Analysis of Food Prices In Different Regions, Divisions and Market Groups

Introduction:

Food is the most basic necessity of the life. Analyzing food prices for different regions gives us knowledge about the demographics of a state and the living conditions. It also provides an idea of the economic condition of the families and the food habits of the region.

Team Members:

This work is done by the team comprising the following people:

* Sravanthi Bondugula
* Xi Chen Stephen

Data:

The analysis of food prices has been done on the data collected from the United States Department of Agriculture(http://www.ers.usda.gov/Data/). The Quarterly Food-at-Home Price Database (QFAHPD) provides food price data to support research on the economic determinants of food consumption, diet quality, and health outcomes.

QFAHPD-1 database consists of food prices for 52 food groups. These Food groups are defined by the Dietary Guidelines for Americans in 2005 (http://www.health.gov/dietaryguidelines/dga2005/document/default.htm). This data is collected on both UPC-coded and random-weight purchases. Quarterly prices are available for this data for the following regions:

- 26 metropolitan markets for 1999-2006 and
- 9 non-metropolitan markets for 2002-2006 and 4 non-metropolitan areas for 1999-2001

The data can be downloaded from http://www.ers.usda.gov/Data/QFAHPD/qfahpd1.htm

The data comprises the total expenditure spent by a region on each food group in each quarter for the years 1999 to 2006. The price of the food group is also specified quarterly for all the years in the range. Apart from these, number of households participated in this survey, standard error of the data and aggregate amount is available.

For our analysis, we have categorized 52 food groups into the following 10 categories:

- Fruits
- Vegetables
- Grains
- Dairy
- Meat
- Nuts
- Eggs
- Fats
- Beverages
- Prepared Foods
Analysis:

Food Prices have been analyzed across regions, divisions and market groups for 52 food groups based on food category and food groups.

Los Angeles does not favor Vegetables

What is more interesting in this data is to find where each food group is bought more than other food groups. This graph shows the highly spent regions for each food group based on the total expenditure.

From the treemap, we can see that Metro South Regions buy large amounts of food and therefore are the major sellers for most of the food groups. Los Angeles spends as much high as others in atleast one of all the food categories except in vegetables. Metro Midwest regions follow the consumption line after the metro south region.

Be Wealthy to be Healthy in San Francisco

It is not only interesting to see which regions are the top buyers of each food group, it is also interesting to see where these food groups are costlier. The hierarchy of food category, food group and highly priced regions are shown in this tree map.
San Francisco is the highest priced city in 75% of the food groups. Followed by San Francisco, the other priced regions are Hartford, Non Metro New England, Urban NY, Non Metro South Atlantic and Non Metro Pacific.

East North Central is the major seller – Metropolitans are food-pricey

The previous visualizations highlight the prices and spending habits of the people for each food group. This graphic gives an over all picture of the price and spending habits of regions/division/market group and food groups.

Each division is placed in the tree map based on the average of the total expenditure spent on the food category. Colors are associated based on the price. High priced regions are displayed with a pink color and Low priced regions are displayed with orange color.

East North Central Regions buy majority of the food products in high quantities, followed by West North Central and Middle Atlantic. East North Central Region is comprised of: Non Metro Mid West 1, Metro Ohio, Chicago and Non Metro East North Central regions. It is very surprising to see that
regions other than the Non Metro Mid West 1 are not the top bought regions in the individual analysis of food categories. But, the aggregate expenditure of all the regions caters to the increase in the overall consumption of this region for each food category.

The following categorization is obtained when market groups are priced hierarchically for all the food groups:

**Highly Priced Market Groups:** San Francisco, HartFord, South Florida, Washington, DC, Non Metro New England, Baltimore, Non Metro West South Central, Los Angeles, Metro California,

**Medium Priced Market Groups:** Philadelphia, North Florida, Atlanta, Non metro Mountain, Metro Mountain., Non Pacific and Noon Metro Pacific.

**Low Priced Regions:** Non Metro West North Central, Salt Lake City, Metro Ohio, Non Metro East North Central, Metro MidWest2, Non Metro Middle Atlantic,Metro South, Metro South 1, Metro South 2, Metro South 3, San Antonio.

Notice that Baltimore is one of the high priced regions in United States.

Food Groups - Division wise

[Image of a chart with food groups categorized by region and price level]
Expenditure and Price are not always negatively correlated

Everyone buys Medium Priced Food Groups. Contrast to what we think, this scatter plot shows that low priced foods are equally bought as the medium priced foods and slightly less than the high priced foods. It is obvious that the high priced foods like Low fat Cheese are not a part of daily consumption and therefore it is bought less often than the low priced foods like Non-alcoholic carbonated beverages and other bread, rolls, rice, pasta and cereal. Regular fat meat is the major consumed product in all the regions.
What is People's favorite in each food category?
The following pie charts gives a glimpse of the most spent food product in each category.

The pie charts show the most spent food group in each food category. The following categories are placed in the figure from left to right, top to bottom: Fruits, Vegetables, Grains, Dairy Products, Meat, Nuts, Fats, Beverages and Prepared Foods. Cheese, Fresh/Frozen Regular Fat Meat, Fresh/Frozen fruit, Raw nuts and seeds, other bread, rolls, rice, pasta, solid fats top the food charts.

Fresh/Frozen Fruit, Fresh/Frozen Starchy Vegetables, other bread, rolls, rice, pasta, cereal, whole and 2% cheese, Fresh/Frozen regular meat, Raw Nuts and seeds, Solid Fats, Non-alcoholic carbonated beverages, Frozen entries and sides are the most spent in each of the corresponding categories. It is very hard to come to a generic preference as people prefer processed food in some categories and Fresh/Frozen entries in other categories.
People spend most on meat while least on fish and vegetable

This line chart shows the trend of the average expenditure of each food group in year 1999 to 2006. From this figure we can see that the “Fresh/frozen regular fat meat” has always been the one people spend most money on, followed by “frozen entrée and sides”, “bread, rice, rolls, cereal etc.” and “packaged snacks”. On the other hand, people spent least on most kinds of canned food, and fresh/frozen nutrient vegetables and fresh/frozen fish.
We have a heat map for the total expenditure of different regions. This map shows that in different areas across the country, people share similar consumption habits on food products. We can see that people send most on “Fresh/Frozen regular Fat Meat” and “Frozen entries and sides”, which canned and vegetable are least favored.

**Software Critique (Spotfire):**

We have worked on the Spotfire and felt that the visualizations built in are enough to highlight the patterns in the categorical hierarchical data. The tool provides a wide collection of visualizations with many features that reduce human intervention. The Details-On-Demand tool box at the right allows you to explore the data and is capable of finding the interesting inferences from patterns. It gives the user an ability to assess the quality of the large data sets. The tool also allows you to find, ignore empty values giving a toggling option between various filter selections.

Several usability features like choice of wide range of colors, multiple columns for various axes, categorical hierarchical options, color by, shape by and size by columns enhance the user's ability to play around with as many options available with minimum effort. The interact-ability features for zooming in, filtering options and visualization properties are a boon to the user for highlighting the graphs and preparing appropriately for the presentation. The tool helps you in finding the right patterns with a minimal interactions and also achieves a very fast and efficient update according to the selections.

Apart from the many interactive and usable features of the tool, we felt that the following improvements can be done to reduce user efforts for data preparation and enhance the visualizations:

Providing color options for sub ranges for large range will allow the user to highlight the differences more prominently. Currently, the user is able to assign only two colors for integer variables and a unique color to a category. This limits the user to highlight the differences for a very large range like 10 to 10000. The middle values are difficult to compare with the upper maximum values. Also, in case of large number of categories available, an easy way of selecting the color for a certain set of categories should be available. It is a very tedious job to search for these categories from the properties.

Highlighting the categories or a range when clicked on a category in the color, size and shape sections will allow the user to take a quick look of the data selected for that category. Although, this is available with the filter options, you cannot see the entire data with the selected category.

Categorical grouping and virtual extension of the data table will be very helpful to the users. In our case, we wanted to group the food groups into categories. Although we could select the food groups belonging to a category at one time, we could not compare them across various categories without creating an extra column in our data. This is a very time-taking task especially when the number of work sheets to work with are more. We have done this for 52 Work sheets. As a visualization tool, it becomes necessary to maneuver the data along with the showing what it is already present.

Mapping of a set of values to categories given a codebook will be a great help to the user. We also would love to see if the tool can automatically identify the map data without asking for a map table. Improving the rotation capability around the two axes like what Google Earth allows will help the user in easy and fast visualization of the plot.
*Let the tool know more about the data:* Automatic picking of the default columns which have high or low correlation and the suitable visualization highlighting the pattern will greatly amaze the user.

With the suggested enhancements and the features available, working with the tool was a great experience.