

1 Introduction

When assessing strategies for viticulture there is a trade-off between the amount of resources invested and the resultant quality and quantity of grapes produced [10]. Due to recent economic, social and environmental pressures, such as: bush fires, tariffs, and shifting attitudes to alcohol consumption; the chosen emphasis taken in balancing the value-productivity tradeoff has become more important [16, 18, 20]. When considering pressures such as those prior listed, the integration of viticulture within the Australian wine industry is a compounding effect.

In this analysis we observed relationships between quantity and quality through the use of linear models. For the purpose of this study; quality was defined by the financial value of grapes per tonne. An extensive amount of research into a variety of factor's effect on grape quality and yield exists. Due to the lack of long-term and in-depth data sources, individual effects are often studied in isolation [1]. The lack of consolidated datasets also restricts the ability to gain statistical insight at a large scale across multiple regions [7, 8]. The dataset used for this analysis includes data collected for the past 10 years from a multitude of vineyards located over a diverse range of Australian winegrowing regions.

2 Methods

We created four linear models to explore the relationships between resource use and vineyard outputs (see table 1). The models looked at the value and quantity of grapes produced by vineyards, as a total and as a scale of area harvested. We compared values to observe how economies of scale affect the use of resources.

2.1 Data

The data used in this analysis was provided by Sustainable Winegrowing Australia, Australia's national wine industry sustainability program. Sustainable winegrowing Australia aims to facilitate grape-growers and winemakers in demonstrating and improving their sustainability [15]. Data recorded by Sustainable Winegrowing Australia is entered manually by winegrowers using a web based interface; with some fields being optional. Two subsets of

Table 1: Summary of models; their predictors, covariates and variable interactions.

	Response	Predictors	Covariates
Model 1	Yield	Water Used, Scope 1 Emissions	Area Harvested, Year, GI
Model 2	$\frac{\text{Yield}}{\text{Area Harvested}}$	Water Used, Scope 1 Emissions	Area Harvested, Year, GI
Model 3	Yield \times Average Sale Price	Water Used, Scope 1 Emissions	Area Harvested, Year, GI
Model 4	$\frac{\text{Yield} \times \text{Average Sale Price}}{\text{Area Harvested}}$	Water Used, Scope 1 Emissions	Area Harvested, Year, GI

this data were defined by vineyards that recorded values for average price of grape sales per tonne, and vineyards who did not. Both subsets contained: region, harvest year, yield, area harvested, water used and fuel used (in litres for diesel, petrol, biodiesel and LPG). To enable comparisons, total fuel was converted to amount of carbon emissions in metric tons.

The first subset of data was used for Model 1 and Model 2 (see Table 1). This subset contained 5298 samples spanning the period from 2012 to 2022, covering 57 GI Regions and 1432 separate vineyards.

The second subset of data, was limited to vineyards that recorded a value for their average sale price of grapes per tonne. This subset was used for Model 3 and Model 4 (see Table 1); and contained 2878 samples spanning the period from 2015 to 2022, covering 51 GI Regions and 944 separate vineyards.

Data was limited to samples that had recorded values for variables used (see Table 1). After reviewing correlation coefficients the data was logarithmically transformed, centred and scaled by standard deviation. Two values for average sale price were removed from the dataset, due to a recording of \$1. Unreported values for average prices per tonne were filled in using regional averages taken from Wine Australia’s annual reports, where they were available [17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27].

Other variables including the use of renewable energy, contractors; and pressures such as frost, fire and disease were also explored. Variables that did not significantly contribute to the prediction of a response variable were excluded.

2.2 Total Emissions

Emissions were calculated from the total diesel, petrol, bio-diesel and LPG used for irrigation and activities within the vineyard. The equations given from the Australian National Greenhouse Accounts Factors, shown as

$$tCO_2e = \frac{Q \times EC \times EF1 + EF3}{1000}, \quad (1)$$

was used to convert the quantity of fuel in litres, Q , using a prescribed Energy Content, EC , and emission factors of scope one, $EF1$, and scope three, $EF3$, to tonnes of Carbon Dioxide equivalent, tCO_2e [3].

The variables were reviewed for correlations by using a Pearson’s Correlation Coefficient (see Tables 1, 2 and 3). This was undertaken for data on the original scale (see Table 1) and for data as a logarithmic transform (see Table 2). All P-values were found to be significant ($p < 2.200E-16$), except the non-transformed values for water used (see Table 3). The logarithmic transforms performed the best due to a skew likely caused by a greater number of smaller vineyards within the dataset (see Table 4).

2.3 Region

The site of a vineyard predetermines several physical parameters such as climate, geology and soil; making location a widely considered key determinant of grape yield and quality [1, 2, 4]. Differences in vineyard locations were captured through the use of Geographical Indicator Regions (GI Regions). Each GI Region has its own unique mixture of climatic and geophysical properties that describes a unique winegrowing region within Australia; these regions were predefined by Wine Australia [6, 11, 13].

The climatic properties of a GI Region are summarised in the Sustainable Winegrowing Australia user manual [14]. The user manual describes climates by rainfall and temperature, creating supersets of Regions of similar climatic properties. The climatic groups were used to illustrate similarities and differences occurring in areas larger than GI regions.

2.4 Analysis

General Linear Models were used as they offered the ability to produce statistical models that were explicit in the relationships between predictors and

response variables. They also allowed the exploration of interactions between predictors and easily comparable differences in the influence and magnitude of relationships.

Data preprocessing, such as logarithmic transforms, was done using the Python programming language [5]. Linear models were created using the R statistical programming language [12]. These models were created iteratively to explore a variety of variable interactions and approaches to modelling the data. Not all explored approaches yielded improvements or accurate models. Alternate approaches included the use of Splines, hierarchical regression, Additive and Generalised Linear Models. Other variables were also explored but not used due to low reporting values such as fertiliser, tractor and contractor use. The use of only scope one emissions was due to the same reason where scope 2 sources were recorded sporadically at best.

2.5 Model Validation

Models were validated using K-fold cross validation calculated through the R Caret Package [9]. K-fold cross validation works by removing a subset of data from the sample used to train models and then predicts those variables to determine how sensitive the model is to changes in the sample data. For this analysis each model was validated using 10 folds, repeated 100 times.

3 Results

3.1 Exploratory Analysis

Simple linear relationships between variables were explored using Pearson Correlation Coefficients. This was undertaken for data on the original scale (see Table 2) and for data as a logarithmic transform (see Table 3). Strong relationships were found to be present, as all P-values were considered significant ($p < 2.200E-16$, see Tables 2 and 3), except for the non-transformed values for water used (see Table 4). The logarithmic transforms showed the strongest correlations, this was likely due to a skew caused by a greater number of smaller vineyards within the dataset (see Table 5).

Table 2: Summary of models, their predictors, covariates and variable interactions.

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{Yield}{Area}$
Yield	1.000E+00	7.440E-01	-4.309E-03	7.290E-01	3.500E-01
Area	7.440E-01	1.000E+00	-5.331E-03	8.921E-01	7.854E-02
Water Used	-4.309E-03	-5.331E-03	1.000E+00	-1.929E-03	-5.600E-03
Scope One Emissions	7.290E-01	8.921E-01	-1.929E-03	1.000E+00	9.357E-02
$\frac{Yield}{Area}$	3.500E-01	7.854E-02	-5.600E-03	9.357E-02	1.000E+00
Average Price Per Tonne	-2.262E-01	-1.178E-01	-3.562E-02	-9.422E-02	-4.849E-02
$\frac{Average Price per tonne}{Area}$	-1.644E-01	-2.042E-01	-2.669E-02	-1.933E-01	-1.698E-01

Table 3: Pearson correlation coefficients for each logarithmically transformed variable.

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{Yield}{Area}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353E-01
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742E-01
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292E-01
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086E-01
$\frac{Yield}{Area}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000E+00
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625E-01
$\frac{Average Price per tonne}{Area}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.070E-01

Table 4: P-values for the non-transformed water used variable's Pearson correlation coefficients.

Variable	Water Used
Yield	7.538E-01
Area	6.981E-01
Scope One Emissions	8.883E-01
$\frac{Yield}{Area}$	6.836E-01
Average Price Per Tonne	5.600E-02
$\frac{Average Price per tonne}{Area}$	1.522E-01

Table 5: Summary statistics for each variable on the original scale..

Variable	Yield	Area	Water Used	Scope One Emissions	$\frac{Yield}{Area}$
Yield	1.000E+00	8.822E-01	8.245E-01	7.617E-01	9.353E-01
Area	8.822E-01	1.000E+00	7.750E-01	8.311E-01	6.742E-01
Water Used	8.245E-01	7.750E-01	1.000E+00	6.668E-01	7.292E-01
Scope One Emissions	7.617E-01	8.311E-01	6.668E-01	1.000E+00	6.086E-01
$\frac{Yield}{Area}$	9.353E-01	6.742E-01	7.292E-01	6.086E-01	1.000E+00
Average Price Per Tonne	-4.591E-01	-1.911E-01	-4.881E-01	-1.559E-01	-5.625E-01
$\frac{Average Price per tonne}{Area}$	-8.918E-01	-8.474E-01	-8.300E-01	-7.063E-01	-8.070E-01

Table 6: Model 1 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Year	9	7.060E+01	7.800E+00	8.353E+01	<2.20E-16 ***
GI Region	54	1.507E+03	2.790E+01	2.972E+02	<2.20E-16 ***
Area Harvested	1	3.211E+03	3.211E+03	3.419E+04	<2.20E-16 ***
Water Used	1	1.040E+01	1.040E+01	1.103E+02	<2.20E-16 ***
Scope One Emissions	1	6.600E+00	6.600E+00	7.056E+01	<2.20E-16 ***

3.2 General Linear Models

Models 1 and 2 showed significant relationships between each of the predictors and their response variable (see Tables 6 and 7). Variables in models 3 and 4 reported similar significance; except for scope 1 emissions (see Tables 8 and 9). Scope one emissions was included in all models to directly compare the response variables as ratios of vineyard size to raw values. Even though not significant within models 3 and 4, when using the Pearson Correlation Coefficients, scope one emissions was strongly correlated to every Model's response variable; this was especially so for Model 1 and 4 (Yield and average price per tonne as a ratio to area harvested, respectively).

The comparison of models performance shows that the average price per tonne of grapes describes a great deal of the relationship between predictor and response when comparing model 2 to model 4 (see Table 10). This relationship between yield and average price was also illustrated in the correlation values between them (see Table 2).

Limitations included overestimating yield for models 1 and 2, (see Figures 1 and 2) and underestimating crop value in models 3 and 4 (see Figures 3 and 4).

Table 7: Model 2 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Area Harvested	1	2.407E+03	2.407E+03	1.080E+04	<2.20E-16
Scope One Emissions	1	3.989E+01	3.989E+01	1.789E+02	<2.20E-16
Water Used	1	5.500E+02	5.500E+02	2.467E+03	<2.20E-16
Area Harvested*Scope One Emissions	1	6.921E+01	6.921E+01	3.104E+02	<2.20E-16
Area Harvested * Water Used	1	1.040E+00	1.040E+00	4.686E+00	3.045E-02 *
Year * GI Region	424	1.144E+03	2.700E+00	1.210E+01	<2.20E-16

Table 8: Model 3 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Year	6	1.324E+01	2.210E+00	8.748E+01	<2.20E-16 ***
GI Region	50	6.498E+02	1.300E+01	5.151E+02	<2.20E-16 ***
Area Harvested	1	2.142E+03	2.142E+03	8.491E+04	<2.20E-16 ***
Water Used	1	3.200E-01	3.200E-01	1.259E+01	3.947E-04 **
Scope One Emissions	1	4.000E-02	4.000E-02	1.492E+00	2.221E-01

Table 9: Model 4 ANOVA summarising variable significance at the .5 level.

Variable	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Area Harvested	1	2.066E+03	2.066E+03	5.700E+04	<2.20E-16
Scope One Emissions	1	6.000E-02	6.000E-02	1.569E+00	2.105E-01
Water Used	1	2.014E+02	2.014E+02	5.557E+03	<2.20E-16
Area Harvested*Scope One Emissions	1	5.246E+01	5.246E+01	1.448E+03	<2.20E-16
Area Harvested * Water Used	1	7.270E+00	7.270E+00	2.005E+02	<2.20E-16
Year * GI Region	243	4.546E+02	1.870E+00	5.162E+01	<2.20E-16

Table 10: Comparison of Model Residuals

	Df	Sum Sq	Mean Sq
Model 1	5231	4.913E+02	1.000E-01
Model 2	4868	1.085E+03	2.200E-01
Model 3	2818	7.111E+01	3.000E-02
Model 4	2629	9.528E+01	4.000E-02

Table 11: Comparison of Model performance.

	RSE	R2	Adjusted R2	F-statistic	P-Value
Model 1	3.065E-01	9.072E-01	9.061E-01	7.753E+02	<2.2e-16
Model 2	4.722E-01	7.951E-01	7.770E-01	4.403E+01	<2.2e-16
Model 3	1.589E-01	9.753E-01	9.748E-01	1.885E+03	<2.2e-16
Model 4	1.904E-01	9.669E-01	9.638E-01	3.095E+02	<2.2e-16

Reviewing the data to uncover reasons for this included the use of binary variables such as the utilisation of renewable energy, contractors, and the occurrence of disease, fire and frost; however none of these variables were able to explain why some vineyards produced less, or why other vineyards sold at higher prices than predicted. A wide variety of these influences were likely already explained within the use of year and GI Region, or the interaction of both variables. The change between some regions was dramatic, with particularly warmer and drier regions producing much higher volumes of grapes at lower prices (See Figures 5 and 6). The use of other variables and methods, specifically splines, were able to create a more normally distributed set of residuals but at a drastically reduced accuracy when comparing R2 and RSE. The introduction of known average prices per tonne also helped increase R2 values a small amount; it is important to not that it is common practice for wineries to purchase grapes at a regional average rate, likely resulting in much less variance within a region.

The relationship between scope one emissions and the response variables that included average sales price

It is possible that the relationships between scope one emissions and the response variables were closely tied to a vineyards area. This possibility could be explained through the emissions

Noting that irrigation systems use fuel and that the application of water was a significant variable in each model scope one emissions' lack of significance and contribution given its F-statistics (See Tables 7 and 8), indicated that it is possible other vineyard activities requiring fuel are not as determining factors for a vineyards grape quality.

References

- [1] Philippe Abbal et al. ‘Decision Support System for Vine Growers Based on a Bayesian Network’. In: *Journal of agricultural, biological, and environmental statistics* 21.1 (2016), pp. 131–151. ISSN: 1085-7117. DOI: 10.1007/s13253-015-0233-2.
- [2] Eduardo Agosta, Pablo Canziani and Martín Cavagnaro. ‘Regional Climate Variability Impacts on the Annual Grape Yield in Mendoza, Argentina’. In: *Journal of Applied Meteorology and Climatology* 51.6 (2012), pp. 993–1009.
- [3] Department of Climate Change, Energy, the Environment and Water. *Australian National Greenhouse Accounts Factors*. Nov. 2022. URL: <https://www.dcceew.gov.au/sites/default/files/documents/national-greenhouse-accounts-factors-2022.pdf>.
- [4] Helder Fraga, Ricardo Costa and João A. Santos. ‘Multivariate Clustering of Viticultural Terroirs in the Douro Winemaking Region’. In: *Ciência T c. Vitiv.* 32.2 (2017), pp. 142–153. URL: <https://doi.org/10.1051/ctv/20173202142>.
- [5] G. van Rossum. *Python Tutorial, Technical Report CS-R9526*. Version 3.10.6. Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica (CWI), 1995.
- [6] J. C. (James C.) Halliday. *Australian Wine Encyclopedia*. VIC: Hardie Grant Books, 2009. ISBN: 978-1-74273-245-9.
- [7] Keith Jones. *Australian Wine Industry Environment Strategy*. 27 Sept. 2002.
- [8] Hanne Knight et al. ‘Firm Resources and the Development of Environmental Sustainability among Small and Medium-Sized Enterprises: Evidence from the Australian Wine Industry’. In: *Business Strategy and the Environment* 28.1 (1 Jan. 2019), pp. 25–39. ISSN: 0964-4733. DOI: 10.1002/bse.2178. URL: <https://doi.org/10.1002/bse.2178> (visited on 13/02/2023).
- [9] Max Kuhn. ‘Building Predictive Models in R Using the Caret Package’. In: *Journal of Statistical Software, Articles* 28.5 (2008), pp. 1–26. ISSN: 1548-7660. DOI: 10.18637/jss.v028.i05. URL: <https://www.jstatsoft.org/v028/i05>.

- [10] Johnny V. Montalvo-Falcón et al. ‘Sustainability Research in the Wine Industry: A Bibliometric Approach’. In: *Agronomy* 13.3 (2023). ISSN: 2073-4395. DOI: 10.3390/agronomy13030871.
- [11] D.P. Oliver et al. ‘Review: Soil Physical and Chemical Properties as Indicators of Soil Quality in Australian Viticulture’. In: *Australian Journal of Grape and Wine Research* 19.2 (1 June 2013), pp. 129–139. ISSN: 1322-7130. DOI: 10.1111/ajgw.12016. URL: <https://doi.org/10.1111/ajgw.12016> (visited on 13/02/2023).
- [12] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021. URL: <https://www.R-project.org/>.
- [13] C.J SOAR, V.O SADRAS and P.R PETRIE. ‘Climate Drivers of Red Wine Quality in Four Contrasting Australian Wine Regions’. In: *Australian journal of grape and wine research* 14.2 (2008), pp. 78–90. ISSN: 1322-7130. DOI: 10.1111/j.1755-0238.2008.00011.x.
- [14] Sustainable Winegrowing Australia SWA. *Sustainable Winegrowing Australia User Manual*. Aug. 2021. URL: <https://www.awri.com.au/wp-content/uploads/2019/07/SWA-user-manual.pdf>.
- [15] Sustainable Winegrowing Australia SWA. *Sustainable Winegrowing Australia*. Case Studies. 2022. URL: <https://sustainablewinegrowing.com.au/case-studies/> (visited on 05/11/2022).
- [16] John W. Toumbourou et al. ‘Student Survey Trends in Reported Alcohol Use and Influencing Factors in Australia’. In: *Drug and Alcohol Review* 37.S1 (1 Apr. 2018), S58–S66. ISSN: 0959-5236. DOI: 10.1111/dar.12645. URL: <https://doi.org/10.1111/dar.12645> (visited on 16/04/2023).
- [17] Wine Australia. ‘National Vintage Report 2019’. In: (2019). URL: https://www.wineaustralia.com/getmedia/807bf053-3692-448a-9ed5-c0084a47e1bb/Vintage-report-2019_full-version.pdf.
- [18] Wine Australia. ‘National Vintage Report 2020’. In: (2020). URL: https://www.wineaustralia.com/getmedia/7dcd66a7-86b5-4606-9ab7-416086077099/MI_VintageReport2020.pdf.

- [19] Wine Australia. ‘National Vintage Report 2021’. In: (2021). URL: [https://www.wineaustralia.com/getmedia/100ffcfcd-5672-49eb-8b5e-ac31cd2f57b3/MI_VintageReport2021_full-\(1\)_1.pdf](https://www.wineaustralia.com/getmedia/100ffcfcd-5672-49eb-8b5e-ac31cd2f57b3/MI_VintageReport2021_full-(1)_1.pdf).
- [20] Wine Australia. ‘National Vintage Report 2022’. In: (2022). URL: https://www.wineaustralia.com/getmedia/b61c7a0e-9e17-4142-a98b-945e3f18873e/MI_VintageReport2022_full.pdf.
- [21] Winemakers’ Federation of Australia. ‘National Vintage Report 2012’. In: (2012). URL: <https://www.agw.org.au/assets/vintage-reports/PDFs/WFA-Vintage-Report-2012.pdf>.
- [22] Winemakers’ Federation of Australia. ‘National Vintage Report 2013’. In: (2013). URL: <https://www.agw.org.au/assets/vintage-reports/PDFs/WFA-Vintage-Report-2013.pdf>.
- [23] Winemakers’ Federation of Australia. ‘National Vintage Report 2014’. In: (2014). URL: <https://www.agw.org.au/assets/vintage-reports/Vintage-Report-and-Production-Profitability-2014.pdf>.
- [24] Winemakers’ Federation of Australia. ‘National Vintage Report 2015’. In: (2015). URL: <https://www.agw.org.au/assets/vintage-reports/Vintage-Report-and-Production-Profitability-2015.pdf>.
- [25] Winemakers’ Federation of Australia. ‘National Vintage Report 2016’. In: (2016). URL: <https://www.agw.org.au/assets/vintage-reports/Vintage-Report-2016.pdf>.
- [26] Winemakers’ Federation of Australia. ‘National Vintage Report 2017’. In: (2017). URL: <https://www.agw.org.au/assets/vintage-reports/NationalVintageReport-2017.pdf>.
- [27] Winemakers’ Federation of Australia. ‘National Vintage Report 2018’. In: (2018). URL: https://www.wineaustralia.com/getmedia/fce5bf18-468b-4e1b-a478-d6ca231d40f6/VintageReport2018_full.pdf.