

# COMP 7507 Project: What's up, US Foodies

Group E: Yi Jue 3035420352, Zhu Qingyu 3035419080, Shi Kexin 3035455278

## Introduction

What kind of questions may “bother” most people around us every day? Though it may not be that bothering, but we do spent countless times thinking what to eat. Thanks to the improvement of material quality and living standards, people tend to focus more on what we are eating today. As a group made up of food lovers, we are trying to explore the food culture of the United States.

In this project, we started from the popular restaurants in America, to look at what kind of food and restaurant are preferred by Americans in different dimensions. We also dig into the eating habits in respect to times. The data we used in this part is from “Yelp”, a popular customer-review-collection platform.

Then by looking at the data from USDA, we then explore where does food comes from. We derive two maps to present the countries that supply the most imported food and the states in U.S. produces the most homegrown food. We also discover the volume and value tendency of these foods as time goes by.

As food type has been mentioned, we cannot escape from talking about whether the food is healthy or not. Data of American daily calories intake from USDA could show us clearly the tendency on calories intake movement from 1970 to 2010. We also presented the detailed daily calories intake of different food type. Based on all these data, we could draw a conclusion on the changing life concepts nutrition.

Most of our visualization is done by using Tableau, and some of them are built by applying D3.js and XDAT. There are many kinds of visualization methods used in this project. For chart types, we have line charts, bar charts, pie charts, tree maps, maps, and so on. Highly integrated information is achieved by representing data in different format, such as the color, the size, the axis, and even animation that are varying with time series. In order to highlight the effect of different visualization, we may manipulate the data in different forms sometimes.

Now let's go on our food exploring trip!

## Part 1: preference on food and restaurant

At the very beginning, to give our audience a general picture of such kind of information, we start from a map view showing the restaurant numbers.

As you could see, each city is corresponds to a circle, where its color represents the states it lies in and its size shows the number of restaurants it owns.



Figure 1

So what are the features of these restaurants? Here we used a tree map to list the most common food and restaurant in US. In this tree map, the larger the rectangle is, the more restaurants of this type that US has. It is clear to see that the top four are "Nightlife", "Bars", "Americans (New)", and "Americans (Traditional)". It is not so hard to imagine the chasing of nightlife among Americans.

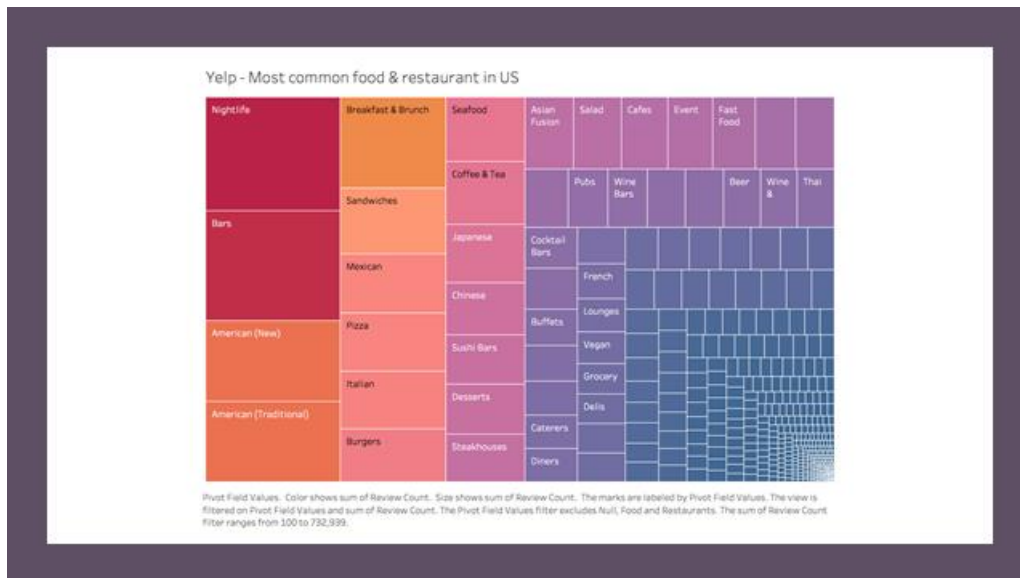


Figure 2

Here we have another form, text visualization, to display the common food & restaurant in US. Even though the differences are not clear here, we could still pick out the top 10 common foods & restaurant in this picture. And also, by using text visualization, more text information is kept here.



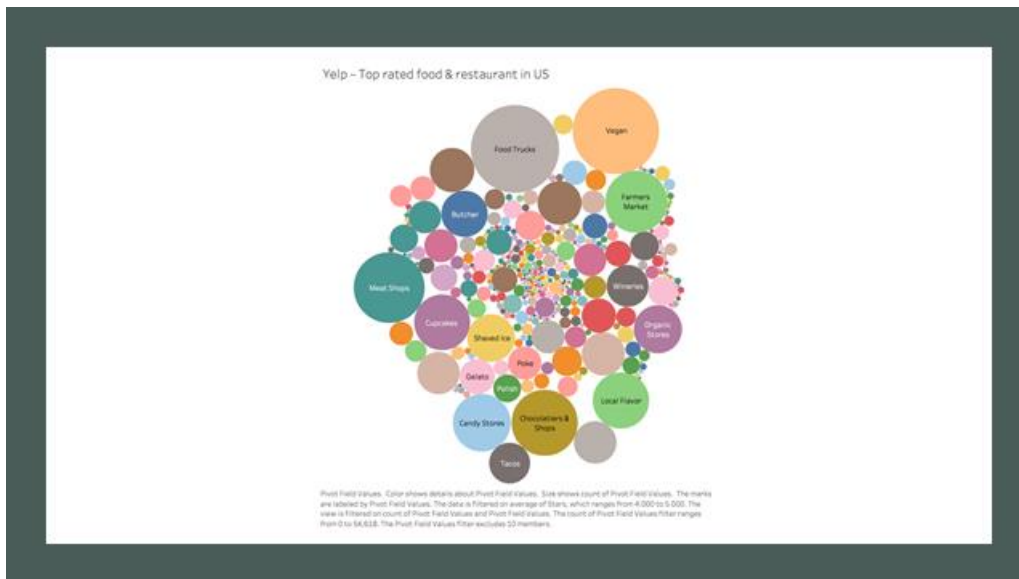


Figure 5

Then we move on to locate inside the food type instead of the restaurant type. We use bar chart to list the top 23 rated foods & restaurant in US. As you can see, “shaved ice”, “Cheese shops”, “poke”, “Gelato” are listed in the top four here. No doubt sweets make people happier, generally speaking.

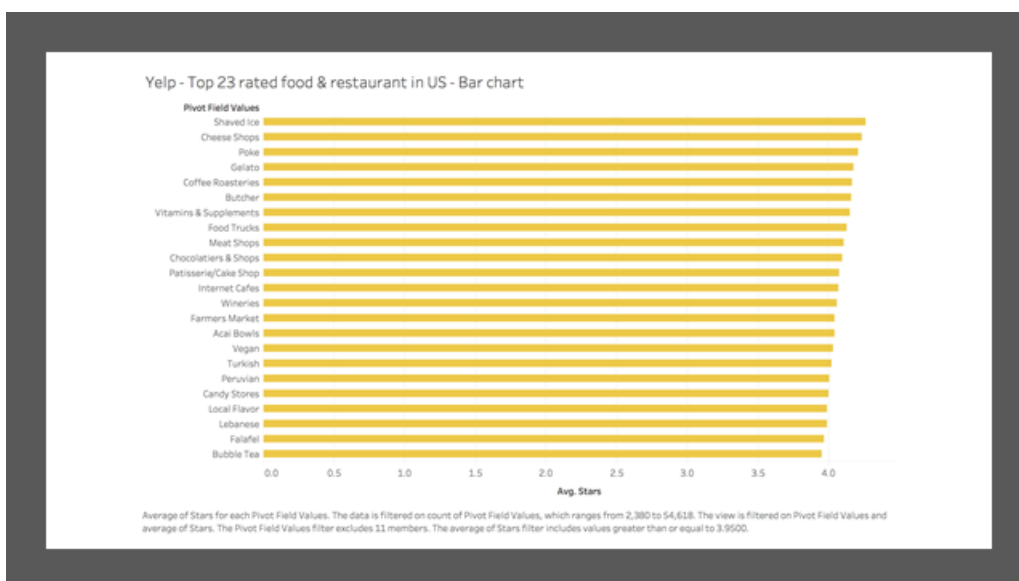


Figure 6

For an extra information, we also give a picture about how each state differs when their people rate food & restaurant. Since the difference of populations for each state may introduce noise, we first normalized the data, so that each state’s rating distribution is based on its own population, in other words, the added up area for that state should be all equal among all other states. Then we find out average ratings are above 3.5 generally. State Alaska generally gives more fixed ratings, ‘2.5’ ‘3.5’ ‘4.5’ and ‘5.0’ stars only.



Figure 7

Up to now, we have already be sure about the preferred food & restaurant type in US. To give more about eating habits, we do the following work.

Firstly, we plot a bar chart about when do American go to restaurants during a day. What surprises us in the beginning is that the US people seem to hang out for food at very late night, during 12pm to 4am. Then we look back to the 'Most common food/restaurant in US' tree map, we got to understand the phenomenon: they are probably having food or drinks in the nightclub. And in this picture you could also see that there is little difference among days from Sunday to Monday at a specific time.

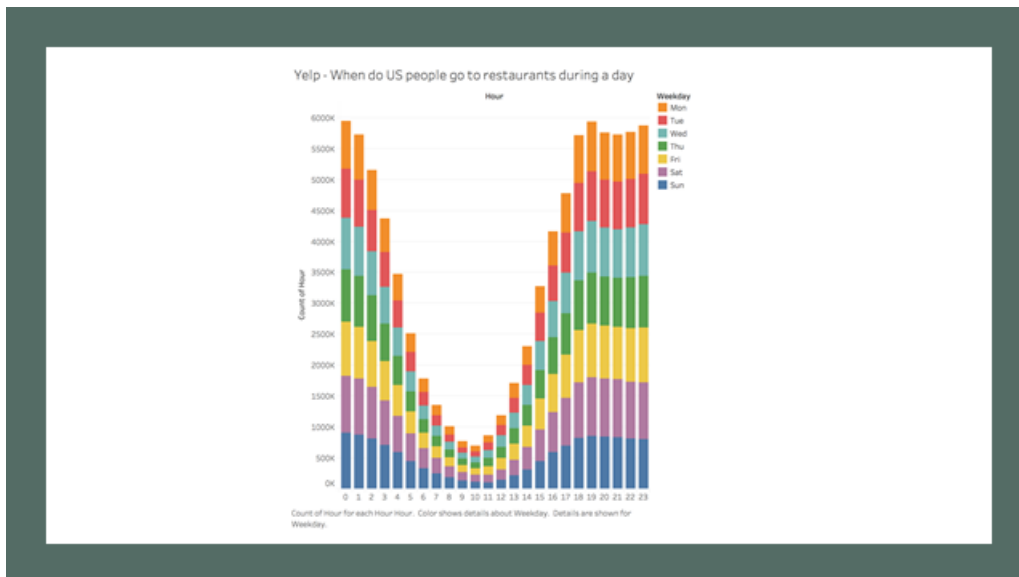


Figure 8

Then we begin to look at what food & restaurant do American go for in the morning. In the tree map, "Sushi bars", "Diners", "Specialty food", "Salad", "Asian Fusion" and "Chicken Wings" are the popular selections for American.

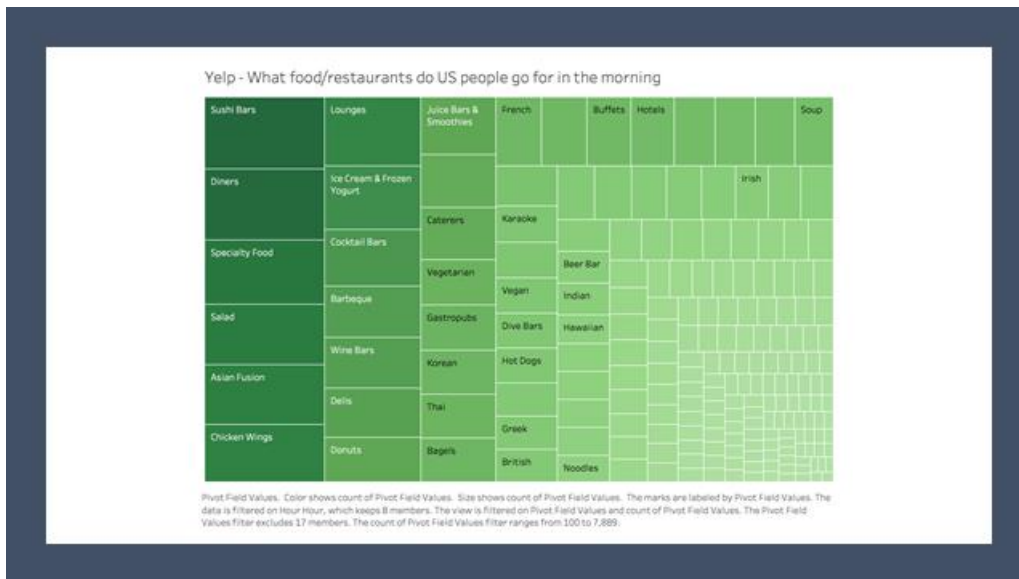


Figure 9

How about afternoon? In this tree map, “Breweries”, “Latin American”, “Food Delivery Services”, “Noodles”, “Tea rooms”, “British”, “Dim Sum” are preferred. Here we can easily pick out “Food Delivery Services”. Could it be an explanation to the low volume in the afternoon in the bar chart? Maybe it could. And we can also get the point that Chinese food culture does influence that of American, such as the “Noodles”, “Tea rooms”, “Dim Sum”.



Figure 10

At last, food & restaurants preferred by American at night is shown below. In the tree map, as you can see, Bars still occupies the top two, or even more. Here we could also see the common food & restaurant “Food trucks”, “Shaved Ice”, “Chocolatier & shops”, “cupcakes” as we have mentioned before.





Figure 11

These 3 three maps are really interesting as they clearly show us how different their preference of food are at different time slots of everyday. We must say, it's not too healthy to have chocolates and cupcakes at night.

Now that we have pointed out the preferred food type in US, we already have an overall idea of what American like to eat most. In fact, Americans enjoys a wide range of food, from Italian food to México food, from Chinese noodles and Japanese sushi to Thailand taste, from Chinese tea to Germany beer. Though, the answer is still simple, Americans like to go to bars and eat fast food (sandwiches and salad) and desserts (shaved ice and cupcakes). Meat such as beef and poke are also their favorite.

According to the result, Americans are really a country of appetite indulgence. If it is true, who serves such a country, or in other word, where does all these food come from? Now let's move on to talking about food supply in US.

## Part 2: food supply

In this part, we start considering the food supply sources. Food supply involves two parts, one is imported product, and another is homegrown product. We will consider these two one by one.

First of all, we plot a picture about the import volume for 14 categories from 1999-2014. In the line chart, it could be clearly seen that except several products (like Nuts and Dairy) that have no significant changes, most food products' import volume got increased. Among all of these increasing ones, "fruits" comes to the first, followed by "cereals and bakery", and then "vegetables oils". However, "Live meat animals" seems to be different to other products because it is decreasing after 2007.

Besides, at the point of 2009, import volume for almost all products went down. We conjectured that the financial crisis of 2007-2008 had a big impact on the food-importing.

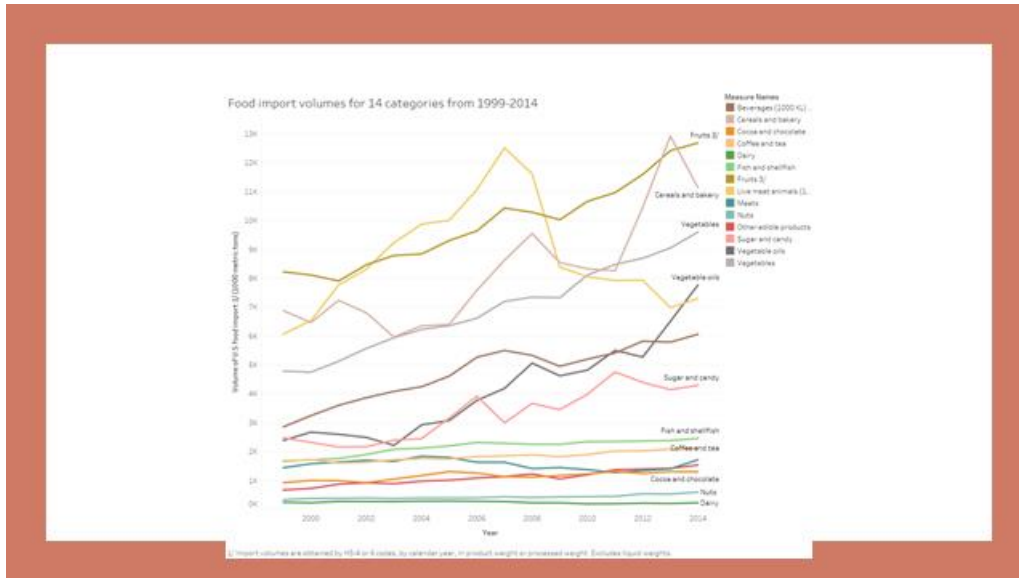


Figure 12

For all these import products, their sources also vary a lot. The map and the bar chart below show the top 64 US food import sources from 1999-2014 based on the value. In the map, the darker area provides more import value. Their specific import values are displayed in the bar chart below. From this group of graphs, you can get the idea that from 1999-2014, most of imported food of US are from Canada and Mexico, who are much closed to United State.

However, in recent years, some Asian countries, like China and India, have significant improvement on import value. These facts may be consistent with the important roles of political and economic factors in food-importing. For example, China becomes a member of WTO in 2001 which signified China's deeper integration into the world economy.



Figure 13

Then, we plot a dashboard about the agricultural production value per state in US from 1999-2016. Generally two major colors are used, red and green, to represent animal products and crops respectively. Again, the darker the color, the more production value the state contributed. Even with a glance, we could find that



California provides the majority of crops, and Texas provides the most animal products. And most crops sources lie in the west coast and northern part of US, while the animal sources are mostly distributed in the middle part and east coast of US. Also, another clue we can get from this visualization work is California is also the NO.1 state regarding to the value of agricultural production in US in 2016.

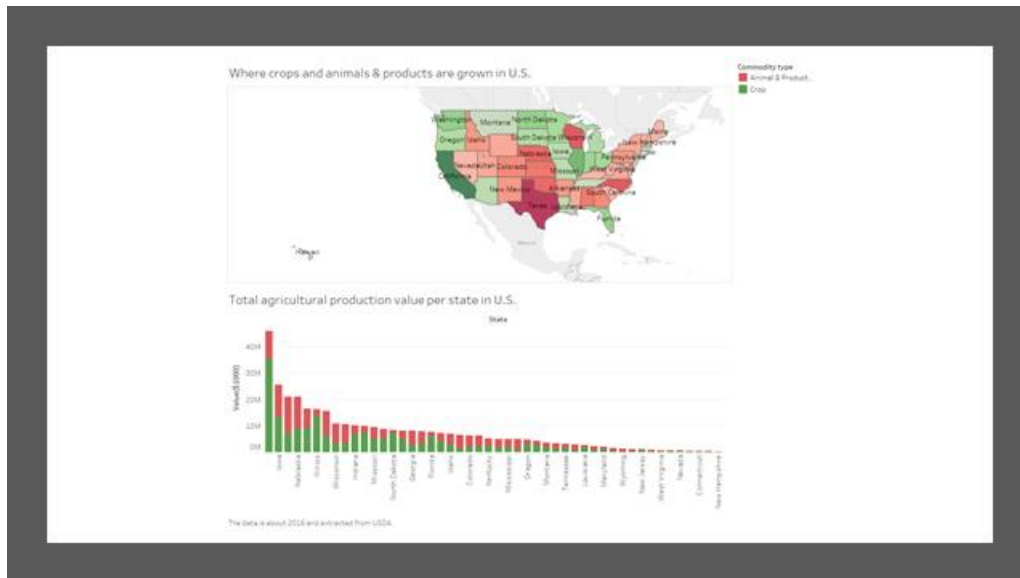


Figure 14

The more specific visualization from the point of homegrown product income at each food type (and subtypes) is given below. This chart follows the color format used in the previous one, with red representing animal products and green representing crops. To show the hierarchy structure of the food types, a solar structure is used here. By tapping a specific area in the solar chart, you could clearly see its corresponding tendency on the right hand side with a line chart. This group could highly integrate the information into a tiny area, which increases the efficiency of the space used. We can see that,

1. Corn is king! How much money the US's harvests bring in, and the conclusion is unmistakable: corn, corn, corn.
2. Contrary to "Live meat animals" import volume, which is keeping decreasing after 2007, "Meat animals" self-production value is keeping going up from 2009 and 2015 and is responsible for about a half of animal & products income in 2016. We can conclude that after financial crisis in 2007-2008, the "meat animals" in U.S. are almost from their own country.

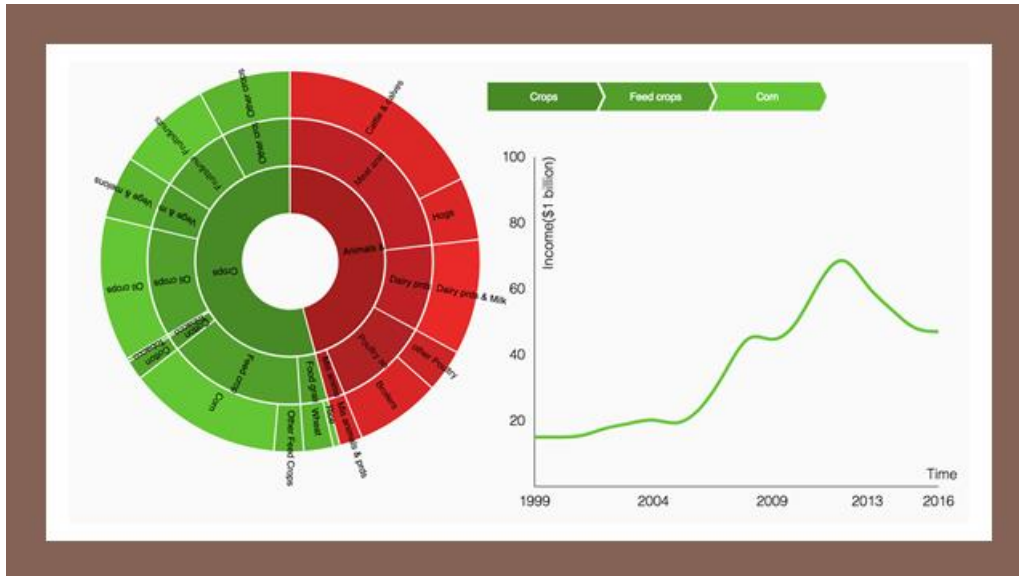


Figure 15

In a nutshell, there are several points in this part:

1. American food is mostly from neighboring countries such as Canada and Mexico
2. Import food types are concentrated in fruits, cereals and bakery, vegetables, vegetable oils, live animal meat, beverages, and sugars. Live animal meat imported value is decreasing these years.
3. The largest crop production state is California and the largest animal production state is Texas.
4. Crop production states are mostly near the western coast or in the north, and the animal production states are major gathered in the south of US.
5. US itself is the sources of large amount of corns, cattle, and dairy.
6. These years US sales on oil crops and fruits are gradually increasing.

Up to now we have known what Americans do like to eat, where the sources of these foods are. Then it is time to consider the healthiness of these eating habits.

## Part 3: nutrition

In this part, we are considering whether these food consumed by American are healthy or not.

To make it easier to understand such kind of abstract topic, we are from the nutrition's perspective, by looking at the calories intake for an American's daily life.

First of all, we plot a group of charts about the total calories intake per food type per American a day from 1970-2009. There are three charts, a pie chart showing the percentage of calories from each food type, a bar chart showing the detailed amount of calories intake from each type, and another bar chart with only 1 bar to show the total calories intake based on such a composition. All food are classified into 7 types: "meat, eggs, and nuts", "flour and cereal products", "sum of added fats and oil", "sugar and sweeteners", "dairy", "vegetables", and "fruit". By locating the "year" attribute, the daily calories intake for an average American is displayed.

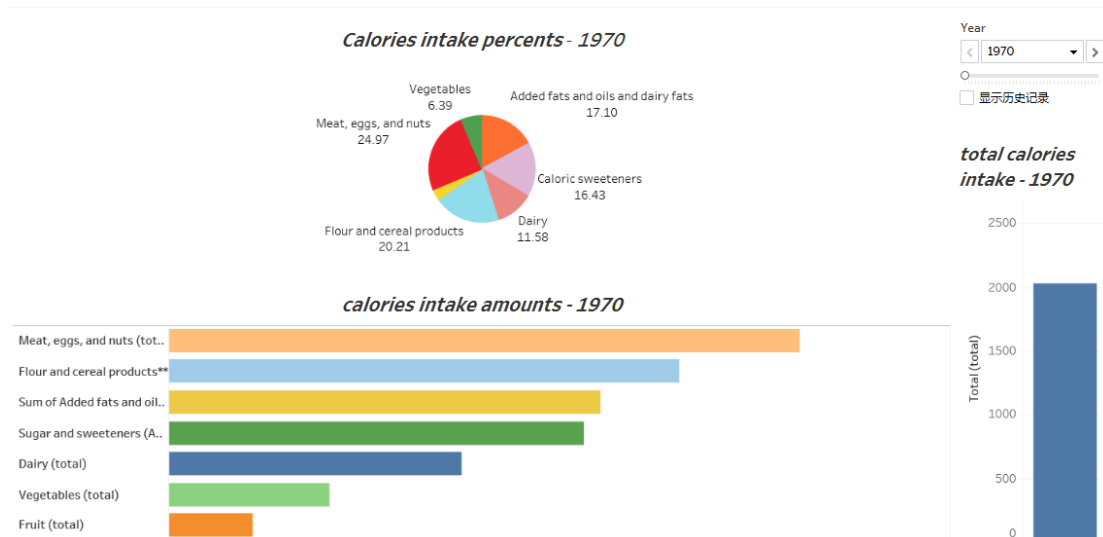


Figure 16

Then the following charts are giving the subtype information of each bar in the first bar chart "Calories intake amounts". For all of them, a common format is used, that is with x-axis separating the "year" information, and y-axis showing the calories amount. We could look at them one by one.

*a. Calories intake by Meat group: (figure 17)*

Here, for simplicity, we merged the meat group with some subsidiary agricultural products, like eggs and nuts. One obvious tendency is that the amount calories intake from red meat is decreasing while that from poultry is in an opposite direction. Also, the calories intake from nuts is slightly increasing, while the others are almost keeping static.

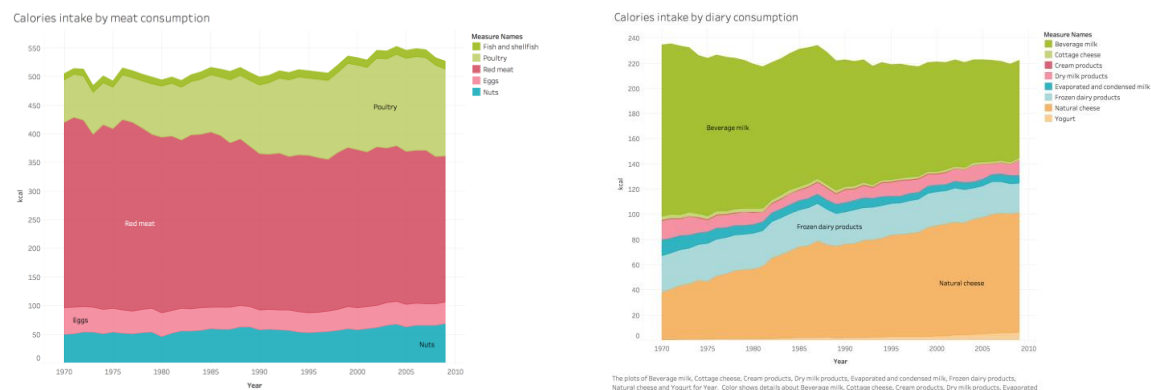
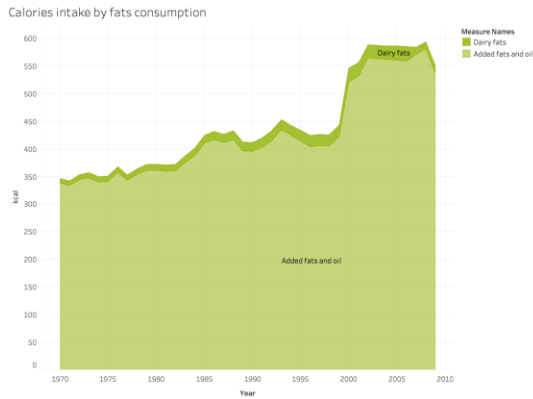


Figure 17

Figure 18

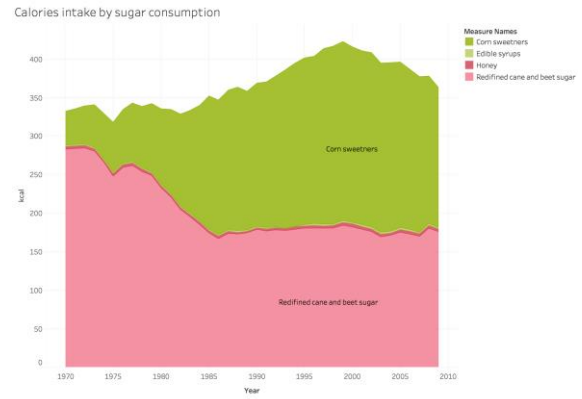
*b. Calories intake by grains group: (figure 18)*

As the major calories intake source in 2009, this chart tells us that it is not happened to be so. With the top 2 composition of grains group, wheat flour and corn products both growing, the total tendency is definitely growing, too. Barley products, which is regarded as one of the most healthy food by nutritionists of America, is still a minority in this group, though.



The plots of Dairy fats and Added fats and oil for Year. Color shows details about Dairy fats and Added fats and oil. The marks are labeled by Dairy fats and Added fats and oil.

Figure 19



The plots of Corn sweeteners, Edible syrups, Honey and Redefined cane and beet sugar for Year. Color shows details about Corn sweeteners, Edible syrups, Honey and Redefined cane and beet sugar. The marks are labeled by Corn sweeteners, Edible syrups, Honey and Redefined cane and beet sugar. Details are shown for Corn sweeteners, Edible syrups, Honey and Redefined cane and beet sugar.

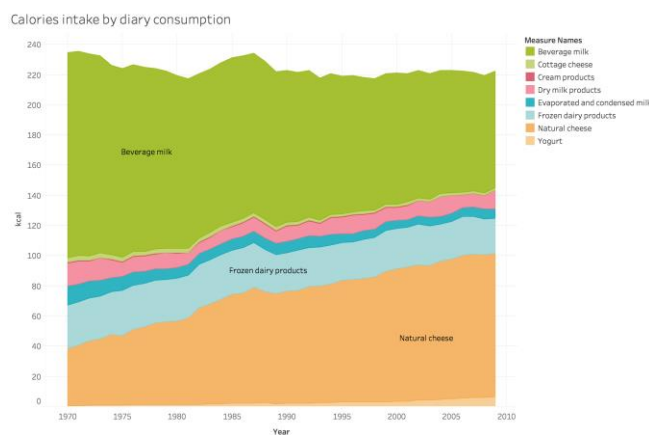
Figure 20

*c. Calories intake by fats group: (figure 19)*

Here two fat sources are listed, “added fats and oil”, and “diary fats”. One notable feature of this chart is a sudden rise in around 2000, where the “added fats and oil” surprisingly jumped up over 150 kcal units. Here, “dairy fats” is representing the fats human obtained from dairy, “added fats” includes butter and other animal fat. It is true that these years’ nutritionists are advocating the healthiness values of added fats and oils from vegetables due to their rich of Monounsaturated Fatty Acids. In this case, the data here could not split the healthy “oils” from those animal fats. However, according to the total amount, we could say that an American in 2010 is eating much more fat than another American living in 1970.

*d. Calories intake by sugar group: (figure 20)*

As a nation whose desserts is usually too sweet for most Asian people, there is no doubt that the calories from sugar are still at an amazing amount. The interestingness in this chart is the different trend of the 2 major compositions, where “corn sweeteners” is increasing while “redefined cane and beet sugar” is decreasing. From around 1998, we could see the total amount of calories from sugar is decreasing. Maybe it is an evidence to prove that the eating habits of American got influenced by nutrition.



The plots of Beverage milk, Cottage cheese, Cream products, Dry milk products, Evaporated and condensed milk, Frozen dairy products, Natural cheese and Yogurt for Year. Color shows details about Beverage milk, Cottage cheese, Cream products, Dry milk products, Evaporated and condensed milk, Frozen dairy products, Natural cheese and Yogurt. The marks are labeled by Beverage milk, Cottage cheese, Cream products, Dry milk products, Evaporated and condensed milk, Frozen dairy products, Natural cheese and Yogurt.

Figure 21

*e. Calories intake by dairy group: (figure 21)*

In this chart, you could see that, “Natural cheese” is providing more and more calories these years in contract with “beverage milk”. One thing need to notice is that, it is widely accepted that natural cheese could provide more nutrition such as protein

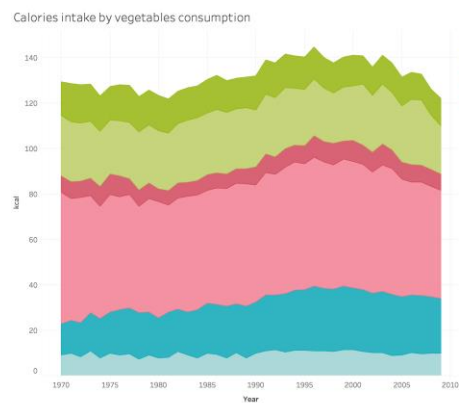
and calcium than beverage milk. Except that, natural cheese is easier to store and keep quality. Apart from these two, if you look at the bottom, you would catch the gradually increasing “yogurt”. In fact it is within our imagination, because most healthy diet involves yogurt nowadays.

*f. Calories intake by fruit group: (figure 22)*

Look at this volatile graph! We could only say that the calories provide by fruits did have a climb from 1970 to 2009, especially the fresh fruits, juice, and frozen fruits. Apart from them, the canned fruits are obviously decreasing and the dry fruits seem unchanged. Here, I have to point out that canned products are regarded as instant products that lack nutrition. This may also be the result from people’s changing attitudes on nutrition.

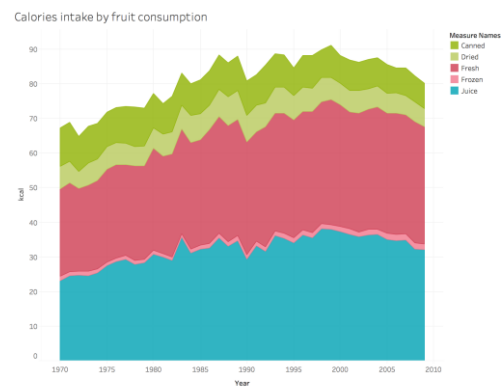
*g. Calories intake by vegetables group: (figure 23)*

Very similar with the fruit group, canned vegetables are providing less and less calories, while “frozen vegetables” are providing more and more calories.



The plots of Canned vegetables, Chips and shoestring potatoes, Dehydrated vegetables, Fresh vegetables, Frozen vegetables and Legumes for Year. Color shows details about Canned vegetables, Chips and shoestring potatoes, Dehydrated vegetables, Fresh vegetables, Frozen vegetables and Legumes.

Figure 22



The plots of Canned, Dried, Fresh, Frozen and Juice for Year. Color shows details about Canned, Dried, Fresh, Frozen and Juice.

Figure 23

So, is an American’s diet healthy in 2010? Probability we could not derive such a conclusion from these data above. In order to make supplement and assist analyzing, we added another database. This new dataset is about the food intake by different groups of persons, like children, adults, lower incomes, higher incomes, from different sources, such as “at home”, “fast food”, “restaurant”, “school”, “other”. The counting is based on the average daily food amount for one person from 2007-2010. This dataset also gives us a recommendation for food intake, by counting the expected food unit per 1000 calories total intake. In the charts below, you could clearly see that, no matter which group, the solid fats and added sugar are much more than their recommended volumes. On the other hand, grains, dairy, fruits and vegetables are much lower than expected. Another interesting point is that, the standard oils amounts are relatively higher than the actual intake ones for all groups here. Maybe it has something to do with the nutritionists’ view that oils are much healthier than solid fats or animal fats, as has been mentioned before.

Or if you change another point of view, such as sources of food, you can also find something interesting. “Fast food” and “Restaurants” contain more solid fats than homemade food and “other” places. “School” for children seems far away from recommended. One usually eats more meat in a “restaurant” than other sources. “Other” provides most added sugars.

From group's perspective, for example, children and lower incomes should eat less added sugar.



Figure 24

Though straightforward, the charts above cannot explain the detail, especially when there is tiny difference between actual and recommended volumes, or when the proportion itself is rather small and not distinguishable. Such as the protein, vegetables and fruits, etc. To overcome this, we build another graph below, a parallel charts built in XDAT. The biggest advantage of this kind of chart is that, it could show details and differences in a more proper and accurate way. For instance, from the charts above, we have no idea about how much amount of protein or whole grains should people of one group eat more or cut down to meet standard. Now it is easy with this parallel graph. First you need to identify the group color, dark red for adults, yellow for children, blue for higher incomes, and light blue for lower incomes. By locating to the axis “total\_protein\_food(ounces/1kcal)” and “standard\_protein(ounces/1kcal)”, you can easily find that, children and lower incomes are with positive slope, while adults and higher incomes are with negative slope, that means the former two need to eat more protein and the latter two need to eat less protein. This is simply an example, if you look at it, you will find more details. Generally, there is many lines from actual to recommended with positive slopes, which means people should eat more, such as fruits, vegetables, grains, etc. Another clue I got from this chart is that, higher incomes tends to be closer to recommended food amount compared to groups of lower incomes, because the corresponding lines from actual to recommended for higher incomes are usually with lower slopes than those for lower incomes.

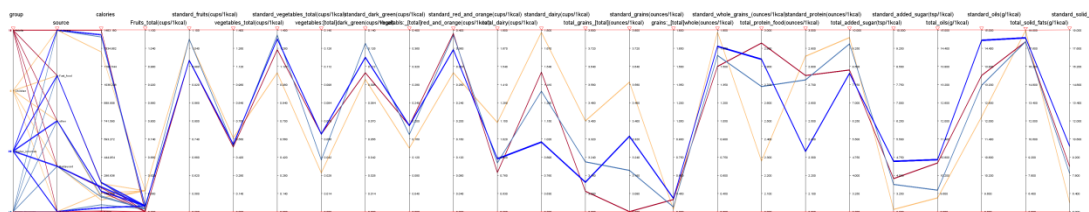


Figure 25

In a nutshell, there are some characteristics according to all the nutrition visualizations above are that, from 1970 to 2010:



*The good points are:*

1. Americans tend to eat more high-protein meat (poultry) than they used to do, and eat less high-fat meat (red meat).
2. Grains consumption is roughly rising, and the proportion of corn products gradually increased. Corn products are relatively healthier grains than wheat flour.
3. Dairy intake is gradually occupied by natural cheese, who has more calcium than beverage dairy products.
4. Fruits consumption is gradually increasing.
5. Nuts consumption is gradually increasing, who are rich of Monounsaturated Fatty Acids and vitamins.

*The bad points are:*

1. Fat consumption is crazy kept in a high level. And most of them are still solid fats, which are recommended reducing by nutritionists.
2. Sugars consumption is relatively high. “Corn sweeteners” has beaten “refined cane and beet sugars”, though either of them is healthy.
3. Though fruits and vegetable consumptions are still under standard level.

For extra, we also prepared a diagnosed diabetes rates map of US from 1994 to 2015. In this map, the darker orange the area is, the higher diagnosed diabetes rate it has. From this map, you can easily pick out the higher diabetes rate states, which are mostly lie in the southern part of US. The highest rate is Mississippi on the mainland.

American diagnosed diabetes rate - 2015

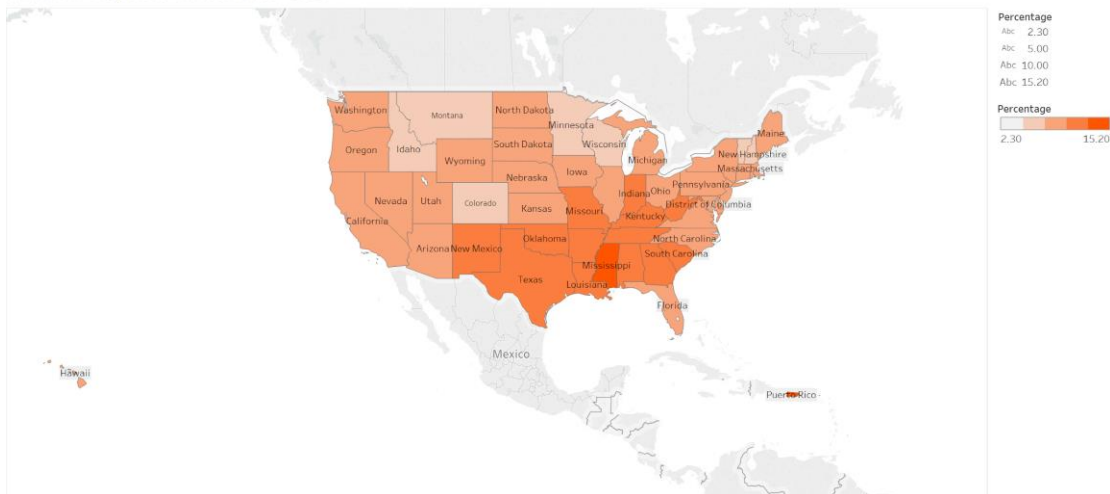


Figure 26

If you still remember what we obtained from part 2, you will find something more interesting. Here, for simplicity, I put the two map together and used a color filter to focus mainly on the meat production area. (The upper one is homegrown food volume, and the lower one is the diabetes rate map)

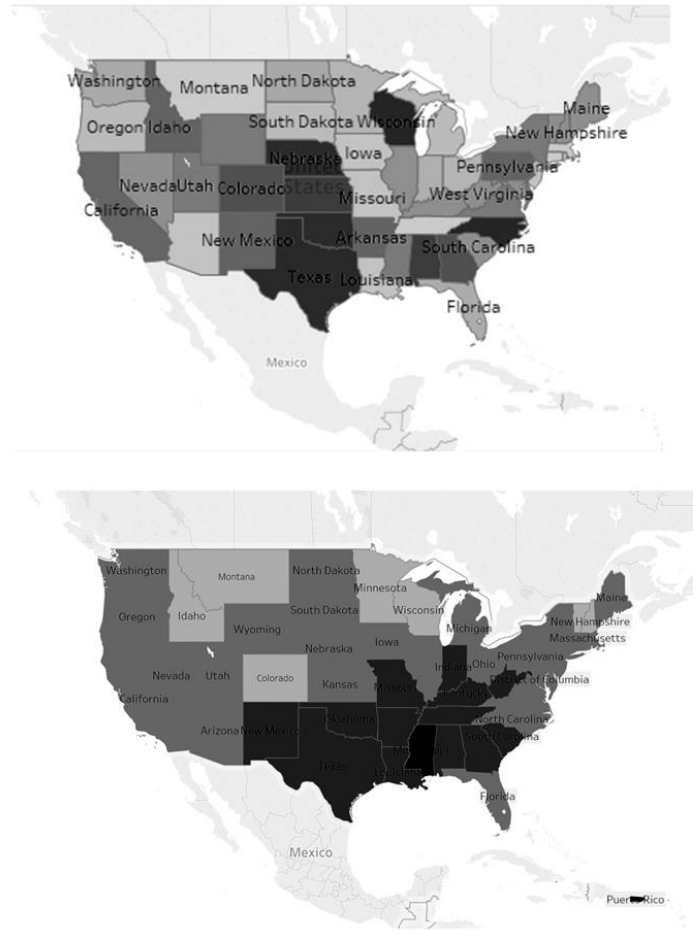


Figure 27

The result could give us a rough result that, meat and diabetes rates are correlated to each other. So if you want to stand away from diabetes, we strongly recommended you eat less meat, or leave southern America. It is for sure that diabetes rates are not such simple to be explained, local culture, neighboring countries, climates, religious and many other issues could hold a weight on it. Here we are doing this part under the theme of our topic.

In a nutshell, there are some key points in this part:

1. Americans are eating healthier than ever before in some aspects, such as consuming more fresh fruits, whole grains, nuts, and oils, and the total calories intake for an average American is decreasing a little bit. On the other hand, though the consumptions of added fats, meat, and added sugars are decreasing these years, they are still at a very high level, even doubled the required amount.
2. Homemade food seems to be healthier than other places, considering the related low added solid fats. And School meal seems not healthy at all. Higher incomes groups are slightly more close to healthy standard than lower incomes groups.
3. The diagnosed obesity rate of US is crazy increasing from 1994 to 2015. Among all states, southern parts have relatively higher rates than northern parts in US.
4. Meat production seems to have relationships with obesity rates.

## Conclusion

American, born from an immigration culture, is also a country of variety of food cultures. There are countless of food type available in this country, no matter it origins from Europe or Asia, or southern America.

Though with such a diversified food culture, the distinctive features are easily captured. They are loving nightlives, bars, food tracks, fast food, dessert, meat and chocolates. They enjoy eating out at night or even midnight. They are eating food from all over the world, the cattle from Canada, and the sugars from south Asia. They also love their homegrown products very much, such as the corns, beef, hog. The long history tells us that American are eating less healthy year by year, but these years this seems untrue. They are increasing their consumption of fruits and whole grains, nuts and vegetables oils, decreasing the intake of added fats and added sugars. It is true that their foods are too greasy or too sweet for nutritionists, and the obesity rates are increasing, but observable changes of food consumption have been made. Hopefully this country could get a balance between excess food and healthy diet soon.

In this project, we tried first introduce the data (descriptive effective), such as most of our graphs obtained. Then we did do some extension work. We tried to form a logic to prove the relationships between eating habits, food sources and life concepts (nutrition). In this logic, we expanded our database and added more datasets in and finally connect all these three big topics. When we are analyzing the food intake from different groups, the data seems to give us more than what we have expected, such as the income level, the food sources, which are all matters to food intake. Using slope helps us to find more information than simply using the data. We also got some surprising results after visualization, which is very common over time. We are stoned by the eating habits of Americans, the prolific California, the fats and sugar consumption of an averaged American, the huge difference between actual intake and standard intake amount of food, etc. Besides, we prepared more materials when analyzing than visualization. Because visualization is a powerful assistant for analyze.

## Difficulties we have met:

Find proper dataset that could be used on a specific tool to do visualization is hard. There are so many kinds of data we could obtain online, and they vary a lot, which lacks possibilities to do cross analysis. Also, without data the project cannot begin. Data preprocessing is really a troublesome work. Most data downloaded online cannot be used directly. Such as the “food intake from different sources by different groups”, it is a dataset with a viewer-friendly view and hard for tableau or Xdat. We have to modify it line by line. It was really time consuming and labor consuming. Finding interesting story from the raw data is hard. Here, part 2 and 3 are using data from USDA, whose data are very suitable for a straightforward report rather than interesting stories telling.

The Yelp dataset has collected detailed longitude and latitude of each restaurant and complete addresses with street, city, and state. We thought this will help us get a great map view of the data points but the coordinate happened to be have something wrong and all the data points on the map are highly grouped together, which looks not making sense as they should be splattering within the states but they are not.

Then we made use of another data set zip\_codes\_states.csv. It helps assign the correct coordinates to each restaurant according to its city. So we have a city-level location of each restaurant on the map.

The Yelp dataset labeled each restaurant with more than one classes within the same column. We were having a hard time to find the way to reorganize and analyze the number of each class while keeping its linkage to corresponding restaurant. This has been overcome at last.

Yelp actually not only lists restaurants but also other kinds of shops, such as home supplies, leisure life centers, etc. We filtered out the shops with 'food' and 'restaurant' labels and assume they should be the restaurant list we want. But when we are doing the word cloud, we noticed there are remaining shops that should be included. We removed such kind in this visualization to present a better view of restaurant components.

Also, upload a dataset with larger than 15M size is difficult. Using terminal to upload local web file to git.io is not very easy.

We found it hard to choose the better choice. This is because different choices are usually from different views, which could tell different details. In this case, we keep our try in the report and justifying them respectively.

Finally we think that we are choosing a quite large topic. In fact, even a sub-topic could stand on its own. If time is allowed, we could explore more on the detailed state level information about tastes, food consumption in volumes, and different standard levels. Except time, limited accessible dataset is also a problem. Datasets from different sources, even same sources, could be in different dimensions and time slot.

The existing tool such as tableau is easy to use, but kind of rigid. Sometimes it cannot meet your expectation, such as a man shape object could be difficult to build here. And importing data must follow its rule. Xdat also puts strict requirements on input data (.txt). d3 requires knowledge of web coding, but it is rather flexible.

## Division of labor:

Here we first want to thank our teacher Dr. Choi, and our TA Mr. Lin and Mr. Chen. Thanks to their patient teaching work. All tools used are within the range of their instruction.

		Yi Jue 3035420352	Zhu Qingyu 3035419080	Shi Kexin 3035455278
proposal		✓	✓	✓
Data collection/ preprocessing/ /visualization	Part 1	✓		
	Part 2		✓	
	Part 3			✓
Taking Video		✓		
Building Website			✓	

Analysis and report	✓	✓	✓
Presentation	✓	✓	✓