

NOTEBOOK PYTHON ICI

Introduction

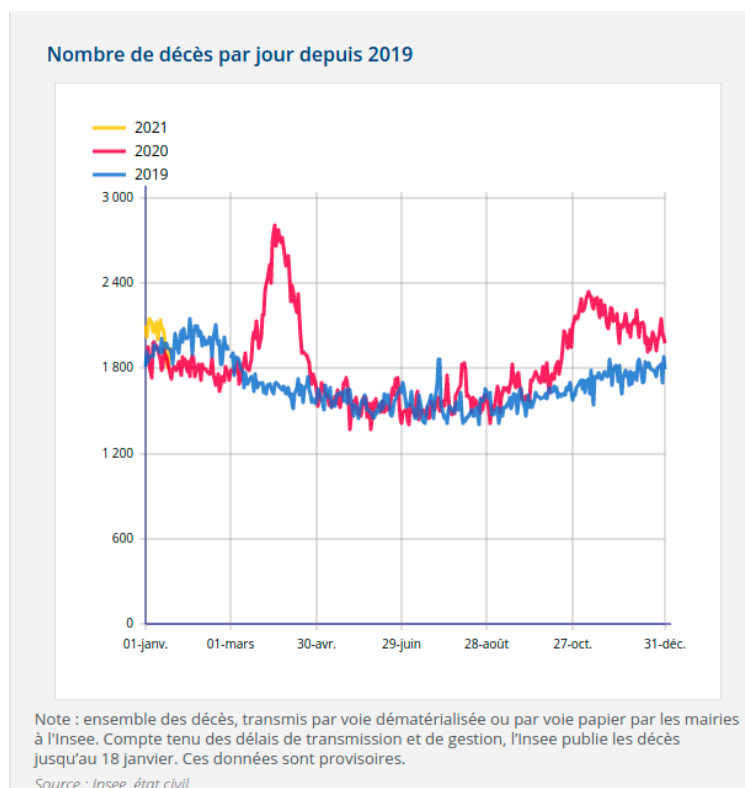
Sur le plan sanitaire, l'année 2020 n'a pas été facile pour la France.

Il y a eu de nombreux décès et maladies graves dus à la propagation du SARS-COV-2. Le système de soin a connu une situation alarmante, avec les hôpitaux publics exceptionnellement chargés d'un côté, et la médecine de ville qui n'a pas pu jouer son rôle. S'ajoute à ceci l'inquiétude générale dans la population, qui a vécu au rythme des mesures édictées par le gouvernement sous un état d'urgence, et vu sa santé mentale se dégrader, notamment chez les jeunes.

N'oublions pas enfin l'arrêt d'activité économique dans pleins de secteurs affectés par les mesures gouvernementales, avec les problèmes de précarité - et donc de santé - que cela génère.

L'un des indicateurs principaux d'un problème de santé publique est la surmortalité, c'est-à-dire le nombre de décès constatés dans l'année par rapport au nombre attendu vues les années précédentes.

Le graphique publié dans [cette page par l'INSEE](#) montre de toute évidence une surmortalité en 2020 par rapport à 2019, avec un premier pic le 1er avril et un 2ème pic le 7 novembre



Objectif

Dans cet article, je m'inspire d'une vidéo publiée par la chaîne YouTube "Décoder l'éco" pour analyser les décès 2020 par rapport aux années précédentes à partir du [fichier national des personnes décédées](#).

L'idée centrale de la vidéo est que la comparaison entre les nombres absolus de décès d'une année à l'autre ne suffit pas :

- Il faut d'abord prendre en compte la croissance (ou la décroissance) de la population entre les deux années : plus une population est nombreuse, plus le nombre de décès annuel est grand
- Il faut surtout prendre en compte la différence de la pyramide d'âge entre les deux années : à population constante, une pyramide d'âge plutôt âgée connaîtra plus de décès annuels qu'une pyramide d'âge plutôt jeune, il y a également une sensibilité selon le sexe.

Ceci m'amène à comparer un indicateur plus précis : le taux de mortalité par âge.

Il s'agit plus précisément, par âge A et année N, du pourcentage, parmi les personnes d'âge révolu A au début de l'année N, des personnes décédées au courant de l'année N.

Le but de cet article est de comparer cet indicateur entre l'année 2020, qui a connu une situation sanitaire exceptionnelle, et les années 2017, 2018 et 2019.

Lecture des données

Le fichier des personnes décédées est consolidé par l'INSEE et publié sur le portail data.gouv.fr

On y trouve un fichier par année depuis 1970, et un fichier par mois et par trimestre depuis 2019.

L'information de chaque ligne du fichier est structurée ainsi :



A noter qu'un décès survenu en année N n'est pas forcément déclaré dans le fichier de l'année N. Il peut, à cause de retard des déclarations papier ou par voie électronique, être déclaré dans le fichier de l'année N+1, voir ultérieurement.

Voici quelques statistiques qui permettent d'appréhender ces données :

Fichier de décès	# Décès (en milliers)	% survenus en France métro.	# Lignes inexploitable	Part des décès métropolitains de l'année du fichier déclarés dans les fichiers des années ultérieures	Part des décès métropolitains hors décembre de l'année du fichier déclarés dans les fichiers des années ultérieures
Fichier 2020	679.9	99.1%	3	NA	NA
Fichier 2019	625.3	98.9%	2	98.08%	99.82%
Fichier 2018	620.1	98.9%	5	97.74%	99.72%
Fichier 2017	612.9	98.7%	5	97.63%	99.59%
Fichier 2016	603.3	98.7%	4	98.11%	99.74%

On voit donc que ça reste représentatif de se limiter aux décès survenus en France métropolitaine, ce qui a pour avantage de comparer les années au regard du même système de soins.

On voit également que moins de 0.4% des décès survenus hors décembre d'une année sont absents du fichier de cette année, ce qui me pousse à raisonner en année glissante du 1er décembre au 30 Novembre, plutôt qu'en année civile.

Vérifions si le raisonnement par année glissante n'altère pas substantiellement la perception du graphique publié par l'INSEE :

	# décès année glissante calculé dans cet article (en milliers)	# décès du graphique ci-dessus publié par l'INSEE (en milliers)	Ecart absolu
2020	665.3	670.2	0.74%
2019	619.6	615.5	0.66%

Je re-définis donc mon indicateur comme ceci :

Par âge A et année N, le pourcentage, parmi les personnes d'âge révolu A au début de l'année N, des personnes décédées entre le 1er décembre de l'année N-1 et le 30 novembre de l'année N+1.

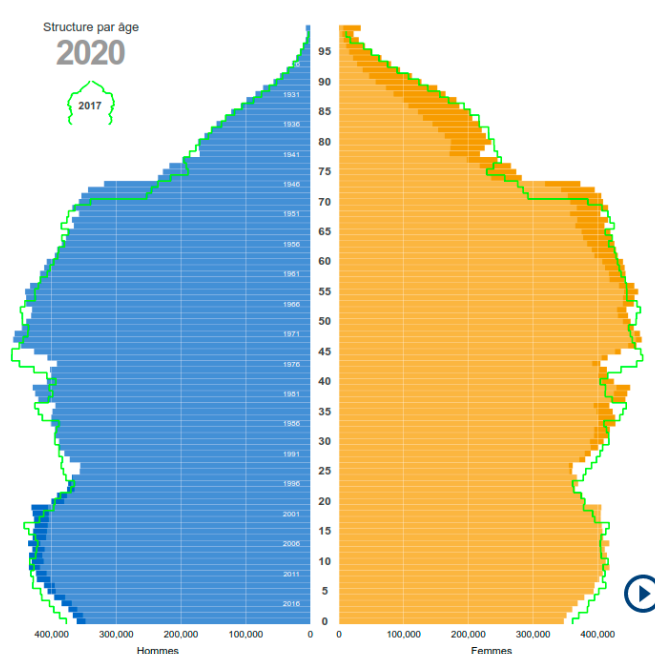
Enfin, pour éviter le biais des années bissextiles, je re-définis l'année 2020 comme la série de 365 jours qui se termine le 30 novembre, et ainsi de suite.

Ceci me donne la base suivante en nombre de décès :

Année glissante	# Décès Femmes en France métropolitaine (en milliers)	# Décès Hommes en France métropolitaine (en milliers)
2020	329.2	330.6
2019	308.1	305.3
2018	309.1	306.8
2017	306.5	302.0

Les pyramides d'âge pour ces quatre années en France métropolitaine sont également [publiées par l'INSEE sur cette page](#)

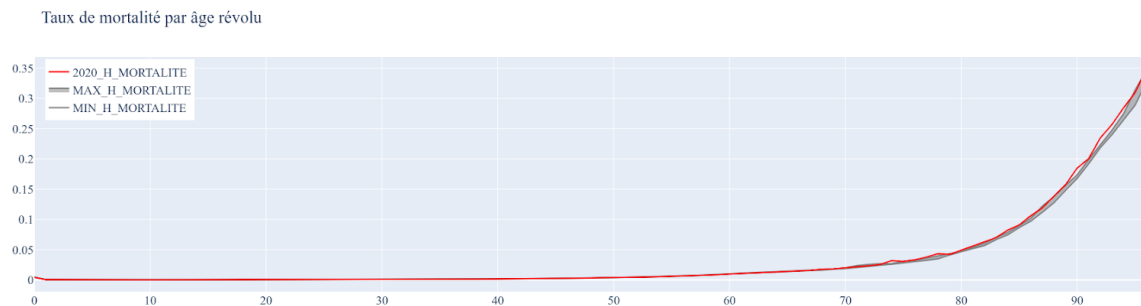
L'INSEE publie également une [pyramide des âges interactive](#) qui permet notamment de voir l'effet "baby boom" : Il y a eu beaucoup plus de naissances en 1946 par rapport à 1945 (année de guerre). Ces baby boomers avaient moins de 71 ans en 2017, mais jusqu'à 75 ans en 2020. Or - comme on le verra plus tard, la mortalité augmente fortement à partir de cette tranche d'âge.



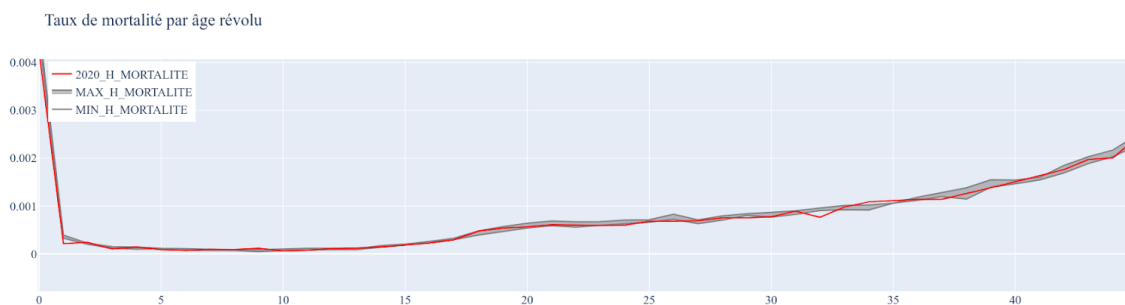
Calcul et comparaison qualitative

Le calcul du taux de mortalité par âge révolu est assez simple, puisqu'il consiste à diviser le nombre de personnes décédées pour chaque âge et chaque année glissante par le nombre de personnes du même âge recensées dans la pyramide d'âge de la même année.

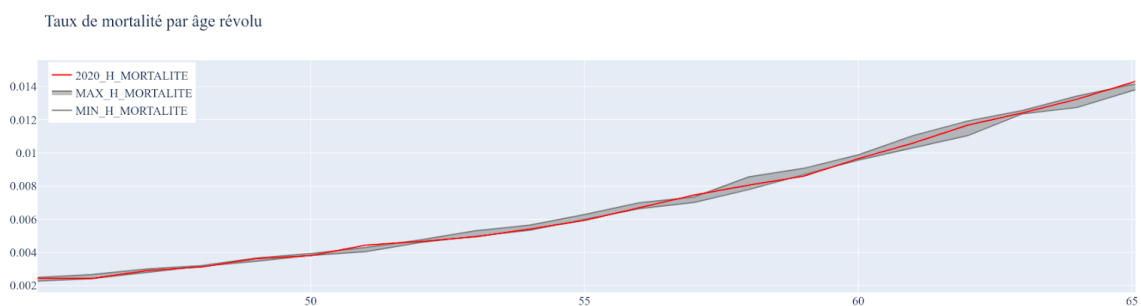
Afin d'avoir un aperçu graphique, j'affiche - pour les hommes - les taux de mortalité de 2020, ainsi que l'enveloppe minimum-maximum constatée sur les années 2017 à 2019 :



En dessous de 45 ans, la mortalité 2020 est plutôt inférieure à l'enveloppe minimum-maximum constatée sur les années 2017 à 2019 :

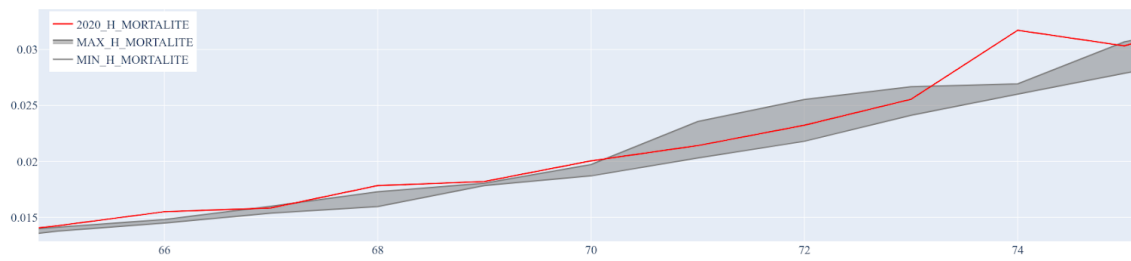


Entre 45 et 65 ans, la mortalité 2020 est plutôt dans l'enveloppe :



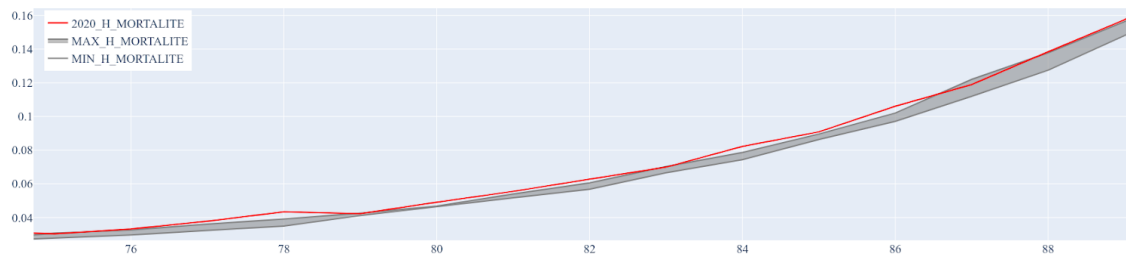
Entre 66 et 75 ans, la mortalité 2020 est dans l'enveloppe pour les âges 71,72,73,75 avec un dépassement significatif pour l'âge 74 et un léger dépassement de l'enveloppe pour les autres âges :

Taux de mortalité par âge révolu



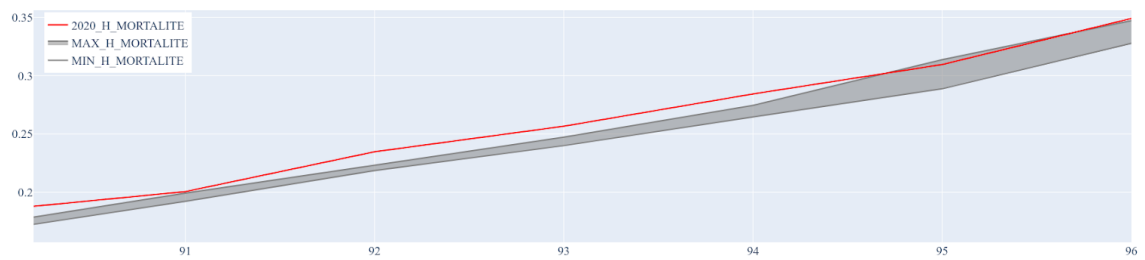
Même constat pour les âges 76 à 89 ans, le seul dépassement significatif est à 78 ans :

Taux de mortalité par âge révolu



Pour les âges 90 et plus, le dépassement de l'enveloppe est clair pour tous les âges :

Taux de mortalité par âge révolu



Comparaison quantitative

Une métrique de comparaison serait la différence, par tranche d'âge, entre le nombre réel de décès 2020 et celui qui aurait eu lieu si le taux de mortalité était semblable aux années précédentes.

Je résume le résultat de la comparaison dans le tableau suivant :

SEXE	AGE RÉVOLU	DÉCÈS 2020 (en milliers)	ENVELOPPE MIN-MAX DEPUIS 2017 (en milliers)	ECART MOYENNE DEPUIS 2017 (en milliers)	REMARQUE
Femmes	0 à 44	6.5	6.3 à 7.2	-0.2	En bas d'enveloppe
Femmes	45 à 64	27.8	27.2 à 28.4	0	Dans l'enveloppe
Femmes	65 à 74	36.6	34.3 à 36.9	1.2	En haut d'enveloppe
Femmes	75 à 89	130.7	123.0 à 130.9	3.7	En haut d'enveloppe
Femmes	90 à 96	97.2	92.1 à 98.5	1.9	En haut d'enveloppe
Hommes	0 à 44	12.7	12.5 à 13.7	-0.3	En bas d'enveloppe
Hommes	45 à 64	52.9	51.6 à 54.4	-0.1	Dans l'enveloppe
Hommes	65 à 74	66.8	62.1 à 66.9	2.7	En haut d'enveloppe
Hommes	75 à 89	139.3	128.0 à 136.3	6.9	Dépasse l'enveloppe
Hommes	90 à 96	50.5	47.0 à 48.9	2.4	Dépasse l'enveloppe

Synthèse

A travers une analyse fine par tranche d'âge de la mortalité 2020 en France métropolitaine je trouve que, par rapport aux années 2017-2018-2019 :

- Chez les hommes de 75 ans ou plus, il y a 9300 décès au dessus de la moyenne, soit +5%
- La mortalité entre 65 et 74 ans est aussi au-dessus de la moyenne (+3900 soit +3.8%) mais reste dans l'enveloppe min-max, il en est de même pour les femmes de 75 ans et plus (+5600 soit +2.5%)
- Pour les 64 ans et moins, la mortalité est 0.8% en dessous de la moyenne

Je pense qu'il est très difficile d'établir un lien clair entre la surmortalité observée dans cette analyse et le nombre de décès imputés à la COVID (~70000 depuis mars 2020 dont 93% de plus de 65 ans d'après Santé Publique France). Le sujet est complexe et plusieurs facteurs difficiles à mesurer doivent être pris en compte.

[ADDENDUM]

L'analyse avec plus d'années d'historique donne les résultats suivants :

SEXE	AGE RÉVOLU	DÉCÈS 2020 (en milliers)	ENVELOPPE CENTILE 25 à 75 DEPUIS 2012 (en milliers)	ECART MÉDIANE DEPUIS 2012 (en milliers)	ECART MOYENNE DEPUIS 2012 (en milliers)	ECART CENTILE 75 DEPUIS 2012 (en milliers)
Femmes	0 à 44	6.5	6.5 à 7.2	-0.3	-0.3	-0.6
Femmes	45 à 64	27.8	27.3 à 28.5	-0.1	-0.1	-0.7
Femmes	65 à 74	36.6	34.3 à 35.8	1.7	1.3	0.8
Femmes	75 à 89	130.7	127.7 à 134.9	0.5	-0.4	-4.1
Femmes	90 à 96	97.2	95.3 à 102.4	-0.9	-4.3	-5.1
Hommes	0 à 44	12.7	12.8 à 14.0	-0.5	-0.6	-1.2
Hommes	45 à 64	52.9	53.6 à 57.0	-2.5	-2.5	-4
Hommes	65 à 74	66.8	62.9 à 66.3	2.4	1.8	0.4
Hommes	75 à 89	139.3	133.5 à 141.9	2.7	1.8	-2.5
Hommes	90 à 96	50.5	47.7 à 50.7	1.6	0.1	-0.2

Je referais l'analyse en avril 2021 sur la base d'années glissantes de mars à mars puisque les décès imputés à la covid commencent en mars 2020.