

Geoducation

Antoine Drouhin, Aurélien Garret, Cécile Hu, Lucas Morel

Introduction

Base de données

Notre base de données a été trouvée sur le site Data.gouv. Nous avons croisé deux jeux de données distincts. Le premier concerne des données sur la performance des lycées en France (taux de réussite etc). Le second présente des données géographiques pour l'ensemble des établissements scolaires français (Coordonnées GPS, etc.).

<https://www.data.gouv.fr/fr/>

Nous avons choisi cette base car elle présente une approche intéressante sur la compréhension d'un sujet qui nous concerne tous, l'éducation. L'approche géographique des questions de réussite scolaire nous a semblé être un champs suffisamment complet pour permettre l'utilisation des méthodes d'analyse vue en cours.

L'intérêt et le questionnement général porte sur la réussite scolaire de la France en fonction de la position géographique des établissements scolaires. Y'a-t-il une corrélation entre la situation géographique des établissements et la réussite scolaire des étudiants ?

Plusieurs problématiques en découlent comme :

Y'a-t-il des différences de réussites entre le top 10 des grandes villes en France et les villes de province ? Paris bénéficie-t-elle d'une réussite supérieure au reste de la France ? Quelles régions de France semblent réussir mieux que les autres ?

Initialisation de la base de donnée

```
read.csv2("geoducation-data2.csv", sep=";", header=TRUE, na.strings = "")->bdd
#exists('bdd')
```

Khi-Deux

```
bddKhiDeux = bdd[, c('Académie', 'Effectif.Présents.série.L', 'Effectif.Présents.série.ES', 'Effectif.Présents.série.M')]
# Petit clean des datas (Antoine)
bddKhiDeux[is.na(bddKhiDeux)] <- 0

# Cette portion de code suppose que bdd est ordonné par nom d'académie. (Antoine)

tableKhiDeux <- data.frame()
indiceCourant = 1
aca <- bddKhiDeux[1, "Académie"]

for(i in 1:nrow(bddKhiDeux)){
  if(aca != bddKhiDeux[i, "Académie"]){
    indiceCourant = indiceCourant + 1
    aca <- bddKhiDeux[i, "Académie"]
  }
}
```

```

if (length(rownames(tableKhiDeux)) != 0 && !is.na(tableKhiDeux[indiceCourant,"Académie"])) && bddKhiDeux[i,"Effectif.Présents.série.L"] != 0 {
  tableKhiDeux[indiceCourant,"ES"] <- tableKhiDeux[aca,"ES"] + bddKhiDeux[i,"Effectif.Présents.série.L"]
  tableKhiDeux[indiceCourant,"L"] <- tableKhiDeux[aca,"L"] + bddKhiDeux[i,"Effectif.Présents.série.L"]
  tableKhiDeux[indiceCourant,"S"] <- tableKhiDeux[aca,"S"] + bddKhiDeux[i,"Effectif.Présents.série.S"]
} else {
  tableKhiDeux <- rbind(tableKhiDeux, data.frame(Académie = aca,ES = bddKhiDeux[i,"Effectif.Présents.série.L"],L = bddKhiDeux[i,"Effectif.Présents.série.L"],S = bddKhiDeux[i,"Effectif.Présents.série.S"]))
}
}

```