Projet de statistiques

Production vinicole française en 2015 par département

# Introduction

La France est le 2e producteur mondial de vin derrière l’Italie et devant l’Espagne. En 2015, la production française représentait un peu plus de 17% de la production mondiale, soit environ 48 millions d’hectolitres. De plus, comme environ un tiers de cette production est exportée, le secteur des vins et spiritueux est le second secteur d'exportation bénéficiaire en France derrière l’aéronautique. Il s’agit donc d’un secteur économique très important pour le pays.

Cette production est très hétérogène car il y a en France de nombreux terroirs différents, et donc de nombreux types de vins. Pour protéger ces spécificités, il existe des dizaines d’appellations sous forme d’AOP ou d’IGP. Elles permettent de contrôler l’origine du produit et, dans le cas des AOP, le respect d’un cahier de charges.

Les données statistiques relatives à la production vinicole française sont éditées tous les ans par la douane, qui les met à disposition sur son site internet <http://www.douane.gouv.fr/>. Nous allons ici analyser une partie de ces données afin d’essayer de dégager certaines caractéristiques de cette production.

# Analyse des données

## Caractéristiques principales des données

On dispose de 23 variables, dont 2 qualitatives qui correspondent au noms des départements et des régions :

* *departement* : Nom du département
* *nombre\_declarations* : Nombre de de producteurs déclarés dans ce département
* *superficie* : Superficie totale des vignobles du département (ha)
* *superficie\_aop* : Superficie des vignobles du département contrôlés par une Appellation d’Origine Contrôlée (ha)
* *superficie\_cognac* : Superficie des vignobles du département destinés à la production de cognac (ha)
* *superficie\_igp* : Superficie des vignobles du département contrôlés par une Indication Géographique Protégée (ha)
* *superficie\_vsig* : Superficie des vignobles du département non contrôlés par une quelconque indication géographique (ha)
* *qte\_aop\_blanc*, *qte\_aop\_rouge*, *qte\_aop\_rose* : Production vinicole du département sous une AOP, correspondant respectivement à du vin blanc, du rouge et du rosé (hl)
* *qte\_vci* : Volume Complémentaire Individuel (hl)
* *qte\_igp\_blanc*, *qte\_igp\_rouge*, *qte\_igp\_rose*, *qte\_igp\_vci* : Production vinicole du département sous une IGP, correspondant respectivement à du vin blanc, du rouge, du rosé, et du VCI (hl)
* *qte\_vsig\_blanc*, *qte\_vsig\_rouge*, *qte\_vsig\_rose* : Production vinicole du département sans indication géographique, correspondant respectivement à du vin blanc, du rouge et du rosé (hl)
* *qte\_cognac* : Production de cognac du département (hl)
* *prod\_commerciale* : Production du département commercialisée (hl)
* *prod\_noncommerciale* : Production du département non commercialisée. Cela correespond aux jus de raisins produits par les viticulteurs, les lies et/ou les vins en dépassement de rendement (hl)
* *total* : Production totale du département
* *region* : Région correspondant au département

On remarque que la variable *total* correspond à la somme de toutes les variables exprimant le volume d’un type de production – celles commençant par « qte » - mais aussi la somme des variables *prod\_commerciale* et *prod\_noncommerciale*.

> summary(vins$total)

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

4 4202 72390 1196390 519220 47855617 1

## Réalisation d’une analyse en composantes principales

Notre étude est basée sur un nombre conséquent de variables. De plus, mis à part les variables de superficies et de productions, et la variable « total » qui est une somme de variables ; il est difficile de trouver des liens corrélations évidentes entre ces variables. Nous avons donc décidé de réaliser une ACP afin d’avoir une première approche des corrélations potentielles.