

Visual Analysis of Grocery Expenses to Interpret Dietary Habits

Akshay Mandke*

University of Maryland, College Park

Nupur Kale**

University of Maryland, College Park

ABSTRACT

Data visualization provides a method to identify relationships between dietary and behavioral factors. Our study and analysis will inform the audience about relationships between expenditure on food items, nutritional value, cost savings and dietary trends. We believe that targeting the two important aspects: dietary patterns and monthly expenditure can help us make our subjects aware of their food habits and bring to their notice what food choices they should restrain from to maintain a healthy diet and a balanced spending. Our analysis is based on the data collected from two apartments each uniquely consisting of male and female members. The visualization report mainly focuses on three parts: (1) Individual Expense Trends (2) Expenditure Comparison (3) Annual Intake Trend.

Keywords: Expenditure, diet, health.

1 INTRODUCTION

Diet, nutrition and expenditure on food are issues of special interest among young adults. Tracking expenditure and food intake is necessary to set health goals and increase an individual's awareness and understanding of their dietary habits. Our project focuses on drawing patterns and finding out trends of eating habits of individuals in the age group of 23-28. This analysis concentrates on three parts: (1) Individual Expense Trends (2) Expenditure Comparison (3) Annual Intake Trend. We collected data over the period of one year on grocery expenditure of two houses. Interestingly, the two houses are distinct based on gender. Our purpose is to create a visualization report of the expenditure of individuals on types of food items consumed over the course of one year and provide insights into the expenditure of every individual on specific food items based on pre-defined categories. Our secondary aim is to find patterns in eating habits between females and males that will help us distinguish between eating habits of the two genders and also get insights into differences between their dietary habits.

Our data model consists of 7 tables in MySQL database having appropriate relationships. The source of our data is the expense tracking mobile application- 'Splitwise' and personally maintained Excel spread sheets. The data that we have collected from two households includes the spending of each individual on categories of items such as bakery, beverage, condiments, dairy, frozen, fruits, grains, meat, seasoning, starch and vegetables. Data wrangling has been done on the extracted data to transform it suitably to fit in our model and load it for the purpose of visualization. We aim to create a report, which will visually represent the expenditure of each person and enable comparison between the spending of each person. We also aim to get insights on the type of food most consumed by each person.

* amandke@umd.edu

**nkale@umd.edu

2 RELATED WORK

Based on our research in this field we would say that raising awareness about healthy dietary habits has been a topic of utmost importance. The ever-increasing obesity continues to be a difficult challenge in the United States. It has led to several weight-related comorbidities (e.g. cancer, type 2 diabetes, cardiovascular disease) and early mortality. Obesity leads to loss of productivity, increased health care costs, and decreased quality of life. These effects significantly threaten the stability of our society [1]. An interesting study reveals that, environmental factors (package size, plate shape, lighting, variety, or the presence of others) can increase food consumption volume far more than people may realize [2].

There have been studies to detect grocery patterns by monitoring store-shopping data to target producing and supplying food articles that are demanding in the market [4]. Analysts and health researchers have stressed in finding out about the food intake trends in the population especially focusing on fresh produce of fruits and vegetables to find patterns of healthy eating habits in the people. The study is specifically done to stress upon healthy eating habits and targeting issues like obesity [3].

Focusing on the other aspect of our project, we collected information on how spending habits are tracked and what are the current trends of grocery spending habits of the population in the United States. According to U.S. Grocery Trends, a majority of shoppers (91%) prepare at least one home-cooked meal each week, but during the same period significant numbers also eat at fast-food establishments (24%), full-service restaurants (25%), or eat ready-made meals at home (22%). The Organic Trade Association (OTA) estimates organic food sales more than tripled between 1997 and 2005, growing from \$3.5 to \$13.8 billion. During that same period, organic food sales averaged an annual growth rate of between 15 and 21 percent, while total U.S. food sales grew only in the range of two to four percent [9]. Major trends in the amount of food available for consumption in the United States between 1970 and 2005 are drawn in this article using data from the ERS Food Availability (Per Capita) Data System. Findings show that Americans do not meet the Federal dietary recommendations and in order to meet them, Americans would need to substantially lower their intake of added fats, refined grains, and added sugars and sweeteners and increase their consumption of fruits, vegetables, whole grains, and low-fat milk and milk products [11].

Another study describes interesting investigation on how households spend additional income on food. The authors of this study also delineate that spending is not the same as consumption and that low-income households are not likely to spend much additional income on fruits and vegetables. (Frazao, Andrews, Smallwood, Prell, 2007) The average household expenditure in Sri Lanka in 2012 is Rs. 40,887 in which Rs. 15,358 (38%) is the expenditure on food consumption and Rs. 25,529 (62%) is on non-food items and services [10].

The reason why this data is important is because it helps us introspect on our choice of eating and makes us realize the amount of fast food we consume which is in turn making us more

and more unhealthy. It is necessary to make people aware of what is affecting their purchases in the stores. Tracking expenditures helps us find out whether specific stores make us spend more over other options (due to reasons like wholesale purchases) [6]. Personal tracking of food consumption is an efficient way to monitor and control our dietary decisions [5].

Based on the lines of all the research, we base our project to find general and distinct trend in the grocery consumption habits of individuals from two households. We aim to create a report based on the data we have collected that will make the subjects aware of their food consumption and also detect any excessive food consumption (leading to unnecessary expenditure) that occurred over the course of one year. Since we have data of two households that distinctly consisted of male individuals in one house and only females in the other, another goal of our project is to compare the food eating habits of these two genders and find out insights about how dietary behaviour varies across the two genders.

Data visualization provides a method to identify relationships between dietary and behavioural factors. These findings will inform the design of a study exploring data visualization to identify relationships between food consumption, reasons for engaging in specific food-related behaviours, relevant contextual factors, and weight and health statuses in diverse populations [1]. Studies have proved that computer assisted analysis is the most effective way to track personal visualization [5]. Data visualization elements need to be applied in the most efficient ways targeting the context of the theme of analysis and study to help users interpret the information effectively [7].

We believe that targeting the two important aspects: dietary patterns and monthly expenditure (to find out unwanted spending) can help us make our subjects aware of their food habits and bring to their notice what foods they should avoid and what food choices they should restrain from to maintain a healthy diet and balanced spending.

3 IMPLEMENTATION

3.1 Database Design

In order to stay consistent and logical with our data, we designed a database schema using MySQL. Figure 1 describes our data model in detail with all the tables and its attributes along with the primary key and foreign keys for each table. All the entities have appropriate relationships with each other.

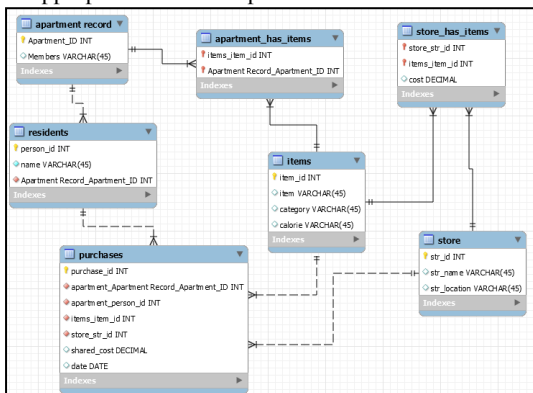


Figure 1: Data Model

3.2 Data Population

We populated the database by importing the spreadsheets (csv) we had as data source for this project. Some of the data has been extracted from expense tracking mobile application 'Splitwise' and stored in the form of CSV. Data Wrangling has been performed to fit the semi structured source data as per the data model. This involved, adjusting the date formats, converting 3 decimal values into 2 decimal values for shared costs, categorizing grocery items, generalizing similar items bought between two different apartments under one unique name, finding and updating unit price of each item specific to each store etc.

3.3 Tableau

We used Tableau to build our Dashboards for the final product of our research. Tableau is a desktop application that enables integration of data and automates the visualization process by self-evaluating the input fields to a graph. It provides with a large array of visualization techniques including bar graphs, tree charts, maps, polygon chart, pie chart, line chart, tabular data representation, etc. [13]

The dataset has huge amounts of data, which could be interpreted using different perspectives. Tableau dashboards are built based on the three themes that we abided by for our study:

1. Individual Expenditure Trends
2. Expenditure habits and Comparison (Male vs. Female)
3. Annual trend

These themes are explained in detail in the report further.

The basic design principles that we followed while creating our dashboards in Tableau are as follows:

1. Simplicity in design
2. Dashboards should be easy to interpret
3. Seamless comparison of entities
4. Good interactivity

3.3.1 Advantages of Tableau

1. You can create visuals quickly and switch between types easily to find the model that best represents your message.
2. Tableau defaults are based on best practices, so your initial result contains good color combinations and layout.
3. The user interface is well organized so you can customize the view with a few clicks vs. multiple menus, etc.

3.3.2 Disadvantages of Tableau

1. It is not a good choice if you want to do a list report (say a report with 20+ columns).
2. The software does not support out of the box solutions and customizations are very limited/almost null.

3.4 Highcharts

Highcharts is a JavaScript framework used to build interact charts and visualizations. It is built to work standalone on the JavaScript and jQuery platform. Highcharts is the simplest yet most flexible charting API on the market. Highcharts currently supports line, spline, area, area spline, column, bar, pie, scatter, angular gauges,

area range, area spline range, column range, bubble, box plot, error bars, funnel, waterfall and polar chart types. [12]

3.4.1 Advantages of Highcharts

1. Highcharts is solely based on native browser technologies and doesn't require client side plugins like Flash or Java. Furthermore you don't need to install anything on your server. No PHP or ASP.NET. Highcharts needs only two JS files to run: The highcharts.js core and either the jQuery.
2. Many of the Highcharts supported charts can be combined in one chart to get a custom design of your visualization.
3. Setting the Highcharts configuration options requires no special programming skills. The options are given in a JavaScript object notation structure, which is basically a set of keys and values connected by colons, separated by commas and grouped by curly brackets.

3.4.2 Disadvantages of Highcharts

1. Highcharts takes the data in a JavaScript array. Even though data can be fed to it through any source there is an extra overhead of data processing to convert to JavaScript arrays.
2. No support for custom graphics and text in the generated charts, e.g. some additional text to the left of a chart. It is possible to implement this functionality by adding a bit of extra code to the framework.

3.5 Website development

The website serves as an interactive report of the research and the dashboards. The dashboard is made using Bootstrap for the UI elements and is backed with HTML, CSS, JavaScript and JQuery. The website holds all the visualization elements created throughout our research project.

4 ANALYSIS

4.1 Individual Expense Trends

4.1.1 Total Individual Expenditure

We used the bubble chart to uniquely determine the total expenditure of each individual. The size of the bubble is in direct proportion with the value of the expenditure, which means that larger the bubble more is the expenditure, and vice versa. The color-coding distinguishes two apartments where teal represents the apartment with males and light orange represents the apartment with females. The highest individual annual expenditure within the 2 groups is from the apartment consisting of males with a total cost of \$825 and similarly the lowest is also from the same apartment with a total cost of \$525.

4.1.2 Expenditure based on calorie

We found the bar chart to be an effective way to visualize each individual's data by splitting the bar into 2 sections describing high and low calorie consumption. Orange denotes high calorie while green being low calorie. The low calorie consumption across both the apartments fall in the range of \$400-\$450 indicating that low calorie diet of all individuals is similar. Interestingly, the high calorie consumption of 3 males (average of

\$340) is twice to that of 5 females (average of \$162) with one of the male members being an outlier with a cost of \$117 on high calorie expenditure.

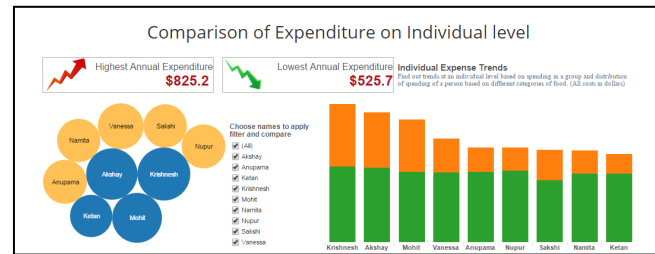


Figure 2: Dashboard 1

4.1.3 Expenditure across categories of items

The series of bar graphs show the distribution of expenditure across categories of items bought by the individuals. We can see that the individual who was an outlier in the previous analysis has no expenditure on meat and very low expenditure on frozen, high calorie snack and dairy items. This justifies his low expenditure on high calorie food products and we can conclude that bakery products and meat tend to expensive or unhealthy if consumed in large numbers. If we filter on the basis of individuals who do not consume meat, we see that the total individual expenditures are similar with an average of \$530. Our data can quite reliably answer that individuals who are vegetarian have similar dietary habits.

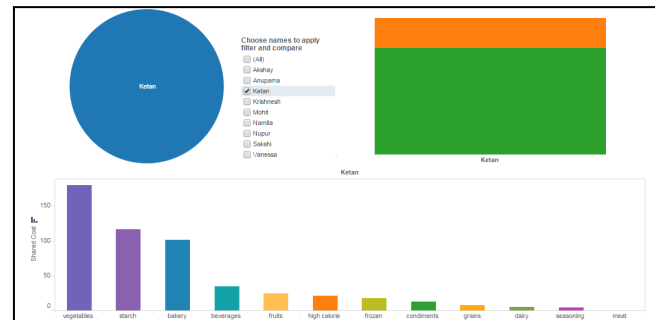


Figure 3: Individual Expenses over Categories

4.2 Expenditure Comparison

House level expenditure comparison across categories of items

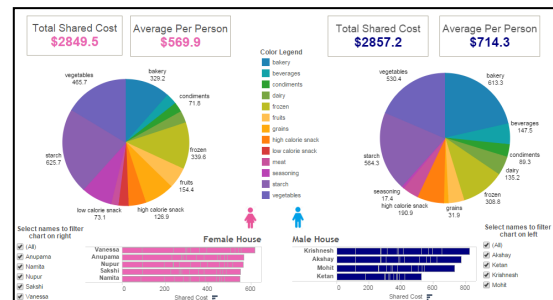


Figure 4: Dashboard 2

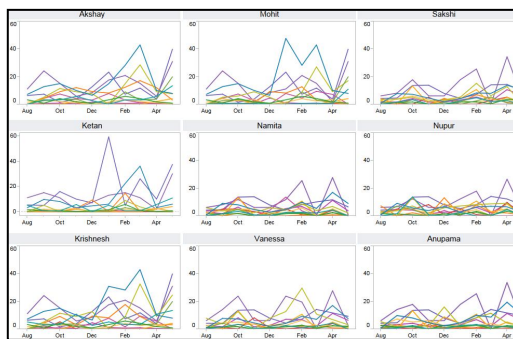
We found that the total expenditure from each apartment is an average cost of \$2853, which is very similar. However the female apartment consists of 5 individuals as compared to 4 from the apartment with males. To dig in further, we plotted the category-based expenditure for individuals from both apartments and

applied a common filter to our dashboard. We found that average per person expenditure for the male gender is \$714 whereas that of the female gender is \$570. Thus we can conclude that even though the total cost spent by both the apartments on groceries is almost same, there are marginal differences between the individual expenditures of males to those of females.

4.3 Annual Intake Trend

To dig deeper into our data, we plotted the individual expenditures across different categories of food items over a period of 10 months starting from August 2014 to May 2015. We plotted our visualizations in the form of a 3x3 matrix where each cell describes each individual's expenditure trends across these 10 months. Several interesting insights that we got found are:

1. Individuals with no or less expenditure on meat tend to spend more on vegetables, starch and bakery products.
2. For individuals with greater high calorie consumption, the trends of consuming frozen and dairy products increase with time indicating a pattern in their dietary habits.
3. During certain months within this period such as December, the line chart steeps down indicating low or no expense at all. It is an interesting result since the individuals tend to travel or are on vacation during this



period and do not spend much on household groceries.

Figure 5: Dashboard 3

5 CONCLUSION

Even though the total shared expenditure for both the male and female houses is similar (\$2857 and \$2850 respectively), the average male expenditure is 11% more than that of females. Hence, food intake of male is more compared to females. Trends shows that the male subjects eat more of high calorie food compared to the females. We found that the male apartment has the highest expenditure on 'Bakery' items (\$613) whereas the female apartment has spent most on 'Starch' items (\$626). The consumption of starch and vegetables is more in both the apartments (\$1190 and \$996 respectively). Since these items form the staple diet of the subjects, it is quite explanatory that they spend more on them. Our data shows that 'low calorie snacks' are consumed only in the female apartment. Interestingly, we found the trend lines to be rising considerably for every individual after the month of December on their expenditures. Since our subjects are students, it is quite possible that they used grocery items brought from home for a temporary period until they started spending extra. For most of the food categories, a general shape of the trend remains the same for a given house. We tend to share things that are commonly bought in the house and hence the

spending trend is seen to be similar for all members of both the house.

6 FUTURE SCOPE

Include additional columns in our dataset for nutritional value of all items. On the basis of the dietary plans we aim to match the nutritional value of items consumed and predict healthier diet for an individual. Locate all stores from where the bills originated. Find out trends in shopping locations over the year. This can also be useful in drawing insights into which outlets help you save money and which ones make you spend more

7 ACKNOWLEDGEMENT

This research was supported by the subjects involved in the study Akshay Mandke, Mohit Juneja, Krishnesh Pujari, Ketan Deopujari, Nupur Kale, Sakshi Jain, Namita Talwai, Vanessa D'souza, Anupama Arora. We thank our colleagues from University of Maryland's iSchool who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper. We thank Adil Yalcin for assistance with techniques in Tableau and suggestions in visualization principles to be followed. We would also like to show our gratitude to Prof. Niklas Elmqvist for sharing his pearls of wisdom with us during the course of this research, and we thank 3 "anonymous" reviewers for their so-called insights. We are also immensely grateful to Sugghosh Veerambuddhi for his assistance in designing the database for our project. Any errors in this project are our own and should not tarnish the reputations of these esteemed persons.

REFERENCES

- [1] Hingle, M., Yoon, D., Fowler, J., Kobourov, S., Schneider, M. L., Falk, D., & Burd, R. (2013). Collection and visualization of dietary behavior and reasons for eating using Twitter. *Journal of medical Internet research*, 15(6).
- [2] Wansink, B. (2004). Environmental factors that increase the food intake and consumption volume of unknowing consumers*. *Annu. Rev. Nutr.*, 24, 455-479.
- [3] Pot, G. (2015). Trends in food consumption over 30 years in British Focus on food and vegetables intake. *The Scientific Newsletter*, (101), 2-2. doi:JUNE 2015.
- [4] Inland Institutional Capital Partners Corporation. (2014). *Grocery Trends 2014*.
- [5] Whisner, C., Porter, A., Pecor, N., & Maciejewski, R. (2015). A survey of personal visualization in mHealth Nutrition Apps. In *IEEE ViS 2015*. IEEE VIS 2015.
- [6] The Hartman Group Inc. (2014). *U.S. Grocery Shopping Trends 2014*.
- [7] Shneiderman, B., Plaisant, C., & Hesse, B. W. (2013). Improving health and healthcare with interactive visualization methods (Vol. 1, pp. 1-13). *HCIL Technical Report*.
- [8] Frazao, E., Andrews, M., Smallwood, D. & Prell, M. (2007). *Food Spending pattern of low income households, United States* Department of Agriculture, Economic Information Bulletin Number 29-4.
- [9] Arlington Economic Development (2007). *Grocery Store trends and opportunities in Arlington and Virginia*.
- [10] Department of Census and Statistics (2003). *Household Income and Expenditure Survey – 2012/13*, ISSN-2279-3968.
- [11] Wells, H. F., & Buzby, J. C. (2008). *Dietary assessment of major trends in US food consumption, 1970-2005*. US Department of Agriculture, Economic Research Service.
- [12] Highcharts. Retrieved from <http://www.highcharts.com/>
- [13] Tableau. Retrieved from <http://www.tableau.com/>