A Multivariate Analysis of Brewery Draft Beer Sales in Bozeman, MT

Christopher R. Barbour

May 1, 2014

Contents

1 Introduction

1.1 Motivation and Data Collection

Local breweries provide a relaxing atmosphere for friends, family, and colleagues to gather and share stories, collaborations, and memories. It is in the interest of the local breweries to provide types of beer that appeal to the community at all times of the year. To inform Bridger and Bozeman Breweries about which months corresponded to higher sales in their seasonal products, I used Principal Component Analysis (PCA) and

Cluster Analysis to see which groups of months had similar sales profiles and which months had higher proportions of seasonal sales.

Bridger Brewery opened in March of 2013 and has steadily become a hangout spot for local residents and college students alike. They serve many varieties of craft beer and pizza, either whole or by the slice. They are also involved in the community and host *Pints with Purpose*, a fundraiser each week where the profits go to a local non-profit agency. Bozeman Brewery opened in 2001 and is an established brewery in the local community. They distribute products throughout Montana and serve seasonals and trademark products in their tasting room. Bozeman Brewery also hosts a Sunday FUNDay, where a local non-profit in Montana receives a portion of the sales from the pints sold.

I collected monthly data for 2013 from Bridger Brewery and Bozeman Brewery. The Bozeman Brewery data set contained the sales for its Hefe, Porter, Amber, IPA products and, a final group of all other products grouped into "Seasonal". The Bridger Brewery data set contained sales on all the different kinds of beer they sold since they opened in March of 2013. A number of these were excluded from the analysis. For example, Abbey Ale is a mistake in the system they use to ring up sales, so this was excluded. Batch 8 was supposed to be a strong, dark IPA that, due to equipment failure, never reached maturity, so it was excluded. In total, 22 beer products were retained for the analysis. I will also consider a subset of the Bridger Brewery data, where all the products that are not sold year round (all but the Blonde, Brown, and IPA) are combined into a Seasonal variable.

Looking at the total sales at Bridger Brewery in Figure 1 over the past year, June and July were the slowest months (could be due to a drop in population on campus). The hours of operation for Bridger Brewery remained the same throughout these slower months. Popularity seems to have increased this fall, with higher beer sales associated with October through January.

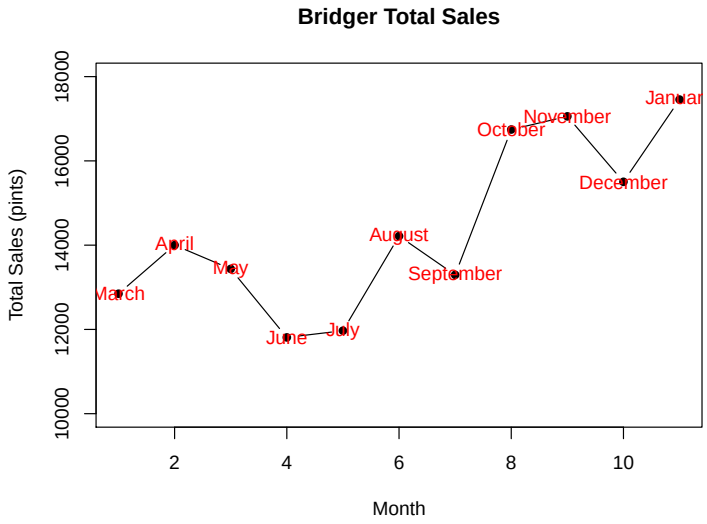


Figure 1: Time series plot of total sales(pints)

1.2 Graphical Methods

One of the primary issues using graphical methods in multivariate analysis is with displaying data that are in at least four dimensions. There have been a number of plotting techniques that have been developed to address this issue, such as Chernoff-faces, Andrew plots, and parallel coordinates plots (see Everitt and Hothorn, 2011 and Hardle and Simar, 2012).

The graphical methods used to display the final results of the cluster analysis incorporate a time series scatter plot of the total sales versus the months the data was collected, as well as the Parallel coordinates plot, which re-scales all variables between 0 and 1, and then plots each observation as a line going across the different variables. With only 11 months of observations for the Bridger Brewery and 12 for the Bozeman

Brewery, this seems to be a reasonable display. With too many observations, these plots can seem very messy and hard to extract the interesting information. These plots are also sensitive to the ordering of the variables, so incorporating information from other analyses, such as Principal Component Analysis and clustering algorithms, can enhance the information available from the plot.

2 PCA

2.1 PCA Introduction

The main idea of principal components analysis (or PCA) is to take q variables $x_1, \ldots x_q$ and define a new set of variables $x_1^*, \ldots x_u^*$ with $u \leq q$ that are linear combinations of our original variables that explain most of the variation observed in the original explanatory variables. Each of the new principal components will be uncorrelated with the others. PCA can be useful in situations where our variables are correlated (which is the case in our data, because we would expect months close together in time to be more correlated). In the full Bridger data set, we have 11 months and 22 types of beer, so it is possible to get some significant dimension reduction in the case where we are examining (clustering) months based on the types of beer sold that month.

2.2 Calculating Principal Components

Since we are finding linear combinations of our original variables, the first principal components will be:

$$x_1^* = c_{11}x_1 + c_{12}x_2 + \ldots + c_{1q}x_q$$

where we choose the c_{1i}'s to maximize the variation of x_1^* as well as satisfy the constraint $\sum_{i=1}^{q} c_{1i}^2 = 1$. Using lagrange multipliers and linear algebra results, this is achieved by finding the eigenvector of the correlation matrix $\mathbf{\Sigma}$ associated with the largest eigenvalue (Everitt and Hothorn, 2011 and Hardle and Simar, 2012). Likewise, the subsequent components are calculated from the eigenvectors associated with the decreasing eigenvalues.

It can also be shown that the proportion of the variance explained in the data by the jth principal component is simply the eigenvalue associated with that component divided by the sum of all the eigenvalues of $\mathbf{\Sigma}$.

2.3 Using the prcomp function

The singular value decomposition (SVD) in matrix theory (see Meyer 2000) can be used to calculate the principal components. In the situation where a matrix X is symmetric and square, the singular value decomposition is analogous to solving the eigenvalue problem (see Wall et al. 2003). Therefore, the SVD for a covariance matrix $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = U\Lambda U^T$$

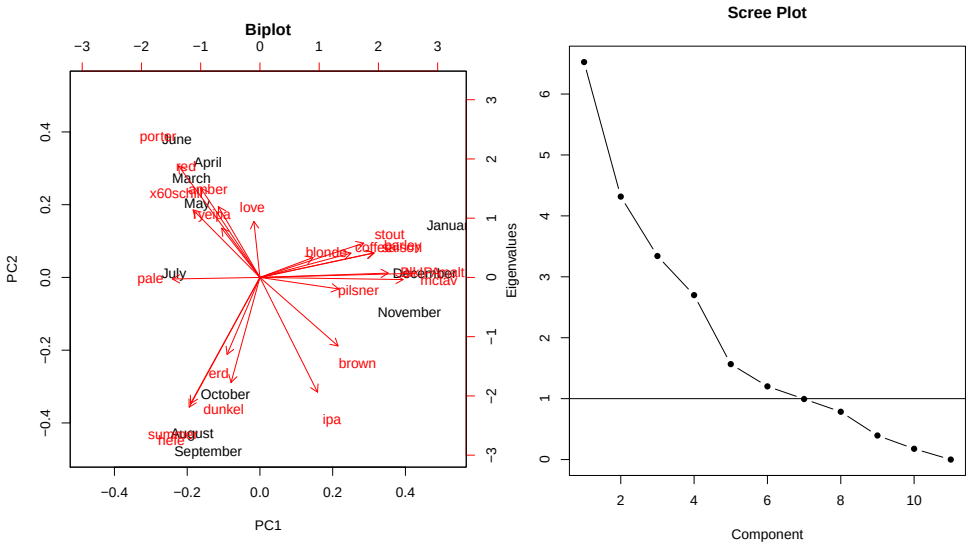
where Λ has the eigenvalues of $\mathbf{\Sigma}$, in order from largest to smallest (traditionally) on the diagonal, and zeros on the off diagonal. The matrix U contains eigenvectors in its columns that are associated with the eigenvalues in Λ. The R function that I used to perform PCA is `prcomp`, which uses a singular value decomposition of the covariance matrix $\mathbf{\Sigma}$ to calculate the principal components.

2.4 The Scree Plot and the Principal Components Biplot

Choosing the number of principal components to retain for interpretation and future analysis is a non-trivial task. There are a number of rules that have been developed in an attempt to answer this question (see Everitt and Hothorn 2011 and Raiche et al. 2006). One of the earliest developed is the *scree plot*, which is simply a plot of the eigenvalues of the PC's versus the number of the component. Since these eigenvalues represent the proportion of the variation that is explained by that PC, we can consider looking for the PC where the slope to the subsequent eigenvalue becomes "shallow" instead of "steep". Another term for this

is looking for the "elbow" of the scree plot. A scree plot of the PCA for the Bridger data is shown in Figure 2. At the 5th component, we see that the steepness of the lines connecting components starts to become slightly "less steep". These 5 components also account for approximately 84% of the variation in the original variables.

A plot that can be used to display the observations and the variables in two dimensions is the principal components biplot (see Venables and Ripley 2002.). I will focus on the interpretation of this plot in the context of the problem. Details of how this plot is constructed will be left to the reader.




```
## Importance of components:
##                           PC1   PC2   PC3   PC4    PC5    PC6    PC7    PC8   PC9    PC10     PC11
## Standard deviation     2.554 2.078 1.828 1.643 1.2516 1.0961 0.9969 0.8858 0.628 0.42096 5.36e-16
## Proportion of Variance 0.297 0.196 0.152 0.123 0.0712 0.0546 0.0452 0.0357 0.018 0.00805 0.00e+00
## Cumulative Proportion  0.297 0.493 0.645 0.767 0.8386 0.8932 0.9383 0.9740 0.992 1.00000 1.00e+00
```

Figure 2: The principal components biplot (left) and the scree plot (right) is shown.

Further discussion of the following "rules" for interpreting these plots can be found in Venables and Ripley (2002), Everitt and Hothorn (2011), and Gabriel (1971). The direction of the vectors for each variable in the biplot shows the contribution of the variable to the principal components. The length of the vectors represent the standard deviation of that variable with longer vectors being associated with variables with more variation in the observations. The cosine of the angle θ between two vectors represents the correlation between the two variables ($cos\,(\theta) \approx \rho$), so small angles represent strong positive linear associations and angles close to 180^0 reflect strong negative linear associations. From this plot, we see that the sales of Erd (strawberry Hefe), dunkel, summer, and hefe were highly linearly associated in the first 11 months of operation at Bridger. If we think about running a line from an observation to the vector of a variable that is perpendicular to that vector, we can approximate the value of that variable for the observation relative to the other observations in the data. Therefore, we can see that the months of September, October, and November had the highest sales of erd, dunkel, summer and hefe products.

The euclidean distance between the observations in the biplot is an approximation of the distance between them in the multivariate space, giving us a representation of how far apart the observations are in multivariate space. We see three distinct groups of points coming out of the biplot. August, September, and October are very close to one another on the plot. Similarly, we see groups of points in the upper left (March through July) and in the left side (November through January). Although this has been an exploratory method thus far, it has allowed us to begin to examine the structure of the data and some possible relationships between the products sold and the time of the year.

2.5 Interpreting Principal Components

When interpreting the principal components, a cutoff of $|0.3|$ was used for the loadings of each variable to aid in interpretation. The first principal component (PC1) had relatively large absolute loadings for malt liquor (0.35) and mctavish scotch ale (0.33). This can be thought of as a weighted average between malt liquor and mctavish ale and large values of these products correspond to a large PC1. The second principal component (PC2) had relatively large absolute loadings for hefe (-0.39), dunkel (-0.32) summer ale (-0.37), and porter (0.3). This can be thought of as a weighted contrast between hefe, dunkel, and summer ale versus porter and low sales of the hefe, dunkel, and summer ale and high sales of porter correspond to small values of PC2.

The third principal component (PC3) had relatively large absolute loadings for coffee stout (-0.43) and love ale (-0.32). This represents a weighted average between coffee stout and love ale. The fourth principal component (PC4) had relatively large absolute loadings for stout (-0.32), amber (-0.39), pale ale (-0.31), and 60 schilling (0.44). Likewise, PC4 corresponds to a weighted contrast stout, amber, and pale ale versus 60 schilling. The fifth component (PC5) had relatively large absolute loadings for red ale (-0.42) and ryeipa (0.63). PC5 represents a weighted contrast between red ale versus rye ipa. Although exact interpretations of these PCs may be difficult, we can use the variables with large loadings to possibly distinguish differences in the sales profile across the months to look for similarities.

3 Distance and Multidimensional Scaling (MDS)

Multidimensional scaling (MDS) is a method of projecting information in three or more dimensions in a lower dimensional representation. The goal is to preserve the multidimensional distances, or dissimilarities, in the multivariate observations and construct a new, often two-dimensional map, to display. Multidimensional scaling starts with a distance, or "dissimilarity", metric between the observations. Choosing a distance measure is no small task. There are many that can be used such as the Minkowski distance metric to metrics that can incorporate missing data. We will be using Euclidean distance to define our distance between observations. There are two types of MDS considered in this paper, metric MDS (also known as classical MDS) and non-metric MDS (specifically Kruskal's non-metric MDS). For more information on these, see Everitt and Hothorn (2011) and Cox and Cox (2008). Since we are working with Euclidean distance, metric MDS will produce the same map of new coordinates as seen in the principal components biplot (Hardle and Simar 2012, page 404), while non-metric MDS assumes a less rigid relationship between observed dissimilarities and distances between coordinates in the new mapping and can sometimes reproduce differences in higher dimensions better in two-dimensions than classical MDS.

3.1 Metric MDS

Metric Multidimensional scaling is concerned with finding a mapping in q dimensions from the original observations in n dimensions, preferably with $q < n$, such that the distance between the mapped observations are close to the original distance matrix. To perform this, we take our matrix of observations X, perform a spectral decomposition of the outer product matrix XX^T, resulting in

$$XX^T = V\Lambda V^T$$

where Λ is a matrix with eigenvalues on the diagonal and V being the matrix of corresponding eigenvectors. In fact, since we are working with Euclidean distance, these are the same eigenvalues and eigenvectors seen in the principal components analysis. The new coordinates are computed as $X = V\Lambda^{0.5}$. Similar to PCA, we can compute the variation of the original variables explained in a subset of the dimension $m < q$ by computing $\frac{\sum_{m<q} \lambda_m}{\sum_n \lambda_n}$. However, the metric MDS is restrictive in the assumption of the direct relationship between dissimilarities and constructed distances (Everitt and Hothorn, 2011 p. 406). Non-metric MDS was developed in order to assess this issue.

3.2 Non-Metric MDS

In Non-metric MDS, the rank order of the dissimilarities is the focus. This assumes a rank order of the original dissimilarities as

$$\delta_{i_1 j_1} < \ldots < \delta_{i_n j_n}$$

for $i \neq j$ where $\delta_{i_1 j_1}$ corresponds to the pair of observations with the smallest dissimilarity and $\delta_{i_n j_n}$ corresponds to the pair with the largest dissimilarity. The idea of non-metric MDS is to map the observations into new coordinates to where the distances between the new coordinates match the rank order of the original dissimilarities

$$d_{i_1 j_1} < \ldots < d_{i_n j_n}$$

where d_{ij} represents the distance between observations i and j in the new coordinate system. In order to to achieve this monotonic requirement, we model the distances as:

$$d_{ij} = \widehat{d_{ij}} + \epsilon_{ij}$$

where $\widehat{d_{ij}}$ are called the disparities. This is achieved with isotonic regression, which was originally used by Kruskal. In isotonic regression, we first need an initial configuration of d_{ij}, which can be achieved using metric MDS. We then compare the rank order of the dissimilarities with these distances. At each step of the algorithm, we partition the dissimilarities into blocks and average over the partitions where our montotinicity is violated. This process is repeated until we have achieved a set of disparities that does not violate the monotonicity. (See Cox and Cox, 2008)

Non-metric MDS then looks to minimize the "stress" between two variables, which in essence is a measure of agreement between disparities and distances (dissimilarities). The stress is defined as:

$$STRESS\left(\mathbf{d}, \widehat{\mathbf{d}}\right) = \frac{\sum_{i<j}\left(\widehat{d_{ij}} - d_{ij}\right)^2}{\sum_{i<j} d_{ij}^2}.$$

`isoMDS` is an R function from the `MASS` package (Venables and Ripley 2002) that uses istonic regression in its algorithm to perform non-metric MDS. The basic structure of the algorithm is:

1. Rank order the initial dissimilarities.
2. Compute the set of distances between the points in an initial configuration using classic MDS.
3. Use the isotonic regression algorithm described above to compute disparities. Compute the initial value of the stress.
4. Use an iterative gradient search to find a configuration of new distances d_{ij} to decrease the stress measure.
5. Repeat steps 3 and 4 until the minimum stress value is found between the observations.

The results from the Non-Metric MDS is shown in the section below alongside the classical MDS solution.

3.3 Minimum Spanning Trees

Another method that can enhance MDS is a minimum spanning tree. A spanning tree is a set of line segments connecting all the points in (potentially) multiple dimensions that uses the following rules:

- Straight lines connect the points.
- Every point is visited once.
- There are no closed loops, i.e., there exists a unique path between any two points in the tree.

The minimum spanning tree is the spanning tree that has the minimum total segment length. A minimum spanning tree can be used to find possible distortions in the original dissimilarity matrix for a 2D MDS representation. In Figure 3, the metric MDS and non-metric MDS solution is given with a minimum spanning tree. The `mst` function in the `ape` package (Paradis et al. 2004) was used to produce the minimum spanning tree and the `cmdscale` function was used to perform the metric MDS. For the non-metric MDS, although we can see that April is very close to June and May in this representation, the tree runs from April to July and April to March. This indicates that the distance between April and July in this representation may not adequately portray the dissimilarity in the original distance metric. In the metric MDS, we see the distance between the April and July is relatively close to the distance between April, at least more so than in the non-metric MDS. This, while not a perfect portrayal of the dissimilarities between the months, gives us a map to examine the observations and eventually compare the solutions from our cluster analysis (See Everitt and Hothorn, 2011).

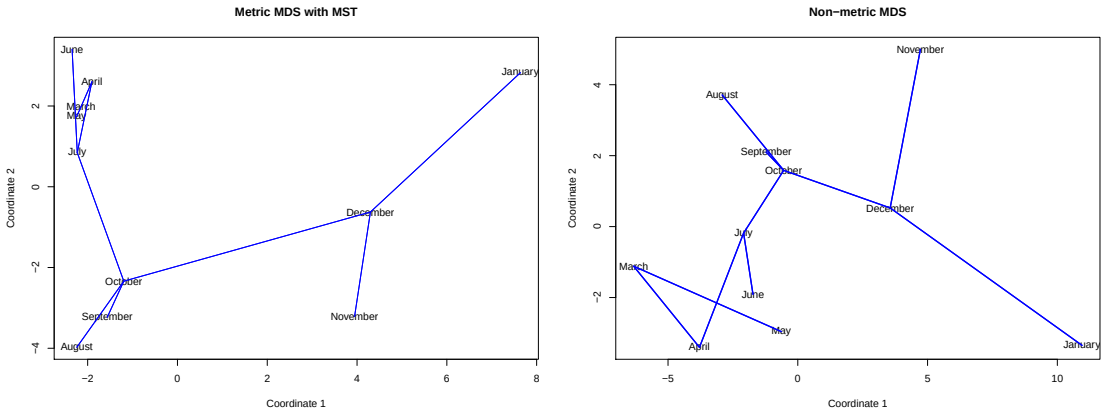


Figure 3: Classical MDS with Minimum Spanning Tree

4 Cluster Analysis

Cluster analysis is an exploratory technique to group observations that are similar and separate observations that are different (in some sense). Sometimes cluster analysis can be an useful method when groups are suspected in the data but are not identifiable by a single variable. For example, we may be interested in grouping potential voters for an election based on a number of characteristics without having access to their voting results.

Recently, a study used cluster analysis in forming different groups among college students dependent on there alcohol, tobacco, and drug use at different universities in the US (Primack et al., 2012). Sometimes these methods can be criticized because they impose groups even if none exist. The usefulness and interpretation of the cluster solution can be the ultimate judgment (Everitt and Hothorn, 2011). Two forms of cluster analysis were used in the following analysis. The first examined is Hierarchical Agglomerative Clustering, which creates partitions of the data starting with each item as its own cluster and running to a single cluster of all observations. This is a simple clustering method which allows us to see group membership for different number of total clusters on the same plot (see section 4.1.1). The second we will examine is a divisive method called Partitioning Around Medoids. This method has an objective function that is maximized for a pre-specified number of clusters and gives an actual observation as a representative for each cluster.

The clustering algorithms discussed are all based on a distance, or dissimilarity matrix between the observations. To help ensure that our cluster solution is not solely based on the increase in total sales, I converted the observations to a proportion of the monthly sales attributed to that product.

4.1 Hierarchical Agglomerative Clustering

The basic Hierarchical clustering algorithm begins with a dissimilarity matrix D and using the algorithm:

1. Start with every item as a cluster (N clusters).
2. Find the pair of clusters, A and B, nearest to one another using a measure of distance between clusters d_{AB}.
3. Merge clusters A and B and update your distance matrix.
4. Repeat 2 and 3 until only one cluster is left.

How we define our distance d_{AB} can have profound impacts on the cluster solutions. Some common linkage methods that are used are single, complete, average, and Wards. For single linkage we use $d(A, B) = min(d_{ij})$ for observations $i \in A$ and $j \in B$. These d_{ij} are the measurements from the original distance matrix. Similarly, the complete linkage method uses $d(A, B) = min(d_{ij})$ for observations $i \in A$ and $j \in B$. The average linkage method incorporates more information; summing over all the distances between clusters A and B and dividing by the product of the sample sizes for each cluster.

Wards method utilizes a reducing sum-of-squares approach in combining clusters. The sum-of-squares for each cluster is the sum of the squared deviations from the centroid of the cluster. At the beginning, each observation is its own cluster and the total sum-of-squares is 0. Each step combines the clusters that produce the smallest increase in the total sum-of-squares. See Everitt and Hothorn (2011) and Hardle and Simar (2012) for more details.

4.1.1 Displaying Results with a Dendrogram

Then we can display these results using a *dendrogram* (Figure 4) to assist in deciding on the number of clusters to keep. We can cut the dendrogram at a particular height in order to get each group membership. For example, a four cluster solution was found for each of the linkage methods. An informal way of picking the number of clusters is to cut the dendrogram at where there is an appropriate difference in the height (Everitt and Hothorn, 2011). This is more of a subjective choice and it may be more beneficial to cut where it will create a number of groups that is believed to be present in the data. For example, the data were collected over four seasons, indicating that it may be appropriate to cut the dendrogram to achieve a 4 cluster solution. The cluster solution presented in the results section used a different clustering algorithm, which is described in the next section.

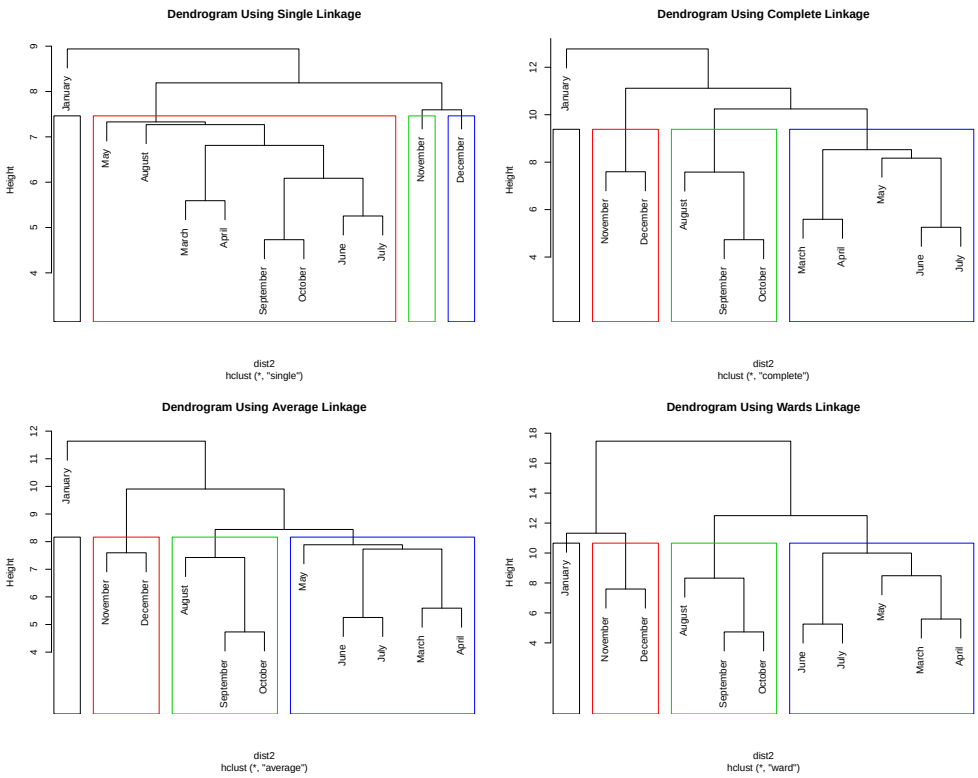


Figure 4: Dendrograms of 4 cluster solutions using single (upper left), complete (upper right), average (lower left), and Wards (lower right) linkage methods

4.2 Partitioning Around Medoids (PAM)

Partitioning around Medoids, or PAM, is a divisive clustering algorithm used for clustering observations into a pre-specified number of clusters k. This implies that we are looking for ways to split the data based on the overall dissimilarities in the observations. For PAM, the clustering algorithm is based on centering the observations around observations in the data that are representative of the clusters, called medoids by Kaufman and Rousseeuw (1990). These representatives can be more robust measures of center than centroids.

The `pamk` function in the `fpc` package (Hennig 2014) performs this clustering algorithm, which has two steps. The first is the BUILD PHASE, which chooses the medoid (representatives) for a pre-specified number of clusters. The first object picked is the observation where the sum of the dissimilarities to all other objects is as small as possible. For each new representative, the following algorithm is used:

- Consider an observation i that has not been selected.
- For each j, compute the dissimilarity between j and the previous representative that j is most similar to. Call this d_j^*.
- Compute the dissimilarity between i and j. Call this d_{ij}.
- Compute $D_{ij} = max\left(d_j^* - d_{ij}, 0\right)$.
- Do this for every non-selected object j and compute $\sum_j D_{ij}$.
- Choose the observation i that has the maximum $\sum_j D_{ij}$.

The second phase is called the SWAP PHASE, where observations are grouped into the clusters that minimize the sum of the distance (or dissimilarity between) between observations and the representative of the group. Details of this algorithm can be found in Kaufman and Rousseeuw (1990).

4.2.1 Average Silhouette Width

Since k has to be specified in advance to employ PAM, it is useful to have diagnostics to aid the decision making process regarding how many clusters to use. A method suggested to measure the fit of a cluster solution is called the average silhouette width. Let's assume that observation i is an element of cluster A. Let $a(i)$ be the average dissimilarity between i and all other observations in cluster A. For every other cluster B compute the average distance between observation i and observations in cluster B. The minimum of these distances we will call $b(i)$. The silhouette width for each observation is defined as

$$s(i) = \frac{b(i) - a(i)}{max(a(i), b(i))}.$$

If the $s(i)$ are close to 1 indicates that the observation is well classified in its cluster. If $s(i)$ is close to 0, then it is not clear which cluster the observation belongs to. If $s(i)$ is close to -1, than observation i would be better classified in the neighbor cluster rather than the current cluster. Kaufman and Rousseeuw suggest using the overall average of these silhouettes can be used to determine the optimal number of clusters.

For the full Bridger Brewery data, the left panel of Figure 5 displays the plot of average silhouette width versus number of clusters k, constructed for $k = 2, \ldots, 8$ on the x-axis and the average silhouette width of all observations $\overline{s}_k$ on the y-axis. We can see that the maximum occurs when the number of clusters equals 3. We can also see the individual silhouettes for each observation in the silhouette plot (right panel). This 3-cluster solution for the Bridger Brewery data is discussed below in Results and Discussion section.

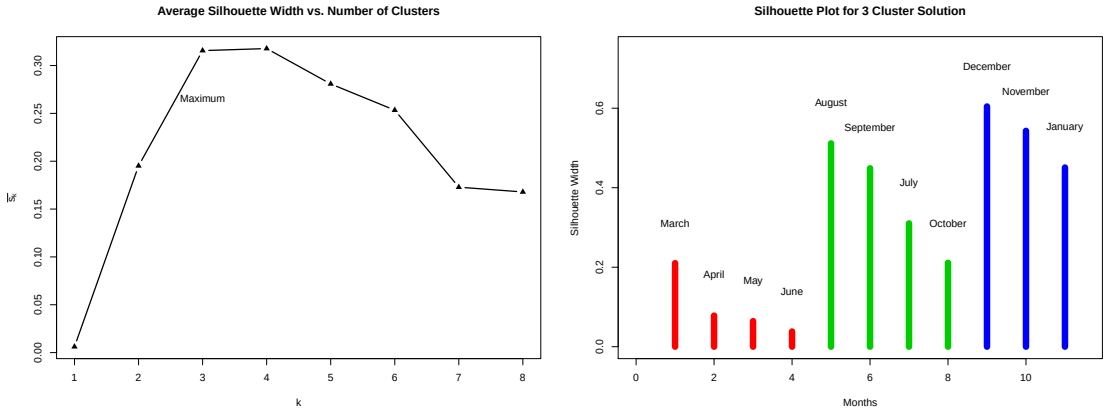


Figure 5: Average silhouette width versus cluster size (left) and silhouette plot for 3-cluster solution (right).

5 Results and Discussion

5.1 Bridger Brewery

5.1.1 Full Data

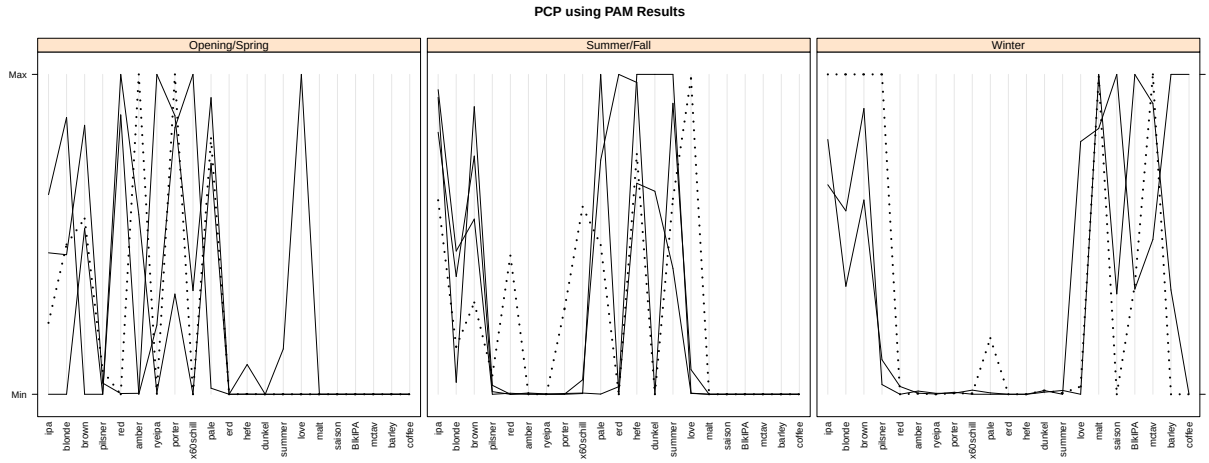


Figure 6: Cluster Analysis Solution Displayed with a PCP

Using the PAM results, we display the data with a PCP, separating each cluster with a different panel (Figure 6). Initially, this was a very messy plot that was hard to digest. However, using the information from the principal component analysis, I was able to sort the variables in a way that a pattern between the clusters started to emerge.

The "Opening and Spring" cluster that encompasses March through June are associated with pilsners, ambers, and red ale, rye ipa, and porters. This was a period when the brewery was first opening and were trying to get a foot in the door with customers in the community. Many of these products are no longer available.

The "Summer and Fall" cluster that encompasses July through October seems to be associated with higher sales of summer ale, hefe, pale ales, and dunkels. This was also the period where Erd was available, which is a strawberry wheat beer that is traditionally German brewed. This specific product was the creation of the owner. These are lighter, less bitter beers that one might associate with warmer, sunnier weather. This is also around the time that the new head brewer, Daniel Pollard, began working at the brewery.

The "Winter" cluster corresponding to November, December, and January is associated with higher sales in the Mctavish scotch ales, coffee stout, stout, malt liquor, barley wine, and black IPA. These are darker, stronger beers with higher bitterness and alcohol content. These are traditionally beers you would associate with colder temperatures and times with less daylight.

We can also see that the sales of the three year-round products (blonde, brown, and ipa), increase through the clusters, staring smaller in the opening months and increasing all the way up to winter. We also see that the total sales increase over time, after a small dip in June and July. This seems to be a good indication for business. Word of mouth advertisement seems to be helping the brewery and, hopefully, is an indication of increasing future sales.

Figure 7 shows the time series plot of total pint sales with colors corresponding to the clusters. It will be interesting to see if the total sales slow down in the months of June and July this upcoming summer. Eventually it will be possible to estimate these seasonality effects when the brewery has been open long enough and enough information is available. These cluster solutions give a sense of what products were offered at what times in the year. Now, using the reduced data set, we can attempt to address which months had higher sales of seasonals compared to year-round products.

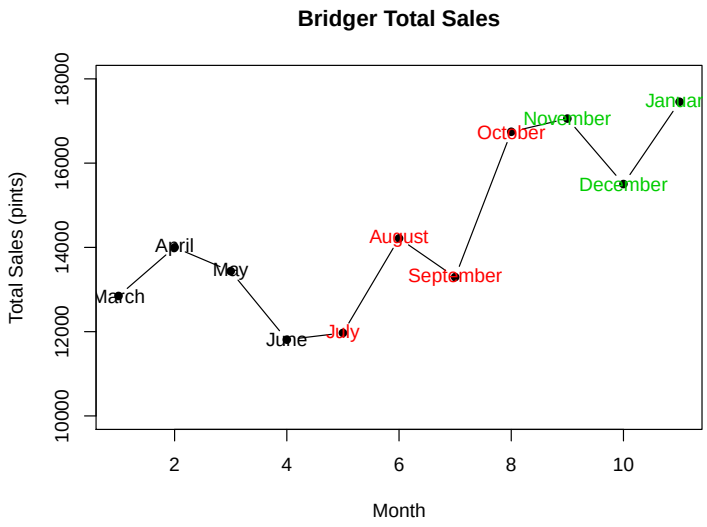


Figure 7: Time series plot of Bridger Brewery total pint sales with colors corresponding to 3-cluster solution

5.1.2 Reduced Data

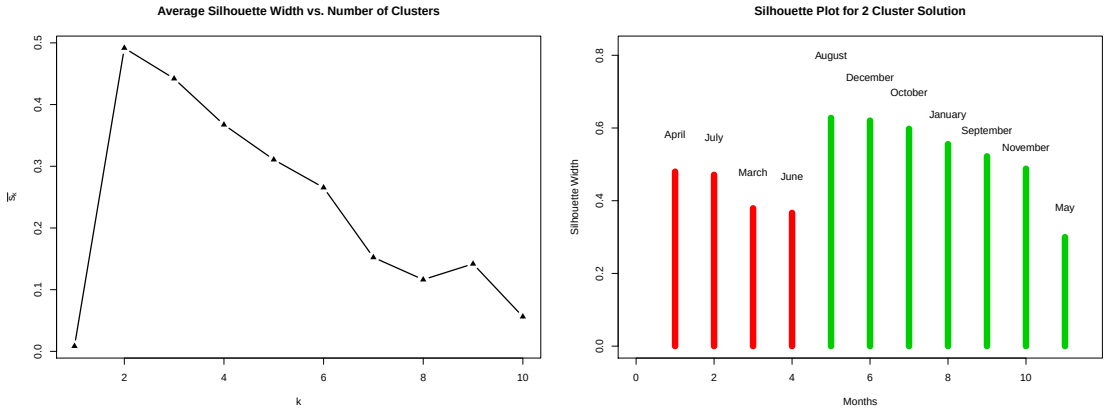


Figure 8: Average silhouette width versus cluster size (left) and silhouette plot for full data 2-cluster solution (right).

The first principal component in the reduced Bridger data was represented as a weighted contrast between IPA, Brown, and Blonde versus Seasonals. Figure 8 displays the average silhouette width by number of clusters (left) and the silhouette plot for the 2-cluster solution (right). The PAM cluster solution that yielded the largest average silhouette width was the two cluster solution. Cluster 1, which I called "High Seasonal Sales" encompassed March, April, June, and July. Cluster 2, which I refer to as "Low Seasonal Sales" encompasses May, and August through January. As the PCP shows, the proportion of monthly sales that seasonals account for are higher in High Seasonal Sales and lower in Low seasonal Sales. In the lower seasonal sales cluster, the products that are in stock year round (IPA, Bobcat Brown, and Bridger Blonde) all have a higher proportion of the months total sales. This information is displayed in a PCP with different panels for the separate clusters (Figure 9).

Many of the products that were available during the periods of higher seasonal proportion of monthly sales are no longer available. It will be interesting to see if the same pattern emerges in the upcoming

months, or if this was a chance phenomenon that we observed. Regardless of which months sell a higher proportion of seasonals, the total sales seem to be on the rise, and this may not be a concern to the brewery.

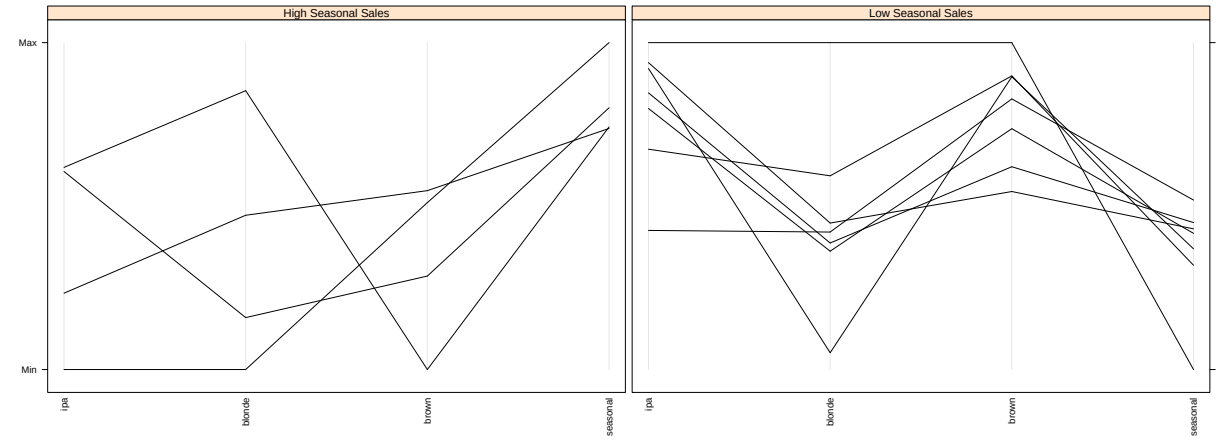


Figure 9: Cluster Solution for reduced data with PCP

5.2 Bozeman Brewery

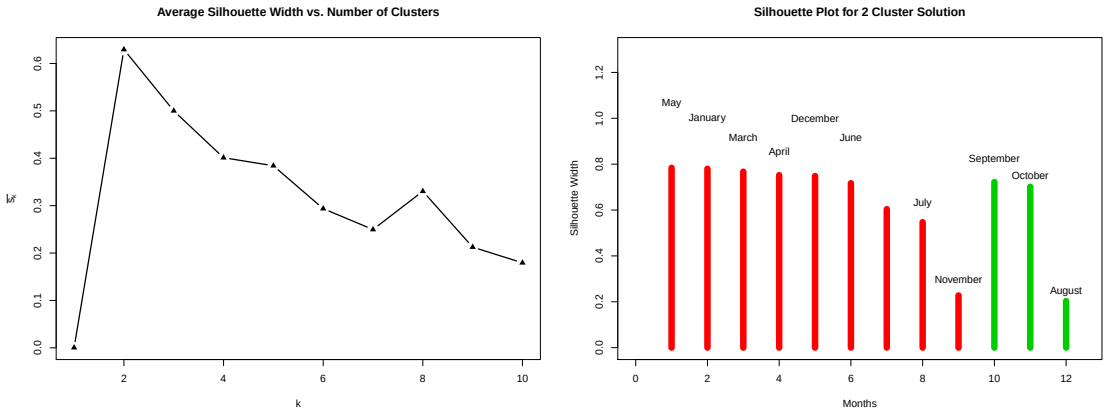


Figure 10: Average silhouette width versus cluster size (left) and silhouette plot for Bozeman Brewery 2-cluster solution (right).

For the Bozeman Brewery data, the total volume of sales was used instead of the proportion of the monthly sales. The reason for this was interpret ability, none of the cluster solutions using the proportion of sales data found noticeable differences between the clusters. The PAM cluster solution that yielded the largest average silhouette width was the two cluster solution. The 2-cluster solution is displayed using a PCP in Figure 10. Cluster 1, which I called "Low Sales" encompassed all the months except for June, July, and August. Cluster 2, which I refer to as "High Sales" encompasses June, July, and August. The PCP did not provide any useful information other than that the three months in the "High Sales" cluster have increased total sales. This is seen in the time series plot of the total pints sold (Figure 11). The clusters have limited information about differences in the months, other than total sales. Further data collection, particularly regarding the seasonals sold, might allow more information to be gained from cluster analysis.

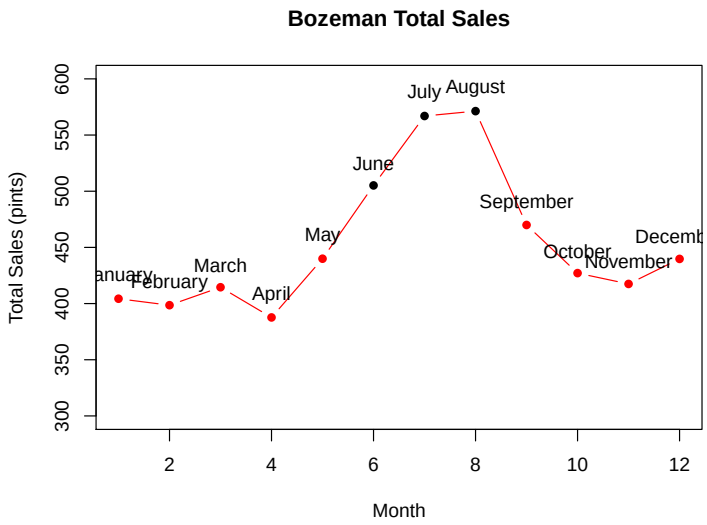


Figure 11: Bozeman total pints sales versus month with colors corresponding to 2-cluster solution

6 Improvements and Future Work

The previous analyses treat the months as the subjects and the sales of the different product varieties as the responses, generating a high dimensional, low sample size setting. It would also be interesting to think about the data set in reverse and treat it as a time series where the observations are the types of beers and months, or time, are the explanatory variable. This would allow for grouping different products according to sales over time as well as estimating the trends over time as more data become available.

Initially my plan was to perform a seasonal trend analysis to see which types of beer products were more popular during different times of the year. The problem I encountered was that I only observed each season (even each month) only one time for both of the breweries. When Bridger has been open long enough, I will set up time series models for different "styles" of beer. A model that could be incorporated to simultaneously estimate probabilities that a product is offered each month as well as the estimated sales given that it is offered is a hurdle model. This uses a zero-inflated binomial probability to estimate the probability that an observation is 0 and, conditional on the observation not being 0, another distribution is used to estimate the average response. It would also be beneficial for Bozeman Brewery to perform a cluster analysis of sales throughout counties in Montana. This could examine differences in volumes of sales as well as taste preference for their customers throughout the state.

There are also model-based clustering methods based on finite mixture distributions of multivariate normals (developed by Banfield). When more observations are collected it could be beneficial to: 1) understand the mathematics behind estimating clusters based on actual probability models and 2) see how different clustering methods and algorithms change the interpretation and estimation of clusters.

Finally, I would be very excited to use factor analysis, a method of examining latent factors from the manifest explanatory variables, to construct a mechanism for measuring taste preferences in beer drinkers. Similar mechanisms have been constructed to measure food consumption factors (Ryman et al 2013), ecological characteristics (Batalha and Carvalho 2013), and in many other applications. If an instrument exists for this, collecting information from patrons could help Bridger Brewery tailor product varieties to customers seasonal taste preferences.

Although many of the methods used in this paper are exploratory in nature, they can be effectively used to gain knowledge about the association between products and different periods of the year without using traditional time series methods. These data are limited, the brewery sales is a new process that will be undergoing many changes over this initial start up period. Even so, these methods can still be used to gain an understanding about the increase in the sales over recent months, identify sales profiles that were

evident in the first year, and see how the proportion of monthly sales in each product change according to the season. As more data become available, we can use these techniques as a starting point in understanding the variables we are working with and their associated patterns across time.

7 References

1. Batalha, M.A., and Carvalho, G.H.. "The drivers of woody species richness and density in a Neotropical savannah." *Biology Letters*: 2013. Print.

2. Cox, M.A.A, and Cox, T.F. "Multidimensional Scaling." *Handbook of Data Visualization.* : Springer, 2008. 316-347. Print.

3. Everitt, B., and Hothorn, T. *An Introduction to Applied Multivariate Analysis with R.* Springer, 2011. eBook.

4. Gabriel, K.R. ."The biplot graphic display of matrices with application to principal component analysis." *Biometrika*, 1971 58:3, 453-467. Print.

5. Hardle, W. and Simar, L. *Applied Multivariate Statistical Analysis.* 3rd ed. Springer, 2012. eBook.

6. Hennig, C. *fpc: Flexible procedures for clustering. R package* version 2.1-7. http://CRAN.R-project.org/package=fpc 2014.

7. Kaufman, L. and Rousseeuw, P. *Finding Groups In Data: An Introduction to Cluster Analysis.* John Wiley and Sons, 1990. 68-108 Print.

8. Meyer, C. D..*Matrix Analysis and Applied Linear Algebra.* Philadelphia: Society for Industrial and Applied Mathematics, 2000. 279-429 Print.

9. Paradis, E., Claude, J., and Strimmer, K. "APE: analyses of phylogenetics and evolution in R language". *Bioinformatics* 2004 20: 289-290.

10. Primack, B.A., Kim, K.H., Shensa, A., Sidani, J.E., Barnett, T.E., and Switzer, A.E.: "Tobacco, Marijuana, and Alcohol Use in University Students: A Cluster Analysis", *Journal of American College Health*, 2012 60:5, 374-386.

11. Raiche, G., Riopel, M, and Blais, J.G. *Non Graphical Solutions for the Cattell's Scree Test*, Universite du Quebec a Montreal, International Meeting of the Psychometric Society, IMPS 2006, HEC, Montreal June 16th.

12. Ryman, T.R., Austin, M.A., Hopkins, S., Philip, J., O'Brien, D., Thummel, K., Boyer, B.B. "Using exploratory factor analysis of FFQ data to identify dietary patterns among Yup'ik people." *Public Health Nutrition* 2013 17: 510-518. Print.

13. Venables, W. N., and Ripley, B.D.. *Modern applied statistics with S.* 4th ed. New York: Springer, 2002. Print.

14. Wall, M., Rechtsteiner, A, and Rocha, L. "Singular Value Decomposition and Principal Component Analysis." Trans. Array *A Practical Approach to Microarray Data Analysis.* Kluwer Academic Publishers, 2003. 91-109. Web. 8 Apr. 2014.