Shengjie Quan

Professor: Praveen Kumar

CSE 5234

December 5, 2016

CSE 5234 Paper

In an enterprise, a main challenge arises these days is that having possession of massive amount of data accumulated during business, seldom can the enterprise utilize them to extract useful information from the data. The major obstacle besides the massive quantity being challenge to manipulating on the data is that data most likely in numerical form is good for computer to process but bad for human to interpret. Data visualization is thus introduce into enterprise computing to solve this problem. Data visualization is in fact a data mining strategy in the sense of extracting useful information from data while abstracting ''in some schematic form, including attributes or variables for the units of information'' thus providing result that can be easily interpreted by human (Friendly, 2009). Data visualization intuitively belongs to data domain of enterprise computing; however, it is not being isolated in data domain. The result produced can have great impact on the business operation of an enterprise. Thus data visualization is more and more engaged in reporting and business intelligence.

The most current data visualization technology is quantitative messages which includes the communication of the following messages(Few, 2004):

- Time-series
- Ranking
- Part-to-whole
- Deviation

- Frequency distribution

- Correlation

- Nominal comparison

- Geographic or geospatial

Despite the complex definition or the concept of any of the terms above, one fact falls into our observation is that all of them are statistics. Data visualization currently is implemented by statistic plots such as bar chat, scatter plot, etc which carries statistic information above. This implementation approach is quite nature due to the fact that statistics summarize the data which reduce the quantity and statistic plots are traditional way to present the statistics summary of the data. A usable implementation however is typically far more complex than just displaying a bar chat (or just a statistic plot). The complexity is laid on the fact that without choosing the correct parameters (e.g. range, etc) to display a plot, the information and pattern associated with the information may not be revealed. Even choosing the type of the plot may also impact the presentation of the information with in the data. Also automate the process of choosing a proper plot type and plot with proper parameter is of great complexity.

In terms of available commercial off-the-shelf (COTS) software, below is a list:

- Sisense

- Zoho Reports

- Dundas Bi

- Idashboards

- Salesforce Analytics Cloud

- Domo

- Chartio

● Gooddata

Since the current approach of data visualization being statistics, any statistic tools such as *R* can also be used as a data visualization software. One newly released software of particular interests is Amazon AWS Quicksight. Figure 1 is a screenshot of the software. The reason AWS
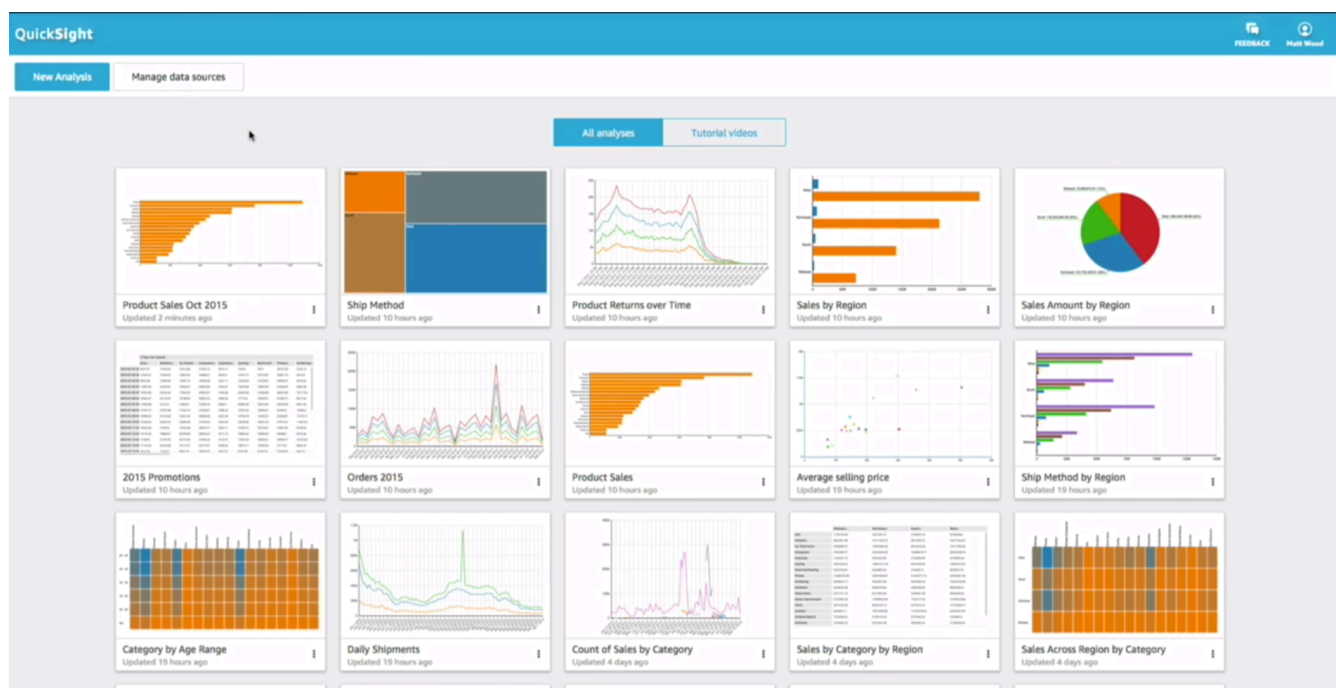

Figure 1: Amazon AWS Quicksight Screenshot

Quicksight being competitive among software of similar functionality is not only because of its deep integration with already exist database software which company may store massive data in, but also because of the fact that AWS Quicksight has ''built-in suggestion engine that provides you with recommended visualizations based on the properties of the underlying datasets" (Amazon Web Services, Inc., n.d.). As I mentioned previously, automating the process of choosing a proper plot type and plot with proper parameter is of great complexity. Moreover, such process is valuable because it fact out the bias when performing the process by a person. Thus AWS Quicksight should in theory produce more objective result.

To sum up, data visualization is an attempt by enterprise to mine information from data and present the information in a meaningful way. It is a data domain problem that falls in reporting and business intelligence category. Data visualization is challenging because both the fact that calculation involves massive data and the fact that produce proper presentation of data efficiently is complex. Data visualization is heavily relied on statistics and converge with statistics in many areas. COTS software is widely available and statistics softwares are acceptable alternatives with AWS Quicksight being quite competitive in this field.

References

Amazon Web Services, Inc. (n.d.). *Amazon quicksight.*

Few, S. (2004, 9). Eenie, meenie, minie, moe: Selecting the right graph for your message. *Perceptual Edge*.

Friendly, M. (2009, 8). Milestones in the history of thematic cartography, statistical graphics, and data visualization.