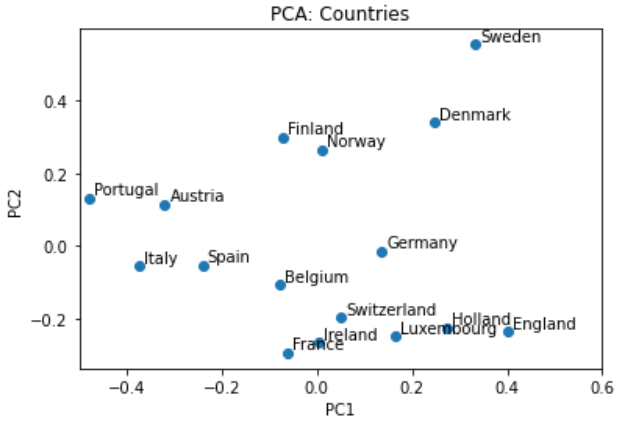
PCA: Food consumption in European countries

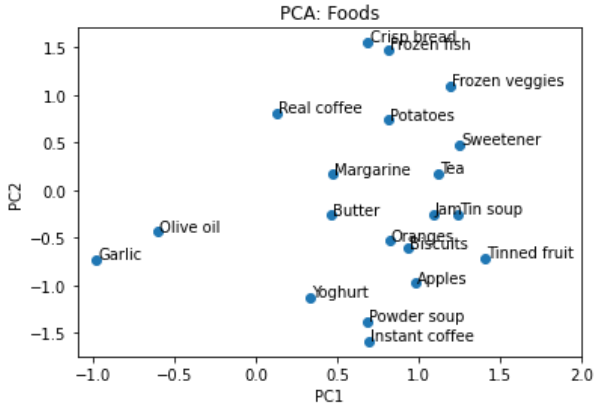
For this problem, the data matrix consists of *m* different countries which correspond to the rows and *n* different food items which correspond to the columns, compromising a data matrix which gives the consumption totals of 20 different food items for 16 different countries.

To perform PCA on this data, first feature scaling was performed on the data to ensure that all food consumption totals were given on the same range. Next, the data was normalized, the covariance matrix was computed, eigen decomposition was performed on the covariance matrix, and the first two principal components were found. After this analysis, the first two eigenvectors were associated with the first two principal components of the countries. The scatter plot of the two-dimensional representations of the countries using their two principal components can be seen below:



Notably, it can be seen that some countries are grouped based on the first two principal components. Italy, Spain, Belgium, and France are close to each other and have negative correlations, meanwhile Norway, Denmark, and Sweden are roughly grouped and have positive correlations. This may suggest that countries grouped similarly have a similar diet, such as Italy, Spain, Belgium, and France, and countries grouped oppositely to the first group have contrasting diets, such as Norway, Denmark, and Sweden. Additionally, it is notable that each of these groups of countries also share a physical proximity, which may add to their similarity in diet.

Using the same PCA analysis framework as before, now the projection of the first two eigenvalues correspond to the first two principal components of the food items. The scatter plot of the two-dimensional representations of the food items using their two principal components can be seen below:



Notably, this plot provides information about how the consumption of certain foods relates to the consumption of other foods for the diets of different countries. For example, where more crisp bread is consumed more frozen fish is consumed, or alternatively where more olive oil is consumed so is garlic. These two examples are also roughly inversely correlated to each other as well, meaning where more olive oil is consumed less frozen fish tends to be consumed. When looking at the location of food items in comparison to the countries in the previous plot, it can be seen how different countries diets differ. Using the same example, countries such as Norway, Denmark, and Sweden may tend to eat more crisp bread and frozen fish and less olive oil and garlic, with the opposite effect for Italy, Spain, Belgium, and France.