

# Mise au point d'un modèle prédictif lors d'un accident de la route pour jeunes conducteurs

*Hackathon IAMD- 2014*

*Sujet Proposé Par:*



Rémi SELLEM

*Réalisé par :*

*Ayoub MHARZI*

*Vincent DI MARIA*

*Yasmina HMAMOUCHE*

*Zineb OUALI ALAMI*

## Sommaire :

Préambule .....	3
I- Jeu de données.....	3
II- Calculs statistiques.....	4
III- Visualisation et conclusions.....	5
Annexe.....	

## Préambule

Dans le cadre de la première édition du Hackathon IAMD organisée par l'école TELECOM Nancy et en collaboration avec le crédit agricole et inTech, notre équipe a travaillé sur le sujet proposé par le Crédit Agricole. Il s'agit de la mise au point d'un modèle prédictif lors d'un accident de la route pour les jeunes conducteurs. Ce document regroupe les différents calculs statistiques réalisés avec l'outil R sur un jeu de données d'Etalab, ainsi que les conclusions tirées de ces calculs.

## I- Jeu de données

La base de données retenue est extraite du fichier national des accidents corporels de la circulation administré par l'observatoire interministériel de la sécurité routière.

Pour chaque accident corporel (soit un accident survenu sur une voie ouverte à la circulation publique, impliquant au moins un véhicule et ayant fait au moins une victime ayant nécessité des soins), les saisies d'information sont effectuées par l'unité des forces de l'ordre qui est intervenue sur le lieu de l'accident. La base répertorie l'intégralité des accidents corporels de la circulation intervenus de 2006 à 2011 en France, avec leur description simplifiée. Cependant pour notre étude on a décidé de retenir que les accidents survenus en Île de France.

Cette base comporte 454 372 accidents décrits par 39 variables.

Après une analyse des données (étude de fiabilité et persistance de données) on en a retenu que les variables suivantes:

- Dep et Com : qui représentant les codes de départements et de commune de chaque accident. Seuls les Dep = 75,77,78,91,92,93,94,95 ont été retenus.
- Ttue, Tbg, Tbl : Nombre de personnes tuées, blessées gravement et légèrement dans chaque accident
- Grav : Indice de gravité, calculé selon le coût des atteintes aux victimes.
- Lum : Luminosité graduée de 1 à 5:
  - 1 – Plein jour
  - 2 – Crépuscule ou aube
  - 3 – Nuit sans éclairage public
  - 4 - Nuit avec éclairage public non allumé
  - 5 – Nuit avec éclairage public allumé
- Atm : Conditions atmosphériques lors de l'accident, graduée de 1 à 9 :
  - 1 – Normale
  - 2 – Pluie légère
  - 3 – Pluie forte
  - 4 – Neige - grêle
  - 5 – Brouillard - fumée
  - 6 – Vent fort - tempête
  - 7 – Temps éblouissant
  - 8 – Temps couvert
  - 9 – Autre

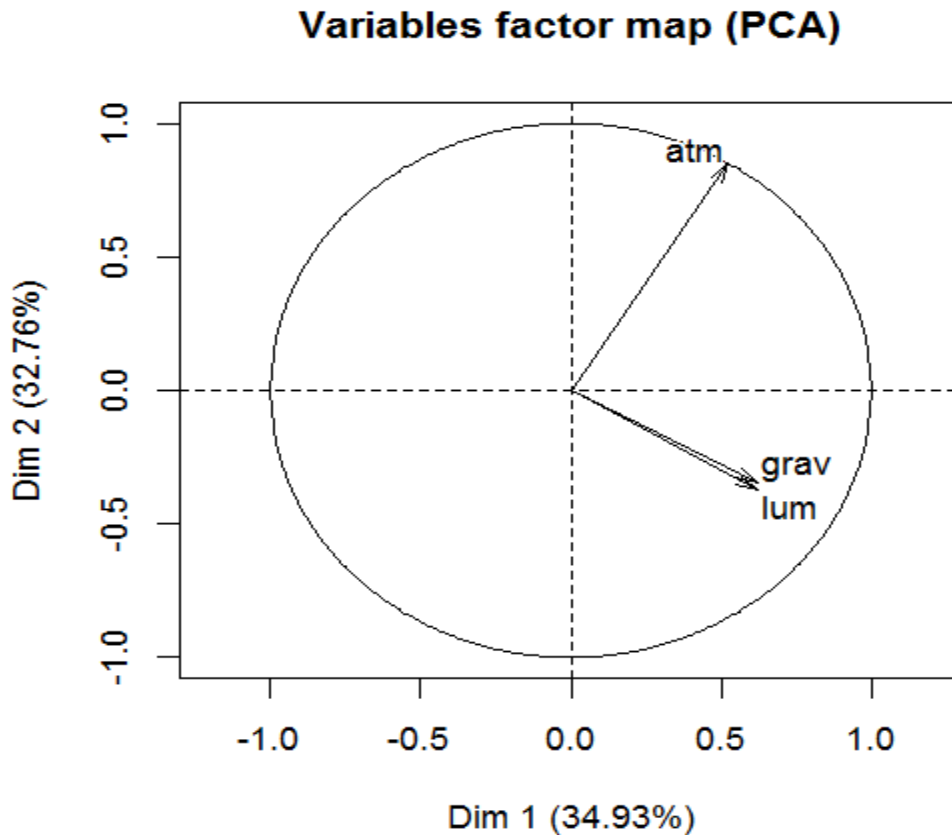
## II- Calculs statistiques

Une fois le jeux de données choisi , nous avons décidé de définir un scénario test pour réaliser les différents calculs. Notre choix s'est porté alors sur l'île de France c'est à dire les départements ( 75,77,78,91,92,93,94 et 95 ), nous avons alors réalisé des scripts sous le logiciel R afin de filtrer ces données et aussi les soumettre aux différents calculs. Pour cela nous nous sommes aidés de la librairie “**dplyr**” qui facilite la manipulation des données grâce aux différentes fonctions qui y sont intégrées. Les scripts déjà cités permettent de récupérer deux tables représentant le nombre d'accident par commune de chaque département et aussi le niveau de gravité de chacun. Les valeurs NA existantes dans les deux tables représentent les communes inexistantes ou dont on a pas de valeurs.

Une fois les zones de gravité définies, il a fallu réaliser des tests pour définir l'impact des conditions météorologiques sur la gravité et la fréquence des accidents. La première étape était de définir la corrélation entre les différentes variables représentant la météo et l'indice de

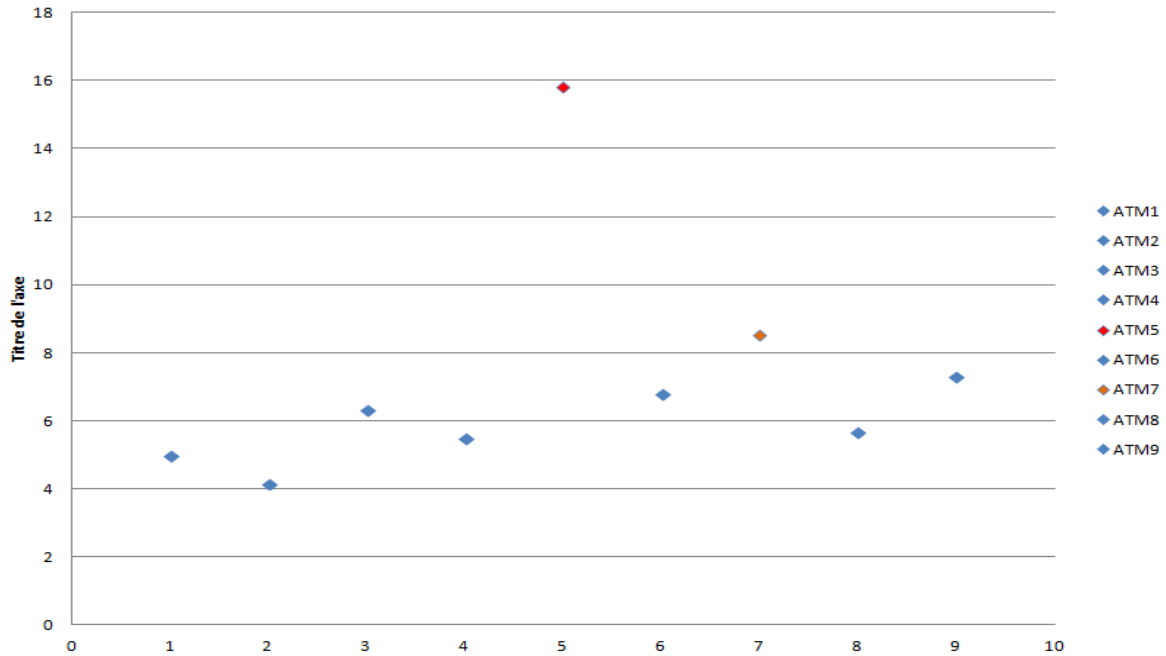


gravité fournis dans la base de données. Pour cela nous avons réalisé une Analyse des Composantes Principales (ACP) sur trois valeurs de notre base de données et qui sont : **atm**, **grav**, **lum**. On obtient alors le schémas suivant.



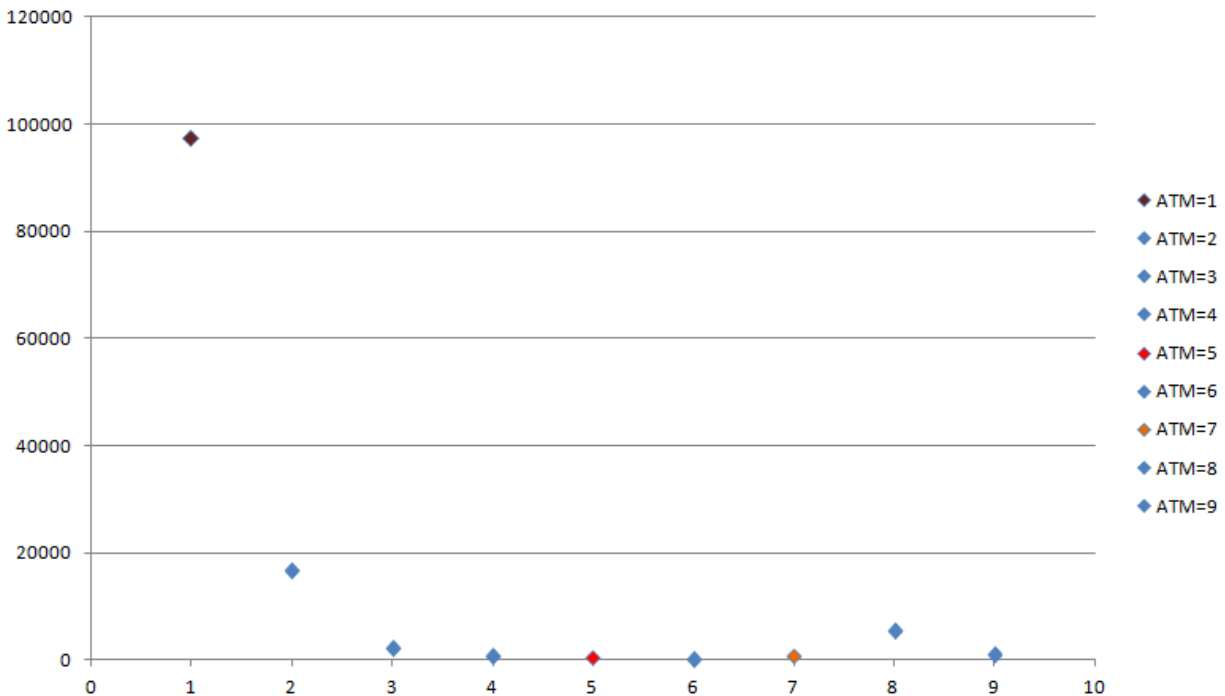
On peut constater du schéma ci-dessus que la gravité d’un accident est liée avec l’indice de luminosité mais que la variable **atm** n’affecte pas la gravité des accidents dans ce plan. Pour avoir une vision plus claire de l’impact de la variable “atm” sur l’indice de gravité nous avons décider de calculer l’indice moyen sur les différentes catégories présentes et le représenter dans le graphe suivant.

## Indice de gravité



On remarque un pic pour la valeur atm=5 (qui désigne brouillard et fumé). Il a fallu aussi définir la relation entre les conditions de la météo et la fréquence des accidents. Le calcul réalisé nous a permis de tracer la figure suivante.

## Nombre d'accident



On remarque que d'après ce schéma la plus grande partie des accident dans la région Parisiennes est commise dans des conditions météo et une luminosité normale et non pas comme on peut le croire dans une mauvaise météo. Cela peut être interpréter de plusieurs façons notamment que les conducteurs sont plus prudents et respectent plus les signalisations ce qui n'est pas toujours le cas dans les jours normaux.

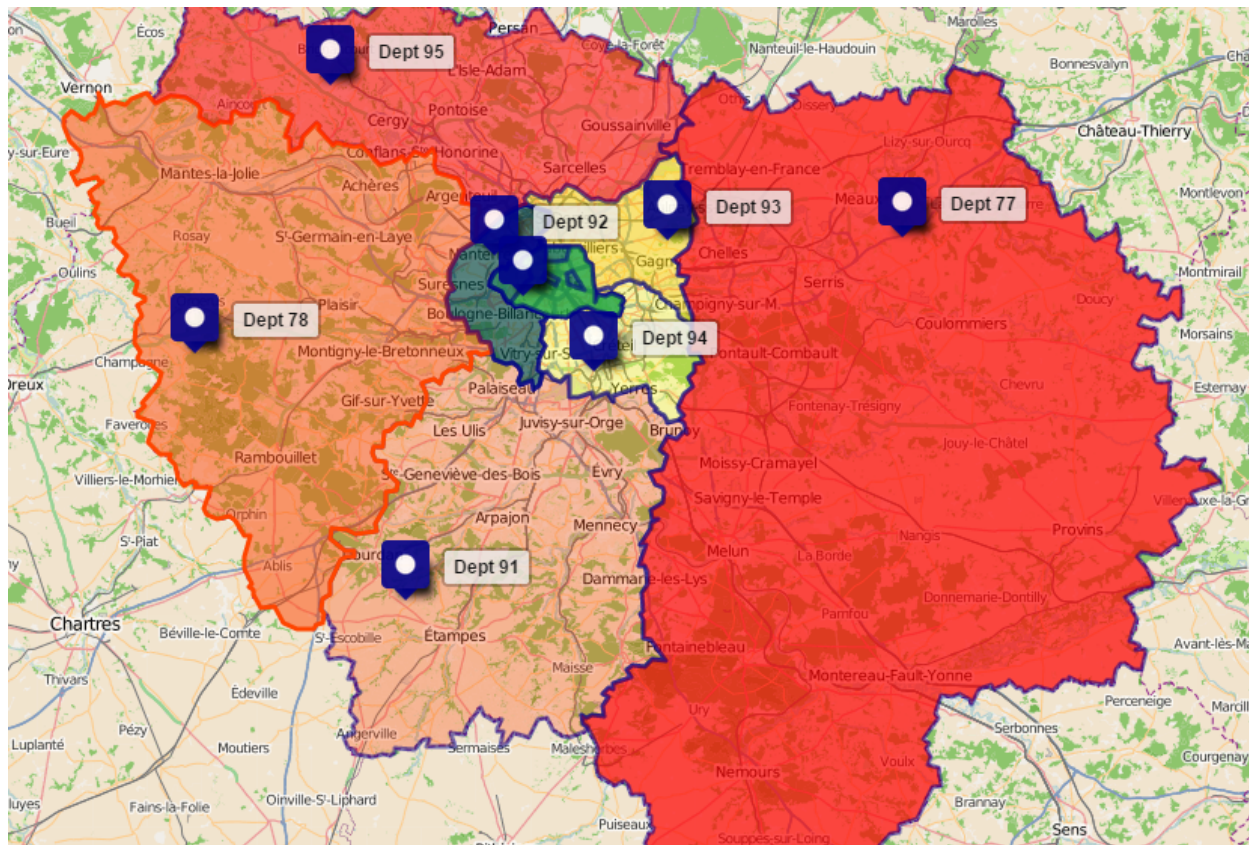
### III- Visualisation et conclusions

Après les calculs réalisés pour analyser les données, et en prenant en compte le temps restant, nous avons jugé pertinent de s'arrêter au niveau des calculs et de réaliser une maquette pour la visualisation des résultats.

Les maps ci-dessous sont tirées d'OpenStreetMap. Pour une visualisation pertinente, elles sont graduées du rouge (plus important) au vert. Aussi les délimitations géographiques des communes sont obtenues d'un fichier GeoJson.

**1 ère conclusion :**

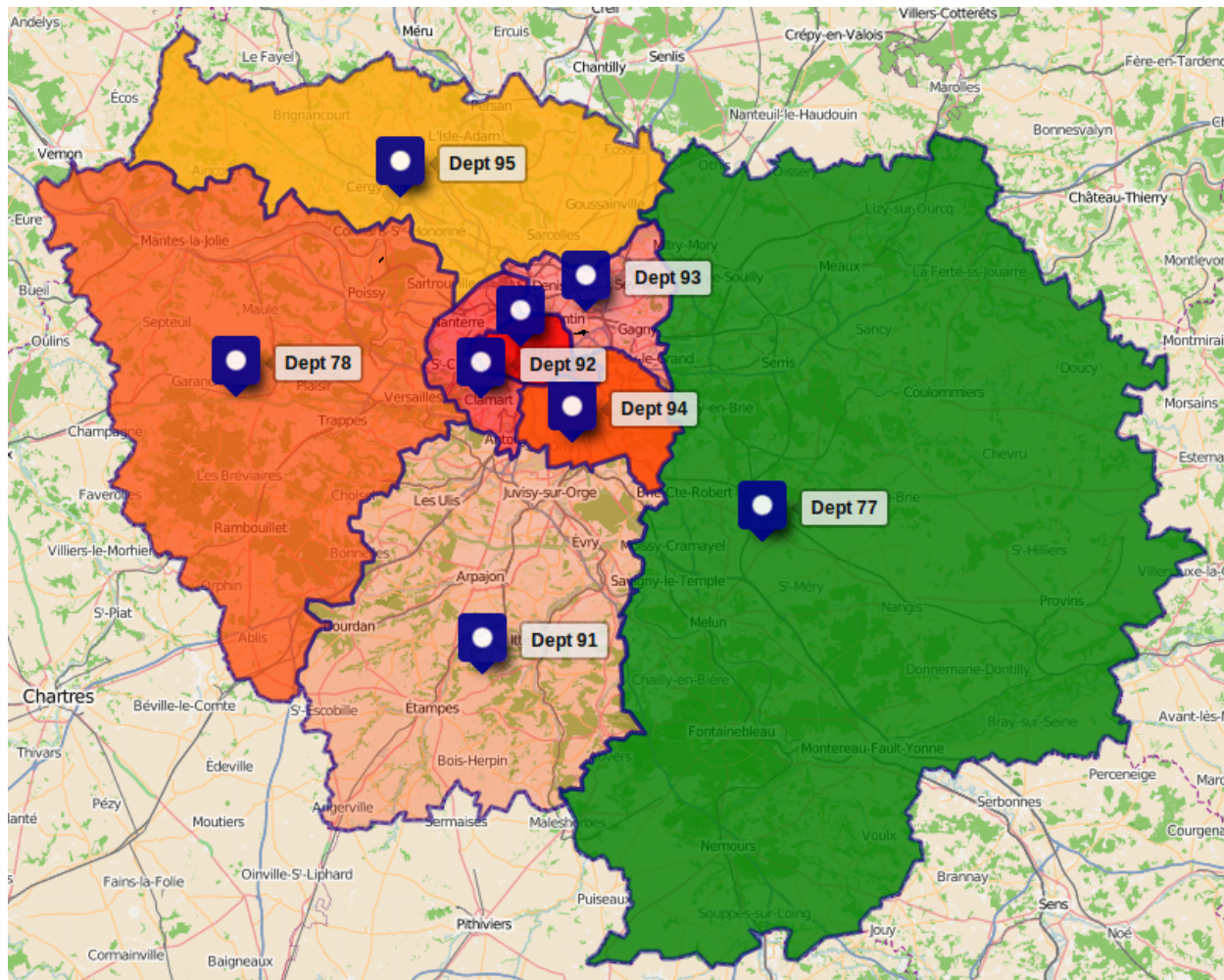
La gravité de l'accident est plus importante en Île de France contrairement au centre de Paris.





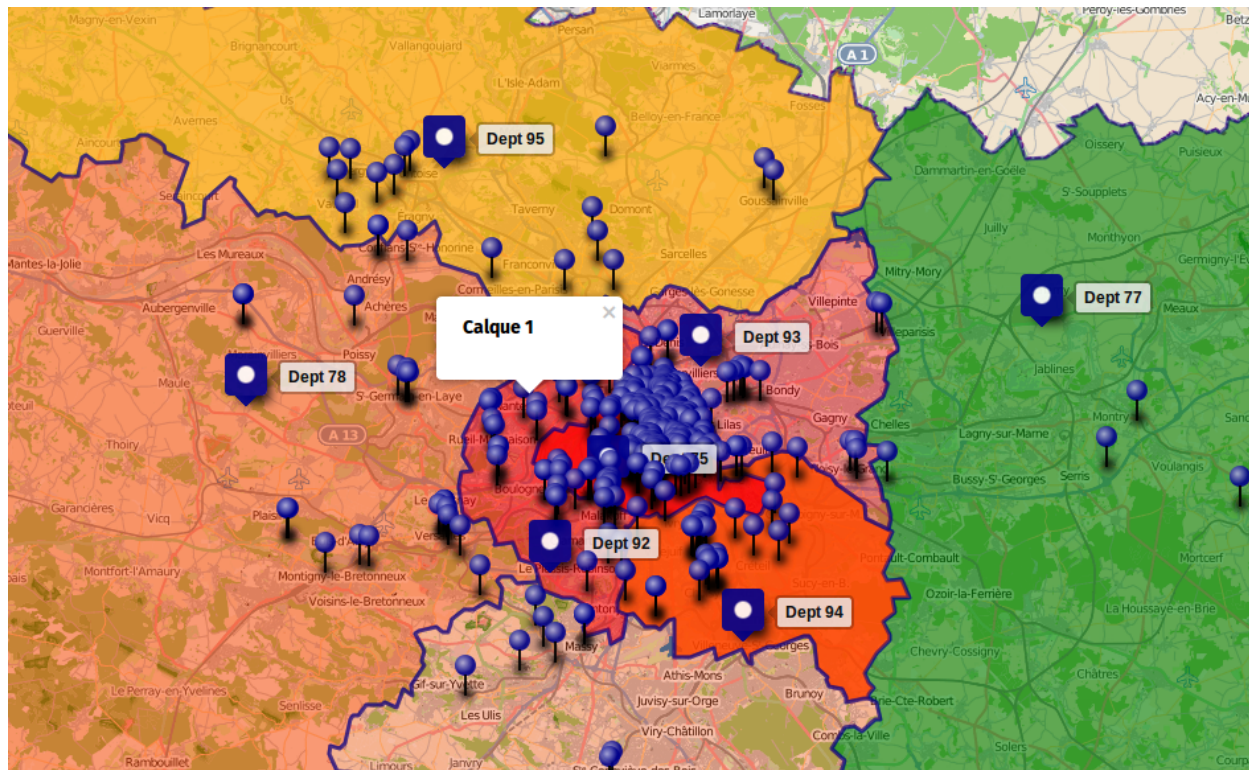
**2 ème conclusion :**

Le nombre d'accidents augmente plus au centre de paris.



**3 ème conclusion :**

Plus il y a de boites de nuits, plus le nombre d'accidents augmente dans la région.





# Annexes

Table nombre d'accidents:

	75	77	78	91	92	93	94	95
1	1292	18	113	289	0	1037	32	1
2	795	14	95	0	376	0	259	9
3	825	17	50	100	0	0	396	169
4	1180	12	8	1	720	0	190	21
5	1360	49	55	26	0	886	0	23
6	1276	42	75	3	0	409	0	232
7	1447	56	28	12	335	540	0	0
8	3036	205	113	2	0	675	0	11
9	1500	27	27	9	160	0	0	10
10	2267	26	22	0	0	584	0	66
11	2711	39	124	22	0	0	296	6
12	3913	15	0	79	1149	0	0	25
13	3106	41	18	0	0	206	0	87
14	2477	174	12	338	108	276	0	6
15	2876	52	39	57	0	37	113	0
16	4747	37	3	90	0	0	239	5
17	3512	45	254	11	0	0	998	0
18	2989	100	74	4	0	0	463	1272
19	2900	19	0	9	256	0	188	242
20	3064	105	77	0	362	0	0	2

Table indice de gravité:

	75	77	78	91	92	93	94	95
1	1.711772446	29.70888889	6.579292035	4.101868512	NA	9.179450338	5.8234375	10.8
2	1.557433962	17.38642857	7.614842105	NA	5.102739362	NA	4.029459459	28.41333333
3	1.615575758	28.64941176	16.9238	7.4002	NA	NA	5.045681818	1.62739645
4	2.92859322	14.95583333	35.96125	0.43	3.492388889	NA	6.678	14.56190476
5	2.134132353	29.12693878	4.137454545	22.88615385	NA	8.805553047	NA	14.84652174
6	2.513197492	21.47	5.363333333	44.13333333	NA	6.352933985	NA	5.957327586
7	1.527187284	24.85357143	8.244285714	24.98166667	4.331343284	7.5585	NA	NA
8	2.007928195	11.18595122	4.684867257	0.645	NA	7.581125926	NA	25.25
9	1.847686667	35.84444444	17.99851852	5.134444444	3.96025	NA	NA	17.646
10	1.920185267	37.50615385	25.50409091	NA	NA	7.029828767	NA	10.10757576
11	1.79272962	20.12512821	7.339274194	34.70454545	NA	NA	3.662939189	42.405
12	2.485987733	46.35066667	NA	5.843417722	2.864795474	NA	NA	8.49
13	2.247330972	12.73097561	33.99277778	NA	NA	8.897475728	NA	12.04436782
14	1.638526443	11.14465517	16.72	4.186065089	6.128148148	8.9575	NA	25.95333333