Write-up – Create a Tableau Story

by Florina Georgescu

Summary

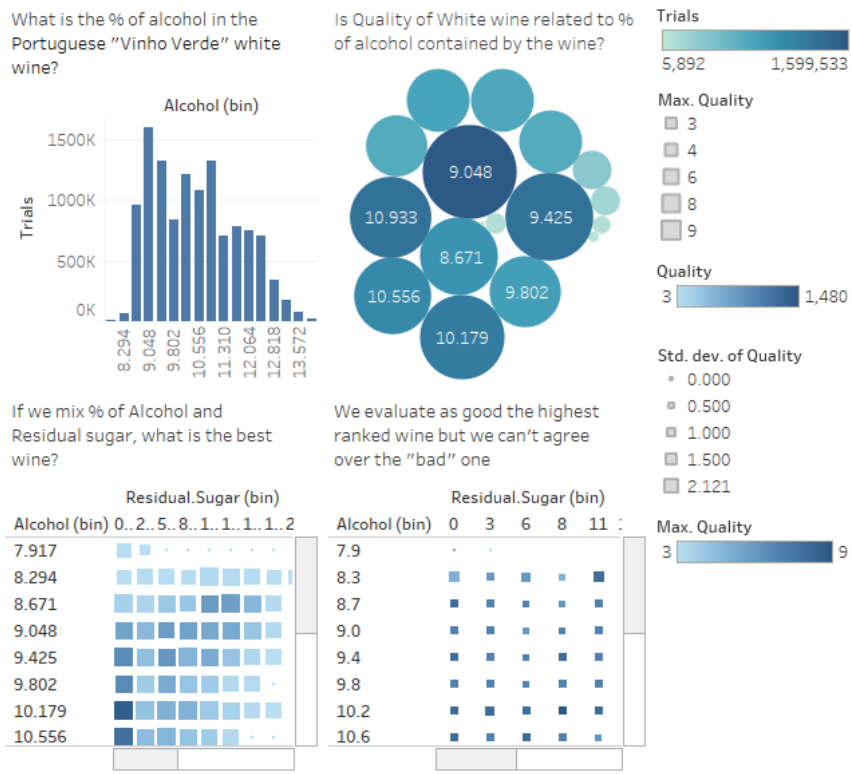
What are the right proportions for the chemical properties of a white wine in order to obtain a tasty one? This is the question I want to answer with my Tableau story.

This tidy data set contains 4,898 white wines with 11 variables on quantifying the chemical properties of each wine. At least 3 wine experts rated the quality of each wine, providing a rating between 0 (very bad) and 10 (very excellent).

I want to identify at least two chemical components that are differentiate the bad wine from the excellent one.

Design

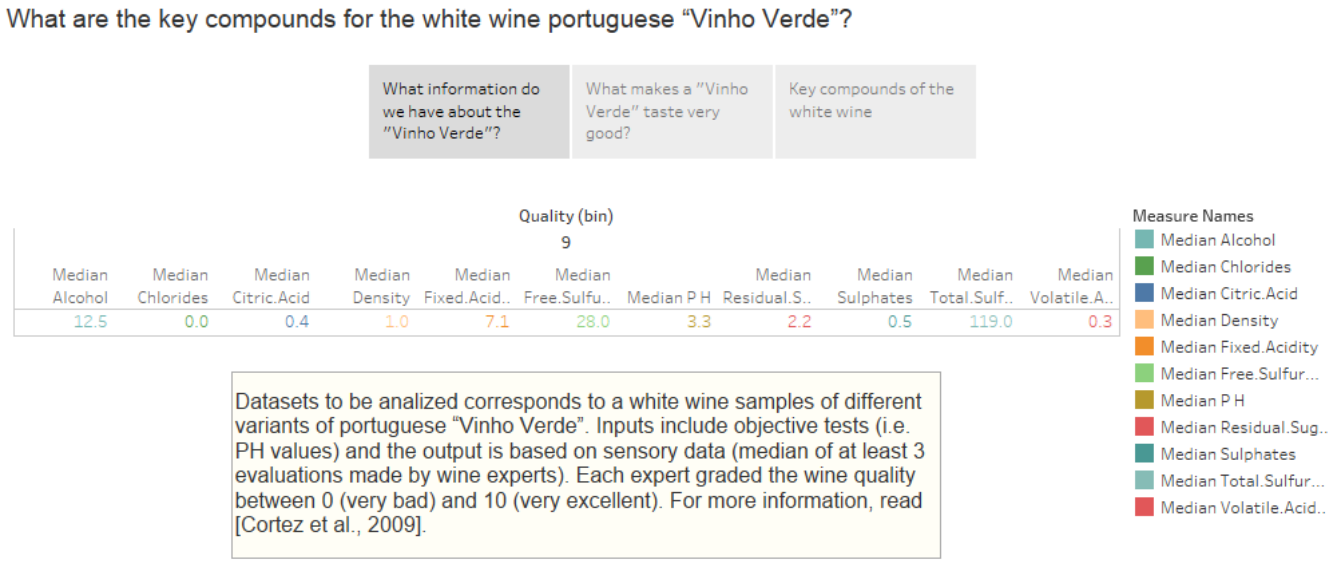
This is my first sketch.



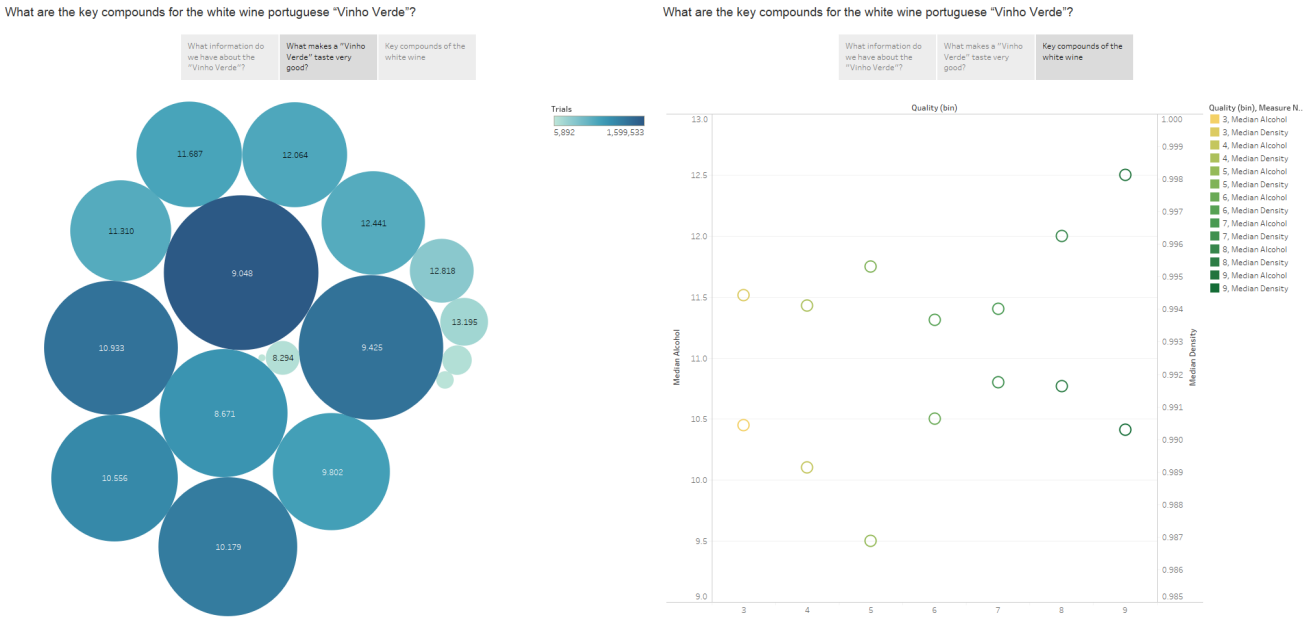
I started my study by looking at the raw data of the 11 parameters. The first idea that came into my mind was to make a histogram or a bubble chart for the alcohol that I knew previously from my R Exploratory Data Analysis that variated with the quality of the wine. I tried a different view for the Residual sugar (bar charts). Each time I have set the quality value for the colour legend.

As seen in the course the colour has the highest impact on the reader and I wanted to stay in a neutral palette and not to use to may flashy colours. I hoped to catch the attention of the reader with the variating the size (e.g. size increases with the level of the quality).

The second option I thought it was to give first the reader an overview of the average value of the parameters and then after to focus on a main parameter such as alcohol.



The “story” tab was a perfect option as Tableau has this predefined and it helps you focus on a story with a short summary for each tab.



When I added this bubble chart I had a very clear idea of what I wanted: to see the number of wines tested and their quality, but now I it does not seem intuitive at all. The Scatter plot was the option for showing in an aggregated way the increase and decrease of alcohol and density parameter for each level of quality.

Feedback

I posted on our Udacity forum a request for feedback (link here) but after 8 days no feedback at all. For this reason, I decided to ask colleagues at work what they thought knowing that they have no idea about the data set and they are not familiar with data analysis.

The first feedback I got was that people could barely understand the last two graphics. The first slide of the story seemed to colorful but at least understandable. They gave me the advice to make it more as a high management summary for a wider public with or without knowledge about statistics. Also, one other suggestion was to integrate the story in one unique slide (kakemono stile).

Therefore, my new design of the story is a set of three bubble charts and 1 line plot showing:

- Minimum value vs quality categories for all parameters
- Average value vs quality categories for all parameters
- Maximum value vs quality categories for all parameters
- Line plot for quality categories vs average value of all parameters

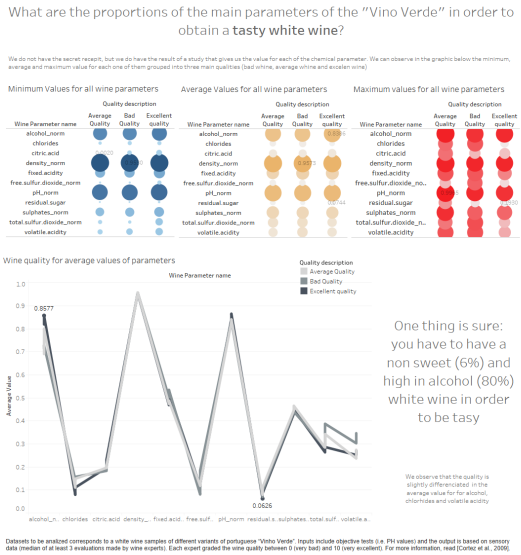
I avoided box plots and histograms (although I made tests with this types of graphics) as they are not understandable for all type pf public.

The colour for the 3 bubble charts I have chosen them as blue (suggesting depth), orange for average and red (suggesting a high value) for maximum.

The line plot has a grey scale for quality (the highest the quality, the darkest the grey).

When people hover over the line graph and choose from the legend let's say the "excellent" quality all the excellent quality values for all parameters from the bubble chart highlights.

I order to achieve these views I had to create an excel file with a different format and to calculate the minimum, average and maximum values. Also I defined arbitrary a bin from bad wine (3-4 quality), average wine (5-7 quality) and excellent wine (8-9 quality).



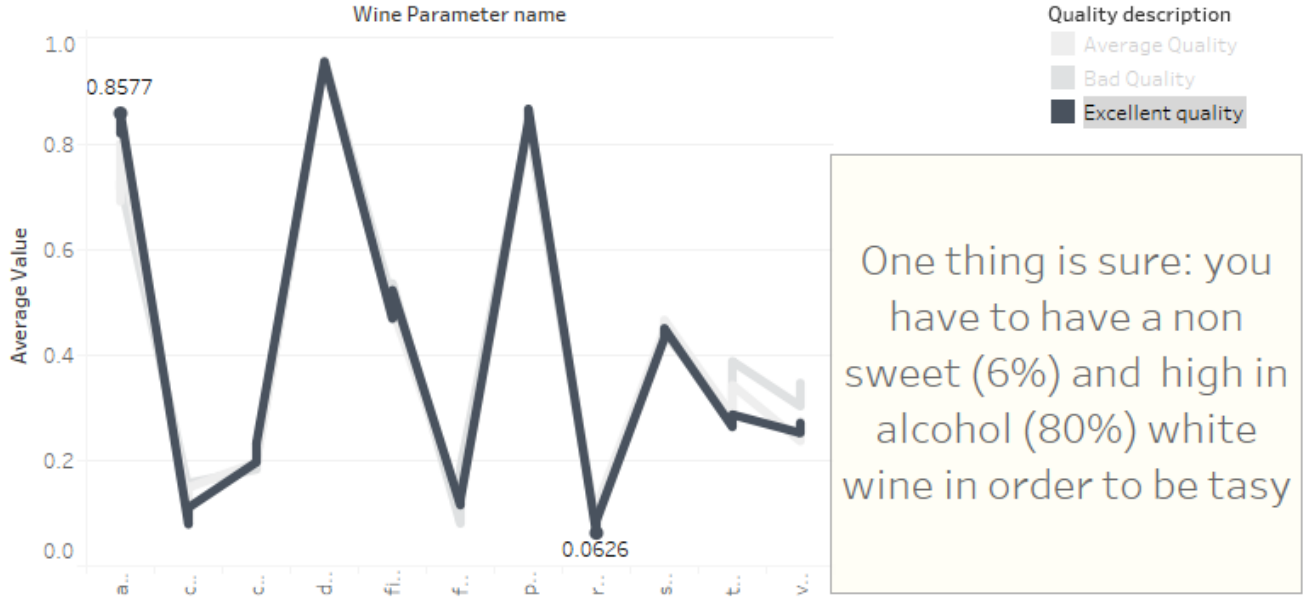
After Udacity's first project review I have changed the dashboard into a story as suggested. I preferred to use the laptop resolution for the layout of the story as is easier to browse also with the mobile phone.

What are the proportions of the main parameters of the "Vino Verde" in order to obtain a **tasty white wine**?

The graphic shows the average, value for the chemical parameters grouped into three main qualities (bad whine, average whine and excelent wine).

The graphic shows the maximum, value for the chemical parameters grouped into three main qualities (bad whine, average whine and excelent wine).

We observe in the graph below that the quality is slightly differenciated in the average value for for alcohol, chlorhides and volatile acidity.



One thing is sure: you have to have a non sweet (6%) and high in alcohol (80%) white wine in order to be tasy

Resources

Tableau story link (first Tableau Story before feedback):

https://public.tableau.com/profile/florina.georgescu#!/vizhome/Vinho_Verde_white_wine_test3/whitewine

Tableau story link (final after feedback of Udacity's reviewer):
https://public.tableau.com/profile/florina.georgescu#!/vizhome/White_wine_final_story/White_wine_story

Raw data link:

https://s3.amazonaws.com/udacity-hosted-downloads/ud651/wineQualityWhites.csv

Resources for visualizations:

https://public.tableau.com/profile/dlima#!/vizhome/DemocracianasAmricas/Dashboard1

https://public.tableau.com/profile/mat.leonard#!/vizhome/shared/8B7RH9JPC

http://geog.uoregon.edu/datagraphics/color_scales.htm

Data Files

- the final data set used to create the visualization: White_wine_summary.csv

Appendix - Readme

```
Citation Request:
  This dataset is public available for research. The details are described in [Cortez et al.,
2009].
  Please include this citation if you plan to use this database:

  P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.
  Modeling wine preferences by data mining from physicochemical properties.
  In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.

  Available at: [@Elsevier] http://dx.doi.org/10.1016/j.dss.2009.05.016
                [Pre-press (pdf)] http://www3.dsi.uminho.pt/pcortez/winequality09.pdf
                [bib] http://www3.dsi.uminho.pt/pcortez/dss09.bib

1. Title: Wine Quality

2. Sources
   Created by: Paulo Cortez (Univ. Minho), Antonio Cerdeira, Fernando Almeida, Telmo Matos and
Jose Reis (CVRVV) @ 2009

3. Past Usage:

  P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.
  Modeling wine preferences by data mining from physicochemical properties.
  In Decision Support Systems, Elsevier, 47(4):547-553. ISSN: 0167-9236.

  In the above reference, two datasets were created, using red and white wine samples.
  The inputs include objective tests (e.g. PH values) and the output is based on sensory data
  (median of at least 3 evaluations made by wine experts). Each expert graded the wine quality
  between 0 (very bad) and 10 (very excellent). Several data mining methods were applied to
model
  these datasets under a regression approach. The support vector machine model achieved the
  best results. Several metrics were computed: MAD, confusion matrix for a fixed error
tolerance (T),
  etc. Also, we plot the relative importances of the input variables (as measured by a
sensitivity
  analysis procedure).

4. Relevant Information:

   The two datasets are related to red and white variants of the Portuguese "Vinho Verde"
wine.
   For more details, consult: http://www.vinhoverde.pt/en/ or the reference [Cortez et al.,
2009].
   Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output)
variables
   are available (e.g. there is no data about grape types, wine brand, wine selling price,
etc.).

   These datasets can be viewed as classification or regression tasks.
   The classes are ordered and not balanced (e.g. there are munch more normal wines than
   excellent or poor ones). Outlier detection algorithms could be used to detect the few
excellent
   or poor wines. Also, we are not sure if all input variables are relevant. So
   it could be interesting to test feature selection methods.

5. Number of Instances: red wine - 1599; white wine - 4898.

6. Number of Attributes: 11 + output attribute
```

Note: several of the attributes may be correlated, thus it makes sense to apply some sort of
feature selection.

7. Attribute information:

For more information, read [Cortez et al., 2009].

Input variables (based on physicochemical tests):
1 - fixed acidity (tartaric acid - g / dm^3)
2 - volatile acidity (acetic acid - g / dm^3)
3 - citric acid (g / dm^3)
4 - residual sugar (g / dm^3)
5 - chlorides (sodium chloride - g / dm^3
6 - free sulfur dioxide (mg / dm^3)
7 - total sulfur dioxide (mg / dm^3)
8 - density (g / cm^3)
9 - pH
10 - sulphates (potassium sulphate - g / dm3)
11 - alcohol (% by volume)
Output variable (based on sensory data):
12 - quality (score between 0 and 10)

8. Missing Attribute Values: None

9. Description of attributes:

1 - fixed acidity: most acids involved with wine or fixed or nonvolatile (do not evaporate readily)

2 - volatile acidity: the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste

3 - citric acid: found in small quantities, citric acid can add 'freshness' and flavor to wines

4 - residual sugar: the amount of sugar remaining after fermentation stops, it's rare to find wines with less than 1 gram/liter and wines with greater than 45 grams/liter are considered sweet

5 - chlorides: the amount of salt in the wine

6 - free sulfur dioxide: the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine

7 - total sulfur dioxide: amount of free and bound forms of S02; in low concentrations, SO2 is mostly undetectable in wine, but at free SO2 concentrations over 50 ppm, SO2 becomes evident in the nose and taste of wine

8 - density: the density of water is close to that of water depending on the percent alcohol and sugar content

9 - pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic); most wines are between 3-4 on the pH scale

10 - sulphates: a wine additive which can contribute to sulfur dioxide gas (S02) levels, wich acts as an antimicrobial and antioxidant

11 - alcohol: the percent alcohol content of the wine

Output variable (based on sensory data):
12 - quality (score between 0 and 10)