

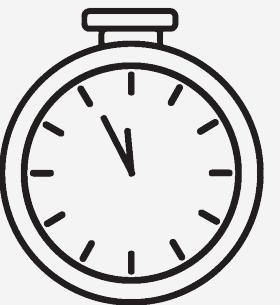


CODE Analytics

# Exploring Compositions of Red and White Wines

in collaboration with  
**CODE Analytics**





# Agenda

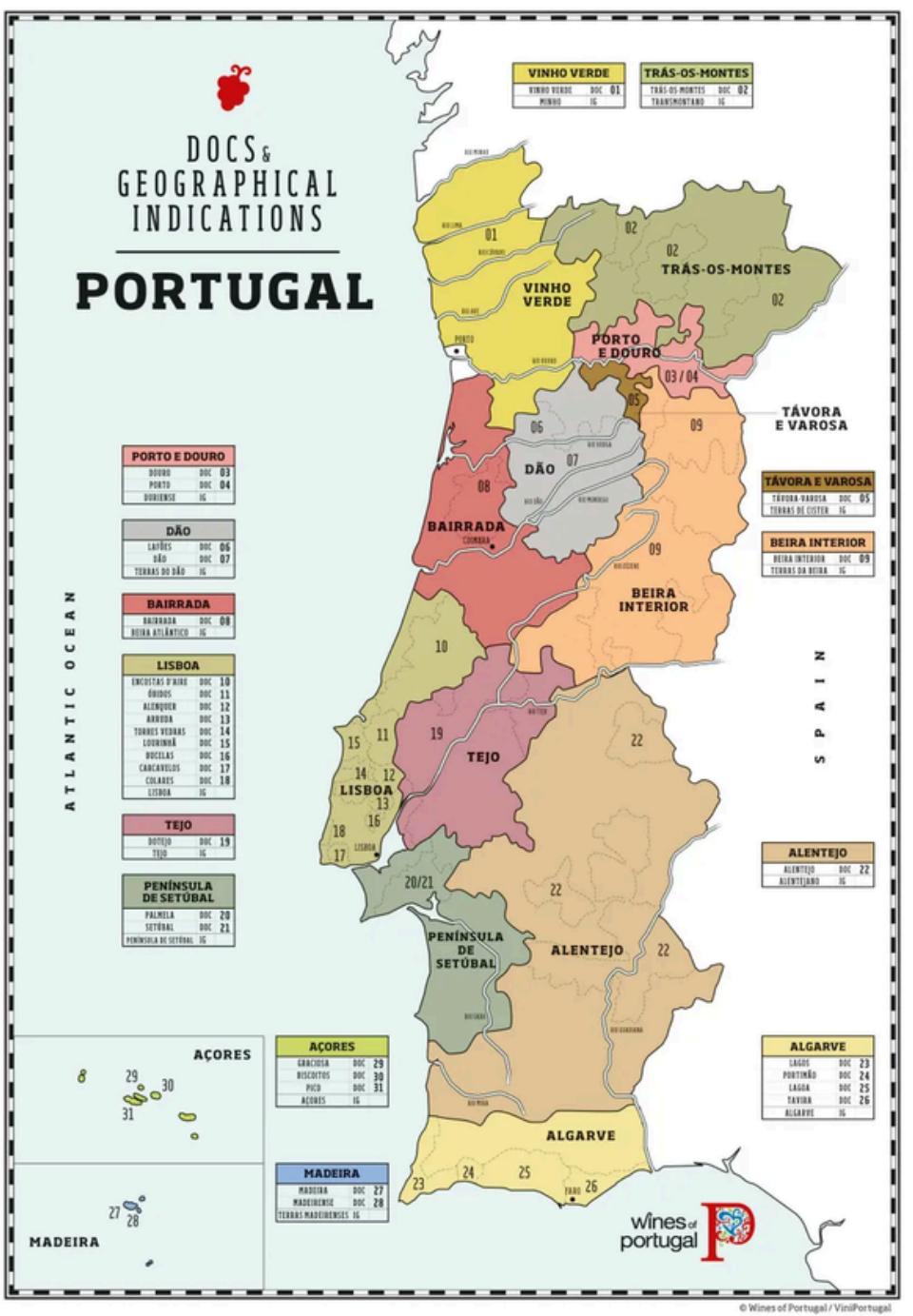
- **Vinho Verde**
- Wine Domain Knowledge
- **Industry Analysis**
- Wine Composition
- **Low vs. High Quality Red Wine**
- Low vs. High Quality White Wine
- **The Data**
- Sulfur Dioxide Trends
- **Sugar Trends**
- Volatile Acid Trends
- **Alcohol Content Trends**
- Problems and Proposed Solutions
- **Next Steps**
- Questions



# BlueBerry Winery

2

# Vinho Verde



# Grape Types



# White Wines

# Rosé/Red Wines

Loureiro

# Vinhão

# Arinto

# Amaral

# Trajadura

# Azal Tinto

# Avesso

# Espadeiro

# Azal

## Source



BlueBerry Winery



3

# Wine Domain Knowledge



## Vinho Verde

- Wine from the dissolved, northern **Minho province**
- A **DOC**: mandating to make wine within **3-6 months after harvest**
- Often slightly **fizzy**

## White Wines

- **Alvarinho** and **Loureiro** sell the best
- Alvarinho's produce low yields but reach much higher alcohol levels

## Red/Rosé Wines

- **Vinhão**, then **Amaral / Azal Tinto** and **Espadeiro** perform best
- These produce deep purple coloring and **peppery notes**

[Source](#)



# Defining **Red** and White Wine

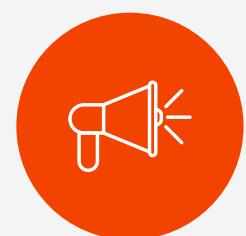
**Stats - Facts - Analysis**

[BACK TO AGENDA PAGE](#)



# Industry Analysis

84.9 million litres in 2022



**Sample Size  
4898**

**White Wine**



**Sample Size  
0**

**Rosé Wine**

**Not included** in this analysis, although a promising market player.



**Sample Size  
1599**

**Red Wine**

Briefly elaborate on what you want to discuss.



**N.2 Market:**  
Bulk-order  
Supermarkets



**Fastest Market:**  
34% Growth  
3.8-5.2 Mil, 2022



**N.1 Market:**  
>10Mil Litres/year

[BACK TO AGENDA PAGE](#)

**Source**

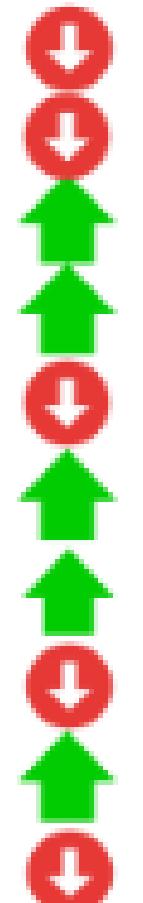


# Wine Composition

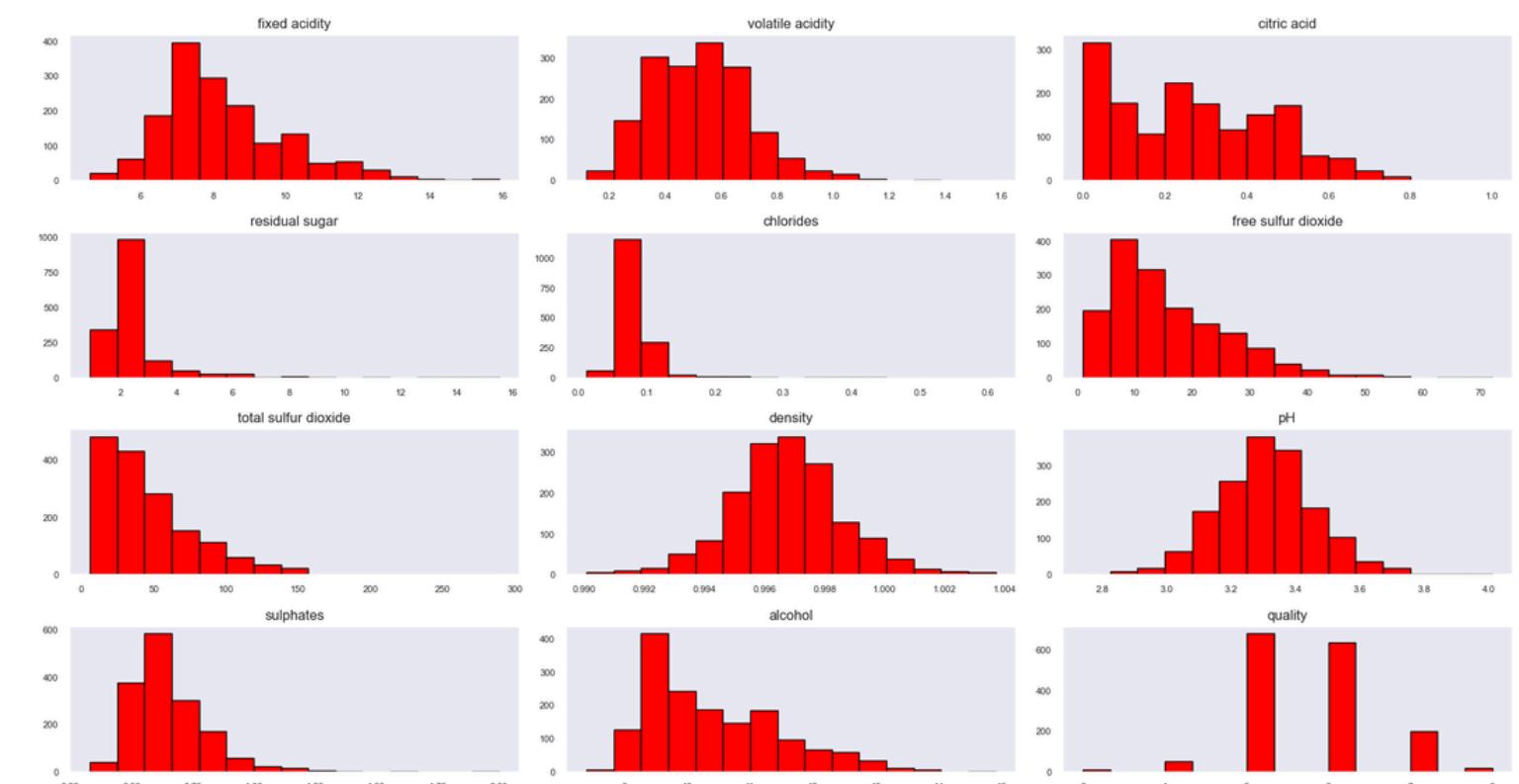
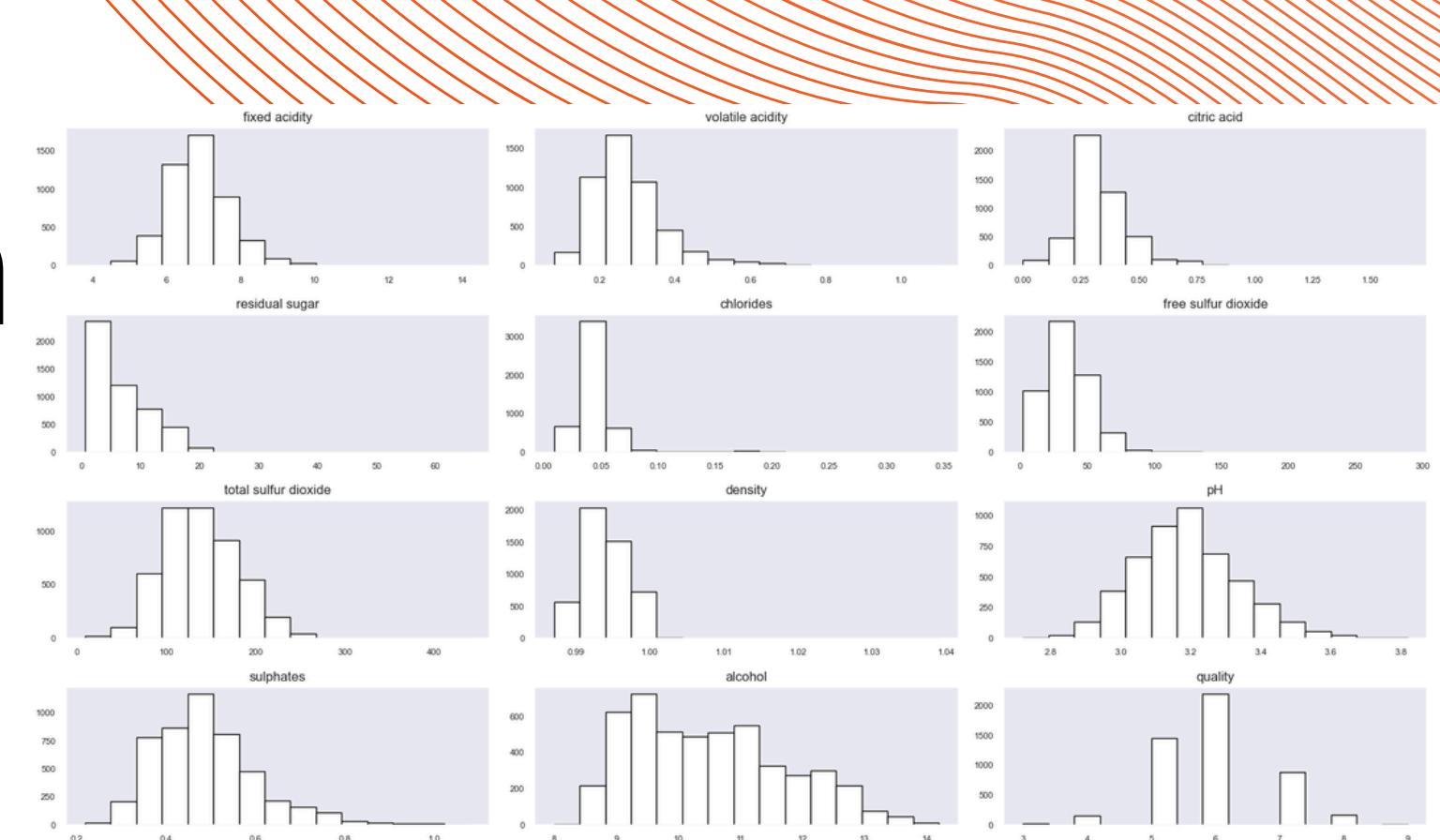
## Properties

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- chlorides
- free sulfur dioxide
- total sulfur dioxide
- density
- sulphates
- alcohol

## White Wine



## Red Wine



[BACK TO AGENDA PAGE](#)



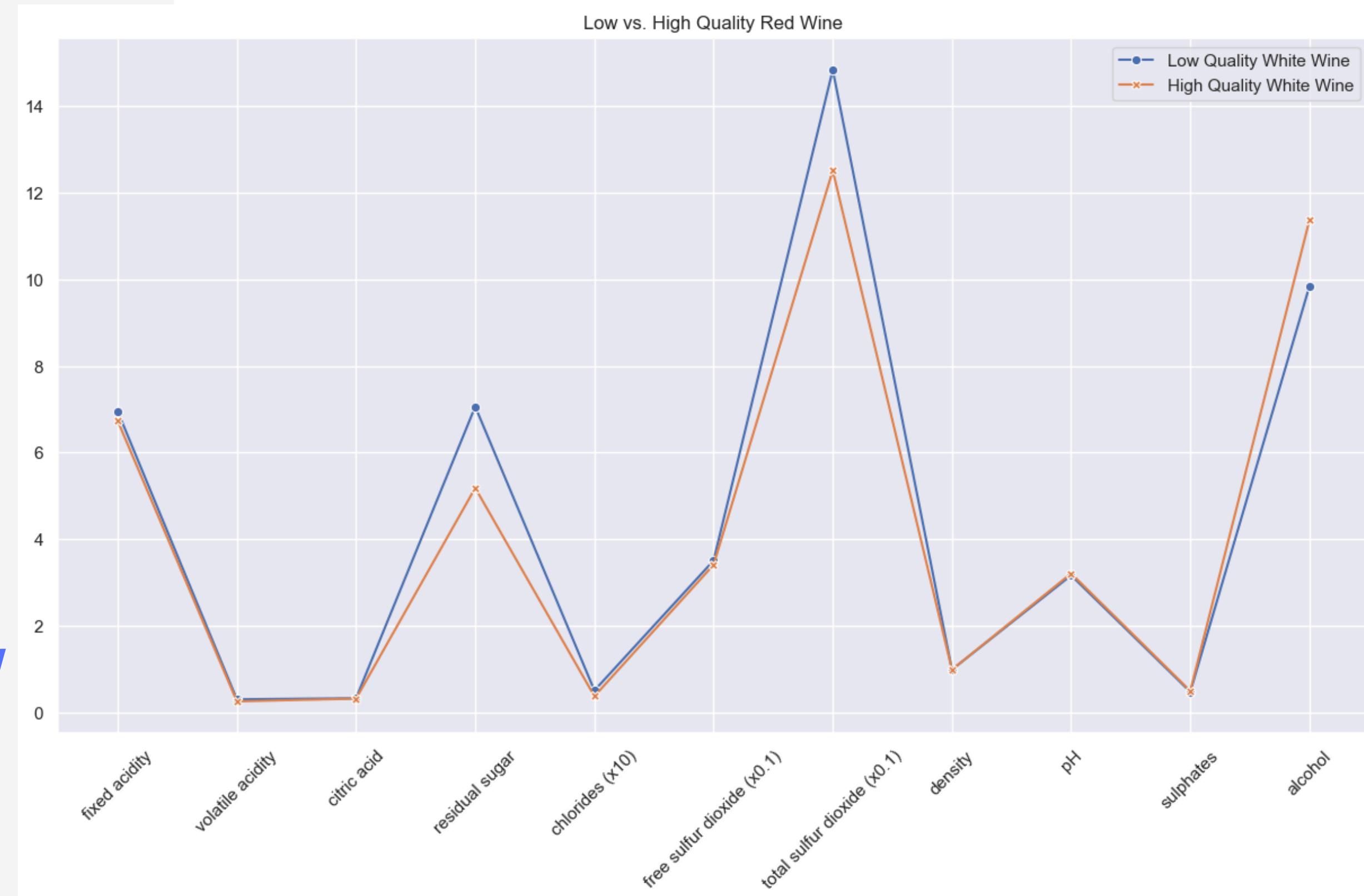
# Low vs. High Red Wine

Orange: High

Blue: Low

Good Red Wine has much less sugar and less total sulfur dioxides. High levels of alcohol do better.

[BACK TO AGENDA PAGE](#)





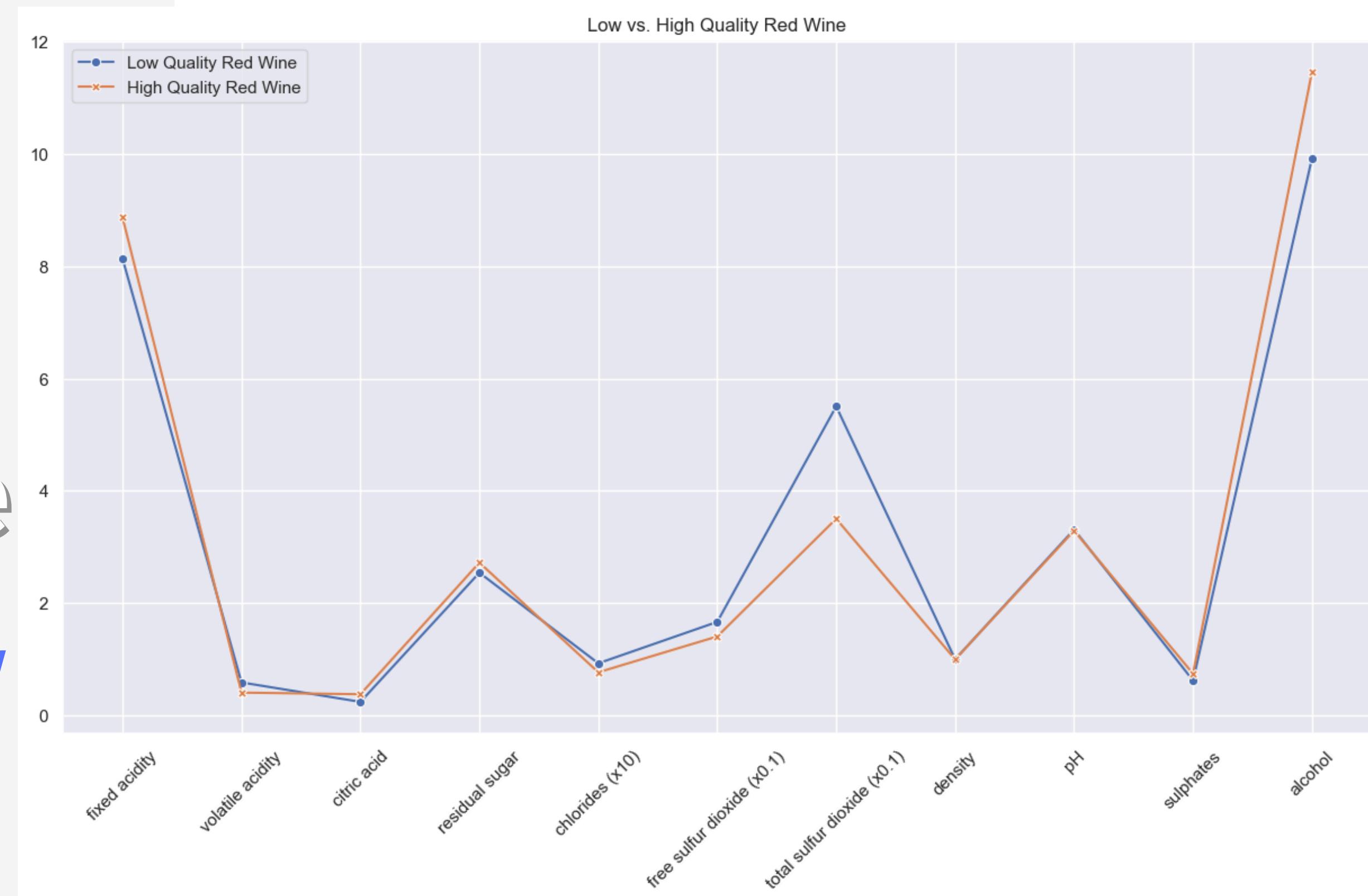
# Low vs. High White Wine

Orange: High

Blue: Low

Good white wine is a little more acidic and also has less total sulfur dioxides. More alcohol always helps.

[BACK TO AGENDA PAGE](#)





# The Data

## Low vs. High Quality Red & White Wine

	Fixed Acidity	Residual Sugar	Total Sulfur Dioxide	Alcohol
Low Red				
High Red				
Low White				
High White				

[BACK TO AGENDA PAGE](#)

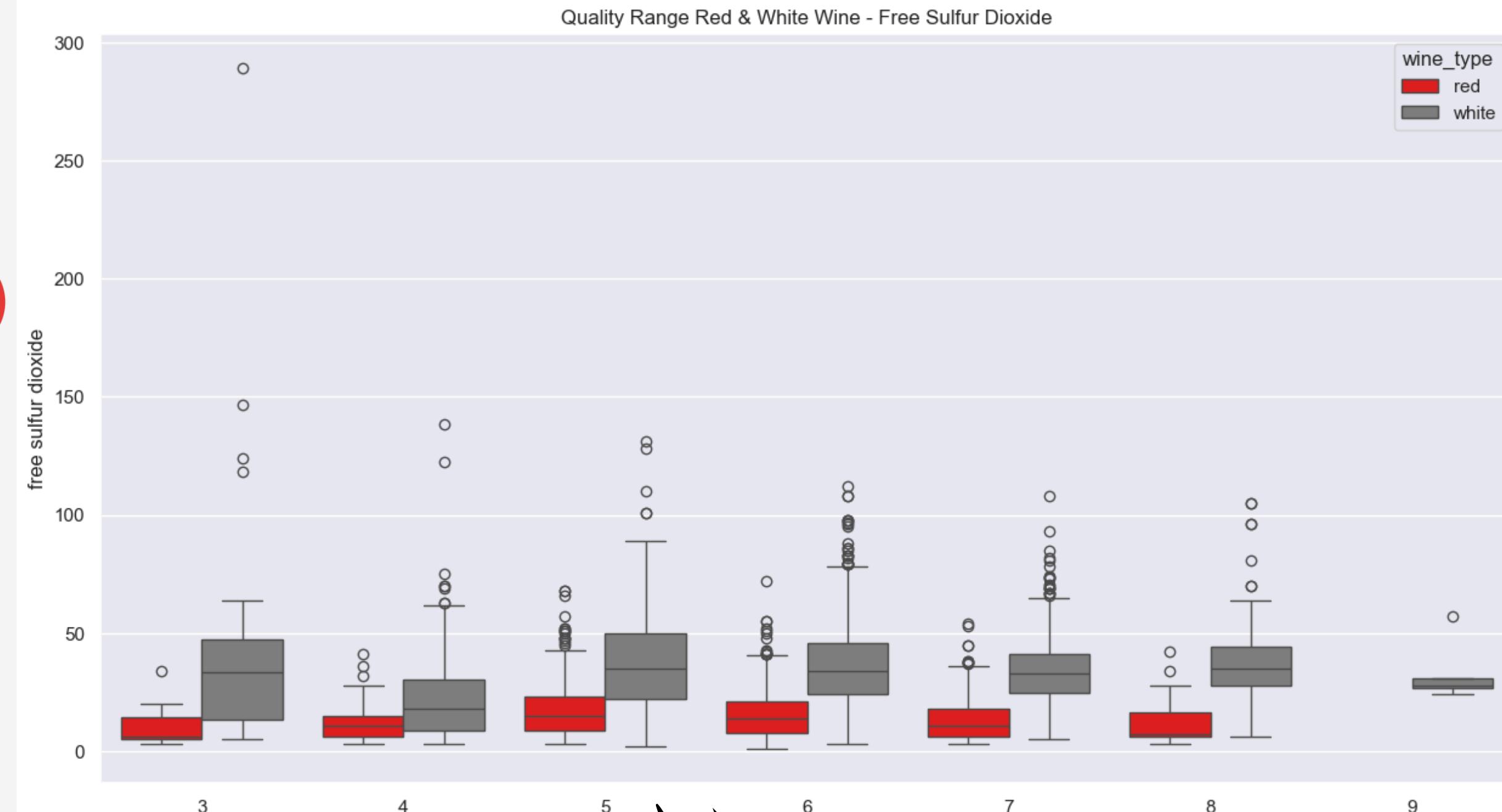


# Sulfur Dioxide with Quality

Red: Red Wine Grey: White Wine

White wine varies significantly more in sulfur dioxide content (note variance).  
Red Wine has much lower levels.

[BACK TO AGENDA PAGE](#)



quality  
significant for White:  
p value: 0.0025  
Not for Red:  
p value: 0.165



# Sugar with Quality

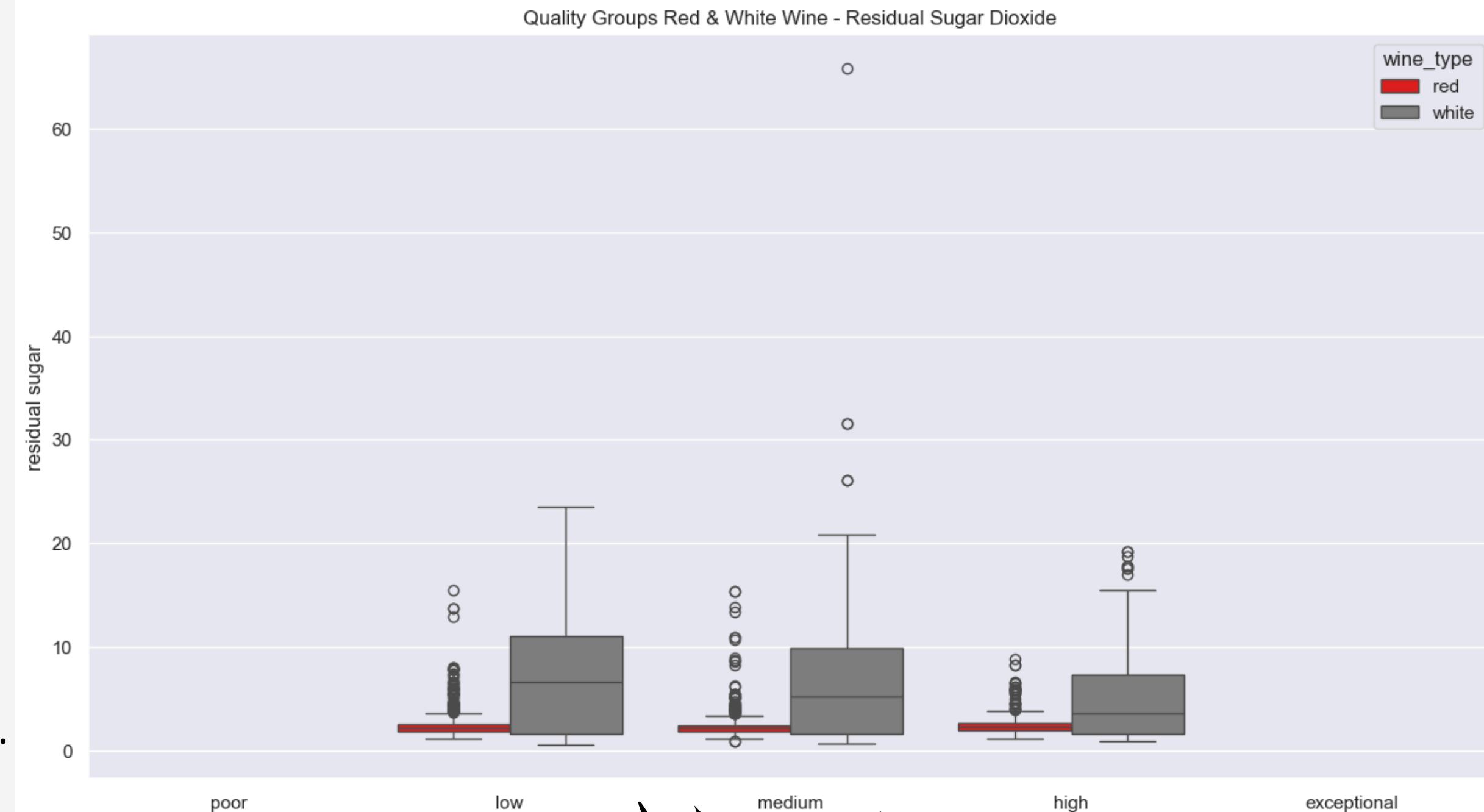
Red: Red Wine Grey: White Wine

Red wine barely varies in sugar content.

White wine has more sugar.

Variance  for both high end wines.

[BACK TO AGENDA PAGE](#)



NOT for White:  
p value: 0.105  
NOT for Red:  
p value: 3.25



BlueBerry Winery



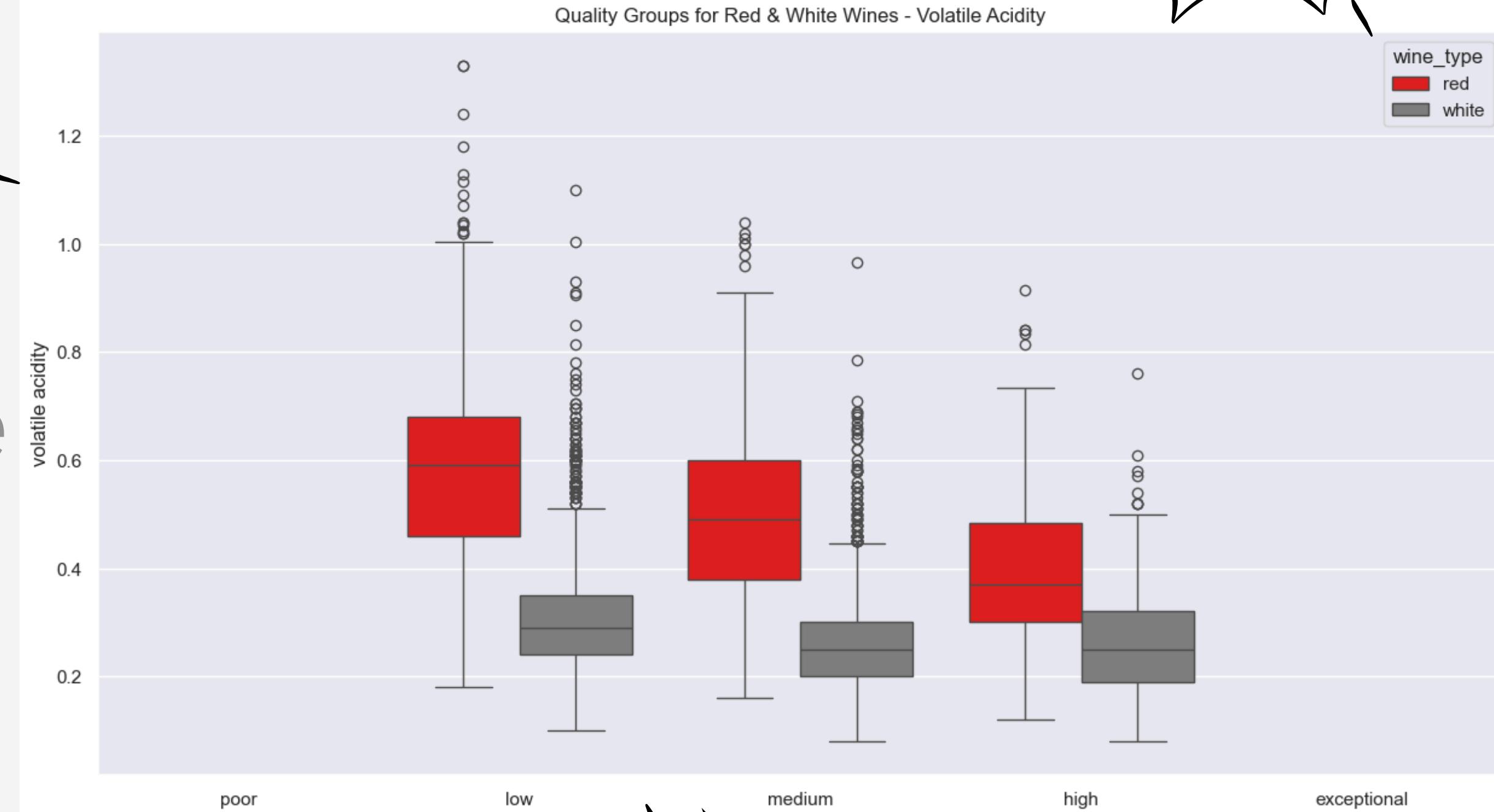
12

# Volatile acidity with Red Wine Quality

Red: Red Wine Grey: White Wine

Mean volatile acidity differs across different wine qualities, variance overlap is far too great to infer a trend.

[BACK TO AGENDA PAGE](#)



F - Statistic for Red:  
242.72 - but not significant

NOT for Red:  
p value: 2.961  
Red variance too large!



BlueBerry Winery



13

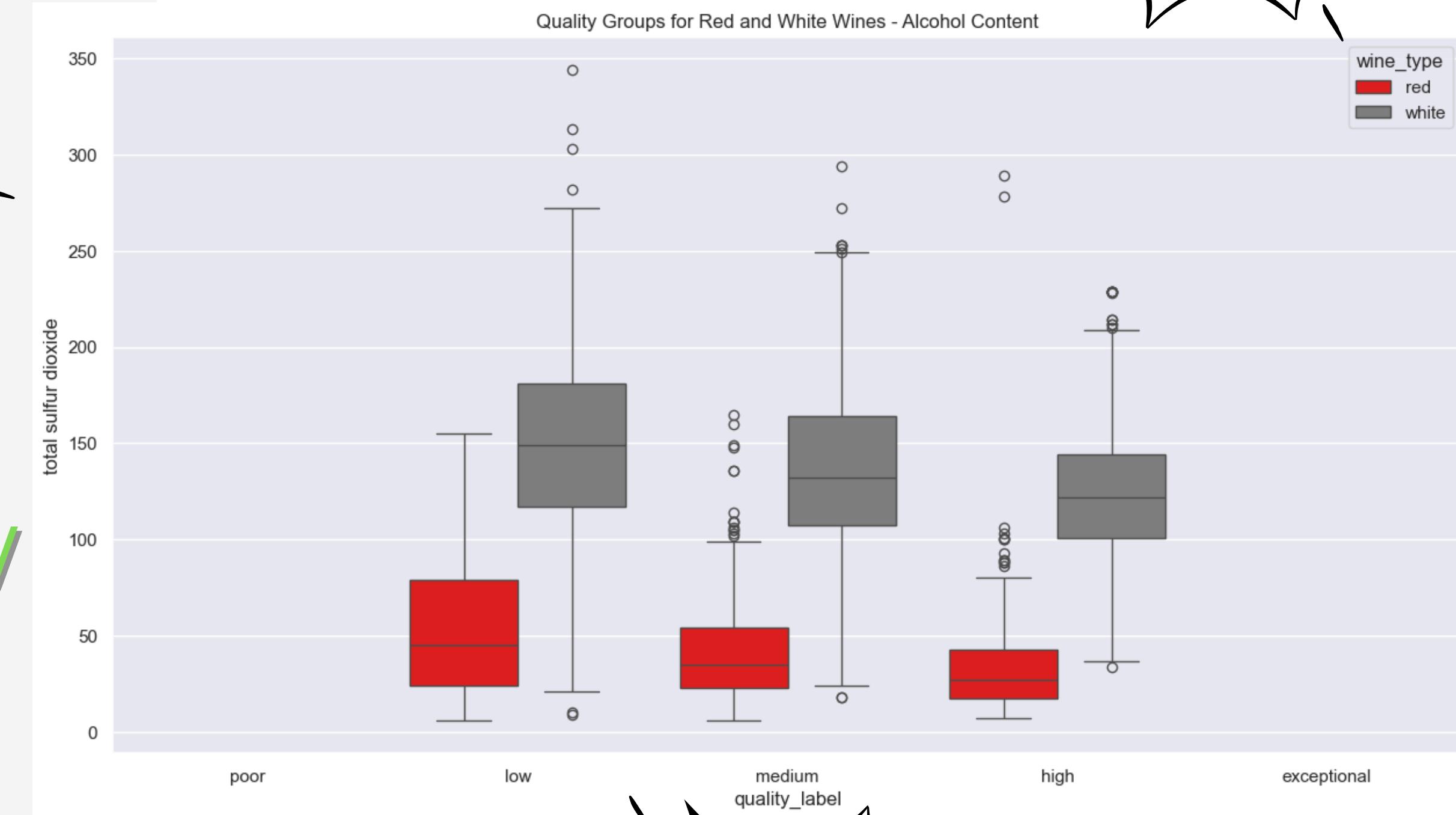
# Alcohol ↓ with Red & White Wine Quality

Red: Red Wine Grey: White Wine

Mean alcohol decrease seen, but variance overlap is far too great, data is **not significant**

[BACK TO AGENDA PAGE](#)

**False correlation**



NOT for **BOTH**:  
p value :3.29, 6.58  
**not significant**

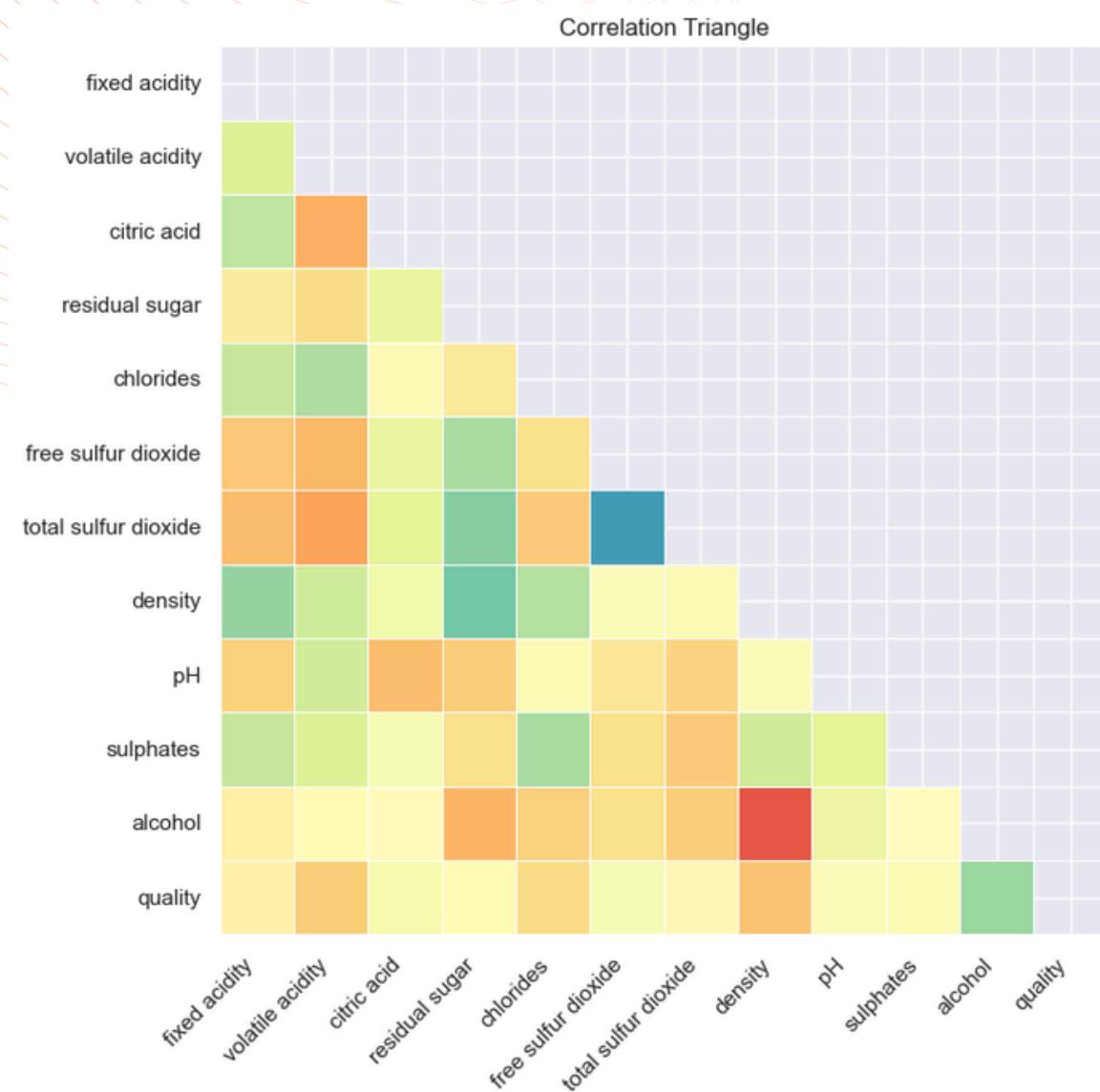
F - Statistic  
**177.67, 48.09** -  
Variance is  
too large!



# Conclusion Problems and Proposed Solutions

[BACK TO AGENDA PAGE](#)

Problems	Solutions
There was a massive <b>disparity</b> in the <b>datapoints</b> provided for <b>White</b> wine and <b>Red</b> wine samples	<b>Normalize the data</b> or generate <b>synthetic data</b> based on the parameters of the provided samples
<b>F Statistic</b> and <b>P values</b> for the data tended to be <b>unacceptable</b> due to the nature of the <b>data points</b>	<b>Investigate outliers</b> and set boundaries for exclusion/ <b>weigh</b> models accordingly to account for their <b>statistical impact</b>
<b>Red</b> and <b>White</b> wine are different products and different <b>compositions</b> , and are <b>regionally</b> dependent due to <b>grape species</b>	Make a <b>business decision</b> on <b>White or Red</b> wine market - <b>Compare</b> either red or wines based on <b>grape species</b> and <b>reinvestigate</b> data



## Conclusions

- White wine is the more established market & is showing the **steady annual growth**
- The northern Portuguese region is best suited for **white wine** production
- Preliminary data, although skewed, indicates that **quality/price prediction** based on wine composition is **possible**

## Objectives

- Collect data on the **performance** of wines based on **grape type**
- Combine **machine learning** with **sales data** of wines to **predict price** margins based on the wine composition of wine **products**

Next Steps:  
Machine Learning

# Thank You For Your Attention

## Questions?

Leon Breitenborn

leon1breitenborn@gmail.com

Junior Data Analyst - Analytics & Research

by your **CODE Analytics Team**

