



HAL
open science

Les données DVF en open data comme matériel de recherche pour les géographes : propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise

Boris Mericskay, Florent Demoraes

► To cite this version:

Boris Mericskay, Florent Demoraes. Les données DVF en open data comme matériel de recherche pour les géographes : propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise. Journée d'études du CIST - Session 1 – Géographie de la mobilité et de la population - Accès et utilisation des données en libre accès en géographie de la population, de la santé et des mobilités, Dec 2021, Paris, France. halshs-03644136

HAL Id: halshs-03644136

<https://shs.hal.science/halshs-03644136v1>

Submitted on 18 Apr 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 Attribution 4.0 International License



Les données DVF en *open data* comme matériel de recherche pour les géographes

Propositions méthodologiques à travers l'exemple de la Bretagne et la métropole rennaise

[Boris Mericskay](#)

[Florent Demoraes](#)

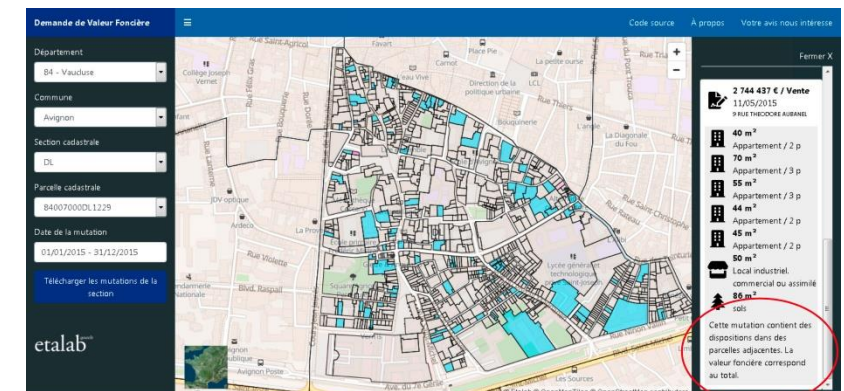
Données sur les marchés immobiliers

- Les données relatives aux marchés fonciers et immobiliers sont multidimensionnelles, nombreuses et présentent une grande diversité, tant au niveau de leurs formes que de leurs producteurs
 - Une vingtaine de bases de données produites et mises à jour par différents acteurs publics et privés, lesquels se concurrencent et se complètent à la fois
- Les bases BIEN et PERVAL des notaires comme « références »



La base « Demandes Valeurs Foncières »

- Projet de centralisation de données sur les transactions immobilières produites par l'administration fiscale et de mise à disposition en *open data* (une démarche, un service et une base de données)
 - Prémices de DVF en 2006, mis en ligne en 2011 les données pour les ayant droits et en 2018 pour le grand public
 - Référentiel commun *quasi* national 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



DVF en open data

- Une même source mais trois bases de données

Demandes de valeurs foncières DVF

Mis à jour le 20 octobre 2021 — [Licence Ouverte / Open Licence version 2.0](#)

Producteur Métadonnées Actions



Ministère de l'économie, des finances et de la relance

Le ministère de l'économie, des finances et de la relance est chargé de mettre en œuvre la politique du Gouvernement économique, financière, budgétaire et fiscale dans des champs d'action variés tels que l'industries, l'innovation mais aussi...

440 jeux de données 1 réutilisations

MINISTÈRE DE LA TRANSITION ÉCOLOGIQUE
Cerema
Datafoncier, données pour les territoires

Accueil Données Outils Usages Accompagnement

DVF+ open-data

Demandes de valeurs foncières géolocalisées DVF

28 favoris

géolocalisées

Mis à jour le 26 octobre 2021 — [Licence Ouverte / Open Licence version 2.0](#)

Producteur Métadonnées Actions



Etalab

Etalab est un département de la direction interministérielle du numérique (DINUM), dont les missions et l'organisation sont fixées par le décret du 30 Octobre 2019. Il coordonne notamment la conception et la mise en œuvre de la stratégie de l'État dans le domaine de la donnée. Ouverture (open...

56 jeux de données 33 réutilisations

SOMMAIRE

1. Description
2. Ressources ⁽¹⁾
3. Ressources communautaires ⁽⁰⁾
4. Réutilisations ⁽²⁹⁾
5. Discussions ⁽⁵⁹⁾

géolocalisée

Les "Demandes de valeurs foncières" (DVF) sont disponibles en open data.

Cela permet d'accéder librement à cette même donnée géolocalisée aisément exploitable pour...

sur le modèle de données partagé dit "Licence Ouverte" national DVF et qui existe depuis 2013. Ce modèle permet notamment une table des mutations dans les communes à la localisation d'une transaction.

Les données du Plan cadastral informatisé également...

✓ Télécharger les données

Espace de téléchargement

✓ Installer les données

Procédure d'installation

✓ Accéder au dictionnaire DVF+/DV3F

Dictionnaire des variables

Synthèse des données DVF en open data

Avantages	Limites
<ul style="list-style-type: none">- Gratuites et accessibles en open data- Disponibles sur l'essentiel du territoire national- Mises à jour deux fois par an- Historicité des données depuis 2014- Réputées exhaustives- Globalement fiables- Géolocalisées à la parcelle	<ul style="list-style-type: none">- Transactions non disponibles pour 3 départements- Complexes à exploiter- Structuration des transactions multiventes- Pas d'identifiant de transaction- Difficilement intelligibles pour les non experts- Présence d'erreurs de saisie- Pas d'information sur les vendeurs et les acquéreurs (sauf dans la version DV3F)

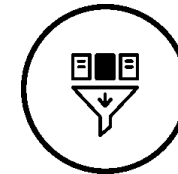
Ce ne sont pas des données « prêtes à l'emploi »

Elles nécessitent un temps de compréhension et la mise en place de méthodes adaptées

Objectifs de notre démarche

- 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

- Formaliser une méthode de préparation des données DVF en open data (apurement, filtrage, restructuration, enrichissement)



- Analyser à plusieurs échelles ces données en mobilisant différentes méthodes (agrégation, lissage, classification) ;



- Restituer (carto)graphiquement ces données au moyen de plusieurs types de rendus (graphiques, cartes choroplèthes, carte catégorisées, cartes lissées, carroyage et cartogramme)



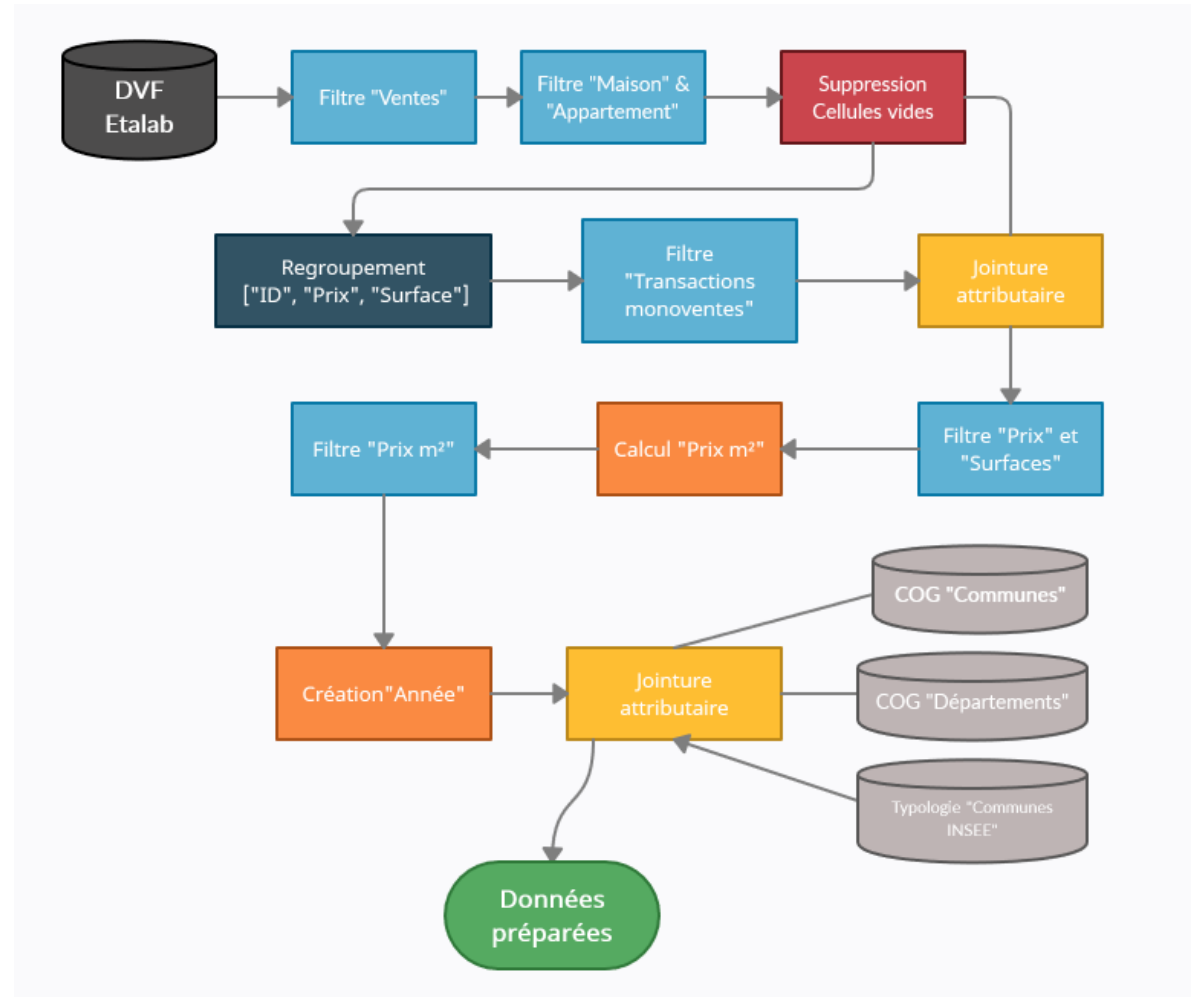
Méthode automatisée reproductible

- Démarche de science ouverte
 - **Répétabilité** de la démarche : refaire les analyses dans les mêmes conditions
 - **Réplicabilité** de la démarche : refaire les analyses avec de nouveaux critères (périmètres d'analyses, variables, seuils de prix, types de biens...)
 - **Reproductibilité** de la démarche : remobiliser les approches pour les adapter à d'autres terrains d'étude comme une autre région ou une ville
- Langage de programmation et le logiciel libre R
 - Nombreux packages complémentaires
 - *tidyverse, sf, cartography, cluster, spatstat, cartogramR...*
 - Chaînes de traitements automatisées
 - Mise en place de 4 scripts en RMarkdown



Préparation des données

Apurement, nettoyage
Restructuration
Enrichissement



Préparation des données

Mutation monovente « monoligne » : une disposition, 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 monovente « multiligne » : une disposition, un bien unique, un prix et une surface mais une redondance du bien au sein de la base (si 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	Maison	250000	150

Mutation multivente « multiligne » : une disposition, 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

Préparation des données

446 818 mutations (1 043 052 lignes)

247 646 Mutations (247 646 lignes)

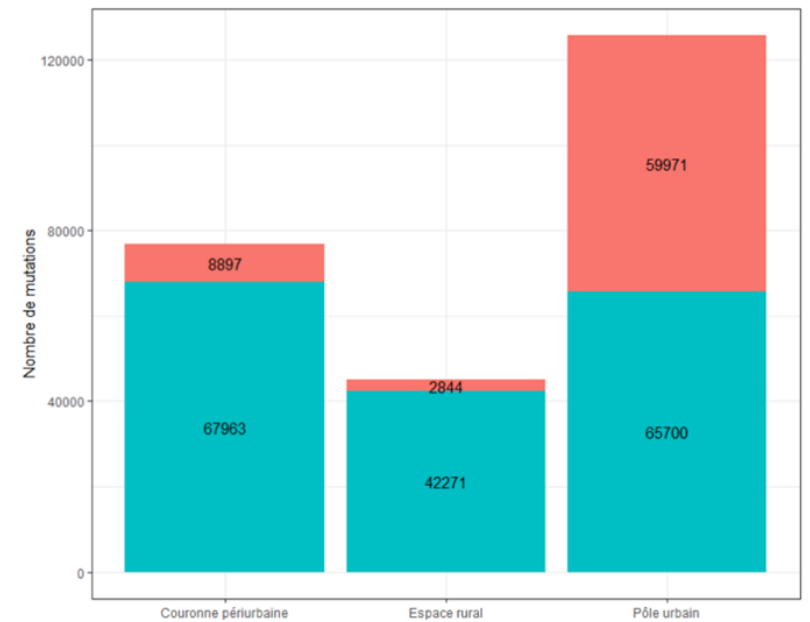
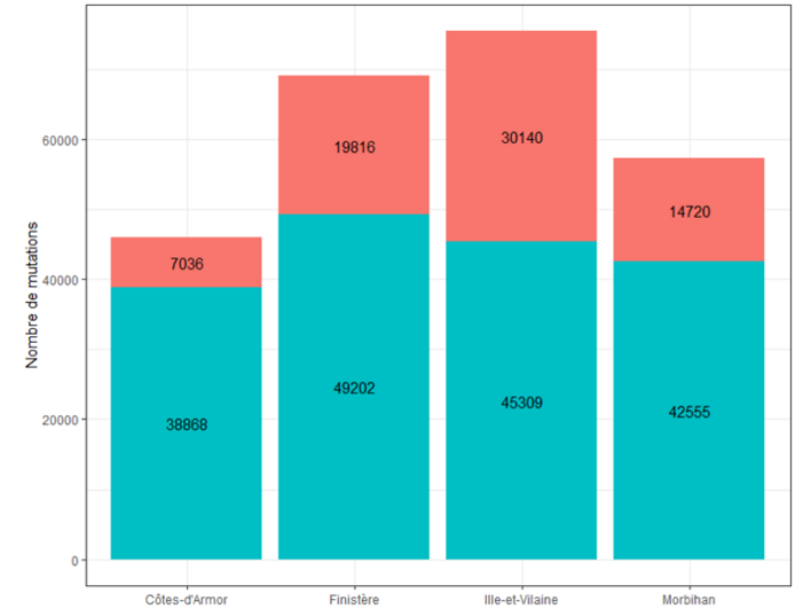
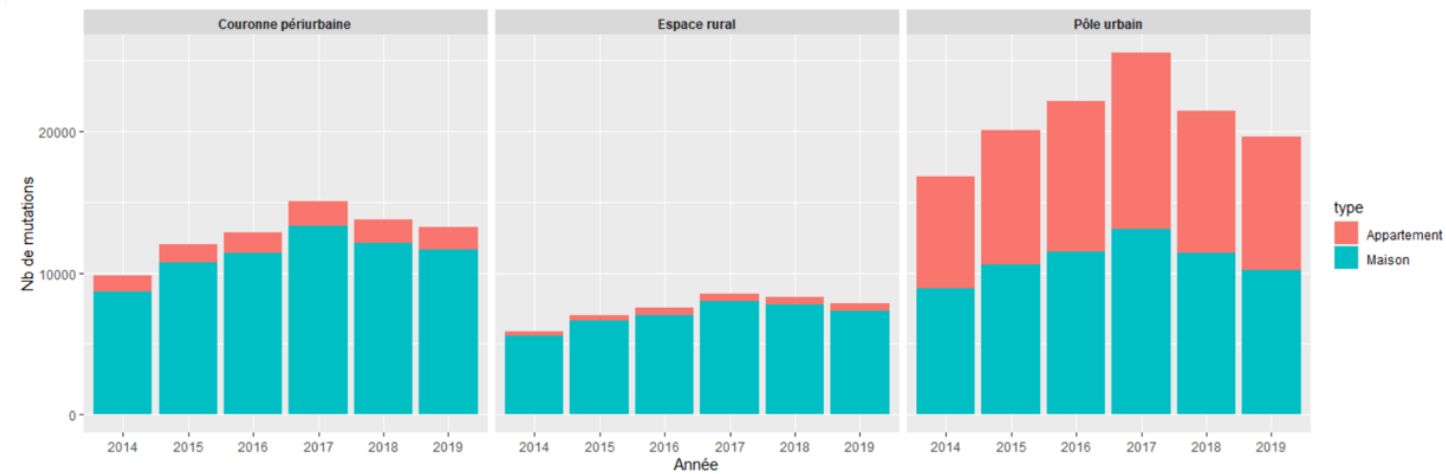
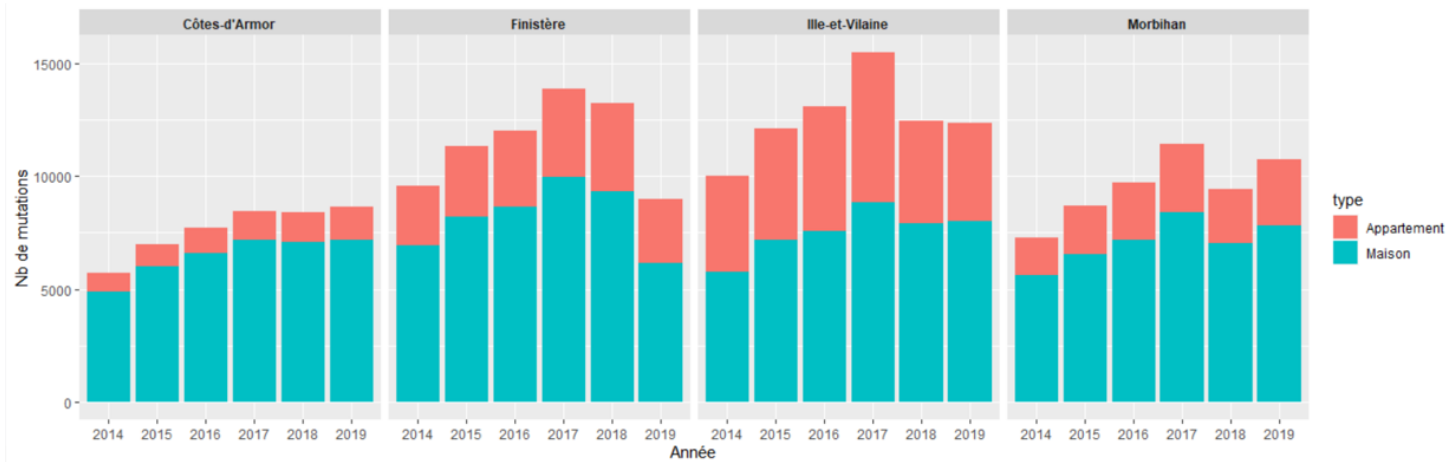
lot4_surface_carrez	lot5_numero	lot5_surface_carrez	nombre_lots	code_type_local	type_local	surface_reelle_bati	nombre_pieces_principales	code_nature_culture
NA	NA	NA	0	NA		NA	NA	AG
NA	NA	NA	0	NA		NA	NA	AB
NA	NA	NA	0	NA		NA	NA	AB
NA	NA	NA	1	3	DÃ©pendance	NA	0	
NA	NA	NA	1	3	DÃ©pendance	NA	0	
NA	NA	NA	1	2	Appartement	40	2	
NA	NA	NA	0	NA		NA	NA	T
NA	NA	NA	1	3	DÃ©pendance	NA	0	
NA	NA	NA	2	2	Appartement	40	2	
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	L
NA	NA	NA	0	NA		NA	NA	T



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	Tréguieux	Côtes-d'Armor	CA Saint-Brieuc Armor Agglomération	Pôle urbain	48.47386	-2.754416

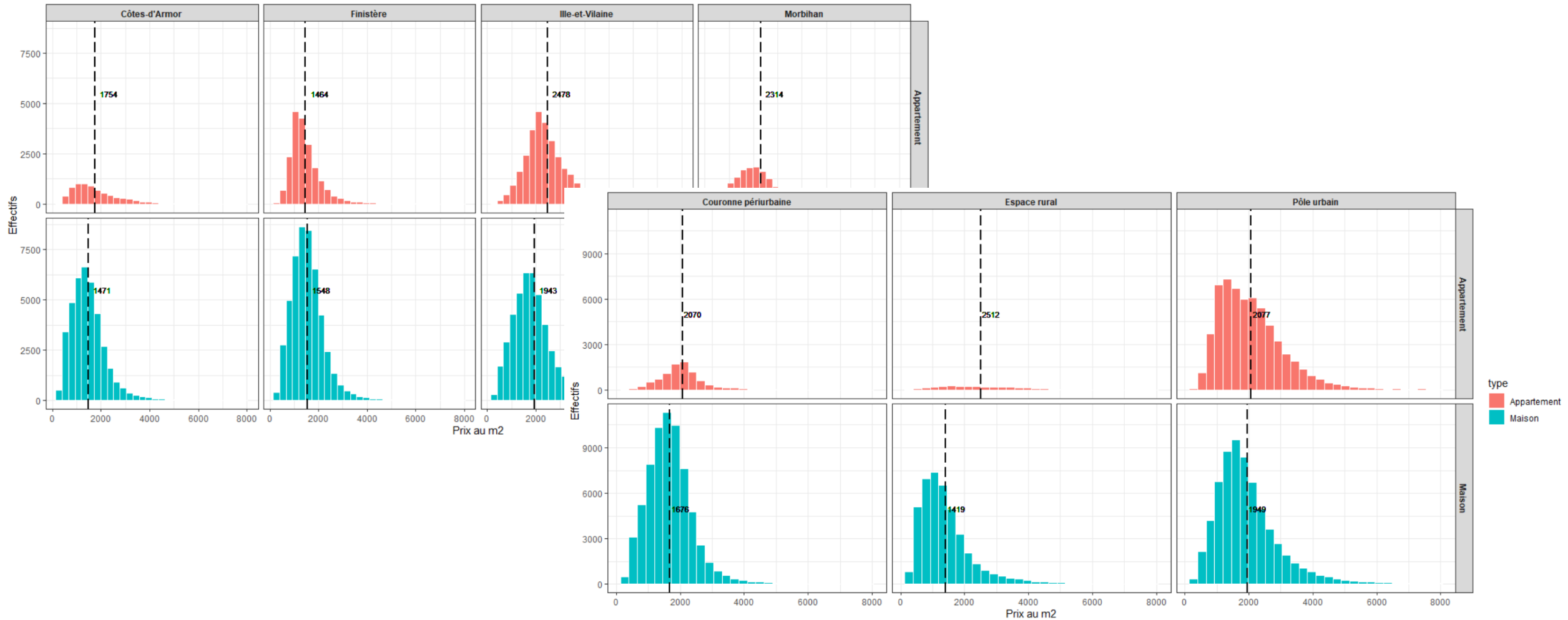
Exploration graphique

Volumes des transactions



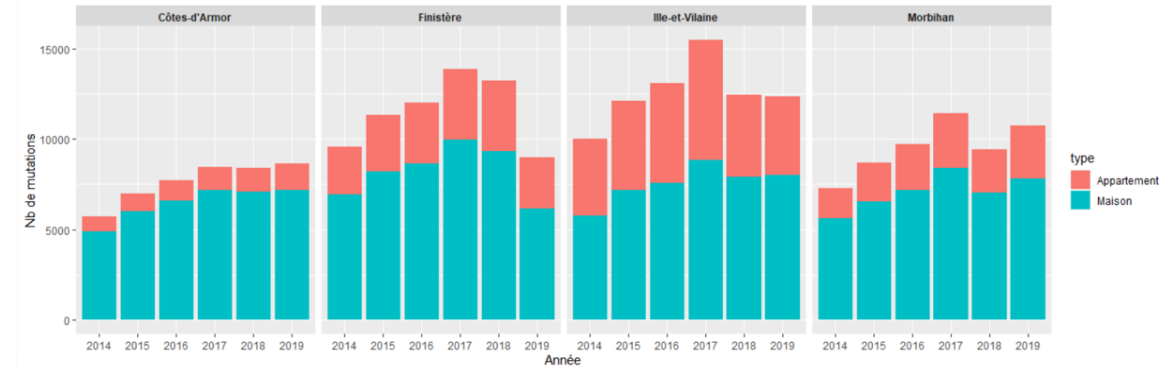
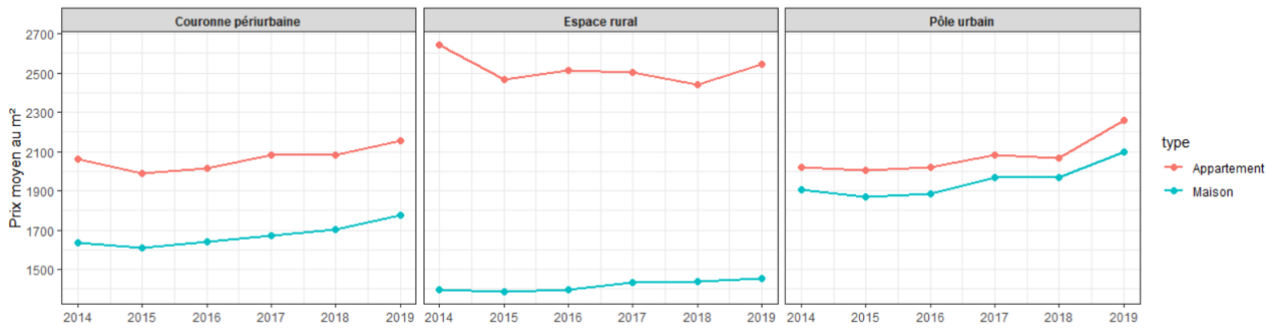
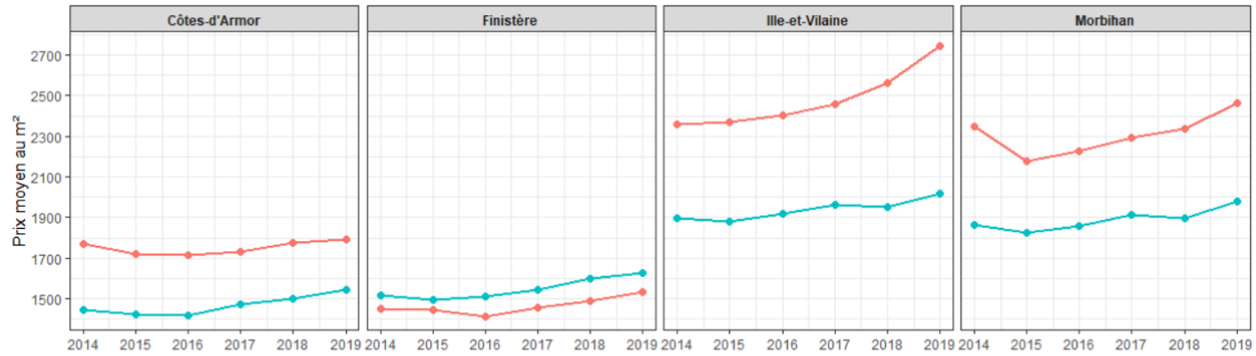
Exploration graphique

Volumes et prix au m² des transactions



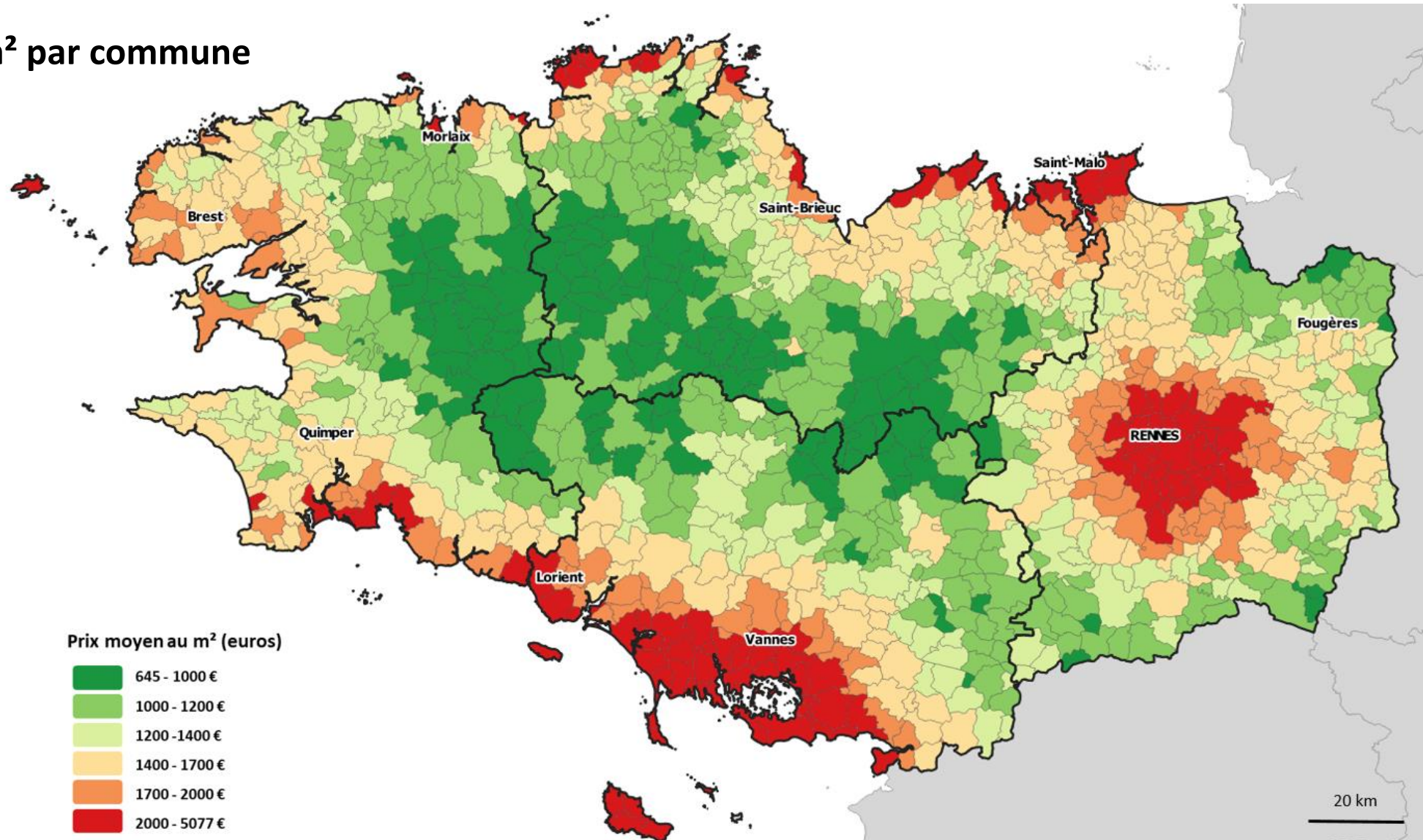
Exploration graphique

Evolutions volumes et des prix au m² des transactions



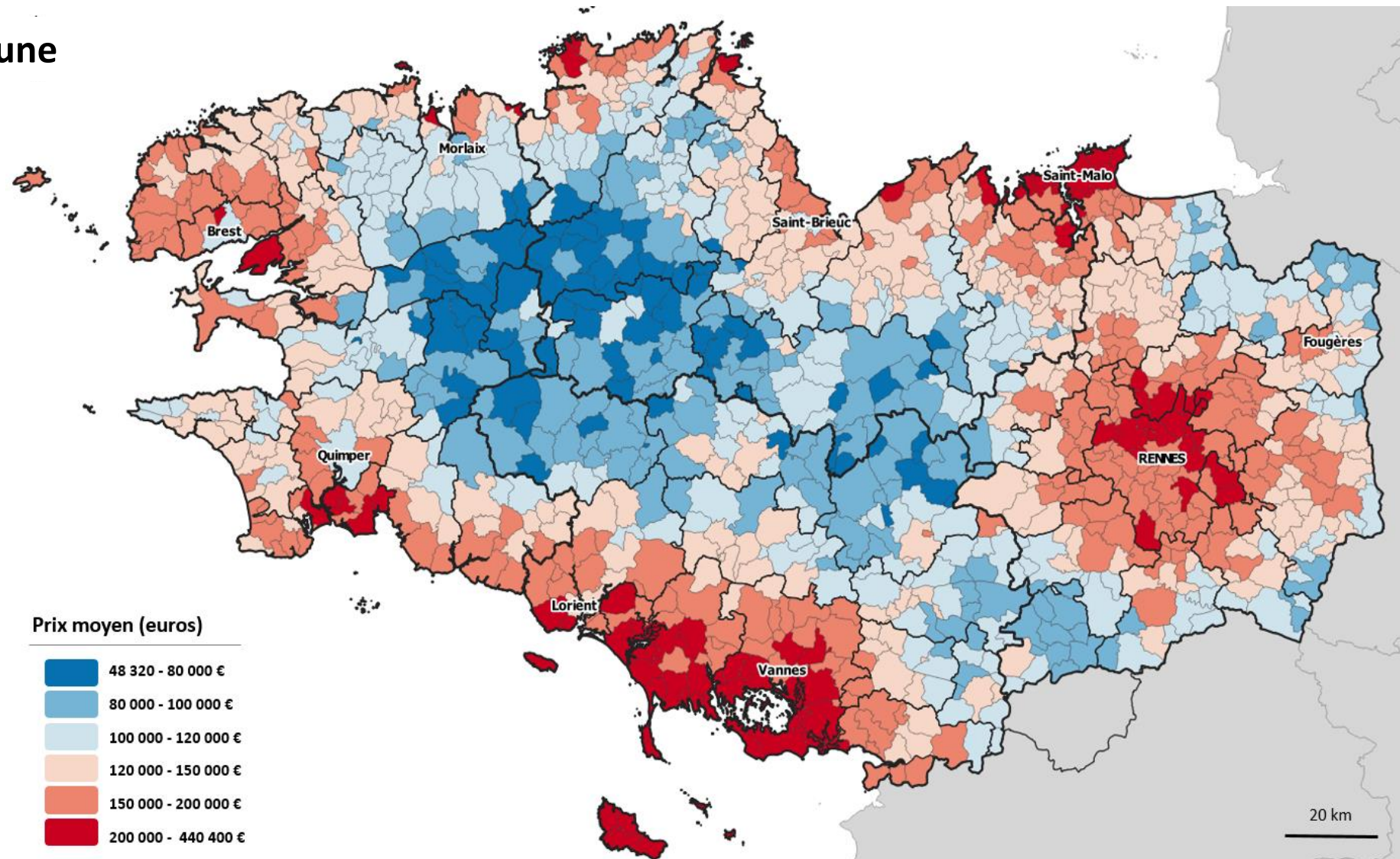
Exploration spatiale

Prix moyens au m² par commune



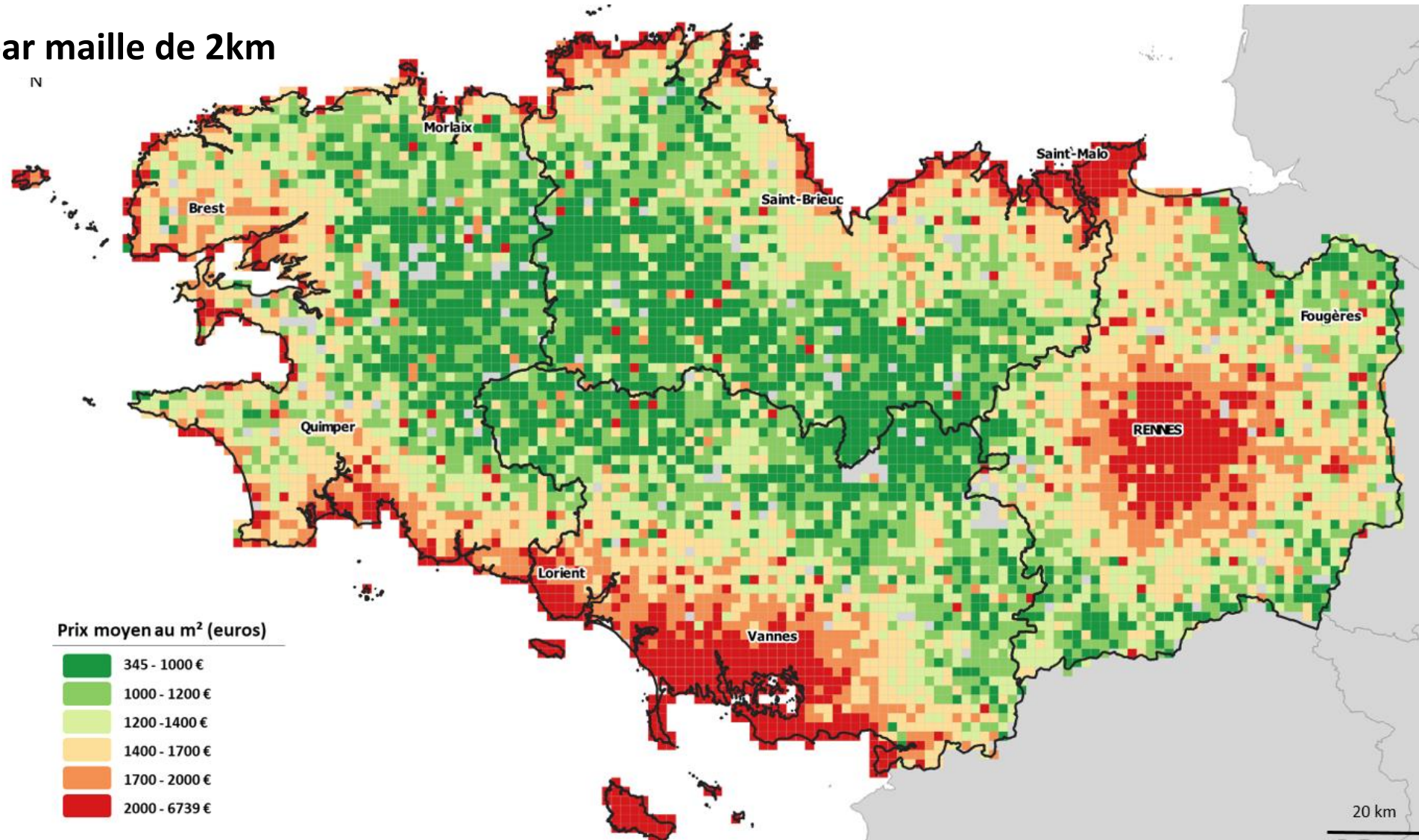
Exploration spatiale

Prix moyen par commune



Exploration spatiale

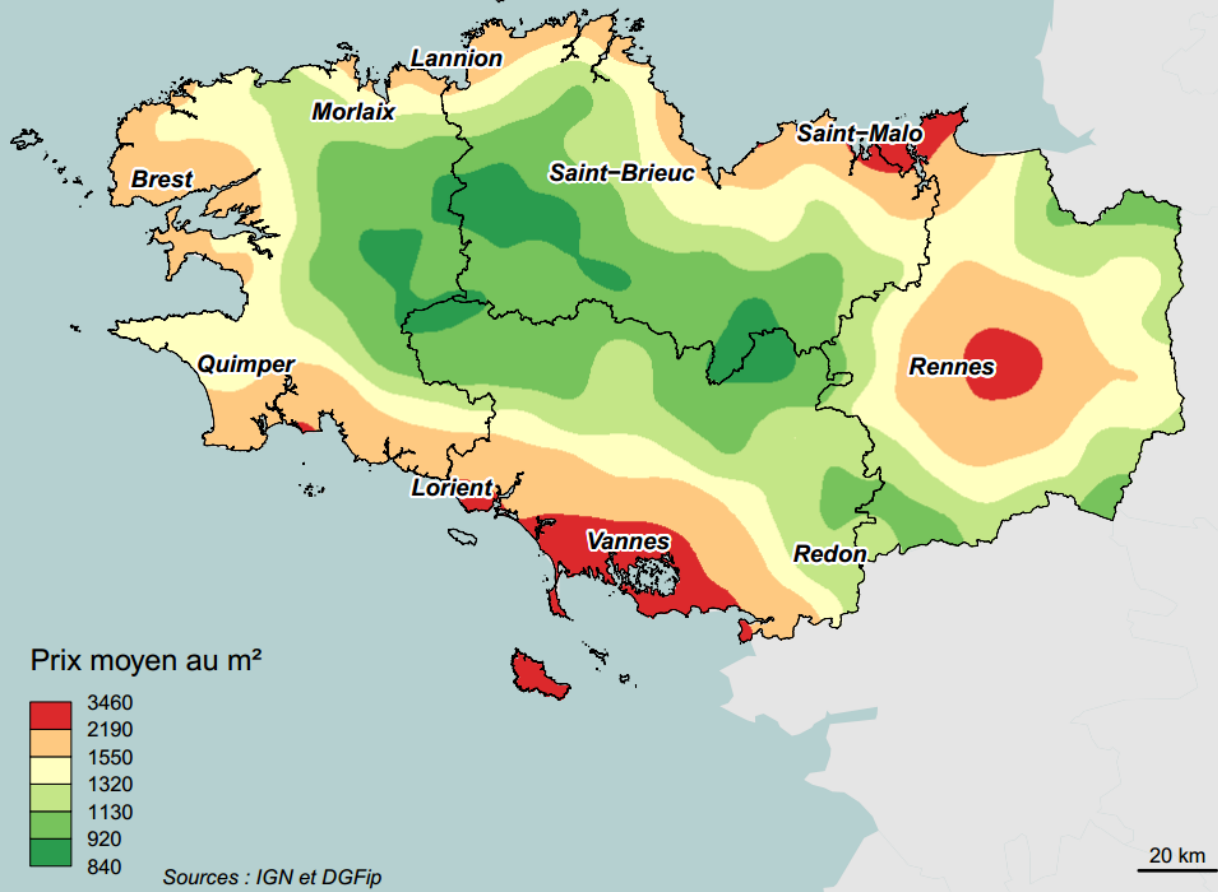
Prix moyen au m² par maille de 2km



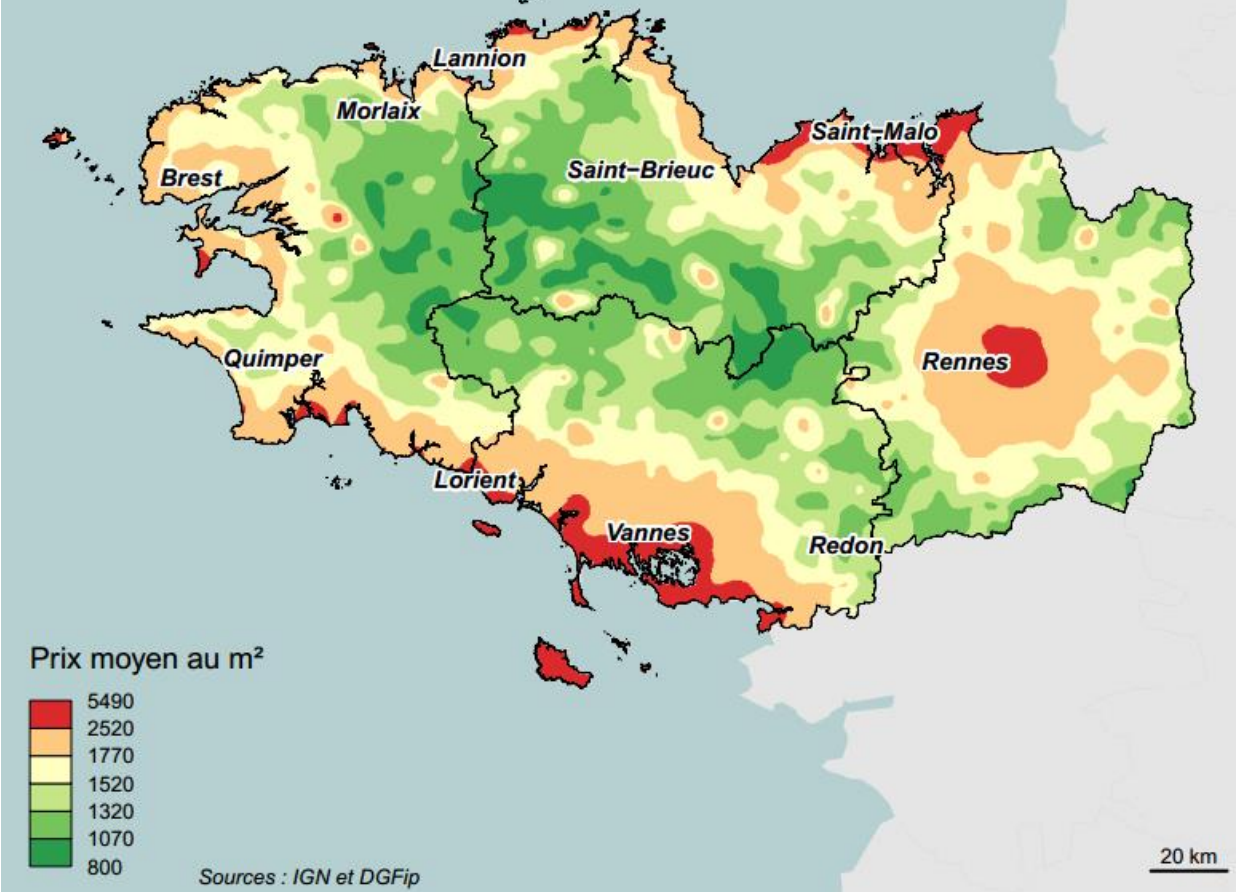
Exploration spatiale

Lissage spatial des prix au m²

Mutations agrégées



Mutations individuelles



Exploration spatiale

Cartogramme et lissage spatial des prix au m²

Ici à l'échelon des EPCI

Permet de croiser volume des mutations et
logique de prix

Package **cartogram**^R

November 17, 2021

Version 1.0-5

Date 2021-11-17

Title Continuous Cartogram

Description Procedures for making continuous cartogram. Procedures available are:
flow based cartogram (Gastner & Newman (2004) <doi:10.1073/pnas.0400280101>),
fast flow based cartogram (Gastner, Seguy & More (2018) <doi:10.1073/pnas.1712674115>),
rubber band based cartogram (Dougenik et al. (1985)
<doi:10.1111/j.0033-0124.1985.00075.x>).

Depends R (>= 3.5.0)

Imports sf, data.table

Suggests lwgeom

SystemRequirements FFTW (>=3.3.1); possible package: fftw-devel (rpm),
libfftw3-dev (deb) or fftw (brew).

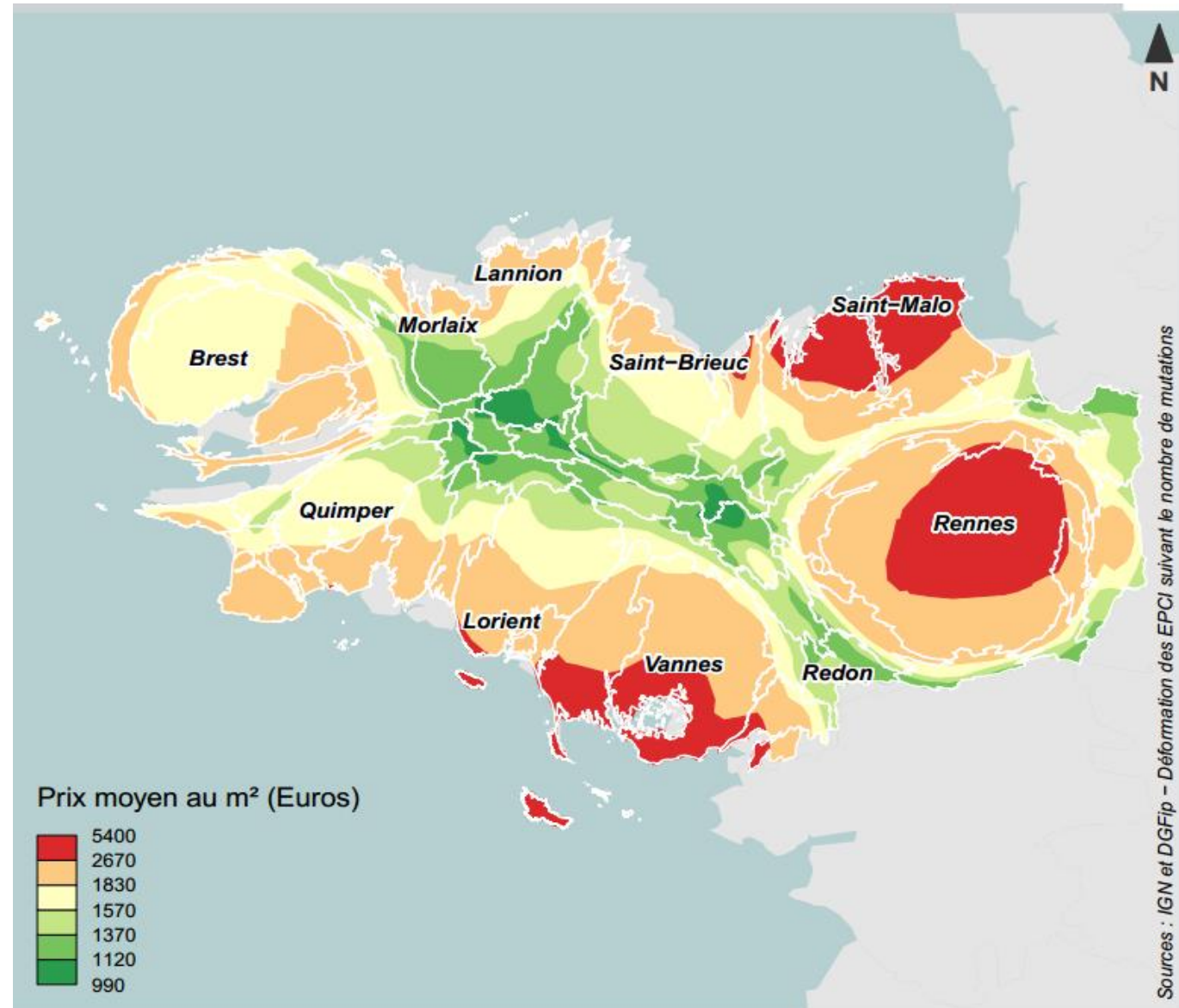
License MIT + file LICENSE

RoxygenNote 7.1.1

Encoding UTF-8

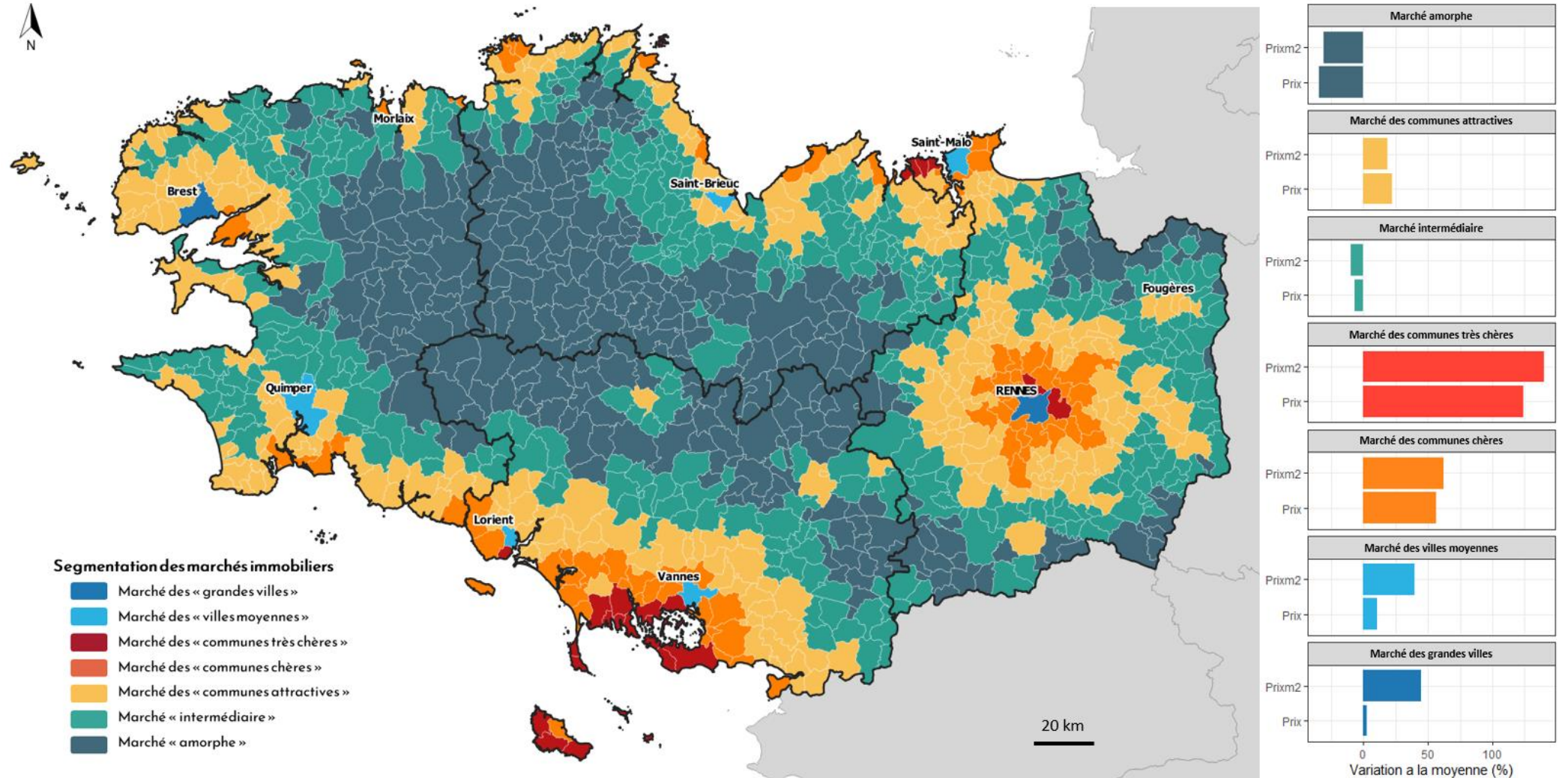
NeedsCompilation yes

Author Pierre-Andre Cornillon [aut, cre],
Florent Demoraes [aut],
Flow-Based-Cartograms [cph] (Author of core C code for gsm and gn
procedures)

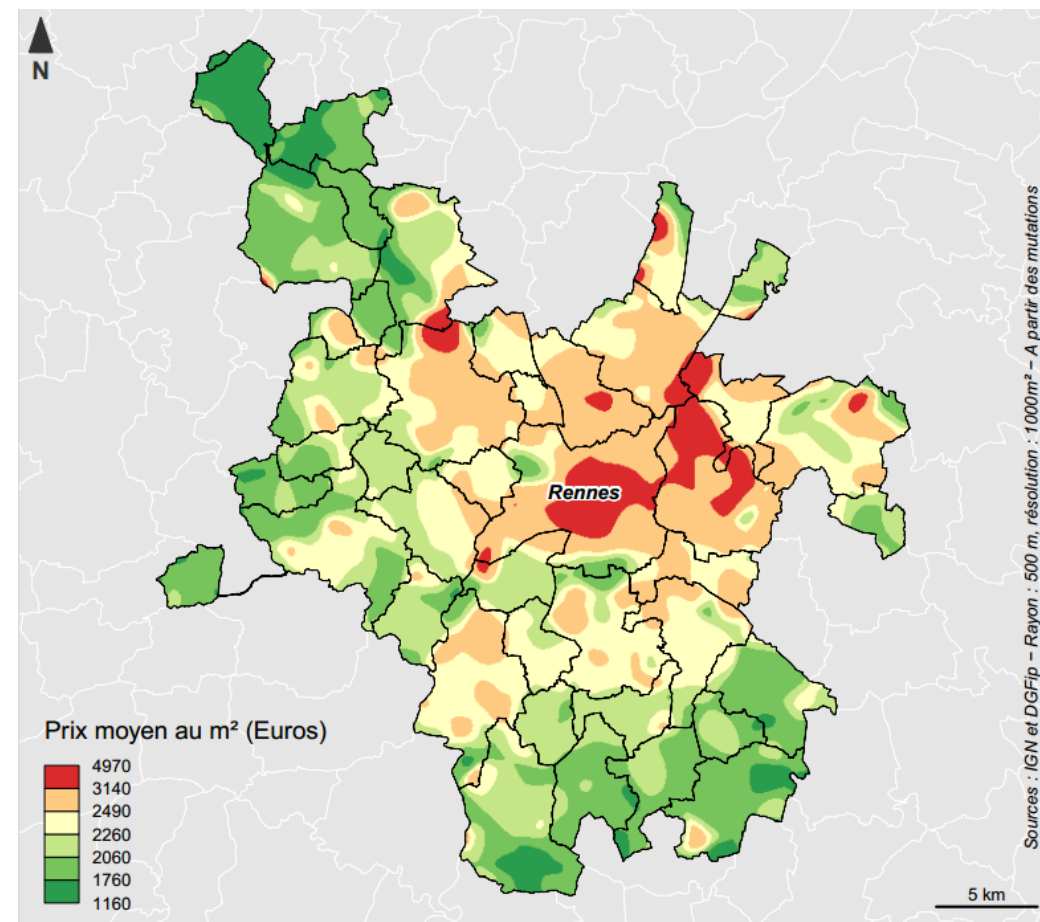
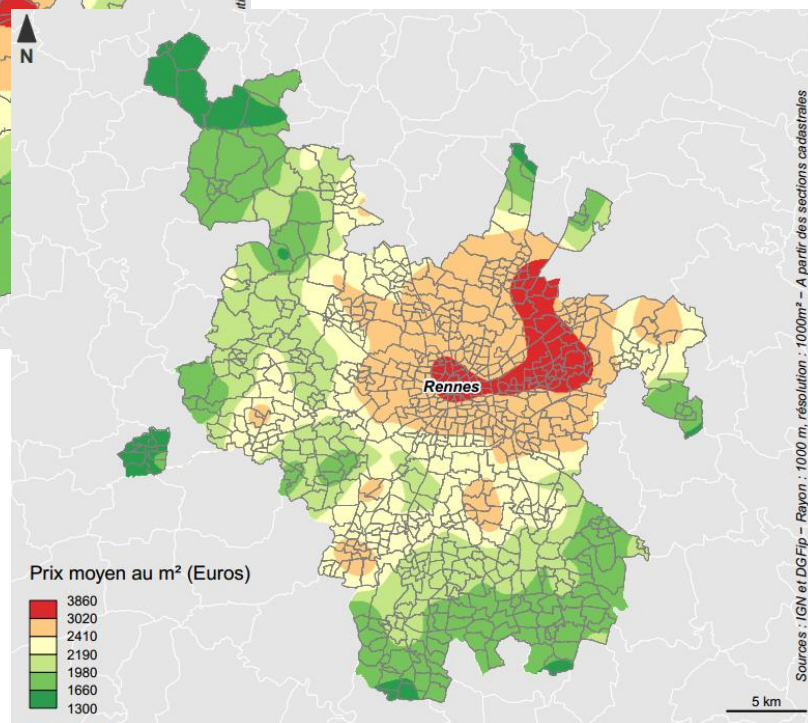
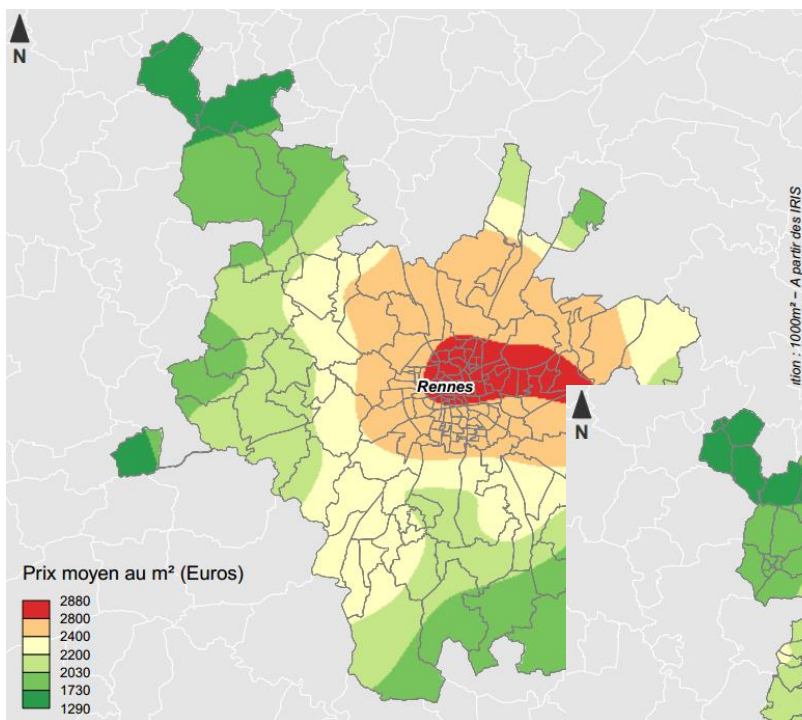


Exploration spatiale

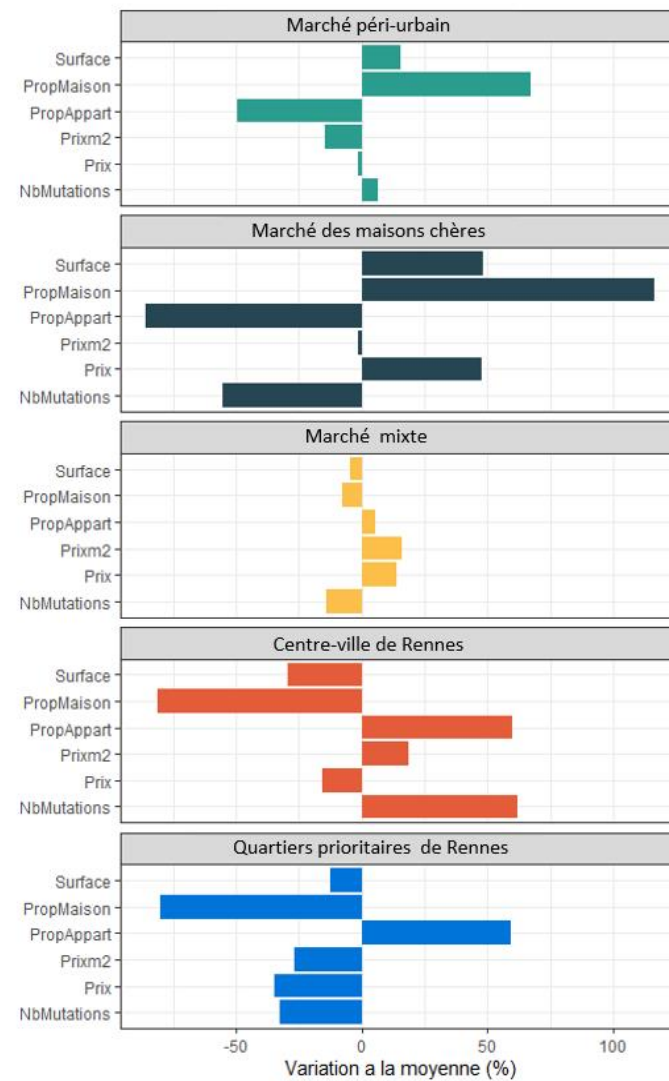
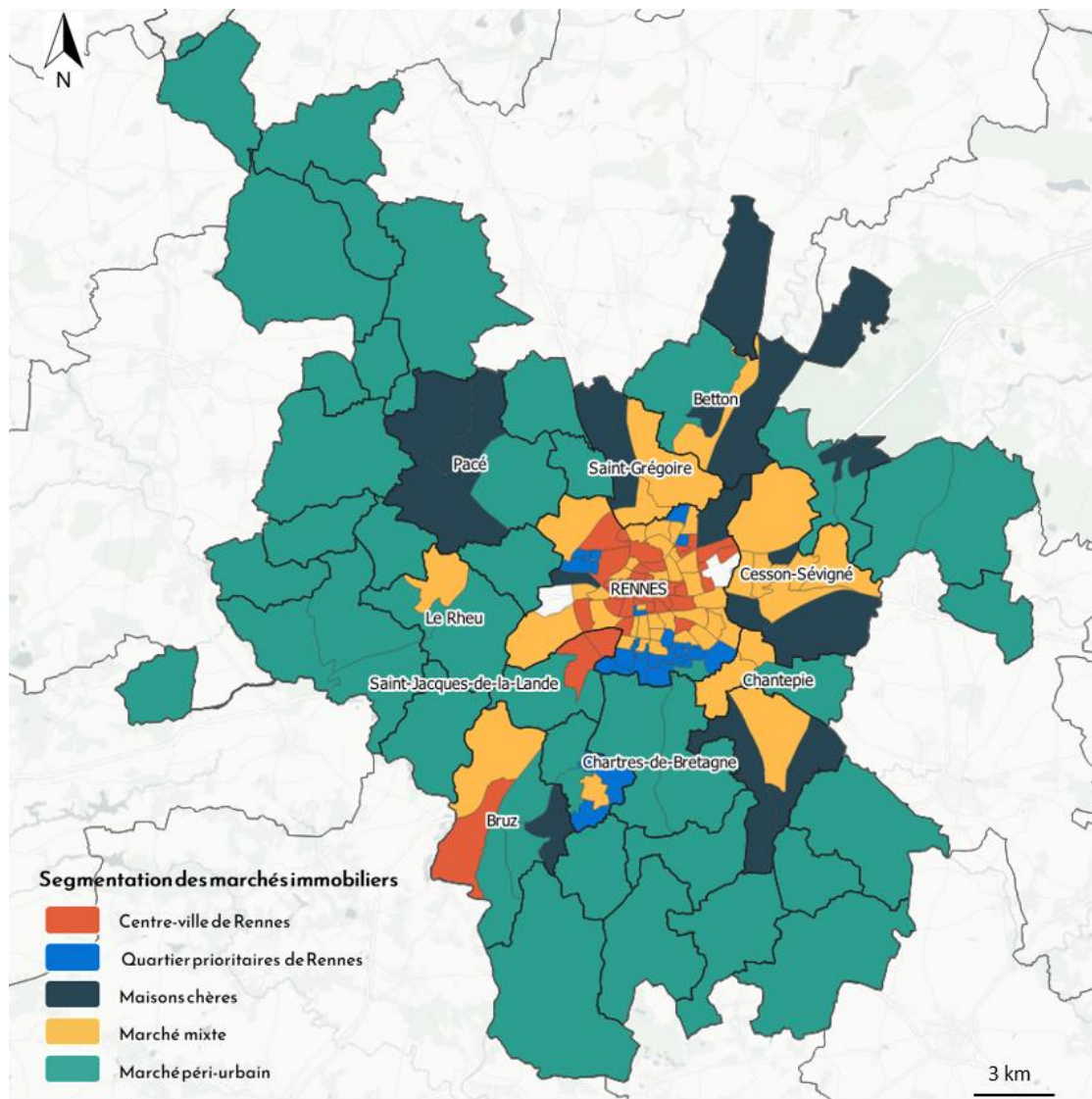
CAH pour identifier des sous marchés



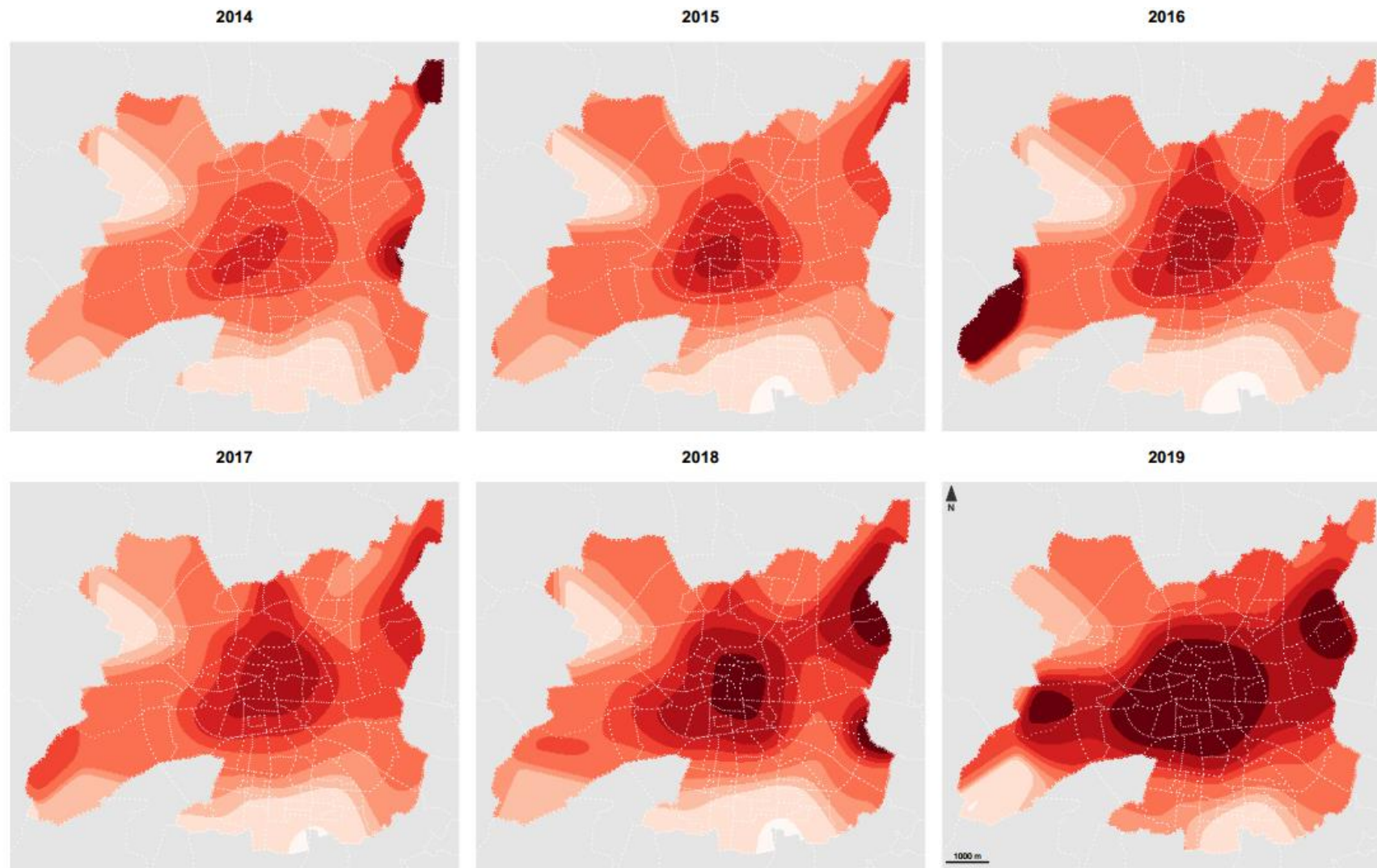
Changer d'échelle : la métropole



Changer d'échelle : la métropole



Changer d'échelle : la ville (commune de Rennes)



Prix moyen au m² des appartements à Rennes



Perspectives

- Des potentialités intéressantes pour observer les dynamiques de l'immobilier et du foncier dans le temps et dans l'espace
- Un « ticket d'entrée » méthodologique et technique non négligeable
- Besoin de travaux exploratoires
 - Une méthodologie de préparation et d'analyse en libre accès
 - Une réflexion sur la granularité des données et les échelles d'analyse
 - Une exploration autour des modes de représentation (carto)graphiques
- De nombreuses perspectives thématiques (foncier, PLU, INSEE,...)
- Une source complémentaire aux données des notaires
- Une appropriation future par les chercheurs ?

