Plan du projet de statistiques

# Analyse descriptive des données

* Caractéristiques des données
* ACP
* Graphique des individus
* K-means + détermination auto du nombre de cluster

# Les déterminants de la production non commerciale

Le but est d’essayer de savoir quelle est la cause cette production non commerciale qui n’est pas négligeable. Pour cela, nous voulons déterminer à quel type de vin correspond majoritairement cette production.

* Tests de corrélation entre la proportion de production non commerciale et la proportion de chaque appellation :
  + 1. Corrélation AOP, IGP, VSIG avec production non commerciale
  + 2. Au sein du groupe corrélé, tester corrélation de la production par couleur avec la production non commerciale.

# Modélisation de la production

La variable la plus importante est probablement la production finale. Nous voulons créer un modèle efficace pour cette production.

* Créer une variable qualitative avec les clusters (des groupes de départements)
* Le modèle pourrait être composé de :
  + La superficie totale des vignobles, ou les superficies par appellation si c’est plus précis
  + La variable qualitative issue des clusters qui pourraient peut-être permettre de diminuer le nombre de variables du modèle