



**HAL**  
open science

## **Stage d'analyse de données individuelles avec R et professionnalisation. Expérience d'utilisation des fichiers détail du recensement de la population en Master 2**

Pierre Pistre, Hadrien Commenges, Claude Grasland

### ► **To cite this version:**

Pierre Pistre, Hadrien Commenges, Claude Grasland. Stage d'analyse de données individuelles avec R et professionnalisation. Expérience d'utilisation des fichiers détail du recensement de la population en Master 2. Journées d'études "Compter, mesurer, calculer...raisonner? Enseigner le quanti en sciences sociales dans le supérieur : pratiques et défis", Jun 2021, Nantes, France. <hal-03522917>

**HAL Id: hal-03522917**

**<https://hal.science/hal-03522917v1>**

Submitted on 12 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## **Stage d'analyse de données individuelles avec R et professionnalisation : expérience d'utilisation des fichiers détail du recensement de la population en Master 2**

Pierre Pistre (1), Hadrien Commenges (2), Claude Grasland (1)

(1) Université de Paris / UMR 8504 Géographie-cités

(2) Université Paris 1 Panthéon-Sorbonne / UMR 8504 Géographie-cités

Notre proposition consiste en un **retour d'expérience sur un enseignement d'analyse de données en Master 2 de géographie** qui est dispensé, sous cette forme, depuis 2015. Son format (stage intensif d'une semaine), sa philosophie (utilisation de R, exploitation des fichiers détail en accès libre du recensement...) ou encore son public (étudiant.e.s de deux masters plus ou moins familiers du traitement de données) pourront permettre de contribuer à plusieurs axes des journées d'étude : sur les objectifs des enseignements de quanti notamment de professionnalisation, sur les spécificités des disciplines et des publics, sur les choix de logiciels enseignés, sur des « trucs et astuces », etc.

Précisons d'abord le contexte institutionnel. Il s'agit d'un **cours mutualisé entre deux formations**, d'une part le **Master Carthagéo**<sup>1</sup> qui prépare principalement aux métiers de la géomatique, et d'autre part le **Master Géoprisme**<sup>2</sup> qui associe des profils thématiques et méthodologiques (quanti et/ou quali) plus variés. Il est obligatoire pour les étudiant.e.s du premier master (une vingtaine) et optionnel pour le deuxième (une dizaine en moyenne). Ces deux formations sont par ailleurs adossées à plusieurs tutelles universitaires (Université Paris 1 Panthéon-Sorbonne, Université de Paris et/ou Ecole Nationale des Sciences Géographiques), et les enseignants sont membres de l'UMR Géographie-cités qui a une tradition ancienne dans l'analyse quantitative des données géographiques.

Ce cours, aussi bien son contenu que son évaluation, se déroule sur **une semaine complète** au mois d'octobre, c'est-à-dire en début d'année universitaire mais une fois que des cours de mise à niveau ont eu lieu sur le langage R et sur l'analyse de données.

Un premier objectif du stage consiste précisément à **approfondir la connaissance des étudiant.e.s dans l'exploitation de données statistiques détaillées dans l'espace** avec R. Cela passe déjà par l'utilisation de données individuelles volumineuses que ce sont les fichiers détail du recensement de la population<sup>3</sup> et qui, malgré leur richesse, restent sous-exploités en géographie. Il s'agit aussi de mettre l'accent sur des enjeux spécifiques à l'analyse de données dans l'espace, par exemple sur le choix des mailles d'étude (IRIS, communes, intercommunalités, etc.) et l'enseignement de méthodes d'analyse et de statistique spatiales (cartographie thématique, autocorrélation spatiale, etc.). Ces dimensions méthodologiques occupent les deux premiers jours de formation, qui incluent une présentation détaillée des données, leur prise en main par chaque étudiant (sélections, agrégations, statistiques univariées) et des approfondissements (statistiques bivariées et spatiales).

Le deuxième objectif est étroitement lié aux modalités d'évaluation du stage. Il s'agit d'appliquer les approfondissements méthodologiques dans le cadre de **projets collectifs par petits groupes (4 ou 5 étudiant.e.s) pour répondre à une commande « fictive »**. Par exemple en octobre 2020, suite à la victoire de plusieurs listes écologistes aux élections municipales dans différentes grandes villes

---

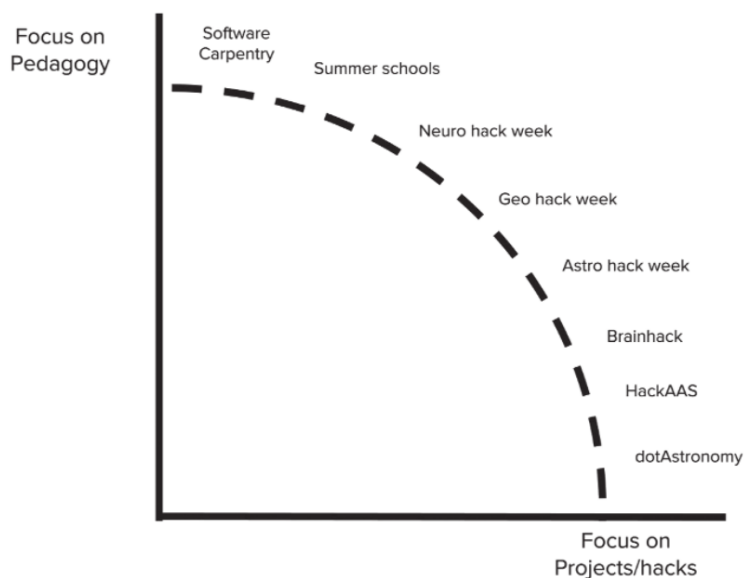
<sup>1</sup> Pour plus d'informations sur le Master Carthagéo : <https://www.pantheonsorbonne.fr/diplomes/master-professionnel-carthageo/>

<sup>2</sup> Pour plus d'informations sur le Master Géoprisme : <https://mastergeoprisme.wordpress.com/>

<sup>3</sup> Au cours des dernières années, le stage a porté successivement sur le fichier des logements ordinaires (LOGEMT), des mobilités résidentielles entre communes (MIGCOM) ou des mobilités quotidiennes (MOBPRO). Pour accéder aux données en *opendata* sur le site de l'Insee : <https://www.insee.fr/fr/statistiques/4508161>.

françaises (Lyon, Bordeaux, Poitiers, Tours, etc.), la consigne prenait la forme d'une demande urgente d'évaluation de la dépendance automobile des ménages dans ces métropoles<sup>4</sup> – c'est-à-dire pendant la seule durée du stage, et en autonomie mais avec un accompagnement personnalisé par groupe pendant 2 jours. L'originalité tient ici surtout au **double format de restitution des résultats** en fin de semaine : 1) la production d'**un 4 pages de type « Insee Analyses »** (cf. définition du sujet, analyses statistiques rigoureuses, vulgarisation des résultats, etc.) ; 2) **une restitution orale en « jeu de rôles »** - les étudiant.e.s dans la peau des auteurs de l'étude et l'équipe enseignante dans le rôle de différents acteurs (surtout représentant Insee et élus locaux) pour varier le registre des questions : sur l'utilisation des données, sur la compréhension et l'implication des résultats...

En définitive, ce stage intensif qui prenait initialement une forme plus classique – une journée étant dédiée à une méthode, avec du cours magistral et des exercices d'application – a progressivement donné plus de place à des projets sur l'ensemble de la semaine, et à la participation (plus) active des étudiant.e.s dans le cadre de collectifs associant différents profils de compétences. Ainsi, si on considère le domaine de la pure programmation informatique (même si la programmation n'est pas le seul intérêt de cette formation, elle y joue un rôle important), la forme stabilisée du stage se situe au milieu de ce schéma et correspond à la **hack week**. Elle en a les attributs : intensité et continuité du travail sur un temps court, travail d'équipe, nécessité de techniques de débrouille.



Source : Huppenkothen D. (2018) "Hack weeks as a model for data science education and collaboration", PNAS, vol.115, n°36, pp.8872-8877.

<sup>4</sup> L'étude portait essentiellement sur l'exploitation du fichier détail des logements ordinaires qui contient notamment des informations sur l'équipement automobile des ménages et le mode principal de transport de la personne de référence pour le trajet domicile-travail des actifs occupés.