CPSC583

Assignment 2

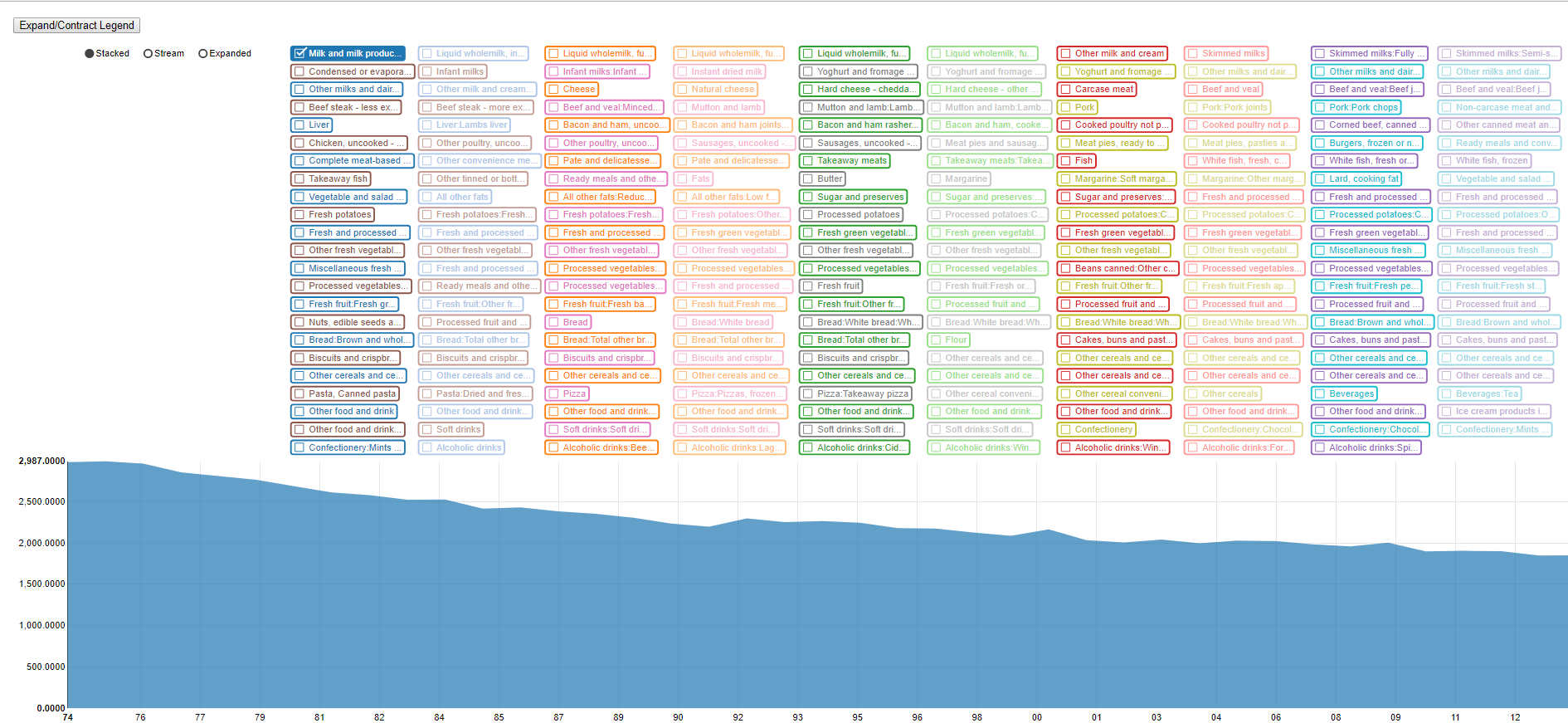
UK Food Trends

Brian Nguyen

<https://github.com/uofcbrian/CPSC583A2>

<http://pages.cpsc.ucalgary.ca/~brian.nguyen/stackedAreaWithFocusChart.html>

Resources taken from http://nvd3.org/



# Data Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| desc1 | 2000 | 2005 | 2010 | 2014 |
| Milk and milk products excluding cheese | 2164 | 2027 | 1897 | 1849 |
| Cheese | 109 | 116 | 118 | 111 |
| Carcase meat | 235 | 226 | 211 | 195 |
| Non-carcase meat and meat products | 779 | 821 | 805 | 760 |
| Fish | 144 | 167 | 151 | 144 |
| Eggs | 2 | 2 | 2 | 2 |
| Fats | 193 | 183 | 183 | 158 |
| Sugar and preserves | 167 | 129 | 126 | 109 |
| Fresh and processed fruit and vegetables, including potatoes | 3338 | 3290 | 2982 | 2846 |
| Fresh and processed potatoes | 1002 | 842 | 742 | 671 |
| Fresh and processed fruit and vegetables, excluding potatoes | 2336 | 2448 | 2240 | 2176 |
| Fresh and processed fruit | 1189 | 1292 | 1133 | 1096 |
| Bread | 782 | 701 | 634 | 555 |
| Flour | 69 | 60 | 58 | 52 |
| Cakes, buns and pastries | 187 | 168 | 153 | 147 |
| Biscuits and crispbreads | 189 | 165 | 162 | 162 |
| Other cereals and cereal products | 470 | 532 | 556 | 560 |
| Soft drinks | 1699 | 1718 | 1718 | 1546 |
| Confectionery | 151 | 123 | 131 | 130 |
| Alcoholic drinks | 725 | 739 | 762 | 675 |

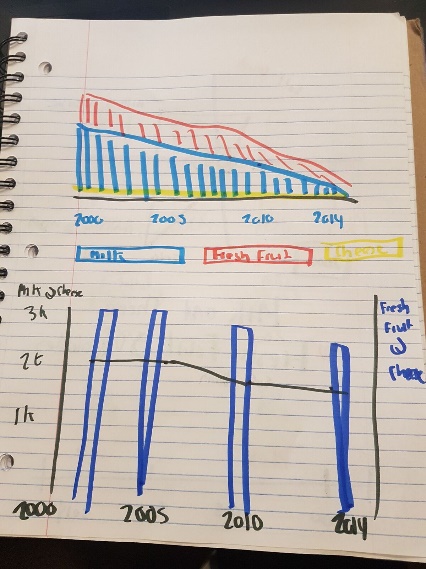
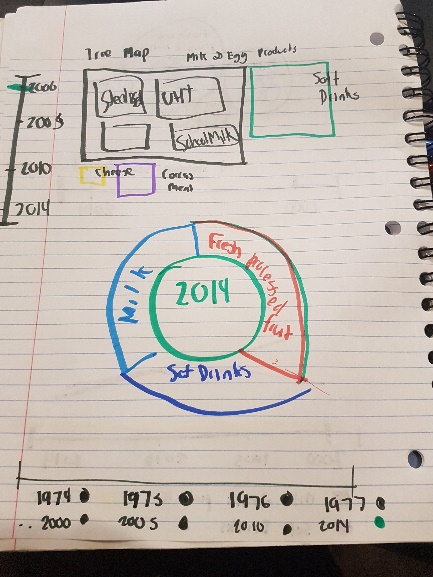
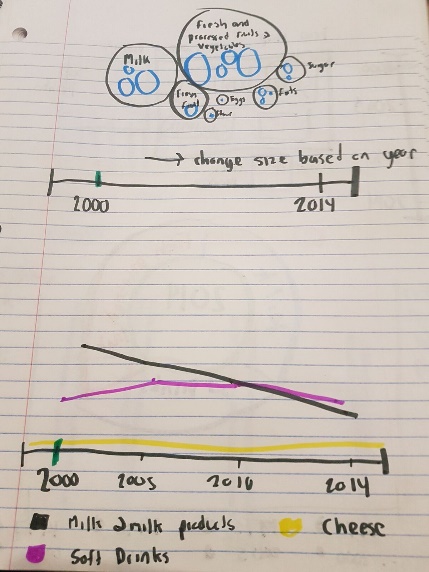
Rationale

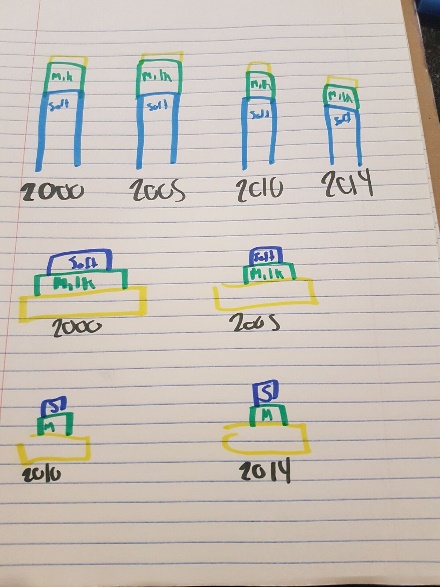
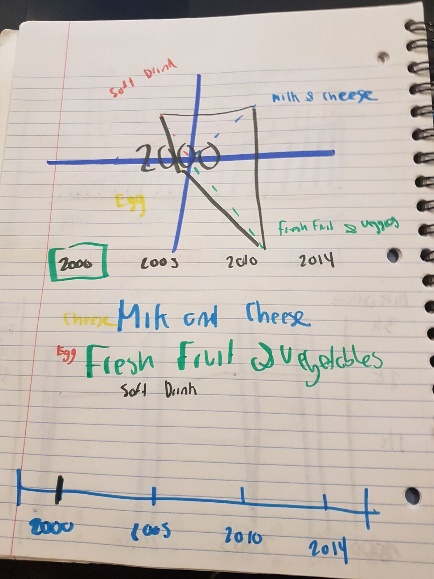
I selected years every ~5th year starting from 2000. Even though I did not start from 1974, I still feel like a 15 year period will show significant trends if there are any. I choose every 5th year because the changes between 5 years are more dramatic.

For the rows, I selected the rows where the first occurrence of the name appears in description one, because every item with the same description one that follows suit is a subset of the former. The first occurrence of such item is therefore representative of the entire set.

From here I am able to see certain trends like for example, a significant decrease in **Milk and milk products excluding cheese** in every 5th year, whereas **eggs** remain consistent. The subset that I have chosen has trends, and by sketching these trends using various ideas will help me figure out the best visualization for showing trends.

# Sketches





I felt that the stacked area graph was the best way to represent trends across multiple food items so I decided to go with that idea.

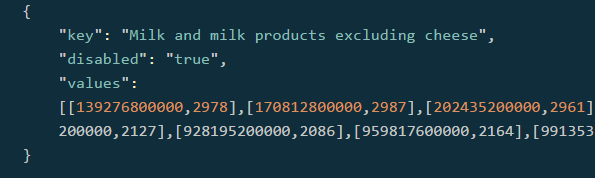
# Preface

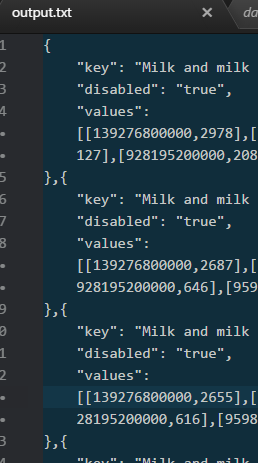
The data given is the average quantity of food and drink purchased per UK consumer per week for each food and drink category. The volume of this information is high so it is impossible to process all of it intuitively by purely observing the numerical values. By visualizing this data set, it allows us to predict the UK food market and profile the diet of an average UK consumer. With this information, public health campaigns can be launched to direct UK consumers in the right direction at the right time if need be.

# Directions and Change

I discovered a D3 library that fit my requirements for a stacked area graph, however the input data had to be in JSON format. There was no quick and easy way to convert the data into the format that was required because additional information had to be appended to the given data set.

To elaborate, each data value had to be appended to a Unix time stamp for the library to generate the correct year for the axes. I had to calculate the initial year (1974) as a Unix time stamp and add the number of seconds in a year to the initial time to get each consecutive year. Next, I had to append the appropriate time stamp to the given data value for each year.

I manually created a JSON entry which was extremely tedious so I wrote a parser in java that read each .csv line and appended the correct Unix time stamp to each data value. It also generated the rest of the JSON entry for me as well by some string operations.

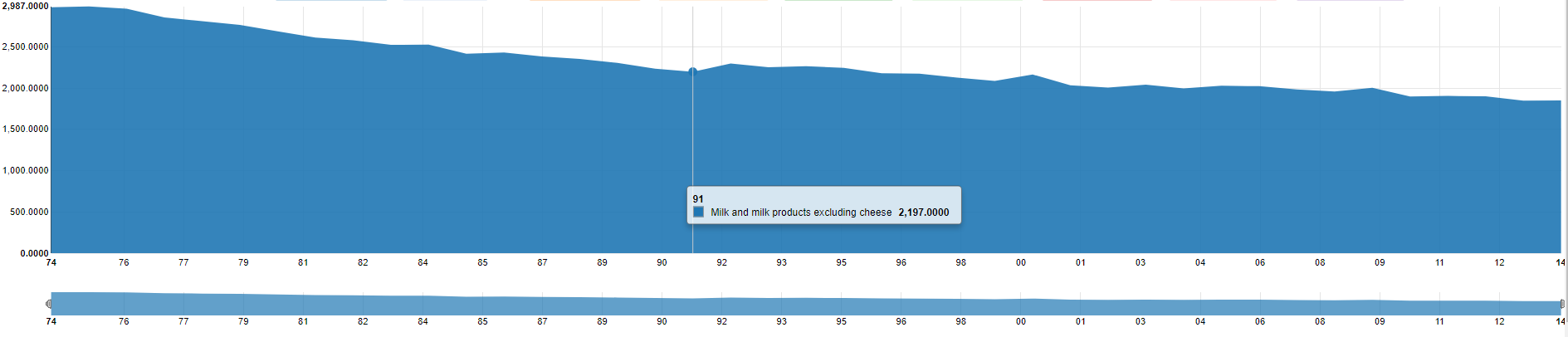
I finally had an output file of the data values in JSON format and simply had to copy paste them into the .html file for the library to do its work. After observing the first iteration of the visualization. I realized that the legend was extremely difficult to read. I simply appended all of the descriptions together to form the key. This means the legend values trim off the excess string and only display the first couple of characters of the key value. You have to hover over the value to get the entire string. I had to trim the string somehow without losing much of the meaning.

This was most difficult part for me because I had to take into consideration what to keep and what to trim off in the string. I also had to be consistent with how I trimmed the string as well.

Finally, I decided to trim some of the data entries from the original data. After playing with the visualization, I realized that there were several entries with such small data values that they were simply dwarfed by the bigger values upon comparison. When comparing popular food item like **milk products** against something small **like sandwiches from takeaway**, it did not really provide any meaningful insights. I removed sets of rows where the data values were < 15 and remained so throughout the 40 years. I did however keep food items that started around ~10 – 15 but rose to the hundreds. I removed items that contained ‘other’ in their description. Several of these food items remained around 5 to 10 in value throughout the 40 years. Keeping these values cluttered my legend and added little insight in the grand scheme.

# Representation





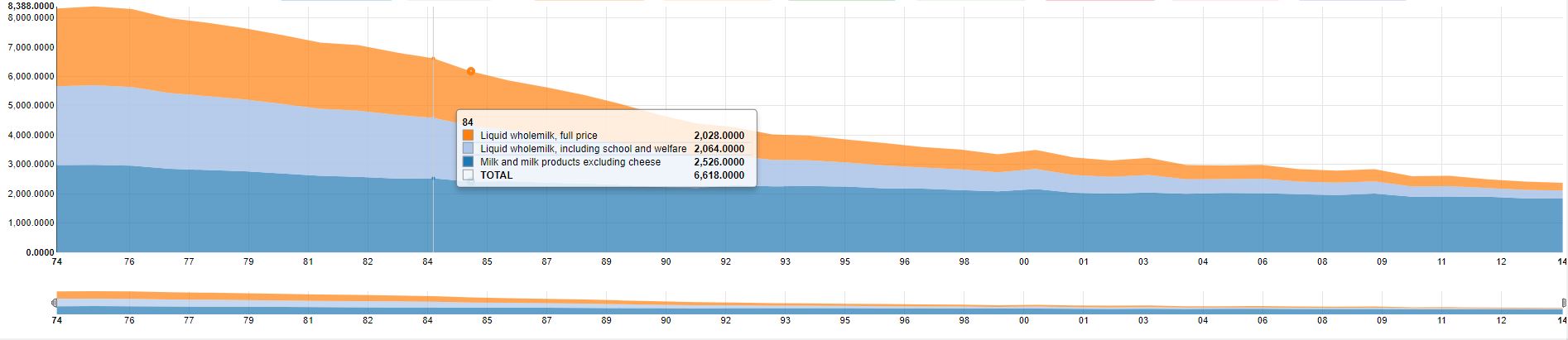
The data for each year is taken and appended to a Unix time stamp. The Unix time stamp is parsed by the library and generates the appropriate year for that data value. The horizontal axis is the year whereas the vertical axis is the given data of food quantity for that particular year. I have selected the **Milk and milk products excluding cheese** and we can see that the area graph depicts the time-series relationship. The area between the axis and the line is emphasized with color for legibility.

Next, we will look at area charts with more categories selected in the next section and discuss how the information is organized spatially.

# Presentation

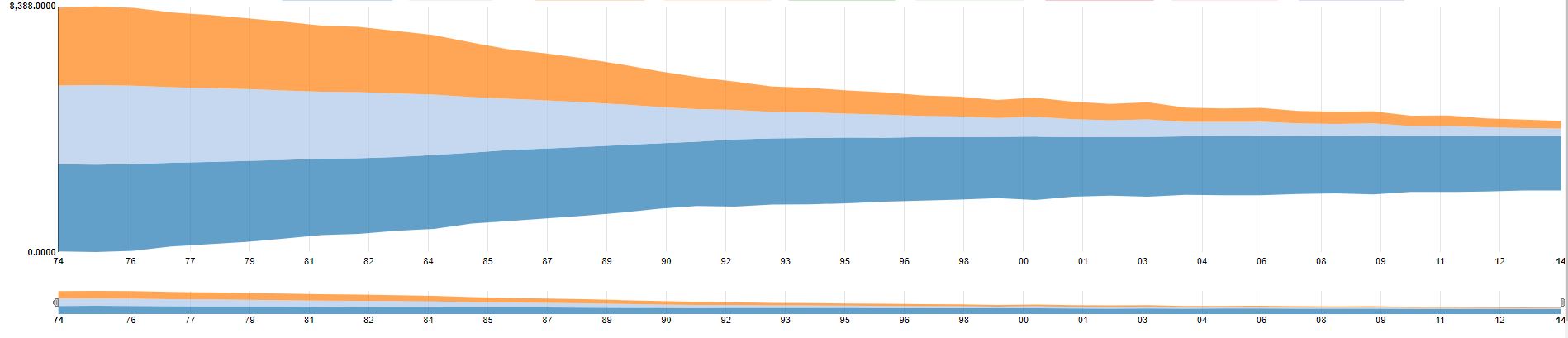
**Stacked Area Chart**

Here I selected the next couple of milk and cream items categories. We see that it communicates the overall trend as opposed to individual values, however upon hovering the graph it will show you the discrete values for that year. Stacking the areas of the categories is useful for comparing quantitative progression over time.



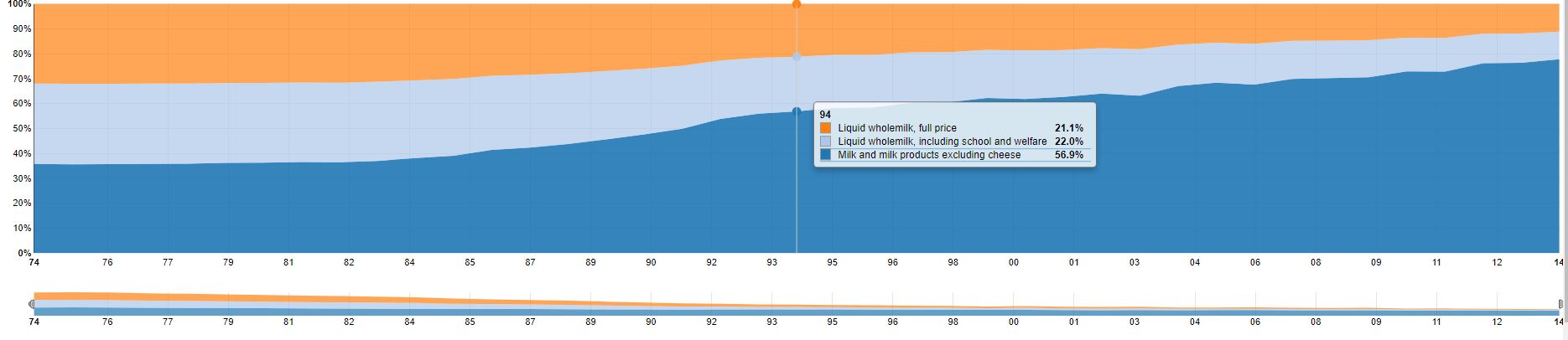
**Streamed Area Chart**

Here is a variation of the stacked area graph but instead of plotting the values against a straight axis, this graph has values displaced around a central baseline. This graph displays the change of data over time through flowing organic shapes. The size of each stream is proportional to the values in each category. The axis of this graph flows parallel to and is used for the time scale. Peaks and troughs in a year by year basis can suggest that there is some underlying periodic factor.



**100% Stacked Area Chart**

This chart as the name implies is where the volume of each category takes up the entire graph. This is used to show the distribution of categories as a parts of a whole relationship where the cumulative total is unimportant.



Interaction

Some of the interaction has been demonstrated above but I will quickly elaborate. The visualization at a basic level is an area chart for a set of categories that are selectable. Selecting multiple items allows you to compare and contrast different volumes of categories, and allows you to stack the areas on each other.

The visualization allows you to dynamically change the view of the area charts in 3 ways: stacked, streamed, and expanded.

Another interaction is having the ability to change the range of years that is currently being displayed. In the example above, the right picture is showing the data from a range of 2006 to 2014 whereas the left picture is showing ranges from only 1985 to 1987.

Finally, hovering the area chart will display the values of all categories selected at that certain year you have hovered onto.

There are some inconspicuous interactions like the ability to select every single category by **selecting just a single category and then clicking on it again.** When selecting multiple categories, you can double click any selected category to deselect all categories except the one you double clicked. This means that if every category is selected, double clicking any category will deselect all categories.

# C:\Users\Brian\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Capture5.jpg

# C:\Users\Brian\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Capture6.jpgC:\Users\Brian\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Capture7.jpg

# Positive Features

* ability to isolate the data through the means of selecting specific items as well as selecting the range of years to display
* ability to compare and contrast selected data through three different views: stacked, streamed, and expanded
* ability to easily communicate the overall trend of selected food items
* ability to see periodic patterns by observing troughs and peaks
* ability to see explicit values for each selected item at each year by simply hovering over the graph