

Rapport du Projet Statistique sur la COVID-19

Membre du groupe:

Komlan DANTODJI

Oumar NIAN

Master 1 Big Data et Fouilles de données:
Université Paris 8

Plan:

1. Introduction
2. Problématique
3. Objectifs
4. Données
5. Technologies utilisées
6. Analyse des résultats
7. Conclusion

Lien github du projet:

https://github.com/jmkd0/Project_M1/blob/master/ProbaSta/projet_covid/analyse_covid.py

1) Introduction

Le virus identifié en janvier 2020 en Chine est un nouveau coronavirus, nommé SARS-CoV-2. La maladie provoquée par ce coronavirus a été nommée COVID-19 par l'Organisation mondiale de la Santé - OMS. Depuis le 11 mars 2020, l'OMS qualifie la situation mondiale du COVID-19 de pandémie ; c'est-à-dire que l'épidémie est désormais mondiale.

Les coronavirus sont une famille de virus qui provoquent des maladies allant d'un simple rhume (certains virus saisonniers sont des coronavirus) à des pathologies plus sévères (comme les détresses respiratoires du MERS, du SRAS ou de la COVID-19).

En France, les premiers cas sont survenus en février 2020. Ainsi nous allons nous intéresser à la progression de la covid-19 depuis son arrivée jusqu'à aujourd'hui.

2) Problématique

Dans le but de limiter la propagation de la covid-19 on a besoin de suivre son évolution en faisant des analyses statistiques.

3) Objectifs

Sur ce projet ,il consiste à faire des analyses statistiques avec les données qui contiennent toutes informations relatives à la covid-19 en france.

Pour cela on a besoin de mettre en place des graphes de suivi selon le sexe, la classe d'âge ,la période, le nombre de personne hospitalisées , le nombre de personne en réanimation, le nombre de personnes décédées en fonction de l'âge etc...

4) Données

Les données utilisées dans notre données viennent de la plateforme officielle du gouvernement .

Nous avons deux catégories de données:

- Données relatives aux résultats des tests virologiques COVID-19

https://www.data.gouv.fr/en/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/#_

Preview									×
reg	↕ jour	↕ P_f	↕ P_h	↕ P	↕ T_f	↕ T_h	↕ T	↕ cl_age90	↕
1	2020-05-13	0	0	0	0	0	0	9	
1	2020-05-13	0	0	0	1	0	1	19	
1	2020-05-13	0	0	0	1	0	1	29	
1	2020-05-13	0	0	0	5	1	6	39	
1	2020-05-13	0	0	0	2	1	3	49	
1	2020-05-13	0	0	0	5	2	7	59	
1	2020-05-13	0	0	0	1	1	2	69	
1	2020-05-13	0	0	0	1	4	5	79	
1	2020-05-13	0	0	0	1	0	1	89	
1	2020-05-13	0	0	0	0	0	0	90	
1	2020-05-13	0	0	0	17	9	26	0	
2	2020-05-13	0	0	0	3	1	4	9	
2	2020-05-13	0	0	0	3	4	7	19	
2	2020-05-13	0	0	0	7	4	12	29	
2	2020-05-13	0	0	0	5	4	9	39	

- Données hospitalières relatives à l'épidémie de COVID-19

https://www.data.gouv.fr/en/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/#_

Preview							×
dep	↕ sexe	↕ jour	↕ hosp	↕ rea	↕ rad	↕ dc	↕
01	0	2020-03-18	2	0	1	0	
01	1	2020-03-18	1	0	1	0	
01	2	2020-03-18	1	0	0	0	
02	0	2020-03-18	41	10	18	11	
02	1	2020-03-18	19	4	11	6	
02	2	2020-03-18	22	6	7	5	
03	0	2020-03-18	4	0	1	0	
03	1	2020-03-18	1	0	0	0	
03	2	2020-03-18	3	0	1	0	
04	0	2020-03-18	3	1	2	0	
04	1	2020-03-18	3	1	0	0	
04	2	2020-03-18	0	0	2	0	
05	0	2020-03-18	8	1	9	0	
05	1	2020-03-18	1	0	7	0	
05	2	2020-03-18	7	1	2	0	

Au début, on avait des données brutes dont on a besoin de faire quelques traitements au préalable.

5) Technologies utilisées

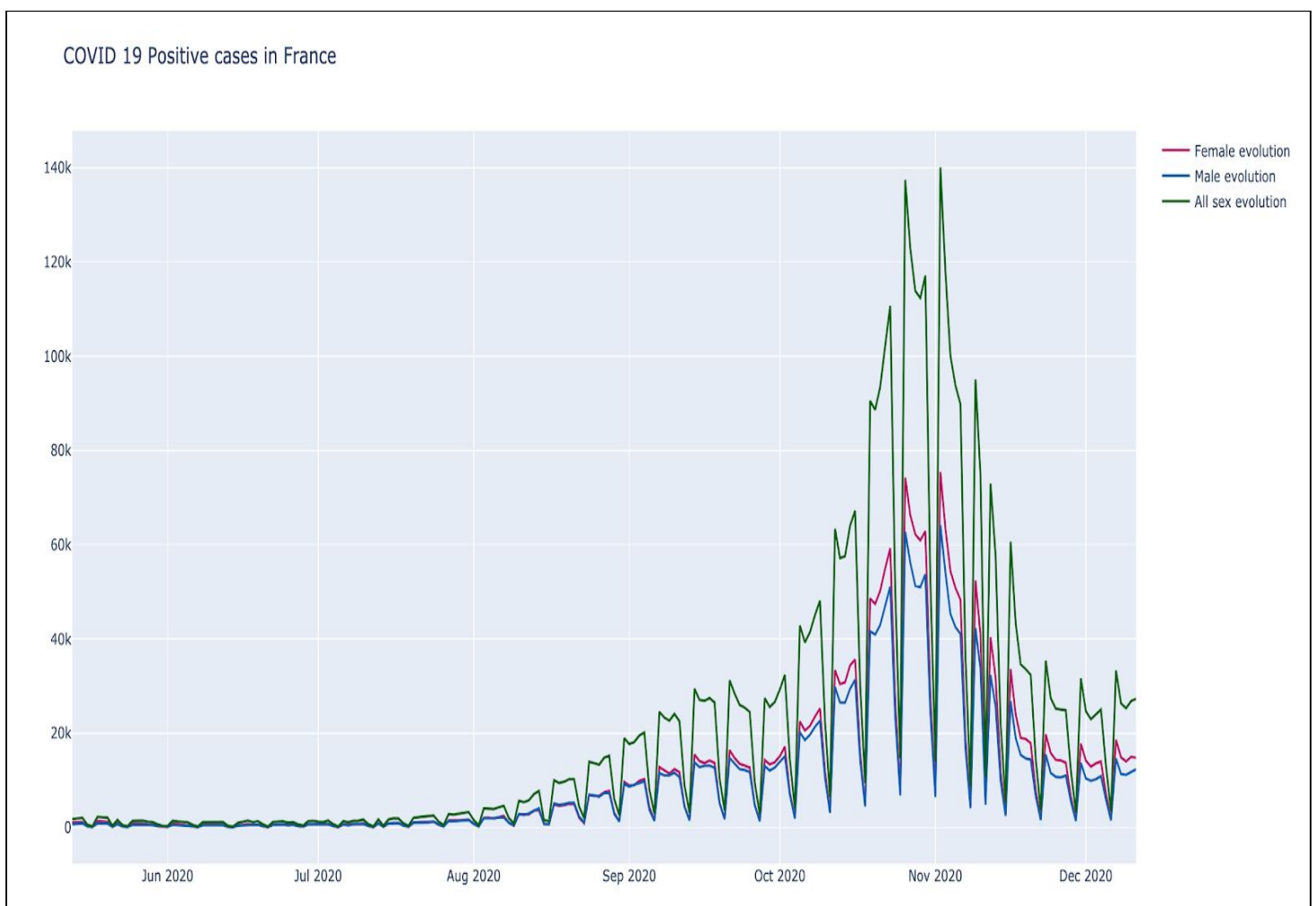
Concernant les technologies on a utilisé le langage de programmation **python** avec une librairie **Pandas** ,**Plotly**.

- **Plotly** est une librairie python qui nous permet de tracer des graphes.
- **Pandas** est une librairie de manipulation de données avec python.

6) Analyse des résultats

A) Analyse des données relatives aux résultats des tests virologiques COVID-19

a) Evolution des cas positifs en fonction de la période:



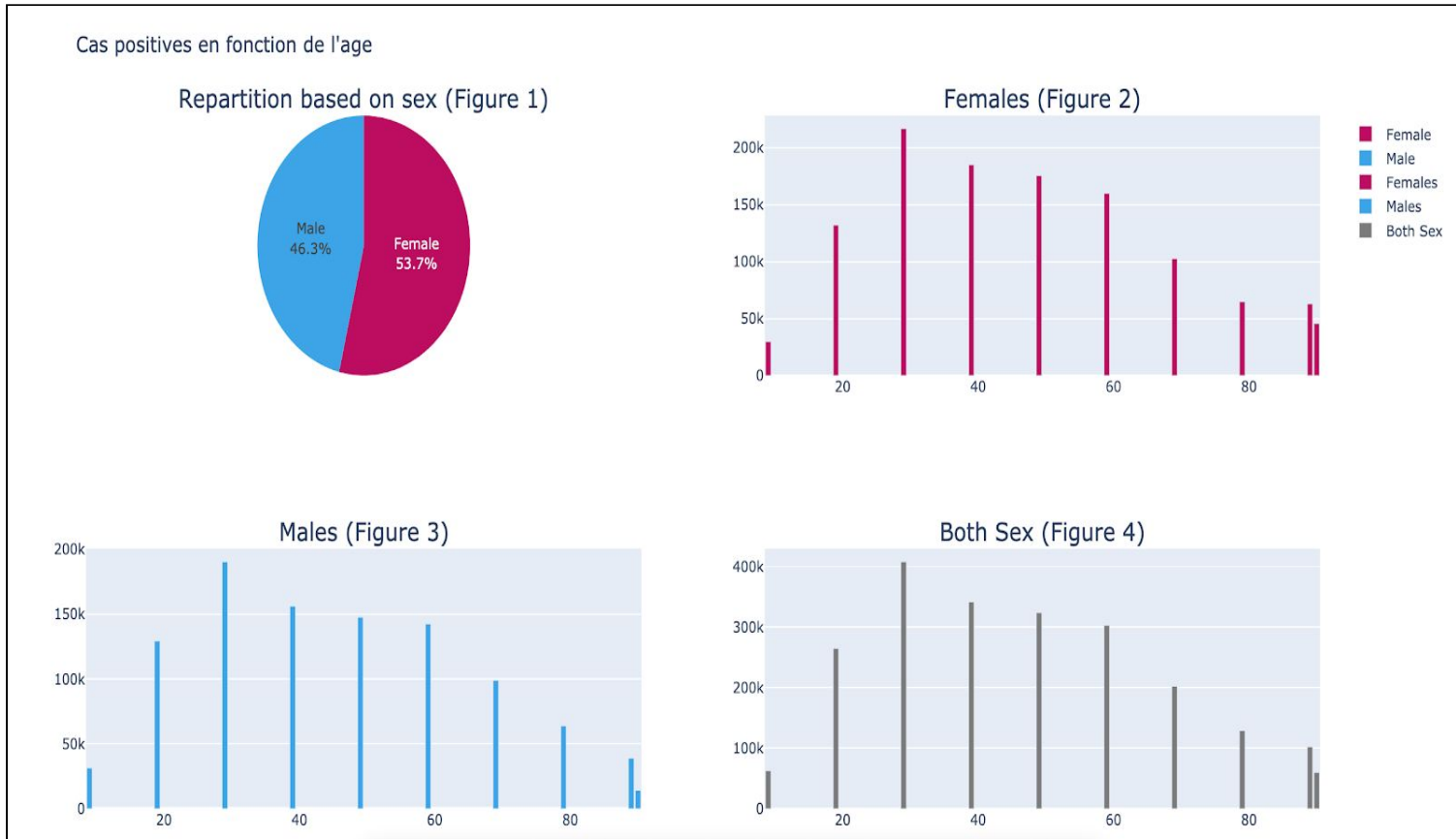
Ce graphes ci-dessus montre l'évolution des cas contaminés en fonction de la période et du sexe.

- Le diagramme en rouge représente l'évolution des cas positifs chez les femmes.
- Le diagramme en bleu représente l'évolution des cas positifs chez les hommes.
- Le diagramme en vert représente l'évolution des cas positifs de manière générale (homme ou femmes).

On constate qu'à partir du mois d'août une augmentation des cas positifs en France qui est la deuxième vague a atteint le pic en novembre.

On constate aussi une baisse des cas positifs à partir du mois de décembre.

b) Evolution des cas positifs en fonction du sexe et de la classe d'age:

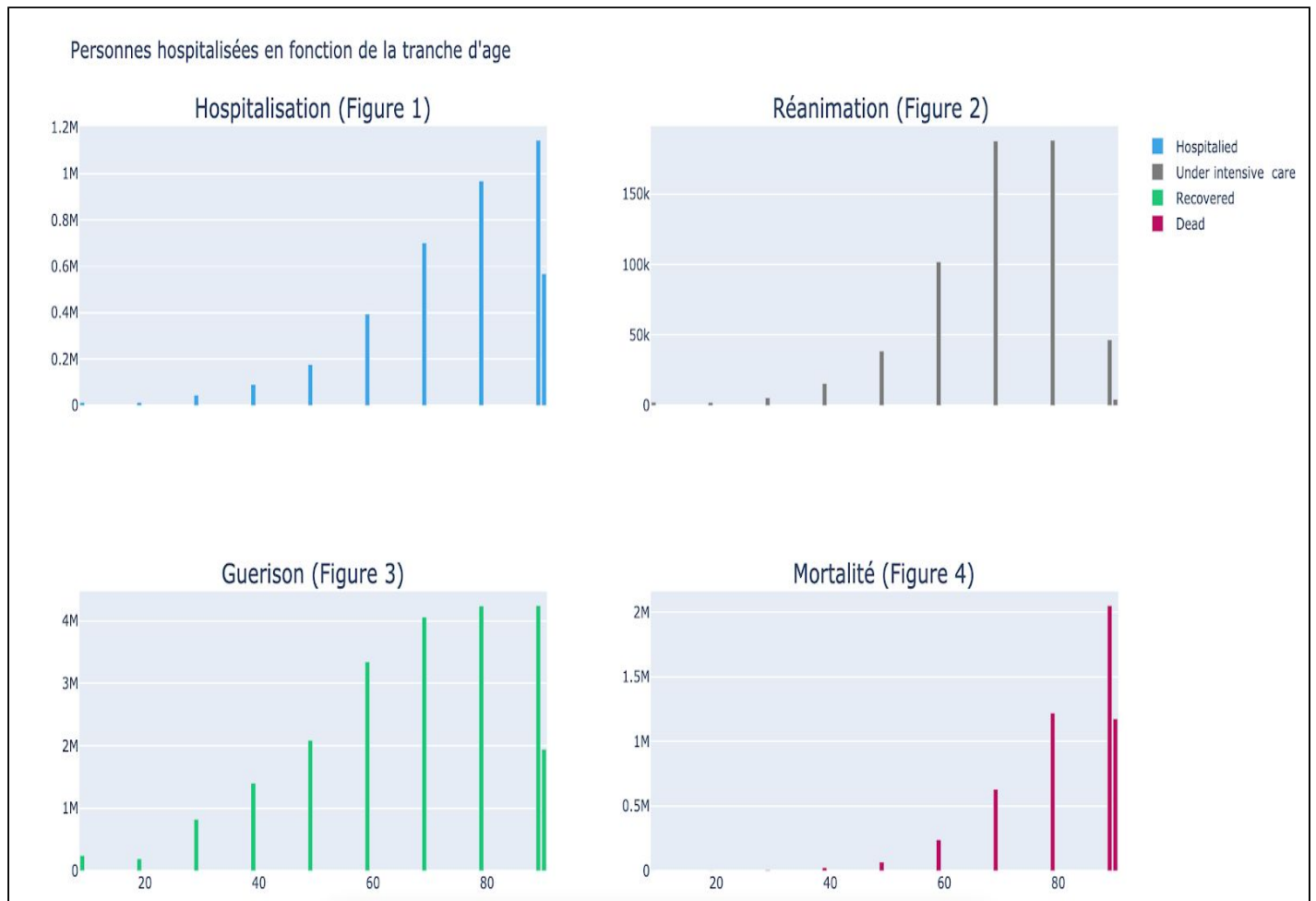


Ce graphe ci-dessus montre l'évolution des cas contaminés en fonction de la classe d'age et du sexe.

- Pour la figure 1 on remarque 53,7% des cas contaminés sont des femmes. 46,3% des cas sont des hommes. Les femmes sont donc plus exposées que les hommes en France.
- Pour la figure 2 on constate que chez les femmes, les jeunes sont plus exposés au virus.
- De même pour la figure 3 chez les hommes, les jeunes sont plus exposés au virus.

En somme, que ce soit chez les hommes ou chez les femmes, les jeunes sont plus exposés au virus.

B) Analyse des données hospitalières relatives à l'épidémie de COVID-19 :

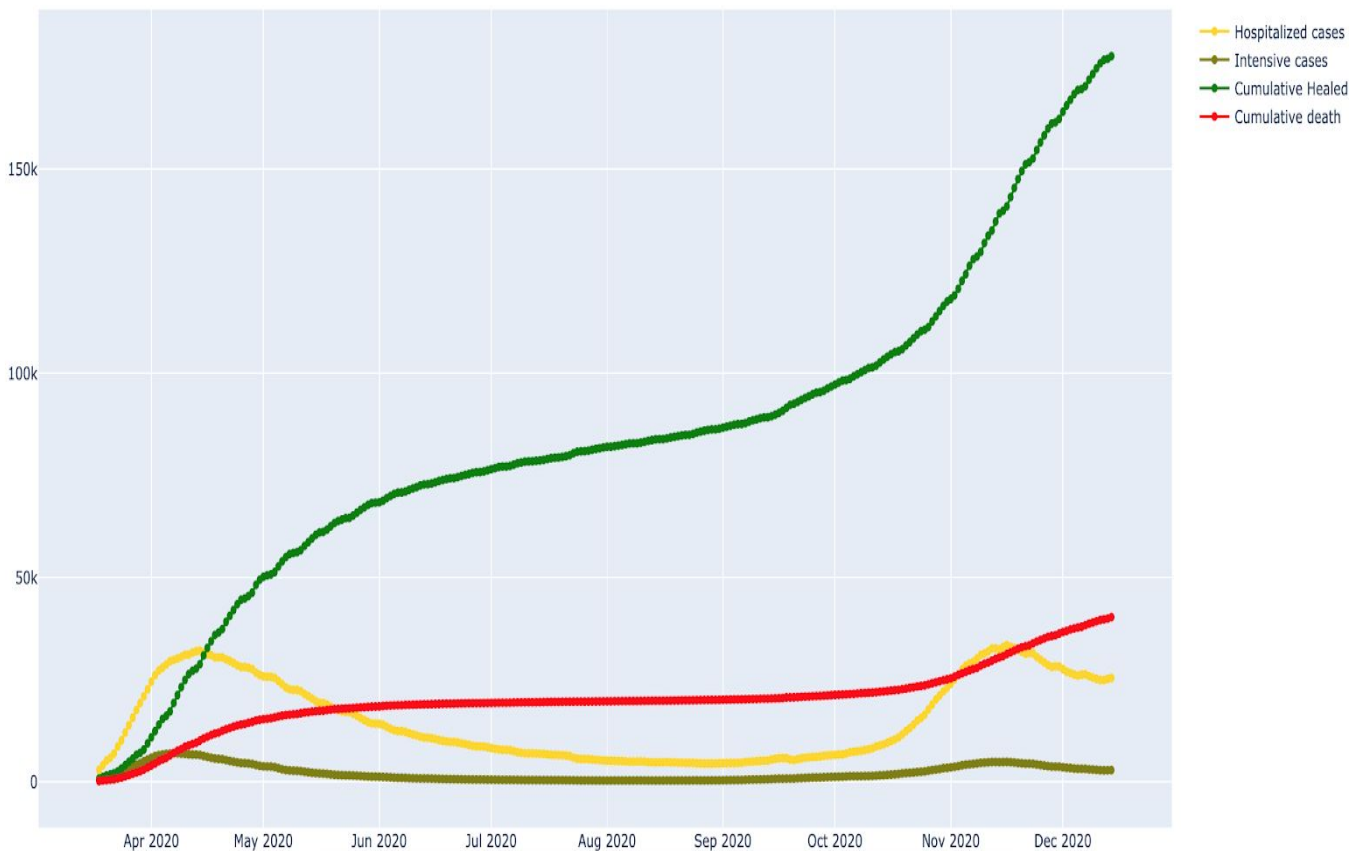


Ce graphe ci-dessus montre l'évolution du nombre de personnes hospitalisées, guéries, décédées et en soins intensifs en fonction de la tranche d'âge.

- Pour la figure 1 on remarque qu'il y a plus d'hospitalisation chez les personnes âgées.
- La figure 2 montre qu'il y a plus de personnes âgées en soins intensifs.
- Pour la figure 3 on constate que le nombre de personnes guéries chez les personnes âgées est plus important car les jeunes sont moins hospitalisés.
- De même pour la figure 4 on constate qu'il y a plus de décès chez les personnes âgées.

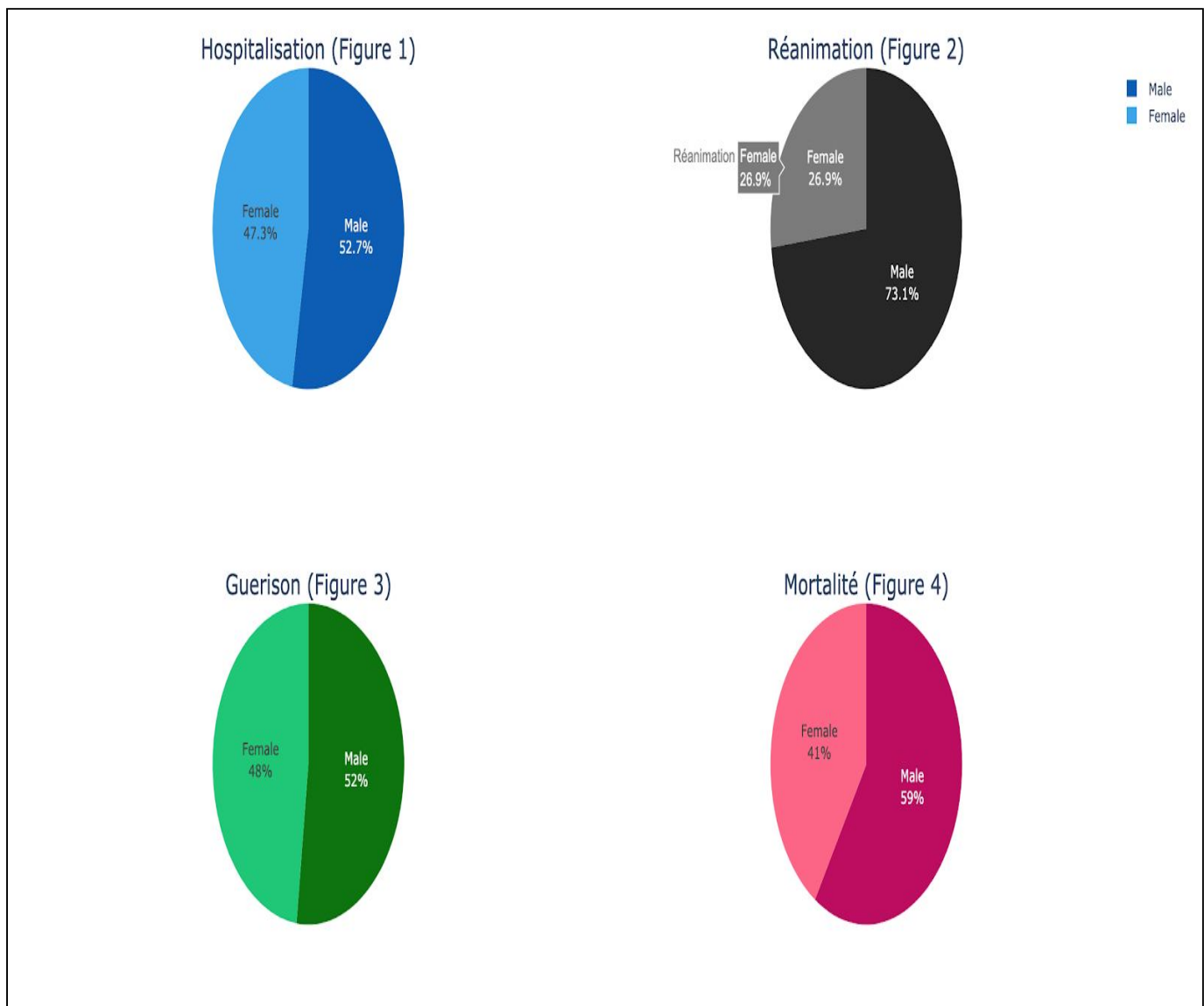
En somme, on constate que les personnes âgées sont des personnes à risque face à ce virus.

COVID 19 Hospital information in France



Ce graphe ci-dessus montre l'évolution de la covid-19 dans les hôpitaux.

- La courbe jaune montre le cas des personnes hospitalisées en France quotidiennement. On a deux pics en avril et novembre correspondant aux deux vagues.
- La courbe en vert sombre montre le cas passés en soins intensifs (réanimations) en France quotidiennement. Elle suit l'évolution de la courbe jaune du nombre de personnes hospitalisées, car plus il y a de personnes hospitalisées, plus le nombre de personnes en réanimation augmente.
- La courbe verte décrit l'évolution cumulative du nombre de personnes guéries après être hospitalisées.
- La courbe décrit l'évolution cumulative du nombre de personnes décédées après hospitalisation.



Ce graphe ci-dessus montre l'évolution de la covid-19 dans les hôpitaux en fonction du sexe.

- La figure 1 montre la repartition suivant les sexes des cas hospitalisés
On remarque que les hommes sont plus hospitalisés à 52.7% contre 47.3% chez les femmes.
- La figure 2 montre qu'il y a plus de personnes en réanimation chez les hommes à 73.1% contre 26.9% chez les femmes.
- La figure 3 montre qu'il ya beaucoup plus de guéris chez les hommes du à un taux élevé d'hospitalisation .
- La figure 4 montre qu'il ya beaucoup plus de décès chez les hommes dus à un taux élevé d'hospitalisation et de réanimation.

CONCLUSION

La France enregistre aujourd'hui deux vagues dont la première en avril et la deuxième en Novembre.

Les femmes sont plus contaminées que les hommes. Par ailleurs, les personnes âgées manifestent plus la maladie que les jeunes qui ont généralement des cas moins sévères.

Bien que les femmes soient plus contaminées par la covid-19, on constate une hospitalisation, réanimation et mortalité plus conséquente chez les hommes que chez les femmes.

Les personnes âgées sont les personnes à risques. Les jeunes sont beaucoup plus résistants.