INSIGHT INFUSION

PROCESS BOOK

COM480 - Data Visualization

June 2nd 2023

Presented by

Rein Gundersen Bentdal

Laurent Faucher

Jean Siffert

PATH

Path

We knew we wanted to tackle a subject that reaches a wide audience on a regular basis, so we decided to analyze food products sold in supermarkets. We already heard of the Yuka [1] application, which lets you analyze food products by scanning them. We chose to use the same database: OpenFoodFacts [2]. After a few difficulties in preparing the data (explained later in the challenges), we began discussing the angles from which we wanted to approach this subject.

From the data, we identified three main areas: Environment, Health and Nutrition. We then started to think about interesting visualizations to make, and began to draw some simple diagrams. The nutrition axis is interesting, but very broad, and we decided to restrict ourselves to an analysis of the environment and health.

Now that we had two main directions and lots of ideas for visualization, it was time to start coding! We have created several simple visualizations with D3.js, inspired mainly on the course notes and examples in the D3.js documentation.

The dataset we use lacks precise data on CO_2 emissions related to product manufacturing and distribution. We therefore decided to add data by exploiting the Agribalyse database set up by ADEME (French Environment and Energy Management Agency) [3], which provides a wealth of information on a product's CO_2 emissions by step (agriculture, processing, transport, distribution, etc.). At this point we decided to split the team: Laurent and Rein are working on the environmental aspect with the 2 datasets and Jean is working on the health aspect and more specifically on the impact of additives.

So we started developing several visualizations on our respective subjects. We regularly shared our progress, both in terms of visualizations and data extraction, to facilitate our progress. For milestone 2, we pooled our creations, but the result was not uniform. So, after milestone 2, we started thinking about the overall structure of the project and how to integrate our different parts in a fluid way. We also defined the site's graphic identity.

We continued to create and improve our visualizations until the final result was achieved. We also worked on the quality of our code.

Finally, we added text elements to provide context and explain our visualizations. The aim is to guide the user through our project so that he or she can be clearly informed.

INSIGHT INFUSION

CHALLENGES

Data cleaning and preprocessing

The first challenge we encountered was the cleaning and preprocessing of the dataset. Indeed, we used data from OpenFoodFacts which is a crowdsourced database about food products. This dataset contains a vast amount of information, but the quality and format of the data is highly variable. We therefore had to carry out an important pre-processing step to ensure a consistent format. This included translating country names into English and linking more precise locations to the corresponding country (e.g. “Lausanne” to “Switzerland” or “Canton de Vaud” to “Switzerland”). We have also created new fields from unformatted data, such as those indicating whether the product is vegetarian, contains palm oil or is halal, for example.

Learning D3.js

Once we had a dataset we were ready to work with, we started exploring the different visualizations we could do and how to build them with D3.js.

We all had a little knowledge of HTML, CSS and JavaScript, but learning how D3.js worked was a technical challenge for us. Indeed, beyond understanding how the various D3 components (axes, scales, hierarchies, etc.) work individually, the real difficulty lay in assembling these components to create meaningful visualizations.

We were very impressed by some of the visualizations we have in courses or on the internet, but reproducing similar results adapted to our problem wasn't always easy.

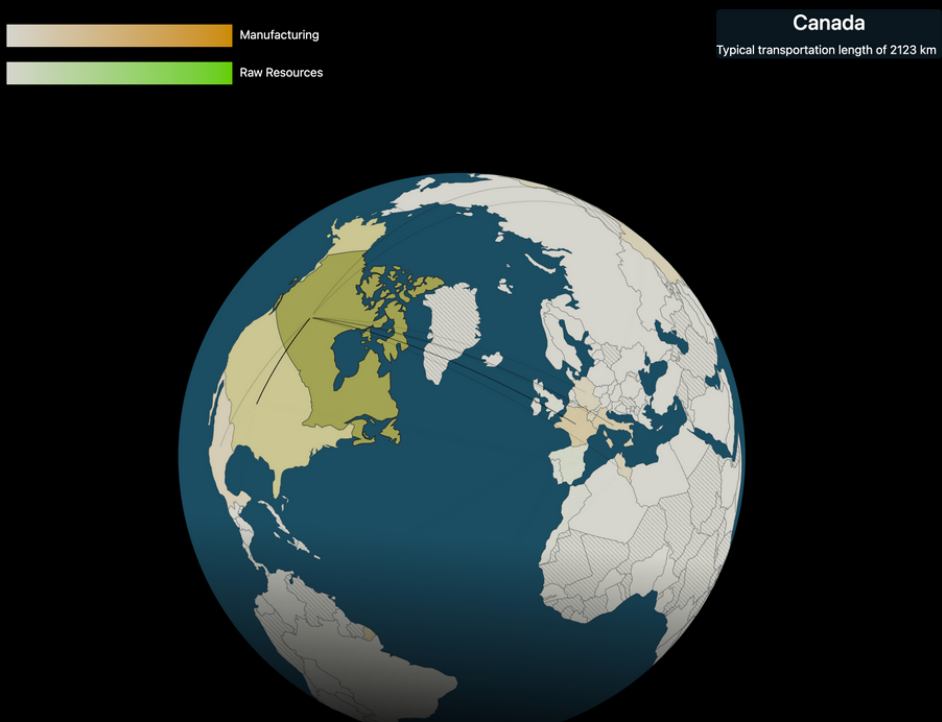
Content selection

Our dataset contains a wide variety of information, and one of our challenges was to choose which part to focus on, to the detriment of others. For example, we decided not to tackle the nutritional aspect of products (salt, sugar, fat, etc.) or the societal aspect (vegetarian, vegan, halal, kosher, etc.).

For our world map we focused only on the fields which contained information about the products raw ingredient origin countries as well as where the product are manufactured. We did this to maximize the number of useable products.

INSIGHT INFUSION

SKETCHES



Transportation of products

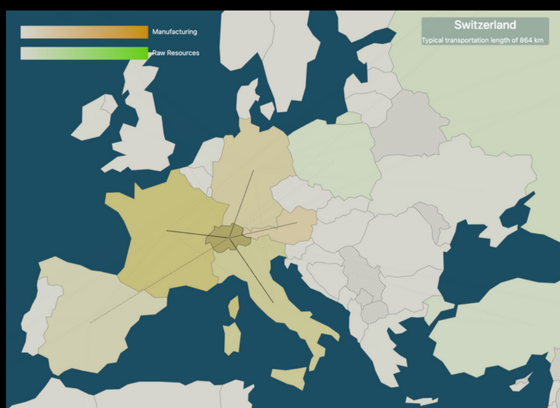
Our interactive world map serves as a visual narrative, illustrating the voyage of food from its source to local grocery stores. Through a detailed and intuitive visual representation, our objective is to make the abstract concept of global food transport tangible and comprehensible.

While traditional charts and graphs are valuable informational tools, they may not provoke a profound shift in consumer behaviour. Our map aims to bridge this gap. By rendering the journey of food products in an engaging, interactive format, we hope people will think more about where their consumed food originates from, and to source products more locally.

To illustrate, understanding that a food item originates from Egypt may not necessarily resonate with the consumer's perception of distance. The journey that food undertakes before reaching our plates often remains an abstract concept.

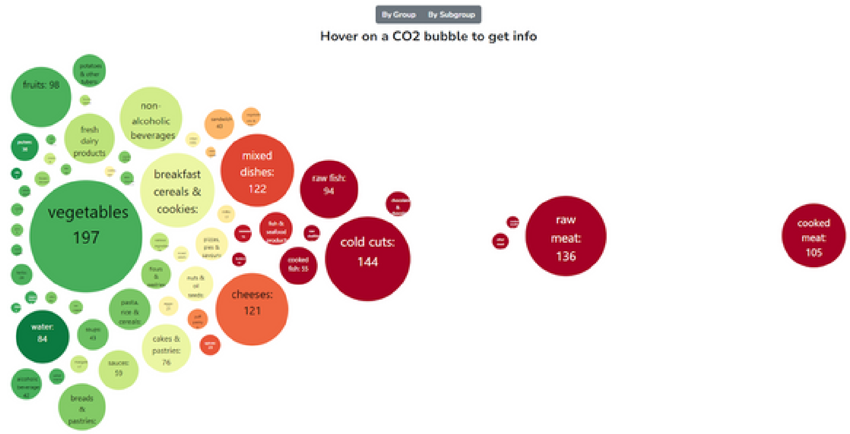
Our map is designed to render these vast distances more comprehensible, elevating the appreciation for the extensive journey many food items undertake.

Utilizing D3.js, a JavaScript library known for its powerful data visualization capabilities, the map employs a 3D projection technique to create a lifelike representation of the world [4]. This method enhances the visual authenticity of the map, rendering a more immersive user experience. Further, the map is fully interactive, allowing users to explore the globe and zoom in for more detailed insights.



Environment - CO2

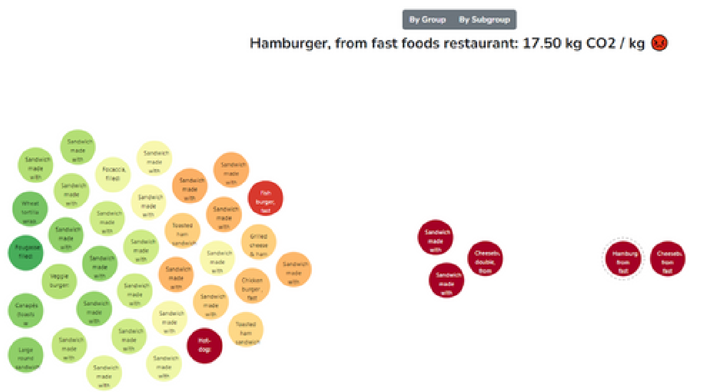
In the environmental part of our project, we were eager to exploit and represent CO2 data on different food groups and products. This first CO2 bubble chart echoes the CO2 particles floating around us. The graph ranks the sub-groups from least to most polluting in terms of CO2 emissions. The only change we made in this graph since the last milestone is adjusting the colors to make it more appealing to the eye.

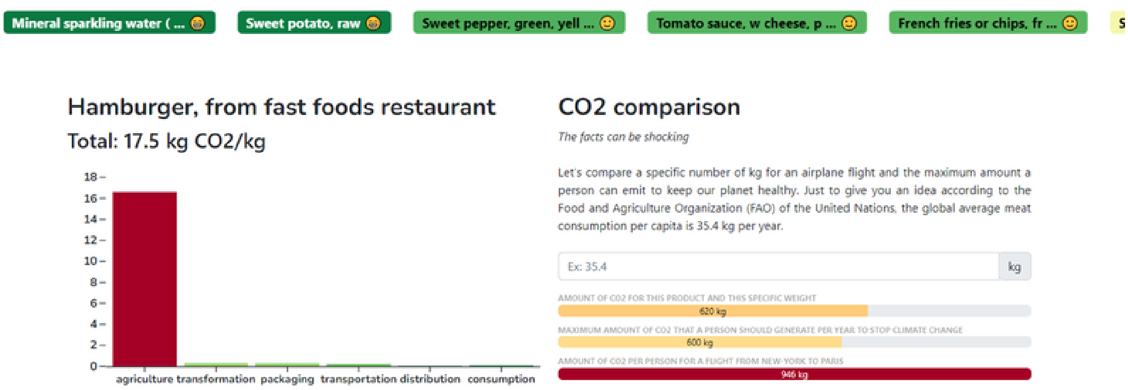


This chart is inspired by Tiffany France's chart on the Observable website [5] but has been modified and is now much more interactive. Hovering on a bubble will give you its full name and CO2 emissions. Clicking on it will give you all the products in that subgroup ranked by the same criteria.

In the previous graph, clicking on the sandwich subgroup makes the following graph appear.

This graph represents the sandwich subgroup but it could have also been any subgroup. We can see in this screenshot that we are hovering on the Hamburger bubble and that its full name and CO2 emissions are displayed above. By clicking on the hamburger bubble the graph below is generated and the web page automatically scrolls down to it.





The top bar of the previous graph is a list of random products going from left to right. We implemented this feature for us to explore products that we do not necessarily think of but that could be interesting to us.

These two bar graphs look different but are in fact intertwined. The left one is inspired from the zoomable bar chart from the official D3.js website [6]. This graph showcases the production sector that is the most harmful to the environment. In the case of the hamburger we clearly see that agriculture is the most harmful since cattle farming accounts for a large proportion of greenhouse gas emissions. The chart on the right is new since MS2. We wanted to add a comparison to shock the users. To create this chart we used bootstrap components [7]. We reused a basic input and 3 progress bars.

By default we wanted to choose a food quantity that meant something so we chose to multiply the total CO_2 emissions of the product with the average meat consumption weight per capita: 35.4 kg per year according to the Food and Agriculture Organization (FAO) of the United Nations [8]. We then compared this product to the maximum amount of CO_2 that a person should generate per year to stop climate change (600 kg) [9] and also a flight from New-York to Paris (946 kg) [10]. In this case 35.4 kg of Hamburgers produces 620 kg of CO_2. The input lets the user choose any weight of the product to compare and keeps this chart highly interactive.

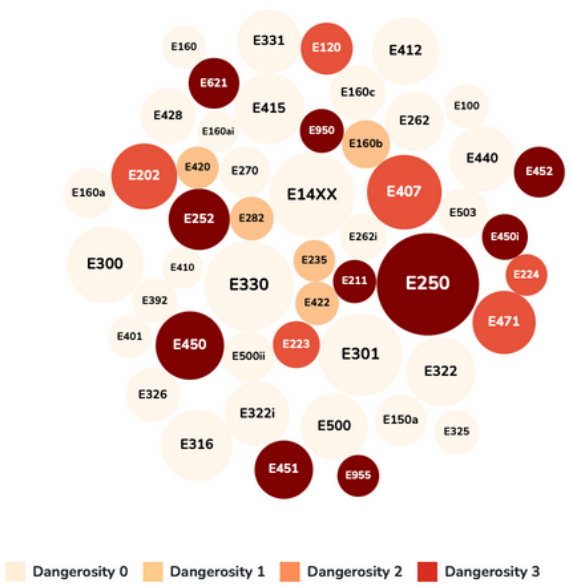
Health - Additives

Overview of additives

This packed bubble chart is a first touch with the distribution of additives in food products. It shows the 50 most present additives with a radius proportional to their use and a color corresponding to their dangerosity level (as shown in the legend).

The labels contained inside the bubble is the code of the additives and the user can hover a buble with the mouse to display the full name.

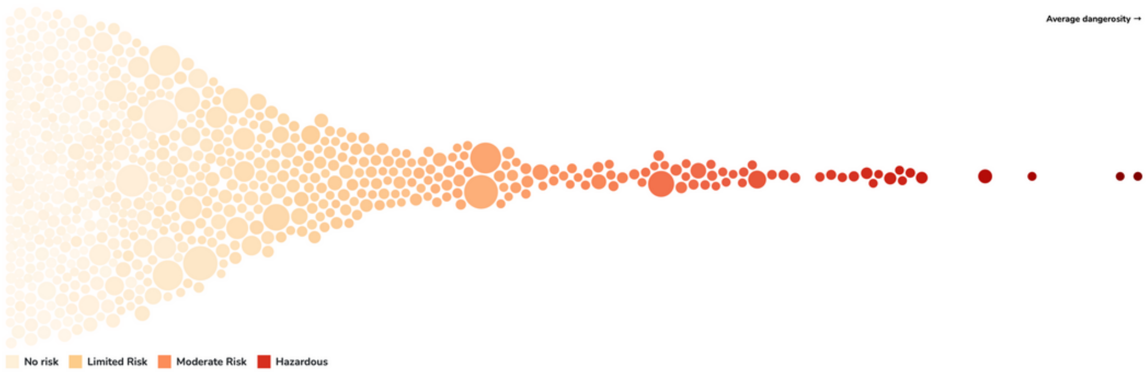
The backbone of this visualization is inspired from the example "Circle Packing" from the official D3.js documentation [11].



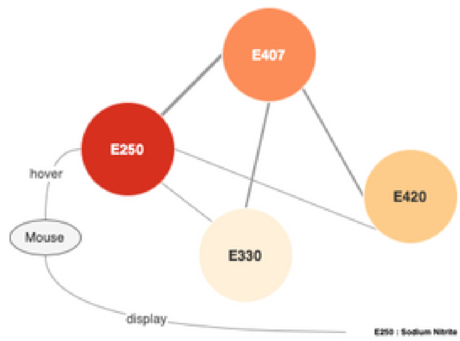
Food categories

This introductory visualization was reused to build a more advanced graph by exploiting the force layout of D3.js to align bubbles on an axis according to some data. This chart shows the distribution of number (radius) and danger level (color) of additives by food categories.

The bubbles are aligned on the X-axis according to their danger level to better appreciate the distribution in addition to the color. This idea was inspired by a visualization by Christian Burkhart [12]



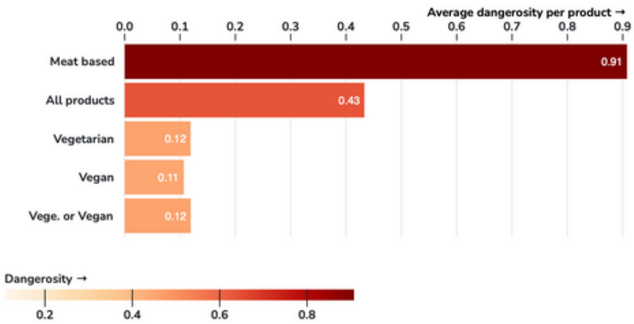
Force-directed graph

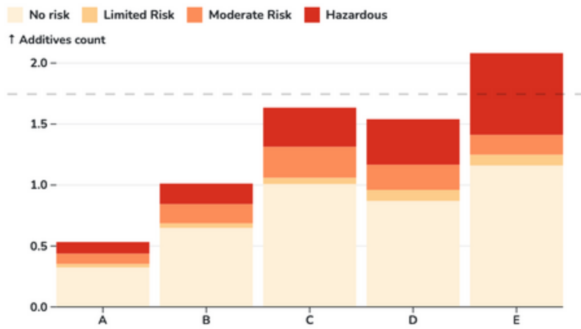


In the milestone 2, we also showed a force-directed network that we wanted to use in order to show the links between additives (additives that are the most often together in products, danger levels by category of products, ...). We have abandoned this idea because they were too many interconnections between nodes and we couldn't come up with a clear and comprehensible result that would add value.

Vegetarian & Vegan

The next parts of this section compare the additives with other variables associated with food products. We first analyzed the impact of vegan and vegetarian products on the count and danger levels of additives. To do so we used a simple bar chart that we made horizontally to fit our layout.





NOVA Index & Nutriscore

We did one chart for the count and another for the danger level. For the other comparisons variables used : Nutriscore and NOVA index, we decided to build stacked bar charts [13]. Indeed, it allows us to plot on the same chart the count (size of the total bar) and the danger level (splitted section of certain color) with respect to the analyzed variable (categorical).

CONCLUSION

Peer assessment

At the beginning of the project, we met regularly to choose the dataset and brainstorm on our ideas for problems and visualizations. We then analyzed the dataset as a team and separated the cleaning and preprocessing tasks. At the same time, we each continued to seek ideas and we regularly shared insights, images of visualizations, interesting links, etc. in a conversation created for this purpose.

Then we separated the stages more clearly: Laurent and Rein were going to work on the environmental aspect and Jean on the health part. We each coded on our own and we organized regular meetings to follow the progress and share our ideas.

Some aspects of D3.js can sometimes be hard to fully understand. So we decided to meet in order to code together to be able to help each other more easily on the technical aspects. This also allowed us to pool our parts for milestone 2. At the end of the project, we all checked our code and met to work on the final product.

Conclusion

This project enabled us to experiment with the process of visualizing insights from a dataset. We were amazed at the extent to which clear visualizations can help us better understand data compared to statistics (mean, medians, quartiles). As Ben Scheneiderman, professor at University of Maryland said, “visualization gives you answers to questions you didn’t know you had”.

We're a team of 3 exchange students currently at EPFL, and it was interesting to see our different ways of approaching the project, each bringing our own experiences to bear.

As mentioned above, we were unfortunately unable to process the entire dataset, but this leaves open the possibility of further developing the project to better understand what we consume on a daily basis.

Important note

As the dataset is built by the community, we believe it contains a bias in favor of lower-quality products (more likely to be added by the community). So some of the conclusions drawn in the section on additives are probably exaggerated and should be treated with caution.

INSIGHT INFUSION

REFERENCES

[1] Yuka (2023). Make the right choice for your health [Online]. Available : https://yuka.io/en/

[2] OpenFoodFacts (2023). OpenFoodFacts - World [Online]. Available : https://world.openfoodfacts.org/

[3] ADEME (2023). Agribalyse - Portail ADEME [Online]. Available : https://agribalyse.ademe.fr/

[4] D3 in Depth (2023). Making maps with D3 [Online]. Available : https://www.d3indepth.com/geographic/

[5] France T. (2020). Topics by Gender using the Smithsonian API (d3.force()) [Online]. Available : https://observablehq.com/@tiffylou/topics-by-gender-in-the-smithsonian-api

[6] Bostock M. (2021). Zoomable Bar Chart [Online]. Available: https://observablehq.com/@d3/zoomable-bar-chart

[7] Bootstrap. (2021). Bootstrap v5.0 components [Online]. Available: https://getbootstrap.com/docs/5.0/components/

[8] FAO. (2021). OECD-FAO Agricultural Outlook 2021-2030 [Online]. Available: https://www.fao.org/3/cb5332en/Meat.pdf

[9] MyClimate. (2023). Calculate your flight emissions! [Online]. Available: https://co2.myclimate.org/en/flight_calculators/new

[10] Kaye, N. & Hartley, Andrew & Hemming, Deborah. (2012). Mapping the climate: guidance on appropriate techniques to map climate variables and their uncertainty. Geoscientific Model Development. 5. 10.5194/gmd-5-245-2012.

[11] Bostock M. (2021). Circle Packing - Bubble Chart [Online]. Available : https://observablehq.com/@d3/bubble-chart

[12] Burkhart C. (2020). Bubble Chart - Split my Runnings by year [Online]. Available : https://observablehq.com/@ch-bu/bubble-chart-split-my-running-by-year

[13] Bostock M. (2021). Stacked bar chart [Online]. Available : https://observablehq.com/@d3/stacked-bar-chart