Group Project Final Report

• the difficulties that you have encountered;

First, we have many interesting ideas about the food topic, for example, the relation between food habit and weather, the relation between fruit’s color and popularity. However, we find it’s very hard for us to find a detailed dataset in a specific aspect.

Second, when we collect some relevant datasets, the format of datasets varies in lots of styles, it’s unlikely to directly use these raw datasets in some visualization tools (e.g. Tableau or XDAT). We have to normalize the dataset by programming to modify the file.

Third, different visualization tools need different formats of data, so if we want to use more than one visualization tools on one dataset, we have to modify the format to fit the different tools, respectively.

Fourth, some tools have friendly UI and are easy to visualize a given dataset, others don’t. For example, Google Chart and Rgraph have no UI to use, we have to learn how to use these kind of tools and to code to realize the visualization.

• the different visualization approaches that you have tried;

Tools: Tableau, XDAT, Google Charts, Rgraph, Gephi, D3.js, Python Plot

Algorithm: k-means

• the visualization techniques that are realized in your results;

Tools: Tableau, XDAT, Google Charts, Rgraph, Gephi

Algorithm: k-means

• the explanation for your choice of techniques;

Tableau: Tableau is a powerful visualization tool to visualize data on global map.

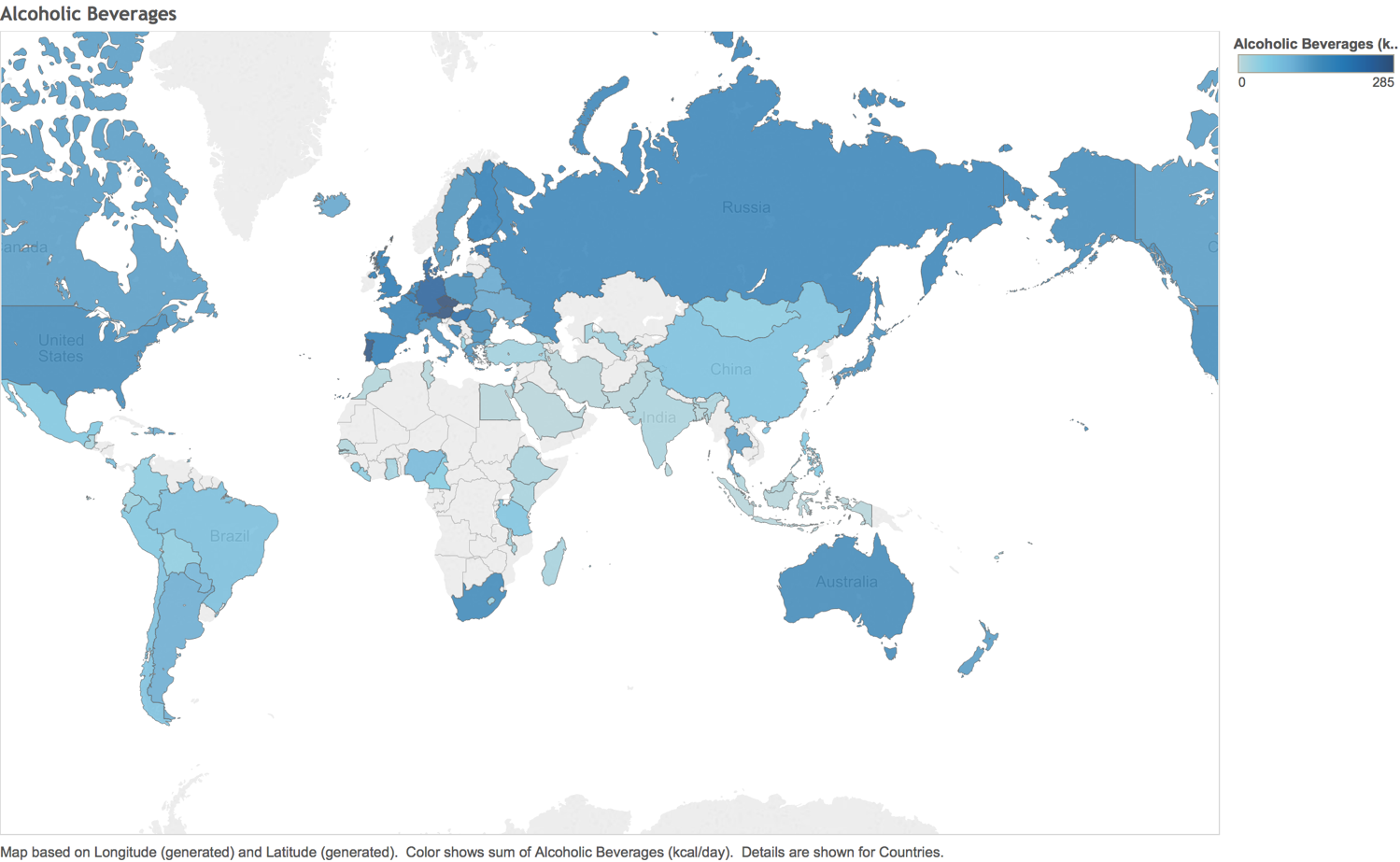
XDAT: XDAT is useful to find the correlation between to factors.

Google Charts: It’s convenient to generate an interactive Graph.

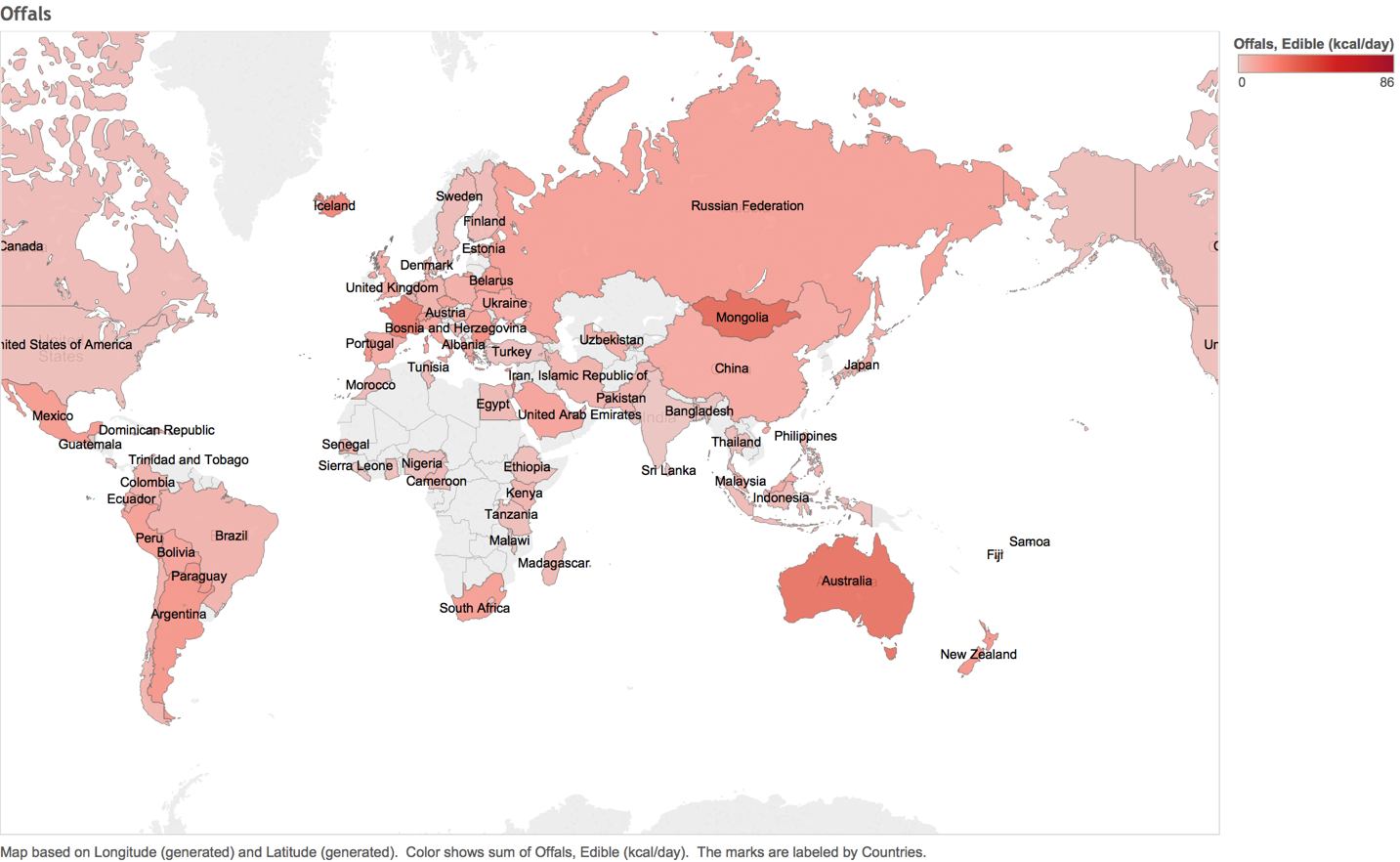
Rgraph: It’s very clear to show the tree structure data.

Gephi: It can be use to mine valuable information in a network structure data.

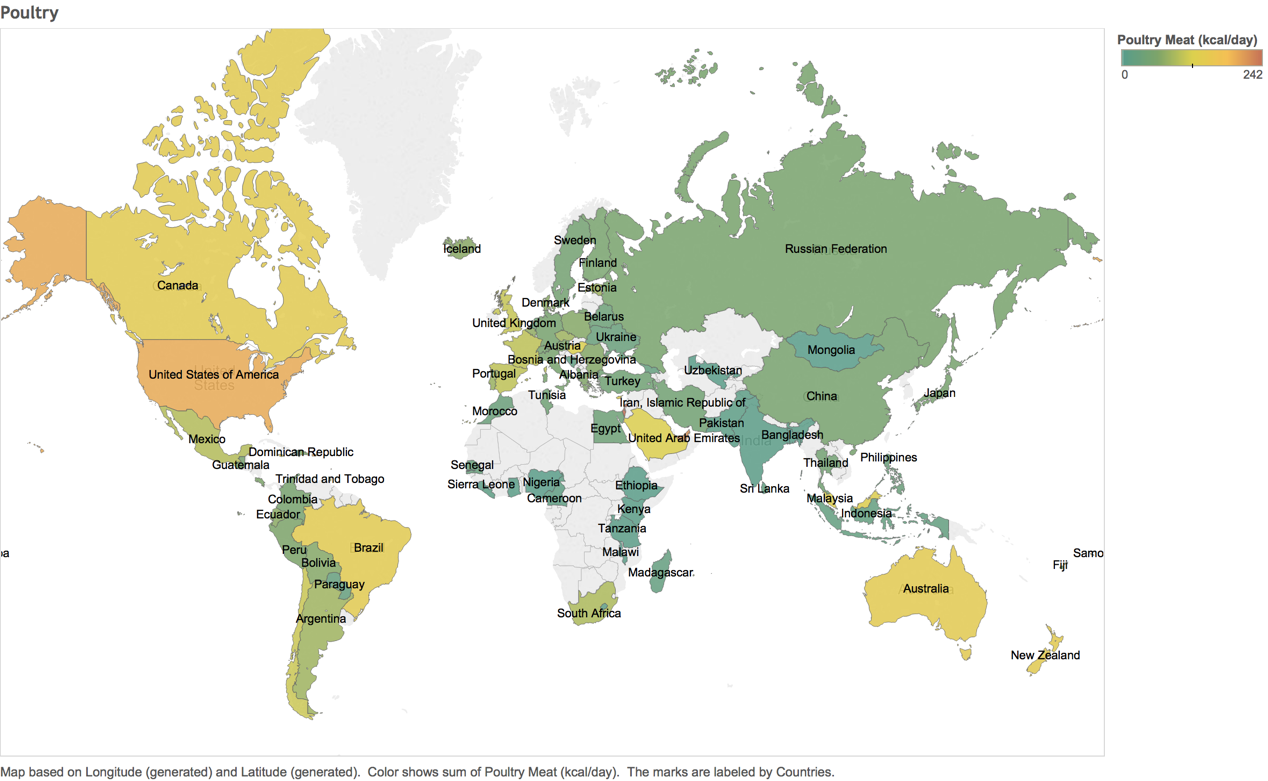
• how your visualization can help answer the queries or questions you have on the data set;

**Question 1**: For a specific kind of food, which county love to eat (or drink) it most?

We visualize the **Alcoholic Beverage** intake on a world map, the blue is darker means the more Alcoholic Beverage drunk. As the graph shows, obviously, European Countries love to drink Alcoholic Beverage most.

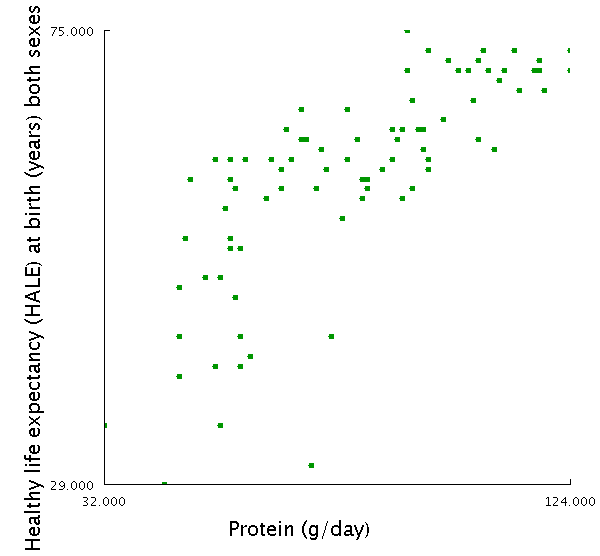
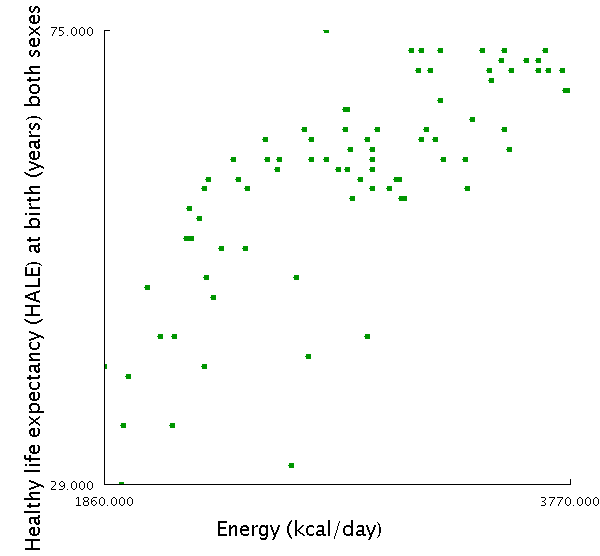


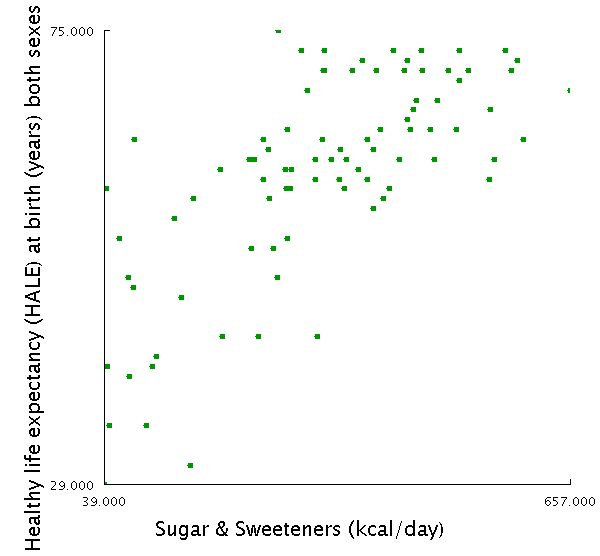
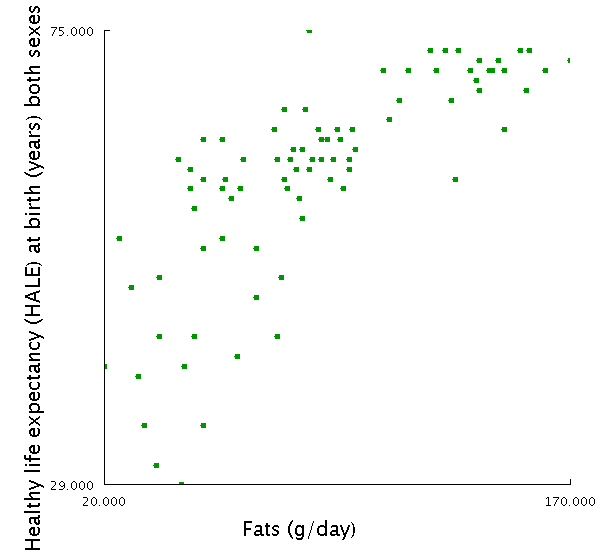
We visualize the **Offal** intake on a world map, the red is darker means the more offal food eaten. As the graph shows, Mongolia and Australia love to eat offal most.



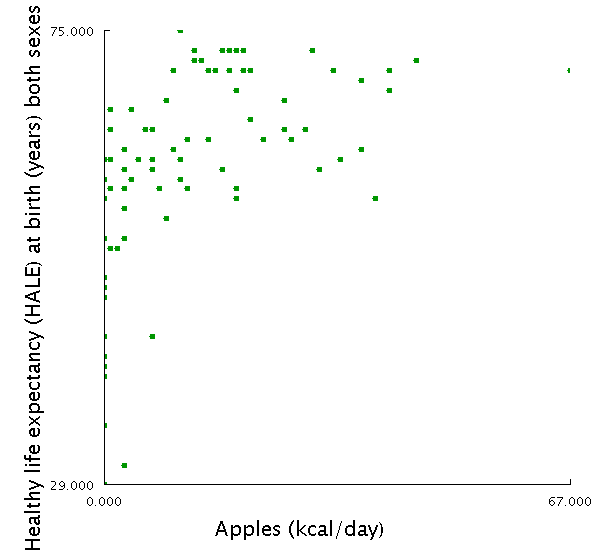
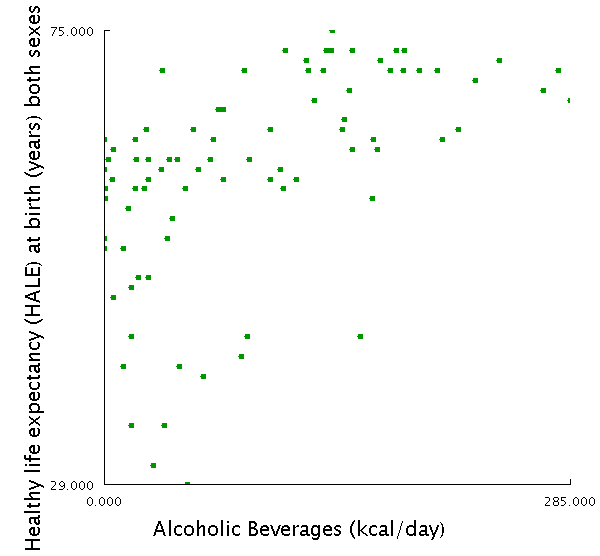
We visualize the **Poultry** intake on a world map, the color is hotter means the more Poultry eaten. As the graph shows, USA loves to eat poultry most.

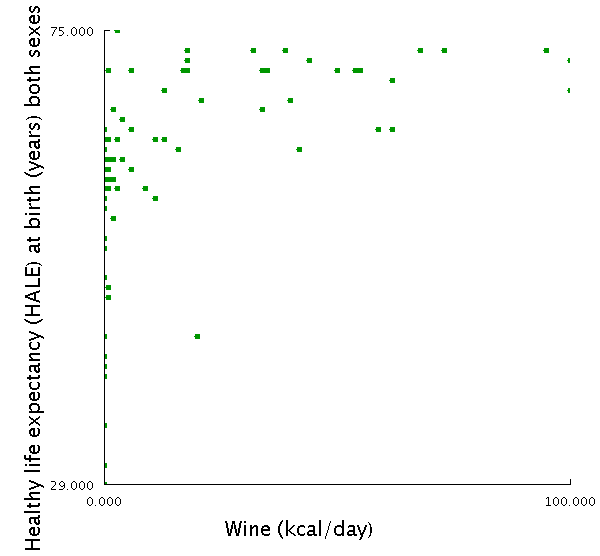
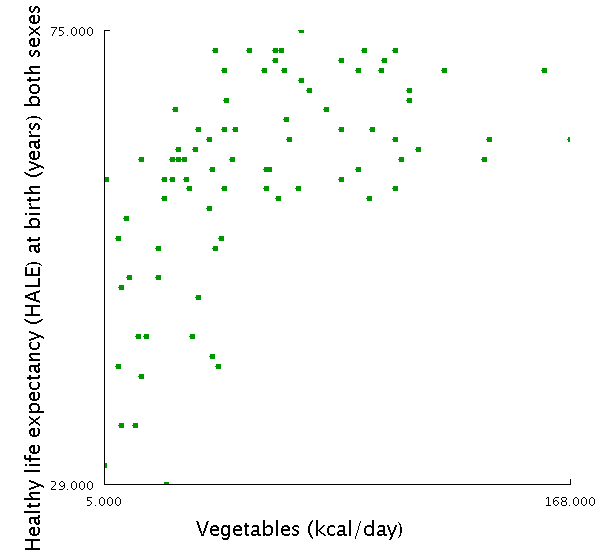
**Question 2**: Is there any apparent correlation between eating habit and health?





We plot the 2D Scatter Charts between 86 Countries’ **Healthy Life Expectancy at birth (HALE)** and some nutrients they intake which we found there’s a correlation. We assume that higher HALE means this Country has a better healthy situation. Although it may not be the root reason, the data shows that **Energy**, **Protein**, **Fats** and **Sugar** intake every day hold a positive correlation between HALE, respectively.



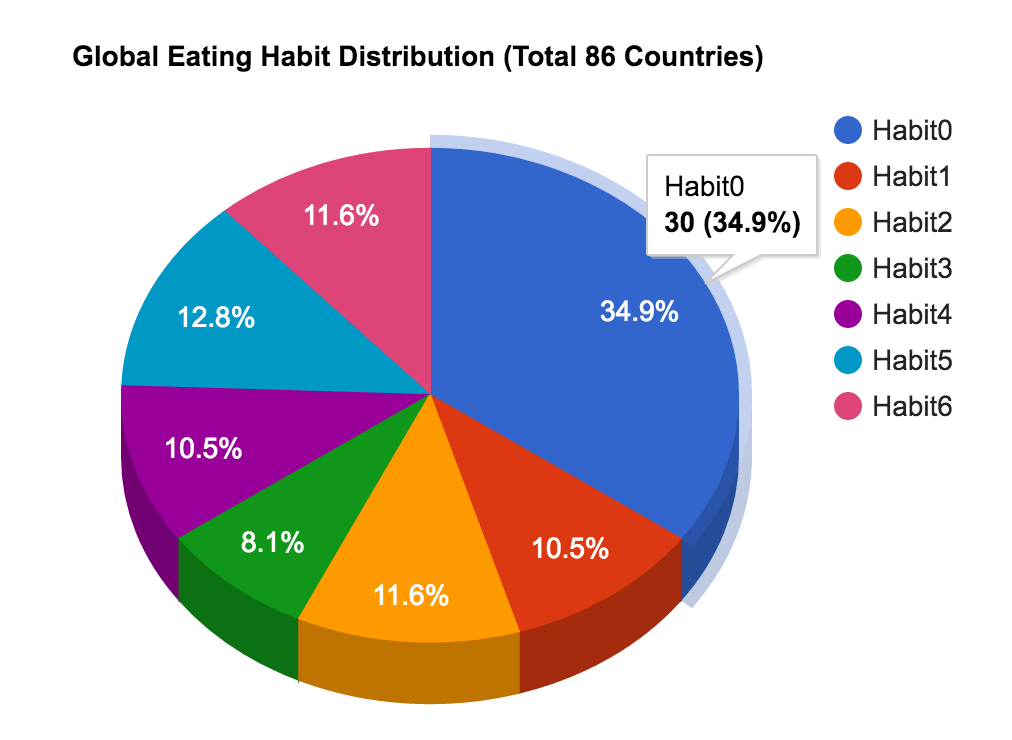


Then, we plot the 2D Scatter Charts between Countries’ **HALE** and 4 specific kind of food, which we usually associate with health, Alcoholic Beverage, Apple, Vegetables, Wine. Violating to our common sense, it shows that the countries drinking more **Alcoholic Beverage** and **Wine** are more likely to have higher **HALE**. However, as the saying goes, an apple a day, keep doctor away, the higher **Apple** intake means higher HALE, as well as **vegetables**.

**Question 3**: Is some stereotype in our mind correct? (e.g. Alcohol is bad for health)

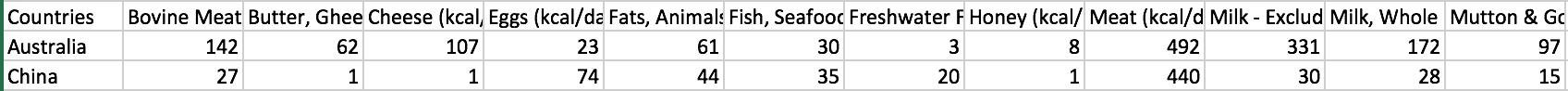
Mentioned in Question 2. Violating to our common sense, it shows that the countries drinking more **Alcoholic Beverage** and **Wine** are more likely to have higher **HALE**.

**Question 4**: What’s every countries eating habits? Is there any similarity among countries, around the world?



We tried to cluster the 86 countries into 7 clusters, according to their different eating habits. We do the clustering by this way:

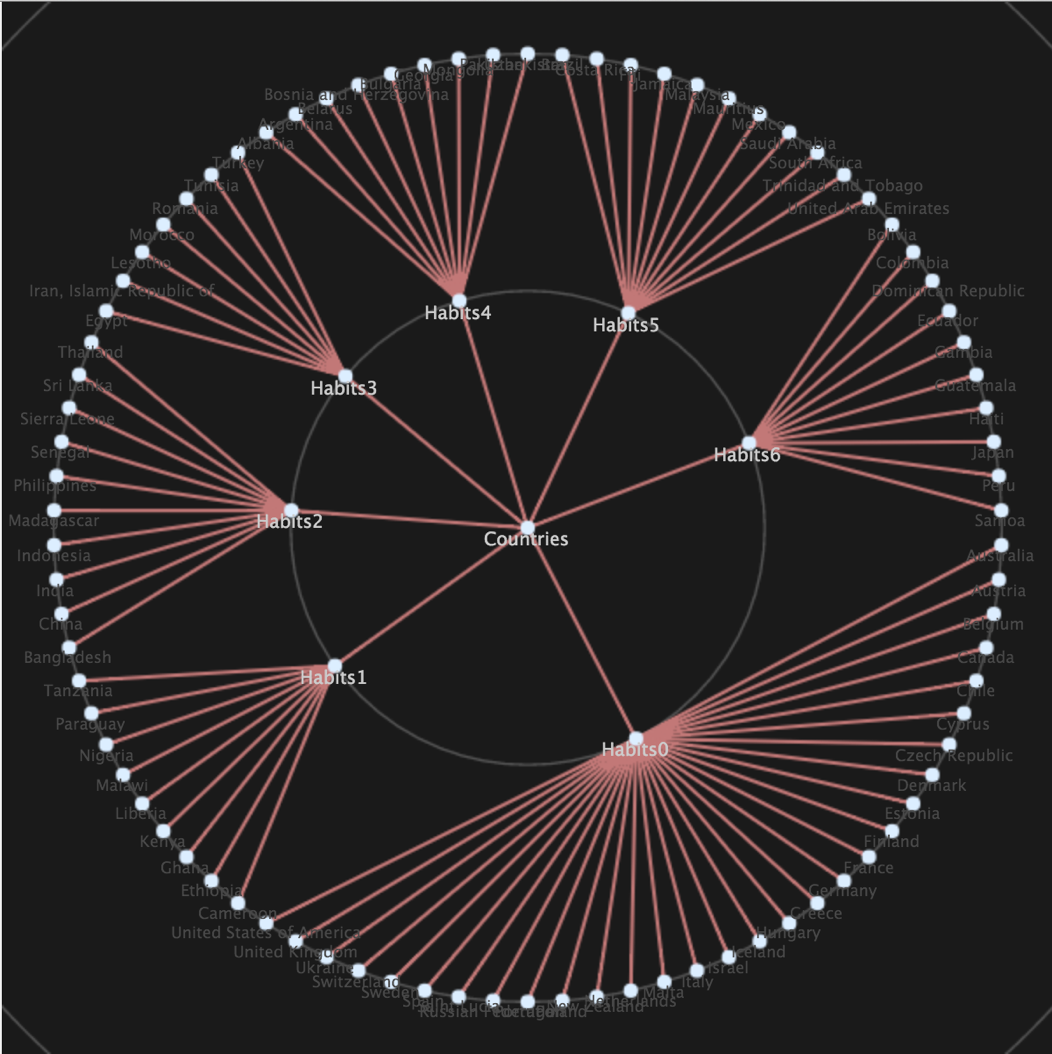
1. we take 43 kinds of food as features, and for each country, there’s a feature vector. For example:

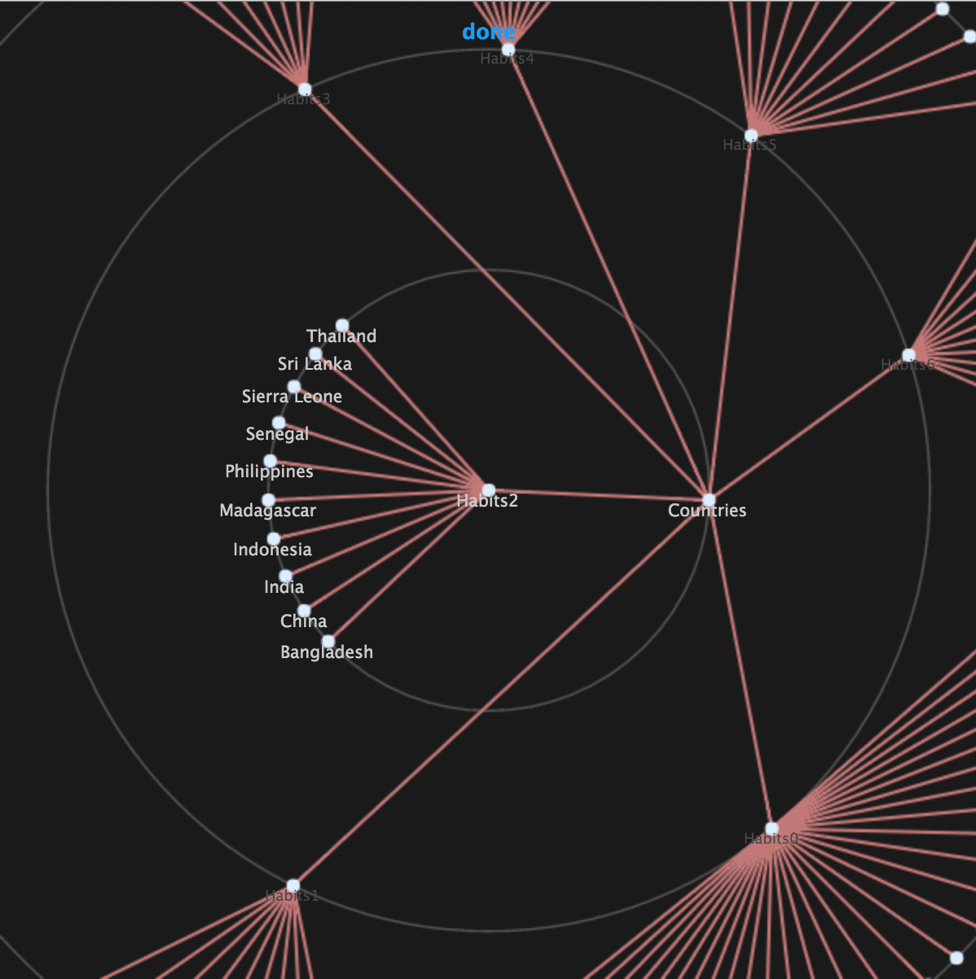


The vector of Australia is [142, 62, 107, 23, 61, 30 ……] (43 dimensions)

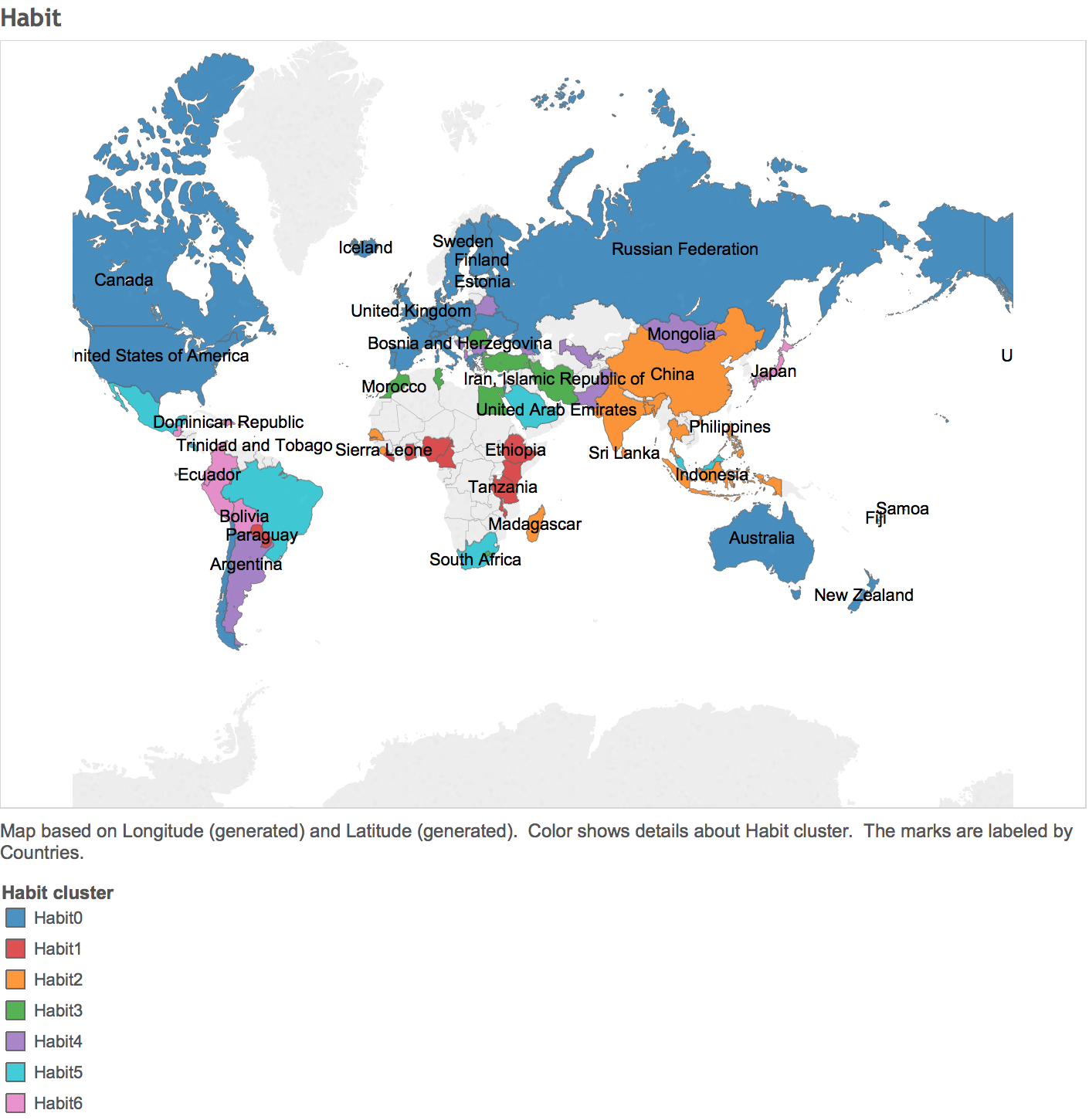
The vector of China is [27, 1, 1, 74, 44, 35 ……] (43 dimensions)

2. Use the k-means algorithm to cluster all 86 vectors into 7 clusters, called Habit0, Habit1, …, Habit6. And this is the result (visualized via Rgraph):

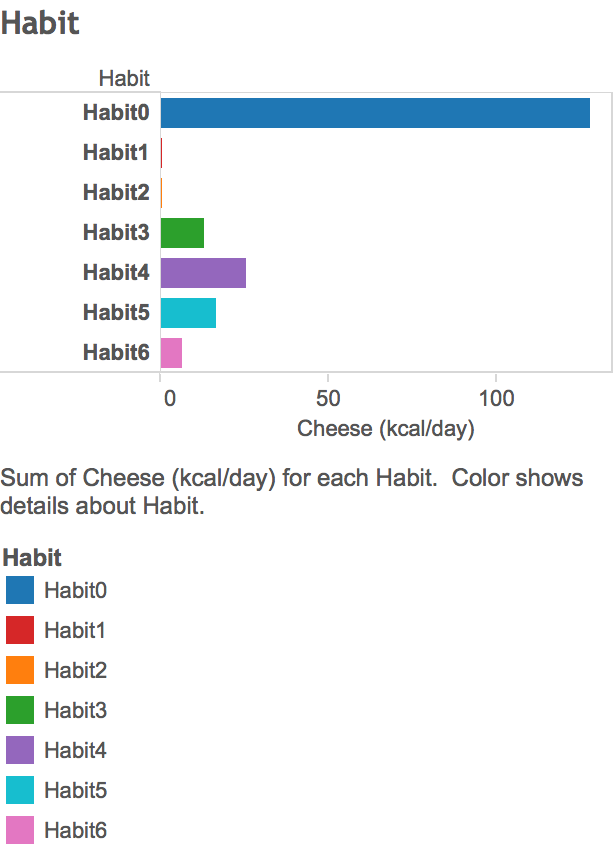
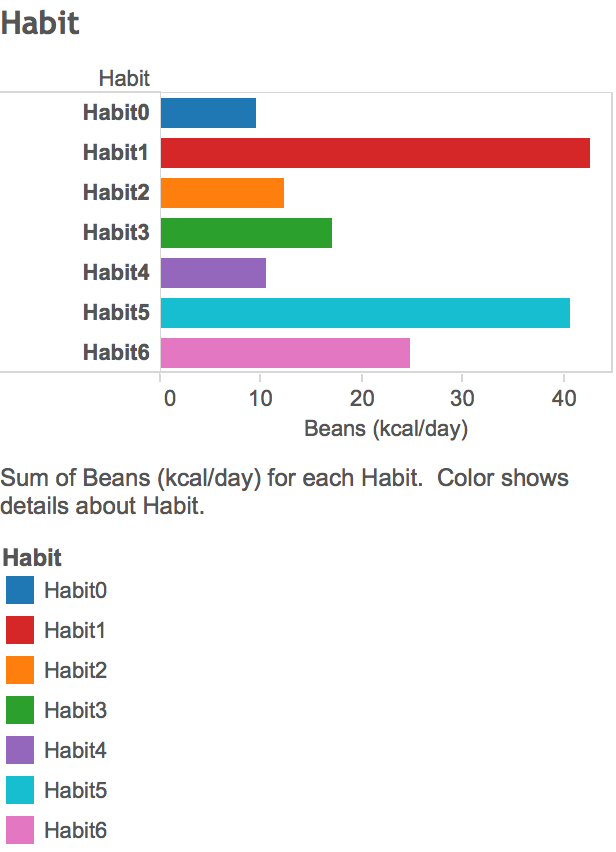


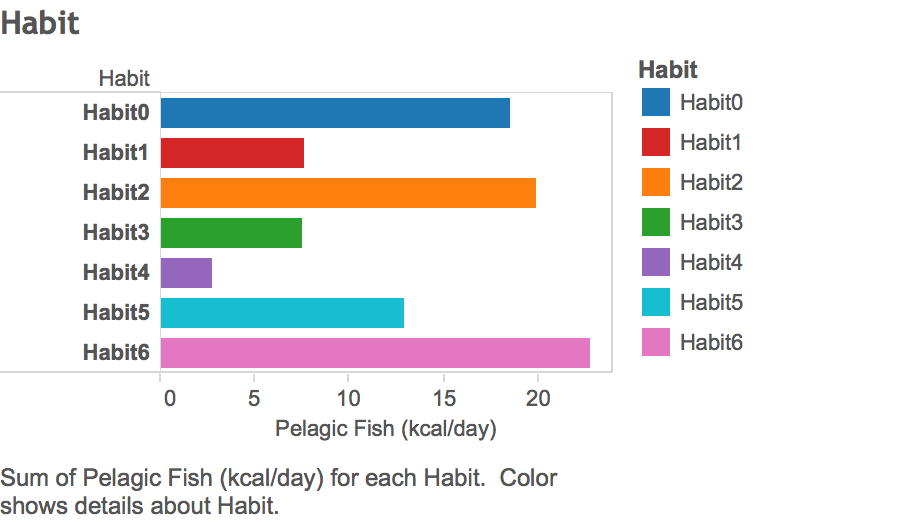
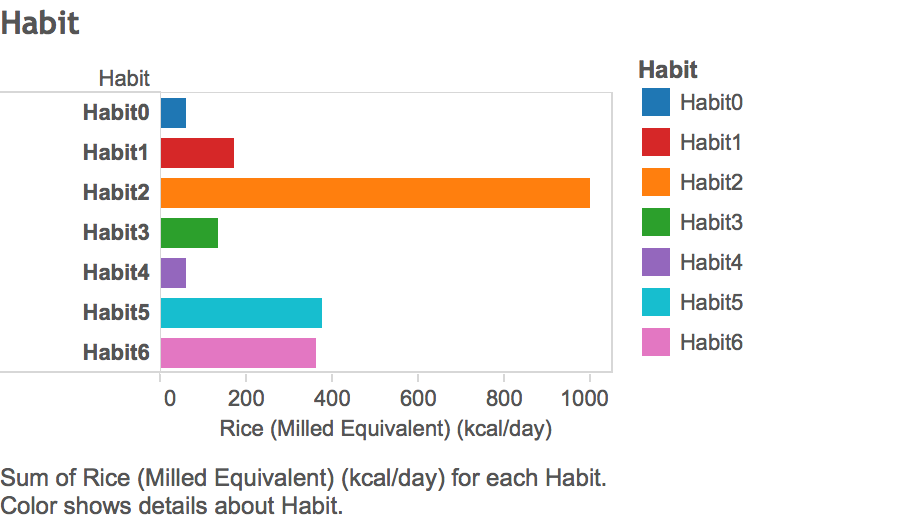


3. To analyze the correlation between eating habits similarity and geography, we visualize the cluster result on Tableau Map View:



4. Analyzing 7 clusters, each cluster has a very distinctive eating habit:





Let’s take 4 examples above: Beans, Cheese, Rice, Pelagic Fish.

The quantity is the average quantity among all countries in this eating habit, respectively. As you can see, Habit0 Countries love to eat Cheese, but don’t like to eat Rice. In this way, we summarize all 7 eating habits’ feature.

--For Countries which hold the Eating **Habit0**,

They **Love** to eat (or drink): Butter, Cheese, Eggs, Honey, Meat, Milk, Pig meat, Alcoholic Beverage, Apple, Coffee, Nuts, Olive Oil, Potatoes, Sugar, Vegetable Oils, Wine;

They **Hate** to eat (or drink): Beans, Cereals

--For Countries which hold the Eating **Habit1**,

They **Love** to eat (or drink): Beans, Palm Oil, Pulses, Roots & Tuber, Starchy Roots

They **Hate** to eat (or drink): Butter, Cheese, Eggs, Meat, Pig meat, Poultry Meat, Apples, Coffee, Sugar, Vegetable, Wheat

--For Countries which hold the Eating **Habit2**,

They **Love** to eat (or drink): Freshwater Fish, Coconut Oil, Rice

They **Hate** to eat (or drink): Bovine Meat, Honey, Milk, Mutton, Offal, Fruit, Nuts, Olive Oil, Potatoes, Vegetable, Wine

--For Countries which hold the Eating **Habit3**,

They **Love** to eat (or drink): Vegetable, Cereals, Wheat

They **Hate** to eat (or drink): Alcoholic Beverages, Bananas

--For Countries which hold the Eating **Habit4**,

They **Love** to eat (or drink): Bovine Meat, Milk, Mutton & Goat Meat, Offal

They **Hate** to eat (or drink): Fish, Seafood, Freshwater Fish, Pelagic Fish, Vegetable, Coconut Oil, Palm Oil, Pulses, Rice, Soya bean Oil

--For Countries which hold the Eating **Habit5**,

They **Love** to eat (or drink): Poultry Meat, Soya bean Oil, Sugar

They **Hate** to eat (or drink): Roots & Tuber, Starchy Roots

--For Countries which hold the Eating **Habit6**,

They **Love** to eat (or drink): Seafood, Pelagic Fish, Bananas, Fruits

They **Hate** to eat (or drink): Nothing.

Question 5: Why do some clusters of countries hold similar eating habits? Because of weather, culture or geography?

According to the data analysis and visual analysis, we believe the eating habits is influenced by comprehensive factors.

For example, as for Countries (USA, Canada, European countries, etc.) hold the Habit1, although they are distributed throughout the world, they share the similar culture — Western Culture.

As for Countries (China, Thailand, Indonesia, Philippines, etc.) hold the Habit2, obviously, they share the similar culture – Chinese Culture or Rice Culture. Moreover, they hold similar climate and geography.

As for Countries (Japan, Haiti, Colombia, etc.) hold the Habit6, they love to eat fish very much. As the map shows, they all close to sea or ocean, so it’s natural for them to love fish, this is a typically geographic factor.

Question 6: What’s the similarity among many kinds of Fruits or Vegetables? How similar they are? In terms of nutrient.

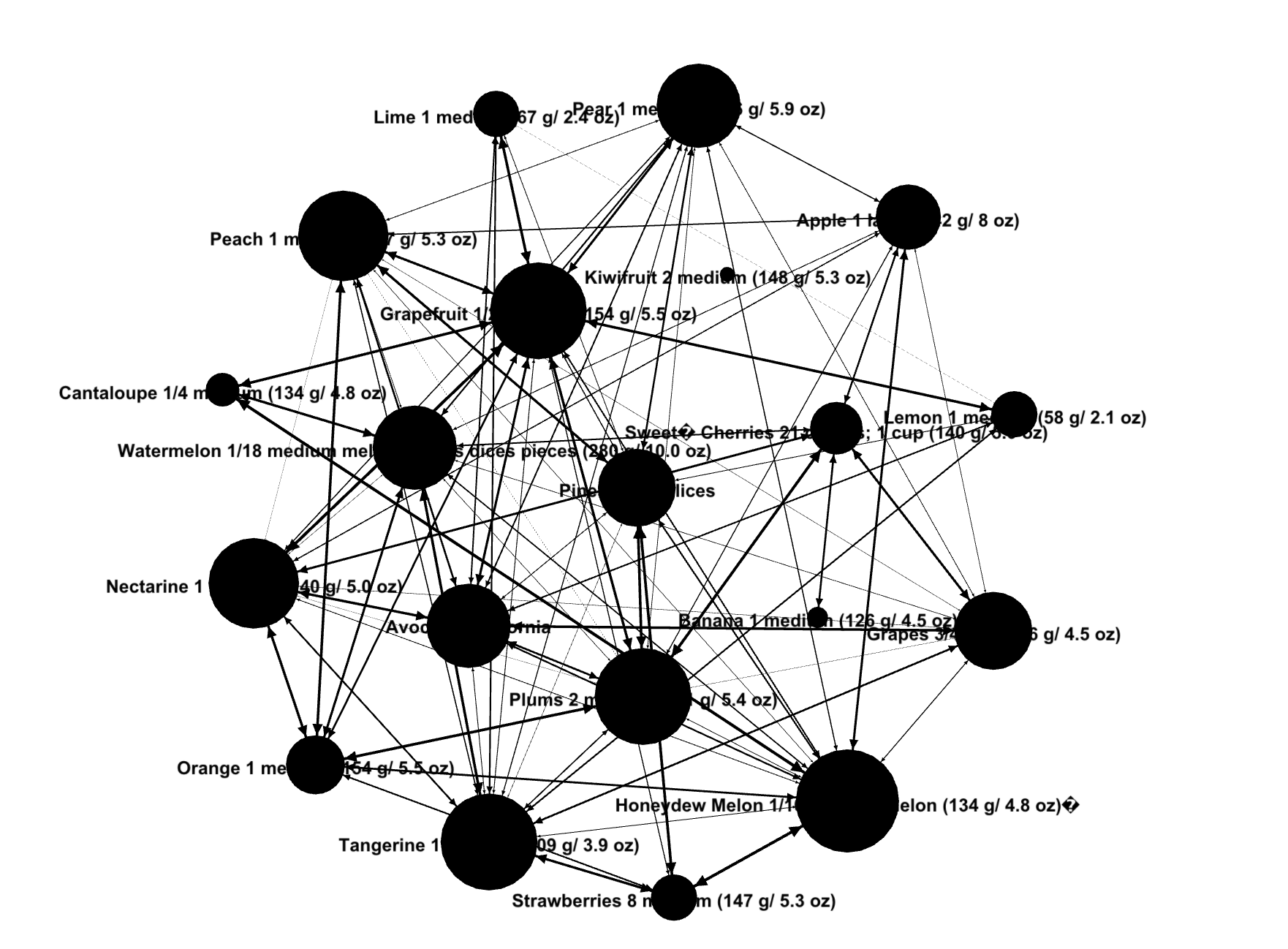
We generate 2 visualized graph. Let’s explain what does these two graph mean.

Take the Fruit Similarity Map as an example. First, we consider one kind of nutrient as a feature, such as Fats, Sodium, Potassium, Carb, Vitamin A, Vitamin B, etc. Then, we compute the Euclidean Distance between every two fruits. If the Euclidean Distance is less than a given threshold (say one-third of average), we regard these two fruits as similar fruits, and the Euclidean Distance is their Similarity Value.

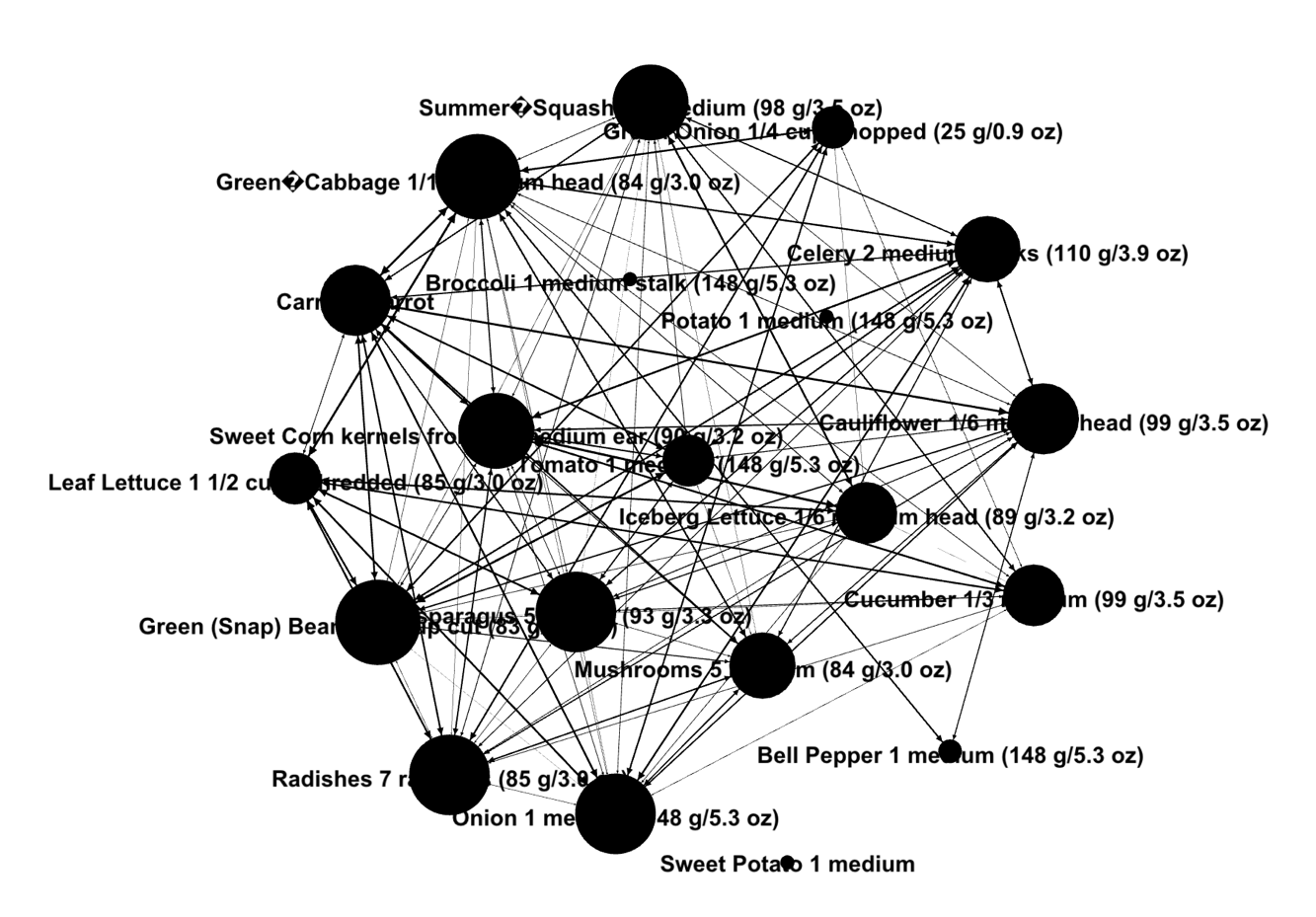
Let’s look at the Graph above, the thicker and darker edge means Larger Similarity Value. And the size of a node is positive correctional to its degree.

As the graph shows, the Grape’s nutrient is similar to many other fruits, and we can say its nutrient composition is “common”. While, Lime is similar to less other fruit, so it’s reasonable to say Lime’s nutrient composition is “unique”.

We analyze the Vegetable in the same way.



Fruits Similarity Map



Vegetables Similarity Map

• any new insights on the data set that you found using your visualization;

1. Obesity is less relevant to Fats intake, but more relevant to Annual insolation.

2. It is a positive correlation between HALE and Energy, Protein, Fats, Animal Products, Sugar.

3. There is no significant correlation between HALE and Alcohol, Fruit, Vegetable.

4. There is no significant correlation between cholesterol and egg, seafood, offal

5. Clustered Food Habit summary.

• any visualization or functionalities that you think should have included, but have not

done so.

D3.js to realize some novel but intuitive visualization.