

Projet 8 : L'obésité dans le monde

Projet personnel



Sommaire

Introduction

Analyse exploratoire

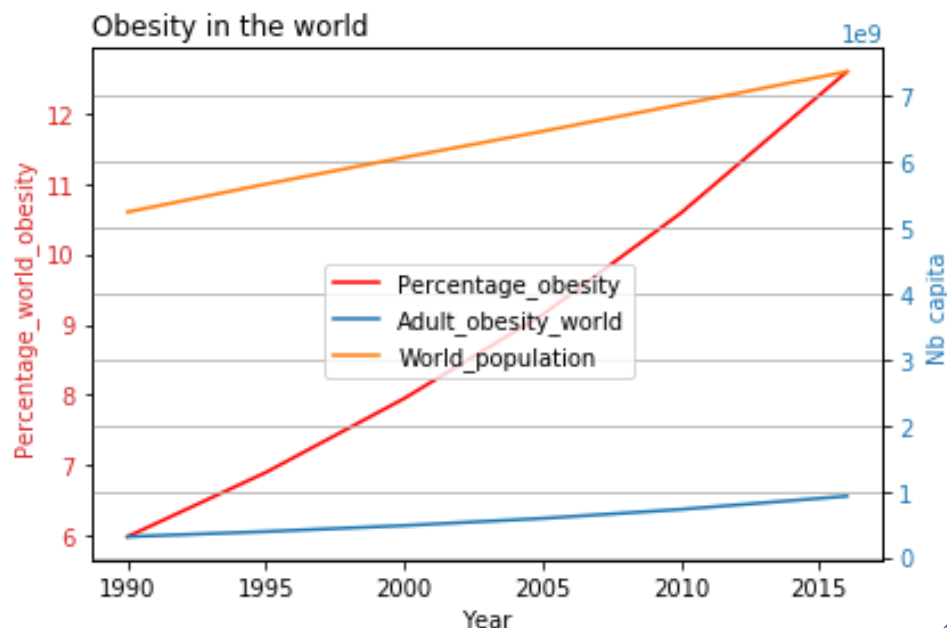
Optimisation

La réalité des variables

Conclusion

INTRODUCTION

Origines du projet



Obésité mondiale
multipliée par 2
en 25 ans

En 2016 près d'un
milliard de personnes
obèses

Projection de l'OMS
en 2045
Obésité mondiale = 25 %

Sélection des variables

Alimentation



- Taux de graisse dans l'alimentation
- Part alimentation d'origine animale
- Consommation d'alcool
- Sous-alimentation
- Densité fast-food

Démographie



- Population
- PIB/hab
- Continent
- Manque de sport adultes et enfants
- Urbanisation
- Taux d'obésité adultes et enfants
- Pauvreté

Santé



- Consommation de produits psychotropes

Table de traitement

Variables

Fusion



Doublons importants :
Plusieurs appellations pour un même pays

Algorithme fuzzy
matching



Manque de précision

Solution
Alternative



Fichier de transition
manuel (Excel)

Maximisation nombre
de pays et variables

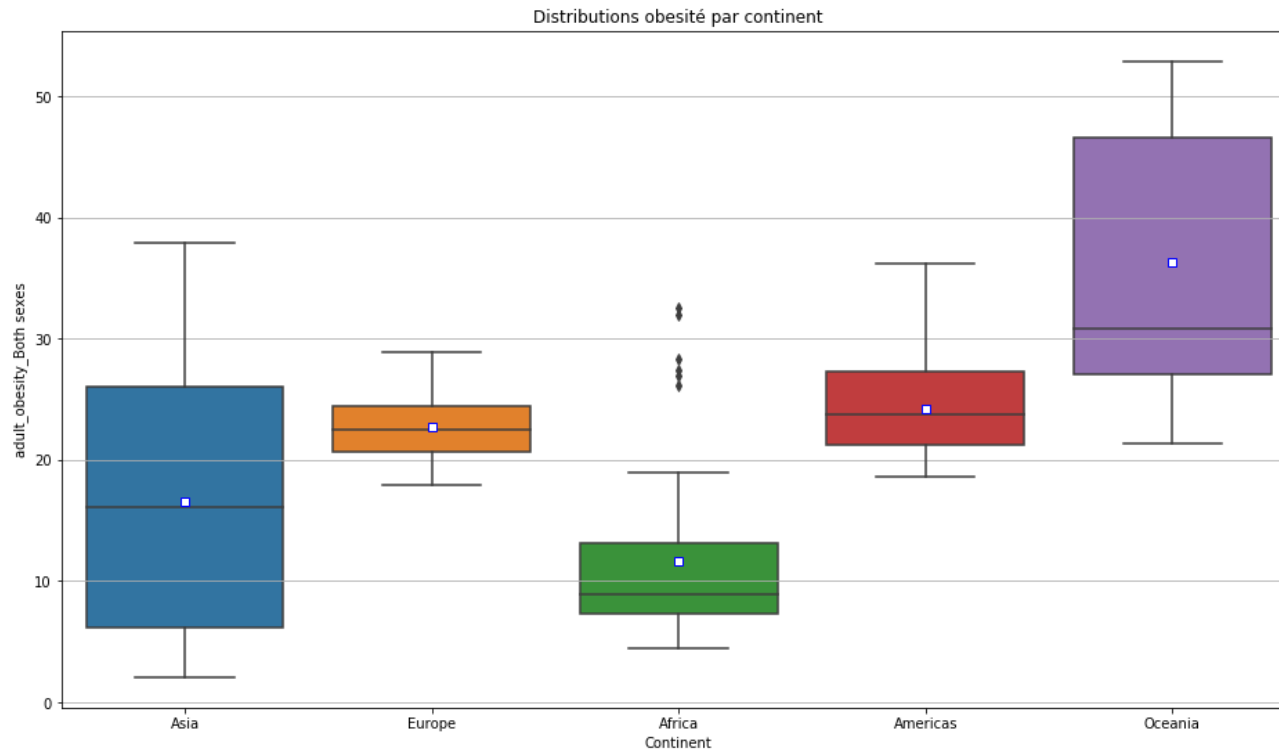


Table de traitement
Année 2016
98 pays
Représentativité 70 %

ANALYSE EXPLORATOIRE

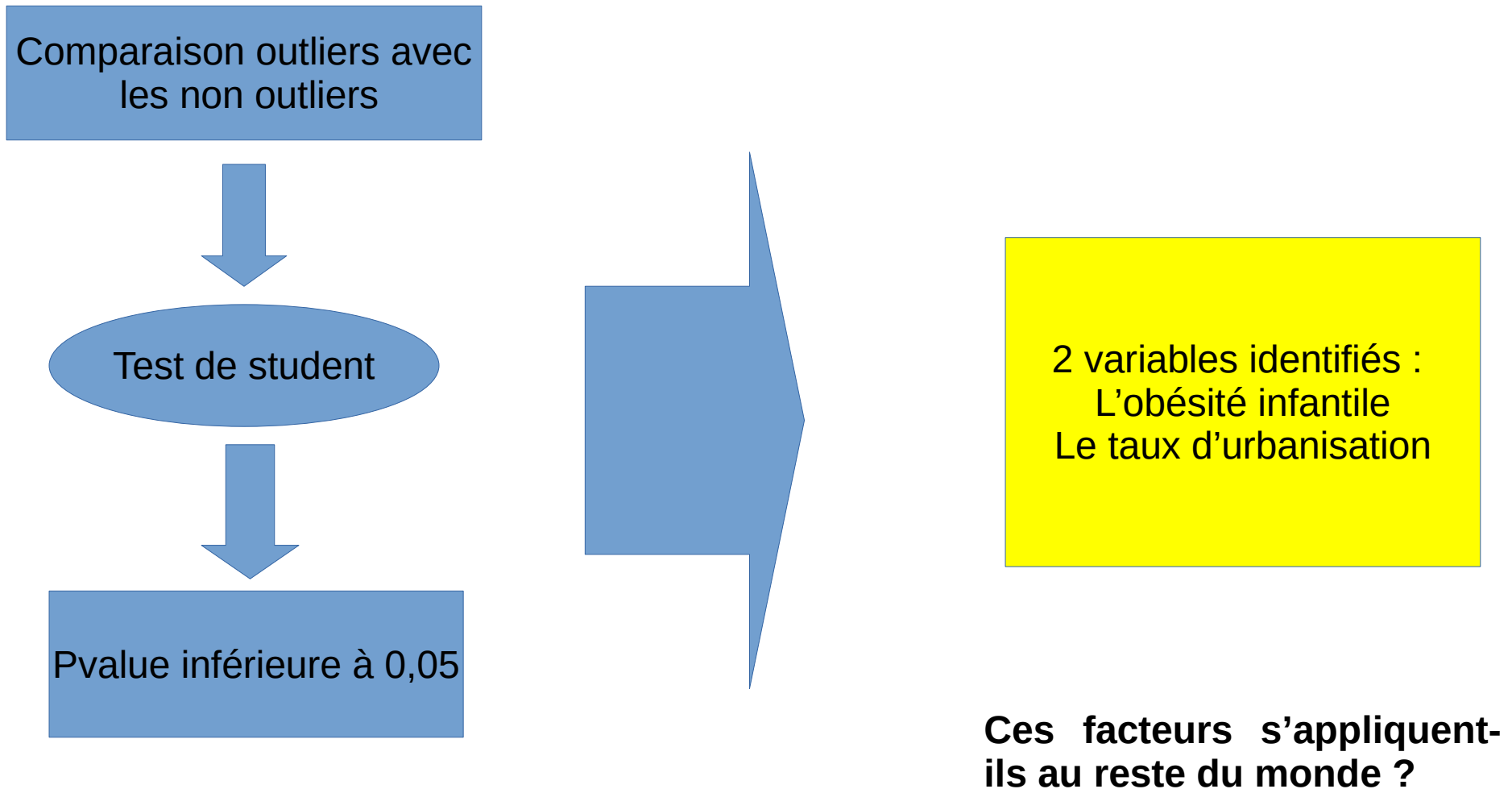
Première approche des données

Peut-on classer l'obésité par continents ?



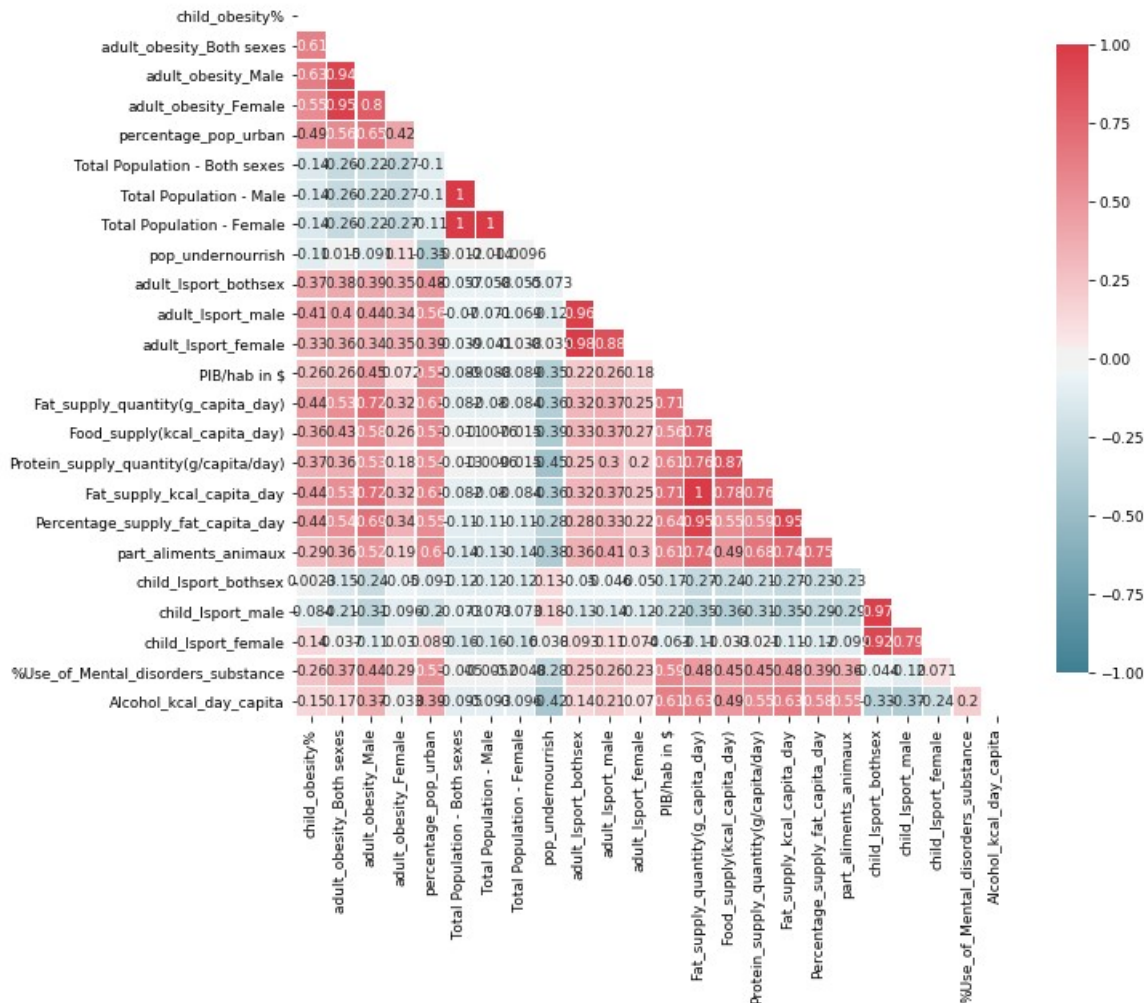
L'Asie présente une plage d'obésité qui couvre tous les continents, sauf l'Océanie. En Afrique, nous identifions plusieurs pays outliers. Nous allons étudier ces outliers en détail.

L'analyse des outliers



Matrice des corrélations

Matrice des corrélations



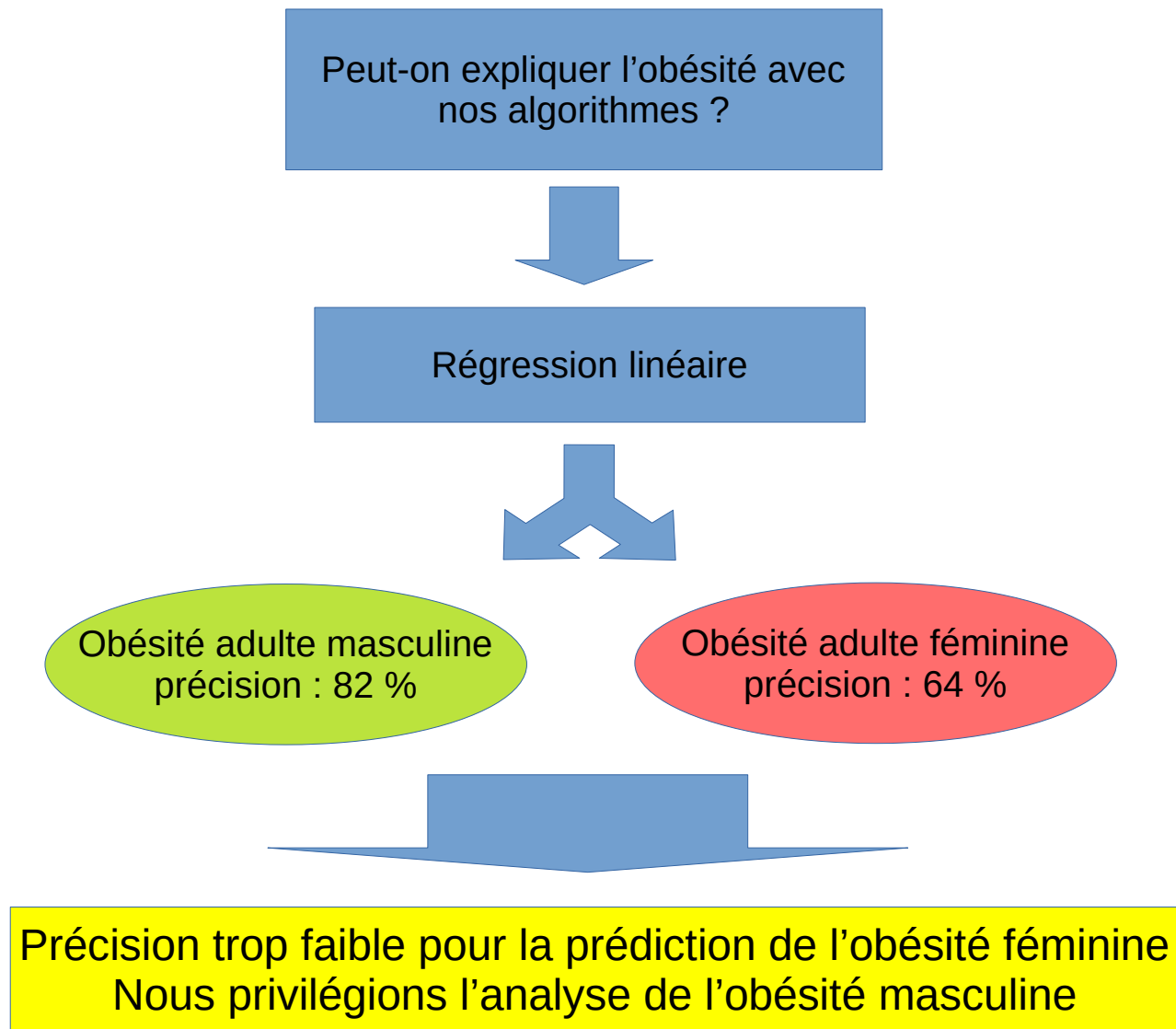
Plusieurs variables présentant des corrélations notables avec l'obésité :

- Child_obesity (présente outliers)
- percentage_pop_urban (présente outliers)
- percentage_supply_fat_capita_day
- Part_aliment_animaux
- adult_lsport_male
- %Use_of_mental_disorders_substance

Cependant, ces corrélations sont surtout pour l'obésité masculine. Les corrélations sont nettement plus faible avec l'obésité féminine.

Peut-on malgré cela prédire l'obésité féminine uniquement avec ces variables ?

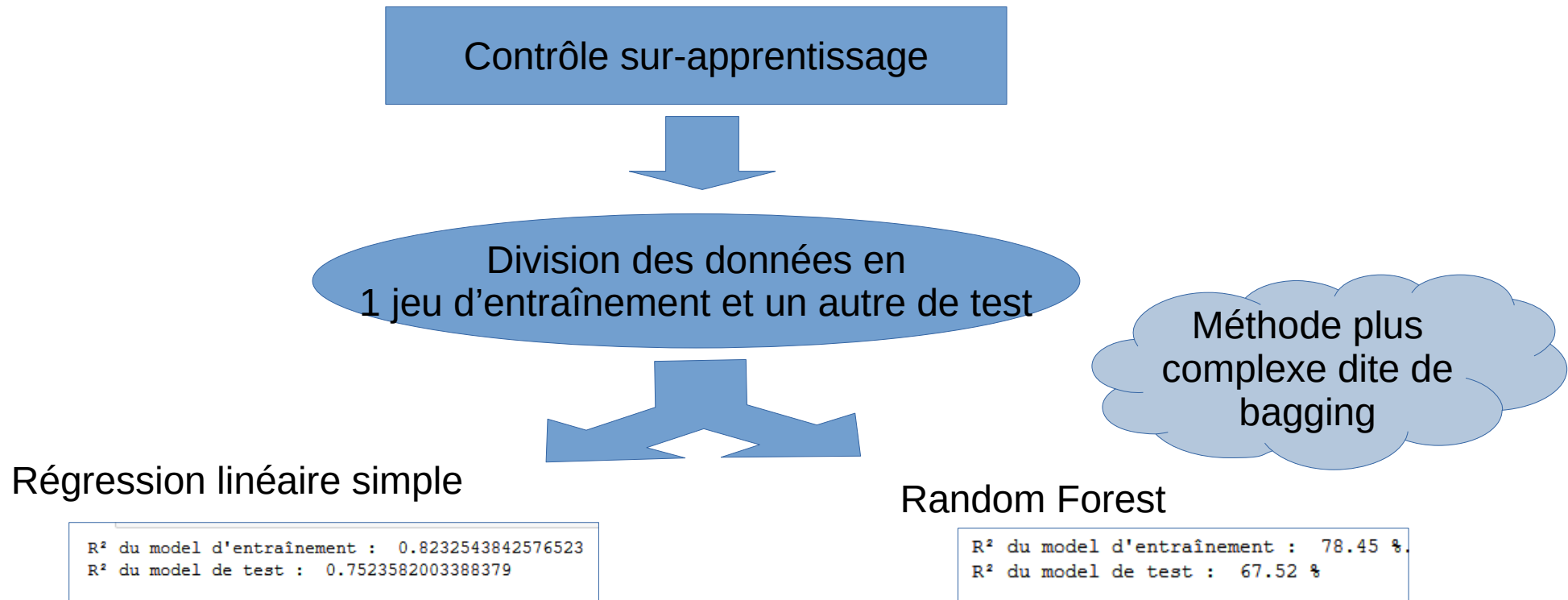
Régression linéaire



OPTIMISATION

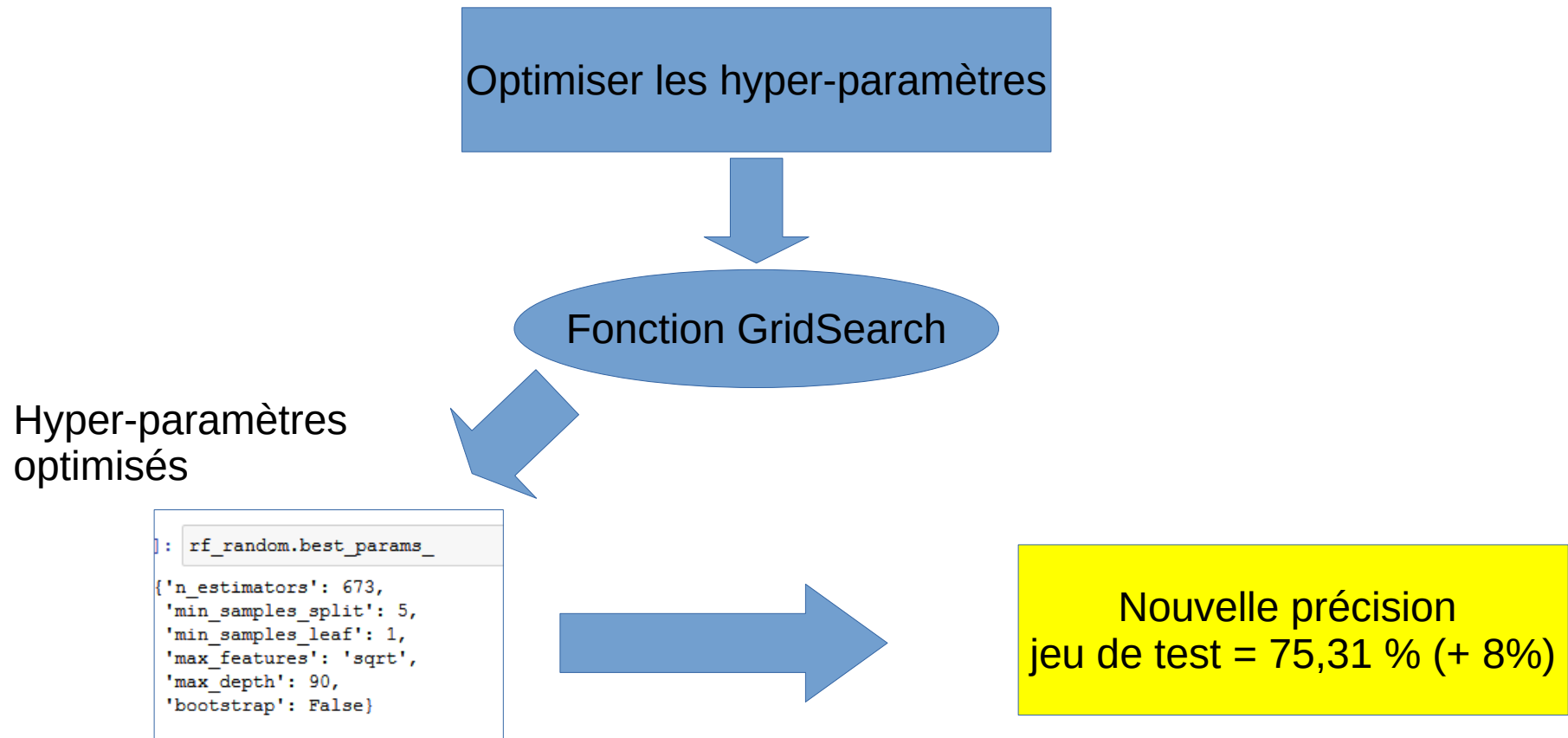
Vérification des modèles

Nous analysons la pertinence de notre modèle et nous le comparons à un modèle plus complexe = Random Forest.



Les résultats du Random Forest sont inférieurs à la régression.
Étant complexe il demande une optimisation des hyper-paramètres

Optimisation des hyper-paramètres



Analyse des poids des variables

Régression simple – importance des variable ordre décroissant :

Taux urbanisation
Continent européen
Obésité infantile
Manque de sport chez les enfants mâle
Taux de graisse dans l'alimentation

Random forest – importance des variable ordre décroissant :

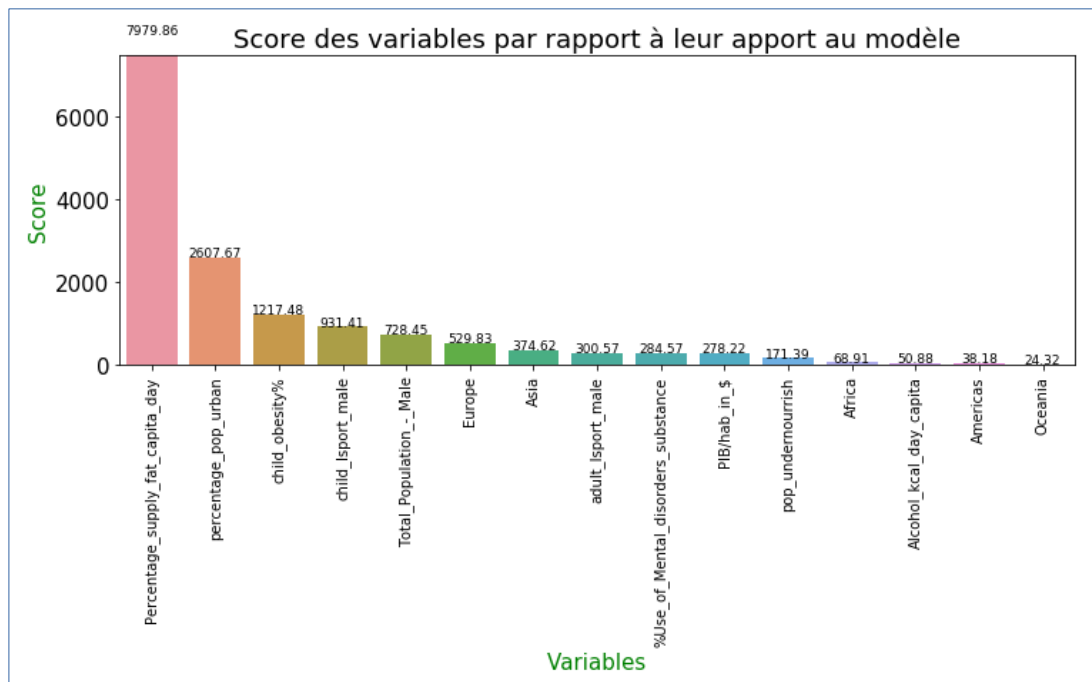
Taux de graisse dans l'alimentation
Taux d'urbanisation
Obésité infantile
PIB par habitant
Total population masculine

3 variables sont présentent dans les deux algorithmes :
Taux de graisse dans l'alimentation
Obésité infantile
Taux d'urbanisation

**Nous allons entraîner un dernier type de modèle pour confirmer
nos observations = modèle Xgboost.**

L'algorithme du XGBOOST

Algorithme qui utilise la méthode de descente du gradient pour diminuer l'erreur du modèle et ainsi optimiser la précision



Précision

Jeu d'entraînement : 91,08 %

Jeu de test : 76,34 %
(+1 % Random forest)

Confirmation de nos observations, les 3 variables sont aussi les meilleurs du XGBOOST

LA RÉALITÉ DES VARIABLES

Classification déterminations nombre de clusters

Nous avons identifié les 3 des facteurs de l'obésité adulte masculine.

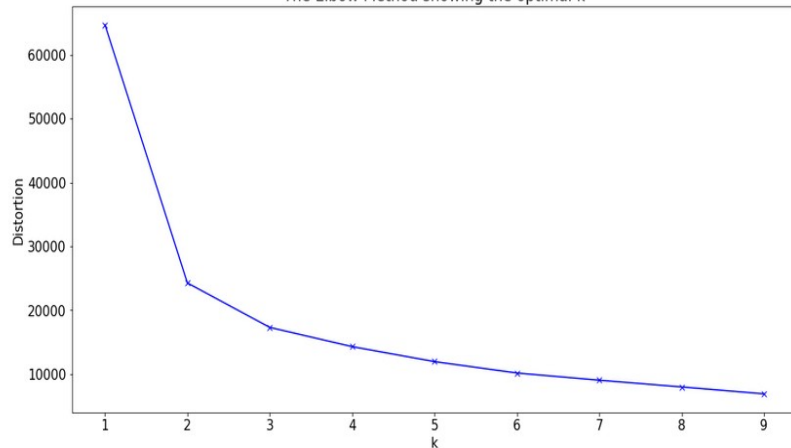
L'objectif est donc désormais de créer une nouvelle carte du monde de l'obésité en fonction de ces variables.

Détermination
nombre clusters

Méthode
du coude

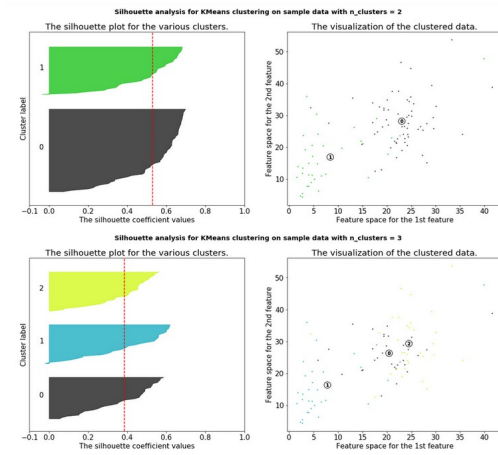
Indice
silhouette

The Elbow Method showing the optimal k



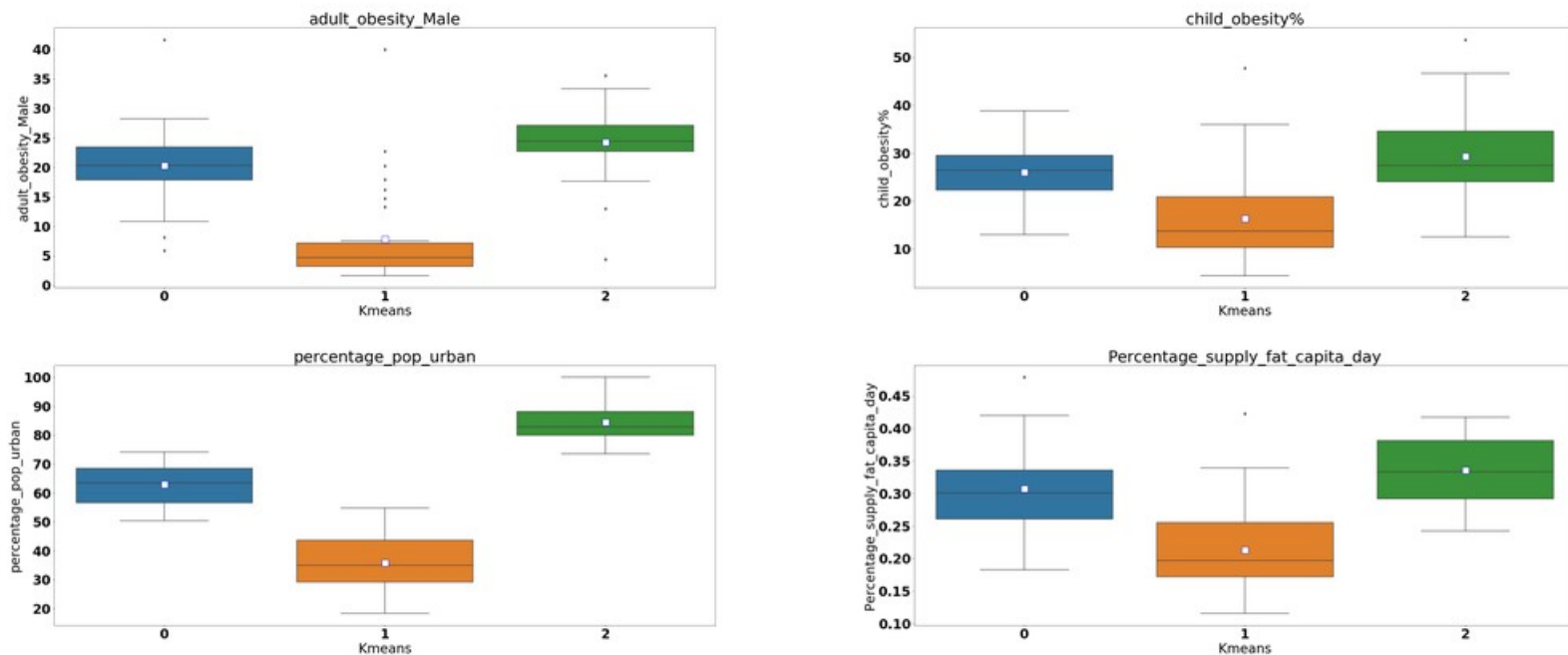
Optimal 2 clusters
Opérationnel 3 clusters

```
Automatically created module for IPython interactive environment
For n_clusters = 2 The average silhouette_score is : 0.5315020114819231
For n_clusters = 3 The average silhouette_score is : 0.3869828339956205
```



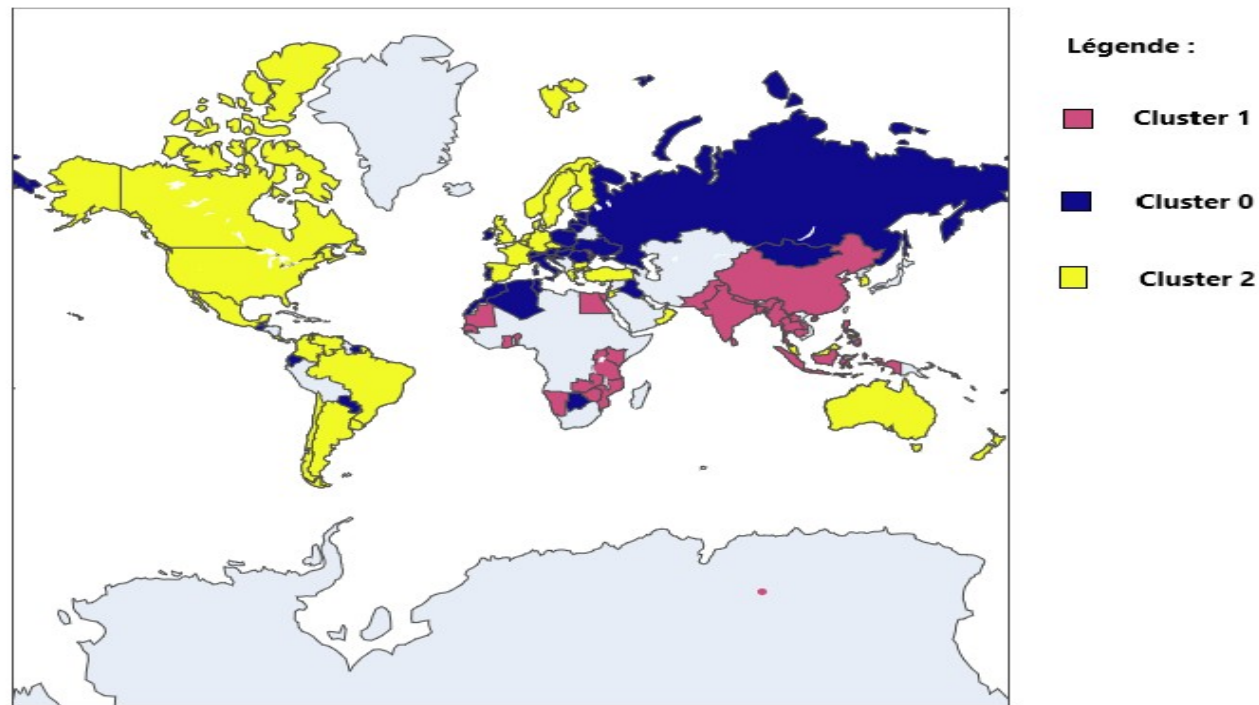
Kmeans

Opérationnellement, nous décidons d'utiliser 3 clusters



La variance inter-classes est bien optimisée, même si outliers sur cluster 1
La variable qui différencie le mieux les clusters est « percentage_pop_urban »

Nouvelle carte de l'obésité



Le cluster 1 rassemble l'Asie et l'Afrique, correspond bien aux continents avec faible taux obésité. Et à l'inverse le cluster 2 concentre bien les continents avec les taux d'obésité les plus important.

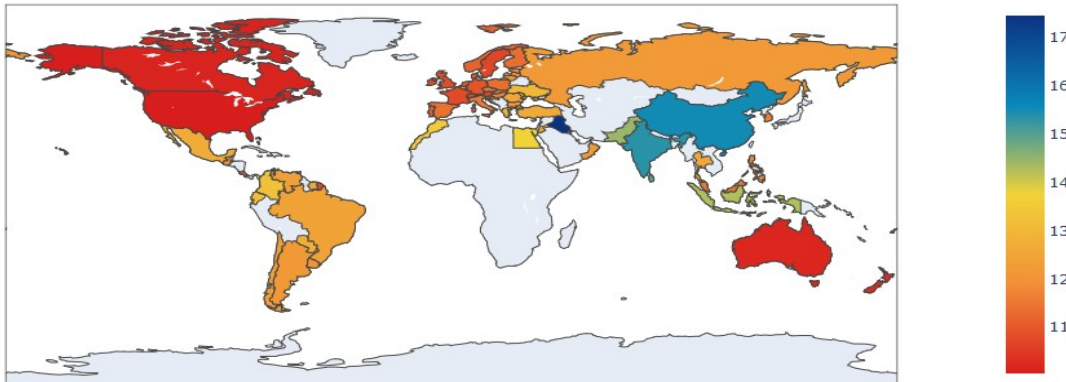
La carte caractérise mieux les pays par rapport à l'obésité masculine.

Et si on analysait plus en détails nos 3 variables ?

Le taux de graisse

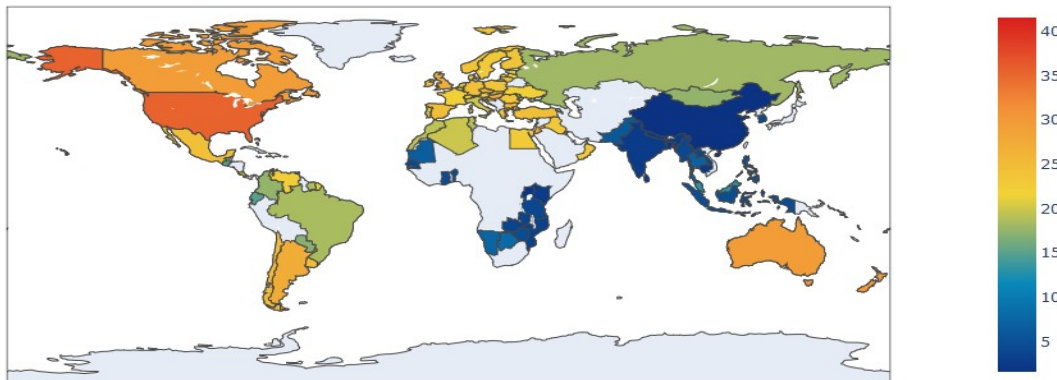
Pour analyser le taux de graisse, nous décidons d'intégrer une nouvelle variable :
Densité des fast food par pays

Nombre d'habitant par restaurant Mac Donald



Caractérisation du taux de graisse
par rapport à la densité des
restaurants Mac Donald

Taux d'obésité masculine dans le monde





On remarque une correspondance
entre taux d'obésité et densité
d'enseigne Mac Donald

Les aliments gras une cible de l'obésité des instances étatiques

D'après le CDC américain, l'alimentation est un problème de comportement qui est une des causes de l'obésité

Behavior

Healthy behaviors include regular physical activity and healthy eating. Balancing the number of calories consumed from foods and beverages with the number of calories the body uses for activity plays a role in preventing excess weight gain.¹ The [Physical Activity Guidelines for Americans](#)  recommends adults do at least 150 minutes of moderate intensity activity or 75 minutes of vigorous intensity activity, or a combination of both, along with 2 days of strength training per week.

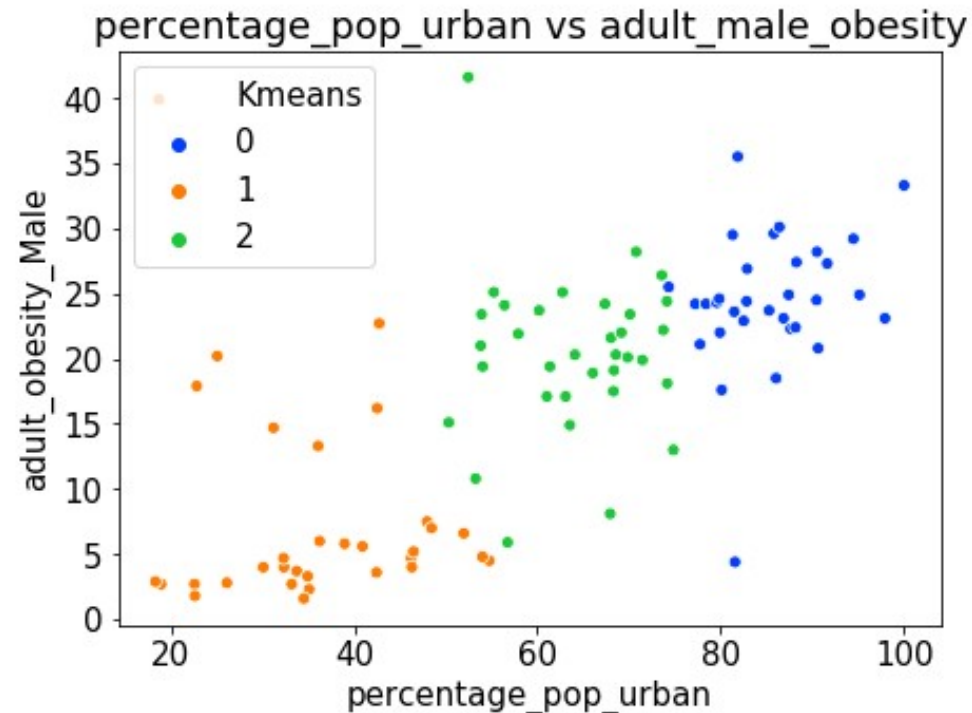
A healthy diet pattern follows the [Dietary Guidelines for Americans](#)  which emphasizes eating whole grains, fruits, vegetables, lean protein, low-fat and fat-free dairy products, and drinking water.

L'important est d'équilibrer sa consommation calorique par rapport à l'intensité de nos activités quotidiennes.

Mais la calorie est dépendante de la physiologie des individus, donc on lui préfère le taux de graisse de l'alimentation plus fiable.

La recommandation du CDC démontre la pertinence de notre variable

Taux d'urbanisation



On confronte en ordonnée notre taux d'obésité à notre taux d'urbanisation en abscisse

Plus le taux d'urbanisation est élevé, plus le taux d'obésité est élevé.

Urbanisation, facteur de progrès mais aussi d'obésité

This publication focuses on physical activity and how it can be supported through urban planning. The focus on physical activity is explained by the fact that inactivity today accounts for an increasing proportion of deaths and disability worldwide and is associated with significant health care costs and productivity losses.² Action to increase rates of physical activity will be necessary to achieve global targets on the prevention of premature mortality from noncommunicable diseases – the leading cause of death worldwide – and to halt the rise in obesity. With more than 80% of the European population expected to live in urban areas by 2030, cities play a pivotal role in promoting and protecting health and well-being.³ As cities continue to expand in population, there is a growing need to develop ways of supporting physical activity in dense urban settings.

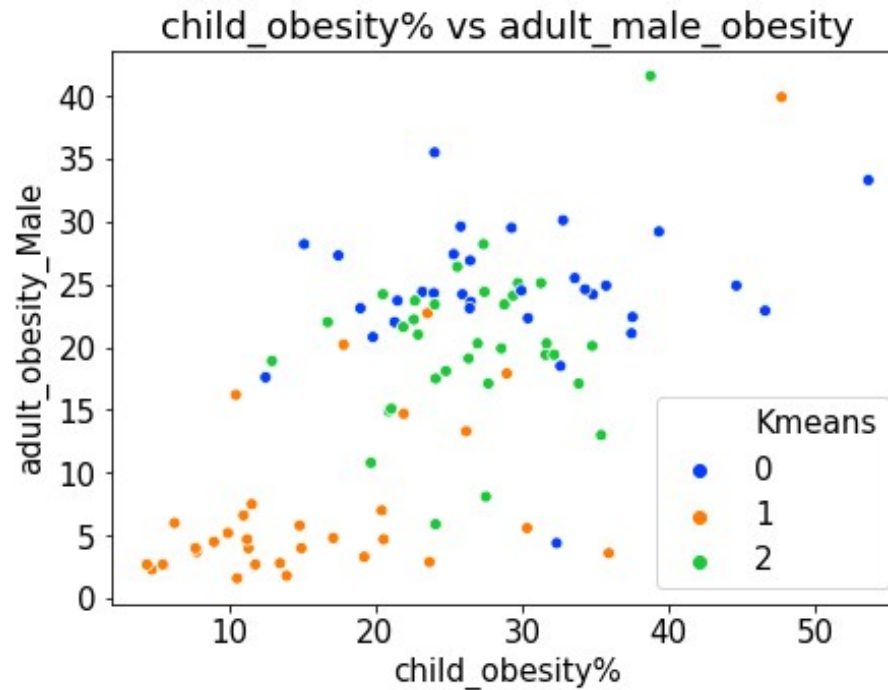
Ici une étude sur la ville de Copenhague où l'activité physique a réinvestie la vie urbain et diminué l'obésité

La cohabitation urbaine doit obliger les villes à aménager des espaces pour l'activité physique

Nous avons donc bien identifié un facteur de l'obésité

Extrait rapport OMS 2017 lien urbanisation et obésité

L'obésité infantile



Même opération que pour l'urbanisation, sauf qu'en abscisse nous avons l'obésité infantile.

On remarque un lien entre obésité infantile et obésité adulte masculine

Un effort de prévention à faire du côté des enfants

On repart pour l'approche opérationnelle

Results

The percentage of non-obese children who grew up to be non-obese adults was 62.6%, and that of obese children who grew up to be obese adults was 80.0%. There was a significant association between childhood body mass index (BMI) and adulthood BMI with a β regression coefficient of 3.76 [95% confidence interval (CI): 1.36-6.16], and between childhood obesity and adulthood obesity with an odds ratio of 5.76 (95% CI: 1.37-24.34). There was no statistical difference between parental obesity at baseline and children's adulthood obesity, after adjustment of confounders. Male participants and those aged 10.0-13.0 years had a higher risk of adulthood obesity with odds ratios of 2.50 (95% CI: 1.12-5.26) and 3.62 (95% CI: 1.17-11.24), respectively.

Conclusion

Childhood obesity is an important predictor of adulthood obesity.

Résultat étude CDC chinois de 2019

Dans cette étude du CDC chinois deux résultats suivent nos observations

62,6 % des enfants non obèses sont devenus des adultes non obèses

En revanche, 80 % des enfants obèses le sont restés une fois adultes

Comme dit dans l'article l'obésité infantile est un important prédicteur l'obésité adulte

Conclusion

Obésité facilement identifiable chez l'homme, au contraire de la femme où les causes semblent multiples et plus profondes

L'urbanisation est un enjeu important de l'obésité notamment par rapport à la place qu'il laisse à l'activité physique

L'accès à une nourriture saine, notamment moins riche en graisse est aussi un facteur déterminant.

Mais la prévention doit surtout s'anticiper dès l'enfance comme car un enfant obèse a plus de chance de devenir obèse à l'âge adulte...

Tous les facteurs identifiés dans nos modèles de régression font échos aux politiques d'aujourd'hui tournées vers l'équilibre alimentaire, la promotion du sport et la prévention chez les enfants.

Cependant les modèles n'expliquent pas toute la variance. Il y a des facteurs manquants dans cette analyse comme par exemple le poids de la génétique ou les maladies chroniques.