

Atelier

« Traitements statistiques avec le logiciel R pour des applications en démographie historique »

**Proposé par l'axe « Concepts et méthodes »
du réseau de l'INED - IN-HOPPE**

International Network - Historical and Osteoarchaeological Past Populations Exploration
<https://in-hoppe.site.ined.fr>

**Et organisé par Arnaud Bringé
(Service Méthodes statistiques de l'Ined)**

Campus Condorcet – Aubervilliers (93) -

La nécessité de cet atelier est apparue face aux multiples demandes d'aide d'historien.ne.s confronté.e.s. à l'exploitation de données nominatives anciennes, multi-sources et souvent biaisées. Leur analyse suppose de répondre aux questions que posent l'hétérogénéité de données issues de documents aux finalités différentes, et leur complétude plus ou moins grande selon les époques et les lieux (normalisation des informations textuelles par exemple). Que ces corpus concernent l'époque moderne ou contemporaine, voire soient plus anciens encore, leur traitement informatique et leur exploitation statistique posent de fait des problèmes méthodologiques et techniques très semblables. La comparaison de données internationales nécessite d'avoir également une réflexion en amont sur la nature des sources et les objectifs qui sous-tendent leur compilation.

L'analyse de ce type de sources a été au centre de plusieurs séances de travail et journées d'étude coordonnées par l'axe « Concepts et méthodes » du réseau IN-HOPPE (*International Network - Historical and Osteoarchaeological Past Populations Exploration*).

Sur la base de l'expérience et de l'expertise acquises au fil des années, nous ouvrons aujourd'hui un atelier de perfectionnement dédié à ces questions, et centré sur les problématiques habituellement rencontrées en démographie historique. L'un des atouts de cet atelier sera d'intégrer des données de différentes natures et de diverses époques, afin de confronter les méthodologies forcément hétérogènes, car reposant sur des sources et des problématiques très variées.

Nous vous proposons de participer à un Atelier e 4 jours, **du 24 au 27 mars 2020 sur le Campus Condorcet à Aubervilliers.**

Objectifs

Cet atelier s'adresse avant tout aux chercheurs, ingénieurs et doctorants en sciences humaines, habitués à travailler sur des sources historiques, qu'il s'agisse de données statistiques ou pré-statistiques, nominatives ou non, mais exploitées dans une perspective socio-démographique.

Il vise à donner des bases informatiques à travers la maîtrise de fonctionnalités dédiées du logiciel R, à sensibiliser à quelques problèmes récurrents en démographie historique et à proposer des solutions permettant de les résoudre, et à offrir la maîtrise d'outils de dénombrement et de visualisation de données historiques. Cela conduira à s'interroger notamment sur la construction des bases de données et à l'introduction de la mesure dans la critique historique dans le cadre de données quantitatives.

L'objectif est que les participants puissent ainsi acquérir deux types de compétences complémentaires, concernant la structuration des données en vue de leur analyse statistique d'une part, et la manipulation aisée du logiciel R d'autre part.

Il est fortement souhaitable que les participants aient déjà un corpus constitué, afin de pouvoir présenter leurs sources et leur problématique, avant d'en analyser progressivement les différentes facettes, tout au long de cette formation. Néanmoins, merci de nous contacter si tel n'est pas votre cas, afin que nous réfléchissions ensemble, en fonction de vos motivations, à la possibilité de vous intégrer à cette formation en binôme avec un collègue disposant de données d'analyse.

Programme indicatif

La formation sera déclinée en une session de 4 jours, avec des données utilisateurs. Les 2 premiers jours seront consacrés à la maîtrise du logiciel R.

Pour ce premier atelier, le nombre de places est limité à 12 personnes maximum.

Une priorité sera donnée aux membres du réseau IN-HOPPE. Un atelier complémentaire pourrait être proposé ultérieurement en fonction du nombre de personnes intéressées.

- JOUR 1 : Présentation des participants / Rappels de statistiques élémentaires. Introduction au logiciel R (Concepts, environnement de programmation, importation et lecture de données, premiers traitements descriptifs). Présentation des différentes sources utilisables en démographie historique (BMS, Recensement, Listes de liens, sources archéologiques, etc.) avec des exemples
- JOUR 2 : Introduction au *tidyverse*, ensemble de packages R destinés à la structuration de données et à la *datavizualisation*

- JOUR 3 : Problèmes posés en démographie historique : Identification des individus, Mise en commun de sources hétérogènes, cohérence des sources
- JOUR 4 : Outils statistiques à disposition des historiens : Analyse de réseaux, Suivi statistique des individus, Analyses multivariées (ACM, CAH), Graphiques évolués ...

Prérequis :

- Aucun pré-requis en statistique, ni concernant le logiciel R. Néanmoins il est préférable d'avoir une expérience de l'utilisation d'un logiciel statistique ou a minima de traitements sous Excel
- Avoir chargé le logiciel R (<https://cran.r-project.org/>)
- Comme indiqué plus haut, il est vivement souhaitable de disposer d'une base de données constituée et avoir défini sa problématique

Modalités pratiques :

- Coûts et prise en charge : en contrepartie de la gratuité de la formation, les frais de transport, d'hébergement et de repas sont à votre charge.
- Restauration du midi à la cantine de l'Ined (prix restauration collective)
- Ordinateur portable **indispensable**

Inscriptions :

Nous vous remercions de répondre à notre invitation avant le **20 janvier**, et de nous faire parvenir, par mail, une brève présentation de vos données et de votre problématique.

Pour de plus amples informations, contacter [Arnaud Bringé](#).