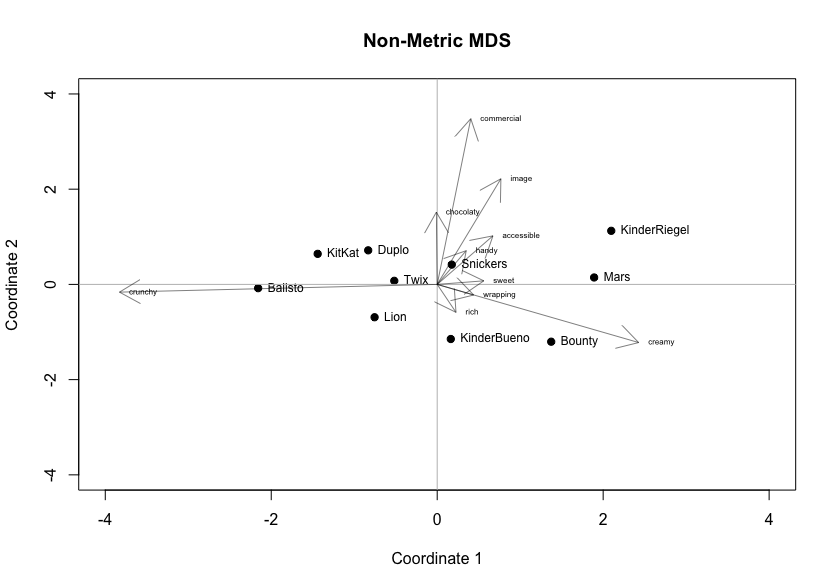
Special Work Performance 2

Giorgi Modebadze – Radoslav Evtimov – Ecenaz Bal

The purpose of this report is to analyze survey data of 50 persons about 10 different chocolate brands in the German market. Each brand is assessed with 13 different attributes evaluated on a scale 1 to 5(1-lowest,5-highest) We used Kruskal's Non-metric Multidimensional Scaling and Principal Component Analysis to make the data more interpretable. For the purposes of data analysis, the missing values were replaced with the mean value for the given product.

SWP2a:

For Multidimensional Scaling we chose non-metric scaling method, which attempts to represent, as closely as possible, the pairwise dissimilarity between objects in a low-dimensional space. The reason for using non-metric scaling is that the survey data is non-metric, they do not possess a meter with which distance between scale values can be measured, meaning for that data only ranking matters and not actual differences between points. On the way to finding NMDS first the Euclidean distance is calculated using *daisy* formula from cluster library. Based on the result of Euclidean distance a two-dimensional perceptual map is created using *metaMDS* from vegan library. The formula is more robust than *isoMDS,* it also performs Nonmetric Multidimensional Scaling (NMDS), but unlike isoMDS it provides infrastructure to do the random starts and comparison of configurations for convergence. The comparison between these two methods on two-dimensional space gave drastic difference in terms of stress values. For *isoMDS* the stress value is 3.99, but as the rule of thumb it is never a good idea to use solutions with stress value more than 0.2 and a stress value approaching 0.3 indicates that the ordination is arbitrary. With every additional dimension, isoMDS also decreases stress value, but addition dimensionality makes interpretations more challenging. On the other hand, *metaMDS* using centring scale and PC rotation, after twenty random start found two convergent solutions with stress value of 0.04.



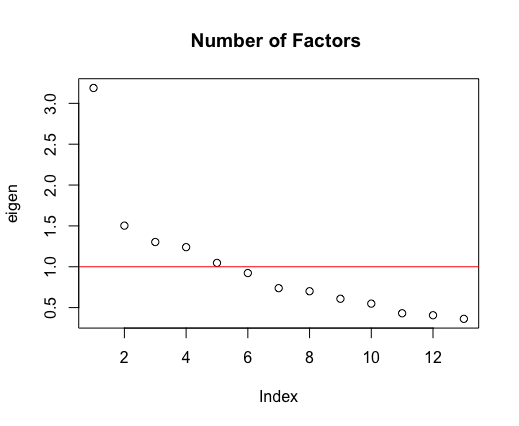
As it can be seen on the map, Balisto and KinderRiegel are the chocolate brands those have a lot more difference comparing to other brands. These brands are located far from each other which only says they are not similar.

However, it is not possible to interpret what attributes caused this dissimilarity unless doing Property Fitting. That is why, an another method is needed.

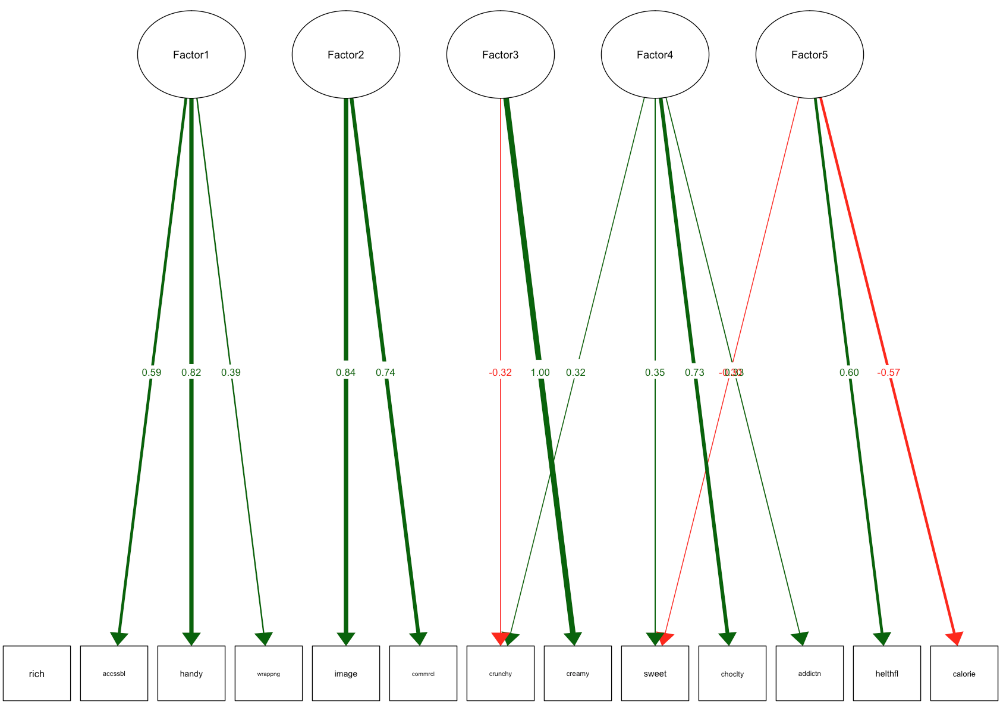
Running a regression of every attribute

To be able to interpret the position of the brand on the perceptual map, it is needed to know how the attributes are influencing this location.

SWP2b:



Yet another way to decearse dimensionality and make results more readable is Factor Analysis or Pricncipal component Analysis. For our survey results PCA was chosen over FA, due to several reasons. One of the main reason behind is that the survey was not constructed to test any theoretical model of latent variables. Despite it FA still was run over the data, to prove our point empirically. As the analysis of eigenvalues of the data showed 5 factors are used to replicate the original data. This is due to the fact that, these 5 factors are above 1 which means each of them can explain more variance than original variable. Although using so many factors, it only explains 46% of the cumulative variance which is not acceptable.



As this graph shows, there can be a significant logical link between the variables inside each of the factors. This proves there are no latent variables that should be taken into consideration.

Moving on to Principal Component Analysis

SWP2c

We generally recommend PCA as a more informative procedure than MDS for typ- ical metric or near-metric (e.g., survey Likert scale) data. However, PCA will not work with non-metric data. In those cases, MDS is a valuable alternative.

MDS may be of particular interest when handling text data such as consumers’ feed- back, comments, and online product reviews, where text frequencies can be con- verted to distance scores. For example, if you are interested in similarities between brands in online reviews, you could count how many times various pairs of brands occur together in consumers’ postings. The co-occurrence matrix of counts—brand A mentioned with brand B, with brand C, and so forth—could be used as a mea- sure of similarity between the two brands and serve as the distance metric in MDS