Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Power-Consumption Analysis through Web-Based Visual Data Exploration

Daniel Pérez * **Ignacio Díaz** * **Abel A. Cuadrado** *
Francisco J. García-Fernández * **Alberto B. Diez** *
Manuel Domínguez **

* *Electrical Engineering Department, University of Oviedo, Campus de Viesques s/n, Gijón, Spain, 33204, (e-mail: {dperez, idiaz, cuadrado, fjgarcia, alberto}@isa.uniovi.es).*
** *Instituto de Automática y Fabricación, University of León, Campus de Vegazana, León, 24071(e-mail: manuel.dominguez@unileon.es)*

Abstract:
In recent years, the increasing capabilities of current technologies have made the acquisition and storage of large datasets an easy task. However, most times their unmanageable size, along with their complexity makes it a challenge to handle and analyze them. The emerging field of *visual analytics* relies on visualization principles and interactive interfaces to provide ways for amplifying human cognition, so that the information processing is improved and a better understanding of these datasets is achieved. In addition, current technologies to develop visual interfaces allow to develop powerful interaction mechanisms that enable the user to be an active part in the process. In this paper, data visualization foundations and web-based methods are considered for user-driven supervision tasks in decision-support systems. A real case consisting in the analysis of electric power demand in university buildings is presented via a web application developed using the recent data visualization library called D3. Time-series visualization and similarity maps are represented along with interactive techniques, allowing a dynamic data exploration.

Keywords: smart grids, monitoring and performance assessment, visual pattern recognition, supervision, energy expenditure, electric power systems, decision support systems

1. INTRODUCTION

Energy plays a key role in our society and makes it move. A proper management of energy, involving generation, distribution and demand, has a great impact in economy, but also on environment and health. Despite many efforts are being done in sustainable, renewable and clean energy production, acting on demand through a rational use of energy is probably one of the best niches to improve the energy balance. Most of the times, energy waste is motivated by a lack of control about how, when and where it is consumed.

Recent technologies and lower costs have led to an increasing number and quality of installed sensors that measure the power demand. However, most of the times this information ends up in a database or gets plotted in a simple trend graphic that shows only aggregate information and gives no insight about inefficiencies on the demand. In many cases, important decisions to achieve an efficient use of energy are taken with tight time by politicians or staff in charge of it, and long reports or spreadsheets are often misread. Information and knowledge about the patterns of demand in a public building or facility should be presented in a clear and concise way, so that the end user can quickly assess the demand condition on a given period and be assertive to make decisions or take corrective actions.

Currently, manipulating and analyzing *big data* are challenging tasks in several fields. One interesting approach to deal with large datasets is *visual analytics* that focuses on analytical reasoning facilitated by interactive visual interfaces, as is explained in Thomas and Cook [2005]. The use of certain visualization techniques supports the powerful human cognition to perform an efficient analysis. Visualization allows humans to extract knowledge from data, and reveal patterns (to support hypotheses or even to create new ones). Moreover, it provides not only a better understanding of the data but also an excellent via for communicating ideas to the rest, see Shneiderman [1996], Card et al. [1999], Ware [2012].

In addition, by means of interaction techniques, the integration of the human in the data analysis loop allows the use of the available domain knowledge and his/her own perception; an example is shown in Ahlberg and Shneiderman [1994]. A user-driven data analysis by means of interaction techniques helps the user to focus on the interesting aspects of data and improves the efficiency of the exploration with respect to a static view.

In the last years, web-based methods have become a powerful framework for implementing interactive displays by using the *document object model* (DOM). Recently, in Bostock et al. [2011] a novel approach, consisting of a javascript library called D3, was proposed allowing the direct manipulation of DOM, binding data to its elements, and animating them easily. It provides an efficient basis to develop dynamic visualizations on the web.

This paper aims at drawing the attention of the community of control to consider data visualization principles and techniques in applications such as supervision of complex processes or decision-support systems using web-based methods. An example built using D3 for power demand analysis in public buildings is presented here as a validation of the proposed approach. The remainder of the paper is: in Section 2 the requirements for an efficient analysis of power consumption are detailed; in Section 3 several works and visualization concepts are described; in Section 4 some tools are briefly reviewed, and the one used here is explained; in Section 5 a real case is shown for validating the previously explained techniques, and finally in Section 6 conclusions are summarized.

2. DESIGN REQUIREMENTS

Several features are desirable in order to exploit the user's knowledge for an efficient interpretation of power demand data.

A common problem in the analysis of power demand of a public building is the concurrence of different types of periodicities as well as many non-regular periods that stem from social activity, such as special events —like football matches, strikes, . . . — , holidays, etc . In addition, the power demand in a building can suffer variations due to weather or periods related to the nature of the activities carried out in it —e.g. examination periods in an educational institution. All this results in changes of the 24 hour day patterns of demand. The interface should provide an interactive, smooth and visual way to combine views showing different kinds of periodicities. For instance, it is desirable to know at a glance the hourly or yearly distribution of the demand on a given weekday. Despite this can be done using database query methods or spreadsheet operations, they imply a textual interaction that makes the exploratory process less interactive and far more ineffective.

The user should also be able to deal simultaneously with several views that tell different aspects of the problem. Existing visualizations do this mainly in two ways: a) showing all the views in the same screen, with a limited capacity to show many of them or b) allowing the user to add, remove views, or commute between them, which requires the user to locate again the interesting points or landmarks identified in the previous view. Despite methods such as *selection* and *brushing* can help to overcome this problem, changing between views remains being a "discontinuous jump" that obliges the user's mind to recompute visual landmarks that point to interesting information.

Finally, in approaches that use cluster analysis methods, like Van Wijk and Van Selow [1999], the selection of prototype day patterns is done in an automated way, with a low degree of human intervention. Since in power demand analysis many kinds of knowledge and information have to be managed –the structure of tariffs, the particular type of activity in the analyzed building, etc.– a more user-centric approach in the classification of daily demand profiles is a desirable feature.

In light of this, the interface should efficiently help the user to:

- *Identify regular temporal patterns.* The user should be able to quickly deploy a view showing the demand organized by hours of a day, by weekday or along the whole year. The user should also be able to get *combined views* as, for instance, hourly demand for each of the seven days of the week.
- *Identify special temporal patterns* such as holidays or singular events. One requirement for this is the use of specialized views such as a *calendar layout* where the spatial organization of time helps in finding special events (e.g. Easter holidays).
- *Identify groups of similar day patterns* of demand. Despite clustering methods can efficiently partition the data into groups of similar day patterns, the user may lose perspective on which clusters are similar to others and to what extent. Projections of the data on a 2D map can reveal continuous variations in the daily demand organized according to its similarity. This is performed by estimating manifold structures in data that the cluster methods would not describe properly.
- Establish mental *links* or *connections* between views. This can be achieved by multiple selection being displayed in multiple views. However, it should additionally strengthen these connections by means of a visual tracking mechanism that eliminates in a natural way the need to recognize specific points or selections of interest across the different views.

3. FOUNDATIONS

There are several works that have studied how the human perceive information. Bertin [1983], Cleveland and McGill [1984], Mackinlay [1986] are some representative examples in this field. Some books have also been written about this topic, some of them are Ware [2012] that shows a description of how human perception and cognition work; in Tufte [2001], several design principles for visual displays are proposed that are widely used by visualization practitioners; and in Few [2006], where dashboard design is exposed in detail.

There are different types of attributes inside a set of data, such as categorical, ordinal, or quantitative. Many visual channels can encode information such as position, color, or size, etc. Spatial position is the most accurate channel for the three data attributes, but the accuracy for other visual channels strongly depends on the data attribute to be represented.

3.1 Interaction and animation

Interaction is the essential part in the conception of data visualization interfaces, that provides dynamics to a static picture, allowing the user to manipulate representations in several ways and focus on the interesting aspects of data.

In Shneiderman [1996], the author proposed a taxonomy of low-level interaction techniques, including his mantra:

"*Overview first, zoom and filter, then details on demand*". There are more different works that put attention to interaction Yi et al. [2007]. Since reviewing the existing mechanisms is beyond the scope of this paper, we aim to show a brief description of the most relevant ones, used in this work:

- *Zooming* and *panning*, are simple affine transformations that allow the exploration of a large dataset by changing the scope and view in the same visual encoding.
- *Context information* allows the information related to one or several items to be checked by placing the mouse over them.
- *Brushing* relates the selection of a subset of elements in order to highlight them by a visual channel (color is usually used).
- *Linking* allows to provide a connection between the visualizations by highlighting in different views a subset of points or other visual elements –selected, for example by means of brushing in one of the views.
- *Animated transitions* is a method for perceiving changes between different visual encodings. Motion not only engages the viewer to different points of interest but also allows to track objects by means of their changes in order to communicate relationships between them. In Tversky et al. [2002] principles for effective animation are suggested and its benefits and drawbacks are discussed. In addition, in Heer and Robertson [2007] animated transitions are explored in order to improve graphical perception in statistical data graphics. Recently a framework that combines continuous transitions between 2D visual encodings with several interaction techniques was proposed in Diaz-Blanco et al. [2012]. Using this approach, a transition that depends on a single parameter $\lambda \in [0, 1]$, is possible between two views revealing meaningful intermediate states between them.

3.2 Reduction of data

Large datasets can produce visual cluttering that makes the user interpretation very difficult. One approach for reducing the size of data is *aggregation*, that consists in a single visual element representing a summary of many items. In cases of many dimensions, other strategy is *dimensionality reduction* (DR), that estimates the underlying structure of the multidimensional data into a reduced group of dimensions, that may be a combination of the original variables. For a reduction to a two dimensional space, the data can be represented in a projection of all the samples preserving similarity.

Many DR techniques have been developed, firstly the widely used principal component analysis (PCA) that linearly reduces the dimensions with maximum variance. Later, several methods have been used by non-linear approaches, such as neural networks like in Hinton and Salakhutdinov [2006] and probabilistic computations like in van der Maaten and Hinton [2008]. A detailed description of these methods can be found in Lee and Verleysen [2007].

4. METHODS

4.1 Data visualization tools

There are different types of data visualization tools, for instance software applications that produce interactive visualizations where your own set of data can be used, such as Tableau –Tableau [2013] –or sites like Many Eyes – Viegas and Wattenberg [2013]. Also Java-based graphic libraries, like the toolkit presented in Fekete [2004], Prefuse in Heer et al. [2005], or the Processing programming language, originally proposed in [Fry, 2004], can be used for supporting the implementation of advanced visualizations.

Regarding web-based tools, several toolkits have been developed. Examples are Processing.js (Processingjs [2013]), directly related to Processing and designed for the web, or Flare (Flare [2013]), that run in the Adobe Flash Player and was adapted from Prefuse. Moreover, Protovis which was proposed in Bostock and Heer [2009], provides a framework to map data attributes to visual elements. These examples show a variety of proposals for the development of web-based applications.

There are many advantages in web applications over desktop products:

- They are platform independent and no installation is needed, requiring only a modern web browser to run.
- They can combine the use of different technologies, such as widely accepted standards (HTML5, CSS3 or SVG). This allows the programmer to harness all their potential and resources, including lots of tunable controls, properties and events, to build the interface, as well as to benefit from a huge variety of open source libraries, ranging from date manipulation to numerical analysis.
- The highly optimized javascript interpreters built in today's browsers make the performance of these applications comparable to equivalent desktop applications.

On the downside, web-based approaches that run in the client side make it difficult to hide data and other resources to the end user in case they are confidential.

Similar to Protovis, in Bostock et al. [2011], the recent data visualization library called Data-Driven Documents (D3) extends this approach. D3 enables the pure manipulation and transformation of the standard Document Object Model (DOM) by mapping the data directly in it. In addition, D3 operators allow actions like modify content, select elements in correspondence with data, and the use of event listeners that enables interaction, and animated transitions can be derived by a collection of several interpolators over time.

Despite D3 can become slow because of the manipulation of large number of elements, most visualizations seldom require drawing simultaneously a huge number of visual elements on the screen. While future work is needed —e.g. a set of methods related to statistics would be worthy— D3 provides a standard representation improving the performance of previous approaches.

5. USE CASE

Here we explain an example for a data analysis using the principles and techniques explained above. A web prototype, implemented using D3, is shown for the exploration of power consumption in two buildings at the University of Oviedo. Firstly a description of data is detailed, then different data representations are explained, and finally the included interaction techniques are described.

5.1 Power demand dataset

The original dataset was retrieved from the data logging system, covering a timespan of one year, with a sample period of 15 minutes. A 4:1 sample reduction of the dataset was done by averaging the power consumption in an hourly basis, resulting in 8760 records (365 × 24). The data variables consisted of active, P, reactive, Q and apparent S power consumption in two university buildings. In addition, the power factor, $\cos\varphi$, and a residual, R, were computed as

$$\cos\varphi = \frac{P}{S} \tag{1}$$

$$R = S^2 - P^2 - Q^2 \tag{2}$$

Note that the residual R should be identically zero under ideal conditions ($S^2 = P^2 + Q^2$), with no harmonic distorsions. This attribute helps in revealing deviations from ideal conditions such as nonconventional loads.

5.2 Visual encodings

The main visualization presented here is performed by using different scatterplots, where relationships are revealed efficiently. Let's consider an *encoding* as a set of N points [1] ,

$$\mathbf{p}_A = \{\mathbf{p}_A(1), \mathbf{p}_A(2), \cdots, \mathbf{p}_A(N)\}, \quad \mathbf{p}_A(i) \in \mathbb{R}^m \tag{3}$$

whose points $\mathbf{p}_A(i)$ are descriptive of interesting information on sample i to be spatially described, such as, for instance, the 2D position in a wall calendar or in a clock-like representation of the timestamp of sample i. Thus, this position of the points encodes similarity information but also its size and color can also encode a different attribute. The encodings used in the case presented in this paper are:

"Clock" encodings We considered daily, weekly and yearly point sets, $\mathbf{p}_D$, $\mathbf{p}_W$, $\mathbf{p}_Y$, containing 2D points distributed in a circular way, with a period of one day, one week and one year, respectively:

$$\mathbf{p}_D(i) = \left[\cos\left(2\pi\frac{h(i)}{24}\right), \sin\left(2\pi\frac{h(i)}{24}\right)\right] \tag{4}$$

$$\mathbf{p}_W(i) = \left[\cos\left(2\pi\frac{d(i)}{7}\right), \sin\left(2\pi\frac{d(i)}{7}\right)\right] \tag{5}$$

$$\mathbf{p}_Y(i) = \left[\cos\left(2\pi\frac{h(i)}{365\cdot 24}\right), \sin\left(2\pi\frac{h(i)}{365\cdot 24}\right)\right] \tag{6}$$

[1] We shall typically consider *spatial encodings* consisting of 2D points in this paper ($m = 2$). However encodings of higher dimensions can be used in case the point set describes a 3D scatter or if other visual encodings, such as color and size, are under consideration.

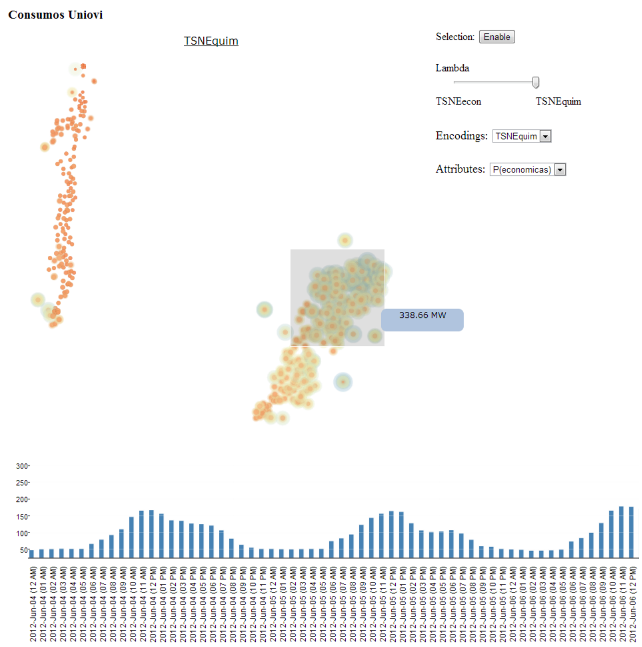

Fig. 1. Screenshot of the web-based tool showing scatterplot and barchart views

This way to encode time, already considered in related contexts in the literature —see Ohlsson et al. [1994], González and J. [2005], Shen and Ma [2008]— allows to display the data in a clock-like fashion that, on one hand, is a widely accepted convention of time representation and on the other hand provides a natural way to aggregate periodic events.

Calendar encoding A calendar encoding assigns each sample its position in a conventional wall calendar. All power demand samples corresponding to a same day will lay in the same point of the calendar in this view. The user can easily get an aggregate value of one or more days by simply selecting the points in the calendar.

The specific calendar view accounts for specific social time granularities such as the irregular number of days in a month, weekends and holiday periods. These irregular periods of time are extremely relevant to power demand analysis and cannot be properly described by classical analytical methods, such as Box-Jenkins or Fourier based methods.

365 × 24 matrix encoding Another useful encoding to represent data is a matrix representation of the items where rows show each day in the year and columns show the hours of that day. This allows the user to get, in a snapshot, variations in the 24-hour patterns of demand along the year.

Dimensionality reduction encoding The positions are the result of a 2D projection computed using a DR technique. In this case, *t-Distributed Stochastic Embedding* (t-SNE) method was used, that is an effective technique for visualizing high-dimensional data, and can be found in van der Maaten and Hinton [2008].

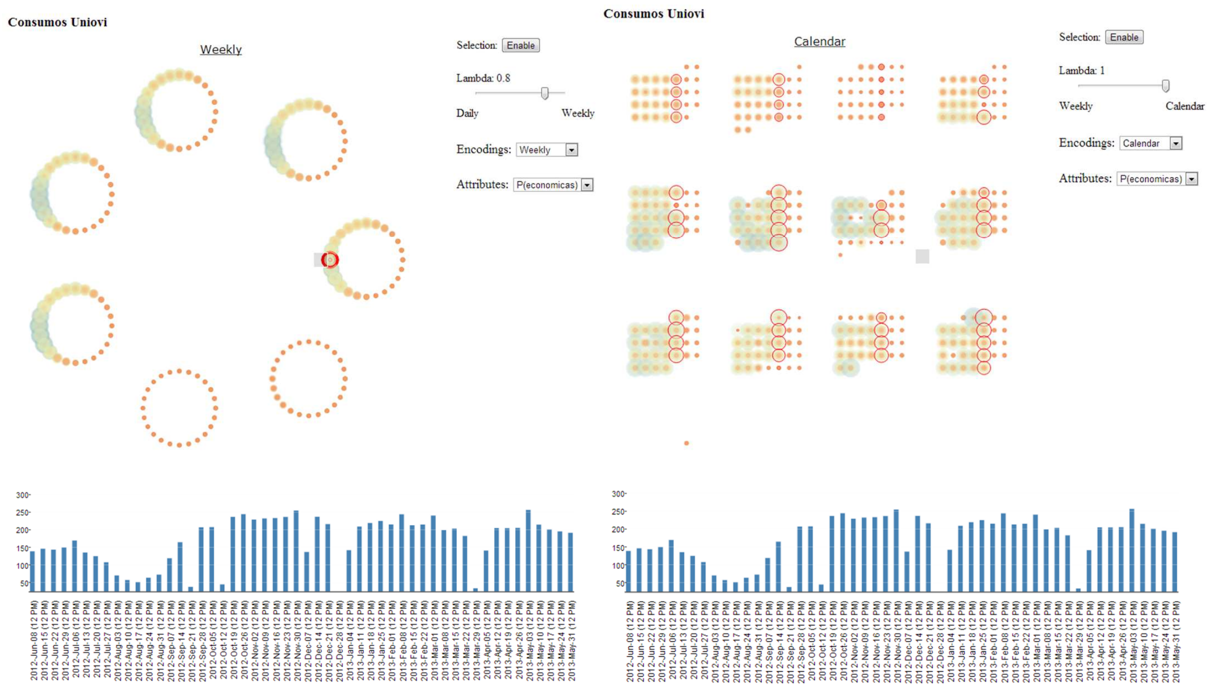

Fig. 2. Intermediate state between weekly and daily mappings (left) and calendar view with a subset of points selected (right)

5.3 Interaction techniques

D3 makes low-level interaction techniques easily available to be implemented. Transformations of the views, such as zooming and panning, improve a detailed exploration of the points. Furthermore, the user can choose several points on the map by selecting a brushing area, then a barchart visualization of the selected points is represented to perform a detailed comparison between them. A tooltip gives information about the point. If there is a selection of several points, another tooltip provides the sum of the current attribute, e.g. active power, for the selected points (see Fig. 1).

The morphing operation, explained in Diaz-Blanco et al. [2012], consists in mixing two or more encodings into a new encoding. The morphing operation between two encodings is quite straightforward. Let $\mathbf{p}_A(i)$ and and $\mathbf{p}_B(i)$ be two different encodings for $i = 1, \cdots, N$. Let also $\lambda \in [0, 1]$ be a mixing coefficient. The morphing operation between $\mathbf{p}_A$ and $\mathbf{p}_B$ would be

$$\mathbf{p}(i,t) = \lambda(t)\mathbf{p}_A(i) + (1 - \lambda(t))\mathbf{p}_B(i) \tag{7}$$

for $i \in \{1, \cdots N\}$. Here an input control is integrated in the interface so that the user can change $\lambda(t)$ to produce a variable "mixture" of two different encodings $\mathbf{p}_A$, $\mathbf{p}_B$ by evaluating the combination for all the N points ($N = 8760$). In other words, the user can steer animated transitions by combining any two encodings with manually tunable proportions, allowing to navigate between different representations in a smooth way. This enables the user to keep visual track of the elements during the transition, establishing links between them. Moreover, a proper mixed selection results in a meaningful intermediate state. For instance, mixing a "weekday encoding" with a "daily encoding" gives rise to a new "hour of a weekday" encoding, as shown in Fig. 2 (left). This encoding, shows all 7×24 combination of days and weekdays so a given point would aggregate all power consumption along the year for a particular weekday at a particular hour.

The mix of the different visual mappings and the integration of the interaction techniques give the capability of analyzing on demand different periodical time intervals by means of visual queries. For example, in Fig. 2 (left) a visual selection (Fridays at 12 pm) can be seen with the points highlighted in red for the exploration of the consumption in these points. Furthermore, keeping this highlight in an encoding change allows linking between encodings for the selected points, see Fig. 2 (right) calendar view with the former points selected.

6. CONCLUSION

In this paper, we consider dynamic data visualizations for the exploration of complex processes and decision-support systems using web-based methods. Several visualization principles and techniques are explained for the analysis of multidimensional data. A proper visual representation of the information joined with different interaction mechanisms improve the data exploration tasks allowing the user to perform a quick analysis of large datasets. Moreover, the use of web-based tools makes the access easy for more people, and flexible for using with modern browsers. The recent javascript library D3 enables direct DOM manipulation, the native integration with different standards webs and simplifies the use of animation and interaction mechanisms.

A use case is shown using real data of power consumption acquired from two university buildings. An interactive web interface, implemented using D3, is presented where several scatterplots for the data are available. These views represent different visual encodings, such as time-series or similarities, depending on the positioning of the points whose size and color encode one certain and selectable variable of the data.

Furthermore several interaction techniques, such as zooming or brushing, are included in the application. The possibility that different information related to the selection of a subset of data can be displayed, and that animated transition between views can be controlled gradually by the user, show an efficient exploratory experience.

ACKNOWLEDGEMENTS

This work has been financed by the Spanish Ministerio de Economía y Competitividad (MINECO) and FEDER funds.

REFERENCES

Christopher Ahlberg and Ben Shneiderman. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 313–317. ACM, 1994.

Jacques Bertin. *Semiology of graphics.* University of Wisconsin Press, 1983. ISBN 0299090604.

Michael Bostock and Jeffrey Heer. Protovis: A graphical toolkit for visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1121–1128, 2009.

Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D^3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.

Stuart Card, Jock Mackinlay, and Ben Shneiderman. *Readings in information visualization. Using vision to think.* Morgan Kaufmann Publishers, San Francisco, 1999.

William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, 1984.

Ignacio Diaz-Blanco, Manuel Dominguez-Gonzalez, Abel Cuadrado-Vega, Alberto Diez-Gonzalez, and Juan Fuertes-Martinez. MorphingProjections: Interactive Visualization of Electric Power Demand Time Series. Diaz-Blanco et al. [2012], pages 121–125. doi: 10.2312/PE/EuroVisShort/EuroVisShort2012/121-125.

J-D Fekete. The infovis toolkit. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 167–174. IEEE, 2004.

Stephen Few. *Information dashboard design.* O'Reilly, 2006.

Flare. `http://flare.prefuse.org/`, November 2013.

Benjamin Jotham Fry. *Computational information design.* PhD thesis, Massachusetts Institute of Technology, 2004. AAI0806331.

P. González and Zamarreño J. Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings*, 37:595–601, 2005.

Jeffrey Heer and George G Robertson. Animated transitions in statistical data graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1240–1247, 2007.

Jeffrey Heer, Stuart K Card, and James A Landay. Prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2005.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

John A. Lee and Michel Verleysen. Nonlinear dimensionality reduction, 2007.

Jock Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

M.B.O. Ohlsson, C.O. Peterson, H. Pi, T.S. Rognvaldsson, and B.P.W. Soderberg. Predicting system loads with artificial neural networks–methods and results from" the great energy predictor shootout". *ASHRAE Transactions-American Society of Heating Refrigerating Airconditioning Engin*, 100(2):1063–1074, 1994.

Processingjs. `http://processingjs.org/`, November 2013.

Z. Shen and K.L. Ma. Mobivis: A visualization system for exploring mobile data. In *Visualization Symposium, 2008. PacificVIS'08. IEEE Pacific*, pages 175–182. IEEE, 2008.

Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

Tableau. `http://www.tableausoftware.com/`, November 2013.

James J Thomas and Kristin A Cook. *Illuminating the path: The research and development agenda for visual analytics.* IEEE Computer Society Press, 2005.

Edward R. Tufte. *The Visual Display of Quantitative Information.* Graphics Press, second edition, 2001. ISBN 0961392142.

Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International journal of human-computer studies*, 57(4):247–262, 2002.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

J.J. Van Wijk and E.R. Van Selow. Cluster and calendar based visualization of time series data. In *Infovis*, page 4. Published by the IEEE Computer Society, 1999.

Fernanda Viegas and Martin Wattenberg. Many eyes. `http://www.many-eyes.com/`, November 2013.

Colin Ware. *Information visualization: perception for design.* Elsevier, 2012.

Ji Soo Yi, Youn ah Kang, John T Stasko, and Julie A Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6): 1224–1231, 2007.