**MT5758 - Group Project (Time related results)**

We used the 'as.POSIXlt' function in R to convert the time stamp associated with each review from UNIX based to the conventional format. We then created 5 additional variables; 'hour', 'day', 'month', 'season' and 'year' in order to explore if there was any time related variability exhibited by the data.

Exploratory Data Analysis ("EDA") indicated that there was little difference between the number of ratings per month. The lowest proportion of ratings occurred in June (7.6%) and the highest in December (9.6%). Therefore, whilst month alone does not give you much information about the number of ratings, there is a clear pattern of frequent ratings during (US) cold months, as shown by the below table.

Fig.x

Season	Months	Proportion of ratings
Winter	Nov-Jan	27.5%
Autumn	Aug-Oct	25.2%
Summer	May-July	23.7%
Spring	Feb-Apr	25.0%

Our EDA also confirmed that as year increases so too does the number of ratings (Fig x1). The observations were collected between the founding of BeerAdvocate.com in January 1998 up until January 2012. It is therefore not surprising that the number of ratings increases with year as it likely reflects the growth of both the popularity of the website and the growth of internet usage itself.
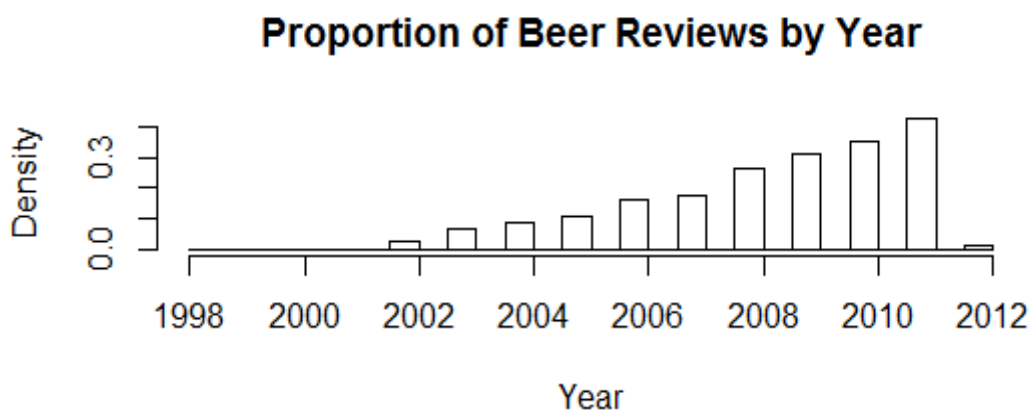
We had intended to investigate the 'hour' variable, however, as there is no time code specified for each observation we could not be confident about our interpretations, given the somewhat international nature of our data.

Based on the results of our aforementioned EDA we undertook two further PCA investigations. The variables included were the same as those of our previous PCAs but with the addition of the scaled 'year' and 'month' respectively. The results can be seen in Figures x2 and x3 below;

Variables	Variable Means on the True Scale	Principal Component 1 Loadings	Principal Component 2 Loadings	Principal Component 3 Loadings	Principal Component 4 Loadings	Principal Component 5 Loadings	Principal Component 6 Loadings	Principal Component 7 Loadings
Standardized ABV	7.043	0.215	-0.547	-0.752	0.193	-0.140	0.154	-0.089
Standardized Appearance Rating	3.85	0.385	-0.004	-0.055	-0.892	-0.127	0.192	0.027
Standardized Aroma Rating	3.745	0.436	-0.008	-0.054	0.006	0.822	-0.289	-0.218
Standardized Palate Rating	3.752	0.448	0.065	0.062	0.104	-0.506	-0.721	-0.070
Standardized Taste Rating	3.802	0.470	0.097	0.076	0.265	0.049	0.234	0.798
Standardized Overall Rating	3.822	0.438	0.202	0.259	0.294	-0.174	0.531	-0.549
Standardized Year	2008	0.050	-0.804	0.593	-0.012	0.009	0.000	0.015
Mardia's Criterion		*0.329*	*0.562*	*0.526*	*0.624*	*0.575*	*0.505*	*0.559*
Equilibrium Contribution		*0.378*	*0.378*	*0.378*	*0.378*	*0.378*	*0.378*	*0.378*
Importance of Components								
Standard deviation		1.918	1.041	0.910	0.740	0.626	0.532	0.431
Eigen Values (Standard Dev. Sq.)		3.680	1.083	0.828	0.547	0.392	0.283	0.186
Proportion of Variance		52.6%	15.5%	11.8%	7.8%	5.6%	4.0%	2.7%
Cumulative Proportion		*52.6%*	*68.0%*	*79.9%*	*87.7%*	*93.3%*	*97.3%*	*100.0%*
Kaiser's Criterion		*1*	*1*	*1*	*1*	*1*	*1*	*1*

Just as was the case for the PCA carried out earlier, the first principal component relates to determinants of good beer rating. For example, a high score on PC1 equates to an observation having above average ratings across all of the review criteria. PC2 relates to a segment of the data that can be described ABV and more importantly year. The large negative loadings indicate that prior to the mean year (2008) people were rating (and thus consuming) beers with higher than average ABV. Therefore a very high score in PC2 would relate to the rating of a strong beer, such as an 'American Double', near the start of our data collection period (1998). PC3 effectively describes the inverse of PC2. An observations' score on this component is pulled in opposite directions by the ABV and year, meaning that a high score is indicative of a rating from nearer the end of the data collection period (2012) of a beer with a relatively low ABV, such as a 'Low Alcohol Beer'. It could be the case that PC1 and PC2 capture a change in consumer preference regarding ABV during the period. Perhaps, people are more aware of the health implications of consuming large amounts of alcohol or it could simply be a change in consumer tastes. As was previously the case, there are two components (PC6 and PC7) that could be capturing the *false reviews* inherent in the data set. This is exemplified by the inverse relationships between the overall rating and palate (PC6) and the overall rating and taste (PC7). Although, both components fall well short of being regarded as important under the Kaiser Criterion, they nonetheless appear to confirm phenomena

It is worth noting that whilst an above average alcohol content (ABV) appears to contribute to a high PC1 score, and thus a positive rating, the ABV loading (0.215) is too low to meet either of our benchmark criteria, and thus is deemed to contribute too little to the component to be included in our analysis.