What's in My Wine? An Interactive d3 Visualization Project Report

Jaishree Ranganathan, Surya Pavan Malireddy
Department of Computer Science
University of North Carolina Charlotte
Charlotte, NC 28223
Email: {jrangan1,smalire1}@uncc.edu

Abstract—In today's world wines are one of the favorite beverages to almost everyone. People go to vineyards and wineries to taste different wines and find out the best wine the suits their taste and interest. According to [1] 'Wine speaks to all the senses: the eyes behold the color, tone, and shade; the nose, the bouquet, the fingers and lips caress the cool crystal; the ears delight in the subtle swishing of the liquid; the tongue rejoices in the reward of a rich harvest'. Wine tasting involves all sensory perceptions of which the most is attributed towards the smell or aroma which describes the flavor of wines. In this project we focus on classifying wines based on the aroma and develop a interactive visualization tool that helps user's select the wines based on various attributes including: country, price range and aroma's to understand the flavor's of that particular wine category.

I. INTRODUCTION

According to [2], if a consumer is buying a specific bottle of wine at the vineyard, at retailers in actual shops, or even online, then they take different information sources into consideration to choose the which wine to buy. Most people decide by watching the wine label on the bottle together with more or less detailed background knowledge of wines in general. Such average wine consumers are interested to know different properties of shortlisted wines, for instance name of the wine, producer, complexity, producing country, region, grape type, color, taste, or price.

Any naive wine taster have difficulty in finding the right choice of wine just by looking at the label or other factors on high level. The following attributes as mentioned above plays an important role to help users better understand. The wine produced country, Title and Vintage, price, points given by experienced wine tasters, and finally the wine description. In this work we used these attributes and developed a interactive visualization tool that could help naive wine tasters.

Author's Kerren et al. [3] say that description of wines written by professional wine tasters have a strict rhetorical structure and consists of three main parts. It basically starts with production facts and ends with an assessment and a recommendation of prime drinking time.

There are many works that are contributed towards analyzing wine reviews and other factors [4], [5] [6]. The reminder of this report explains the methodolgy, use case and conclusion.

TABLE I
DATASET INFORMATION

Attribute	Number of Non Null Entries
country	129908
description	129971
designation	92506
points	129971
price	120975
province	129908
region_1	108724
region_2	50511
taster_name	103727
taster_twitter_handle	98758
title	129971
variety	129970
winery	129971

II. METHODOLOGY

This sections explains the dataset, data classification, external resources, visualization techniques adopted.

A. Data Collection

In this work we used the wine reviews data available in kaggle [7]. This dataset was originally collected from WineEnthusiast website. The dataset labeled winemag-data-130k-v2.csv is used for our project which contains 10 attributes and 130,000 records. Sample data is shown in Fig. 1.

B. Data Cleaning

The data consists of the following attributes country, description, designation, points, price, province, region_1, region_2, taster_name, taster_twitter_handle, title, variety, winery. The Table I shows number of entries and null values.

The original data contains 129971 entries of which 9983 were duplicates. In this process of data cleaning these duplicate records are removed. The field 'description' is used to categorize the data based on 'aromas'. This is a text field and basic text pre-processing is applied in order to process the text for categorization. Python jupyter notebook is used for text processing.

1) Tokenization: Tokenization is the process of splitting into pieces called 'tokens'. In this case the description text is divided into 'tokens' or 'words' separated by space.
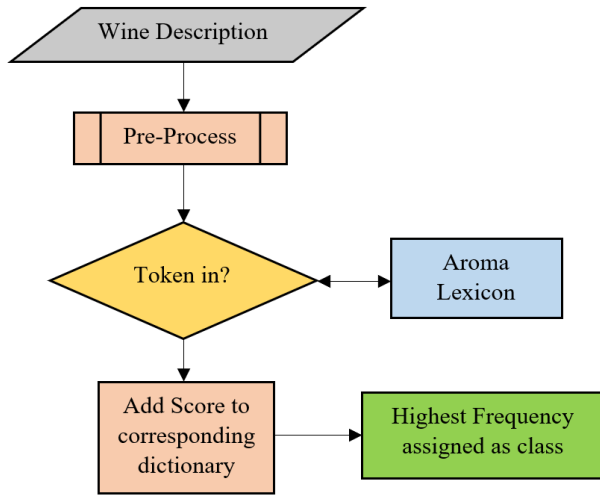
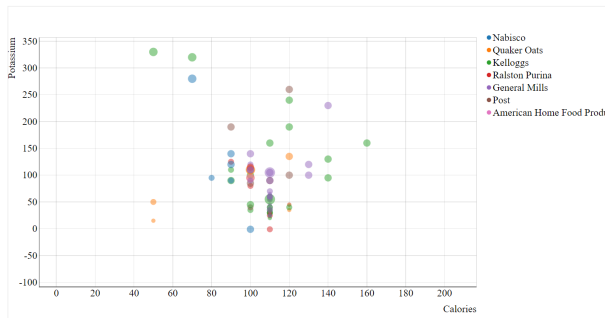| country	description	designation	points	price	province	region_1	region_2	taster_name	taster_twitter_handle	title	variety	winery
Italy	Aromas include tropical fruit, broom, brimstone and dried herb. The palate isn't overly expressive, offering unripened apple, citrus and dried sage alongside brisk acidity.	Vulkà Bianco	87		Sicily & Sardinia	Etna		Kerin O'Keefe	@kerinokeefe	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia
Portugal	This is ripe and fruity, a wine that is smooth while still structured. Firm tannins are filled out with juicy red berry fruits and freshened with acidity. It's already drinkable, although it will certainly be better from 2016.	Avidagos	87	15.0	Douro			Roger Voss	@vossroger	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red	Quinta dos Avidagos
US	Tart and snappy, the flavors of lime flesh and rind dominate. Some green pineapple pokes through, with crisp acidity underscoring the flavors. The wine was all stainless-steel fermented.		87	14.0	Oregon	Willamette Valley	Willamette Valley	Paul Gregutt	@paulgwine	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm
US	Pineapple rind, lemon pith and orange blossom start off the aromas. The palate is a bit more opulent, with notes of honey-drizzled guava and mango giving way to a slightly astringent, semidry finish.	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore		Alexander Peartree		St. Julian 2013 Reserve Late Harvest Riesling (Lake Michigan Shore)	Riesling	St. Julian

Fig. 1. Sample Data

TABLE II
LEXICON - SAMPLE DATA

Word	Aroma	Score
fig	fruit	1
prune	fruit	1
walnut	nutty	1
hazelnut	nutty	1

2) Case-Folding: Since we use the description text in the next step to categorize into various aromas with a lexicon, it is necessary to do case-folding. In case-folding all letters are reduced to lower case.

3) Special Character Removal: Any special characters in the text field apart from lower case letters are removed. This helps is parsing the text more quickly than with the presence of special characters or symbols.

C. Aroma Lexicon

According to Merriam-Webster dictionary [8], lexicon is a a book containing an alphabetical arrangement of the words in a language and their definitions. In this work we created a lexicon containing the words that describe various aroma in wine using the UC Davis aroma wheel[1]. The Davis Wine Aroma wheel contains words that describe fragrances and flavors found in most wine. According to UC Davis aroma wheel[1] unique fragrances found in wine, are due to the grapes being used in the production of the wine, coupled with the soils and terroir or soil the grapes were planted in and the choices made by the wine maker. The base to UC Davis wine aroma wheel is [9], [10] which was proposed using the list of standardized wine terminology developed by the Sensory Evaluation Sub-committee of the technical projects committee of American Society of Enology and Viticulture. The Fig.2 shows the UC Davis aroma wheel.

Sample data in the lexicon is shown in Table II.

D. Aroma Categorization

The developed lexicon is used to categorize the wine dataset. Fig. 3 shows the overall methodology in pre-process, and categorization.

[1]https://www.thewinecellarinsider.com/wine-topics/wine-educational-questions/davis-aroma-wheel/

Fig. 2. UC Davis Aroma Wheel[1]

E. D3 Visualization

This project used d3.js [11] for visualization purpose. We develop an interactive scatter plot and sunburst (wine aroma wheel).

1) Zoomable Scatter Plot: A scatter plot is a type of plot or mathematical diagram using Cartesian coordinates to display values for typically two variables for a set of data[2].

The sample zoomable scatterplot[3]. is shown in Fig. 4.

2) Zoomable Sunburst: Sunburst is similar to treemap, but uses radial layout. The root node of the tree is at the center, with leaves on the circumference [12] and zoomable sunburst d3 example[4]. Fig.5 shows sample sunburst from website4.

III. EXAMPLE USE CASES

1) Scenario 1 - Shortlist best wines from the available list: A customer want to select the wine based on the review points, price and the country of origin.

The Fig. 6 shows the screenshot of the page on load. We see there are options to choose country, price range and aroma.

[2]https://en.wikipedia.org/wiki/Scatter_plot
[3]http://bl.ocks.org/peterssonjonas/4a0e7cb8d23231243e0e
[4]https://bl.ocks.org/mbostock/4348373

Fig. 3.   Aroma Categorization
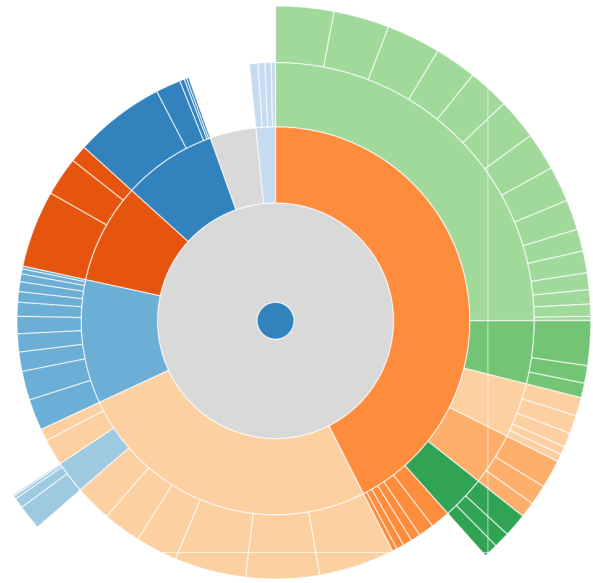

Fig. 5.   Zoomable Sunburst Sample

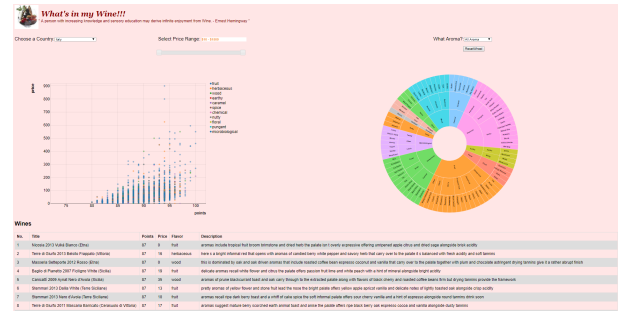Fig. 4.   Zoomable Scatter-Plot Sample3

Fig. 6.   Scenario 1 - On Page Load

It shows the list of descriptions at the bottom of the page. In this Fig. 6 we have the country Italy and list of all the wines available in the dataset.

The Fig. 7 shows how data is updated based on selecting the price range For instance in this case we have the list of wines in the price range $221 to $427.

2) Scenario 2 - What's in my Wine? Wine of my favorite flavor: Based on the shortlisted wines from Scenario 1, the customer then wants to understand the flavors in the wine and choose the one that best suits his/her taste.

The Fig. 8 shows how the scatter plot and wheel gets updated upon clicking on a specific wine in the list. For instance If the user selects a wine with flavor 'herbaeceous', they are able to see the intricate details of the aromas that the wine can contain. The description field is also updated and only displays the specific selected attribute so that the user can read the actual description of the wine by experts.

Fig 9 shows how the page is updated on using the dropdown of aromas. For instance the user wants to see the list of wines with 'woody' flavor. In this case the figure shows the list of wines from Italy, with the price range $221 to $427 and with dominant flavor 'wood'.

IV. CONCLUSION

In this project we developed an interactive d3 visualization that helps naive wine consumers choose the top rated and best wine. Also this tool helps them to understand the wine much better by using the aroma wheel.

V. TEAM MEMBERS CONTRIBUTION

Jaishree Ranganathan - lexicon creation, data cleaning, aroma categorization, sunburst, scatter plot - aroma dropdown,
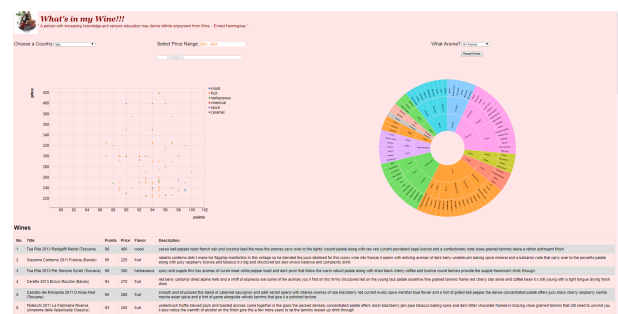
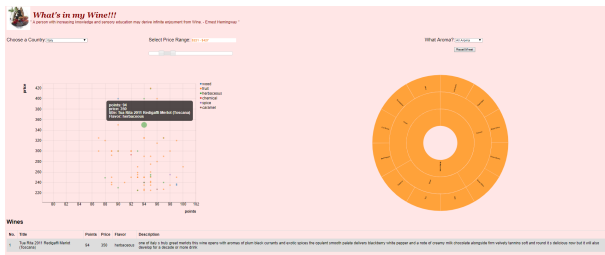Fig. 7.   Scenario 1 - Select Price Range

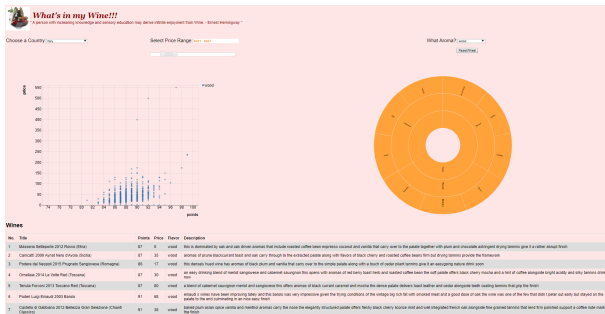Fig. 8. Scenario 2 - Select scatter point and aroma wheel update

Fig. 9. Scenario 2 - Select aroma from dropdown

html page, sunburst and scatter plot integration logical idea, sunburst and scatter plot integration features, presentation, paper report, project proposal.

Surya Pavan Malireddy - scatter plot, sunburst and scatter plot integration features, project proposal.

REFERENCES

[1] M. Pousch, "Wine quotes." [Online]. Available: http://www.enjoyhopewellvalleywines.com/wine-sayings.html

[2] A. Kerren, "Visualization of workaday data clarified by means of wine fingerprints," in *Workshop on Human-Computer Interaction and Visualization*. Springer, 2009, pp. 92–107.

[3] A. Kerren, M. Prangova, and C. Paradis, "Visualization of sensory perception descriptions," in *Information Visualisation (IV), 2011 15th International Conference on*. IEEE, 2011, pp. 135–144.

[4] V. M. Bauer, D. Jankovic, J. Kainer, and S. Schmidt, "Wine and vine statistics visualization tool."

[5] R. Gawel, A. Oberholster, and I. L. Francis, "A mouth-feel wheel: terminology for communicating the mouth-feel characteristics of red wine," *Australian Journal of Grape and Wine Research*, vol. 6, no. 3, pp. 203–207, 2000.

[6] I. Francis and J. Newton, "Determining wine aroma from compositional data," *Australian Journal of Grape and Wine Research*, vol. 11, no. 2, pp. 114–126, 2005.

[7] zackthoutt, "Wine reviews dataset." [Online]. Available: https://www.kaggle.com/zynicide/wine-reviews

[8] M.-W. Dictionary, "Merriam-webster," *On-line at http://www. mw. com/home. htm*, 2002.

[9] A. C. Noble, R. A. Arnold, J. Buechsenstein, E. J. Leach, J. Schmidt, and P. M. Stern, "Modification of a standardized system of wine aroma terminology," *American Journal of Enology and Viticulture*, vol. 38, no. 2, pp. 143–146, 1987.

[10] A. C. Noble, R. Arnold, B. M. Masuda, S. Pecore, J. Schmidt, and P. Stern, "Progress towards a standardized system of wine aroma terminology," *American Journal of Enology and Viticulture*, vol. 35, no. 2, pp. 107–109, 1984.

[11] M. Bostock *et al.*, "D3. js," *Data Driven Documents*, vol. 492, p. 701, 2012.

[12] J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald, "An evaluation of space-filling information visualizations for depicting hierarchical structures," *International journal of human-computer studies*, vol. 53, no. 5, pp. 663–694, 2000.