Analyzing Wine Quality: A Comprehensive Visualization of Red and White Wine Dataset

Noshin Farzana
CS
American International University-Bangladesh
Dhaka, Bangladesh
noshinfarzana681@gmail.com

Avijit Saha
CS
American International University-Bangladesh
Dhaka, Bangladesh
avisaha938@gmail.com

Fabliha Hasnin
CS
American International University-Bangladesh
Dhaka, Bangladesh
fabliha.hasnin17@gmail.com

ABSTRACT

In today's world, where data is playing a significant role, it's important to have easy ways to visualize and understand it. The graphical representation of information and data is known as Data Visualization. In order to make complex datasets easier for users to understand, data is presented in visual formats like charts, graphs, and maps. In many areas, such as business, research, finance, and healthcare, Data Visualization facilitates data analysis and decision-making. This research paper introduces Data Visualization techniques for analyzing wine quality of red and white wine using R language. It illustrates how insights from wine-related data can be easier to understand, when presented visually. A better understanding of the variables influencing wine quality and the winemaking process can be gained by utilizing visuals. Moreover, it will also be examined how visualizations can be used to support analysis of various wine ratings, tastes, and ingredients. The ultimate goal of this paper is to present a complete overview of Data Visualization methods specifically designed for wine quality analysis and prediction, thereby promoting informed decision-making and advancing knowledge in the field.

KEYWORDS

Data Visualization, R language, wine quality, red and white wine

IEEE Reference format:

Farzana, N., Saha, A., and Hasnin, F. (2024). Analyzing Wine Quality: A Comprehensive Visualization of Red and White Wine Dataset. *International Conference on Computing Advancements (ICCA).*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
© 2018 Copyright held by the owner/author(s). 978-1-4503-0000-0/18/06...$15.00

https://doi.org/10.1145/1234567890

1. INTRODUCTION

Wine, a beloved beverage cherished for its diverse flavors and cultural significance, has captivated enthusiasts for centuries. Red wine, characterized by its rich color and robust flavors, typically derives its hue from the skin of dark-colored grapes during fermentation. Conversely, white wine, renowned for its crisp acidity and vibrant aromatics, undergoes fermentation without the presence of grape skins, resulting in a lighter hue and delicate flavor profile. Wine quality is typically assessed through two primary methods: sensory evaluation and physicochemical analysis. Despite the abundance of data available on wine attributes and quality, effectively using this information to make predictions remains difficult. Traditional data analysis methods frequently fail to uncover meaningful insights from complex wine datasets, limiting our understanding of the factors that influence wine quality and production. Stakeholders in the wine business find it difficult to derive meaningful insights from these datasets in the absence of user-friendly visualization methods and approaches. This limits their ability to successfully fulfill customer preferences, improve marketing tactics, and optimize production operations.

Therefore, the issue at hand centers on the requirement for easily understood and practical data visualization methods designed especially for wine quality analysis. By addressing this challenge, we aim to close the knowledge gap between raw data and useful insights, enabling stakeholders to fully utilize wine-related data for innovative thinking and well-informed decision-making in the business.

2. LITERATURE REVIEW

Sadiku et al. (2016) highlighted the importance of data visualization in simplifying complex information across diverse fields. Modern visualization techniques, influenced by computer graphics, offer innovative means to represent data effectively. Visualization aids decision-making by providing visual representations of complex datasets, crucial in extracting insights and making informed decisions. **[1]**

Panimalar et al. (2017) underscored the critical role that data visualization plays in helping to make sense of Big Data, overcome obstacles, and encourage adherence to appropriate practices. Their thorough examination covers different analytics, common mistakes, a wide array of visualization tools, and techniques, highlighting their importance in making informed decisions in today's data-driven world. **[2]**

Tanya et al. (2015) addressed the techniques of sensing free sulfur dioxide in wine. This study presents a method for detecting sulfites in wine using suspended core optical fibers for creating an autonomous, cost-effective, real-time wine monitoring system. The system's sensitivity to detect sulfites without diluting the wine can be adjusted by changing the fiber core size, allowing for accurate average readouts with minimal volume loss. This non-intrusive method reduces the risk of spoilage and has the potential to revolutionize wine analysis by detecting acidity, temperature, and microorganisms. **[3]**

Burcu et al. (2014) mentioned some ways like Clonal selection in viticulture, non-Saccharomyces yeasts in fermentation, and pre-fermentation techniques like nanofiltration to reduce the alcohol concentration in wine. Moreover, reverse osmosis and spinning cone columns are two post-fermentation techniques that successfully lower alcohol content without sacrificing flavor profile.**[4]**

Siddiqui (2021) addressed that various tool like Power-BI, Tableau, Infogram, ChartBlocks, Google Charts, Datawrapper, Plotly, SAP Analytics Cloud, D3.js, and NVD3 are some examples of more sophisticated visualization. These are the resources that can assist people in obtaining an organized and user-friendly Data Visualization. **[5]**

Jordão et al. (2015) discussed the sensorial impact of alcohol levels in wines. A number of strategies, including dealcoholizing processes such as vacuum distillation, spinning cone, and reverse osmosis, are used to manage the increased alcohol content in wines. Although wine qualities are preserved, these methods remove alcohol selectively while taking into account how they will affect the wine's sensory qualities and acceptability to consumers. **[6]**

According to research by Dahal et al. (2021), Gradient Boosting is the best machine learning algorithm for predicting wine quality. If there is a large amount of data in the training dataset, however, ANN produces a better forecast. **[7]**

Gupta et al. (2021) highlighted different machine learning techniques for predicting the wine quality. An enhanced Multiple Regression Model (MP5) was used in the study to compare the Random Forest and Decision Tree algorithms. Red wine classification was one area where MP5 performed better than both approaches, indicating the possibility of quality-enhancing wine production corrections. **[8]**

3. DATASET OVERVIEW

Information about Portuguese Vinho Verde wine, both the red and white, is included in this dataset. It includes information from 4898 samples of white wine and 1599 samples of red wine. **[9]**

The dataset contains 13 attributes in total, which are as follows:

Table 1

Attribute	Description
type	'red', 'white'
fixed_acidity (g/dm³)	The natural acids found in grapes that are fermented to make wine.
volatile_acidity (g/dm³)	A strong vinegar-like unpleasant flavor may arise due to high volatile acidity levels.
citric_acid (g/dm³)	Enhances the flavor of the wine and is present in small amounts.
residual_sugar (g/dm³)	Wines are classified as sweet if they contain more than 45 grams per liter, while most have at least 1 gram. Dry wines are defined as having no sweetness.
chlorides (g/dm³)	Indicate whether wine contains chloride salts, like sodium chloride.
free_sulfur_dioxide (mg/dm³)	This is the type of sulfur dioxide that results from the equilibrium between the bisulfite ion and molecular SO_2 (dissolved gas). It stops oxidation and microbiological growth in wine as a preservative.
total_sulfur_dioxide (mg/dm³)	Indicates the total amount of SO_2 in wine, both bound and free. While SO_2 is normally undetectable in wine at low concentrations, its presence can be detected in the aroma and taste of the wine when free SO_2 levels rise beyond 50 parts per million (ppm).
density (g/cm³)	Measures the thickness of wine juice due to alcohol and sugar. It is slightly thicker than water.
pH	Usually between 3 and 4, shows how acidic the wine is. Stronger pH values indicate lower acidity, and lower pH values suggest higher acidity.
sulphates (g/dm³)	Refers to the amount of potassium sulphate that is added to wine in order to raise the levels of sulfur dioxide (SO_2).
alcohol	The percentage of alcohol in a given volume of wine, is indicated by the amount of alcohol in the wine, or ABV (alcohol by volume). Wine typically has an alcohol content of 5% to 15%.
quality	A wine expert's evaluation of the overall excellence of the wine is expressed as a score that ranges from 0 (Poor) to 10 (Excellent). This is the target attribute of our dataset as we are predicting wine quality.

By analyzing these data, it can be figured out what factors contribute to wine being labeled as either good or bad quality by wine experts. Moreover, by using this dataset, Machine Learning Model can be developed to predict wine quality.

4. DATA VISUALIZATION TECHNIQUES

All the graphs are plotted using RStudio.

4.1 The Dataset at a glance

	type	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	alcohol	quality
1	white	7.0	0.27	0.36	20.70	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
2	white	6.3	0.30	0.34	1.60	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
3	white	8.1	0.28	0.40	6.90	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
4	white	7.2	0.23	0.32	8.50	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
5	white	7.2	0.23	0.32	8.50	0.058	47.0	186.0	0.9956	3.19	0.40	9.9	6
6	white	8.1	0.28	0.40	6.90	0.050	30.0	97.0	0.9951	3.26	0.44	10.1	6
7	white	6.2	0.32	0.16	7.00	0.045	30.0	136.0	0.9949	3.18	0.47	9.6	6
8	white	7.0	0.27	0.36	20.70	0.045	45.0	170.0	1.0010	3.00	0.45	8.8	6
9	white	6.3	0.30	0.34	1.60	0.049	14.0	132.0	0.9940	3.30	0.49	9.5	6
10	white	8.1	0.22	0.43	1.50	0.044	28.0	129.0	0.9938	3.22	0.45	11.0	6
11	white	8.1	0.27	0.41	1.45	0.033	11.0	63.0	0.9908	2.99	0.56	12.0	5
12	white	8.6	0.23	0.40	4.20	0.035	17.0	109.0	0.9947	3.14	0.53	9.7	5
13	white	7.9	0.18	0.37	1.20	0.040	16.0	75.0	0.9920	3.18	0.63	10.8	5
14	white	6.6	0.16	0.40	1.50	0.044	48.0	143.0	0.9912	3.54	0.52	12.4	7
15	white	8.3	0.42	0.62	19.25	0.040	41.0	172.0	1.0002	2.98	0.67	9.7	5
16	white	6.6	0.17	0.38	1.50	0.032	28.0	112.0	0.9914	3.25	0.55	11.4	7
17	white	6.3	0.48	0.04	1.10	0.046	30.0	99.0	0.9928	3.24	0.36	9.6	6
18	white	6.2	0.66	0.48	1.20	0.029	29.0	75.0	0.9892	3.33	0.39	12.8	8
19	white	7.4	0.34	0.42	1.10	0.033	17.0	171.0	0.9917	3.12	0.53	11.3	6
20	white	6.5	0.31	0.14	7.50	0.044	34.0	133.0	0.9955	3.22	0.50	9.5	5
21	white	6.2	0.66	0.48	1.20	0.029	29.0	75.0	0.9892	3.33	0.39	12.8	8
22	white	6.4	0.31	0.38	2.90	0.038	19.0	102.0	0.9912	3.17	0.35	11.0	7
23	white	6.8	0.26	0.42	1.70	0.049	41.0	122.0	0.9930	3.47	0.48	10.5	8
24	white	7.6	0.67	0.14	1.50	0.074	25.0	168.0	0.9937	3.05	0.51	9.3	5
25	white	6.6	0.27	0.41	1.30	0.052	16.0	142.0	0.9951	3.42	0.47	10.0	6
26	white	7.0	0.25	0.32	9.00	0.046	56.0	245.0	0.9955	3.25	0.50	10.4	6
27	white	6.9	0.24	0.35	1.00	0.052	35.0	146.0	0.9930	3.45	0.44	10.0	6
28	white	7.0	0.28	0.39	8.70	0.051	32.0	141.0	0.9961	3.38	0.53	10.5	6
29	white	7.4	0.27	0.48	1.10	0.047	17.0	132.0	0.9914	3.19	0.49	11.6	6
30	white	7.2	0.32	0.36	2.00	0.033	37.0	114.0	0.9906	3.10	0.71	12.3	7
31	white	8.5	0.24	0.39	10.40	0.044	20.0	142.0	0.9974	3.20	0.53	10.0	6
32	white	8.3	0.14	0.34	1.10	0.042	7.0	47.0	0.9934	3.47	0.40	10.2	6
33	white	7.4	0.25	0.36	2.05	0.050	31.0	100.0	0.9920	3.19	0.44	10.8	6
34	white	6.2	0.12	0.34	1.50	0.045	43.0	117.0	0.9939	3.42	0.51	9.0	6
35	white	5.8	0.27	0.20	14.95	0.044	22.0	179.0	0.9962	3.37	0.37	10.2	5
36	white	7.3	0.28	0.43	1.70	0.080	21.0	123.0	0.9905	3.19	0.42	12.8	5

4.2 Univariate Visualization

Investigates one attribute at a time.

4.2.1 Histogram

A histogram is a graphical depiction of the frequency distribution of a continuous series using rectangles. Usually, each bar encompasses a class, or bin, of numerical values. Data points that fall inside the respective bin are represented by a bar's height. When numerical variables have continuous data ranges, histograms work well as graphical representations.

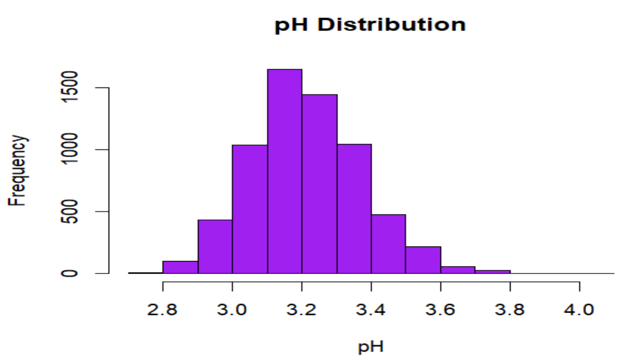
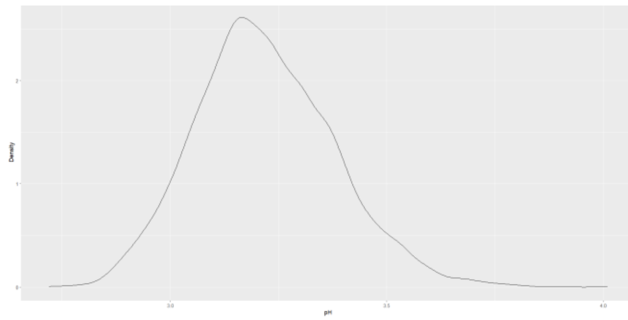


Figure 1: Histogram

The histogram above depicts the frequency distribution for the pH levels of red and white wines. Each bar represents the pH level, and its height represents the number of wine samples with pH values within the corresponding range. Higher bars indicate that there are more wine samples with pH values ranging from 3.0 to 3.2.

4.2.1.1 Line Histogram and Skewness



```
> print(skewness(pH))
[1] 0.3867495
> hist(pH,main = "pH", xlab = "pH", ylab="Frequency" ,col = "purple" )
>
> (mean(pH))
[1] 3.218501
> (median(pH))
[1] 3.21
>
> pHMode <- (names(sort(table(dataset$pH[!is.na(dataset$pH)]), decreasing = TRUE)[1]))
> print(pHMode)
[1] "3.16"
```

Figure 2: Positively Skewed or Right Skewed Histogram

From the output generated by R studio, it can be seen that the graph is not completely symmetrical. It positively skewed or right-skewed histogram as the skewness value is positive and Mean > Median > Mode. Moreover, it has a longer right tail and most of the data falls to the right of the graph's peak.

4.2.2 Bar Graph

A bar graph depicts categorical data as rectangular bars with heights or lengths that correspond to the values they represent. Bar graphs are useful for comparing items or demonstrating how something changes over time.

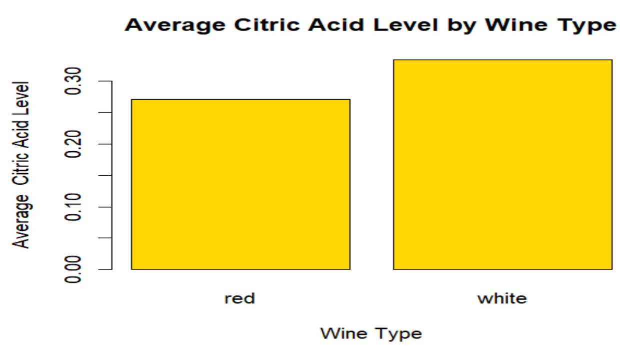


Figure 3: Bar Graph

In this bar graph, the x-axis represents two types of wine, red and white, and the y-axis represents the level of citric acid. Overall, this bar graph shows the levels of citric acid in red and white wines. Observing the length of the bar reveals that white wine contains the most citric acid.

4.2.3 Box Plot

A box plot is a technique for summarizing data that is measured on an interval scale that also displays any outliers.

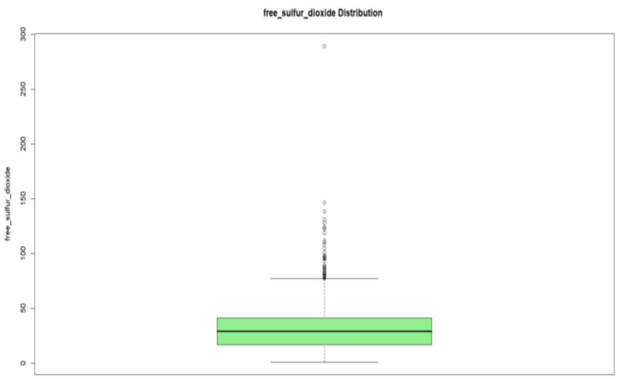


Figure 4: Box Plot

Here, the values belong to outside of the box are Outliers. The dark colored middle line shows the median value.

4.3 Multivariate Visualization

Investigates more than one attribute at a time.

4.3.1 Violin Plot

The black dot in a violin plot denotes the median. The violin shape is generated where there is a large distribution of data. The narrower part represents a lesser spread of data.

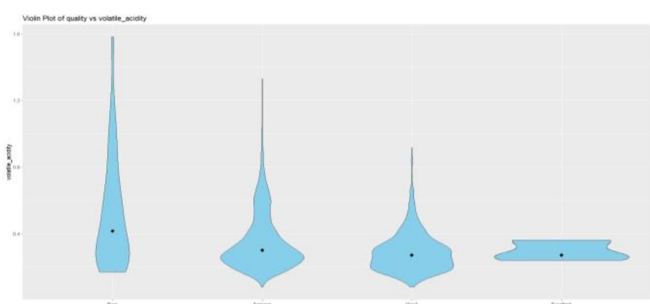


Figure 5: Violin Plot for quality vs volatile_acidity

In Figure 5, the violin plot shows the distribution of volatile_acidity of different quality wine. The quality of different wine has observed based on the volatile_acidity in it. From the graph it can be seen that the median volatile_acidity of Poor quality wine is higher than the median of other quality (Average, Good, Excellent) of wine. That means the wines which have high levels of volatile_acidity is of Poor quality. On the other hand, the median volatile_acidity of Excellent quality wine is lower than the median of other quality (Poor, Average, Good) of wine. That means the wines which have low levels of volatile_acidity is of Excellent quality. The shape of the distribution (extremely skinny on upper end and wide in the middle) indicates the volatile_acidity of Average quality wine is highly concentrated around the median. That means maximum wines contain medium volatile_acidity are of Average quality.

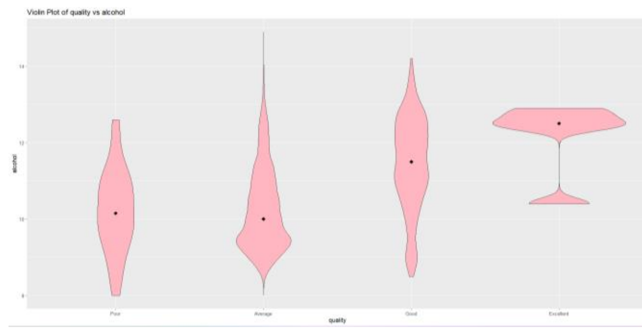


Figure 6: Violin Plot for quality vs alcohol

Here in Figure 6, the violin plot shows the distribution of alcohol of different quality wine. The quality of different wine has observed based on the alcohol (by volume) in it. From the graph it can be seen that the median alcohol of Excellent quality wine is higher than the median of other quality (Average, Good, Excellent) of wine. That means the wines which have high levels of alcohol is of Excellent quality. So, it can be seen that the higher alcohol amount (around 12-15% ABV) maintains richness and intensity of wine. On the other hand, the median alcohol of Poor, Average quality wine is lower than the median of other quality (Good, Excellent) of wine. The lower alcohol levels maintain delicacy in these wines. The shape of the distribution (skinny on lower end and wide in the middle) indicates the alcohol of Excellent quality wines are highly concentrated around the median.

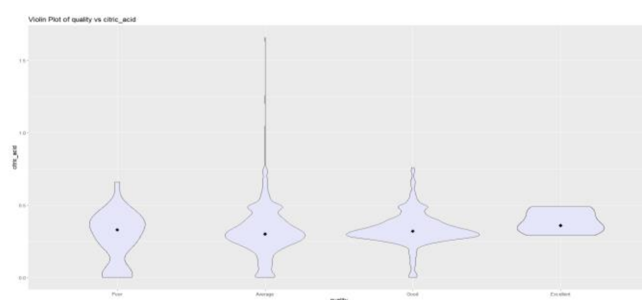


Figure 7: Violin Plot for quality vs citric_acid

In Figure 7, the violin plot shows the distribution of citric_ acid of different quality wine. The quality of different wine has observed based on the citric_acid in it. From the graph it can be seen that the median citric_acid of Poor quality wine is higher than the median of other quality (Average, Good, Excellent) of wine. That means the wines which have high level of citric_acid is of Poor quality. In small amounts, citric acid can add brightness and freshness to the wine's flavor profile, while higher levels contribute to a tart or sour taste. On the other hand, the median citric_acid of Good quality wine is lower than the median of other quality of wine. That means the wines which have small amount of citric_acid is of Good quality. The shape of the distribution (skinny on both end and wide in the middle) indicates the citric_acid of Good quality wine are highly concentrated around the median. That means maximum wines contain small amount of citric_acid are of Good quality.

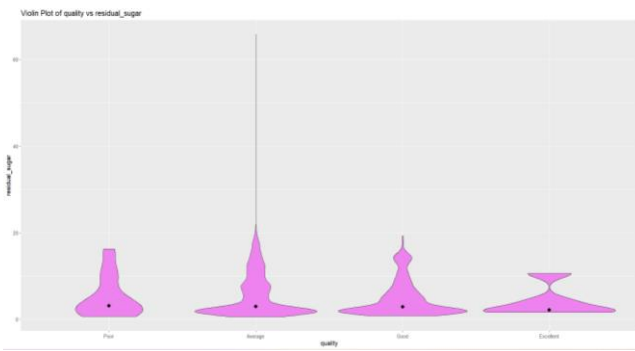


Figure 8: Violin Plot for quality vs residual_sugar

The violin plot in Figure 8, shows the distribution of residual_sugar of different quality wine. The quality of different wine has observed based on the residual_sugar in it. From the graph it can be seen that the median residual_sugar of Poor quality wine is higher than the median of other quality (Average, Good, Excellent) of wine. That means these wines have a higher residual_sugar content, considered as sweet wine. On the other hand, the median residual_sugar of Excellent quality wine is close to 0. That means these wines have very small amount of residual_sugar, considered as dry wine. Moreover, the Average and Good quality of wines contain moderate amount of residual_sugar, considered as medium sweet wine. The shape of the distribution (wide in the middle) indicates the residual_sugar of Average and Good quality wine are highly concentrated around the median. That means maximum Average and Good quality wines are medium sweet.

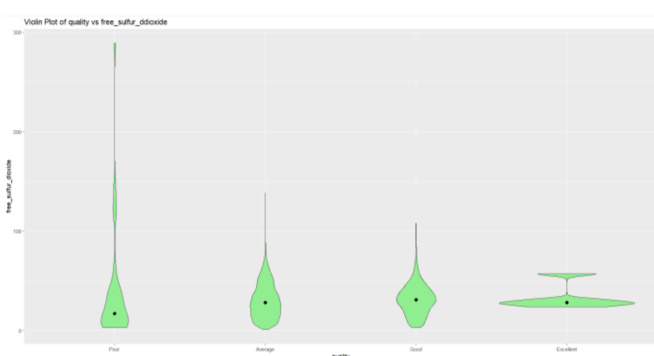


Figure 9: Violin Plot for quality vs sulfur_dioxide

Here, the violin plot shows the distribution of free_sulfur_dioxide of different quality wine. The quality of different wine has observed based on the free_sulfur_dioxide in it. From the graph it can be seen that the median free_sulfur_dioxide of Poor quality wine is lower than the median of other quality (Average, Good, Excellent) of wine. The lower amount of free_sulfur_dioxide indicates that it is unable to protect wine from spoilage microorganisms and chemical deterioration for a long period of time. On the other hand, the median free_sulfur_dioxide of Excellent quality wine is in a standard level. That means it acts as a preservative, extending the shelf life of wine and allowing it to age gracefully. Moreover, the Average and Good quality of wines contain high amount of free_sulfur_dioxide, which may cause "burnt match" aromas.

4.3.2 Scatter Plot

The scatter plot is a powerful data visualization technique. From a graph, the relationship between variables is observed and visually represented by a scatter plot. Dots represent the values of the variables. The positions of each dot on the horizontal and vertical axes represent the values for each individual data point.

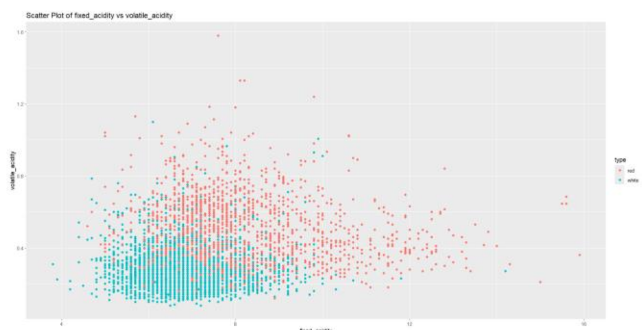


Figure 10: Moderately co-related

From the graph it can be seen that fixed acidity and volatile acidity are moderately co-related.

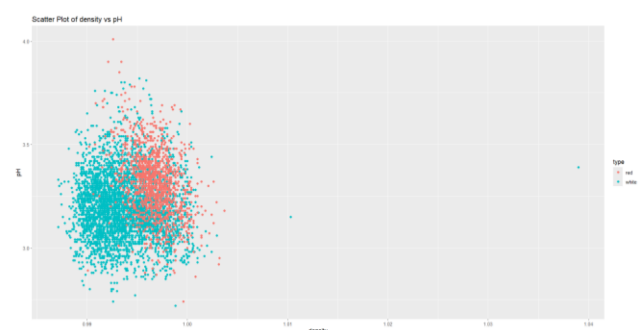


Figure 11: Strongly co-related

From the graph it can be seen that density and pH are strongly co-related as the red and blue data points are very close to each other.

4.3.3 Radar Chart

A radar chart displays multivariate data stacked on an axis with a single central point.

Figure 12: Radar chart of alcohol vs quality (white and red wine)

By comparing the radar chart of white wine and red wine it can be said that the alcohol level varies with quality the more alcohol level the better quality. For white wine quality-9 has the highest alcohol level. And for red wine quality-8 has the highest alcohol level.

Figure 13: Radar chart of residual_sugar vs quality (white and red wine)

From the radar chart it can be seen that, for low quality white wine the residual sugar is high and for high quality one the sugar level is low. As for red wine, the highest residual sugar is in quality-7 along with quality-4 and quality-5 and quality-6 have lower residual sugar.

Figure 14: Radar chart of volatile_acidity vs quality (white and red wine)

From the output it can be seen that, for white wine quality-4 has the highest volatile acidity and other high qualities seems to have lowest volatile acidity, as high level of volatile acidity can lead to unpleasant vinegar-like flavor. As for red wine quality-3 has the highest volatile acidity and other high quality wines have lower volatile acidity.

Figure 15: Radar chart of chlorides vs quality (white and red wine)

In Figure 15, it can be seen that, for both white and red wine the lowest qualities have the highest chlorides content, for high quality wine the chlorides content is decreasing. So, quality-9 from white wine and quality-8 from red wine have the lowest chlorides content. As we know high chlorides can lead to salty or briny flavor.

Figure 16: Radar chart of pH vs quality (white and red wine)

From the output it can be seen that, for white wine quality-9 has the highest pH level and slowly decreases towards low quality wines. For quality-8 the pH level is balanced (approx. 50%). On the contrary, for red wine, quality-3 has the highest pH level and quality-8 has the lowest pH level, but quality-6 has balanced pH level (approx. 50%).

4.3.4 Line Graph

The purpose of a line graph is to observe the up and down of different values based on suitable attributes.

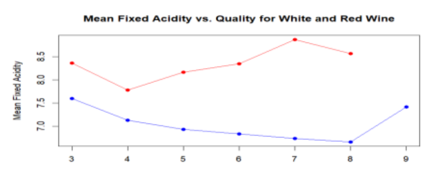


Figure 17: Line graph for fixed_acidity vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Fixed Acidity". The color blue represents white wine and color red represents red wine. From this graph it can be seen that for white wine, from quality-3 to quality-8 the “fixed acidity” level is decreasing and in quality-9 it increased. As for red wine, the “fixed acidity” level decreased at quality-4 but increased from quality-5 to quality-7. Then it decreased again in quality-8.

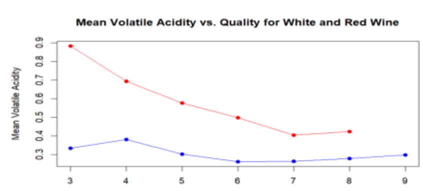


Figure 18: Line graph for volatile_acidity vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Volatile Acidity". The color blue represents white wine and color red represents red wine. From this graph it can be seen that for white wine, from quality-3 to quality-4 the "volatile acidity" level is increased and then there is a subtle flow of decreasing and increasing from quality-5 to quality-9. As for red wine, the "volatile acidity" level decreased from quality-3 to quality-7 and then increased a little at quality-8.

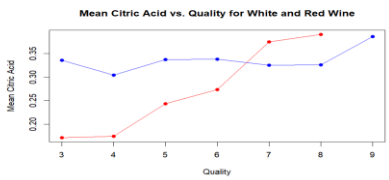


Figure 19: Line graph for citric_acid vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Citric Acid ". The color blue represents white wine and color red represents red wine. From this graph it can be seen that, for white wine, from quality-3 started from closer to 0.35 then decreased to 0.30 at quality-4. Then there is wave increasing and decreasing values from quality-5 to quality-9. As for red wine it is seen there is more like closer to no change between quality-3 and quality-4 and the value is less than 0.20. Then the line seems to move upward from quality-5 to quality-8.

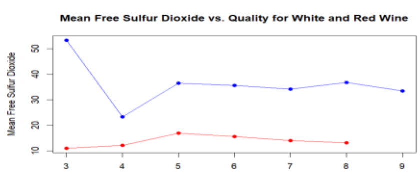


Figure 20: Line graph for free_sulfur_oxide vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Sulfur Dioxide ". The color blue represents white wine and color red represents red wine. From this graph it can be seen that, for white wine quality-3 started from greater than 50 then decreased to quality-4 by the value 22. Then it increased again from quality-5, it shows there is a flow of decreasing and increasing in values till quality-9. As for red wine we can see the value increased from quality-3 to quality-5 and then decreased from quality-6 to quality-8.

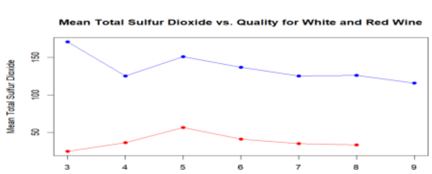
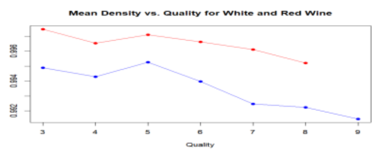


Figure 21: Line graph for total_sulfur_oxide vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Total Sulfur Dioxide ". The color blue represents white wine and color red represents red wine. From this graph it can be seen that, for white wine quality-3 started from greater than 150 then decrease to quality-4. Then it increased again and from quality-5 it shows there is a flow of decreasing and increasing in values till quality-9. As for red wine we can see the value increased from quality-3 to quality quality-5 and then decreased from quality-6 to quality-8. This graph is similar to the graph of free_sulfur_dioxide vs quality.



. Figure 22: Line graph for density vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean Density". The color blue represents white wine and color red represents red wine. Based on the graph, it is evident that for white wine, the quality score of 3 initially ranged between 0.994 and 0.996. Subsequently, it decreased to 0.994 for quality 4, followed by an increase up to quality 5. However, from quality 5 onwards, the values exhibited a declining trend until quality 9. As for red wine, quality-3 commenced significantly above 0.996. It then declined in quality 4, followed by a slight increase in quality 5. Subsequently, the values declined steadily until quality 8.

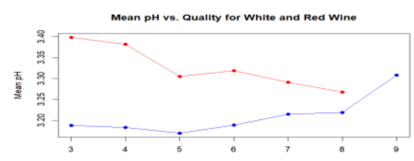


Figure 23: Line graph for pH vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean pH". The color blue represents white wine and color red represents red wine. Based on the graph, it is evident that for white wine, the value of quality-3 initially started at 3.20. then it decreased until quality-5. The there is an impressment till quality-7 and between quality-7 and quality-8 there is almost no change in value. Then increased very high in quality-9. As for the red wine the value of quality-3 initially started at 3.40. Subsequently, it tills quality-5. Then a slight increasement is noticed in quality-6. Then again decreased till quality-8.

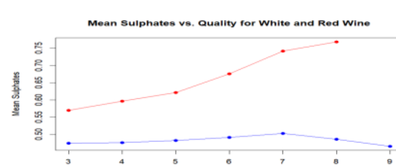


Figure 24: Line graph for sulphates vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean sulphates". The color blue represents white wine and color red represents red wine. Based on the graph, it is evident that for white wine, the value of quality-3 initially started at less than 0.50 then slowly increased until quality-7. Then slowly decreased towards quality-9. As for red wine quality-3 started at sulphate level in between 0.55 and 0.60 then increased until quality-8. In conclusion red wine's sulphate levels are higher than white wine.

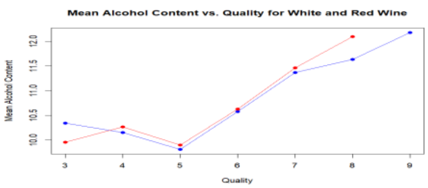


Figure 25: Line graph for alcohol vs quality (white and red wine)

The x-axis label is set to "Quality", the y-axis label is set to "Mean alcohol". The color blue represents white wine and color red represents red wine. Based on the graph, it is evident that for white wine, the value of quality-3 begins around 10.0, decreases until quality-5, then steadily rises to quality-9. Similarly, red wine exhibits a comparable pattern, with quality-3 starting at approximately 10.0, dropping until quality-5, then ascending steadily to quality-9.

5. DISCUSSION

The study uses a variety of visualization techniques to assess wine quality. Histograms show frequency distributions for numerical data, such as pH levels, while highlighting concentration ranges. Bar graphs compare categorical data, such as citric acid levels in red and white wines. The box plot summarizes the distribution of wine attributes by identifying outliers and central tendency. This can help winemakers identify and address variations in wine quality. Violin plots show distributions of data and medians, as well as correlations between wine quality and variables such as volatile acidity, alcohol content. In terms of multivariate visualization, scatter plots show the relationship between various wine attributes. Moderately correlated attributes show some interdependence, whereas strongly correlated attributes indicate a significant relationship that can be used to predict wine quality and refine production processes. Radar charts compare various attributes to wine quality in both red and white wines. Higher alcohol content in both types of wine generally indicates higher quality. Line graphs show changes in quality scores, providing insights into trends and correlations in wine characteristics.

6. CONCLUSION

In summary, Data Visualization emerges as a vital tool for analyzing and understanding the complexities of wine-related data. In the context of wine analysis, data visualization allows for the exploration of relationships between attributes such as acidity levels, alcohol content, and sensory characteristics, providing valuable insights into the factors influencing wine quality. Furthermore, visual representations help in identifying trends, patterns, and outliers within wine datasets, enabling researchers, producers, and enthusiasts to optimize winemaking practices, enhance product quality, and meet consumer preferences effectively. Thus, Data Visualization plays a pivotal role in shaping the future of wine analysis and appreciation.

REFERENCES

[1] Sadiku, M. N. O., Shadare, A. E., Musa, S. M., & Akujuobi, C. M. (2016). Data Visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 02(12), 11.

[2] A. Panimalar, K. M. Khule, S. Karthika, and T. Nirmala Kumari, "Data Visualization Tools and Techniques For Datasets In Big Data," *International Research Journal of Engineering and Technology (IRJET)*, vol. 04, no. 08, pp. 1667, Aug. 2017.

[3] T. M. Monro, R. L. Moore, M.-C. Nguyen, H. Ebendorff-Heidepriem, G. K. Skouroumounis, G. M. Elsey, and D. K. Taylor, "Sensing Free Sulfur Dioxide in Wine," *IEEE Sensors Journal*, vol. 15, no. 1, pp. 318-325, Jan. 2015.

[4] B. Ozturk and E. Anli, "Different techniques for reducing alcohol levels in wine: A review," in *Bio Web of Conferences* 3, 2014.

[5] A. T. Siddiqui, "Data Visualization: A Study of Tools and Challenges," *Asian Journal of Technology & Management Research (AJTMR)*, vol. 11, issue 01, pp. 1-5, Jun. 2021, ISSN: 2249-0892.

[6] A. M. Jordão, A. Vilela, and F. Cosme, "From Sugar of Grape to Alcohol of Wine: Sensorial Impact of Alcohol in Wine," *Beverages*, vol. 1, pp. 292-310, 2015, doi: 10.3390/beverages1040292. ISSN: 2306-5710. [Online]. Available: www.mdpi.com/journal/beverages.

[7] K. R. Dahal, J. N. Dahal, H. Banjade, and S. Gaire, "Prediction of Wine Quality Using Machine Learning Algorithms," in *Open Journal of Statistics*, vol. 11, pp. 278-289, 2021.

[8] M. Gupta and V. C. Vanmathi, "A Study and Analysis of Machine Learning Techniques in Predicting Wine Quality," in *Int. J. Recent Technol. Eng.*, vol. 10, no. 1, pp. 1-5, May 2021. ISSN: 2277-3878 (Online).

[9] "Wine Quality Data Set (Red & White Wine)," *Kaggle*, Available: https://www.kaggle.com/datasets/ruthgn/wine-quality-data-set-red-white-wine. [Accessed: May 7, 2024]