What link can we establish between the nurtriscore and the nutritional values of the dressings and sauces in the Untited States ?

Introduction

The nutriscore is a labeling that we can found on most of the products in supermarkets. It helps us to be informed of the global nutritional quality of products. Yet, it can be not significant if we compare two products of two different food groups. This is why it is important to understand the specificity of the food groups. In this study, we will analyze the dressing and sauces in the United States, to do so we will use the data of 19,413 products from the open data website openfoodfacts. In this document we will first see the repartition of all the dressings and sauces between the five nutriscore grades: A, B, C, D and E. Then we will analyze the impact of the three major nutritional values (carbohydrates, fat and proteins). We will finally see in details which factors are responsible for the nutriscore of our products.

Nutriscore of our products

First of all we need to see the average nutriscore grade of our products. In order to find it, we will make a bar chart with the nutriscore grade on the x-axis and the percentage of products of the database on the y axes. Finally, the bars are going to be colored according to the nutriscore grade they symbolized.

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(count)` instead.
```

Percentage of products per nutriscore grade
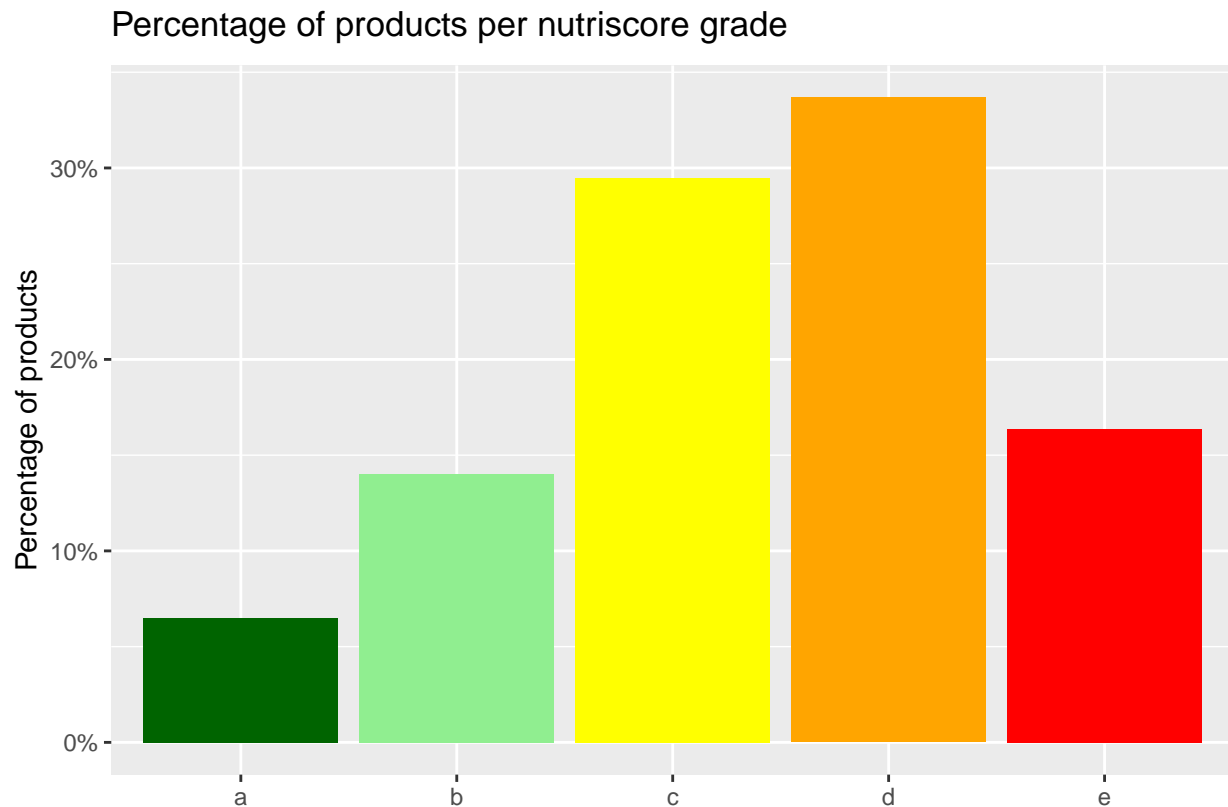
We can see that most of the dressing and sauces are ranked C or D. We will now see what factors are making those products rated so low.

What makes that nutriscore ?

Major categories

To begin with, we need to know which one of the important categories of nutritional values are the most responsible for the nutriscore grades we found. We will analyze the quantity of carbohydrates, fat and proteins and what is their impact on the nutriscore.

Impact of carbohydrates

The first category are carbohydrates. To see how they influence the nutriscore, we will make five box plot representing the amount of carbohydrates (y-axis) in every product of the different nutriscore grade (A, B, C, D and E) (x-axis).

## Quantity of carbohydrates for each nutriscore grade
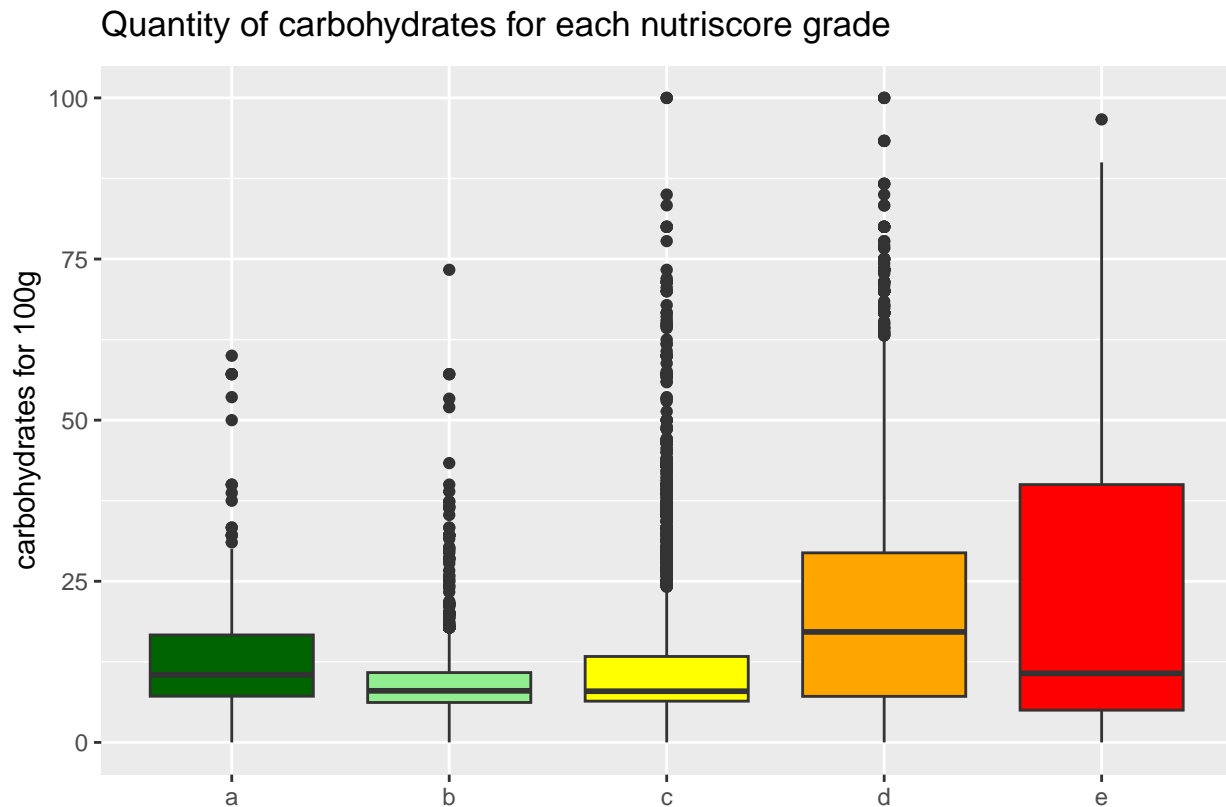
The first observation we can make is that the median of each box plot is significantly low, it means that 50% of every product only have 12g of carbohydrates in their nutritional values. Yet, the box plots of the nutriscore D and E are more than twice as bigger as the other ones, it means that the products with the highest quantity of carbohydrates have a bad nutriscore grade.

Impact of fat

The second category is fat. To see how it influences the nutriscore, we will make the same graphic but with the quantity of fat for 100g on the y-axis.

Quantity of fat for each nutriscore grade
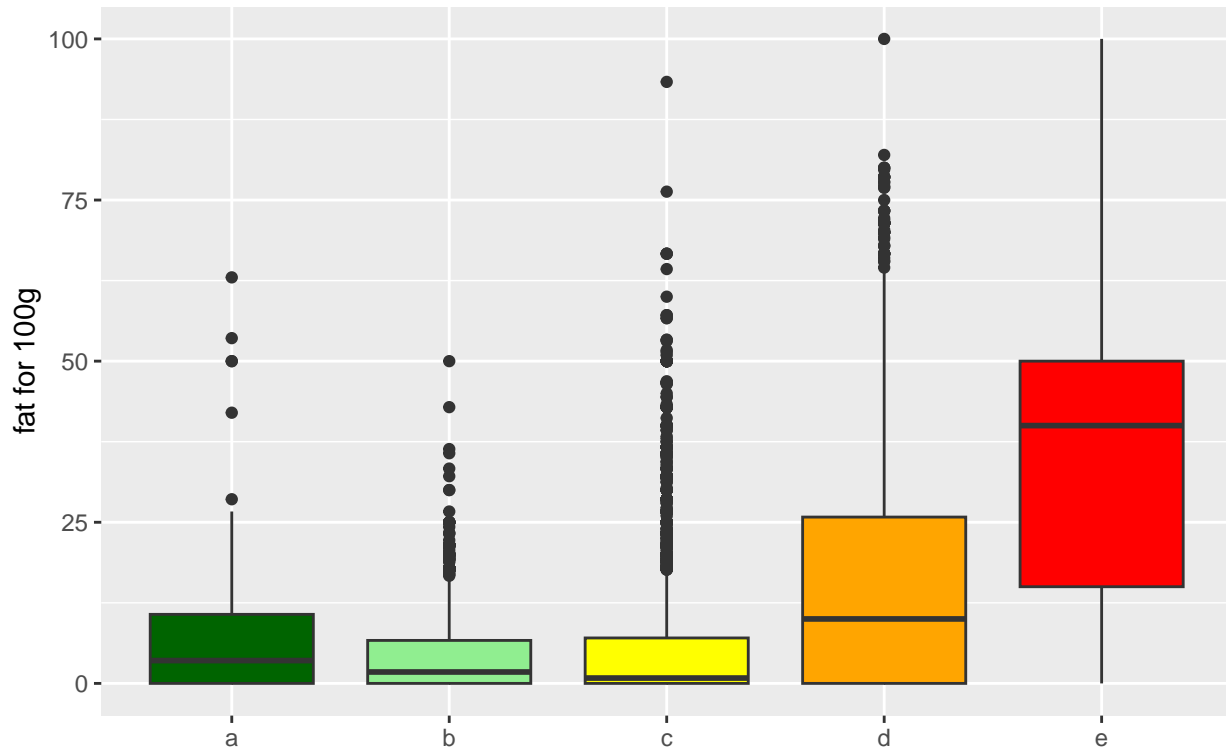
In contrary to carbohydrates, we immediately see that the more fat a product contains, the worst is its nutriscore grade. In fact, the median of the first three nutriscore are extraordinary low compared to the median of the nutriscore E. Moreover, the size of the first three boxes are twice as small as the two last ones. It means that almost all the products of the nutriscore A, B and C have less than 12g of fat in their composition. For the nutriscore D, 75% of the products have less than 25g of fat in their composition. This proportion goes up for the nutriscore E with 50g or less of fat in 75% of its products.

Impact of proteins

The last category are proteins. The analysis of this category will be the same as before to compare them in the best way. The y-axis represents now the quantity of proteins for 100g.

Quantity of proteins for each nutriscore grade
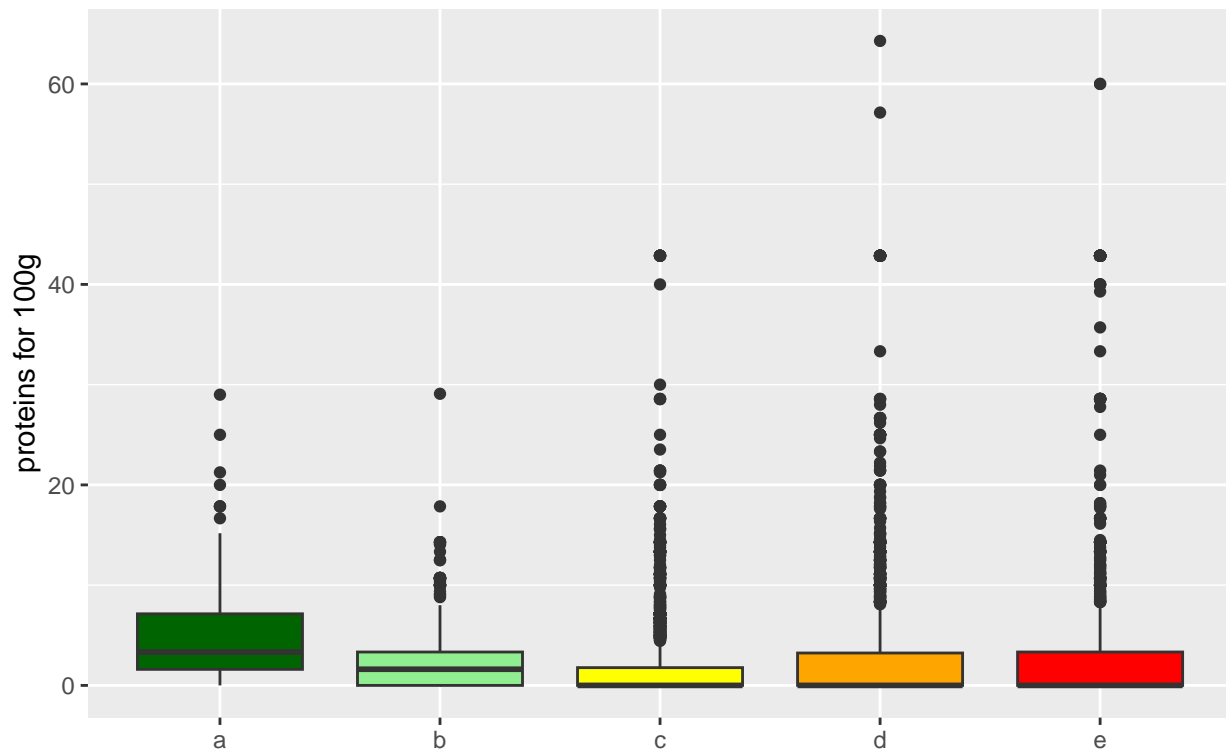
This box plots are significantly different from the other ones we made. In fact, all the nutriscore have a very low quantity of proteins in their composition, almost all the product we are studying have less than 10g of proteins in their composition. Yet, we noticed that the first and third quartile of the nutriscore A are slightly higher than the others, which means that the quantity of proteins for our category of products influences the nutriscore in a good way.

Conclusion

Thus, the big categories we need to analyze in details are fat and carbohydrates because they seem to have a high influence on the nutriscore of the dressings and sauces in the United States. The proteins are not very significant compared to the others, this is why we won't use them anymore.

Other factors

In this part, we will find out what is the nutritional value that have the baddest influence on the nutriscore of the dressings and sauces in the United States. We will analyze the saturated fat, the sugars and the salt.

Impact of saturated fat

In the previous part we saw that fat has a really negative influence on the nutriscore, we need to know now which kind of fat and this is why we chose the saturated fat. We will use the same kind of graphics than before: box plots. The x-axis will be the quantity of saturated fat for 100g and the y axes the nutriscore grade.

Quantity of saturated fat for each nutriscore grade
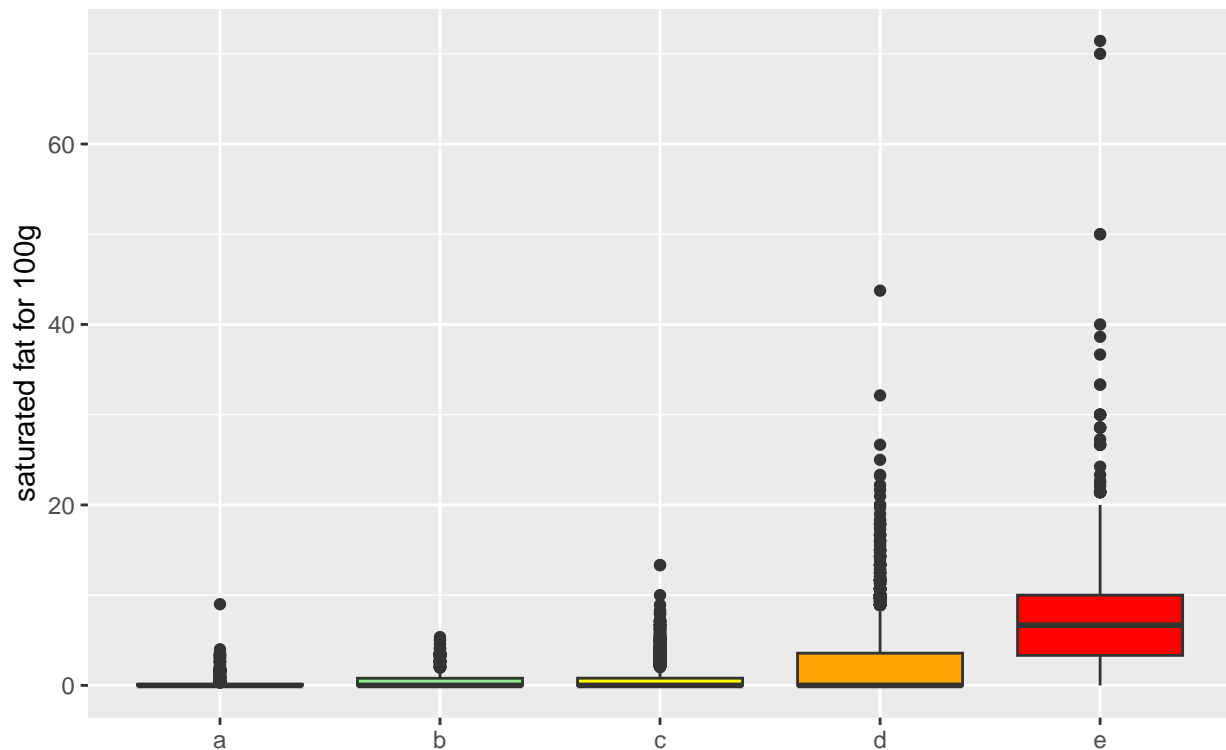
The first observation we can make is that the medians of the nutriscore A, B, C and D are at 0, it means that 50% of products from these nutriscore contains no saturated fat. Yet, the third quartile of the nutriscore D is a lot higher than the others which represent the fact that there are more saturated fat in the products ranked D. Finally, the first quartile of the nutriscore E is the same as the third quartile of the nutriscore D, we can conclude that the quantity of saturated fat in the products ranked E is significantly higher.

Impact of sugars

In the previous part we saw that carbohydrates have a bad influence on the nutriscore, yet we need to know which kind of carbohydrates and this is why we chose the sugars. We will use the same kind of graphics than the previous graph, but the y axes will be the quantity of sugars for 100g.

Quantity of sugars for each nutriscore grade
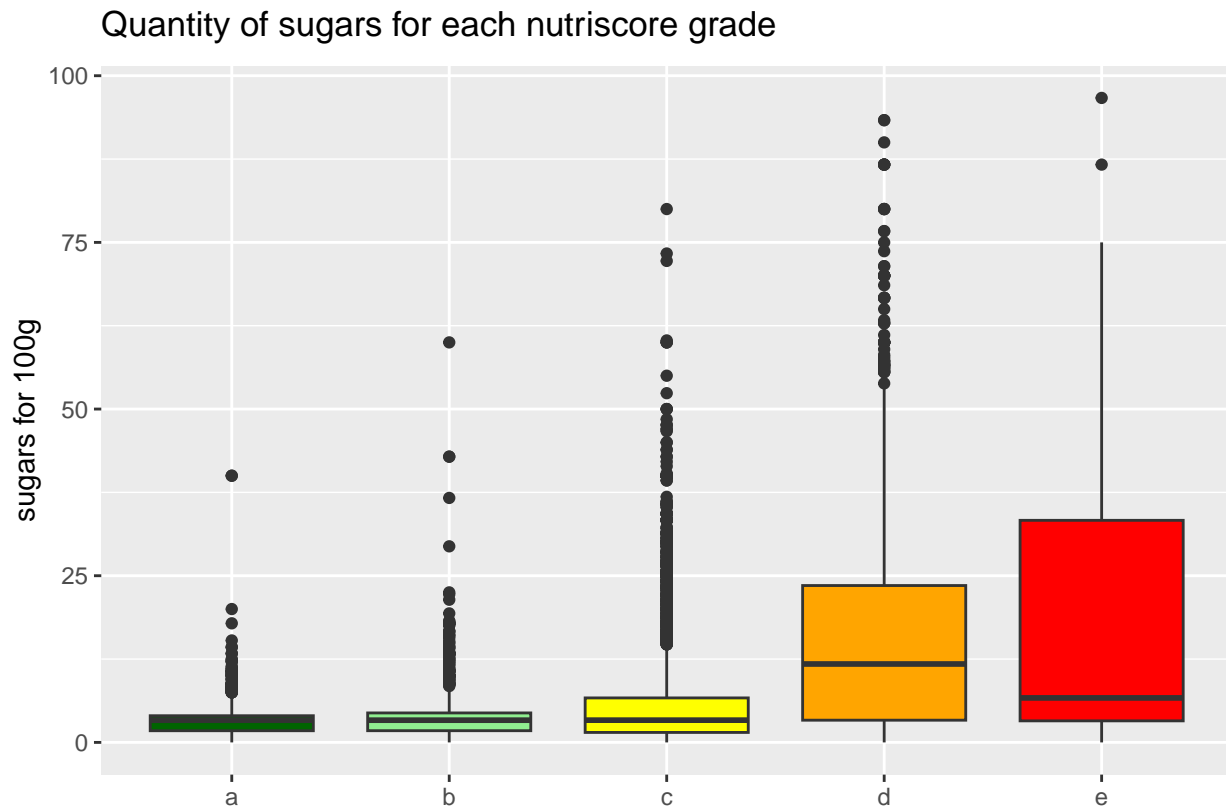
To begin with, the nutriscore A, B and C are not very affected by the quantity of sugars. Now we will see the nutriscore D and E. The first analysis is that the median of the nutriscore D is twice as big as the median of the nutriscore E, in fact 50% of products ranked D have less than 12.5g of sugars and this quantity is 6g for the product ranked E. Yet, the third quartile of the nutriscore D is smaller than the one of the nutriscore E. 75% of products ranked D have less than 25g of sugars and this quantity is approximately 34g of the nutriscore E. Thus, the sugars are influencing the nutriscore in a wrong way.

Impact of salt

Another important factor is the quantity of salt because a high consumption of salt is not healthy. We will use the same kind of graphic with the quantity of salt for 100g on the y axes. Yet, Lots of products seem to have a high quantity of salt (almost 50g sometimes), these values are incoherent, and we have to ignore them. This is why we will only analyze the products with less than 10g of salt per 100g.

Quantity of salt for each nutriscore grade
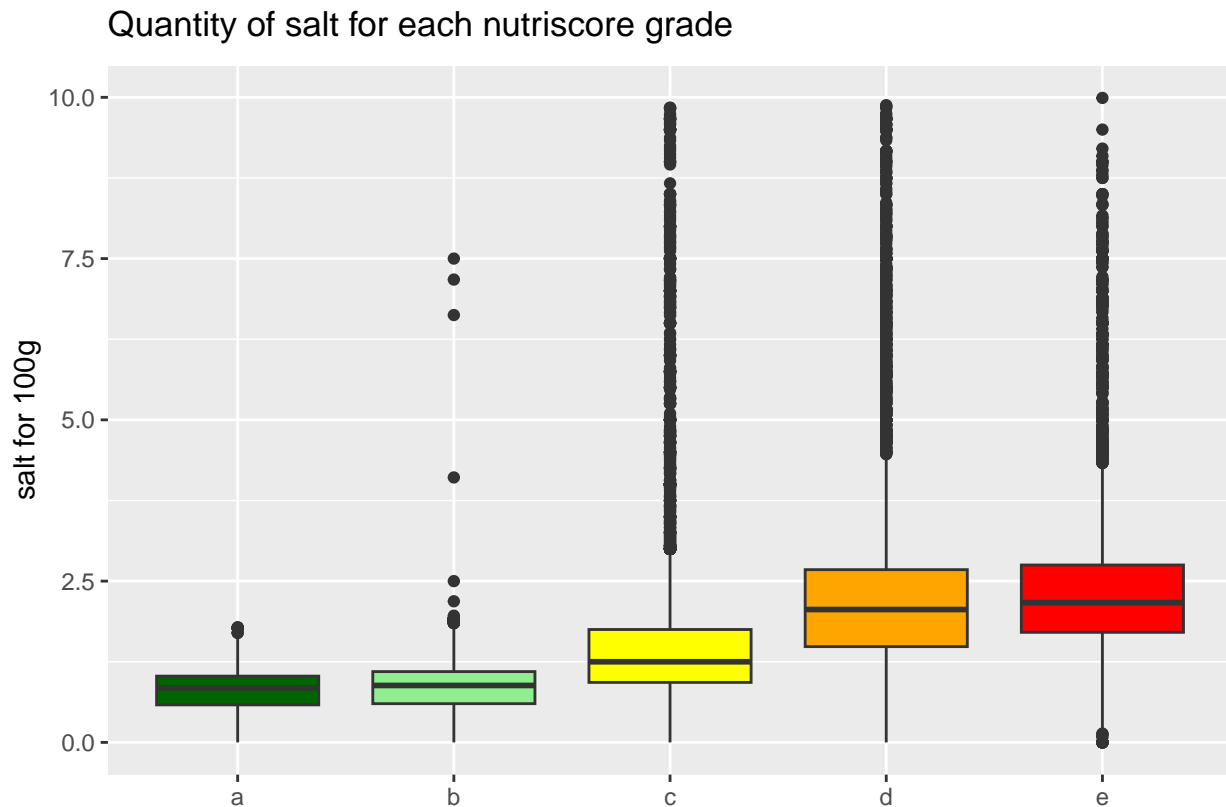
On this graph, we clearly see that the more salt there is in a product, the worst its grade is. Notwithstanding, the quantity of salt in the nutriscore D and E are almost the same.

Salt or sugars

We saw that salt and sugars have an influence on the nutriscore. But our products are sweet and salty, so we need to know if we have more salty products or sweet products to have a better analysis. To do so we made two arrays of data which are counting the number of products for the different levels of salt and sugars. h means high, m means medium and l means low.

```
## # A tibble: 3 x 2
##   level_salt nb_produit
##   <chr>           <int>
## 1 h                9643
## 2 l                 929
## 3 m                8841
```

This array of data shows us that 95% of our product have a high or a medium level of salt.

```
## # A tibble: 3 x 2
##   level_sugars nb_produit
##   <chr>             <int>
## 1 h                  5357
## 2 l                 10332
## 3 m                  3724
```

This array of data shows us that 46% of our product have a high or a medium level of sugar.

This shows us that we have more salted products and consequently that the salt has a bigger influence than the sugars in the nutriscore of the dressings and sauces in the United States.

Identification of the two main factors

Now that, we identify the two major factors of the low nutriscore of our products: saturated fat and salt. We will make a final graph, in the form of a scatter plot, with the quantity of saturated fat in 100g of products on the x-axis and the quantity of salt in 100g of products on the y axes. Each point represent a product and will be colored with the color of its nutriscore.

Quantity of saturated fat in function of the quantity of salt linked with the nutr
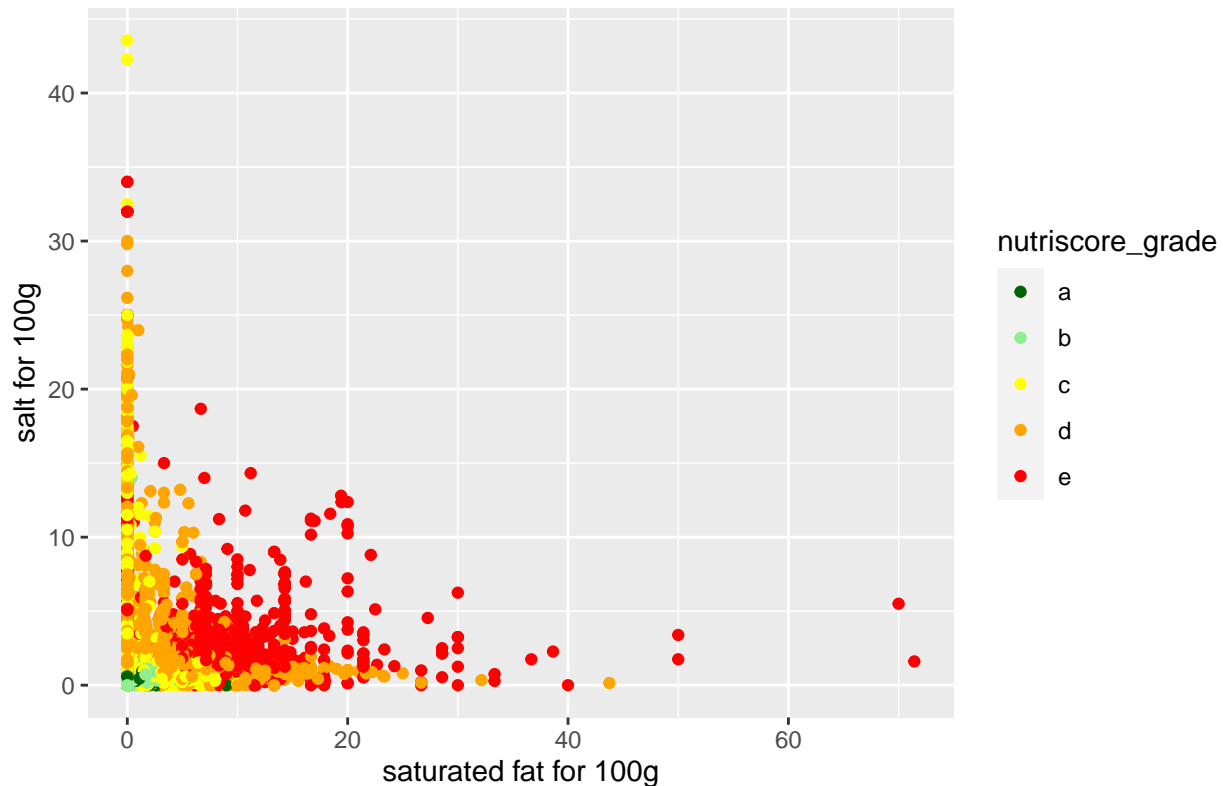
Firstly we see that products with saturated fat and salt are mostly ranked C, D or E. So salty and fatty products have a bad nutriscore. Yet, The saltiest products seem to be ranked C or D and the fattiest products are ranked E. We can conclude that the saturated fat have a bigger influence than the salt on the nutriscore.

General conclusion

The food groups we analysed (dressings and sauces) represents a wide array of products which can be salty, sweet or fatty. After all the graphs we studied, we can conclude that the fat is the factor that influences the most the nutrional rating. Thus, when you buy dressings and sauces, be aware that a bad nutriscore grade indicates a fatty product. Yet, stay aware of the sugar and salt.

Study made by Sarah Sekma and Émilie Dubief