

Wine Quality

Alexandre SALMON, Thomas FRION

13/01/2021

1 Objectif de l'étude

Notre étude va porter sur la qualité des vins portugais, de la région Vinho Verde. Nos objectifs via cette étude sont :

- Déterminer un modèle de prédiction de la qualité d'un vin rouge et d'un vin blanc. Ainsi, connaître le critère physionomique le plus important dans la détermination de la qualité pour chaque type de vin.
- Comparer les deux modèles (blanc et rouge) pour savoir si ce qui fait un bon vin blanc, fait un bon vin rouge

Les conclusions de cette étude pourront permettre aux vignerons d'améliorer la qualité de leurs vins.

2 Analyse descriptive

Pour cette étude nous disposons de deux jeux de données : un pour le vin rouge et 1 pour le vin blanc. Nous avons obtenu ces données sur la page UCI: Wine Quality Data Set

Pour les deux jeux de données nous avons douze variables: onze variables d'entrée qui correspondent aux critères physionomiques du vin et une variable de sortie qui correspond à la qualité du vin en 0 et 10. Les onze critères physionomiques sont les suivantes :

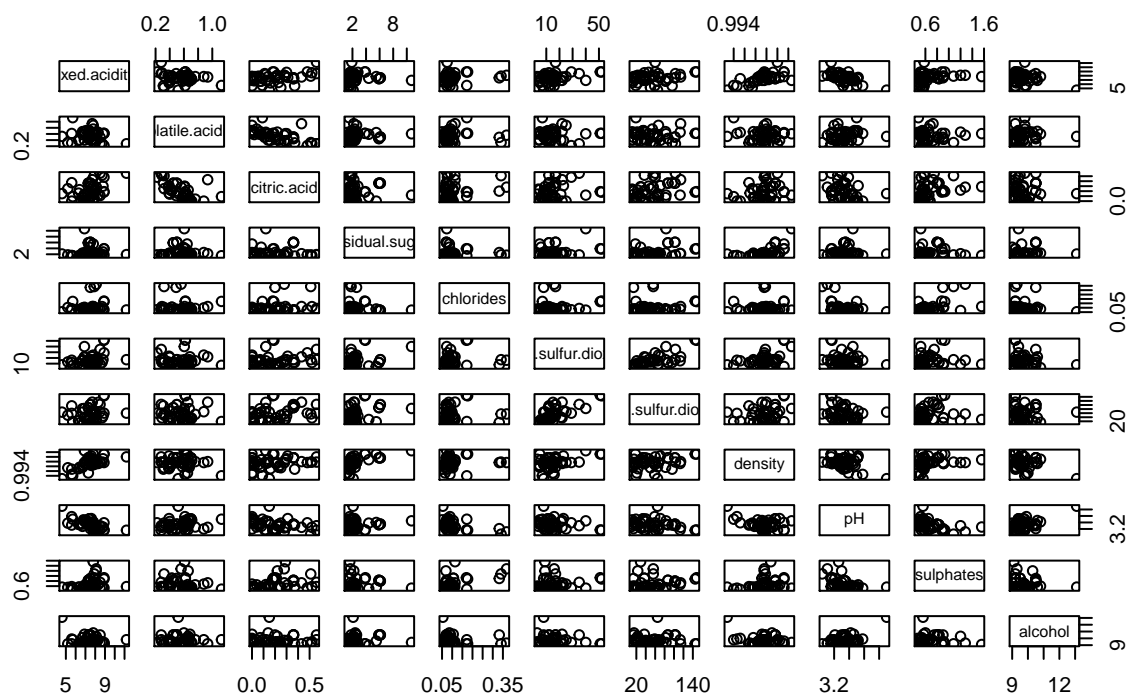
1. Acidité fixe (ou acidité naturelle du raisin)
2. Acidité volatile (teneur d'acides gras)
3. Acide citrique (teneur d'acide citrique)
4. Sucre résiduel (sucres encore présents après la fermentation)
5. Chlorures (teneur des différents chlorures)
6. Dioxyde de soufre (SO_2) libre (teneur du principe actif du SO_2)
7. Total du dioxyde de soufre (SO_2) (teneur de toutes les formes du SO_2)
8. Densité
9. pH
10. Sulfates (teneur du fongicide)
11. Alcool

La variable de sortie correspond à la médiane des notes données par des experts (au minimum trois notes). Si la variable vaut 0 cela signifie que le vin est de très mauvaise qualité. Si la variable vaut 10 alors le vin est de bonne qualité.

Nous avons 1599 vins rouges et 4898 vins blancs.

Afin de connaître les liens entre les variables d'entrée, nous allons prendre l'ensemble des 11 variables d'entrée ainsi que les 50 premiers vins. Nous avons fait le choix de cet échantillon, car si nous faisons le plot de chaque jeu de données complet, le résultat sera illisible.

Plot of 50st Red Wines



Plot of 50st White Wines

