



# **Préparer et analyser les données de "Demandes de valeurs foncières" en open data : proposition d'une méthodologie reproductible**

Boris Mericskay, Florent Demoraes

## **► To cite this version:**

Boris Mericskay, Florent Demoraes. Préparer et analyser les données de "Demandes de valeurs foncières" en open data : proposition d'une méthodologie reproductible. *Cybergeo : Revue européenne de géographie / European journal of geography*, 2022, Cartographie, Imagerie, SIG, Document 1031, <10.4000/cybergeo.39583>. <hal-03792187>

**HAL Id: hal-03792187**

**<https://hal.science/hal-03792187v1>**

Submitted on 29 Sep 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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

---

## Préparer et analyser les données de "Demandes de valeurs foncières" en *open data* : proposition d'une méthodologie reproductible

*Preparing and analyzing the "Demande de valeurs foncières" open database: proposal for a replicable methodology*

*Preparar y analizar datos abiertos para el mercado de suelos francés : propuesta de una metodología reproducible*

**Boris Mericskay et Florent Demoraes**

---



### Édition électronique

URL : <https://journals.openedition.org/cybergeo/39583>

DOI : 10.4000/cybergeo.39583

ISSN : 1278-3366

### Éditeur

UMR 8504 Géographie-cités

Ce document vous est offert par Université Rennes 2



### Référence électronique

Boris Mericskay et Florent Demoraes, « Préparer et analyser les données de "Demandes de valeurs foncières" en *open data* : proposition d'une méthodologie reproductible », *Cybergeo: European Journal of Geography* [En ligne], Cartographie, Imagerie, SIG, document 1031, mis en ligne le 28 septembre 2022, consulté le 29 septembre 2022. URL : <http://journals.openedition.org/cybergeo/39583> ; DOI : <https://doi.org/10.4000/cybergeo.39583>

---

Ce document a été généré automatiquement le 29 septembre 2022.



Creative Commons - Attribution 4.0 International - CC BY 4.0

<https://creativecommons.org/licenses/by/4.0/>

---

## RÉSUMÉ / ABSTRACT / RESUMEN

Longtemps stockées dans différentes bases complexes et peu accessibles, les données concernant les transactions foncières et immobilières sont aujourd'hui plus facilement disponibles depuis l'ouverture en 2019 de la base "Demandes de valeurs foncières" (DVF). Toutefois, l'utilisation de ces ressources informationnelles à des fins analytiques nécessite un temps d'acculturation et un travail de préparation important qu'il ne faut pas sous-estimer. Cet article présente une chaîne de traitements dédiée à l'analyse des données DVF en *open data* que nous avons formalisée dans l'environnement R et documentée dans un souci de reproductibilité. Cette chaîne de traitements illustre le potentiel que représentent ces données pour appréhender les dynamiques spatiales des marchés de l'immobilier résidentiel à plusieurs échelles. À travers une étude de cas portant sur la région Bretagne et la métropole rennaise, l'objectif est de proposer et de discuter un ensemble de pistes méthodologiques relatives à la préparation, l'analyse et la représentation (carto)graphique de ces "nouvelles données" ainsi que les enjeux associés à leur utilisation.

The data concerning land and real estate transactions, which for a long time were stored in various complex and sparsely accessible databases, are now more easily available since the opening of the DVF database ("Demandes de Valeurs Foncières" in French, "Real Estate Transaction Datafiles" in English) in 2019. However, the use of these information resources for analytical purposes requires time for acculturation and significant preparation work that should not be underestimated. This article presents a workflow dedicated to the analysis of open data DVF that we formalized in the R environment and documented in order to ensure replicability. This workflow illustrates the potential of these data to understand the spatial dynamics of residential real estate markets at different scales. Through a case study of the Brittany region and the Rennes urban/metropolitan area, the objective is to propose and discuss some methodological avenues for the preparation, analysis and (carto)graphic representation of these "new data" as well as the challenges and issues associated with their use.

Durante largos periodos, la información referente a transacciones de terrenos e inmuebles, ha sido almacenada en bases de datos complejas y poco accesibles. Actualmente estas se encuentran con mayor disponibilidad y particularmente la base de datos de Solicitudes de valores del suelo (Demandes de Valeurs Foncières, DVF, en francés) se encuentra a disposición desde el año 2019. Sin embargo, la utilización de estos recursos con fines analíticos requiere un tiempo de comprensión y un importante trabajo de preparación que no debe subestimarse. Así, este artículo presenta una secuencia de procesamiento en código abierto aplicada a la base de datos DVF, formalizada en R y documentada para asegurar su reproducibilidad. Tal secuencia, ilustra el potencial de estos datos para comprender la dinámica espacial de los mercados inmobiliarios residenciales a diferentes escalas. A través de un estudio de caso de la región de Bretaña y el área metropolitana de Rennes, se propone y discute una serie de orientaciones metodológicas referentes a la preparación, análisis y representación cartográfica y gráfica de estos "nuevos datos", así como los desafíos y problemáticas asociadas a su uso.

---

# Préparer et analyser les données de "Demandes de valeurs foncières" en *open data* : proposition d'une méthodologie reproductible

*Preparing and analyzing the "Demande de valeurs foncières" open database: proposal for a replicable methodology*

*Preparar y analizar datos abiertos para el mercado de suelos francés : propuesta de una metodología reproducible*

**Boris Mericskay et Florent Demoraes**

---

## Introduction

- 1 Les données immobilières ont connu ces dernières années de profondes transformations à la fois dans leurs modalités de production, de mise à disposition et d'utilisation (Simon, 2009 ; Juillard, 2019). En réponse aux pressions foncières et aux tensions des marchés immobiliers que connaissent de nombreux territoires, les acteurs publics multiplient les dispositifs d'observation et de mesure de l'immobilier pour orienter leurs politiques d'aménagement. En ce sens, les données sur les transactions immobilières et foncières constituent des ressources informationnelles stratégiques, aussi bien pour les gestionnaires, les opérateurs, les négociateurs que les chercheurs (Boulay *et al.*, 2012 ; Chappert *et al.*, 2014 ; Saujot, Erard, 2015 ; Coulondre, 2018).
- 2 Longtemps restée une affaire de spécialistes, dans la mesure où ces informations étaient présentes au sein de bases parcellaires, peu accessibles voire payantes, la mise à disposition en *open data* de la base "Demandes de valeurs foncières" (DVF), produite par la Direction Générale des Finances Publiques (DGFIP) en 2019, a ouvert de nouvelles perspectives. Déjà explorées depuis plusieurs années dans la sphère académique, les données relatives aux transactions foncières, notamment celles issues des bases

notariales, ont fait l'objet de nombreuses études (Buhot, 2006 ; Guérois, Le Goix, 2009 ; Boulay, 2011 ; Despond, Bergel, 2013). Avec l'ouverture des données DVF, de multiples possibilités se dessinent pour de nouveaux publics. L'étendue des informations associées aux données DVF participe à améliorer la connaissance du marché de l'immobilier et l'évaluation des biens qui le composent (Masclat de Barbarin, 2019). Ces nouvelles perspectives de compréhension des marchés immobiliers sont liées à la masse des données à traiter, aux possibilités de moduler les échelles d'observation, aux opportunités d'analyse offertes par la vaste palette d'outils *open source* et à la combinaison de plusieurs modes de représentation.

- 3 Les 17 millions de mutations de biens fonciers-immobiliers collectées depuis 2014, renseignées et géoréférencées dans la base DVF, permettent désormais à des non spécialistes d'explorer, de quantifier et de spatialiser les structures des marchés de l'immobilier à différentes échelles. Toutefois, les bases DVF mises à disposition par les pouvoirs publics en accès libre et sous licence ouverte ne sont pas des données "prêtes à l'emploi" et diffèrent de celles accessibles aux ayants droit qui ne sont d'ailleurs pas plus faciles à utiliser (Casanova *et al.*, 2019). L'utilisation des bases DVF en accès libre à des fins d'analyse nécessite un temps de découverte et de compréhension qu'il ne faut pas sous-estimer.
- 4 C'est dans cette perspective que s'inscrit cet article, dont le but est de proposer une démarche documentée et reproductible à destination des géographes, des aménagistes, des urbanistes ou des data-journalistes pour appréhender spatialement les marchés immobiliers résidentiels à partir des données DVF accessibles en *open data*. Pour ce faire, nous développons à titre illustratif les exemples de la région Bretagne et de la métropole rennaise. L'ambition n'est pas ici de présenter un modèle ni une monographie des marchés bretons, mais plutôt de fournir des pistes méthodologiques issues de la statistique, de l'analyse spatiale et des techniques cartographiques pour explorer et représenter à différentes échelles géographiques des dynamiques de prix et de volume de ventes d'appartements et de maisons. L'objectif est de : (1) présenter une méthode de préparation des données DVF (apurement, filtrage, restructuration, enrichissement) ; (2) analyser à plusieurs échelles ces données à l'aide de différentes méthodes (agrégation, lissage, classification) ; et (3) restituer (carto)graphiquement ces données au moyen de plusieurs types de rendus (graphiques, cartes choroplèthes, cartes catégorisées, cartes lissées, carroyages et cartogrammes).
- 5 La première partie expose le contexte et les principales caractéristiques des données décrivant les transactions immobilières (sources, types de données, accessibilité) et les évolutions qu'elles connaissent (centralisation, ouverture). Toujours dans cette première partie, les données DVF sont également abordées selon leurs modalités de production, de structuration et de mise à disposition. La deuxième partie détaille la phase de préparation des données DVF en *open data*, préalable à tout traitement. La troisième partie présente plusieurs méthodes de représentations (carto)graphiques pour révéler et donner à voir les dynamiques et la structure spatiale du marché immobilier résidentiel à l'échelle de la région Bretagne. Enfin, la dernière partie propose une série d'analyses et de représentations reproductibles sur des périmètres plus restreints et à des échelons plus fins, à travers les exemples de la métropole rennaise et de la ville de Rennes.
- 6 Toute la démarche méthodologique présentée dans le présent article repose sur l'environnement R et s'inscrit dans une logique de reproductibilité et plus largement de

science ouverte. Pour cette raison, nous fournissons l'ensemble des scripts (documentés) que nous avons mis au point pour chacune des étapes. L'intérêt est à la fois de permettre une réutilisation totale de la chaîne de traitements et surtout de pouvoir modifier les scripts en fonction des besoins de chacun. Cette question de l'ouverture des méthodes et de reproductibilité de la recherche est ici envisagée selon trois perspectives (Goodman *et al.*, 2016) :

- Répétabilité de la démarche : reproduire les analyses présentées ici dans les mêmes conditions (utile à des fins de compréhension et d'évaluation) ;
  - Réplicabilité de la démarche : relancer les analyses avec de nouveaux critères (pour adapter la démarche en modifiant par exemple les périmètres d'analyses, les variables analysées, les seuils de prix, les types de biens, etc.) ;
  - Reproductibilité de la démarche : remobiliser les approches présentées ici pour les adapter à d'autres terrains d'étude, comme une autre région ou une ville en particulier.
- 7 L'ensemble des étapes de préparation, d'analyse et de représentation (carto)graphique des données DVF en *open data* sont traçables et reproductibles via quatre scripts R (en *markdown*) accessibles en ligne via une page [Github](#).
- Un [premier script](#) centralise les étapes de préparation des données DVF (apurement, filtrage et restructuration).
  - Un [deuxième script](#) présente la production et la représentation graphique d'indicateurs génériques sur le marché immobilier résidentiel.
  - Un [troisième script](#) propose tous les traitements et les cartes à l'échelle régionale (agrégations spatiales, CAH, lissage spatial et cartogramme lissé).
  - Un [quatrième script](#) rassemble les analyses et les représentations cartographiques réalisées sur la métropole rennaise (agrégations spatiales, CAH, lissage spatial).

## Les données DVF en *open data* : contexte et caractéristiques

### Appréhender les marchés immobiliers par les données de transactions

- 8 Les données relatives aux marchés fonciers et immobiliers sont multidimensionnelles, nombreuses et présentent une grande diversité, tant au niveau de leurs formes, de leurs producteurs que de leurs périmètres (Coulondre, Juillard, 2021). On dénombre en effet une vingtaine de bases produites et mises à jour par des acteurs publics et privés, lesquelles se concurrencent et se complètent à la fois (Driant *et al.*, 2005 ; Coulondre, 2018 ; Juillard, 2019).
- 9 L'accès à ces données reste souvent difficile, le type d'information disponible disparate et le niveau d'exhaustivité hétérogène (Le Corre, 2019). De plus, cette abondance d'informations est trompeuse, dans la mesure où toutes ces données dispersées ne permettent pas la mise en place d'analyses systématisées et comparables entre territoires. Malgré une forte attente des professionnels en données fines, fiables et pérennes sur le foncier et l'immobilier, cet écosystème informationnel segmenté participe à rendre opaques les marchés des ventes et surtout celui des loyers (Boulay *et al.*, 2012 ; Levasseur, 2013) et conduit à une connaissance fragmentée et partielle de ces derniers (Simon, 2009 ; Basciani-Funestre, Darley, 2013).

- 10 Parmi toutes les données existantes sur les ventes immobilières et foncières, les plus connues et les plus utilisées par les chercheurs sont les bases notariales. Depuis les années 1990, les notaires compilent les informations des actes de vente (déclaration du vendeur et de l'acquéreur sur la nature et le prix du bien) au sein de deux bases de données dont ils commercialisent l'accès par l'intermédiaire de structures professionnelles privées. Il s'agit de la base BIEN pour la région Île-de-France (produite depuis 1994 par la Chambre des notaires de Paris) et de la base PERVAL pour le reste du territoire (produite depuis 1990 par les Notaires de France). Les usages de ces deux bases immobilières notariales se voient aujourd'hui questionnés avec la démarche DVF, initiée par la DGFIP, que nous présentons ci-après.

## Les données DVF comme nouvelle base de référence des transactions immobilières et foncières

- 11 En réponse aux besoins de mise en œuvre de politiques territoriales adaptées à la réalité des marchés fonciers et immobiliers, l'État a inscrit à l'agenda politique les prémices de DVF en 2006 à travers la loi ENL<sup>1</sup>, puis mis en ligne en 2011 les données pour les ayants-droits et en 2019 pour le grand public. Cette centralisation de données sur les transactions immobilières produites par l'administration fiscale et cette mise à disposition en *open data* renvoie à la fois à une démarche, un service et une base de données (GnDVF, 2017). Ce projet s'inscrit dans une dynamique de transformation numérique de l'État français dans laquelle la modernisation du traitement des données foncières constitue un enjeu important (Masclet de Barbarin, 2019). On trouve d'ailleurs aujourd'hui d'autres jeux de données sur le foncier librement accessibles, comme le plan cadastral informatisé (PCI), la base des permis de construire (Sitadel) ou les fichiers des locaux et des parcelles des personnes morales.
- 12 Les données DVF sont présentées comme nécessaires à la bonne lecture du fonctionnement des marchés et à l'exercice d'évaluation des politiques publiques (Lagleize, 2019). Elles sont également mises en avant comme un accélérateur de la transparence des marchés par la mise à disposition d'informations sur les caractéristiques et les prix des biens à toutes les parties prenantes (vendeurs, acquéreurs et intermédiaires). Toutefois, la relation entre l'accès à l'information sur les prix de l'immobilier, la transparence et les effets modérateurs sur les prix n'est pas mécanique (Couloudre, Juillard, 2021). De plus, certains auteurs (Casanova *et al.*, 2019) soulignent que la mise à disposition de ces informations produites par l'administration fiscale renvoie aussi à un agenda néolibéral de l'État où les données sont destinées à une utilisation par des acteurs privés qui peuvent en tirer des sources de profit<sup>2</sup>.
- 13 L'un des principaux intérêts de DVF est de fournir un référentiel commun sur un périmètre géographique *quasi national*<sup>3</sup> qui permet de comparer les territoires entre eux, d'en comprendre les spécificités et de produire des études et des indicateurs selon des méthodes homogènes. La base DVF recense en effet toutes les mutations à titre onéreux ayant fait l'objet d'un acte notarié<sup>4</sup>. Pour constituer cette base, la DGFIP compile à la fois des informations renseignées par le service de la publicité foncière dans la base FIDJI (Fichier Informatisé des Données Juridiques Immobilières) et des éléments qui proviennent de la base de données MAJIC (Mise à Jour des Informations Cadastreales) (Casanova *et al.*, 2017 ; Cerema, 2020).

- 14 Même si les données issues des bases PERVAL/BIEN et DVF proviennent de la même source, le dispositif de renseignement n'est pas identique (GnDVF, 2017). Comme l'observent Casanova *et al.* (2017), ces deux bases renvoient à deux chaînes de production de données distinctes et à des objectifs différents qui déterminent la structure et le contenu de chacune des bases. En premier lieu, la base DVF, exhaustive par nature, couvre des parts du marché immobilier plus étendues que les bases des notaires (AUDAP, 2013 ; Casanova *et al.*, 2017 ; GnDVF, 2017). La DGFIP veille en effet, pour des raisons fiscales, à centraliser chacune des mutations. Toutefois, les données DVF sont bien moins riches en informations que les bases des notaires, limitant ainsi la portée analytique de leurs traitements. On retrouve en effet au sein des bases BIEN/PERVAL un éventail d'informations supplémentaires qui concernent les caractéristiques des biens (époque de construction, distinction entre neuf et ancien, etc.), les vendeurs et les acquéreurs (âge, catégorie socioprofessionnelle, modalités de financement, commune d'origine, etc.).
- 15 De plus, en raison des objectifs visés par les producteurs de ces bases et des cultures métiers associées, la structuration comme les modalités de remplissage des champs diffèrent entre les données DVF et les bases BIEN et PERVAL. Les bases des notaires sont conçues pour l'étude des marchés et proposent des informations retravaillées, à la différence des données DVF qui renvoient à des informations "brutes" (Casanova *et al.*, 2017 ; Coulondre, 2018). Par exemple, les bases notariales comportent une ligne par bien (surface, prix, acquéreur, vendeur, etc.) afin de faciliter les analyses, à la différence des données DVF. Les données DVF sont en ce sens plus complexes à traiter et nécessitent des pré-traitements spécifiques.
- 16 Enfin, la dernière particularité des données DVF par rapport aux bases des notaires réside dans leur mise à disposition en *open data* depuis avril 2019<sup>5</sup> dans une version non équivalente et appauvrie de celle de la DGFIP (Casanova Enault *et al.*, 2019). La puissance publique assure la publication en ligne de ces ressources par l'intermédiaire de différentes bases, et plusieurs canaux de diffusion.
- Une première base, intitulée "Demandes de valeurs foncières", mise à disposition par la DGFIP, rassemble les transactions foncières sous forme de données brutes<sup>6</sup>.
  - Une deuxième base, intitulée "DVF géolocalisées", retravaillée et géolocalisée par Etalab<sup>7</sup> sous une forme enrichie, permet un accès aux données par commune ou département afin de faciliter leur prise en main<sup>8</sup>. C'est cette base qui est utilisée dans l'article.
  - Enfin, une troisième base, intitulée "DV3F", est constituée et diffusée par le Cerema depuis 2011. Destinée aux acteurs publics, cette base est structurée suivant une logique différente et associe des données foncières complémentaires permettant des analyses plus poussées<sup>9</sup>. Cette base est disponible en deux versions, l'une intitulée "DVF+" et l'autre "DV3F" depuis 2014. La version DVF+, qui contient moins d'informations, est disponible en *open data* sur le site Data Foncier.

Tableau 1 : Avantages et limites des données DVF en *open data*

Avantages	Limites
-----------	---------



<ul style="list-style-type: none"> <li>- Gratuites et accessibles en <i>open data</i></li> <li>- Disponibles sur l'essentiel du territoire national</li> <li>- Mises à jour deux fois par an</li> <li>- Historicité des données depuis 2014</li> <li>- Réputées exhaustives</li> <li>- Globalement fiables</li> <li>- Géolocalisées à la parcelle</li> </ul>	<ul style="list-style-type: none"> <li>- Transactions non disponibles pour 4 départements</li> <li>- Complexes à exploiter</li> <li>- Structuration des transactions multivalentes</li> <li>- Pas d'identifiant de transaction</li> <li>- Difficilement intelligibles pour les non experts</li> <li>- Présence d'erreurs de saisie</li> <li>- Pas d'information sur les vendeurs et les acquéreurs (sauf dans la version DV3F)</li> </ul>
--	---

Mericskay, Demoraes, 2022

- 17 L'ouverture des données DVF est souvent mise en avant comme un "succès", si on s'en tient aux grands nombres de téléchargements et de consultations de la carte interactive mise en place par Etalab. Cela étant, dans les faits, la mise à disposition de ces informations n'a au final donné lieu qu'à assez peu d'exploitations (Juillard, 2020). Cette faible appropriation s'observe également sur le plan académique. On constate en effet que les données DVF demeurent encore peu utilisées par les chercheurs. Seuls Casanova Enault *et al.* (2019) ont exploré les données DVF en *open data* pour en proposer une réelle analyse critique. Le tableau 2 présente un panorama non exhaustif de recherches et d'études ayant exploré ou exploité des données DVF.

Tableau 2 : Panorama de travaux ayant utilisés des données DVF

Thématique	Zone d'étude	Référence
Croissance démographique et marché immobilier	Région Languedoc Roussillon	Gidrol, Canonéro, 2016
Analyse comparative des bases DVF et PERVAL	Aire urbaine d'Avignon	Casanova <i>et al.</i> , 2017
Analyse des profils des propriétaires du littoral	Littoral provençal et azuréen	Casanova Enault, 2017
Analyse des marchés immobiliers locaux	Bordeaux Métropole	Coulondre, 2017
Mobilités résidentielles des ménages modestes	Nantes Métropole	Chesnel <i>et al.</i> , 2017
Mobilités résidentielles et parc locatif	Région Nouvelle Aquitaine	Béoutis, Joubert, 2018
Mobilités résidentielles des couples	Maine-et-Loire	Chaillot, Loiseau, 2019
Analyse des prix moyens des transactions immobilières	Région Bourgogne Franche Comté	Pilarski, Zemis, 2019
Comparaison des données DVF de la base d'origine de la DGFIP avec les versions <i>open data</i>	Département du Vaucluse	Casanova Enault <i>et al.</i> , 2019
Analyse des marchés immobiliers locaux	Nantes Métropole	Coulondre, 2019

Analyse des marchés résidentiels aux abords des futures gares	Région Île-de-France	Coulondre, 2020
Analyse de l'évolution du marché résidentiel	Aire urbaine de Nantes	Mericskay <i>et al.</i> , 2020
Comparaison des données DVF avec des données issues de sites d'annonces immobilières	Région Provence-Alpes-Côte d'Azur	Boulay <i>et al.</i> , 2021
Analyse des prix des logements	Région Hauts-de-France	Eblé <i>et al.</i> , 2021

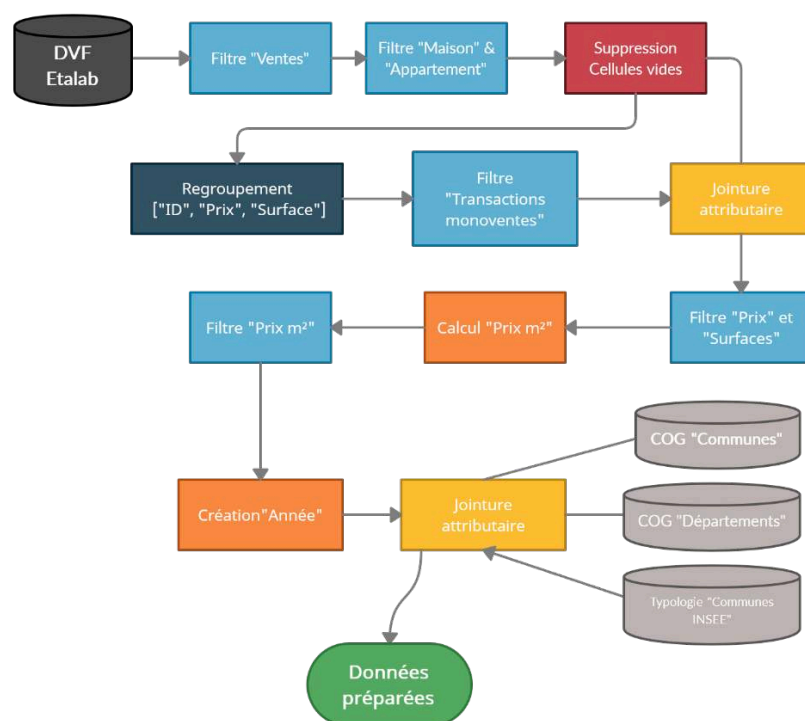
Mericskay, Demoraes, 2022

- 18 Il est important de souligner qu'au-delà des bases des notaires et des données DVF, le paysage informationnel relatif aux marchés immobiliers (vente comme location) se renouvelle, notamment avec la multiplication de plateformes spécialisées dans les petites annonces (comme Le Bon Coin) et les annonces immobilières (comme Se Loger, PAP ou Meilleurs Agents), dont certaines données peuvent être récupérées (par *web scraping* notamment) et analysées (Coulondre, Juillard, 2021 ; Chapelle, Eyméoud, 2022). Plusieurs travaux de recherche récents commencent d'ailleurs à explorer ces nouvelles sources pour à la fois cerner l'intérêt de ces nouvelles données et analyser certains marchés immobiliers (Boulay *et al.*, 2021 ; Le Goix *et al.*, 2021).

## La préparation des données DVF en *open data*, une étape complexe et essentielle

- 19 Dans le cadre de cet article, ce sont les données "DVF géolocalisées", enrichies par Etalab, qui sont utilisées. La base initiale pour les quatre départements bretons (Côtes-d'Armor, Finistère, Ille-et-Vilaine et Morbihan) recense entre le 1<sup>er</sup> janvier 2014 et le 31 décembre 2019, 446 818 mutations distinctes renvoyant à 1 043 052 lignes et 40 colonnes dans la base de données. Si de prime abord, ces données revêtent de multiples potentialités, le volume et la richesse d'informations présentes engendre une série de difficultés d'exploitation.
- 20 Les bases DVF en *open data* renvoient en effet à des données "brutes", correspondant à des fichiers lourds et difficilement exploitables par des personnes ne maîtrisant pas les outils et les langages de préparation et de traitement de données comme R, Python ou le SQL (Juillard, 2020). De plus, la version DVF en libre accès est moins complète que celle mise à disposition des ayants-droits. L'absence de certaines informations sur les transactions et la présence d'erreurs de saisie ou d'incohérences implique un temps non négligeable de compréhension préalable de leur contenu ainsi qu'un important travail d'apurement et de préparation en fonction des objectifs visés (Casanova Enault *et al.*, 2019 ; Coulondre, 2019). Malgré le travail effectué en amont par Etalab pour faciliter la prise en main des bases DVF, il demeure indispensable de nettoyer, filtrer, restructurer et enrichir ces données pour constituer une base permettant la mise en place d'analyses (figure 1).

Figure 1 : Chaîne de préparation des données DVF Etalab



Mericskay, Demoraes, 2022

- 21 L'intégralité de la chaîne de traitement des données DVF présentée dans le présent article a été mise en œuvre dans l'environnement R. Ce choix technique se justifie par plusieurs raisons. Il s'agit tout d'abord d'une solution logicielle libre et gratuite, en phase avec la démarche de science ouverte dans laquelle nous nous inscrivons. Cette solution offre en outre un grand nombre d'extensions (*packages*<sup>10</sup>) qui permettent de manipuler et d'analyser des données, notamment volumineuses et complexes (Wickham *et al.*, 2019), de réaliser des graphiques (Wickham, 2016), d'effectuer des opérations d'analyse spatiale (Lovelace *et al.*, 2019 ; Bivand, 2020) et de produire des cartes (Giraud, Lambert, 2016). De plus, la possibilité d'exécuter la totalité des traitements au sein d'un même environnement permet d'éviter les ruptures logicielles et facilite *in fine* la reproductibilité de la démarche à d'autres terrains où à d'autres échelles (Giraud, Lambert 2017).
- 22 La première étape a consisté à garder uniquement les mutations sur lesquelles nous souhaitions centrer notre analyse, à savoir les ventes de maisons et d'appartements<sup>11</sup>. De fait, ont été exclues les mutations relatives aux échanges, aux expropriations et aux adjudications ainsi que les ventes de terrains, de locaux industriels et commerciaux. Par ailleurs, un contrôle de qualité des données a été effectué et a permis d'exclure des mutations mal renseignées<sup>12</sup>.
- 23 Le principal verrou dans l'emploi des données DVF en *open data* est lié au fait qu'une même mutation peut concerner plusieurs biens et donc être associée à plusieurs lignes dans la base (figure 2). Cette particularité s'explique par l'emboîtement de plusieurs niveaux au sein de la base (Casanova *et al.*, 2017 ; Casanova Enault *et al.*, 2019). Les transactions correspondent à la totalité du ou des biens fonciers/immobiliers achetés le même jour par un même acquéreur. Ces transactions peuvent être décomposées en

plusieurs mutations (étudiées dans cet article), c'est-à-dire des parties de la transaction. Enfin, les mutations peuvent à leur tour être composées de plusieurs biens (appartement, maison, terrain, etc.).

Figure 2 : Synthèse des types de mutations modélisées dans la base DVF d'Etalab

**Mutation monovente** : une mutation, un bien unique, un prix et une surface

ID_mutation	Date	Nature	Type	Prix	Superficie
2019-168833	10/01/2019	Vente	Appartement	135000	35

**Mutation multivente « cohérente »** : une mutation, plusieurs biens différents, un prix et une surface mais une redondance du bien au sein de la base (si un ou plusieurs terrains associés par exemple)

ID_mutation	Date	Nature	Type	Prix	Superficie
2019-168400	10/01/2019	Vente	Maison	250000	150
2019-168400	10/01/2019	Vente	Terrain	250000	150

**Mutation multivente « homogène »** : une mutation, plusieurs biens différents, avec plusieurs surfaces mais un prix unique (pour le lot)

ID_mutation	Date	Nature	Type	Prix	Superficie
2019-168835	10/01/2019	Vente	Appartement	450000	45
2019-168835	10/01/2019	Vente	Appartement	450000	50
2019-168835	10/01/2019	Vente	Appartement	450000	75

**Mutation multivente « incohérente »** : une mutation, plusieurs biens différents de différents segments, avec plusieurs surfaces mais un prix unique (pour le lot)

ID_mutation	Date	Nature	Type	Prix	Superficie
2019-233321	10/01/2019	Vente	Appartement	450000	45
2019-233321	10/01/2019	Vente	Maison	450000	50
2019-233321	10/01/2019	Vente	Terrain	450000	75

Mericskay, Demoraes, 2022, inspiré de Casanova *et al.*, 2017

- 24 Dans le cas de la Bretagne, on retrouve ainsi 340 801 lignes au sein du jeu de données qui renvoient à 266 082 mutations. Par ailleurs, il n'existe aucun champ de référence d'identification des transactions, à la différence de la base originale produite par la DGFIP. On trouve en revanche un champ d'identifiants des mutations construit à partir de la date et de la disposition. Cette modélisation particulière des données DVF librement accessibles nécessite la mise en place de traitements préalables pour la construction et l'exploitation d'indicateurs fiables (Casanova *et al.*, 2017 ; Coulondre, 2018 ; Coulondre, 2019 ; Casanova Enault *et al.*, 2019).
- 25 La solution adoptée ici a consisté à ne travailler que sur les ventes n'impliquant qu'un seul bien (mutations monoventes) et les biens qu'il semble logique de considérer ensemble, comme une maison et un terrain, ou un appartement et une dépendance (mutations multiventes cohérentes). Nous avons donc écarté les mutations multiventes "homogènes" (lot de biens d'un même type) et les mutations multiventes "incohérentes" (lot de biens de plusieurs segments) pour lesquelles il est impossible d'obtenir les caractéristiques individuelles comme le prix de chaque bien sans traitements spécifiques<sup>13</sup>. Les deux sous-populations de mutations analysées dans cet article représentent 95 % du total des mutations<sup>14</sup>, soit un ordre de grandeur similaire à la moyenne nationale (GnDVF, 2017)<sup>15</sup>. Malgré cette limite, cette méthode demeure la plus simple à mettre en place et convient aux objectifs exploratoires et méthodologiques visés dans cet article.
- 26 La base DVF contient également de nombreuses erreurs de saisie (prix ou surface incohérents) ou des mutations "déguisées"<sup>16</sup>. Afin de retirer de notre échantillon ce

type de valeurs que nous qualifierons "d'aberrantes", en ce sens qu'elles ne reflètent pas la réalité des prix du marché, il semblait nécessaire, pour ne pas fausser les analyses, de supprimer certaines mutations. Nous avons ainsi fixé des seuils minimaux et maximaux en ce qui concerne les prix, les prix au m<sup>2</sup> et les surfaces des mutations simples. Il est possible de définir ces seuils de manière relative (en choisissant des percentiles dans la distribution statistique des variables concernées) ou de manière absolue en appliquant des valeurs minimales et maximales de référence. Chacune de ces méthodes présente des avantages et des inconvénients, les percentiles pouvant, selon l'échelle d'analyse, aboutir à considérer comme aberrants des sous-marchés très locaux, les valeurs absolues pouvant varier d'un analyste à l'autre. Dans le cadre de cette étude, nous avons opté pour une hybridation de ces deux méthodes<sup>17</sup>. Dans une optique de reproductibilité à d'autres territoires, la définition de seuils est personnalisable<sup>18</sup>. Elle nécessite d'explorer au préalable le jeu de données et de les adapter aux objectifs visés et aux caractéristiques des marchés étudiés, dans la mesure où ces choix relèvent d'une décision cruciale et potentiellement impactante sur la suite des analyses.

- 27 Une fois le travail de filtrage, d'apurement et de restructuration effectué, la dernière étape a consisté à enrichir les données pour élargir les possibilités d'analyse par la suite. Nous avons ajouté à chacune des mutations des descripteurs géographiques (nom du département, de la commune et de l'EPCI (Établissement Public de Coopération Intercommunale) et le type de commune selon une typologie de l'Insee), afin de mieux les resituer dans leur contexte territorial et de faciliter certaines opérations d'agréations statistiques à différents échelons d'analyse.
- 28 Au final, le jeu de données retravaillé recense pour les quatre départements bretons, du 1 janvier 2014 au 31 décembre 2019, 247 646 mutations décrites en lignes. Chacune de ces mutations est renseignée par 14 variables qui peuvent être nominales (type, commune, EPCI, département, typologie de l'Insee), quantitatives (prix, surface, prix au m<sup>2</sup>) ou spatiales (coordonnées géographiques) comme l'illustre la figure 3.

Figure 3 : Structure des données DVF de la base Bretagne, à l'issue de la phase de préparation

id	date	annee	type	prix	surface	prixm2	codecommune	commune	Dep	EPCI	Typo_INSEE	latitude	longitude
2014-185931	02/01/2014	2014	Maison	185000	82	2256	22171	Plaintel	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Couronne périurbaine	48.41036	-2.822322
2014-185933	09/01/2014	2014	Maison	140000	83	1687	22106	Langueux	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.50325	-2.718138
2014-185936	08/01/2014	2014	Appar...	80000	78	1026	22278	Saint-Brieuc	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.50548	-2.773664
2014-185937	06/01/2014	2014	Maison	165000	109	1514	22081	Hillion	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Couronne périurbaine	48.48662	-2.638281
2014-185939	15/01/2014	2014	Maison	83000	52	1596	22215	Ploufragan	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.48867	-2.816707
2014-185941	09/01/2014	2014	Maison	175000	91	1923	22106	Langueux	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.49156	-2.710499
2014-185946	03/01/2014	2014	Maison	140000	63	2222	22215	Ploufragan	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.49069	-2.804397
2014-185948	18/01/2014	2014	Maison	280000	108	2593	22210	Ploubazlanec	Côtes-d'Armor	CA Guingamp-Paimpol Agglomération de l'Ar...	Pôle urbain	48.80104	-3.014141
2014-185950	20/01/2014	2014	Appar...	102000	44	2318	22055	Binic-Étable...	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Couronne périurbaine	48.62955	-2.830727
2014-185954	08/01/2014	2014	Maison	103000	105	981	22360	Trégueux	Côtes-d'Armor	CA Saint-Brieuc-Armor Agglomération	Pôle urbain	48.47386	-2.754416

Mericskay, Demoraes, 2022

- 29 Il convient de souligner que nous avons choisi de conserver les mutations de 6 années (2014-2019) dans la même série statistique. Ce choix s'appuie sur le fait qu'il n'existe pas sur la période de variations interannuelles importantes en ce qui concerne le nombre de mutations ou le prix (prochaine section). De plus, traiter les données sur l'ensemble de la période permet, d'une part, de gommer les mutations exceptionnelles (encore présentes malgré l'apurement) qui pourraient fausser localement l'interprétation des dynamiques et, d'autre part, d'augmenter les effectifs de mutations et *in fine* de garantir la robustesse statistique des résultats. Toutefois, comme tout choix méthodologique, ce dernier n'est pas parfait dans la mesure où il peut exister des marchés locaux où les variations interannuelles de prix sont fortes.

## Les données DVF en graphiques et en cartes pour saisir les grandes tendances de l'immobilier à l'échelle régionale

### Appréhender les dynamiques du marché de l'immobilier résidentiel breton par les graphiques

- 30 L'une des premières pistes d'analyse des dynamiques de volume et de prix de l'immobilier résidentiel réside dans une exploration des mutations à l'aide de méthodes de statistiques descriptives et de représentations graphiques. Le calcul d'indicateurs simples, comme le nombre de mutations, les moyennes ou les médianes des prix ou des surfaces, permet d'obtenir un premier portrait synthétique du marché immobilier résidentiel breton exploré ici (tableau 3).

Tableau 3 : Récapitulatif des principaux indicateurs du marché immobilier résidentiel breton sur la période 2014-2019 (source : DVF)

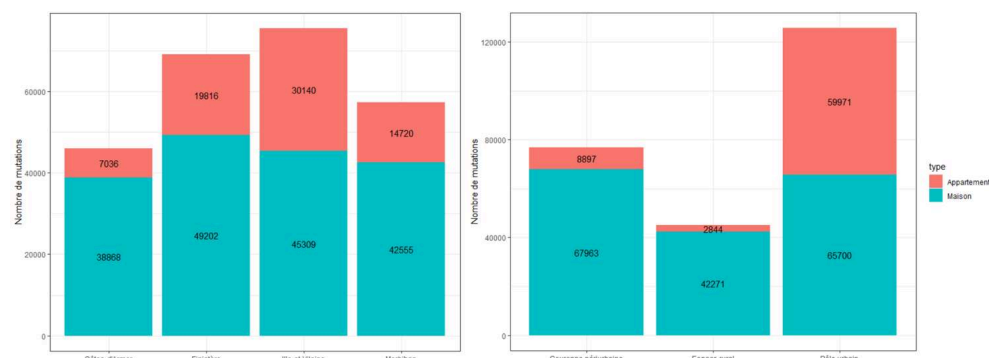
	Mutations monoventes	Part	Prix médian	Prix moyen	Surface moyenne	Prix/m <sup>2</sup> médian	Prix/m <sup>2</sup> moyen
<b>Appartements</b>	71 712	29 %	100 000 €	116 575 €	58 m <sup>2</sup>	1 941 €/m <sup>2</sup>	2 093 €/m <sup>2</sup>
<b>Maisons</b>	175 934	71 %	152 000 €	173 367 €	101 m <sup>2</sup>	1 573 €/m <sup>2</sup>	1 716 €/m <sup>2</sup>
<b>Total</b>	247 646	100 %	135 000 €	156 920 €	89 m <sup>2</sup>	1 663 €/m <sup>2</sup>	1 825 €/m <sup>2</sup>

Mericskay, Demoraes, 2022

- 31 En complément des tableaux, l'exploration de l'ensemble des mutations peut, dans un premier temps, s'appuyer sur la représentation de leurs caractéristiques via des graphiques. La visualisation de données sous des formes variées (histogrammes, diagramme en barres, boîtes à moustache, courbes, etc.) se prête particulièrement bien à la description et à l'analyse des données DVF pour comprendre des tendances et donner à voir des liens entre les différentes informations de la base par la distribution, la comparaison ou les relations (Yau, 2013). L'idée est de pouvoir croiser et représenter graphiquement plusieurs variables, à la fois quantitatives (nombre de mutation, prix, prix au m<sup>2</sup>) et qualitatives (type de bien, type de commune, département). Ces graphiques ont la particularité de ne pas être dépendants des caractéristiques du jeu de données. Le script permettant de les produire peut ainsi être adapté pour d'autres territoires d'étude dans une optique de reproductibilité.
- 32 En premier lieu, les diagrammes en barres empilées permettent de combiner plusieurs types de variables pour comparer et mettre en évidence des logiques de volume de vente. Si l'on s'intéresse aux parts des mutations par département ou par type de commune par exemple, il ressort des logiques territoriales bien marquées. Par exemple,

les départements du Finistère et de l'Ille-et-Vilaine regroupent à eux deux 60 % des mutations, dont 70 % des ventes d'appartements (figure 4). Quand on regarde la répartition des volumes des mutations monoventes selon le type des communes (typologie Insee urbain/rural), des divergences entre communes apparaissent aussi (figure 4).

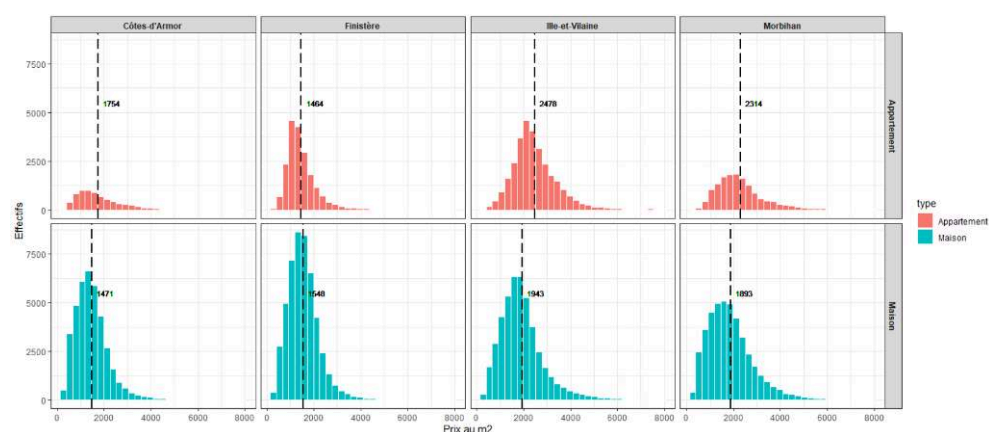
Figure 4 : Répartition des mutations en fonction du type de bien, du département et du type de commune (Insee) sur la période 2014-2019



Mericskay, Demoraes, 2022. Source : DVF

- 33 À la différence des diagrammes en barres utilisés pour les données catégorielles, les histogrammes permettent l'exploration de variables continues comme les prix, les prix au m<sup>2</sup> ou les surfaces des biens vendus. L'utilisation d'histogrammes est très utile en phase d'exploration des données. De plus, la production de planches comportant plusieurs graphiques (*small multiples*)<sup>19</sup> permet d'explorer un grand nombre de dimensions, d'identifier rapidement des tendances ou des spécificités et de les comparer facilement (Van den Elzen, Van Wijk, 2013). La figure 5 offre une vision d'ensemble qui permet d'identifier facilement les différences de volume et les logiques de prix entre les départements et en fonction des biens vendus.

Figure 5 : Histogrammes de distribution des prix moyens au m<sup>2</sup> pour les maisons et les appartements dans les 4 départements bretons sur la période 2014-2019

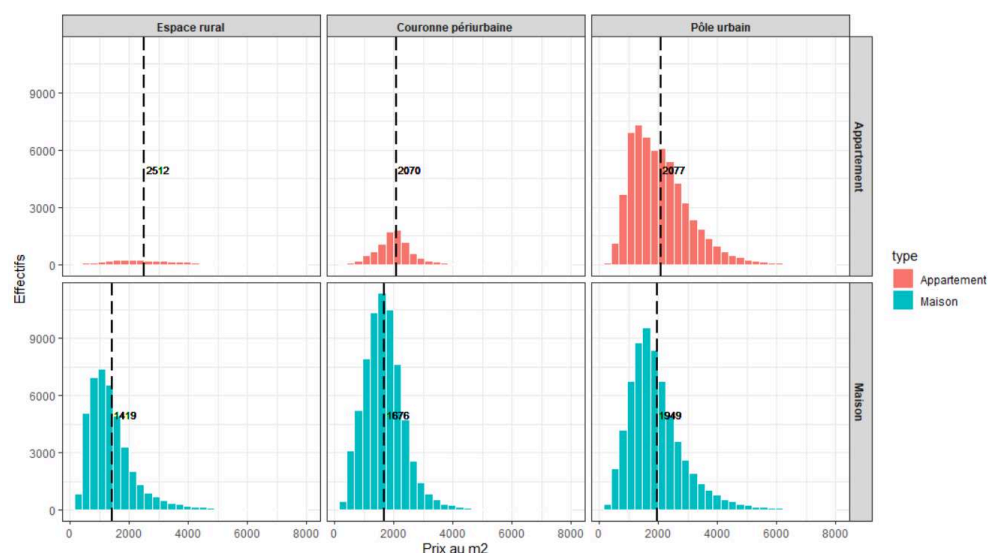


Mericskay, Demoraes, 2022. Source : DVF

- 34 Envisager les volumes des mutations et les prix au m<sup>2</sup> selon le type de bien et le type de commune suivant la typologie urbain/rural de l'Insee permet une autre lecture des dynamiques en présence par une entrée thématique (figure 6). Cette distinction offre

une lecture synthétique pour identifier rapidement les spécialisations des espaces urbains, périurbains et ruraux.

Figure 6 : Histogrammes de distribution des prix moyens au m<sup>2</sup> pour les maisons et les appartements en fonction de la typologie urbain/rural de l'INSEE sur la période 2014-2019

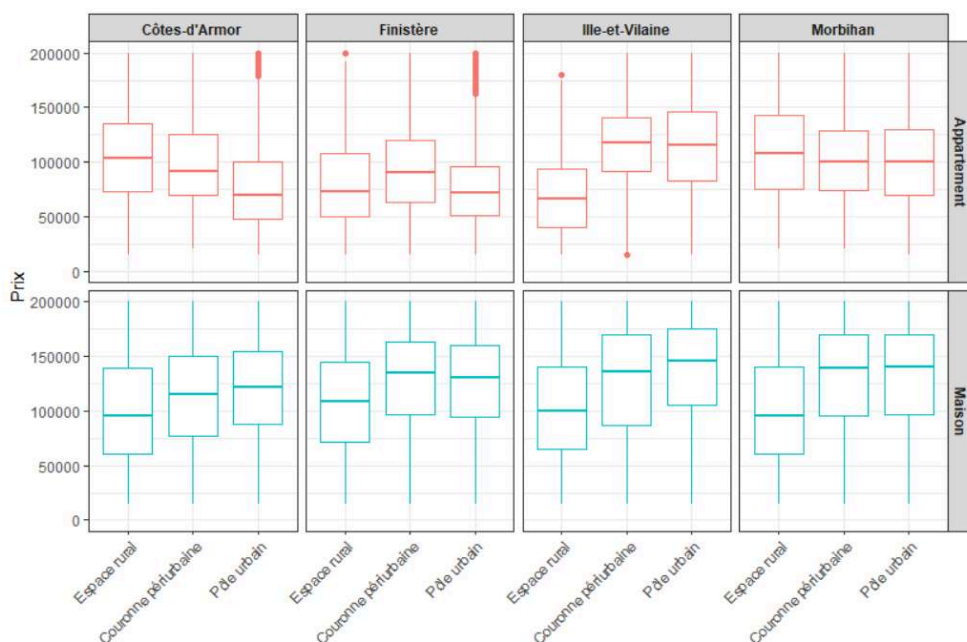


Mericskay, Demoraes, 2022. Source : DVF

- 35 De son côté, la représentation en "boîte et moustaches" fournit une représentation visuelle simple de la médiane, des quartiles et de l'étendue de variables comme le prix ou la surface des biens. Leur affichage sous forme de planche offre en complément la possibilité de comparer un même caractère dans des populations de tailles différentes. La figure 7 permet par exemple d'identifier rapidement certaines logiques de prix récurrentes et d'autres plus originales par la combinaison de trois dimensions (type de bien, type de commune et département).



Figure 7 : Boîtes à moustaches des prix en fonction du département, du type de bien et du type de commune (Insee) sur la période 2014-2019



Mericskay, Demoraes, 2022. Source : DVF

- 36 La disponibilité des mutations depuis le 1er janvier 2014 permet également une exploration temporelle des dynamiques des marchés de l'immobilier étudiés. Le recours à des diagrammes en barres empilées offre une lecture temporelle comparative des évolutions en présence. La figure 8 met par exemple en lumière des logiques assez similaires avec une augmentation générale des volumes de ventes entre 2014 et 2017 dans les 4 départements. Concernant les évolutions des prix (graphique présent dans le script), on constate une tendance similaire d'augmentation jusqu'en 2017 puis de décroissance dans les couronnes périurbaines et pôles urbains, et dans une moindre mesure dans les espaces ruraux.

Figure 8 : Évolution du nombre de ventes de maisons et d'appartements par année, par département et par type de commune (Insee) sur la période 2014-2019



Mericskay, Demoraes, 2022. Source : DVF

## Appréhender les dynamiques du marché de l'immobilier résidentiel par l'analyse spatiale et les cartes

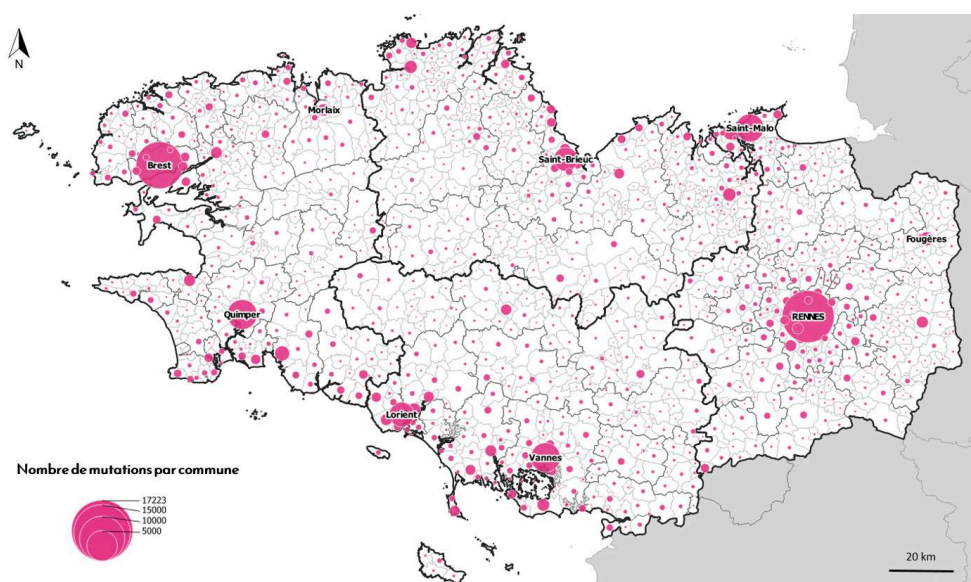
- 37 En complément à l'exploration des données DVF au moyen de graphiques, l'entrée géographique constitue une approche complémentaire pour mettre en lumière des dynamiques et des structures spatiales de marchés immobiliers résidentiels (Boulay, 2011 ; Le Goix, 2013 ; Brémond, Marie, 2016 ; Le Corre, 2019 ; Le Goix *et al.*, 2019). Comme le souligne Palm (1978), la première démarche du géographe s'intéressant aux dynamiques spatiales de l'immobilier consiste à s'interroger sur les échelles géographiques les plus pertinentes. Il est en effet possible d'analyser les données DVF à plusieurs échelons territoriaux (région, département, EPCI, commune) et aussi dans d'autres types de mailles. Ces dernières peuvent être associées à des découpages statistiques (IRIS) ou construits de toutes pièces aux fins de l'analyse (carroyage). En fonction des objectifs visés et des échelles d'observation, les mutations peuvent être envisagées de manière individuelle ou agrégées dans plusieurs niveaux d'unités spatiales (Boulay, 2011), dont découlent des visions différentes et complémentaires du message transmis.

### L'agrégation spatiale comme étape préalable à la représentation cartographique

- 38 La première opération par laquelle on peut commencer consiste à dénombrer le nombre de mutations par commune (figure 9). Ce type de représentation cartographique, qui met en jeu une variation de taille, permet de représenter de manière lisible le volume des transactions dans l'espace d'étude à travers une maille administrative familière. Le marché de l'immobilier résidentiel breton, en termes de volumes de ventes, apparaît largement polarisé par les espaces les plus peuplés et les

plus dynamiques de Bretagne, c'est-à-dire près du littoral et à proximité des grands pôles urbains. En complément, et comme nous le développons dans le script, il est possible de décliner ce type de cartes en différenciant les maisons des appartements pour faire affiner les explorations et faire ressortir des spécificités locales.

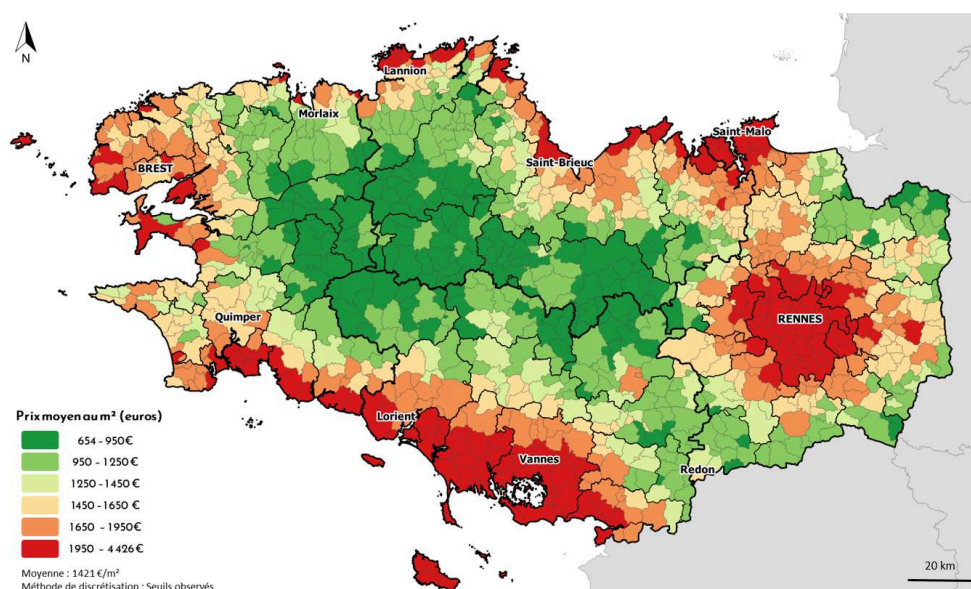
Figure 9 : Nombre de mutations immobilières résidentielles par commune en Bretagne sur la période 2014-2019



Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 39 De son côté, l'exploration spatiale des prix au m<sup>2</sup> agrégés à différents échelons offre une autre lecture des dynamiques territoriales du prix de l'immobilier résidentiel. La figure 10, qui présente le prix moyen au m<sup>2</sup> des ventes d'appartements et de maisons agrégés par commune, permet de mettre en lumière de grandes tendances régionales (effet littoral de gradient, logique centre-périphérie autour de la capitale régionale). Représenter un indicateur "grand public" sous forme de carte choroplèthe est un moyen facile de visualiser la façon dont la variable de prix au m<sup>2</sup> varie dans une zone géographique<sup>20</sup>. De plus, sur l'ensemble d'une région, le découpage communal permet de visualiser une information précise à un échelon connu des lecteurs et donne la possibilité de croisement avec des données externes comme celles de l'Insee (non développé ici).

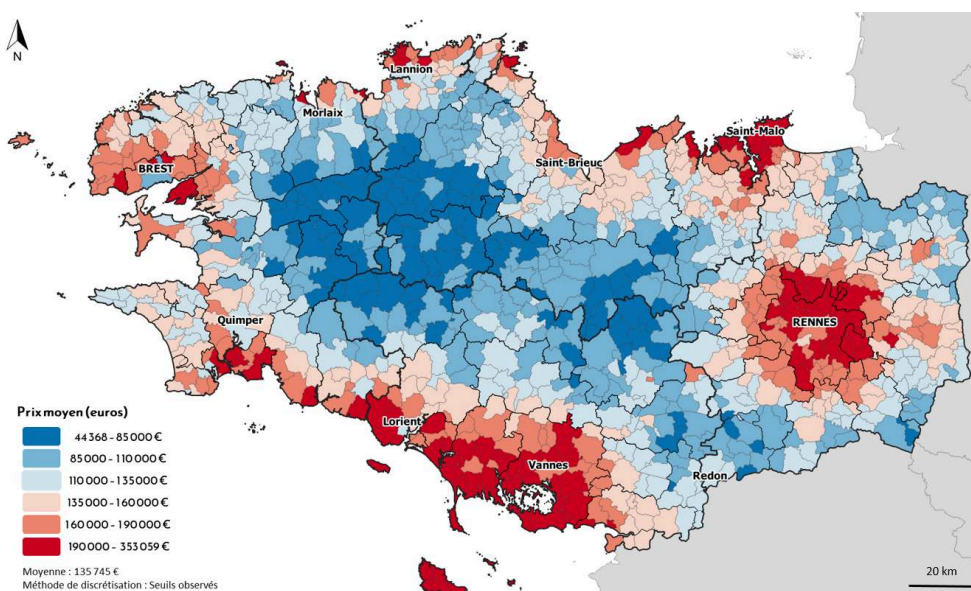
Figure 10 : Prix moyen au m<sup>2</sup> de l'immobilier résidentiel par commune en Bretagne sur la période 2014-2019



Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 40 Afin d'illustrer l'accessibilité financière (abordabilité) au marché "vécu" par les ménages qui disposent d'un budget rapporté au prix des biens (et non aux prix au m<sup>2</sup>), nous avons représenté sur la figure 11 les prix nominaux moyens (des ventes de maisons et d'appartements) par commune. Cette carte choroplèthe<sup>21</sup>, complémentaire à celle des prix moyens au m<sup>2</sup>, fait ressortir des logiques spatiales assez similaires à cette dernière, mais permet un regard plus fin sur la localisation des communes les plus chères sur le territoire breton, comme autour du golfe du Morbihan, à proximité de Saint-Malo ou en périphérie de Rennes (au nord notamment).

Figure 11 : Prix nominal moyen de l'immobilier résidentiel par commune en Bretagne sur la période 2014-2019

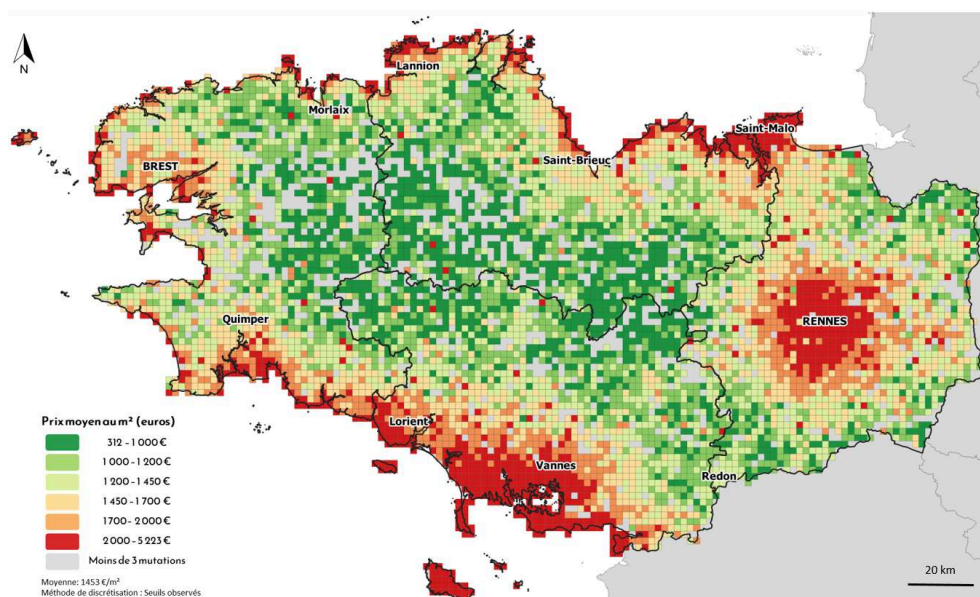


Mericskay, Demoraes, 2022. Sources : DVF/IGN



- 41 Si ces deux cartes en aplats de couleurs restituant une information à l'échelon communal (figures 10 et 11) offrent une lecture aisée des prix nominaux et des prix au m<sup>2</sup> de l'immobilier en Bretagne, dans le sens où il est facile de comparer les communes entre elles, il convient néanmoins de souligner que les prix des biens vendus peuvent répondre à des logiques indépendantes d'un quelconque maillage politico-administratif. Comme le souligne Le Corre (2019) dans sa thèse sur les prix de l'immobilier en Île-de-France : *"il ne s'agit pas de rejeter d'emblée des limites administratives mais de proposer une réelle réflexion sur les biais de l'agrégation spatiale"*.
- 42 Pour s'affranchir de l'arbitraire des découpages territoriaux et limiter l'effet du MAUP - *Modifiable Area Units Problem* - (Openshaw, Taylor, 1979)<sup>22</sup>, une première piste méthodologique consiste à recourir à un carroyage vectoriel comme référentiel spatial d'agrégation (Lajoie, 1992 ; Pumain, Saint-Julien, 2010). Souvent utilisé dans des travaux sur les marchés immobiliers (Boulay, 2011 ; Bermond, Marie, 2016 ; Le Corre, 2021), le partitionnement de l'espace d'étude à l'aide d'un maillage régulier présente plusieurs avantages, comme celui d'offrir une certaine neutralité et d'éviter les biais liés à la taille et à la forme des unités administratives (Oliveau, 2020). De plus, cette méthode donne la possibilité de définir des granularités fines et personnalisables (en termes de forme et de dimension des mailles) en fonction de la zone d'étude et des objectifs visés.
- 43 Dans notre cas, nous avons retenu des mailles carrées, de 2 km<sup>23</sup> de côté, pour obtenir une image infra-communale des prix moyens au m<sup>2</sup> sur l'ensemble de la Bretagne (figure 12). Cette forme d'agrégation dans des unités spatiales homogènes fournit un niveau géographique d'observation plus détaillé des grandes tendances déjà mises en avant dans la figure 11. On peut par exemple bien appréhender l'effet du littoral sur les prix moyens au m<sup>2</sup>, en distinguant un gradient spatial assez classique. Cette forme de représentation met également en lumière des zones plus chères de la Bretagne intérieure, comme celles à proximité des villes moyennes telles que Loudéac, Pontivy ou Carhaix.

Figure 12 : Prix moyen au m<sup>2</sup> de l'immobilier résidentiel par maille de 2 km de côté en Bretagne sur la période 2014-2019

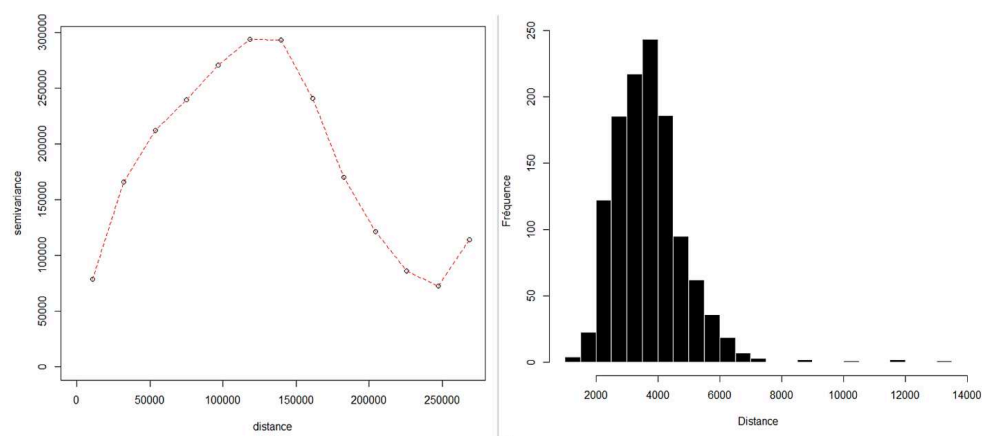


Mericskay, Demoraes, 2022. Sources : DVF/IGN

## Le lissage spatial, une technique éprouvée pour révéler des tendances et des gradients

- 44 Si l'agrégation des valeurs au sein d'un carroyage vectoriel fournit un premier niveau de généralisation offrant une lecture assez fine des variations de prix dans l'espace, la juxtaposition de carreaux voisins de petite taille et de couleurs parfois très différentes reste néanmoins difficile à assimiler sur une carte qui couvre l'ensemble d'une région. Pour pallier cette limite, nous avons utilisé une autre technique cartographique, celle du lissage spatial calculé à partir de la méthode des noyaux (Silvermann, 1998). Ces cartes lissées sont également appelées "cartes de chaleur"<sup>24</sup>.
- 45 Le lissage spatial présente l'avantage de schématiser de manière continue la distribution spatiale d'une variable (Genebes *et al.*, 2018 ; Aschan-Leygonie *et al.*, 2019) et de rendre lisibles aussi bien les spécificités locales d'un phénomène que les grandes tendances, en s'affranchissant des limites administratives (Grasland, 1999 ; Lambert, Zanin, 2016). La puissance évocatrice des cartes choroplèthes qui viennent d'être présentées peut ainsi être augmentée par les méthodes de lissage, en conférant aux cartes ainsi obtenues une capacité de révélation de structures spatiales sous-jacentes (Banos, 2001).
- 46 Comme le démontrent Guérois et Le Goix (2009), Le Goix *et al.* (2016), Pilarski et Zemis (2019) ou Le Corre (2019), les cartes lissées sont particulièrement efficaces pour donner à voir les grandes tendances de prix de l'immobilier. Suivant l'échelle d'analyse, la taille de la zone d'étude et le degré de généralisation souhaité, le lissage spatial peut porter directement sur le prix au m<sup>2</sup> des mutations localisées à la parcelle ou sur un prix moyen au m<sup>2</sup> par unité spatiale, comme la commune. Dans le cadre de l'approche régionale, pour illustrer les possibilités de variation du degré de généralisation, nous avons choisi de calculer deux lissages ; le premier portant sur le prix moyen au m<sup>2</sup> préalablement agrégé par commune, le deuxième sur les mutations elles-mêmes. Il est à noter que le choix du rayon et la fonction de lissage constituent des paramètres importants, dans la mesure où le résultat est très dépendant de ces derniers (Zaninetti, 2005).
- 47 Pour la première carte (figure 14), nous avons tout d'abord adopté comme rayon de lissage la portée spatiale du phénomène qui se lit sur le semi-variogramme empirique (figure 13) et qui est un paramètre couramment utilisé pour calibrer la fonction de lissage (Zaninetti, 2005). Dans le cas présent, nous avons dans un premier temps retenu une portée de 13 km, qui correspond à la distance au-delà de laquelle les prix moyens au m<sup>2</sup> par commune ne sont plus corrélés entre eux. Ceci étant, compte tenu de l'étendue de la Bretagne (environ 140 km du Nord au Sud), ce rayon a abouti à un premier résultat beaucoup trop lissé. Nous avons donc utilisé un autre paramètre fréquemment employé (Zaninetti, 2005), qui correspond à la distance la plus fréquente entre unités spatiales adjacentes (figure 13). Au final, nous avons retenu un rayon de 5 km, qui découle directement de la taille des unités spatiales (ici les communes) contenant les valeurs à lisser.

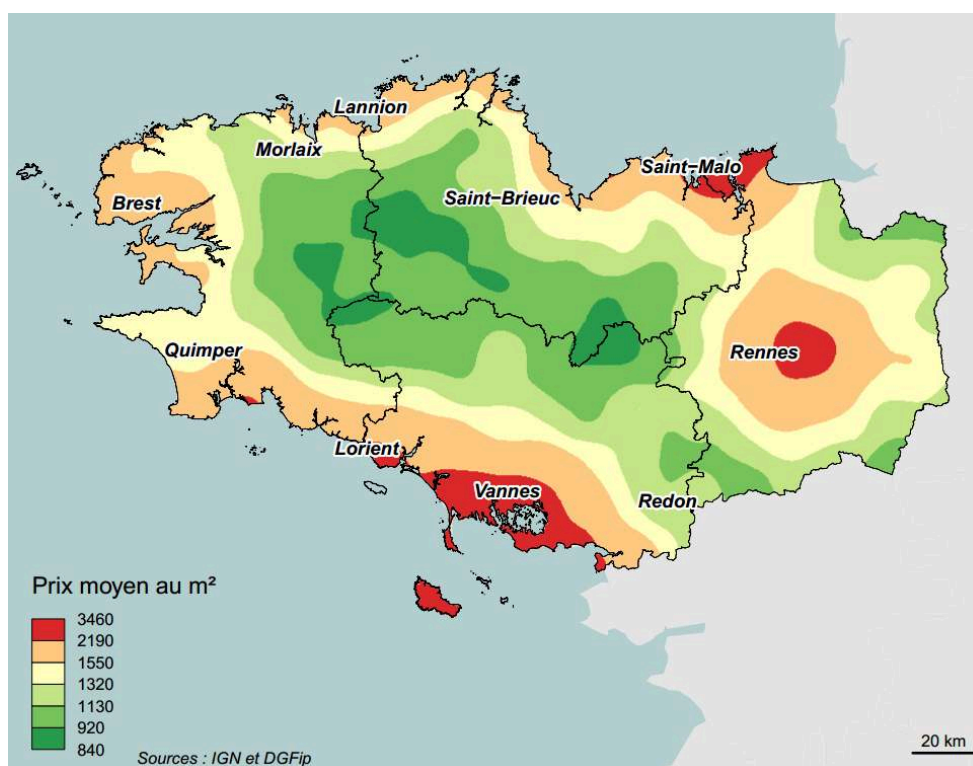
Figure 13 : Semi-variogramme empirique du prix au m<sup>2</sup> de l'immobilier résidentiel par commune en Bretagne sur la période 2014-2019 en fonction de la distance (à gauche) et histogramme des distances entre communes bretonnes voisines, de centroïde à centroïde (à droite)



Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 48 Pour ce qui est de la fonction de lissage, nous avons opté pour une fonction gaussienne qui permet de généraliser l'information sans pour autant estomper complètement les variabilités locales qui peuvent exister entre des communes limitrophes (Grasland, 1999 ; Zaninetti, 2005). Le résultat de ce premier lissage des prix moyens agrégés par communes livre une vision très schématique des prix au plan régional (figure 14)<sup>25</sup>. On retrouve les grandes logiques déjà mises en avant dans les cartes précédentes mais de façon plus généralisée, comme l'effet littoral ou le gradient centre-périphérie de la métropole rennaise. Ce type de carte constitue un outil de communication efficace à destination des médias ou du grand public, dans la mesure où il restitue un message synthétique et compréhensible de façon plus expressive que des cartes choroplèthes, dans lesquelles le nombre élevé d'unités spatiales peut complexifier la lecture.

Figure 14 : Surface lissée des prix moyens au m<sup>2</sup> de l'immobilier résidentiel agrégés par commune en Bretagne sur la période 2014-2019

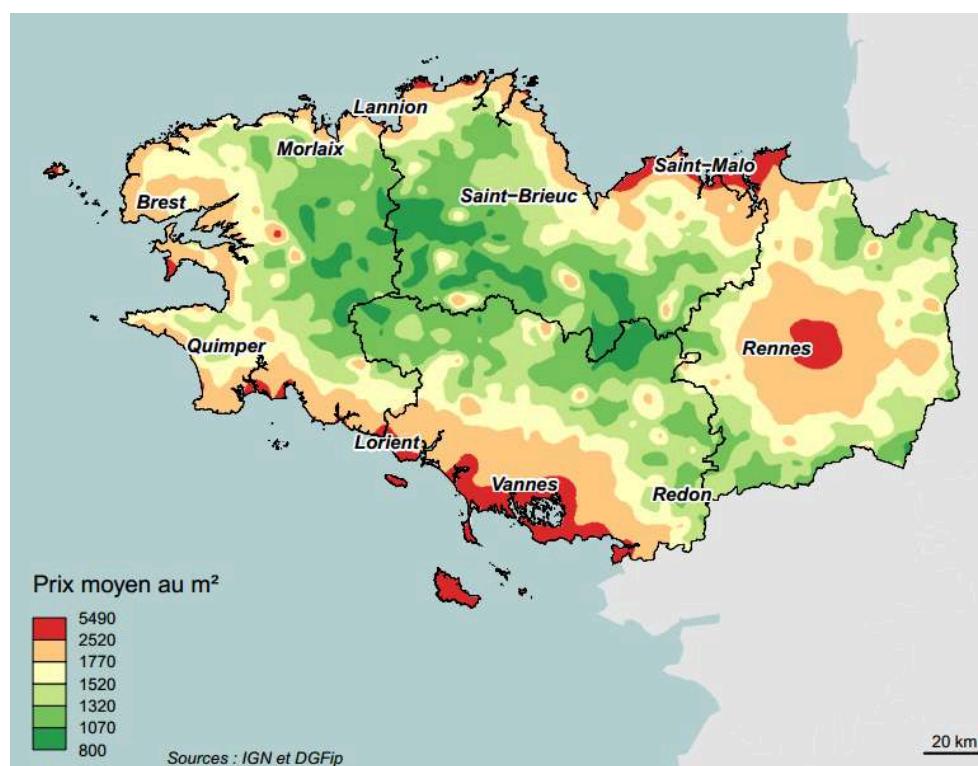


Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 49 La figure 15 représente les prix moyens au m<sup>2</sup> obtenus à partir du lissage appliqué directement sur les 247 646 mutations (sans agrégation préalable dans un découpage spatial). Le lissage est calculé au plus près de la valeur de chaque mutation<sup>26</sup>. Le rayon de lissage adopté ici (2 km)<sup>27</sup> est plus petit que dans l'exemple précédent, afin de mieux faire ressortir les variabilités locales sans trop obérer la généralisation d'ensemble. C'est aussi la fonction gaussienne qui a été utilisée. Le résultat apporte plus de nuances. Si l'on retrouve les principaux gradients, on distingue également des poches avec des prix localement plus élevés qui correspondent aux villes moyennes (Loudéac, Pontivy, Lamballe, Fougères, Vitry, etc.).



Figure 15 : Surface lissée des prix moyens au m<sup>2</sup> calculés à partir des 247 646 mutations immobilières résidentielles en Bretagne sur la période 2014-2019

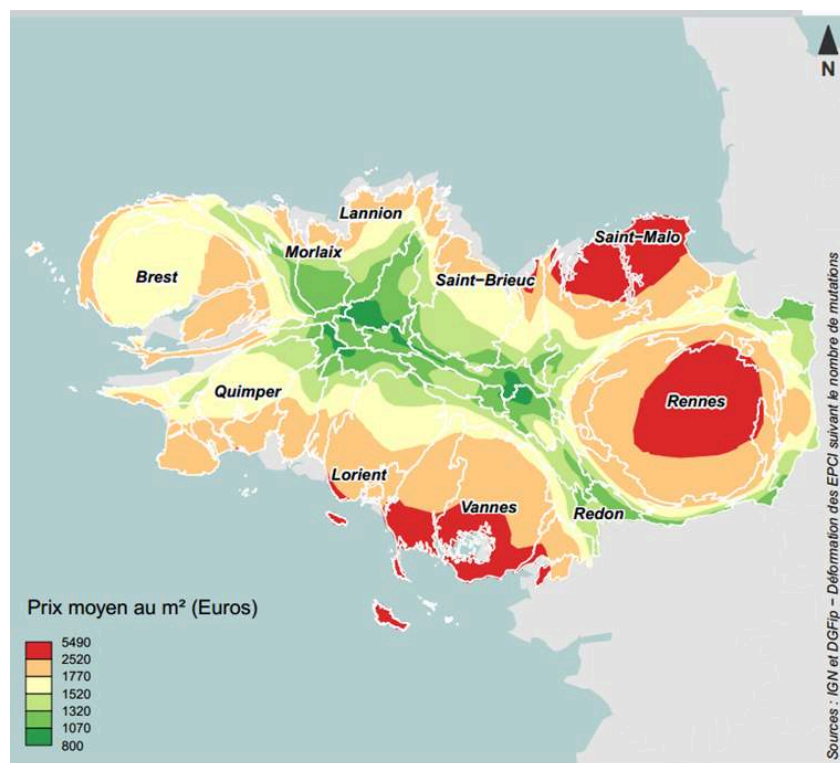


Mericksay, Demoraes, 2022. Sources : DVF/IGN

### Le cartogramme lissé, ou l'intérêt de ne pas perdre de vue les volumes en jeu

- 50 Sur les cartes de prix, lissées ou non, les volumes des mutations n'apparaissent pas. Pour restituer les volumes qui attestent du dynamisme du marché de l'immobilier, tout en gardant l'information simplifiée des prix au m<sup>2</sup>, nous avons combiné à titre exploratoire deux techniques cartographiques déjà employées conjointement dans le travail de Demoraes *et al.*, 2021. Nous avons dans un premier temps construit un cartogramme qui déforme les unités spatiales en fonction du nombre de mutations. Nous avons dans un deuxième temps lissé les valeurs des mutations. La surface lissée obtenue a ensuite été déformée avec les mêmes paramètres que ceux ayant servi à construire le cartogramme, de telle sorte que la surface lissée se superpose *in fine* parfaitement au cartogramme. Pour calculer le cartogramme, nous avons utilisé le tout nouveau package *cartogramR* (Cornillon, Demoraes, à paraître)<sup>28</sup>. Nous avons tout d'abord déformé les 1 270 communes bretonnes en fonction du nombre de mutations qu'elles ont enregistrées, mais le résultat était assez attendu et peu lisible : les communes de Rennes et de Brest occupaient une place prépondérante écrasant le reste. Pour observer ce qui se passe également en dehors de ces deux principales métropoles, nous avons donc changé d'échelon d'analyse et agrégé le nombre de mutations par EPCI (figure 16)<sup>29</sup>.

Figure 16 : Cartogramme des prix moyens au m<sup>2</sup> lissés, calculés à partir des 247 646 mutations immobilières résidentielles en Bretagne sur la période 2014-2019



Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 51 Ce type de carte permet de restituer visuellement deux variables au sein d'une même représentation : le nombre de mutations par EPCI à travers la déformation de ces derniers et les prix moyens au m<sup>2</sup> sous forme de surfaces lissées. Il est intéressant de noter que le nombre de mutations n'est pas forcément corrélé au prix à cet échelon-là. Les volumes élevés échangés correspondant à une demande forte aussi bien à Rennes qu'à Brest sont associés dans le premier cas à des prix élevés, dans le second à des prix proches de la moyenne régionale.

### La cartographie des sous marchés régionaux par CAH (classification ascendante hiérarchique)

- 52 Pour compléter cette exploration du marché résidentiel à l'échelle régionale, l'identification et la délimitation géographique de sous marchés regroupant des communes présentant des caractéristiques similaires permet de saisir plus en profondeur les dynamiques immobilières et leur inscription spatiale. De nombreux travaux de recherche (Guérois, Le Goix, 2009 ; Boulay, 2011 ; Casanova Enault, 2017 ; Le Brun, 2019 ; Le Corre, 2019) et de rapports opérationnels (Coulondre 2017, 2018, 2020, 2021) ont montré l'intérêt de recourir à certaines méthodes d'analyses statistiques (ACP, CAH) pour construire et cartographier des typologies de sous marchés ou de profils d'acquéreurs.
- 53 Nous proposons ici une segmentation du marché immobilier résidentiel breton à l'échelon communal, qui repose sur une méthode de classification ascendante hiérarchique (CAH)<sup>30</sup>. Cette technique de partitionnement statistique permet de révéler

les différences spatiales régionales en regroupant les 1 270 communes bretonnes en classes homogènes. La CAH permet d'affiner les explorations par l'analyse simultanée de plusieurs variables. Le choix des variables s'est ici porté uniquement sur des indicateurs présents dans la base DVF, à savoir le nombre de mutations, le prix nominal moyen et le prix moyen au m<sup>2</sup> des biens vendus par commune, afin de proposer une classification simple<sup>31</sup>.

- 54 La CAH mise en œuvre<sup>32</sup> sur les mutations monoventes réalisées entre 2014 et 2019 dans les communes bretonnes permet de distinguer sept sous marchés<sup>33</sup>, dont les caractéristiques sont détaillées dans le tableau 4.

Tableau 4 : Les sept sous marchés immobiliers des communes bretonnes sur la période 2014-2019

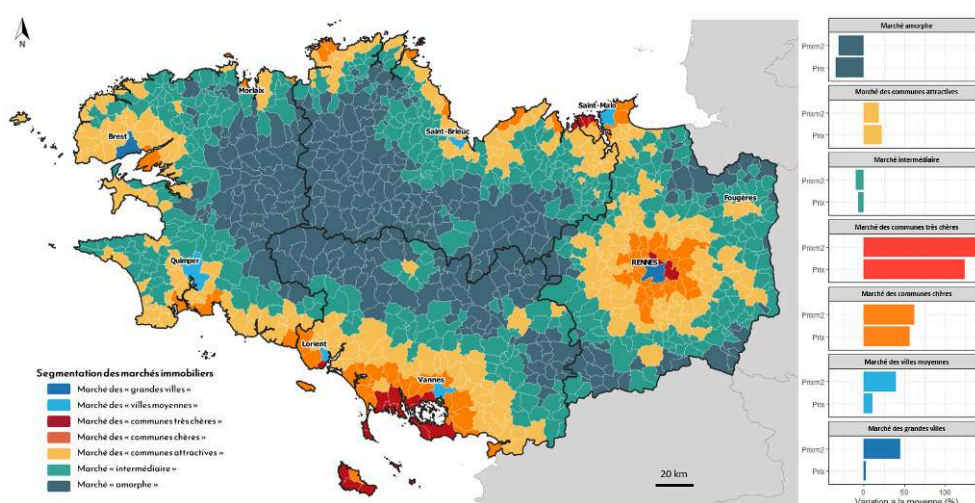
	Nombre de communes	Nombre moyen de mutations	Prix moyen	Prix moyen au m <sup>2</sup>
<b>Marché des grandes villes</b>	2	15 795	140 800 €	2087 €/m <sup>2</sup>
<b>Marché des villes moyennes</b>	5	5 417	151 400 €	2020 €/m <sup>2</sup>
<b>Marché des communes très chères</b>	28	427	307 200 €	3460 €/m <sup>2</sup>
<b>Marché des communes chères</b>	82	430	215 050 €	2339 €/m <sup>2</sup>
<b>Marché des communes attractives</b>	302	242	167 750 €	1700 €/m <sup>2</sup>
<b>Marché classique</b>	443	94	127 520 €	1300 €/m <sup>2</sup>
<b>Marché amorphe</b>	344	76	90 090 €	996 €/m <sup>2</sup>

Mericskay, Demoraes, 2022. Sources : DVF

- Un premier groupe très restreint, incluant Rennes et Brest, renvoie au marché résidentiel "des grandes villes", caractérisé par de très gros volumes de mutations, des prix des biens dans la moyenne et des prix au m<sup>2</sup> élevés.
- Le deuxième groupe, associé au sous marché des "villes moyennes", se caractérise par un assez grand volume de mutations, des prix nominaux dans la moyenne et des prix au m<sup>2</sup> élevés.
- Le troisième groupe est composé de 28 communes "très chères", avec des volumes de mutations assez importants et surtout des prix nominaux et des prix au m<sup>2</sup> très élevés.
- Le quatrième groupe est composé de 82 communes "chères", avec des volumes de mutations assez importants et des prix nominaux et des prix au m<sup>2</sup> élevés.
- Le cinquième groupe, avec plus de 280 communes, renvoie au sous marché des communes "attractives", qui se caractérisent par des volumes de mutations assez importants, des prix des biens et des prix au m<sup>2</sup> élevés.

- Le sixième groupe correspond au sous marché des communes "intermédiaires", qui réunit plus de 40 % des communes bretonnes, avec des prix et des volumes situés dans la moyenne régionale.
  - Le septième groupe, qui regroupe un tiers des communes de la région, renvoie au sous marché le moins dynamique, que l'on peut qualifier "d'amorphe", avec peu de mutations et des prix nominaux et au m<sup>2</sup> bien en-dessous de la moyenne régionale.
- 55 La représentation des 7 groupes identifiés par la CAH sous la forme d'une carte catégorisée révèle une forte structure spatiale des marchés immobiliers qui s'inscrit globalement dans les dynamiques démographiques régionales et fait ressortir certaines spécificités locales (figure 17). Ce type de représentation cartographique met ainsi en évidence une série de logiques spatiales, comme un effet centre-périphérie autour de Rennes, des logiques de gradients près des littoraux et le centre de la Bretagne caractérisé par un marché résidentiel "amorphe".

Figure 17 : Typologie des marchés immobiliers par commune en Bretagne et variation par rapport à la moyenne régionale pour chacun des 7 sous marchés<sup>34</sup> sur la période 2014-2019



Mericskay, Demoraes, 2022. Sources : DVF/IGN

## Appréhender les données DVF à des échelles locales : les exemples de la métropole et de la commune de Rennes

### Exploration du marché métropolitain

- 56 Cette dernière partie analyse, à titre illustratif là-encore, les marchés de l'immobilier résidentiel sur un périmètre plus restreint. Nous développons ici les exemples de Rennes Métropole et de la ville de Rennes. D'autres échelles d'analyse peuvent être envisagées pour étudier à des niveaux plus locaux les dynamiques et les structures spatiales du marché de l'immobilier résidentiel (aire urbaine fonctionnelle, aire urbaine, bassin de vie, etc.).
- 57 Sur l'ensemble de la métropole, on dénombre pour la période d'étude 31 722 mutations monoventes, dont les principales caractéristiques sont présentées dans le tableau 5.

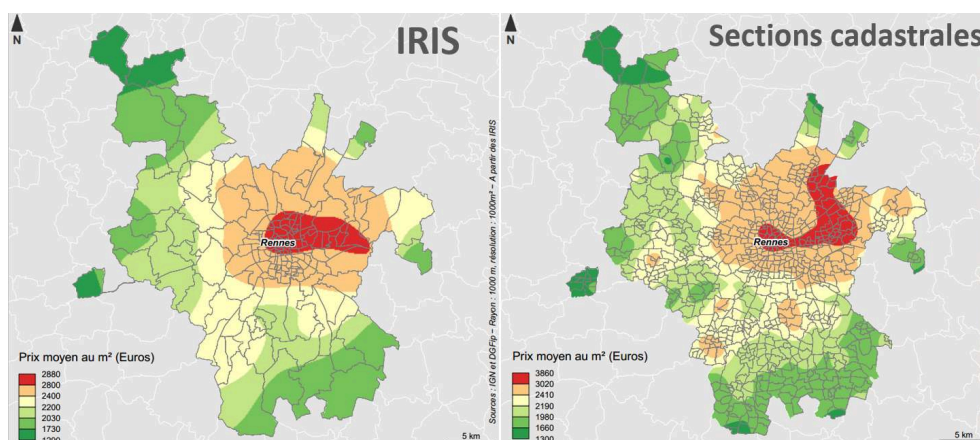
**Tableau 5 : Récapitulatif des principaux indicateurs du marché immobilier résidentiel de Rennes Métropole sur la période 2014-2019**

	Nombre	Part	Prix médian	Prix moyen	Surface moyenne	Prix/m <sup>2</sup> médian	Prix/m <sup>2</sup> moyen
<b>Appartements</b>	20 555	65 %	127 800 €	141 860 €	60 m <sup>2</sup>	2346 €/m <sup>2</sup>	2470 €/m <sup>2</sup>
<b>Maisons</b>	11 167	35 %	250 000 €	277 520 €	111 m <sup>2</sup>	2570 €/m <sup>2</sup>	2570 €/m <sup>2</sup>
<b>Total</b>	31 722	100 %	159 850 €	189 620 €	78 m <sup>2</sup>	2350 €/m <sup>2</sup>	2500 €/m <sup>2</sup>

Mericskay, Demoraes, 2022. Source : DVF

- 58 L'exploration spatiale des données DVF sur des périmètres plus restreints peut être envisagée selon différents référentiels d'analyse (communes, IRIS, sections cadastrales, maillage, etc.). Nous avons fait le choix de travailler sur deux découpages du territoire métropolitain, l'un d'origine statistique, l'autre fiscale. D'un côté les IRIS<sup>35</sup>, au nombre de 172, correspondent à un échelon pivot qui permet de croiser les données DVF avec d'autres variables de l'Insee notamment (revenu, âge, CSP, etc.). De l'autre, les sections cadastrales, au nombre de 835, constituent un autre échelon d'analyse pertinent car ce découpage possède un lien direct avec l'objet d'étude (les dynamiques immobilières)<sup>36</sup>.
- 59 Ces deux référentiels ont fait l'objet de plusieurs analyses présentées dans le script (aggrégations, dénombrement, CAH). Nous proposons de revenir plus en détail sur les traitements et les échelons d'analyse utilisés pour produire les cartes lissées. L'objectif est de montrer différents degrés de généralisation de l'information qu'il est possible d'obtenir.
- 60 Pour les trois cartes suivantes, afin de réduire l'effet de bord<sup>37</sup>, le prix au m<sup>2</sup> des biens vendus dans les communes limitrophes de Rennes Métropole a été pris en compte, dans la limite du rayon de lissage. C'est aussi la fonction de décroissance gaussienne que nous avons retenue pour les calculs. Par ailleurs, pour les trois cartes, les prix au m<sup>2</sup> ont été discrétisés en 6 classes<sup>38</sup> et la même double gradation harmonique de couleurs a été utilisée. Sur la figure 18, en tenant compte de la portée, de la taille des unités spatiales et de l'étendue de Rennes Métropole, nous avons opté pour un rayon de lissage de 1 km. Sur la figure 19, nous avons changé la focale, compte tenu de la finesse de l'agencement du semis de points en entrée (les mutations), et opté pour un rayon de 500 m pour être au plus près des valeurs initiales et ne pas estomper les variations qui peuvent exister localement.

Figure 18 : Carte lissée des prix au m<sup>2</sup> de l'immobilier résidentiel (IRIS et sections cadastrales) dans Rennes Métropole sur la période 2014-2019

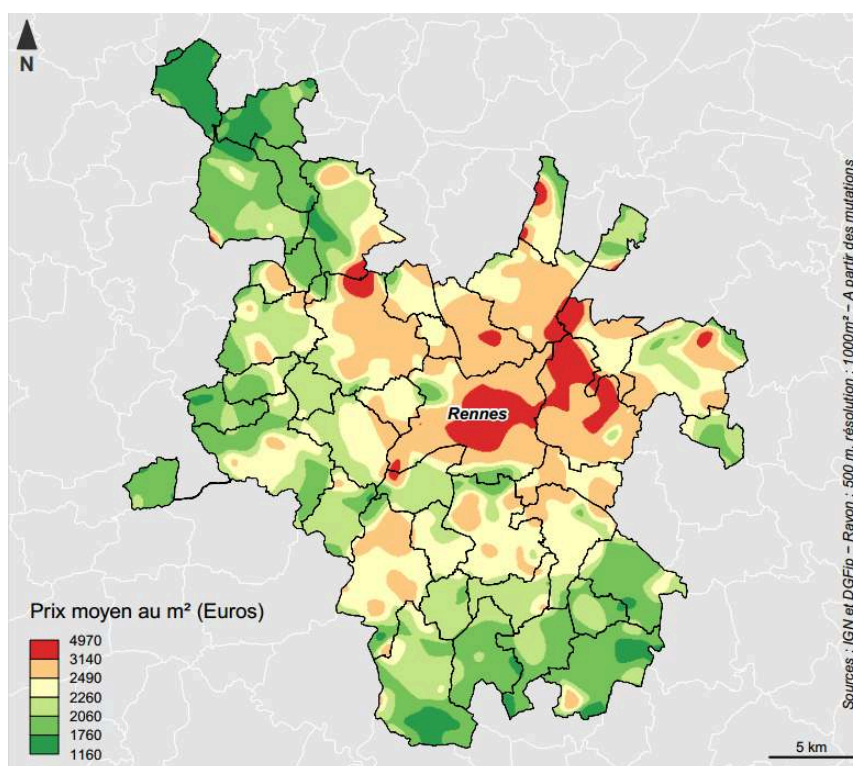


Mericskay, Demoraes, 2022. Sources : DVF/IGN

- 61 La figure 18 donne ainsi à voir les contrastes qui existent dans les prix de l'immobilier sur l'ensemble de Rennes Métropoles. Les prix au m<sup>2</sup> s'organisent suivant une logique centre-périphérie et vont décroissant au fur-et-à-mesure qu'on s'éloigne du centre. Ces deux cartes lissées illustrent bien l'incidence de l'échelon d'agrégation des données initiales sur le résultat cartographique. À partir du même jeu de données initial (les 31 722 mutations enregistrées sur la période), la carte de gauche, dont le lissage a été calculé à partir des prix au m<sup>2</sup> moyennés au préalable par IRIS, restitue une information très généralisée par rapport à la carte de droite, établie à partir des valeurs agrégées par section cadastrale. L'amplitude des prix est logiquement plus faible sur la carte des IRIS (davantage moyennée) que sur la carte des sections cadastrales, bien plus nombreuses.
- 62 De son côté, le lissage spatial calculé à partir des 31 722 mutations individuelles (figure 19) offre une lecture plus précise et davantage de nuances que les précédentes, notamment en dehors de la commune de Rennes. La figure 19 fait ainsi davantage ressortir les prix au m<sup>2</sup> plus élevés dans les centres villes de communes connexes à Rennes comme Betton, Cesson-Sévigné ou Saint Grégoire, mais également au sein des centres historiques de communes périphériques comme Acigné, Chartres-de-Bretagne ou Vern-sur-Seiche.



Figure 19 : Carte lissée des prix au m<sup>2</sup> calculés à partir des 31 722 mutations (maisons et appartements) dans Rennes Métropole sur la période 2014-2019



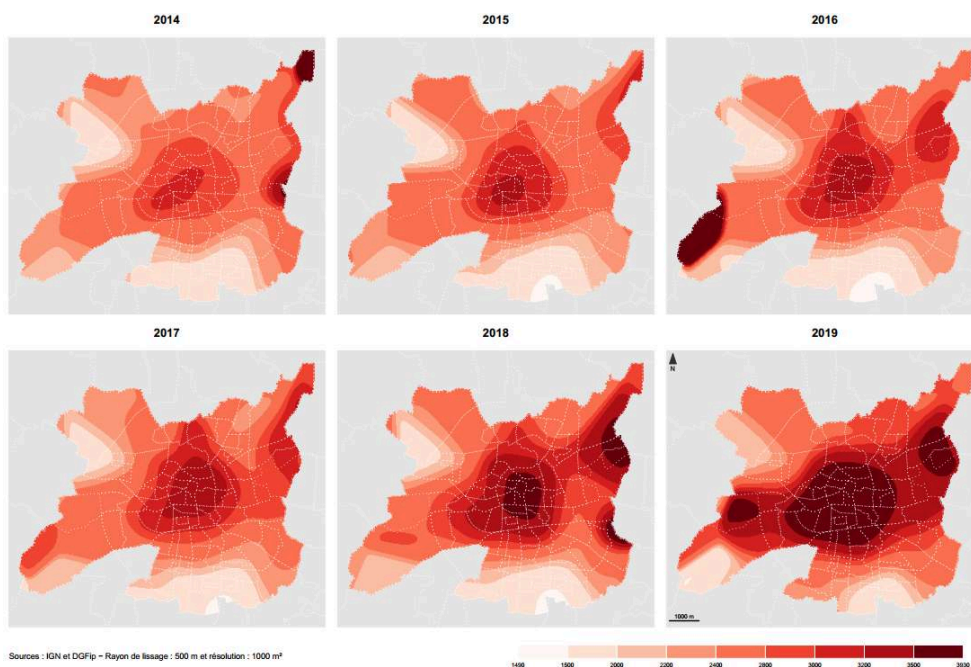
Mericksay, Demoraes, 2022. Sources : DVF/IGN

## Les données DVF pour retracer la trajectoire des prix dans le temps et dans l'espace : l'exemple de la commune de Rennes

- 63 Nous proposons dans cette dernière partie de montrer un autre potentiel des données DVF en libre accès. Nous explorons ici les trajectoires d'évolution des prix au m<sup>2</sup> des appartements dans la commune de Rennes entre 2014 et 2019. Si ces prix n'ont pas connu de grands changements sur l'ensemble de la Bretagne entre 2014 et 2019 (figure 9), ils ont en revanche beaucoup augmenté dans la métropole rennaise sur la même période. Cette évolution est d'ailleurs plus marquée pour les appartements, qui ont connu une hausse de plus de 326 €/m<sup>2</sup> (+13,8 %) en six ans au plan régional. Les appartements y concernent par ailleurs plus de 87 % des mutations et ont connu à Rennes une augmentation de 530 €/m<sup>2</sup> (+22 %) en six ans, hausse qui justifie de s'y intéresser plus en détail.
- 64 Pour donner à voir l'évolution temporelle de "la géographie des prix de la ville" (Guérois, Le Goix, 2009), nous avons produit une planche de six cartes lissées, une par année (figure 20)<sup>39</sup>, calculée à partir des mutations individuelles non agrégées des appartements. Nous avons utilisé pour les six cartes la même discrétisation (reposant sur des amplitudes égales<sup>40</sup>) d'une année à l'autre pour pouvoir les comparer et appréhender visuellement les changements de prix. La figure 20 met efficacement en lumière la diffusion en "tache d'huile" des valeurs immobilières les plus élevées, à partir d'un noyau central qui s'élargit au fil des années et qui s'étend également en

direction de l'est. Les cartes lissées, année par année, mettent aussi en exergue les disparités intra-urbaines qui s'accroissent sur la période.

Figure 20 : Cartes lissées des prix au m<sup>2</sup>, calculés à partir des mutations d'appartements à Rennes sur la période 2014-2019



Mericksay, Demoraes, 2022. Source : DVF

- 65 L'utilisation du lissage spatial peut toutefois poser certains problèmes, notamment en faisant émerger des structures artificielles, induites par la méthode elle-même et les données mobilisées (Banos, 2001). En ce sens, le contrôle des effets de bordure, surtout à une échelle fine, constitue un enjeu méthodologique important à prendre à compte dans ce type d'approche. Malgré la correction de l'effet de bord apportée ici, on observe certains comportements qui peuvent paraître erratiques sur les limites extérieures de la commune et qui découlent de mutations survenues dans des espaces très peu denses, où les ventes sont peu nombreuses et correspondent certaines années à des biens très onéreux.

## Perspectives

- 66 Nous l'avons vu dans cet article, à travers l'exemple de la Bretagne, les données DVF disponibles en *open data* représentent une source informationnelle à fort potentiel pour appréhender, objectiver et saisir à plusieurs échelles les dynamiques des marchés immobiliers résidentiels selon différents angles. Outre leur accès sous licence ouverte, elles constituent un socle de données normalisées, mises à jour deux fois par an et localisées à la parcelle. Leur disponibilité depuis le millésime 2014, avec une structure stable dans le temps sur la quasi-totalité du territoire national, permet aujourd'hui d'envisager la conception de protocoles d'exploitation standardisés pour en tirer le meilleur parti, pour le plus grand nombre de réutilisateurs potentiels et pour toute zone d'étude.



- 67 Cet article a été l'occasion de relever certains défis méthodologiques liés au volume et à la façon dont les données DVF en *open data* sont structurées. Comme nous l'avons vu, la question de la préparation des données est une étape cruciale. En terme analytique, le choix des périmètres d'étude (région, métropole et ville), des échelons d'analyse (EPCI, communes, IRIS, sections cadastrales) et les opérations d'agrégation sont centrales dans ce type de réflexion et méritent d'être davantage investiguées et questionnées. Les différentes explorations ont également permis de tester, de discuter et de repenser les modes de représentation pour trouver les manières les plus adéquates de donner à voir des logiques de prix ou de volumes de mutations.
- 68 Une des clefs de l'originalité et de l'efficacité des cartes ne se trouve donc pas uniquement dans la représentation des données mobilisées ici mais surtout dans la préparation et l'analyse de ces dernières, en amont, qui guident l'intention cartographique. Élaborer une carte de communication ne se limite pas uniquement à la "simple" mise en forme de données spatiales, mais renvoie aujourd'hui à des compétences plus larges. Cela confirme bien la forte dimension méthodologique intimement associée à la cartographie thématique. Cela est d'autant plus vrai face à l'instrumentation contemporaine, à l'image du langage R et de sa multitude de packages qui offrent de nombreuses possibilités en perpétuelle évolution tant pour la représentation cartographique que pour l'analyse spatiale.
- 69 Cet exercice de multireprésentation (carto)graphique reposant sur plusieurs méthodes illustre la panoplie de traitements et de rendus possibles tout comme leur complémentarité pour révéler les dynamiques immobilières à différentes échelles. Ainsi, en fonction du contexte d'utilisation, des objectifs et du public visé, il est possible de faire appel à un mode de représentation ou à un autre pour restituer sous la forme la plus adaptée les structures spatiales de marchés immobiliers résidentiels. Le tableau 7 résume les avantages et limites des méthodes cartographiques expérimentées dans cet article.

Tableau 7 : Synthèse des méthodes d'analyse spatiale et de représentation cartographique expérimentées dans l'article

Méthode	Avantages	Limites
<b>Agrégation à un échelon communal ou infra-communal</b>	<ul style="list-style-type: none"> <li>- Information précise dans un découpage "familier".</li> <li>- Applicable à plusieurs variables (prix, prix au m<sup>2</sup>, volume)</li> <li>- Possibilité de croiser avec des données statistiques de l'Insee</li> </ul>	<ul style="list-style-type: none"> <li>- Généralisation de l'information</li> <li>- Effet MAUP</li> <li>- Découpage non homogène dans le cas des IRIS</li> </ul>
<b>Carroyage</b>	<ul style="list-style-type: none"> <li>- Effet MAUP limité</li> <li>- Granularité fine et personnalisable</li> </ul>	<ul style="list-style-type: none"> <li>- Lisibilité/moirage, question du nombre minimal de mutations par maille</li> <li>- Difficulté à se repérer précisément sans toponyme</li> </ul>
<b>Lissage à partir de données agrégées</b>	<ul style="list-style-type: none"> <li>- Aspect lisible et esthétique</li> </ul>	<ul style="list-style-type: none"> <li>- Généralisation un peu excessive de l'information</li> </ul>

<b>Lissage direct à partir des mutations</b>	- Aspect lisible et esthétique restituant plus de nuances	- Traitements assez longs
<b>Cartogramme couplé à une surface lissée</b>	- Représentation originale qui permet de ne pas perdre de vue les volumes - Permet de mettre en balance volumes et prix (pas toujours de corrélation, localement)	- Complicé à lire si la déformation est trop forte, indispensable d'ajouter des points de repères pour le lecteur
<b>Carte restituant une CAH</b>	- Spatialisation d'une typologie (sous marchés aux caractéristiques propres)	- Généralisation de l'information

Mericskay, Demoraes, 2022

- 70 Une autre plus-value de notre article réside dans la mise à disposition de tous les scripts qui formalisent et documentent les chaînes de traitements des données DVF en *open data* que nous avons mis en œuvre intégralement dans R, un environnement libre et gratuit. Ce choix vise à faciliter la compréhension et l'évaluation de travaux de recherche qui constituent aujourd'hui un enjeu important dans le développement de la science ouverte (données libres, logiciels libres, méthodes ouvertes, publications en libre accès), et notamment en géographie (Commenges *et al.*, 2014 ; Nüst, Pebesma, 2020). Ainsi, quiconque souhaite reproduire et adapter tout ou une partie des étapes de préparation, d'analyse et de représentation (carto)graphique présentées dans l'article, peut le faire. La méthode présentée est perfectible et surtout personnalisable à chaque contexte d'utilisation. Toutefois, les choix des seuils dans la préparation des données comme des échelles ou des rayons de lissage au niveau des analyses sont cruciaux et demandent une attention particulière.
- 71 Enfin pour les administrations publiques qui produisent et diffusent les données DVF, ce travail participe à la fois à illustrer des formes de réutilisations possibles ainsi que leurs limites dans des contextes de recherche. Depuis l'appel lancé en 2012 pour l'ouverture des bases DVF (Boulay *et al.*, 2012) et leur mise à disposition en 2019, la communauté scientifique s'approprie progressivement ces données et cet article en est une illustration. Certes, les informations présentes au sein de la base DVF en libre accès se concentrent uniquement sur les caractéristiques des biens et non sur les profils des vendeurs et des acquéreurs comme au sein des bases des notaires. Certes, leur structuration n'autorise pas l'analyse des transactions en bloc et leur utilisation nécessite des compétences techniques spécifiques. Malgré cela, les données DVF en libre accès sont gratuites, accessibles et le volume comme l'exhaustivité des mutations permettent d'analyser dans l'espace et dans le temps les dynamiques immobilières sur un large éventail de territoires. Gageons ainsi que nombreux seront les chercheurs, les analystes, les chargés de missions à explorer davantage les potentialités de cette mine de données dans de futurs travaux.

## BIBLIOGRAPHIE

- Aschan-Leygonie C., Cuntz C., Davoine P.A., 2019, *Introduction aux systèmes d'information géographique*, Paris, Armand Colin, 224 p.
- AUDAP, 2013, "Les mutations foncières et immobilières : analyse croisée DVF et Perval", *Note de l'observatoire de l'Agence d'urbanisme Atlantique & Pyrénées*.
- Banos A., 2001, "À propos de l'analyse spatiale exploratoire des données", *Cybergeo : European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeo/4056>
- Basciani-Funestre M.A., Darley A., 2013, "Des marchés fonciers toujours opaques ?", in : Guelton S. (dir.), *Le foncier en Île-de-France : Retour sur 10 idées reçues*, Paris, Adef éditions, 31-42.
- Béoutis A., Joubert M., 2018, "Parc locatif : un rôle majeur dans les trajectoires résidentielles", *Insee Analyses Nouvelle-Aquitaine*, No.68, [en ligne]. URL : <https://www.insee.fr/fr/statistiques/3652857>
- Bermond M., Marie M., 2016, "Différenciation socio-spatiale et migrations résidentielles dans l'aire urbaine de Caen", *Revue d'Économie Régionale & Urbaine*, vol. 4, 817-846.
- Bivand R.S., 2020, "Progress in the R ecosystem for representing and handling spatial data", *Journal of Geographical Systems*, 1-32.
- Boots B.N., Getis A., 1988, *Point pattern analysis*, Newbury Park, Sage.
- Boulay G., 2011, *Le prix de la ville : le marché immobilier à usage résidentiel dans l'aire urbaine de Marseille-Aix-en-Provence (1990-2010)*, Thèse de doctorat, Université de Provence Aix-Marseille 1, 565 p.
- Boulay G., Guérois M., Le Goix R., 2011, "Acquéreurs et vendeurs dans l'inflation immobilière : une analyse des trajectoires locales à Paris et Marseille (1996-2006)", in : Pumain D., Mattei M.F. (dir.), *Données Urbaine* vol. 6, Paris, Anthropos/Economica, 167-179.
- Boulay G., Buhot C., Fournier J.L., 2012, "Les chercheurs exclus de l'open data ? Appel à signatures lancé à la communauté de la recherche", *Cybergeo : European Journal of Geography*, [en ligne]. URL : <http://journals.openedition.org/cybergeo/25520>
- Boulay G., Blanke D., Casanova Enault L., Granié A., 2021, "Moving from Market Opacity to Methodological Opacity: Are Web Data Good Enough for French Property Market Monitoring?", *The Professional Geographer*, vol. 73, No.1, 115-130.
- Buhot C., 2006, *Marché du logement et division sociale de l'espace dans les îles du Ponant*, Thèse de doctorat en géographie, Université de Bretagne occidentale.
- Casanova L., Helle C., 2012, "Ce que les dynamiques foncières révèlent du devenir des territoires : éléments de prospective du sud-est français", *L'Espace géographique*, vol. 41, No.2, 111-127.
- Casanova Enault L., 2017, "Des acteurs de poids sur le littoral : les petits propriétaires immobiliers. Application au littoral provençal et azuréen", *Cybergeo : European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeo/28394>
- Casanova L., Boulay G., Gérard Y., Yahi L., 2017, "Deux bases de données, aucune référence de prix", *Revue d'Économie Régionale & Urbaine*, vol. 4, 711-732.
- Casanova Enault L., Boulay G., Coulon M., 2019, "Une aubaine pour les géographes ? Intérêts des fichiers open DVF sur les transactions foncières et immobilières et précautions d'usage",

- Cybergeog : *European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeog/33602>
- Cerema, 2017, "Présentation de la base de données DVF ", *Cerema Club SIG*, [en ligne]. URL : [http://www.pays-de-la-loire.developpement-durable.gouv.fr/IMG/pdf/2017-11-27\\_dv3f\\_blg\\_2.pdf](http://www.pays-de-la-loire.developpement-durable.gouv.fr/IMG/pdf/2017-11-27_dv3f_blg_2.pdf)
- Cerema, 2020, "Guide de prise en main de DVF3F", [en ligne]. URL : <http://doc-datafoncier.cerema.fr/dv3f/tuto/>
- Chaillot P., Loizeau P., 2019, "Maine-et-Loire : des couples entre 25 et 64 ans quittent les villes-centres pour acheter une maison", *Insee Analyses Pays de la Loire*, No.70, [en ligne]. URL: <https://www.insee.fr/fr/statistiques/3711825>
- Chapelle G., Eyméoud J.B., 2022, "Can big data increase our knowledge of local rental markets? A dataset on the rental sector in France", *PloS one*, vol. 17, No.1, [en ligne]. URL : <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0260405>
- Chappert A., Kabla-Langlois I., Friggit J., Laporte P., 2014, "L'organisation du service statistique dans le domaine du logement", *Rapport du Conseil Général de l'Environnement et du Développement Durable et de l'Insee*, [en ligne]. URL : <https://www.vie-publique.fr/rapport/34272-lorganisation-du-service-statistique-dans-le-domaine-du-logement>
- Chesnel H., Le Graët A., Fouin L., 2018, "Les habitants modestes du cœur de la métropole nantaise deviennent plus souvent propriétaires lorsqu'ils s'éloignent", *Insee Analyses Pays de la Loire*, No.55, [en ligne]. URL : <https://www.insee.fr/fr/statistiques/3148471>
- Commenges H., Beauguitte L., Buard E., Cura R., Le Néchet F., Le Texier M., et al., 2014, *R et espace : Traitement de l'information géographique*, Groupe ElementR, Framabook, Paris.
- Cornillon P.A., Demoraes F., (à paraître), "cartogramR: A Package to Quickly Create Accurate Density-equalizing Maps in R", *Journal of Statistical Software*.
- Coulondre A., 2017, "Le potentiel des données DVF dans l'analyse des marchés immobiliers locaux : le cas d'une typologie des micro-marchés du logement à Bordeaux Métropole", *Rapport de recherche - Bordeaux Métropole et LIFTI*, [en ligne]. URL : <https://hal.archives-ouvertes.fr/hal-01658515>
- Coulondre A., 2018, "Ouvrir la boîte noire des marchés du logement", *Métropolitiques*, [en ligne]. URL : <https://hal.archives-ouvertes.fr/hal-01838393/>
- Coulondre A., 2019, "L'interopérabilité des données DVF pour l'analyse des marchés immobiliers locaux : Un guide méthodologique pour l'étude du résidentiel illustré par le cas de Nantes Métropole", *Rapport de recherche - Nantes Métropole, Explore et LIFTI*, [en ligne]. URL : <https://hal.archives-ouvertes.fr/hal-02362109>
- Coulondre A., 2020, "Les marchés résidentiels en Île-de-France : diversités locales et évolutions aux abords des futures gares", *Rapport de recherche - ORF et DRIEA*, [en ligne]. URL : <https://hal.archives-ouvertes.fr/hal-02873106>
- Coulondre A., Juillard C., 2021, *Ce que la data fait au marché*, Rapport de recherche : PUCA ; BPCE ; Crédit Logement ; LIFTI, [en ligne]. URL : <https://hal-enpc.archives-ouvertes.fr/halshs-03321023v1>
- Demoraes F., Bouquet M., Mericskay B., 2021, "L'efficacité visuelle des cartogrammes animés en question – Une piste d'amélioration à travers l'exemple de la ségrégation à Bogotá (1993-2005)", *M@ppemonde*, vol. 131, [en ligne]. URL : <https://journals.openedition.org/mappemonde/5813>
- Desponds D., Bergel P., 2014, "« La mauvaise réputation ». Distance aux ZUS et prix de l'immobilier : application d'un modèle hédonique à six intercommunalités franciliennes",

- Cybergeog : *European Journal of Geography*, [en ligne]. URL: <https://journals.openedition.org/cybergeog/26283>
- Diggle P.J., 1979, "Statistical methods for spatial point patterns in ecology", in: Cormack R.M., Ord J.K., *Spatial and temporal analysis in ecology*, Fairland, International Cooperative Publishing House, 95-150
- Donnelly K.P., 1978, "Simulations to Determine the Variance and Edge Effect of Total Nearest-Neighbor Distance", in: Hooder I. (ed.), *Simulations methods in Archeology*, London, Cambridge University Press
- Driant J.C., Gomez E., Herbert B., Martin Y., 2005, "La Connaissance des marchés locaux de l'habitat : les principales sources d'information existantes et leurs limites d'utilisation", *Rapport de l'Agence Nationale pour l'Amélioration de d'Habitat*, [en ligne]. URL : [www.silvereco.fr/wp-content/uploads/2013/12/Connaissance-des-march%C3%A9s-locaux-de-lhabitat-Anah-2005.pdf](http://www.silvereco.fr/wp-content/uploads/2013/12/Connaissance-des-march%C3%A9s-locaux-de-lhabitat-Anah-2005.pdf)
- Eblé S., Fievet A., Gamblin V., 2021, "Un dynamisme et des prix des logements plus élevés à Lille, sur le littoral et au sud de la région", *Insee Analyses Hauts-de-France*, No.124, [en ligne]. URL : <https://www.insee.fr/fr/information/5385935>
- Genebes L., Renaud A., Sémécurbe F., 2018, "Le lissage spatial", in : Loonis V. (dir.), Manuel d'analyse spatiale, *Insee Méthodes*, No.131, [en ligne]. URL : <https://www.insee.fr/fr/statistiques/fichier/3635442/imet131-l-chapitre-8.pdf>
- Gidrol J.C., Canonéro B., 2016, "Croissance démographique et marché immobilier : 100 000 transactions en trois ans sur les territoires du Languedoc-Roussillon", *Insee Analyses Occitanie*, No. 32, [en ligne]. URL: <https://www.insee.fr/fr/information/2511191>
- Giraud T., Lambert N., 2016, "Cartography: Create and Integrate Maps in your R Workflow", *Journal of Open Source Software*, vol. 1, No.4, 54.
- Giraud T., Lambert N., 2017, "Reproducible cartography", in: Peterson M.P. (ed.), *Advances in Cartography and GIScience: Selections from the International Cartographic Conference 2017*, Cham, Springer, 173-183.
- GnDVF, 2017, "Guide d'accompagnement DVF", *Rapport du Groupe national DVF*, [en ligne]. URL: [http://www.groupe-dvf.fr/wp-content/uploads/2018/10/DVF-GuideAccompagnement\\_2017-12-01.pdf](http://www.groupe-dvf.fr/wp-content/uploads/2018/10/DVF-GuideAccompagnement_2017-12-01.pdf)
- Goodman S.N., Fanelli D., Ioannidis J.P., 2016, "What does research reproducibility mean?", *Science translational medicine*, vol. 8, No.341.
- Guérois M., Le Goix R., 2009, "La dynamique spatio-temporelle des prix immobiliers à différentes échelles : le cas des appartements anciens à Paris (1990-2003)", *Cybergeog : European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeog/22644>
- Grasland C., 1999, "Lissage cartographique et animation spatio-temporelle : quelques réflexions théoriques et méthodologiques", in : Domingo J. (dir.), *Travaux de l'Institut Géographique de Reims*, vol. 26, No.101-104, 83-104.
- Juillard C., 2019, "Le tournant numérique des données immobilières : permanences et mutations", *Rapport d'étude du iread, du LIFTI, du PUCA et d'Urbanics*, [en ligne]. URL : <https://www.i-read.io/apps/etude/le-tournant-numerique-des-donnees-immobilierees/>
- Juillard C., 2020, "Produire des données de prix et de loyers à l'heure de la PropTech : Quel rôle pour l'État ?", *Working Paper - Chaire "Villes et numérique" | Sciences Po École urbaine*, [en ligne]. URL : [https://www.sciencespo.fr/ecole-urbaine/sites/sciencespo.fr.ecole-urbaine/files/2020\\_01%20-%20Juillard%20copie.pdf](https://www.sciencespo.fr/ecole-urbaine/sites/sciencespo.fr.ecole-urbaine/files/2020_01%20-%20Juillard%20copie.pdf)

Lagleize J.L., 2019, *Rapport au Premier ministre sur la maîtrise des coûts du foncier dans les opérations de construction*, Assemblée nationale.

Lajoie G., 1992, *Le carroyage des informations urbaines : une nouvelle forme de banque de données sur l'environnement du Grand Rouen*, Mont-Saint-Aignan, Presses universitaires de Rouen et du Havre.

Lambert N., Zanin C., 2016, *Manuel de cartographie : principes, méthodes, applications*, Paris, Armand Colin, 224 p.

Le Brun P., 2019, "La désaccumulation sélective : la place des personnes âgées sur le marché immobilier résidentiel du Grand Lyon (2006-2015)", *Espace, populations, sociétés*, vol. 3, [en ligne]. URL : <https://journals.openedition.org/eps/9310>

Le Corre T., 2019, *Paris à tous prix. Analyse des inégalités par une géographie de l'investissement sur le marché immobilier résidentiel en Île-de-France*, Thèse de doctorat, Paris, Université Paris 1 - Panthéon Sorbonne, [en ligne]. URL : <https://tel.archives-ouvertes.fr/tel-02298354>

Le Corre T., 2021, "Une base de données pour étudier vingt années de dynamiques du marché immobilier résidentiel en Île-de-France", *Cybergeo : European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeo/37430>

Le Goix R., Boulay G., Callen D., Derambure-Dutel F., Reys A., Giraud T., Cura R., 2016, "Marchés immobiliers et zones critiques : une analyse du changement social en relation avec la dynamique des valeurs immobilières", *Actes du CIST2016*, Grenoble, 268-276.

Le Goix R., Giraud T., Cura R., Le Corre T., Migozzi J., 2019, "Who sells to whom in the suburbs? Home price inflation and the dynamics of sellers and buyers in the metropolitan region of Paris, 1996-2012", *PloS one*, vol. 14, No.3, [en ligne]. URL: <https://dx.plos.org/10.1371/journal.pone.0213169>

Le Goix R., Ysebaert R., Giraud T., et al., 2021, "Unequal housing affordability across European cities. The ESPON Housing Database, Insights on Affordability in Selected Cities in Europe", *Cybergeo: European Journal of Geography*, [en ligne]. URL : <https://journals.openedition.org/cybergeo/36478>

Levasseur S., 2013, "Éléments de réflexion sur le foncier et sa contribution au prix de l'immobilier", *Revue de l'OFCE*, No.128, 365-394.

Lovelace R., Nowosad J., Muenchow J., 2016, *Geocomputation with R*, Boca Raton, CRC Press

Masclet de Barbarin M., 2019, "L'ouverture au public des bases de données foncières de l'administration fiscale", *Ingénierie patrimoniale*, No.2, 69-74.

Mericskay B., Rivière J., Loret S., 2020, "Les mutations de la géographie des prix de l'immobilier résidentiel", *Atlas social de la métropole nantaise*, [en ligne]. URL: <https://asmn.univ-nantes.fr/index.php?id=480>

Nüst D., Pebesma E., 2021, "Practical reproducibility in geography and geosciences", *Annals of the American Association of Geographers*, vol. 111, No.5, 1300-1310.

Oliveau S., 2020, "Pavage, carroyage, tessellation", *Hypergeo*, [en ligne]. URL: <https://hypergeo.eu/pavage-carroyage-tessellation/>

Openshaw S., Taylor P.J., 1979, "A million or so correlation coefficients: three experiments on the modifiable areal unit problem", in: Wrigley N. (eds.), *Statistical applications in spatial sciences*, Londres, Pion, 127-144

Palm R., 1978, "Spatial segmentation of the urban housing market", *Economic Geography*, vol. 54, No.3, 210-221.

- Pilarski C., Zemis G., 2019, "Le prix moyen des transactions immobilières varie du simple au double entre les territoires de franges de Bourgogne-Franche-Comté", *Insee Flash Bourgogne-Franche-Comté*, No.82, [en ligne]. URL : <https://www.insee.fr/fr/statistiques/4127730>
- Pumain D., Saint-Julien T., 2010, *Analyse spatiale : les localisations*, Paris, Armand Colin, 192 p.
- Saujot M., Erard T., 2015, "Les innovations de la ville intelligente au secours de la ville durable ? Décryptage à partir des enjeux de données", *Working Paper IDDRI*, No.2, [en ligne]. URL: <https://www.iddri.org/fr/publications-et-evenements/document-de-travail/les-innovations-de-la-ville-intelligente-au-secours>
- Silverman B.W., 1998, *Density estimation for statistics and data analysis*, Londres, Routledge, 176 p.
- Simon A., 2009, "Une nouvelle ère dans le domaine de l'information immobilière : La révolution de la géolocalisation", *L'observateur de l'immobilier*, [en ligne]. URL: <https://halshs.archives-ouvertes.fr/halshs-00551304>
- Van den Elzen S., Van Wijk J.J., 2013, "Small multiples, large singles: A new approach for visual data exploration", *Computer Graphics Forum*, vol. 32, No.3, 191-200.
- Wickham H., 2016, *ggplot2: elegant graphics for data analysis*, Houston, Springer, 260 p.
- Wickham H., Averick M., Bryan J., Chang W., McGowan L.D.A., François R., Yutani H., 2019, "Welcome to the Tidyverse", *Journal of Open Source Software*, vol. 4, No.43, 1686.
- Yau N., 2013, *Data visualisation : De l'extraction des données à leur représentation graphique*, Paris, Éditions Eyrolles, 355 p.
- Zaninetti J.M., 2005, *Statistique spatiale, méthodes et applications géomatiques*, Paris-Londres Hermès Science Publishing Lavoisier, 320 p.

## Packages R

<https://cran.r-project.org/web/packages/cartogramR/index.html>

<https://cran.r-project.org/web/packages/cartography/index.html>

<https://cran.r-project.org/web/packages/cluster/index.html>

<https://cran.r-project.org/web/packages/geoR/index.html>

<https://cran.r-project.org/web/packages/ggplot2/index.html>

<https://cran.r-project.org/web/packages/maptools/index.html>

<https://cran.r-project.org/web/packages/raster/index.html>

<https://cran.r-project.org/web/packages/rgdal/index.html>

<https://cran.r-project.org/web/packages/rgeos/index.html>

<https://cran.r-project.org/web/packages/rmapshaper/index.html>

<https://cran.r-project.org/web/packages/sf/index.html>

<https://cran.r-project.org/web/packages/sp/index.html>

<https://cran.r-project.org/web/packages/spatstat/index.html>

<https://cran.r-project.org/web/packages/spdep/index.html>

<https://cran.r-project.org/web/packages/tidyverse/index.html>

## NOTES

1. Loi ENL (Engagement National pour le Logement) donne la possibilité aux collectivités qui en font la demande de disposer de l'ensemble des informations foncières détenues par l'administration fiscale au cours des 5 dernières années.
2. Plusieurs entreprises spécialisées proposent aujourd'hui des applications et des services utilisant les données DVF pour consulter et comparer les prix de l'immobilier et estimer la valeur de biens.
3. La base DVF couvre l'ensemble de la France métropolitaine à l'exception des départements du Bas-Rhin, du Haut-Rhin et de Moselle, où les données les concernant relèvent du "livre foncier". La base DVF couvre aussi les départements et régions d'outre-mer, à l'exception de Mayotte.
4. Dans la pratique, les notaires procèdent à l'enregistrement des actes au sein de deux systèmes d'information distincts. Une copie de l'acte est obligatoirement transmise au service de la publicité foncière de la DGFIP, chargé de l'enregistrement des actes de vente, qui assure l'opposabilité au tiers de toute mutation et qui sert à alimenter la base DVF. L'acte est également enregistré (sur la base du volontariat) au sein du système des notaires de la société Min.not qui alimente les bases PERVAL et BIEN.
5. Conformément au décret n°2018-1350 du 28 décembre 2018 relatif à la publication sous forme électronique des informations portant sur les valeurs foncières déclarées à l'occasion des mutations immobilières.
6. Les données brutes sont proposées par année sous forme de fichier .txt.
7. Etalab est un département de la Direction interministérielle du numérique qui coordonne la conception et la mise en œuvre de la stratégie de l'État dans le domaine des données (*open data*, algorithmes publics, schémas de données, etc.).
8. Ces données sont accessibles en *open data*, sous la forme de fichier plat (.csv). Ces données "retravaillées" sont encodées en UTF8, normalisées avec des données de référence à jour (COG 2019, PCI 2019) et enrichies des coordonnées géographiques (par géocodage) des parcelles (WGS84).
9. La base DV3F propose une structure sous la forme d'une base de données relationnelle (PostgreSQL). Elle est enrichie d'informations issues des fichiers fonciers (géolocalisation des parcelles, informations sur les acquéreurs et les caractéristiques des biens) et propose des indicateurs facilitant l'observation. L'objectif visé est de proposer une base qui ne présente pas uniquement des données fiscales brutes mais des données d'observations foncières sous la forme de décomptes, de classifications, de typologies ou de segmentations (Cerema, 2017).
10. En complément des fonctions de base de R, plusieurs packages sont utilisés dans cet article. L'apurement et la préparation des données DVF reposent principalement sur le package *tidyverse*. Les analyses statistiques et la création des graphiques font appel aux packages *ggplot2* et *cluster*. La manipulation des données spatiales est mise en œuvre avec les packages *sf*, *sp*, *raster*, *rgdal*, *rgeos*, *spdep*, *geoR* et *rmapshaper*. Les cartes choroplèthes sont produites avec les packages *cartography* et *maptools* et les cartes lissées avec le package *spatstat*. Enfin les cartogrammes sont conçus avec le package *cartogramR*.
11. Sélection des ventes de maisons et d'appartements > 270 617 mutations distinctes (350 024 lignes).
12. Suppression des mutations avec des cellules vides (type, prix, surface, coordonnées géographiques) > 266 082 mutations distinctes (340 801 lignes).
13. Il existe quelques méthodes permettant d'analyser les mutations multiventes, comme le découpage du prix total de la transaction entre biens homogènes proportionnellement à leur superficie (Casanova *et al.*, 2017), la création de proxy d'identifiants de bâtiment (Casanova Enault *et al.*, 2019) ou l'agrégation des caractéristiques des biens au niveau des transactions (Coulondre, 2019).



14. Dans notre base, les mutations monoventes représentent 81 % des mutations (216 735 mutations) ; les mutations multiventes cohérentes représentent 14 % des mutations (36 867 mutations) ; les mutations en lots représentent moins de 5 % des mutations (12 480 mutations). Ces dernières renvoient à davantage de biens, puisqu'elles intègrent les ventes en lots pour lesquelles une mutation peut correspondre à la vente de plusieurs biens, jusqu'à 79 dans notre cas.

15. Ce chiffre doit toutefois être relativisé dans la mesure où, dans certaines zones d'études, les mutations multiventes peuvent constituer une part importante des mutations (Casanova Enault *et al.*, 2019).

16. Dans le cadre de successions par exemple, des vendeurs peuvent céder des biens à des prix modiques, un biais déjà identifié dans les bases PERVAL/BIEN (Boulay, 2011 ; Le Corre, 2019).

17. La définition des seuils minimaux du prix nominal et du prix au m<sup>2</sup> est relative à la distribution de ces variables. Elle s'appuie sur la méthode des percentiles. Nous avons exclu la valeur du 1er percentile, soit respectivement 15 000 Euros et 330 Euros/m<sup>2</sup>. Le seuil minimal de la surface est quant à lui absolu et a été fixé à 10 m<sup>2</sup>, superficie en-dessous de laquelle il est difficile d'imaginer habiter. Pour ce qui est des seuils maximaux concernant les prix nominaux, les prix au m<sup>2</sup> et les surfaces, leur définition ne peut reposer sur la méthode des percentiles car certaines valeurs extrêmes, qui ne renvoient pas forcément à des erreurs de saisie, doivent être conservées dans le jeu de données (comme des biens très chers ou très grands). Nous avons donc utilisé la méthode du seuillage manuel et adopté le seuil naturel arrondi le plus élevé qui apparaît sur les histogrammes de distribution. Pour le prix nominal, le seuil a été fixé à 5 millions d'Euros, pour le prix au m<sup>2</sup> à 8000 Euros/m<sup>2</sup> et pour les surfaces à 350 m<sup>2</sup> pour une maison et à 200 m<sup>2</sup> pour un appartement.

18. Cette adaptation reste simple à mettre en place puisqu'il suffit de changer les valeurs de ces paramètres dans le script.

19. Les *small multiples* sont une série de graphiques ou de tableaux similaires qui fractionnent un graphique en plusieurs versions de lui-même en utilisant la même échelle et les mêmes axes.

20. La méthode de discrétisation utilisée est celle des seuils observés sur l'histogramme de fréquence avec des bornes de classe définies manuellement. Cette méthode permet d'identifier les ruptures dans la série statistique et de distinguer des groupes de communes dont les prix au m<sup>2</sup> sont assez semblables entre eux et différent des autres groupes. Le passage d'une teinte de couleurs à l'autre au sein du nuancier renvoie à la moyenne observée.

21. La méthode de discrétisation utilisée est la même que celle de la figure 10.

22. Le MAUP désigne l'influence du découpage spatial (effets d'échelle et effets de zonage) sur les résultats de traitements statistiques ou de modélisation.

23. Le choix de mailles carrées de 2 km représente un compromis entre rigueur statistique (3 mutations au minimum par maille) et lisibilité cartographique (granularité pas trop fine pour une couverture cartographique régionale). Au total, le carroyage comporte 7 456 carreaux dont 16 % contiennent moins de 3 mutations et 6 % aucune mutation.

24. La technique du lissage spatial s'applique avant tout à des données considérées comme exhaustives ou supposées représentatives de l'ensemble de la distribution étudiée. C'est pourquoi le lissage par la méthode des noyaux est plus adapté aux spécificités de phénomènes économiques et sociaux (concentration d'équipements, d'événements localisés, ventes de bien localisés, etc.) et ne doit pas être confondu avec les méthodes d'interpolation utilisées dans les sciences de l'environnement, dont le but est d'estimer la valeur en tout point d'une zone d'étude à partir d'échantillons de mesures ponctuelles (Zaninetti, 2005).

25. La méthode de discrétisation utilisée est appelée Q6, une variante de la méthode des quantiles qui consiste à isoler et mettre en évidence les deux classes extrêmes. Cette méthode de discrétisation peut s'utiliser avec n'importe quelle forme de distribution. Elle permet de comparer les cartes entre elles (Lambert, Zanin, 2016).

26. La méthode de discrétisation utilisée est également celle appelée Q6.
27. La distance moyenne qui sépare les 247 646 mutations est de 66 m (moyenne des distances au plus proche voisin). Un rayon de 2 km garantit donc qu'un nombre important de mutations soit pris en compte dans le calcul. Le résultat est donc particulièrement robuste d'un point de vue statistique.
28. <https://cran.r-project.org/web/packages/cartogramR/>
29. La méthode de discrétisation utilisée est également celle appelée Q6.
30. Une CAH vise à constituer des groupes d'individus statistiques regroupés en classes les plus homogènes possibles. Dans notre cas, compte tenu du faible nombre de variables en entrée (quatre), il n'a pas été nécessaire de réaliser, en amont de la CAH, une analyse en composantes principales (ACP) pour réduire les dimensions.
31. Il est tout à fait envisageable d'ajouter d'autres variables aux analyses comme les surfaces moyennes, la part des maisons et des appartements, la population, la densité, le revenu médian, etc.
32. La classification a été calculée à l'aide de l'algorithme de regroupement hiérarchique ascendant AGNES (*Agglomerative Nesting*), en utilisant la méthode de Ward et la distance euclidienne. Cette classification regroupe les individus de façon à ce que les classes constituées soient les plus homogènes possibles, c'est-à-dire avec une variance intra-classe minimale et une variance interclasse maximale.
33. On a ici opté pour une partition en sept classes qui, à lecture de l'arbre hiérarchique et du graphe des gains d'inertie, semblait le meilleur compromis entre diversité et cohérence des catégories obtenues.
34. Pour des questions de visibilité, le taux de variation autour de la moyenne du nombre de transactions, qui monte à + 7 500 % pour le groupe des "Grandes villes", a été enlevé du graphique.
35. Les IRIS (Îlots Regroupés pour l'Information Statistique) sont les plus petites unités spatiales statistiques, inférieures à l'échelon communal. Ils constituent la brique de base en matière de diffusion de données statistiques infra-communales.
36. Les sections présentent en effet une certaine homogénéité sur le plan du bâti et de la morphologie urbaine. Il s'agit donc d'un maillage idoine si l'on s'intéresse par exemple aux discontinuités spatiales des prix de l'immobilier, thème qui dépasse le cadre du présent article.
37. L'effet de bord décrit le fait que lorsqu'on cible l'analyse sur un espace donné, les unités spatiales ou les points situés en limite de la zone d'étude ont mécaniquement moins de voisins que ceux localisés au centre. Cela pose problème notamment pour les opérations de lissage spatial calculées sur un rayon donné. Pour corriger cet effet, plusieurs solutions ont été proposées (Donnelly, 1978 ; Diggle, 1979 ; Boots, Getis, 1988). La plus couramment utilisée consiste à intégrer dans le calcul les valeurs des unités spatiales ou des points situés sur la frange extérieure de la zone d'étude, lorsque ces valeurs sont disponibles bien sûr. C'est la solution que nous avons retenue ici.
38. Méthode de discrétisation Q6. Les ordres de grandeur donnés à voir sur les cartes sont donc comparables. La première classe occupe la même proportion de surface d'une carte à l'autre, de même pour les classes suivantes.
39. Rayon de lissage : 500 m. Fonction de décroissance gaussienne. Pour limiter l'effet de bord, le prix des appartements vendus dans les communes limitrophes de Rennes a été pris en compte, dans la limite du rayon de lissage.
40. Cette méthode de discrétisation permet de valoriser les valeurs extrêmes.

## INDEX

**geographyun** 908, 926, 250

**Keywords** : exploratory data analysis (EDA), real estate market, spatial analysis, cartography, open data

**Mots-clés** : analyse exploratoire des données (EDA), marché immobilier, analyse spatiale, cartographie, données ouvertes

**Palabras claves** : análisis exploratorio de datos, mercado inmobiliario, análisis espacial, cartografía, datos abiertos

## AUTEURS

### BORIS MERICKSKAY

Université Rennes 2 – UMR Espaces et Sociétés 6590 (ESO-Rennes), France

Maître de conférences en géographie

Boris.merickskey@univ-rennes2.fr

<https://perso.univ-rennes2.fr/boris.merickskey>

### FLORENT DEMORAES

Université Rennes 2 – UMR Espaces et Sociétés 6590 (ESO-Rennes), France

Enseignant-chercheur

florent.demoraes@univ-rennes2.fr

<https://perso.univ-rennes2.fr/florent.demoraes>