

IUT Grand Ouest Normandie

Bachelor Universitaire de Technologie
Science des Données
Campus de Lisieux

Science des Données 1 - SAE 2.06
Analyse de données, Reporting et Dataviz

Thématique

Analyse des décès en France entre 2019 et 2022



Auteurs

BOUCHER Nathan
DIOP Mandir
GAMONDELE Maxime

Table des matières

Introduction	3
1 Les décès en France	4
1.1 Répartition des décès selon le sexe en fonction de l'année	4
1.2 Répartition des décès en 2022 selon la région	4
1.3 Répartition des décès en 2022 selon le sexe par région	5
1.4 Répartition des décès selon l'âge des individus sur la période 2019 à 2022	5
1.5 Répartition des décès selon l'âge des individus par région	6
2 Les décès en Normandie	7
2.1 Répartition des décès en Normandie selon le département par année	7
2.2 Répartition des décès en Normandie selon le sexe par département par année	7
2.3 Répartition des décès en Normandie selon l'âge des individus en fonction du sexe	8
2.4 Répartition des décès en Normandie selon l'âge des individus en fonction du sexe par département	8
Conclusion	9
3 Annexe	10
Table des figures	11

Introduction

Le gouvernement français met à disposition des concitoyens sur le site data.gouv.fr, des fichiers de données regroupant l'ensemble des décès survenus sur le territoire depuis 1970.

[Site du gouvernement](#)

Ces fichiers, disponibles sous la dénomination `deces-YYYY.txt` où `YYYY` désigne l'année, sont des fichiers à largeur fixe, c'est-à-dire que chacune des variables est définie sur un nombre de colonnes bien spécifique (2 à 10 par exemple). Nous avons ci-dessous (cf. Fig. 1) une capture d'écran nous informant des variables contenu dans les fichiers.

Contenu du fichier
Chaque enregistrement est relatif à une personne décédée et comporte les zones suivantes :

- le nom de famille
- les prénoms
- le sexe
- la date de naissance
- le code du lieu de naissance
- la localité de naissance en clair (pour les personnes nées en France ou dans les DOM/TOM/COM)
- le libellé de pays de naissance en clair (pour les personnes nées à l'étranger)
- la date du décès
- le code du lieu de décès
- le numéro d'acte de décès

Fig. 1. Variables du jeu de données

Ainsi, on se propose de réaliser une analyse exploratoire de ces données portant sur les décès survenus de 2019 à 2022. Ce fichier de données contient 2 621 178 observations. Notre analyse sera scindée en deux parties distinctes, une première se focalisant sur les décès à l'échelle de la France puis une seconde partie sur les décès à l'échelle de la Normandie.

1 Les décès en France

Cette première partie abordera l'analyse des données concernant les décès en France. Nous nous pencherons sur l'étude des sexes, des nouvelles régions françaises, des années et des âges de décès.

1.1 Répartition des décès selon le sexe en fonction de l'année

Sur les années de 2019 à 2022, nous pouvons observer (cf. Fig. 2) que pour chacune des années, nous trouvons une répartition des décès par sexe d'environ 50 %. Nous supposons qu'il y a donc une faible association entre les variables sexe et année de décès. Suite au calcul du V de Cramer (qui nous permet de mesurer l'association entre des variables qualitative contre qualitative), nous obtenons la valeur de : 0,56 %. Cela confirme notre analyse et nous indique une association presque nul entres ces 2 variables.

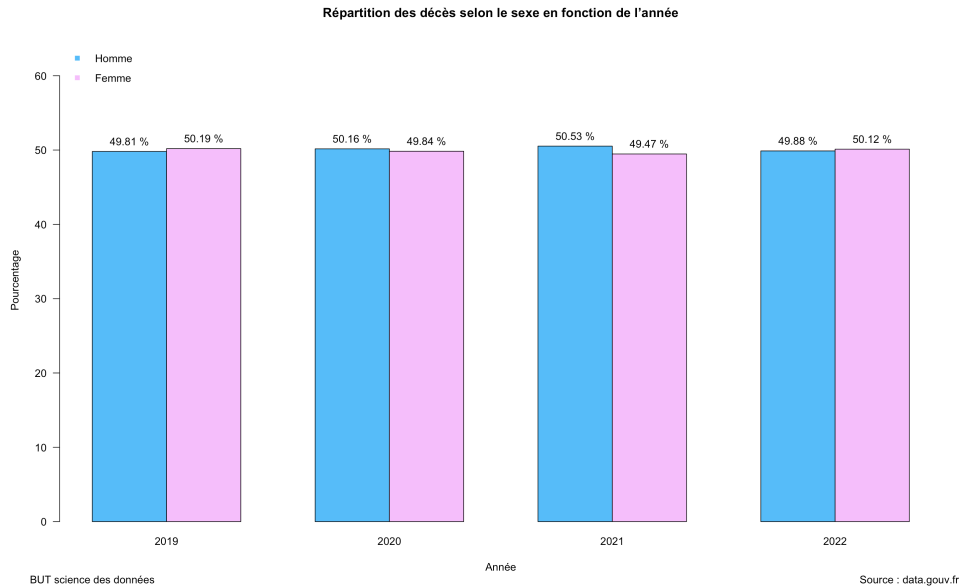


Fig. 2. Répartition des décès selon le sexe en fonction de l'année

1.2 Répartition des décès en 2022 selon la région

Ce graphique (cf. Fig. 3) nous permet d'identifier les régions avec le plus de décès. On constate que la région avec le plus grand nombre de décès en 2022 est la région Île-de-France. La région avec le moins de décès est la Corse. Cela est notamment explicable par le nombre d'habitants par région.

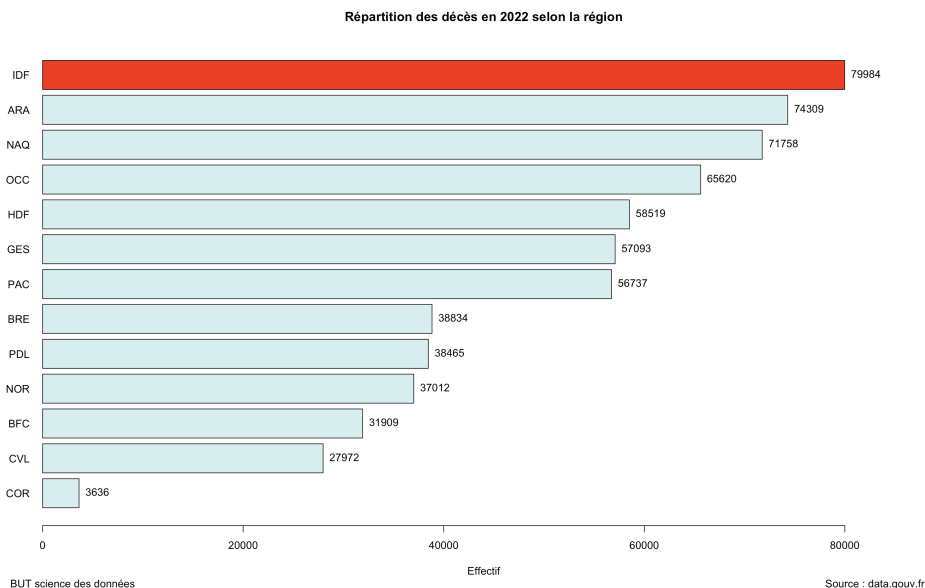


Fig. 3. Répartition des décès en 2022 selon la région

1.3 Répartition des décès en 2022 selon le sexe par région

Ci-dessous (cf. Fig. 4), le diagramme à barres empilées verticales nous montre la répartition des décès en 2022 selon le sexe par région. Nous constatons, que comme sur le premier graphique 1 (cf. Fig. 2) les proportions de décès par sexe en fonction des régions tendent vers 50 %. De ce fait, nous supposons que l'association entre la variable sexe et la variable région de décès est très faible. La valeur du V de Cramer est de 0,47 %. Cela confirme la faible association de ces deux variables.

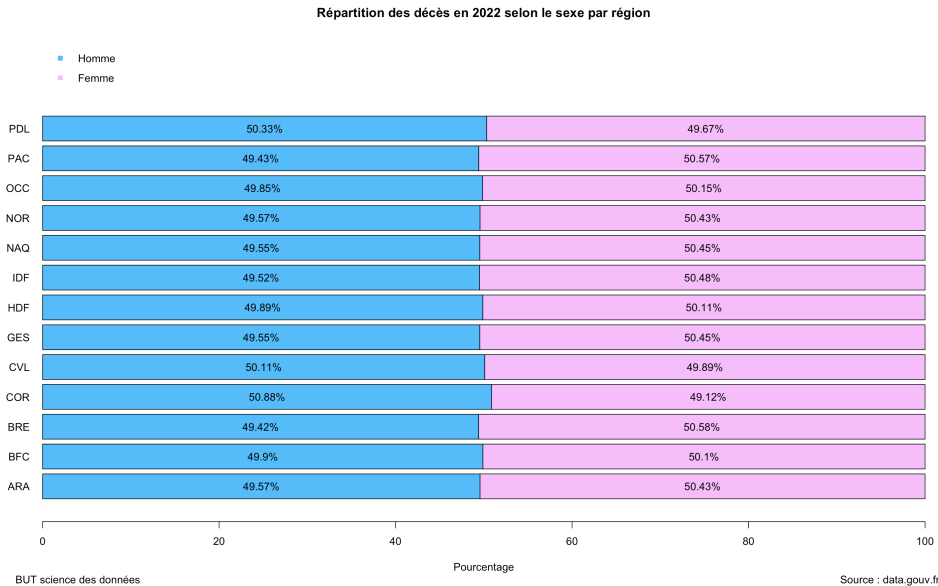


Fig. 4. Répartition des décès en 2022 selon le sexe par région

1.4 Répartition des décès selon l'âge des individus sur la période 2019 à 2022

Ci-dessous (cf. Fig. 5), cet histogramme permet de mettre en avant la répartition des décès sur la période de 2019 à 2022. Nous constatons que 50 % des décès occurred avant 85 ans et les 50 % autres après 85 ans. Il est intéressant de noter que la moyenne est à 80 ans (à gauche de la médiane) ce qui est explicable par un nombre conséquent de décès avant 80 ans. De plus, on constate un pic sur l'intervalle de 0 à 5 ans. On en conclut les individus se trouvant dans cette tranche sont plus susceptible de décéder. Cela est notamment dû au fait de la fragilité d'un enfant à cet âge.

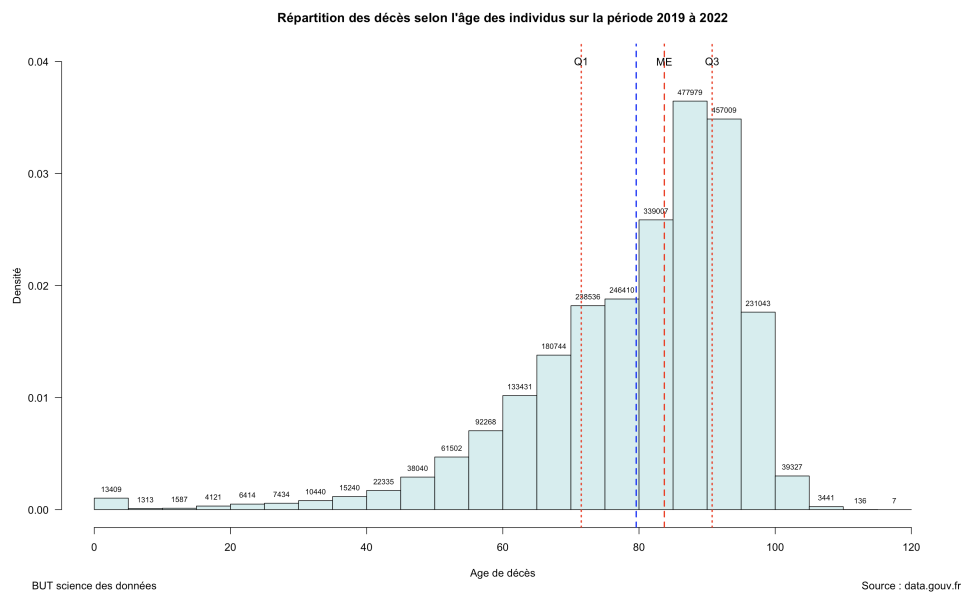


Fig. 5. Répartition des décès selon l'âge des individus sur la période 2019 à 2022

1.5 Répartition des décès selon l'âge des individus par région

Le graphique ci-dessous (cf. Fig. 6), nous permet de comparer les distributions conditionnelles entre la variable âge de décès et région. Sur ces boxplot parallèles, les moyennes qui sont misent en avant par les croix rouge sont relié par une ligne grise nous permettant de visualiser l'intensité de l'association, qui ici, semble faible. On notera qu'à l'aide du rapport de corrélation (qui nous permet de mesurer l'association entre des variables quantitative contre qualitative) nous obtenons une association de 0,57 % qui est synonyme d'une faible association. Cela signifie que les régions n'affectent pas l'âge de décès des individus.

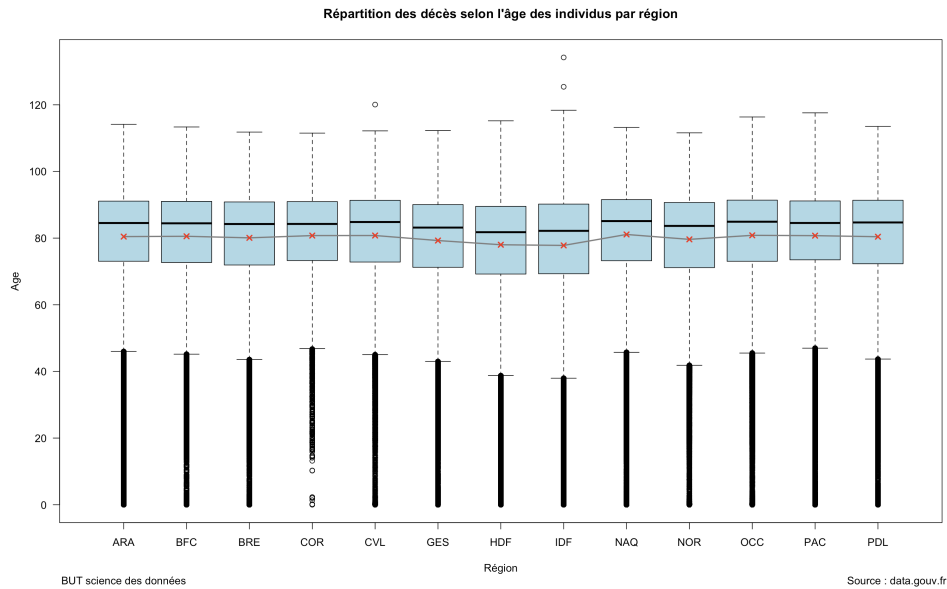


Fig. 6. Répartition des décès selon l'âge des individus par région

2 Les décès en Normandie

Après avoir étudié les décès sur l'ensemble du territoire français, nous allons, à travers cette sections nous concentrer sur l'étude des décès en Normandie et ses départements.

2.1 Répartition des décès en Normandie selon le département par année

À travers ce diagramme à barres juxtaposées (cf. Fig. 7), nous pouvons observer la répartition des décès dans les départements de Normandie pour les années de 2019 à 2022. On constate que les départements avec le plus de décès sont les mêmes sur les 4 ans avec le département de Seine-Maritime en premier, suivi du Calvados, en troisième position nous retrouvons la Manche, puis l'Eure et enfin l'Orne. Il semble y avoir une très faible association entre les départements et les années, car pour chaque année, nous retrouvons des valeurs similaires. Le V de Cramer entre les variables départements de décès et années est de 1,05 % ce qui confirme notre hypothèse précédente d'une très faible association entre les deux variables. Comme vu auparavant dans le graphique 2 (cf. Fig. 3) le diagramme reflète les effectifs de décès sans prendre en compte le nombre d'habitants du département, c'est pour cela que l'on retrouve cet ordre qui est similaire à celui du nombre d'habitants par département pour la Normandie.

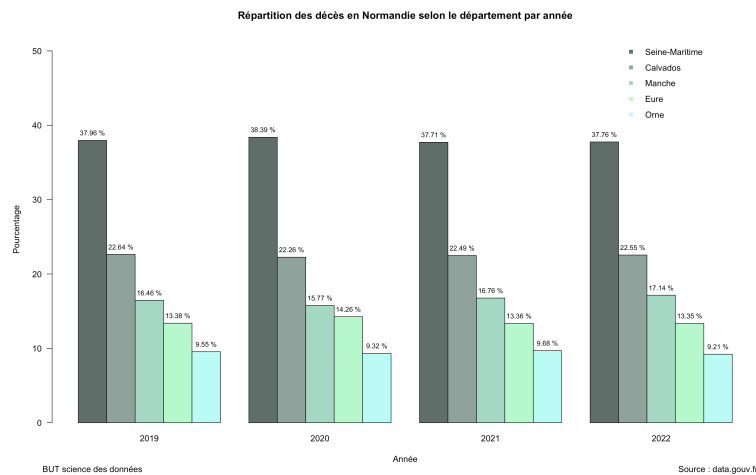


Fig. 7. Répartition des décès en Normandie selon le département par année

2.2 Répartition des décès en Normandie selon le sexe par département par année

Le diagramme à barre juxtaposé ci-dessous (cf. Fig. 8), nous permet de comparer les distributions des décès selon le sexe par département par année. De ce fait, nous observons que chaque année les proportions de décès en fonction du sexe pour chaque département normand se rapprochent de 50 %. De ce fait, nous supposons qu'il y a une faible association entre la variable sexe et département. Les V de Cramer obtenus pour les années 2019, 2020, 2021 et 2022 sont respectivement 1,08 %, 1,12 %, 0,87 % et 0,72 %, ce qui souligne la très faible association entre les variables.

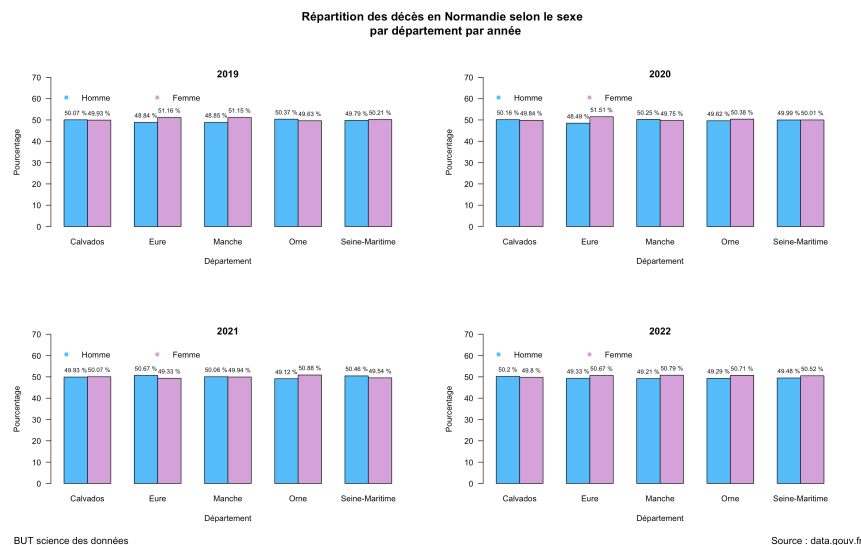


Fig. 8. Répartition des décès en Normandie selon le sexe par département par année

2.3 Répartition des décès en Normandie selon l'âge des individus en fonction du sexe

Les boxplot parallèles ci-dessous (cf. Fig. 9) nous offrent une vision claire de la distribution de l'âge de décès en fonction du sexe. On constate que les femmes semblent vivre en moyenne plus longtemps que les hommes. Le rapport de corrélation de 6,36 % nous indique une faible association entre ces variables. De plus, on constate que la tranche d'âge de décès des femmes est plus faible que celles des hommes, comme si ces derniers avaient plus de chances de mourir à n'importe quel âge contrairement aux femmes.

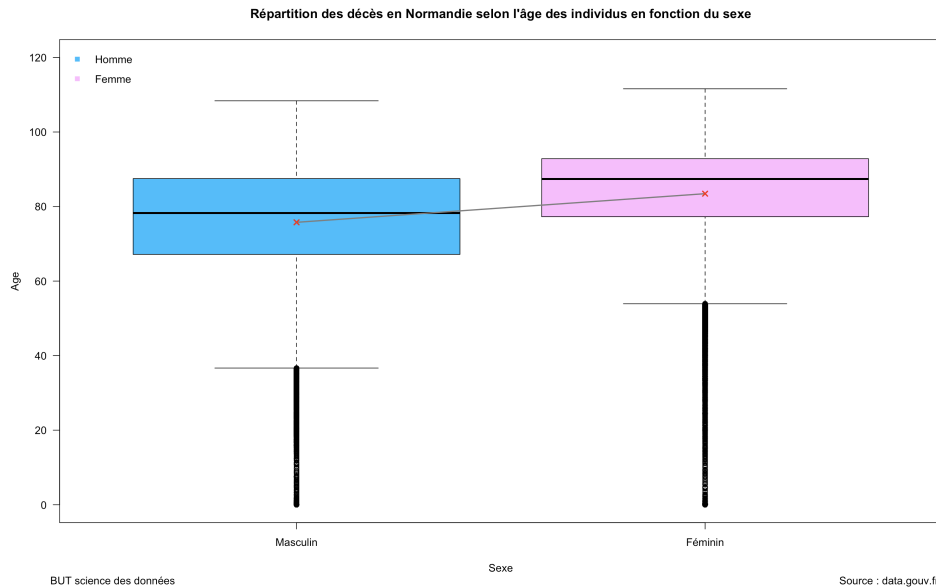


Fig. 9. Répartition des décès en Normandie selon l'âge des individus en fonction du sexe

2.4 Répartition des décès en Normandie selon l'âge des individus en fonction du sexe par département

Les graphiques ci-dessous (cf. Fig. 10), mettent en évidence les âges de décès en fonction du sexe pour chaque département de Normandie. Comme précédemment, les femmes semblent vivre plus longtemps peu importe le département. L'intervalle d'âge de décès des hommes est plus important que les femmes. De plus, on notera que pour les départements du Calvados, l'Eure, la Manche, l'Orne, la Seine-Maritime ont un coefficient de détermination avec la variable sexe de : 5,88 %, 7,61 %, 7,47 %, 6,70 %, 5,88 %, soit une association faible entre les variables sexes et âge de décès.

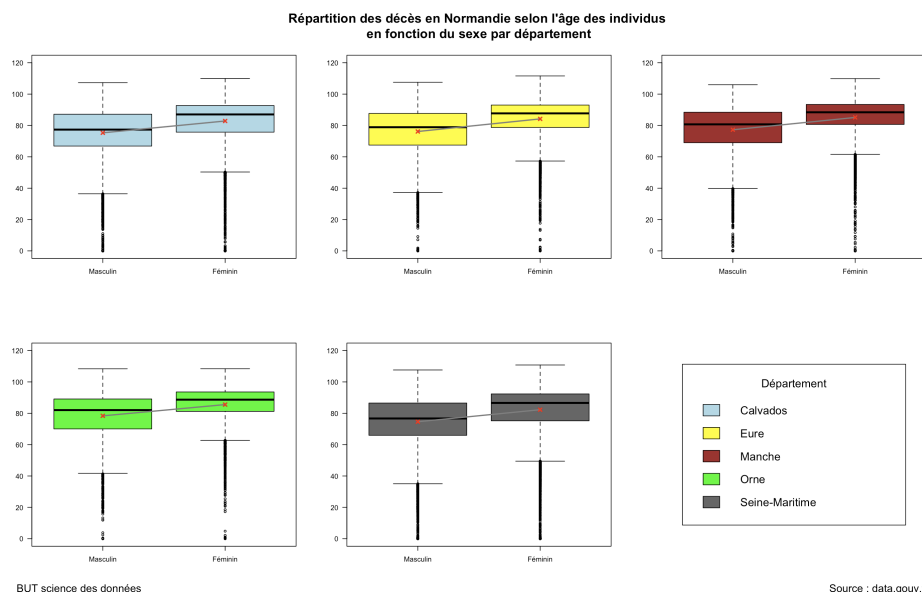


Fig. 10. Répartition des décès en Normandie selon l'âge des individus en fonction du sexe par département

Conclusion

À travers ce rapport, nous proposons une analyse des décès de la population française sur son territoire au cours des années de 2019 à 2022. Cette étude se divise en deux parties distinctes, la première montrant les caractéristiques des décès sur l'ensemble de la métropole, tandis que la seconde offre une analyse détaillée des décès survenus en Normandie.

La première partie met en avant l'analyse des décès en France. Premièrement, sur les 4 années, la proportion d'hommes est de femmes décédés tends vers 50 %, soit une répartition homogène de la variable sexe. De plus, on a pu observer que les proportions de décès entre les sexes sont similaires à travers les régions de la France. Nous avons observé que la région avec les plus de décès est l'Île-de-France et celle avec le moins de décès est la région Corse. En France, la moyenne de décès est de 79,59 ans et la médiane est de 83,72 ans.

Si l'on se concentre sur la Normandie, on observe que quelque soit l'année, c'est le département de la Seine-Maritime où il y a le plus de décès. Comme pour l'observation au niveau national, il y a une répartition homogène des décès en fonctions des sexes pour chaque département. L'âge de décès médian des femmes est supérieur à celui des hommes. On en conclut que les femmes vivent plus longtemps que les hommes, et cela, quels que soient les départements.

3 Annexe

Sur la page dédiée du site data.gouv.fr, nous avons un dessin d'enregistrement permettant de visualiser : le type, le format, la taille et d'autres informations complémentaires du fichier de données comme ci-dessous (cf. Fig. 11) :

Dessin d'enregistrement
Le fichier est fourni au format txt
<i>Nom et Prénom</i> - Longueur : 80 - Position : 1-80 - Type : Alphanumérique La forme générale est NOM*PRENOMS
<i>Sexe</i> - Longueur : 1 - Position : 81 - Type : Numérique 1 = Masculin; 2 = féminin
<i>Date de naissance</i> - Longueur : 8 - Position : 82-89 - Type : Numérique Forme : AAAAMMJJ - AAAA=0000 si année inconnue; MM=00 si mois inconnu; JJ=00 si jour inconnu
<i>Code du lieu de naissance</i> - Longueur : 5 - Position : 90-94 - Type : Alphanumérique Code Officiel Géographique en vigueur au moment de la prise en compte du décès
<i>Commune de naissance en clair</i> - Longueur : 30 - Position : 95-124 - Type : Alphanumérique
<i>DOM/TOM/COM/Pays de naissance en clair</i> - Longueur : 30 - Position : 125-154 - Type : Alphanumérique
<i>Date de décès</i> - Longueur : 8 - Position : 155-162 - Type : Numérique Forme : AAAAMMJJ - AAAA=0000 si année inconnue; MM=00 si mois inconnu; JJ=00 si jour inconnu
<i>Code du lieu de décès</i> - Longueur : 5 - Position : 163-167 - Type : Alphanumérique Code Officiel Géographique en vigueur au moment de la prise en compte du décès
<i>Numéro d'acte de décès</i> - Longueur : 9 - Position : 168-176 - Type : Alphanumérique
NOTA : Certains enregistrements peuvent contenir en toute fin des caractères non significatifs. Il est donc important, pour lire correctement ce champ, de bien respecter sa longueur ou sa borne de fin.

Fig. 11. Dessin d'enregistrement du fichier de données

En complément des fichiers sur les décès, nous avons récupéré sur le site de l'INSEE, les codes officiels des régions (mis à jour le 1er janvier 2023) et des départements. (cf. Fig. 12)



Fig. 12. Extrait du site internet de l'INSEE

[Site internet de l'INSEE](#)

Table des figures

1	Variables du jeu de données	3
2	Répartition des décès selon le sexe en fonction de l'année	4
3	Répartition des décès en 2022 selon la région	4
4	Répartition des décès en 2022 selon le sexe par région	5
5	Répartition des décès selon l'âge des individus sur la période 2019 à 2022	5
6	Répartition des décès selon l'âge des individus par région	6
7	Répartition des décès en Normandie selon le département par année	7
8	Répartition des décès en Normandie selon le sexe par département par année	7
9	Répartition des décès en Normandie selon l'âge des individus en fonction du sexe	8
10	Répartition des décès en Normandie selon l'âge des individus en fonction du sexe par département	8
11	Dessin d'enregistrement du fichier de données	10
12	Extrait du site internet de l'INSEE	10