

# Recipes around the world: What are the major differences?

## Data Scraping

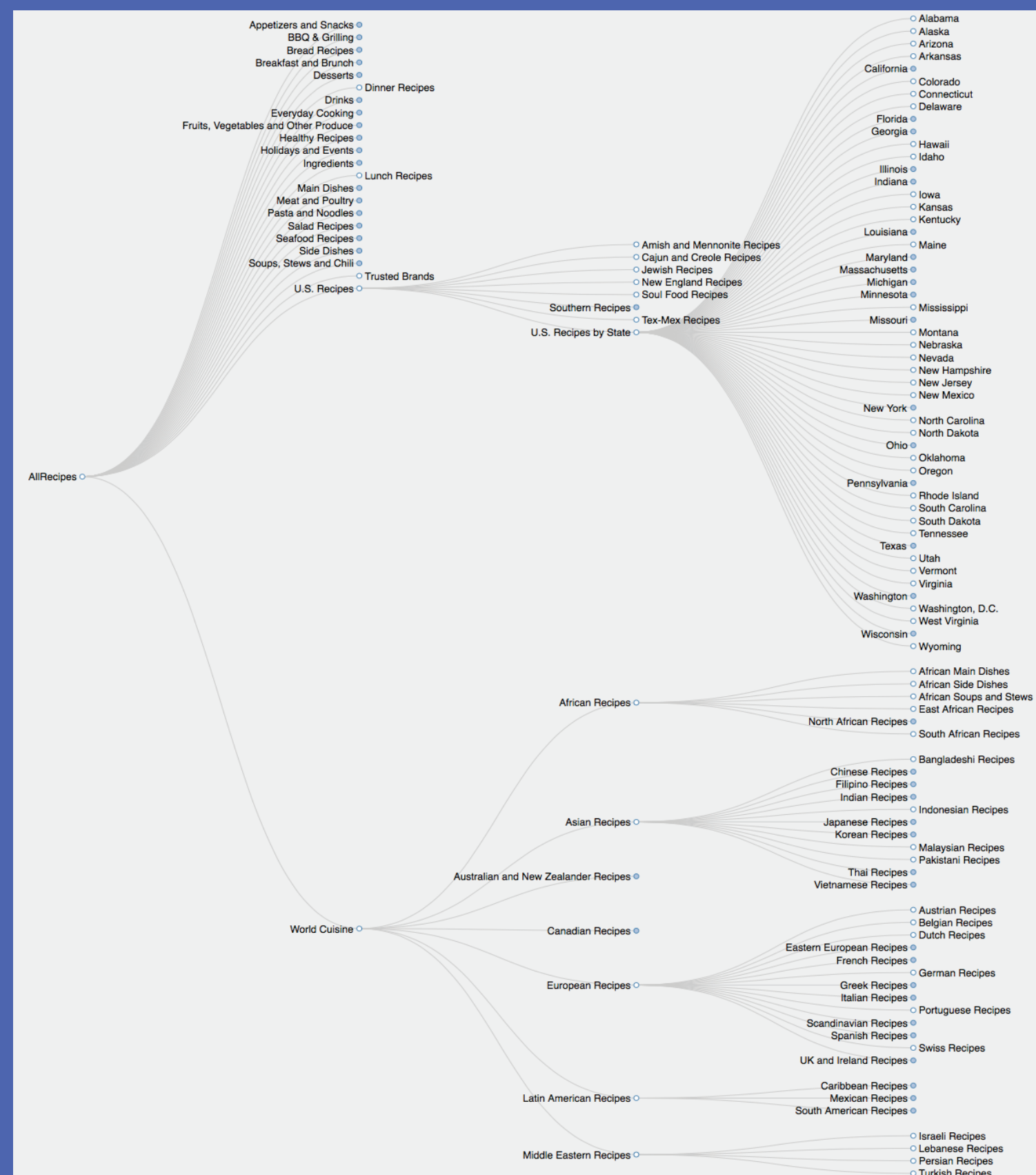
### Introduction

A good visualization of recipes requires a rich amount of information. We focused our work on the specialities of each world region in order to observe the cultural differences of the population through an analysis of what they eat!  
Let's see if geographical situation has an influence on countries specialities!

### Scraping data from *allrecipes.com*

The first challenge of our project is to get a lot of recipes including their geolocalization. The data scraping process consists in a series of complex steps that transform the raw data we got from the website *allrecipes.com* into structured information:

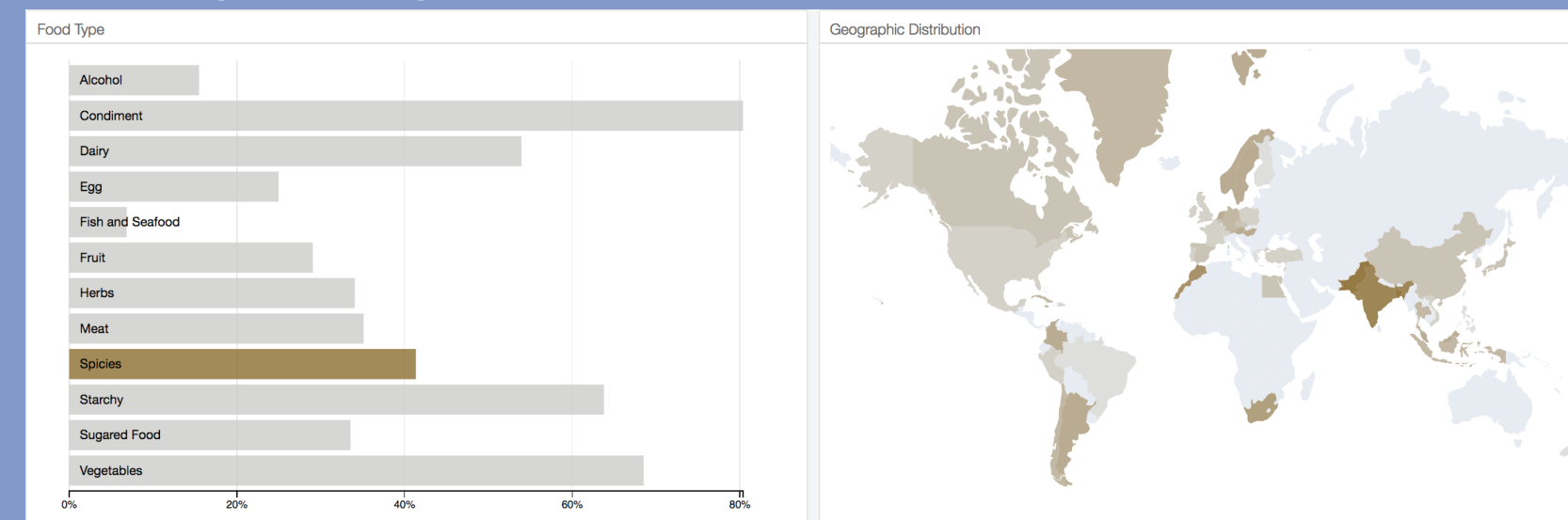
1. Get all categories of the website (2,538 categories) and select only the geographical groupings. (See the image)
2. Per category, get all linked recipes and their recipe identifiers (ID).
3. Get the HTML page of each recipe and scrap the maximum information.
4. Perform additional requests (targeting directly the hidden web service) to get the nutritional information of the recipes and the ingredient clusters.



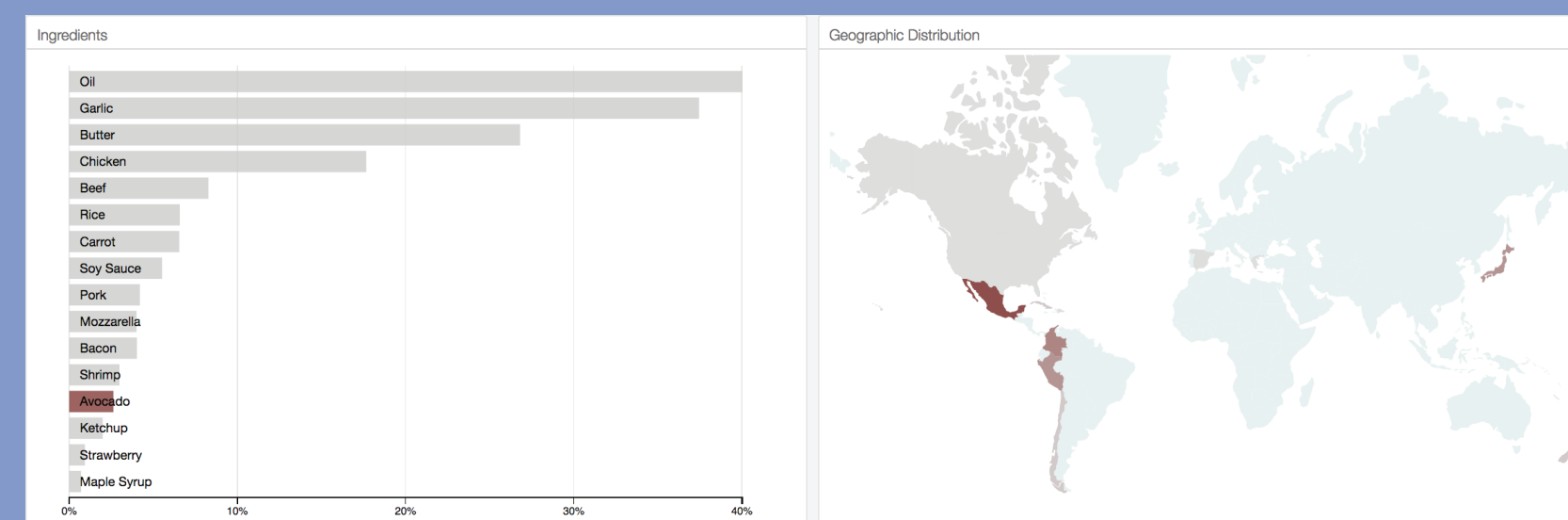
15,481 Recipes  
with country geolocalisation

## Data Analysis

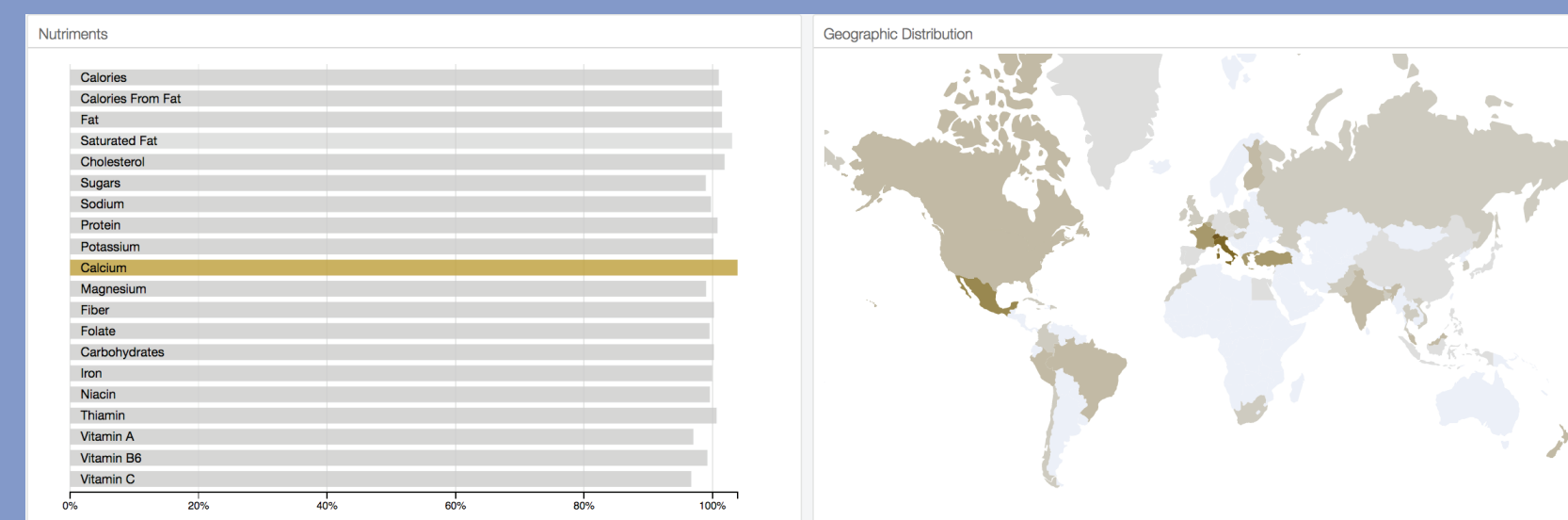
### Who puts spices in their meal?



### Which country has specialities with avocado?



### Is cheese consumption related to larger calcium intake?



### Data treated as dimensions

After obtaining all the recipes, our second main goal is to standardize the data to extract useful information. For this, countries and US states are associated to a geocode using the Google API. Then, we perform text analysis on each recipe's description to define ingredient labels and group them by food types. During these processes, different issues of missing or incorrect data have to be handled.

Finally, the cleaned information is aggregated by four categories and saved as a JSON file:

1. Geolocalization
2. Ingredient
3. Type of food
4. Nutritional intake

The web API (D3) reads the JSON file treating each such category as a dimension. A self organizing map (SOM) then leverages the visualization of each such category as a 2D image (on a world map).

## Visualization

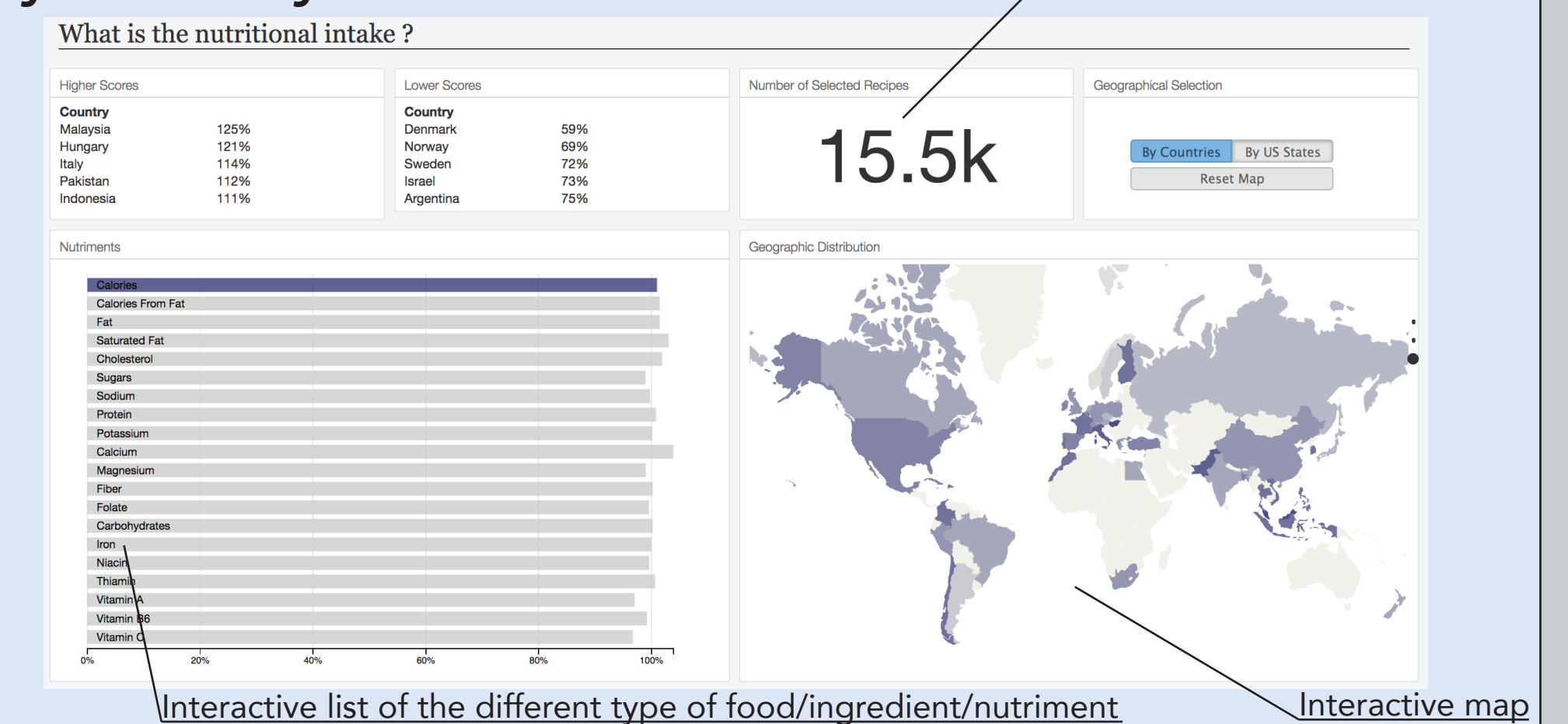
### Interactive Visualization

The last challenge to tackle in this project is to find an intuitive way to display all the information. We decided to create an interactive and easy to use visualization on a website. (See the link)

1. The visualization is fully interactive. It adapts the map and the displayed statistics according to the user choices.
2. The user has the possibility to observe the differences in specialities between the countries or the US states interactively on a map.
3. There are three different visualizations:
  - a. Type of food: Observe which region consumes more meat, fish or vegetables.
  - b. Ingredient: Observe which region eats more rice or maple syrup.
  - c. Nutrient: Compare the nutritional intake between countries or US states.

<http://cocotte-minute.ovh/ada/>

### By country



### By US states

