Descriptive analysis of beer styles:

Analysis of the styles based on the location:

Since the BeerAdvocate dataset contains ratings from users and for beers all around the world, it is interesting to investigate where each beer style is mainly produced and where it is mostly rated.

The histogram in figure ?? shows the distribution of production countries for the Belgian-Style ales. In fact, they are mostly produced in the United States, Belgium and Canada. The histogram in figure ?? shows the distribution of rating countries for the Belgian-style ales. We notice they are also mostly rated in the same three countries where they are mostly produced. This makes sense because users from a given country are more likely to rate beers from their own country. We can then suppose that they will also most likely prefer beer styles that are produced in their own country.

Categorization of users based on their level of experience and country of origin:

Having a very large number of users in our dataset, it is likely that some users might resemble each other. The only information we have about a certain user are his ratings and his birth-country. The birth country is likely to influence a user’s ratings, since a user will probably rate beers from his own country, or in the surrounding countries. The number of ratings of a user indicates a certain level of experience. In fact, users having rated more than 500 different beers might have a different palate than a user having rated less than 10 beers for example. Thus, to determine whether it is relevant to categorize users based on these 2 features, we conducted a few analyses.

First, for each user a “favorite style” was determined based on a user’s top-rated beers. We then use a regression analysis to predict the probability or likelyhood? that a certain beer style will be a user favorite based on his country of origin and number of ratings. Figure?? shows the regression analyses for four different styles in four different countries. We can see for example that a user from the United States or Canada is more likely to prefer IPA’s than a user from Germany or Belgium. However, a user from Belgium is much more likely than any of the three other countries to prefer Belgian-style ales. We also notice that the probability of a user to prefer stouts increases with the number of ratings. Thus, stouts are more appreciated by experienced users than novice users. From these observations we can conclude that the preferences of a user are likely to depend on his country of origin.

In fact, users tend to rate mostly beers from their own country. Thus, they are more likely prefer beer styles that are brewed in their country.

To determine whether the preferences of a user evolve as his experience increases, a comparative analysis was conducted. In fact, we compared the distributions of “favorite styles” for users having done more than 500 ratings between their favorite beer style among their 50 first ratings and their favorite style among their 50 last ratings. Figure ?? shows these two distributions. We observe that the general “trend” is similar, but we can still observe some differences. For example, experienced users seem to like stouts and wild/sour beers a lot more than non-experienced users, while non-experienced users tend to like IPA’s and Belgian-style ales more. These observations are consistent with the regression analysis. To determine how significant these differences are, we conduct a paired statistical test called the Wilcoxon sign-rank test. It tests the null hypothesis that the two related paired samples come from the same distribution. The p-value is between 0.05 and 0.1, thus we don't reject the null hypothesis at a confidence level of 5% but we can reject it at a confidence level of 10%.

These analyses show that the preferences of a user most likely depend on his birth country and his level of experience.

Thus, in the shelve arrangement of the supermarket, we should separate beer preferences between experienced and non-experienced users, and depending on the country, the preferred beer styles will vary.