

2IV35 Visualization Set 2

Jeroen van Oorschot Student number: 0721913
j.v.oorschot@student.tue.nl

Mart Pluijmaekers Student number: 0753117
m.h.l.pluijmaekers@student.tue.nl

February 1, 2015

Contents

1	Dataset	3
2	Analysis of the data	3
3	Visualizations	3
3.1	Scatter plot	3
3.2	Interval tree	3
3.3	Median tree	3
4	Results	4
5	Conclusion	4

1 Dataset

Many people in the world take joy in drinking wine although many of them do not have a clue about what they are actually drinking and what makes their wine taste better than others. Many people go to wine-connaisseurs for advice, or enter liquor stores where usually someone has some knowledge about wines, to help give advice on the purchase. However, has many properties which can be measured other than the year. For example, sweetness, acidity and even chlorides play a part in the taste and quality of the wine. Not even the best and most experienced connaisseurs can possibly have tasted and formed an opinion on every wine in existence. Which means that it can be beneficial to have a way to objectively analyze wines to help determine which wines are expected to be good and which are not.

2 Analysis of the data

TODO

3 Visualizations

The visualizations listed in this section are implemented to help analyze the dataset

3.1 Scatter plot

Since we are interested in correlation of data, it makes sense to use a scatter plot since it can visualize this correlation. As the dataset already has a quality parameter available, we want to (possibly) correlate all other parameters to the listed quality. Using this analysis it will become possible to see which parameters influence the quality more and which influence it less.

Usefulness of this visualization would be limited if all differences in measurements are very small and therefore the datapoints would lie close to each other. On the other hand, if the differences are big, but no correlation is found, scatter plots are useless too.

3.2 Interval tree

The interval tree is a specialized visualization for this dataset. The root of the tree has children for all columns and each of these children has children for every existing quality. When creating the interval tree, input of the user is required in the form of a number (n) of interval requires, the program then creates n equal divisions represented as children. Those children then get colored, the lowest is red, while the highest is green, all other children are colored using a linear interpolation from red to yellow and yellow to green.

The size of the children depicts the

3.3 Median tree

The median tree is very similar to the interval tree, except that it only has two divisions for every quality. The first for the data bigger than the median, while the latter is for all data

smaller than the median.

This is useful since when the measurements are very unbalanced, assume one measurement is ten times bigger than all other. If then the program has to divide the total interval in 3 sections, we get 1 green leaf, while all other leaves are red and have a size of almost 100%, which obviously does not help.

4 Results

TODO

5 Conclusion

Using our visualizations we see that a top wine usually has a relatively high alcohol and sulfur dioxide concentration while the concentration of chlorides, total and residual sulfur dioxide.