



HAL
open science

PACTES Chaleur : construction d'un outil cartographique d'identification des populations vulnérables à la chaleur

François Carbonnel, Grégoire Mercier, Elise Chin, Layana Caroupaye-Caroupin, Emilie Agrech, Camille Debrock, Imen Kadri, Mireille Abraham, Paul Basso-Bert, Manal Ahikki

► To cite this version:

François Carbonnel, Grégoire Mercier, Elise Chin, Layana Caroupaye-Caroupin, Emilie Agrech, et al.. PACTES Chaleur : construction d'un outil cartographique d'identification des populations vulnérables à la chaleur. Commissariat général au développement durable (CGDD). 2023, pp.gd4h.ecologie.gouv.fr/defis. hal-04260855

HAL Id: hal-04260855

<https://hal.science/hal-04260855v1>

Submitted on 26 Oct 2023

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.



Open licence - etalab

Green Data for Health

Data papers du Challenge Green Data for Health

PACTES Chaleur : construction d'un outil cartographique d'identification des populations vulnérables à la chaleur

Octobre 2023

Mots clés

Santé environnementale, vague de chaleur, canicule, approche populationnelle, atlasanté, sirsé

Environmental health, hot temperature, extreme heat, population-based approach, atlasanté, sirsé

Auteurs

François Carbonnel^{1,2,3}, Grégoire Mercier^{1,4,5}, Elise Chin⁶, Layana Caroupaye-Caroupin⁶, Emilie Agrech⁶, Camille Debrock⁶, Imen Kadri⁶, Mireille Abraham³, Paul Basso Bert³, Manal Ahikki⁷

François Carbonnel et Grégoire Mercier ont contribué de façon équivalente à ce travail.

¹ Institut Desbrest d'Epidémiologie et de Santé Publique, UMR UA11, University of Montpellier-INSERM, Montpellier, France

² MSPU Avicenne, Cabestany, France

³ University Department of General Practice, UFR Médecine Montpellier-Nîmes, University of Montpellier, Montpellier, Montpellier, France

⁴ Service de santé publique, CHU de Montpellier, Montpellier, France

⁵ KanopyMed, Montpellier, France

⁶ Bénévole dans le cadre du Challenge GD4H

⁷ CHU de Montpellier

SOMMAIRE

- Résumé
- Éléments de contexte et motivation
- Description des données
- Intérêt des données
- Collecte et traitement des données
- Plus-value et perspectives
- Bibliographie

Résumé

Le changement climatique augmente la fréquence, l'intensité, la durée et l'amplitude temporelle des vagues de chaleur. Leur impact sur la santé est donc un enjeu majeur. Au-delà des facteurs de risques individuels, la vulnérabilité des populations face aux vagues de chaleur permet d'orienter des politiques publiques d'adaptation et de résilience. C'est pour cette raison que l'équipe PACTES Chaleur a défini un **indice de vulnérabilité à la chaleur** qui agglomère des données à l'échelon de la commune pour la France métropolitaine sur la température au sol, la défavorisation sociale, l'équipement en climatiseurs et la couverture végétale. Ce data paper décrit ces données, leur intérêt scientifique et les conditions techniques et juridiques de cette réutilisation.

Éléments de contexte et motivation

Contexte

En décembre 2022, des premières discussions entre la Maison de santé pluriprofessionnelle universitaire Avicenne et le Département universitaire de médecine générale de la Faculté de médecine de Montpellier-Nîmes ont porté sur la construction d'un outil cartographique de diagnostic territorial environnemental de santé. L'objectif était de pouvoir identifier les déterminants de santé d'origine environnementale autour d'une future communauté professionnelle territoriale de santé (CPTS) nommée Ruscino, dans la perspective de l'élaboration de son projet de santé.

Rapidement, cet outil a été nommé PACTES pour Pacte Cartographie Territorial Environnemental en Santé, en écho au rapport « Dessiner la santé publique de demain » du Haut conseil de la santé publique qui invitait à redéfinir les liens entre santé publique et médecine clinique (Chauvin 2022).

Le pôle santé environnementale de la Direction de la Santé Publique de l'ARS Occitanie a ensuite suggéré de participer au challenge Green Data For Health (GD4H).

Nous avons alors créé une équipe projet, emmenée par le Dr. Grégoire Mercier, médecin de santé publique au CHU de Montpellier, particulièrement investi sur la question des vagues de chaleur et le Dr. François Carbonnel, constituée par Mme Portero-Espert (DT-ARS66), Mme Karen Asnar (IREPS 66) et l'équipe de la MSPU Avicenne avec Karolina Griffiths pour le DUMG, en collaboration avec Dataactivist.

Très rapidement, nous avons décidé de nous restreindre à un déterminant de santé pertinent avec le territoire des Pyrénées-Orientales, les vagues de chaleur. Il avait été en effet noté par Santé publique France une surmortalité de 32,4 % sur ce département pendant l'été 2022 (Santé publique France 2022). Nous avons pu nous appuyer sur des travaux publiés sur la susceptibilité à la chaleur des quartiers de la métropole de New York (Madrigano 2015).

Nous avons également mis en place un partenariat avec Atlasanté, qui développe des outils publics de cartographie, dont un dédié à la santé environnementale, SIRSE.

Le projet, alors nommé PACTES Chaleur, a été mis en œuvre avec une équipe de bénévoles via le GD4H et Dataactivist, composée de Data Scientists (Elise Chin, Imen Kadri), de Data Analysts (Layana Caroupaye Caroupin), de géomaticiennes (Émilie Agrech et Camille Debrock), de Manal Ahikki, impliquée dans l'équipe de Grégoire Mercier et deux internes en médecine générale (Paul Basso-Bert et Mireille Abraham), pendant 6 semaines.

Motivation

En raison du changement climatique, la fréquence des jours de chaleur extrême devrait augmenter et ces événements sont associés à une augmentation de la morbidité et de la mortalité. Bien que la plupart des recherches sur le changement climatique aient été menées à grande échelle, il est de plus en plus reconnu que les schémas de vulnérabilité et les stratégies d'adaptation correspondantes sont plus utiles à conceptualiser au niveau local. Pour allouer efficacement les ressources, les acteurs et décideurs locaux doivent comprendre quelles sont les caractéristiques de la population et des quartiers qui augmentent la vulnérabilité (Madrigano 2015).

Des analyses cartographiques s'intéressent généralement aux déterminants physiques du risque de chaleur (par exemple l'exposition à la température et la durée), aux caractéristiques de vulnérabilité qui devraient contribuer à la morbidité ou à la mortalité liées à la chaleur sur la base d'études épidémiologiques antérieures ou aux deux à la fois. Cependant, ces analyses et études sont circonscrites à des métropoles à l'extérieur de la France.

En France, aucun outil d'analyse de la susceptibilité de la population *a priori* à une échelle géographique fine n'existe.

C'est pour cette raison que nous avons défini un « **indice de vulnérabilité à la chaleur** », créé en compilant des données à l'échelon de la commune pour la France métropolitaine. Cet indice agglomère des données :

- de température au sol d'origine satellitaire (Agence spatiale européenne) à l'été 2022 ;
- de défavorisation sociale via FDep ;
- d'équipement en climatiseurs via les diagnostics de performance énergétique lors des achats/ventes/locations immobilières ;
- de couverture végétale via le CORINE Land Cover.

Cet indice se basait sur une méthodologie appliquée à la ville de New-York (Madrigano 2015).

L'objectif de ce data paper est de décrire ces données et leur intérêt scientifique afin d'en susciter des réutilisations, et de détailler les conditions techniques et juridiques de cette réutilisation.

Description des données

L'ensemble des données sont disponibles sur GitLab dans le dossier data/ :

https://gitlab.com/data-challenge-gd4h/pactes-chaleur/-/tree/main/data/processed?ref_type=heads

Description des fichiers

L'indice de vulnérabilité face aux vagues de chaleur agglomère quatre facteurs sociaux et environnementaux qui contribuent au risque de chaleur dans les communes, à savoir :

- la température au sol ;
- l'accès à l'air conditionné dans les domiciles ;
- les espaces verts ;
- l'indice de défavorisation sociale.

Le jeu de données, décrit dans le tableau 1, comprend :

- le fichier de référentiel des communes ;
- quatre fichiers, un par indicateur ;
- un fichier regroupant tous les indicateurs, ainsi que leur z-scores ;
- le fichier d'indice de vulnérabilité ;

Green Data for Health
Data papers Challenge Green Data for Health

- le fichier contenant les données de morbidité.

Dans le cadre du Challenge GD4H, le référentiel des communes de l'Insee de l'année 2022 a été utilisé, auquel a été supprimé les territoires d'outre-mer. Il y a ainsi un total de 34 869 communes.

Tableau 1 : description des caractéristiques principales de chaque fichier de données

Nom du fichier	Description	Nombre de lignes	Intitulé des variables
comm_arr_hors_dom_2022.csv	Référentiel des communes de la France métropolitaine (Insee 2022).	34 869	libelle_commune_insee, code_commune_insee
comm_arr_hors_dom_2022_temperature_moyenne_ete_C.csv	Données de température au sol d'origine satellitaire (Agence spatiale européenne) à l'été 2022.	34 869	libelle_commune_insee, code_commune_insee, temperature_moyenne_ete_C
comm_arr_hors_dom_2022_batiments_avec_climatisation_pourcentage.csv	Données des résidences principales équipées en climatisation	34 877	libelle_commune_insee code_commune_insee total_batiments batiments_sans_climatisation batiments_avec_climatisation batiments_avec_climatisation_pourcentage
comm_arr_hors_dom_2022_tx_cv.csv	Données de couverture végétale (CORINE Land Cover)	33 552	code_commune_insee tx_cv
comm_arr_hors_dom_2022_fdep15.csv	Données d'indice de défavorisation sociale de 2015 pour la France métropolitaine.	34 869	libelle_commune_insee code_commune_insee fdep15
comm_arr_hors_dom_2022_all_indicators.csv	Données regroupant l'ensemble des indicateurs et leur z-scores	33 549	libelle_commune_insee code_commune_insee fdep15 batiments_avec_climatisation_pourcentage temperature_moyenne_ete_C tx_cv z_score_fdep15

Green Data for Health
Data papers Challenge Green Data for Health

			z_score_batiments_avec_climatisation_pourcentage z_score_temperature_moyenne_ete_C z_score_tx_cv
comm_arr_hors_dom_2022_indic_vuln.csv	Données d'indice de vulnérabilité	33 549	libelle_commune_insee code_commune_insee indic_vuln
morbidity.csv	Données de morbidités	523	code_commune_insee, nb_iCanicule, nb_hospits_iCanicule

Description des variables

Dans la suite, nous ne décrivons les colonnes (libelle_commune_insee, code_commune_insee) que pour le premier fichier. Ces colonnes sont identiques dans tous les fichiers.

Tableau 2 : dictionnaire des variables

Nom du fichier	Nom variable	Type de variable	Description
comm_arr_hors_dom_2022.csv	libelle_commune_insee	String	Nom de la commune selon le référentiel de l'Insee
	code_commune_insee	String	Code Insee de la commune
comm_arr_hors_dom_2022_temperature_moyenne_ete_C.csv	temperature_moyenne_ete_C	Float	Température moyenne de l'été 2022 (juin, juillet, août), en °C
comm_arr_hors_dom_2022_batiments_avec_climatisation_pourcentage.csv	total_batiments	Integer	Nombre total de bâtiments dans la commune
	batiments_sans_climatisation	Integer	Nombre de bâtiments sans climatisation

Green Data for Health
Data papers Challenge Green Data for Health

	batiments_avec_climatisation	Integer	Nombre de bâtiments avec climatisation
	batiments_avec_climatisation_pourcentage	Float	Pourcentage de bâtiments avec climatisation, rapport entre le nombre de bâtiments avec climatisation et le nombre de bâtiments total
comm arr hors dom 2022 tx cv.csv	tx_cv	Float	Taux de couverture végétale
comm arr hors dom 2022 fdep15.csv	fdep15	Float	Indice de défavorisation sociale de 2015
comm arr hors dom 2022 all indicateurs.csv	z_score_fdep15	Float	Z-score de l'indice de défavorisation sociale de 2015
	z_score_batiments_avec_climatisation_pourcentage	Float	Z-score du pourcentage de bâtiments avec climatisation
	z_score_temperature_moyenne_ete_C	Float	Z-score de la température moyenne de l'été 2022
	z_score_tx_cv	Float	Z-score du taux de couverture végétale
comm arr hors dom 2022 indic vuln.csv	indic_vuln	Float	Indice de vulnérabilité à la chaleur, correspond à la somme des z-scores des quatre indicateurs
morbidity.csv	nb_iCanicule	Integer	Somme des passages aux urgences sur la période d'étude de 2012 à 2022
	nb_hospits_iCanicule	Integer	Somme des passages aux urgences suivis d'une hospitalisation sur la

			période d'étude de 2012 à 2022
--	--	--	--------------------------------

Intérêt des données

Intérêt scientifique des données

Il s'agit de la première étude en France estimant de manière exhaustive la vulnérabilité à la mortalité liée à la chaleur à une échelle géographique fine.

Les perspectives de réutilisation scientifique sont nombreuses :

- enrichissement et perfectionnement du score avec d'autres sources de données ;
- validation externe en analysant la corrélation éventuelle avec des données mesurant l'impact réel observé des vagues de chaleur sur la santé (mortalité et morbidité) ;
- approche incluant une comparaison internationale avec d'autres pays.

Autres sources d'intérêt

Cet outil peut être utilisé par les équipes locales de soins primaires (MSP et CPTS) dans la préparation et dans l'évaluation de leur projet médical. Il peut aussi être intégré par différents organismes publics (ARS, CPAM, DRSM...) dans l'accompagnement des équipes de soins primaires. Enfin, il a toute sa place dans la préparation et le suivi des plans régionaux santé environnement.

Les résultats de ce travail sont utiles pour orienter les efforts de prévention des décès liés à la chaleur, y compris les mesures de planification urbaine, les messages publics pendant les vagues de chaleur, la fourniture de climatiseurs et d'autres actions géographiquement ciblées.

Collecte et traitement des données

Collecte et processus de traitement mis en œuvre

Nous avons tout d'abord adapté le [référentiel de communes Insee](#).

Afin de créer l'indice de vulnérabilité, nous avons dans un premier temps collecté et traité les données des quatre indicateurs afin de calculer par la suite un z-score pour chaque indicateur par commune.

1. Indice de défavorisation

Le jeu de données brut est disponible sur le site cepidc.inserm.fr et le code sur [GitLab](https://gitlab.com).

La base de données contient l'indice de défavorisation sociale (FDep) de 2015 pour la France métropolitaine. Les limites géographiques communales utilisées sont datées au 1^{er} janvier 2017. Le fichier comporte 35 322 lignes et quatre variables.

Les variables contiennent les informations suivantes :

- CODGEO : Code commune Insee ;
- P15_POP : population communale au 1^{er} janvier 2015 ;
- FDep15 : score d'indice de défavorisation sociale, plus le score est élevé plus le désavantage social est important ;
- Q5 : classification de l'indice de défavorisation sociale en quintiles allant du moins désavantagées (Q5 = 1) au plus désavantagées (Q5 = 5).

Les variables utilisées pour la suite de l'étude sont : CODGEO et FDep15.

Traitements des données

La variable CODGEO a été renommée 'code_commune_insee'. Les données ont été ensuite fusionnées avec le référentiel des communes de l'Insee de l'année 2022. À la suite de cette fusion, nous constatons neuf données manquantes du score FDep15 pour neuf communes.

Pour résoudre ce problème, nous avons imputé les données manquantes par la moyenne du score de défavorisation des communes les plus proches de la commune manquante.

2. Couverture végétale

Deux jeux de données sont nécessaires pour le calcul du taux de couverture végétale :

- la [base de données Corine Land Cover \(CLC\)](#), disponible sur le site du gouvernement ;

- la base de données Georef qui recense les communes de France métropolitaine, en veillant à ne pas prendre en compte les départements d'outre-mer.

Pré-traitement des données

Les données CLC doivent être préparées avant traitement afin de minimiser la taille de la couche et par conséquent le temps de traitement. Pour ce faire, il est nécessaire de sélectionner les classes à garder :

- 141. Espaces verts urbains
- 222. Vergers et petits fruits
- 223. Oliveraies
- 231. Prairies et autres surfaces toujours en herbe à usage agricole
- 244. Territoires agroforestiers
- 311. Forêts de feuillus
- 312. Forêts de conifères
- 313. Forêts mélangées
- 321. Pelouses et pâturages naturels
- 322. Landes et broussailles
- 323. Végétation sclérophylle
- 324. Forêt et végétation arbustive en mutation
- 333. Végétation clairsemée

Pour réaliser cette opération, il faut d'abord catégoriser la couche selon le champ « Code_18 », puis ne sélectionner que les numéros de milieux correspondants à la liste ci-dessus. Une fois la catégorisation réalisée, effectuer une simple sélection avec un rectangle en englobant la France métropolitaine (ne pas oublier la Corse).

Traitement des données

Pour le traitement des données, afin de faciliter le traitement par le logiciel SIG QGis (open source), il est préférable de commencer par deux étapes de simplifications :

1. Les données CLC ayant un code à 3 chiffres sont à simplifier en un seul entier. Pour cette étape, nous regroupons les codes par le premier des trois chiffres (ex : 112 devient 1, 213 devient 2, etc.). Un nouveau champ dans la table attributaire est à créer. Nous avons choisi le nom « Niveau 1 ». Il prendra l'expression « to_int (left "Code_18" , 1) ».

2. La seconde étape de simplification de traitement est un « regrouper ». Elle permet de regrouper ensemble les catégories CLC par « Niveau 1 » et de limiter le temps de calcul de l'algorithme utilisé ci-après.

Finalement nous pouvons effectuer le traitement « intersection », pour pouvoir superposer les couches et les faire travailler ensemble. Cette étape permet alors d'obtenir une couche complète comportant le croisement des deux couches d'intérêt et de calculer la surface de chaque « Niveau 1 » par commune. Il faut créer, dans la table attributaire de notre nouvelle couche, un champ « Surf class » qui prendra l'expression « \$area/10000 ».

Nb : en sortie, sur ce fichier, nous obtenons plusieurs valeurs pour une seule commune puisqu'à une commune peut correspondre plusieurs catégories de couverture végétale. Il est donc nécessaire de rassembler ces valeurs pour obtenir une valeur par commune.

Calculs de statistiques

Une fois que l'on a les surfaces de milieu végétal par commune, il faut utiliser l'outil « Statistiques par catégories », en utilisant le champ du code commune. Un fichier sous forme de table attributaire sera disponible, avec un certain nombre de métriques statistiques. Ici, nous nous sommes intéressés à la somme. En effet, la somme permet d'obtenir pour chaque commune la somme de toutes les surfaces de milieu végétal.

3. Température de surface

Pour récupérer les températures de surface des communes métropolitaines françaises, nous avons utilisé les données de Copernicus, à partir de l'[API Historical Weather](#), qui permet de requêter les températures quotidiennes moyennes pour des coordonnées géographiques et une plage de date données. Le code est disponible de traitement de ces données est disponible sur [GitLab](#).

Traitement des données

Le traitement des données s'est effectué en trois étapes :

1. Extraction des départements de chaque commune à partir des codes Insee (preprocessing/department.py)

En effet pour trouver les bonnes coordonnées géographiques de chaque commune, le libellé Insee ne suffit pas, il faut également leur numéro de département (et non le code département de l'Insee) :

- Pour la majorité des communes, ce numéro correspond aux deux premiers chiffres du code commune Insee.
- Pour les communes de la Corse, on remplace « 2A » et « 2B » par « 20 ».
- Il y a une commune dont le numéro ne correspond pas aux deux premiers chiffres du code commune Insee, mais cela ne pose pas de problème dans la recherche.

2. Extraction des coordonnées géographiques de chaque commune à partir du nom de la commune et de son département (preprocessing/coordinates.py)

Pour cela, nous utilisons la librairie geopy de Python.

- Pour les geocoders, on utilise tout d'abord IGNFrance, puis BANFrance si on ne trouve rien avec le premier, puis Nominatim (OpenStreetMap) si on ne trouve rien avec les deux premiers.
- Pour les villes avec arrondissements (Paris, Lyon, Marseille), on cherche par le nom de la ville sans arrondissement et on regarde dans la liste des localisations qu'on obtient, celle qui correspond à la ville avec l'arrondissement.
- Quelques points d'attention :
 - Les geocoders donnent en sortie différents formats d'adresses. Nous avons écrit un code personnalisé pour chaque format observé, pour extraire le nom de la commune et le numéro du département.
 - Comme il existe des communes de départements différents (e.g. Héricourt dans le 62, et 70), on les différencie par le numéro du département, d'où l'impossibilité de se reposer uniquement sur le libellé de la commune pour trouver ses coordonnées géographiques.
 - Pour les villes avec arrondissements, on les trouve que si on enlève « [...] arrondissement » dans le libellé de la commune (e.g. « Paris 1^{er} arrondissement » → « Paris »). Ensuite pour bien récupérer la latitude, longitude de l'arrondissement, on fait un *matching* du libellé complet avec le libellé qu'on trouve avec geocoder.
 - Parfois dans l'adresse de geocoder, il n'y a pas de département, donc on regarde seulement le *matching* avec le libellé.

3. Requête pour récupérer des moyennes de températures quotidiennes et calcul de la moyenne estivale (compute_temperature.py)

Limite de la méthode

Cette méthode fonctionne dans le cadre du challenge, mais n'est pas scalable car nous réalisons une requête pour chaque commune et nous traitons une grande quantité de données.

Pour donner un ordre de grandeur, nous avons 34 868 communes métropolitaines au total, 30 + 31 + 30 de données quotidiennes à récupérer pour chacune d'entre elles, donc au total 3 172 988 à traiter. Cela a pris 5 heures en continu (sans erreur et arrêt de notre propre machine) et beaucoup plus longtemps en pratique.

Une solution de stockage des données et un accès à un autre fournisseur de données météo (par exemple : Météo France) permettrait de pérenniser le traitement des données de température.

4. Résidences principales équipées en climatisation

Cette analyse de données a été réalisée à partir de la base de données nationale des bâtiments (BDNB), disponible sur le site data.gouv.fr. Cette base de données combine plusieurs autres bases de données pour fournir des informations détaillées sur tous les bâtiments en France.

Les données ont été regroupées par commune pour obtenir le nombre total de bâtiments avec et sans climatisation, ainsi que le taux de présence de climatisation par commune. Le code est également disponible sur [GitLab](#).

Traitement des données

Nous avons tout d'abord téléchargé les données « BDNB - Export france - csv ». Trois tables de cette base de données ont été utilisées :

1. *Batiment_groupe* : cette table comprend l'ID du bâtiment, le code du département, le code de la commune, et le code de l'IRIS.
2. *Batiment_groupe_log_type* : cette table contient également l'ID du bâtiment, une variable booléenne indiquant la présence de climatisation, ainsi que le type de logement (appartement, maison, non résidentiel, logement collectif).
3. *Batiment_groupe_bd_topo* : cette table comprend également l'ID du bâtiment et le type d'usage du bâtiment (principal, secondaire, agricole, annexe, commerce, etc.).

Les données pertinentes ont été sélectionnées à l'aide d'une requête SQL, en croisant les informations de ces trois tables. Les critères de sélection étaient les bâtiments de type maison, appartement et logement collectif dont l'usage principal est résidentiel.

```
SELECT
  bgdrft.code_departement_insee,
  bgdrft.presence_climatisation,
  bgdrft.type_batiment,
  bgr.code_commune_insee,
  bgr.libelle_commune_insee,
  bto.l_usage_1,
  bto.l_nature,
  bto.l_usage_2
FROM
  work_db.public.bat_grp_reduit AS bgr
  INNER JOIN work_db.public.batiment_groupe_dpe_logtype   bgdrft   ON
  bgr.batiment_groupe_id = bgdrft.batiment_groupe_id
  INNER JOIN work_db.public.batiment_groupe_bdtopo_bat   bto      ON
  bgr.batiment_groupe_id = bto.batiment_groupe_id
WHERE
  type_batiment in (
    'Maison', 'Appartement', 'Logements collectifs'
  )
  AND l_usage_1 LIKE '%Résidentiel%';
```


Pour traiter les communes avec arrondissements, à savoir Paris, Lyon et Marseille, nous avons réalisé une autre manipulation :

1. Télécharger les données des bâtiments du département de la commune sur data.gouv.fr. Ce sont les fichiers « BDNB - Export dep<numero_dept> », en remplaçant numero_dept par le numéro du département de la commune (69 pour Lyon par exemple).
2. Extraire les mêmes informations que précédemment, en gardant également l'iris et la commune. Voici la requête SQL lancée pour chacune des trois communes :

```
SELECT
  bgdrft.code_departement_insee,
  bgdrft.presence_climatisation,
  bgr.code_iris,
  bgdrft.type_batiment,
  bgr.code_commune_insee,
  bgr.libelle_commune_insee,
  bto.l_usage_1,
  bto.l_nature,
  bto.l_usage_2
FROM
  work_db.public.bat_grp_reduit as bgr
  INNER JOIN work_db.public.batiment_groupe_dpe_logtype   bgdrft   ON
  bgr.batiment_groupe_id = bgdrft.batiment_groupe_id
  INNER JOIN work_db.public.batiment_groupe_bdtopo_bat   bto     on
  bgr.batiment_groupe_id = bto.batiment_groupe_id
WHERE
  type_batiment in (
    'Maison', 'Appartement', 'Logements collectifs'
  )
AND l_usage_1 LIKE '%Résidentiel%';
```

3. En utilisant la table de correspondance IRIS - commune de l'Insee, disponible à ce [lien](#), nous avons pu déduire à partir des iris les données par arrondissement.

Enfin, un dernier fichier est nécessaire pour traiter les communes de Corse. Le fichier est un filtre sur la Corse des données sur les cantons qu'on peut retrouver sur le site de l'Insee ([Code officiel géographique au 1^{er} janvier 2022 - Liste des communes](#))

Nous traitons séparément les communes de la Corse et hors Corse pour des raisons de traitement de chaînes de caractères dans les codes communes de la Corse.

Gestion des données manquantes

Après cette sélection, le nombre de communes résultant ne dépasse pas 32 000. Il reste donc environ 4 000 communes pour lesquelles les données sont manquantes.

Pour résoudre ce problème, une méthode d'imputation basée sur l'appartenance des communes à un canton spécifique a été utilisée. Les données manquantes ont été remplacées par la moyenne du canton correspondant. Par exemple, pour chaque canton, la moyenne du taux de présence de climatisation a été calculée, puis utilisée pour remplacer les données manquantes des bâtiments appartenant au même canton.

Grâce à cette méthode, un fichier de données complet contenant les données par commune a finalement été obtenu.

5. Morbidité

Les données ont été récupérées auprès de Santé publique France. Le fichier contient 6 205 lignes et 6 colonnes.

Les variables contiennent les informations suivantes :

- rge_code_etablissement : code unique du service d'urgence ;
- annee_entree : année de l'entrée du passage aux urgences ;
- nb_iCanicule : nombre de passages aux urgences pour l'indicateur chaleur ;
- nb_hospits_iCanicule : nombre de passages aux urgences suivi d'une hospitalisation pour l'indicateur chaleur ;
- Libelle : libellé de l'établissement ;
- code_insee : code Insee de la commune de l'établissement.

Traitement des données

La somme des passages aux urgences et passages aux urgences suivis d'une hospitalisation a été calculée par commune sur la période d'étude de 2012 à 2022.

Fichier final

Une fois l'ensemble des données des indicateurs traitées, l'indice de vulnérabilité par commune a été calculé. C'est la somme des z-scores s'il existe pour ces quatre indicateurs, sinon il est nul.

Conditions d'accès et de réutilisation

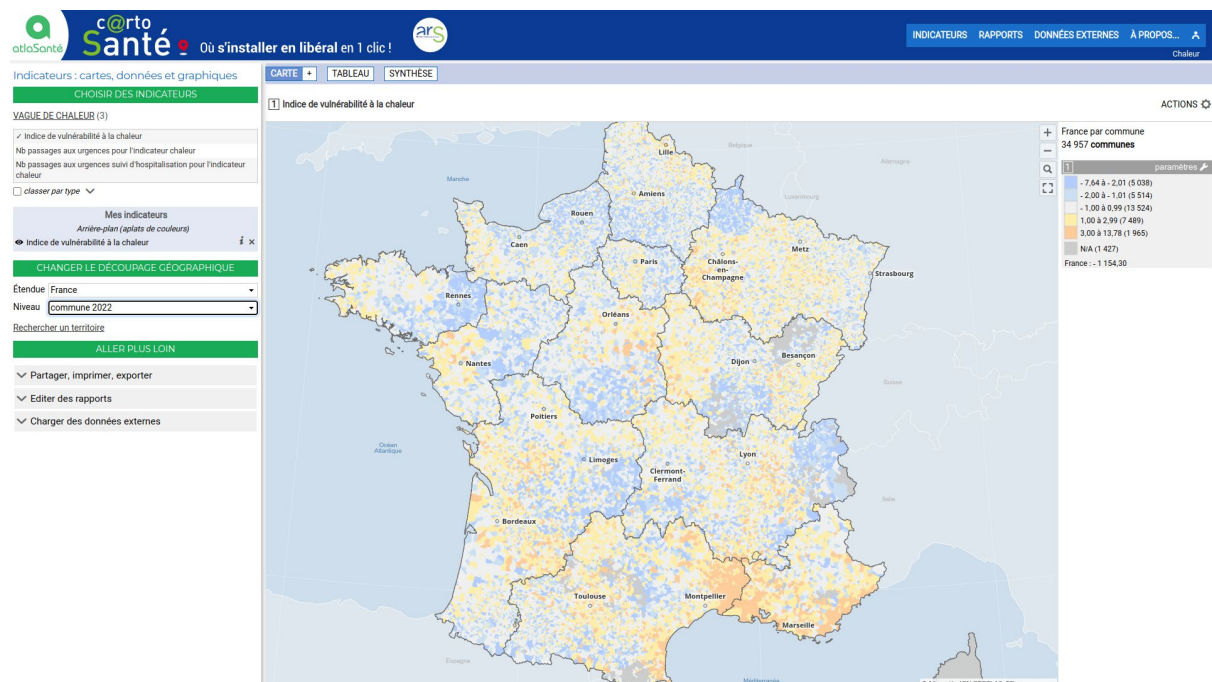
L'ensemble des ressources est en accès libre. Le code est publié sous licence [MIT](#). Les données sont publiées sous [Licence Ouverte 2.0](#).

Plus-value et perspectives

Ce travail collaboratif a permis de créer un jeu de données original, validé, exhaustif, réutilisable qui sera accessible librement sur data.gouv.fr.

En outre, trois indicateurs sont d'ores et déjà intégrés dans la base de données socio-sanitaires et environnementales territorialisée SIRSE. Cette dernière est accessible ici : https://sirse.atlasante.fr/#c=indicator&i=chaleur.indic_vuln&view=map1

Indice de vulnérabilité à la chaleur :



Nombre de passages aux urgences pour l'indicateur chaleur :

Green Data for Health

Data papers Challenge Green Data for Health

1 Nombre de passages aux urgences pour l'indicateur chaleur



Nombre de passages aux urgences suivis d'hospitalisation pour l'indicateur chaleur :



L'un des enjeux est la pérennisation de cet outil, en continuant le partenariat avec les acteurs mobilisés et le GD4H et en s'appuyant sur la modélisation en cours.

Une perspective importante est le développement d'un outil PACTES plus riche, permettant un diagnostic territorial environnemental de santé complet, adapté aux besoins des CPTS/MSP/CSP.

Bibliographie

- Chauvin 2022 : [Dessiner la santé publique de demain - Ministère de la Santé et de la Prévention.](#)
- Madrigano 2015 : [A Case-Only Study of Vulnerability to Heat Wave-Related Mortality in New York City \(2000-2011\).](#)
- Santé publique France 2022 : [Bilan canicule et santé : un été marqué par des phénomènes climatiques multiples et un impact sanitaire important.](#)



Green Data for Health

Data papers du Challenge Green Data for Health

Le Green Data for Health

Le Green Data for Health (GD4H) est une offre de service incubée au sein de l'Écolab, laboratoire d'innovation pour la transition écologique du Commissariat général au développement durable (ministère de la Transition écologique et de la Cohésion des territoires et ministère de la Transition énergétique).

Action phare du 4^e Plan national santé environnement, le GD4H vise à permettre une meilleure mobilisation et valorisation, par les acteurs de la recherche et de l'expertise, des données environnementales au service de la santé-environnement.

Pour en savoir plus : gd4h.ecologie.gouv.fr

Le Challenge Green Data for Health

Le Challenge Green Data for Health est une initiative portée par l'Écolab du Commissariat général au développement durable (ministères chargés de l'Écologie, de l'Énergie et des Territoires). Dans ce cadre, 15 équipes ont été constituées pour mener 15 projets visant à mieux mobiliser la donnée environnementale au service de la santé-environnement.

Pour en savoir plus : gd4h.ecologie.gouv.fr/defis