

Rapport sur la corrélation potentielle entre certaines propriétés chimiques et la qualité du vin

MONTERIN Maxime

LA QUESTION DONNEE

Existe-t-il une corrélation forte entre certaines propriétés chimiques et la qualité du vin ?

GENERATION DE GRAPHIQUES POUR REPONDRE A LA QUESTION (vin rouge)

Choix d'un scatter plot pour montrer une corrélation entre un élément chimique et la note d'un vin

Chargement des données

```
redwine <- read.csv("~/projet-if36-p25-morphe-utt/data/red_wine.csv", header = TRUE, sep = ';')
whitewine <- read.csv("~/projet-if36-p25-morphe-utt/data/white_wine.csv", header = TRUE, sep = ';')
```

Etape 1 : On ne garde que les variables qui sont reliées à de la chimie

```
variables_chimiques <- c("fixed_acidity", "volatile_acidity", "citric_acid",
                        "residual_sugar", "chlorides", "free_sulfur_dioxide",
                        "total_sulfur_dioxide", "density", "pH", "sulphates", "alcohol")
```

Etape 1.5 : Ajout des unités aux catégories gardées

```
unit_labels <- c(
  "fixed_acidity"      = "Acidité fixe (g/dm³)",
  "volatile_acidity"   = "Acidité volatile (g/dm³)",
  "citric_acid"        = "Acide citrique (g/dm³)",
  "residual_sugar"     = "Sucre résiduel (g/dm³)",
  "chlorides"          = "Chlorures (g/dm³)",
  "free_sulfur_dioxide" = "SO2 libre (mg/dm³)",
  "total_sulfur_dioxide" = "SO2 total (mg/dm³)",
  "density"            = "Densité (g/cm³)",
  "pH"                 = "pH",
  "sulphates"          = "Sulfates (g/dm³)",
  "alcohol"            = "Alcool (% vol.)"
)
```

Etape 2 : On factorise nos différentes variables et leurs valeurs dans un tableau à 2 dimensions.

ça va permettre de générer pleins de graphiques en une seule fois plutôt que d'en générer plusieurs à la main

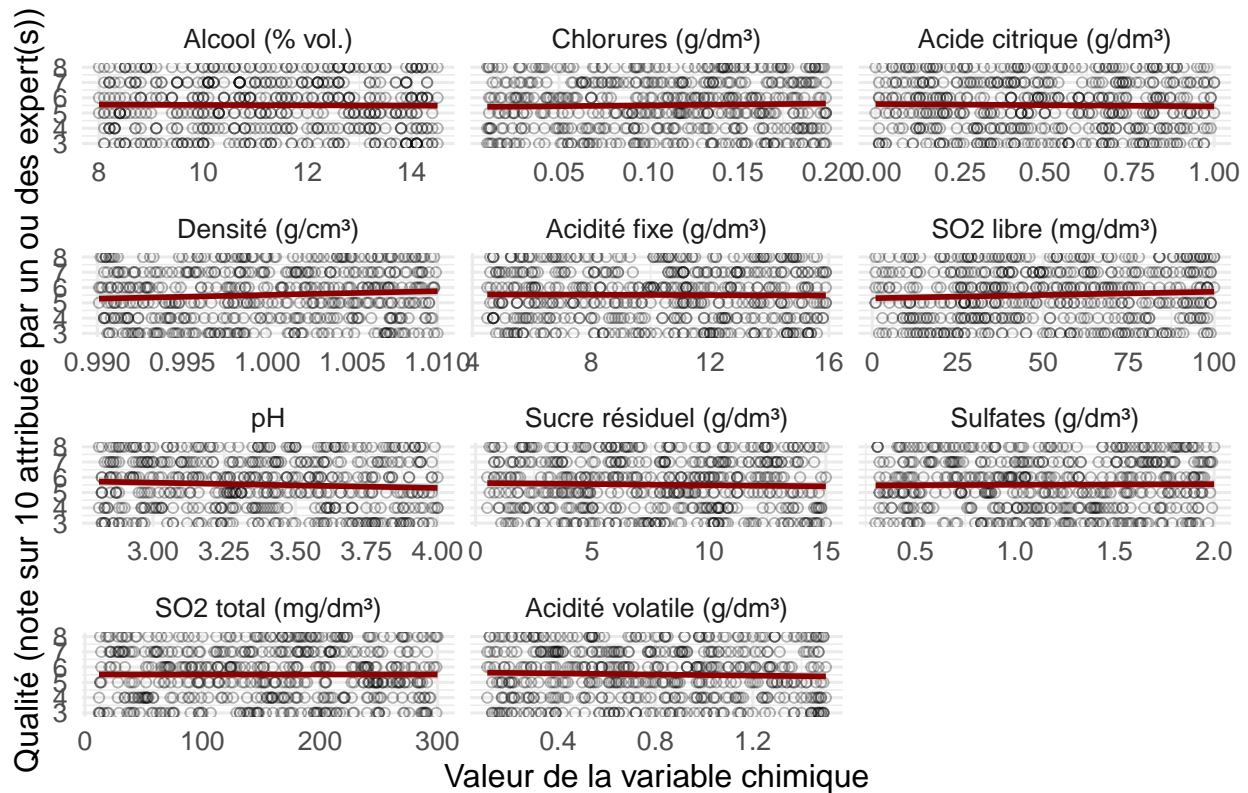
```
redwine2 <- redwine %>%
  select(quality, all_of(variables_chimiques)) %>%
  pivot_longer(
    cols = all_of(variables_chimiques),
    names_to = "variable",
    values_to = "valeur"
  )
```

Etape 3 : Génération des scatters plots avec le tableau factorisé B)

```
ggplot(redwine2, aes(x = valeur, y = quality)) +
  # Génération des points
  geom_point(alpha = 0.3, shape = 1) +
  # Génération de la droite de regression avec son intervalle de confiance
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  # Sert pour la création de nos graphiques, c'est ce qui permet la répartition sur 3 colonnes
  # mais aussi le fait que les abscisses soient libre
  facet_wrap(~ variable, scales = "free_x", ncol = 3, labeller = labeller(variable = unit_labels)) +
  # Utilise un thème épuré, pour faire ressortir la droite de regression je trouve ça cool
  theme_minimal(base_size = 12) +
  # Gestion des labels
  labs(
    title = "Relations entre une variable chimique d'un vin rouge et sa note de qualité",
    x = "Valeur de la variable chimique",
    y = "Qualité (note sur 10 attribuée par un ou des expert(s))"
  )
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

Relations entre une variable chimique d'un vin rouge et sa note de qua



GENERATION DE GRAPHIQUES POUR REpondre A LA QUESTION (vin blanc)

Etape 2

```
whitewine2 <- whitewine %>%
  select(quality, all_of(variables_chimiques)) %>%
  pivot_longer(
    cols = all_of(variables_chimiques),
    names_to = "variable",
    values_to = "valeur"
  )
```

Etape 3 :

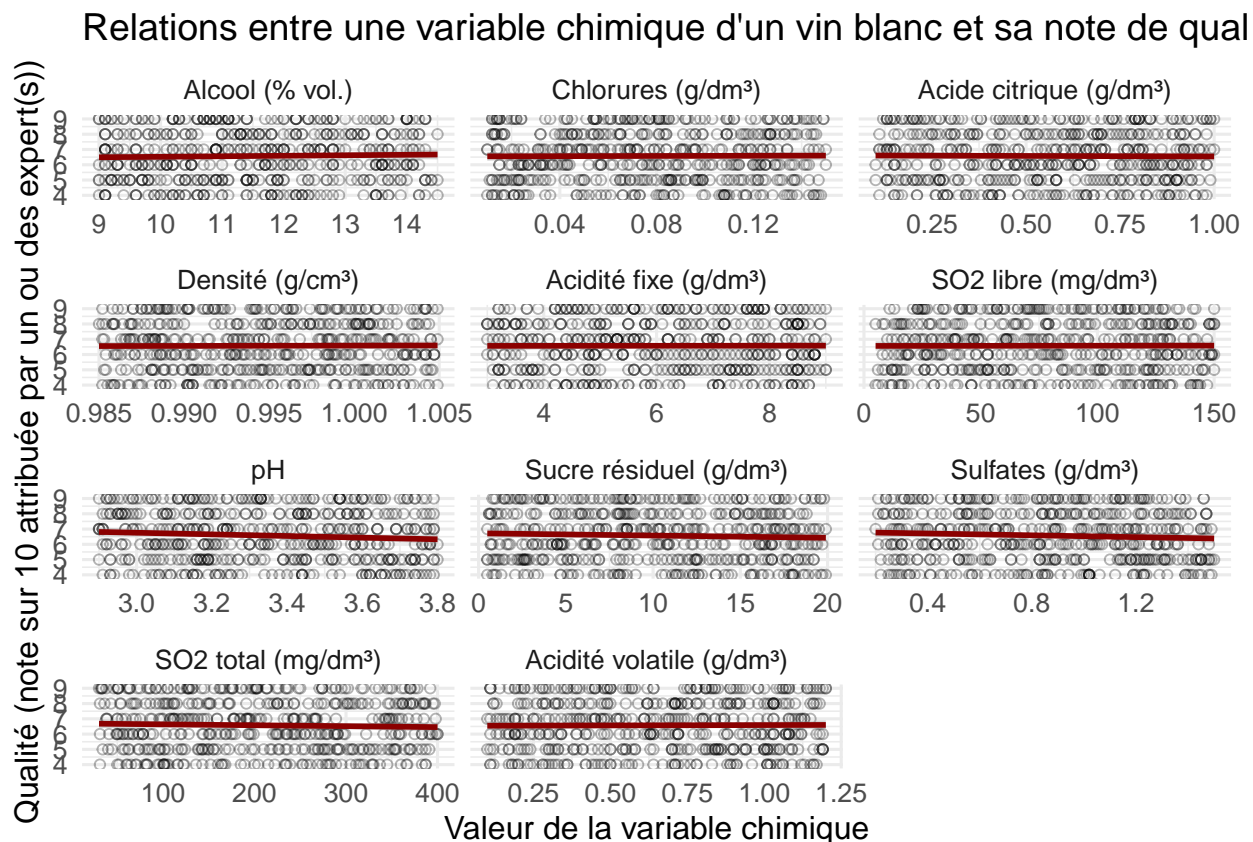
```
ggplot(whitewine2, aes(x = valeur, y = quality)) +
  # Génération des points
  geom_point(alpha = 0.3, shape = 1) +
  # Génération de la droite de regression avec son intervalle de confiance
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  # Sert pour la création de nos graphiques, c'est ce qui permet la répartition sur 3 colonnes
  # mais aussi le fait que les abscisses soient libre
```

```

facet_wrap(~ variable, scales = "free_x", ncol = 3, labeller = labeller(variable = unit_labels)) +
# Utilise un thème épuré, pour faire ressortir la droite de regression je trouve ça cool
theme_minimal(base_size = 12) +
# Gestion des labels
labs(
  title = "Relations entre une variable chimique d'un vin blanc et sa note de qualité",
  x      = "Valeur de la variable chimique",
  y      = "Qualité (note sur 10 attribuée par un ou des expert(s))"
)

```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



ANALYSE ET REPONSE A LA QUESTION DONNEE

VIN ROUGE : La droite de regression est stable quasiment de partout, on peut donc en conclure qu'il n'y a pas de lien fort entre les éléments chimiques et la note du vin. Une des hypothèses qu'on pourrait tirer serait que la note pourrait être attribuée en fonction du ressenti des experts. Par exemple lors d'une dégustation, une note peut-être attribuée en fonction du gout, de l'odorat, voire aussi du visuel si il y a des dépôts en surface etc... Certainement que la provenance du vin joue aussi.

VIN BLANC : Concernant le vin blanc la même analyse peut-être appliquée, la droite de regression est stable quasiment partout. Cependant un lien plus fort peut-être observé entre le pH et la note, ainsi que le sucre résiduel et les sulfates. En effet la droite de regression tend de manière négative, donc plus elle descend dans les notes et plus ces valeurs chimiques sont élevées dans le vin. Donc si on reprend notre analyse

concernant le vin rouge, on pourrait en déduire que ces variables jouent plus sur les variables de ressenti des experts (gout, visuel etc...) que le reste. Plus elles sont élevées dans un vin, plus elles impactent sur la qualité du vin, et plus ce dernier a une note médiocre

CONCLUSION

En conclusion : Il existe bel est bien une corrélation entre certains éléments chimiques et la note du vin, puisque l'évolution des éléments chimiques impactent la note et inversement. En revanche il ne peut pas être catégorisé comme lien fort, puisque aucune variable chimique à elle seule impacte de manière majoritaire la note d'un vin en particulier. Avec cette analyse, on comprend surtout que les variables chimiques d'un vin forment un tout, et que c'est leurs associations qui font varier un vin et sa qualité !