Visualizing the Nutritional Landscape of Food: An NIH Codeathon Project

Authors
Tarek Antar
antarte@nih.gov
National Institute on Aging, 9000 Rockville Pike, Bethesda, MD 20892

Lauren Chan
chanl@oregonstate.edu
Oregon State University, 2520 SW Campus Way, Corvallis, OR 97331

Eva Jason
jasoner@nih.gov
Bioinformatics and Scientific Programming Core, NICHD, IRP, Bethesda, MD 20892

Dustin Moraczewski
dmoracze@nih.gov
Data Science and Sharing Team, NIMH, IRP, Bethesda, MD 20892

Kyle Pu
kyle.pu@ucla.edu
University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095

Eric Ruan
erruan@cs.stonybrook.edu
Stony Brook University, 100 Nicolls Rd, Stony Brook, NY 11794

Adam Thomas
adamt@nih.gov
Data Science and Sharing Team, NIMH, IRP, Bethesda, MD 20892

Abstract

Making nutrition and dietary choices are among the most critical aspects of daily life, and as such are subject to conflicting influences from research and media. Past research projects have focused on making recommendations for foods based on nutrient content, but the ability to apply these findings are limited by the lack of tools to visualize and interpret this information for the general public. For this reason, our group worked to tackle this issue at the Summer Intern Program Codeathon hosted by the National Institutes of Health (NIH).

Our team developed an interactive dashboard using nutrition and Dietary Reference Intake information that was obtained from the United States Department of Agriculture. Using Streamlit, we produced a nutrition data visualization tool including three unique views. The Individual Foods view offers users the ability to see the nutrition content for a single food item, including macronutrients and 20 different micronutrients. The Comparison view allows for the selection of various foods and reports nutrient content of each item with color-coded charts for direct comparison. The Diet with Recommendations view allows users to select multiple foods common to their own diet and pick micronutrients of interest to create a personalized visualization of the data. Additionally, in this view users will receive three recommendations for foods that will improve the nutrient content of their diet.

This initial model for interactive nutrient data visualization offers customizable views and individualized dietary recommendations. We hope to further refine our dashboard in the future to improve functionality and customization options including to 1) increase the number of foods and nutrients our dashboard can reference, 2) automate updating of data from USDA FDC, 3) include macronutrients within the Diet with Recommendations view, 4) include all age bracket options for dietary reference intakes, 5) include activity level and dietary limitation specific customizations 6) allow for custom serving sizes, and 7) pursue user testing and elicit feedback.

Keywords

Nutri-informatics, nutrient content, dietary intake, data visualization, nutrition

Introduction

Dietary choices are highly individual and often fluctuate on a daily basis. Variables like access to food, work schedules, taste preferences, and culture can all weigh heavily on individual food choices. Moreover, the decision of what foods to eat is further impacted by a person's knowledge and understanding of diet and nutrition. In the current age of the Internet and social media, identifying factual and unbiased information regarding nutrition has become more challenging for individuals[1,2].

Nutrient content is a quantifiable metric that can be used as an unbiased approach to nutrition. The UDSA provides recommendations on two main subsets of nutrients. First, there is macronutrient content which refers to carbohydrates, protein, and fat, which provide the calories (energy) required for all bodily functions. Second, there are micronutrients which encompass vitamins and minerals. These nutrients are also necessary for all of the functions of living, but are needed in smaller quantities. Information about nutrient content can be found in places like the Nutrition Facts panel on food packaging or in databases published online. The combination of this quantifiable information regarding nutrient content in food and the variables impacting food selection like taste, accessibility, and cultural influences[3] all contribute to a person's dietary patterns. While it is largely difficult for scientific methods to account for all of the individual variations of diet preferences and access, information regarding nutritional content of food can be leveraged to provide the general population data-driven interpretations of the nutritional value of their diet and recommendations for improvements.

Databases from the United States Department of Agriculture (USDA) such as Food Data Central (FDC) (https://fdc.nal.usda.gov/) are robust resources for nutrition information in foods. The FDC is a data system that offers nutrient profile information with references to experimental research and agricultural data[4]. In 2015 it was leveraged as a data source for a nutrition informatics study by Kim et al[5].

This study, entitled *Uncovering the Nutritional Landscape of Food,* focused on evaluating nutrition content from the USDA FDC for thousands of raw or minimally processed foods. Within this paper, the authors describe the development of a hierarchical Food to Food Network to coordinate foods with similar nutritional content. This clustering coordinated foods into four major categories, protein-rich, fat-rich, carbohydrate-rich, and low-calorie. The authors then focused on identifying foods that had an optimal balance of nutrients in reference to USDA Dietary Recommended Intakes (DRIs), specifically trying to find the smallest set of foods that could satisfy all of an individual's nutritional needs. The authors also describe the phenomenon of "bottleneck nutrients", which are those nutrients that are either in limited quantity or in excess within a food, leading to a lower nutrient favorability. A Nutrient to Nutrient was then developed to depict the coordination of various nutrients with one another, such as a positive correlation between vitamin A and K within plant based foods.

The data presented from Kim et al. provides substantial methods for food clustering, combined and single food dietary optimization, and food ranking based on nutrient content. However, there are also limitations largely related to the usability of the data from a layperson perspective. In particular, a substantial amount of nutritional data is used to identify optimally nutritious foods, but the distinct nutrition features of each food are not easily identifiable or described. Additionally, while this is meaningful in prioritization of highly nutritious foods, most average individuals eat more than four foods per day and also not all individuals have access or means to obtain some of the highest ranked foods on the list (e.g. cherimoya). Tools for visualization of nutrient content of a person's diet and easily interpretable recommendations are essential to allow users to interact with this data and draw inferences regarding their personal dietary intake.

In efforts to expand on the work of Kim et al., our group participated in a four day codeathon to visualize nutrient data from the USDA FDC. During our codeathon we, 1) reviewed the data and code presented by Kim et al. 2) produced a user friendly interface for depicting the individual nutritional content of foods and comparing nutrition content between foods and 3) developed a recommendation tool for our interface to highlight foods that pair well nutritionally with the food(s) selected by the user.

Methods

Data for this project came from three primary sources, USDA FDC[6], USDA Dietary Reference Intake tables[7] and the processed data available in Supplement Table 2 from the Kim et al. project[5]. The USDA FDC hosts nutritional content data on thousands of different foods. This wide array of nutritional information is publicly available for download and offers unique food identification numbers, as well as nutrient identification numbers for easy coordination across data sets. From this data set, we extracted nutritional information for 2,257 foods including details about 25 nutrients.

Our second data source was the USDA Dietary Reference Intake tables that document expert-reviewed, federally-defined nutrient intake requirements for males and females of various age brackets. We extracted data regarding Recommended Dietary Allowance and Adequate Intake references for males and females in the 19-30 year old age bracket for our project. Unfortunately, due to time constraints we were unable to utilize reference values for other age brackets.

The data files we utilized from Kim et al. included a detailed list of over 1,000 raw foods indicating their major category (protein-rich, fat-rich, carbohydrate-rich, and low-calorie) and subcategories within each major category (e.g. milk, or fish/shellfish).

We harmonized the three data sources with R to generate our primary data file, a tab-separated values file, which is available in our GitHub repository[8]. Our data harmonization can be broadly split into two steps. The first step was to match the USDA data on foods and nutrients with the clusters found by Kim et al. Second, we added dietary reference values for individual nutrients to the dataset. We removed any nutrients which did not have clear dietary reference intake values from the USDA. All food quantities were standardized to 100g servings, which were used to calculate the percent of daily intake for visualization.

All visualization is freely available on our web app, developed with Python and the Streamlit software library, which streamlines data oriented website creation. Other packages used include Plotly for interactive bar graph and pie chart diagrams, as well as Pandas for data manipulation and analysis.

The recommendation system analyzes nutritional content of user's currently selected food choices to determine which vitamins do not yet meet daily recommended values. Total percent of daily recommended value for each vitamin in the current diet is calculated and used to determine which vitamin is most lacking. We then rank and recommend foods which satisfy the nutritional requirement and continually update recommendations as users select additional foods to better balance their nutritional needs.

Implementation

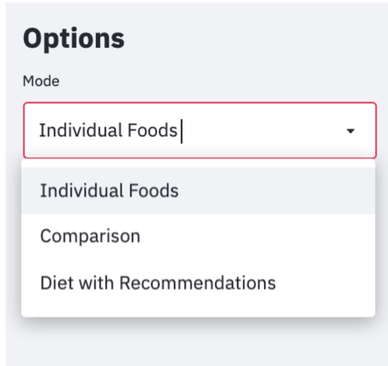
With a complete dataset describing foods and their contributions towards daily nutrient intake, we created an interactive dashboard using Streamlit[9] for visualizing the data. Streamlit is an open-source framework for creating web applications using the Python programming language. We selected this library as it streamlines the production of visually appealing web apps. In our visualization, we grouped the nutrients into three broad categories; macronutrients, vitamins, and minerals. We then created a visualization for each category focusing on displaying the degree to which foods satisfied dietary needs individually and as part of a diet. The 3 macronutrients were primarily displayed using pie charts, while the categories with more nutrients were displayed in bar graph format for improved readability.

Operation

Our dashboard app[10] is easily accessible for any user with Internet access. Switching between different views is accomplished using the sidebar. The user is given the ability to tailor the recommendations based on their sex and dietary limitations. Currently, activity level can be selected but will not have an impact on the recommendations. We plan to update this as a future direction. We leveraged the Plotly library to produce our visualizations, which allows for more options for interaction within graphs. Items on the legend can be selected or deselected without making changes in the food multiselect box, allowing for faster comparisons. It is also possible to zoom in on specific sections of the bar graph as well as hover over sections of the pie and bar charts to display more information.

Use Cases

The Kim et al. study highlighted some of the most optimally nutritious foods as ocean perch and almonds. We sought to visualize these optimally nutritious foods within our dashboard. Our dashboard offers three different views for nutrient visualization: Individual Foods, Comparison, and Diet with Recommendations.



Left sidebar on the dashboard

1) Individual Foods

If a user is interested in viewing the nutritional content for a single food item, they can select the Individual Foods option. Within this option, a food such as ocean perch can be evaluated for its carbohydrate, protein, and fiber content in comparison to USDA DRIs as well as a selection of nutrients including: Vitamins A, D, E, K, and C, B vitamins, calcium, copper, iron, magnesium, phosphorus, potassium, selenium, sodium, zinc, and more.

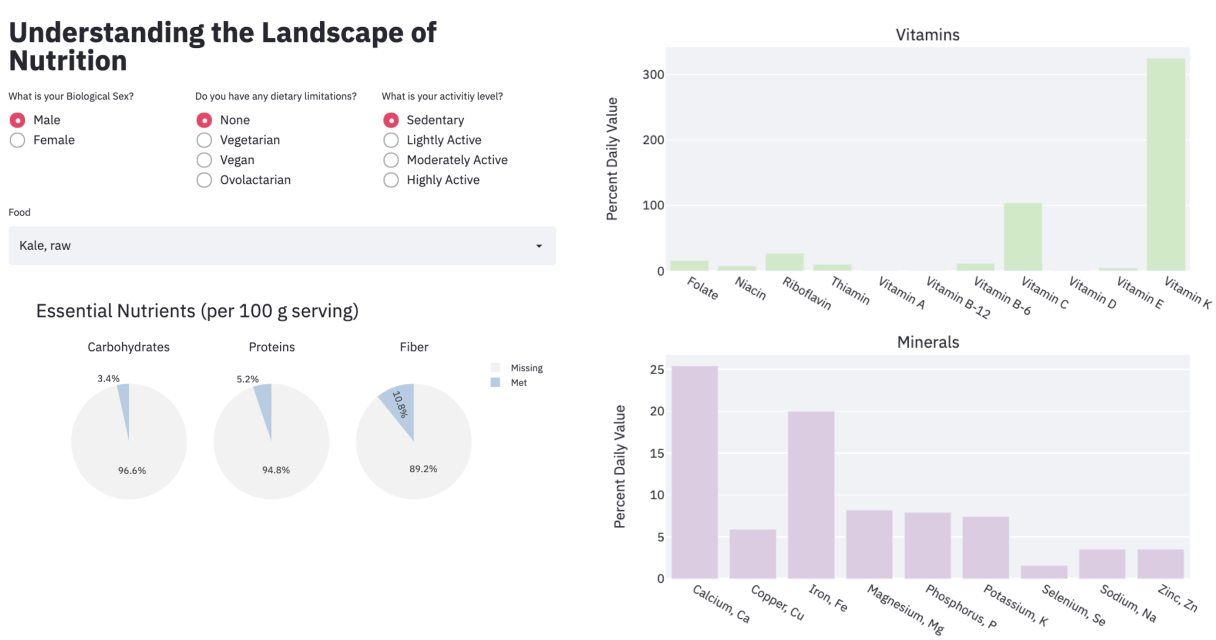


Figure 1. *An example of the Individual view of our dashboard, displaying nutrient content for kale.*

2) Comparison

Users interested in viewing a nutrient comparison between multiple foods may utilize the Comparison view to select multiple foods. All nutrient options are consistent with those available in the Individual view. Color coding enables consistent identification of how each food item contributes to the overall nutritional content.

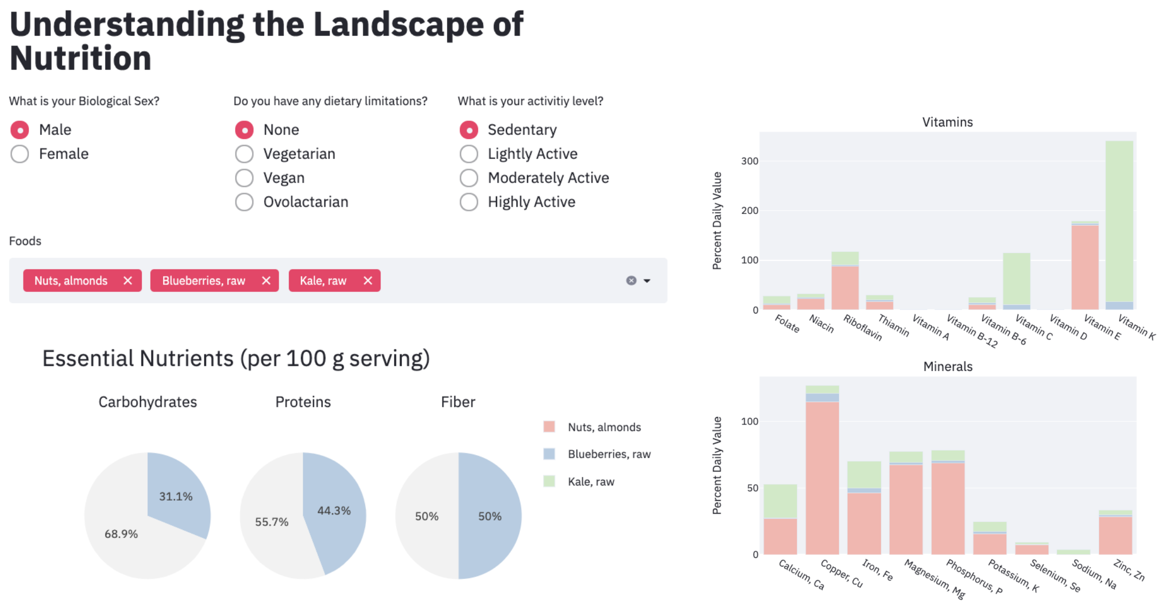


Figure 2*. An example of the Comparison view of our dashboard with nutrient content for almonds, blueberries, and kale.*

3) Diet With Recommendations

This view is for individuals interested in eliciting feedback regarding suggested foods to consume given their current dietary intake. Users can select multiple foods and their micronutrients of interest to generate a visualization of their data. Additionally, users receive three recommended foods specifically selected to improve the nutrient content of their listed diet to meet USDA DRI suggested dietary reference ranges. Dietary limitation and activity level information is not currently functional within this view, and macronutrient data is unavailable.

Figure 3*. An example of the Recommendations view showing kale and salmon being evaluated for Vitamin A, Riboflavin, Iron, and Potassium content. Recommended food products populate automatically.*

Conclusion

Throughout this four day codeathon project, our team developed a working prototype for a nutritional reference tool. The tool displays the nutritional information for a single food. Users can select several foods to compare nutrient composition. Additionally, users can select several foods and receive recommendations that will help reach their DRIs. Some challenges faced during this project include the limited capacity of our reference data sets, as they included many raw foods and not always their cooked or processed counterparts. As most individuals eat a combination of raw and cooked foods, using only raw values for scoring nutritional content limits the capacity to assess a more complete diet for adequacy. Additionally, some nutrients of interest (e.g. iodine) were not quantified within the USDA FDC dataset, limiting our ability to provide comprehensive nutrient data. As previously noted, the time constraints of this project hindered our ability to include Dietary Reference Intake values for age brackets other than

19-30 years old, and we additionally were unable to facilitate the differentiation of different activity levels to support individualization on our dashboard. Regardless of these limitations, our successful codeathon project provides a substantial foundation for this multifunctional nutrition application. We hope to continue refining our dashboard to include greater functionality and customization for users in a future codeathon. Our next steps for this project are to 1) increase the number of foods and nutrients our dashboard can reference, 2) automate updating of data from USDA FDC, 3) include macronutrients within the Diet + Recommendation view, 4) include all age bracket options for dietary reference intakes, 5) include activity level specific customizations 6) allow customization of serving sizes, and 7) pursue user testing and elicit feedback. We look forward to the opportunity to continue growing the functionality of this application.

Data and Software Availability

Code and aggregated data sets from this project are available within our GitHub repository (https://github.com/STRIDES-Codes/Uncovering-the-Nutritional-Landscape-of-Food/blob/main/README.md). This code has also been archived on Zenodo http://dx.doi.org/10.5281/zenodo.5504204

References

1. Chang K-L, Elliott LM, Sand S, Dailey R, Blachford S. *A Study of the Impacts of Social Media Outlets on Generation-X and Millennial Consumers' Beef Consumption, with an Emphasis on the Importance of Nutrition Information*. ageconsearch.umn.edu; 2014. https://ageconsearch.umn.edu/record/169648/files/Chang_Elliott_Sand_Angell_Dailey.pdf.

2. Jung EH, Walsh-Childers K, Kim H-S. Factors influencing the perceived credibility of diet-nutrition information web sites. *Comput Human Behav*. 2016;58:37-47.

3. Caswell JA, Yaktine AL, Committee on Examination of the Adequacy of Food Resources and SNAP Allotments, et al. *Individual, Household, and Environmental Factors Affecting Food Choices and Access*. National Academies Press (US); 2013.

4. McKillop K, Harnly J, Pehrsson P, Fukagawa N, Finley J. FoodData Central, USDA's Updated Approach to Food Composition Data Systems. *Current Developments in Nutrition*. 2021;5(2):596-596.

5. Kim S, Sung J, Foo M, Jin Y-S, Kim P-J. Uncovering the nutritional landscape of food. *PLoS One*. 2015;10(3):e0118697.

6. FoodData Central. https://fdc.nal.usda.gov/. Accessed August 12, 2021.

7. Dietary reference intakes (DRIs): Estimated average requirements. https://www.nal.usda.gov/sites/default/files/fnic_uploads/recommended_intakes_individuals.pdf. Accessed August 12, 2021.

8. *Uncovering-the-Nutritional-Landscape-of-Food*. Github https://github.com/STRIDES-Codes/Uncovering-the-Nutritional-Landscape-of-Food/releases/tag/v0.1. Accessed Sept 13, 2021.

9. Streamlit: The fastest way to build and share data apps. https://streamlit.io/. Accessed August 12, 2021.

10. Understanding the Landscape of Nutrition Dashboard. https://share.streamlit.io/tantar/uncovering-the-nutritional-landscape-of-food/main/UNLF.py. Accessed August 12, 2021.