# Affordability of Diets Around the World

Kristina Fujimoto*
Data Science and Biology Student

Emily Wang†
Data Science and Marketing Analytics Student

Xinyu Wu‡
Economics and Math Student

**ABSTRACT**

People around the world cannot afford healthy diets, which keep people healthy and reduces risk of diseases. These visualizations show the diet cost data and compare it to population percentages which cannot afford these diets. As food, heath, and nutrition are applicable to everyone, this visualization is meant to inform the public about diet costs and advocate for affordable healthy options worldwide. The three diets we focused on were energy sufficient diet, nutrient adequate diet, and the recommended healthy diet.

Our GitHub Repository: `https://github.com/DS4200-S22/final-project-nutrition`

## 1 INTRODUCTION

Our project focuses on the topic of the cost of nutritional meals and its connection to affordability in each country. From articles pointing out food scarcity in some countries to large percentages of populations in poverty, informing about the cost of healthy diets is important in every single way. With many people seeking cheaper food, it is important to make sure these people receive healthy foods and are able to maintain a healthy diet with their current financial state. Not everyone can afford the best products and the most varied diet with vegetables and protein. To help combat this issue, we plan to create visualization tools that allow users to clearly understand the cost of three different healthy diets and the percentage of populations which cannot afford these healthy diets. Specifically we wish to focus on observing these values for each country in the energy sufficient diet, nutrient adequate diet, and recommended healthy diet. By informing the public and raising awareness of this issue, we hope to bring attention to the need for healthy food programs and for healthy diets to be more affordable for the whole population. This will lead to less diseases caused by unbalanced diets and other problems related to the lack of healthy food.

## 2 RELATED WORK

**Automated Food Ontology Construction Mechanism for Diabetes Diet Care**

This journal ties in well with our topic of observing different food nutrition information and informing the public about healthiness. This journal focuses on diabetes and customizing a diet that is specific to each diabetic individual to help balance their diet in relation to sugar intake and insulin amounts in the body. This journal will be very useful in helping us tie together our data to be able to make connections between different topics related to food and diet. [5]

**Fighting Food Poverty**

This journal looks at how climate change is affecting food production and seeing how it raises prices of all different kinds of food, causing higher percentages of the population to be unable to afford food. This journal is particularly interesting because it shows the correlation between the two topics and is similar to how we are looking at cost of diets and percentages of the population which can afford it. This has a use case of informing the public like ours but is slightly different in its focus. The general topic is similar in some ways as food poverty is a rising issue. [2]

**Measurement of Health, Human Capital, and Economic Welfare at The Poverty Level Influences using Heuristic Network Approach**

This journal focuses on the modeling of poverty for developing countries and how many different aspects of life are affected by poverty. This journal looks to model this in an interesting way and we can use this as a guide for how we should model our data in visualizations. The analysis is also interesting as it compares health to poverty, which is what our topic does in the sense of affording a healthy diet. [7]

**Food Clustering Analysis for Personalized Food Replacement**

This journal was interesting and related to our topic in food personalization in relation to observing the health of food and wanting to increase healthiness. This journal can be helpful like the article Automated Food Ontology Construction Mechanism for Diabetes Diet Care [5] where it relates to food personalization, which is important when considering diet types. We will take this useful information as a guide in how to build a good visualization tool for in formatting the public. [6]

**DIETOS: A recommender system for adaptive diet monitoring and personalized food suggestion**

This journal is very helpful in how we can create a visualization that can personalize food based on nutritional value. This journal can help guide us in creating a visualization that is useful in informing about food and cost. Here we liked the use of a map visualization so we decided to implement one that was of the world and depicted the cost of the nutrient adequate diet for each country. [1]

**A Guide to Modern Data Visualization**

This journal will be a general useful guide to how we want to create the visualization and what elements we want to include to make it complex and in-depth in order to properly inform our target group. While this doesn't relate to our topic of nutrition vs disease, we are hoping this can guide us with the general data science coding portion of our project in order to create the visualization. This resource had interesting visualizations which inspired us to create the scatter plots that compared the data. [4]

**Using Tableau to Visualize Data and Drive Decision-Making**

This visualization will be very helpful to us in seeing how to create our visualization well. We can use this to create a draft of our visualization in Tableau before making our final version on a different platform. We can also use this as a guide in how to make a productive visualization which will achieve our final goal of informing the public and will have useful tips of what to include. We used this resource to see how we should organize our data and decided to remove some rows that had missing data based on this resource. [3]

**Toward Understanding Representation Methods in Visualization Recommendations through Scatterplot Construction Tasks**

This research article as part of the Eurographics Conference on Visualization is particularly useful for the scatter plots that we will implement as part of our webpage. This research looks at different possible visual encodings within a scatter plot, including the opacity of the dots, an interactive future allowing users to filter the data, different sizes of dots, using an outline vs filled dot, etc. The

*e-mail: fujimoto.k@northeastern.edu
†e-mail:wang.emi@northeastern.edu
‡e-mail:wu.xinyu2@northeastern.edu

researchers found that each had its pros and cons. For example, the dots that are opaque allowed users to easily identify high vs low density areas but made it difficult to see some of the dots that were outside of the higher density areas (since they were closer in color to the white background). This research is relevant to our work and will allow us to easily know which sort of scatter plot to produce that will be the most effective in allowing users to gain insights from our graphs. [9]

**Narrative Visualization: Telling Stories with Data**

This paper published as part of IEEE Transactions on Visualization and Computer Graphics details how data visualizations are able to use different methodologies to tell a wide variety of stories. For example, details-on-demand and timeline sliders are both really great interactivity features that can allow a user to develop more of a story line when looking at and using the visualization in the sense that they are able to dive deeper (either get more granular information, or be able to retrieve data from a different time period). For our visualization, this research paper is applicable because we hope to implement data-on-demand functionalities for both our world map and also scatter plots. This will allow users to have a better experience in terms of exploring the 'story' behind our data. [8]

## 3 Use Case

One important use case would be someone looking to move to another country and comparing the cost of the healthy diet in their current country and in their future country. This can help prepare them for how much they need to spend to maintain their current diet in this new country that they're moving to. It can also inform them about the situation of the population in the new country is and where the population is in a socioeconomic sense.

Another important use case for this visualization would be for people who are advocating to the UN or other world organizations about food poverty and the need for food programs in developing countries. They could use this informational tool to show that a majority of populations in some developing countries cannot afford healthy diets and advocate for the importance of eating healthy for the sake of maintaining the health of the population and reducing diseases and death due to malnutrition.

## 4 Data

**Data on Healthy Diet Costs and Affordability in Different Countries**

`https://sites.tufts.edu/foodpricesfornutrition/methods/data-code/`

This dataset contains everything we need to create our informational visualization. It contains many rows representing countries and columns which indicate the cost of a energy sufficient diet, a nutrient adequate diet, and a recommended healthy diet as well as affordability indexes which observes the percentage of the population which can afford these diets as it is greater than 65 percent of their income. The researchers collect the data through a new index that uses least-cost diets to track food system performance and monitor food prices and percentage of populations. There might be some bias towards developed countries as those have lower food prices and more data possibly coming in from sources.

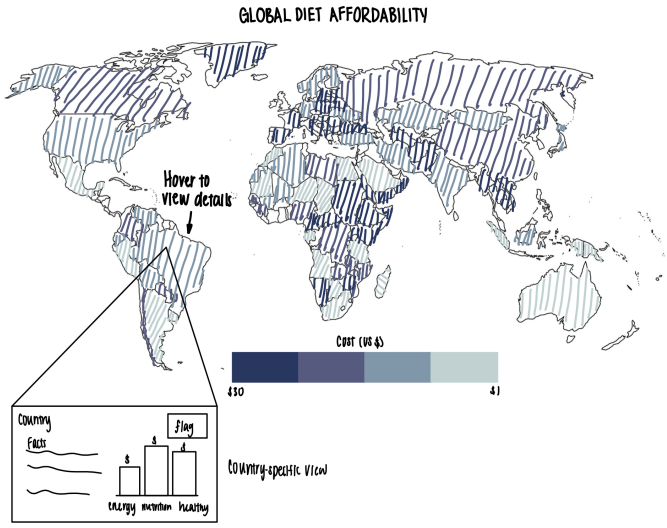
## 5 Task Analysis

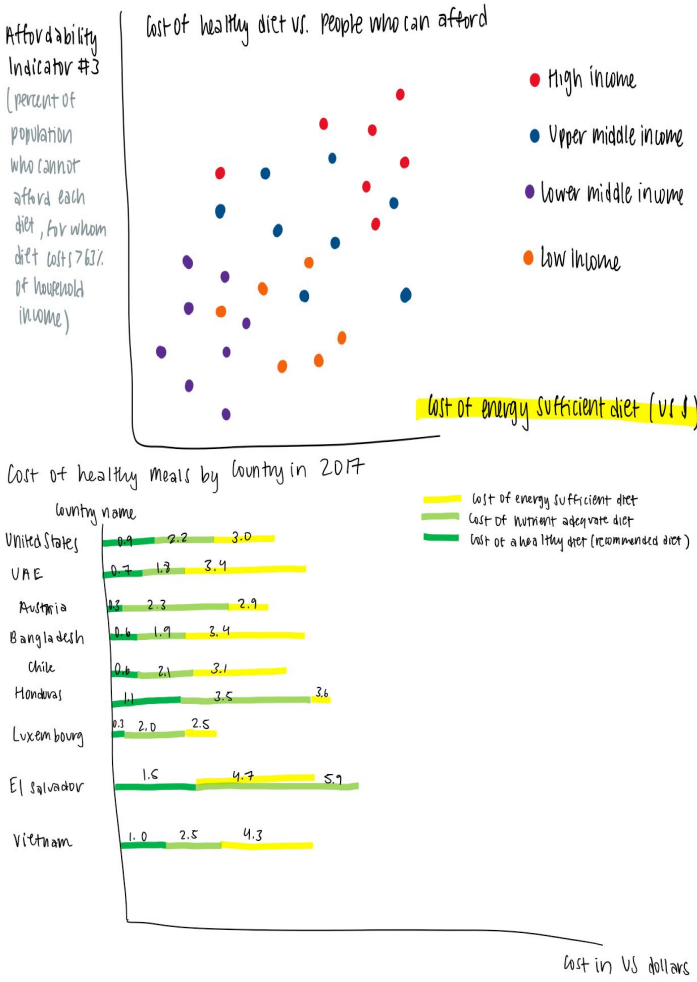
| Domain Task | Analyze Task (high level) | Search Task (mid-level) | Analytic Task (low-level) |
|---|---|---|---|
| I want to use my visualization to see which diet is most expensive overall. | Consume, Discover | Browse | identify |
| I want to use my visualization to see which countries are most expensive for food. | Produce, Derive | Browse | identify |
| I want to show that a great percentage of people can't afford healthy food. | Consume, Discover | Lookup | summarize |
| I want to see which countries are wealthier. | Produce, Derive | Locate | identify |

The primary consumer of our visualization could be a number of people. Whether that's a government official, advocate group, or an individual looking to move to a different country. The primary consumption type will be Discover, in which the end user will be able to learn more about something they might've not known before. They are able to ask questions like "which country would be the cheapest to move to" or "which country has the highest/lowest cost of a nutritional diet". It can aid in other research, and be used to guide personal, political, or other decisions.

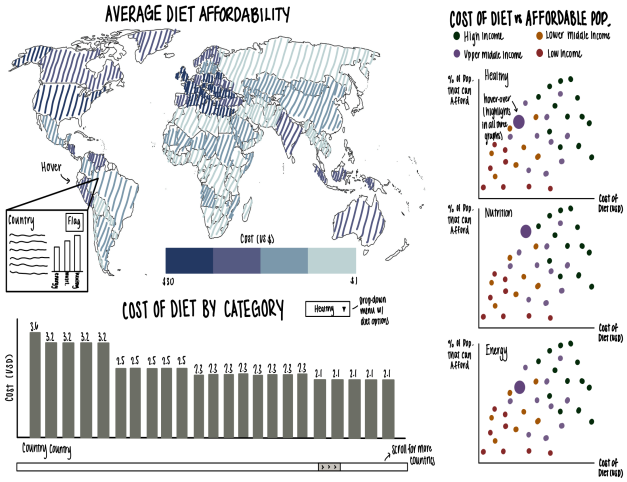
## 6 Design Process

Rough Sketches:






Digital Sketch:



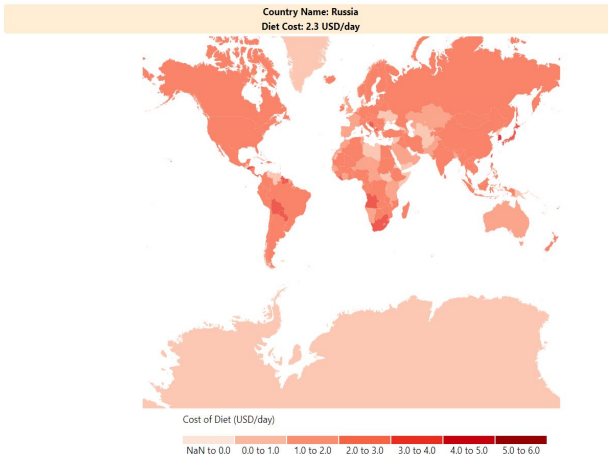

For our final project, we originally planned on have the following visual encodings: 1) a world map with the average diet affordability for each country, 2) a set of three scatter plots depicting affordability of the three different diet categories, and 3) a bar chart that represents each country's diet cost as a singular bar. For the first graph, we originally wanted to have a highlighting effect for each country in which a user hovers over a country on the map, and there will be a popup visualization for that specific country. We had planned for the pop-up visualization to include: the name of the country, the country's flag, some sort of written facts about the country, and a barchart showing the costs of the energy-sufficient, nutrient-sufficient, and healthy diets. For the final visualization however, we had made some changes to this visualization. Instead of depicting the average diet cost, we choose to use the nutrient-adequate diet as a proxy (otherwise the data would be skewed for a number of countries). Furthermore, we didn't end up getting the pop-up for each country to work. Instead, we implemented a label above the map that shows the country name and the cost of the diet. This makes sense to us because it still shows the necessary information but doesn't overwhelm the user with too much content in a small area. The second major visualization component that we implement is the three scatter plots. The scatter plots also have a brushing and linking function to connect them to one another. When a user selects one or many dots on any one of the plots, the corresponding countries' data points in the other two scatterplots will also be emphasized (the dot will become enlarged). This plot was largely implemented as we had imagined in our original designs. One difference is that the final design has the three scatterplots placed vertically versus horizontally. For the last bar plot, we will use filtering to select what data is depicted on the barchart. In our initial designs, the user would select from a drop down menu which category of diet they are looking at. Then, the corresponding costs of the diet for each country will show up on the bar chart. Instead of using a drop down menu, we decided to implement three buttons (one for each diet type). This was a change that we felt made sense because it allows users to easily see which other diet types there are rather than having to open the drop down menu. The required components for our visualization are the three primary charts — the world map, set of three scatter plots, and the bar chart. The different interactions are also important components (hovering in the map, brushing and linking between the scatter plots, the filtering of the bar chart) since they make the visualization more user-friendly and are able to help users complete the tasks that we had previously outlined. The nice-to-have components of our original designs were things like the flag of each country in the world map when a singular country is highlighted and also the additional details for each country that aren't in our current dataset. This ended up not being implemented as we ran out of time and had some difficulties with other functionalities that held us back. In order to implement this visualization, we do not need any APIs. We will use d3 combined with JavaScript, CSS, and HTML.

During our usability tests, we had users complete tasks such as identifying which country with the least affordardable for each diet on our scatterplot and finding the diet costs for the United States. After navigating our visualizations and completing these tasks, users found that the graphs were intuitive to use overall. They expressed that one major issue with our visualization was that there was too much scrolling necessary to peruse our webpage. They also identified a few other minor issues such as being confused with what the visualizations were showing at first glance due to missing graph titles, labels, and legends. To remedy these issues, we changed our implementation to include a tab for each visualization rather than having all the visualizations on one page so users are not overwhelmed with information, and added legends, graph titles, and labels for each of our graphs to allow users to better understand the context of the visualizations.

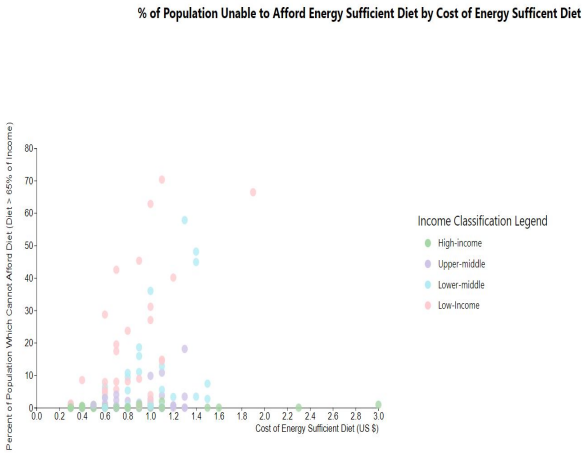
## 7 Visualization Design

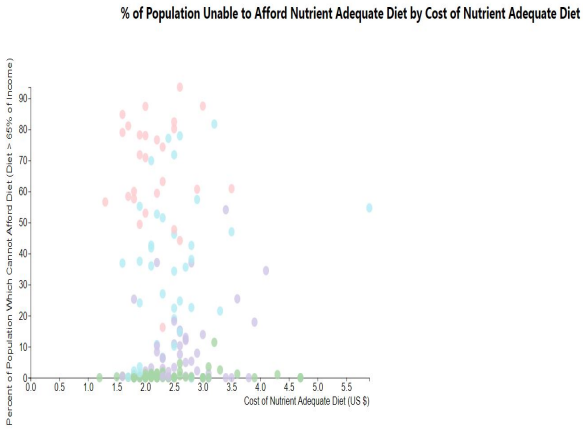
We chose to make this visualization to inform the public and advocate for more affordable healthy diets. Our visualization focuses on our goal by implementing a world map depicting the

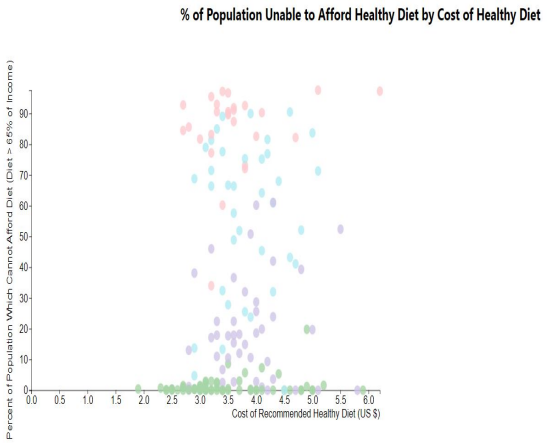
average cost of the three diets, a bar chart visualizing the cost of the three diets separately for each country, and scatter plots to compare the affordability to the cost of the three diets.




The purpose of implementing the world map is to give a broad overview of our data and provide a general understanding of the cost of healthy diets in different countries. This also gives the user more information about which country provides more affordable healthy meals compared to others. The user could also hover over each country to receive more information that lists the cost in US dollars of the nutrient adequate diet cost, which is the middle tier of healthy diet, as well as the name of the country. We will implement this hover feature in order to give the user more detail about these costs and where the average was calculated from. We are using a scale of reds to color each country that indicate the cost of the nutrient adequate diet which becomes darker as the cost is higher, and becomes lighter as the cost is lower. This map helps users see more clearly the data of cost of a healthy diet in each country. There is also a legend at the bottom which shows the scale of the colors in US dollars. The marks in this visualization are primarily areas (each country), and channels include color representing how expensive the average diet is, shapes representing each country, and area/length in the bar chart representing more specifically how much each diet costs. This user can use this visualization by hovering on the different countries and reading the US dollar value of the cost of the diet that appears at the top of the visualization. They can also see visually how dark red countries are which indicate the cost level of the diet.

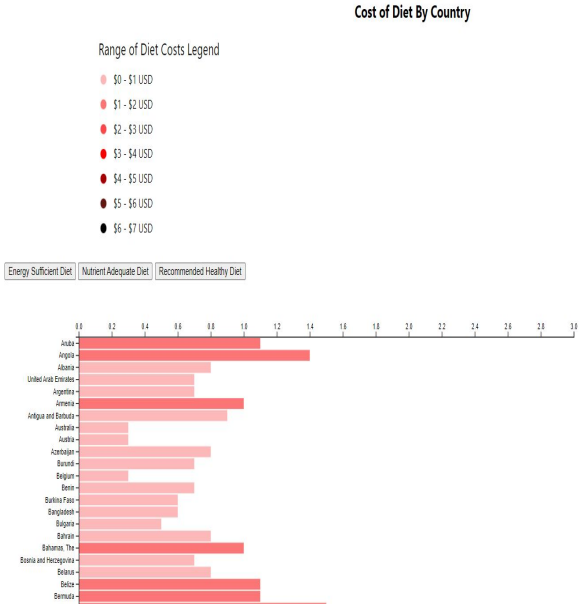



The purpose of the scatter plots in our visualization is to be able to clearly compare how much the diet type costs to what percentage of the population can afford this diet, as well as indicate the income classification for each country. Each diet type has its own scatter plot in order to visualize more clearly by distinguished type. The diets include the energy sufficient, nutrient adequate, and recommended healthy diet. In order to show income classification, we have colored the dots of each country based on the classification in the dataset and indicated in the legend what each color means in order to get a better understanding whether countries with higher incomes can afford each diet more or less as well as see whether the diet type costs more or less compared to other countries. Each scatter plot indicates on the x-axis the cost in US dollars for the diet type and on the y-axis the percent of the population which can afford the diet type. These scatter plots also implement brushing and linking which will allow users to compare for one or more country the difference in affordability and cost of all the diets. When a user selects an area of some countries on one scatter plot, on the other two scatter plots the same countries selected will show up and reveal the differences between the diet types. The purpose of this portion of our visualization is to give the user a better understanding of the affordability of these diet types and compare them to each other and between different countries. The marks are represented by points. Each points indicates a country on this scatter plot and the location indicates both how much the cost of an energy sufficient diet is as well as how much of the population can afford this diet. Channels include horizontal and vertical position as well as color. The horizontal and vertical position each have a different significance, both showing the comparison of the country's cost of diet and the percentage of the population which can afford that diet. The colors indicate the income classification determined for that country and helps visualize the cost of meals and affordability

of the population for countries in the same income category. The tasks that this visualization completes are look at comparing which countries have the most expensive diet, what percentage of the population can afford these meals, as well as which countries are wealthier. For the user to use these scatter plots, they can hover over the dots which each represent a country and see by the color what its income classification is as well as the information that pops up which informs the user about the cost of the diet type in that country as well as the country name and the percent of the population which cannot afford that diet. In addition, the user can select a region of dots and those country dots will show up bolded in the remaining two scatter plots to easily compare the three diet types costs and percentage of population which cannot afford for a country or groups of countries.




The bar chart at the bottom has a purpose of giving the user a better understanding of the cost of each diet type in each country. While the map provides an overall average between the three diets and a general background to the cost of these diet types, this bar chart allows the user to dive deeper and understand the cost in each country individually and compare them to each other. A user can select whether they want to view the cost of a energy sufficient, nutrient adequate, or recommended healthy diet for each country. The bar chart will show all the countries in descending order as well as the value at the end of the bar indicating the exact US dollar amount for the cost. The user can then get a better understanding of the costs and see the distribution of each data type. The marks in this visualization are 2D areas represented by the lengths of the bars. Channels in this visualization include shape, color (more expensive being darker) and also area of the bars. The user can click on the three different diet type buttons to see how expensive they are for each country and the legend helps indicate how expensive it is based on how dark the color is.

## 8 Discussion

In developing this visualization, we've had the opportunity to learn about the costs of different diets around the world and the proportion of the population that can afford the three diet types. This has made us realize that the affordability of food, particularly healthy diets, is still a very prevalent problem that much of society has to face on a day-to-day basis. Many countries (Liberia, Togo, etc.) have ¿90 percent of the population which cannot afford a healthy diet. A surprising finding from this project is that the different income classifications of countries (low, lower, upper, high) do not correspond to the cost of diet; the countries in each income bracket still have a wide range of diet costs. However, these income classification groups do determine the percentage of the population that is unable to afford a given diet (i.e. higher income classification countries have lower percentages of people who can't afford each diet type).

In terms of this project as a whole, we were able to learn a lot about different aspects and interactions that a single visualization can have. A lot of times one small change can make a big impact. For example, implementing the three different tabs made our visualization much more user friendly and easy to navigate (versus having to scroll multiple pages to view the entire visualization). Furthermore, having a balance between interactivity and simplicity is something that we worked on throughout the project after noticing that having too many interactions would overwhelm a user and too few would be boring. Furthermore, this project truly showed the endlessness of data visualization and how important it can be in communicating data and telling stories.

When reflecting on this project, a few limitations that we identified include: being able to zoom in on the map and show each country individually, having the map labels show up where the mouse is rather than on top of the entire world map, and developing a way for users to filter the bar chart. The first two limitations would help the map be more interactive and allow users to more clearly identify individual countries. We had made significant progress on both of these features but ended up having some difficulties. For next steps, we would implement more interactivity for the world map and the bar chart, and also improve on what the map depicts (ideally, we could allow the user to choose whether the map shows the energy-sufficient, nutrient-adequate, or suggested healthy diet). Another next step would be publishing this writeup and the visualization so other people can gain insights from our work. Whether that's food inequality activists, policymakers, or other potential users, we believe that this data allows people to learn more about costs and affordability of diets around the world.

## 9 Conclusion

With the topic of food affordability in mind, we began our project by searching for a data set with data such as costs of various diets and population income levels. To utilize the data, we conducted data cleaning using the Pandas library in Python by getting rid of unnecessary columns and rows with too many null values. Finally, we worked on using D3, Javascript, CSS and HTML to create three visualizations. This included a chloropleth map depicting various countries' nutrient adequate diet costs with an interactive hover feature showing the country's name and diet cost, three scatterplots showing the relationship between diet cost and the percentage of the population able to afford that diet (with brushing and linking), and a bar graph depicting the costs of the diet per country and adjusts based on the diet type the user selects. After conducting usability tests, we also improved our visualizations with adjustments including adding tabs to view each visualization separately (to prevent users from becoming overwhelmed) and adding graph titles, axis labels, and legends. In the future, we could improve on areas such as adding a zoom function in the map and having the map labels show near the country the user is hovering over so that users can clearly identify the countries. Furthermore, we also would like to include three separate maps, one for each diet, with a similar functionality to our bar chart where the map adjusts accordingly based on the user's selection. For future work, we hope to publish our visualization tool and write-up so we can raise awareness of food cost inequality between countries.

## 10 Acknowledgements

We would like to acknowledge Professor Ab Mosca for her guidance on our project and teaching us the conceptual and technical knowledge that formed the basis of our understanding necessary

to complete this project. We would like to thank the members of the "Visualizing the Recipe" group for being our test users who completed the tasks in our usability testing and provided us with helpful feedback that allowed us to make major improvements on our visualization. Finally, we would like to thank the TAs of DS 4200 for answering our questions during office hours and providing the support that allowed us to excel on this project.

## REFERENCES

[1] G. Agapito, B. Calabrese, P. H. Guzzi, M. Cannataro, M. Simeoni, I. Caré, T. Lamprinoudi, G. Fuiano, and A. Pujia. Dietos: A recommender system for adaptive diet monitoring and personalized food suggestion. In *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, pp. 1–8, 2016. doi: 10.1109/WiMOB.2016.7763190

[2] S. Davies. Fighting food poverty. *Engineering Technology*, 7(11):48–49, 2012. doi: 10.1049/et.2012.1111

[3] J. Hoelscher and A. Mortimer. Using tableau to visualize data and drive decision-making. *Journal of Accounting Education*, July 2018. doi: 10.1016/j.jaccedu.2018.05.002

[4] A. King, N. Chew, A. Jay, A. MacLean, and A. Bargagliotti. A guide to modern data visualization. *Math Horizons*, 28, Jan. 2021. doi: 10.1080/10724117.2020.1768775

[5] H.-C. Li and W.-M. Ko. Automated food ontology construction mechanism for diabetes diet care. In *2007 International Conference on Machine Learning and Cybernetics*, vol. 5, pp. 2953–2958, 2007. doi: 10.1109/ICMLC.2007.4370653

[6] H.-C. Li, W.-M. Ko, and H.-W. Tung. Food clustering analysis for personalized food replacement. In *NAFIPS 2007 - 2007 Annual Meeting of the North American Fuzzy Information Processing Society*, pp. 382–386, 2007. doi: 10.1109/NAFIPS.2007.383869

[7] Muliati, I. Gani, E. U. A. Gaffar, Haviluddin, and A. F. O. Gaffar. Measurement of health, human capital, and economic welfare at the poverty level influences using heuristic network approach. In *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, pp. 367–373, 2018. doi: 10.1109/EIConCIT.2018.8878595

[8] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, Nov 2010.

[9] D. S. Sehi L'Yi, Youli Chang and J. Seo. Toward understanding representation methods in visualization recommendations through scatterplot construction tasks. *Eurographics Conference on Visualization (EuroVis) 2019*, 38(3), 2019. doi: 10.1111/cgf.13682

## 11 APPENDIX A: GROUP CHARTER

- **Group Purpose**

  We decided to work together as a group of three after seeing how well we worked together on the homework assignments. We found it very convenient to communicate between the three of us and the dynamic is very collaborative without anybody having to overcompensate for lack of work. The purpose of our group is to work together to create this visualization related to cost of healthy food in relation to affordability and learn more about this interesting topic.

- **Group Goals**

  Our group's project goal is to create an effective nutrition visualization which the general public could use to inform themselves about which healthy diets are more expensive and in which countries as well as see how much of the population cannot afford these diets. This can help inspire people to advocate for more food programs and affordability of healthier food.

  The group's process goal is to be meeting consistently and working effectively towards our project end result and goal. We are hoping to complete all the assignments at a timely manner before the deadlines and complete them to our best abilities. We are planning on putting in a lot of time and effort into making this a good quality visualization and project as this topic is important to all three of us.

  The group's quality goals is to create a high quality visualization with plenty of analysis and explanations. We hope to be able to go to office hours and make sure that our project not only has all the important characteristics of a successful project for a good grade but also one that we could potentially make useful to people in the future as health concerns and diseases related to diet are becoming a pressing issue.

  We are all willing to commit as much effort as possible into making this a great visualization. We think that by choosing this topic it will help keep us engaged and we have already established can work together really well. We are aiming to get an A in this course as we try to always do our best and are gaining a lot of knowledge throughout this course already.

- **Group Member Roles/Responsibilities**

  For our group member roles, we will have Emily Wang act as the group leader, Xinyu Wu act as the meeting facilitator and information manager, and Kristina Fujimoto act as the documentation coordinator. Emily will oversee the project throughout the duration of the project and ensure that the project deliverables are completed in a timely manner. She will also consistently check in on the project members to confirm that they are working toward previously established goals and checkpoints. Xinyu will coordinate meetings and be responsible for finding relevant datasets pertaining to cost of a healthy diet and affordability. She will also be responsible for the preliminary planning of our visualizations. Kristina will be responsible for documenting the data we use as well as consistently reviewing our completed work to prevent possible errors. She will also act as a point of communications between our group and the TAs and the professor when we need to ask questions, etc regarding the project. All three of us will agree to work on the technical aspects of this project from writing project code to developing visualizations on D3 together. We agree to consistently communicate with each other to make decisions regarding visualization elements and techniques as well as all other aspects of our work during this project.

- **Ground Rules**

  We will plan to meet weekly (or more often if needed) over Zoom to discuss our project deliverables and our plan of action. Overall, all members of our group agree to contribute equally to each project milestone, and will communicate actively using our group chat. If we run into problems, have a particularly busy week, or aren't able to meet during a given week, we will be honest and open with each other. Discussions will be largely the same as they have been. We will respect each other's opinions, and decide everything as a group. We will logic and discuss our differences when dissenting views come up, although we believe this is unlikely to happen. Accountability just means that we check-in with each other every couple of days. We each are willing to commit as many hours as necessary to complete the project to the best of our ability. We expect anywhere between 1 to 4 hour commitment per week. Each member will actively participate in all meetings and during all aspects of the project.

- **Potential Barriers and Coping Strategies**

  One potential barrier that we may run into during this process include making sure to make time to meet as a group each week, and getting work done before the deadline. As each of us are highly involved on campus and have many other commitments, we will need to communicate and plan ahead to avoid any last-minute work. This will include setting a consistent meeting time and actively using the text group. Otherwise, we don't see any barriers as we have worked with each other in the past. Some high-friction group work experiences that we've had in the past include being 'ghosted' by team members and members not doing high quality work. We don't believe this will affect us, but we know that if this does come up, we just need to discuss with one another and talk through any challenges. We will maintain open, honest relationships with one another.

- **Mid Project Update**

  Yes, we have all been meeting up consistently together and work together well. We are all working hard together to complete each assignment by doing our roles and keeping each other motivated throughout the process and on top of things. There are no problems group-wise, but we are having potential issues implementing some ideas which we would go to the professor or TAs for help so we can get more resources in order to learn how to do it. Both of my group members contribute greatly, I have loved seeing Xinyu make really detailed sketches and come up with great ideas. I have loved seeing Emily help keep us organized and helped find great resources for background info and coming up with the idea. They both also contribute greatly to all assignments and have such positive outlooks towards achieving our goals and tasks.

  Our group works well together, and we are good at communicating in order to finish project assignments on time and split up work evenly. We all work on the different requirements of the project, and are willing to be flexible and contribute to build upon each of our own ideas. Emily works diligently on each section and question, contributing ideas to make our project better. Kristina is good at making sure we don't miss parts of the assignment, responding to each question in full, and aggregating everything for the submission.

  Our group works very well together. Each member completes responsibilities that we set in advance, contributes ideas, and communicates well so we understand each other's schedules and can set up meetings accordingly and make project decisions without conflict. Everyone has an even role within the

project and contributes with enthusiasm, which makes for a pleasant and effective working environment. Kristina works conscientiously by combining our efforts and checking each final deliverable to ensure we have completed each section of the project milestone. She also contributes ideas and acts enthusiastically at each meeting. Xinyu creates impressive visualizations that are visually representative of our goals and look aesthetically pleasing, while also contributing ideas and ensuring that our group stays on task. Overall, our group works well, with each group member being respectful and communicating with the others while ensuring that we put our best efforts toward the project.

## 12 APPENDIX B: DATA EXPLORATION

### 12.1 Data Collection

**Data Types:** There are many data types including items which are the rows of the excel file. One other common data type are attributes which are the columns. Below is a description of each column and the type of attribute it is (categorical or ordered). This is done for both sheets of our data, one that focuses on cost of each diet type per country and another that focuses on the nutritional values needed for each nutrition.

| | |
|---|---|
| country_name | Name of each country listed, categorical |
| country_code | Code for each country listed, categorical |
| macroregion | One of the 7 regions of the world, categorical |
| subregion | Each region split up further to describe the country's location better, categorical |
| income_classification_2017 | Income class of a majority of the country in 2017, ordinal (ordered) |
| population_millions | Population in millions of the country, quantitative (ordered) |
| Cost_energy (cost of energy sufficient diet) | Cost of least expensive, locally available starchy staple to sustain human energy in US dollars per day, quantitative (ordered) |
| Cost_nutrition (cost of nutrition adequate diet) | Cost of least expensive, locally available foods to meet daily nutrient requirements in US dollars per day, quantitative (ordered) |
| Cost_healthy (cost of healthy diet) | Cost of recommended diet, least expensive locally available items in each food group needed to reach recommended diet per day, quantitative (ordered) |
| Percent_poverty_energy<br>Percent_poverty_nutrition<br>percent_poverty_healthy | Diet costs as percentage of global ($1.9) poverty line available for food expenditures ($1.2/day) for each diet type, quantitative (ordered) |
| Percent_food_energy<br>Percent_food_nutrition<br>percent_food_healthy | Diet costs as percentage of country average food expenditures per capita from national accounts data for each diet type, quantitative (ordered) |
| Percent_unafford_energy<br>Percent_unafford_nutrition<br>percent_unafford_healthy | Percent of population who cannot afford each diet, for whom diet costs >63% of household income for each diet type, quantitative (ordered) |
| Num_unafford_energy<br>Num_unafford_nutrition<br>num_unafford_healthy | Number of people who cannot afford each diet, for whom diet costs >63% of household income in millions for each diet type, quantitative (ordered) |
| 100_percent_unafford_energy<br>100_percent_unafford_nutrition<br>100_percent_unafford_healthy | Allows food budget up to 100% of income for each diet type, quantitative (ordered) |
| 100_num_unafford_energy<br>100_num_unafford_nutrition<br>100_num_unafford_healthy | Allows food budget up to 100% of income, quantitative (ordered) |
| Mean_percent_unafford_energy<br>mean_percent_unafford_nutrition<br>mean_percent_unafford_healthy | Allows food budget up to mean for the country's WorldBank income classification, quantitative (ordered) |
| Mean_num_unafford_energy<br>Mean_num_unafford_nutrition<br>mean_num_unafford_healthy | Allows food budget up to mean for the country's WorldBank income classification, quantitative (ordered) |
| 98_cost_nutrition | CoNA, using RDA instead of EAR for nutrient requirements, quantitative (ordered) |
| cost_flex | Cost of Flexitarian Diet in US dollars per day, quantitative (ordered) |
| cost_pesc | Cost of Pescatarian Diet in US dollars per day, quantitative (ordered) |
| cost_veg | Cost of Vegetarian Diet in US dollar per day, quantitative (ordered) |
| cost_vgn | Cost of Vegan Diet in US dollar per day, quantitative (ordered) |

| | |
|---|---|
| nutrient | Name of each nutrient, categorical |
| unit | Abbreviation for measurement unit, categorical |
| avg_req | ARs: Average requirement of each nutrition, quantitative (ordered) |
| rec_allow | RDAs or AIs: Recommended dietary allowance or adequate intake for each nutrition, quantitative (ordered) |
| reduce_risk_lower | AMDR lower: Acceptable Macronutrient Distribution Range (AMDR) is the range of intakes for a particular energy source that is associated with reduced risk of chronic disease while providing adequate intakes of essential nutrients, quantitative (ordered) |
| reduce_risk_upper | AMDR upper: Acceptable Macronutrient Distribution Range (AMDR) is the range of intakes for a particular energy source that is associated with reduced risk of chronic disease while providing adequate intakes of essential nutrients, quantitative (ordered) |
| tolerable_upper_intake | UL: Tolerable Upper Intake Level is the highest level of nutrient intake that is likely to pose no risk of adverse health effects for almost all individuals in the general population, quantitative (ordered) |

**Initial Data Review:** Generally the data fits our expectations. Something to note is that there is a lot of spread between the 75th percentile and the max for a lot of these values. This might make it difficult to create certain visualizations since it will leave a lot of blank space in between higher up values. The nutrition data is also very much expected values and distributions.

| | num_unafford_nutrition | mean_num_unafford_healthy | cost_flex | cost_pesc | cost_veg | cost_vgn |
|---|---|---|---|---|---|---|
| count | 143 | 143 | 171 | 171 | 171 | 171 |
| mean | 22.76325417 | 32.34862544 | 3.62985738 | 3.722881982 | 3.4780046 | 3.396904483 |
| std | 89.05863296 | 123.9124957 | 0.863845706 | 0.93592585 | 0.949335187 | 1.067537722 |
| min | 0.000853498 | 0.002102443 | 1.854946904 | 1.895973838 | 1.801109445 | 1.748946755 |
| 25% | 0.64439815 | 1.212861595 | 3.015239439 | 3.064099174 | 2.880508518 | 2.76862119 |
| 50% | 3.559770756 | 4.84089142 | 3.464443935 | 3.568079884 | 3.289229395 | 3.125358987 |
| 75% | 15.3469208 | 20.67365692 | 4.072363087 | 4.226308309 | 3.852758023 | 3.762073915 |
| max | 942.7648082 | 1176.350467 | 7.416760012 | 7.777582649 | 7.987439383 | 8.458398585 |

| | R | S | T | U | V | W | X |
|---|---|---|---|---|---|---|---|
| | 100_percent_unafford_nutrition | 100_num_unafford_energy | 100_num_unafford_nutrition | 100_num_unafford_healthy | mean_percent_unafford_energy | mean_percent_unafford_nutrition | mean_num_unafford_energy |
| count | 143 | 143 | 143 | 143 | 143 | 143 | 143 |
| mean | 0.14343971 | 0.809674871 | 4.951955107 | 12.20541031 | 0.12009538 | 0.43103623 | 4.398716214 |
| std | 0.20353841 | 3.581148305 | 15.99765291 | 47.74860799 | 0.18326222 | 0.29005572 | 14.20250851 |
| min | 0 | 0 | 0 | 0 | 0 | 8.66141E-05 | 0 |
| 25% | 0.001393803 | 0.000526432 | 0.025941245 | 0.068852021 | 0.005556463 | 0.1386605 | 0.040020052 |
| 50% | 0.03275637 | 0.021564422 | 0.273126867 | 0.767132832 | 0.03493116 | 0.4427491 | 0.307908485 |
| 75% | 0.2307357 | 0.272888928 | 3.755952027 | 8.354748921 | 0.14652415 | 0.6901597 | 2.749383073 |
| max | 0.8458904 | 38.62573401 | 154.7643559 | 523.6418378 | 0.7909634 | 0.9623644 | 130.7808383 |

| | percent_food_nutrition | percent_food_healthy | percent_unafford_energy | percent_unafford_nutrition | num_unafford_energy | num_unafford_nutrition | num_unafford_healthy | 100_percent_unafford_energy |
|---|---|---|---|---|---|---|---|---|
| count | 171 | 171 | 143 | 143 | 143 | 143 | 143 | 143 |
| mean | 0.593938625 | 0.932114071 | 0.067019286 | 0.251733572 | 1.919581105 | 11.85214074 | 20.77472621 | 0.031377272 |
| std | 0.454583782 | 0.843413013 | 0.140026761 | 0.284684288 | 7.829598204 | 50.40863049 | 86.48069122 | 0.081300792 |
| min | 0.167431001 | 0.266600255 | 0 | 0 | 0 | 0 | 0 | 0 |
| 25% | 0.310757469 | 0.446779064 | 0.000794972 | 0.011036151 | 0.003108121 | 0.068633995 | 0.172435216 | 8.03426-05 |
| 50% | 0.445574633 | 0.663824923 | 0.005873047 | 0.1142699 | 0.04899294 | 0.946174216 | 1.617526968 | 0.001701342 |
| 75% | 0.736646297 | 1.032552375 | 0.054688585 | 0.4731357 | 1.122518828 | 7.552362631 | 11.35626853 | 0.01497706 |
| max | 3.912985269 | 6.64809753 | 0.7025125 | 0.935762 | 85.63882305 | 571.1156332 | 956.9492372 | 0.44839 |

| | population_millions | cost_energy | cost_nutrition | cost_healthy | percent_poverty_energy | percent_poverty_nutrition | percent_poverty_healthy | percent_food_energy |
|---|---|---|---|---|---|---|---|---|
| count | 171 | 171 | 171 | 171 | 171 | 171 | 171 | 171 |
| mean | 42.09414175 | 0.831018644 | 2.455916329 | 3.671937589 | 0.694251164 | 2.051726257 | 3.067617034 | 0.213354886 |
| std | 151.8707277 | 0.365158043 | 0.655645983 | 0.752789857 | 0.305061022 | 0.547741005 | 0.628897123 | 0.200524398 |
| min | 0.029577 | 0.265348333 | 1.172840304 | 1.887692082 | 0.221677806 | 0.979816461 | 1.577019283 | 0.03225538 |
| 25% | 2.086704 | 0.607675596 | 2.03142607 | 3.184915821 | 0.507665494 | 1.697097803 | 2.660748388 | 0.084095241 |
| 50% | 8.797566 | 0.788297947 | 2.342294252 | 3.561060854 | 0.658561359 | 1.956803887 | 2.974988182 | 0.155900018 |
| 75% | 28.885239 | 1.013898239 | 2.72470131 | 4.100957017 | 0.847032781 | 2.276275113 | 3.426029254 | 0.259742031 |
| max | 1386.395 | 3.034385921 | 5.867017476 | 6.206968342 | 2.534992416 | 4.901434817 | 5.185437211 | 1.439768989 |

| | reduce_risk_lower | reduce_risk_upper | tolerable_upper_intake |
|---|---|---|---|
| count | 3 | 3 | 15 |
| mean | 124 | 224.3 | 1139.733333 |
| std | 119.5543391 | 145.0406495 | 1447.22061 |
| min | 51.8 | 90.6 | 5 |
| 25% | 55 | 147.2 | 30 |
| 50% | 58.2 | 203.8 | 300 |
| 75% | 160.1 | 291.15 | 2250 |
| max | 262 | 378.5 | 4000 |

**Data Issues:** Yes, we did notice a few issues with our data. There are some values missing for certain countries in the cost of each diet type as well as affordability indicators. This provides not a holistic view of each country for the cost of diets as it excludes certain countries due to lack of data availability. For the solution, we will make sure to make it known in our visualization that there is no data available for those specific countries by marking them on a map with a grey color or another indicator. We are a little confused about the meanings of certain columns, such as the bounds on certain indicators. We will work on this by reading the description of the data carefully and doing more research on the original web source of the data as well as research on certain nutrition terms to fully understand and be able to analyze this data properly.

## 12.2 Processing and Data Exploration

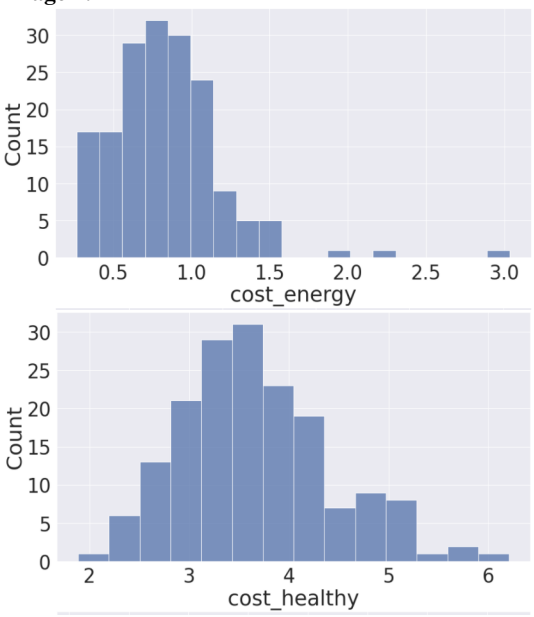
**Data Trends:** Some of the trends we see is that poverty is a pressing issue in many countries as the percentage and number of people who cannot afford healthy meals is quite high. And we can see that these percentages and number of people who cannot afford food only increases as income classification goes from upper class to lower class. Another trend we see is that with subregions, countries tend to have similar values of food cost and percentage of population which cannot afford those diets.

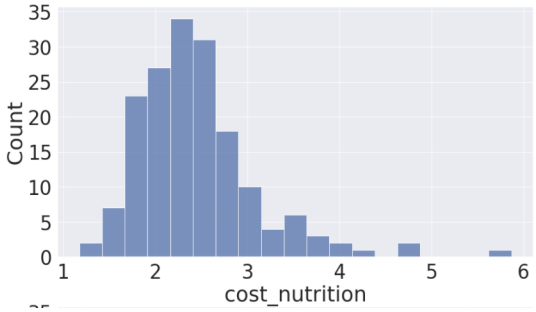
We were really surprised that the values of people unable to afford these types of healthy meals is so high. As we understand that a big part of the world doesn't receive the nutrition that is necessary for humans, we knew these values would be high, we just didn't expect it to be to this extent.

Another trend we noticed that we also expected, was that specialty diets such as the ones described in the data such as flexitarian, pescatarian, vegetarian, and vegan, all have higher costs per day than any type of healthy meal. This is due to the restrictions that these diets have on food products individuals have to buy and causes individuals to receive certain nutrients from other sources than from foods they're unable to eat.

**Screenshots**

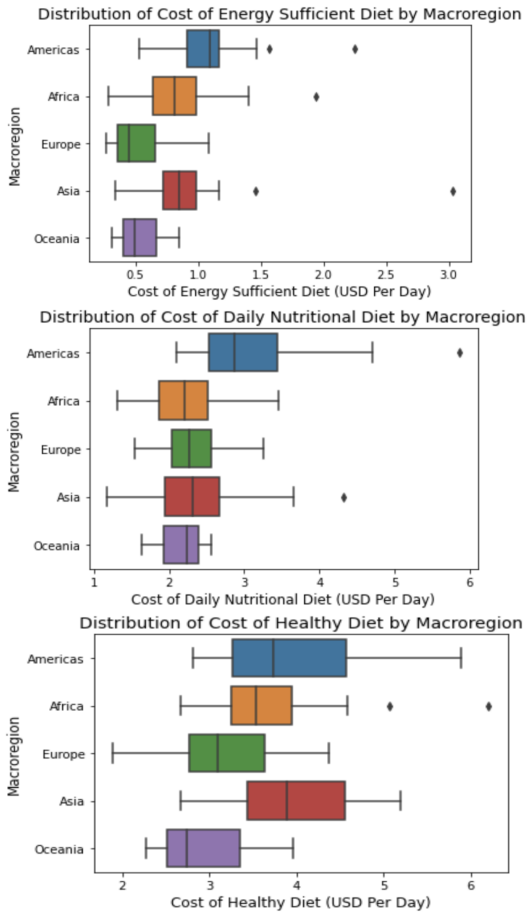
**Image 1:**





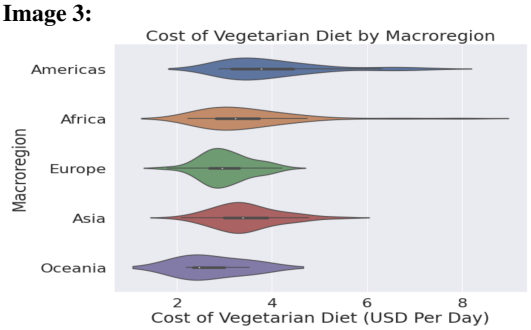

These three distribution plots show the skew of the costs of energy-sufficient, nutrition-sufficient, and healthy diets. More specifically, it depicts the number of data points at each cost level as a bar plot, including area marks and channels depicting length. The visualizations suggest that there is a greater skew to the right for the energy-sufficient diets and less skew for the healthy diets (i.e. the costs of healthy diets are more normalized). This is not surprising since energy-sufficient diets tend to be cheaper, and require less than healthy diets.

**Image 2:**




The three boxplots demonstrate the distribution of the energy-sufficient, nutrition-sufficient and healthy diets (as measured by cost in US dollars per day) grouped by macroregion, which includes Americas, Africa, Europe, Asia and Oceania. All three boxplots are visually encoded with marks including points, which represent outlier values of higher diet cost of certain countries), lines, which represent the range of costs in that macroregion, the middle 50% of costs and the median cost, and areas, which represent the middle 50% cost of the diets in their prospective macroregions. For channels, the marks including the lines and points are positioned horizontally, as traditional boxplots are typically shown. The areas are positioned both vertically and horizontally. The color of the areas are used to distinguish different macroregions, with blue representing the Americas, orange representing Africa, green representing Europe, red representing Asia, and purple representing Oceania. From the first boxplot, the distribution of cost of energy sufficient diets are relatively similar across macroregions with many regions' middle 50% ranging between $0.30 to $1.00 and the Americas having the highest median cost. The second boxplot showing the distribution of cost of daily nutritional diet demonstrates a similar middle 50% of costs for Africa, Europe, Asia and Oceania, which all overlap between $1.90 and $2.80 while the Americas has significantly higher costs with the middle 50% ranging between $2.50 and $3.50. The third boxplot depicting the distribution of costs of healthy diets show the Americas have a more similar middle 50% of costs to Africa and Asia, which range approximately from $3.25 to $4.50. Oceania and Europe's middle 50% of costs of healthy diet is lower with medians around $3.00.

**Image 3:**




The violin plot demonstrates the distribution of the cost of a vegetarian diet by macroregion which is primarily the continents. This violin plot is visually encoded with marks including points and area, which represent the max and min cost, the range, of vegetarian food per day in US dollars in each macroregion, as well as where most of the distribution sits, by a wider area of the violin. For channels, the marks including the lines are positioned horizontally, the way violin plots usually are represented. The area is shown vertically and horizontally as a distribution of cost of vegetarian diet. The colors of the different areas represent different macroregions and distinguish each violin. The visible trends from this violin plot is that the average cost for a vegetarian meal per day is pretty much similar with each macroregion, but the range is greater in the Americas and Africa compared to other macroregions. The max for these two macroregions is also much higher than those of the other macroregions.

### 12.3 Data Snippet

| | country_name | country_code | macroregion | subregion | income_classification_2017 | population_millions | cost_energy | cost_nutrition | cost_healthy |
|---|---|---|---|---|---|---|---|---|---|
| 0 | Aruba | ABW | Americas | Latin America and the Caribbean | High-income | 0.105366 | 1.129429 | 2.835770 | 3.964442 |
| 1 | Angola | AGO | Africa | Sub-Saharan Africa | Lower-middle | 29.816748 | 1.401920 | 3.230905 | 4.582558 |
| 3 | Albania | ALB | Europe | Southern Europe | Upper-middle | 2.873457 | 0.759639 | 2.588742 | 4.281676 |
| 4 | United Arab Emirates | ARE | Asia | Western Asia | High-income | 9.487203 | 0.745475 | 1.805732 | 3.412669 |
| 5 | Argentina | ARG | Americas | Latin America and the Caribbean | High-income | 44.044811 | 0.652094 | 2.624992 | 3.496882 |

| | nutrient | unit | avg_req | rec_allow | reduce_risk_lower | reduce_risk_upper | tolerable_upper_intake |
|---|---|---|---|---|---|---|---|
| 0 | Energy | kcal | 2329 | 2329 | NaN | NaN | NaN |
| 1 | Protein | g | 37.6 | 46 | 58.2 | 203.8 | NaN |
| 2 | Lipids | g | NaN | NaN | 51.8 | 90.6 | NaN |
| 3 | Carbohydrate | g | NaN | NaN | 262.0 | 378.5 | NaN |
| 4 | Calcium | mg | 750 | 1000 | NaN | NaN | 2500.0 |
| 5 | Iron3 | mg | 22.4, 11.2 | 22.4, 18 | NaN | NaN | 45.0 |

### 12.4 Task Analysis

| Domain Task | Analyze Task (high level) | Search Task (mid-level) | Analytic Task (low-level) |
|---|---|---|---|
| I want to use my visualization to see which diet is most expensive overall. | Consume, Discover | Browse | identify |
| I want to use my visualization to see which countries are most expensive for food. | Produce, Derive | Browse | identify |
| I want to show that a great percentage of people can't afford healthy food. | Consume, Discover | Lookup | summarize |
| I want to see which countries are wealthier. | Produce, Derive | Locate | identify |

The primary consumer of our visualization could be a number of people. Whether that's a government official, advocate group, or an individual looking to move to a different country. The primary consumption type will be Discover, in which the end user will be able to learn more about something they might've not known before. They are able to ask questions like "which country would be the cheapest to move to" or "which country has the highest/lowest cost of a nutritional diet". It can aid in other research, and be used to guide personal, political, or other decisions.

## 13 Appendix C: Interview

### 13.1 End User Personas

I am a researcher for the World Health Organization, and I've been tasked to investigate nutrition and diet across the globe, and report to my senior researcher who is writing an academic paper about starvation in developing countries. As an experienced researcher working for a globally-renowned organization, I have a strong background in using and interpreting visualizations. I will be referring to this tool as a source for my research, and using it to compare how much a healthy diet costs in countries that are in different stages of economic development. I'll want to know which countries are the most in need of aid when it comes to supporting healthy diets, and which countries do not need help from the WHO. Without the visualization tool, I would have to go find this data myself, and likely use Excel to sort the data. This would be a difficult and long process, and would take up more time than I have. The visualization will allow me to not see the raw data but still gain the insights that I'm looking for.

I was recently laid off from work and have been depending on US government-supported food stamps to survive. These have not been sufficient, and the labor market has been particularly tight due to COVID. Therefore, I'm hoping to move to a different country where the cost of food is less, and will ideally have more government support for me to avoid hunger. I've been looking at potential countries to move to just by Googling. However, there are just too many potential places to look at. I need to know how much a nutritious diet will cost me in each of these countries. This visualization will help immensely in showing me where I can get the most out of the very little money that I have left. It'll also show me which countries would be most economically sustainable for me to move to.

### 13.2 Interview Script

Hi, nice to meet you, could you briefly introduce yourself so I can better understand what background you bring?
May I ask what your current approach is to researching food and nutrition in underdeveloped countries?
So in terms of our visualizations, what specific goals are you hoping to achieve?
Please describe three pain points in your current process of research.
Who is the intended end user for these visualizations, who do you want to be able to use them?
Do you have a current vision of what you would like it to include?
Are there certain questions you want to be able to answer based off of the visualization?

### 13.3 Interview Notes

- Visualization will be used to report research outcomes and be seen by many who want to be informed on the topic
- Needs to be catered towards informing/finding answers to questions based on the data
- Currently manually searches for the answers in Excel spreadsheet: need to make it simpler to find min/max values etc.
- Would like something that is represented on map, with hover features
- Have different colors for socioeconomic standing
- Possibly include bar graphs to represent cost of different diets listed by countries

### 13.4 Interview Results

Visualization Designer: Hi, nice to meet you, could you briefly introduce yourself so I can better understand what background you bring?
End User (WHO researcher): Hello! I'm a researcher at the World Health Organization. My team is writing a paper related to poverty and starvation around the globe, particularly in developing countries.

Visualization Designer: May I ask what your current approach is to researching food and nutrition in underdeveloped countries?
End User (WHO researcher): I'm mostly looking at academic papers and databases. I usually read through them and look at what data sources they use to go do my own research.

Visualization Designer: So in terms of our visualizations, what specific goals are you hoping to achieve?
End User (WHO researcher): I'm primarily looking to get a general overview of costs of diets across the global. I want to be able to compare between different countries, but don't want the hassle of having to read through long papers.

Visualization Designer: Please describe three pain points in your current process of research.
End User (WHO researcher): I really dislike having to go through the data on excel myself. And I also hate not being able to look at the data visually on a chart or some sort, but it's on an excel file and all the data meshes together. I also feel like having to find the max and min of a column is ridiculous to find the data every time and whenever people ask me to find answers for them, it takes forever.

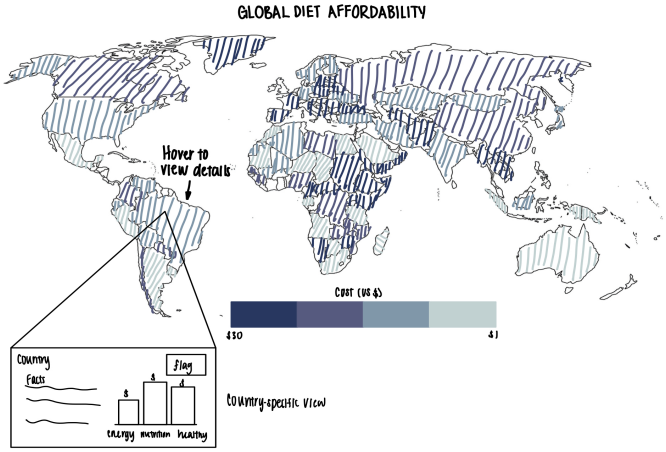
Visualization Designer: Who is the intended end user for these visualizations, who do you want to be able to use them?
End User (WHO researcher): My senior scientists as well as other people who would like to inform themselves about this topic. I want to be able to use it to lookup answers for myself and others based on the data.

Visualization Designer: Do you have a current vision of what you would like it to include?
End User (WHO researcher): Yes, I want it to have a map with different countries and clicking on or hovering which would allow my data to show up. I also want to be able to have it be color coded by a certain column indicator I have about whether the country is on the wealthier or poorer side of socioeconomic status.

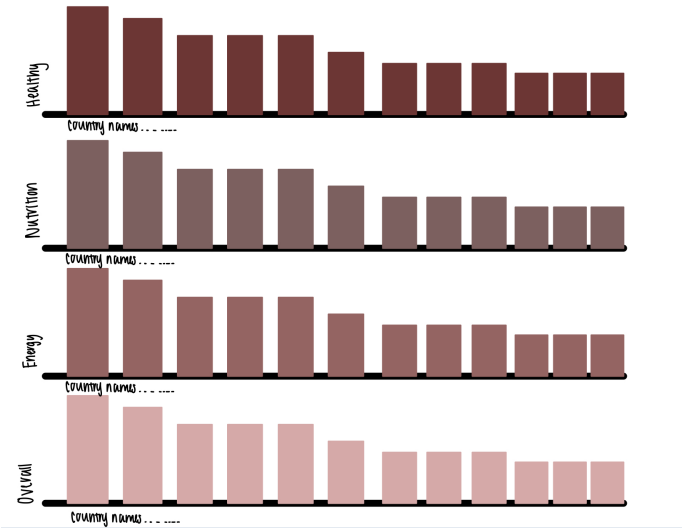
Visualization Designer: Are there certain questions you want to be able to answer based off of the visualization? End User (WHO researcher): Yes, I would like to answer all different types of questions of the data but here are a few especially to keep in mind. I want to be able to compare different socioeconomic statuses of countries, or be able to see which ones belong to which status based on color. I also want to be able to see which countries have the most expensive healthy diet costs in a bar chart on the side or be able to see the max and min for each one. I also want to be able to see which healthy diet costs the most averaged between countries. I also want to see the percentages of populations which can't afford food because it's too expensive for them. These are only a few things I would like to be able to see.

## 14 Appendix D: Design Sketches

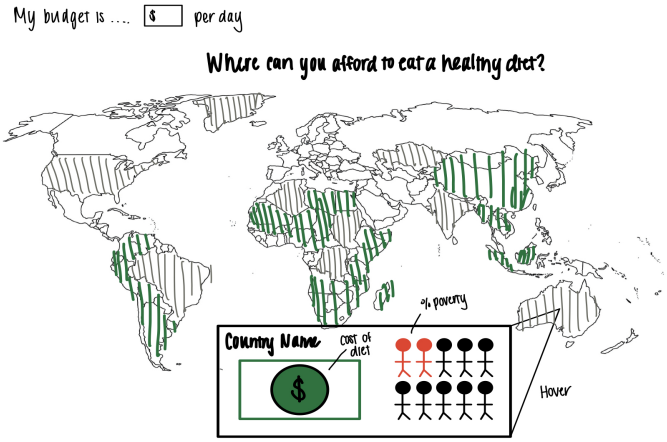
### 14.1 Xinyu




Favorite: This graph depicts a world map with each country colored based off of the cost of a diet. More expensive diets are darker in color while cheaper diets are lighter in color. Furthermore, when a country's shape is hovered over, there is a pop-up box that contains the country name, country flag, some basic facts, along with the cost of the three different diet levels (energy-sufficient, nutrition-sufficient, healthy). The marks in this visualization are primarily areas (each country), and channels include color representing how expensive the average diet is, shapes representing each country, and area/length in the bar chart representing more specifically how much each diet costs.

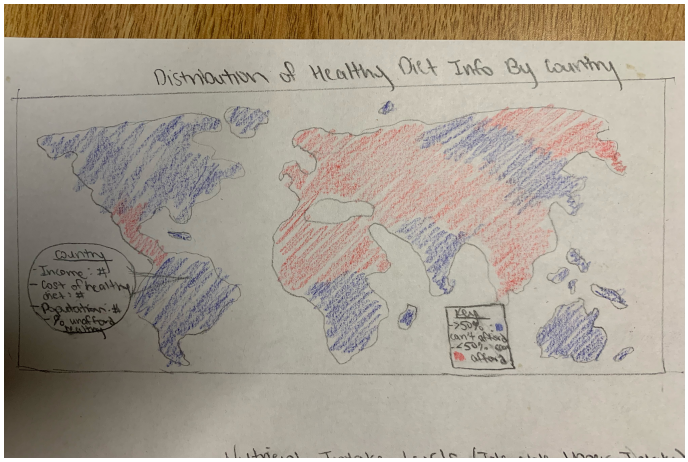



Favorite: This is a set of bar charts, with each bar chart representing a different diet (energy-sufficient, nutrition-sufficient, healthy diet, and overall). The bars are for each individual country, and the height represents the cost of that diet in the given country. The 'overall' bar chart would likely be an average of the other three diets. The marks in this visualization are 2D areas represented by the height of the bars. Channels in this visualization include shape, color (healthy being the darkest color) and also area of the bars.

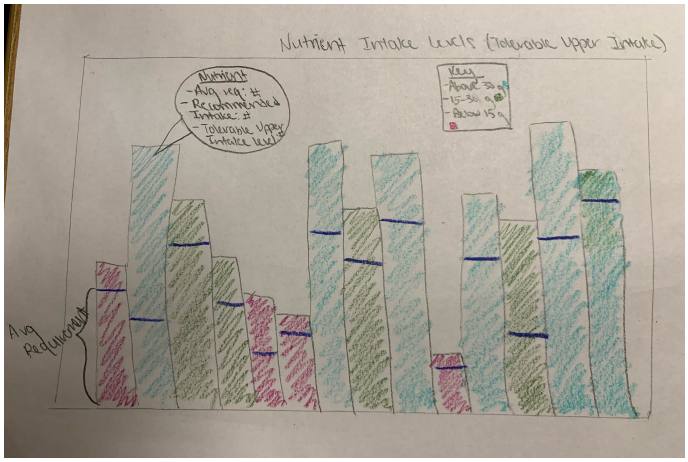



This visualization includes a place for the user to enter their budget per day. Then, the visualization will show areas where they would be able to afford a healthy diet. When a country's region is hovered over, then there would be a separate visualization that pops up with the exact cost of the diet along with a visualization showing the percentage of population that lives in poverty in that given country. Marks used in this graph are 2D areas of each country, channels like colors are used to represent the places where the given budget is sufficient, and for the poverty percentage.

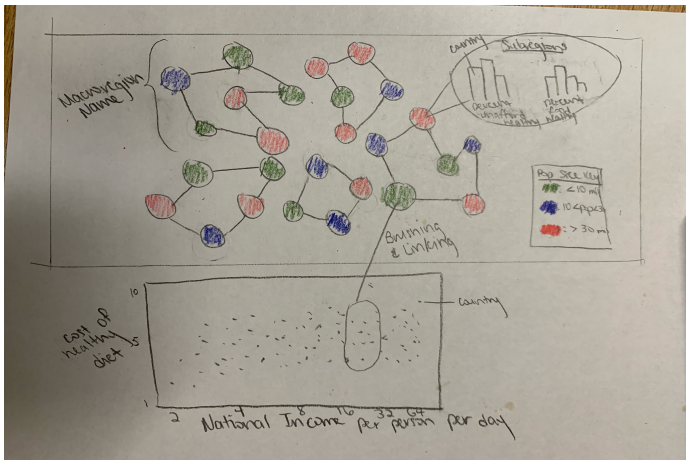
### 14.2 Emily




The visualization above depicts a map of the world with the marks as areas (items) representing countries. There are channels including the color, which represents whether the percentage of the population can afford a healthy diet (based on a certain threshold such as 50%), and shape, which represents the actual physical shape of the country. When the user hovers over a specific country/place, there is a bubble that pops up depicting the country's income and metrics such as the country's population, the cost of a healthy diet and the percentage of the population that are unable to afford a healthy diet. I chose these marks and channels because I wanted to visually show each country's metrics by allowing the user to study a visualization that is familiar (a world map) and display their geographical locations while showing how these metrics compare between countries in different or similar locations. This addresses domain tasks including showing which countries are most expensive for food, and showing the percentage of people that cannot afford a healthy diet for each country.

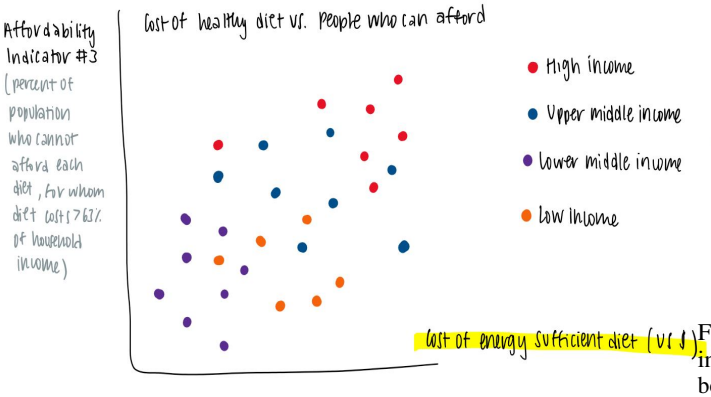



The visualization above depicts a bar graph with the marks as areas (items), each representing a specific nutrient. There are channels including the color, which represents whether the highest tolerable intake level for that nutrient is within a certain threshold (i.e. above 30 grams, between 15-30 grams, etc), and size, with the height of each overall bar representing the highest intake level that someone can consume without any adverse effects and the height of the smaller stacked bar representing the average requirement for the nutrient. When the user hovers over a specific bar, there is a bubble that pops up depicting the nutrient's average intake requirement, a recommended intake and the tolerable upper intake level. I chose these marks and channels because I wanted to visually show each nutrient and how the average requirements and tolerable upper intake levels compare between nutrients. This addresses domain tasks including showing how people can maintain a healthy diet based on how much of each nutrient they should/can consume.

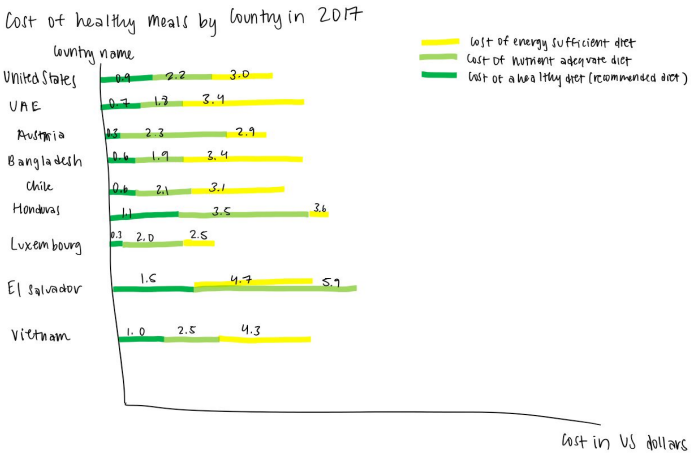



The visualization above depicts a connected node graph with the marks as nodes (items), each representing a specific subregion and each group of nodes that are connected representing a macroregion. There are channels including the color, which represents whether the population size of each subregion falls in specific range (i.e. greater than 30 million people), and position (which group/where the node is located), which shows the macroregion that a particular subregion or group of countries falls within. When the user hovers over a specific node, there is a bubble that pops up depicting a bar graph with the distribution of metrics for countries within that subregion such as percentages of populations who can't afford a healthy diet or an adequate nutritious diet, or the cost of a healthy diet. Hovering over a node also leads to a brushing and linking feature to a scatterplot below which selects all the countries in the scatterplot. The scatterplot shows the costs of a healthy diet in relation to the national income per person per day for each country (represented by a point). The marks are points while the channel of position of each point is used to demonstrate how expensive a healthy diet is in relation to income for each country. I chose these marks and channels because I wanted to visually show how different distributions of diet unaffordability vary by subregions and regions and allow users a visual way of comparing these metrics. I also wanted to show how the cost of healthy diets is related to the national income by country so I show this in greater detail with a brushing and linking feature in the scatterplot. This addresses domain tasks including showing which countries are most expensive for healthy diets, when taking average income level into account.

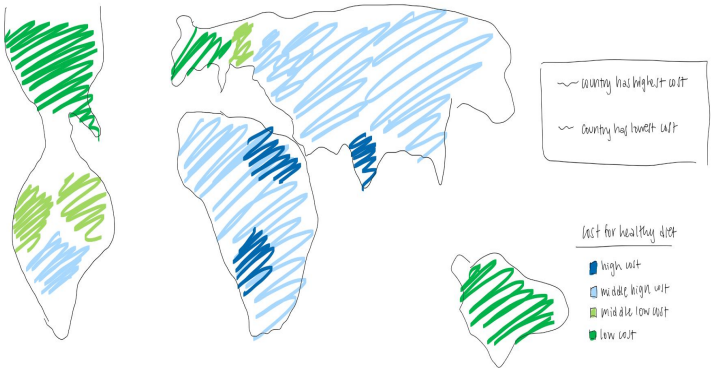
### 14.3 Kristina




Favorite: For marks, I chose to represent this with points. Each points indicates a country on this scatter plot and the location indicates both how much the cost of an energy sufficient diet is as well as how much of the population can afford this diet. For channels, I chose horizontal and vertical position as well as color. The horizontal and vertical position each have a different significance, both showing the comparison of the country's cost of energy sufficient diet and the percentage of the population which can afford that diet. The colors indicate the income classification determined for that country and helps visualize the cost of meals and affordability of the population for countries in the same income category. The tasks that this visualization completes are 2, 3, and 4, which look at the comparing which countries have the most expensive diet, what percentage of the population can afford these meals, as well as which countries are wealthier.




For marks of this visualization, I chose lines to help show the cost of each different type of healthy diet from our data in each country because it can be easier to compare between each country. In reality this would also be able to select each cost diet and order it from most expensive to least expensive to get a better understanding of the data. For channels, I chose to implement color and size (length). The color indicates the different categories of healthy diets and the length indicates to cost in US dollars of the healthy diet. This completes task 1 and 2, which are: seeing which diet is most expensive overall and which country has the most expensive healthy diets. This visualization does this through showing the costs for each country and costs for each diet and be able to compare them all.

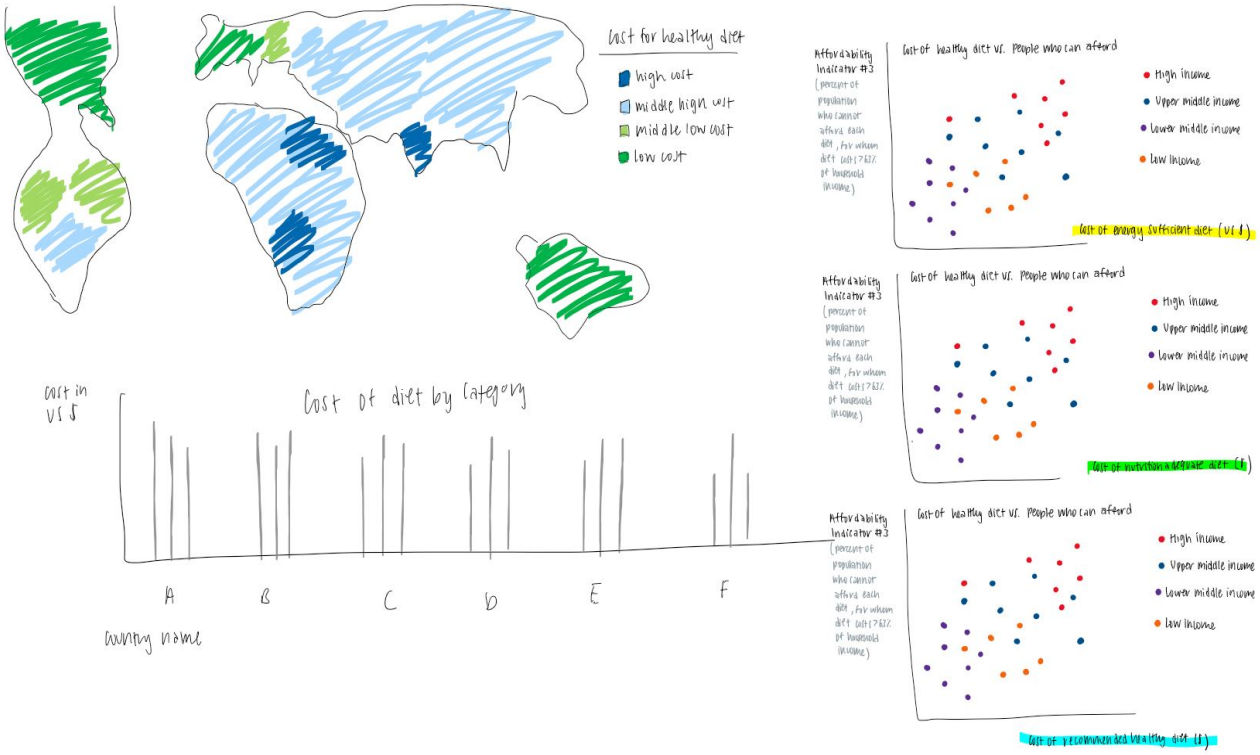



For marks, I chose to represent this with points. Each points indicates a country on this scatter plot and the location indicates both how much the cost of an energy sufficient diet is as well as how much of the population can afford this diet. For channels, I chose horizontal and vertical position as well as color. The horizontal and vertical position each have a different significance, both showing the comparison of the country's cost of energy sufficient diet and the percentage of the population which can afford that diet. The colors indicate the income classification determined for that country and helps visualize the cost of meals and affordability of the population for countries in the same income category. The tasks that this visualization completes are 2, 3, and 4, which look at the comparing which countries have the most expensive diet, what percentage of the population can afford these meals, as well as which countries are wealthier.

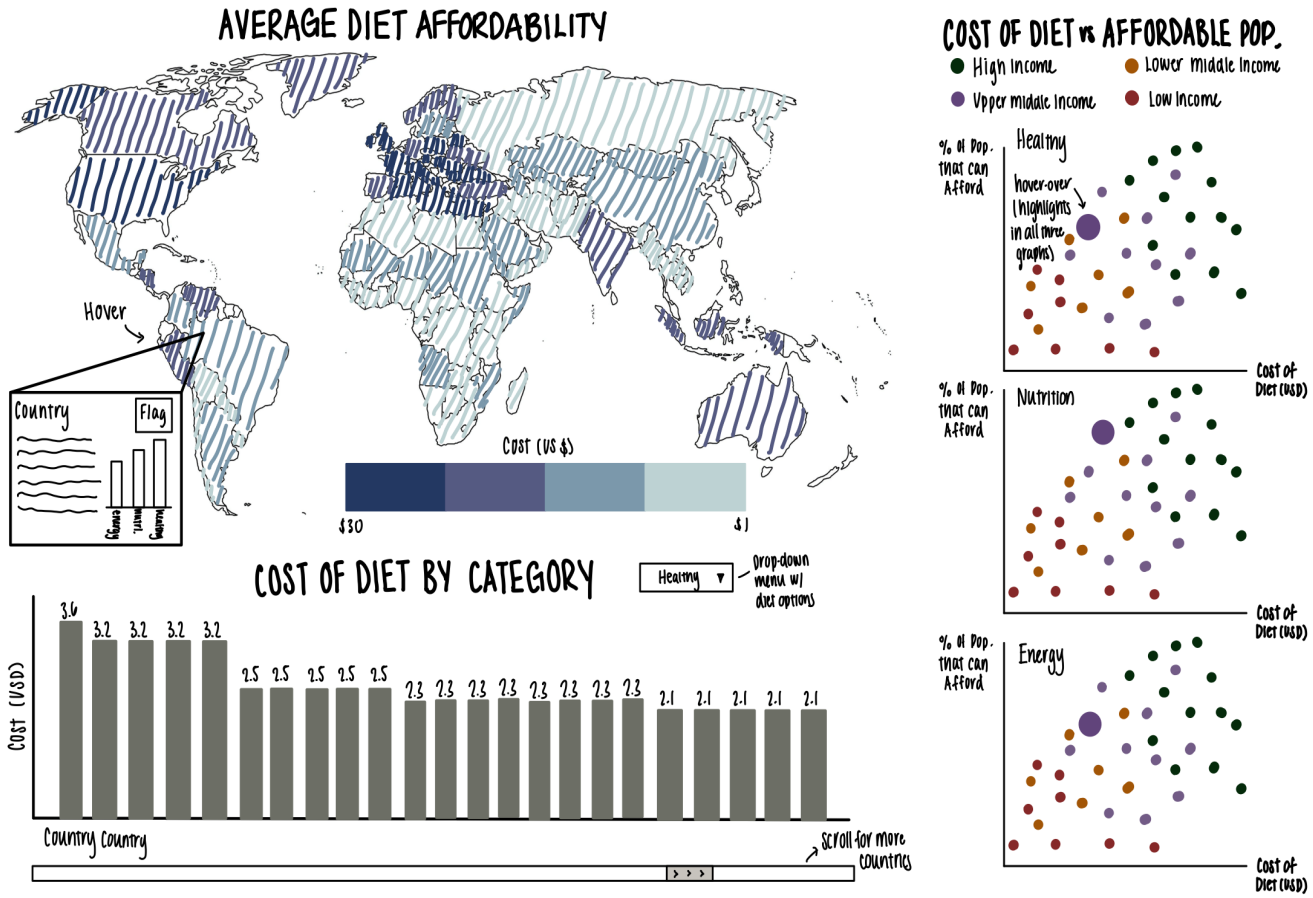
### 14.4 Group Sketch Selection

These visualizations are our favorite because we really enjoyed the detail and the different tasks that each of these complete. These can help us make a visualization tool that is the most descriptive of our dataset and the informative goal we want to accomplish. These visualizations each hit at least one of our task abstractions that we chose from the last project task and we really think these will help provide more insight about the data and world condition of healthy diets and cost.

For our first favorite, we chose the best map one we drew because it has the most detail. As this one was common between us we feel as if it would be best for representing some of our most crucial data. This would represent the countries and their cost for a healthy diet and indicate which countries have more expensive healthy diets versus which ones have more affordable healthy diets. This would complete our second task which observes which countries have more expensive healthy diets. For our second favorite, we chose a bar graph to help represent all the different countries and different types of healthy diets. This will help us compare each country individually and see on a better perspective of how much each diet costs as you can visually see the labeled bars that indicate the different diets and different countries. This completes tasks one and two which looks at which diets are most expensive as well as which countries have more expensive diets. For our third favorite, we chose a scatter plot which will help represent the different healthy diets in comparison to what percentage of the population can afford these diets. This helps us compare cost vs affordability of these diets and gives us a better understanding of how we need to create an impact which lowers the cost of healthy diets. This completes tasks three and four which look at the percentage of people which can't afford healthy diets and looks at which countries are wealthier by the color indicator of the dots of countries on the scatter plot.




Our final sketch will include these visualizations which includes a map, some scatter plots for each diet type, and a bar chart with the different diet types. We chose these for the same reasons as we did in our individual sketches. For the map, we chose a mark of area that represents the country, and we chose a channel of color to represent the cost of healthy diets in each country. For the bar chart, we chose lines to represent the different diets and countries, and we chose color and length to represent the type of diet and the cost of the diet. For the scatter plots, we made one for each different kind of diet and will incorporate brushing and linking. We chose points to represent countries, as well as position to show cost vs affordability as well as color to indicate country income classification. This tool helps us identify which countries are most expensive for food, represented by data in all of them. This tool also helps us see which diets are most expensive as shown in the bar chart and comparing the scatter plots. This tool also completes the task of determining population which can't afford the meals which is shown in the scatter plots. And finally, we can determine which countries are more expensive through the scatter plots as well.




Our visualization primarily has three parts: 1) the world map, 2) a set of three scatter plots, and 3) a bar chart with each country. The world map gives a general overview of diet affordability by region, and provides a easy visual overview of which countries are more or less affordable. It also allows users to interact and find more details by hovering over a given region. This then results in a visualization that details the name, flag, and more specific data for the given country. A user can explore how affordable a given country is using this world map. Particularly if they are looking to move to a certain region (for work, study abroad, or other purposes), they can learn more about expected costs of diets by simply hovering over the country on the map. The set of three scatter plots shows the energy-sufficient, nutrition-sufficient, and healthy diets in relationship to the percentage of population that can afford each diet. It also depicts the different socioeconomic classes by coloring the dots. If the user highlights a given country on any one of the three plots, the corresponding data points in the other two plots will also be enlarged/highlighted. A user is therefore able to learn more about the general affordability depending on the income group. This could be just out of curiosity, or to learn more about what to expect in a given economy, given a user's income level. The final plot, below the world map, allows users to more easily compare the affordability between countries. A user can select from the drop-down menu which diet they want to compare (energy-sufficient, nutrition-sufficient, and healthy) and the bar chart will appear with each country representing a different bar. Furthermore, the user can scroll using the bottom scroll bar in order to view countries that are not in the current view. This chart allows for easy comparison between different countries and the three categories of diets. Therefore, a user can identify which country would be most vs least affordable.

## 16 Appendix F: Usability Testing

### 16.1 Preparation

Our visualization informs the public about the cost of different diets and allows the user to understand cost versus affordability of these diets. Our visualization is intended to inform and educate on this topic of nutrition and cost and help advocate for healthier meals to be more affordable. Our visualization consists of a map, a bar chart, and three scatter plots. The purpose of the map is to provide a general overview of the average cost of the three types of diets: energy sufficient, nutrient adequate, and recommended healthy diet. This provides an understanding of the cost in each country for a healthy diet. The purpose of the bar chart is to view each diet type cost individually for all the countries. It allows the user to select the diet type and view the cost of that diet in all the countries in the dataset. This provides the user with a comparison between the different diet types and to see more specifically which countries are more expensive. The purpose of the scatter plots is to compare the cost of each diet type to the percentage of the population in each country which can afford it. The user can understand more clearly the distribution of cost and affordability for each diet type as well as the income classification for each country as it is depicted by different colors of dots.

By looking at the scatter plot, which country is the least affordable for each diet.
We are testing for this outcome as this involved interacting with all the scatter plots (currently the one with the most interactive aspects implemented). It involves using the hover feature to observe the data more closely which is something we want to work properly so the user can be informed properly about the data.

By looking at the scatter plot, which country has the highest cost for each diet type.
Similarly to the previous task, we want to test how well the interactive features of the scatter plots work and see whether obtaining this type of information is easily done so for any type of user. This will help us understand whether it is working properly.

By looking at the scatter plot, compare and contrast each diet type's distribution.
This task will allow the user to see the big picture and compare how expensive versus how affordable each diet type is for the countries. It is just to get an overall understanding of how the countries are distributed on the plots. This will help us understand if our data is easy to interpret/understand by any type of user.

For the United States, what is the affordability and cost for each diet type.
This would require the user to use the brushing and linking feature we implemented (if we have it ready). It will help us test to see how easy it is to use as well as how it connects the scatter plots.

### 16.2 Results

Overall, by conducting usability testing with another group, we were able to receive helpful feedback to implement on our visualization tool. One major issue with our visualization tool was a lack of graph titles and labels. Since our scatterplots and map visualization did not have any titles, it was not intuitive for our test users to understand the scope of the data that was being presented within each visualization. Additionally, another issue identified included the lack of legends for the scatterplots and map visualization. For instance, the test users were confused by the different colored points within each scatterplot and could not tell what the colors represented. Moreover, some of the points at the bottom of the scatterplot did not include percentage values, as the values were missing in the original data. The test users also identified being overwhelmed by the bar chart, as it shows a bar for each country so there are many bars. On the contrary, the test users thought the overall map tool was effective as it visually depicts where countries are and their corresponding diet costs. They also expressed that the visualizations were, overall, relatively intuitive and easy to navigate, with the hovering and brushing and linking functions making the overall visualization tool more effective.

For our first task, in which users had to identify which country with the least affordardable for each diet, our test users were able to complete this task with ease by hovering over the leftmost point in each scatterplot. This illuminated the effectiveness of our hovering function, which allows users to intuitively hover their mouse over any point on the scatterplot to view more information such as a country's diet cost. Since users were able to find this objective easily, a change to this design is unnecessary.

For our second task, in which users had to compare and contrast each diet type's distribution, our test users completed this task quickly by locating the rightmost point in each scatterplot. This usability task showed the effectiveness of the hovering function in conjunction with the scatterplot as well, with users being able to easily identify the highest cost. Therefore, we do not need to change this design.

For our third task, in which users had to identify which country has the highest cost for each diet type, our test users were able to give general descriptions of comparisons between diet type costs after briefly looking at each scatterplot. Since they were able to show an overall understanding of the affordability of each diet type for the countries, this usability task showed the effectiveness of the way we designed our scatterplots. One issue, they pointed out, was to include titles and axis labels to make it even easier to understand. Therefore, slight changes to the scatterplots are necessary.

For our fourth task, in which users were instructed to find the affordability and diet cost for the United States, our test users were able to easily identify this with the brushing and linking function. This usability task highlighted the effectiveness of the brushing and linking in conjunction with the scatterplots since users were able to easily understand how the diet costs of the U.S. compares with other countries for each type of diet, as well as between diet types, without having to search for the point representing the U.S. in each scatterplot. Therefore, a change to this design is unnecessary.

To modify our visualization tool after conducting these usability tests, we will add titles and clear axis labels to each of our scatterplots to allow users to have a better understanding of the data that is being presented. Furthermore, we will add legends, such as one for the scatterplots that demonstrate how the different colors correlate to different income classification levels (i.e. upper-middle class, low-income class, etc.) as well as a legend for the map that depicts how each shade of red correlates with a different diet cost range (i.e. $1-$2, $3-$4, etc.). This will allow users to easily interpret how the diet costs of countries in similar areas of the world compare. To help users obtain a more concise understanding of the bar chart without being too overwhelmed, we plan to color code each of the bars to represent a specific diet cost range (i.e. $1-$2, $3-$4, etc.) so that users can mentally group different countries' costs by looking at the bars' colors. Finally, we will get rid of points with missing values in the scatterplots so points without metrics such as percentage of population unable to afford an energy sufficient diet will not show.

## 17 APPENDIX G: REFLECTIONS

### 17.1 Kristina Fujimoto

Three things that I feel I did well on during this project was to communicate well with my group members and meet often/for long periods of time, help debug problems we were facing and finding resources for our visualization code, and keeping everyone on track by helping divide work and making sure we were turning it in on time. Three things I would change would be to have more team members as this would lessen the load for us, find a way to hover on the bar chart or make it more concise as there was too many countries on it, and possibly start earlier on project assignments as we would stay together working on it late into the night. The most valuable lesson I learned about design through this course is that there are lots of resources online to help you come up with a visualization that you have in your mind and that it takes a lot of tweaking and debugging to get code to work especially for interactive features.

### 17.2 Emily Wang

Three things that I feel I did particularly well for this project include actively communicating with my group members clearly and often throughout the project, checking and making sure our group followed all the instructions and small details mentioned in the pms, and learning to figure out technical problems in our code by being persistent and taking time to research resources. One thing I would change includes planning further in advance and starting assignments earlier because we would sometimes underestimate the time it would take to finish an assignment and work much later into the night to complete it. Another thing I would change would be being more realistic about the many features we wanted to implement in our visualizations given our limited time constraints from the beginning. Although we were able to implement many of the features we initially wanted to include, it would have been advantageous to strongly consider which features would be the most effective that we should prioritize before coding the visualizations to successfully implement all of the most effective features. Many of these features would focus on greater user interactivity since there were a few additional features we were close to figuring out but did not have time to fully implement. A third thing I would change is working with a greater number of people, as the project milestones for this course project included many components from coding the visualizations to writing in the report and having more members would help alleviate the amount of work necessary for each group member and help accelerate the process. The most valuable lesson I learned throughout the course of this project is that there is a great amount of thought and effort that goes into visualization design, from understanding design principles to considering the effectiveness of users' interactions with the visualization to the advanced technical skills necessary to execute these designs. Completing this project has taught me to have a greater appreciation for many of the visualizations I come across daily whether it may be in news articles or advertisements.

### 17.3 Xinyu Wu

Three things that I believe I did well for this project were: being flexible to changing our design as we proceeded along the project development process, actively communicating and working with my team members, and learning to find resources online when we didn't know how to implement a specific functionality or ran into errors. In terms of changes for the future, I would have a much more nuanced approach to our original designs. In hindsight, it didn't make sense to have so many different moving parts to our visualization on a single webpage, unless you're creating a dashboard where users can click on each part and zoom in. Another thing I would change would be to spend more time working on interactions. We were very close to having a few other interactions such as a zoom function on the map and also the hover over function on the map. We underestimated how difficult these interactions would be to implement. Lastly, I would spend more time on the data cleaning portion of this project, rather than having to go back and clean the data multiple times as we discovered different problems with the data throughout the design process. In terms of the most valuable lesson, I now have a much deeper appreciation for the different visualizations that we see in our daily lives — whether that's in the NYTimes front page or on some random website that we just so happen to scroll to. I now have a true understanding of how much technical skill and understanding is required to build even seemingly simple visualizations.

| Project Name | Specific Aspect of Visualization tool or presentation you liked the most | Follow-up question that you have for the group |
|---|---|---|
| Basketball Shot Percentages | The bar chart comparing shooting percentages with brushing and linking | Would there be a way to see the brushing and linking without having to scroll down? That might be easier for the user to identify relationships between the different charts? Is there a better way to consolidate the data? |
| Wordle Wonders | The word cloud showing popular starting words with a hovering feature in the line graph that shows the word highlighted in the word cloud and changing graphs that show the statistics behind each word in the cloud | Since wordle is a daily activity, how would you change your visualization to update with the new data everyday? |
| Visualizing Gender Equality in Film | Violin plot that shows runtime and rating | What other factors are there besides the Bechdel Test that could provide a deeper analysis on the representation of women in film? |
| An Analysis of MLB Opening Day Player Payroll vs Team Success | Bracket visualization that shows team spending on salary with color gradient scale | Were there any problems with your tooltip being so large in terms of brushing over the graph? |
| LEARN | Side by side maps of Massachusetts by county with hovering feature showing | Will this project be used by FirstByte? Did they have any requests that you believe your visualization weren't able to implement in full? |
| Plastic Pollution | Interactive chloropleth map of waste generation that shows waste generation metrics | Do you think this visualization could be used to advocate for less plastic pollution? |
| Visualizing Basketball Salaries | Bar charts that compare the two players and its interactiveness with being able to use the dropdown menu with different players | How would you consolidate the scatter plots in the future, I feel as if there are too many scatter plots and data that makes it confusing? Is there a better way to visualize these data? |
| Visualizing Congressional Stock Trading | Treemap showing most commonly traded stocks with hovering feature that changes the bars and points in the other two graphs | It seems like all of these companies are just your general blue-chip large cap stocks that everyone trades on the daily; how do you differentiate between a trade being placed based off of 'insider' knowledge versus just general shifts in portfolio allocation? |
| Covid-19 Dashboard | Line graph that shows the Covid cases over time with the drop-down menu selecting states and U.S. territories | How do you think future Covid-19 trends and rates will affect the view of your dashboard? |
| Premier League | Scatterplot showing expected goals for and against each team in the league with the icons of each team representing a point and a filtering feature with a bar chart | Do you think your visualization is really able to get insight into the data and can it be used to predict stuff? |

| | | |
|---|---|---|
| Once Upon a Vis | The bubble chart is very creative and linking it to the other charts are really interesting | How can you adapt this visualization for more books to be included without it overcrowding the graphs? |
| MBTA Ridership Visualization | I like how their map has a zoom function, allowing users to look more closely at specific stops on the MBTA. This is useful because otherwise it gets a little crowded | If you had more time/knew how to implement everything, what would be the one other visualization aspect that you would implement to best improve upon what you have currently? |
| Visualizing the Recipe | I liked their unique idea and how they visualized calories and other elements of food, it is quite interesting | Is there a way to make this more versatile by including cost of ingredients so you can see how much the meal will cost roughly? |
| Exoplanet Visualization | I like how this group implemented both a filtering and brushing and linking to allow users to better interact with the visualization | Were there any trends that you expected to notice but didn't end up finding to be true? |
| Which city should I live in? | This project was unique in how they approached answering the domain problem. They took a top down approach where they first looked at the continent, then the specific cities within that continent, and lastly comparing the cities | Were you able to test your visualizations and see which cities you'd personally be 'best fit' for? Any exciting/surprising insights? |
| Crypto | Pie chart that shows distribution and is really interesting how one takes almost half | Can you adapt this to include more information for people who don't know what crypto is? |
| Bike Allocation Optimization for Citi Bike | Map that shows the most popular starting and ending stations for Citi Bikes | How do you think the different end users you mentioned will use your visualization differently? Are there specific aspects that you believe certain users would find more useful than others? |
| CoOp'd | I found it interesting that they were able to identify these different data biases that may limit the usability/accuracy of their visualization | How would you go about getting rid of/correcting for these biases? |
| Food Desert Visualization | I like how this group's visualization allows users to zoom in and select/deselect different counties. This makes it much more interactive, and was something we had originally wanted to implement but had run into some errors when doing so | You have a lot of intractability which may be a little too complicated for ease of use. What would you get rid of to improve this? |
| Bar Hopping | This visualization is very interesting because it allows you to select specific bars and learn more about that bar by itself without being cluttered by other information | Is there a way to see the different names of bars without having to click on each dot on the map? I think this would be useful in a lot of use cases. |