

# Présentation du projet de statistiques

Antoine DROUHIN  
Aurélien GARRET  
Cécile HU  
Lucas MOREL

## Base de données choisie

La base de données sur laquelle nous allons travailler est une base publique provenant du site Data.gouv.fr

Deux jeux de données distincts ont été croisés :

- des données sur la performance des lycées en France (le taux de réussite, etc.)
- des données géographiques pour l'ensemble des établissements scolaires français (coordonnées GPS, etc.).

Pour accéder à notre base :

<https://www.dropbox.com/s/n2w3q3y3cs3iynit/Sauvonlecole.csv?dl=0>

## Pourquoi ?

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 champ suffisamment complet pour permettre l'utilisation des méthodes d'analyse vue en cours.

## Méthodes statistiques envisagées

La base comporte de nombreuses variables, de type quantitatives et qualitatives, nous permettant d'appliquer de nombreux tests, offrant ainsi divers indicateurs.

## La problématique

L'intérêt et le questionnement général portent 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 sous-questionnements 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 ?

### Méthode du Khi-deux :

- pour comparer les filières aux régions (2 variables qualitatives)

### Méthode de Student :

- pour mettre en évidence les résultats des lycées publics par rapport aux privés (2 modalités présentes)

### Méthode d'Anova :

- pour les 10 villes les plus représentées face aux réussites (qualitatif et quantitatif avec plus de 2 modalités)

### Méthode de corrélation :

- pour constater les effectifs et les taux de réussite (2 variables qualitatives)