

RAPPORT FINAL



Premier compte-rendu BI

La criminalité en France

BUSINESS INTELLIGENCE

Promo 2016
Professeur encadrant :
Laure SOULIER

Auteurs : Ince UMIT
Bouhrara ASSYA
Khattab DINA

Février 2017 — Mars 2017

Présentation de la problématique

Dans ce projet nous allons chercher à déterminer les facteurs pouvant favoriser la hausse de criminalité dans certaines régions. Nous allons notamment analyser des données par région telles que le chômage, le potentiel fiscal et le taux d'éducation. Nos recherches se centralisent sur de potentielles corrélations avec le taux de crimes, en cherchant à répondre aux questions :

- Les dispositifs de surveillance sont-ils réellement dissuasifs ?
- Le nombre de personnes à charge d'un policier est-il un indicateur de sureté ?
- Le niveau d'éducation global ainsi que celui du chômage impacte-t-il sur le taux de criminalité ?

C'est le sentiment d'insécurité à l'échelle nationale qui nous a encouragé à choisir ce sujet.

Les données

Nous avons utilisé plusieurs sources pour récolter des bases de données.

Géoclip : nous a permis de créer nous même les bases à partir des critères que l'on souhaitait analyser, notamment le chômage et le nombre de personnes non diplômées. De plus cette ressource nous permettait d'analyser l'évolution sur le temps de ces différents facteurs.

Wikipédia : pour le nombre de détenus et le PIB par pays c'est par capture web de pages Wikipédia que les bases ont été créées. Le reformatage des données XML a été fait par scrapping en Python.

Data-Gouv : nous a fourni une base de crimes organisée par département. Nous avons donc procédé à une fusion des données pour obtenir une base organisée par région.

INSEE : nous y avons récupéré une base pour l'éducation organisée par type de diplôme et selon le genre. De même pour la base chômage.

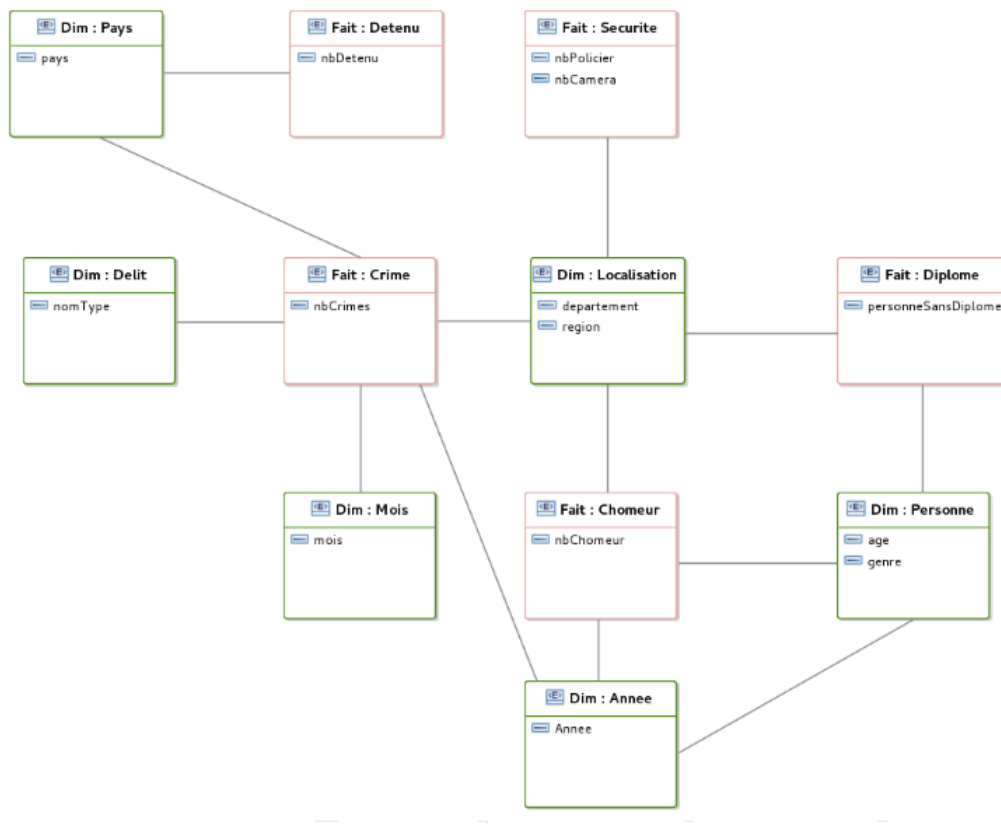
Un point traitement de données : Plusieurs traitements ont dû être appliqués sur les bases par le biais de Pentaho pour rendre les données plus facilement exploitables.

Il s'agissait notamment de supprimer des bases les Dom-Tom car leurs valeurs trop faibles allaient être insignifiantes et peu porteuse de sens. On a également fusionné les données des arrondissements d'une ville donnée pour avoir une seule valeur pour la ville, procédure notamment faite pour Marseille et Lyon.

Les crimes étaient stockés dans une base Excel où chaque page correspondait à un département. Pour pouvoir répartir les données par région on a dû faire un fichier *ktr* pour chacune des régions. Toutefois cette procédure fut très laborieuse le fichier Excel étant très volumineux et de plus on ne pouvait extraire que trois départements par transformation sur Pentaho. D'ailleurs Pentaho nous a également permis de trier les crimes par catégories *homicide*, *vol* et *autres*.

Présentation du schéma en constellation

Nous avons ici défini cinq faits pour vraiment couvrir une analyse assez large :



Sur le diagramme les classes vertes correspondent aux dimensions et les roses aux faits. Et le schéma logique correspondant est donné par :

Faits :

- Détenu(idDet, #idP, nbdetenu)
- Sécurité(idS, #idlocal, nbpolicier, nbcamera)
- Diplôme(idD, #idlocal, #idPers, personneSansDiplome)
- Chômeur(idC, #idPers, #idAn, #idlocal, nbChomeur)
- Crime(idCr, #idMois, #idP, #idlocal, #idDelit, nbCrimes)

Dimension :

- Pays(idP, pays)
- Localisation(idLocal, département, region)
- Personne(idPers, age, genre)
- Année(idAn, année)
- Mois(idMois, mois)
- Délit(idDélit, nomtype)

Quelques remarques :

Remarque 1 : Nous avons pris parti de considérer le PIB par pays comme étant un critère externe. Il n'est pas défini comme étant une dimension dans le schéma en constellation. Nous l'avons gardé pour avoir une idée au niveau mondiale de nos déductions.

Remarque 2 : Par ailleurs, un axe intéressant d'étude aurait été d'analyser l'impact du genre sur la criminalité. Nous avons trouvé une base concernant le niveau d'éducation par genre et une autre concernant le chômage par genre. Mais notre base de crimes n'est pas organisée par genre, on ne pouvait donc pas mettre en relation ces différents paramètres.

Remarque 3 : Enfin en ce qui concerne les dimensions temporelles, nous avons été dans l'obligation de faire deux dimensions distinctes. Une pour le 'Mois' et l'autre pour 'Année', car notre base

'Chômage' nous donne le nombre de chomeurs par année et non par mois.

La démarche de construction des cubes OLAP

La démarche de construction des cubes Olap consiste à importer la(les) table(s) de donnée(s) nécessaire(s) depuis phpMyAdmin pour un fait donné et puis à créer une vue qui permettra de visualiser les mesures selon les différentes opérations possibles sur les dimensions (rotation, rollup, drilldown,...)

Pour commencer, concernant le fait Sécurité, nous avons la table Localisation contenant les départements, le fait contient deux mesures : le nombre de policiers par département ainsi que le nombre de caméra de vidéosurveillance. Le cube OLAP obtenu est le suivant :

Département	Mesures		
	• NbCamera	• Personnes pour 1 policier	• Policier total
▣ All Dim localisation2.Départements	1 900	172 596,5	23 102
1	18	1 931,8	189
2	14	2 196,2	74
3	6	1 974,5	68
4	7	1 046,8	88
5	6	935,7	87
6	62	635,6	1 582
7	20	2 347,6	54
8	16	2 027,6	68
9	3	3 067,9	23
10	4	1 648,2	97
11	12	1 227,2	189
12	3	2 277,6	45
13	62	1 334,9	1 457
14	16	1 127,4	225
15	6	2 290	19
16	5	2 309,3	62
17	32	1 601	240
18	8	1 587,1	95
19	5	1 494,3	36
21	10	1 373,7	206
22	7	2 455	103
24	3	1 613	74
25	19	2 345,3	113
26	18	1 937,3	157
27	15	1 958,2	112
28	38	1 555,2	151
29	6	3 819,4	78
30	40	997,5	538
31	10	1 931,2	493
33	35	1 973,9	518
34	59	996,3	915
35	6	2 329,8	243
36	4	2 089,7	48

Ensuite concernant le fait Education, nous avons trois dimensions à savoir la localisation, le genre et la date (représentation des non diplômés par genre, département et date). Nous avons pu faire une visualisation a deux niveaux (drill down sur le genre) comme on peut le constater sur les deux images ci-dessous :

		Mesures
Date	Genre	• Nb sansdip
▢ All Dim date4.Dates	▢ All Dim genre2.Genres	38 842 220,8
2010	▢ All Dim genre2.Genres	8 108 198
2011	▢ All Dim genre2.Genres	7 994 719,6
2012	▢ All Dim genre2.Genres	7 896 755,7
2013	▢ All Dim genre2.Genres	14 842 547,5

		Mesures
Date	Genre	• Nb sansdip
▢ All Dim date4.Dates	▢ All Dim genre2.Genres	38 842 220,8
	femme	21 407 546,7
	homme	17 434 674,1
2010	▢ All Dim genre2.Genres	8 108 198
	femme	4 359 231,8
	homme	3 748 966,2
2011	▢ All Dim genre2.Genres	7 994 719,6
	femme	4 291 431,8
	homme	3 703 287,8
2012	▢ All Dim genre2.Genres	7 896 755,7
	femme	4 229 877
	homme	3 666 878,7
2013	▢ All Dim genre2.Genres	14 842 547,5
	femme	8 527 006,1
	homme	6 315 541,4

Fait chômage : trois dimensions genre, date et département. Un cube a été fait pour le chômage selon les départements, la date ainsi que le genre :

		Mesures		
		Chomage		
		Genre		
Date	• ▢ All Dim genre.Genres •	Femmes	• Hommes	
▢ All Dim date3.Dates	32 613	17 718,2	14 894,8	
1996	1 852,2	1 081,7	770,5	
1997	1 877,5	1 082,7	794,8	
1998	1 803,8	1 051,4	752,4	
1999	1 750,6	1 016,5	734,1	
2000	1 498,4	891,1	607,3	
2001	1 368,4	817,3	551,1	
2002	1 375,1	778,7	596,4	
2003	1 476,7	829,6	647,1	
2004	1 539,1	859,4	679,7	
2005	1 549,8	859,8	690	
2006	1 552,1	851,5	700,6	
2007	1 405,8	755,7	650,1	
2008	1 311,6	702,1	609,5	
2009	1 625,6	830,1	795,5	
2010	1 641,6	846,3	795,3	
2011	1 636,9	852,6	784,3	
2012	1 757,2	882,6	874,6	
2013	1 856,7	924,4	932,3	
2014	1 855,8	907	948,8	
2015	1 878,1	897,7	980,4	

Et enfin le fait crime s'exprimant selon les dimensions localisation, date et type de délit (avec les codes 1 : Homicides, 2 : Autre et 3 : Vols).

Nous avons choisi d'afficher les données qui nous ont semblées les plus parlantes de Janvier 2014 à Janvier 2017, c'est à dire ici le nombre de vol par mois

Puis le nombre d'homicides depuis 2012 par mois.

Voici un exemple de requête MDX permettant d'afficher les homicides (delit 1) en fonction du mois à partir de Janvier 2012 :

MDX Query Editor

Editeur MDX

```

select NON EMPTY {[Measures].[Delit], [Dim delit.Key-delit].[1]]} ON COLUMNS,
NON EMPTY {[Dim date.Date].[2012_01], [Dim date.Date].[2012_02], [Dim date.Date].[2012_03], [Dim date.Date].[2012_04],
[Dim date.Date].[2012_05], [Dim date.Date].[2012_06], [Dim date.Date].[2012_07], [Dim date.Date].[2012_08], [Dim date.Date].[2012_09], [Dim date.Date].[2012_10], [Dim date.Date].[2012_11], [Dim date.Date].[2012_12], [Dim date.Date].[2013_01], [Dim date.Date].[2013_02], [Dim date.Date].[2013_03], [Dim date.Date].[2013_04], [Dim date.Date].[2013_05], [Dim date.Date].[2013_06], [Dim date.Date].[2013_07], [Dim date.Date].[2013_08], [Dim date.Date].[2013_09], [Dim date.Date].[2013_10], [Dim date.Date].[2013_11], [Dim date.Date].[2013_12], [Dim date.Date].[2014_01], [Dim date.Date].[2014_02], [Dim date.Date].[2014_03], [Dim date.Date].[2014_04], [Dim date.Date].[2014_05], [Dim date.Date].[2014_06], [Dim date.Date].[2014_07], [Dim date.Date].[2014_08], [Dim date.Date].[2014_09], [Dim date.Date].[2014_10], [Dim date.Date].[2014_11], [Dim date.Date].[2014_12], [Dim date.Date].[2015_01], [Dim date.Date].[2015_02], [Dim date.Date].[2015_03], [Dim date.Date].[2015_04], [Dim date.Date].[2015_05], [Dim date.Date].[2015_06], [Dim date.Date].[2015_07], [Dim date.Date].[2015_08], [Dim date.Date].[2015_09], [Dim date.Date].[2015_10], [Dim date.Date].[2015_11], [Dim date.Date].[2015_12], [Dim date.Date].[2016_01], [Dim date.Date].[2016_02], [Dim date.Date].[2016_03], [Dim date.Date].[2016_04], [Dim date.Date].[2016_05], [Dim date.Date].[2016_06], [Dim date.Date].[2016_07], [Dim date.Date].[2016_08], [Dim date.Date].[2016_09], [Dim date.Date].[2016_10], [Dim date.Date].[2016_11], [Dim date.Date].[2016_12], [Dim date.Date].[2017_01]]} ON ROWS
from [crimef]

```

Accepter Remettre

Globalement, la démarche a été de représenter le cube olap avec l'ensemble des dimensions pour chacun des faits, ainsi à partir de ce cube il fut possible de faire certains choix de représentation : affiner ou pas le grain pour l'étude d'une caractéristique. Pour l'éducation par exemple, il a été possible d'augmenter l'expressivité des données en représentant le genre, si le nombre de chômeur nous interesse plus que le nombre de chômeurs par genre, on choisira de ne pas mettre en evidence le genre. (principe du drilldown/rollup)

Nous avons cherché à hiérarchiser, mais nous n'y somme pas parvenu. Nous avons constaté que sous JPivot, nous pouvions arranger les hiérarchies nous même. L'idée fut alors de dénormaliser les départements pour les avoir en colonne, devenant alors des données pouvant être ensuite groupables par région à la main sur JPivot. Par contre à l'affichage nous avons les régions et les departements aux mêmes niveaux et cela n'aurait eu aucun sens d'avoir par ligne les régions avec tous les departement sur colonnes.

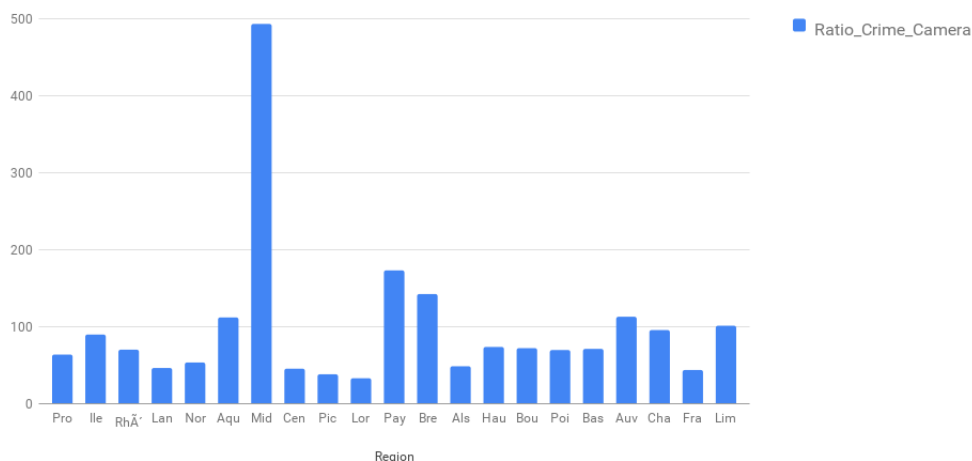
Les résultats & interprétations

Les graphes suivants ont été obtenus à l'aide de GOOGLE charts et les données ont été formatées pour obtenir la sortie qui est dans les fichiers *html*. deux d'entre eux proviennent du cube OLAP.

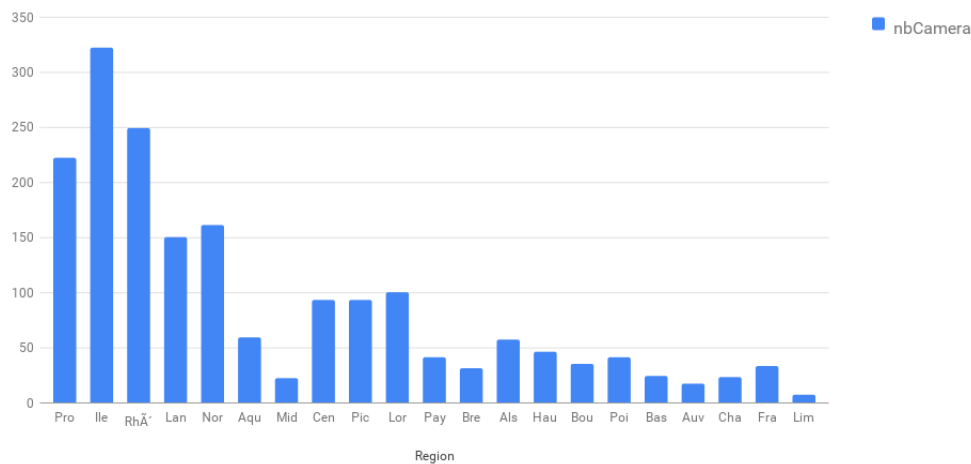
Sécurité : Dans un premier temps nous avons cherché à vérifier si les dispositifs de surveillance, tels que la vidéosurveillance avaient un réel impact sur le nombre de crimes. Pour ce faire nous avons étudié le ratio

$$r_1 = \frac{\text{nombre de crimes}}{\text{nombre de camras}}$$

pour chaque région. L'histogramme qui récapitule les résultats est présenté ci-dessous :



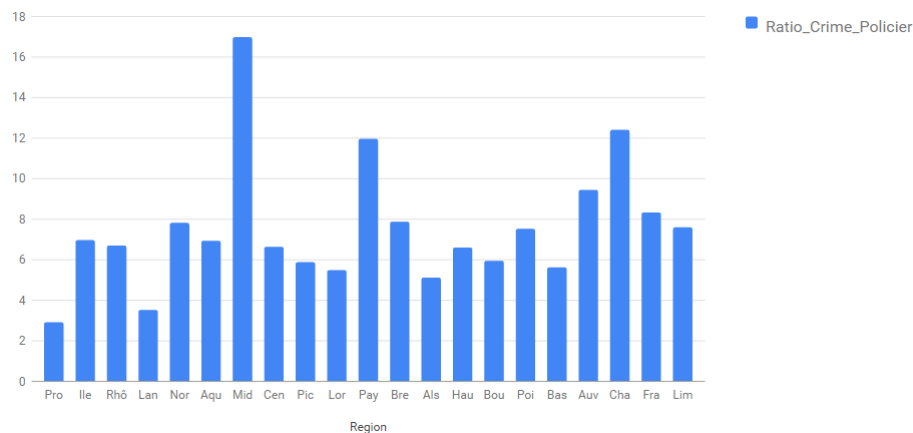
On constate un fort ratio pour les Midi-Pyrénées. De façon parallèle sur l'histogramme des caméras



on note de façon évidente que le nombre de caméras en Midi-Pyrénées est très faible. D'ailleurs dès que le ratio r_1 est élevé on constate un nombre assez faible de caméras. Ainsi l'augmentation du niveau de vidéo-surveillance d'une région donnée influe sur la diminution du nombre de crimes qui y est constatée.

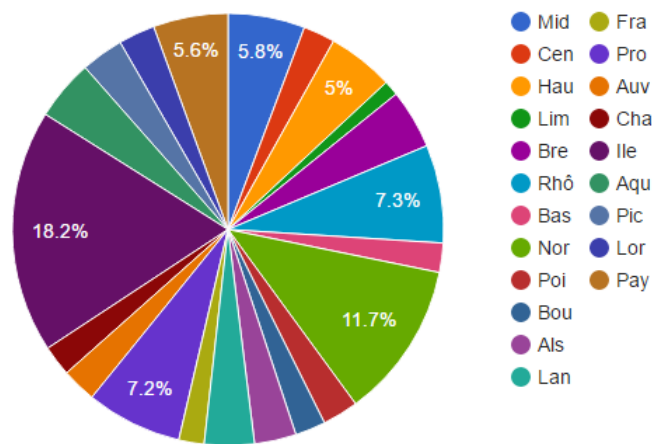
Par ailleurs pour confronter le taux de crimes avec les effectifs de policiers on analyse pour chaque région le ratio :

$$r_2 = \frac{\text{nombre de crimes}}{\text{nombre de policiers}}$$



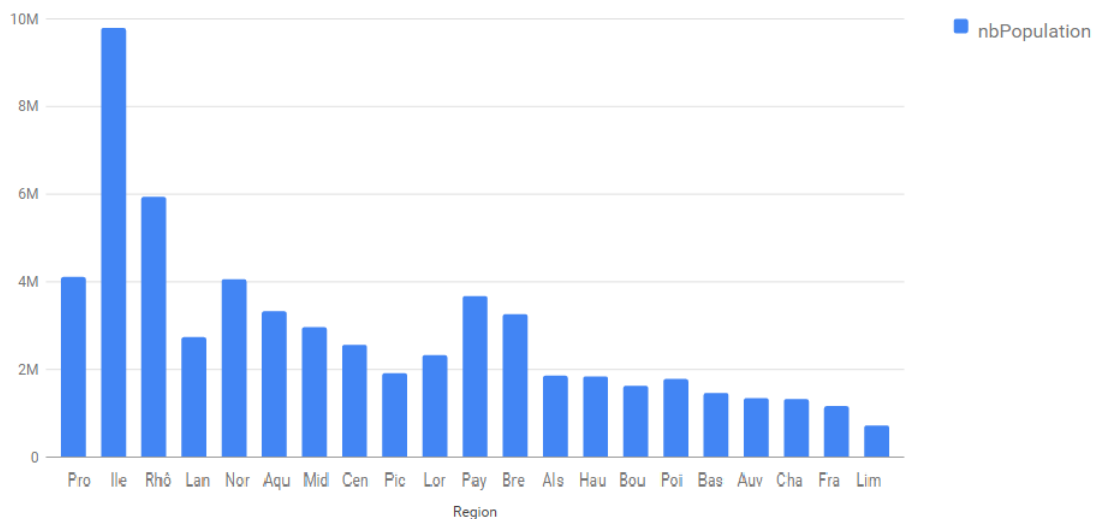
On peut noter que les régions qui ont un ratio plus élevé correspondent à celles qui possèdent également un ratio pour r_1 plus élevé. Ce qui amène à déduire que les régions avec un faible effectif de policiers et moins d'équipements de vidéo-surveillance sont confrontées à un plus grand taux de criminalité.

Les ratios ne nous permettant pas d'avoir un ordre de grandeur des valeurs réelles, nous avons modélisé un diagramme circulaire pour évaluer la proportion de vols d'une région donnée à l'échelle nationale. On se retrouve alors avec le diagramme suivant :



Une corrélation étrange se dessine alors sur ce diagramme. En effet, on voit que les régions associées aux parts les plus grandes correspondent aux régions possédant un système de vidéo-surveillance assez bien déployé. Intuitivement on devrait s'attendre au résultat inverse.

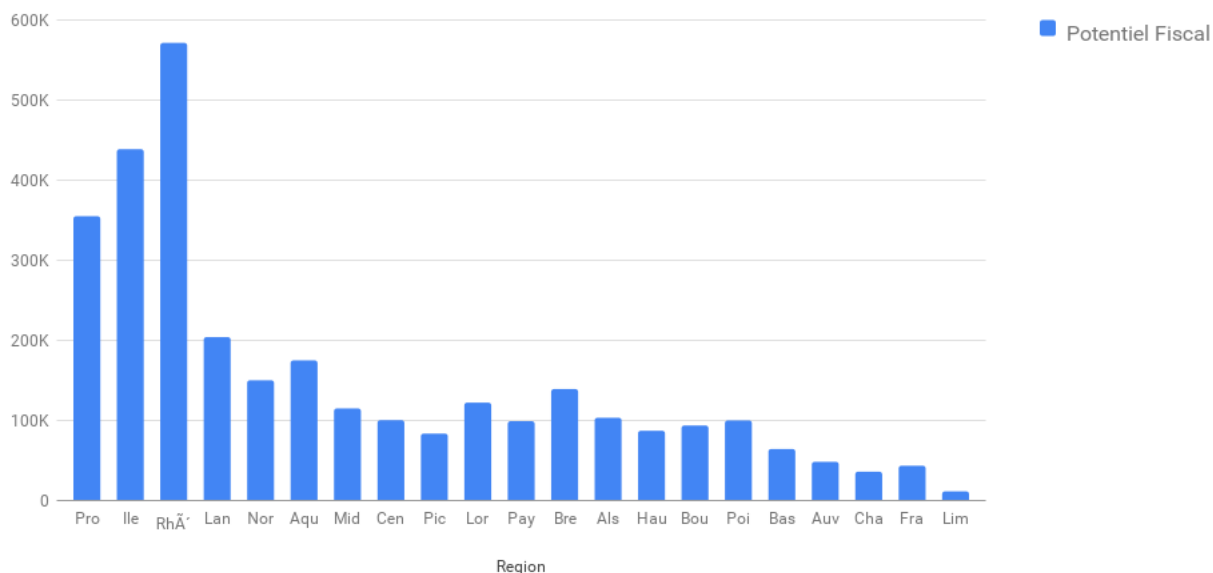
En comparant avec la population par région modélisée dans le graphe ci-dessous :



on peut essayer d'évaluer le taux de criminalité. Par exemple, la Normandie compte 2 millions d'habitants de moins que Rhône-Alpes alors qu'elle représente 4 points de plus concernant le nombre de crimes à l'échelle nationale. D'ailleurs la région PACA qui possède exactement le même nombre d'habitants que la Normandie a également 4 points de moins au niveau du diagramme circulaire. Ainsi on peut en juger que relativement à sa population la Normandie a une forte concentration de crimes contrairement à l'Ile-de-France, qui relativement à sa population possède un taux de criminalité relativement peu élevé. Ainsi le facteur "population de la région" était à prendre en compte pour interpréter la pertinence du précédent diagramme circulaire.

On peut donc bien conclure qu'une augmentation du nombre de caméras et de policiers implique une diminution du nombre de crimes.

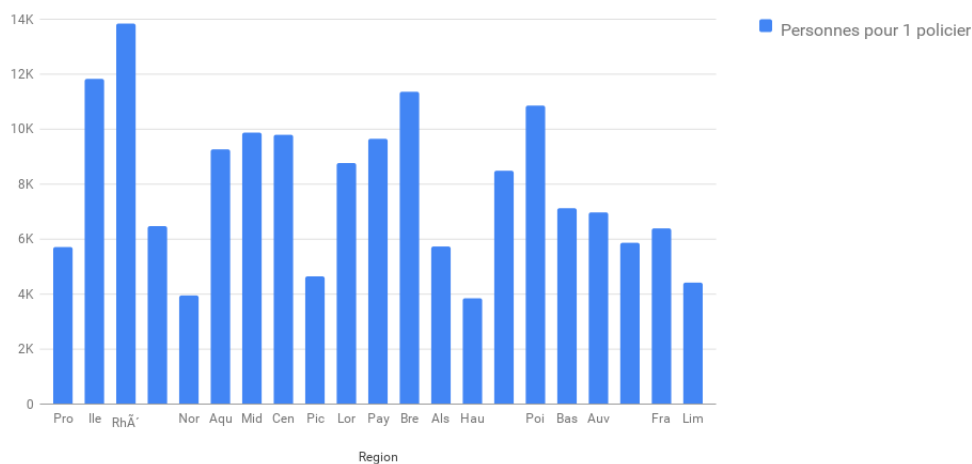
Indicateur de richesse : la valeur qui nous permet d'évaluer la richesse d'une région est donnée par le potentiel fiscal. L'histogramme représentatif de cette valeur est donné par l'histogramme :



On a cherché à voir les répercussions de ces chiffres sur le ratio :

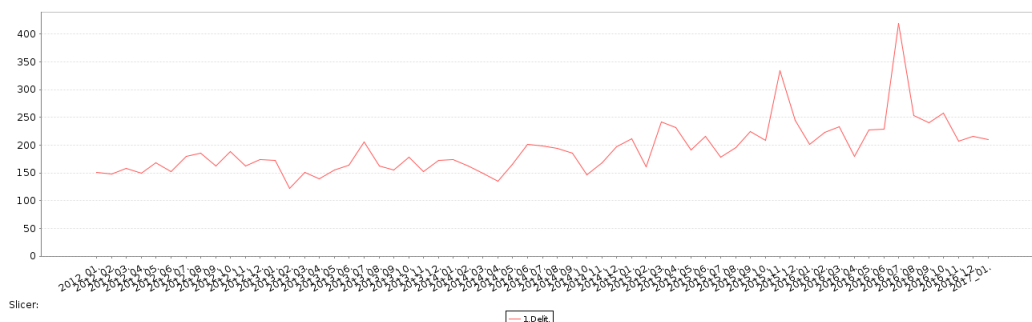
$$\frac{\text{nombre d'habitants}}{\text{nombre de policiers}}$$

par région, dont l'histogramme est le suivant : une conséquence semble se dessiner. En effet on re-



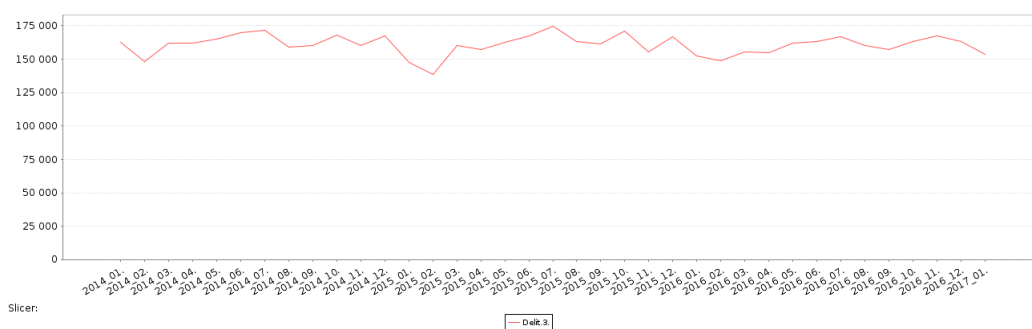
marque que plus le potentiel fiscal est élevé et plus le nombre de personnes par policier est élevé. On peut donc en déduire que les régions où cette corrélation est forte doivent voir leur nombre de policiers augmenter dans le but de diminuer le taux de criminalité.

Evolution temporelle : Dans un premier temps on a regardé l'évolution du nombre de meurtre par mois au cours du temps.



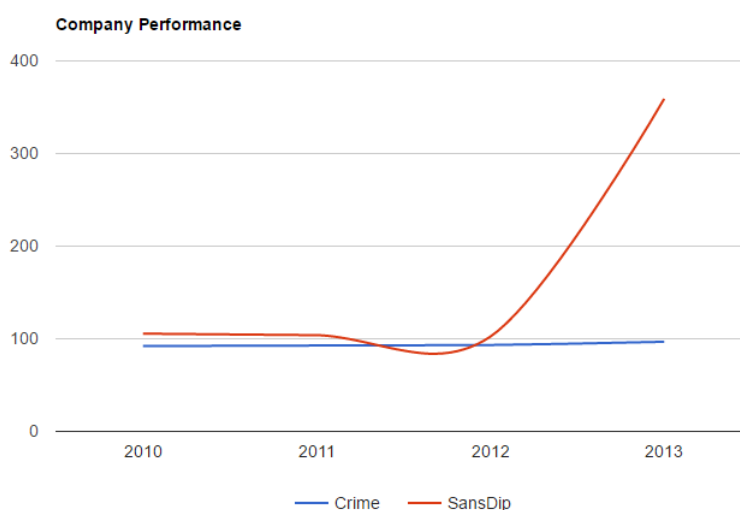
On note clairement que c'est toujours entre les mois de février et avril que le minimum est atteint et le maximum de meurtres est toujours atteint durant les vacances d'été (excepté le pic en novembre 2015 qui fait echo aux regrettés attentats du Bataclan). On voit de façon évidente que le nombre de meurtre a fortement augmenté durant les dernières années.

Par la suite nous avons analysé l'évolution temporelle du nombre de vols.



On remarque qu'il y a une forme de périodicité et que pour chaque année le minimum est atteint aux environs du mois de février et que sans surprise la plus forte hausse est visible durant les vacances. Toutefois le nombre de vols reste relativement constant durant la période étudiée. On peut donc par exemple, proposer un plus grand renforcement sécuritaire pour les périodes où il y'a un pic.

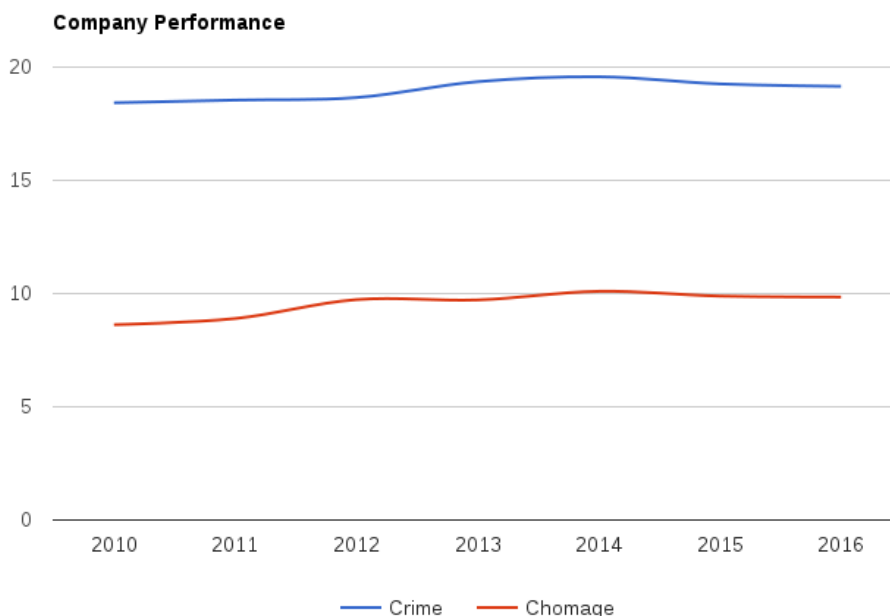
Éducation & chômage : A présent on cherche à analyser l'évolution au niveau national du nombre de non-diplômés et du nombre de crimes entre 2010 et 2013.



A première vue le niveau d'éducation ne semble pas être un facteur amplificateur de hausse de criminalité. Toutefois en regardant de façon plus attentive les données, on constate qu'aux environ de

2010 le nombre de crimes est de l'ordre de 92 000, alors qu'en 2013, au moment où le nombre de personne sans diplôme a augmenté, le nombre de crime est passé à 96 000. Cette hausse du nombre de crimes est tout de même assez significative, c'est uniquement la différence d'échelle entre les deux types de données qui donnait l'impression d'une non corrélation. Ainsi les précédents résultats semblent inciter à penser que les deux faits sont liés.

Nous avons ensuite cherché une corrélation entre le chômage et le nombre de crimes.



on constate que les deux courbes se suivent parfaitement. Cela semble indiquer que le nombre de crimes suit la tendance définie par le chômage.

Par notre étude partielle, l'éducation et le chômage apparaissent comme des facteurs amplificateurs de la hausse de criminalité. Les chiffres restent assez préoccupants car la tendance reste à la hausse durant les deux dernières années, ce qui est un indicateur de la non performances des réformes en cours sur le sujet. Nous aurions pu d'avantage aboutir notre analyse dans ce sens, toutefois voyant que la démarche allait être similaire à celle pour la sécurité nous avons préféré ne pas alourdir ce rapport en graphiques.

Cartographie à l'échelle mondiale

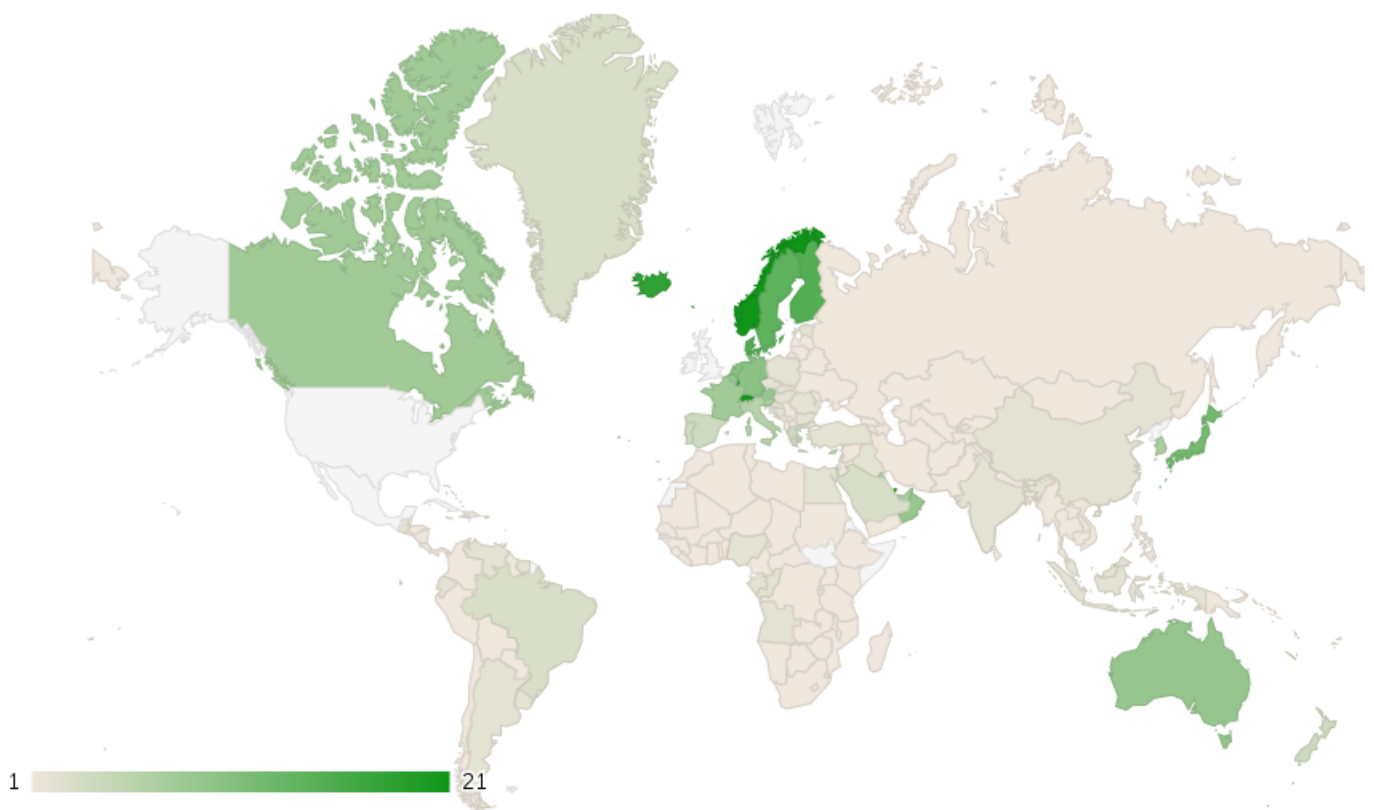
Afin d'évaluer explicitement si "le PIB du pays crée un quelconque enjeu au niveau de la hausse de criminalité" nous avons étudié le ratio :

$$\frac{PIB}{nombre\ dtenus}$$

Plus la teinte est foncé et plus le ratio est élevé.

Au moment du scraping les données associées aux Etats-Unis n'ont pas pu être collectées pour une raison à première vue de formatage de balise XML.

Afin d'avoir des données intéressantes à analyser, les données ont été discrétisées par intervalles. De la sorte les fortes valeurs ne viennent pas biaiser les résultats. On voulait confirmer le fait que "un fort PIB était indicateur d'un faible nombre de détenus" et donc un indicateur pour un taux de criminalité faible.



Clairement on voit les pays développés comme étant de teinte plus foncée et cela nous incite à penser que notre préalable hypothèse était bien fondée.

Conclusion

Sachant que notre analyse n'offrait pas énormément de dimensions, nous avons cherché à élargir notre étude sur les faits à analyser. L'ensemble de nos résultats porte à croire que le déploiement de système de surveillance à un pouvoir assez dissuasif. On a pu constater que dans plusieurs régions les équipements de vidéo-surveillance devaient se revoir à la hausse pour avoir un ratio crimes policiers d'avantage dans les normes.

En effet, juste pour citer un exemple, la Normandie représente à l'échelle nationale une proportion seulement 6% plus faible que l'Ile de France pour une population 1.5 fois moins nombreuse. Leur effectif de policiers semble tout autant dissuasifs étant donné que leurs ratios crimepoliciers sont du même ordre de grandeur. Pourtant en Normandie un policier à trois fois moins de personnes à charge qu'un policier en Ile de France. Ainsi augmenter le nombre de policiers ne semble pas être la solution adéquat, mais investir dans un équipement de surveillance semble plus pertinent vu le fort écart entre ces deux régions sur ces chiffres(cf. graphe caméras).

Notre étude nous porte à penser que l'éducation et le chômage apparaissent comme facteurs amplificateurs de la hausse de criminalité. Nous n'avons pas d'avantage approfondi l'analyse dans ce sens car nous n'aurions pas pu exprimer un plan d'action concret pour recourir à une hausse du chômage et diminution d'instruction. Toutefois ces résultats nous auraient permis de définir les zones prioritaires.

D'un point de vu mondial, la France ne semble pas être la plus touchée par une population carcérale

élevée. Est-ce réellement un indicateur d'une plus faible criminalité ? La question nous restera sans réponse car on sait pertinemment que les systèmes juridiques ne sont pas les mêmes à travers le monde. En effet un pays autorisant la peine de mort ne peut pas réellement comparer ses effectifs carcéraux avec un pays qui ne la pratique pas et cela vient biaiser nos résultats.

Nos sources

Data.Gouv.fr :

- <https://www.data.gouv.fr/fr/datasets/chiffres-departementaux-mensuels-relatifs-aux-crimes-et-delits-enregistres-par-les-services-de-police-et-de-gendarmerie-depuis-janvier-1996/>
- <https://www.data.gouv.fr/fr/datasets/polices-municipales-et-richeesse-des-communes/>
- <https://www.data.gouv.fr/fr/datasets/villes-sous-videosurveillance/>

Insee.fr :

- <https://www.insee.fr/fr/statistiques/2386698consulter>
- <http://www.bdm.insee.fr/bdm2/choixCriteres?codeGroupe=1820codeRetour=2>

Wikipedia.fr :

- https://fr.wikipedia.org/wiki/Liste_des_pays_par_population_carc%C3%A9rale
- https://fr.wikipedia.org/wiki/Liste_des_pays_par_PIB_nominal

Geoclip.fr :<https://franceo3.geoclip.fr/index.php?profil=FD>