

Introduction

Pourquoi une suite ?

Ce nouvel ouvrage fait suite à « *Traiter des données historiques* » paru en 2001 dont il se veut le prolongement et un complément. Plusieurs sollicitations ont concouru pour nous remettre à l'ouvrage.

Ce furent d'abord des comptes rendus de lecture dans lesquels les auteurs pointaient l'absence de telle ou telle méthode qu'ils jugeaient fertile dans le champ historique. Nous ne pouvions qu'être réceptifs à ces suggestions, mais le périmètre que nous nous étions fixés à l'époque pour ce premier ouvrage obéissait à la règle suivante : n'aborder que des méthodes et outils que nous avons expérimentés dans le cadre de nos enseignements universitaires. À partir de là, l'idée d'une suite a commencé à cheminer et a pris forme pour répondre, d'abord, aux besoins des étudiants qui exploraient des chantiers historiques ouverts récemment. Ce fut aussi la lecture du livre de Claire Lemerrier et Claire Zalc, *Méthodes quantitatives pour l'historien* qui ouvrait de nombreuses pistes et l'appétit pour les explorer.

Cet ouvrage résulte d'une collaboration bicéphale : histoire et mathématiques. La partie historique du tandem avait des sources et soumettait hypothèses et questionnements... à la partie « matheuse ». Laquelle, pour y répondre, suggérait outils et méthodes, et s'empressait de les mettre en œuvre. Restait alors la nécessité de décrire toutes ces démarches, en les illustrant au besoin d'exemples pédagogiques, sous une forme comestible pour des historiens.

Boîte à outils, mais pas boîte noire

Notre ambition est d'offrir une panoplie de méthodes, en exposant pour chacune :

- ✓ Les concepts de base sous-jacents, avec le vocabulaire qui les décrit. Au fil du texte, nous avons choisi de donner des explications aussi simples que possible en reléguant, pour les lecteurs intéressés, dans des « enclos mathématiques » des présentations plus formalisées (i.e. avec des formules mathématiques).
- ✓ Le type des données qu'elles traitent et le formatage approprié pour être soumis à un logiciel.
- ✓ L'usage pratique détaillé d'un logiciel pour opérer les traitements.
- ✓ Les aides à l'interprétation des résultats.

En bref, notre souci a été de concilier théorie et pratique...

Chacune des parties correspond à une famille d'outils :

- ✓ La première partie est consacrée à des compléments sur les bases de données. Le premier chapitre s'adresse à des lecteurs désireux d'organiser et de mettre éventuellement en ligne une base de données historiques. Le second expose comment extraire la substantifique moelle d'une base de données via l'interrogation en langage SQL.
- ✓ La deuxième partie est vouée à l'analyse de données multidimensionnelles. Autrement dit, comment traiter des données qui décrivent les individus statistiques par de nombreuses variables (dans le premier ouvrage, nous nous étions limités à 2). Sont étudiées les méthodes de classification, l'Analyse en Composantes Principales (pour des variables quantitatives), l'Analyse des Correspondances Multiples (pour les variables qualitatives). À ces méthodes exploratoires, nous avons adjoint une méthode de nature « explicative » : la régression logistique.
- ✓ La temporalité s'introduit dans la troisième partie sous l'appellation d'analyse longitudinale. Le chapitre Analyse des Séquences vise à comparer et classer des parcours temporels décrits comme une succession d'états. Le chapitre Event History

Analysis s'intéresse au rythme de propagation d'un événement au sein d'une population.

- ✓ La quatrième partie est consacrée à l'Analyse des Réseaux Sociaux, application de la théorie mathématique des graphes à l'étude de liens entre acteurs sociaux. Dans cette partie, nous nous sommes limités à trois thèmes : la cohésion, la centralité, les rôles et positions.
- ✓ La cinquième partie présente un SIG et l'usage que peut en faire un historien. Nous présentons alors le géoréférencement, la vectorisation des fonds et l'analyse des données.

L'acquisition d'une « boîte à outils » n'implique pas l'usage de tous les outils de la boîte. Nous nous sommes donc efforcés de rendre, autant que faire se peut, indépendants les différents chapitres afin qu'un lecteur intéressé par telle méthode puisse trouver dans le chapitre qui lui est consacré tous les éléments nécessaires. Ce parti pris de rédaction a une contrepartie : d'inévitables redites dans certains chapitres. Et des exceptions : nous avons dû commettre quelques infractions à cette règle :

- ✓ Le chapitre III « Nuage de points » est une introduction utile aux trois méthodes d'analyse multidimensionnelle présentées dans cette partie.
- ✓ Le chapitre IV d'introduction au logiciel de Statistique R est indispensable pour aborder les chapitres V, VI, VII, VIII, IX, X qui recourent systématiquement à ses services.
- ✓ Le chapitre XI d'introduction à la théorie des graphes et à l'analyse des réseaux, le chapitre XII de présentation du logiciel Pajek, sont requis pour les autres chapitres de cette partie.
- ✓ Les méthodes de classification exposées au chapitre V sont utilisées dans les chapitres IX et XV.

Les exemples

Chaque méthode est illustrée par le traitement complet d'un ou deux exemples relevant du champ historique. Certains, comme le réseau des familles florentines, sont bien connus et peuvent être considérés comme du domaine public, d'autres sont issus de données originales.

Les deux bases de données servant d'exemple au chapitre II sont interrogeables via des requêtes SQL sur le site :

<http://jacquescellier.fr/histoire>

Tous les fichiers sources des exemples (formats divers : excel, R, pajek...) y sont aussi téléchargeables, ainsi que d'autres documents.

Les logiciels

Contrairement aux méthodes présentées dans le premier ouvrage qui pouvaient être mises en œuvre avec des logiciels dits de « bureautique », les méthodes exposées dans ce deuxième tome, requièrent des logiciels spécialisés. Pour d'évidentes raisons, nous nous sommes orientés vers des logiciels, sinon libres au sens « officiel », mais du moins gratuit. Parfois, le choix s'est imposé « naturellement » : MySQL pour la gestion des bases de données, R pour tous les traitements statistiques abordés dans les parties II et III. Pour l'analyse des réseaux sociaux et la cartographie, le choix fut plus épineux, mais il fallut bien trancher.

Les sujets non abordés

Avec une boîte à outils se pose naturellement le choix des outils dont on veut la garnir. Faute d'opérer une sélection toujours discutable mais nécessaire, il ne s'agit plus d'une boîte mais d'un entrepôt... Il existe donc des thèmes auxquels nous avons délibérément renoncé pour des raisons variées : absence de compétence de notre part sur le sujet, trop grande complexité de mise en œuvre, nécessité de longs développements

pour les aborder, parfois l'ombre d'un doute sur la solidité des fondations, ou tout simplement l'existence d'ouvrages traitant très bien du sujet. C'est ainsi, par exemple, que nous avons renoncé à présenter des outils utilisés pour les méthodes d'analyse de discours ou pour l'édition partagée des textes, comme ARCANE, par exemple.