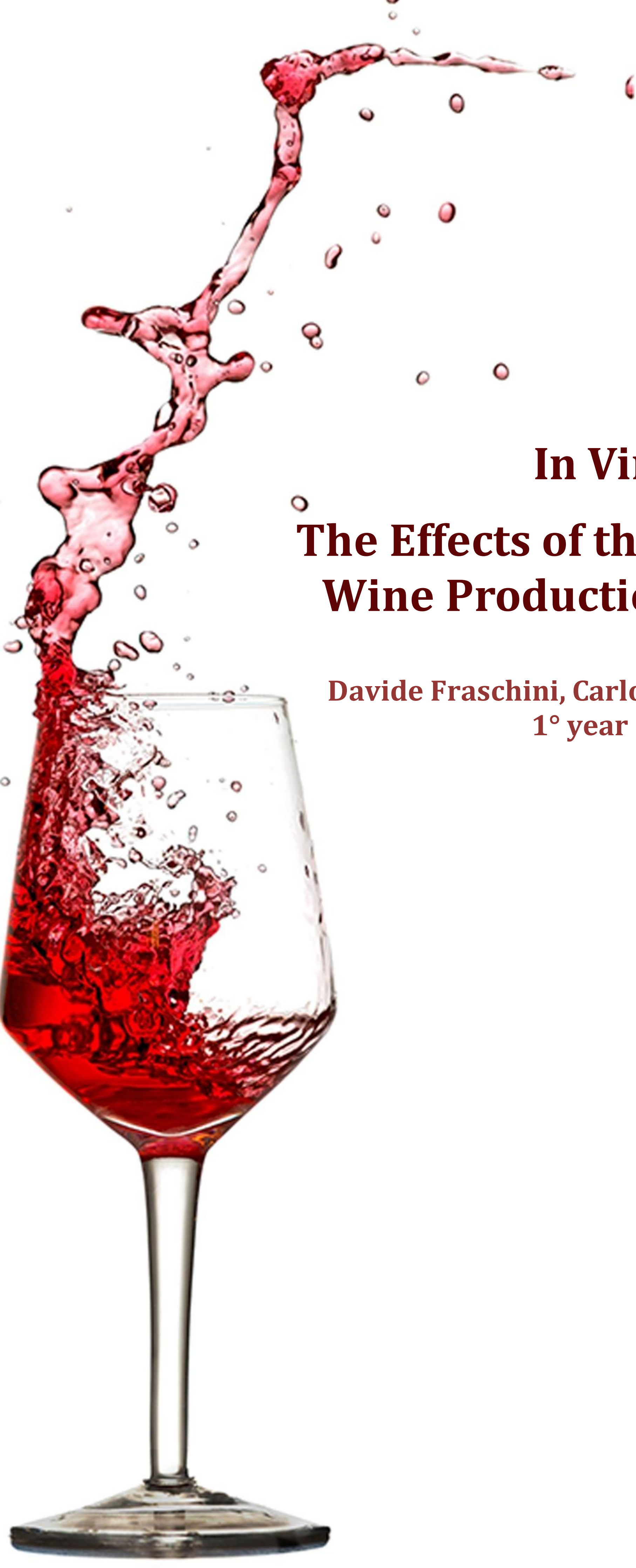


In Vino Veritas:

The Effects of the Pandemic on Global Wine Production and Consumption

Davide Fraschini, Carlotta Greco and Viviana Locatelli
1° year BESS 2023-2024

Introduction and Motivation

Wine, a beverage steeped in tradition and cultural significance, has long been a staple of social gatherings and culinary experiences. However, recent years have witnessed the wine industry undergo significant changes due to the impact of the COVID-19 pandemic.

Our statistical analysis focuses on global wine production and consumption trends, particularly in 2018 and 2020. By examining these pivotal years, we aim to highlight the correlations between the two variables before and after the pandemic. The year 2018 serves as our baseline for understanding the prevailing conditions and trends in the global wine market, free of disruption.

In contrast, 2020 stands out as a transformative year, marked by the emergence of the COVID-19 pandemic and its extensive repercussions. Is it the case that regions with higher wine production volumes are also the ones consuming the most? How do shifts in production patterns influence consumption trends, and vice versa? By examining data sets spanning multiple years, sourced from reputable industry reports such as the International Organisation of Vine and Wine, we aim to provide a thorough understanding of the global wine market.

Our Dataset

Our dataset consists of the global wine production and consumption across 209 countries during the critical years of 2018 and 2020. Through meticulous scrutiny of the OIV website, we collected data measured in millions of hectoliters, for each nation. Initially, our univariate analysis embraced 71 nations in production, and 176 in consumption, providing a comprehensive overview of individual trends. Furthermore, the variables were labelled into three qualitative categories: small, medium, and large. This categorization aimed to simplify the analysis and ease interpretation of the data.. For a clearer grasp regarding the countries included into the three subgroups of the univariate analysis, refer to the attached document. This file contains a detailed list of all countries categorised according to the three variables, aiding in a clearer comprehension of their respective classification: Classification of countries

Nevertheless, it's essential to recognize that our study faces some limitations. Indeed, it is constrained by the limited or imperfect availability of data on the OIV website, furthermore our analysis focused on countries where production and consumption exceeded 1 million hectoliters. This selective approach excluded from our database nations which did not contribute to the market.

Univariate Analysis

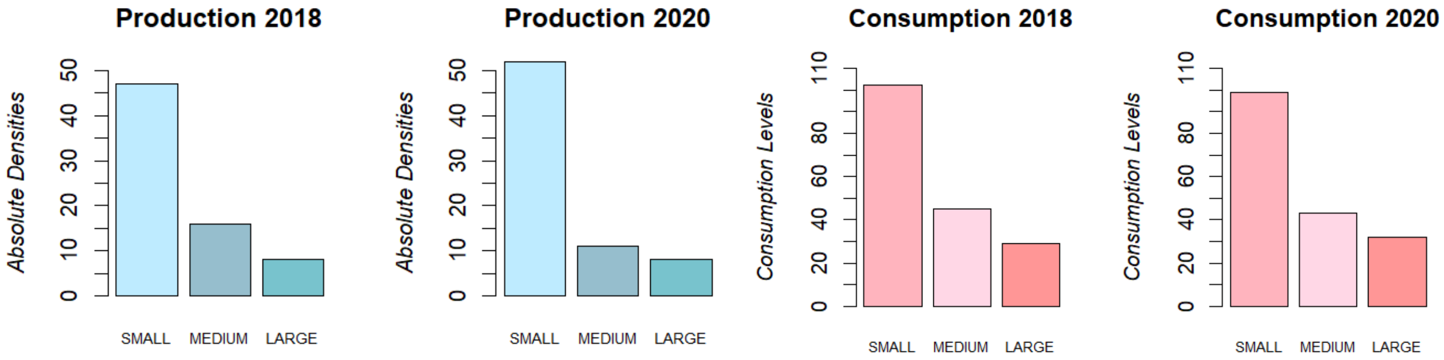
Addressing the considerable distance between individual data points in the dataset, the values have been treated as discrete qualitative. Therefore, they have been categorised into six groups: small, medium, large producers and small, medium, large consumers. This approach aimed to enhance interpretability and mitigate the wide variation between observations. All the countries with production lower than 1,000 hl/mil entered the class of small producers (47 countries in 2018, 52 in 2020), measurements below 10,000 hl/mil were grouped into medium producers (16 countries in 2018, 11 in 2020) and the last category included all values above 10,000 hl/mil as large producers (8 countries in 2018, 8 countries in 2020). Similarly, for consumption, we have segmented consumers into small, when consumption was below 100 hl/mil (102 nations in 2018, 99 nations in 2020), medium, when lower than 1,000 hl/mil (44 nations in 2018, 42 nations in 2020) and lastly in large consumers, when exceeding 1,000 hl/mil (29 nations in 2018, 32 nations in 2020).

From the following tables, that record the absolute and relative frequencies for production and consumption in both observed years, it is possible to agree that overall there seems to be a slight decrease in the consumption trends from 2018 to 2020 across all consumer categories, except for large consumers. As a result, the absolute frequencies for the small and medium categories have slightly decreased, remaining, however, relatively stable between the two years. The same conclusion can be applied to relative frequencies. Therefore, the distribution across consumer categories remains fairly consistent. As for the production trend, it seems that it has increased from 2018 to 2020 for small producers, while it has decreased for medium producers and remained stable for large producers. This result is reflected into both the absolute and relative frequencies.

	Absolute Frequency		Relative Frequency	
Labels	Consumption 2018	Consumption 2020	Consumption 2018	Consumption 2020
Small Consumer	102	99	0,5795454545	0,5689655172
Medium Consumer	45	43	0,2556818182	0,2471264368
Large Consumer	29	32	0,1647727273	0,183908046
	Absolute Frequency		Relative Frequency	
Labels	Production 2018	Production 2020	Production 2018	Production 2020
Small Producer	47	52	0,661971831	0,7323943662
Medium Producer	16	11	0,2253521127	0,1549295775
Large Producer	8	8	0,1126760563	0,1126760563

Measures of Centrality

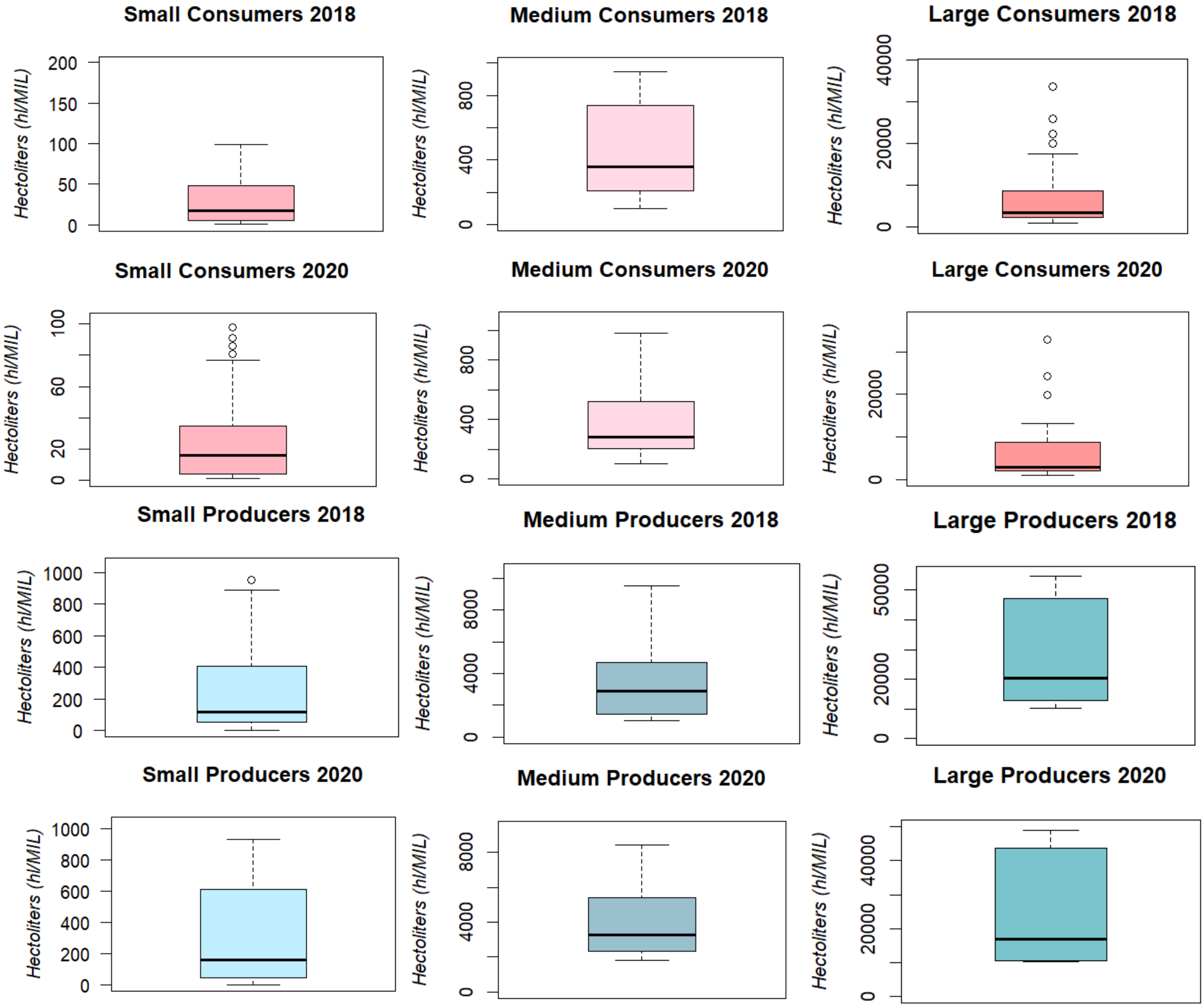
The analysis of our research data reveals a significant number of extreme values, which substantially influence the sample mean, both in the case of production and consumption. On the other hand, the median is an effective way of representing the sample data. The bar plots reveal a notable positive skewness, illustrated by the mean that is higher than the median for both variables in the observed years. Such skewness highlights a lack of symmetry in the data, signifying that the majority of observations cluster towards the lower end of the scale, with fewer extreme high values pulling the mean upwards, but still not managing to distance it from the lowest point of the range.



Measures of Variability

	Consumption		Production	
	2018	2020	2018	2020
Mean	1361,227273	1328,258621	4149,887324	3690,971831
Median	65	55	406	438
Sample range	33717	32853	54782	49065
First Quartile	13	13	88,5	81
Third Quartile	549	498,5	2067,5	2010,5
Sample Interquartile Range	536	485,5	1979	1929,5
Sample Variance	18934158,88	17987842,97	112306651,3	92842192,8
Sample Standard Deviation	4351,339894	4241,207725	10597,48325	9635,465365
Coefficient of Variation	3,204466695	3,193058685	2,553679757	2,610549689

The table above allows for a deeper understanding of the data distribution and their dispersion from the mean value, as evident from the boxplots. With the sample interquartile range significantly greater than the median, the variability of the dataset is also high. Therefore, in the central portion of the data, dispersion is significant. The middle 50% of the data is spread out over a wide range, even if the median is not particularly extreme. The results for the variance and standard deviation confirm the great scattering; data values substantially vary from the sample average. Between consumption and production, the coefficient of variation, compared to the mean, indicates that the former exhibits greater dispersion.



From a careful observation of the boxplots, it can be noted that there is a consistent presence of outliers for large consumers between 2018 and 2020, whereas in the case of small consumers, the presence of outliers increased only after the pandemic. This could be attributed to lockdown measures, stockpiling, bulk purchases, and increased spending on specific categories like comfort food and beverages. Conversely, in production, market dynamics during the COVID-19 pandemic might have compelled a reduction in production for some countries classified as small-scale producers.

Bivariate analysis

In our bivariate analysis, a key consideration was the discrepancy in the number of observations between wine production and consumption. With 71 nations included in the production dataset and 176 nations in the consumption dataset, a direct comparison between the two sets would have been impractical and could have compromised the effectiveness of our study. To address this disparity, we opted to study a sample of the consumption data by shortening it to match the nations present in the production dataset. This approach allowed us to focus on a subset of the consumption that was directly comparable to the production, ensuring a more effective examination of the relationship between the two variables across the selected observations.

Relationship between the two variables

	Year 2018	Year 2020
Sample Covariance	52637420,5	46092985,74
Sample Correlation Coefficient	0,770773328	0,767480656

Analysing the correlation coefficient from the table above, the high positive correlations observed in both 2018 (0.7708) and 2020 (0.7675) indicate a solid connection between the two variables. The positive covariance values also seem to support this hypothesis, suggesting that as wine production increases, consumption levels tend to rise simultaneously, and vice versa, showing how the variables, within the wine market, are closely linked. The slight decrease of the correlation coefficient in the second period could be explained by disruption in the supply chain, change in consumer behaviour and the economic impact of COVID-19 restrictions and regulations.

Simple Linear Regression Model

To further deepen the previously mentioned hypothesis, the relationship between the two variables was examined in a linear regression model. By looking at the scatterplots, production and consumption were considered respectively as the independent and dependent variable, resulting in two equations:

$$\widehat{y_1} = 1292,7 \ + \ (0,4687) \cdot \widehat{x_1} \qquad \widehat{y_2} = 1288,9 \ + \ (0,4965) \cdot \widehat{x_2}$$

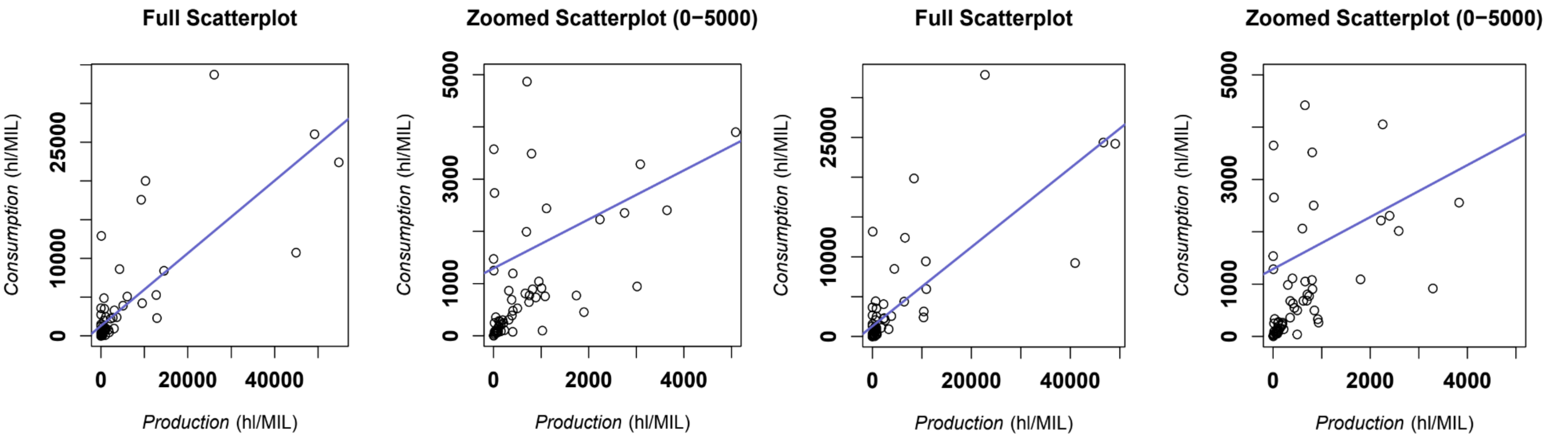
$\widehat{y_1}(\widehat{x_1})$ represent the 2018 model

$\widehat{y_2}(\widehat{x_2})$ represent the 2020 model

The point of these equations is to serve as predictive tools. *De facto*, looking at the coefficients of determination in the table below, the accuracy of the models seem to be moderate, if not high: about 59% in 2018 and 52% in 2020. Further insight is gained by examining the sum of squares (SST, SSR, SSE), which clarifies the variance within the dataset and the explanatory power of the regression model. In 2018, the total sum of squares (SST) amounts to 2906888696, with a significant portion of the variance explained by the regression model (SSR = 1726957934). Whereas, in 2020, the total sum of squares (SST) slightly diminishes to 2719489934 and the proportion of variance justified by the regression model (SSR = 1427649740) also attenuates. The drop in the coefficient of determination suggests that either the relationship between production volumes and consumption levels changed between 2018 and 2020, or that other external factors influenced the correlation during the studied time periods.

	Year 2018	Year 2020
beta 1 (parameter)	0,468693705	0,496465932
beta 0 (parameter)	1292,706329	1288,882175
Coefficient of Determination	0,5940915236	0,5249696725

SSR	1726957934	1427649740
SSE	1179930762	1291840194
SST	2906888696	2719489934



2018 **2020**

Conclusion

Looking at the data and the consequent interpretation, the following findings can be derived: firstly, in the univariate analysis focusing on country categorization by production and consumption scales, it was observed that from 2018 to 2020, there was a marginal decline in consumption trends in almost all consumer segments. During the same period, production showed an increase for small-scale producers and remained steady for large-scale producers. Small producers likely exhibited greater flexibility and were able to adjust their production strategies to meet evolving consumer demands. They were able to tap into niche markets where demand remained stable or increased. Indeed, during times of crisis, consumers may have prioritised supporting local businesses. On the other hand, larger producers may have encountered challenges in their supply chains due to disruptions in sourcing raw materials or logistical constraints, leading them to exercise caution in scaling up production. Additionally, small and medium producers may have prioritised maintaining or improving the quality of their wines over increasing volume, resonating with consumers seeking higher quality products during turbulent times.

Secondly, the bivariate analysis revealed a strong positive correlation between wine production and consumption, with a slight decline in the correlation coefficient noted in 2020, potentially attributed to disruptions induced by the pandemic. The linear regression model suggested modest accuracy in predicting consumption levels based on production volumes for both 2018 and 2020, although there was a decrease in the coefficient of determination observed in 2020. This decline implies a possible alteration in the relationship between production and consumption, or external factors influencing the correlation.

The analysis focused on countries actively engaged in wine production, providing a more targeted and representative sample for our investigation. Excluding nations with zero production levels helped to streamline the dataset, eliminating irrelevant data points that could potentially skew the results, meanwhile it may have overlooked valuable insights from countries with minimal wine production. Finally, our analysis has provided compelling evidence supporting the hypothesis that there exists a strong correlation between wine production and consumption levels.

References:

https://www.oiv.int/what-we-do/data-discovery-report?oiv