

Etat d'avancement du stage – 13/06 – 17/06

Bérénice Le Mao

Sujet : Quelle carte multi-échelle est utile aux gestionnaires de risque après un séisme ?

I) Tâches réalisées cette semaine :

1) Processus de généralisation – routes de Nice

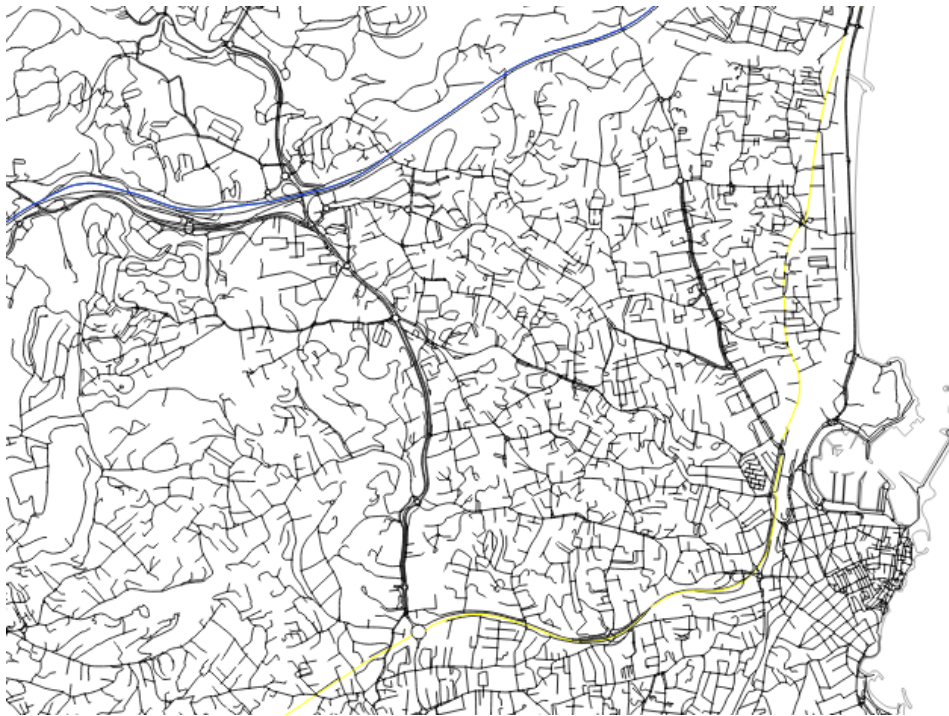
J'ai pu terminer la généralisation des routes, que j'ai séparées en 5 niveaux de détail.

> reseau_routier_3

- même chose que la précédente généralisation mais commencer à s'appuyer plus sur l'autre bdd (celle avec les tronçons de route)



- Nice est au centre mais aucune route dans bdd ne semble se connecter alors que ce n'est pas le cas dans la réalité, donc je le fais à la main en suivant l'autre bdd coupée par dep (tronçons de route) pour de courtes distances. Sinon, je sélectionne directement depuis l'autre bdd et je regroupe sur la nouvelle couche.
- trier en fct de la densité

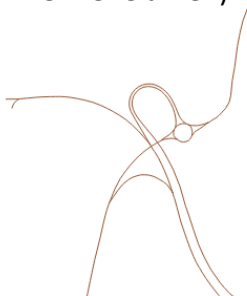


> privilégier le trait le plus central



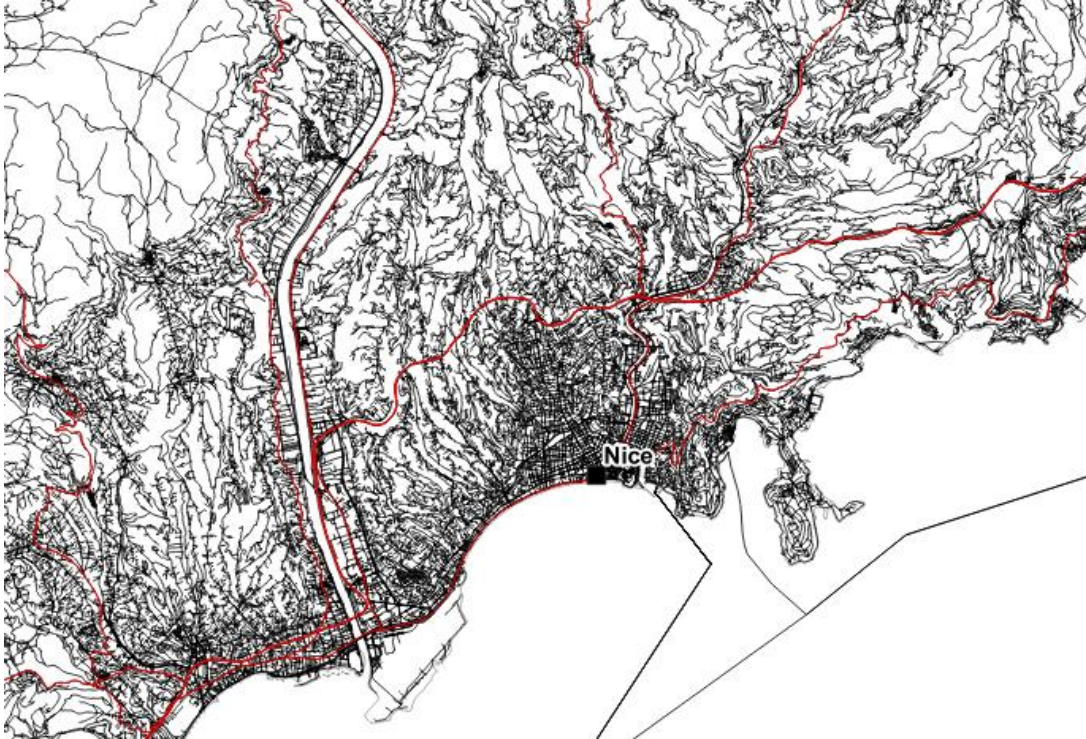
> reseau_routier_4

- sélectionner les routes des généralisations précédentes sur la bdd des routes numérotées et nommées pour garder les précisions du réseau (ronds-points, séparation des routes) mnt devenues importantes comme point de repère
- sélectionner un réseau plus dense encore
- ne pas prendre toute la bdd sans changement : la nettoyer (ex ci-dessous : bras qui ne mène à rien)



> reseau_routier_5

- utiliser uniquement les tronçons de route à Nice : pas besoin de découper les vecteurs une fois sélectionnés mais le processus de sélection est plus long car chaque intersection coupe la ligne
- s'appuyer également sur les routes sélectionnées par géoportail

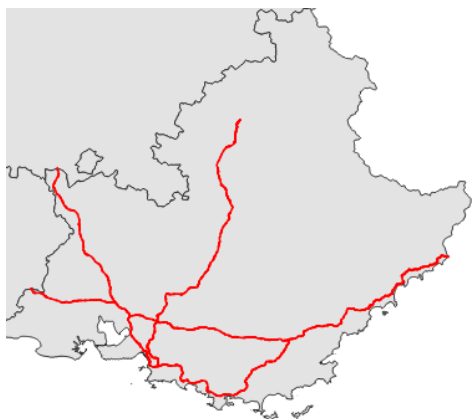


- très dense donc commencer par trier à l'aide de la table attributaire et enlever
 - escaliers
 - pistes cyclables
- prendre les formes géométriques et les points de repères saillants (ex place ronde)



Comme résultat pour chaque donnée, j'ai :

A) Les routes



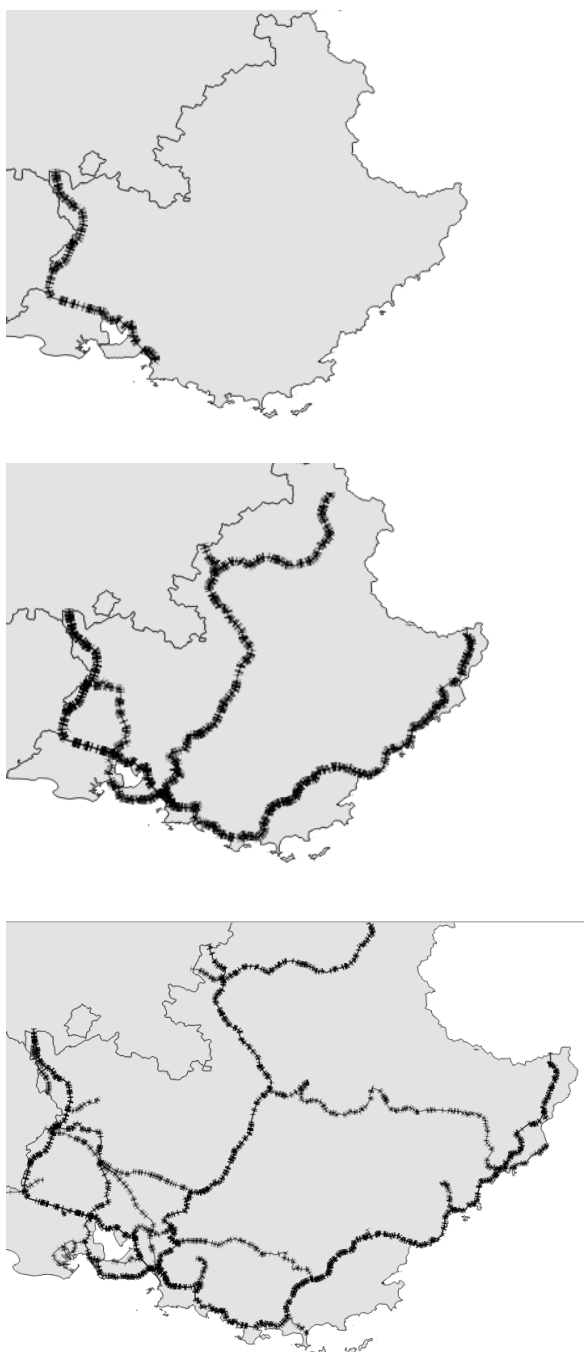


B) Les cours d'eau





C) Les voies ferrées



2) Nettoyer les données du BRGM

Ces données étant issues du BRGM, j'ai d'abord dû les nettoyer. Après avoir remplacé les « . » par des « , » j'ai pu arrondir les valeurs pour ne garder que les entiers. Mais j'étais encore face au problème des répétitions dans la base de données ; problème auquel j'ai déjà été confronté et que j'ai tenté de résoudre manuellement à l'aide de la formule SOMME dans excel, avant de me rendre compte que le fichier excel des ilots IRIS comprenait plus de 7000 entités et que les parcourir manuellement prendrait beaucoup trop de temps. J'ai ainsi décidé d'utiliser des méthodes automatiques. J'ai alors pensé à deux solutions : la première

était de passer par une base de données postgres afin de rentrer une requête SQL pour regrouper les lignes avec une même valeur. L'autre solution était de passer par python.

A. Utiliser postgres

J'ai d'abord créé une nouvelle table dans la base de données puis j'ai rajouté les champs présents dans le fichier csv et ainsi que leur type de données (integer, json...)

- pays_code → text
- field → integer
- nbhabitants → //
- D0
- D1
- D2
- D3
- D4
- D5
- D3+4+5
- D1+2+3+4+5
- dcom_iris → TEXT (corse > A+B)
- P0
- P1
- P2
- P3
- P4
- P2+3
- reg_code
- dpt_code
- id_enjeu
- commune_code
- iems → integer

The screenshot shows the 'Columns' tab in pgAdmin for a table named 'csv_iris'. The table has 24 columns. The first 14 columns are: pays_code (text), field (integer), nbhabitants (integer), D0 (integer), D1 (integer), D2 (integer), D3 (integer), D4 (integer), D5 (integer), D3+4+5 (integer), D1+2+3+4+5 (integer), dcom_iris (text), P0 (integer), P1 (integer), P2 (integer), P3 (integer), P4 (integer), P2+3 (integer), reg_code (text), dpt_code (text), id_enjeu (text), commune_code (text), and iems (integer). The 'Not NULL?' and 'Primary key?' columns are all set to 'No'. The 'Default' column is empty for all columns. The 'Inherited from table(s)' dropdown is set to 'Select to inherit from...'. The 'Columns' tab is selected, and the 'General' tab is also visible. The 'Close', 'Reset', and 'Save' buttons are at the bottom right.

Name	Data type	Length/Precision	Scale	Not NULL?	Primary key?	Default
pays_code	text			<input type="checkbox"/>	<input type="checkbox"/>	
field	integer			<input type="checkbox"/>	<input type="checkbox"/>	
nbhabitants	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D0	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D1	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D2	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D3	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D4	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D5	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D3+4+5	integer			<input type="checkbox"/>	<input type="checkbox"/>	
D1+2+3+4+5	integer			<input type="checkbox"/>	<input type="checkbox"/>	
dcom_iris	text			<input type="checkbox"/>	<input type="checkbox"/>	
P0	integer			<input type="checkbox"/>	<input type="checkbox"/>	
P1	integer			<input type="checkbox"/>	<input type="checkbox"/>	
P2	integer			<input type="checkbox"/>	<input type="checkbox"/>	
P3	integer			<input type="checkbox"/>	<input type="checkbox"/>	
P4	integer			<input type="checkbox"/>	<input type="checkbox"/>	
P2+3	integer			<input type="checkbox"/>	<input type="checkbox"/>	
reg_code	text			<input type="checkbox"/>	<input type="checkbox"/>	
dpt_code	text			<input type="checkbox"/>	<input type="checkbox"/>	
id_enjeu	text			<input type="checkbox"/>	<input type="checkbox"/>	
commune_code	text			<input type="checkbox"/>	<input type="checkbox"/>	
iems	integer			<input type="checkbox"/>	<input type="checkbox"/>	

- importer les données

Import/Export data - table 'csv_iris'

Options Columns

Import/Export ☒ Import ☐ Export

File Info

Filename

Format

Encoding

Miscellaneous

OID ☐

Header ☒

Delimiter

Specifies the character that separates columns within each row (line) of the file. The default is a tab character in text format, a comma in CSV format. This must be a single one-byte character. This option is not

- ATTENTION csv normalement avec des virgules mais moi je suis en ; donc bien renseigner le délimiteur

resultat_propre - Bloc-notes

Fichier Edition Format Affichage Aide

```
pays_code;field;nbhabitants;D0;D1;D2;D3;D4;D5;D3+4+5;D1+2+3+4+5;dcom_iris;P0;P1;P2;P3;P4;P2+3;reg_code;dpt_code;id_enjeu;commune_code;
FR;2092;1161;432;7;1;0;0;0;0;8;40010000;1161;0;0;0;0;93;4;51874;4001;4
FR;2093;189;70;1;0;0;0;0;0;1;40010000;189;0;0;0;0;93;4;51874;4001;4
FR;2096;502;390;12;2;0;0;0;0;13;40040000;502;0;0;0;0;93;4;51875;4004;4
FR;2097;21;16;0;0;0;0;0;0;1;40040000;21;0;0;0;0;93;4;51875;4004;4
FR;2098;145;119;9;1;0;0;0;0;11;40050000;145;0;0;0;0;93;4;51876;4005;5
FR;2099;645;1170;36;5;0;0;0;0;42;40060000;645;0;0;0;0;93;4;51877;4006;4
FR;2100;65;63;5;1;0;0;0;0;6;40070000;65;0;0;0;0;93;4;51878;4007;5
FR;2094;1068;565;50;9;1;0;0;0;1;60;40080000;1068;0;0;0;0;93;4;51879;4008;5
FR;2095;20;27;2;0;0;0;0;0;2;40090000;20;0;0;0;0;93;4;51880;4009;4
FR;2101;82;31;1;0;0;0;0;1;1;2;40120000;82;0;0;0;0;93;4;51881;4012;4
FR;2102;24;9;0;0;0;0;0;0;0;40120000;24;0;0;0;0;93;4;51881;4012;4
FR;2103;551;247;5;1;0;0;0;0;6;40130000;551;0;0;0;0;93;4;51882;4013;4
FR;2105;7;3;0;0;0;0;0;0;0;40130000;7;0;0;0;0;93;4;51882;4013;4
FR;2106;49;50;2;0;0;0;0;0;2;40160000;49;0;0;0;0;93;4;51883;4016;4
FR;2107;96;90;3;0;0;0;0;0;3;40170000;96;0;0;0;0;93;4;51884;4017;4
FR;2108;1010;568;10;1;0;0;0;0;12;40180000;1010;0;0;0;0;93;4;51885;4018;4
FR;2110;16;9;0;0;0;0;0;0;0;40180000;16;0;0;0;0;93;4;51885;4018;4
FR;2111;1400;582;21;3;0;0;0;0;24;40190000;1400;0;0;0;0;93;4;51886;4019;4
FR;2087;1287;535;19;3;0;0;0;0;22;40190000;1287;0;0;0;0;93;4;51886;4019;4
FR;2109;143;100;5;1;0;0;0;0;6;40200000;143;0;0;0;0;93;4;51887;4020;4
FR;2112;136;62;2;0;0;0;0;0;3;40210000;136;0;0;0;0;93;4;51888;4021;4
FR;2113;2;1;0;0;0;0;0;0;0;40210000;2;0;0;0;0;93;4;51888;4021;4
FR;2115;434;284;18;3;0;0;0;0;22;40220000;434;0;0;0;0;93;4;51889;4022;4
```

Ln 1, Col 1 100% Windows (CRLF) UTF-8

- importer les colonnes

Import/Export data - table 'csv_iris'

Options Columns

Columns to import

pays_code x nbhabitants x field x D0 x D1 x D2 x
D3 x D4 x D5 x D3+4+5 x D1+2+3+4+5 x dcom_iris x
P0 x P1 x P2 x P3 x P4 x P2+3 x reg_code x
dpt_code x id_enjeu x commune_code x iems x

An optional list of columns to be copied. If no column list is specified, all columns of the table will be copied.

NULL Strings

Specifies the string that represents a null value. The default is \N (backslash-N) in text format, and an unquoted empty string in CSV format. You might prefer an empty string even in text format for cases where you don't want to distinguish nulls from empty strings. This option is not allowed when using binary format.

Not null columns

Do not match the specified column values against the null string. In the default case where the null string is empty, this means that empty values

- ⇒ problème : l'import n'a pas fonctionné et même en regardant les forums, je ne trouvais pas la réponse
- ⇒ Le problème venait très probablement du proxy de l'IGN qui bloque beaucoup d'accès, mais comme je ne savais pas que le problème pouvait venir de là, je suis directement passée à la deuxième solution : python !

B. Utiliser python

Pour modifier un fichier excel en utilisant python, il existe plusieurs manières de faire. La plus conseillée est de passer par anaconda, une distribution libre et opensource des langages de programmation Python. De là, j'ouvre « Jupyter Notebook », une application web qui permet d'utiliser plus de 40 langages de programmation dont python, et de réaliser des calepins (ou notebooks) contenant les programmes. Dessus, j'importe dans un premier temps pandas, un outil qui permet de manipuler des structures de données. Puis j'importe mon fichier csv.

```
Entrée [1]: import pandas as pd

Entrée [3]: filedata = pd.read_csv(r"C:\Users\ble-mao\Documents\Stage M2\espace de travail\donnees retravaillées\Nice\Scénario Nice - M6")
filedata.head()

Out[3]:
```

	pays_code	field	nbhabitants	D0	D1	D2	D3	D4	D5	D3+4+5	D1+2+3+4+5	dcom_iris	P0	P1	P2	P3	P4	P2+3	reg_code	dpt_code	id_enjeu	commune_code	iems
0	FR	2092	1161	432	7	1	0	0	0	0	0	0	0	0	0	0	0	0	93	4	51874	4001	4
1	FR	2093	189	70	1	0	0	0	0	0	0	0	0	0	0	0	0	0	93	4	51874	4001	4
2	FR	2096	502	398	12	2	0	0	0	0	0	0	0	0	0	0	0	0	93	4	51875	4004	4
3	FR	2097	21	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	93	4	51875	4004	4
4	FR	2098	145	119	9	1	0	0	0	0	0	0	0	0	0	0	0	0	93	4	51876	4005	5

```
Entrée [ ]:
```

Ici, le fichier ne s'affiche pas bien car le fichier est lu comme un csv standard, séparé par des virgules, et non par un point-virgule. Je transforme alors le csv en fichier excel.

Jupyter notebooks_iris Dernière Sauvegarde : il y a 4 heures (modifié)

File Edit View Insert Cell Kernel Widgets Help

Exécuter

```
Entrée [1]: import pandas as pd

Entrée [4]: filedata = pd.read_excel(r"C:\Users\ble-mao\Documents\Stage M2\espace de travail\donnees retravaillées\Nice\Scénario Nice - M6.2\retravail\resultat_propre_2.xlsx")
filedata.head()

Out[4]:
```

	pays_code	field	nbhabitants	D0	D1	D2	D3	D4	D5	D3+4+5	...	P1	P2	P3	P4
0	FR	2092	1161.171666	431.583022	7.095112	0.856463	0.073707	0.002676	0.000011	0.076394	...	0.012776	0.001470	3.632145e-06	0.000355
1	FR	2093	188.828334	70.156196	1.177529	0.142537	0.012298	0.000448	0.000002	0.012748	...	0.002128	0.000245	6.104022e-07	0.000059
2	FR	2096	502.182819	398.209138	11.537073	1.547140	0.145114	0.005751	0.000025	0.150890	...	0.010998	0.001256	4.036210e-06	0.000302
3	FR	2097	20.817181	16.487437	0.494905	0.066854	0.006311	0.000252	0.000001	0.006564	...	0.000477	0.000054	1.784218e-07	0.000013
4	FR	2098	145.000000	118.621941	8.934172	1.471635	0.164379	0.007828	0.000044	0.172252	...	0.010518	0.001190	6.471750e-06	0.000285

5 rows x 23 columns

```
Entrée [ ]:
```

Je me rends alors compte que lorsque j'ai nettoyé la bdd (arrondi les décimales), les valeurs restaient les mêmes ; seule l'apparence changeait. J'ai donc copié collé les valeurs dans un autre fichier excel.

```
Entrée [1]: import pandas as pd

Entrée [5]: filedata = pd.read_excel(r"C:\Users\ble-mao\Documents\Stage M2\espace de travail\donnees retravaillées\Nice\Scénario Nice - M6.2\retravail\resultat_propre_2.xlsx")
filedata.head()

Out[5]:
```

	pays_code	field	nbhabitants	D0	D1	D2	D3	D4	D5	D3+4+5	...	P1	P2	P3	P4	P2+3	reg_code	dpt_code	id_enjeu	commune_code	iems
0	FR	2092	1161	432	7	1	0	0	0	0	...	0	0	0	0	0	93	4	51874	4001	4
1	FR	2093	189	70	1	0	0	0	0	0	...	0	0	0	0	0	93	4	51874	4001	4
2	FR	2096	502	398	12	2	0	0	0	0	...	0	0	0	0	0	93	4	51875	4004	4
3	FR	2097	21	16	0	0	0	0	0	0	...	0	0	0	0	0	93	4	51875	4004	4
4	FR	2098	145	119	9	1	0	0	0	0	...	0	0	0	0	0	93	4	51876	4005	5

5 rows x 23 columns

Les données s'affichent correctement et je rentre la formule pour regrouper les lignes correspondant aux mêmes valeurs dans le champ « dcom_iris »

```
Entrée [6]: filedata.groupby(by="dcom_iris").sum()
```

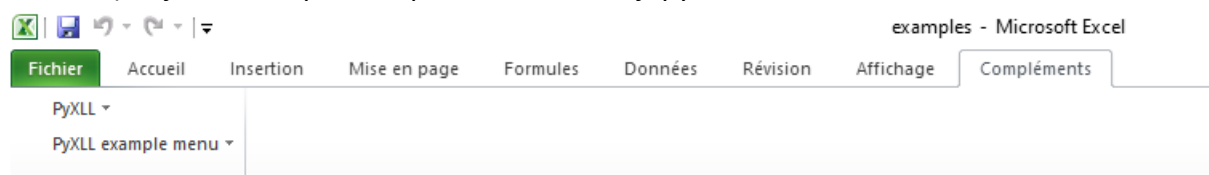
```
Out[6]:
```

	field	nbhabitants	D0	D1	D2	D3	D4	D5	D3+4+5	D1+2+3+4+5	P0	P1	P2	P3	P4	P2+3	reg_code	id_enjeu	iems
dcom_iris																			
40010000	4185	1350	502	8	1	0	0	0	0	9	1350	0	0	0	0	0	186	103748	8
40040000	4193	523	414	12	2	0	0	0	0	14	523	0	0	0	0	0	186	103750	8
40050000	2098	145	119	9	1	0	0	0	0	11	145	0	0	0	0	0	93	51876	5
40060000	2099	645	1170	36	5	0	0	0	0	42	645	0	0	0	0	0	93	51877	4
40070000	2100	65	63	5	1	0	0	0	0	6	65	0	0	0	0	0	93	51878	5
...
2B3560000	6073	137	99	1	0	0	0	0	0	1	137	0	0	0	0	0	94	64227	3
2B3610000	18250	294	215	3	0	0	0	0	0	3	294	0	0	0	0	0	282	192684	9
2B3640000	57	29	38	0	0	0	0	0	0	1	29	0	0	0	0	0	94	64229	3
2B3650000	6088	148	190	1	0	0	0	0	0	2	148	0	0	0	0	0	94	64230	3
2B3660000	6083	97	131	1	0	0	0	0	0	1	97	0	0	0	0	0	94	64231	3

3588 rows × 19 columns

Le code fonctionne, il ne me reste plus qu'à exporter les résultats sous fichier excel. Or aucun bouton de téléchargement ne permet d'enregistrer les données sous ce format.

Je décide alors d'intégrer jupyter notebook sur excel où la formule changera le fichier directement. J'installe donc l'extension pyxll sur excel mais je fais face à plusieurs problèmes : python est en 64 bits alors qu'excel est en 32 bits. J'ai donc réinstallé python en 32 bits mais ça ne fonctionne toujours pas car le problème vient d'anaconda qui lui est en 64 bits et utilise donc un python en 64 bits. Je réinstalle donc anaconda en 32 bit mais un autre message d'erreur s'affiche : le module python n'a pas réussi à être importé. Le fichier dll n'est pas lu car il n'est pas présent dans le même chemin que l'extension de python. Ce changement de chemin est sûrement dû au fait que j'ai utilisé deux sources d'installation : pip et anaconda, ce qui a changé le chemin. Anaconda est ainsi placé dans un autre chemin que pyxll et python. Pourtant, même en changeant le chemin, l'extension s'affiche mal (ci-dessous) et je n'arrive pas à importer le module jupyter notebook.



Je procède alors autrement : je reviens sur jupyter et je décide de télécharger les données traitées directement sur la plateforme. Or il n'y a pas de bouton pour enregistrer les données sous format excel donc je procède par étapes. Je ne peux pas télécharger les résultats directement car pour faciliter la lecture, jupyter cache les colonnes et les lignes qui dépassent une certaine longueur. Donc je dois tout afficher dans un premier temps sinon le fichier excel produit ne gardera que les résultats affichés

```
pd.set_option('display.max_rows', 3588)
filedata.head(filedata.groupby(by="dcom_iris").sum())
```

Ensuite je copie colle dans un fichier excel et je place chaque entité séparée par un espace dans une colonne différente. Je fais de même pour les fichiers de commune. Mon fichier est

maintenant propre : je peux commencer à représenter les dommages matériels et humains par échelle.

3) Afficher les dommages matériels et humains

Comme un des besoins récoltés lors de l'entraînement Séisme de Grenoble était d'avoir deux cartes différentes avec d'un côté les dégâts prioritaires et de l'autre, les dégâts secondaires, je décide de faire deux types de regroupements pour chaque échelle : prioritaire et secondaire.

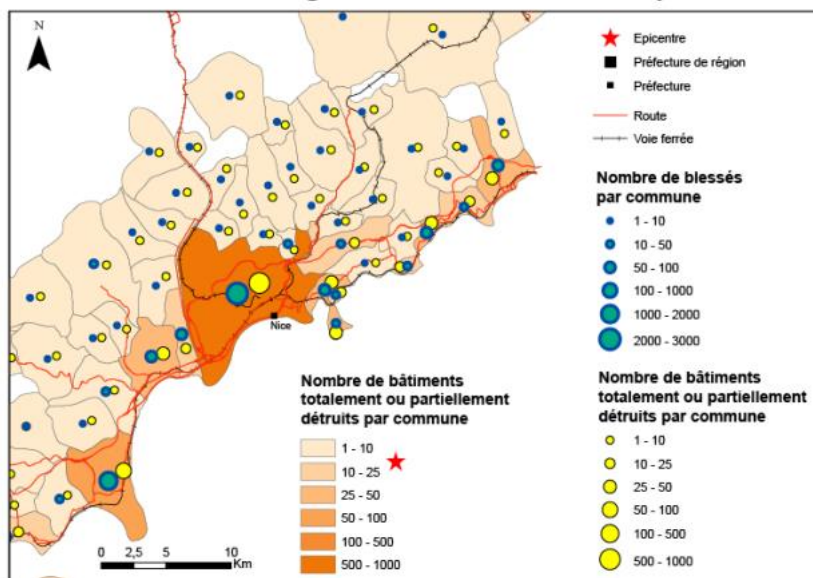
- prioritaires
 - P2+P3
 - D4+D5
- secondaires
 - D1+2+3

A. Représentation des dommages matériels

Les choix de Thileli

- afficher de deux manières une même donnée (nb de bâtiments totalement ou partiellement détruits par commune)

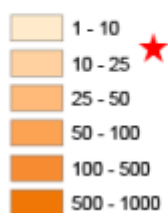
Estimation des dommages humains et matériels après un séisme



> je vais choisir de représenter les dégâts matériels comme surface et les dégâts humains comme point

- lors de la classification, elle a mis plus de précisions dans les dégâts moindre, avec des écarts d'une dizaine de bâtiments détruits tandis que pour les gros nombres, elle a mis de plus grands écarts, allant jusqu'à 500

**Nombre de bâtiments
totalement ou partiellement
détruits par commune**



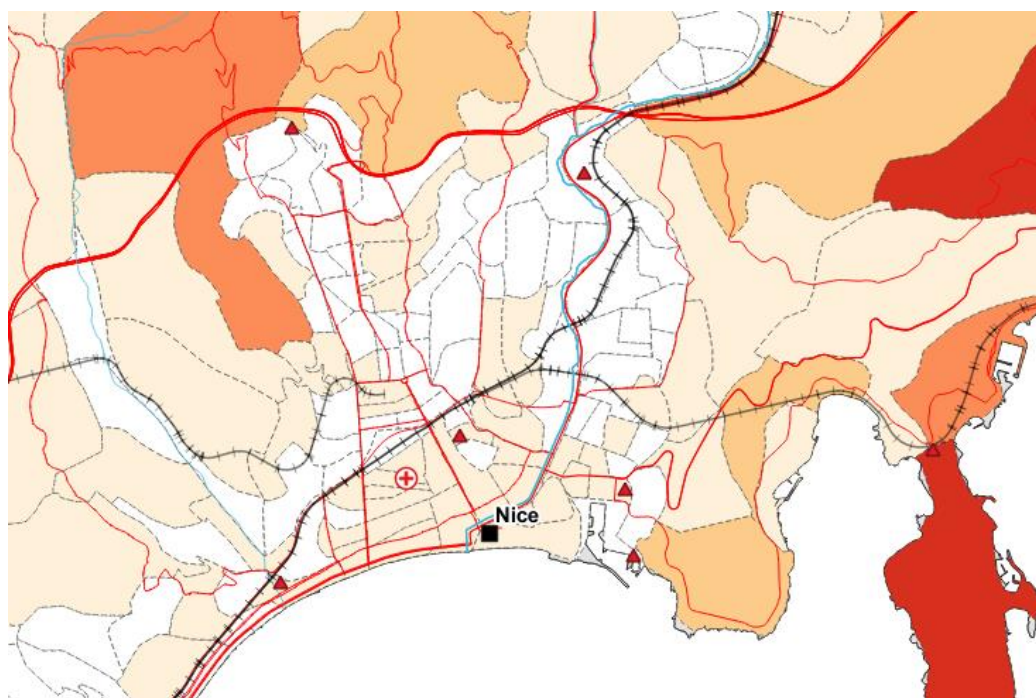
> j'ai décidé de faire l'inverse : je donne plus d'importance aux plus gros nombre donc ils ont besoin d'avoir plus de précision spatiale

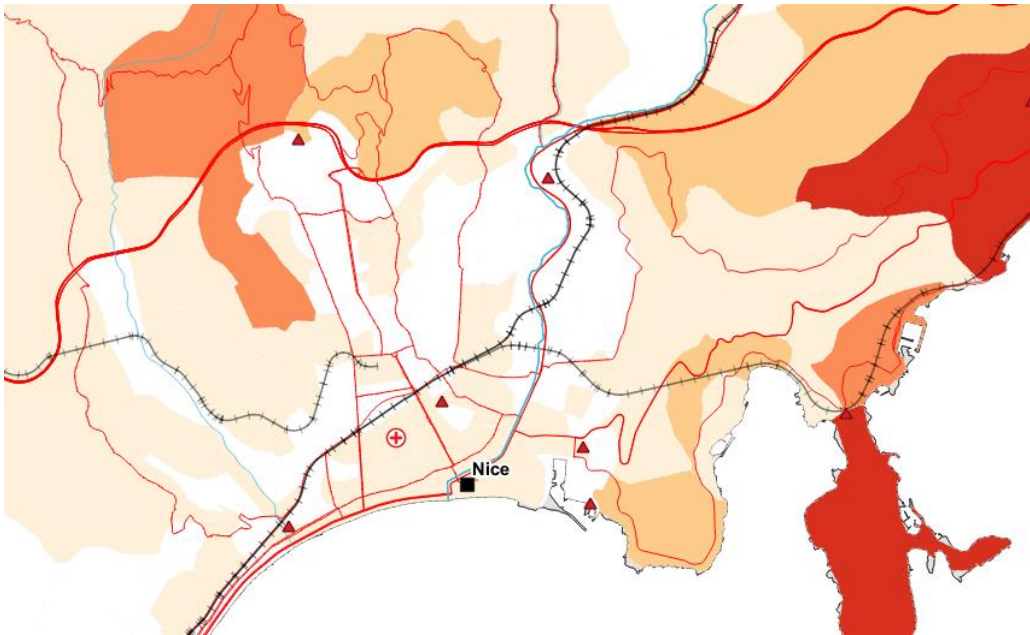
Symbole	Valeurs	Légende
✓	0,00 - 200,00	0 - 200
✓	200,00 - 400,00	200 - 400
✓	400,00 - 500,00	400 - 500
✓	500,00 - 600,00	500 - 600
✓	600,00 - 700,00	600 - 700

- Thileli mettait deux niveaux de généralisation des dégâts pour chaque échelle en jouant sur le nombre de classes mais je choisis de ne garder pour chacune des échelles qu'un niveau de généralisation car je considère qu'il est important d'avoir un niveau de détail assez rapidement pour avoir une réponse rapide

Mes autres choix

- malgré mon travail de généralisation, je remarque qu'il y a tjrs trop de données sur ma zone, notamment dû aux limites administratives. Je décide donc de les représenter en blanc pour ne pas surcharger la carte (ci-dessous les limites des ilots IRIS)

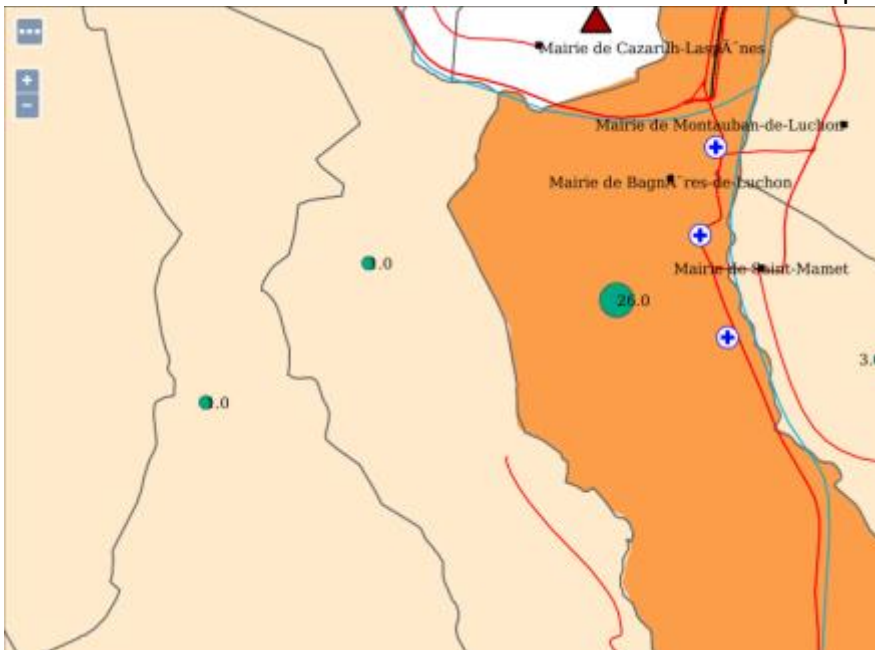




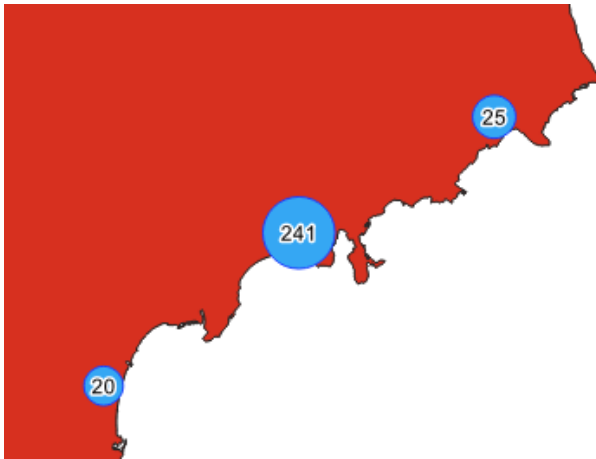
B. Représentation des dommages humains

Les choix de Thileli

- les blessés sont créés en rond avec le numéro de blessés par-dessus

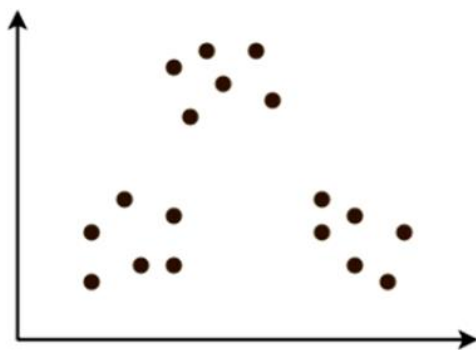


> je vais garder le cercle mais l'étiquette n'est pas bien placée donc je modifie la taille des clusters de manière à ce qu'il y ait une différence de taille notable mais pas non plus proportionnelle, pour laisser la place à l'étiquette que je centre alors. Je rajoute également un tampon car une information foncée sur du clair attire le regard et c'est ce que je cherche

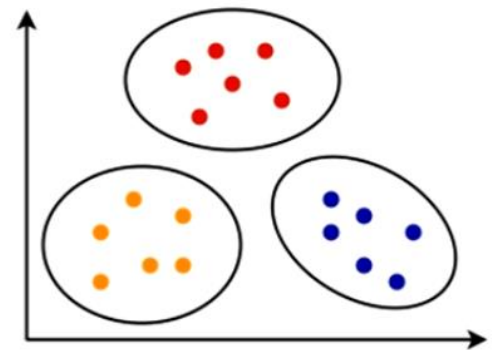


Mes choix

- Comme Thileli, je crée des points à l'aide de l'outil QGIS "centroïdes" et je supprime les entités ne possédant pas de perte humaines au niveau P2+P3. J'utilise ensuite l'outil "partitionnement en K-moyenne" de QGIS car il s'agit de la méthode par défaut pour l'initialisation de clusters. Il permet de créer un identifiant commun pour chaque groupe de points en fonction de la distance 2D pour chaque entité en entrée. En d'autres termes, il vise à partitionner les entités en k clusters dans lesquelles chaque entité appartient au cluster ayant la moyenne la plus proche.
 - A cluster is defined as a collection of data points exhibiting certain similarities.

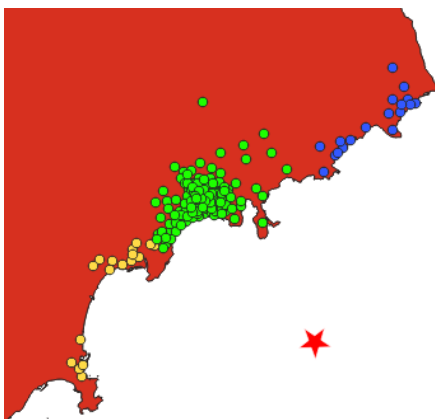


Before K-Means

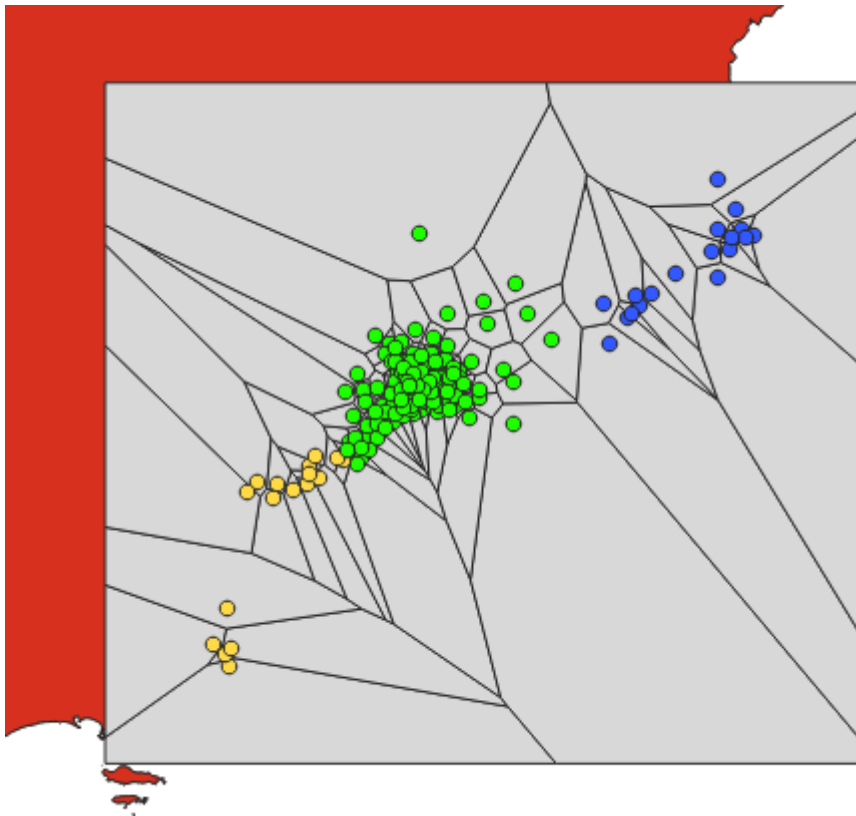


After K-Means

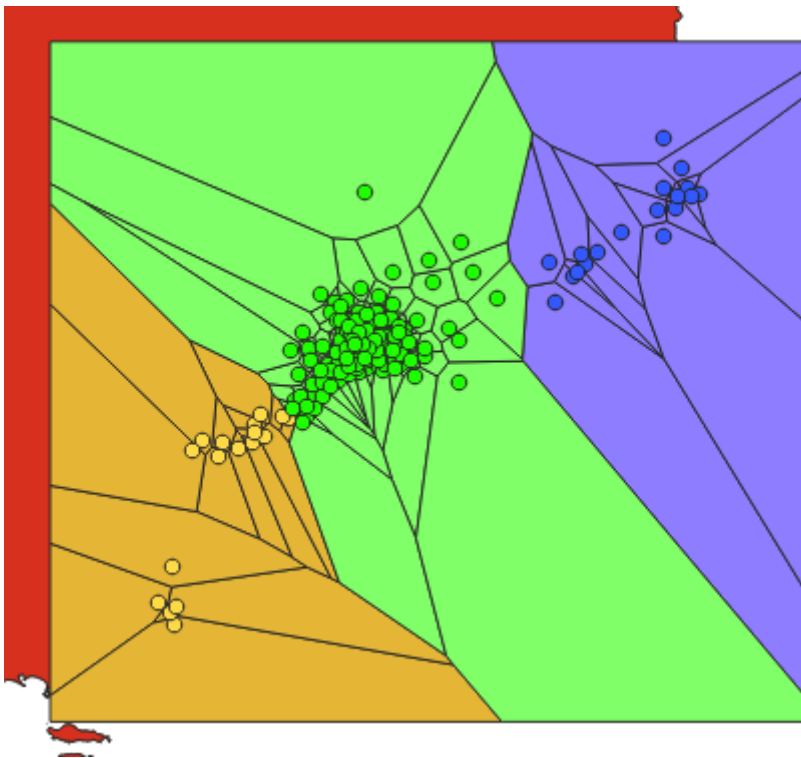
⇒ Cluster départemental : 3 clusters



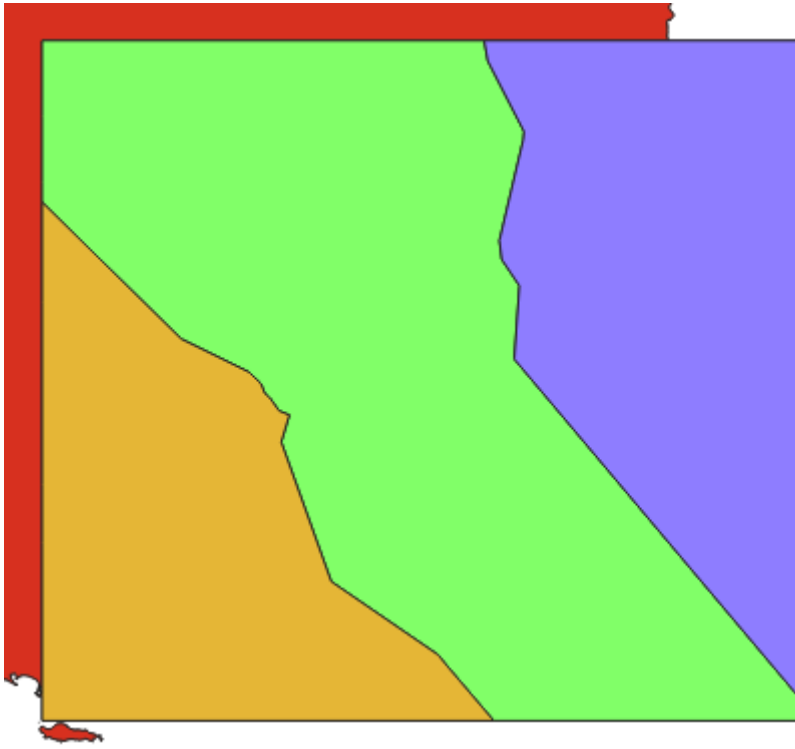
- création des polygones de voronoi autour de ces points



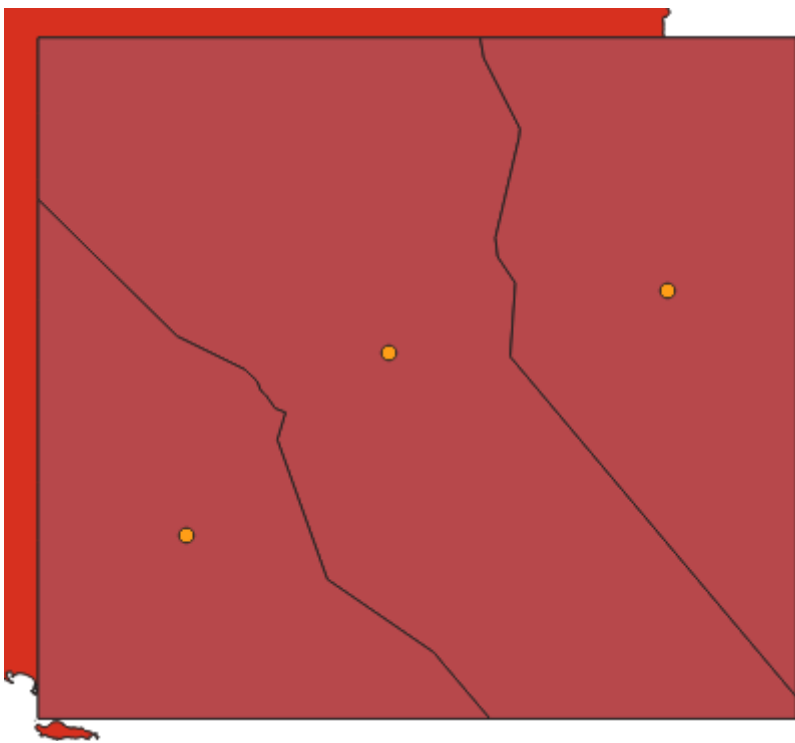
- sélectionner et exporter les polygones de Thiessen correspondant aux identifiants des clusters



- créer un champs ou j'additionne les valeurs des blessés par identifiant (à l'aide de l'outil "statistiques basiques pour les champs")
- regrouper et fusionner



- créer le centroïde de ces nouveaux polygones



- les clusters semblent parfois dans l'eau donc je les déplace manuellement pour les rapprocher du continent



- 2e cluster > régional = prendre le plus important des trois du cluster départemental et l'exporter
- 3e cluster > communal = 6 clusters
- 4e cluster > îlots IRIS = je choisis de procéder autrement que par localisation. Je préfère utiliser la table attributaire et regrouper les petites valeurs selon leur proximité. C'est donc une étape manuelle dans laquelle je peux prendre en compte les points isolés et mieux adapter mes choix

Objectifs de la semaine qui suit

1) Ecrire le code Openlayers

Je dois écrire un premier code sur lequel je vais importer toutes mes couches pour pouvoir définir pour chacune d'elles l'échelle adaptée. Le choix doit également se réfléchir en lien avec les autres données, et dans la fluidité de l'ensemble.

Malgré tout, un tel code charge une quantité assez importante de données pour chaque zoom, ce qui entraîne de la latence. C'est pourquoi je dois également écrire un deuxième code avec pour chaque échelle une agrégation de couches préalablement groupées sur geoserver. Je dois également m'assurer que ces agrégations soient mosaïquées pour permettre une meilleure fluidité de la carte.

a) Faire passer mes premiers entretiens

Mon premier entretien est mardi à 14h. Les autres ne sont pas encore fixés.