

# Première phase : Filtre des 3V

## Rapport d'étape

### Table des matières

Présentation.....	2
Objectif.....	2
Mise en œuvre.....	2
Identification des pôles.....	2
Traitement des données.....	3
Indice.....	4
Annexe 1 – Création des pôles générateurs.....	6
Présentation.....	6
Gares ferroviaires.....	6
Localisation.....	6
Quantification.....	6
Gares routières.....	6
Localisation.....	6
Quantification.....	6
Habitations.....	7
Localisation.....	7
Quantification.....	7
Pôles activité issus de Sirene.....	8
Écoles élémentaires, collèges et lycées.....	8
Localisation.....	8
Quantification.....	8
Enseignement supérieur.....	8
Localisation.....	9
Quantification.....	9
Annexe 2 – Traitement des données.....	10
Créer la couche tampons.....	10
Création de plusieurs tampons en SQL simple.....	10
Création de plusieurs tampons en SQL.....	10
Ajouter les champs nécessaires.....	10
Mettre à jour les champs.....	11
Découper les 3V en portion et récupérer la synthèse des données des tampons.....	11
Découpage.....	11
Ajouter l'identifiant unique.....	11
Conversion en points.....	12
Jointure spatiale.....	12
Jointure attributaire.....	12
Calcul de l'indice.....	12
Préparation.....	12
Calcul du nombre d'individus par tranche.....	12
Calcul du potentiel par distance.....	13
Calcul du potentiel final.....	13

# Présentation

## Objectif

L'objectif est de **produire un indice pour chaque portion de 3V** qui témoigne de leurs **distances** aux pôles, du **nombre d'individus** qu'ils représentent et de la **diversité** des types de pôles. Puis, en fixant un seuil à cet indice, nous pouvons filtrer les portions de 3V trop éloignées et/ou représentant trop peu d'individus.

Pour illustrer : une portion proche de pôles représentant un grand nombre d'individus aura une valeur d'indice forte ; une portion éloignée qui concerne un grand nombre d'individus pourra avoir la même valeur d'indice qu'une portion proche qui implique un faible nombre d'individus..

Nous tenons compte également de la diversité de pôles pour éviter qu'une portion impliquant un grand nombre d'individus mais essentiellement de type *habitant* et peu de type *activité* ne soit pas surestimée.

## Mise en œuvre

Après avoir localisé et quantifié les *pôles générateurs de trafic* en termes d'individus, nous avons « transféré » les informations des pôles ainsi que leurs distances aux 3V. Pour cela nous les avons au préalable découpées en **portions de 100 m**.

Pour chaque portion nous avons donc des informations sur le nombre total d'habitants (pôles *habitat*), de travailleurs/écoliers/étudiants (pôles *activité*), et d'usagers des transports en commun (pôles *transport*). Cette information du nombre d'individus est décrite par distance, soit à 2 km, 4 km, 6 km, 8 km et 10 km (nous avons défini une distance maximum de 10 km) :

portion	hab 2 km	hab 4 km	hab 6 km	hab 8 km	hab 10 km	act 2 km	act 4 km	act 6 km	act 8 km	act 10 km
1	1 500	4 800	8 600	15 000	40 000	300	700	800	1 400	2 000
2	0	900	2 500	8 100	12 300	0	0	500	900	1 100

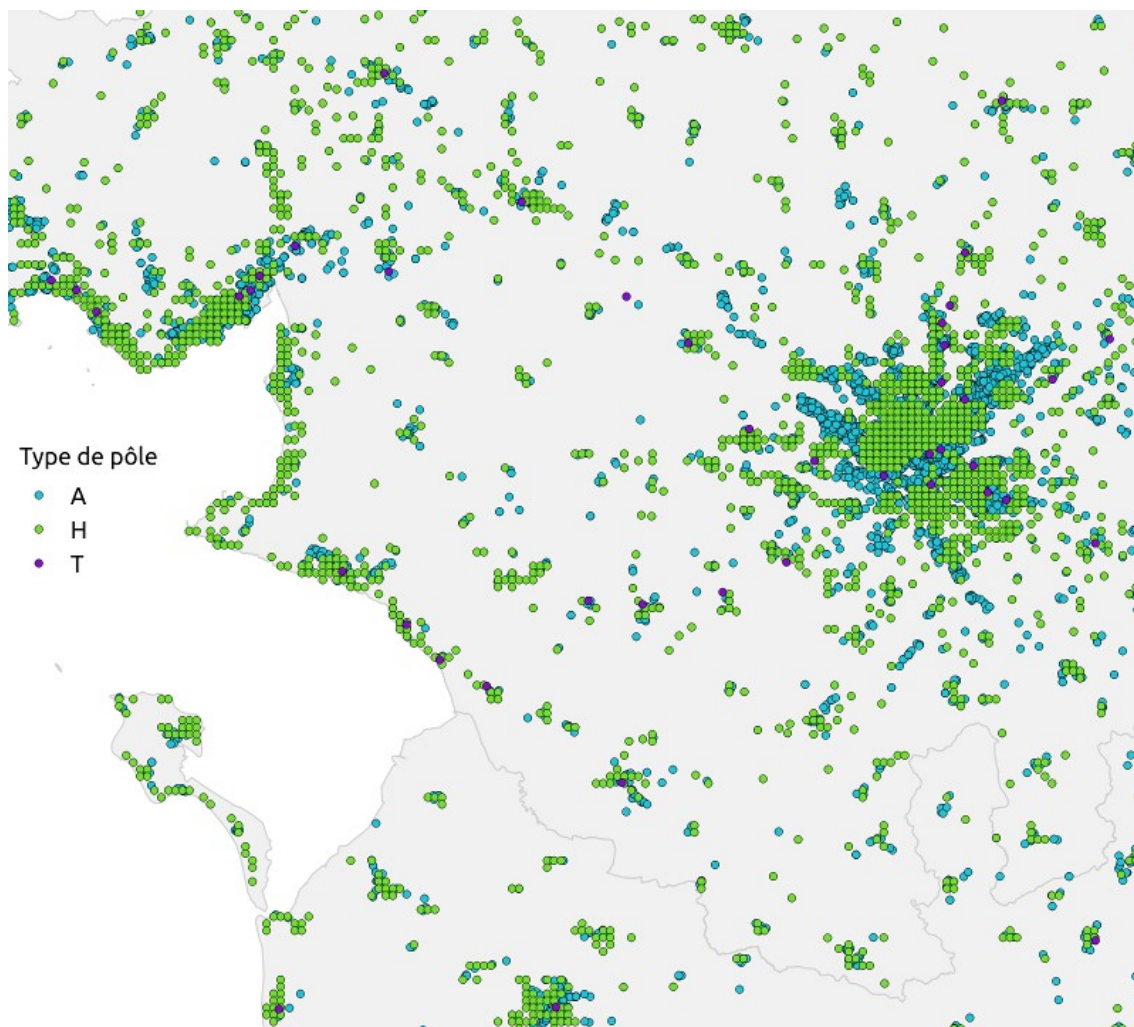
À partir de ces informations nous avons défini un **indice** qui témoigne à la fois de nombre d'individus et de leur distance à la portion de 3V. Nous pouvons ainsi fixé un **seuil** de cet indice pour distinguer et **filtrer les portions de 3V trop éloignées et/ou représentant trop peu d'individus**.

## Identification des pôles

*Le détail de la création des pôles est présenté en Annexe 1.*

Pour identifier les pôles nous avons essentiellement utilisé des sources de **données libres** (*Open Data*) à l'exception des données utilisées pour localiser les pôles *habitat* qui ont nécessité les données de la BD Carto de l'IGN (les données OpenStreetMap auraient également pu être utilisées).

Pour chaque type de données nous avons eu la possibilité d'avoir une **information quantitative** permettant de travailler avec un **nombre d'individus suffisamment précis** au lieu d'une simple distinction pôle majeur / pôle simple proposée initialement.



## Traitement des données

*Le détail du traitement des données est présenté en Annexe 2.*

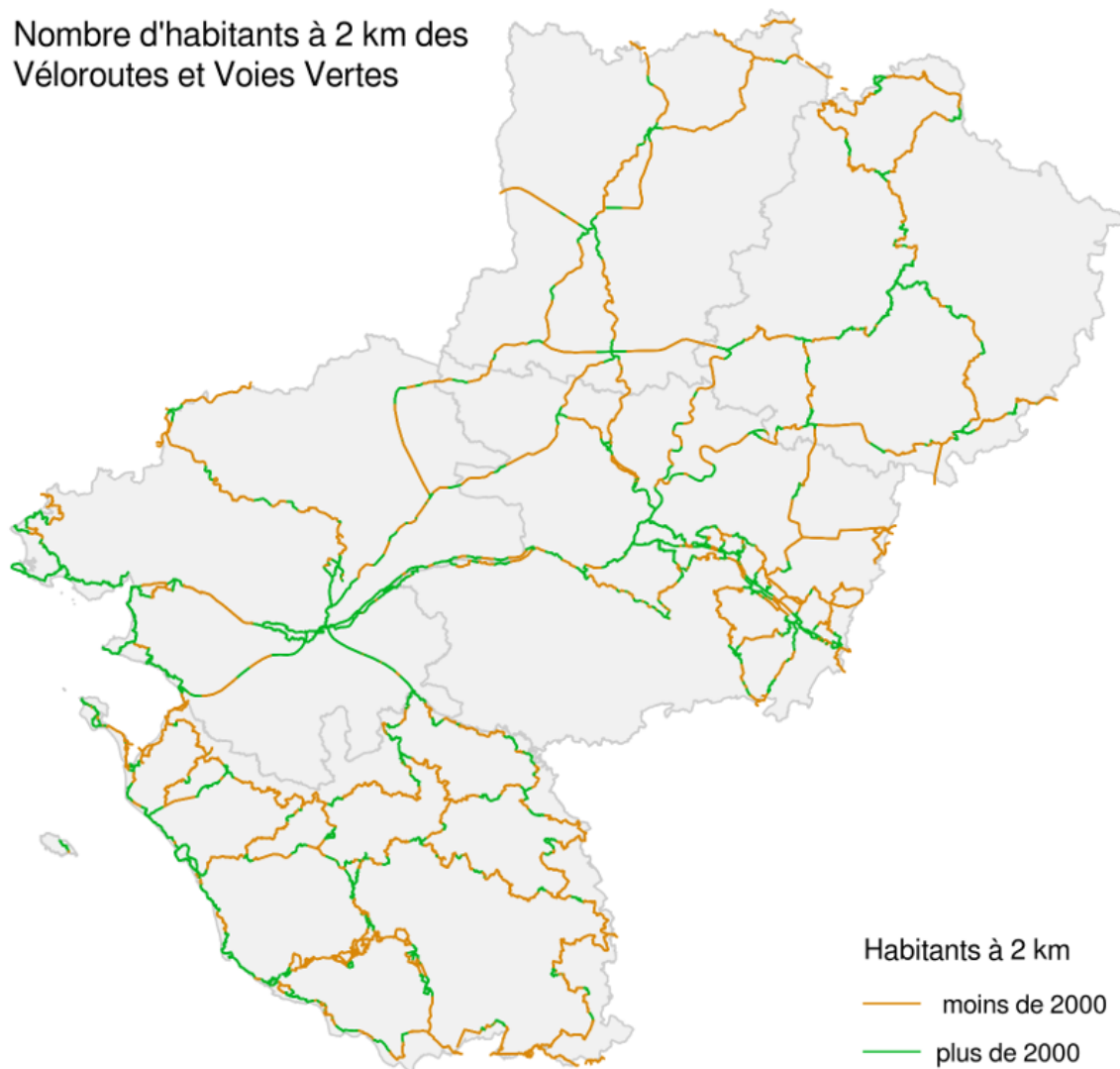
Le traitement des données est l'étape qui a permis de passer d'une information contenue dans les pôles à une information contenue dans les 3V sous la forme d'une synthèse des individus, des distances et des types de pôle (cf. tableau).

Pour cela les 3V ont été découpées en portions de 100 m et nous avons récupéré les données contenues dans les pôles grâce à des **traitements géomatiques** de type **tampons** et **jointures spatiales**.

Pour des raisons de **performance** et de **reproductibilité** nous avons privilégié l'approche par tampons, c'est-à-dire des cercles de pas de 2 km autour des pôles plutôt que l'approche par iso-distance initialement prévue.

Ce résultat intermédiaire offre déjà des informations exploitables, par exemple la carte suivante présente le nombre d'habitants situés à moins de 2 km des 3V ; la représentation se base sur une valeur seuil pour simplifier la lecture :

## Nombre d'habitants à 2 km des Véloroutes et Voies Vertes



### Indice

Le détail du calcul de l'indice est présenté en Annexe 2.

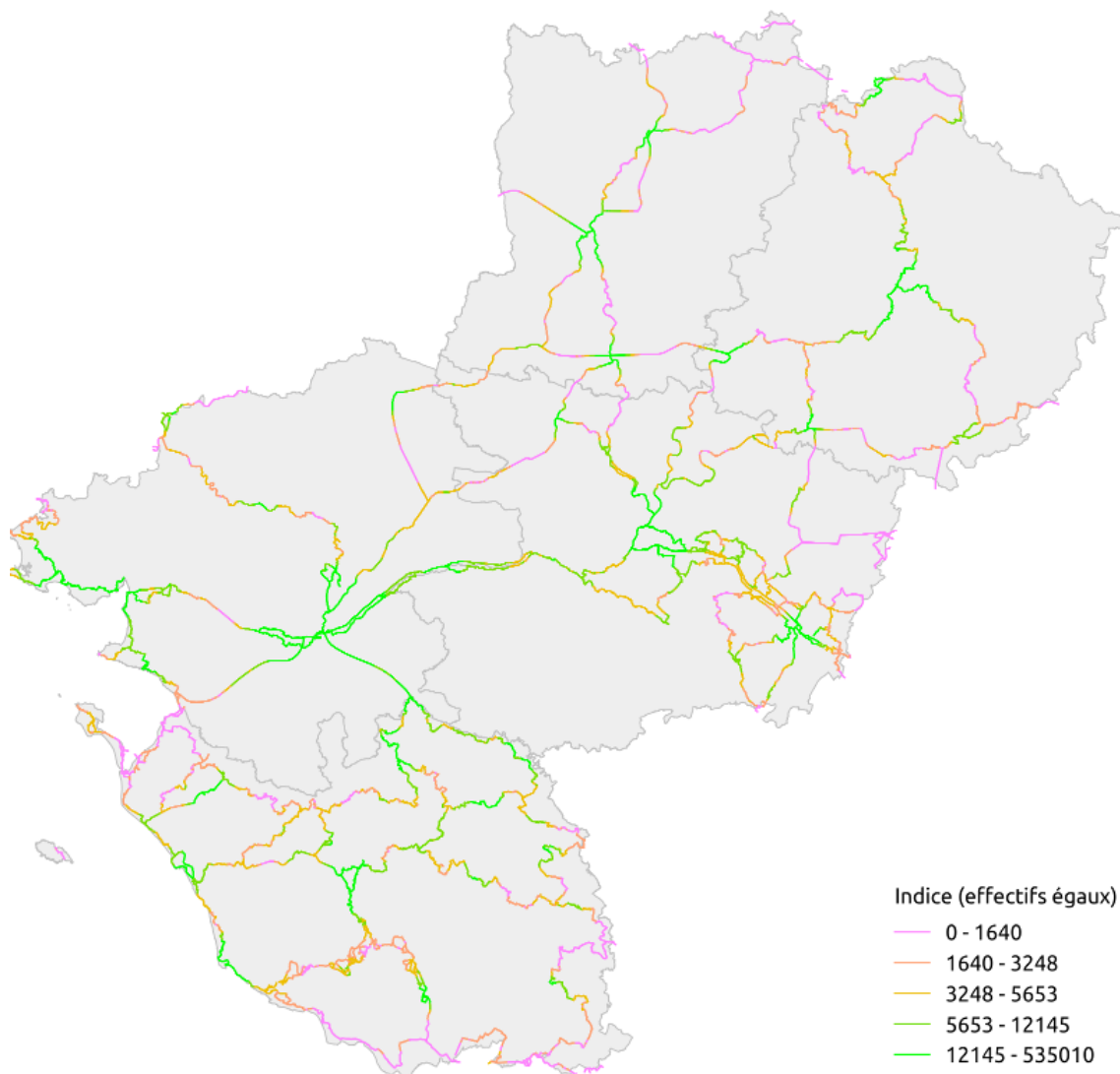
L'indice retenu tient compte de la distance et du nombre d'*individus* soit le **nombre de déplacements potentiels**.

Le nombre de déplacements potentiels est calculé par tranche de distances : les déplacements potentiels de 2 km, de 4 km, de 6 km, de 8 km et de 10 km.

Le calcul est basé sur un principe du **plus petit dénominateur commun**, c'est-à-dire que si une portion présente 500 *habitants* et 200 *actifs* dans son rayon de 2 km alors le potentiel sera de 200 déplacements.

Ainsi le potentiel a été calculé pour chaque distance. Puis nous avons réalisé une **somme pondérée des potentiels de chaque distance**. Il s'agit d'une pondération linéaire arbitraire :

$$p_{\text{final}} = 1 \times p_{2\text{km}} + 4/5 \times p_{4\text{km}} + 3/5 \times p_{6\text{km}} + 2/5 \times p_{8\text{km}} + 1/5 \times p_{10\text{km}}$$



Un des avantages de cette approche est qu'elle nous permet de décrire d'autres indices, par exemple :

- Rapport densité d'habitants à 4 km/densité PdL
- Rapport densité moyenne à 10 km/densité PdL
- Indicateur de rabattement *transport* à 2 km

# Annexe 1 – Création des pôles générateurs

## Présentation

Toutes les opérations sont réalisées sous QGIS. Pour chaque ensemble, il est généralement question de 2 étapes principales à savoir la localisation et la quantification qui sont généralement présentes dans des sources de données différentes.

Toutes les sources de données, à l'exception de la BD Carto utilisée pour localiser les pôles *habitat*, sont des données libres et gratuites afin d'assurer une reproductibilité de la méthode.

## Gares ferroviaires

Il s'agit des pôles principaux de type *transport*.

### Localisation

Couche des gares de France OpenData SNCF (*referentiel-gares-voyageurs*) :

[https://ressources.data.sncf.com/explore/dataset/referentiel-gares-voyageurs/export/?sort=intitule\\_gare](https://ressources.data.sncf.com/explore/dataset/referentiel-gares-voyageurs/export/?sort=intitule_gare)

Isolement des gares des PdL et à proximité avec les limites de région et un tampon de 10km : Création d'un tampon de 10 km autour de la région puis **sélection spatiale**.

### Quantification

Données de fréquentation des gares OpenData SNCF :

[https://ressources.data.sncf.com/explore/dataset/frequentation-gares/export/?sort=nombre\\_personnes](https://ressources.data.sncf.com/explore/dataset/frequentation-gares/export/?sort=nombre_personnes)

Jointure attributaire avec la couche des gares par le **code UIC**. Subtilité : il faut traiter les code *UIC* d'une base ou de l'autre avec un traitement pour tronquer les digits inutiles ; exemple d'expression QGIS pour réaliser ce traitement : `substr("uic", 5, 6)`

## Gares routières

Elles complètent les gares ferroviaires pour les pôles *transport*.

### Localisation

Le tableur *Implantations 2016 lieux publics PdL* contient les coordonnées géographiques de nombreux types de lieux publics dont les gares routières des PdL. Pour l'exploiter il faut l'importer dans QGIS et sélectionner les *gares routières* avec l'expression suivante : `"SS_CATEG" LIKE 'Gare routière'`

### Quantification

Il n'y a pas de données disponibles concernant le nombre de montées ; nous pourrions prendre une **valeur moyenne journalière arbitraire** de 10 montées.

# Habitations

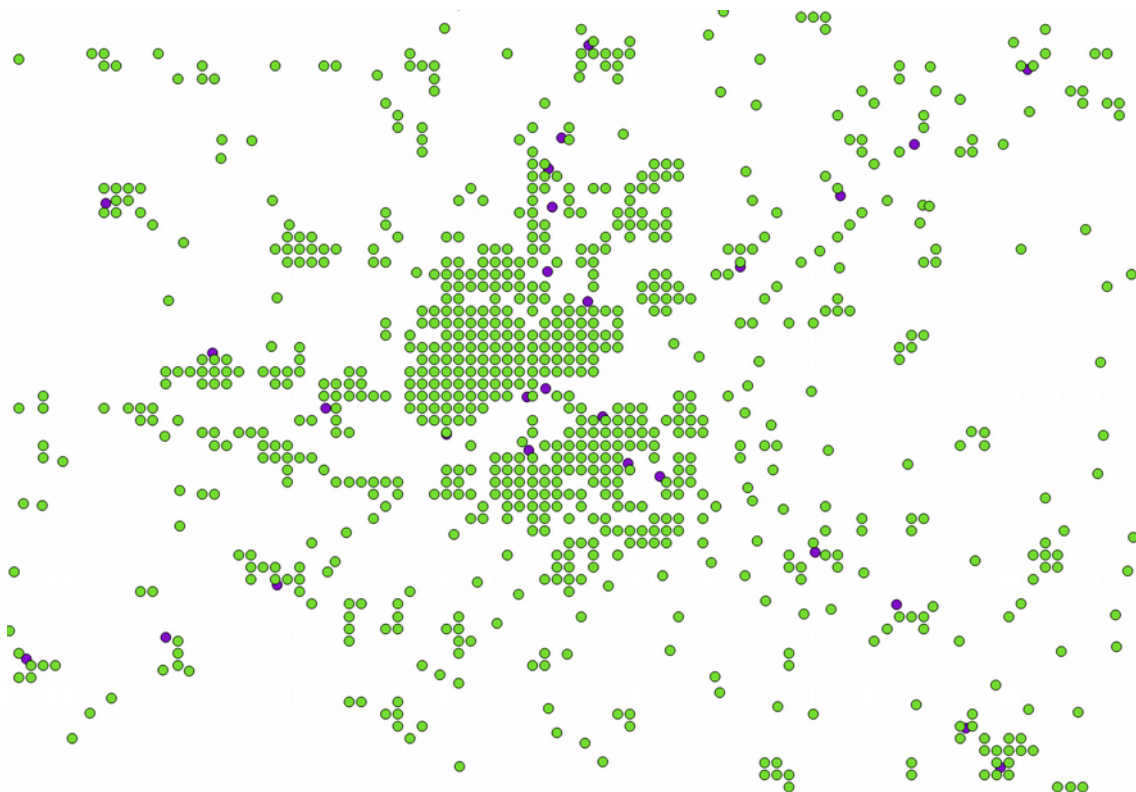
## Localisation

**Prétraitements** des couches BD Carto des départements : regroupement des 5 départements (*merge*) puis isolement par sélection des polygones d'habitation par l'expression "occupation\_du\_sol" LIKE 'bâti'.

Les pôles *habitat* seront de type point comme tous les pôles mais pour des raisons de **prise en compte de la surface des grandes zones habitées** on distinguera 2 cas en fonction de la surface :

- Centroïde du polygone pour les polygones inférieurs à 0.25 km<sup>2</sup>
- Grille de points pour les polygones supérieurs 0.25 km<sup>2</sup>

L'illustration suivante présente les pôles *habitat* en vert (en mauve, les pôles *transport*) :



## Quantification

Nombre d'habitants de plus de 6 ans : Carroyage Insee :

<https://www.data.gouv.fr/fr/datasets/donnees-carroyees-a-200-m-sur-la-population/>

- 1ère étape : récupérer le nombre d'habitants pour chaque polygone par jointure spatiale de type intersection avec somme des valeurs de la couche INSEE (cette approche a pour effet négligeable de surestimer le nombre d'habitants ; elle peut être affinée en travaillant au prorata de la surface de la cellule INSEE intersectée).
- 2ème étape : transformer les polygones de moins de 0.25 km<sup>2</sup> en centroïde
- 3ème étape : générer une grille de point de pas 500 m
- 4ème étape : jointure spatiale de la grille « vers » les polygones de plus de 0.25 km<sup>2</sup> ; les polygones héritent ainsi du nombre de points de la grille présents à l'intérieur de chaque

polygone ; ce nombre permet de calculer le nombre d'habitants par point (de la grille) qui est propre à chaque polygone.

- 5ème étape : jointure spatiale des polygones « vers » la grille : affecte aux points de la grille la valeur du nombre d'habitants calculée à l'étape précédente. Les points en dehors des polygones sont évacués lors de cette opération.

Nous avons conservé la valeur d'habitants de la tranche d'âge supérieure à 6 ans.

## Pôles activité issus de Sirene

Les établissements des entreprises et leurs effectifs issus de la base Sirene sont les éléments principaux des pôles de types *activité*.

La base de données Sirene a été libérée le 1er janvier 2017 et est disponible à cette adresse : <http://www.data.gouv.fr/fr/datasets/base-sirene-des-entreprises-et-de-leurs-etablissements-siren-siret/>

Une version géocodée avec le géocodeur de la Base Adresse Nationale (<http://adresse.data.gouv.fr>) et des adresses présentes dans la base Sirene nous a été fournie.

La base de données présente une donnée d'effectif arrondie mais suffisamment précise pour être utilisée avec les autres valeurs d'effectif. Cette information est contenue dans le champ *efetcent*.

Le travail consiste donc à filtrer les entreprises présentant un effectif inférieur à 10 personnes et les entreprises sans effectifs.

## Écoles élémentaires, collèges et lycées

Les écoles, collèges et lycées complètent les pôles de type *activité*.

### Localisation

*Adresse et géolocalisation des établissements d'enseignement du premier et second degrés :*

<https://www.data.gouv.fr/fr/datasets/adresse-et-geolocalisation-des-etablissements-denseignement-du-premier-et-second-degres/>

### Quantification

La donnée *Effectifs des écoles du premier degré public et privé sous tutelle du ministère en charge de l'Éducation Nationale* est accessible en Open Data :

<https://www.data.gouv.fr/en/datasets/effectifs-des-ecoles-du-premier-degre-public-et-privé-sous-tutelle-du-ministere-en-charge-de-leducation-nationale/>

La donnée *Effectifs des établissements du second degré public et privé sous tutelle du ministère en charge de l'Éducation Nationale* est également accessible en Open Data :

<https://www.data.gouv.fr/fr/datasets/effectifs-des-etablissements-du-second-degre-public-et-privé-sous-tutelle-du-ministere-en-charge-de-leducation-nationale/>

**2 jointures** entre les 2 données d'effectifs et la couche de géolocalisation sur le *UAI* permettent d'obtenir l'effectif de chaque établissement.



## **Enseignement supérieur**

Les établissements d'enseignement supérieur complètent les pôles de type *activité*.

### **Localisation**

*Principaux établissements d'enseignement supérieur :*

<https://www.data.gouv.fr/fr/datasets/principaux-etablissements-d-enseignement-superieur-mesr/>

### **Quantification**

Les effectifs sont les valeurs issues des sites des établissements universitaires concernés.

## Annexe 2 – Traitement des données

### Créer la couche tampons

cf. PostGIS – Utilisation de base

```
CREATE TABLE veloroutes.tampons (  
  TYPE          VARCHAR(1),  
  valeur        INTEGER,  
  distance      INTEGER,  
  geom          geometry(Polygon,2154)  
);
```

### Création de plusieurs tampons en SQL simple

Les **3 opérations** suivantes permettent de créer la couche des tampons avec les valeurs attributaires à partir d'une **couche de pôles générateurs** structurée simplement avec les **champs type et valeur**. Les types de pôle doivent être de 3 types codifiés par les lettres **H, A** et **T** pour Habitat, Activité et Transport.

### Création de plusieurs tampons en SQL

L'objectif est d'ajouter des lignes dans la couche *tampon* à partir de tampons créés depuis la couche *poles* avec comme valeurs :

- le type de pôles
- la valeur
- la distance du tampon

La géométrie du tampon est créée grâce à la fonction `ST_Buffer` et est transférée dans le champ *geom*.

```
INSERT INTO tampons(TYPE, valeur, geom, distance) SELECT TYPE, valeur,  
ST_Buffer(geom, 2000), '2000' FROM poles;  
INSERT INTO tampons(TYPE, valeur, geom, distance) SELECT TYPE, valeur,  
ST_Buffer(geom, 4000), '4000' FROM poles;  
INSERT INTO tampons(TYPE, valeur, geom, distance) SELECT TYPE, valeur,  
ST_Buffer(geom, 6000), '6000' FROM poles;  
INSERT INTO tampons(TYPE, valeur, geom, distance) SELECT TYPE, valeur,  
ST_Buffer(geom, 8000), '8000' FROM poles;  
INSERT INTO tampons(TYPE, valeur, geom, distance) SELECT TYPE, valeur,  
ST_Buffer(geom, 10000), '10000' FROM poles;
```

### Ajouter les champs nécessaires

```
ALTER TABLE tampons  
ADD COLUMN h2 INTEGER,  
ADD COLUMN h4 INTEGER,  
ADD COLUMN h6 INTEGER,  
ADD COLUMN h8 INTEGER,  
ADD COLUMN h10 INTEGER,  
ADD COLUMN a2 INTEGER,  
ADD COLUMN a4 INTEGER,  
ADD COLUMN a6 INTEGER,  
ADD COLUMN a8 INTEGER,  
ADD COLUMN a10 INTEGER,  
ADD COLUMN t2 INTEGER,  
ADD COLUMN t4 INTEGER,  
ADD COLUMN t6 INTEGER,
```

```
ADD COLUMN t8 INTEGER,  
ADD COLUMN t10 INTEGER;
```

## Mettre à jour les champs

```
UPDATE tampons  
  SET h2 = valeur WHERE TYPE = 'H' AND distance = 2000;  
UPDATE tampons  
  SET h4 = valeur WHERE TYPE = 'H' AND distance = 4000;  
UPDATE tampons  
  SET h6 = valeur WHERE TYPE = 'H' AND distance = 6000;  
UPDATE tampons  
  SET h8 = valeur WHERE TYPE = 'H' AND distance = 8000;  
UPDATE tampons  
  SET h10 = valeur WHERE TYPE = 'H' AND distance = 10000;  
UPDATE tampons  
  SET a2 = valeur WHERE TYPE = 'A' AND distance = 2000;  
UPDATE tampons  
  SET a4 = valeur WHERE TYPE = 'A' AND distance = 4000;  
UPDATE tampons  
  SET a6 = valeur WHERE TYPE = 'A' AND distance = 6000;  
UPDATE tampons  
  SET a8 = valeur WHERE TYPE = 'A' AND distance = 8000;  
UPDATE tampons  
  SET a10 = valeur WHERE TYPE = 'A' AND distance = 10000;  
UPDATE tampons  
  SET t2 = valeur WHERE TYPE = 'T' AND distance = 2000;  
UPDATE tampons  
  SET t4 = valeur WHERE TYPE = 'T' AND distance = 4000;  
UPDATE tampons  
  SET t6 = valeur WHERE TYPE = 'T' AND distance = 6000;  
UPDATE tampons  
  SET t8 = valeur WHERE TYPE = 'T' AND distance = 8000;  
UPDATE tampons  
  SET t10 = valeur WHERE TYPE = 'T' AND distance = 10000;
```

## Découper les 3V en portions et récupérer la synthèse des données des tampons

Il s'agit d'une étape intermédiaire qui va permettre de récupérer une synthèse des informations contenues dans la couche tampon. Les étapes consistent à :

- Découper les 3V en portions de 100 m
- Affecter un identifiant unique aux portions de 3V
- Créer un point au milieu de la portion avec l'identifiant unique de la portion
- Récupérer dans les points les données cumulées de la couche tampon par jointure spatiale
- Réaffecter les données des points aux portions de 3V par jointure attributaire

### Découpage

Le découpage des 3V en **portions de 100 m** peut être réalisé grâce à la fonction de GRASS GIS *v.split*. Cette fonction est directement accessible sous QGIS.

### Ajouter l'identifiant unique

Ajouter un identifiant unique à chaque portion de 3V va être utile avant de créer des points milieu pour récupérer les informations de ces points (cf. étapes suivantes).

Pour ajouter un champ avec un identifiant unique sous QGIS nous pouvons utiliser la calculatrice de champs avec le paramètre \$rownum.

## Conversion en points

Importer les portions dans PostGIS

```
SELECT ST_Line_Interpolate_Point(geom, 0.5)AS geom, id FROM veloroutes_split
```

## Jointure spatiale

La jointure spatiale entre les points représentant les portions de 3V et la couche tampon va permettre de réaliser une synthèse par point des informations contenues dans la couche tampon. C'est-à-dire que chaque point héritera des *valeurs* cumulées par type de pôle et par distance (h2, a2, t2, h4, a4, t4, etc.).

Pour avoir une synthèse il faut utiliser la fonction *Join attributes by location* de QGIS et choisir l'option *Prendre un résumé des entités intersectées* et ne conserver que la somme (*sum*).

## Jointure attributaire

Une fois les informations de synthèse affectées aux points, nous pouvons les transférer aux portions de 3V par jointure attributaire en utilisant l'identifiant unique créé aux étapes précédentes.

## Calcul de l'indice

### Préparation

Le calcul de l'indice est effectué dans un **tableur** par souci de souplesse. Les données sont copiées depuis la table attributaire sous QGIS dans un tableur et nous ne conservons que le champ *id* et les champs des valeurs par distance et par type.

### Calcul du nombre d'individus par tranche

À ce stade nous disposons d'une couche des portions de 3V qui contient pour chaque portion le nombre d'individus par distance et par type de pôle :

portion	hab 2 km	hab 4 km	hab 6 km	hab 8 km	hab 10 km	act 2 km	act 4 km	act 6 km	act 8 km	act 10 km
1	1 500	4 800	8 600	15 000	40 000	300	700	800	1 400	2 000
2	0	900	2 500	8 100	12 300	0	0	500	900	1 100

À partir de cette donnée nous calculons par simple soustraction la part d'individus par tranche de distance. Par exemple le nombre d'*habitants* de la tranche (anneau) de 4 km est le nombre total d'*habitants* à 4 km moins le nombre (total) d'*habitants* à 2 km, le nombre d'*habitants* de la tranche 6 km est égal au nombre total d'*habitants* à 6 km moins le nombre total d'*habitants* à 4 km, etc.

portion	h2	h4	h6	h8	h10	a2	a4	a6	a8	a10
1	1 500	3 300	3 800	6 400	25000	300	400	100	600	600
2	0	900	1 600	5 600	4 200	0	0	500	400	200

## Calcul du potentiel par distance

Le potentiel par tranche est calculé par une méthode du plus petit dénominateur commun. D'autres méthodes telles que des moyennes (harmoniques, pondérées, etc.) auraient également pu être utilisées mais nous avons jugé et observé que l'approche du plus petit dénominateur commun était la plus rationnelle.

Le plus petit dénominateur commun peut être calculé grâce à la fonction de minimum du tableur.

Pour le potentiel de chaque distance il faut tenir compte des différents cas de figure :

$$p2km = \text{MIN}(h2; a2)$$

$$p4km = \text{MIN}(h2; a4) + \text{MIN}(h4; a2)$$

$$p6km = \text{MIN}(h2; a6) + \text{MIN}(h4; a4) + \text{MIN}(h6; a2)$$

$$p8km = \text{MIN}(h2; a8) + \text{MIN}(h4; a6) + \text{MIN}(h6; a4) + \text{MIN}(h8; a2)$$

$$p10km = \text{MIN}(h2; a10) + \text{MIN}(h4; a8) + \text{MIN}(h6; a6) + \text{MIN}(h8; a4) + \text{MIN}(h10; a2)$$

## Calcul du potentiel final

Ainsi le potentiel a été calculé pour chaque distance puis nous avons réalisé une somme pondérée des potentiels de chaque distance. Il s'agit d'une pondération linéaire arbitraire :

$$p_{\text{final}} = 1 \times p2km + 4/5 \times p4km + 3/5 \times p6km + 2/5 \times p8km + 1/5 \times p10km$$

Le résultat peut être observé sur la carte d'indice présentée en début de rapport.