

# AllerVis: Geospatial Visualization of National Per Capita Consumption of Primary Food Allergens

José R. Carvajal, Alireza Iranpour, and Lucca Siaudzonis

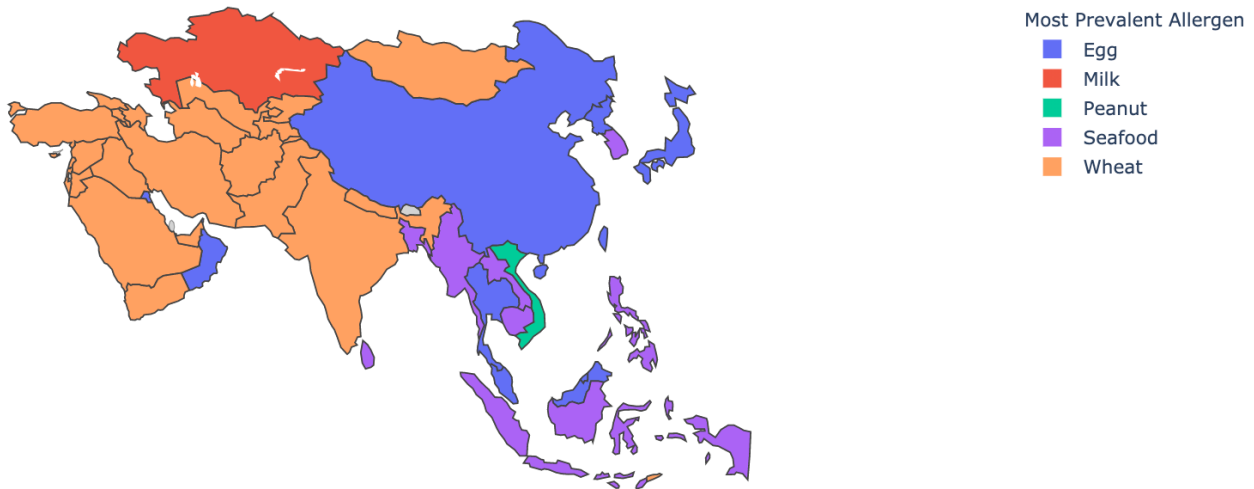


Fig. 1. Choropleth map of the most prevalent primary food allergens in Asia excluding soy.

**Abstract**—Mindful of the rise in air travel for leisure, and aware of the significant segment of the world population affected by food allergies, we have developed a tool to help users anticipate the challenges they may face travelling abroad due to their food allergies. We present: AllerVis, an interactive visualization tool developed for users who wish to be informed about the prevalence of primary food allergens of their concern at a destination of their choosing. Drawing from the available information visualization literature, and the tools at our disposal, AllerVis encodes per capita consumption of primary food allergens, for each country where data was available, using geospatial visualization idioms, and a stacked bar chart, to showcase the prevalence of select allergens in each country to inform travellers' decision making and preparation.

**Index Terms**—food consumption, allergens, choropleth, geospatial data.

## 1 INTRODUCTION

In summation and following the ‘what-how-why’ model [15], AllerVis is an interactive visualization web tool that encodes per capita consumption of primary food allergens, for each country where data was available to the authors, for exploratory purposes. AllerVis allows a user to select from a range of allergens and regions and presents the available information using both geospatial idioms and a stacked bar chart allowing the user to see prevalence of their selected allergens in each country and in relation to the region they selected. This is done to aid users in making informed decisions about where to travel given their dietary restrictions.

The purpose of this project was to create an interactive web tool that allows users to answer the question of: given a choice, where is it safest for me, my children, or a partner, to travel in light of our food allergies? Currently, and driving the motivation for this project, is the lack of an existing tool that serves a similar purpose as corroborated by both our research and the life experience of people close to us who live

with food allergies.

At this time, our tool aims to respond to the following realities.

First, according to the World Allergy Organization (WAO), an estimated 220 to 520 million people around the world live with food allergies [16]. To put this into perspective, and to stress the need for our tool, according to a study by Loh and L.K. Tang, the percentage of the world population challenged by food allergies can be as high as 10%, with a great prevalence noted amongst young children [9]. To provide context to this figure, in the United States, food allergies account for 2,000 hospitalizations and 200 deaths per year [16]. Although, at present, 170 foods have been known to cause allergic reactions due to consumption, eight main allergens: milk, eggs, fish, crustacean shellfish, tree nuts, peanuts, wheat, and soybean are responsible for the majority of adverse food reactions [5]. As such, this provides a manageable scope for our tool by delimiting the possibilities of foods that we wish to find and present the prevalence for to the main eight food allergens with a possibility of expansion in the future.

Second, our globalized world has seen a dramatic expansion in leisure air travel [2] and in the consumption of non-local goods due to greater interconnectedness of goods, cultures, and capital [6]. As such, the task at hand becomes relevant given that users, just like the present authors, may ask themselves the following questions: in mind of my severe food allergies, and with the available knowledge of the most common allergens present in each country's dietary habits, what countries are relatively safe for me to travel to? And where can I access this information for free, in an easy to consume, easily personalized,

- José R. Carvajal is with the Department of Geography at the University of British Columbia. E-mail: joc08@mail.ubc.ca.
- Alireza Iranpour is with the Department of Computer Science at the University of British Columbia. E-mail: alireza.iranpour@cs.ubc.ca.
- Lucca Siaudzonis is with the Department of Computer Science at the University of British Columbia. E-mail: lsiau@cs.ubc.ca.

and informative tool?

Third, and finally, at a global scale, while the mosaic of world cuisines is complex, at a national level it is reasonable to say that certain ingredients compromise the basis of that nation's cuisine. For instance, for historical, colonial, and environmental reasons, well beyond the scope of this project, traditional Costa Rican food consists of the following ingredients: rice, beans, coffee, potatoes, plantains, bananas, pork, maize, coconut, beef and fish [3]. As such, despite being a famous travel destination due to its natural resources, it also boasts of a national traditional cuisine that is relatively allergen friendly and bland: in terms of spice.

In light of these realities presented above, the aim of this tool is to inform end users in their decision to travel and how to prepare if they decide to travel in light of their dietary restrictions. The tool is not meant to homogenize cultures or to make destinations unfriendly. Rather, as an end user, it should help me decide what measures I should keep in mind if I decide to travel to a country with a high presence of allergens that affect me in their traditional cuisine. It is our hope that by developing this tool that we contribute to creating an environment in which information on food allergens is not unattainable or overwhelming when making travel decisions. Thus, below, following Munzner's Nested Model for Visualization Design and Validation, drawing from existing available literature, and using the tools at our disposal, we present AllerVis.

## 2 DATA ABSTRACTION

Our tool visualizes tabular data containing the following variables outlined below. Information about how this data was processed and obtained is found in Section 5.

### 2.1 Categorical Variables

- **Country:** A total of 184 countries were identified in the data gathering process. These also act as key attributes in our data table. Implicit to this variable, yet relevant to our visualization, is the association between country and fixed geo-spatial position.
- **Region:** Each country belongs to one of six different regions: Europe, Asia, Oceania, Africa, N. America and S. America. The user is able to filter by region to customize his visualization.
- **Allergen type:** Consumption data pertaining to a total of nineteen allergens was retrieved for this tool. Each of these allergen types act as a key attribute in our data table.

### 2.2 Quantitative Variables

- **Per capita consumption:** The per capita consumption of food allergen  $x$  for each country that information was available was retrieved for this project. These items are not explicitly encoded in our tool. However, it is part of our multidimensional table and was used to derive the relative prevalence of allergens within a country.
- **Allergen prevalence:** Derived sequential data representative of relative prevalence of a certain allergen[s] per country.

## 3 TASK ABSTRACTION

### 3.1 Use Case

There are several features for the tool that we want to implement. To guide our design, we made our decision around our user case:

#### 3.1.1 "I'm dangerously allergic to X, Y, and Z, and I have a choice of where to travel"

Our ideal user type is a person who is allergic to a given food, or set of foods, and has a choice of where to travel. For instance, such a person would say: "I wish to travel to Asia next summer, however, I am aware that I am allergic to egg, peanuts, and shellfish. I wonder what countries is safest for me to travel to given my allergies?" To inform their decision about destination selection, the user would use our tool and compare South East Asian countries, for instance, on their

allergen prevalence and find that Laos, for example, has a relatively low prevalence of their selected allergens.

### 3.2 Requirements

There are also a set of requirements that we considered essential in our implementation.

#### 3.2.1 Encode allergen-prevalence of a country.

Using the available data, the tool should be able to encode an allergen-prevalence score for each country derived from the *per capita* consumption data to support the user's tasks.

#### 3.2.2 Handle complexity and support search or filter options.

Due to limited space, and in mind of the size of the data set, the tool must be able to represent the tabular data in a fashion that is neither cognitively or visually overwhelming. It should also be interactive and support search or filter options that so the user is able to input their personal set of allergens and easily visualize which countries consume the most of them and have an idea of how safe it is to travel to them.

#### 3.2.3 Handle incomplete data (missing values).

When recent relevant data is missing for either an allergen type or country, the tool should still be capable of representing the available data and make clear which values are missing.

#### 3.2.4 Easily add new food attributes.

The implementation for our solution should be scalable. It should be easy for the maintainer to add support for new allergens when the data for these is provided or becomes available.

### 3.3 Tasks

Our tool, then, is designed to support the following tasks:

#### 3.3.1 Actions

- **Present:** For the authors, as designers, the tool is design to show-case food consumption data succinctly using geo-spatial idioms and a stacked bar chart.
- **Discover:** For the user, the tool is designed for the user to consume the already generated consumption data and find new knowledge that was previously unknown to them. In this case, to learn food allergen prevalence for individual countries.
- **Enjoy:** Hypothetically, this tool is also designed for a casual interaction in which the user finds the web tool yet is not allergic to anything. Features such as "All" and "World view" are designed to support this feature.
- **Browse:** The tool should also be able to support users who are aware of their allergies yet are free in terms of where to wish to travel. To this end, the tool allows the user to browse.
- **Look up:** The tool is designed for a user who knows what they are allergic to and have some sense of where they want to travel. As such, the tool is designed to support the search task of look up.
- **Identify:** The tool is designed to help a user identify allergen prevalence both quantitatively and relationally within a select region or view.
- **Compare:** The tool is designed to aid in comparison and visualize allergy prevalence for each country but also present it in a view that allows for comparison with nearby countries.
- **Explore:** The tool is also designed for users that are simply having a casual encounter with the tool and filter options that are meant designed to deal with complexity may also serve the purpose of exploration.

## 4 RELATED WORK

Presently, the authors have found no similar work or comprehensive tool that addresses the our intended tasks. Below, however, we discuss relevant related academic and non-academic work that informed our design choices or provided motivation for our tool.

### 4.1 Non-academic Work

As authors, we explored non-academic work to inform our tasks abstractions and user needs. In this process, we found that despite the existence of certain blogs or articles that show best practices on how to travel in mind of your food allergies [1, 4, 10, 13] there exists no data base or interactive tool that links together the eight most common food allergens and their prevalence or presence in traditional national cuisines or food consumption habits. As such, during this exploratory work we found both the user base for this potential tool and motivation for this work given the themes explored in these blogs and websites.

In these sites, it is clear that despite the rather unique problem this presents for people, “can I travel in mind of my allergies?”, allergens are a consideration when making decisions to travel for people with severe food or environmental allergies [10, 13]. For instance, in Scott [13], the author devotes a whole section of his website to the issue of travel, leisure and allergies. When travelling, due to his daughter’s severe allergies to shellfish, peanuts, and eggs, he makes clear that distance to hospital, local food labelling practices, and ability to fluently speak the local language inform his decisions to travel. For users like Scott, Laura, Missy and Megan [8, 9] when making travel plans for leisure, an important consideration is the predominance of allergens in the national consumption habits of their destination. As a blog said, you make the decision to travel to Italy despite being allergic to gluten [13]. As such, we find this project relevant in its ability to provide a useful service that is currently being unmet. We point to the first imperative for travelling with allergies, as outlined in all of these blogs and web articles: research where you’re going [1, 4, 10, 13] as the ideal end user for our tool.

### 4.2 Academic Work

In this section, we refer loosely to related work as academic work and literature pertaining, mostly, to geo-spatial data visualization that was surveyed by the authors in an effort to inform our visualization choices and to validate these against existing choices.

At a very macro level, Visualization Analysis and Design by Tamara Munzner [15] was instrumental in providing:

1. An overview for navigating data and task abstraction.
2. An overview of existing visualization idioms for spatial data.
3. Information on how to visualize tabular data.
4. Information on how to use the color channel appropriately.
5. Insights into how to handle complexity and visual overload associated often with geo-spatial visualizations.

Similarly, the following works, we do not discuss in detail but they were used as foundations to explore themes that we believed were important to our design.

First, Hermann et al. [14] offered useful insights, and syntheses of background work, into the appropriate use of graphic representations of data. To this end, this paper explores the advantages of stacked charts to encode, and easily compare, quantitative attributes of different items pertaining to the same categorical variable. Stacked charts, the study holds, allow for easy comparison between separate items showcasing different values for the same attributes. This aided, then, in our decision and consideration of the stacked bar chart as the idiom of choice to complement the map projections.

Second, in acknowledgement of the limitations associated with geo-spatial idioms due to clutter, Johnson and Cozart [7], and their study of interactive data visualization, were particularly helpful in elucidating the benefits of complementing geo-spatial visualizations with both

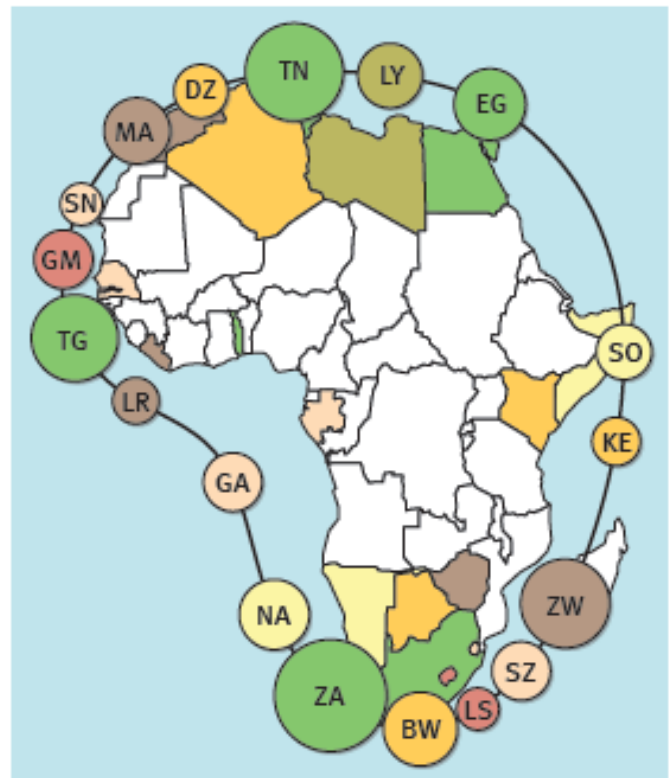


Fig. 2. Internet users in Africa in 2002 per thousand inhabitants. The image is from [17].

interactive features and a connected alternative view of the data. This way, the spatial position of the data is not lost and is presented on the map. However, this is complemented by a connected view that represents the data alternatively allowing, for instance, in our case, for the allergen prevalence data to also be easily compared and read through the accompanying stacked bar chart offering the user a more holistic representation of the data.

Third, the work of Lunterova, Spetko and Palamas [11] and the compiled work explored in Karimi and Karimi [8] was surveyed in an effort to better understand important considerations for the handling of geo-spatial data and for real life applications of these idioms. For the most part, particularly useful in these works was the illustration of common errors of cluttering and visual overload present in some of the studies and visualization choices and we kept these in mind as we built our tool in mind of these common pitfalls.

Finally, we discuss the following three works in the literature in greater detail as they pertained very closely to the goals of our design and we owe large credit of our own choices to theirs. All three of these works explore the visualization of quantitative data with spatial distribution and for categorical attributes. In other words, they present quantitative thematic maps.

Bettina Speckmann and Kevin Varneek’s exploration of necklace maps [17], provided great inspiration for our project. First, it provided an overview of the most common geospatial idioms: choropleth maps, cartograms, and proportional symbol maps, that was useful in our consideration of benefits and disadvantages of each of these when making our design choices. The authors of this work, propose a Necklace Map (see Fig. 2) that addresses the common visual cluttering of the aforementioned idioms. A necklace maps “visualizes well data sets which are not proportional to region sizes and which do not contain data for all regions of the input map.” As such, it was an idiom that we were very interested in as authors in exploring further given the nature of our data set. At this time, we were unable to retrieve the code they used to implement this solution. However, it is an idiom that could be explored in future work or iterations of this tool. Overall, we find

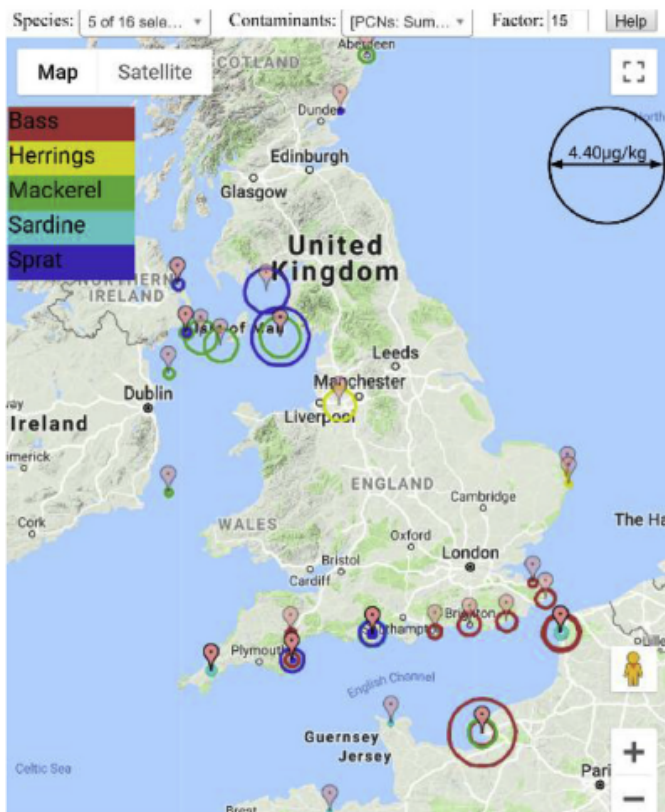


Fig. 3. Spatial distribution of sum PCN congeners for locations around the UK. The image is from [18].

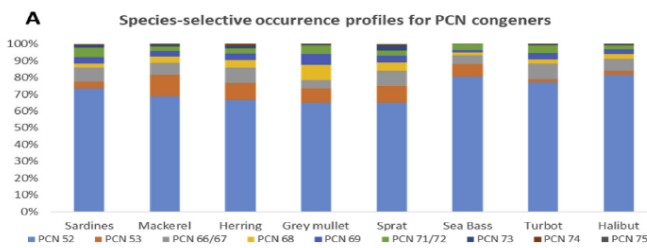


Fig. 4. Average pattern of PCN congener occurrence. The image is from [18].

their study useful to our tool to inform our decision on how to handle missing data, our decision to enable filter by region options and in our assessing of the advantages and disadvantages of different world map projections.

Zhihua et al. [18], in their geospatial visualization of food contaminant distributions informed our decision making in the following ways. Their study, similar to our tool, uses marks encoded through size and color channels, and superimposed on map projections, to visualize spatially the distribution of their data. In their case, PCN concentrations in the UK (See Fig. 3). While we vary from this study by choosing not to use the Google Maps projection, in mind that topography and city distribution were not relevant variables to our study, the choice to encode using area and color channels influenced our visualization choices for our bubble map. Similarly, their choice to use a stacked bar chart, paired with the spatial idioms, to provide a break down of the different PCN congeners per species (See Fig. 4) informed our decision to provide our own stacked bar chart that gave further information to the user of what exactly ‘allergen prevalence’ was in that country and how it compared to other countries.

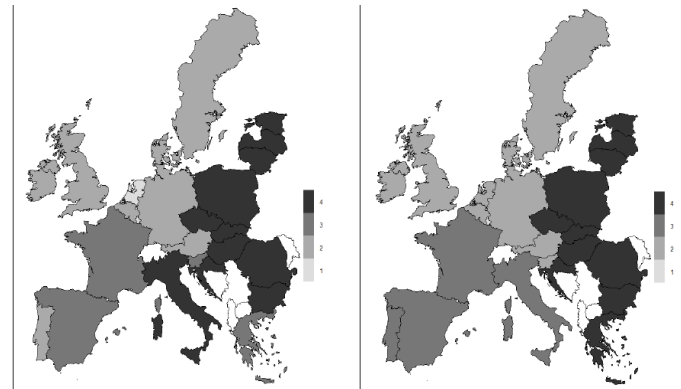


Fig. 5. Intensity of food consumption by the degree of urbanization in 2010 (left) and 2015 (right), Rural Areas. The image is from [12].

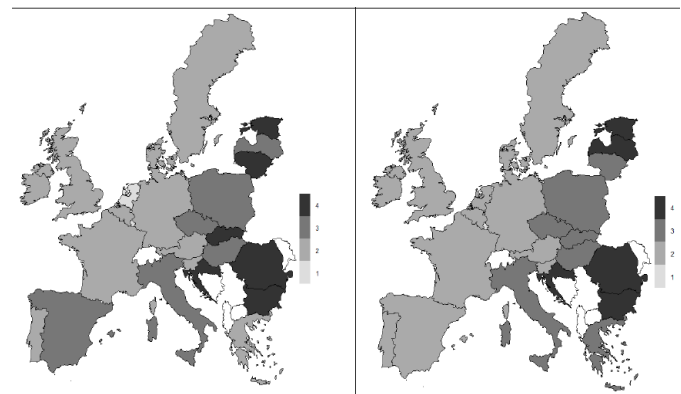


Fig. 6. Intensity of food consumption by the degree of urbanization in 2010 (left) and 2015 (right), Cities. The image is from [12].

Finally, Malička [12], in her study of degree of food consumption in relation to degree of urbanization in EU countries, uses a choropleth to visualize degree of consumption and a side by side view that allows for easy comparison between time, and across the different variables: rural, urban, and suburban (See Fig. 6 and Fig. 5). As such, from this study we drew two important considerations. First, in terms of our own choropleth, to delimit the range of saturation ‘bins’ in such a way that comparison using the color channel was still meaningful. Similarly, so that if the viewer printed the sequential data choropleth map, the differences would still be discernible in grey scale. Finally, their side by side view invited a conversation in our groups about the possibility of filtering by multiple regions, allowing for a side by side view. In the end, we believed this would take away from the advantage of the filter by region option that made smaller countries more visible but acknowledge the efficacy of the side by side view to deal with memory overload comparing between views.

## 5 DATA PROCESSING

Our main static dataset consists of a single table. It consists of:

- 184 rows and 23 columns, totalling 4,462 cells;
- each row represents a country;
- three columns contain categorical data representing country name (such as Canada), code (such as CAN), and continent (such as NAM);
- one column contains quantitative data with a country’s population (e.g. 37,058,856); and



- nineteen columns containing quantitative data representing allergen prevalence (explained below), which is a real number from 0 to 1. The nineteen allergens follow.
  - Nuts: almond, hazelnut, macadamia, cashew, peanut, pecan, pine nut, pistachio, and walnut.
  - Cereals: barley, corn, oat, rye, rice, and wheat.
  - Animal-based: beef, seafood, egg, and milk.

## 5.1 Allergen prevalence

Allergen prevalence is calculated from consumption data. We believe overall consumption of an allergen, such as milk, is a good proxy for estimating how popular it is within a country’s culinary culture. We have gathered and wrangled each country’s consumption data for each allergen in **kilograms per capita, per year**. Only the most recent year was used. From that, allergen prevalence was calculated as:

$$prevalence_{country,allergen} = \frac{consumption_{country,allergen}}{\max_{all\ countries} \{consumption_{allergen}\}}$$

When a user selects multiple allergens, their *combined prevalence* was calculated. This is calculated dynamically after a users selection. The combined prevalence of a set of allergens is simply the sum of the individual allergens’ prevalence.

## 5.2 Data gathering

The main data source we used was the UN Food and Agriculture Organization (FAO). However, that usage was indirect. We gathered the data from two sources that had it more easily accessible and organized. These were:

- *International Nut & Dried Fruit*, where we obtained data about nuts; and
- *Our World in Data*, where we obtained data about the other allergens.

## 5.3 Data Wrangling

There was significant work necessary to get the data in the correct format.

The data from *International Nut & Dried Fruit* was in a PDF. We converted that to a CSV file by writing a parser in Python and fixing mistakes in the parser manually. We used Numpy and Pandas for miscellaneous data processing.

There were three ways we which we handled missing values.

- For countries that were missing data on all allergens, we just left it as such. These countries were later shown in grey in AllerVis.
- For the nuts-related allergens, we filled the data for the missing countries with the minimum value amongst the other countries. The reasoning behind this is that the data available to us was for the top consuming countries.
- For the other allergens, we filled the fields for missing countries with the median value amongst the other countries.

## 6 SOLUTION AND RESULTS

We now take the opportunity to guide the reader through our solution choices, their justification, and provide figures of the results. In mind that many of the readers have the ability to follow the tool live, available at <http://allervis.herokuapp.com/>, we discuss our encoding choices as they are visible on the tool starting from the top left portion of the monitor, then the top right, and then the lower portion.

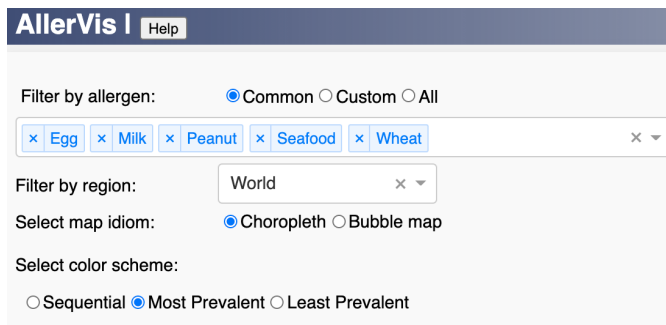


Fig. 7. AllerVis Control Panel.

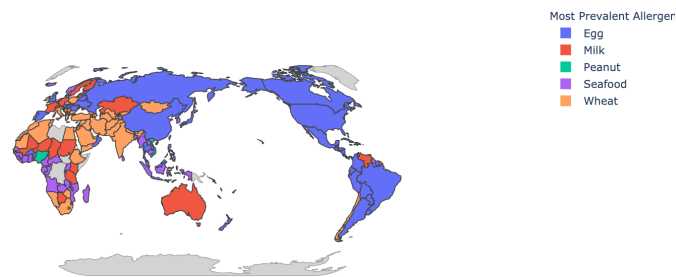


Fig. 8. Default geographical visualization depicting most prevalent allergen out of: Egg, Peanut, Seafood, Wheat, and Milk.

## 6.1 Components

Our tool is comprised of three main divisions. At the top left, we have the control panel, where the user can manipulate the data and change the views. On the right side, in the map area, we visualize the spatial aspect of our data using two different map idioms (choice of choropleth and bubble map). Finally, at the bottom, we show an overview of the entire data as a stacked bar chart.

### 6.1.1 Control panel

In the control panel, the user can: select a set of allergens, restrict the geographical scope, specify the displayed map idiom, and choose the color scheme. See Fig. 7. We have purposely selected the default view to showcase the main allergens of Egg, Peanut, Seafood, Wheat, and Milk. This translates to a world view map and bar chart that depict the most prevalent of these allergens. The world map will show a choropleth of the Mercator Projection that uses hue to encode the most prevalent allergen for each country. See Fig. 8. We believe that this default view is illustrative of what is possible with the tool and provides a useful overview for casual encounters with the tool that is neither too specific nor too general.

### 6.1.2 Updating...

While the tool updates upon filtering, in mind of possible lag, the browser Tab will show itself as "Updating..." (See Fig. 9) so the user is aware that his action is being processed. Default view.

## 6.2 Help button

However, if upon finding the tool the user is unsure about what is possible in this tool, the users can click on the help button to read



Fig. 9. AllerVis Updating...

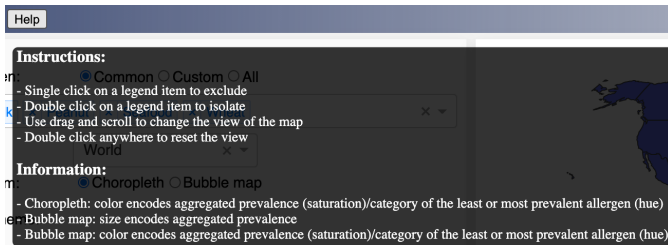


Fig. 10. Help! Instructions and key to visualization on Allervis.

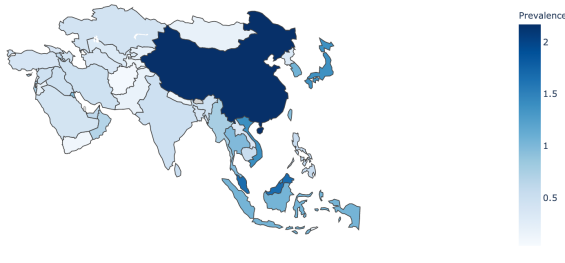


Fig. 11. Choropleth depicting sequential data for prevalence of egg, seafood and peanut in Asia.

instructions on how they can interact with the visualizations. See Fig. 10.

### 6.2.1 Choices of allergen

The default configuration has been set to consider only the most common allergens. However, the user can choose the custom option to select one or multiple allergens of their choice. Alternatively, the ‘all’ option can be selected to include all of the possible allergens, but this is only considered for exploratory purposes.

### 6.2.2 Choices of region

Initially, the region is set to show the entire world, but using the drop-down menu, it can be narrowed down to any of the six continents (Europe, Asia, Africa, North America, South America, and Oceania). This was intended to provide more specificity and to decrease information overload as users may want to only compare countries within the same region without having to look at the rest of the world. Once the user selects a region, that region is isolated on the map. The stacked bar chart is also updated accordingly to only include the constituent countries.

### 6.2.3 Color schemes

**Sequential** In the sequential color scheme, we use saturation to encode the aggregated prevalence of the selected allergens. Countries with higher values appear darker than those with lower values. For instance, if a country consumes large amounts of each allergen, it will

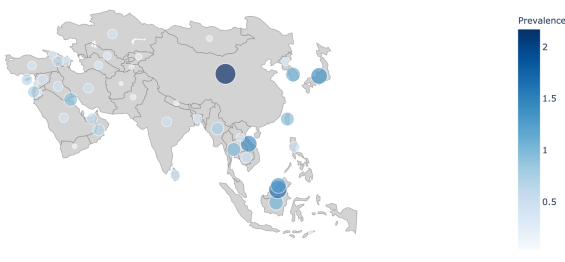


Fig. 12. Bubble map depicting sequential data for prevalence of egg, seafood, and peanut in Asia.

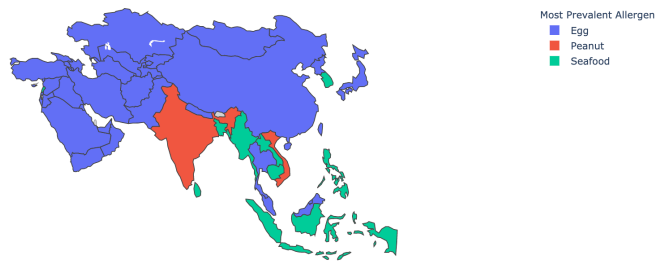


Fig. 13. Choropleth depicting data for most prevalent allergen out of egg, seafood, and peanut in Asia.

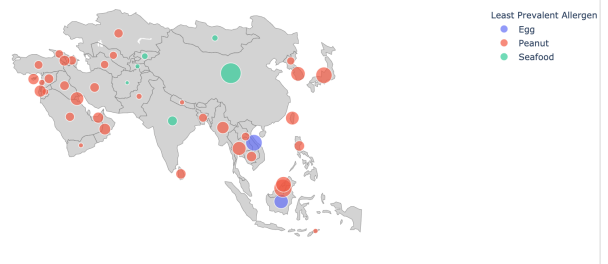


Fig. 14. Bubble map depicting data for least prevalent allergen out of egg, seafood, and peanut in Asia.

appear darkest. See Fig. 11 and Fig. 12. By using saturation, we allow for easy visual comparison of allergen prevalence and also ensure that some of the tool is still useful if it is printed in black and white. We chose the color blue in awareness of the common red/green color blindness and to ensure that our tool was interpreted as prevalence rather than a ‘danger score.’

**Most prevalent** In addition to the sequential scheme, the user can also show the most prevalent allergen for each country. In this option, the color channel (hue) is used to encode allergen type where each country’s color represents the most prevalent allergen in that country. See Fig. 13. Given that categorical nature of the variable allergen type, we believe hue is an appropriate channel to encode this. However, due to the amount of allergen types the tool supports, 19, there is a possibility of color repetition. To this end, we have ensure that the stacked bar chart and the legend for the map never repeat the same color next to each other for clarity. Similarly, we do not anticipate this to be a significant problem since most users would not be allergic to that many allergens.

**Least prevalent** Similarly, the user is capable to sorting by Least Prevalent. This option is similar to the previous one. However, this time, the color of each country indicates the least prevalent allergen in that country. See Fig. 14.

## 6.3 Geospatial Idioms

### 6.3.1 Map projection

For map projection, we chose natural earth as as this would give a more realistic representation of the areas. Moreover, the user could interactively drag the map around to feel the natural curve of the earth and its effect on the projection. This projection, however, is similar to a Mercator projection in that it gives visual advantage to Northern and larger countries. However, we anticipate the ability to filter by region to fix this issue.

## 6.4 Map Types

To illustrate the spatial aspect of our data, we have considered two idioms.

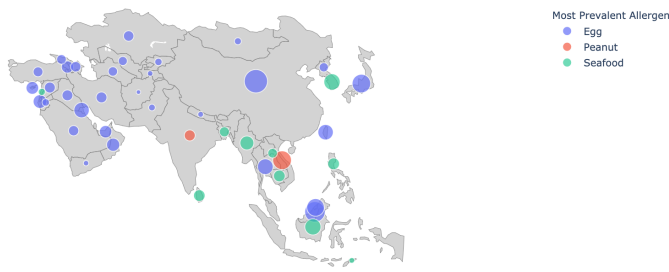


Fig. 15. Bubble map depicting data for most prevalent allergen out of egg, seafood, and peanut in Asia.

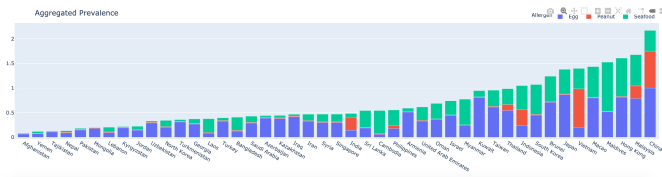


Fig. 16. Stacked bar chart depicting prevalence of egg, seafood, and peanut in each country in Asia.

**Choropleth** In the choropleth map, as seen in Fig. 11 and Fig. 13 depending on the selected color scheme, each country is represented with a level of saturation or the corresponding color of either the most or least prevalent allergen. One limitation of this idiom is that due to countries' irregular shapes and unequal areas, it may not be easy to distinguish levels of saturation. As a result, we considered an alternative idiom where we use bubble marks in lieu of countries' actual geographical area to provide a more uniform representation.

**Bubble map** In the bubble map, in addition to the color channel, we also have the size channel which can be used to encode more information. Although, in the sequential color scheme, both channels encode the same information (aggregated prevalence), in the other two scheme options while size continues to encode aggregated prevalence, the color channel (hue) is used to encode allergen type. See Fig. 15. That is larger bubbles indicate countries with higher aggregated consumption and the color of the bubble shows, for each country, either the most or the least prevalent allergen.

## 6.5 Stacked Bar Chart

To provide an overview of the aggregated prevalence of the selected set of allergens as well as their proportions for each individual country, we are using a stacked bar chart where each color represents an allergen type. See Fig. 16. The chart is always sorted from low to high prior to rendition for easier comparison. The chart has also been encoded in such a way that even if the user sorts by all allergens the colors will not repeat themselves next to each other on the chart. For clarity of view, for the default view, given the number of countries presented, it does not provide a horizontal legend for each country. However, a hover function will allow for easy identification of any selected bar chart. Finally, the bar chart supports a select function that allows the user to use their mouse to select a sub set of the bar chart and zoom into it.

## 6.6 Actions

### 6.6.1 Hovers

In each of the map idioms, hovering over a given country shows some information about that country including the country's name and the aggregated prevalence of the selected allergens for that country which is the value that determines saturation. Hovering is also supported in the stacked bar chart, in which it is possible to see the numeric value of prevalence in each allergen and country.

### 6.6.2 Clicks on map legend

**Single click** A single click on a given legend item excludes that legend from the list and hides the data associated with it from the visualization.

**Double click** A double click on a given legend item isolated that item while hiding all other items.

## 6.7 Missing Countries

Some of the countries were missing from our data set. To indicate such cases, we left those countries as a grey area on the choropleth visualizations. In the case of the bubble map idiom, and the stacked bar chart, these countries are simply missing a bubble to indicate a missing value. See Fig. 8.

## 6.8 Supported Platforms

As of now, AllerVis runs the totality of its functions in computers and in mobile form, though it was designed to be primarily used in a desktop.

## 7 IMPLEMENTATION

Our solution has been deployed as a highly interactive Heroku app. It can be publicly accessed in <http://allervis.herokuapp.com>. The repository with the implementation is <https://github.com/Alireza-Iranpour/allervis>.

AllerVis was entirely written in Python. The HTML and CSS components of the webtool were also technically written in Python, using Dash. All the plots were rendered with Plotly.

## 8 DISCUSSION AND FURTHER WORK

### 8.1 Milestones

We proposed Milestones in our project update, and now, in Table 1, see below, we compare them with what we actually experienced. We do this to provide some insight into our process, the division of tasks, and the our experience with time management.

### 8.2 Limitations and Further Work

The tool presented was evaluated not just by the present authors but was also discussed with the instructor, Tamara Munzner, and a handful of users, approximately ten, with food allergies who were kind enough to provide feedback on the tool. In recognition of the useful feedback provided by these people and our own reflections, we discuss, conclusively, the current limitations of our method and implementation to invite the opportunity for further work.

#### 8.2.1 Data gathering, processing, and management

In relation to our data gathering, processing and management, these are our current limitations and opportunities for further work.

First, is the problem of missing values and data. For the allergens that we were able to retrieve consumption data for, we acknowledge that for some countries there was no recent relevant data available. As such, further work would include the retrieval of this data for the countries with missing values.

Second, we were unable to find data for whole sections of allergens. For instance, while we found information for 'sea food consumption' this was not divided into 'shellfish' or 'fish' which are different allergens. Similarly, the common allergen of 'soy', because it is most often not consumed directly in the form of 'soy bean' but rather in other products in which it is an ingredient in, or base of, we were unable to find data for its consumption altogether that we thought was appropriate for visualization.

Third, for the purpose of this project we used consumption *per capita* data as a proxy for allergen prevalence in national food consumption habits. However, we acknowledge that these are not the same.

Fourth, we invite further work that adds geographical nuance to this tool. The available data was available for countries. However, as authors we acknowledge that national food consumption habits vary, often, within regions inside each country and the tool would benefit from access to such data to add further specificity and nuance.

Table 1. List of milestones, alongside both our estimated time and actual time needed for completion. We did not keep accurate measurements of the time to complete each task, so the numbers presented for that column are our best estimates based on hindsight, subject to rounding.

Task	People	Estimated time (hours)	Actual time (hours)	Estimated date of completion	Actual date of completion
Gather and clean all data	Lucca, Jose	30	35	November 20th	December 3rd
Related work, part 1: annotated bibliography	Jose	10	12	November 14th	November 17th
Related work, part 2: literature review	Jose	20	15	November 30th	December 8th
Explore use cases	Lucca, Jose, Alireza	15	15	November 17th	November 17th
Tool design prior to MVP	Lucca, Jose, Alireza	15	10	November 30th	November 28th
Implement MVP	Alireza	20	20	December 2nd	November 30th
Deploy interactive tool	Alireza	20	12	December 7th	December 1st
Tool design after MVP (Discussing idioms, maps, colors, etc.)	Alireza, Lucca, Jose	15	30	December 3rd	December 8th
Implement tool after MVP	Alireza, Lucca	10	30	December 7th	December 8th
Read documentation for tools	Alireza, Lucca	20	30	December 9th	December 7th
Prepare demo and presentation	Lucca, Jose, Alireza	10	14	December 10th	December 10th
Write report	Lucca, Jose, Alireza	35	30	December 14th	December 14th

## 8.2.2 Solution and implementation

Finally, in our implementation, we recognize the following limitations and opportunities for further work:

First, the tool often presents significant lag when it updates. We implemented an ‘Updating’ label so the user was aware that the action requested was being processed. However, further work should be devoted to a faster update.

Second, and this was a recurrent feedback we received, the geo-spatial projections and the bar chart are not currently interactive between themselves. For instance, while the tool supports a select function in the stacked bar chart, for instance, this does not translate also to a selection within the map. This would be a useful feature to implement.

Third, because the tool visualizes information for nineteen allergens, if a user filters by ‘All’ there will be repetition of colors that are used to encode an allergen type. However, we made sure these would never be next to each other in the legend or stacked bar chart. Furthermore, we do not anticipate a user to filter by all. The All feature is exploratory and interesting. However, given our user case, it is unlikely that someone will be allergic to enough allergens for this color repetition to be an issue.

Fourth, the map projection chosen has the advantage that it capable of zooming in and out and dragging. However, the Mercator projection, and projections similar to this one such as the one chosen for this tool, occlude smaller countries and tend to give visual prevalence to Northern and larger countries. However, we anticipate that the filter by region option will address this issue. Further projections in future work should be explored.

Finally, cartograms or similar geo-spatial visualizations that encode area were not explored. Primarily, given the inconsistency of available data for certain allergens it did not seem like a viable option. As authors, however, we recognize that given the perfect data set a cartogram would have been an appropriate idiom for visualization.

## ACKNOWLEDGMENTS

No conflict of interest were reported by the authors. The authors wish to thank Rhys Ker who provided the idea for this tool.

## REFERENCES

- [1] Asthma and A. Allergy Foundation of America. Traveling with asthma and allergies. <https://www.aafa.org/traveling-with-asthma-allergies/>, Sep 2015.
- [2] D. Banister. The tip of the iceberg: Leisure and air travel. *Built Environment (1978-)*, 26(3):226–235, 2000.
- [3] E. Foley, B. Cooke, and T. Ryckman. *Costa Rica (Cultures of the World)*. Cavendish Square Publishing, 2017.
- [4] F. Food & Allergy Canada. Living with allergies: Managing in a different environment. <https://foodallergycanada.ca/living-with-allergies/ongoing-allergy-management/managing-in-different-environments/travel/>.
- [5] F. Food Allergy Research & Education. Facts and statistics. <https://www.foodallergy.org/resources/facts-and-statistics>.
- [6] C. Hawkes. Globalization and the nutrition transition: A case study. *Food Policy for Developing Countries: Case Studies*, 2007.
- [7] M. O. Johnson, T. Cozart, and R. D. Isokpehi. Harnessing knowledge for improving access to fruits and vegetables at farmers markets: Interactive data visualization to inform food security programs and policy. *Health Promotion Practice*, 21(3):390–400, Oct. 2019. doi: 10.1177/1524839919877172
- [8] B. Karimi. *Geospatial Data Science Techniques and Applications (1st edition)*. CRC Press, Taylor Francis Group, Oct. 2017. doi: 10.1201/b22052
- [9] W. Loh and M. Tang. The epidemiology of food allergy in the global context. *International Journal of Environmental Research and Public Health*, 15(9):2043, Sep 2018. doi: 10.3390/ijerph15092043
- [10] L. Lopuch. Traveling abroad with food allergies is totally doable. <https://www.foodallergy.org/resources/facts-and-statistics>.
- [11] A. Lunterova, O. Spetko, and G. Palamas. Explorative visualization of food data to raise awareness of nutritional value. In *HCI International 2019*



– *Late Breaking Papers*, pp. 180–191. Springer International Publishing, 2019. doi: 10.1007/978-3-030-30033-3\_14

- [12] L. Maličká. Consumption of food in the EU by the degree of urbanization: data visualization and cluster analysis of the EU sample. *Potravinarstvo Slovak Journal of Food Sciences*, 14:343–350, June 2020. doi: 10.5219/1282
- [13] S. McKenzie. Scott’s rules for traveling with food allergies. <https://www.allergicliving.com/2010/06/30/allergies-travel-rules-for-food/>, Aug 2020.
- [14] T. O. Melanie R. Herrmann, Duncan P. Brumby and X. M. P. Gilbert. Does data visualization affect users’ understanding of electricity consumption? *Building Research Information*, 46:238–250, Aug. 2017. doi: 10.1080/09613218.2017.1356164
- [15] T. Munzner. *Visualization Analysis and Design*. AK Peters Visualization Series. CRC Press, Dec. 2014.
- [16] R. Pawankar, G. W. Canonica, S. T. Holgate, and R. F. Lockey. *World Allergy Organization (WAO) white book on allergy*. WAO, 2011.
- [17] B. Speckmann and K. Verbeek. Necklace maps. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):881–889, 2010. doi: 10.1109/TVCG.2010.180
- [18] L. Zhihua, Y. Gong, M. Holmes, X. Pan, Y. Xu, X. Zou, and A. R. Fernandes. Geospatial visualisation of food contaminant distributions: Polychlorinated naphthalenes (PCNs), potentially toxic elements (PTEs) and aflatoxins. *Chemosphere*, 230:559–566, Sept. 2019. doi: 10.1016/j.chemosphere.2019.05.080