

# Analyse de la situation de la France suite à l'épidémie de COVID-19

Clustering using K-means

Etudiants : Owen ALLEMANG & Antoine BURGLIN

Date : 12/05/2020

Le code Python est disponible ici [https://github.com/SkirOwen/ciri\\_machine](https://github.com/SkirOwen/ciri_machine)

Le but de cette étude est d'analyser l'évolution du COVID-19. Plus particulièrement d'observer la situation de la France face à d'autres pays du monde avant son confinement et après le confinement.

Pour ce faire, nous avons créé deux bases de données à partir des rapports journaliers issue du site de la WHO (World Health Organization), NYTimes et (CSSE) Johns Hopkins University.

Nous avons créé bases de données avant et après le confinement, le 2020 – 03 – 17 pour la France.

Chaque base de données comporte six attributs :

pays, cas, morts, growth factor, nouveaux cas, nouvelles morts

Ensuite nous avons utilisé un algorithme de clustering (using K-means), en usant de la librairie *Sklearn*. Il s'agit d'apprentissage non-supervisé.

Dans cette algorithme nous utilisons  $K = 3$ , ce qui correspond à trois clusters :

- Situation de Faible intensité : en violet - cluster 0,
- Situation de Moyenne intensité : en gris - cluster 1,
- Situation de Forte intensité : en bleu - cluster 2.

Avec notre algorithme de clustering et nos base de données, nous avons entraîné l'algorithme avec tous les pays du monde sauf la France. Une fois cet entraînement terminé, nous avons demandé à l'algorithme de nous fournir un cluster d'appartenance pour la France aux deux dates différentes selon la moyennes des cas et la moyenne de nombre de nouveaux cas :

Avant 2020-03-17 :

- Cases : 6633
- New Cases : 120.6

Après 2020-03-17 au 2020-05-22 :

- Cases : 350 506
- New Cases : 52 945

La figure suivante nous montre l'exécution du programme dans la console :

```
1 CENTROIDS LOCATIONS:
2 [[1.76232044e+02 3.19613439e+00]
3  [8.10330000e+04 1.47332727e+03]
4  [1.52872500e+04 2.77950000e+02]]
5 -----
6 PREDICTION FOR FRANCE before 2020-03-17 :
7 Cases: 6633
8 New Cases: 120.6
9 Cluster : #9b59b6
10 -----
11 CENTROIDS LOCATIONS:
12 [[6.88599429e+03 1.04333247e+02]
13  [1.58030900e+06 2.39440758e+04]
14  [1.96625600e+05 2.97917576e+03]]
15 -----
16 PREDICTION FOR FRANCE after 2020-03-17 :
17 Cases: 172673
18 New Cases: 2616.257575757576
19 Cluster : #95a5a6
20 -----
```

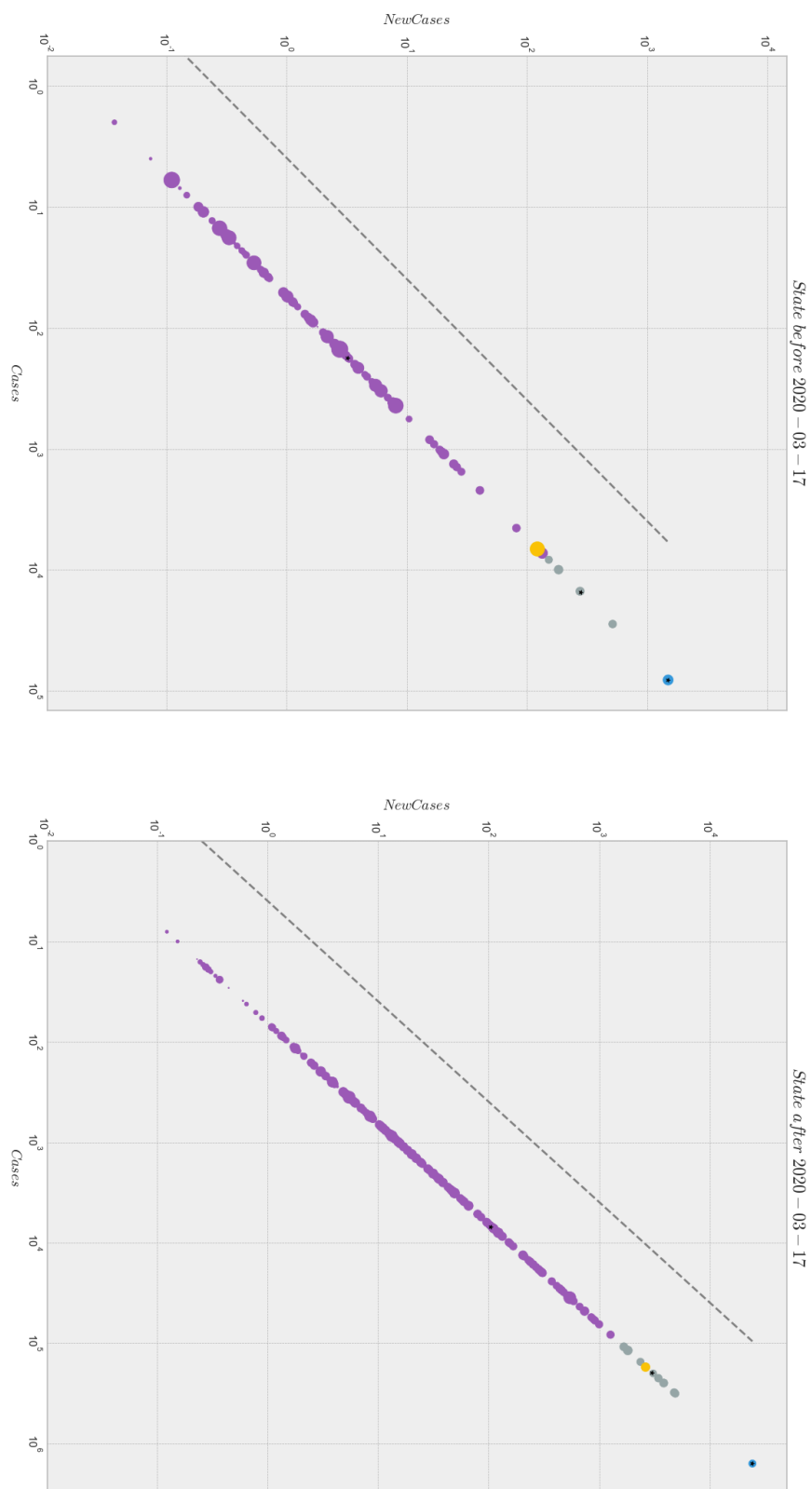


FIGURE 1 – Graph des pays avant et après 2020-03-17, taille correspond au growth factor, log scale

### Interprétation des résultats :

On observe que la France se place, avant le confinement, dans les pays faiblement touchés (en violet). Après la date du confinement le nombre de cas a en effet augmenté, cela n'est pas inattendu puisque certaines personnes ne respectaient le confinement, mais aussi à cause du temps d'incubation et de la "latence" de l'infection. Elle fait partie des pays les plus touchés dans le monde avec l'Italie, l'Espagne et le Royaume-Unis (le nom des pays peut être affiché en changeant le paramètre `label_countries` à `True`).

Au 2020-05-22, on observe que les Etats-Unis se place en secteur bleu (haut) alors qu'au 2020-03-17 celui-ci se plaçait en secteur violet.

C'est le pays le plus largement atteint par l'épidémie au vu du nombre de morts et du nombre de cas infectés extrêmement élevé par rapport aux autres pays.

Les US ont remplacés la Chine qui était le pays le plus touché au 2020-03-17. Suite aux mesures de confinements mis en place le nombre de cas et de nouveaux cas a diminué. Comme le montre notre graphe à l'heure actuelle la Chine est en secteur bleu.

	Cases	New Cases
Avant 2020-03-17		
US	4507	80.48
Chine	81033	1473.33
Après 2020-03-17		
US	1580309	23944.08
Chine	3030	45.90

Dans notre étude, nous utilisons les moyennes du nombre de nouveaux cas et du nombre de cas depuis le début de l'épidémie. Pour obtenir une information sur la "vitesse" de la pandémie, nous avons calculé pour chaque pays le facteur d'évolution.

$$\text{Growth Factor} = \frac{\Delta N_d}{\Delta N_{d-1}}$$

$\Delta N_d$  nombre de nouveau cas au jour  $d$ .

$\Delta N_{d-1}$  nombre de nouveau cas au jour  $d - 1$ .

Ce facteur d'évolution nous donne une indication de l'évolution de l'épidémie dans chaque pays. Si  $\text{Growth Factor} < 1$  le nombre de cas semble diminuer,  $\text{Growth Factor} > 1$  le nombre de cas semble augmenter,  $\text{Growth Factor} = 1$  il s'agit du point d'inflexion.

Par exemple pour la France au 2020-03-17, le growth factor est de 4.10 et au 2020-05-22 celui-ci était de 1,52. Cela montre que les mesures de confinement ont eu un impact sur l'évolution de l'épidémie mais cette valeur reste supérieure à 1. La tendance est à la baisse mais des efforts seront encore nécessaires pour ramener le facteur d'évolution en dessous de 1.

Voici un autre graph montrant l'efficacité de confinement, ici chaque pays à la date de confinement, si un pays n'a pas d'information alors celui-ci est ignoré :

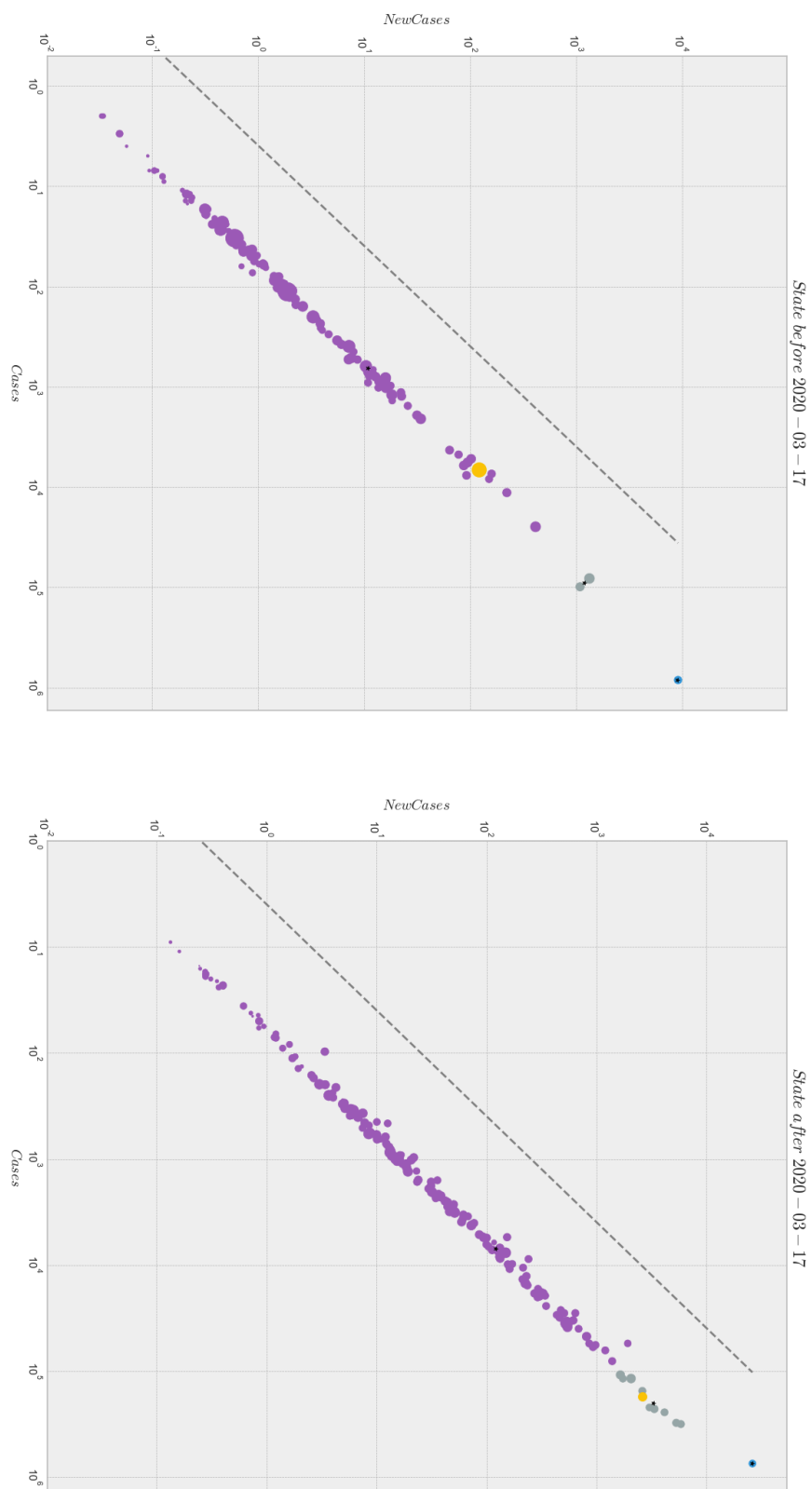


FIGURE 2 – Graph des pays avant et après leur lockdown date, taille correspond au growth factor, log scale