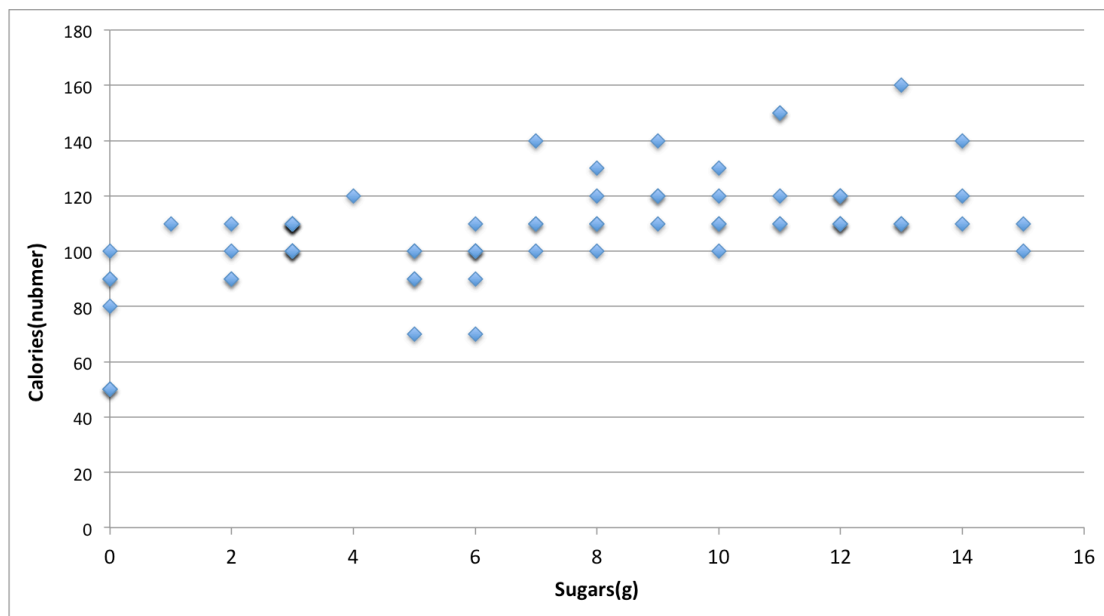


CSE591 Assignment 3
Name:Ersi Zha Id:1206169363

Task1:

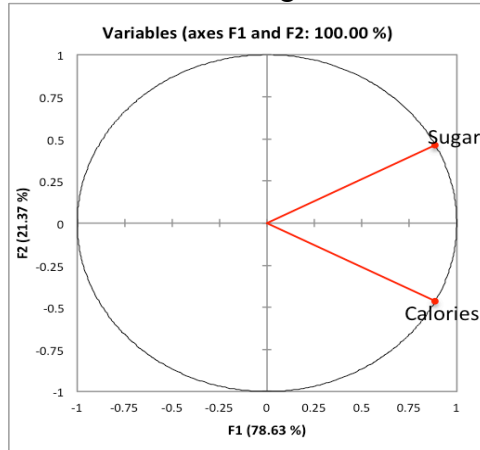
I choose two variables calories and sugars after analyzing the breakfast cereal data set. Why I choose these two variables as the two axes? I think calories and sugars are directly related to the nutritional value of cereal and many people will take these two variables into consideration when they are facing different brands of cereal. For example, some people would prefer the cereal with high calories and sugars in order to concentrate on their daytime work, whereas some people prefer low calories and sugars in order to keep fit and healthy.

I utilize excel as the tool to generate my scatter plot. In my scatter plot, the x-axis represents sugars(g) and y-axis represents calories(number). As shown in the figure, cereals with 90-130 calories account for the most part in the total cereal data set. The statistics of sugars distributed homogeneous in the range of 0 to 16. Furthermore, in the scatter plot, there are four kinds of cereals contain 0 sugars and two kinds of cereals contains 15g sugars but with calories below 110. So we can conclude that although there is a relationship between sugars between calories, the sugars content is not the only factor to influence calories content. Cereal with high sugars doesn't mean that they have high calories.



Task2:

I utilize excel(xlstat) to run PCA on the data and then generate the plot. Xlstat is a plug-in of excel. To better explain my PCA plot, I display the following correlation circle figure in this task. This correlation circle interprets the meaning of axes and shows the trend of sugar and calories. Then I will explain my PCA plot.



In the following PCA plot, the points on the right upper corner represent those cereals with high sugar content and the cereals on the right lower corner represent those cereals with high calories content. Some points present more than one cereal because some cereals have the same sugars and calories content. By analyzing the plot, Golden_Crisp has the highest sugar content 15, whereas Puffed_Rice, Puffer_Wheat and All-Bran_with_Extra_Fiber have the lowest sugar content 0 because they are at the most left upper corner. Also Puffed_Rice, Puffer_Wheat and All-Bran_with_Extra_Fiber have the lowest calories content 50 among all the cereals, on the contrary, we can easily judge that Mueslix_Crispy_Blend contains the highest calories 160.

