

MBAE Hakim  
AID Lamia

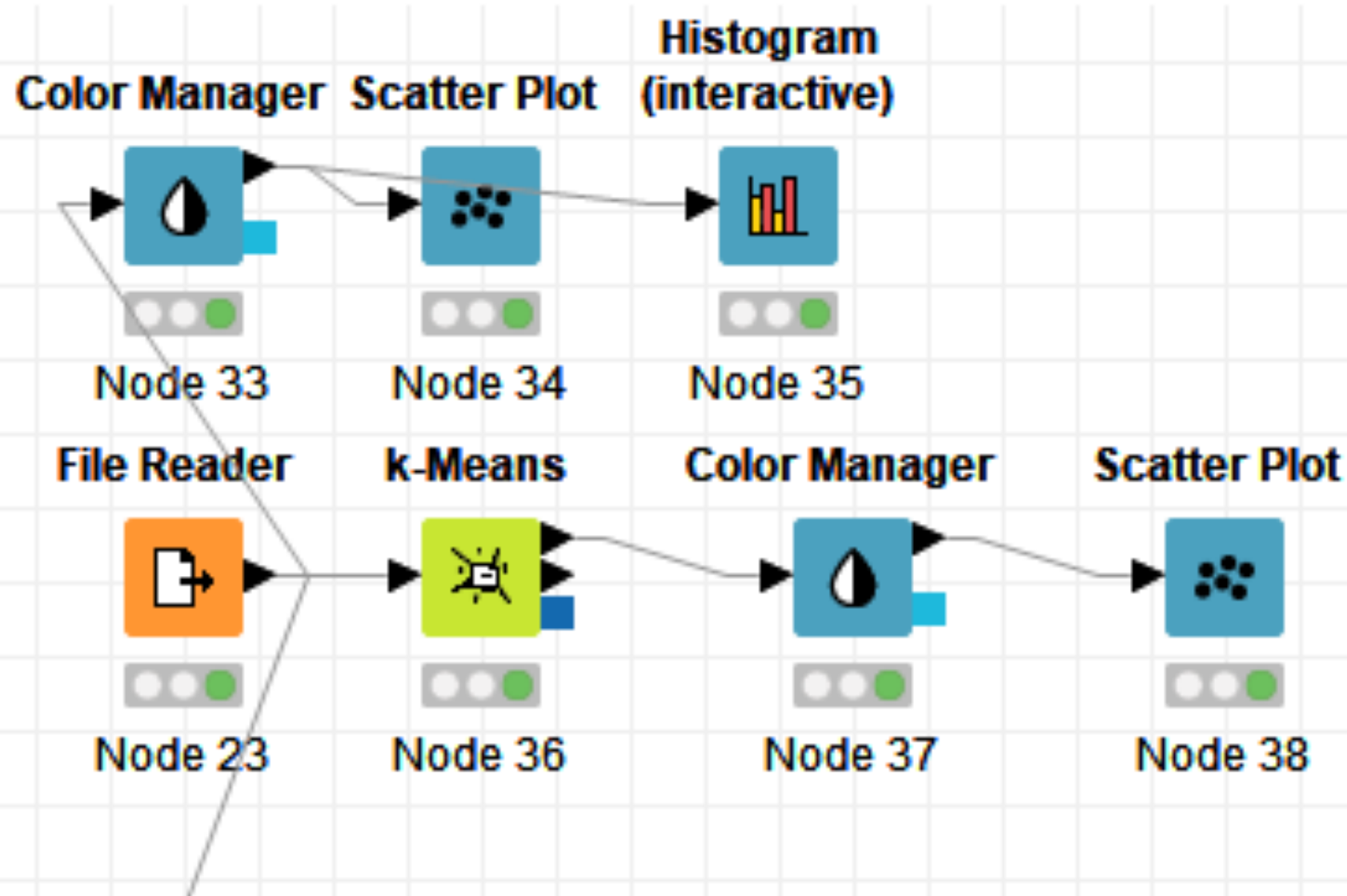
# Projet VDD

(Projet 7 : Données qualité du vin )

2019-2020

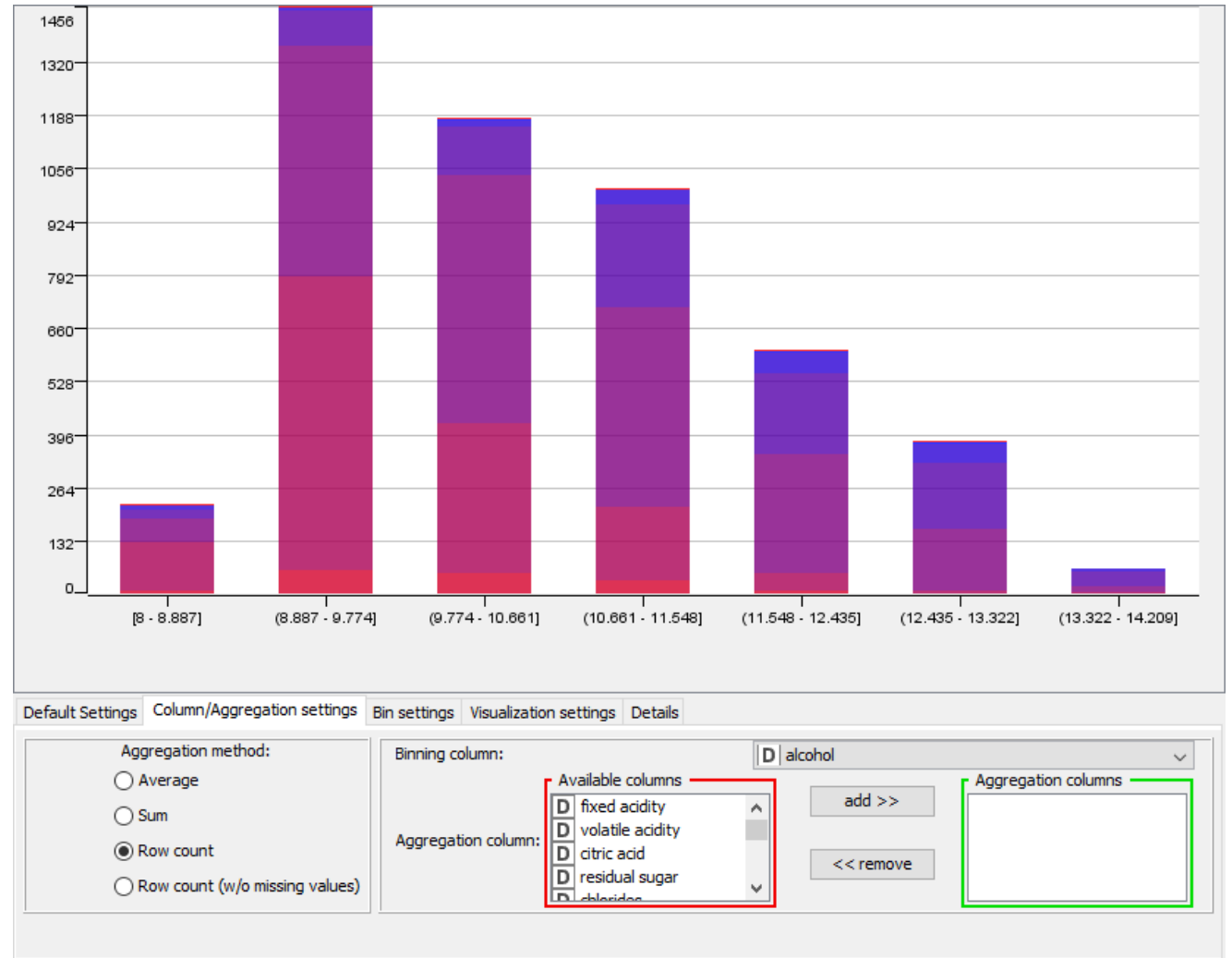


Vin blanc



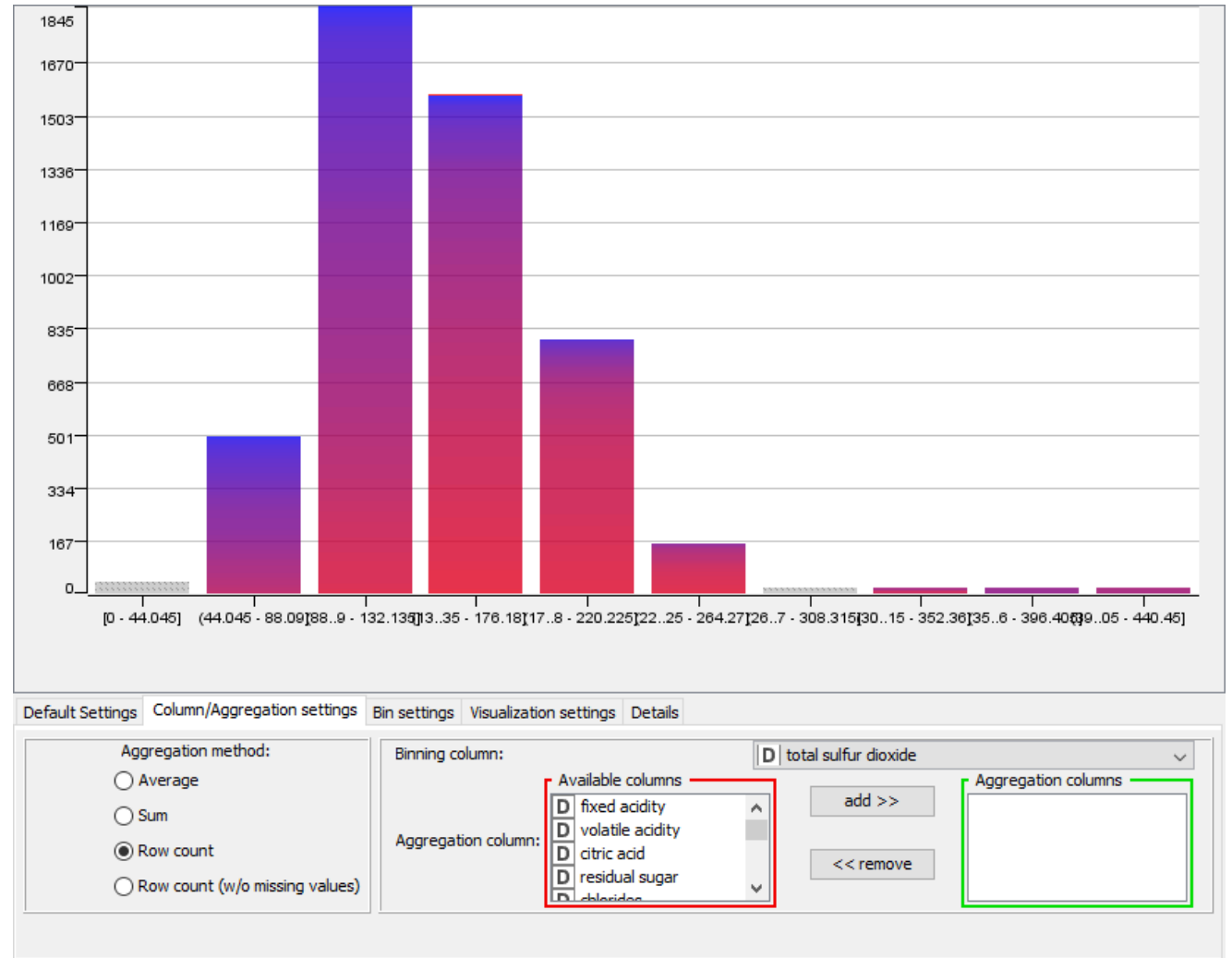
# Alcool et qualité

- Plus le vin est de qualité plus il est bleu, dans le cas contraire il tend vers le rouge.
- Sur l'histogramme ci-joint, on remarque que les vins blancs avec une forte valeur en alcool ont une plus grande probabilité d'être de bonne qualité.



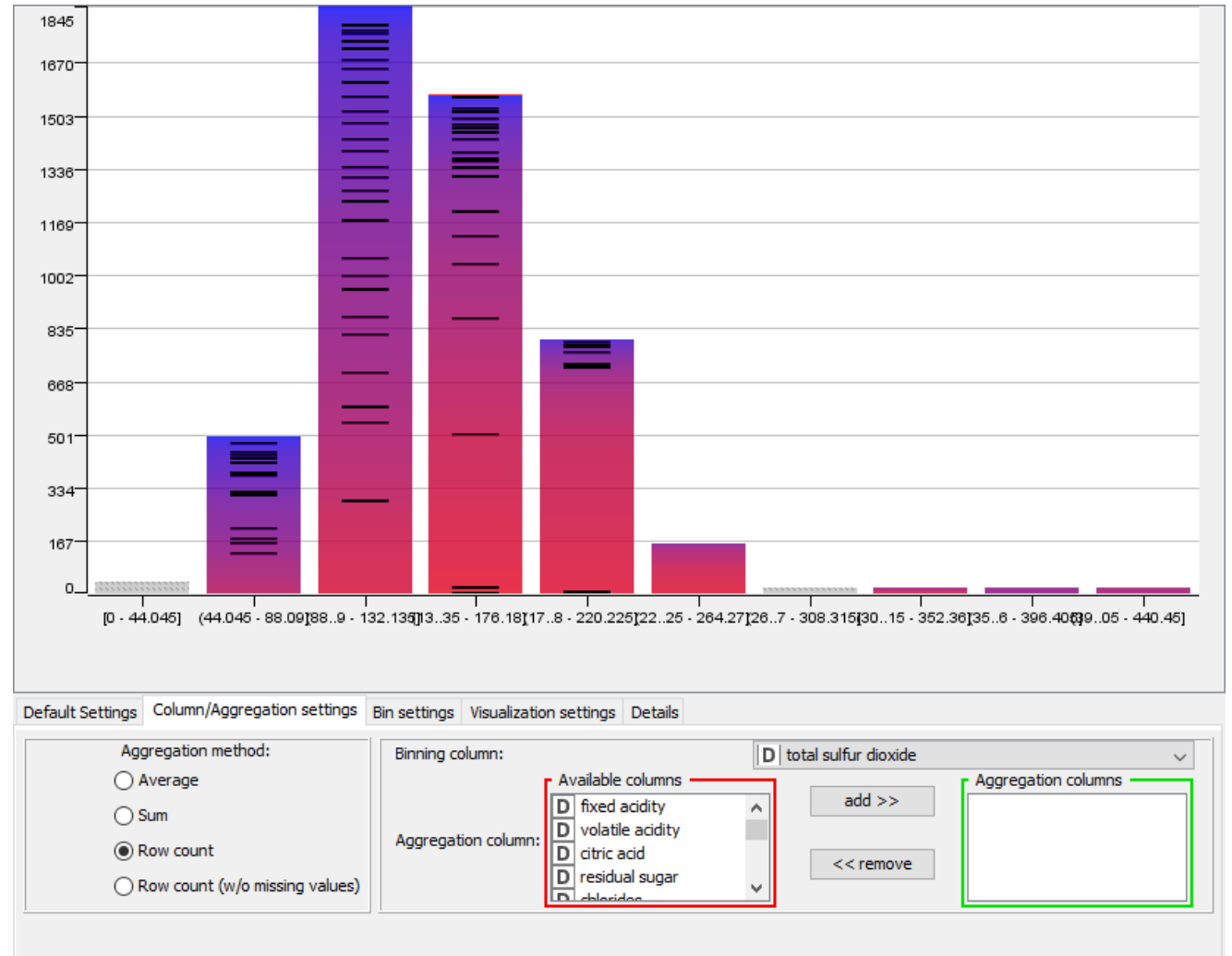
## Alcool et dioxyde de soufre total

- Plus la couleur est bleu plus le vin contient de l'alcool, dans le cas contraire il tend vers le rouge.
- On remarque sur l'histogramme ci-joint qu'en général plus le vin est fort en alcool plus il a de chance d'être allégé en dioxyde de soufre total.



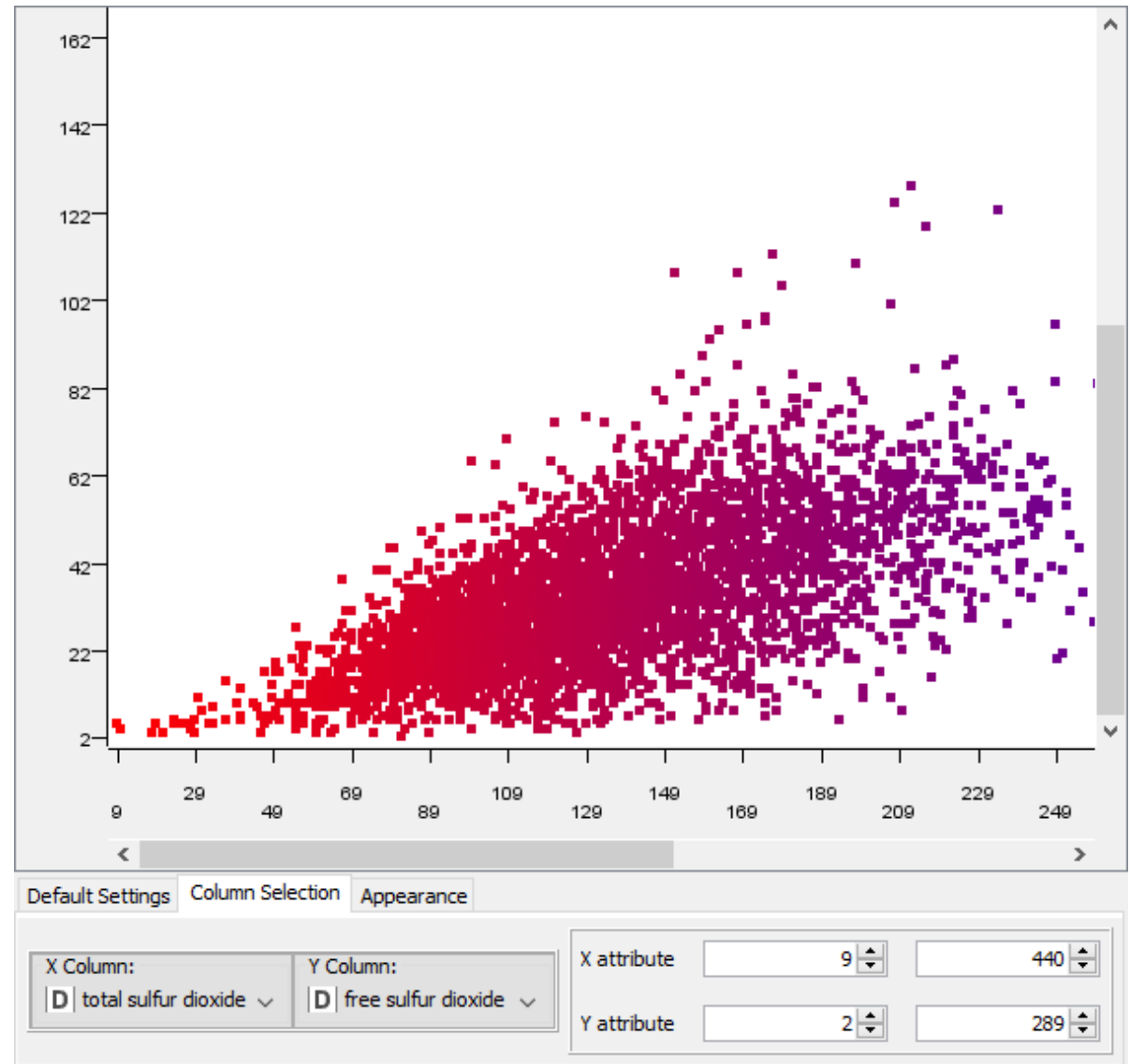
# Qualité et dioxyde de soufre total

- Néanmoins on remarquera que la teneur en dioxyde de soufre total n'influe pas la qualité (ici les meilleurs vins sont sélectionnés en noir). On remarquera quand même qu'aucun vin de qualité n'a plus de 220 en dioxyde de soufre total (je ne m'avance pas sur le lien avec la qualité car l'échantillon au dessus de 220 est trop petit).



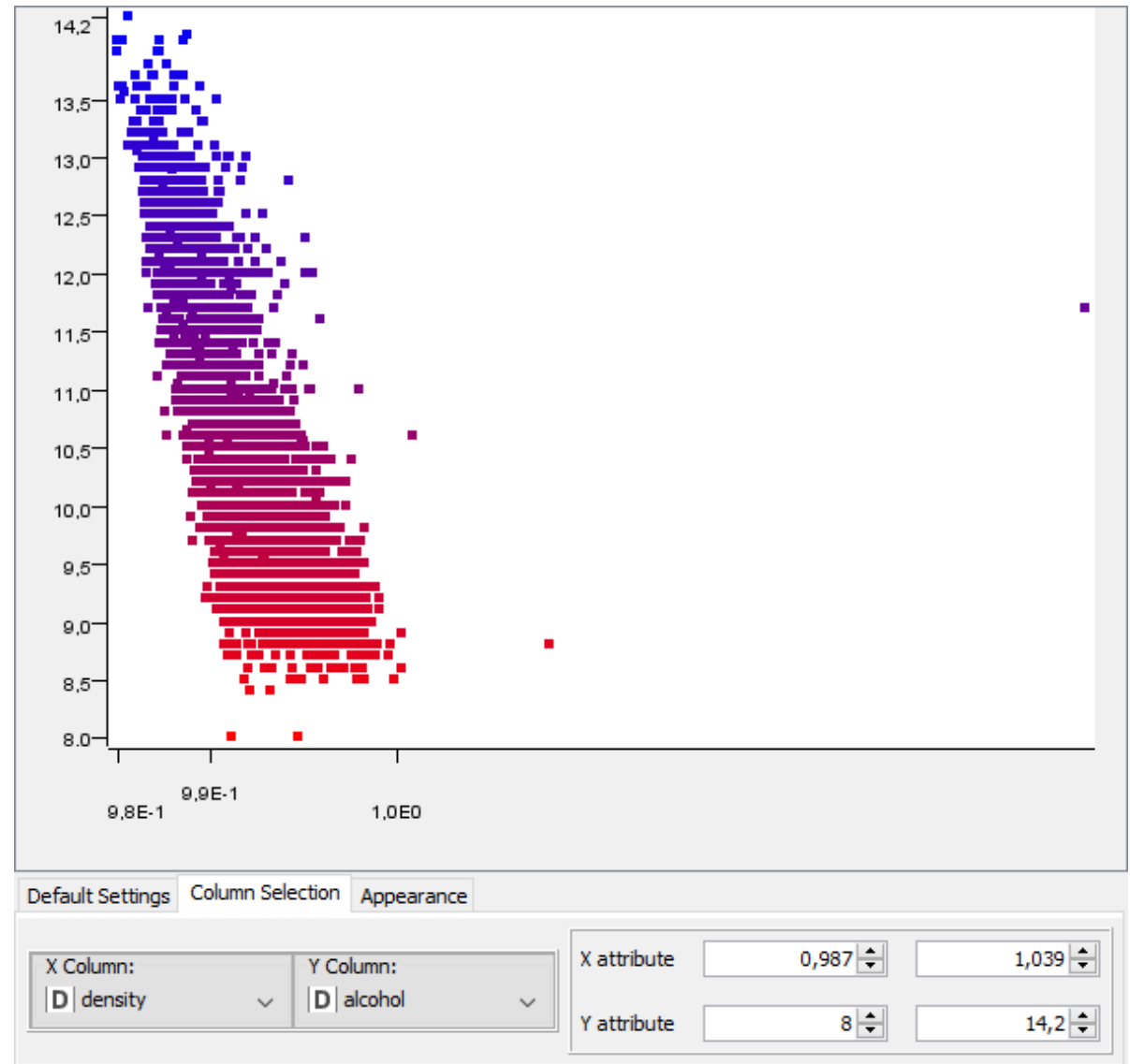
## Quantité de dioxyde de soufre libre et total

- Ici on peut remarquer que la quantité de dioxyde de soufre total borne la quantité de dioxyde de soufre libre.



## Alcool et densité

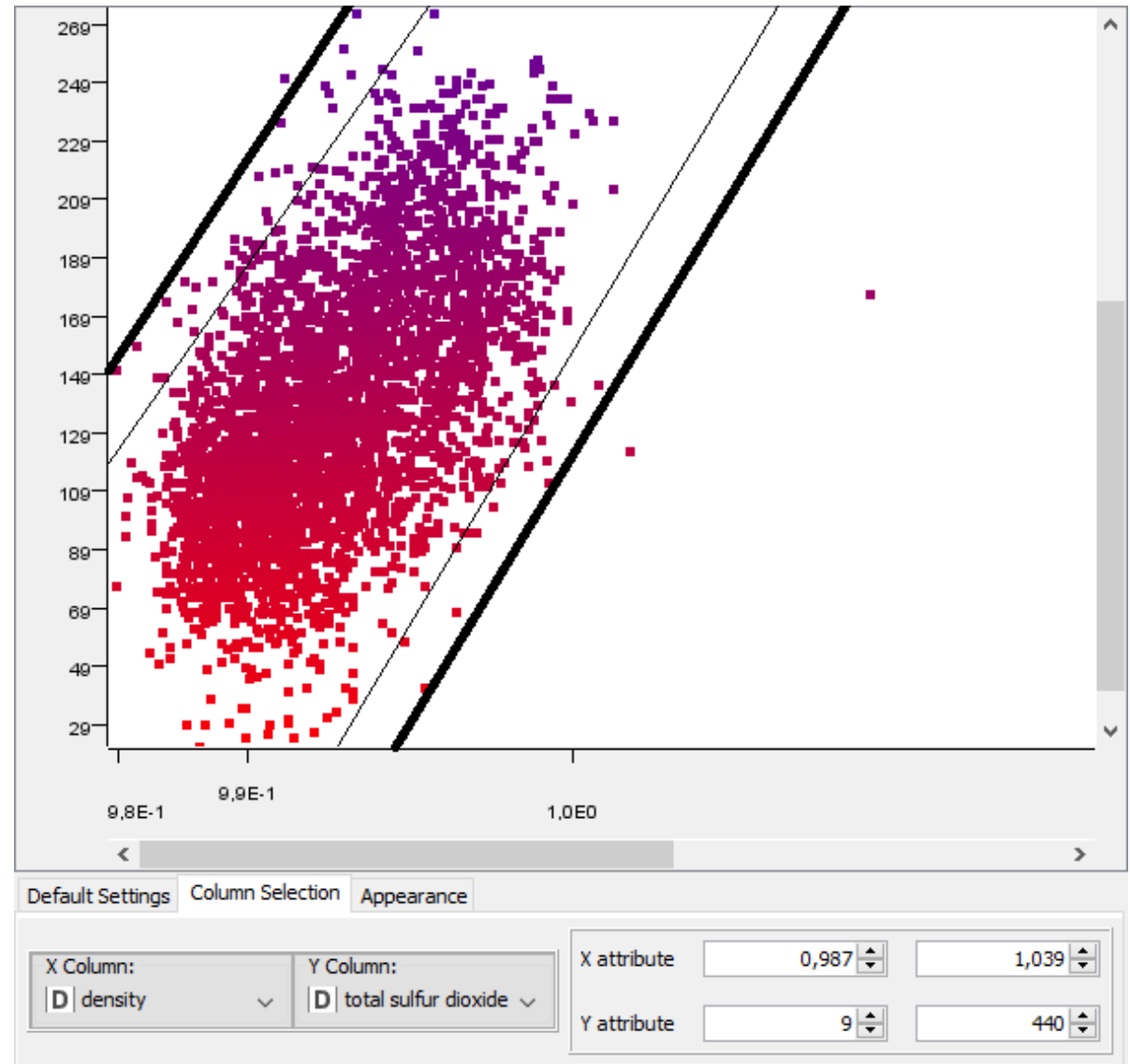
- On remarque ici que quand l'alcool décroît, la densité augmente.





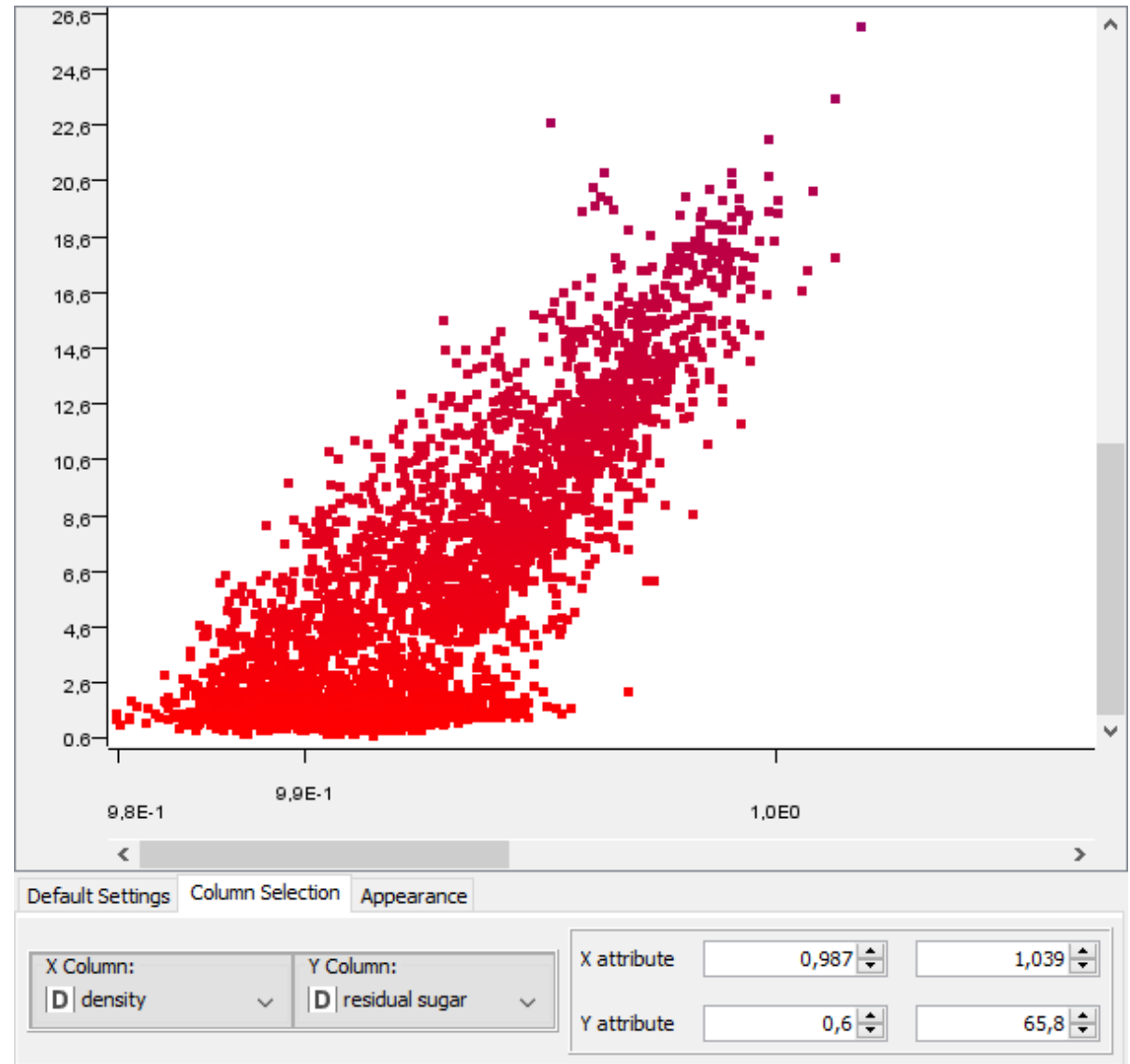
## Densité et dioxyde de soufre total

- Sur ce graphique on voit que la densité et dioxyde de soufre total augmentent ensemble.
- (les courbes ont été tracés en dehors de knime)



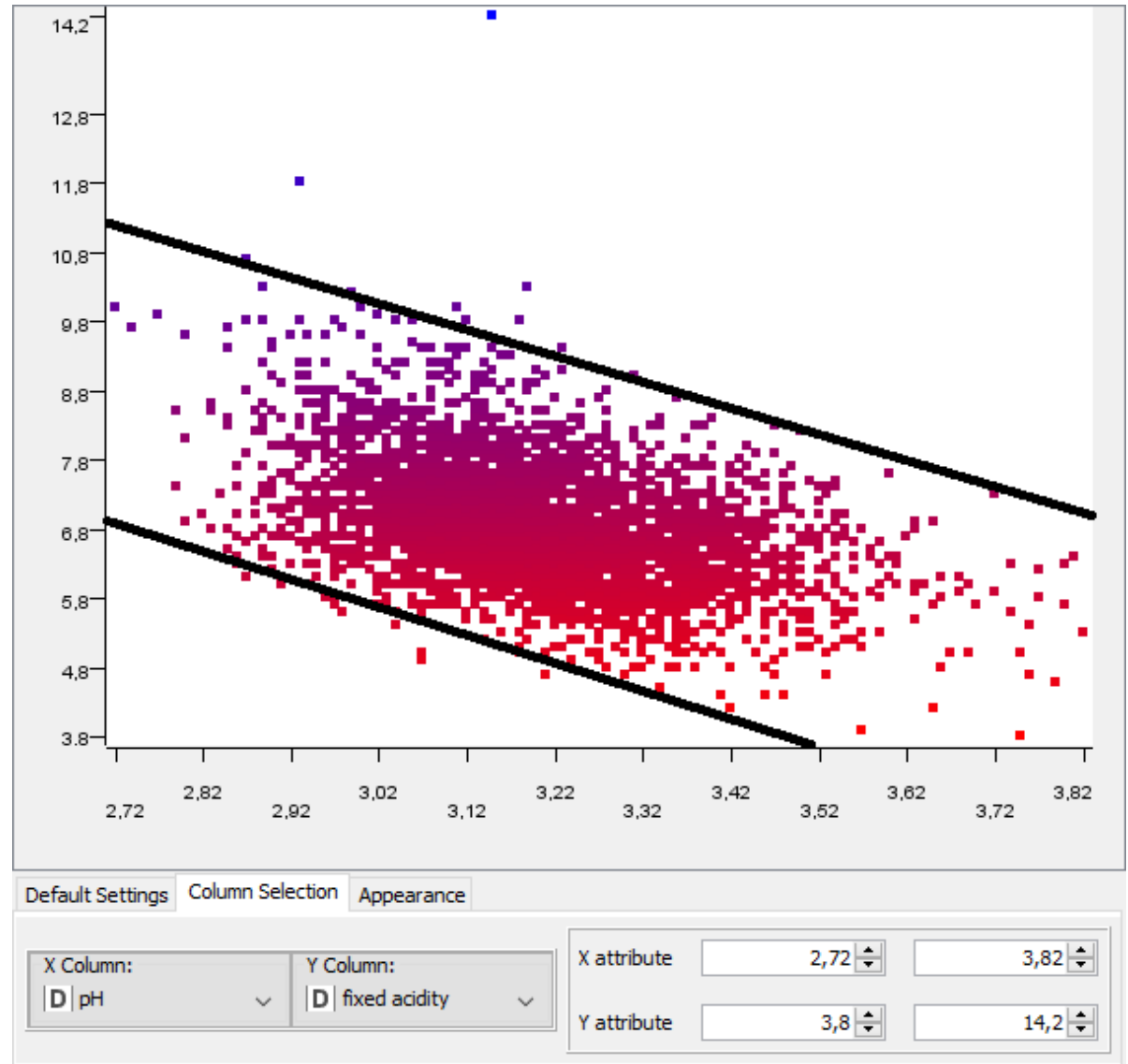
## Densité et sucre résiduel

- Sur cette projection, il semble que la densité borne le sucre résiduel.

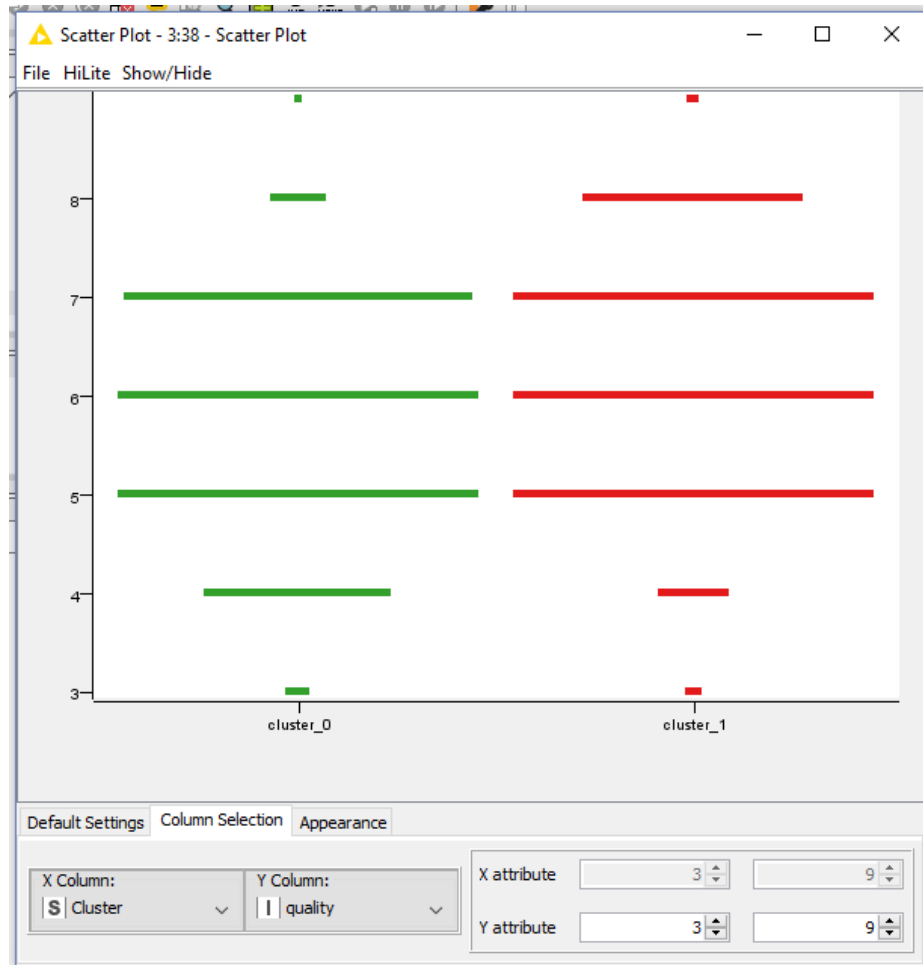


## Acidité fixe et pH

- Ici on remarque que quand l'acidité baisse, le pH semble augmenter. Et ceci paraît logique.



Ici l'essai de clustering le plus concluant. On notera que 50% des vins moyens sont retirés ainsi que 70% des vins mauvais. Néanmoins un peu moins de 20% des bons vins sont écartés.



Dialog - 3:36 - k-Means

File

K-Means Properties Flow Variables Memory Policy

number of clusters: 2

max. number of iterations: 249

Exclude

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- free sulfur dioxide
- total sulfur dioxide
- pH
- sulphates
- quality

Include

- chlorides
- density
- alcohol

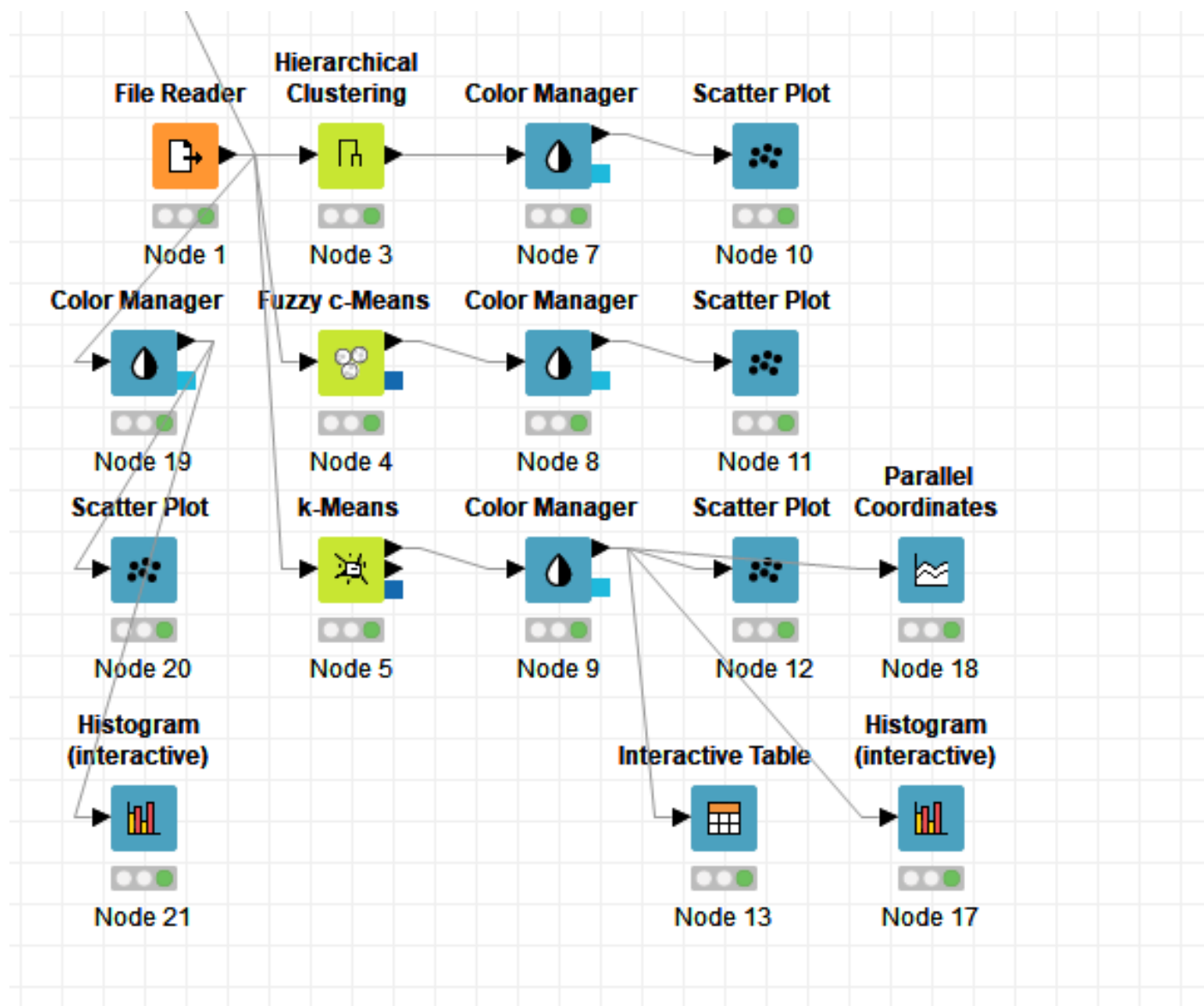
☐ Always include all columns

☐ Enable Hilite Mapping

OK Apply Cancel ?

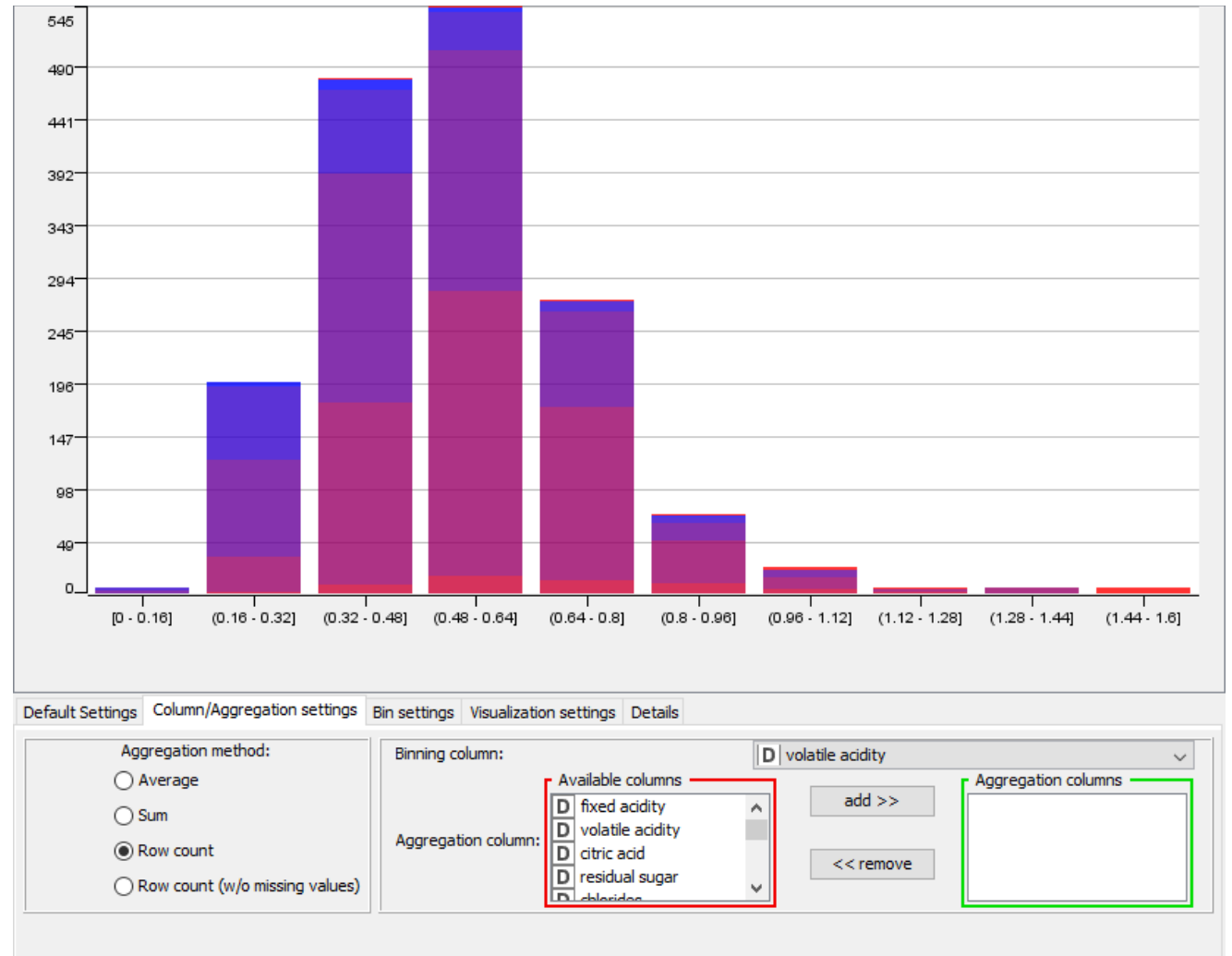


Vin rouge



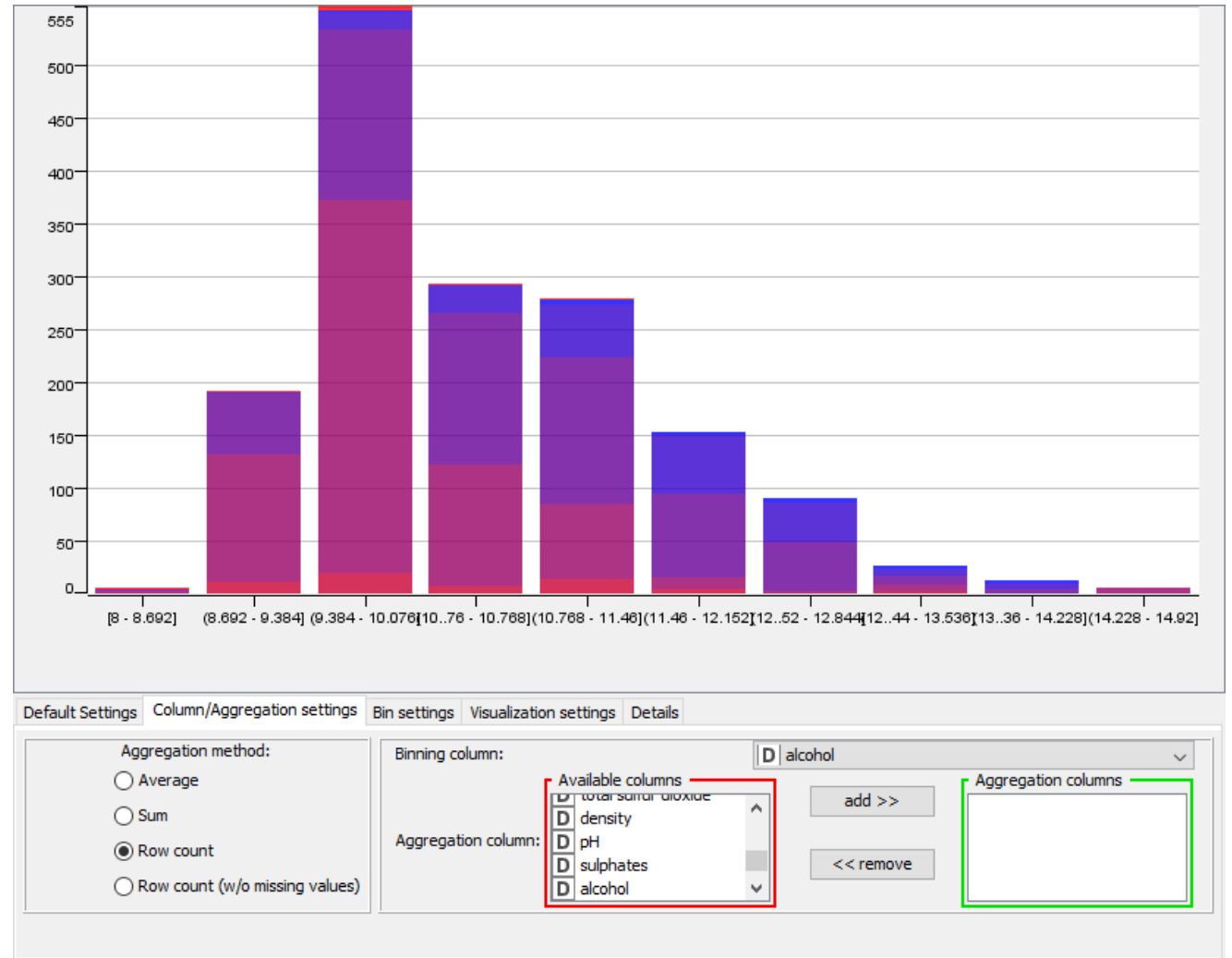
# Qualité et acidité volatile

- On voit ici que les vins rouges au dessus de 0.64 pour l'acidité volatile ont plus de chance d'être de moins bonne qualité que les autres.



## Qualité et alcool

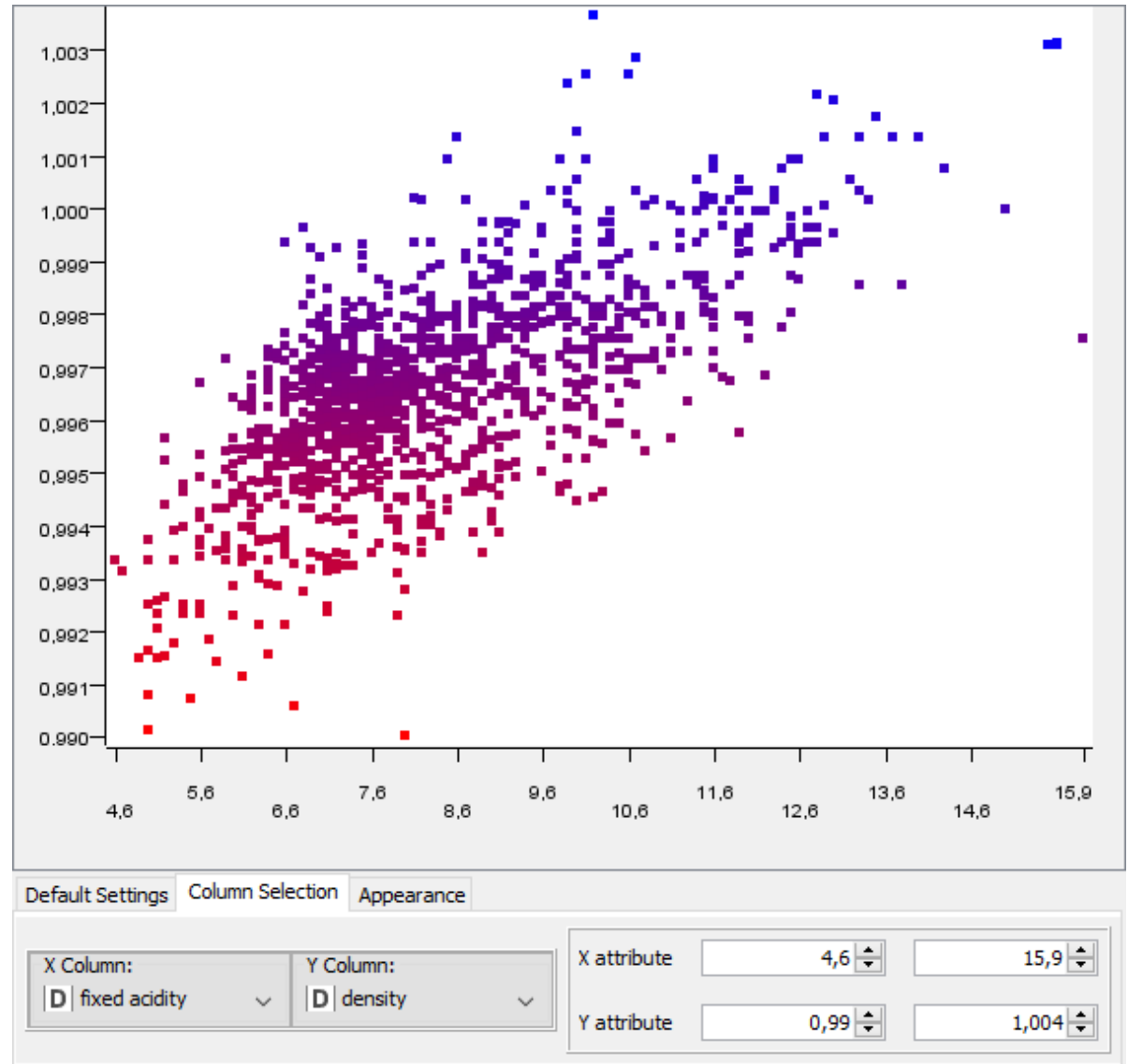
- On voit que le pourcentage de bleu augmente de la gauche vers la droite. Cela signifie que plus la valeur en alcool est importante, plus le vin rouge a de chance d'être de bonne qualité.





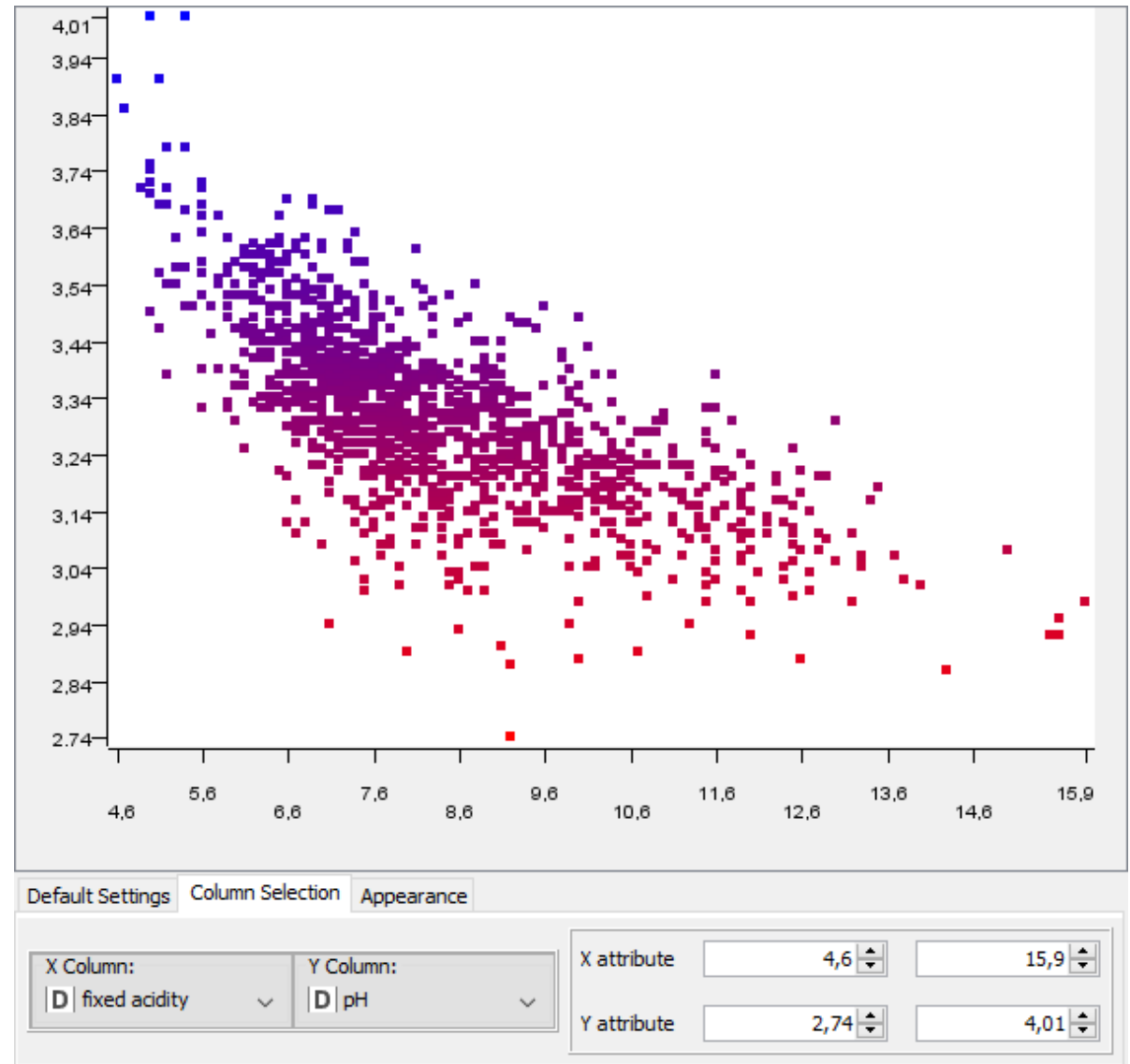
## Acidité fixe et densité

- L'acidité fixe semble augmenter avec la densité.



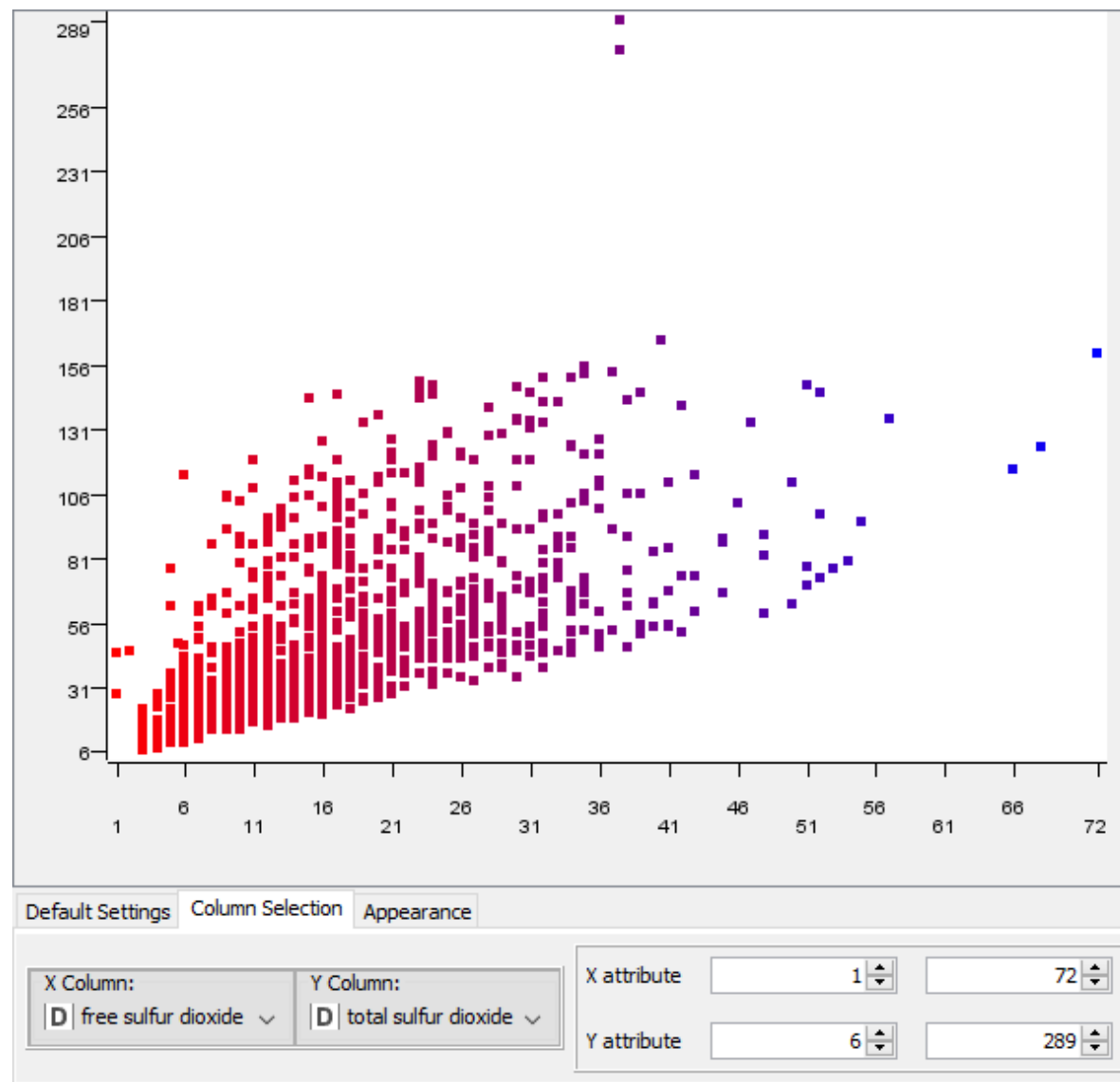
## Acidité fixe et pH

- Comme pour le vin blanc, il semblerait que quand l'acidité fixe augmente, le pH diminue.



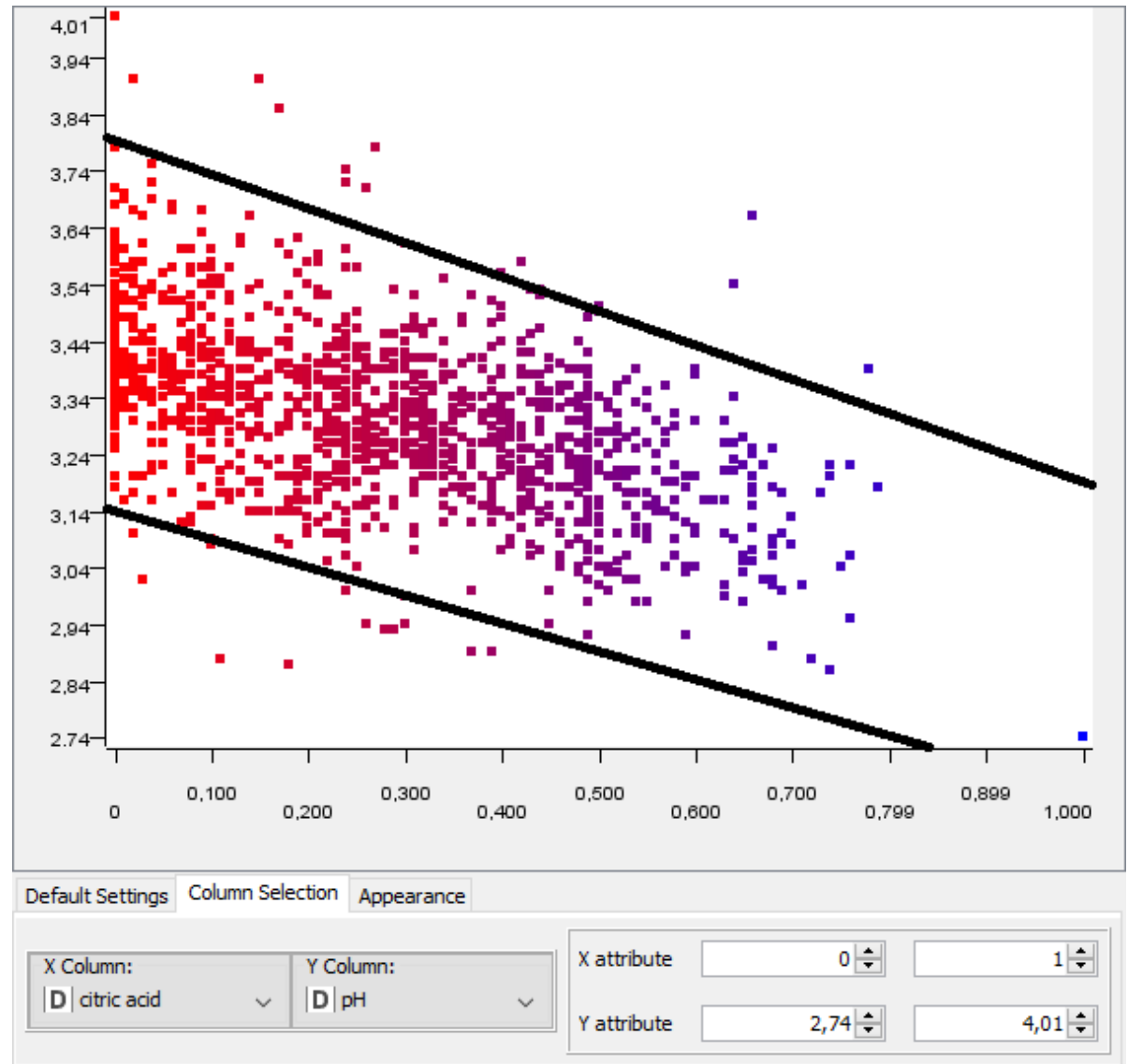
## Dioxyde de soufre libre et total

- Et encore une fois le dioxyde de soufre libre semble borné par le total.



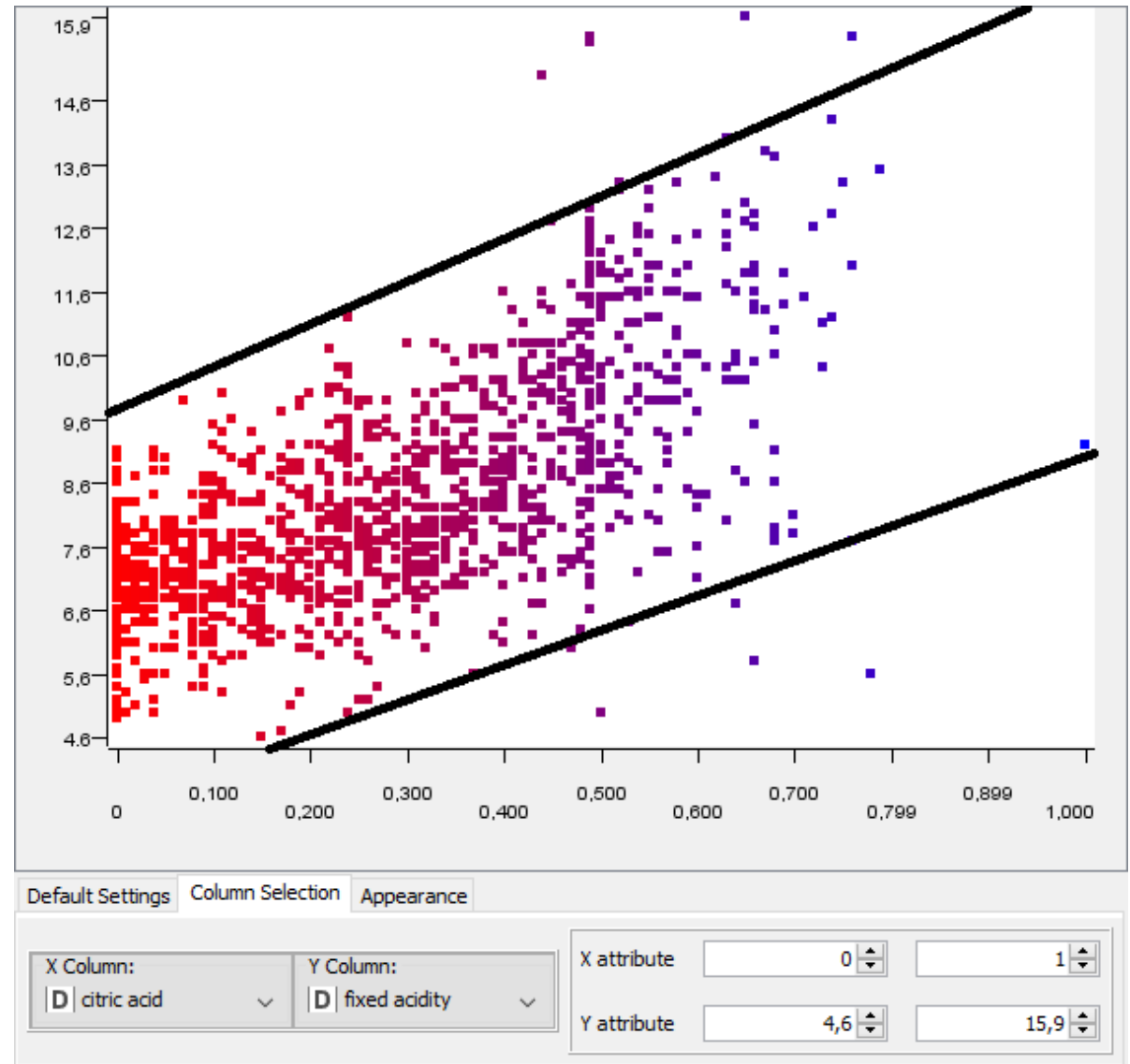
## Acide citrique et pH

- L'acide citrique aussi diminue le pH.

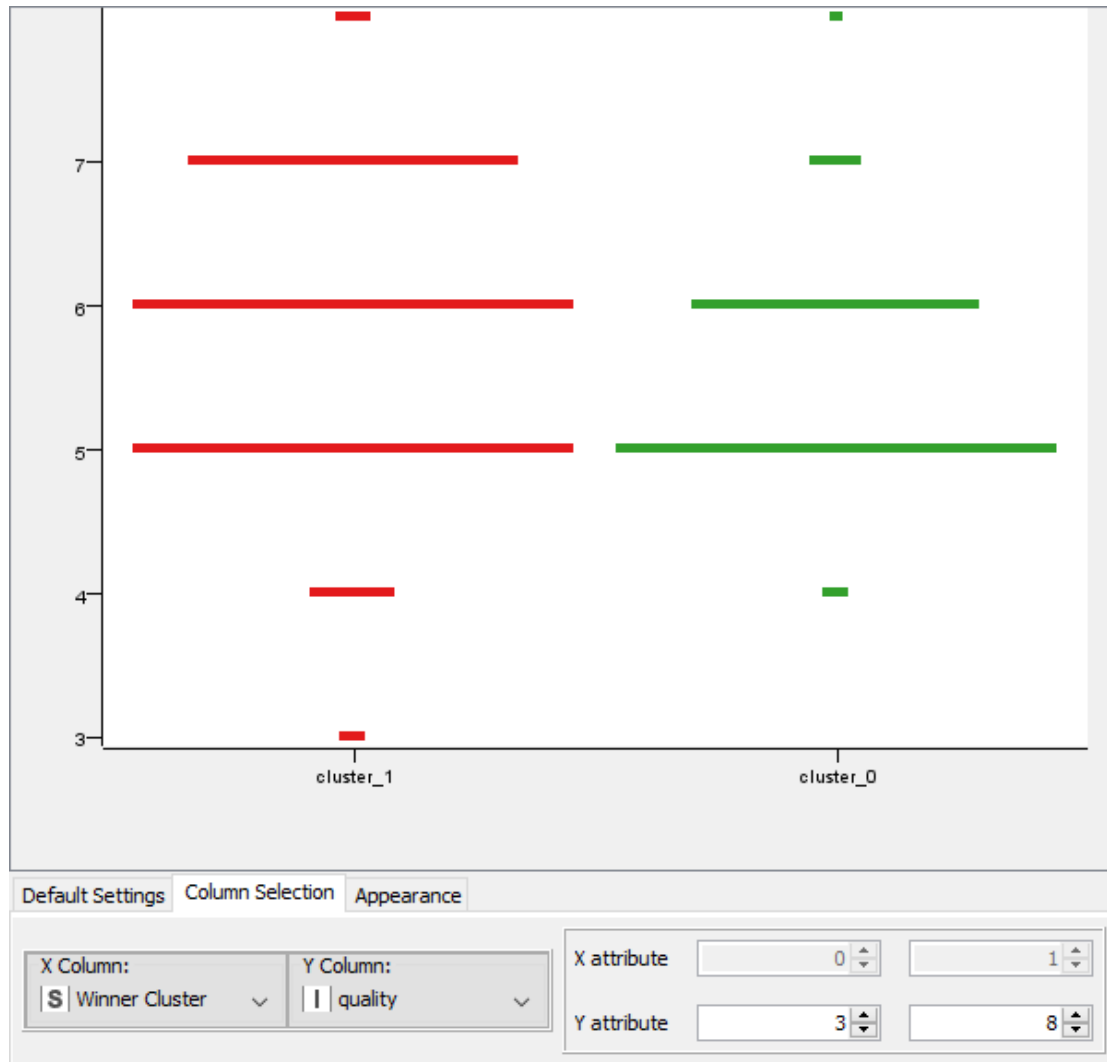


## Acide citrique et acidité fixe

- L'acide citrique semble augmenter avec l'acidité fixe.



Les deux images ci-dessous montrent un essai de cluster où on a pu extirper presque la moitié des vins moyens en ne perdant pas beaucoup de vin de qualité.



Dialog - 3:4 - Fuzzy c-Means

File

Fuzzy c-means Used Attributes Flow Variables Memory Policy

Exclude

Filter

- D fixed acidity
- D volatile acidity
- D citric acid
- D chlorides
- D pH
- D sulphates
- I quality

Include

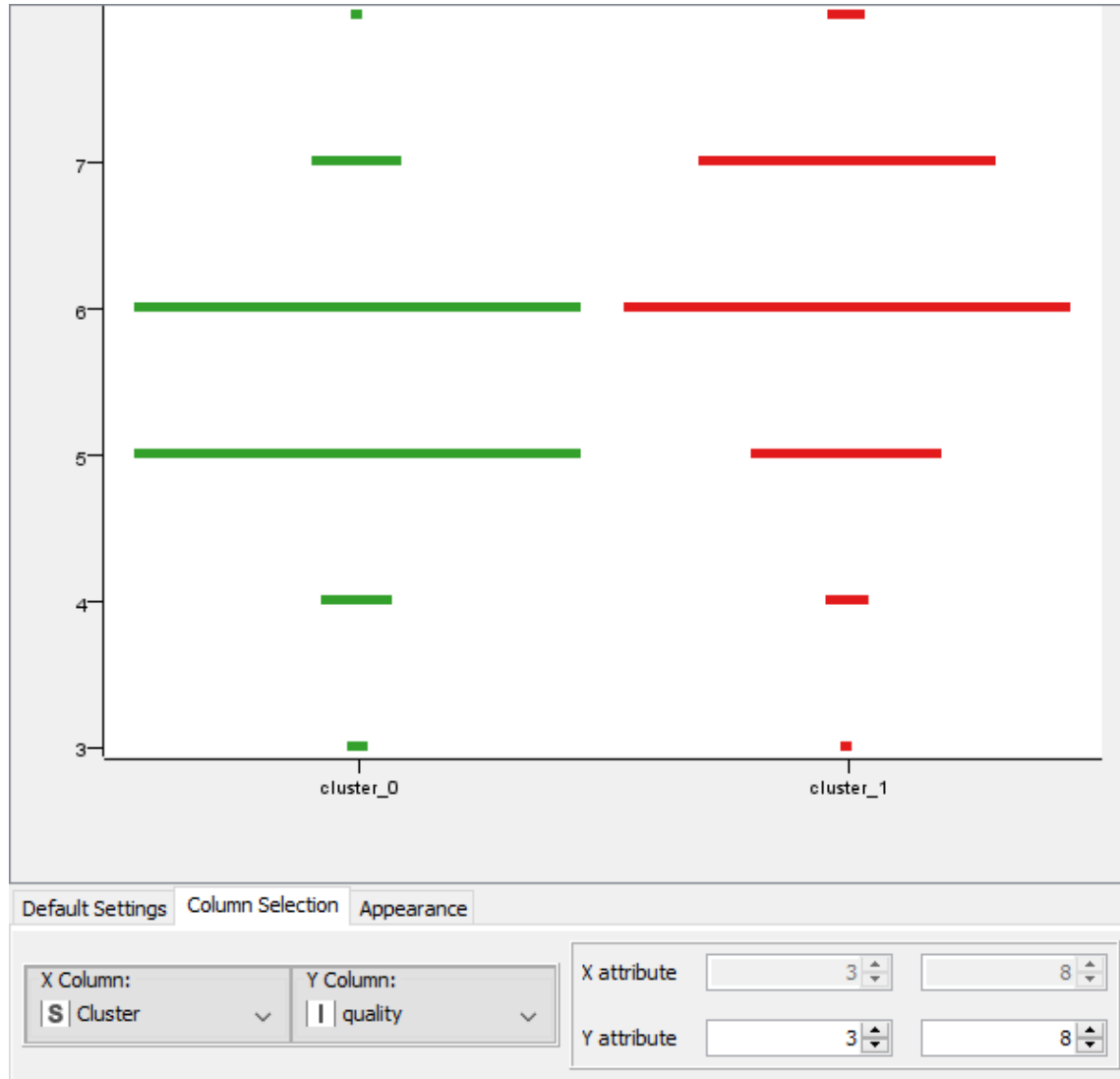
Filter

- D residual sugar
- D free sulfur dioxide
- D total sulfur dioxide
- D density
- D alcohol

Always include all columns

OK Apply Cancel ?

Ici, le même principe sauf que plus de vins ont été retirés (mais naturellement une plus grande proportion de bons vins aussi)



Dialog - 3:5 - k-Means

File

K-Means Properties | Flow Variables | Memory Policy

number of clusters: 2

max. number of iterations: 249

Exclude

Filter

- D fixed acidity
- D volatile acidity
- D citric acid
- D residual sugar
- D chlorides
- D free sulfur dioxide
- D total sulfur dioxide
- D pH
- D sulphates

Include

Filter

- D density
- D alcohol

☐ Always include all columns

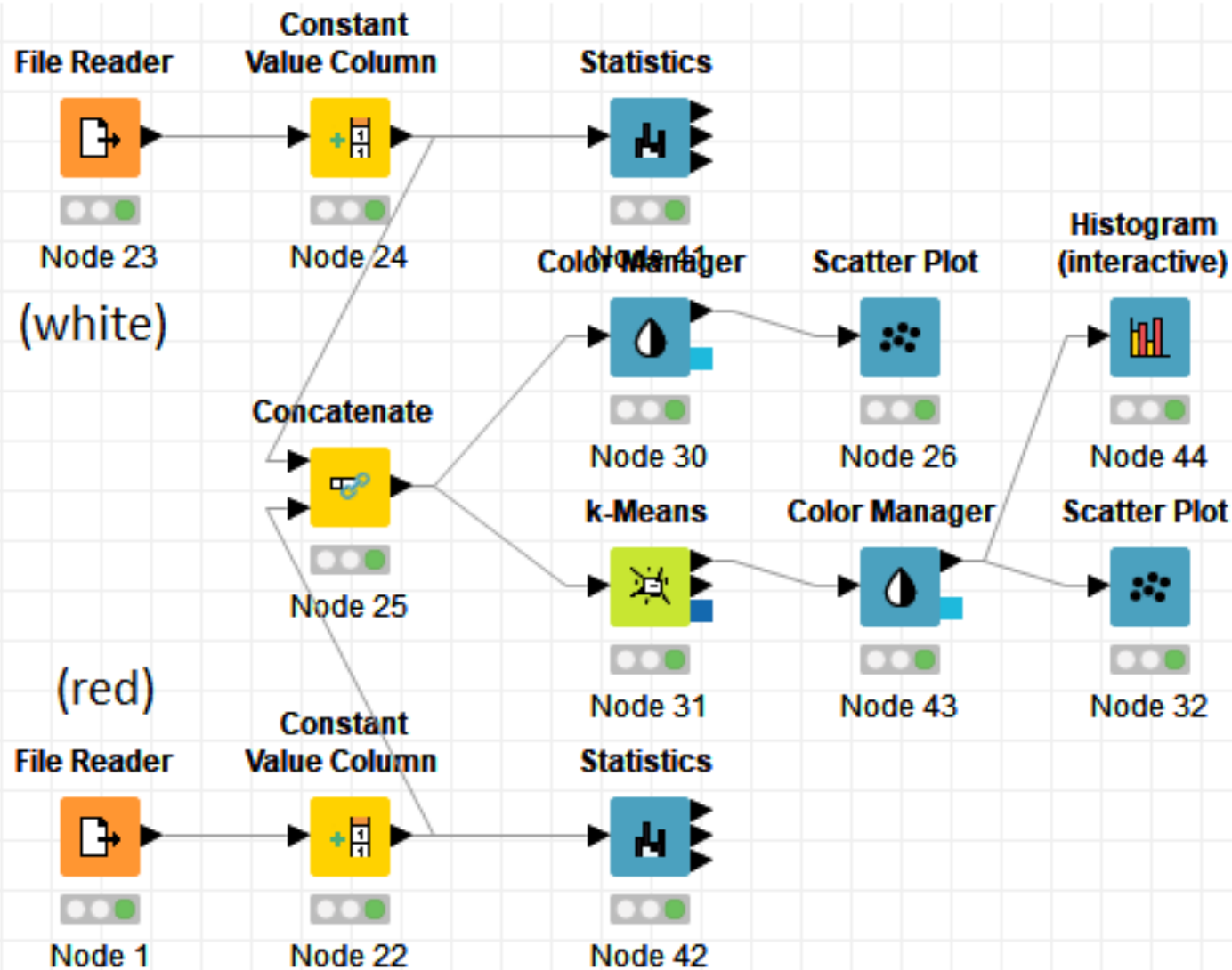
☐ Enable Hilite Mapping

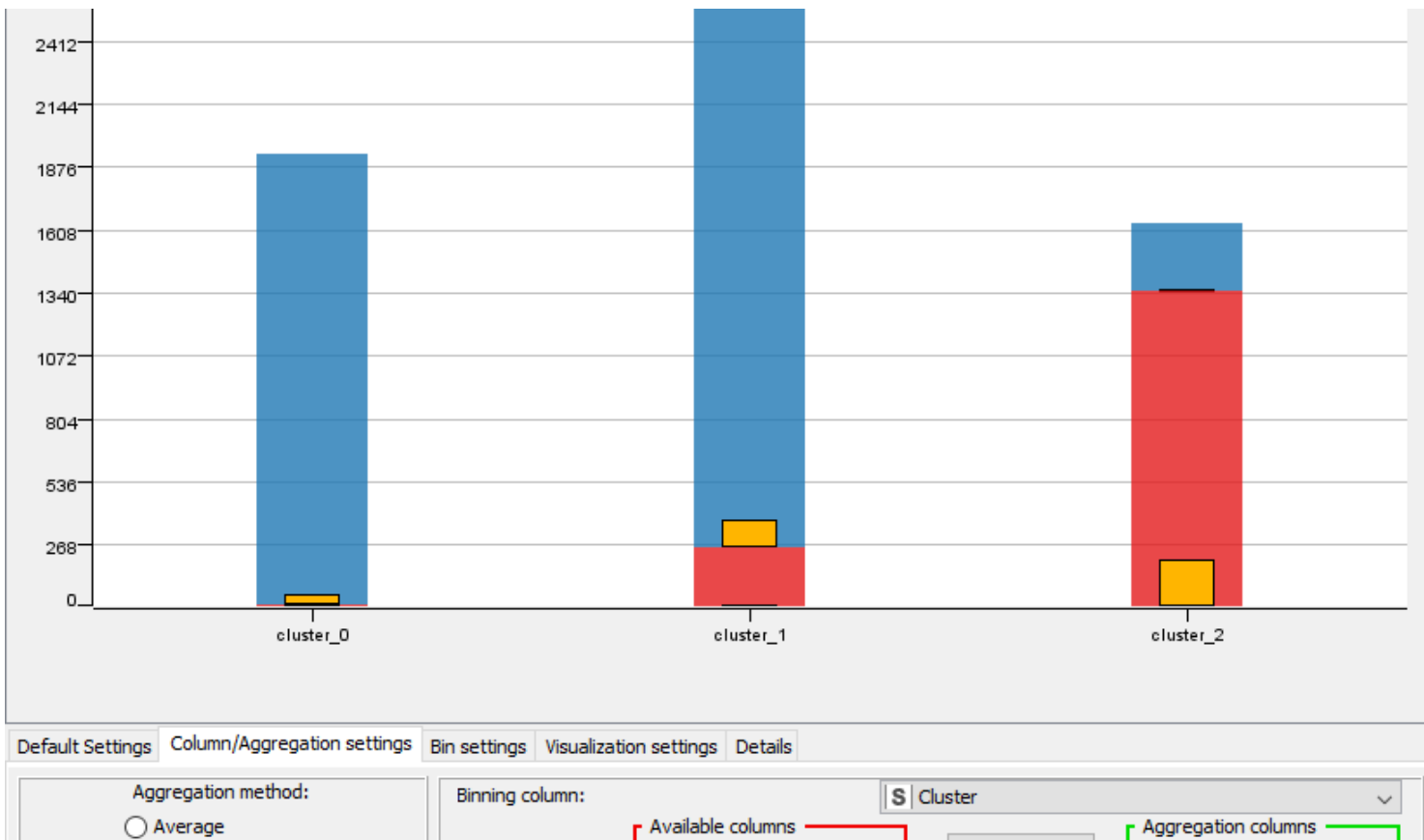
OK Apply Cancel ?



Vin rouge et blanc

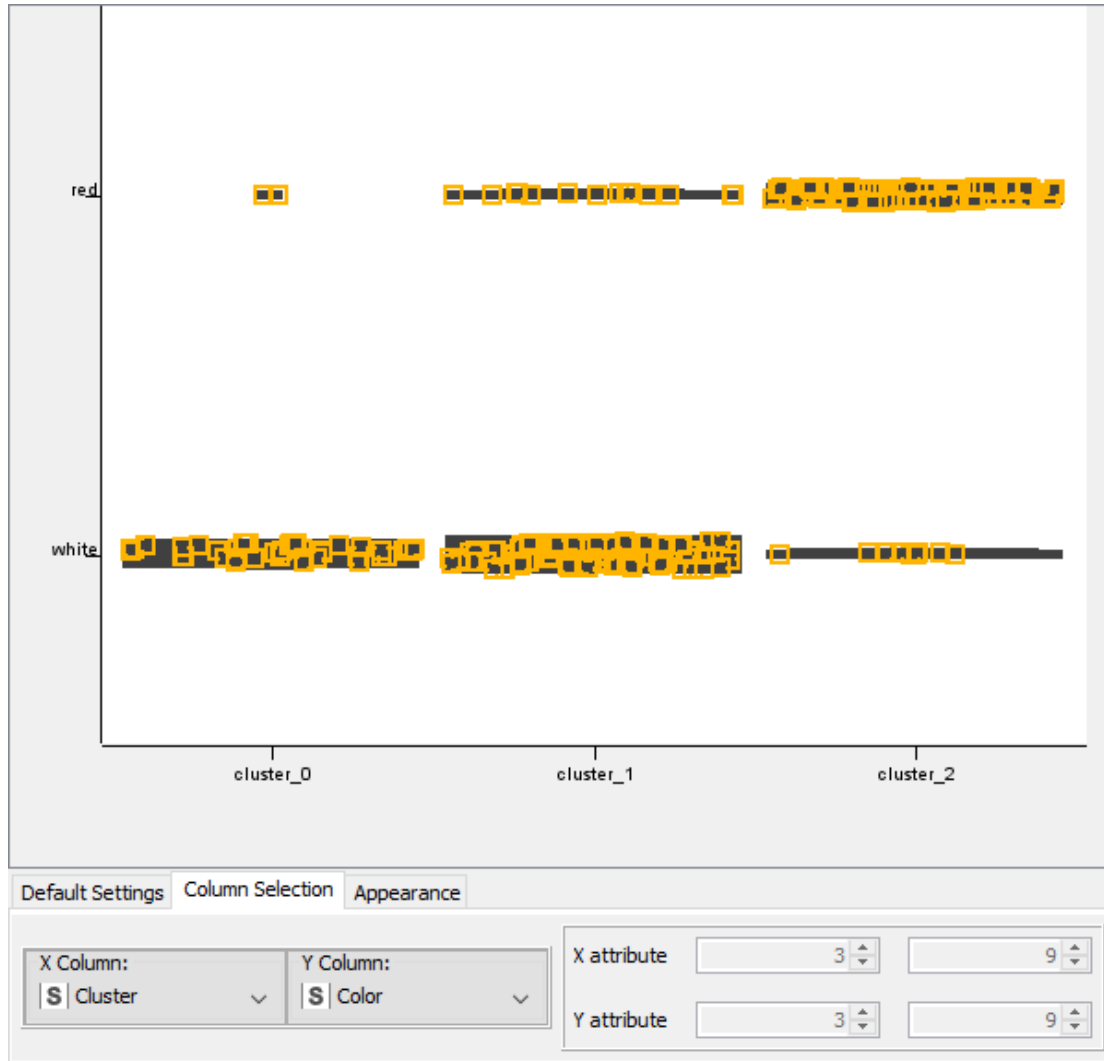






# Cluster

- Ci-joint et après, on a essayé de clusteriser les vins rouges et blancs ensemble, voici un resultat.
- On notera que en jaune ce sont les très bons vins. En rouge les vins rouges et en bleu les vins blancs.
- De plus on remarque qu'il vaut mieux pour un vin blanc de ne pas appartenir au cluster 2 et pour les vins rouges c'est au 1 car on a moins de chance d'être bon dans ces cas.
- En somme on remarque qu'on peut clusteriser les vins et prédire avec un certain degré de certitude si il sera bon ou pas. Et ceci dépendra de sa couleur.
- Mais on remarque aussi qu'on ne peut pas être très précis.



Dialog - 2:31 - k-Means

File

K-Means Properties Flow Variables Memory Policy

number of clusters: 3

max. number of iterations: 249

Exclude

Filter

- D fixed acidity
- D volatile acidity
- D citric acid
- D residual sugar
- D chlorides
- D free sulfur dioxide
- D pH
- D sulphates
- I quality

Include

Filter

- D total sulfur dioxide
- D density
- D alcohol

Always include all columns

Enable Hilite Mapping

OK Apply Cancel ?

## Note a propos des différents clusters:

Le k-means, le fuzzy c-means et le hierarchical clustering donnent les mêmes résultats si on selectionne les même variables.

# Conclusion

Pour conclure on aimerait indiquer que certaines valeurs obtenues sont cohérente avec nos connaissances (pH, acid, ...).

L'alcool et la densité semblent jouer un rôle clé dans le gout du vin.

Néanmoins s'il était si facile de savoir ce qu'il fallait pour faire un bon vin les viticulteurs ne s'appliqueraient pas à du "mauvais" vin.