

Web 3D Data Visualization of Spatio Temporal Data using Data Driven Document (D3js)

Alfred David
Department of Computer Science
Christ University
Hosur Road, Bangalore

Clarence J M Tauro
Center for Research
Christ University
Hosur Road, Bangalore

ABSTRACT

Data visualization is an important tool to analyze complex Spatio Temporal data. The web based data visualization tools are mostly 2D. A versatile tool would be one which will combine 2D and 3D data visualization techniques, as this would give an extra dimension, analytical usability and also be very easy to implement with a simple API.

The current technologies for such visualization are based on SVG, OpenGL, WebGL and other proprietary 3D libraries. These are useful but are difficult and cumbersome to use and the focus shifts from the data in question to be studied to methods and hacks of how the above listed tools can be manipulated to visualize the specific result patterns from the data.

This paper provides a direction of how to use a web based JavaScript library, which encompasses some of the technologies used above and provides an abstract interface to be able to use this effectively to visualize large complex sets of spatio temporal data. This will render a more intuitive, user focused, tactile navigation and selection for specific analysis of large set of

Spatio-Temporal data is illustrated in the choropleth example using D3.js on a spatio temporal database of the Edison electric company data of users and usage for New York City and its suburbs.

Keywords

Data visualization; spatio-temporal data; 3D visualization; visualization framework; D3.js

1. INTRODUCTION

The data visualization is a very important aspect of complex data analysis. It is that part wherein large amounts of complex data are bucketed, segregated, sliced and spliced to intelligent data business rules. These strata of diced and spliced data are spruced up in either a two-dimensional or a three dimensional representation which would enable an analyst or even a layman to understand the apparent interpretation. All this is now even more simplified when this analysis and graphical representation of data is done in a web based environment. Additionally, when the data is represented in three dimensions, the analytical interpretation is intuitive, flexible and actionable. The advancement in the visualization technology has revolutionized industries such as meteorological data, medicine, geographical data, global aid, repatriation etc. There are some inherent new challenges to the theory and the actual application of visualization of spatio temporal data. The reason being, the systems have not reached their full potential in a web-based environment. They are unable to render the current or expected system state information to decision makers in a tangible format. The system state information is usually drawn from multiple data source, which is to be merged. To be able to report the system

information from this large combined data set is a daunting task, which currently big data systems mechanism such as Map-Reduce may be able to solve, but for the decision maker to visualize this information to precisely concur on the data and make decision is a much larger problem. Currently, graphical maps are used but these cannot solve the complex multidimensional data relationships in an effective manner. They fail at the level of capturing temporal information or at the different levels of spatial resolution. Data from cities; of utilities, transportation system, and housing are rich data with significant geospatial and temporal characteristics that consist of large quantities of data from a variety of data sources. Information must be stored efficiently and integrated to a common framework. A mapping between the information elements and their spatial and temporal locations may be used to provide information in meaningful format. This information is nothing but service information. This can be complex, multidimensional, physical or abstract information that is intrinsically difficult to represent, manipulate and deduce actionable information from.

2. WEB 3D DATA VISUALIZATION OF SPATIO TEMPORAL DATA

2.1 Data Visualization

Visualization techniques can be broadly classified by whether the data display is 2D or 3D. 2D data representations of Geo-Spatial temporal data, like geographical and meteorological data are currently being widely used in numerous systems. With the advent of big data analysis, both statically and on the cloud there is an increasing interest in 3D representation of data, especially, a representation that can be universally accessed through the web. Both 2D and 3D visualization techniques have their inherent strength and weakness. In a scenario when large amounts of data has to be presented in a limited space, 2D visualization will bring in a cognitive overload thereby reducing usability and decision making. On the contrary 3D visualization uses an extra spatial dimension to create a virtual world where information is presented. However, delivering 3D content over the web is more technically challenging. The distribution and navigation of 3D content over the Internet are major obstacles that have so far prevented the popularization of 3D websites. A detailed 3D model generation is time consuming, fraught with a lot of manual work and also user interaction in 3D is not always intuitive.

The objective of this visualization framework is to provide a good foundation to integrate various data sources and generate 2D and 3D worlds in a seamless way. At the same time it is an endeavor to provide a highly interactive and intuitive graphical user interface to users. We also try and combine 2D and 3D to improve usability of the user interface to the maximum extent.

The library under consideration here, the D3js provides a good opportunity to display Geo-Spatial data on the web.

A generic framework, for a web based 3D Geo-Spatial temporal data visualization can be represented by the following diagram.

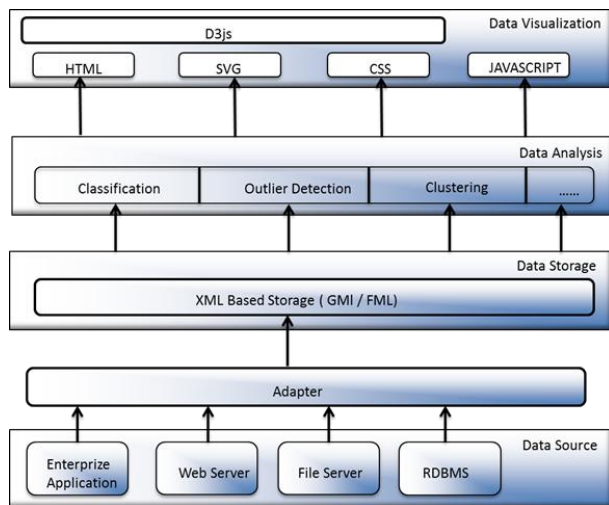


Fig 1: A generic framework, for a web based 3D Geo-Spatial temporal data visualization

This consists of four parts: Data source layer; Data Storage layer; Data Analysis layer; Visualization layer.

Data Source layer: The main challenge of a Geo-spatial data visualization process is to integrate and present multiple data sources and types in a comprehensible manner, so that users can extract some cognition from the data without being required to explicitly manage heterogeneous data. The data sources include enterprise information systems, file system servers, relational or NoSql database systems

Data Storage Layer: Data from different sources are in different formats. Different adapters are applied to them to convert them into a global and consistent format. XML (Xtensible markup Language) provides a common medium of data description and display for diverse systems to come together and understand each other e.g. GML (Geography Markup Language), FML (Financial Markup Language). Both these languages are xml-encoded languages representing geographical and financial data respectively. This encoding helps in storage, exchange and modeling information containing both spatial and non-spatial attributes of data.

Data Analysis Layer: Geo Spatial data is retrieved from the data storage layer and is further analyzed using data mining techniques depending on the requirements. **Visualization Layer:** The analysis results are presented in various visual formats and representations e.g. Chloropleth, Voronoi diagram, Hamiltonian graphs, no contiguous cartograms etc.

2.2 Visualization using D3js

D3js is an open source JavaScript library to manipulate documents based on data. In the background D3js helps bring data to life using HTML, SVG and CSS.

D3 allows you to bind arbitrary data to a Document Object Model (DOM) and then apply data-driven transformations to the document. For example, you can use D3 to generate an HTML table from an array of numbers. Or, use the same data to create an interactive SVG bar chart with smooth transitions and interaction.

D3 is not a monolithic framework that seeks to provide every conceivable feature. Instead, D3 solves the crux of the problem: efficient manipulation of documents based on data. This avoids proprietary representation and affords extraordinary flexibility, exposing the full capabilities of web standards such as CSS3, HTML5 and SVG. With minimal overhead, D3 is extremely fast, supporting large datasets and dynamic behaviors for interaction and animation. D3's functional style allows code reuse through a diverse collection of components and plugins.

D3 is not a new graphical representation. Unlike Processing, Raphaël, or Protovis, the built in features come directly from web standards: HTML, SVG and CSS. For example, you can create SVG elements using D3 and style them with external style sheets. You can use composite filter effects, dashed strokes and clipping. If browser vendors introduce new features tomorrow, you'll be able to use them immediately—no toolkit update required. And, if you decide in the future to use a toolkit other than D3, you can take your knowledge of standards with you!

Best of all, D3 is easy to debug using the browser's built-in element inspector: the nodes that you manipulate with D3 are exactly those that the browser understands natively.

Here the data under consideration for analysis is converted to a JSON format and then rendered to the D3js framework and the specific visualization technique is invoked to represent the data. Please do not revise any of the current designations.

3. CASE STUDY

3.1 Method

For this case study we've take raw data from the NYC data site of the Edison Electric Company and put it to analysis and visualization using the D3js framework. To visualize this data, we have used the Visual technique called chloropleth.

The Edison electric company data is a large set of non-segmented data for the purpose of actionable analysis,

```
"meta" : {  
  "view" : {  
    "id" : "74cu-ncm4",  
    "name" : "Electric Consumption by ZIP Code – 2010",  
    "attribution" : "Mayor's Office of Long Term Planning and Sustainability",  
    "averageRating" : 0,  
    "category" : "Environmental Sustainability",  
    "createdAt" : 1317246310,  
    "description" : "2010 electricity consumption in kWh and GJ, by ZIP code, building type, and utility company.",  
    "displayType" : "table",  
    "downloadCount" : 369,  
    "numberOfComments" : 0,  
    "oid" : 343463,  
    "publicationAppendEnabled" : false,  
    "publicationDate" : 1318360950,  
    "publicationGroup" : 241819,  
    "publicationStage" : "published",  
    "rowClass" : "",  
    "rowsUpdatedAt" : 1318358360,  
    "rowsUpdatedBy" : "txun-eb7e",  
    "signed" : false,  
    "tableId" : 244781,  
    "totalTimesRated" : 0,  
    "viewCount" : 2124,  
    "viewLastModified" : 1320856765,  
    "viewType" : "tabular",  
    "columns" : [ {  
      "id" : -1,  
      "name" : "sid",  
      "dataTypeName" : "meta_data",  
      "fieldName" : "sid",  
      "position" : 0,  
      "renderTypeName" : "meta_data",  
      "format" : {  
      }  
    }, {  
      "id" : -1,  
      "name" : "id",  
      "format" : {  
      }  
    }  
  ]  
  }  
}
```

Fig 2: Sample raw data from the NYC data site of the Edison Electric Company used for the analysis and visualization using the D3js framework

From the data in Fig 3, it is clear that the main challenge for an interactive 3D visualization is to integrate and present multiple data sources and types in a comprehensible manner, so that the users can extract some logical actionable information from the data without being required to explicitly manage the heterogeneous data. In the data of the Edison electric company, the multiple sources of data have been merged and the varied formats have been converted to a uniform JSON format.

This data still needs to be segmented. We've segmented the data as per the population, types of users and zip codes as show in the below Fig 3.



Fig 3: Segmented data as per the population, types of users and zip codes

This segmented data in JSON format is then put to the visualization process and subject to specific representations.

In this case study, I've used the Choropleth as the visual representation for the Edison electric company's segmented data, so that end users can visualize on the following parameters

- Connections based on different regional segments
- Connection densities based on different types of users
- The data can be color coded
- The data can be computed as being a Quantile, Linear, Logarithmic and Square Root.
- The Regional buckets can be varied and represented in the visualization.

In the custom application we choose the data set for analysis as show below Fig 4

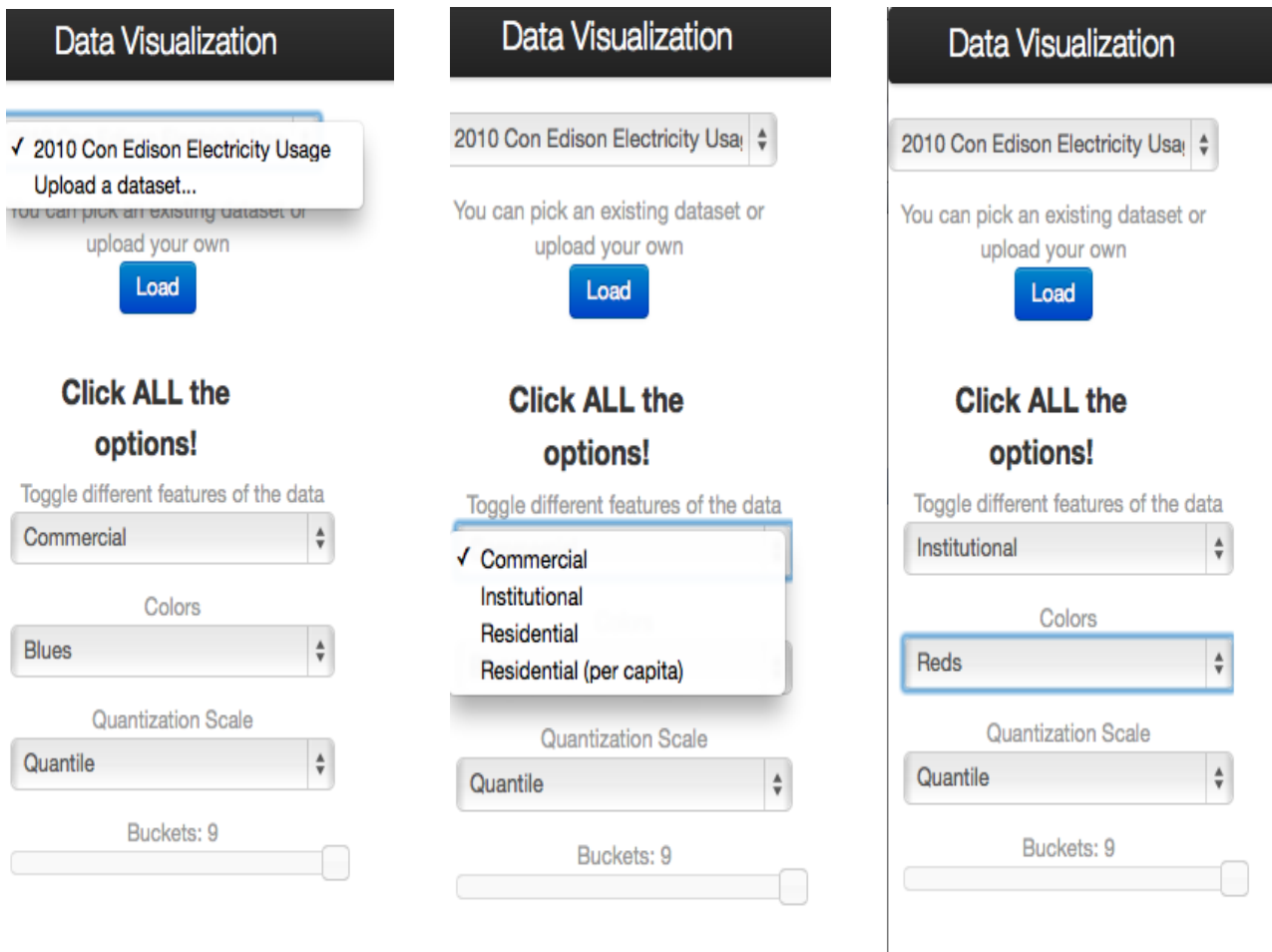


Fig 4: Choice of data set for analysis

Once we have chosen the data set, in this case the Edison electricity usage data; then from the below drop buttons we can choose the data we need to see the visualization of. e.g. Commercial, institutional, residential etc. by default it is commercial, this is customizable.

One can choose the color of the visualization so that the choropleth will be depicted in that color and its hues to be able to show the density modalities of the data set pertaining to the different area subsections of the choropleth map

Once the color of the visualization has been chosen, then it time to choose the quantization scale, which can compute as a quantile, linear, logarithmic or square root as shown in Fig 5.

Lastly, the user or the analyst can choose the degree of buckets he would want to see of the data on the visualization.

So from the above steps, a user has been able to bring in data and setup some visualization customizations, so that his analysis on the data can have multiple dynamic representations of the same data as opposed to a single static representation. Do not use abbreviations in the title or heads unless they are unavoidable.

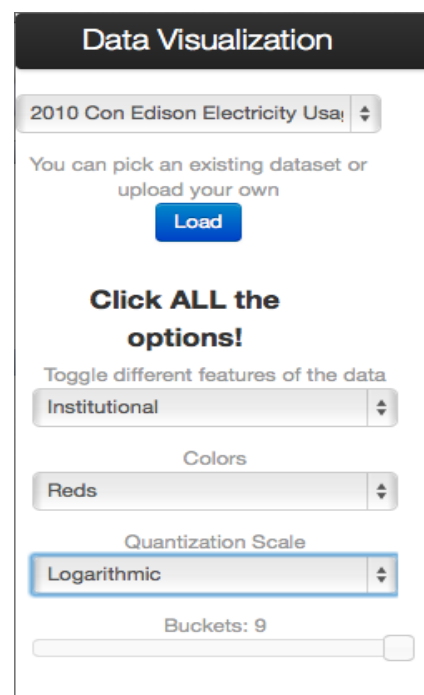


Fig 5: Choice of quantization scale

3.2. Result

As you can see in the below fig 6, there is a depiction of the data and if we look at the parameters of the visualization it is clear that, there are dimensions of the Edison electric company usage data. We are looking at the commercial connections spread across the new region, depicted in blue color visualization, with a quantile scale and depiction in a bucket of seven. These parameters can be changed

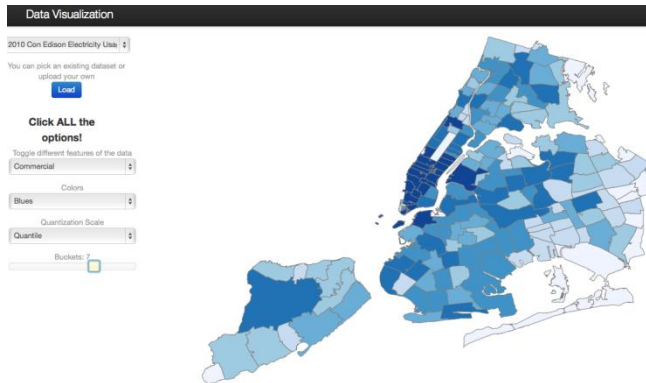


Fig 6. Depiction of the data

Likewise let's look at another snap shot of this data of institutional users and for showing the flexibility of the program lets tweak some of the parameters. As shown in the fig 7 below, here is a snapshot of visualization of the institutional users, represented in Green and its hues, with the quantization scale being square root and the buckets being nine.

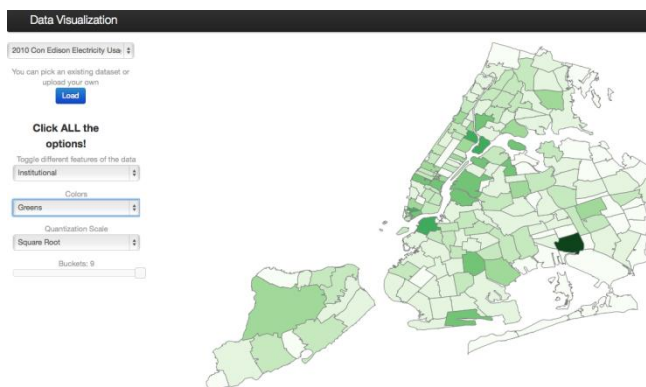


Fig 7. Snapshot of visualization of the institutional users and its hues, with the quantization scale being square root and the buckets being nine

4. CONCLUSION

Visualization of data is reaching its zenith potential, with the onset of big data. 3d visualization and interpretation of data is not only limited to data scientists and researchers alike.

Most organizations of different sizes are embarking on the analysis route to give themselves that extra bit of data, to fine tune their supply chain and market dynamics, thereby being on the cutting edge to minimize costs and maximize profits. In this realm, it becomes imperative that visualization of large sets of spatio temporal data needs to be done at much simpler level of implementation and also be used in a much less complex tool architecture, wherein data visualization can be undertaken even at small scale software services shops or individual data analyst level.

5. REFERENCES

- [1] Dongmei Yua,b Rentao Huang b GuanYan c, "Research on Statistics Data Spatio-Temporal Reasoning Model of Geographical Thematic Element based on Data Mining," Qing Dao University, 16 Qingda Road, Qing Dao, China, 266071,b Wu han University, 129 Luoyu Road, Wu han, China, 430079.c Science Press, Bei Jing, China, 100717.
- [2] Jianghui Ying, Denis Gracanic, and Chang-Tien Lu, "Web Visualization of Geo-Spatial Data using SVG and VRML/X3D," Proceedings of the Third International Conference on Image and Graphics (ICIG'04).
- [3] Hattori, S., "Spatio-Temporal Web Sensors by Social NetworkAnalysis," 2012 IEEE/ACM International Conference on Digital Object Identifier: 10.1109/ASONAM.2012.172.
- [4] Hui Yang ; Parthasarathy, S., "Mining Spatial and Spatio-Temporal Patterns in Scientific Data," Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on Digital Object Identifier: 10.1109/ICDEW.2006.92
- [5] Martin, S.C. ; Whitcomb, L.L. ; Arsenault, R. ; Plumlee, M. ; Ware, C., "A System for Real-Time Spatio-Temporal 3-D Data Visualization in Underwater Robotic Exploration," Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on Digital Object Identifier: 10.1109/ROBOT.2005.1570345.
- [6] Wang Yanbing, "3D GIS Spatial Modeling for City Surface and Subsurface Integration," Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on Digital Object Identifier: 10.1109/IGARSS.2006.391.
- [7] S. Shekhar, C. Lu, P. Zhang, and R. Liu, "Data Mining for Selective Visualization of Large Spatial Datasets", Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence, 2002.
- [8] J. Boye, SVG Brings Fast Vector Graphics to Web, <http://tech.irt.org/articles/js176/#2>.
- [9] A. E. Walsh and M. Bourges-Sévenier, Core Web3D, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.