An Analysis of Beer Styles

Ivan P. Ramler*

Abstract
With the recent surge in the popularity of craft beers and home brewing, the styles of beer that are available to consumers have increased tremendously. However, many beer drinkers are unaware that there are many numerical measurements that are used to describe beer. This paper introduces the numerical measurements used to describe the main styles of beer defined by the Beer Judge Certification Program such as specific gravity, alcohol content, bitterness, and color. An interesting aspect of these characteristics is that they are presented as interval data, for which many common statistical methods don't readily apply. As a result, a novel method for clustering interval data is used to group and describe the similarities in the beer styles.

Key Words: Cluster Analysis, Correlation, Dendrogram, Interval Data

1. Introduction

In the past twenty odd years, there has been rapid growth in the number of craft breweries and craft beer drinkers in the United States. Indeed, in 1980 there were only 8 craft brewers, that number increased to 537 in 1994, and over 1900 in 2011 (Brewers Association 1, n.d., Brewers Association 2, n.d.). This influx of craft breweries has broadened the spectrum of styles of beer that people are familiar with from the common Light American Lagers (such as Bud Light, Coors Light, and Miller Lite) to many different style such as Stouts, Porters, and India Pale Ales that show up in bars and restaurants throughout the United States.

In addition to the growth of the craft beer industry, the homebrewing hobby has grown to over 1 million Americans (Brewers Association 3, n.d.). Perhaps more so than craft brewers, homebrewers explore the many styles of beer. Data from the National Homebrew Conference shows that many different styles are represented including some that are rarely seen from commercial breweries (Jeff, 2012).

With the increasing awareness to the many varieties of beer, many people may not be aware of the many numerical characteristics that can be used to describe a beer. The Beer Judge Certification Program (BJCP) has a database of "key statistics" used to describe typical beers from their recognized styles of beer (http://www.bjcp.org/stylecenter.php). These statistics are used to provide guidance to judges and entrants in homebrew competitions and are generally accepted as a good way to describe styles of beer on a similar scale.

This paper will introduce these key statistics defined by the BJCP, use the guidelines to simulate individual beers in order to describe and rank the perceived bitterness from hops for the styles, and use cluster analysis to group similar styles together. The paper finishes with a discussion on what additional information, beyond the scope of this analysis, that could be used to improve the bitterness rankings and groupings.

2. Beer Statistics and Simulating Individual Beers

The BJCP defines 23 main styles of beer with several substyles in each for a total of 80 styles of beer (http://www.bjcp.org/stylecenter.php). For 72 of these styles (in 20 of the

*St. Lawrence University, 23 Romoda Street, Canton, NY 13617

main styles) they provide guidelines for brewers and judges based on five different numerical characteristics of beer. These characteristics are:

- OG: The **original gravity** of a beer is the relative density (or specific gravity, SG) compared to water before fermentation is started.
- FG: The **final gravity** is the specific gravity of a beer after fermentation is complete.
- ABV: The percent **alcohol by volume** of a beer fermentation is complete.
- IBU: The **International Bittering Unit** is a measure of the bitterness of beer from the amount of hops used during the brewing process.
- SRM: The **Standard Reference Method** is used to measure color intensity and darkness of a beer. Briefly, SRM is defined as 10 times the light absorbency of a beer sample illuminated by a light source with a wavelength of 430 nanometers measured through a half inch glass cuvette measured on a logarithmic scale (Daniels, 1998). Low values such as 2 or 3 appear straw-like in color under normal light, values around 10 - 15 are amber to copper in color, and values of 30 or more tend towards dark brown to black.

There are several other beer characteristics that can be calculated from the BJCP key statistics that are useful for describing beer. They include:

- OE: The **original extract** of a beer is the mass (grams) of sugar in 100 grams of unfermented beer.
- FE: The **final extract** of a beer is the mass (grams) of sugar in 100 grams of fermented beer.
  - Both OE and FE can be calculated from the original and final gravities using De Clerck's equation (De Clerck, 1958):

    $$Extract = -205.347Gravity^2 + 668.72Gravity - 463.37$$

- RA: **Real attenuation** is the percent of sugars converted to alcohol and CO_2 from the yeast in a beer. To some extent, low numbers indicate that the beer will be maltier while high values indicate the beer will be drier. Real attenuation is calculated from OE and FE using Balling's approximation (De Clerck, 1958) of:

  $$RA \approx 1 - (0.1808OE + 0.8192FE)/OE$$

The original five characteristics (from BJCP) will be used to construct groups of similar styles (Section 3) while the latter will be used in conjunction with IBU to calculate and rank the hop bitterness of beer styles (Section 4). One complicating factor is that, for each style of beer, the BJCP gives a range of values that they would consider to be consistent or typical with the style. This doesn't imply that all beers for a given style fall inside the ranges provided by BJCP; instead they've constructed the ranges based on numerous commercial beers from each style. As a result these ranges, which seemingly appear to be interval style data, are essentially marginal distributions of each style for each of the five characteristics. However, because the BJCP doesn't release the information on the actual beers they used to create their guidelines, individual beer information is not easily available. As a replacement to this information, simulation can be used to generate individual beers from each style according to the guidelines.

Unfortunately, it is not as straightforward as just generating observations directly from the marginal distributions because:

Table 1: Correlation Matrix of BJCP Key Statistics based on mid-ranges.

	OG	FG	ABV	IBU	SRM
OG	1.00	0.77	0.97	0.53	0.24
FG		1.00	0.65	0.49	0.46
ABV			1.00	0.49	0.18
IBU				1.00	0.34
SRM					1.00

(a) since only ranges are given, the distributional form is unknown and,

(b) many of the five characteristics are highly correlated with each other.

To work around the first issue, we will assume the distribution for each style, $i = 1 \ldots 72$, follows a five-dimensional multivariate normal distribution with mean vector, μ_i, equal to the mid-range of each BJCP characteristic. Standard deviations for each style and characteristic are calculated using the $range/6$ rule of thumb. To incorporate the correlation between characteristics, we first make the assumption that the correlations between characteristics is the same for all styles. This is then found by obtaining the correlations between the means of each style and are displayed in Table 1.

For the analyses in Sections 3 and 4, 10,000 datasets were simulated with each of the 72 styles being represented exactly once in each simulated dataset.

3. Grouping Similar Beer Styles

Although the BJCP provides groupings of the beer styles using the main-styles, by their admission "The groupings in the Style Guidelines are somewhat arbitrary, and often did not represent a unanimous decision of those on the Style Committee who worked on the document" (BJCP, n.d.). As a result, there may be ways to improve the groupings. One such potential way is to use cluster analysis to construct groups of similar styles based on the style guideline's numerical characteristics. However, given that the guidelines are presented as intervals, standard clustering algorithms such as hierarchical clustering cannot be directly used to incorporate the variability in a style. Of course, one could easily use the mid-range to construct a crude clustering of the styles, but this ignores the additional information provided by the intervals. Instead, a clustering technique is developed that uses simulated datasets (as described in Section 2) to group the styles while still incorporating the information available from the intervals. This method uses hierarchical clustering at two-stages and is described as follows.

3.1 Two-Stage Clustering Algorithm

1. For each simulated dataset, use hierarchical clustering (with complete linkage) and obtain two potentially different classifications by cutting the resulting tree into both a low number and a high number of clusters (the method for obtaining the number of clusters is described in Section 3.2). These classifications are used to create the distance matrix for the second clustering stage.

2. Given the pair of classifications for each simulated dataset from Step 1, construct a distance matrix for the styles where each entry (i.e., distance between a pair of styles)

is the $1 - p_{ij}$, where p_{ij} is the proportion of clusterings where the style i and j were classified in the same cluster. For example, out of a total of 20,000 clusterings (i.e., 10,000 datasets, each clustered twice), American India Pale Ale and English India Pale Ale were in the same cluster about 94.2% of the time (not surprising as these two styles are rather similar) and would have a "distance" of $1 - 0.942 = 0.058$. Similarly, Imperial Stout and a Light American Lager (to very different styles of beer) have a distance of $1 - 0.0004 = 0.9996$ (i.e., they were is the same cluster only 0.04% of the time). Figure 1 displays the pairwise distance matrix for all 72 styles.

3. Using the distance matrix from Step 2, use hierarchical clustering (with average linkage) to construct a hierarchical tree for the beer styles using standard approaches to analyze the results (e.g., dendrograms, estimating optimal number of clusters, etc.).

The rationale to obtaining two sets of classifications for each set of simulated beers is that by having a low number of clusters, the broad sense connections between styles can be more easily identified by allowing many styles of beer to "intermingle" in the same cluster. However, by doing this, any sub-clusters that break beer styles into more focused categories is lost. This is where obtaining a second set of classifications with a large number of clusters is useful. The danger of using only a high number of clusters is that when creating the distance matrix for the second stage, having too many clusters will mean that most beers will never be in a cluster together and the resulting hierarchical tree would give the impression that there is absolutely no connection between many styles of beer. Blending these results together should allow the more interesting sub-cluster structure to remain while still showing a connection between all the beer styles at some level.

The colored dendrogram displayed in Figure 2, constructed using R code from Francois (2005), shows the results of the two-stage clustering approach. Using the maximum average silhouette width (Rousseeuw, 1987), the algorithm resulted in 19 clusters, of which there were five singletons (Imperial IPA, Eisbock, Strong Scotch Ale, Imperial Stout, and Old Ale; black color in Figure 2), and 14 clusters with two or more styles (the colors in Figure 2 correspond to clusters). Many of the resulting clusters make intuitive sense; for example, one cluster contains many English styles such as Mild, Ordinary and Special Bitters along with Scottish Light 60/ and Heavy 70/. While these styles are not categorized in the same main groupings by the BJCP, they all generally can be characterized by having low gravity, fairly low IBUs, and are amber to red in color. Similarly, another cluster contains such lighter beers as Standard and American Lagers, Cream Ale, Witbier and Weissbier, Kolsch, Blonde Ale, and American Wheats. This cluster essentially combines the Light Lager and Light Hybrid Beer categories of the BJCP. Additionally, many of the styles from the Porter, Stout, and Dark Lager BJCP categories are placed together due to their similar color (dark brown to black) and specific gravity (low OG with average FG) structure.

Overall, given that many of the resulting clusters categorize similar beer styles, these results could be useful to individuals interested in trying new beer styles based on their recognition of styles within the same cluster. Further, those looking to try something different can use the results to pick new beers from clusters they are familiar with.

3.2 Determining the Number of Clusters

The Step 1 of the algorithm from Section 3.1 uses the silhouette width in two ways to calculate the number of clusters (K). The first just uses the standard approach of selecting K to be where the maximum average silhouette width is used. The second finds local maximums in the average silhouette width (across all potential values of K), choosing the largest K that occurs at a local maximum. In particular, it uses the following procedure.

1. For $k = 1, 2, \ldots, K_{max}$, calculate the average silhouette width (ASW) by cutting the hierarchical tree at k clusters. (Note that in this application, $K_{max} = 50$.)

2. Use a smoother to help remove some of the excess variation in the ASW across the values of k. One method to do so (implemented for the algorithm in Section 3.1) is to use the locpoly function from the KernSmooth package in R (Wand, 1995; Wand, 2011). This is done to help identify only the clear peaks in the ASW. Figure 3 displays an example of the ASW's from one of the simulated datasets along with the smoothed portion.

3. Find the local maximums from the smoothed ASW (using first and second derivatives), choosing K to be where the last local maximum occurs.

4. If no local maximum can be determined using the derivative method, pick K to be the maximum ASW. (This would be the same as using the standard approach for using ASW to determine K.)

Although for the purpose of clustering beer styles this method gives reasonable results, it has not been extensively tested or evaluated. Further research should be done to investigate whether this method can be used to better identify the optimal number of clusters and whether it could be useful to find any sub-cluster structure within a dataset.

4. Perceived Hop Bitterness

With the increasing popularity of highly hopped beers, many beer manufacturers are beginning to label their products with the amount of IBUs in them to give consumers a sense of the "hoppiness" of their beer. However, while IBUs are a major component in measuring how hoppy a beer tastes, they are not the only factor in the perceived bitterness of a beer. In particular, the bittering effect of hops is less noticeable in beers with higher specific gravities. Instead measures that take into account the gravity (or maltiness) of the beer could be used to give a more accurate way of describing a beers hoppiness. For example, a common and simple guideline used by brewers to determine the balance of hops and malt in their beer is the OG to IBU ratio. Another idea came from Eric Myers of the blog Top Fermented to multiply the IBU:OG ratio by the apparent attenuation, very similar to real attenuation but is defined as $(OE - FE)/OE$, to help account for the remaining non-fermented sugars in the beer (Myers, 2009). A simpler, potentially more intuitive hop bitterness measurement could be just to create a scaled IBU by multiplying the IBU by the real attenuation. This measurement adjusts the amount of hop bitterness from IBUs by the percent of remaining sugars in the beer (i.e., $sIBU = IBU \times RA$).

Figures 4 and 5 show the hoppiness rating for the 72 BJCP beer styles based on the scaled IBU method (Figure 4) and Myers' method (Figure 5). For each style, the 10,000 simulated datasets were used to approximate the distributions of the scaled IBU and Myers' statistic. The intervals displayed in Figures 4 and 5 represent the mean $\pm$ one standard deviation for each style. The color on the intervals show the mid-range SRM of the style and they are sorted in increasing order by their mean values. For ease of comparison across figures, these measures have been standardized to be on a 0 - 100 scale.

Both of these methods have their merits and generally agree with each others rankings. The scaled IBU method shows that there are just a few styles of beer that are much more hoppy than others with the majority of styles overlapping in hoppyness with many other styles. The Myers' hoppiness score shows a much more steady growth in hoppiness with only one style (Imperial IPA) being significantly more hoppy than the others. However, it isn't clear which method is ultimately better. Ideally, a panel of expert beer judges could be used to evaluate these methods and offer input as to which (if either) is better. Unfortunately, this idea may not be overly feasible. Instead, either method could be used as a relatively simple way for beer enthusiasts to get a relative sense of the hoppiness of a beer.

5. Discussion

While the analyses done here provide an interesting way to view the connections between the different beer styles, given the limitations of working with only the intervals provided by the BJCP, only a general description of styles can be done. Preferably, information from individual beers would be available. Unfortunately, it is unlikely that many commercial brewers would be willing to make this information public. One possible solution would be to acquire information from homebrewers. Given the plethora of freely shared beer recipes available online, it would presumably be much easier to acquire a sizable dataset from the homebrew community. In addition to obtaining information on the values that the BJCP use to create their guidelines, knowing the complete breakdown of the ingredients could lead to improved ways to cluster styles or rate bitterness. For example, the bitterness of a beer comes from more than just the amount of hops added. Some malts add a different type of bitterness to a beer; an example of this is roasted barley which is used to give a coffee-like bitterness to beer. Additionally, the use of dry-hopping (i.e., adding hops to partially fermented beer), adds a very marginal amount of bitterness to a beer, but can impart a large amount of hop flavor and aroma which can be associated with "hoppiness". These are just a few of the complex issues that are problematic in any type of bitterness rankings.

As stated in Section 3.1, the cluster algorithm developed to incorporate correlated interval data has not been extensively tested. Further research is needed to be able to better understand the impact of the two-stages and, in particular, the novel approach to estimating the number of clusters using the local maximums of the average silhouette width. However, it shows promise and is general enough to easily be applied to other research problems where interval data is present.

REFERENCES

BJCP (n.d.), Retrieved Jul 18, 2012 from Frequently Asked Questions About the Beer Judge Certification Program, http://www.bjcp.org/bjcpfaq.php#stylefaq.

Brewers Association 1 (n.d.), Retrieved July 16, 2012 from http://www.brewersassociation.org/pages/about-us/history-of-craft-brewing.

Brewers Association 2 (n.d.), Retrieved July 16, 2012 from http://www.brewersassociation.org/pages/business-tools/craft-brewing-statistics/number-of-breweries.

Brewers Association 3 (n.d.), Retrieved July 16, 2012 from http://www.brewersassociation.org/pages/media/press-releases/show?title=worldwide-homebrewing-celebration-continues-growth.

Daniels, R. (1998), *Designing Great Beers; The Ultimate Guide to Brewing Classic Beer Styles*, Brewers Publications, Boulder, CO.

De Clerck, J. (1958), *A textbook of Brewing* Translated by Kathleen Barton-Wright, Chapman & Hall, London, Vol II.

Jeff (July 2, 2012), Retrieved July 16, 2012 from Lug Wrench Brewing Company, http://www.lugwrenchbrewing.com/2012/07/most-popular-beer-styles-2012-nhc-entry.html.

Francois, R. (2005), A2R: Romain Francois misc functions, http://addictedtor.free.fr.

Myers, E. (2009), Developing simplified beer statistics. Part 1 Hoppiness. Retrieved July 20, 2012 from http://www.topfermented.com/2009/08/26/developing-simplified-beer-statistics-part-1-hoppiness.

Rousseeuw, P. J. (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 5365.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall, London.

Wand, M. (2011). KernSmooth: Functions for kernel smoothing for Wand & Jones (1995). R package version 2.23-7. http://CRAN.R-project.org/package=KernSmooth

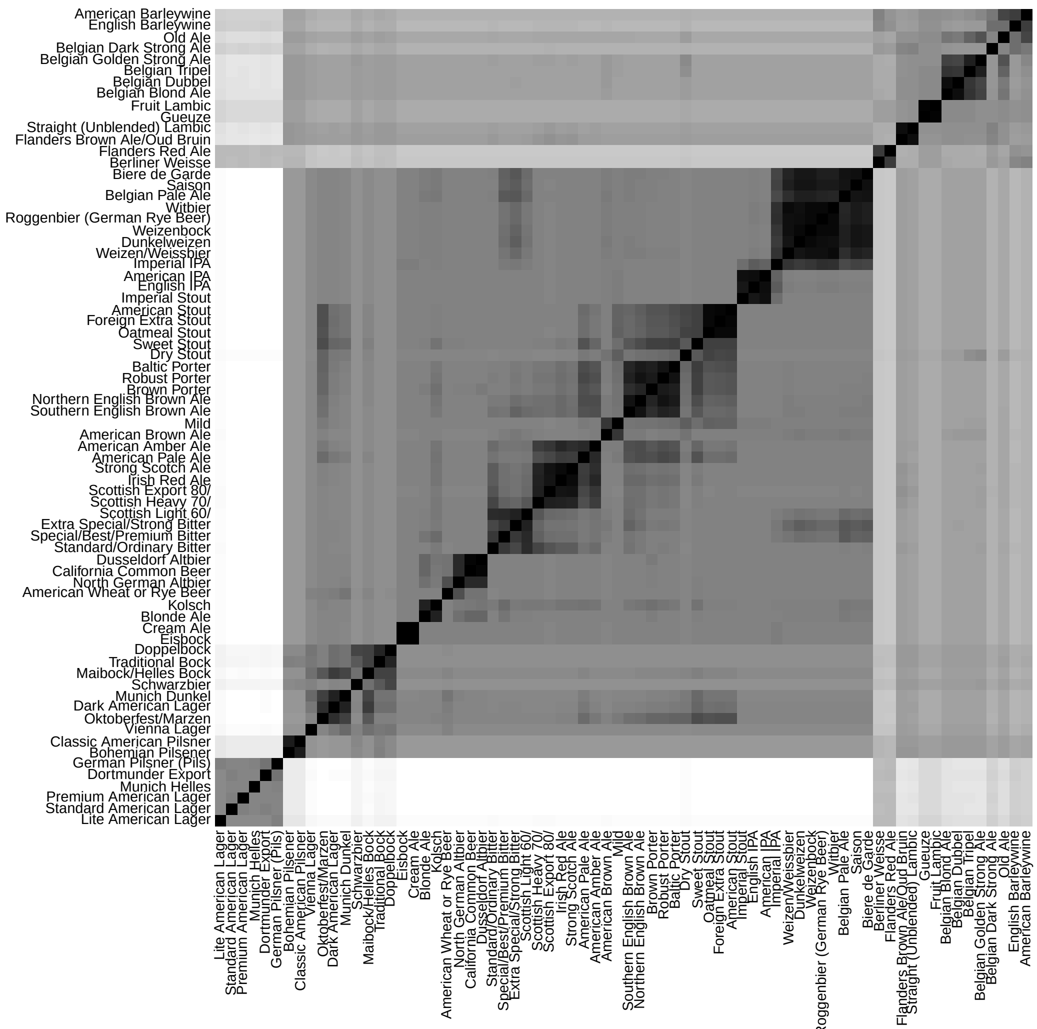

Figure 1: Distance matrix for second stage clustering of the 72 beer styles. The darker the color, the more often the pair of styles appeared in the same clustering.

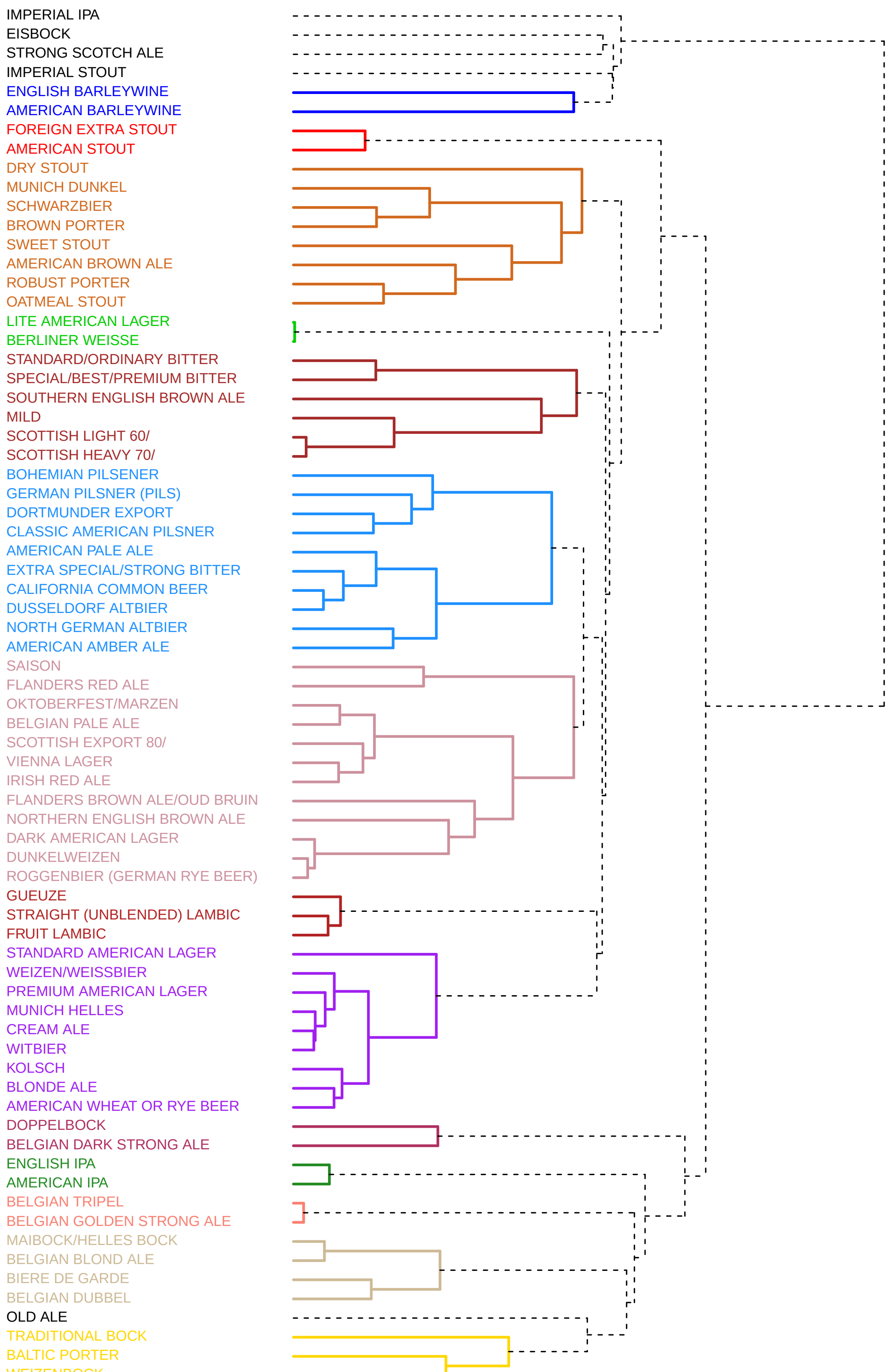

Figure 2: Beer style clustering results using method outlined Section 3.1. Non-black colors denote groups, styles in black represent singletons that were not assigned to groups.

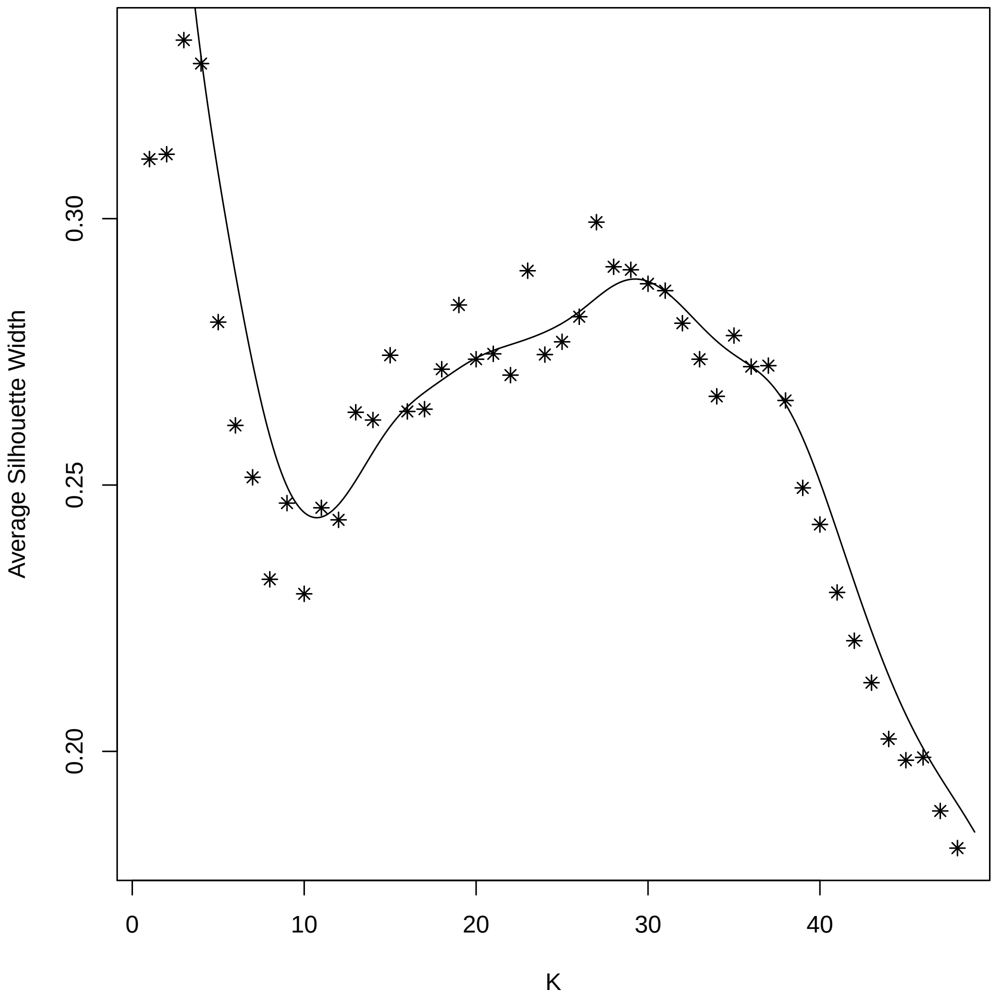

Figure 3: Average silhouette widths of clusters from simulated beer dataset along with local polynomial smoother. Maximum average silhouette width occurs at $K = 3$, while the last local maximum occurs at $K = 30$.

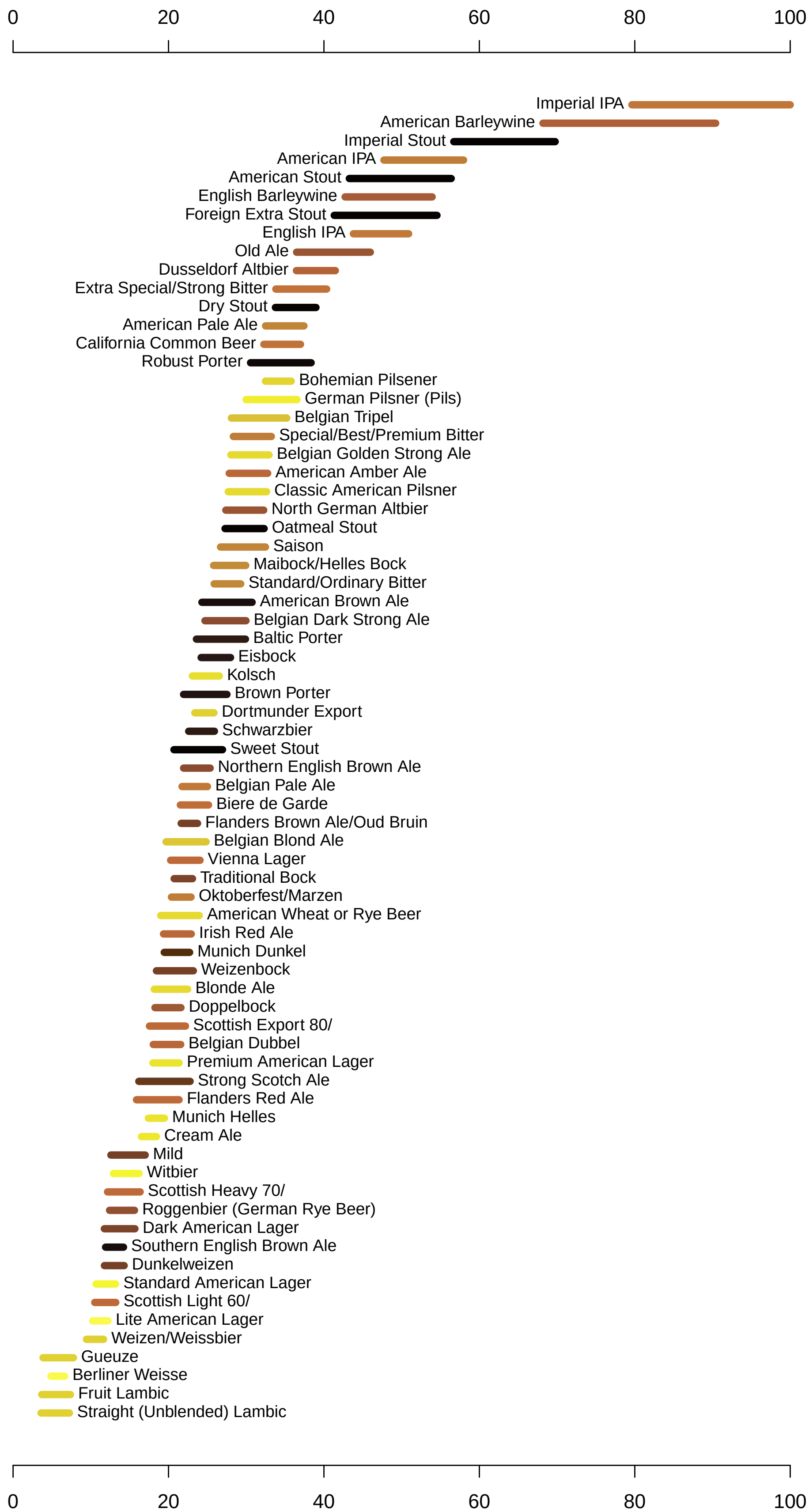

Figure 4: Scaled IBU (mean ± one standard deviation) for BJCP beer styles.

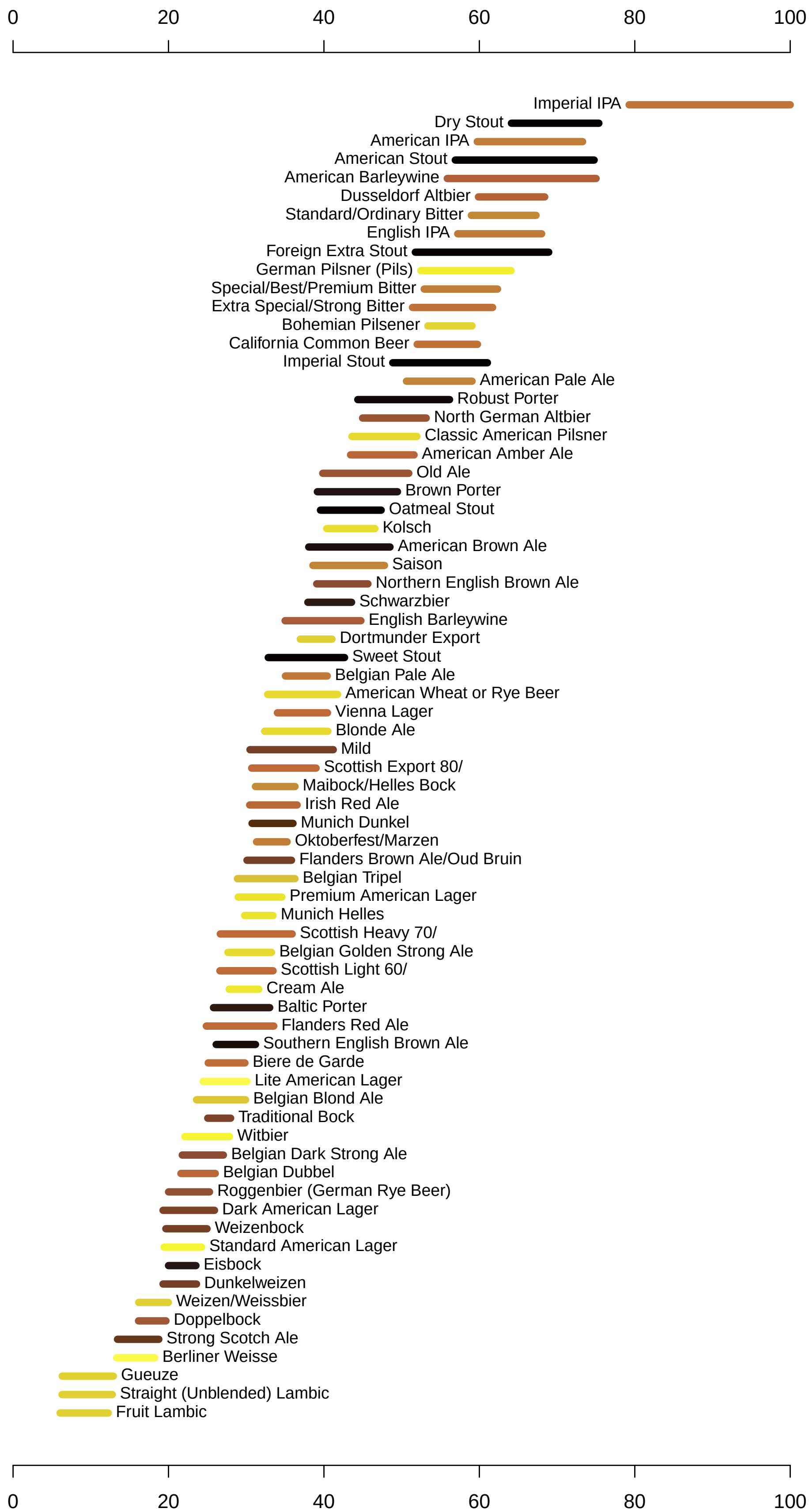

Figure 5: Myers' hoppiness score (mean $\pm$ one standard deviation) for BJCP beer styles.