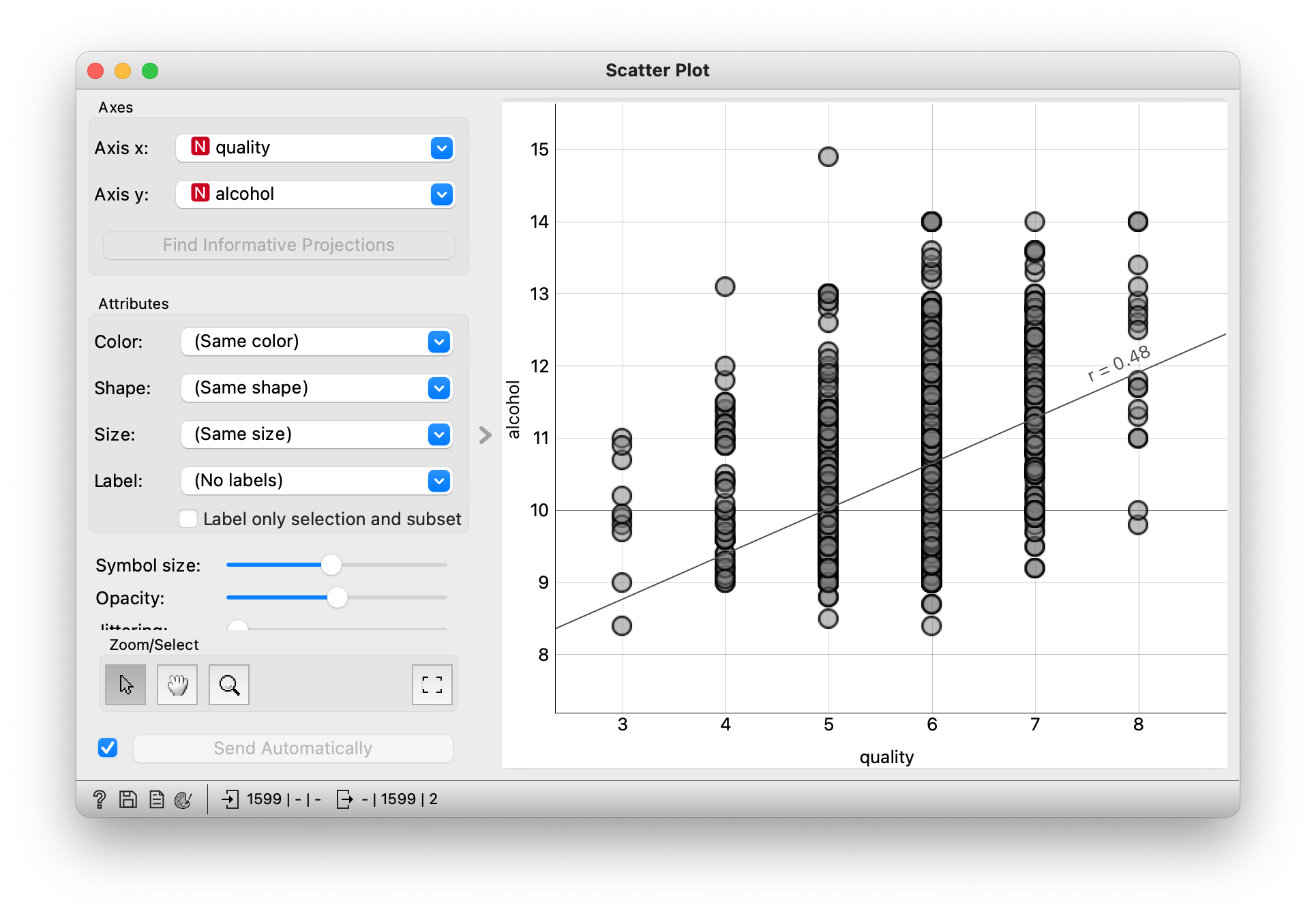
**Introduction - Fabio**

1. Goals: Predict the quality of different Portuguese “Vinho Verde” wine based on variables/attributes provided in the dataset.
2. Dataset/Attributes: *Red Wine Quality:* [*www.kaggle.com/uciml/red-wine-quality*](https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009)*.*

This dataset contains physicochemical and sensory data on red wine variants of the Vinho Verde wine.

Attributes: Input variables: These attributes are variables that are accurately measured and recorded to the dataset as input data. Input attributes are: Volatile acidity, total sulfur dioxide, density, chlorides, pH, free sulfur dioxide, residual sugar, fixed acidity, citric acid, sulphates, alcohol.

Output variable: Quality. This is the target attribute which is a result of manipulating the input data. This variable is the target attribute in the dataset.



| **Volatile Acidity :** Acidic elements of a wine that are gaseous, rather than liquid, and therefore can be sensed as a smell. Volatile acids are produced through microbial action such as yeast fermentation, malolactic fermentation, and other fermentations carried out by spoilage organisms. |
| --- |
| **Total Sulfur Dioxide:** Preserves wine’s freshness and fruit characters by virtue of antioxidant, antimicrobial and anti-enzymatic properties. |
| **Density:** Concentration of alcohol, sugar, glycerol, and other dissolved solids. |
| **Chlorides:** Amount of salt in the wine |
| **Ph:** Measure of the concentration of free hydrogen ions in solution. About 3.0 to 3.4 is desirable for white wines, while about 3.3 to 3.6 is best for reds. |
| **Free Sulfur Dioxide:** The portion of SO2 that is free in the wine plus the portion that is bound to other chemicals in the wine such as sugar. It **prevents the wine from reacting with oxygen** which can cause browning and off-odors (oxidation), and it **inhibits the growth of bacteria and undesirable wild yeasts** in the grape juice and wine. |
| **Residual Sugar:** The natural grape sugars left over in a wine after the alcoholic fermentation is complete. |
| **Fixed Acidity:** The combined sum of titratable and volatile acids present. |
| **Citric Acid:** Citric acid imparts a citric character that enhances the taste of many white and blush *wines*. |
| **Sulphates:** Natural by-product of the fermentation process that work as a preservative against certain yeast and bacteria |
| **Alcohol:** Fortified wines range from 15.5% to 25% ABV, with an average of 18%. Typically about 12% |

**Data - Tyler**

1. Preprocessing
   1. The dataset has already been cleaned, so there was no need to apply additional data cleaning. Only discretization was performed.
   2. The dataset was automatically discretized from a range of numeric attributes into nominal by using the first-last method in WEKA.
   3. Then, I created a new categorical attribute, Quality Category, using the Wine Quality numeric inputs. I used an IF statement in Excel to split the numerical quality values into three categories: Poor, Good, and Excellent.
2. Data Analysis

* correlations

1. Our goal is to identify which attributes are responsible for producing excellent quality wine. We used Orange to calculate the r value of each attribute on wine quality.

* classification

1. Now that we have an understanding of which attributes are correlated with higher wine quality, the next step is to conduct classification experiments to get the bigger picture. We applied 4 different classification methods using Weka.
2. In Weka, the datasets are separated into training and testing sets by using 10-fold cross validation. The training data is portioned into 10 sets of equal size and the algorithms executed 10 times.

**Whats Next - Mehmet**

1. Model Training/Testing (classification methods) explain methods
2. What’s next?
   1. Remove less relevant attributes, and re-run the Random Forest algorithm to try and get a better result.