

Benjamin Baillet  
Alexandra Millot  
Badr El Habti

# Finance et risques climatiques

## Data challenge Caisse des Dépôts et Consignations



Mathématiques  
Appliquées  
& Statistiques

Master  
*iref*

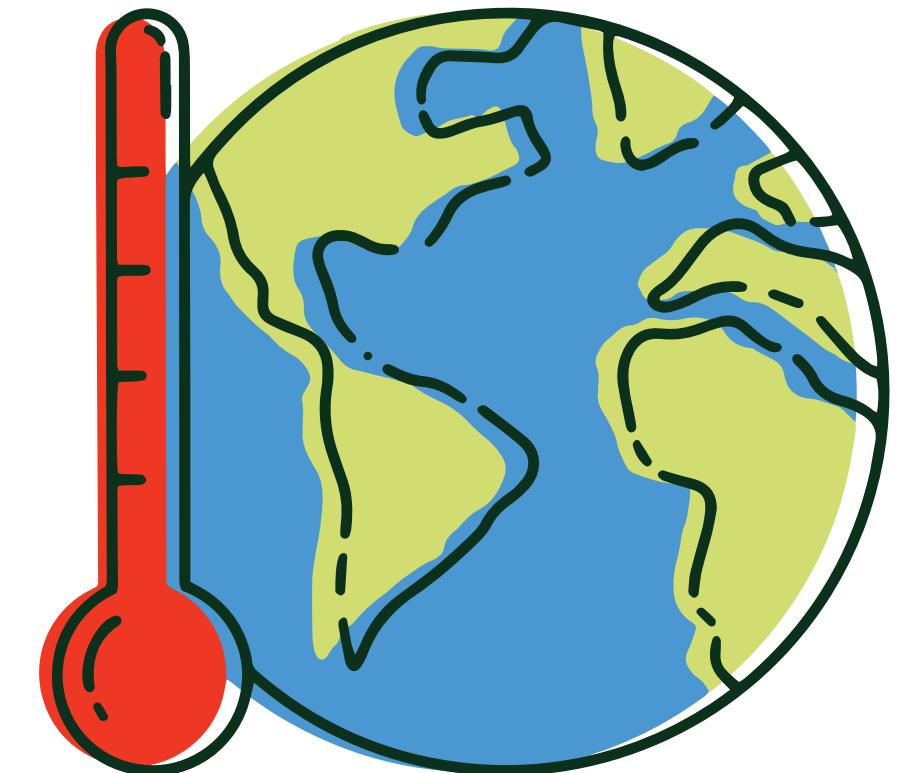
# Introduction Partie 1

Dans le cadre d'un projet de scoring des risques physiques liés au changement climatique, la Caisse des Dépôts (CDC) lance un défi aux participants : évaluer la vulnérabilité d'un portefeuille fictif d'actifs situés en France métropolitaine. L'objectif est de développer une méthode solide pour mesurer et représenter les risques climatiques, en s'appuyant sur des indicateurs pertinents et des outils d'analyse géospatiale.

## Nos objectifs

Nous avons 4 principaux objectifs :

- Choix des aléas climatiques ;
- Choix des indicateurs par aléa ;
- Estimation des scores des risques ;
- Analyse des scores grâce à la cartographie



# 01

## Les sources de données

# Sources de données

Nos données proviennent principalement du site **DRIAS** (Données climatiques de référence pour l'adaptation au changement climatique), où nous avons récupéré deux fichiers CSV essentiels pour notre analyse :

- **Fichier :** indices\_projets\_communes.csv
  - **Description :** Données climatiques détaillées pour les communes françaises.
  - **Utilisation :** Évaluation des indicateurs de sécheresse, inondations, feux de forêt et vagues de chaleur.
- **Fichier :** feux\_foret.csv
  - **Description :** Indicateurs spécifiques liés à la fréquence et à l'intensité des feux de forêt.
  - **Utilisation :** Étude des aléas de feu et intégration dans les modèles de clustering.

# Sources de données

Pour la cartographie, nous utilisons également des **données géographiques** (Shapefile des communes françaises) :

- **Fichier** : COMMUNE.shp
  - **Description** : Il s'agit des délimitations géographiques des communes françaises, qui sont essentielles pour visualiser les zones géographiques sur la carte.
  - **Utilisation** : Ces données sont utilisées comme base pour la cartographie des aléas (inondations, sécheresse, feux de forêt, etc.) à l'échelle des communes. Elles servent à afficher et à superposer les résultats des analyses de risque sur la carte.

Nous avons pour finir les **données du portefeuille d'actifs financiers** :

- **Fichier** : Portefeuille.xlsx
  - **Description** : Données financières des actifs, incluant les secteurs d'activité et les maturités.
  - **Utilisation** : Pondération des risques climatiques par secteur et évaluation des scores de risque.



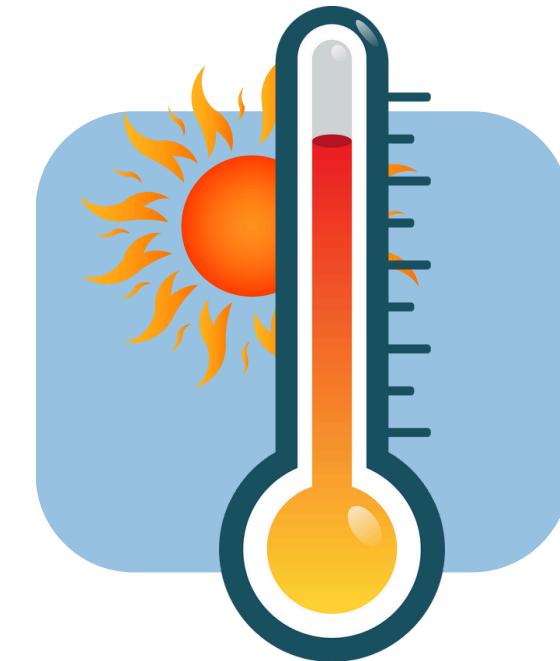
02

Les aléas et leurs indicateurs

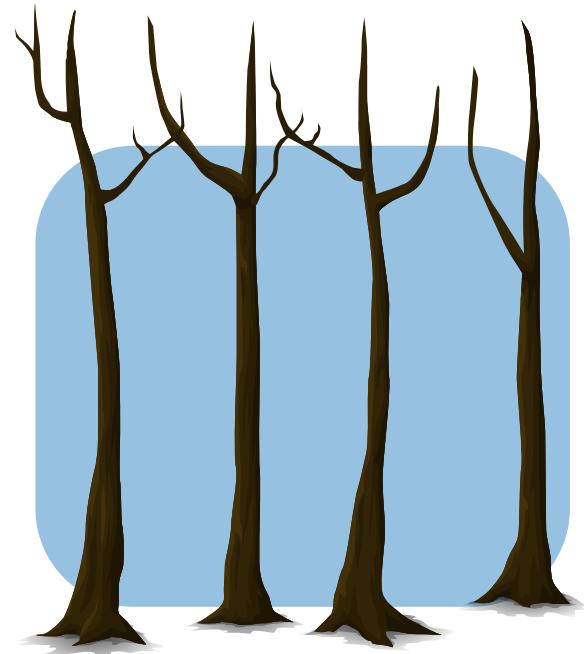
# Choix des aléas



Feux de  
forêts



Vagues de  
chaleur



Sécheresse



Inondations

# Les indicateurs - Feux de forêts

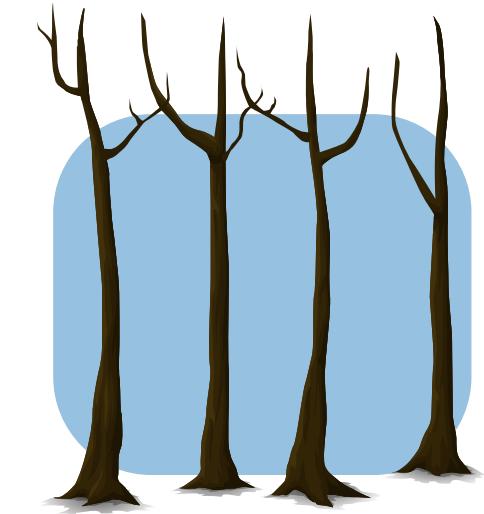


## Indicateurs utilisé :

- NORTX35 / nj\_f\_chal : Mesure la fréquence des jours de forte chaleur. Les fortes chaleurs augmentent le stress hydrique des végétaux, réduisent l'humidité ambiante et créent des conditions favorables à l'inflammabilité des forêts, ce qui en fait un indicateur pertinent pour anticiper les risques.
- NORNSV25 /feu\_forêt : évalue la sécheresse très forte de la végétation vivante, correspondant à des jours où le niveau de sécheresse NSV2 atteint ou dépasse 5. Cet état représente un dessèchement extrême de la strate arbustive, rendant la végétation extrêmement vulnérable au feu et favorisant les conditions de propagation de très grands feux.

En combinant ces deux variables, l'analyse intègre à la fois les conditions climatiques propices aux feux et l'occurrence réelle des incendies.

# Les indicateurs - Sécheresse



## Indicateurs utilisés :

Les variables `sech` (période de sécheresse en jours) et `nj_f_chal` (nombre de jours de forte chaleur) ont été sélectionnées pour leur complémentarité dans la caractérisation des phénomènes de sécheresse.

D'une part, la variable `sech` initialement `NORPXCDD` fournit une mesure absolue en jours, ce qui est essentiel pour quantifier la gravité et la durée réelle des épisodes de sécheresse.

D'autre part, la variable `nj_f_chal` initialement `NORTX35` exprime le nombre de jours de forte chaleur. Cet indicateur est particulièrement adapté pour évaluer la prévalence des sécheresses.

Ensemble, ces deux variables offrent une représentation complète et équilibrée du phénomène.

# Les indicateurs - Inondations



## Indicateurs utilisés :

Les variables `prec_cumul` (cumul de précipitations en millimètres), `max_jourpluie_cons` (nombre maximum de jours consécutifs de pluie) et `sech` ont été choisies pour leur pertinence dans l'analyse des risques liés aux inondations. Ces variables permettent d'évaluer différents aspects complémentaires de ce phénomène climatique.

La variable `prec_cumul` initialement NORRR mesure la quantité totale de précipitations sur une période donnée. Cet indicateur est crucial pour estimer le potentiel d'accumulation d'eau et les risques associés, comme les débordements de cours d'eau ou les saturations des sols, qui peuvent entraîner des inondations.

La variable `sech` initialement fournit une mesure absolue en jours et favorise les inondations. En effet, plus il y a de sécheresse et plus il pleut, plus les risques d'inondations sont élevés.

# Les indicateurs - Inondations

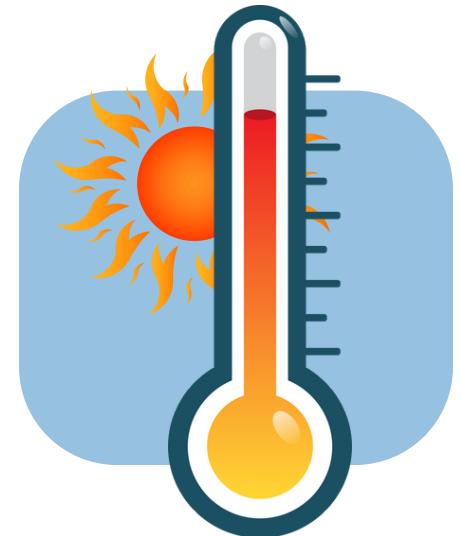


## Indicateurs utilisés :

La variable `max_jourpluie_cons` initialement `NORPXCWD` capture la durée des épisodes de pluie consécutifs. Une période prolongée de précipitations peut significativement augmenter les risques d'inondation, même si le cumul total reste modéré. Cet indicateur est donc essentiel pour analyser la temporalité des épisodes pluvieux et leur impact.

Ensemble, ces 3 variables offrent une représentation complète et équilibrée du phénomène.

# Les indicateurs - Vagues de chaleur



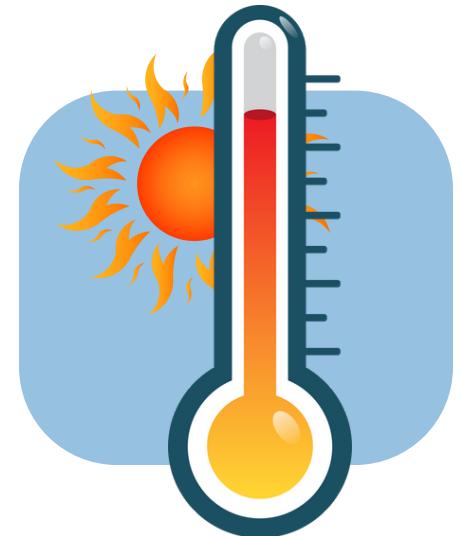
## Indicateurs utilisés :

Les variables `temp_max` (température maximale en °C) et `nj_vdc` (nombre de jours dans une vague de chaleur) ont été sélectionnées pour leur pertinence dans l'analyse des vagues de chaleur.

La variable `temp_max` initialement `NORTXAV`, mesure les températures maximales atteintes sur une période donnée. En analysant les températures maximales, on peut évaluer l'intensité des vagues de chaleur et identifier les régions les plus exposées.

La variable `nj_vdc` initialement `NORTXHWD`, capture la durée des vagues de chaleur, définies comme des périodes où la température maximale dépasse de plus de 5 °C la normale saisonnière pendant au moins 5 jours consécutifs.

# Les indicateurs - Vagues de chaleur



## Indicateurs utilisés :

En combinant `temp_max` et `nj_vdc`, il est possible d'obtenir une vision complète des vagues de chaleur. La mesure en **intensité** et en **durée** permet de mieux comprendre la gravité de ces phénomènes, d'évaluer leurs impacts sur les différentes régions et de modéliser les risques futurs dans un contexte de changement climatique.

Les indicateurs de chaque aléa sont combinés afin de générer un score pour chaque commune, représentant le niveau de risque associé à cet aléa.

# 03

## Création des clusters et Cartographies

# Méthode de création des clusters

Nous avons privilégié le **clustering** pour son approche inductive, qui évite d'imposer des seuils arbitraires et permet aux données de révéler naturellement leurs structures. Cela fonctionne bien pour tous les indicateurs sauf pour inondations.

La méthode de **clustering K-Means** a été choisie pour ce projet en raison de sa pertinence face à la nature des données. Elle présente plusieurs avantages clés :

- **Optimisation des classes** : K-Means minimise la variance intra-classe et maximise la séparation inter-classe, s'adaptant ainsi automatiquement à la distribution réelle des données.
- **Objectivité** : La classification repose sur des critères mathématiques rigoureux, évitant les découpages arbitraires.
- **Gestion des valeurs extrêmes** : K-Means isole les valeurs critiques en créant des clusters spécifiques, permettant de mieux visualiser les **zones à risque** et de valoriser les **données essentielles**.

Cette approche garantit une **représentation analytique et visuelle précise** des variations internes des données.

# Méthode de création des clusters

## K-Means

### Fonctionnement de l'algorithme :

#### Étape 1 : Initialisation

- Choisir un nombre de clusters K (défini par l'utilisateur).
- Initialiser aléatoirement K centroïdes (ou utiliser une méthode optimisée comme K-Means pour sélectionner des centres initiaux bien espacés).

#### Étape 2 : Attribution des points aux clusters

- Pour chaque point de données, calculer la distance entre le point et chaque centroïde (souvent avec la distance euclidienne).
- Assigner chaque point au cluster dont le centroïde est le plus proche.

#### Étape 3 : Mise à jour des centroïdes

- Recalculer les centroïdes des clusters en prenant la moyenne des coordonnées des points appartenant à chaque cluster.

#### Étape 4 : Répétition

- Répéter les étapes 2 (attribution) et 3 (mise à jour) jusqu'à convergence, c'est-à-dire :
  - Les centroïdes ne bougent plus significativement.
  - Ou le nombre maximum d'itérations est atteint.

# Méthode de création des clusters

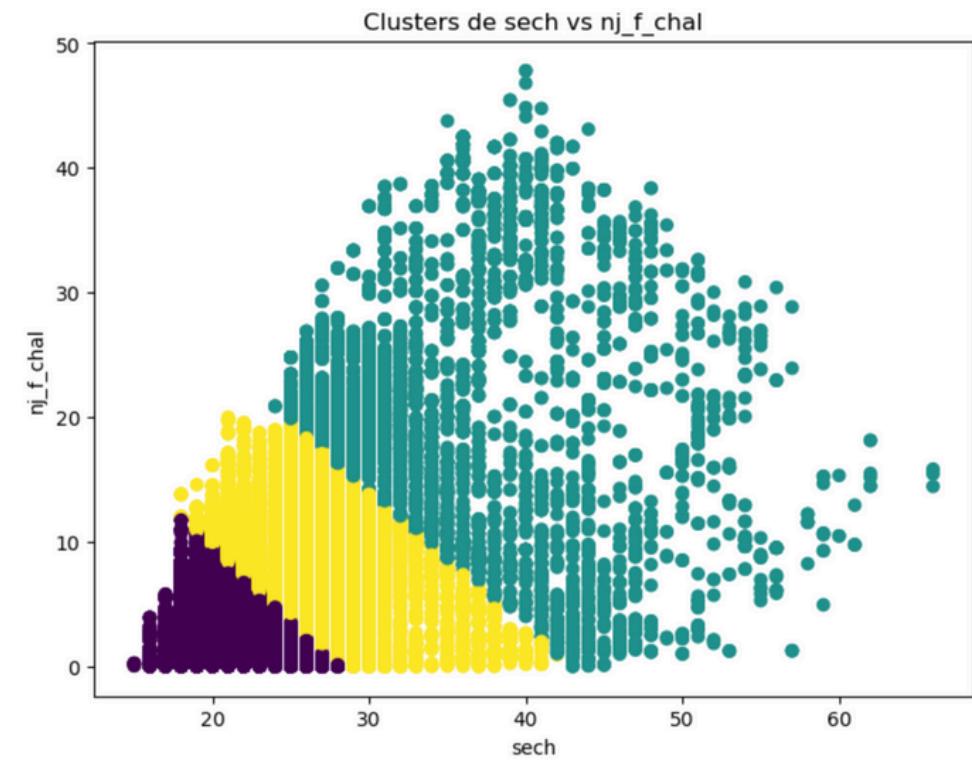
## K-Means

Pour décider combien de clusters (K) utiliser nous employons la **méthode du coude** pour tous les indicateurs:

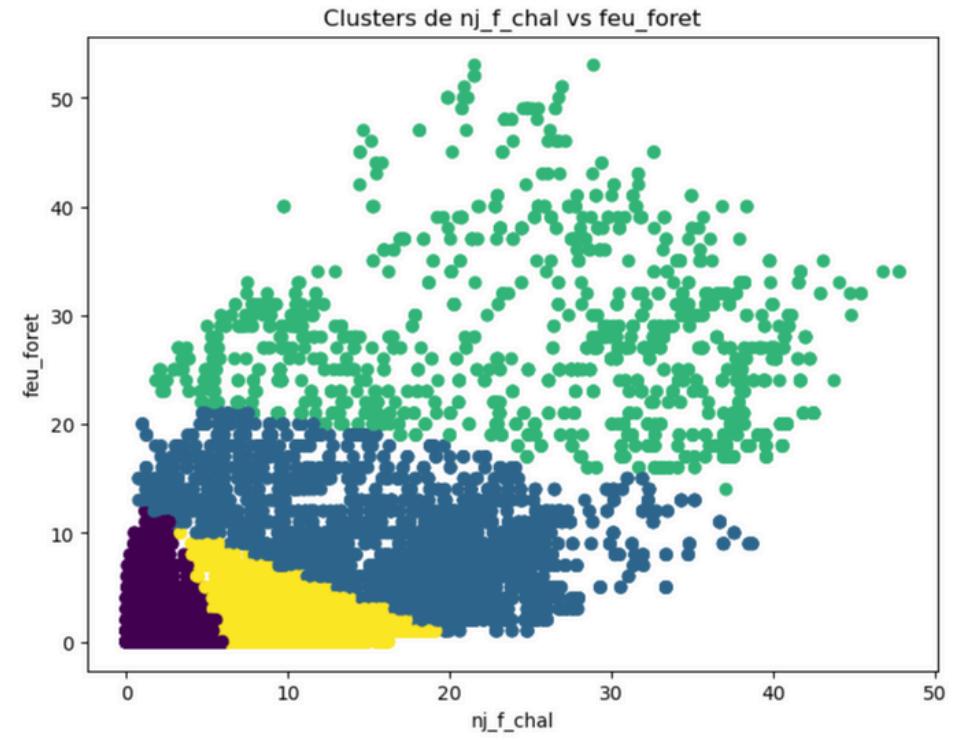
- L'algorithme K-Means est exécuté pour différents nombres de clusters (ici, de 1 à 10).
- Pour chaque valeur de K, on calcule la somme des carrés des distances des points à leur centroïde (appelée inertie intra-classe).
- Les résultats sont tracés sur un graphique.
- Le "coude" de la courbe (où la diminution de l'inertie ralentit) indique le nombre de clusters optimal.

# Résultats des clusters

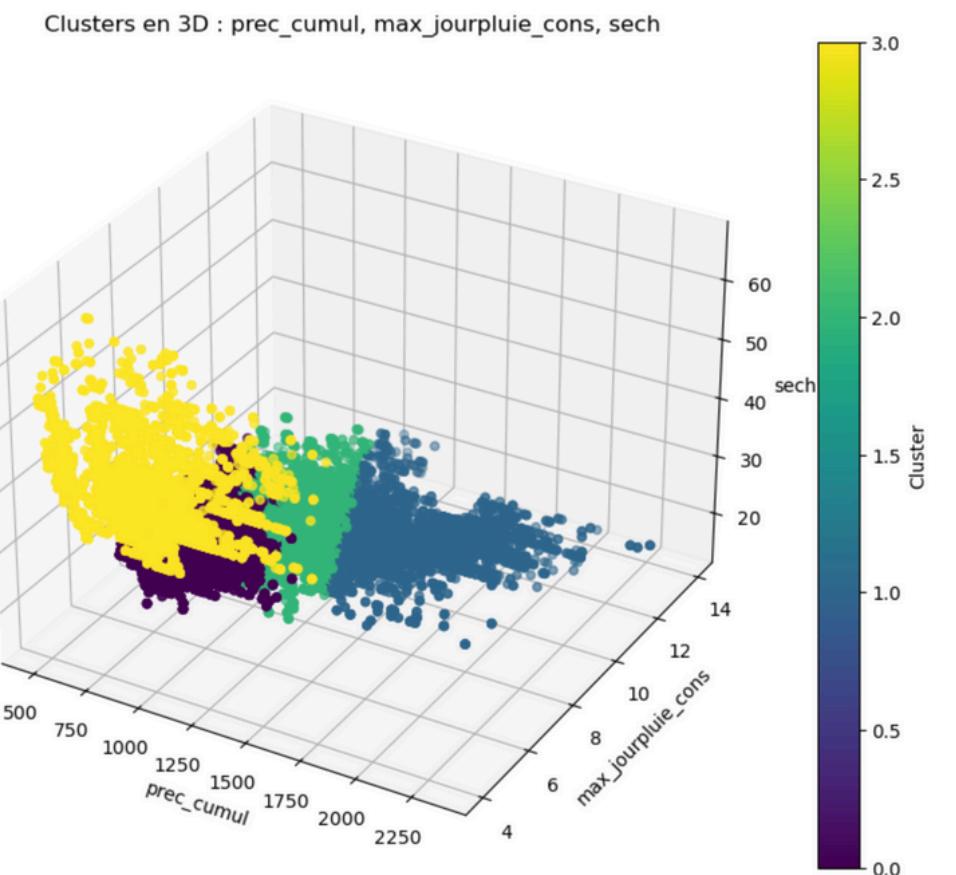
Sécheresse



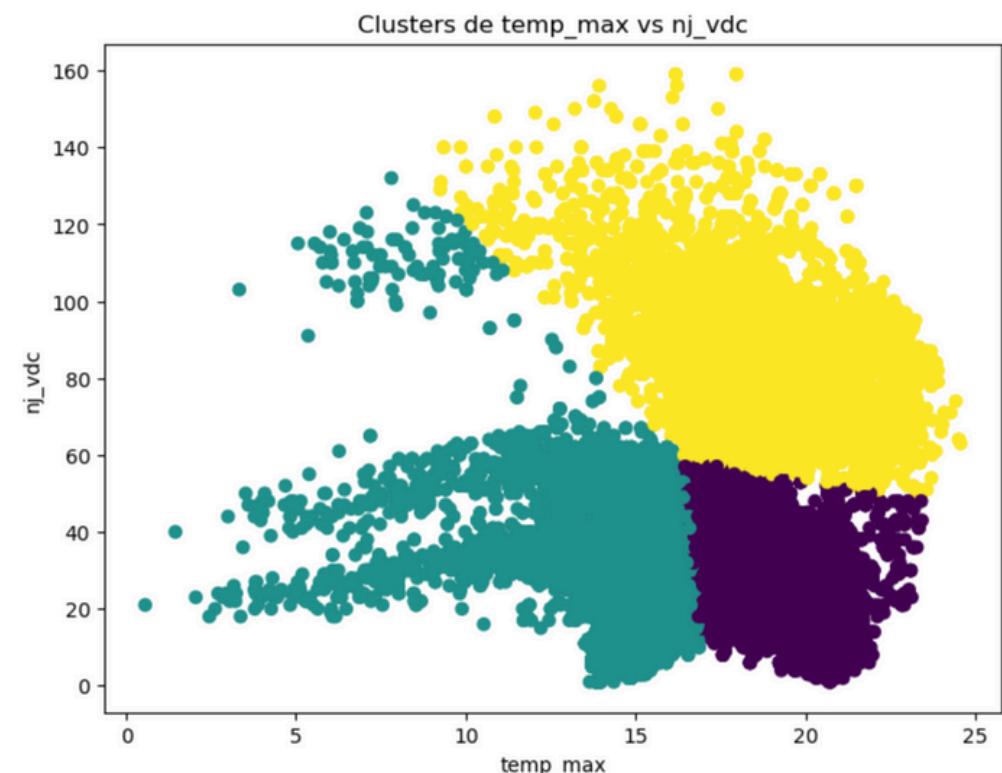
Feux de forêt



Inondations



Vagues de chaleur



# Cartographies

Pour la création des cartes, nous avons décidé d'utiliser **GéoPandas** pour les cartes représentant les différentes segmentations temporelles et utiliser **QGIS** pour les cartes représentant l'ensemble de la période étudiée.

Le niveau de score correspond au niveau de risque :

**Score 0 : Risque faible**

**Score 1 : Risque modéré**

**Score 2 : Risque élevé**

**(Score 3 : Risque critique)**

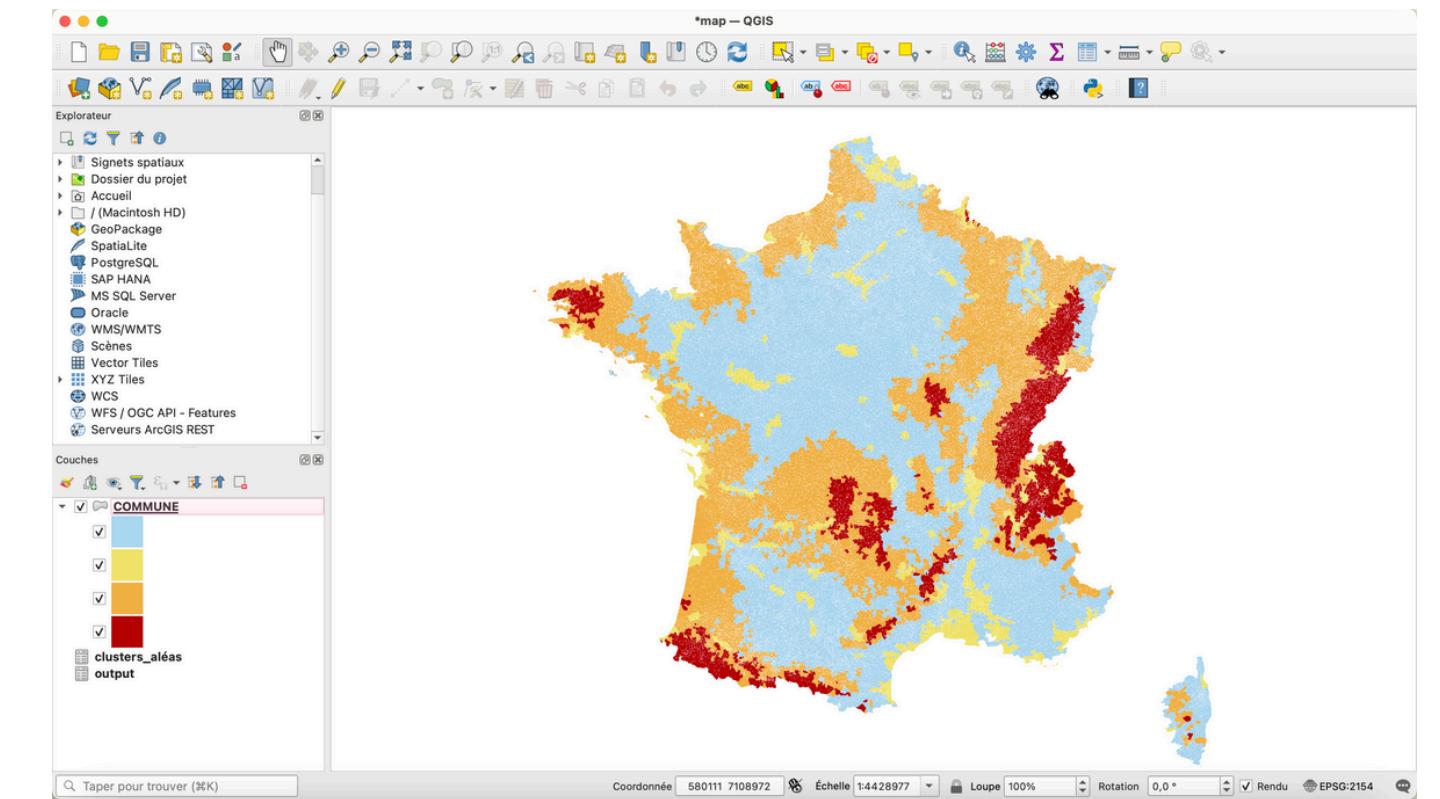
 Pour l'affichage des cartes avec QGIS nous avons directement récupéré les scores obtenus grâce à notre méthode de clustering.

Nous travaillons sur trois horizons temporels :

**H1: 2021-2050**

**H2: 2051-2070**

**H3: 2071-2100**



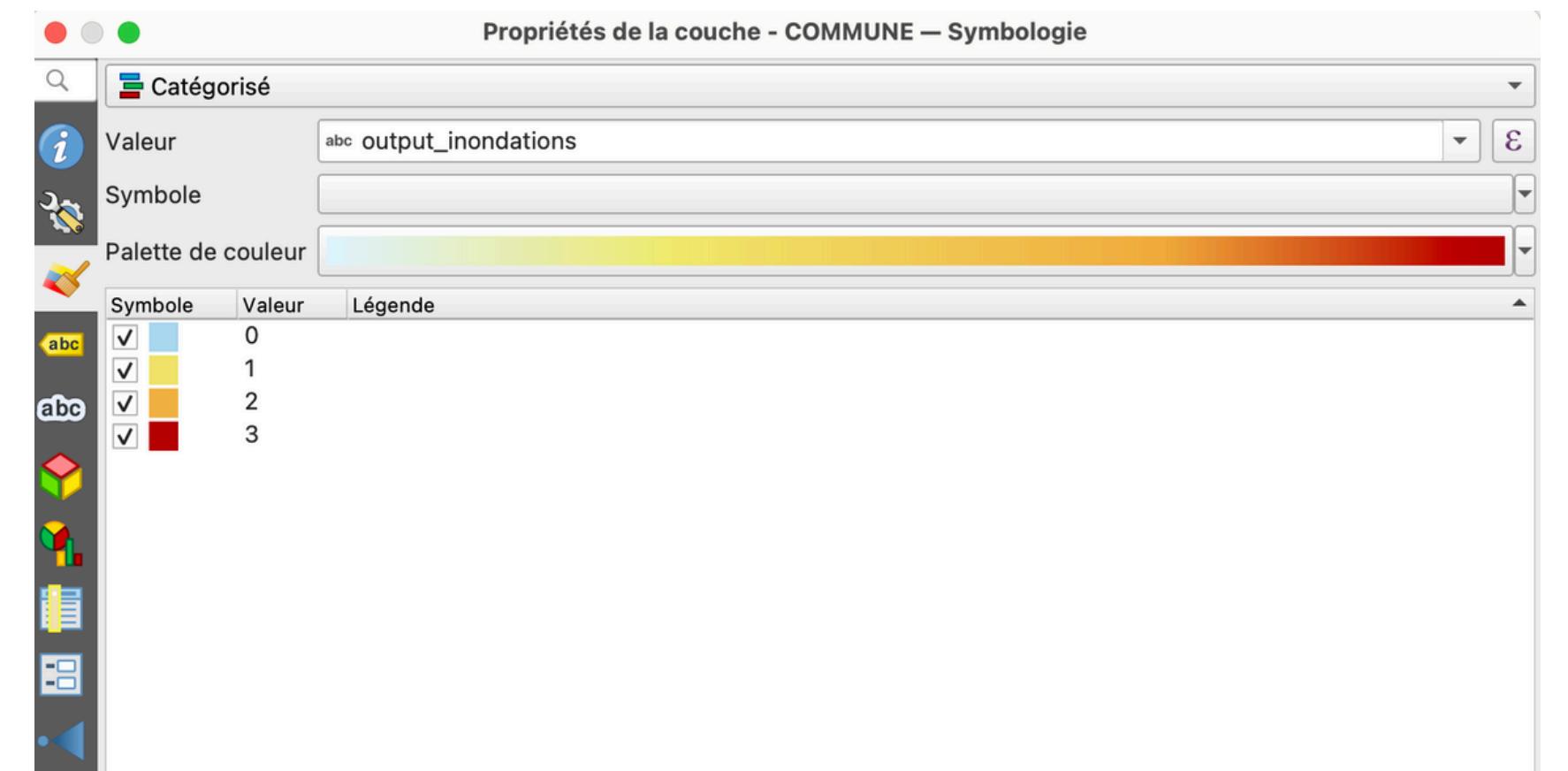
# Méthodologie des cartes QGIS

Pour créer les cartes QGIS, nous utilisons les colonnes de scores pour chaque aléa, obtenues à l'aide de la méthode des k-means.

Nous importons le fichier CSV ainsi que le Shapefile des communes françaises. Ensuite, nous réalisons une jointure entre les deux fichiers en associant les colonnes "insee" du CSV et "INSEE\_COM" du Shapefile via l'onglet *Jointures*.

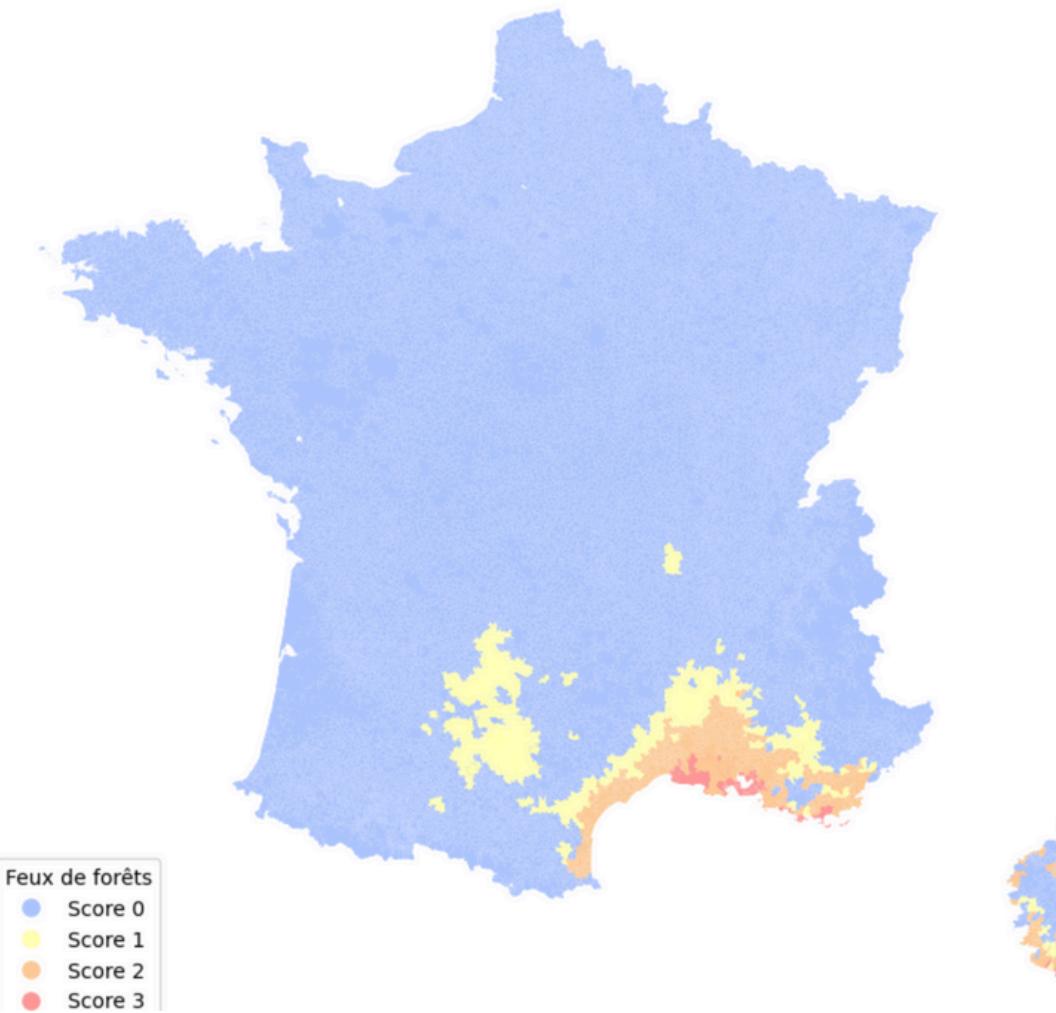
Enfin, dans l'onglet *Symbologie*, il suffit de choisir l'option d'affichage "Catégorisé", de sélectionner la colonne correspondant aux scores de l'aléa à traiter, puis d'attribuer une couleur distincte à chaque catégorie de score.

À partir de là, il ne reste qu'à finaliser la mise en page en ajoutant la légende, le titre et l'échelle.

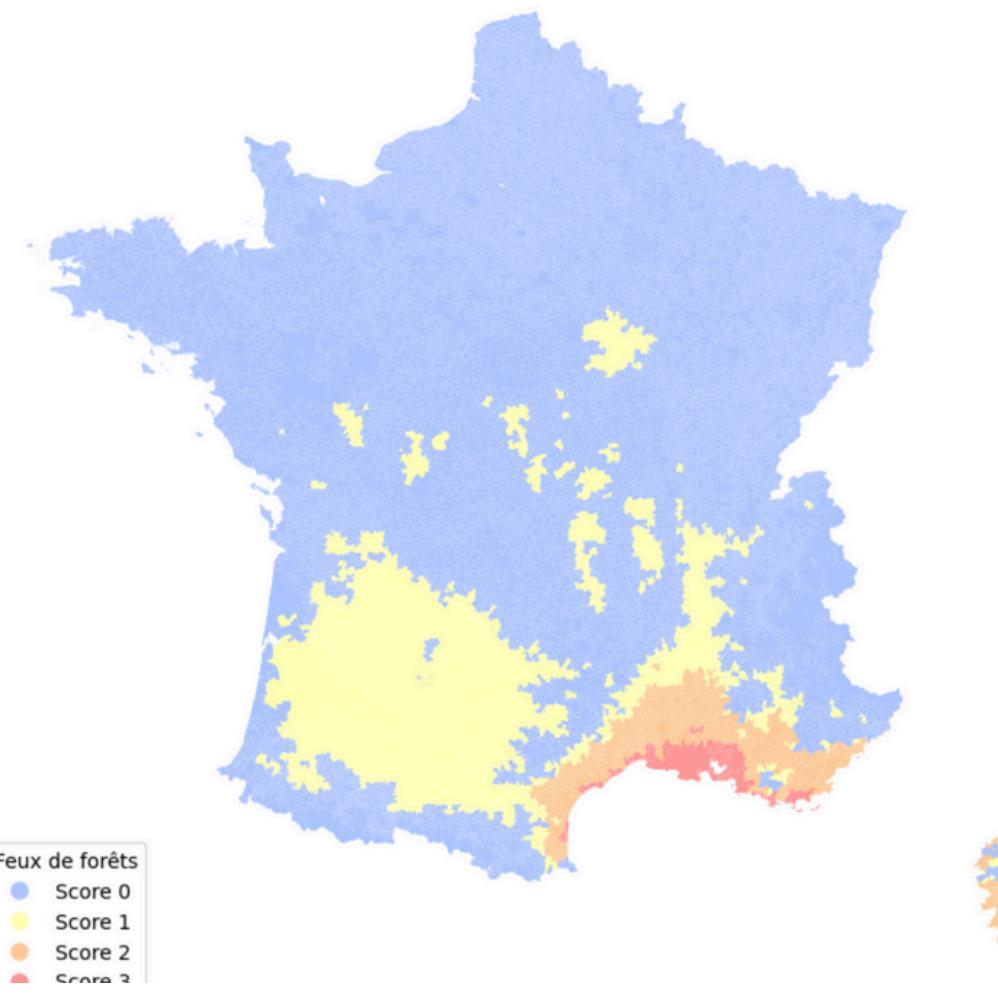


# Cartographie - Feux de forêts

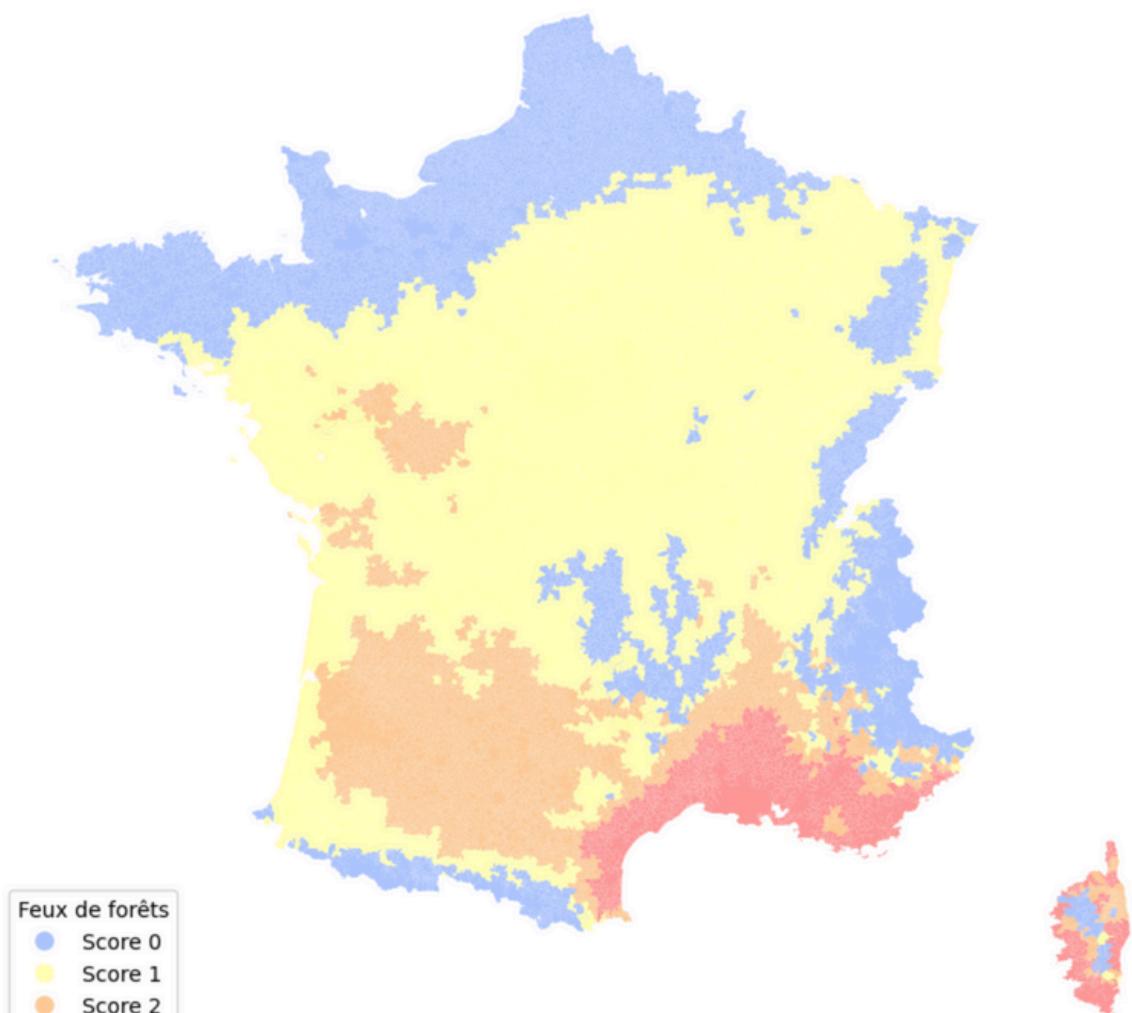
Feux de fôrets en France (Période H1)



Feux de fôrets en France (Période H2)



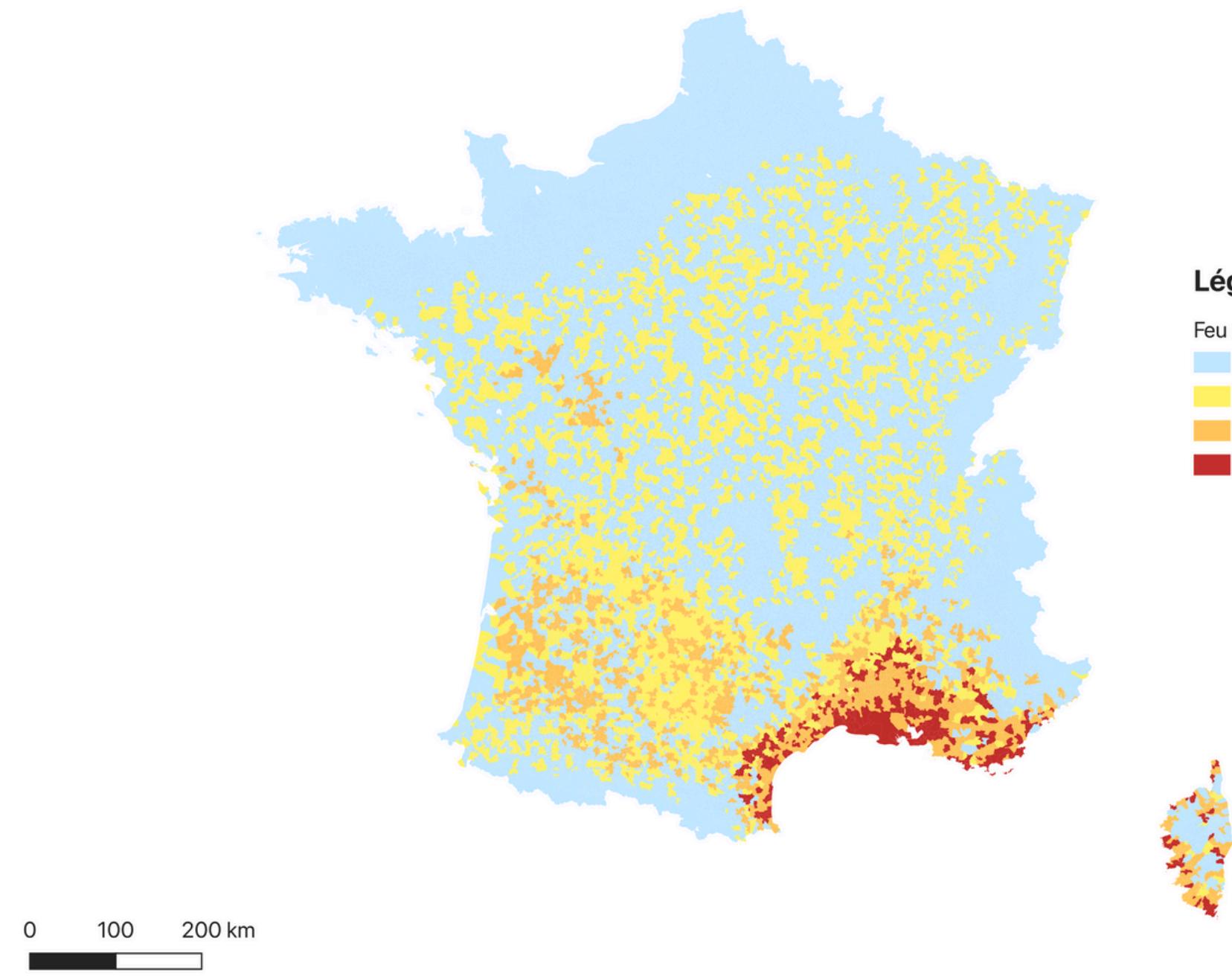
Feux de fôrets en France (Période H3)



L'évolution du risque de feux de forêts en France montre une tendance générale à l'aggravation, avec une augmentation significative du risque modéré, élevé, et critique à mesure que la période avance. Cela souligne la nécessité d'une préparation accrue pour lutter contre les incendies dans les années à venir, particulièrement dans les zones les plus vulnérables du sud et du sud-est du pays.

# Cartographie QGIS - Feux de forêts

Risques de Feu de forêt en France



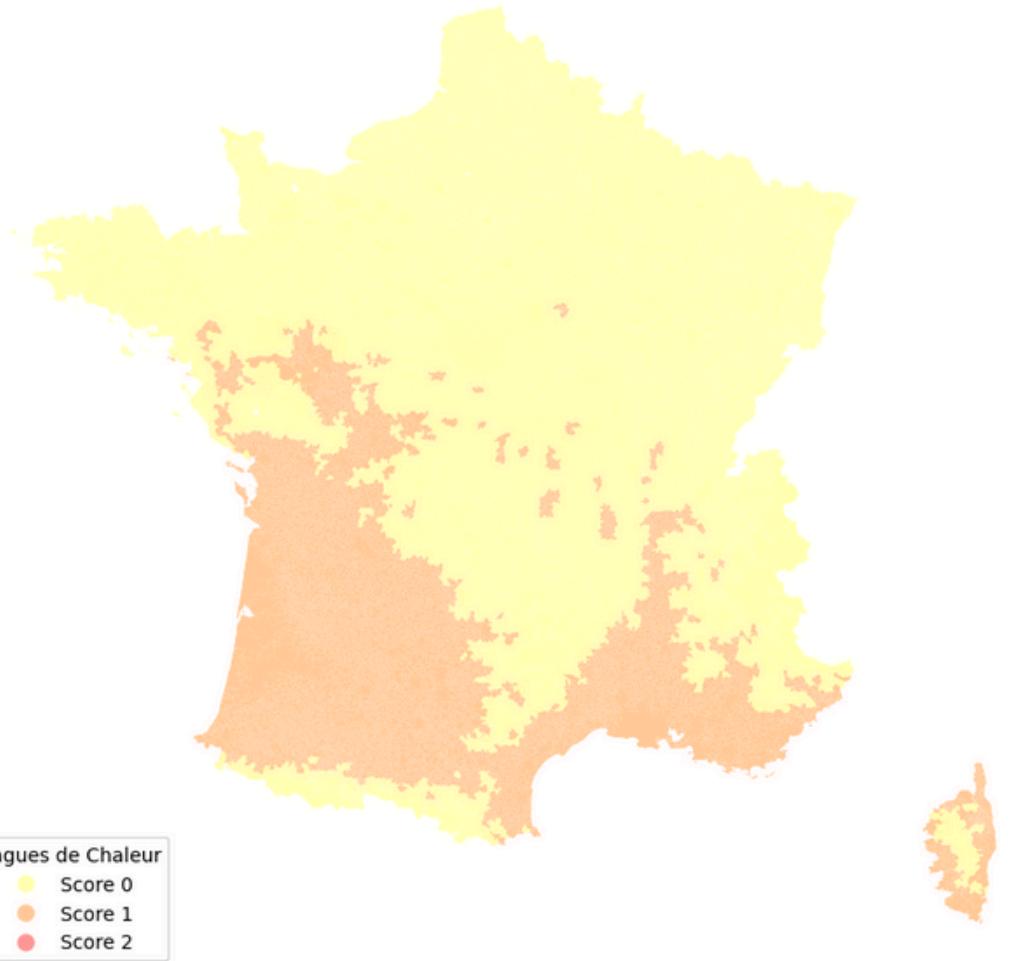
Cette carte illustre les risques de feux de forêt en France sur l'ensemble de la période, offrant une vue globale des zones à risque.

Le sud, notamment la région Méditerranéenne, présente un risque élevé à critique, soulignant la nécessité d'une gestion préventive stricte.

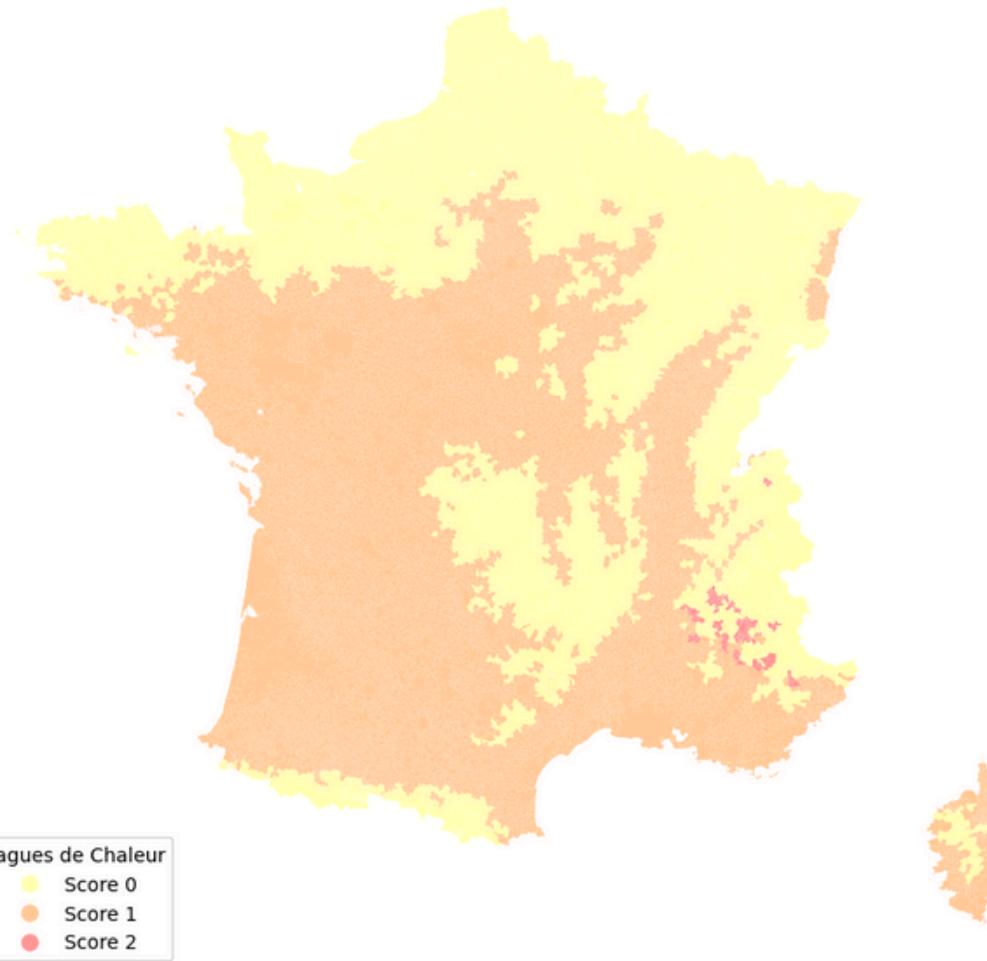
Le reste du pays montre une répartition plus diversifiée, avec des zones à risque faible à modéré, particulièrement dans le nord et l'est, où les conditions sont moins propices à la propagation des feux.

# Cartographie - Vagues de chaleur

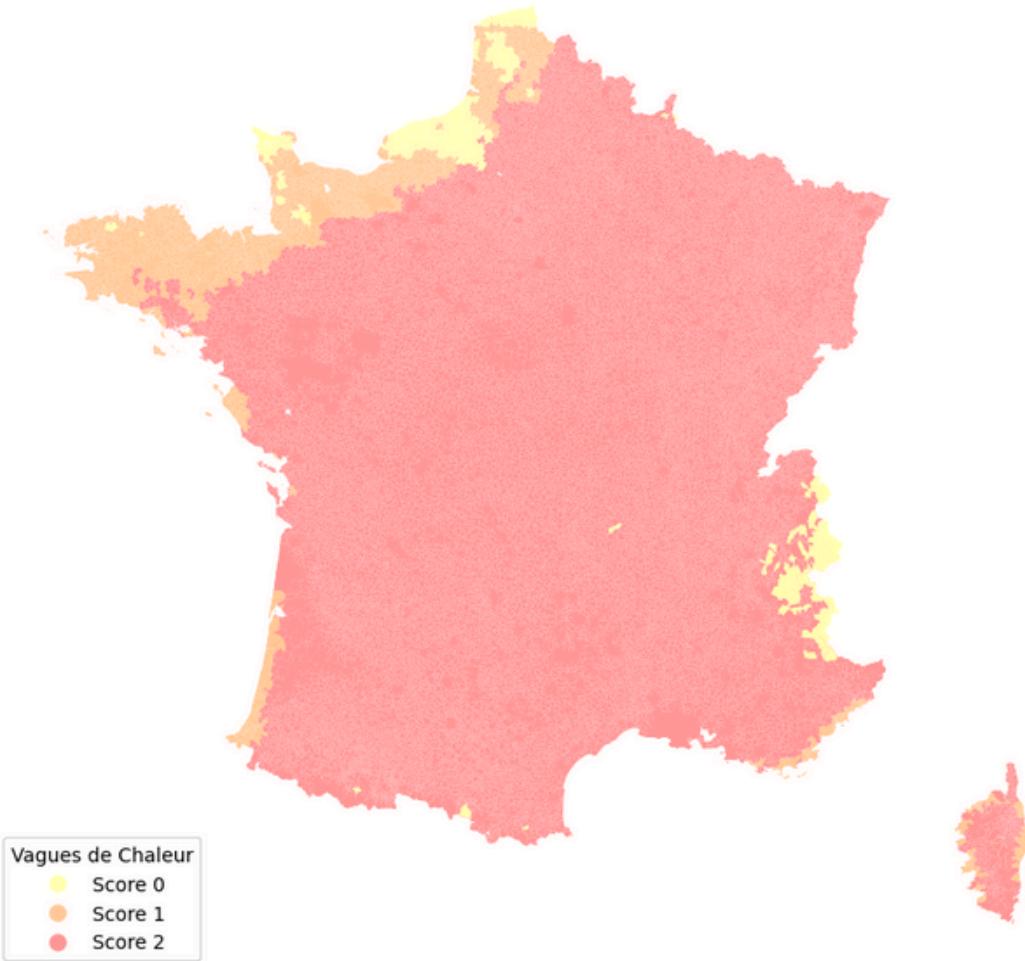
Vagues de Chaleur en France (Période H1)



Vagues de Chaleur en France (Période H2)



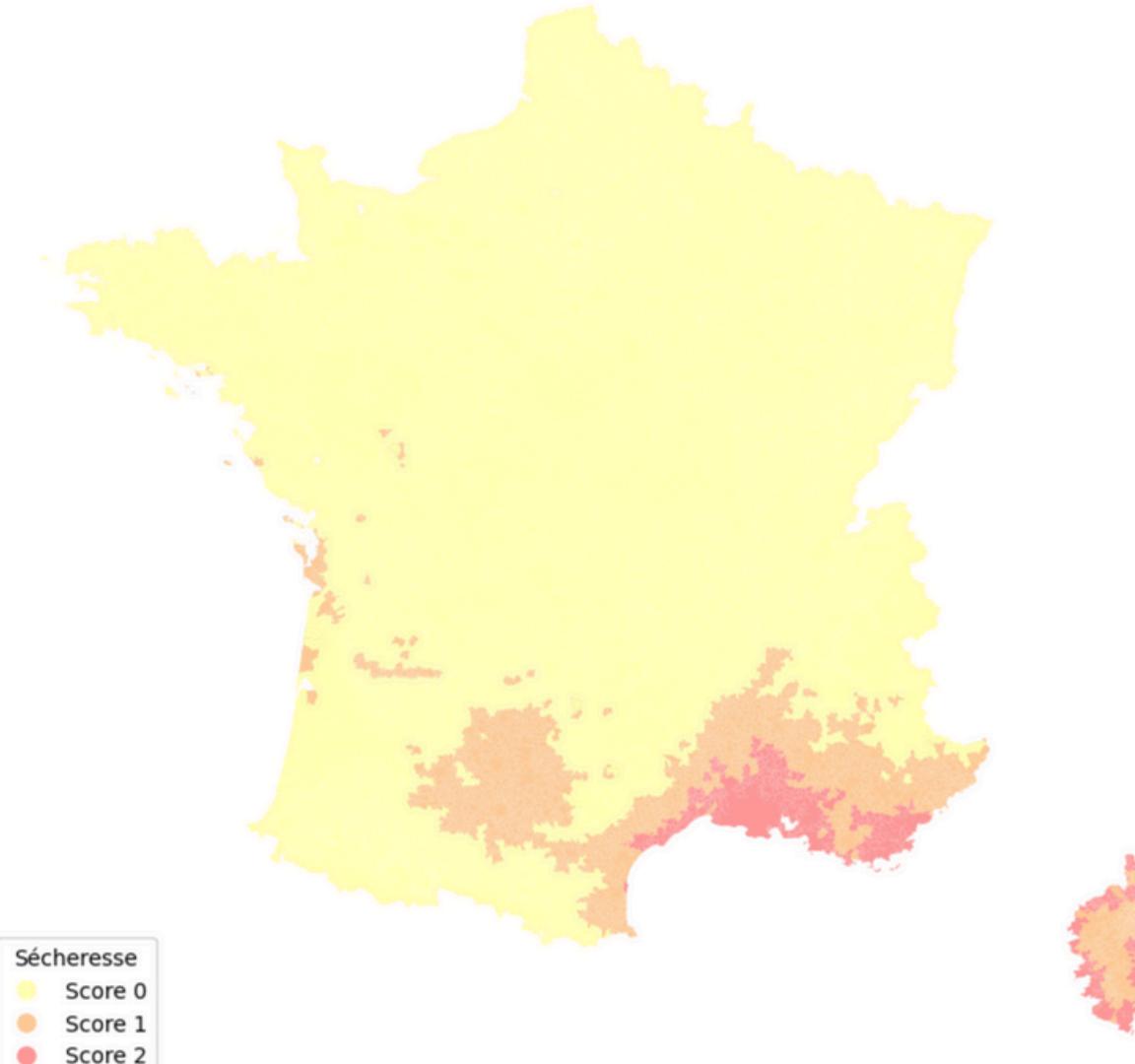
Vagues de Chaleur en France (Période H3)



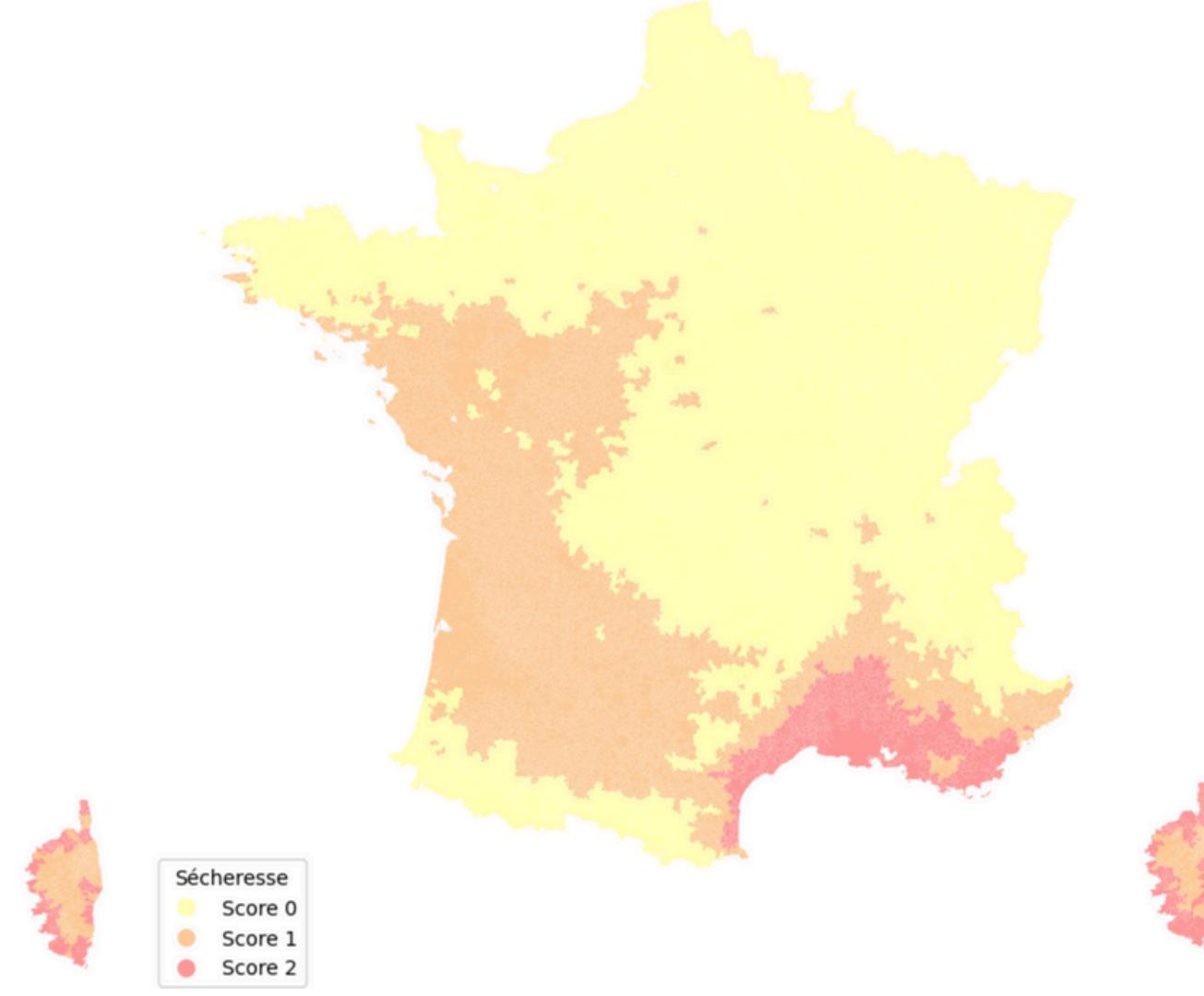
Les cartes montrent une augmentation généralisée du risque de vagues de chaleur à travers les périodes, avec un passage de zones de faible risque à un risque élevé, en particulier dans le sud et le centre du pays.

# Cartographie - Sécheresse

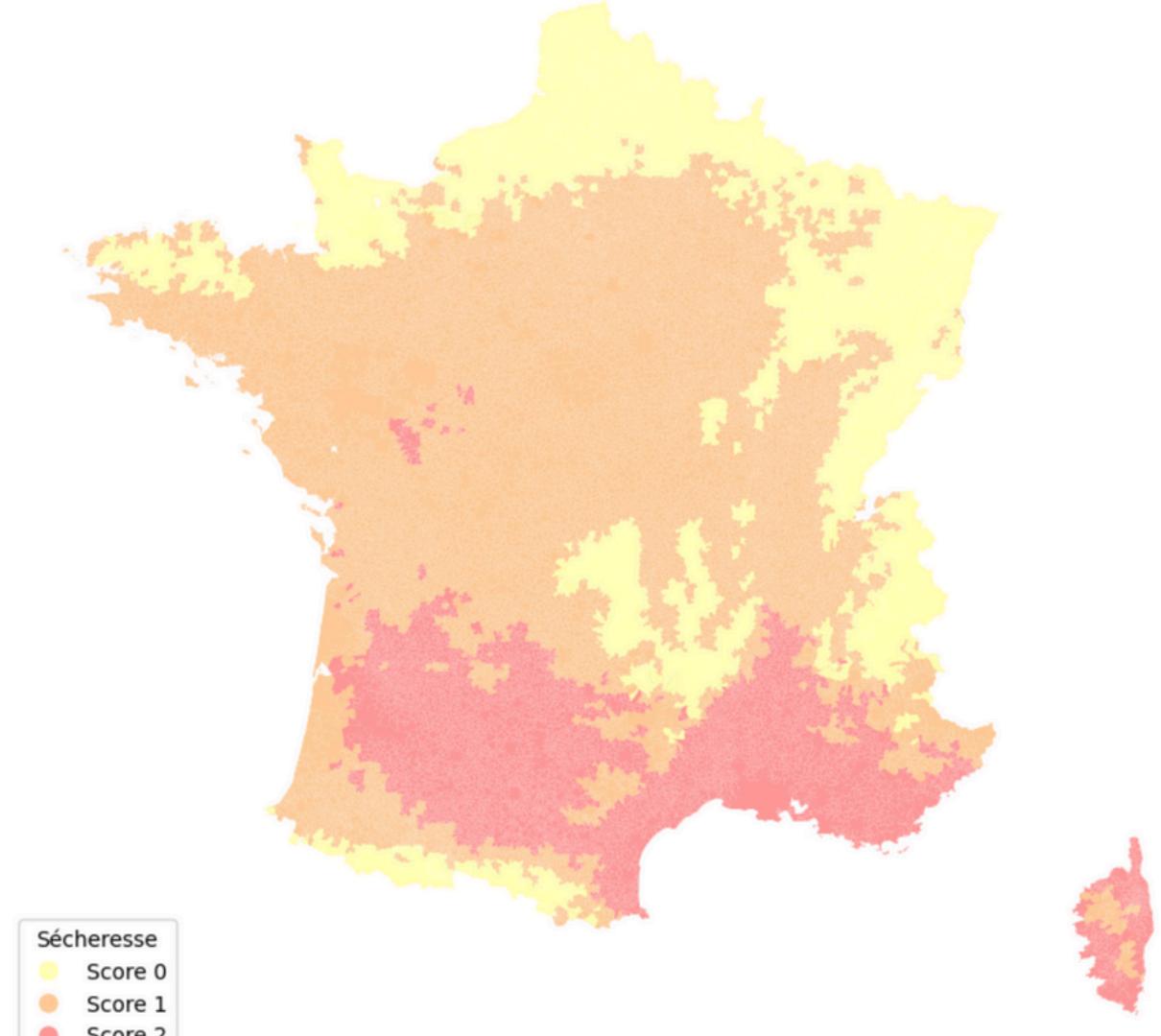
Sécheresse en France (Période H1)



Sécheresse en France (Période H2)



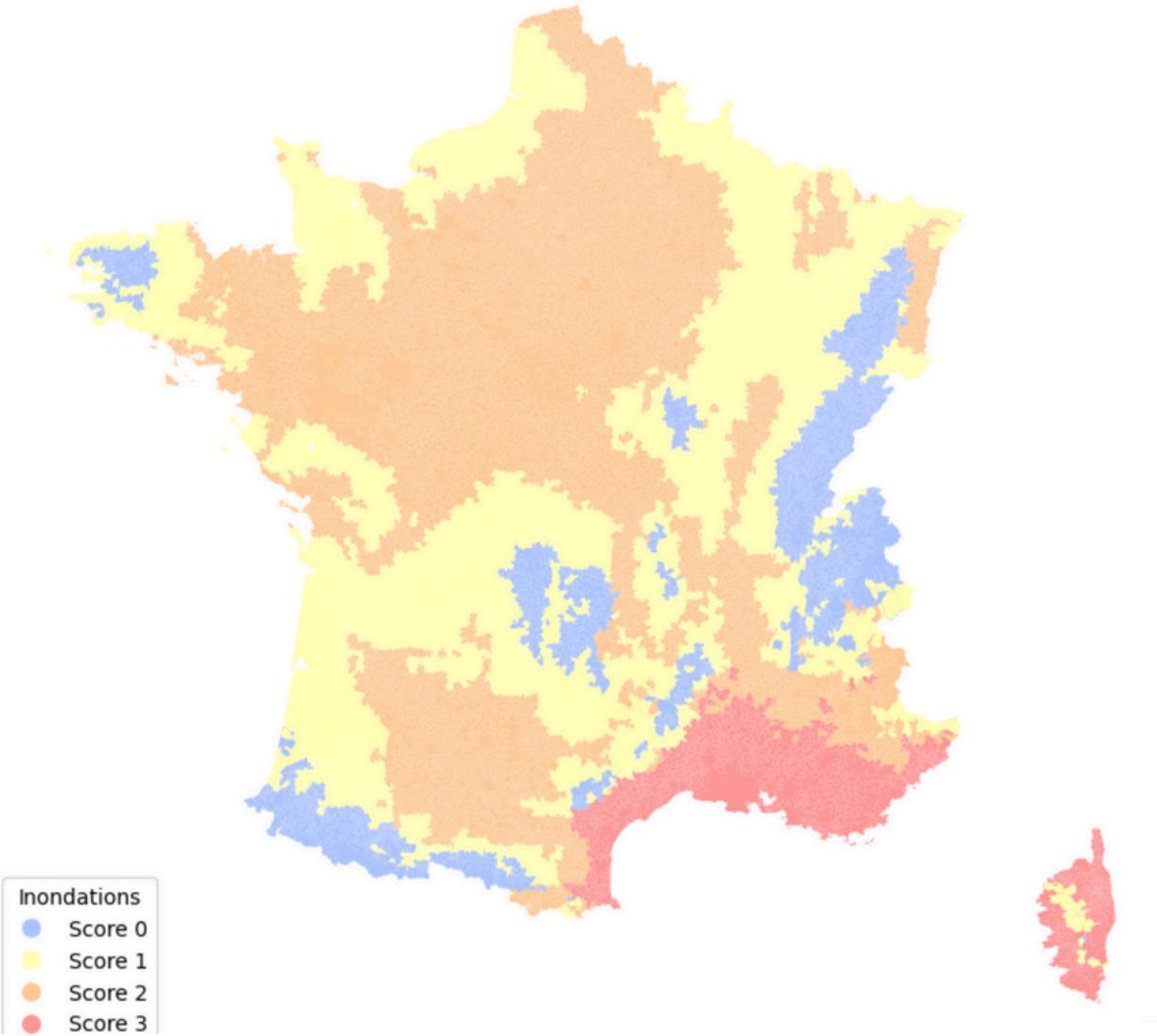
Sécheresse en France (Période H3)



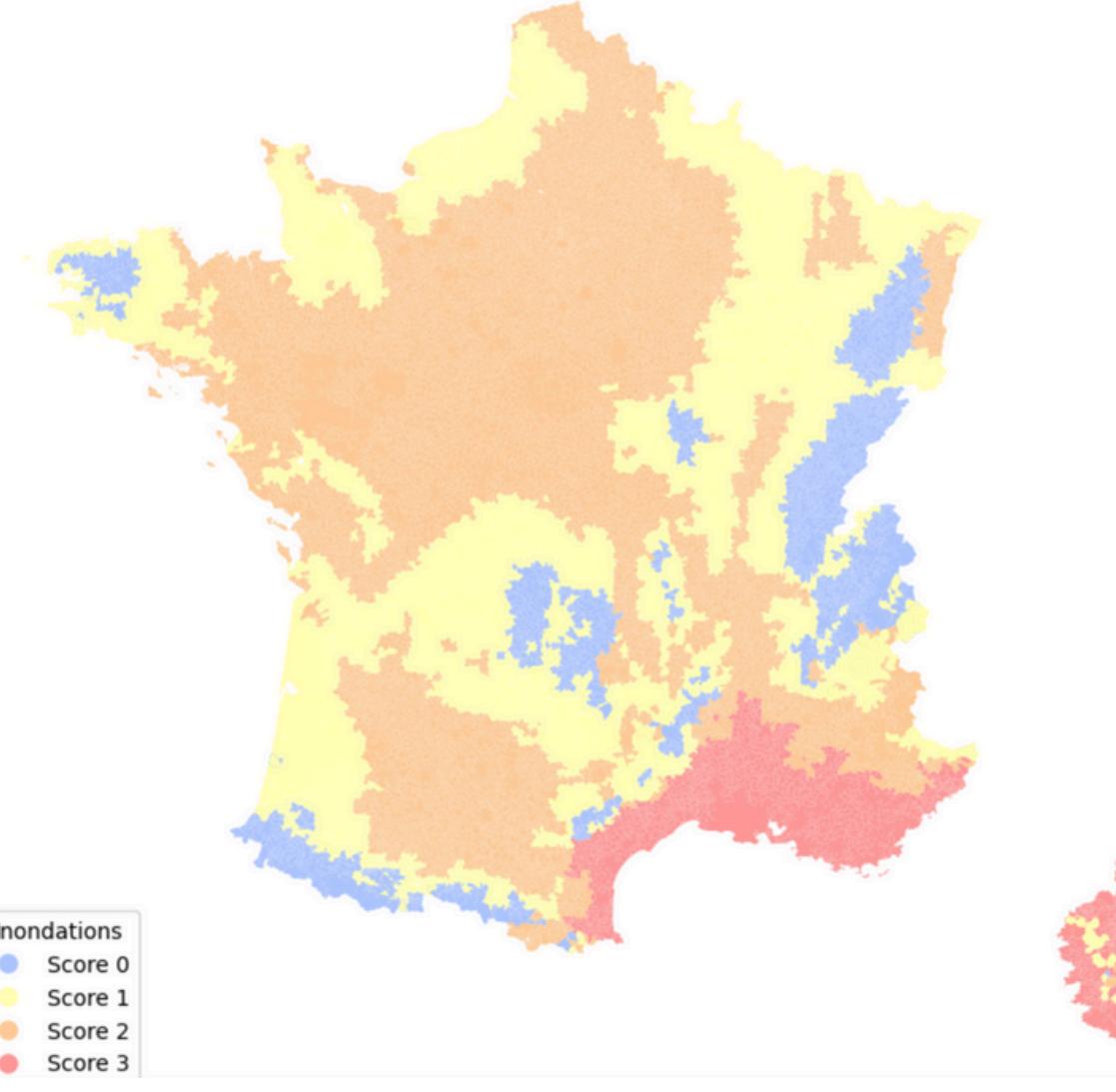
Les cartes montrent une augmentation du risque de sécheresse à travers les périodes, avec un passage de zones à faible risque vers des zones de plus en plus affectées par un risque élevé, en particulier dans le sud et le centre de la France.

# Cartographie - Inondation

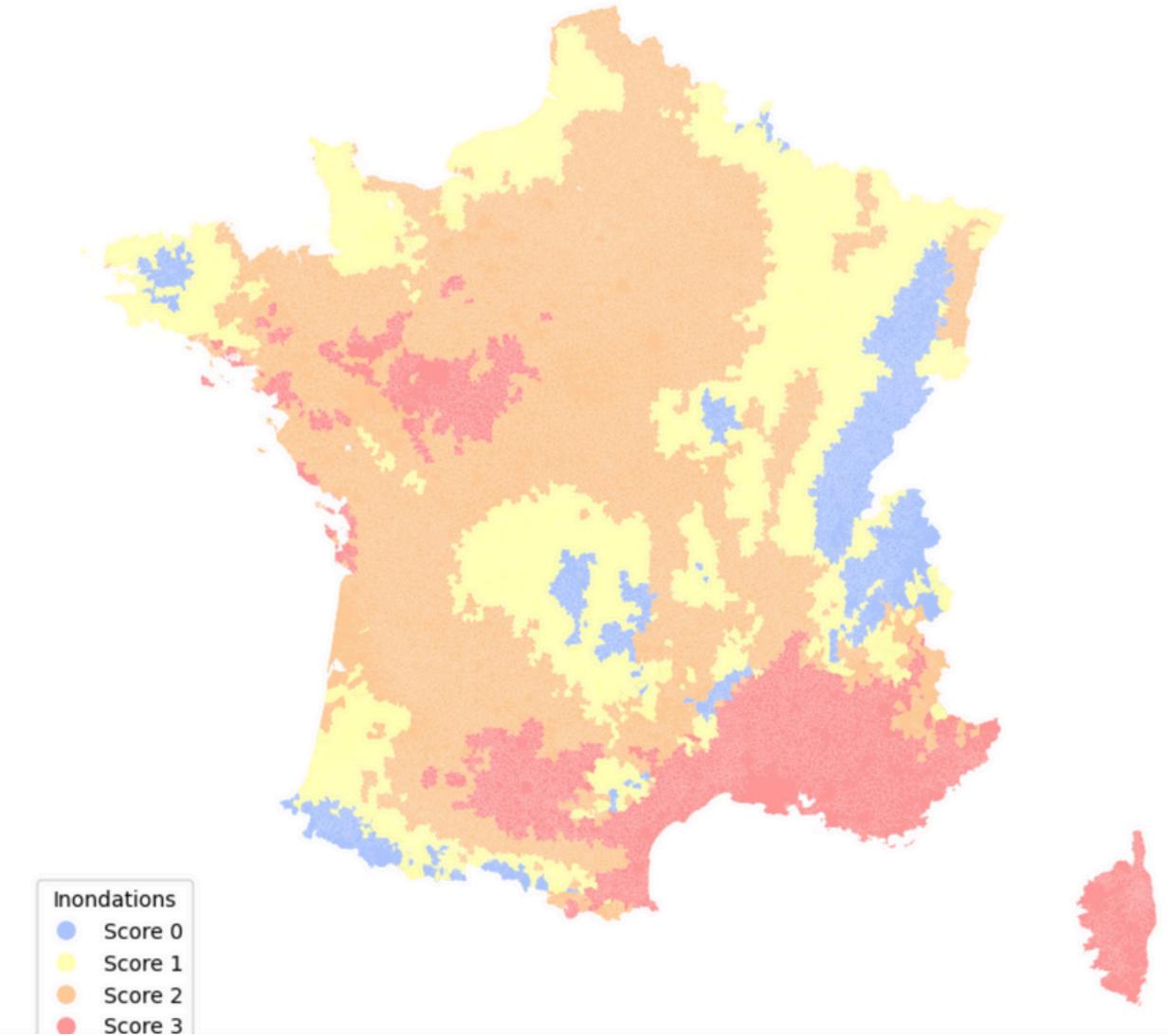
Inondations en France (Période H1)



Inondations en France (Période H2)



Inondations en France (Période H3)



L'évolution sur ces trois périodes révèle une amplification du risque d'inondation dans les zones historiquement sensibles, notamment dans le sud-est et le sud-ouest. Ces tendances mettent en évidence une vulnérabilité croissante, nécessitant des mesures d'adaptation renforcées dans ces régions.

# Conclusion Partie 1

En conclusion, cette première partie a permis de construire une méthodologie robuste pour analyser les **risques climatiques en France métropolitaine**. À travers l'identification des principaux aléas (**feux de forêt, sécheresse, vagues de chaleur et inondations**) et l'utilisation d'indicateurs spécifiques pour chaque phénomène, nous avons produit des scores de risque détaillés par commune. Ces scores, combinés à des techniques de clustering comme **K-Means** et à une visualisation géospatiale avancée via **QGIS**, ont permis de dégager une vision claire et précise des vulnérabilités régionales. Les résultats obtenus montrent une aggravation notable des risques dans certaines zones, notamment dans le sud et le sud-est de la France, appelant à des mesures préventives renforcées.

Désormais nous allons passer à une analyse approfondie du **portefeuille d'actifs fictif** fourni. En intégrant les scores de risque dans une approche financière, nous évaluerons les impacts des aléas climatiques sur les différents secteurs représentés dans le portefeuille. Cette étape consistera à croiser les données de risque avec les caractéristiques des actifs (secteurs d'activité, localisation géographique, maturités) pour estimer leur vulnérabilité globale. L'objectif sera non seulement de quantifier ces impacts, mais également de proposer des stratégies concrètes d'atténuation, adaptées à chaque secteur et à chaque région. L'objectif est d'aider les décideurs à intégrer les enjeux climatiques dans la gestion de leurs portefeuilles et à renforcer leur résilience face à des risques climatiques croissants.

# Introduction Partie 2

Dans le cadre d'un projet de scoring des risques physiques liés au changement climatique, la Caisse des Dépôts (CDC) lance un défi aux participants : évaluer la vulnérabilité d'un portefeuille fictif d'actifs situés en France métropolitaine. L'objectif est de développer une méthode solide pour mesurer et représenter les risques climatiques, en s'appuyant sur des indicateurs pertinents et des outils d'analyse géospatiale.

## Objectifs

Notre étude a trois principaux objectifs :

- Mesurer les risques physiques associés à quatre principaux aléas climatiques.
- Élaborer une stratégie de scoring permettant de quantifier la vulnérabilité des contreparties, même dans le cas de multiples localisations.
- Restituer les résultats sous forme de rapports et d'outils visuels au niveau du portefeuille.

# Présentation des données

Pour cette étude, nous allons nous baser sur trois fichiers CSV du site **DRIAS** (Données climatiques de référence pour l'adaptation au changement climatique) : un contenant les données climatiques détaillées pour les communes françaises, un autre avec les indicateurs spécifiques liés à la fréquence et à l'intensité des feux de forêt, et un dernier avec les indicateurs des précipitations cumulées et des épisodes d'inondations.

Nous utiliserons également deux autres fichiers fournis : le premier contient les délimitations géographiques des communes françaises, essentielles pour visualiser les résultats sur une carte, et le second correspond au portefeuille d'actifs.

# 04

## Scoring des risques du portefeuille

# Présentation du portefeuille

Le portefeuille est constitué de 99 actifs. Pour chaque actif, nous disposons des informations suivantes :

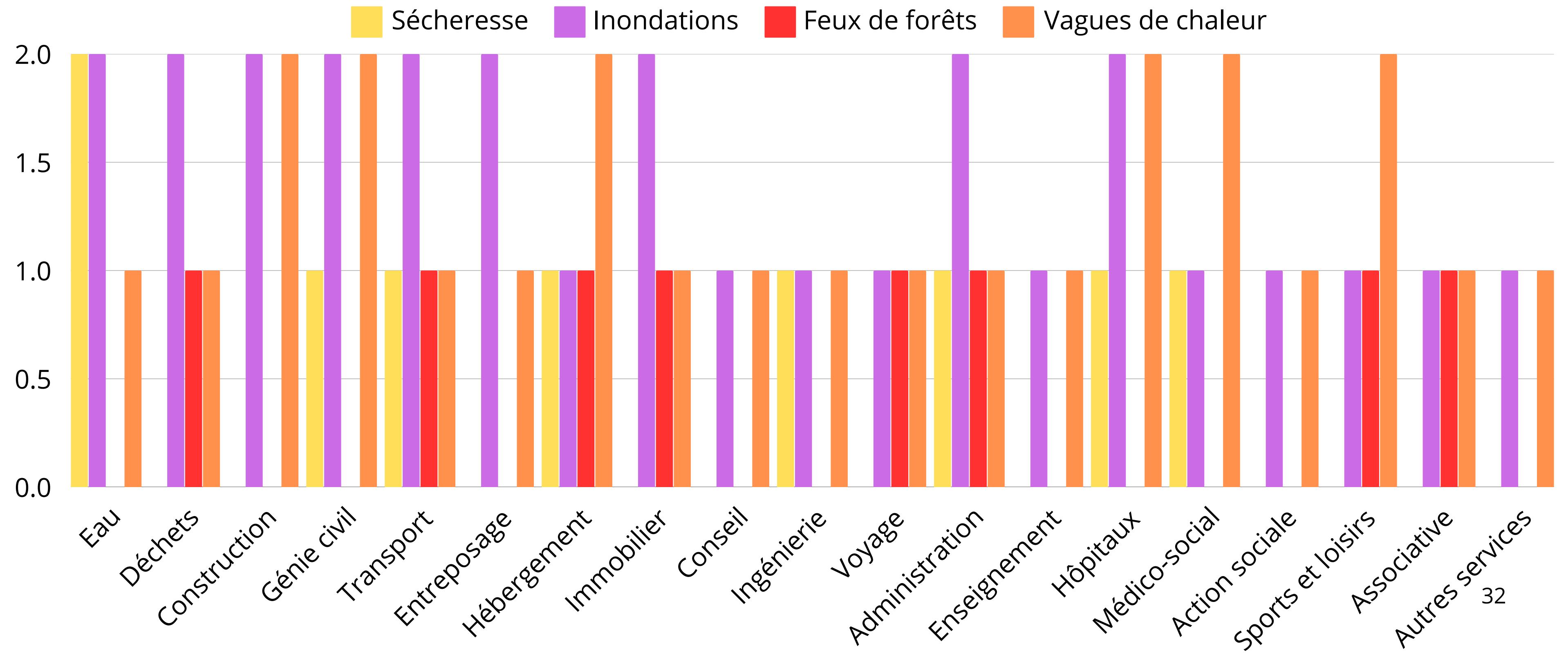
- Identifiant tiers : L'identifiant unique de la contrepartie associée à l'actif.
- Localisation (Commune) : La commune où se situe l'actif.
- Localisation (Code INSEE) : Le code INSEE de la commune.
- Secteur d'activité (code NACE 2) : Le secteur d'activité correspondant à l'actif, basé sur la nomenclature NACE 2.
- Maturité du prêt : L'année d'échéance du prêt associé à l'actif.
- Encours (million EUR) : Le montant restant dû en millions d'euros.

# Prise en compte du secteur d'activité

Pour intégrer le secteur d'activité de la contrepartie dans le scoring de notre portefeuille, des poids spécifiques ont été affectés aux indicateurs en fonction des secteurs. Cette méthode distingue trois niveaux de risque :

- 0 : Absence de risque ou impact négligeable pour ce secteur par rapport à ce type de risque.
- 1 : Exposition modérée à ce risque, caractérisée par un impact limité ou une vulnérabilité moyenne.
- 2 : Forte exposition à ce risque, entraînant un impact significatif ou une grande vulnérabilité.

# Choix des poids par secteur



# Calcul du score total pondéré

Les scores du portefeuille sont également calculés pour les quatre indicateurs, en tenant compte de la maturité du financement (H1, H2 ou H3) et de la localisation de l'actif, identifiée à l'aide du code INSEE. Les plages de scores ont été déterminées à l'aide de la méthode des K-means sauf pour Inondations :

- Sécheresse : de 0 à 2.
- Inondations : de 0 à 3.
- Feux de forêts : de 0 à 3.
- Vagues de chaleur : de 0 à 2.

Pour chaque ligne d'actif, le score total pondéré est calculé comme suit :

- Les scores des 4 indicateurs sont additionnés après avoir été pondérés selon leur poids respectif.
- Le total est ensuite multiplié par l'encours de la ligne, puis divisé par l'encours total du portefeuille.

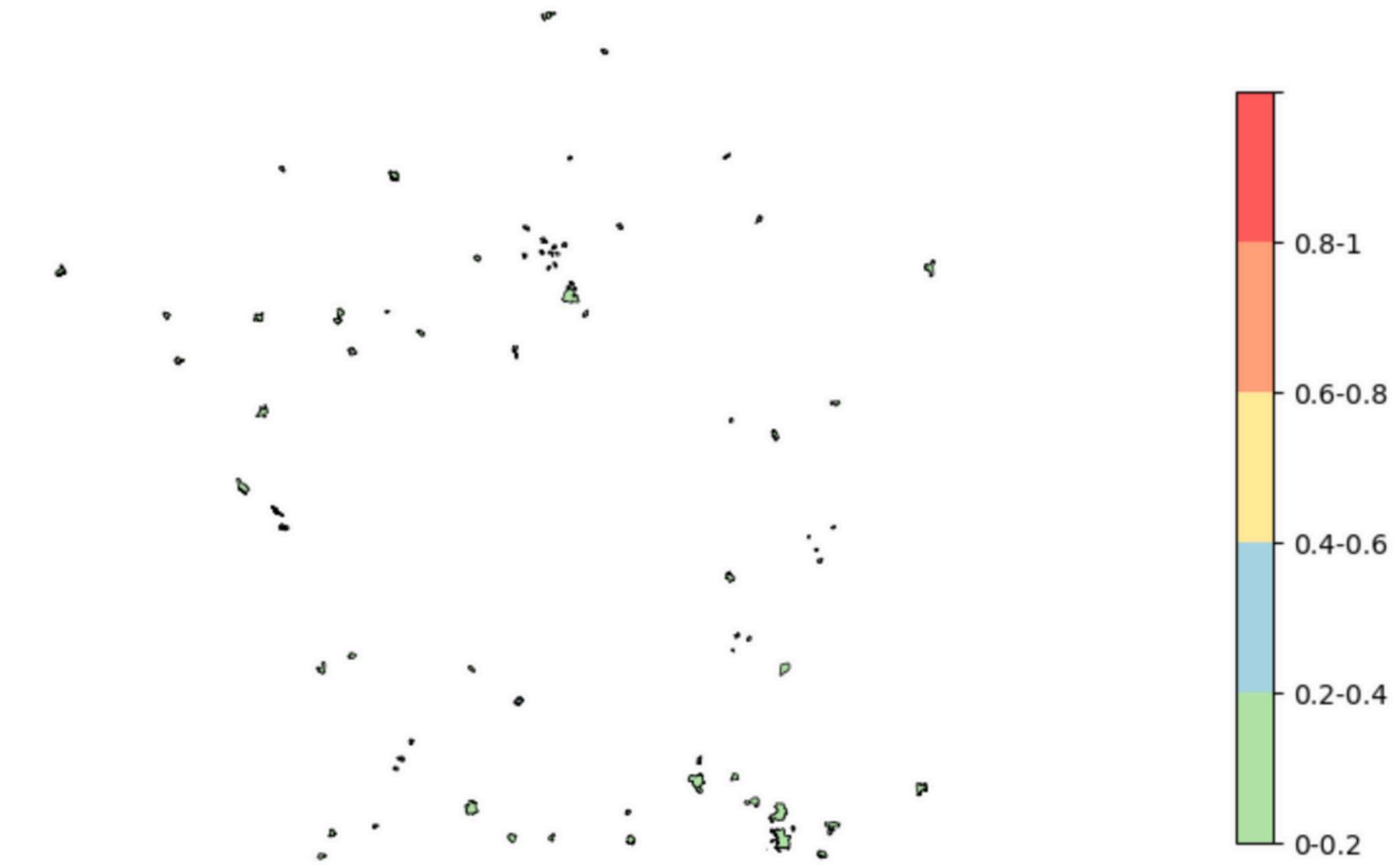
# 05

## Résultats sur le portefeuille

# Risque lié à la localisation

## Carte 1 : Score total de risque par commune

**Carte des scores total commune par commune**



# Risque lié à la localisation

## Carte 2 : Ratio de risque par commune

**Carte des ratios commune par commune**



# Risque lié à la localisation

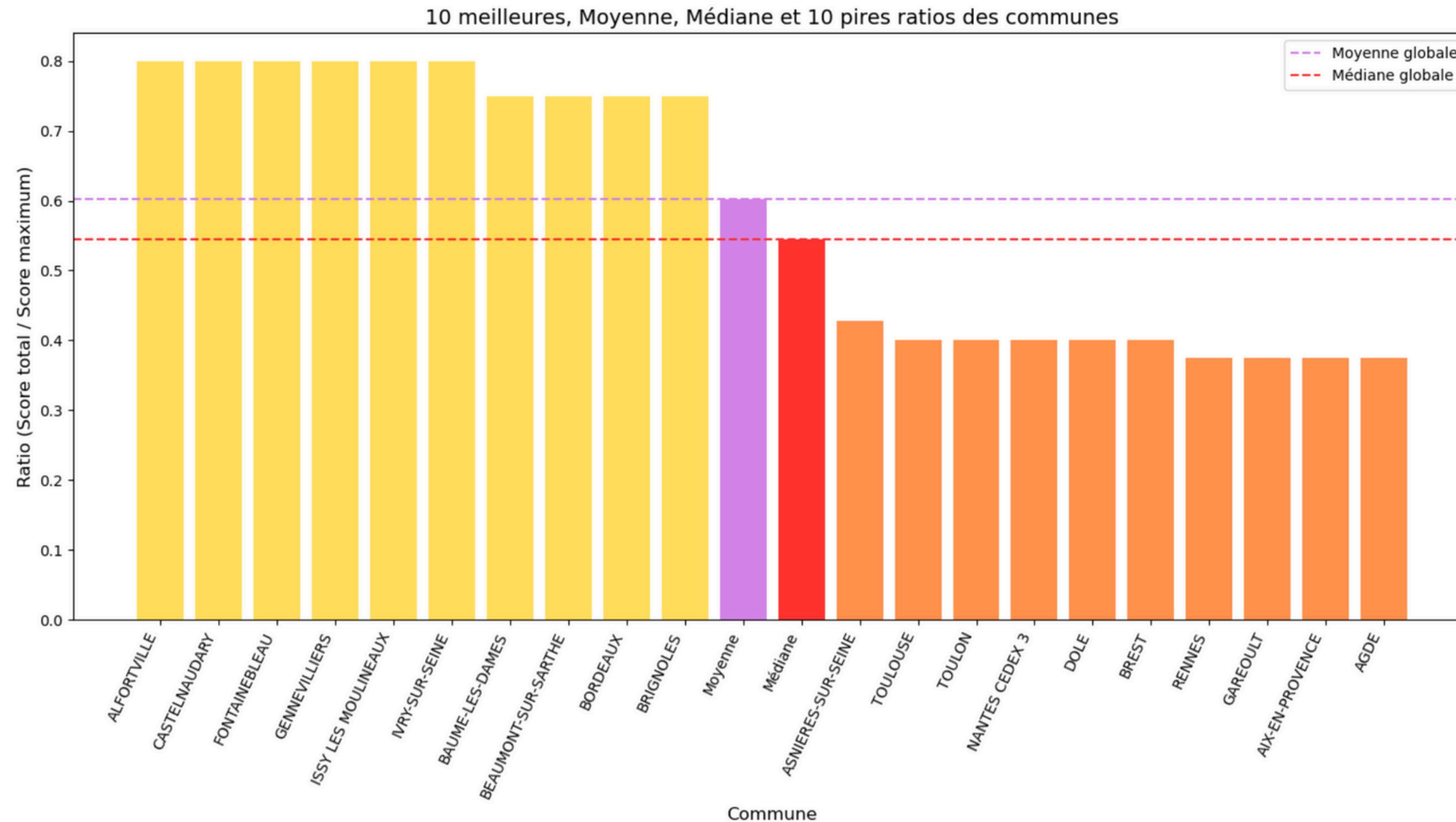
Le score total de risque mesure la somme des risques pour tous les actifs d'une commune, reflétant une concentration absolue de risques.

Cependant, il peut être biaisé par le nombre d'actifs ou leur poids financier dans une commune.

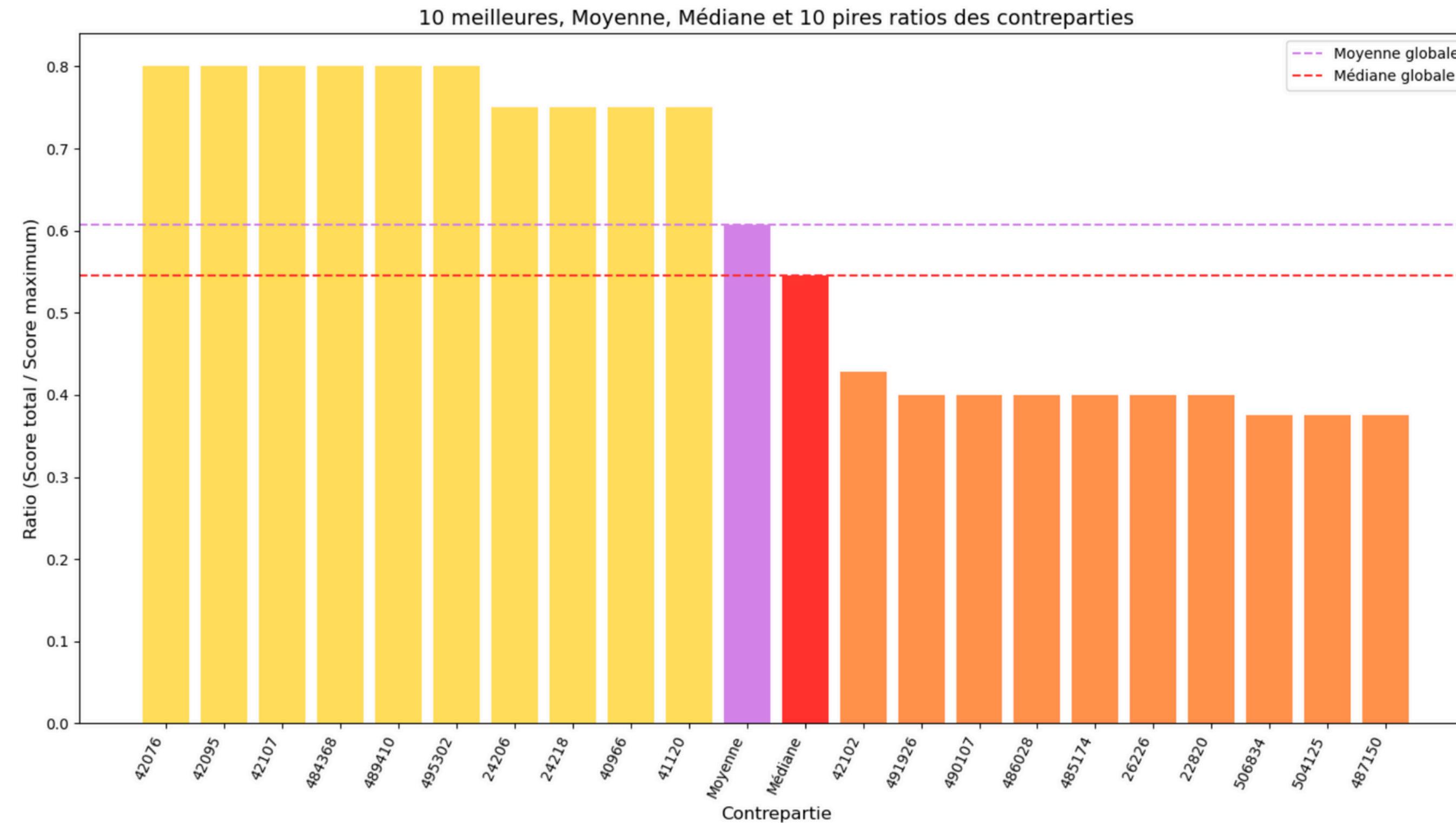
En revanche, le ratio de risque met en perspective le niveau de risque par rapport au maximum possible, offrant une vision relative et standardisée.

Étudier le ratio est donc plus pertinent pour identifier les communes où les actifs sont proportionnellement plus exposés, indépendamment de leur concentration, et permet de cibler efficacement les zones les plus vulnérables.

# Risque lié à la localisation



# Risque lié à la contrepartie

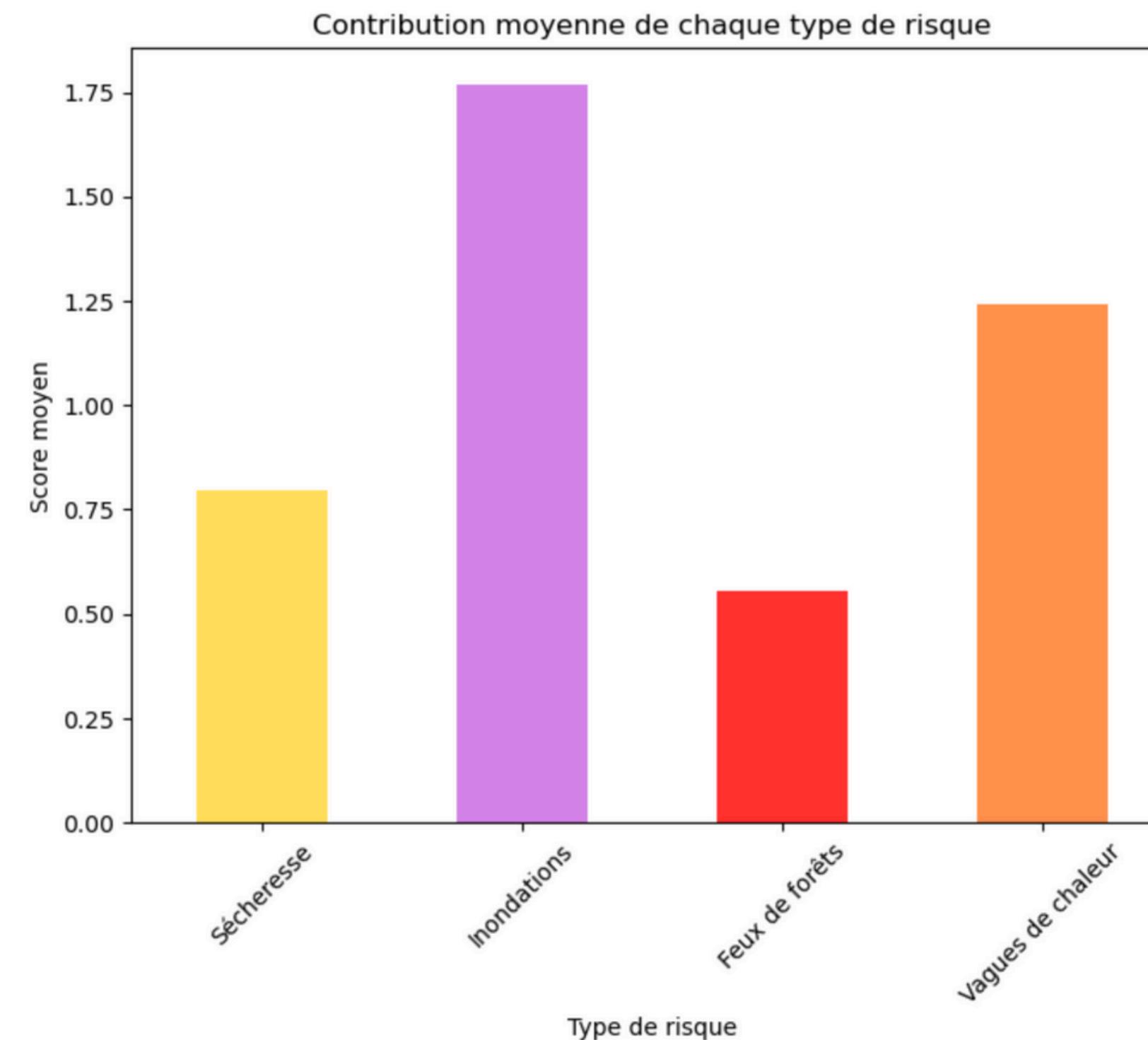


# Risque du Portefeuille

Type de score	Valeur
Score global du portefeuille	6.50471
Score global maximal du portefeuille	10.9908

# Répartition du risque au sein du portefeuille

## Contributions des aléas



# Répartition du risque au sein du portefeuille

## Contributions des aléas

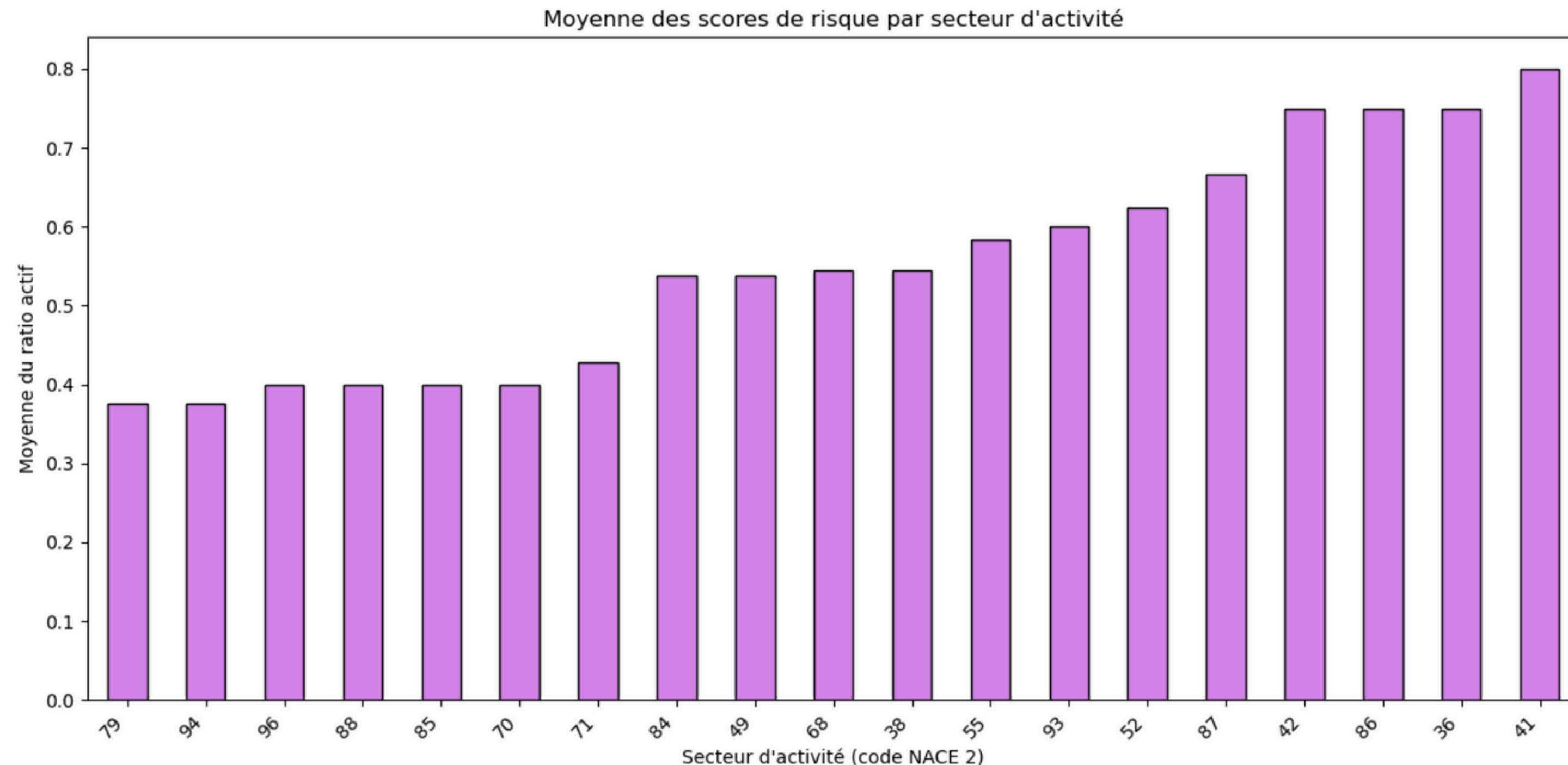
Le précédent graphique montre la contribution moyenne de chaque type de risque au sein du portefeuille.

Le risque lié aux inondations se distingue avec la contribution moyenne la plus élevée, indiquant un impact significatif sur l'ensemble des actifs analysés. À l'inverse, les feux de forêts ont une contribution moyenne relativement faible, ce qui pourrait indiquer un risque plus localisé ou un impact moins généralisé. La sécheresse et les vagues de chaleur se situent à des niveaux intermédiaires, soulignant qu'ils ont également un rôle dans l'exposition globale au risque.

Ces résultats mettent en évidence l'importance de prioriser la gestion des risques liés aux inondations tout en maintenant une attention ciblée sur les autres types de risques.

# Répartition du risque au sein du portefeuille

## Analyse des scores par secteur d'activité



# Répartition du risque au sein du portefeuille

## Analyse des scores par secteur d'activité

1) Analyse des valeurs faibles :

- Les secteurs ayant les plus faibles ratios de risque incluent :
  - 79 (Activités des agences de voyage) et 94 (Activités des organisations associatives).
  - Cela suggère que ces secteurs présentent des niveaux de risque relativement bas, peut-être en raison de leur faible exposition aux grandes pertes d'actifs ou de leur moindre dépendance à des facteurs volatils comme les conditions économiques.

2) Analyse des valeurs moyennes :

- Les secteurs tels que 88 (Action sociale sans hébergement), 85 (Enseignement) et 70 (Activités des sièges sociaux) affichent des ratios de risque modérés.
- Ces secteurs peuvent être considérés comme stables, mais avec une exposition légèrement plus élevée au risque par rapport aux secteurs moins risqués.

3) Analyse des valeurs élevées :

- Les secteurs comme 41 (Construction de bâtiments), 36 (Captage, traitement et distribution d'eau) et 86 (Activités hospitalières) se trouvent parmi ceux ayant les ratios de risque les plus élevés.
- Cela peut être expliqué par leur dépendance à des investissements en capital élevés, leur vulnérabilité aux aléas climatiques, ou encore leur exposition à des réglementations strictes qui peuvent compliquer la gestion des actifs.

# Conclusion Partie 2

- Les inondations sont le risque dominant, nécessitant une gestion prioritaire, tandis que les feux de forêts ont un impact limité et localisé. La sécheresse et les vagues de chaleur ont une contribution intermédiaire.
- Les secteurs comme 36 (Captage, traitement et distribution d'eau), 41 (Construction de bâtiments) par exemple, avec des scores élevés sont particulièrement vulnérables en raison de leur concentration géographique dans des zones à haut risque climatique.
- À l'inverse, des secteurs tels que 79 (Activités des agences de voyage) présentent une faible exposition, traduisant un moindre risque climatique.
- Une gestion des risques efficace doit combiner priorisation des secteurs vulnérables, analyse géographique des actifs, et diversification pour renforcer la résilience du portefeuille.