An Overview of Data Visualization

Mohaiminul Islam

School of Intelligent Technology and Engineering

Chongqing university of Science and Technology

E-mail: mohaiminul100@hotmail.com

Abstract — Data visualization is a general term that describes any effort to help people understand the significance of data by placing it in a visual context. Patterns, trends and correlations that might go undetected in text-based data can be exposed and recognized easier with data visualization software. Data visualization is the presentation of quantitative information in a graphical form. In other words, data visualizations turn large and small data-sets into visuals that are easier for the human brain to understand and process. Data visualizations are surprisingly common in our everyday life, but they often appear in the form of well-known charts and graphs. It can be used to discover unknown facts and trends. Good data visualizations are created when communication, data science, and design collide. Data visualizations done right offer key insights into complicated data-sets in ways that are meaningful and intuitive. In this article , we would like to discuss about data visualization, importance of data visualization, data visualization tools etc.

Keywords— Data Visualization, Tableau, Plot, Data Base, Analytics

1. Introduction

Today more than ever, organizations are using data visualizations, and data tools, to ask better questions and make better decisions. Emerging computer technologies and new user-friendly software programs have made it easy to learn more about your company and make better data-driven business decisions[1].

The strong emphasis on performance metrics, data dashboards, and Key Performance Indicators (KPIs) shows the importance of measuring and monitoring company data. Common quantitative information measured by businesses includes units or product sold, revenue by quarter, department expenses, employee stats and company market share. Data visualization tools go beyond the standard charts and graphs used in Microsoft Excel spreadsheets, displaying data in more sophisticated ways such as infographics, dials and gauges, geographic maps, sparklines, heat maps, and detailed bar, pie and fever charts. The images may include interactive capabilities, enabling users to manipulate them or drill into the data for querying and analysis. Indicators designed to alert users when data has been updated or predefined conditions occur can also be included.

Human brains have a much harder time understanding complex data when it is encoded in numbers and text compared to graphics. Dataviz techniques take advantage of that quality, helping to present massive amounts of information in the most efficient, visual form[2][3].

In fact, visually displayed data is easier to grasp and analyze, making it faster for decision makers to find patterns, including new and hidden, and to understand even difficult concepts. Charts, graphs, maps, dashboards—data visualization can be helpful in identifying issues and deficits, choosing the best product and business operation strategy, forecasting sales volume and stock prices, fine-tuning project management and resource administration, and so on and so forth.

2. Importance of data visualization

Data visualization has become the de facto standard for modern business intelligence (BI). The success of the two leading vendors in the BI space, Tableau and Qlik -- both of which heavily emphasize visualization -- has moved other vendors toward a more visual approach in their software. Virtually all BI software has strong data visualization functionality.

Data visualization tools have been important in democratizing data and analytics and making data-driven insights available to workers throughout an organization. They are typically easier to operate than traditional statistical analysis software or earlier versions of BI software. This has led to a rise in lines of business implementing data visualization tools on their own, without support from IT.

Data visualization software also plays an important role in big data and advanced analytics projects. As businesses accumulated massive troves of data during the early years of the big data trend, they needed a way to quickly and easily get an overview of their data. Visualization tools were a natural fit.

Visualization is central to advanced analytics for similar reasons. When a data scientist is writing advanced predictive analytics or machine learning algorithms, it becomes important to visualize the outputs to monitor results and ensure that models are performing as intended. This is because visualizations of complex algorithms are generally easier to interpret than numerical outputs[4].

3. Data visualization working method

Most of today's data visualization tools come with connectors to popular data sources, including the most common relational databases, Hadoop and a variety of cloud storage platforms. The visualization software pulls in data from these sources and applies a graphic type to the data.

Data visualization software allows the user to select the best way of presenting the data, but, increasingly, software automates this step. Some tools automatically interpret the shape of the data and detect correlations between certain variables and then place these discoveries into the chart type that the software determines is optimal[5].

Typically, data visualization software has a dashboard component that allows users to pull multiple visualizations of analyses into a single interface, generally a web portal.

Below is a chart forecasting tablet sales by operating system.

4. Data visualization tools

Data visualization tools provide data visualization designers with an easier way to create visual representations of large data sets. When dealing with data sets that include hundreds of thousands or millions of data points, automating the process of creating a visualization, at least in part, makes a designer's job significantly easier[6].

These data visualizations can then be used for a variety of purposes: dashboards, annual reports, sales and marketing materials, investor slide decks, and virtually anywhere else information needs to be interpreted immediately.

A. Tableau

Tableau has a variety of options available, including a desktop app, server and hosted online versions, and a free public option. There are hundreds of data import options available, from CSV files to Google Ads and Analytics data to Salesforce data.

Output options include multiple chart formats as well as mapping capability. That means designers can create color-coded maps that showcase geographically important data in a format that's much easier to digest than a table or chart could ever be[7].

The public version of Tableau is free to use for anyone looking for a powerful way to create data visualizations that can be used in a variety of settings. From journalists to political junkies to those who just want to quantify the data of their own lives, there are tons of potential uses for Tableau Public. They have an extensive gallery of infographics and visualizations that have been created with the public version to serve as inspiration for those who are interested in creating their own.

Pros

- Hundreds of data import options
- Mapping capability
- Free public version available
- Lots of video tutorials to walk you through how to use Tableau

Cons

- Non-free versions are expensive ($70/month/user for the Tableau Creator software)
- Public version doesn't allow you to keep data analyses private

Bottom Line

Tableau is a great option for those who need to create maps in addition to other types of charts. Tableau Public is also a great option for anyone who wants to create public-facing visualizations.

B. Infogram

Infogram is a fully-featured drag-and-drop visualization tool that allows even non-designers to create effective visualizations of data for marketing reports, infographics, social media posts, maps, dashboards, and more.

Finished visualizations can be exported into a number of formats: .PNG, .JPG, .GIF, .PDF, and .HTML. Interactive visualizations are also possible, perfect for embedding into websites or apps. Infogram also offers a WordPress plugin that makes embedding visualizations even easier for WordPress users[8].

Pros

- Tiered pricing, including a free plan with basic features
- Includes 35+ chart types and 550+ map types
- Drag and drop editor
- API for importing additional data sources

Cons

- Significantly fewer built-in data sources than some other apps

Fig.1. Infogram visualization

Bottom Line

Infogram is a great option for non-designers as well as designers. The drag-and-drop editor makes it easy to create professional-looking designs without a lot of visual design skill.

C. ChartBlocks

ChartBlocks claims that data can be imported from "anywhere" using their API, including from live feeds. While they say that importing data from any source can be done in "just a few clicks," it's bound to be more complex than other apps that have automated modules or extensions for specific data sources.

The app allows for extensive customization of the final visualization created, and the chart building wizard helps users pick exactly the right data for their charts before importing the data.

Designers can create virtually any kind of chart, and the output is responsive—a big advantage for data visualization designers who want to embed charts into websites that are likely to be viewed on a variety of devices.

Pros

- Free and reasonably priced paid plans are available
- Easy to use wizard for importing the necessary data

Cons

- Unclear how robust their API is
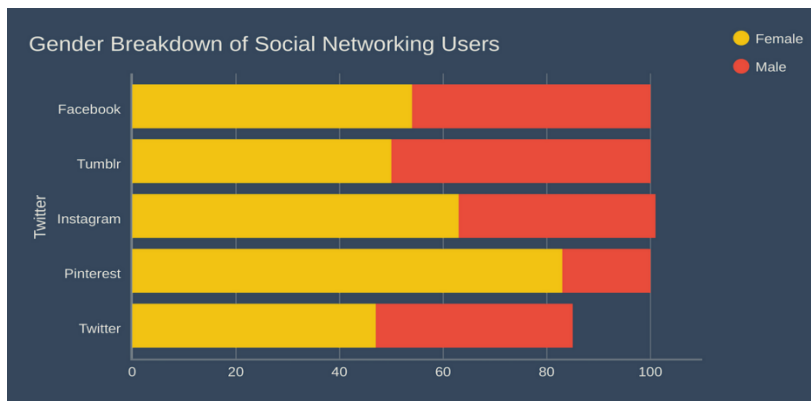- Doesn't appear to have any mapping capability

Fig.2.ChartBlocks visualization

Bottom Line

ChartBlocks has an excellent free plan, which is a big plus. The ease of use for creating basic charts and graphs is also outstanding.

D. Datawrapper

Datawrapper was created specifically for adding charts and maps to news stories. The charts and maps created are interactive and made for embedding on news websites. Their data sources are limited, though, with the primary method being copying and pasting data into the tool.

Once data is imported, charts can be created with a single click. Their visualization types include column, line, and bar charts, election donuts, area charts, scatter plots, choropleth and symbol maps, and locator maps, among others. The finished visualizations are reminiscent of those seen on sites like the New York Times or Boston Globe. In fact, their charts are used by publications like Mother Jones, Fortune, and The Times[9].

The free plan is perfect for embedding graphics on smaller sites with limited traffic, but paid plans are on the expensive side, starting at $39/month.

Pros

- Specifically designed for newsroom data visualization
- Free plan is a good fit for smaller sites
- Tool includes a built-in color blindness checker

Cons

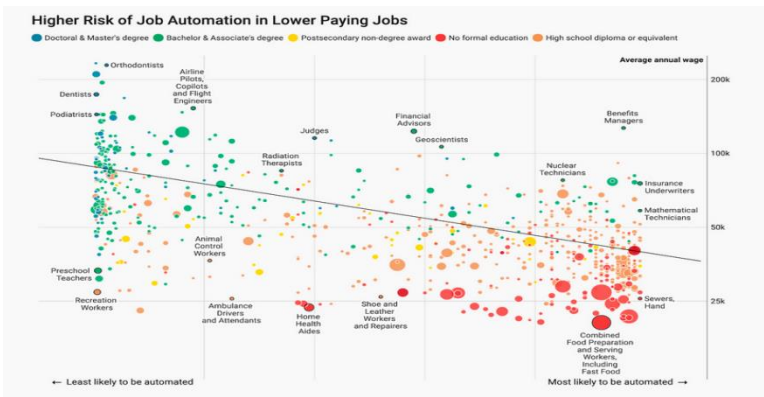- Limited data sources
- Paid plans are on the expensive side

Fig.3.Datawrapper visualization

Bottom Line

Datawrapper is an excellent choice for data visualizations for news sites. Despite the price tag, the features Datawrapper includes for news-specific visualization make it worth it.

E. Google Charts

Google Charts is a powerful, free data visualization tool that is specifically for creating interactive charts for embedding online. It works with dynamic data and the outputs are based purely on HTML5 and SVG, so they work in browsers without the use of additional plugins. Data sources include Google Spreadsheets, Google Fusion Tables, Salesforce, and other SQL databases.

There are a variety of chart types, including maps, scatter charts, column and bar charts, histograms, area charts, pie charts, treemaps, timelines, gauges, and many others. These charts can be customized completely, via simple CSS editing.

Pros

- Free
- Wide variety of chart formats available
- Cross-browser compatible since it uses HTML5/SVG
- Works with dynamic data

Cons

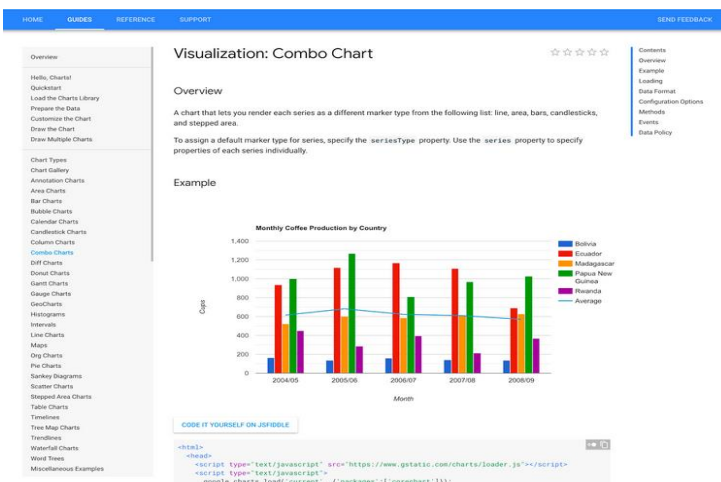- Beyond the tutorials and forum available, there's limited support

Fig.4. Google Charts data visualization

Bottom Line

Google Charts is a great option if a designer is somewhat comfortable with coding and wants a powerful, free solution. Being able to use any SQL database as a data source makes it a good option for large data sets, too[10].

These are the best data visualization tools at present.

5. Data visualization categories

A. Temporal

Data visualizations belong in the temporal category if they satisfy two conditions: that they are linear, and that they are one-dimensional. Temporal visualizations normally feature lines that either stand alone or overlap with each other, with a start and finish time.

Examples of temporal data visualization include:

- Scatter plots
- Polar area diagrams
- Time series sequences
- Timelines
- Line graphs

B. Hierarchical

Data visualizations that belong in the hierarchical category are those that order groups within larger groups. Hierarchical visualizations are best suited if you're looking to display clusters of information, especially if they flow from a single origin point.

The downside to these graphs is that they tend to be more complex and difficult to read, which is why the tree diagram is used most often. It is the simplest to follow due to its linear path[11].

Examples of hierarchical data visualizations include:

- Tree diagrams
- Ring charts
- Sunburst diagrams

C. Network

Datasets connect deeply with other datasets. Network data visualizations show how they relate to one another within a network. In other words, demonstrating relationships between datasets without wordy explanations.

Examples of network data visualizations include:

- Matrix charts
- Node-link diagrams
- Word clouds
- Alluvial diagrams

D. Multidimensional

Just like the name, multidimensional data visualizations have multiple dimensions. This means that there are always 2 or more variables in the mix to create a 3D data visualization. Because of the many concurrent layers and datasets, these types of visualizations tend to be the most vibrant or eye-catching visuals. Another plus? These visuals can break down a ton of data down to key takeaways.

Examples of multidimensional data visualizations include:

- Scatter plots
- Pie charts
- Venn diagrams
- Stacked bar graphs
- Histograms

E. Geospatial

Geospatial or spatial data visualizations relate to real life physical locations, overlaying familiar maps with different data points. These types of data visualizations are commonly used to display sales or acquisitions over time, and can be most recognizable for their use in political campaigns or to display market penetration in multinational corporations.

Examples of geospatial data visualizations include:

- Flow map
- Density map
- Cartogram
- Heat map

6. Data visualization market

The Data Visualization Market is segmented by Organization Department (Executive Management, Marketing, Operations), Deployment Mode (On-premise, Cloud/On-demand), End User (BFSI, IT and Telecommunication, Education), and Geography.
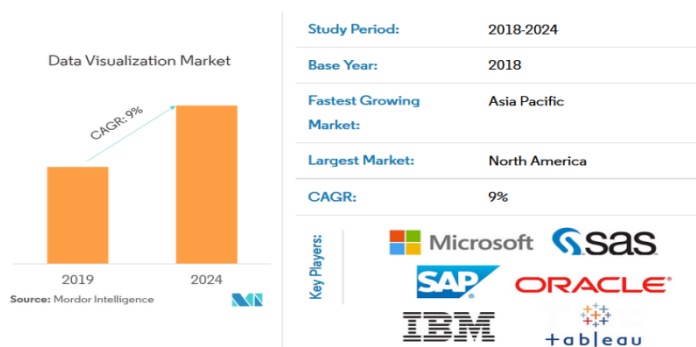
Market Snapshot

Fig.5. Snapshot

Market Overview

The global data visualization market is expected to register a CAGR of over 9% during the forecast period (2019-2024). The emerging nature of data visualization is encouraging a shift toward analytically driven businesses, where users can explore data in various forms of graphical representation, which were initially only available in tabular reports.

- Growth in big data, multi-device access to visualization (mobile, desktop, web, etc.), customized reports for specific roles, ongoing training for business users, and constantly changing optimizing dashboards.
- The increasing adoption of data visualization in small and medium enterprises is expected to fuel the growth of the data visualization market.
- Factors, such as the shortage of skills and budget and the lack of technical skills for report creation, are hindering the growth of the data visualization market[12].

Scope of the Report

Data visualization represents a wide range of business applications. Visualizing data is a concept that is used in different terms and technologies, including dashboards, infographics, scorecards, analytics, big data, and business intelligence. Data Visualization helps extract a meaningful report from the diverse and rapidly increasing volumes of business data. Many users believe that the usage of big data by data visualization helps in better decision making.

Organizational Department	Executive Management Marketing Operations Finance Sales Other Organizational Departments
Deployment Mode	On-premise Cloud/On-demand
End User	BFSI IT and Telecommunication Retail/E-commerce Education Manufacturing Government Other End Users
Geography	North America Europe Asia Pacific Latin America Middle East & Africa

Fig.6.Report

Key Market Trends

Retail to Witness Significant Growth

- The retail sector is witnessing a huge growth in the amount of data being generated from various channels, such as social sites, blogs, and apps. Most of the unstructured data produced is left unused, which if utilized, may provide valuable information.
- The incorporation of the Internet of Things (IoT) in the retail sector has increased over the past few years, through the increased utilization of devices, such as RFID and sensors.
- Additionally, retailers across the globe have been focusing on enhancing their connectivity, communication solutions, and devices which is generating a huge amount of big data[13]. This is further analyzed to study the pattern and purchasing behavior of a consumer which in turn is expected to propel the market growth over the forecast period.

North America to Hold Significant Share

- North America is expected to hold significant share as it has the United States which is a prominent market for data visualization. The US data visualization market is witnessing healthy growth, due to the increasing amount of data being generated in small- and medium-scale firms.
- Data visualization applications in North America are driven by factors, such as growth in big data and the need for analytics in the retail sector.
- An increasing need to track data from various sources and compiling it into various data representations, in order to make important business decisions, can be observed in North America, as this region is a major market for technology-based solutions.

Fig.7. Market Growth

Competitive Landscape

The data visualization market is moderately competitive. The players in the market are innovating in providing strategic solutions to increase their market presence and customer base. This is enabling them to secure new contracts and tap new markets.

- January 2019 - Tableau expands data prep capabilities with Tableau Prep Conductor, a product, which enables organizations to schedule and manage self-service data preparation. Tableau Prep Conductor is part of a new

subscription offering called "Tableau Data Management Add-On". This helps the customers to ensure that governed and timely data is available and ready for analysis[14].

- January 2018 - Tableau launched Hyper, new data engine technology included within Tableau 10.5. With Hyper's ability to slice and dice massive volumes of data in seconds, customers can see up to 5X faster query speed and up to 3X faster extract creation speed[15].
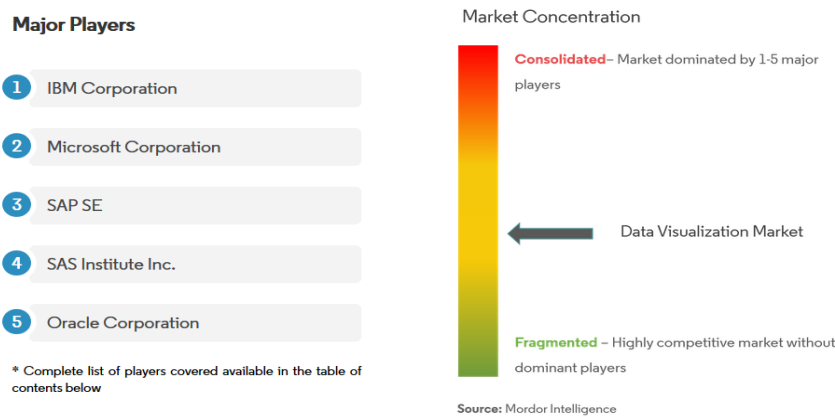

Fig.8, Market Concentration

7. The future of data visualization

Data visualization is entering a new era. Emerging sources of intelligence, theoretical developments and advances in multidimensional imaging are reshaping the potential value that analytics and insights can provide, with visualization playing a key role. The principles of effective data visualization won't change. However, nextgen technologies and evolving cognitive frameworks are opening new horizons, moving data visualization from art to science.

Looking back, much attention has been given to the principles of effective data visualization, such as substance, context and actionability. As timeless tenets that will continue to be important, regardless of medium or format, a brief review seems in order:

- Effective data visualization should be substantive And while creative visuals can enhance interest and memory, embellishment can't make up for lack of substance. According to purist Edward Tufte, "Every single pixel should testify directly to content."

- Visualization should be accurate and contextual. David McCandless's Billion Dollar O'Gram provides an example of how greater meaning can be added by incorporating the bigger picture. According to McCandless, "Absolute figures in a connected world don't give you the whole picture. They're not as true as they could be. We need relative figures that are connected to other data so that we can see a fuller picture."

- More than anything else, data visualization should facilitate decision-making, a goal that is difficult to achieve for many. According to a recent [KPMG study](International 2015), while data and analytics are deemed increasingly important to organizations, generating actionable insights remains a top challenge.

Conclusion

By combining data visualization best practices with modern digital technology, many companies that have to deal with massive amounts of information can quickly analyze it and get data-driven insights in order to streamline various aspects of their operation. All industries, and all companies regardless of size and scope of work can benefit from dataviz and business intelligence it brings to life.

Acknowledgement

References

[1] Manuela Aparicio and Carlos J. Costa (November 2014). "Data visualization". Communication Design Quarterly Review.

[2] Vitaly Friedman (2008) "Data Visualization and Infographics" in: Graphics, Monday Inspiration, January 14th, 2008.

[3] Tukey, John (1977). Exploratory Data Analysis. Addison-Wesley. ISBN 0-201-07616-0.

[4] Tufte, Edward (1983). The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press. ISBN 0-9613921-4-2.

[5] "Data Visualization for Human Perception". The Interaction Design Foundation. Retrieved 2015-11-23.

[6] Friendly, Michael. "A Brief History of Data Visualization". Springer-Verlag. Retrieved 19 November 2017.

[7] Funkhouser, Howard Gray (Jan 1936). "A Note on a Tenth Century Graph". Osiris. 1: 260–262.

[8] Sardar Mohkim Khan (26 January 2011). "DataMarket Expands Horizons: Adds 100 Million Time Series, 600 Million Facts".

[9] Tamara Munzner. "Process and Pitfalls in Writing Information Visualization Research Papers". www.cs.ubc.ca. Retrieved 9 April 2018.

[10] Benjamin B. Bederson and Ben Shneiderman (2003). The Craft of Information Visualization: Readings and Reflections, Morgan Kaufmann ISBN 1-55860-915-6.

[11] G. Scott Owen (1999). History of Visualization. Accessed Jan 19, 2010.

[12] Kosara, Robert (11 November 2013). "A Guide to the Quality of Different Visualization Venues". eagereyes. Retrieved 7 April 2017.

[13] Swayne, Deborah (1999). "Introduction to the special issue on interactive graphical data analysis: What is interaction?". Computational Statistics.

[14] Pavlopoulos, Georgios A.; Iacucci, Ernesto; Iliopoulos, Ioannis; Bagos, Pantelis (2013). Interpreting the Omics 'era' Data. Multimedia Services in Intelligent Environments. Smart Innovation, Systems and Technologies.

[15] Lucić V, Förster F, Baumeister W (2005). "Structural studies by electron tomography: from cells to molecules". Annual Review of Biochemistry. 74: 833–65.