance, vous découvrirez deux packages très utiles et complémentaires : **tidyr** (qui permet de manipuler et modifier des tableaux de données) et **ggplot** (qui permet de créer des figures). On y apprendra à modifier ou créer des données ainsi qu'à gérer tous les paramètres d'une figure (couleurs, taille de police, etc) en utilisant les fonctions spécifiques de ces packages.

3) Projections spatiales automatisées

Cette séance vous permettra de découvrir une fonctionnalité extrêmement utile de R : celle de créer des « boucles », c'est-à-dire de répéter le même code (i.e., les mêmes opérations) sur plusieurs jeux de données différents. Concrètement, cela permet par exemple de créer automatiquement des projections par « tranches » de x cm et de les enregistrer dans un dossier plutôt que de subdiviser vos tableaux et de les projeter à la main.

4) Les SIG sur R

Cette séance vous introduira à l'univers des SIG sur R, à travers la création de cartes avec des données **raster** (i.e., variables continues, comme l'élévation ou la température) et **vector** (i.e., variables discrètes, comme des sites ou des rivières). Nous verrons également comment explorer et analyser sommairement ces données spatialisées.

Nous vous conseillons dans la mesure du possible d'assister aux 4 séances, car les notions importantes seront abordées au fur et à mesure et serviront d'une séance sur l'autre.

Quelques ressources pour aller plus loin

Voici quelques ressources si vous souhaitez vous renseigner sur les possibilités d'utilisation de R :



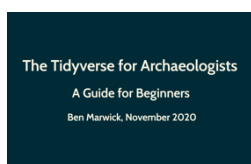
Le blog d'Allison Horst qui « vulgarise » R sous la forme d'illustrations !



Découvrir R et RStudio est un manuel en français réalisé par Thierry Zorn, Murielle Lethrosne, Vivien Roussez, Pascal Irz & Nicolas Torterotot. Il s'agit d'un dispositif de formation proposé par les Ministères de la transition écologique et solidaire (MTES), et de la Cohésion des territoires et des Relations avec les collectivités territoriales du gouvernement français.



Un document présentant la Gestion des données avec R est proposé par Christophe Lalanne et Bruno Falissard. Il y est notamment expliqué sous quelle forme apparaissent les données par rapport à excel (concept de data frame) et comment les importer pour travailler dans Rstudio.



[Ben Marwick](#) est professeur à l'Université de Washington et est l'un des archéologues les plus actifs dans le développement de pratiques liées à l'"ouverture" des données et à la reproductibilité des analyses et codes sources. A voir notamment ce powerpoint introductif, et une [liste de publications dans le domaine de l'archéologie qui incluent des scripts R](#).



Rchaeology



Le groupe Rchaeology est très actif sur Twitter et a une page dédiée aux débutants sur son site internet.

