Red Wine Analysis

Data:
The dataset of interest pertains to Red Wine variants of Portuguese Vinho Verde wines, which I sourced from Kaggle. My choice to work with this particular wine dataset is motivated by several factors. Firstly, the dataset itself offers a valuable opportunity for exploration and analysis due to the diverse range of variables it encompasses. These variables include fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulfates, alcohol, and quality. In my wine analysis, I have chosen to focus primarily on three key variables: fixed acidity, volatile acidity, and pH, along with the overall quality of the red wine. The rationale behind this selection is rooted in the shared characteristics of these variables and their importance in the context of wine quality. Fixed acidity, volatile acidity, and pH are all fundamental attributes that significantly influence the sensory experience and overall quality of red wine. Fixed acidity contributes to the wine's structure and stability, impacting its taste and mouthfeel. Volatile acidity, on the other hand, influences the wine's aroma, and a balanced pH level is crucial for the wine's stability and perceived quality. By focusing on these variables, I aim to gain deeper insights into how the interplay between acidity, pH, and other attributes contributes to the overall quality of red wine. This focused analysis will provide a better understanding of the factors that make certain wines stand out and offer valuable information for both wine enthusiasts and industry professionals.

Research Objective:
To investigate the interrelationship between varying levels of acidity, specifically fixed acidity, volatile acidity, and pH, and their impact on the perceived quality of wine.
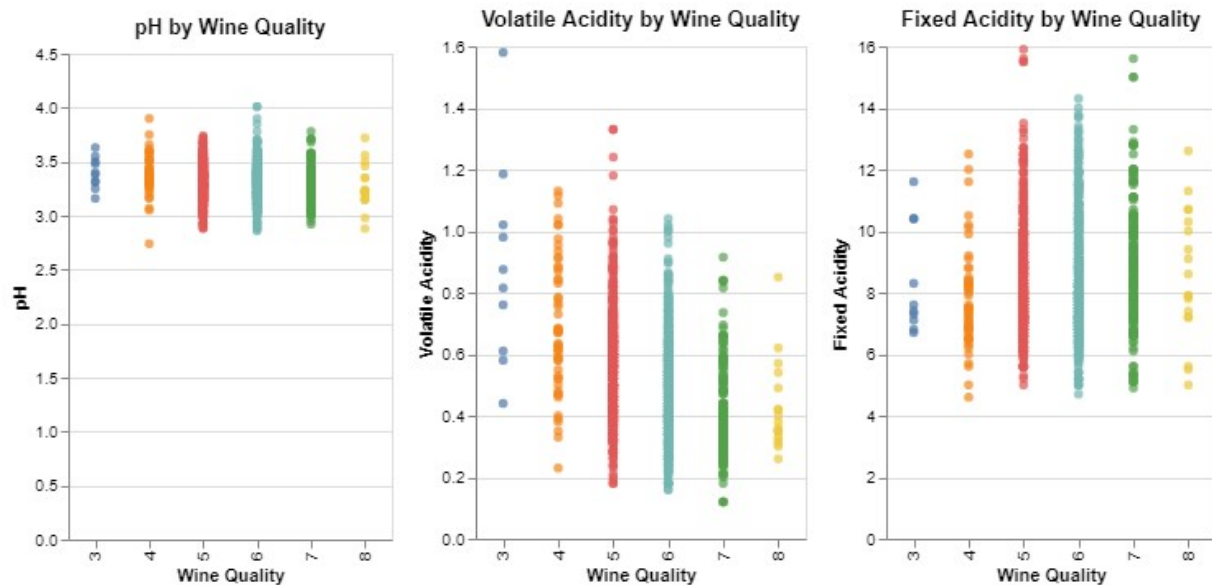
Tasks:
1. **Visual Exploration:**
   a. Create visuals to visually represent the relationship between acidity attributes and wine quality
   b. Interpet the visualizations to identify patterns
2. **Data Analysis:**
   a. Leverage created visuals as a basis to compute summary statistics
   b. Explore trends within the dataset

Primary Phase:

Initially, I developed three distinct scatter plot graphs featuring wine quality on the y-axis and various acidity levels on the x-axis. The objective behind creating these visualizations was to analyze the distribution of scatter points across the graphs for each explanatory variable in relation to the response variable, wine quality.

However, upon seeking advice and feedback from colleagues, I received a valuable recommendation aimed at enhancing the visual clarity and comprehensibility of the graphs. The suggestion was to switch the placement of the y and x-axes, aligning with the notion that it is more intuitive to position the quality of the wine on the y-axis. This adjustment not only improves the readability of the graphs but also offers a more intuitive interpretation of the data.

Secondary Phase:
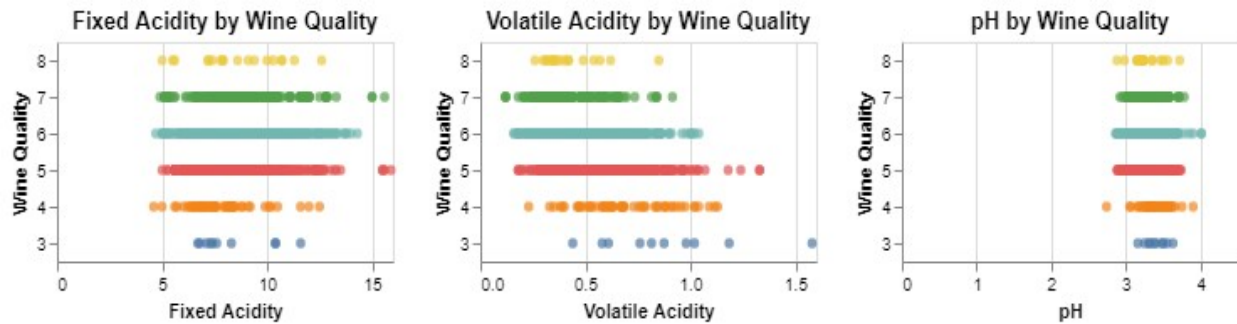
In the secondary phase of our project, I leveraged the invaluable insights gained from previous recommendations to instigate a series of enhancements aimed at perfecting the visual representation of our data. A pivotal and transformative adjustment involved reconfiguring the placement of wine quality from the X-axis to the Y-axis. This strategic shift was created, keeping in mind the goal of elevating the intuitive appeal of our graphs while deepening the interpretative value of our dataset.

The decision to embrace this new approach was influenced by insightful conversations with fellow students. Through these interactions, I gained a fresh perspective and realized the power of aligning our data visualization with the audience's cognitive expectations. This shift, inspired by our discussions, proved to be instrumental in making the data more approachable and comprehensible.

Furthermore, as part of our ongoing efforts to refine the visual representation, I introduced a faceting approach to our graphs. This adjustment enabled a side-by-side display of multiple graphs, contrasting with the previous method where they sequentially appeared during code execution. Beyond enhancing the overall visual presentation, this change considerably simplified the process of comparing and contrasting different aspects of our data.

Wrap-up Phase:
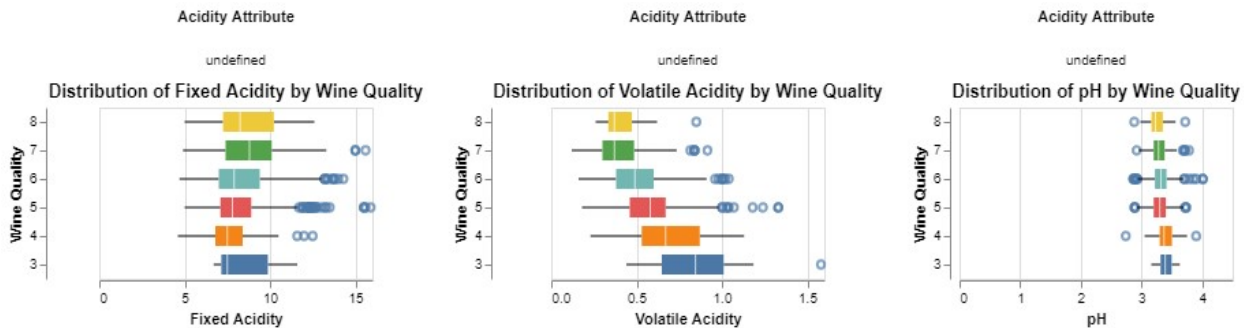
Going back to the insightful suggestions I received from my classmates, it prompted me to reconsider the visualization approach. Upon reflection, it became apparent that a more suitable approach would involve crafting similar-styled graphs employing boxplots. This alternative approach opens up the opportunity to delve into essential summary statistics such as the median, minimum, maximum, and the interquartile range (IQR), allowing for a more comprehensive understanding of the data.

Furthermore, I've incorporated an interactive feature, despite the unfortunate limitation that Google Docs does not support interactive SVGs. This interactive element allows users to hover over the boxplot, delivering a dynamic presentation of the boxplot's critical statistics. The intention behind this interactivity is to foster greater user engagement and enhance comprehension of the dataset, thus offering a more enriching and informative experience.

Key Elements of Design:
● For the final visualization, I decided that the visualization that best answers my research question takes on the form of a histogram
● Color scheme was chosen to differentiate the different wine quality
● Having my y-axis be wine quality and the x-axis being one of the three acidity groups of interest works best
● Implemented faceted plots to facilitate side-by-side comparisons, enhancing user-friendliness in visualizing the data

Potential Future Improvements:
In the event of a subsequent iteration, there are several enhancements that could be explored. Firstly, presenting findings through alternative measures, as SVG output is not supported in Google Docs, could be an area to consider. Furthermore, a more in-depth analysis could be undertaken, including the examination of correlation coefficients and the implementation of regression lines to vividly illustrate the intricate relationship between wine quality and acidity levels.

Conclusion:

These refinements not only elevate the quality of our data visualization but also create a more user-friendly experience. By offering enhanced visual representations and a seamless user experience, our data analysis toolkit empowers users to delve deeper into the nuances of red wine quality, providing valuable insights to both enthusiasts and industry professionals.

Our focused analysis, centered around the interplay of fixed acidity, volatile acidity, and pH, has revealed intriguing patterns within the dataset of Portuguese Vinho Verde wines. When we examined the boxplots for fixed acidity by wine quality and pH by wine quality, it was apparent that there is relatively low variance between different quality levels. This observation highlights the stability and consistent nature of these attributes across various wine quality categories.

However, as we shifted our focus to the boxplot representing volatile acidity by wine quality, a distinct and compelling pattern emerged. Here, we observed a more pronounced staggered distribution, with higher-rated wines consistently exhibiting lower levels of volatile acidity. This discovery not only provides valuable insights into the factors contributing to red wine quality but also underscores the critical role of volatile acidity in shaping the aroma and sensory experience of wine.

In this data exploration, we've gained a deeper understanding of how varying levels of acidity, especially fixed acidity, volatile acidity, and pH, influence the overall quality of red wine. These findings offer a valuable resource for wine enthusiasts and industry professionals seeking to make informed decisions and refine their appreciation of the intricate world of red wine variants. Our research objective, to investigate the interrelationship between these acidity levels and wine quality, has been met with insights that will continue to enhance the appreciation and evaluation of red wines.