

Difference and similarities between the American and French Food Products

Anas Seddiki

anas.seddiki@epfl.ch

Constantin de Bentzmann

constantin.bentzmann@epfl.ch

Clément Grillet

clement.grillet@epfl.ch

Abstract

This document presents a full report on similarities and differences between how the United States and France consume food. We explore the range of food products sold in these two countries, and try to see if there are any significant nutritional differences. We also decide to analyse the ecological footprint of food manufacturers in both of these countries, as ecology is a very relevant matter. We do this by taking a look at the packaging materials used by the major food manufacturers. We will also talk about some limitations that we have encountered in our dataset that might affect our results.

1 Introduction

The Open Food Fact database is taken from world.openfoodfacts.org website. Open Food Facts is a non-profit association of volunteers. They host a database of food products with ingredients, allergens, nutrition facts and all the tidbits of information we can find on product labels. Their data about food is of public interest and has to be open. The complete database is published as open data and can be reused by anyone and for any use.

The goal of this project was to compare food products between countries to see whether there is a nutritional difference, and if the range of available products is a cause of an unhealthy diet.

2 The data set

This is a dataset of food products listing the ingredients and nutritional facts of more than 1,000,000 foods from over 150 countries in the world. The data is meant to be free to the public to help users decode food labels, and make better food choices

in general. We made some assumptions in order to work on this data set and to draw conclusion. Namely, that the entered country of origin of a product was reliable. We are aware that some products sold in the USA can be sold in France as well. Even if, we will see that there is a strong difference in brands between the USA and France. So the intersection of product sold in both country is pretty low.

3 Pre-processing

The dataset we are using has many flaws, so the pre-processing was a major phase before being able to start our analysis. This involved removing features due to too little data, cleaning text and dealing with missing values. Another big downside to this dataset is the origin of the products. More than 50% of the entries are from France and about 20% are from the United States (Figure). This limited the geographical scope of our analysis, which was initially supposed to be global. Indeed, we decided to pivot from a global analysis, to a comparison between the two most represented countries to increase the accuracy of our results.

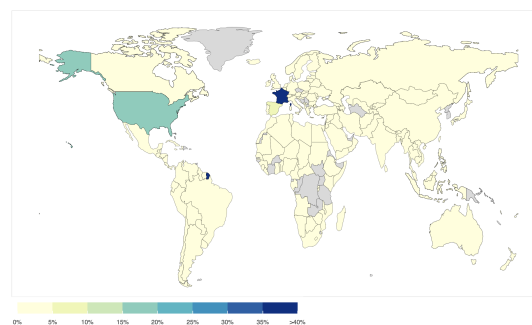


Figure 1: Distribution of product origin in the dataset

4 Food trends and nutrition score

We often stigmatize the American diet by saying that it is unhealthy compared to the rest of the

world. So we decided to look into the nutrition facts of the products sold in the United States compared to France to see whether this stigma is confirmed by data.

4.1 Correlation matrix

We first decide to take a look at the correlation matrix (Figure 2) of the different nutrients to get an idea about their relationship and about the basis of the nutritional score. We can deduce and discover interesting information. For example we can get a visual idea of the formula used to compute the nutrition score. It takes into account many different factors and ranges between -15 for the best products and 40 for the worst ones. Here we have a clearer view of which nutrient impact the most the nutrition score. For example the energy, the fat and trans-fat contained in a product will drastically increase the nutrition score which seems logical. But we can see also more interestingly that the amount fiber has a negative correlation with the nutrition score, the more fiber there is the lower the nutrition score, thus the healthier the product. The sugar and fat are negatively correlated, the fatter the product the less sugar there is in it. Negative correlation can also be seen between protein and sugar, if the product is rich in protein it will likely be low in sugar. Looking at this matrix we can fetch much information that will help explain some of our later results.

4.2 Nutrient content in French and American products

We look at the difference in mean content of nutrients commonly known as 'unhealthy' (fat, sugar, sodium, palm oil, trans fat, saturated fat) in products sold in the US opposed to France.

The results we obtained (Figure 3) go back and forth between the two countries, and do not point to a particular country diet as being healthier. We can see that French products have a higher mean fat content, whereas American products contain more sugar. French products also contain more ingredients that may contain palm oil. The nutritional score is higher (worse) in French products due to the fact that this value is highly correlated with fat. No significant difference is observed in saturated fat and salt content. Trans fats turned out not to be an interesting statistic to look at because not enough French products had data for this field. Not much can be concluded from these results, we cannot say that one country is healthier than the

other. Furthermore, this data set is not very indicative of a country's diet as it does not represent consumption behaviour, but only a range of available products. Healthy and unhealthy products are found in every country, and this is not what determines the country's diet. The diet is rather determined by the consumption behaviour and habits of the country's population : Which products do they buy? How many times do they eat out on average? Do they opt for healthy meals or fast food when they do eat out?

4.3 Nutrition score between the US and France

Here we can see a clear difference between the United States and France. If we take a look at the graph about the nutrition grade (Figure 5), we can see that the US has a higher percentage of products with the lowest grade. Regarding the grades from 'B' to 'D' we do not notice any big differences, even if France has a slightly higher share of 'B' graded product. The main information comes for the product falling in the 'E' grade. Indeed, French product are almost twice as many compared to the US ones. This is directly linked to our previous results. If we take a look at the graph focusing on fat quantity in the products, we see that french products tends to be fatter than American ones. Then, if we remember the correlation matrix we saw that fat is highly correlated with the nutrition score, thus grade. If we take a look at the distribution of the nutrition score we can observe the same trend as the nutrition grade. Though, we can notice a bi-modal distribution for both of the countries. We observe peak around 5 for both of the countries, then another one a little after 10. We notice then that the products present in the data-set are separated in two different clusters. This can be explained that very few products fall in this categories. Products grades tend to be polarized towards the extremes.

To further explain the difference between the US and France in the nutrition grade we could look again at the previous graphs. The correlation matrix shows that fat have a much higher impact on the grade than the sugar. We also see that the US have more sugar than France but less fat. The reason for such a difference can be explained by taking a look at policies that were in place in the second half of the 20th century in the US.

In a now famous scientific paper (Cristin E.

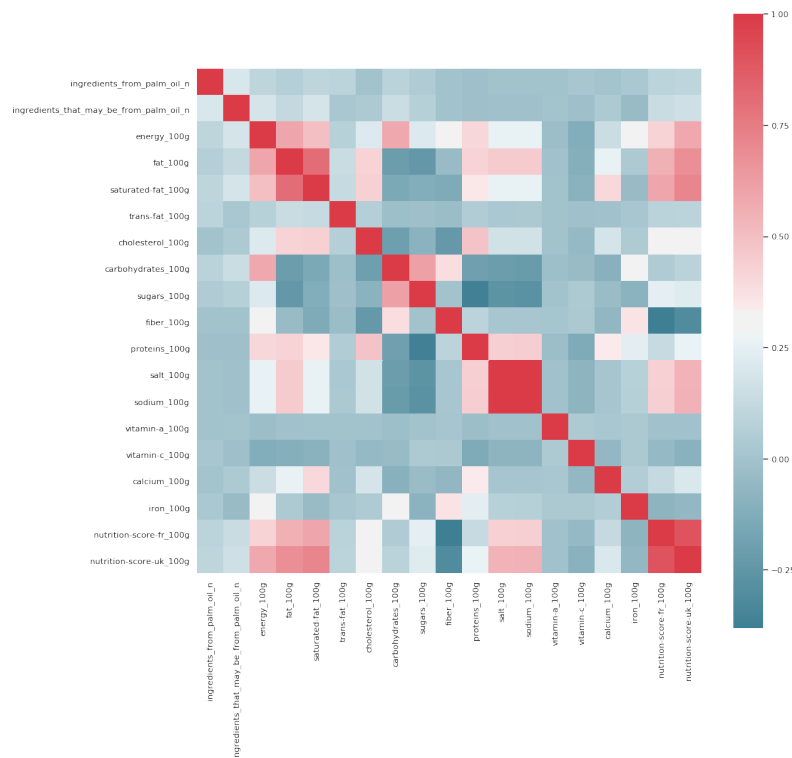


Figure 2: Correlation matrix. The Pearson correlation coefficient was used to obtain the correlation values between nutrients.

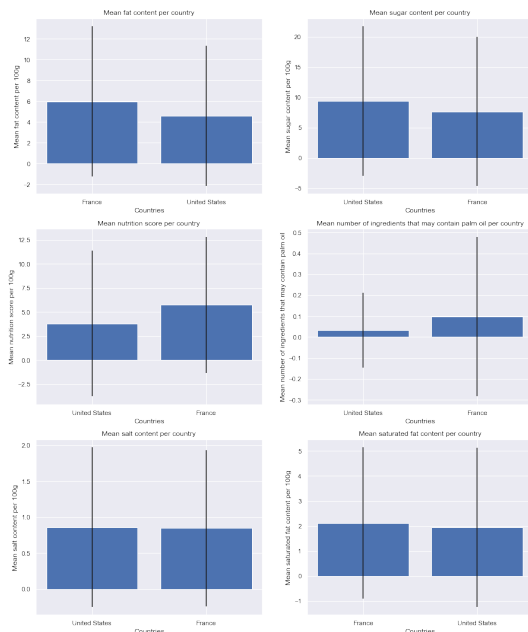


Figure 3: Mean nutrient content in French and US products

Kearns, Laura A. Schmidt and Stanton A. Glantz, 2016) we see that a trade group called the Sugar Research Foundation, known today as the Sugar Association, paid three Harvard scientists the

equivalent of about \$50,000 in today's dollars to publish a 1967 review of research on sugar, fat and heart disease. The studies used in the review were handpicked by the sugar group, and the article, which was published in the prestigious New England Journal of Medicine, minimized the link between sugar and heart health and cast aspersions on the role of saturated fat. "They were able to derail the discussion about sugar for decades," said Stanton Glantz, a professor of medicine at U.C.S.F. and an author of the JAMA Internal Medicine paper. Thus America, for the past five decades started to substitute the fat by sugar in their products. Creating a very large amount of "hidden sugar".

On the other hand, France has always been a country with a fat diet. Cheese is a very important feature of the french diet for example. If we use the numbers given by the report from the "High Council of Public Health" using the QUALI database in 2015 (French high council of public health , 2015) cheeses cannot be graded better than "C", moreover we see that more than 90% of cheeses are "D" or "E". With the same reasoning if we look at delicatessen in the report, it shows that almost 70% of these products are in the "D"

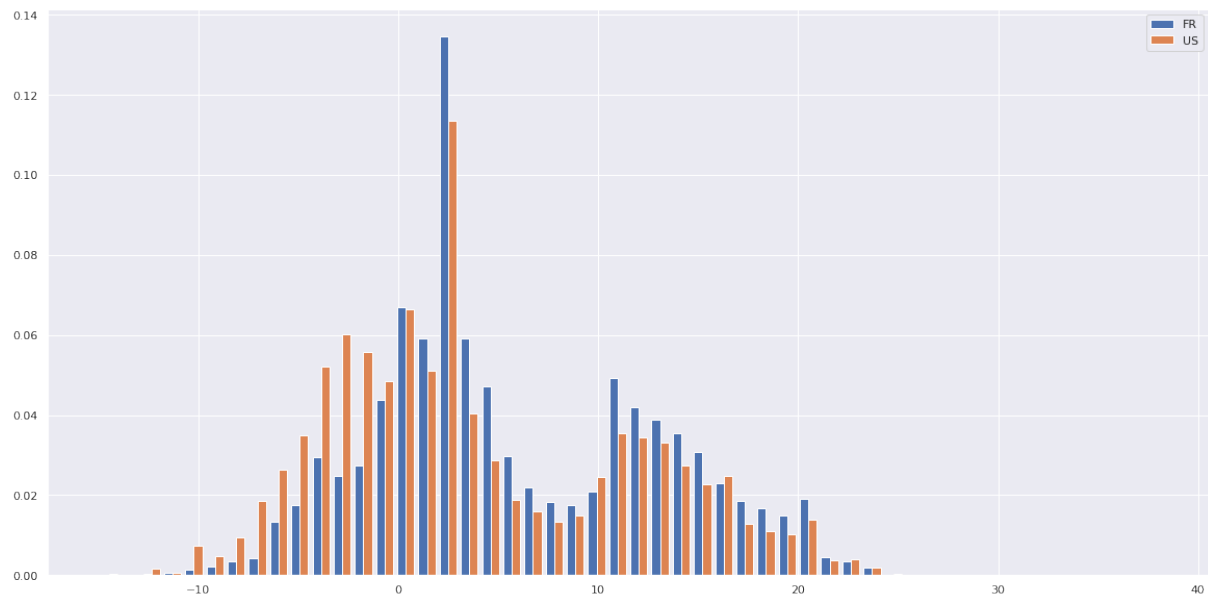


Figure 4: Nutrition score US vs France.

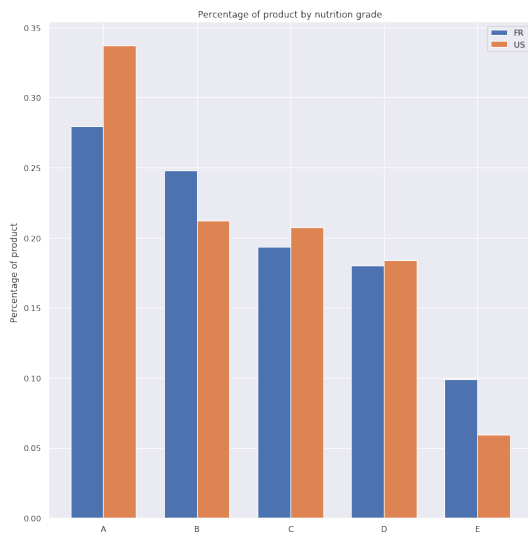


Figure 5: Nutrition grade US vs France.

or "E" grade. France tend to have more traditional food that will fall in the end of the ranking. This explains the high share of "E" and "D" product in France.

5 Study of the ecological footprint of the different major brands

United States

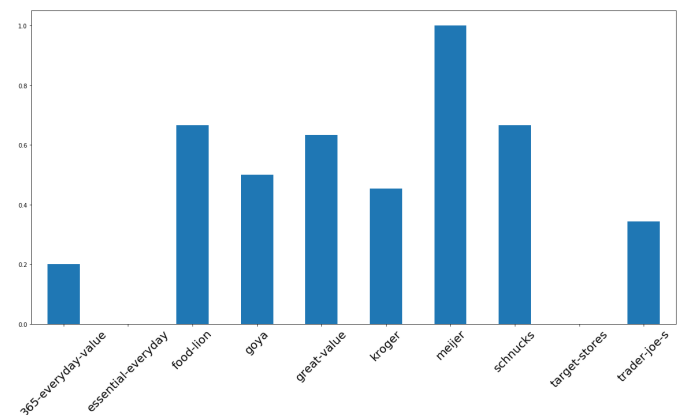


Figure 6: Percentage of plastics in packaging for the top 10 US brands

Here we can see that despite the global effort that seems to start with the environmental issues we know, most of the mass market brands produces packagings full of plastics (Figure 6). Surprisingly, we can note that none of the users that registered food from essential-everyday indicated the packaging type of the product, thus giving this result. Same for target-stores. It is interesting to see these results as an indicator of the will to change the odds concerning environmental issues. However these data consistency errors show that the dataset is not 100% reliable, mostly as it relies

on users good practices.

France

We can see an interesting result, giving the percentage of plastics in the different brands' packaging (Figure 7). This top 20 shows a full range of brands, from very ecological to not at all, in terms of packaging of course. If we compare the two top ten, it is difficult to find a major difference between the Us and France.

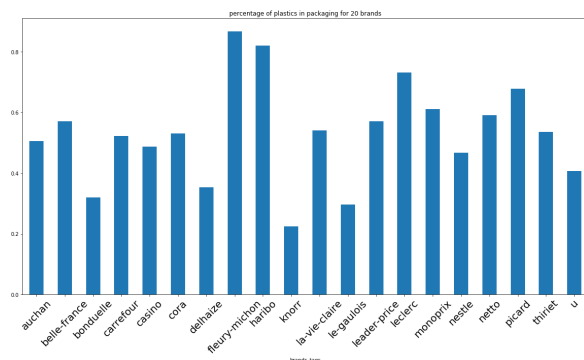


Figure 7: Percentage of plastics in packaging for the top 20 french brands.

6 Conclusion

To conclude this study we can say several things. First it showed us how an open source and collaborative dataset can be messy and thus complex to work with. The way it is filled is prone to errors and misunderstandings from the user. Secondly, we speculate that the data it contains is entered by a certain category of people that are more concerned with their health than the average consumer. This could introduce a bias in the type of products that are represented in the dataset.

With that said, we found interesting results in terms of nutrition and nutrients and also in terms of ecology. The choice to keep only France and the United States for our comparisons is mostly due to representative issues, but we found other results with Switzerland for example, in our notebooks.

We know that the United States has a worse diet than Europe, this can easily be deduced by looking at the obesity rates. This was however not reflected in the availability of products sold in the US versus France. From this we can conclude that the main factor of a bad diet in a country as a whole, is not the range food products sold. We hypothesize that the main driver would be the consumption habits of the different populations. This would have to be confirmed in further behavioral studies.

We would like to thank the staff for the organization of this project which was very instructive. It allowed us to apply what we learned during the semester, which is not only how to extract and visually present interesting statistics from raw data, but also to think like a data scientist and always put your results into question.

References

- Cristin E. Kearns, Laura A. Schmidt and Stanton A. Glantz. 2016. *Sugar Industry and Coronary Heart Disease Research* Philip R. Lee Institute for Health Policy Studies, San Francisco, California.
- French high council of public health. 2015. *notice relating to information on the nutritional quality of food products*. French high council of public health, Paris, France.