INTRODUCTION

The crime rate is always an important factor to describe the security of a town. The formula used in France is amount of crime per year*1000/ population [1]. As we can see from this formula, crimes are always related to population size.

This article tries to find out if we replace population size by other factors such as venue characters and figure out if there is a relation between venues and crime without considering the population effects.

Data sources:

French government publishes every year crimes registered by the gendarmes by commune on the official site. This report divides France into 35414 communes and gives our details the type de every crime. [2]

]:

	Code index	Libellé index \ CGD	CGD BELLEY	CGD BOURG EN BRESSE	CGD GEX	CGD TREVOUX	CGD CHATEAU THIERRY NOGENTEL	CGD LAON	CGD SOISSONS	CGD ST QUENTIN	...	CGD ST BENOIT	CGD ST PAUL	CGD ST PIERRE	C MAR BARTHE
0	Année 2017 - compagnies de gendarmerie	Départements	1.0	1.0	1.0	1.0	2.0	2.0	2.0	2.0	...	974.0	974.0	974.0	978.0
1	1	Règlements de compte entre malfaiteurs	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0

My study doesn't need to considerate crime type; therefore, we calculate the sum of all the crimes by commune.

Also, data.gov.fr offers the GPS address of all the communes registered in France. .[2] We can find longitude and latitude of every commune registered in France.

[3]:

	Code_commune_INSEE	Nom_commune	Code_postal	Libelle_acheminement	Ligne_5	coordonnees_gps
0	90078	PETITEFONTAINE	90360	PETITEFONTAINE	NaN	47.7237763721, 7.00757336099
1	90089	ROUGEMONT LE CHATEAU	90110	ROUGEMONT LE CHATEAU	NaN	47.7460113941, 6.95212889734
2	90091	ST GERMAIN LE CHATELET	90110	ST GERMAIN LE CHATELET	NaN	47.7002701756, 6.96114583666
3	90093	SERMAMAGNY	90300	SERMAMAGNY	NaN	47.687801557, 6.8309146345
4	90105	VILLARS LE SEC	90100	VILLARS LE SEC	NaN	47.4554088507, 6.98803547043

These two files above will be my data source and help me construct my analyses.

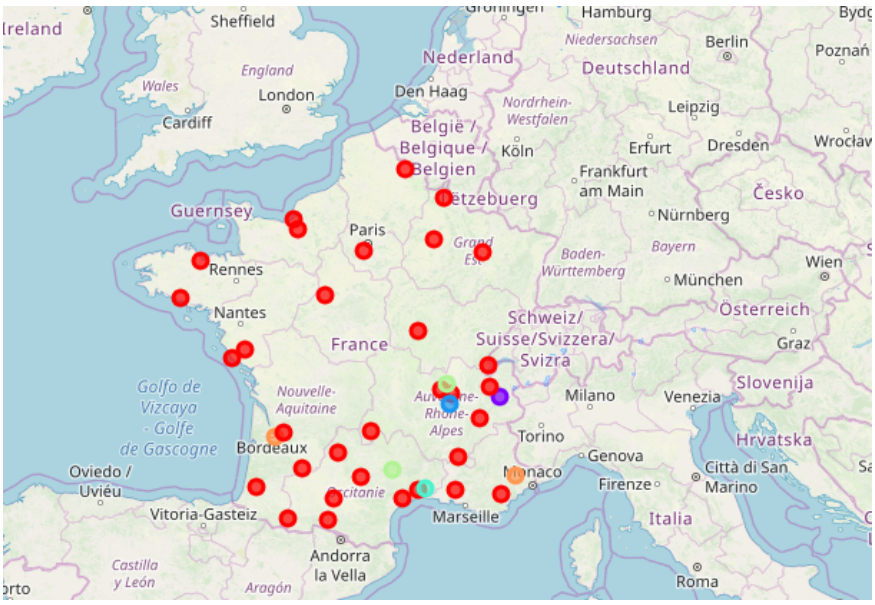
METHODOLOGY

Considering the important amount of data (35414 communes), I've chosen the 25 top and 25 last rankings of crime amount on the list after combing these 2 tables above.

	Nom_commune	Total_Crime	Code_postal	coordonnees_gps	x	y
0	CHARLEVILLE MEZIERES	0.0	8000	49.7752965803, 4.71724655966	49.775297	4.717247
1	NERAC	0.0	47600	44.1321004332, 0.343358299602	44.132100	0.343358
2	LA BOURBOULE	458.0	63150	45.5796236201, 2.7501805247	45.579624	2.750181
3	BARCELONNETTE	462.0	4400	44.3785614205, 6.65215089713	44.378561	6.652151
4	ARGELES GAZOST	698.0	65400	43.006001672, -0.0963626224791	43.006002	-0.096363
5	MAULEON LICHARRE	736.0	64130	43.2166409006, -0.883099093035	43.216641	-0.883099
6	AMBERT	753.0	63600	45.5557839461, 3.75625356567	45.555784	3.756254
7	PUGET THENIERS	853.0	6260	43.9521649219, 6.90609311435	43.952165	6.906093
8	AURILLAC	963.0	15000	44.9245233686, 2.44162453828	44.924523	2.441625
9	ALBI	986.0	81000	43.9258213622, 2.14686328555	43.925821	2.146863
10	CHATEAU CHINON VILLE	989.0	58120	47.0626511564, 3.92696760176	47.062651	3.926968

I've analyzed two factors affecting the crime amount: characters of venues nearby and number of avenues nearby.

On the one hand, to find avenue characters of the town, I used Foursquare API.[3] to figure out the venues around and then group them by category. To simply analyses, I took the 10 most popular avenues and used K-means to cluster. The result is presented then on a map.



Next, I added the amount of crime to this map above; the final map will give us visualization about the relationship between crime and venue characters. Normally, Choropleth map should be used to visualize the amount of crime by commune, however, the considerable number of commune in France containing in geojson file takes a lot of resources and time to download, the result couldn't be downloaded in notebook. Therefore, I chose Bubble map to visualize the number of crime by commune on the map.

On the other hand, this report studied also the relationship between amount of crime by commune and number of venues around. Foursquare API [3]was used to count the number of avenues by commune. Then I calculated the correlation coefficient between crime number, number of venues around and the clusters obtained.

Out[43]:

	Avenu	Cluster_Lables	Total_Crime
0	1	1	5518.0
1	20	0	986.0
2	4	0	8128.0
3	5	0	698.0
4	4	0	963.0

RESULTS

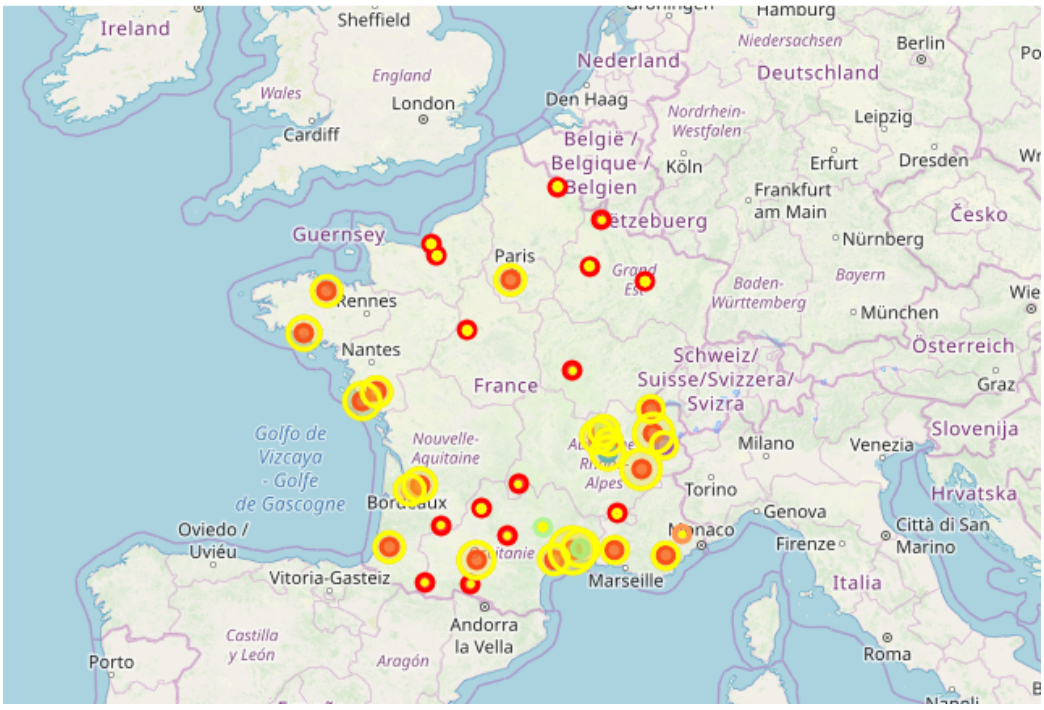
Unfortunately, I didn't find any relationship between them; the correlation coefficient is not significant between amount of crime, the venue clusters and number of venues:

1]:

	Avenu	Cluster_Lables	Total_Crime
Avenu	1.000000	-0.342901	-0.327720
Cluster_Lables	-0.342901	1.000000	0.041644
Total_Crime	-0.327720	0.041644	1.000000

Plus, we can get a conclusion from the map below, that the relation between venues cluster, which are distinguished by colors and number of crime presented by yellow bubble isn't significant:

Or we can say that the effects of venues characters over crime aren't significant.



DISCUSSION

Just like what I mentioned above, this report chose 50 communes from 33527 in the list, the result can't represent all the country.

What's more, several communes chosen are so thinly populated that Foursquare API hasn't enough venues information. We have to remove these communes during my analyses.

Maybe, communes aren't the best choice to be the unit of analysis. Region or department should be better.

In term of classification technique, I've chosen K-means as the method, however, as we know; it's difficult to evaluate if I've chosen the best number of cluster.

CONSLUSION

As a result, there's no significant relation between venues and crimes.

REFERENCES

1.http://medias.lemonde.fr/mmpub/edt/doc/20090831/1233615_29b3_tauxdecriminalitepparville.pdf

2. https://www.data.gouv.fr/fr/

3. https://developer.foursquare.com/places-api