Visualizing the Flavor of Data

Corina Schweller
MAT 259
W06

INTRODUCTION:

There are numerous flavors found in wine that I will visually map in order to find similarities and differences between them. I have chosen 85 flavors from seven wines in order to explore the relationships between the flavors and wines. The wines selected are commonly found varietals, present in America and particularly in the California region. [1] The *flavors* refer to tastes and smells that stem from the grapes and are a result of the climate, soil, winemaking techniques and the aging process. The data will be presented in the form of the Kohonen Self-Organizing Mapping Algorithm (SOMs).

DATA COLLECTION:

The project will commence with a search for the main varieties of wine and the key flavors. The initial wine varieties and flavors will be chosen from winery tasting notes available on location in the wineries or on the Internet, notable books such as "The Wine Encyclopedia," and numerous web sites including winespectator. com. The flavors of the wines will be selected on the basis of applicability to the wine and flavors that are not associated with the specific wine will be recorded with a zero value. After discovering what prevalent varieties of wine are, I will implement the Internet search engine google.com and type in the wine name and flavor to see how many instances that the key flavor is associated with a certain type of wine. The search words will remain consistent in format in order to attempt to eliminate certain margins of error. Searching for these key words will result in numeric whole numbers that range from a few hundred to four million. The key words will be sorted alphabetically and the corresponding amounts listed in a database.

UNITS OF ORGANIZATION:

The wine varieties will be categorized into seven main groups labeled Cabernet Sauvignon, Chardonnay, Merlot, Pinot Noir, Riesling, Sauvignon Blanc, and Syrah. The 85 prominent flavors, such as honey, apple, cedar, cherry, melon, tar, and pepper will comprise the key words. [2,3] Once the database has been built, I will implement the Kohonen Self-Organizing Mapping Algorithm (SOMs). The Kohonen Mapping system (Figure 1) was invented by Professor Tuevo Kohonen and utilizes self-organizing neural networks in order to reduce the data dimensions. [4,5] Kohonen maps can be defined as a structure of connected units that have data and natural location as the parts of a weight vector. The maps compress data while maintaining the relationships and actually become trained by the data in the process.
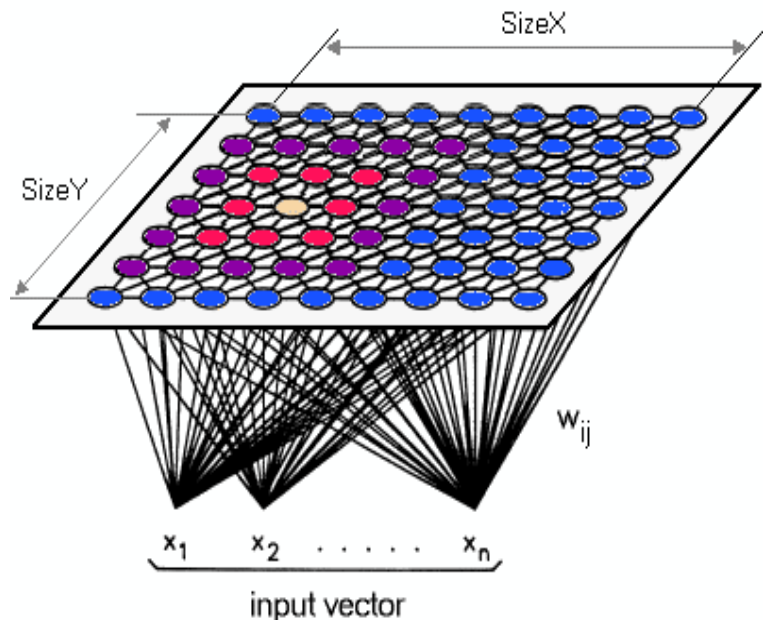

Figure 1 - Simple Kohonen Network [6]

WINE FLAVOR DATABASE

	Anis	Apple	Apricot
CABERNET SAUVIGNON	0	0	0
CHARDONNAY	0	771000	0
MERLOT	0	0	0
PINOT NOIR	0	0	0
RIESLING	0	337000	158000
SAUVIGNON BLANC	0	342000	0
SYRAH	28600	0	0

Asparagus	Basil	Bay Leaf	Bell Pepper	Berry
96400	0	0	53200	454000
0	0	0	0	0
0	0	23100	50800	559000
0	0	0	0	396000
0	0	0	0	0
0	0	0	0	303000
0	41500	0	0	245000

Black Cherry	Black Currant	Blackberry	Butter	Butterscotch
316000	133000	369000	0	0
0	0	0	0	108000
360000	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	194000	0
0	0	199000	0	0

Buttery	Caramel	Caraway	Cedar	Cherry
0	0	0	189000	667000
222000	238000	0	0	0
0	212000	0	261000	0
0	0	915	0	534000
0	0	0	0	0
0	0	0	0	0
0	0	0	921000	346000

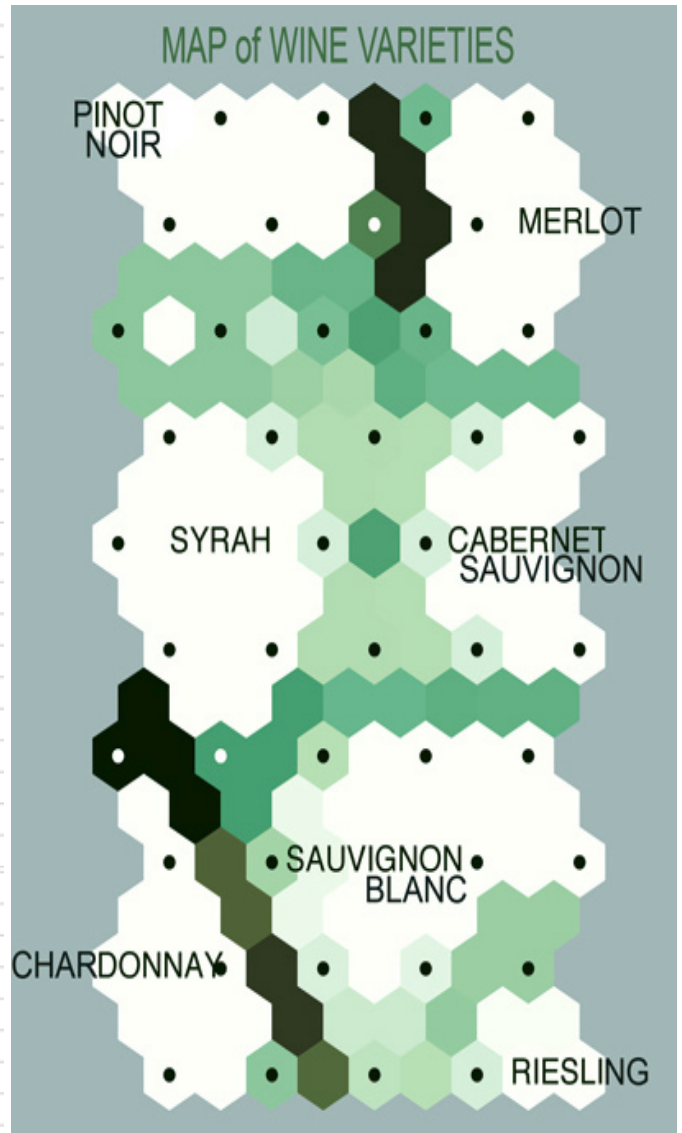*Figure 2 - Database of Wine Flavors and Varieties*

Figure 3 - Kohonen Map of Wine Varieties

KOHONEN VISUALIZATION:

Figure 2 is a segment of the Wine Flavors and Varieties database. The numeric values are from the google.com searches. Flavors that are not part of the tasting experience for a particular wine are recorded with a value of zero. The values obtained in the database will be transferred into the Kohonen Self-Organizing Mapping Algorithm in order to clearly visualize the relationships of the wines to each other. [7] Mapping this data can provide information for exploring the representation of the wine flavor database with other software. Applying the database information to the Kohonen Self-Organizing Map algorithm has garnered the result shown in Figure 3.

READING THE MAP:

The regions denoted in white are 'valleys' and signify similarities between items. The darker regions are referred to as 'hills' and show differences between those items listed within and the rest of the items on the map. Figure 3 shows that Chardonnay is quite different in flavor properties from the other wines. The dark line separating Chardonnay from the rest of the wines indicates that its flavor properties vary from the rest of the wines. There is also a boundary between Pinot Noir and Merlot. Riesling, Sauvignon Blanc, Syrah and Cabernet Sauvignon seem to have the most similar qualities.

WINE FLAVOR DATABASE

7	CABERNET SAUVIGNON	CHARDONNAY	MERLOT	PINOT NOIR	RIESLING	SAUVIGNON BLANC	SYRAH
Anis	0	0	0	0	0	0	28600
Apple	0	771000	0	0	337000	342000	0
Apricot	0	0	0	0	158000	0	0
Asparagus	96400	0	0	0	0	0	0
Basil	0	0	0	0	0	0	41500
Bay Leaf	0	0	23100	0	0	0	0
Bell Pepper	53200	0	50800	0	0	0	0
Berry	454000	0	559000	396000	0	303000	245000
Black Cherry	316000	0	360000	0	0	0	0
Black Currant	133000	0	0	0	0	0	0
Blackberry	369000	0	0	0	0	0	199000
Butter	0	0	0	0	0	194000	0
Butterscotch	0	108000	0	0	0	0	0
Buttery	0	222000	0	0	0	0	0
Caramel	0	238000	212000	0	0	0	0
Caraway	0	0	0	915	0	0	0
Cedar	189000	0	261000	0	0	0	921000
Cherry	0	0	0	534000	0	0	346000
Cigar	0	0	0	140000	0	0	0
Cigar Box	36800	0	41000	0	0	0	0
Cinnamon	0	0	0	140000	0	0	0
Citrus	0	616000	0	0	279000	389000	0
Clove	0	0	99300	0	0	0	0
Cocoa	0	0	0	88800	0	0	0
Coconut	108000	0	181000	0	0	0	0
Coffee	0	0	766000	446000	0	0	194000
Crisp	0	616000	0	0	0	431000	0
Currant	0	0	228000	133000	0	0	0
Diesel	0	0	0	0	29900	0	0
Dust	0	0	0	72100	0	0	0
Earth	224000	0	412000	0	0	0	0
Fig	0	0	0	0	18000	86300	0
Flint	0	0	0	0	0	0	0
Floral	0	399000	340000	0	0	0	0
Fruit	0	1830000	0	1740000	0	0	733000
Ginger	148000	0	0	0	0	0	0
Gooseberry	0	0	0	0	0	25	0
Grape	0	0	0	882000	0	0	0
Grapefruit	0	0	0	0	152000	0	0
Grassy	0	0	0	0	0	184000	0
Green Olive	16900	0	17600	0	0	0	0
Green Peppercorn	12000	0	10600	0	0	0	0
Gunmetal	0	0	0	0	513	0	0
Honey	0	526000	0	0	239000	0	0
Honeysuckle	0	96500	0	0	41900	0	0
Kerosene	0	0	0	0	707	0	0
Leather	173000	0	427000	219000	0	0	0
Lemon	0	644000	0	0	0	359000	0
Licorice	0	0	0	0	0	0	90900
Maple	0	0	0	104000	0	0	0
Melon	0	382000	0	0	0	269000	0
Mineral	0	329000	0	0	0	195000	0
Mint	0	0	0	822000	0	0	0
Mushroom	137000	0	211000	0	0	0	0
Musk	13100	0	0	0	0	0	0
Nutmeg	0	85300	0	0	0	0	0
Oak	1090000	550000	1230000	730000	0	608000	0
Oakey	0	0	65400	0	0	0	224000
Orange	0	0	0	438000	0	0	0
Papaya	0	60400	0	33100	0	0	0
Peach	0	420000	0	0	212000	0	0
Pear	0	382000	0	0	165000	0	0
Pepper	0	0	0	401000	0	315000	0
Pimento	10700	0	0	0	0	0	0
Pineapple	0	301000	0	0	145000	0	0
Plum	0	0	452000	276000	0	0	0
Raspberry	0	0	0	281000	16500	0	179000
Rose Petal	0	0	0	0	0	0	0
Smoke	157000	0	261000	195000	0	0	108000
Smokey	0	189000	0	0	0	0	0
Spice	0	577000	0	0	0	0	0
Steel	0	0	0	0	110000	0	0
Strawberry	0	0	0	251000	0	0	0
Sweet Wood	607	0	572	0	0	0	0
Tar	86500	0	102000	79000	0	0	83200
Tart	0	420000	0	0	0	144000	0
Terpentine	0	0	0	0	95800	0	0
Toast	194000	399000	250000	198000	0	0	101000
Tobacco	0	0	0	0	0	0	106000
Tomato	0	0	0	196000	0	0	0
Tropical Fruit	0	249000	0	0	0	0	0
Truffle	0	0	85500	0	0	0	0
Vanilla	409000	540000	488000	331000	0	0	214000
Violet	0	0	0	952000	76900	0	0
Woodruff	0	0	0	0	489	0	0

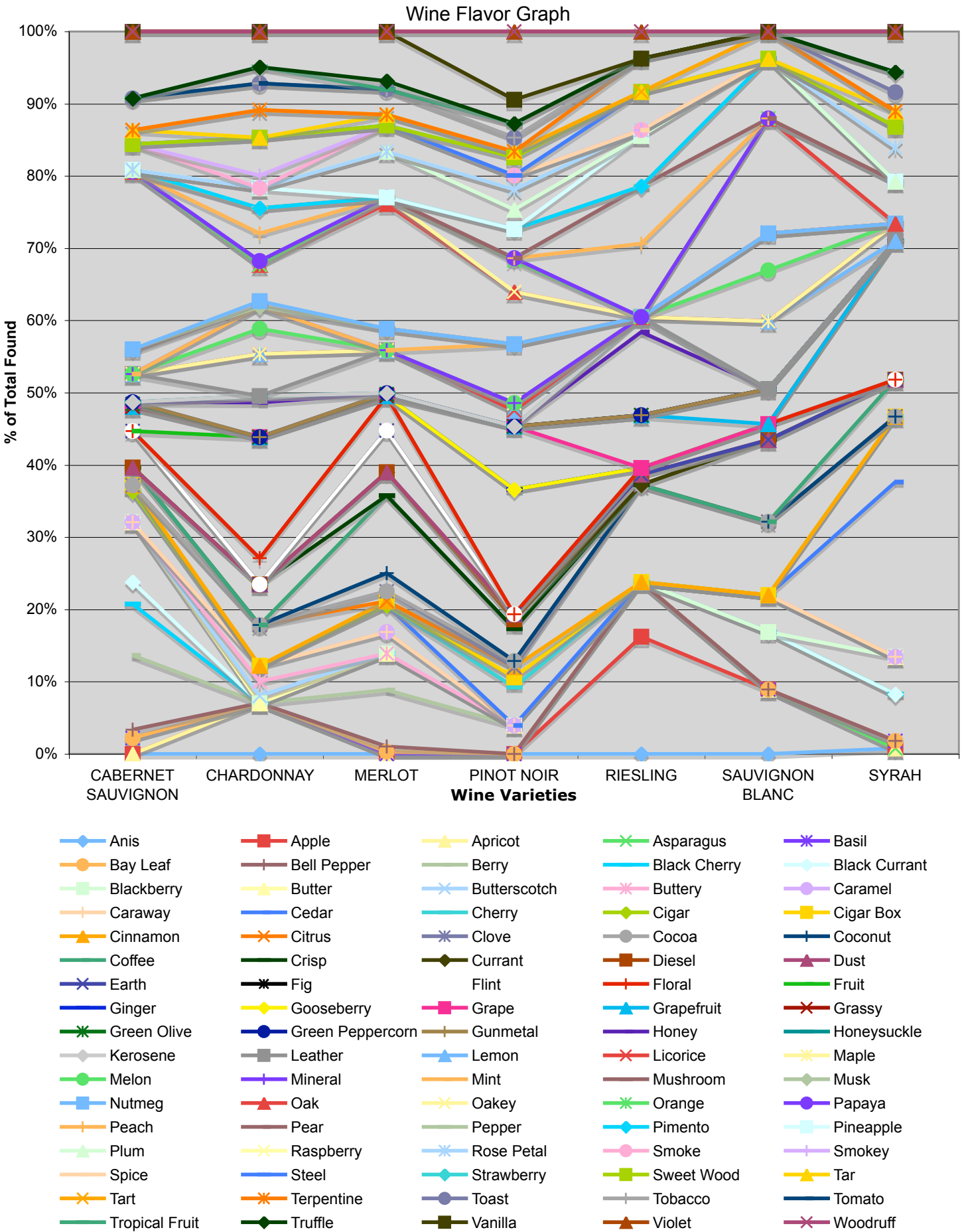*Figure 4 - Database of Wine Flavors and Varieties*

Figure 5 - Graph of Wine Flavors and Varieties

MAPPING FLAVORS:

Changing the database (shown in Figure 4) to map the flavors instead of just the wine varieties shows how the flavors relate to each other across all the wine types. Figure 5 gives a 2D view of the database. The Kohonen map given in Figure 6 shows that the Fruit flavor is distinct from the rest of the flavors. Fruit is found in Chardonnay, Pinot Noir, and Syrah. [8,9] Fruit is associated to Chardonnay 1,830,000 times, 1,740,000 times to Pinot Noir, and 733,000 times to Syrah. It has the highest instance for Chardonnay and Pinot Noir. For all the wines it is present in, Fruit is found a total of 4,303,000 times. The high numeric value for Fruit might be the reason it is in a category of its own, even after running the mapping algorithm many times.

Oak and Coffee are also separated and while Oak is found in most of the wines, at a total of 4,208,000, coffee is only present in the red wines and has a total of 1,406,000 times found. Bell Pepper, Cigar, Bay Leaf, and Asparagus are very similar in value and thus appear next to each other in a row without barriers. Most of these items are green and edible, except for the cigar, and thus this grouping actually makes sense. Anis and Licorice are another grouping that has similarities. They both have a powerfully pungent taste and are both found in Syrah.

Apricot and Steel are an unexpected combination. They both are found solely in Riesling at quantities of 158,000 and 110,000. Another grouping, Melon and Tart are both tasted in Chardonnay and Sauvignon Blanc. I was pleasantly surprised that Peach, Pear and Pineapple clustered together. They are all present in the white wines Riesling and Chardonnay. [10,11]

For all of the wines there is a high amount of Apple, Berry, and Cedar, Toast, Vanilla, Violet, Citrus, Lemon, Cherry, Crisp, Coffee, and Coffee. The wines are associated much less to the flavors Caraway, Anis, Bay Leaf, Basil, Flint, Gooseberry, Kerosene, Musk, Gunmetal, Papaya, Rose Petal, Woodruff, and Sweet Wood.
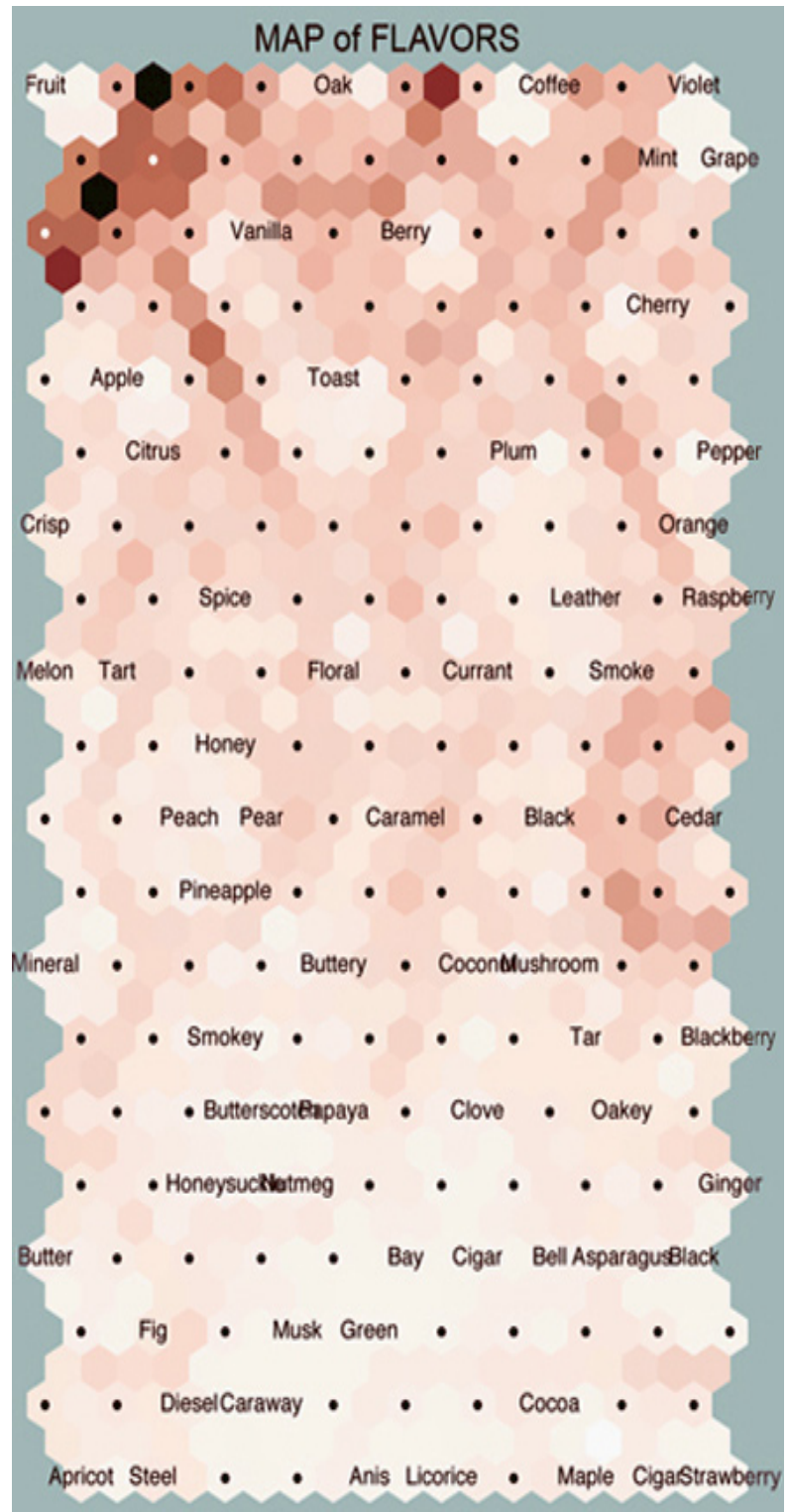

Figure 6 - Kohonen Map of Wine Flavors

There are a lot of similarities in the mappings of the wine flavors using values found on google.com. I will also try different approaches to compiling the wine database and mapping these flavors.

WINE DATABASE (0 and 1)

7	CABERNET SAUVIGNON	CHARDONNAY	MERLOT	PINOT NOIR	RIESLING	SAUVIGNON BLANC	SYRAH
Anis	0	0	0	0	0	0	1
Apple	0	1	0	0	1	1	0
Apricot	0	0	0	0	1	0	0
Asparagus	1	0	0	0	0	0	0
Basil	0	0	0	0	0	0	1
Bay Leaf	0	0	1	0	0	0	0
Bell Pepper	1	0	1	0	0	0	0
Berry	1	0	1	1	0	1	1
Black Cherry	1	0	1	0	0	0	0
Black Currant	1	0	0	0	0	0	0
Blackberry	1	0	0	0	0	0	1
Butter	0	0	0	0	0	1	0
Butterscotch	0	1	0	0	0	0	0
Buttery	0	1	0	0	0	0	0
Caramel	0	1	1	0	0	0	0
Caraway	0	0	0	1	0	0	0
Cedar	1	0	1	0	0	0	1
Cherry	0	0	0	1	0	0	1
Cigar	0	0	0	1	0	0	0
Cigar Box	1	0	1	0	0	0	0
Cinnamon	0	0	0	1	0	0	0
Citrus	0	1	0	0	0	1	0
Clove	0	0	1	0	0	0	0
Cocoa	0	0	0	1	0	0	0
Coconut	1	0	1	0	0	0	0
Coffee	0	0	1	1	0	0	1
Crisp	0	1	0	0	0	1	0
Currant	0	0	1	1	0	0	0
Diesel	0	0	0	0	1	0	0
Dust	0	0	0	1	0	0	0
Earth	1	0	1	0	0	0	0
Fig	0	0	0	0	1	1	0
Flint	0	0	0	0	1	0	0
Floral	0	1	1	0	0	0	0
Fruit	0	1	0	1	0	0	1
Ginger	1	0	0	0	0	0	0
Gooseberry	0	0	0	0	0	1	0
Grape	0	0	0	1	0	0	0
Grapefruit	0	0	0	0	1	0	0
Grassy	0	0	0	0	0	1	0
Green Olive	1	0	1	0	0	0	0
Green Peppercorn	1	0	1	0	0	0	0
Gunmetal	0	0	0	0	1	0	0
Honey	0	1	0	0	0	0	0
Honeysuckle	0	1	0	0	0	0	0
Kerosene	0	0	0	0	1	0	0
Leather	1	0	1	1	0	0	0
Lemon	0	1	0	0	0	1	0
Licorice	0	0	0	0	0	0	1
Maple	0	0	0	1	0	0	0
Melon	0	1	0	0	0	1	0
Mineral	0	1	0	0	1	1	0
Mint	0	0	0	1	0	0	0
Mushroom	1	0	1	0	0	0	0
Musk	1	0	0	0	0	0	0
Nutmeg	0	1	0	0	0	0	0
Oak	1	1	1	1	0	1	0
Oakey	0	0	1	0	0	0	1
Orange	0	1	0	0	0	0	0
Papaya	0	1	0	0	1	0	0
Peach	0	1	0	0	1	0	0
Pear	0	1	0	0	1	0	0
Pepper	1	0	0	1	0	1	0
Pimento	1	0	0	0	0	0	0
Pineapple	0	1	0	0	1	0	0
Plum	0	0	1	1	0	0	0
Raspberry	0	0	0	1	1	0	1
Rose Petal	0	0	0	1	0	0	0
Smoke	1	0	1	1	0	0	1
Smokey	0	1	0	0	0	0	0
Spice	0	1	0	0	0	0	0
Steel	0	0	0	0	1	0	0
Strawberry	0	0	0	1	0	0	0
Sweet Wood	1	0	1	0	0	0	0
Tar	1	0	1	1	0	0	1
Tart	0	1	0	0	0	1	0
Terpentine	0	0	0	0	1	0	0
Toast	1	1	1	1	0	0	1
Tobacco	0	0	0	1	0	0	1
Tomato	0	0	0	1	0	0	0
Tropical Fruit	0	1	0	0	0	0	0
Truffle	0	0	1	0	0	0	0
Vanilla	1	1	1	1	0	0	1
Violet	0	0	0	1	0	0	0
Woodruff	0	0	0	1	0	0	0

Figure 7 - Database of Wine Flavors and Varieties
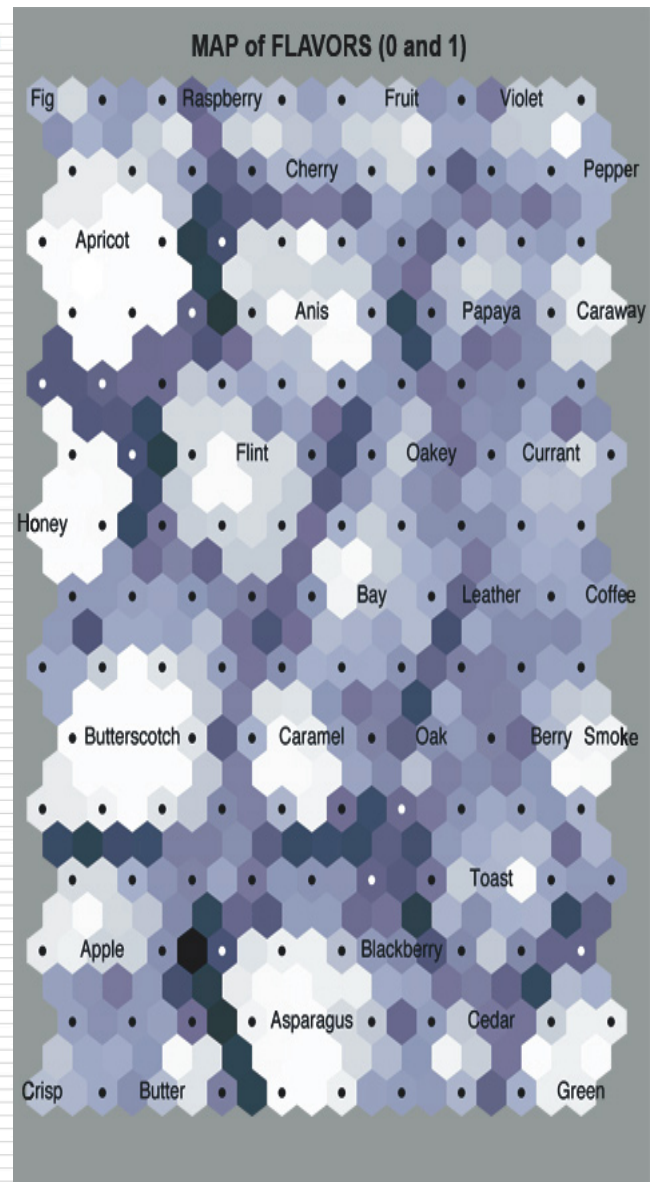
Figure 8 - Kohonen Map of Wine Flavors

0 and 1 MAPS:

The database in Figure 7 reduces the information to simply indicate that a flavor is found in a wine. The "1" indicates that the flavor is actually part of the tasting experience for that specific wine. The "0" signifies that a certain flavor is not part of the wine profile for that particular wine. On the Kohonen maps, lighter colors demonstrate the similarities and darker areas are differences. The maps of the 0 and 1 database (Figure 8 and 9) exhibit an ample amount of differences in the flavors of the wines. This system seems to give a more comprehensive representation of the flavors present in the wines and conveys obvious divisions of the flavors. The problem with this mapping system is that there are too many separations and the connections between flavors are completely lost.
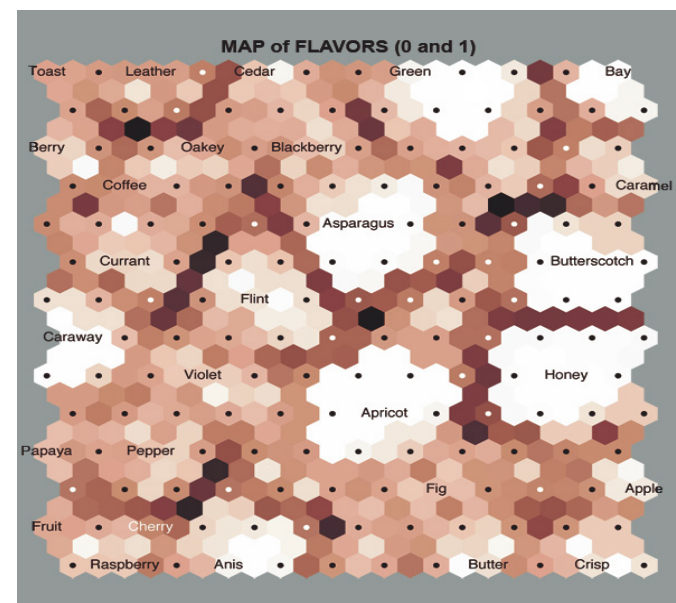

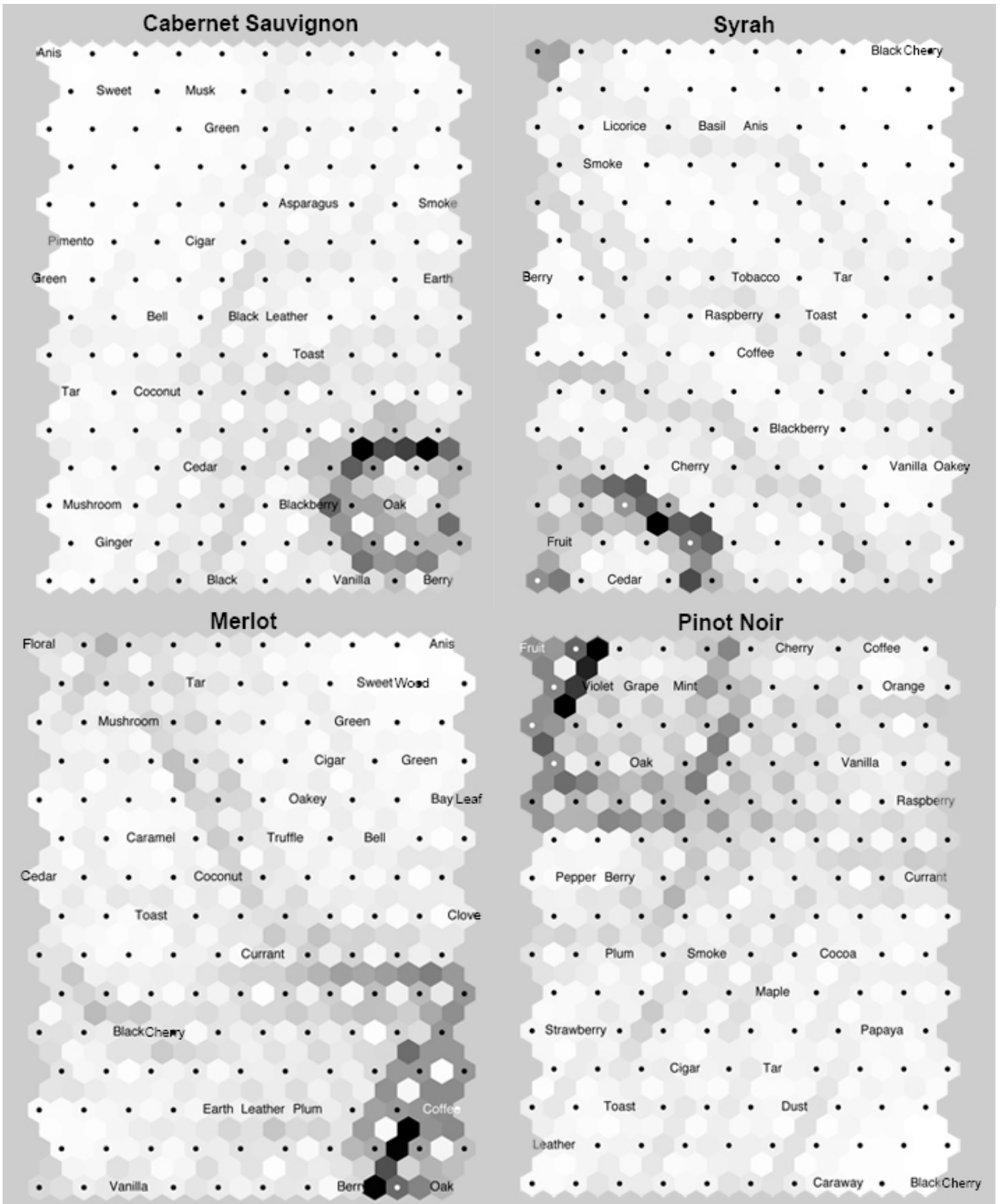*Figure 9 - Kohonen Map of Wine Flavors*

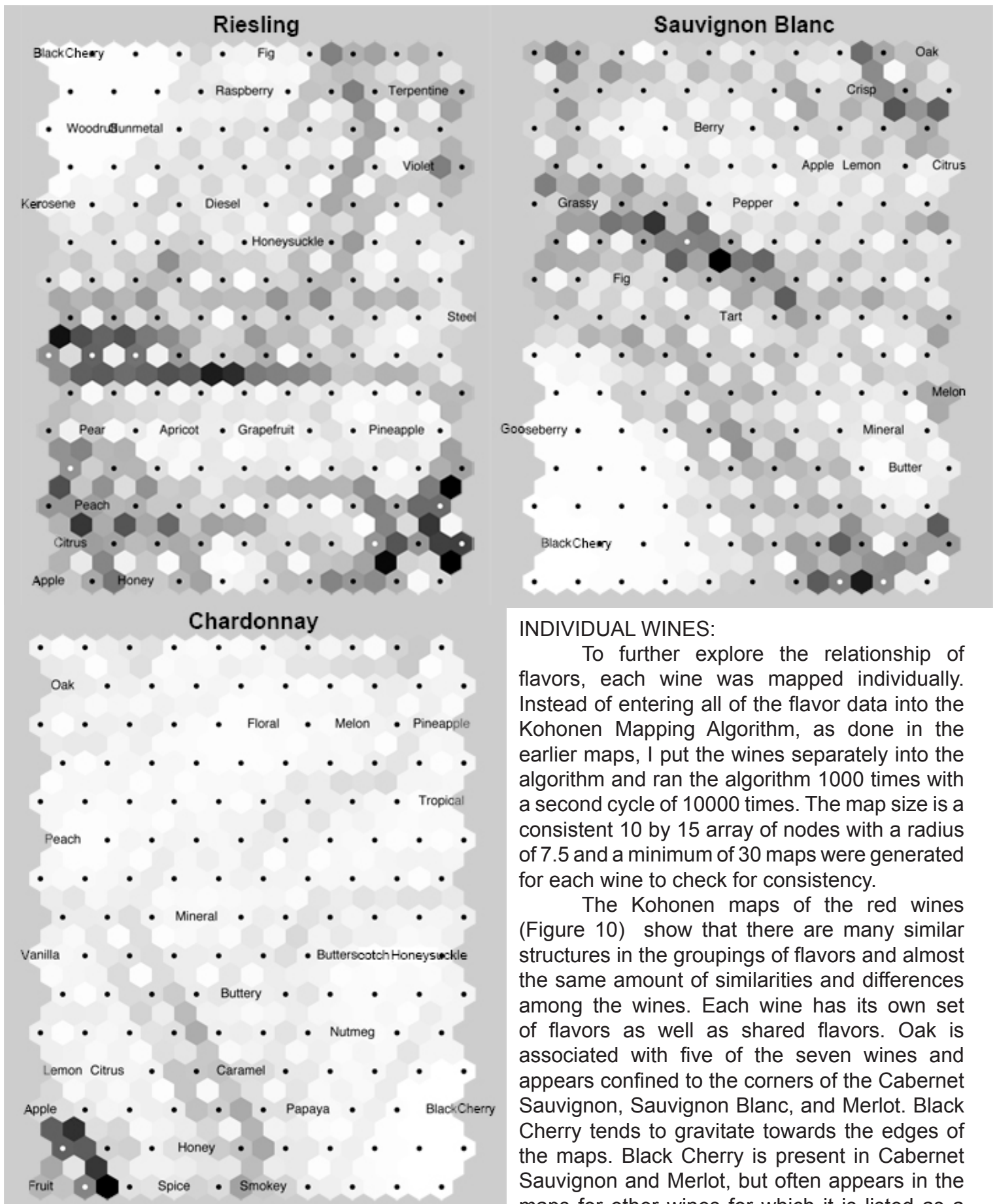*Figure 10 - Kohonen Maps of Individual Red Wine Flavors*

Figure 11 - Kohonen Maps of Individual White Wine Flavors

INDIVIDUAL WINES:

To further explore the relationship of flavors, each wine was mapped individually. Instead of entering all of the flavor data into the Kohonen Mapping Algorithm, as done in the earlier maps, I put the wines separately into the algorithm and ran the algorithm 1000 times with a second cycle of 10000 times. The map size is a consistent 10 by 15 array of nodes with a radius of 7.5 and a minimum of 30 maps were generated for each wine to check for consistency.

The Kohonen maps of the red wines (Figure 10) show that there are many similar structures in the groupings of flavors and almost the same amount of similarities and differences among the wines. Each wine has its own set of flavors as well as shared flavors. Oak is associated with five of the seven wines and appears confined to the corners of the Cabernet Sauvignon, Sauvignon Blanc, and Merlot. Black Cherry tends to gravitate towards the edges of the maps. Black Cherry is present in Cabernet Sauvignon and Merlot, but often appears in the maps for other wines for which it is listed as a zero.
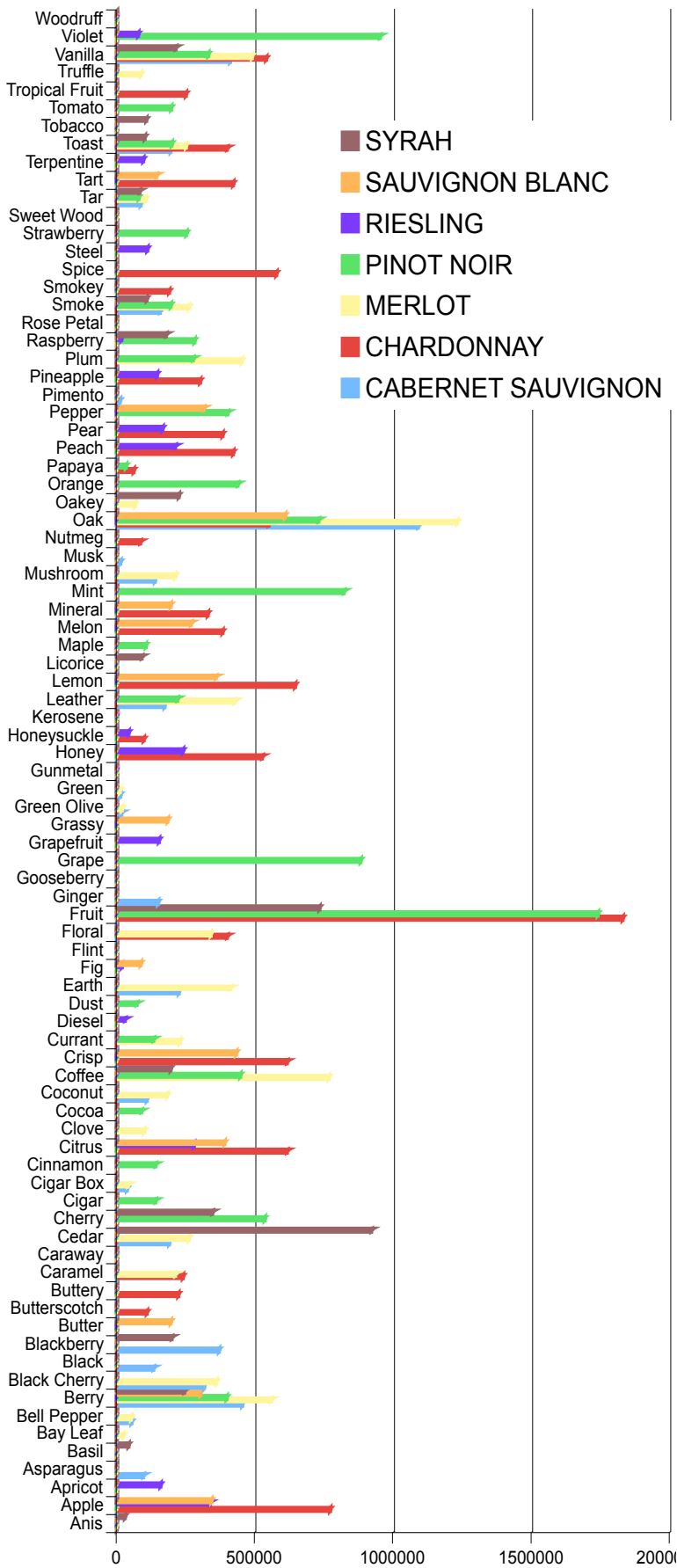
Wine Flavor Graph

Legend:
- **SYRAH**
- **SAUVIGNON BLANC**
- **RIESLING**
- **PINOT NOIR**
- **MERLOT**
- **CHARDONNAY**
- **CABERNET SAUVIGNON**

Flavors (top to bottom): Woodruff, Violet, Vanilla, Truffle, Tropical Fruit, Tomato, Tobacco, Toast, Terpentine, Tart, Tar, Sweet Wood, Strawberry, Steel, Spice, Smokey, Smoke, Rose Petal, Raspberry, Plum, Pineapple, Pimento, Pepper, Pear, Peach, Papaya, Orange, Oakey, Oak, Nutmeg, Musk, Mushroom, Mint, Mineral, Melon, Maple, Licorice, Lemon, Leather, Kerosene, Honeysuckle, Honey, Gunmetal, Green, Green Olive, Grassy, Grapefruit, Grape, Gooseberry, Ginger, Fruit, Floral, Flint, Fig, Earth, Dust, Diesel, Currant, Crisp, Coffee, Coconut, Cocoa, Clove, Citrus, Cinnamon, Cigar Box, Cigar, Cherry, Cedar, Caraway, Caramel, Buttery, Butterscotch, Butter, Blackberry, Black, Black Cherry, Berry, Bell Pepper, Bay Leaf, Basil, Asparagus, Apricot, Apple, Anis

X-axis: 0, 500000, 1000000, 1500000, 2000000

Figure 12 - Bar Graph of Wine Flavors

FLAVORS OF INDIVIDUAL WINES:

The white wines (Figure 11) are much more varied in flavors and groupings. Two of the white wines, Riesling and Sauvignon Blanc, have a lot of differences between the flavors while Chardonnay has many similarities among its flavors. When tasting the wines, you will notice that Chardonnay tends to be smoother in taste and more buttery than Riesling or Sauvignon Blanc. [12] The graph in Figure 12 gives a 2D look at the flavors and supports the groupings created by the Kohonen maps.

SOURCES:

A multitude of sources have provided the research parameters. Tasting Notes from actual wineries were studied along with books such as "Wine Essentials" and "Wine for Dummies." The Internet was a major source in this project, both in defining the Kohonen Self-Organizing Mapping Algorithm and searching for the wine flavors. Well reputed sites, including winespectator.com and winepros.org were perused for building the database key words.

CONCLUSION:

Kohonen Mapping has proven to be an ideal way to visualize the relationships between flavors in wines. Mapping all of the wines and flavors gives an overview of the groupings of flavors. Fruit and Oak are separated from the rest of the flavors and both are present in large quantities in several of the wines. The most insightful maps were those of the individual flavors since they allowed for comparisons between each of the wines. These maps have revealed that the red wines have more in common than the white wines and more importantly, the taste of a wine can be derived from a Kohonen map.

OTHER VISUALIZATION:

Using the information generated with the Kohonen Mapping system, I will create other visualizations of the data by inserting the coordinates from the mapping files into animation software. The flavor data will be displayed in an organic shape that will be placed on a surface based on its relationship to the other flavors. Just like Kohonen Mapping, the distances between shapes will be based upon similarities and differences.

REFERENCES:

[1] Nation Master Encyclopedia.
   Cabernet Sauvignon. http://www.nationmaster.com/encyclopedia/Cabernet-sauvignon
   Chardonnay. http://www.nationmaster.com/encyclopedia/Chardonnay
   Merlot. http://www.nationmaster.com/encyclopedia/Merlot
   Pinot Noir. http://www.nationmaster.com/encyclopedia/Pinot-noir
   Riesling. http://www.nationmaster.com/encyclopedia/Riesling
   Sauvignon Blanc. http://www.nationmaster.com/encyclopedia/Sauvignon-Blanc
   Syrah. http://www.nationmaster.com/encyclopedia/Syrah

[2] Professional Friends of Wine. Tasting Notes.
   http://www.winepros.org/tasting-notes/tastings_list.htm

[3] Wine Spectator Online. Today's Picks.
   http://www.winespectator.com/Wine/Daily_Wine_Picks?lid=DWP&lpos=TopNav

[4] Tom Germano. Self Organizing Maps.
   http://davis.wpi.edu/~matt/courses/soms/#Introduction

[5] Tuevo Kohonen. The Self-Organizing Map (SOM).
   http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm.shtml

[6] Kohonen Networks
   http://www.vias.org/tmdatanaleng/cc_ann_kohonen.html

[7] Kohonen's Self-Organizing Map (SOM).
   http://www.willamette.edu/~gorr/classes/cs449/Unsupervised/SOM.html

[8] E. McCarthy and M. Ewing-Mulligan. Wine for Dummies. Indiana: Wiley Publishing, 2003.
   pg 28-30.

[9] S. Brook and et al. Le Cordon Bleu Wine Essentials. New York: John Wiley & Sons, Inc.,
   2001. page 70-76, 184.

[10] San Francisco Chronicle. THE CHRONICLE'S WINE SELECTIONS
   Monterey County Chardonnay.
   http://sfgate.com/cgi-bin/article.cgi?f=/c/a/2006/01/05/WIGOHGGEE11.DTL

[11] Lynda Abbott. Chardonnay. http://student.santarosa.edu/~labbott/chardonnay.html

[12] Food & Wine. http://www.foodandwine.com/search/wine.cfm?order=&kw=syrah&regions=
   9999_9999&p=9999&q=9999&t=9999&v=9999