Journal of Marketing Management
December 2015, Vol. 3, No. 2, pp. 39-46
ISSN: 2333-6080(Print), 2333-6099(Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/jmm.v3n2a4
URL: http://dx.doi.org/10.15640/jmm.v3n2a4

Data Visualization in Marketing

U. N. Umesh[1] & Martin Kagan[2]

Abstract

In the last twenty years, the volume of data, and particularly marketing data, has increased exponentially. Hardware and software makers have struggled to keep pace with this explosion in volume. Visualization of marketing data is a solution that we suggest as a way to solve this problem. The four V's of data, as noted by strategists at IBM, are Volume, Variety, Velocity and Veracity. Types of data, referred to as hierarchical data, are more difficult to evaluate with standard tables, due to the more complex relationships between levels of the hierarchy. The approach we present here addresses all four areas of data explosion observed in recent years, particularly for hierarchical data. In addition to computers and software that have to deal with data overload, according to Miller's Law, the human mind is easily overwhelmed as it is only able to handle about seven pieces of information when it comes to memory and processing. The overloading problem that the manager's mind is likely to face can be solved, or at least mitigated, by the visualization software that we present in this study. Visualization is particularly important for hierarchical data, where the individual data points are connected in a tree-like structure, with large clusters of data broken into sub-categories. The hierarchical analyses of data suggested here can help people to see relationships between variables and groups, while making it easy to check on data veracity. We demonstrate the approach using sales data for a technology company. The visualization helps to understand the break-up of sales data into categories, sub-categories etc.

Key Words: visualizing data; managerial interpretation of data; big data; marketing management and strategy; strategic insights.

Introduction

The sheer volume of data in recent years has tended to overwhelm human capacity to understand it and make managerial decisions. While large datasets are becoming increasingly prevalent due to firms computerizing all their activities, the varieties of both Big Data and Small Data are increasing and astonishingly varied. In the classic book, *Big Data*, Mayer-Schonberger and Cukier (2013) point out that the search terms used by the public about the flu were a far more accurate predictor of the spread of flu than the medical reports generated by the government health organization, Centers for Disease Control (CDC). While some people might not go to the doctor if their case of flu was mild, or the doctor may take time sometimes to report cases, Google records the search terms instantaneously. The sheer volume of data generated from these searches, when matched with geographic location and analyzed with other matching information, has been demonstrated as a fast and efficient source of information to predict the spread of flu. Thus, rapid categorizing and summarizing of data can provide valuable insights to managers.

[1] Professor of Marketing, Carson College of Marketing, Washington State University, Vancouver, WA 98686.
unumesh@comcast.net
[2] Founder, Data Cocoon LLC, 15810 SW Cardinal Loop, Beaverton, OR 97007.

Traditionally, visualization has been the domain of statistics. A standard textbook in statistics (cf. Berenson, Levine and Szabat, 2015) has a chapter on creating bar charts, pie charts, line charts, histograms, etc. These are simple representations of data that require standard input. Such representations are useful in everyday use and can be observed dozens of times in the daily issue of a business newspaper such as the *Wall Street Journal*. They present a quick idea of a distribution of a variable in different categories, the mean, etc. They complement the tables with summary data. In contrast, with hierarchical data; traditional simple charts are not as efficient at conveying the complex relationships. Instead of being put in neat cells, with hierarchical data, analysts tend to group the sample, and then find sub-groups and further sub-groups within this sample.

However, with the proliferation of types and variety of data, there is a need for more types of analyses and presentations that a) bring out the relationships between different elements b) summarize complex data with simple and easily understood visuals c) simplify the visualization without the loss of the many dimensions of the data and d) at the same time, achieve all this quickly with easy to use analytical tools.

The aims of this study are the following:

1. Demonstrate the need for data visualization in marketing
2. Analyze hierarchical marketing data and provide managerial insights
3. Demonstrate how the insights derived compare to traditional tables

In particular, we apply the analyses for hierarchical data, where the relationship between datapoints has a tree-like structure, i.e., the each large category of datapoints can be broken into two or more sub-categories (cf. Hierarchical Database Model, 2015).

2. Background

2.1 Importance of Data Visualization

The human mind has the ability to process images very effectively and has developed this skill since ancient times. Even those who could not read and write were capable of visualizing information that they observed (and were required to have this skill for survival!). Romer (2015) notes that 65% of humans are visual learners and those humans can process images in milliseconds, while most people process text in a slow linear fashion. Thus, given the short attention span of society today, it makes sense for providing more visualization of information. As an example, a manager may have to make a call while receiving a text and referencing an e-mail. Data have to be evaluated quickly for rapid and efficient decision making. Visualizing marketing data helps to achieve this managerial objective.

Data visualization is viewed as extremely important in many classical areas. For instance, White et al. (1984) did a study of coronary arteriogram and recommended better analytical techniques. Psychologists, Morey and Cowan (2005) found that visual and verbal memories are not the same. They note that there are many instances of interference between verbal and visual materials in working memory and that these are dependent on task and memory-load conditions. Thus, applying data visualization in marketing is not unusual as it has a basis from applications in other fields as well.

2.2 Growth of Data

Chen, Chiang and Storey (2012) note that the Internet business data generation and collection speeds have increased exponentially since the 1990's. The number of PCs in use worldwide had crossed one billion by 2008. All of these Web and PC sources create and use data in one form or another. Those who question the future growth of data due to slowing demand for desktop computers have to only look at the surge of cell phones in the world with far more people being able to afford cell phones than desktop computers. All major forecasters are expecting smartphone-usage worldwide to exceed two billion units by 2016. We can only expect the availability and use of data to keep growing over time. All this growth has attracted the attention of large technology companies such as IBM, Microsoft and Amazon. IBM has come up with the 4V's of Big Data to focus its efforts in the area: Volume, Variety, Velocity and Veracity (IBM 2015). These terms correspond respectively to the sheer volume and quantity of data, the different forms of data, the speed at which data are created and analyzed, and the uncertainty and quality of collected data.

2.3 Hierarchical Data are Difficult to Interpret

Hierarchical data are encountered frequently in marketing (cf. Peltier, Zahay and Krishen, 2013). For instance, sales or advertising for the company can be broken up into regions, and these in turn can be broken up into types of customers. Such data can be considered hierarchical because sales in a region is influenced by total national sales, and in fact the sales in all regions sum up to the national sales level. Raw data or tables of means and variances are not as insightful when hierarchical data are encountered. For instance, sales allocations within one region may be quite different from another region. These patterns may not be clearly visible with traditional tables or charts. Hierarchical visualization tools are needed, and would provide some insights to the manager, as noted in the subsequent sections.

2.4 Human Cognitive Capacity Limitations

Miller (1956) came up with the famous Miller's Law, which quantifies the limit of the mental processing capacity of humans. He noticed the coincidence between the limits of one-dimensional absolute judgment and the limits of short-term memory. The ability to distinguish between more than a certain number of stimuli is limited. Similarly, the memory span that most people have, measured by the ability to repeat a series of items with 50% accuracy, was also similarly limited. The limit was usually seven, and most people were in the range of five to nine. The limitation of judgment and memory has been replicated in many studies and influences the behavior of consumers and managers.

Interestingly, as the data availability grows rapidly there is not a similar growth in human cognitive capacity. Therefore it is important to come up with methods of simplification and presentation so that managers can evaluate and reach decisions based on the torrent of data facing them. There are two ways to do this simplification. One is the summarization of data in terms of means, variances, average costs, etc. The other is to provide a visual representation of the data.

Due to limits of human cognition, it is much easier to present more information in a visual format than provide the information in a table, without exceeding the mental processing limits of managers. Thus, methods such as multidimensional scaling, hierarchical cluster analysis, and correspondence analysis were developed over the years (Carroll and Green, 1997). Multidimensional scaling's (MDS) main advantage was promoted as the easy presentation of data from the judged similarity of stimuli by respondents. In this study, we present the visualization of marketing data using a newly developed program called B4_UR_IZ*. While the technique and the software were originally developed to visualize accounting and financial data, we will show that it can have applications in the marketing field.

3. Data

There are many sources and types of marketing data. Some of the more common forms in marketing are the following:

- Primary data – e.g., surveys of restaurant patrons
- Internal company data – e.g., sales reports filed by sales division
- Social media sources –e.g., Facebook usage
- Purchase Data – e.g., scanners in supermarkets
- Large customer database – Marriott Hotels database of customers
- Patent Data – USPTO.gov
- Government routine data collection – e.g., census department
- Syndicated services – e.g., Nielsen Retail Index

While there are many different forms and sources of data, we illustrate the visualization approach using data that could have been from most of these data sources.

Data are available in two basic formats: non-hierarchical and hierarchical. As an example, if 2014 sales is $5 million, it could be broken up into sales by region. Then, the sales in the Western region of say $1 million can be categorized into the individual sales of the four different product lines of the company, e.g., hardware, software, peripherals and support products.

If the sales of these four product lines are $250K, $300K, $400K and $50K respectively, then together they constitute sales of $1 million for the western region. In statistical terms, sales in the Western region would be the parent and the sales of the Hardware product line would be the child. Similarly, in the dyadic relationship between Western and National regions, the Western-region's sale is the child and National sales are the parent. In hierarchical structures there can be many tiers or dimensions of parent-child relationships. As an example, National sales could be the child of total sales of all firms in the industry. Further Hardware sales of $250K in the Western region could be the parent and broken into sales in specialty stores and general stores, which would be the children.

The system in its current form takes data from either an indexed Excel* file, or an Access* file. The data must have a hierarchical structure or parent-child relationships in order to produce a meaningful visualization.

In this example, we will demonstrate the procedure for hierarchical data. See Table 1.

Table 1: Sales by Product/Type/Year

Description	2007 Reg. 1	2007 Reg. 2	2007 Reg. 3	2007 Reg. 4	2007 Total	2008 Reg. 1	2008 Reg. 2	2008 Reg. 3	2008 Reg. 4	2008 Total	2009 Reg. 1	2009 Reg. 2	2009 Reg. 3	2009 Reg. 4	2009 Total
Revenues	2200	851	1428	4496	8975	7612	790	0	4241	12643	3411	1755	3147	4882	13195
Consumer Revenue	115	142	385	514	1156	373	49	0	303	725	491	243	582	770	2086
Consumer Desktop	87	49	176	298	610	81	38	0	93	212	34	54	1	648	737
Consumer Portable	28	93	209	216	546	292	11	0	210	513	457	189	581	122	1349
Professional Revenue	294	360	582	613	1849	646	199	0	572	1417	673	502	718	1070	2963
Professional Desktop	46	91	66	20	223	103	66	0	230	399	23	132	46	45	246
Tower1	115	142	385	514	1156	373	49	0	303	725	491	243	582	770	2086
Professional Portable	133	127	131	79	470	170	84	0	39	293	159	127	90	255	631
Laptop	1788	308	418	3170	5684	6445	514	0	3310	10269	2227	914	1728	3024	7893
Other Hardware	50	48	21	245	364	106	40	0	133	279	8	108	111	224	451
Tablets	1738	119	49	2378	4284	2197	140	0	1711	4048	1706	173	71	1828	3778
Cell units	0	141	348	547	1036	4142	334	0	1466	5942	513	633	1546	972	3664
Software, Service &Other Revenue	3	41	43	199	286	148	28	0	56	232	20	96	119	18	253
Cost of Sales	1248	512	889	2596	5245	4286	453	0	2419	7158	2007	1034	1877	2892	7810
Consumer Revenue	86	107	289	386	868	280	37	0	227	544	368	182	437	578	1565
Consumer Desktop	65	37	132	224	458	61	29	0	70	160	26	41	1	486	554
Consumer Portable	21	70	157	162	410	219	8	0	158	385	343	142	436	92	1013
Professional Revenue	176	216	349	368	1109	388	119	0	343	850	404	301	431	642	1778
Professional Desktop	28	55	40	12	135	62	40	0	138	240	14	79	28	27	148
Tower1	69	85	231	308	693	224	29	0	182	435	295	146	349	462	1252
Professional Portable	80	76	79	47	282	102	50	0	23	175	95	76	54	153	378
Laptop	983	169	230	1744	3126	3545	283	0	1821	5649	1225	503	950	1663	4341
Other Hardware	28	26	12	135	201	58	22	0	73	153	4	59	61	123	247
Tablets	956	65	27	1308	2356	1208	77	0	941	2226	938	95	39	1005	2077
Cell units	0	78	191	301	570	2278	184	0	806	3268	282	348	850	535	2015
Software, Service &Other Revenue	2	21	22	100	145	74	14	0	28	116	10	48	60	9	127
Gross Profit	952	339	539	1900	3730	3326	337	0	1822	5485	1404	721	1270	1990	5385
Consumer Revenue	29	36	96	129	290	93	12	0	76	181	123	61	146	193	523
Consumer Desktop	22	12	44	75	153	20	10	0	23	53	9	14	0	162	185
Consumer Portable	7	23	52	54	136	73	3	0	53	129	114	47	145	31	337
Professional Revenue	118	144	233	245	740	258	80	0	229	567	269	201	287	428	1185
Professional Desktop	18	36	26	8	88	41	26	0	92	159	9	53	18	18	98
Tower1	46	57	154	206	463	149	20	0	121	290	196	97	233	308	834
Professional Portable	53	51	52	32	188	68	34	0	16	118	64	51	36	102	253
Laptop	805	139	188	1427	2559	2900	231	0	1490	4621	1002	411	778	1361	3552
Other Hardware	23	22	9	110	164	48	18	0	60	126	4	49	50	101	204
Tablets	782	54	22	1070	1928	989	63	0	770	1822	768	78	32	823	1701
Cell units	0	63	157	246	466	1864	150	0	660	2674	231	285	696	437	1649
Software, Service &Other Revenue	2	21	22	100	145	74	14	0	28	116	10	48	60	9	127

Consider the following dataset: Sales of a technology product over three years, and four regions. 1) Revenues, 2) Cost of Sales and 3) Gross Profit are the three parent categories. Each parent category is divided into four sub-categories or children: a) Consumer Revenue, b) Professional Revenue, c) Laptop, and d) Software, Service & Other Revenue. These four are further sub-divided into sub-sub-categories or more children: Consumer Revenue – Consumer Desktop and Consumer Portable; Professional Revenue – Professional Desktop, Tower1 and Professional Portable; Laptop – Other Hardware, Tablets and Cell Units; Lastly, Software, Service and Other Revenue does not have any children (or sub-sub categories). See Table 1 for all the numbers. This level of categorization demonstrates a series of hierarchical relationships.

4. Analyses

First we begin by providing a brief overview of the technique. We then analyze the data and come up with insights. We show how these insights are much easier to obtain from the visualization technique presented here as compared to summaries in tables. B4_UR_IZ is a new product that has just been developed by Data Cocoon, LLC. It is a dynamic visual analysis tool and a data engine. It incorporates a multidimensional data array, with full scripting capabilities, in a cell structure that transforms, as needed, into a unique multifaceted patent pending mapping grid. B4_UR_IZ gives users real time, multidimensional data analysis capability. The software allows companies, big and small, to swiftly analyze the large amounts of complex marketing, financial, statistical and other data available today in an invaluable, intuitive way that is far more effective than looking at numbers or charts/graphs.

An examination of data from Table 1 shows a collection of many numbers whose inter-relationships and sub-groupings are difficult to fathom. While a particular number can be easily accessed and understood, it is difficult to say from the table how it fits in with other numbers. The numbers in this table were input into the B4_UR_IZ program and the resulting Figure 1 demonstrates the relationships and links. First, the figure has three broad branches going in three directions, and representing Revenues, Cost of Sales, and Gross Profit and distinguished by using three different colors. The figure has a tree-like structure with Revenues, Cost of Sales and Gross Profit in the middle where all the individual branches are joined.

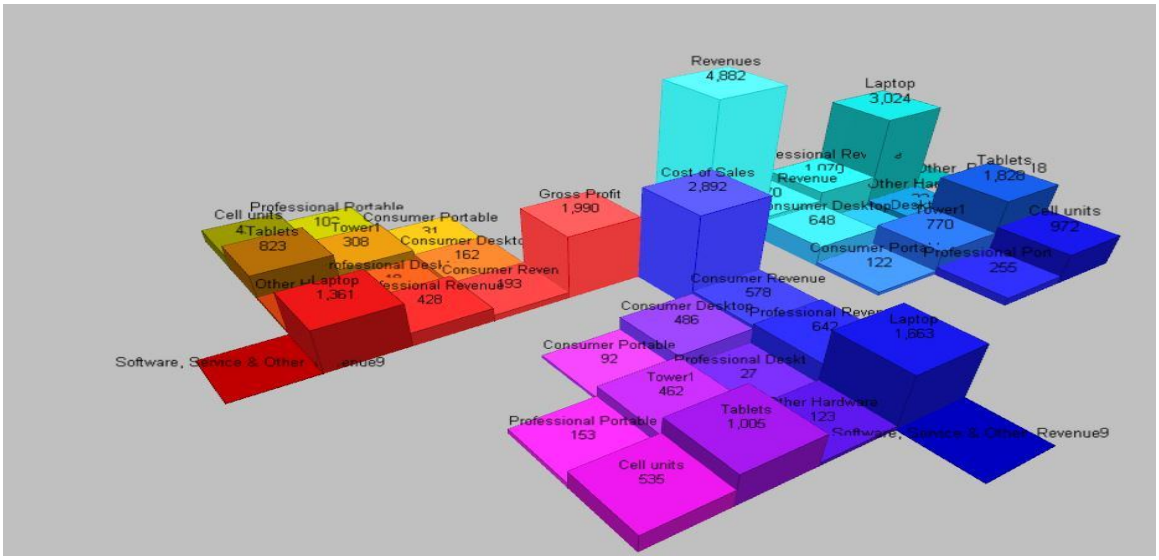

Figure 1: Visualization of the data in hierarchical format**.**

An analyst can understand this dataset further by looking at the sub-categories of products/markets. Subsequent branches divide the three base numbers by product categories, markets, etc. The categories are divided into following sub-categories: Revenue is split into the first category of Consumer Revenue, which is further split into Consumer Desktop, and Consumer Portable; the second category of Professional Revenue, which is further split into Professional Desktop, Tower1 and Professional Portable; the third category of Laptop, which is further split into Other Hardware, Tablets, and Cell units; and the fourth category of Software, Service & Other which is not split as it has no children.

These sub-categories can be viewed as branches within each of the three broad areas of Revenues, Cost of Sales, and Gross Profit in Figure 1. For instance, in the Cost of Sales area, the branch with Laptop provides a visual comparison of their relative importance of the sub-branches or children - Other Hardware, Tablets and Cell Units. Cost of Sales of Tablets exceeds Cost of Sales of Other Hardware. A similar comparison can be made for these four lines, for Gross Profits and Revenues. The software allows rotation of these figures on a computer screen, thus allowing comparison of different branches and their relative sizes.

The system allows the data to be presented in a circular pattern. See Figure 2. The visualization here helps to place the quarterly data along a spine with elevations representing quarterly data as an example. The visualization is far more compact than a typical spreadsheet while providing Revenues, Cost of Sales and Gross Profit that can be easily compared to one another over an extended length of time (multiple quarters). Smaller branches can also be compared based on the pattern differences visualized along the central axis.

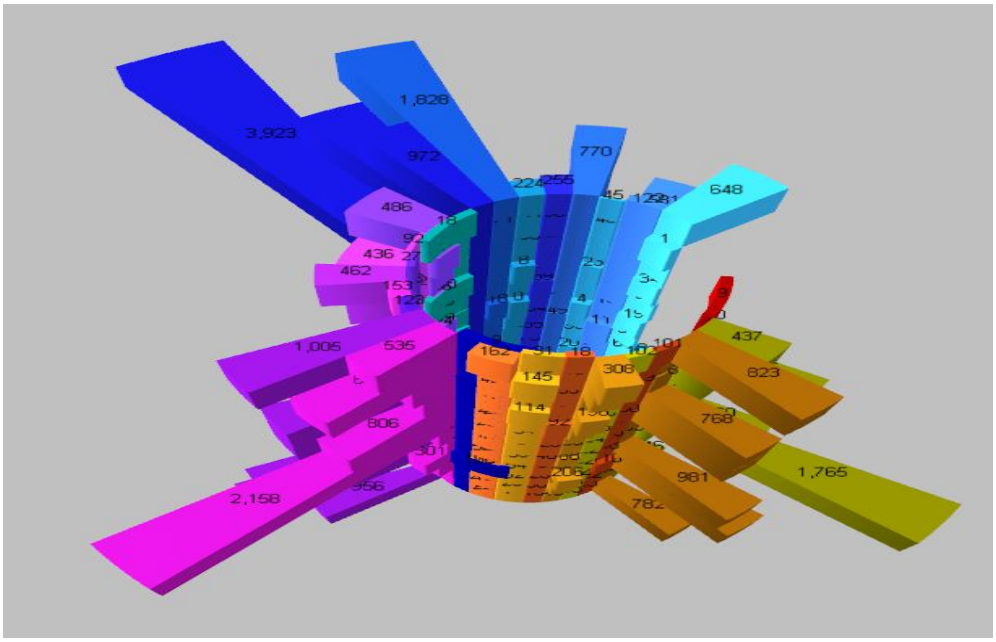

Figure 2: Plotting Data over Time in Multiple Dimension

Visualization helps to detect data entry errors much better than a spreadsheet of data. If the pattern is suddenly broken, then one can easily see that one of the numbers is incorrect. Further examination about the background of the data point can determine if either there is an error that needs to be corrected or the value was unusually low or high because of some unusual event such as a strike, a snowstorm, gain of a large customer, etc. In Figure 3, we can see the error that is shaded to demonstrate the quick visualization of the error.

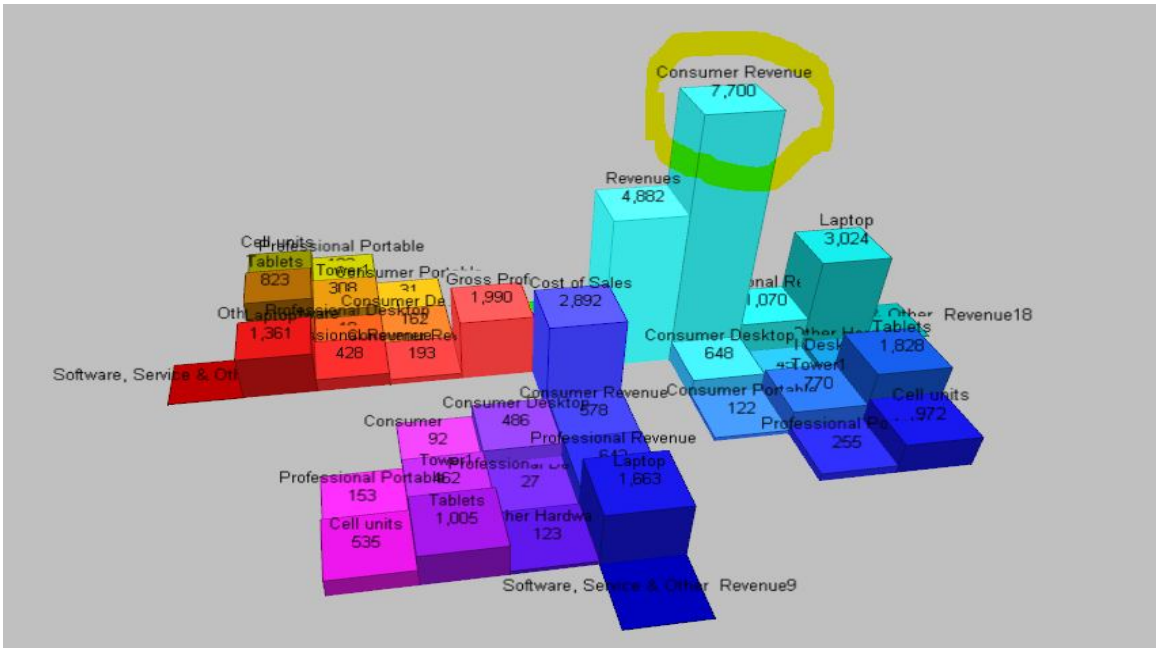

Figure 3: Demonstrating Ability to Detect Data Veracity

5. Discussion

The problem with much of the data that is obtained today is the presence of errors. IBM refers to this problem as Veracity of Big Data. The visualizations discussed here help to improve data quality by helping people to identify errors in data. It is difficult to readily identify errors from an Excel table with numerous rows and columns. While time-consuming analyses can find errors using conventional approaches, a busy manager wanting to make decisions will want an easily usable approach to detecting errors. Our approach is suited for catching these errors by visualization.

In addition to veracity, the other three V's are Volume, Variety and Velocity. Visualization using the software has no problem with any reasonable volume that Excel can handle. Like most software it may not be able to handle extraordinarily large datasets, but a manager on the go who wants to make a quick decision may not encounter these. However, in those cases, the restriction may also be the processing capacity of the available computer, and may also hold for most other software that analyze data. As Intel* and Microsoft* continue to expand computing capacity with every new generation of their releases, the problem will recede – but to only some extent – as the increasing volumes of data are matched by processing capacity. The increasing Variety of data should not have much relevance to our approach to analyzing marketing data. If the data structure is hierarchical, there should not be a problem of analyses irrespective of the type and variety of the marketing data. Variety also implies the structure of data and format; all of which are compatible hierarchical structure and our type of analysis. Lastly, the velocity is relevant both to the creation of data streams and the quick analyses of data. The need of the hour is speed of analyses, and by implication, ease of use of tools that can help achieve this speed. In software use, the ease of use, the speed of running of software *and* ease of analyses and interpretation are all relevant for facing velocity. Overall, the procedure presented here appears to be suitable for handling Veracity, Volume, Variety, and Velocity.

6. Conclusion

Overall, the visualization of data is helpful to understand complex marketing details quickly. As the need for quick decision-making keeps rising in marketing, particularly with the advent of the Internet, visualization of data will become increasingly important. E-Commerce and online purchases requires rapid reaction to price changes. For instance, airlines change fares multiple times a day along some part of their route system. In these cases, rapid understanding through visual representation of the effect of marketing variables on strategy will help in improving profitability.

We have demonstrated that data visualization is helpful for marketing decisions. However, spreadsheets and other non-visual data are very important and cannot be done away with. It is best to provide a marketing analyst both visual and non-visual data so that sound marketing decisions can be made. Some managers are best at understanding numbers and others are mostly visual; hence both must be provided to managers for making sound decisions.

Heer and Shneiderman (2012) state that multiple, linked visualizations are important for providing meaningful insights into multidimensional data rather than isolated visualization of the same data. Their point is that a single image cannot answer all questions. While this point may appear tautological, their conclusion is quite valid and useful to marketing strategy because a) the quantity of data that can be presented in a single image is limited and b) inter-relationships between variables and data sets cannot be entirely presented with a simple image. Multiple images or multidimensional images are essential for presenting data, particularly in the field of marketing, which is the crux of our contribution to the field of marketing.

References

Berenson, M., Levine, D., & Szabat (2015). *Basic Business Statistics*. 13th edition, Boston, MA: Pearson.

Carroll, J. D. & Green, P.E. (1997). Psychometric methods in marketing research: Part II, Multidimensional Scaling, *Journal of Marketing Research* 34(2): 193-204.

Chen, H., Chiang, R.H.L., &Storey, V.C. (2012). Business intelligence and impact: From big data to big impact, *MIS Quarterly*, 36 (4): 1165-1188.

Heer, J. &Shneiderman, B. (2012). Interactive dynamics for visual analysis. *Queue*, 10(2): 1-26

Hierarchical Database Model, 2015, https://en.wikipedia.org/wiki/Hierarchical_database_model

IBM 2015 http://www.ibmbigdatahub.com/infographic/four-vs-big-data

Mayer-Schonberger, V. & Cukier, K. (2013), *Big Data*. London, England: John Murray Publishers.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information, *Psychological Review*, 63: 81-97

Morey, C. & Cowan, N. (2005). When do visual and verbal memories conflict? The Importance of Working-Memory Load and Retrieval, *Journal of Experimental Psychology*, 31(4): 703-713.

Peltier, J., Zahay, D. & Krishen, A.S. 2013. A hierarchical data integration and measurement framework and its impact on CRM system quality and customer performance, *Journal of Marketing Analytics,* 1: 32-48.

Romer, B. (2015). A picture is worth more than a thousand words, http://exchangemagazine.financial.thomsonreuters.com/articles/a-picture-is-worth-more-than-a-thousand-words?adbid=619572992046043136&adbpl=tw&adbpr=14412844&cid=social_20150710_48979916 *Exchange Magazine*, Thomson Reuters.

White, C. W., Wright, C.B., Doty, D.B., Hiratza, L.F., Eastham, C.L., Harrison, D.G., & Marcus, M.L. 2004. Does visual interpretation of the coronary arteriogram predict the importance of a coronary stenosis? *The New England Journal of Medicine*, 310 (March): 819-824.