Study of comparaison between the nutriscore and the composition of a product for cereals in France

Tristan Fontanière B2

Emile Dechenaud B2

The nutriscore is a logo given to dietary products since 2017, its goal is to inform customers about the nutritional value of a product. It is based on a color code to make it easy to understand, it goes from green for A graded products to red for E graded products. When a product is richer on fibers and proteins, it usually get a better nutriscore than a product which contains a high amount of fat, sugars or salt. However, the nutriscore is not always reliable, in fact it doesn't take in account if the product contains added nutrients or the degree of transformation. What is important to note is that A doesn't mean "good for health" and E doesn't mean "bad for health".

The goal of this study is to highlight a link between the nutriscore grade which is given to a product and the composition of a product. Here we will make a focus on all the products that contains cereals (such as pasta, rice, cereals...) in France. The red thread we will follow is :

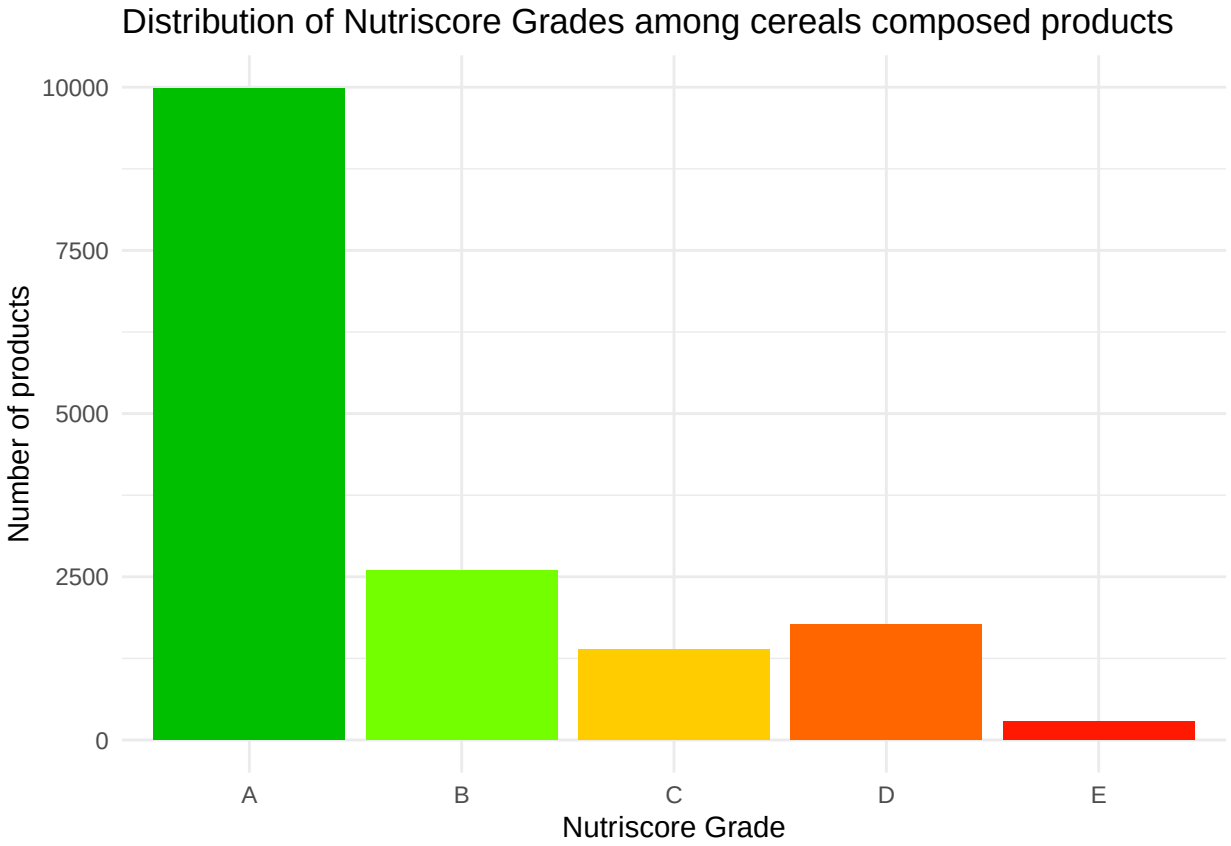
What correlation can we establish between the nutriscore and the nutritional variables ?

After filtering all the products that are referenced in the nutriscore grade data base, we ended having 16 043 datas to study. They were cleaned and filtered through cereals composed products and France and we also deleted the datas that were considered wrong to make this study more accurate. So to speak the lines that were marked as untruthful or had unreal values were deleted, as well as the duplicated values.

There are differents variables in the data set we used, for each products we can look at its bar code, its web page, its name, its brand, its stores, its owner, its labels and more importantly its composition. In the composition we have for each 100g the amount of differents nutrients the fats, the sugars, the salt... We also have created columns that indicate if the product is vegetarian, organic, vegan and also if it contains palm_oil. There is also an indicator of the level of each nutrients.

There will be different parts in this study, first we are going to see the distribution of cereal based products throughout the different nutriscore. Then, we will look at the percentage of vegan and vegetarian products. After that, we will focus on the composition of the products to understand how its composition impact its nutriscore. And finally, we will look at the level of each nutrients as a conclusion.

To begin with, let's look at the number of product in each grade. In this bar chart we will represent the amount of products in each nutriscore grade, each color represents a nutriscore grade based on the official colors. The x axis is each nutriscore grade and the y axis is the amount of products.



What we can see throughout this graphic is that there are four times as many A graded products as B graded products. In fact, the A graded is the highest grade given among the cereals based products. On the other hand, the C and D graded products have both very low amounts, there are roughly around 2000 products that have this grade. But the lowest amount is the E products, there are around a couple hundreds of them.

Next, let's look at the different labels that are useful to understand the products. We are going to look at the proportion of products for different labels that are organic, vegetarian and vegan. It's very important to note that one product can have two different labels. The table contains the different labels and the proportion of these in percent.

```
##     Category Proportion
## 1      Vegan       35.7
## 2 Vegetarian       40.1
## 3    Organic       32.8
```

What we can see here is that many products are vegetarian, in fact 40% of all the products we selected. Moreover, the vegans and organic products are also quite high with respectively 35.5% and 32.5%. Which actually make sense, indeed the majority should be vegetarian because it's made of cereals, but a lot of products are badly given from people who entered the datas on the site.

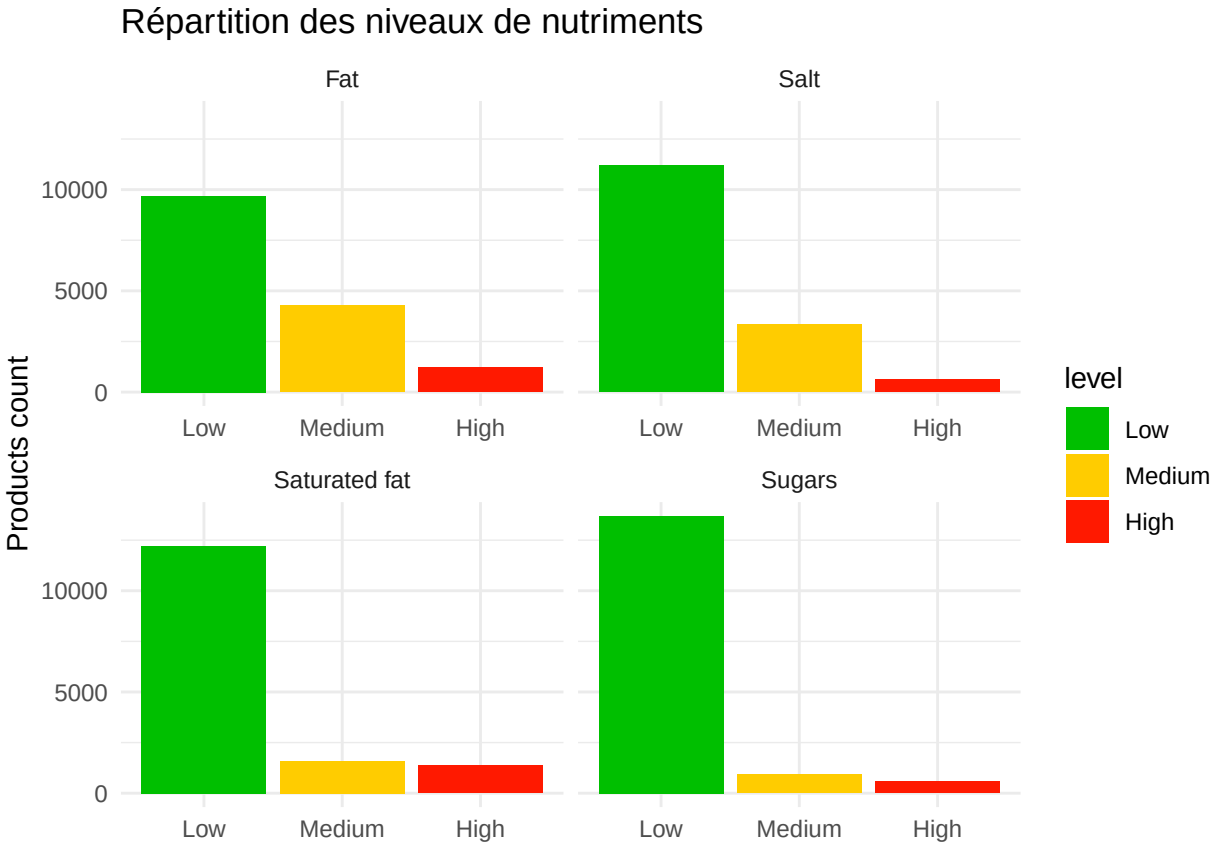
Then, let's look at the different nutrients in each products. We are going to look at for each ingredients, the minimum amount in a product, the maximum, the first and third quartil and the median.

```
##        Variable Min   Q1 Median   Q3   Max
## 1           Fat   0  1.3      2  4.9 100.0
## 2      Proteins   0  7.2     11 13.0  85.5
## 3 Carbohydrates   0 45.4     69 73.0  96.0
```

What we can see here is that there are strange values, in fact all nutrients have a minimu of 0 and a maximum around 90. The median is also strange for the fats and the proteins, but it can make sense so we are not going

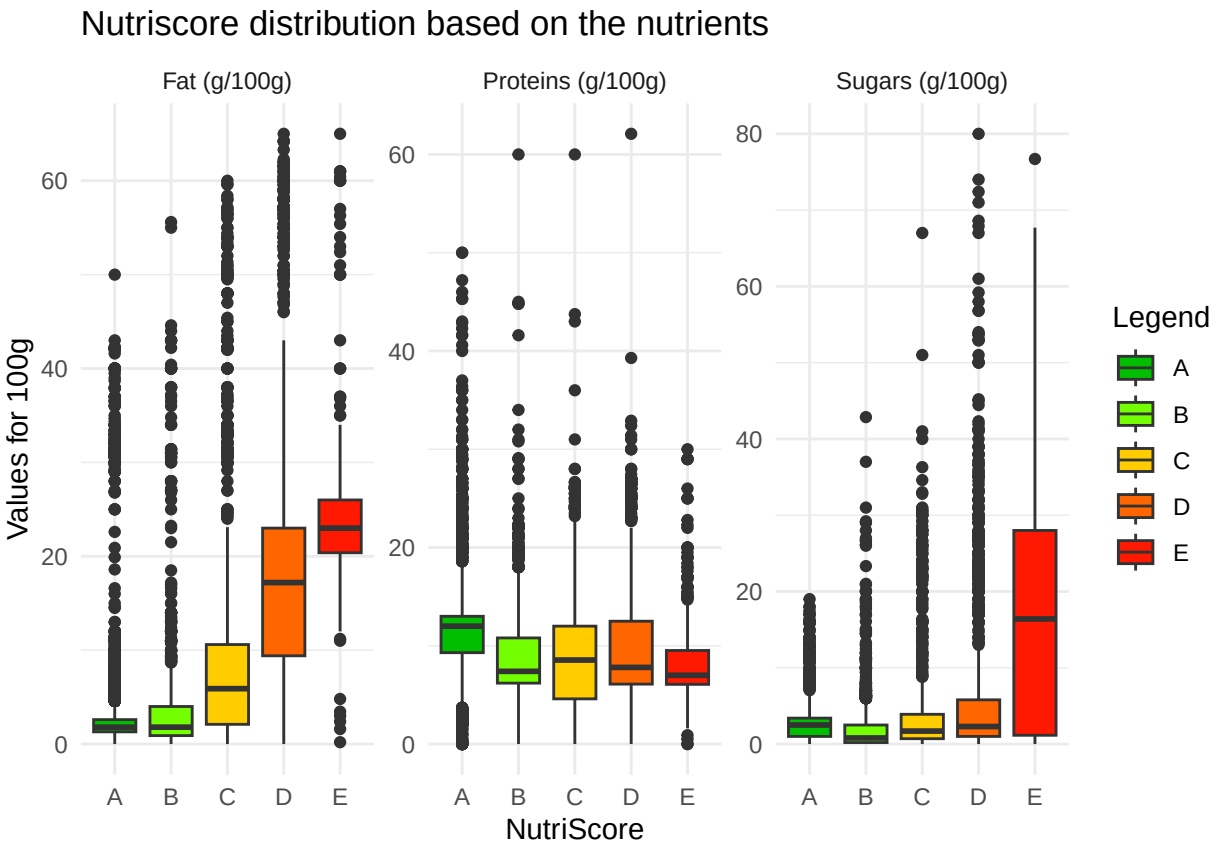
to mind it. The first and third quartils for each of them seems also realistic. For the rest of this study we are going to filter a bit more our values and take only : fats under 80g, proteins under 70g and carbohydrates under 80g.

Now, that the data are better filtered let's look at the level of each nutrients in the composition of products. With again a bar chart we will represent the amount of products on the y axis and the level of nutrients on the x axis. There will be four different graphs that represent each nutrients.



On these four graphs, what is instantly eye-catching is that the level of each nutrients is most of the time quite low, the amount of products that have low level of each nutrients are always high. There is a slight difference for the level of fats, there are twice as much products that have a medium level of fat than products that have medium level of saturated fat. Also, while the medium level of salt is around 2500, the medium level of sugars is around 1000. The amount of products that have low level of nutrients is always between 10 000 and 15 000. Indeed, almost all the products have low level of sugars. Therefore, what we can tell here is that the majority of products do not contain things that can underrate their nutriscore such as sugars and fats.

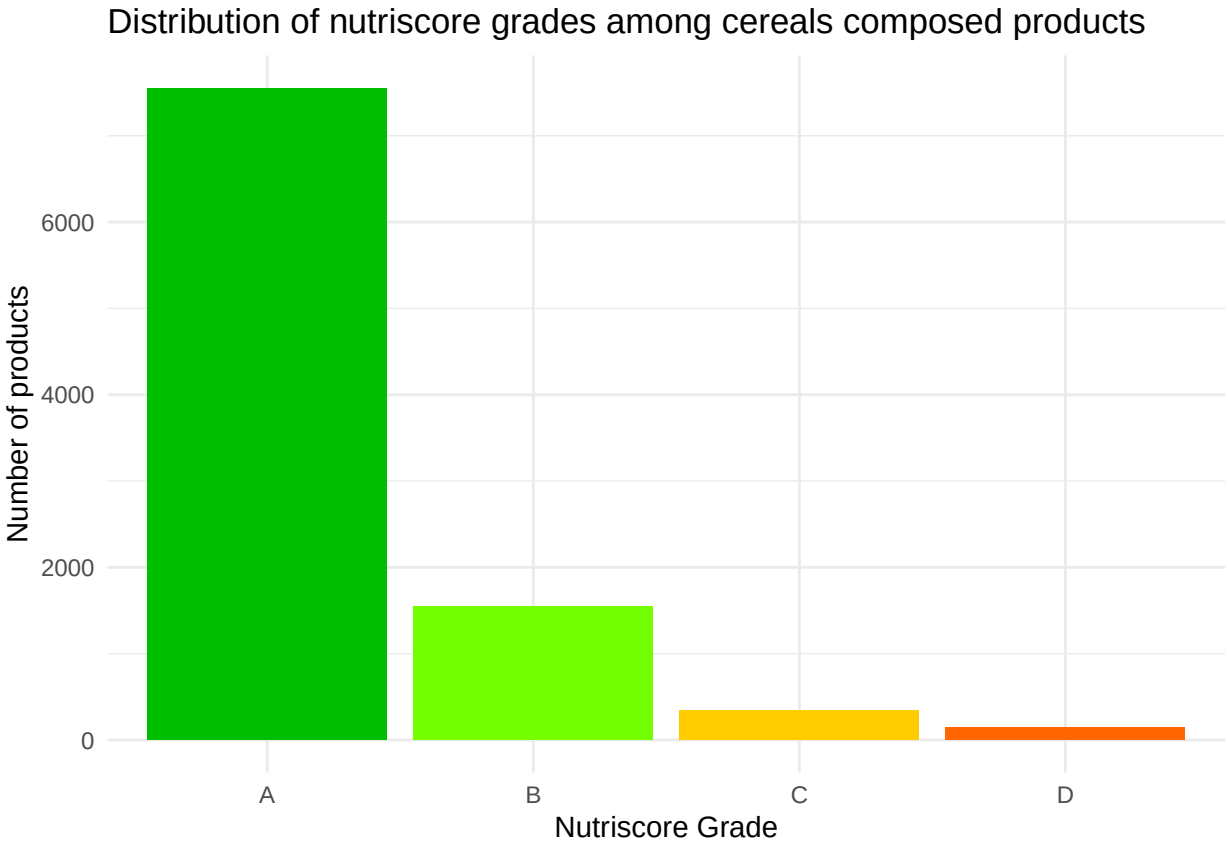
Before seeing the real link between nutrients and nutriscore grade, let's just quickly look at the distribution of nutriscore grade for each nutrient values. We will represent it with boxplots and we will only take the amount of fat, proteins and sugars because the amount of saturated fat contains a lot of aberrant values.



Let's study these boxplots one by one, there are a lot of aberrant points because some products are badly informed but we can not mind them. The first one which represent the nutriscore grade compared to the amount of fat simply shows that the higher the amount of fat, the more the nutriscore tends to E. The distribution of nutriscore compared to the amount of proteins is likely to be average, it seems like it doesn't change impact the nutriscore that is given to the product. And the most obvious one is the level of sugars, when a product goes above 15g of sugars most of them are E graded.

Now that we have seen that, for the conclusion we will once again filter the datas and now we only take in account the products that have low level of each nutrients.

On the next graph, we will look at the amount of products by nutriscore grade once again.



What we can see is that there are no E graded products here, on a total of 8,000 products approximately, 7,000 are A graded, 1,000 are B graded and the others are C or D graded.

In conclusion, we have seen that for cereals based products in France, the most important factors are the sugars level and the fats level. Then, we have seen that a lot of products are either vegan or vegetarian. And after that we have seen that almost all the products have very low level of each nutrients especially the level of fat and sugars again. When taking this into account, when we filter with all of that, the result we have is that most of the products are A graded. Finally, we can say that the link between nutrients and nutriscore for these products are the level of fat and sugars, and that without high level of these variables the products are almost all A graded.