



Année 2019-2020

Atelier « Structures et quantification des faits sociaux et économiques »

Contact :

Pierre Blavier (pierre.blavier@univ-lille.fr)

Fabien Eloire (fabien.eloire@univ-lille.fr)

Séance 1 : « Introduction à la cartographie sous R », Jessie Lerousseau (Clersé)

Jeudi 7 novembre 2019 – 10h30-12h (C107, SH3)

L'idée de cette séance, qui se veut interactive, est de nous amener à voir comment traiter une analyse cartographique sous R. Deux moments principaux rythmeront cette séance. Dans un premier temps, nous aborderons la question de l'importation de données géolocalisées et de leur appariement à des données numériques. Dans un second temps, nous reproduirons cet exercice en nous focalisant sur la manipulation de données géolocalisées et les possibilités offertes par le « zonage à façon » – agrégation de données géolocalisées à une échelle géographique supérieure. Nous prendrons l'exemple du taux de chômage de la métropole européenne de Lille, que nous cartographierons par Iris (maillage infracommunal) et par commune. La séance sera ouverte à toutes formes de suggestions qui pourraient faciliter la compréhension de l'exercice.

Séance 2 : « Le calage sur marges facile », Alexandre Mairot (CDSP)

Jeudi 12 décembre 2019 10h30-12h (C108, SH3)

Le calage sur marge est une procédure nécessaire pour redresser un échantillon par pondération des individus en utilisant une information auxiliaire disponible sur un certain nombre de variables, appelés variables de calage. Le redressement consiste à "caler" les marges du tableau croisant toutes les variables des effectifs connus. Durant cette séance, nous aborderons rapidement le principe général du calcul du calage sur marge et nous nous arrêterons plus longuement sur sa mise en oeuvre par l'utilisation de la procédure SAS, CALMAR développé par l'INSEE ainsi que le package ICARUS sous R.

Liens utiles :

Principe du calage sur marge : <https://www.insee.fr/fr/information/2387498>

Présentation package ICARUS : <https://cran.r-project.org/web/packages/icarus/index.html>

Présentation procédure CALMAR :

<https://www.insee.fr/fr/statistiques/fichier/2021902/doccalmar.pdf>

Séance 3 : « Le traitement de verbatim courts (injures) », Sébastien Delarre (Clerse)

Jeudi 30 janvier 2020 – 8h30-10h (C107, SH3)

La séance présente une façon de traiter des sources textuelles multiples mais courtes (du type de celles qu'on récupère avec les champs ouverts des questionnaires), à la différence de ce qui se fait traditionnellement en textométrie (textes longs mais relativement peu nombreux). Elle présente la constitution d'un dictionnaire, l'étiquetage des items, la décomposition des verbatim en mots, l'appariement avec le dictionnaire, la ré-agrégation au niveau de l'observation, et des exemples d'exploitation (sur les injures raciales et sexistes). Les algorithmes seront présentés en version SAS et Python.

Séance 4 : « Analyse des correspondances multiples : positions et prises de position », Fabien Éloire & Timothée Verley

Jeudi 19 mars 2020 – 10h30-12h (C107, SH3)

Dans cette séance, nous présentons deux enquêtes qui mobilisent l'analyse des correspondances pour étudier la structure sociale de deux sous-champs : les galeries d'art brut et les acteurs de la controverse économique sur le "tournant de la rigueur" (au sommet de l'Etat en France en 1983). Nous montrons comment nous choisissons les variables actives, et interprétons les résultats obtenus en termes de « position » des agents. Puis nous montrons comment il est possible d'ajouter des variables supplémentaires afin d'étudier les « prises de position » (opinions, choix, pratiques) des agents selon la place qu'ils occupent dans la structure.

Argumentaire de l'atelier

L'atelier « **Structures et quantification des faits sociaux et économiques** » poursuit deux objectifs.

Premièrement, il vise à présenter des travaux de recherche qui s'intéressent au "fait économique", entendu au sens large - et interdisciplinaire - de tout ce qui peut relever, de près ou de loin, de la sphère productive : travail, emploi, entreprise, profession, marché, consommation, organisation, etc.

Deuxièmement, partant du postulat que nombre d'objets de recherche en sciences sociales possèdent une dimension structurale et peuvent potentiellement être étudiés en mobilisant une forme ou une autre de quantification, il se propose d'être un espace d'échanges autour de cette approche.

L'atelier n'est pas réservé aux quantitativistes confirmés, il ne vise pas spécialement à présenter des outils et méthodes sophistiqués, mais il ne s'en tient pas non plus à la "critique des chiffres" (même s'il la trouve nécessaire et salutaire). Il s'adresse à tout chercheur désireux d'introduire une dimension structurale et quantitative dans son travail, et qui souhaite en discuter à partir d'un jeu de données concret. Il traite de bases de données, de variables, de codages, de choix de la population (échantillonnage), de pourcentages et d'interprétations, bref de ce qui fait le quotidien des travaux de recherche, dans leur aspect plus quantitatif et structural. Il présente, de préférence, des études en cours, encore inachevées, où il y a matière à discuter, à présenter des matériaux bruts, à entrer dans la "boîte noire" des données empiriques, de leur collecte systématique, de leurs sources et de leur codage pour les analyser.

L'atelier est donc conçu comme un espace de discussion ancré dans des méthodologies et des *pratiques de la quantification* en sciences sociales. Sans même parler du

big data et des possibilités infinies offertes par le numérique, il prend acte du développement de l'accès à une importante offre de jeux de données quantitatives, issus de "grandes enquêtes" ou du courant de la "science reproductible". Il s'appuie aussi sur le constat que la palette des méthodes, de l'analyse de réseaux à l'analyse de séquences en passant par les méthodes économétriques de régression ou les méthodes factorielles, n'a cessé de s'élargir ces dernières décennies.

Il part enfin du principe que l'opposition entre méthodes qualitatives et quantitatives n'a plus lieu d'être, tant il apparaît aujourd'hui que les deux domaines sont complémentaires, et tant il a été montré que les possibilités d'articulation sont nombreuses et fructueuses. Cependant, il tient aussi compte du fait que les approches en termes de quantification ont leur propre *matérialité*, autrement dit qu'elles impliquent des *conditions sociales et matérielles* de mise en oeuvre : maîtrise d'une certaine technicité, et nécessité de maîtrise de certains outils bureautiques et informatiques, tels le tableur ou les logiciels de traitement statistique. Bref, il tient compte du fait que la quantification ne met pas tout le monde à l'aise, et s'efforce, par son existence même et son organisation, de diffuser cette culture commune, et d'être espace de socialisation *qui à ce jour n'existe pas au Clersé*, autour de ces questions.

L'atelier est ouvert à toutes-tous, membres internes du laboratoire, ou externes sur sollicitation des organisateurs. Il implique une présentation des résultats empiriques, certes, mais surtout des données qui ont permis d'y aboutir. La présentation doit être courte pour laisser place à la discussion. La séance fait l'objet d'une préparation conjointe, en amont, entre l'intervenant et les organisateurs.

Comité d'organisation : Placide Abasabanye, Pierre Blavier, Ornella Boutry, Anne Bustreel, Benoît Cart, Ada Chmilevski, Catherine Comet, Sébastien Delarre, Marie-Aude Depuiset, Alexandre Guerillot, Guillemette de Larquier, Fabien Éloire, Anne Fretel, Thomas Houssoy, Arthur Jatteau, Clotilde Lemarchant, Jessie Lerousseau, Lucy Marquet, Marion Mauchaussee, Simon Nadel, Martine Pernod-Lemattre, Héloïse Petit, Manuel Schotté, Adèle Sébert, Marie-Hélène Toutin-Trelcat, Timothée Verley, Michael Zemmour.