# Le corpus de données

Les données proviennent du US Census de 1990, des données policières du US LEMAS survey de 1990, et des données sur le crime du FBI UCR de 1995. Elles ont été collectées par Michael Redmond, chercheur du département informatique de l’université La Salle de Philadelphie. Ce même Michael Redmond en a fait don à la science en 2009. La notion de crime regroupe les meurtres, les viols, les vols et les agressions.

# Signification des caractères

Chaque individu du corpus de données représente une ville américaine qui va être décrite selon 101 variables (toutes normalisées telles que chaque variable soit entre 0 et 1) et une variable à expliquer : le nombre de crimes violents pour 100k personnes.

Le fichier excel contient deux feuilles. La première feuille contient tout le corpus de données. La deuxième feuille est une extraction de la première feuille. On vous donne dans le tableau ci-dessous, la signification des caractères de la deuxième feuille.

|  |  |
| --- | --- |
| **Libellé** | **Signification** |
| Population | Population |
| householdsize | Nombre de personnes au sein du foyer |
| racePctBlack | % de noirs dans la population |
| racepctwhite | % de blancs dans la population |
| racePctAsian | % d'asiatiques dans la population |
| racePctHisp | % d'hispaniques dans la population |
| agePct12t21 | % de personnes ayant entre 12 et 21 ans |
| agePct12t29 | % de personnes ayant entre 12 et 29 ans |
| agePct16t24 | % de personnes ayant entre 16 et 24 ans |
| agePct65up | % de personnes ayant plus de 65 ans |
| pctWWage | % de foyers dont les revenus proviennent de salaires |
| pctWFarmSelf | % de foyers dont les revenus proviennent de l’agriculture ou de l’auto-entreprenariat |
| pctWInvInc | % de foyers dont les revenus proviennent d’investissements ou de loyers |
| pctWSocSec | % de foyers dont les revenus proviennent de la sécurité sociale |
| pctWPubAsst | % de foyers dont les revenus proviennent des aides de l’état |
| pctWRetire | % de foyers dont les revenus proviennent de la retraite |
| perCapInc | Revenu par personne |
| PctNotHSGrad | % de personnes de plus de 25 ans ayant un niveau d’éducation qui ne dépasse pas le bac |
| PctBSorMore | % de personnes de plus de 25 ans ayant fait des études supérieures |
| PctFam2Par | % de familles avec enfants avec les deux parents présents |
| PctWorkMom | % de mères actives ayant des enfants de moins de 18ans à charge |
| PctWorkMomYoungKids | % de mères actives |
| PctIlleg | % d’enfants nés hors mariage |
| PersPerOccupHous | Nombre moyen de personnes par foyers |
| PctPersDenseHous | % de personnes vivant dans des foyers denses (plus d’une personne par pièce) |
| PctHousNoPhone | % de foyers n’ayant pas le téléphone |
| PctWOFullPlumb | % de foyers n’ayant pas une plomberie complète |
| PctForeignBorn | % de personnes nées à l'étranger |
| PctBornSameState | % de personnes nées dans le même état |
| PctSameHouse85 | % de personnes dans la même maison depuis 1985 |
| PctSameCity85 | % de personnes dans la même ville depuis 1985 |
| PctSameState85 | % de personnes dans le même état depuis 1985 |
| PopDens | Densité de la population |
| ViolentCrimesPerPop | Nombre total de crimes violent pour 100K habitants |