# Creating a Thesis for Fun and Profit

BY

Carmen St. Jean

B.S., University of New Hampshire (2010)

THESIS PROPOSAL

Submitted to the University of New Hampshire
in partial fulfillment of
the requirements for the degree of

Master of Science

in

Computer Science

May 2014

# Table of Contents

# List of Figures

# Chapter 1

# Introduction

Fishery managers have only one "lever" to pull when it comes to fishery management: the ability to set harvest quotas. Fishermen work within these quotas by exerting various levels of fishing effort. Both managers and fisherman can better understand the ecosystem they work within and the implications of their decisions with the assistance of a production model, which is a system of differential equations designed to predict outcomes. A visualization may enhance such a model by making its inner workings more explicit and may be useful for decision making. We hope to explore design alternatives and evaluate the effectiveness of different modes of portrayal and interaction to make a visualization of the MS-PROD model that will be a valuable tool to the modelers and stakeholders alike.

# Chapter 2

# Background

## 2.1 The Model

Both the short- and long-term effects of human exploitation on an ecosystem such as a fishery are not easily understood since experiments which would allow ecosystem managers to investigate the impact of different levels of exploitation over many years are either impractical or impossible to conduct on a large scale. Fortunately, ecosystem models can be used instead to help gain a better understanding of an ecosystem.

Ecosystem models are abstract representations of an ecological system, which could mean anything from an individual species to an entire community of species. A classic example is the Lotka-Volterra model, which is a pair of differential equations for describing the non-linear interactions between a predator species and a prey species. [16, 23]:

$$\frac{dN_1}{dt} = N_1 (\alpha - \beta N_2) \tag{2.1}$$

$$\frac{dN_2}{dt} = -N_2 (\gamma - \delta N_1) \tag{2.2}$$

where $N_1$ is the number of prey, $N_2$ is the number of predator, $t$ is time, $\alpha$ is the prey's growth rate, $\beta$ is the rate at which the predator destroys the prey, $\gamma$ is the death rate of the predator, and $\delta$ is the rate at which the predator increases from consuming the prey. The model can be generalized to discuss an arbitrary number of species rather than just a single pair.

The Lotka-Volterra model can be modified to take competition instead of predation into account, as in the Rosenzweig-MacArthur model [19] and the Leslie-Gower model [15].

These adaptations to the model also consider carrying capacity, which is the maximum number of a species that can sustained indefinitely:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \left(\frac{N_1 + \alpha_{12} N_2}{K_1}\right)\right) \tag{2.3}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \left(\frac{N_2 + \alpha_{21} N_1}{K_2}\right)\right) \tag{2.4}$$

where $r_i$ is the growth rate for species $i$, $K_i$ is the carrying capacity for species $i$, and $\alpha_{ij}$ is the effect species $j$ has on species $i$. As with Lotka-Volterra, this model concerns only two species, but it can be generalized to include more than two.

Both Lotka-Volterra and Leslie-Gower do not incorporate a factor which is critical when discussing fisheries management: the effect of harvest. The Schaefer model adds a term to account for the effect of harvest on an individual species [21]:

$$\frac{dN}{dt} = rN \left(1 - \left(\frac{N}{K}\right)\right) - qEN \tag{2.5}$$

where $N$ is the number (or biomass) of the species, $r$ is the growth rate, $K$ is the carrying capacity, $q$ is the catchability coefficient, and $E$ is the fishing effort.

Simple models, when available and correct, are generally preferred; since fewer components are needed to describe their real-world counterparts, they are more easily understood and implemented. All three of these models are subjectively simple in that they only consider a few ecological factors each. However, ecosystems are complex systems which require management that recognizes them as such [3]. Thus, a more holistic approach called ecosystem-based fishery management (EBFM) has been advocated [17]. However, this approach has not often been implemented due to a lack of models which consider all necessary ecological factors. Gamble and Link developed a multispecies production model (MS-PROD) to fill this gap [7].

The MS-PROD model is built upon the Schaefer production model by also including Lotka-Volterra terms for predation, Leslie-Gower terms for competition, and carrying capacities for functional groups ($K_G$) as well as for the entire system ($K_\sigma$):

$$\frac{dN_i}{dt} = r_i N_i \left( 1 - \frac{N_i}{K_G} - \frac{\sum_{j=1}^{g} \beta_{ij} N_j}{K_G} - \frac{\sum_{j=1}^{G} \beta_{ij} N_j}{K_\sigma - K_G} \right) - N_i \sum_{j=1}^{P} \alpha_{ij} N_j - H_i N_i \tag{2.6}$$

where $N_i$ is the number (or biomass) of species $i$, $t$ is a unit of time, $r_i$ is growth rate for species $i$, $\beta_{ij}$ is the interaction of species $j$ on $i$, $\alpha_{ij}$ is the predation of species $j$ on $i$, $H_i$ is the harvest rate on species $i$, $g$ is the number of species within species $i$'s group, $G$ is the number of groups, and $P$ is the number of predators.

This model is distinguished from other multispecies production models by describing stocks with explicit ecological and harvest factors. Ten key species were chosen from the Northeast United States Continental Shelf Large Marine Ecosystem (NEUS LME) from four major functional groups. Given an input parameter set of initial biomass values, a predation matrix, an interaction matrix, catchability values, and harvest effort values, the MS-PROD model runs simulations for 30 years with an annual time step to predict individual biomasses. While this outputted information is potentially valuable to fishery managers, it was lacking an interactive graphical user interface.

## 2.2 Visualization Methods

### 2.2.1 Time Series

Fisheries management is focused on the sustainability of choices concerning fish stocks. A main purpose of ecosystem management is to ensure future generations can enjoy the same natural resources [3]. As such, the MS-PROD model provides biomass forecasts for 30 years. Therefore, time-oriented visualization techniques must be explored.

Frank discussed the different types of time-oriented data by defining three criteria: a) linear vs. cyclic, b) time points vs. time intervals, and c) ordered time vs. branching time vs. time with multiple perspectives [5]. The MS-PROD data consist of discrete time points with definite starting and ending points, falling under the linear time and time points

categories. As for the third criterion, the model outputs one ordered time result for a single execution, however fisheries managers may like to compare alternative scenarios. Therefore, our domain falls into the branching time category, though in a limited sense. Aigner et al. point out that techniques for visualizing branching time and time with multiple perspectives are unfortunately uncommon [1], but a line chart—a technique frequently used for ordered time data—serves as a good starting point.

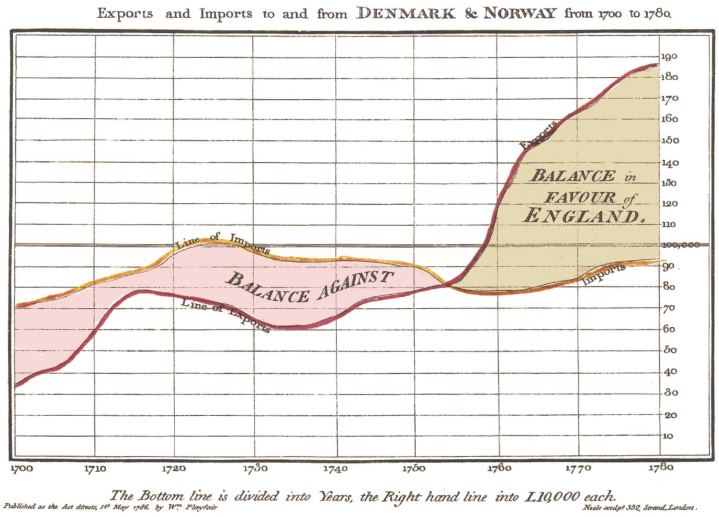
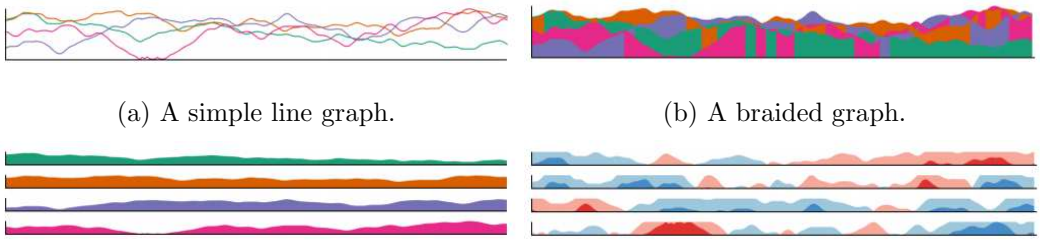



Figure 2-1: Playfair's original time series chart [18].

The line chart was first invented by William Playfair in 1786 to communicate time series data, seen in Figure 2-1 [18]. Today, it remains a common method for visualizing time-oriented data in many fields, including science, economics, planning, and engineering to name a few. Line charts typically encode time on the horizontal axis, progressing from left to right, and some time-varying value on the vertical axis. Points in the chart are connected by line segments such that the slope of the line indicates the rate of change between time steps.

Multiple time series can be part of a single line chart; each series needs only to be distinguished by a color and/or line style. However, as the number of time series on a single



(a) A simple line graph.

(b) A braided graph.

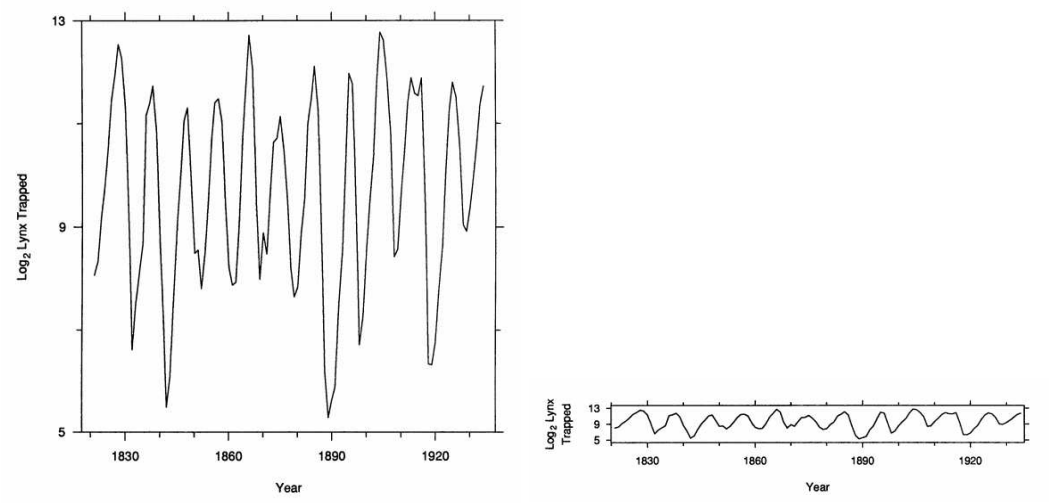
(c) Small multiples.

(d) Horizon graphs.

Figure 2-2: Four possible methods for visualizing multiple time series [13].

line chart increases, it becomes more difficult to identify an individual series. Javed et al. studied different plotting techniques for multiple time series, as seen in Figure 2-2 [13]. The first of their techniques is the "simple line chart," which was Playfair's original line chart with all series plotted together. A slight variation on that is "small multiples," where each series had its own line chart though all charts share the same axis scales. Horizon graphs, originally developed by Saito et al., wrap around a baseline in two color tones to save space [20]. Lastly, braided graphs feature all series on one chart with the coloring under the curves alternating as series intersect each other. The user evaluation by Javed et al. revealed that a simple line graph with all time series on one plot or a single graph for each time series is better suited to a variety of tasks than a horizon graph or a braided graph. They also found that users complete tasks more correctly when there is more display space allocated to the graphs. They did not recommend using a higher number of simultaneous time series—their study used eight at the most—because it also leads to a decline in correctness of task completion.

Some line charts are more effective at conveying the nature of the data than others because of the way different drawing techniques affect interpretability. Cleveland et al. found the shape of a line chart—defined as the height of the chart divided by the width of the chart—to be a critical factor [4]. Shape of the chart directly impacts the slopes of line segments, which viewers interpret in order to understand the dependence of the $y$ variable on the $x$ variable. Figure 2-3a, a time series of Canadian lynx trapping data, features a



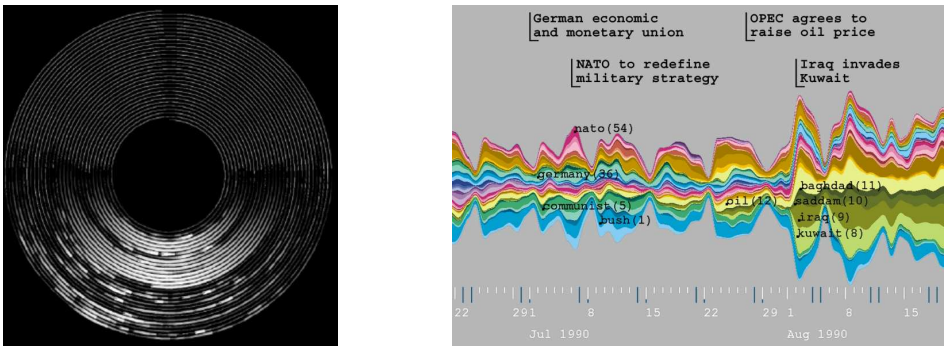

(a) Shape parameter of 1.0.

(b) Shape parameter of 0.074.

Figure 2-3: Two effect of chart shape on Canadian lynx data [4].

shape of 1.0 and seems to imply rapid increases and decreases in the population. On the other hand, Figure 2-3b has a shape of 0.074 and shows more clearly that the population rises somewhat steadily and declines somewhat rapidly, which Figure 2-3a failed to show. Their user evaluation found that judgment of two slopes is influenced by the orientation midangle, defined as the average of the minimum slope orientation and the maximum slope orientation. They proprosed line chart shape should be selected such that orientations are as close to $\pm 45°$ as is possible, like in Figure 2-3b.

There are many alternatives to and variations of Playfair's original time series. One example is Weber et al.'s spiral time series, seen in Figure 2-4a [25]. The spiral time series was designed for cyclic data. Cycles are emphasized in a properly-parameterized spiral visualization, however it may be difficult to describe periodic behavior in unknown datasets or determine if that behavior even exists. Another example is the ThemeRiver by Havre et al, seen in Figure 2-4b [10]. Each "current" in the ThemeRiver represents an entity or subject and must be of a distinctive color. Positioning along the y-axis is meaningless, instead the abundance of the entity or subject over time is indicated by the width of the current.




(a) Spiral time series.

(b) ThemeRiver time series.

Figure 2-4: Two alternative time series visualizations.

### 2.2.2 Networks

The input parameters to the MS-PROD model includes predation and interaction matrices. These are relationships which may be better understood if incorporated into the visualization. Relationships are often visualized through a node-link diagram, which typically represents entities as nodes and links as relationships between the nodes they connect. There are many types of node-link diagrams used for illustrating networks, of which only a few are discussed in the following sections.

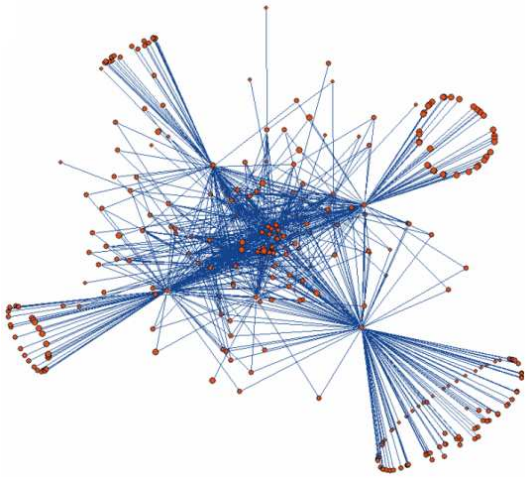


Figure 2-5: A force-directed visualization of a food web of Gulf of Alaska data [6].

One option for showing fish species interactions would be to use a force-directed layout as Gaichas and Francis did, seen in Figure 2-5 [6]. Here, the nodes represent an individual species in the Gulf of Alaska, while the links represent a predator-prey interaction. In a force-directed layout, nodes repel each other, while related nodes become pulled toward each other by links [11]. The result is an aesthetically pleasing layout where there are relatively few link crossings and links are of approximately similar length. The color of the node can be used to indicate group membership, while the size can represent the magnitude of some property of the node. Likewise, the drawing style of the link can be varied to encode different types of relationships.

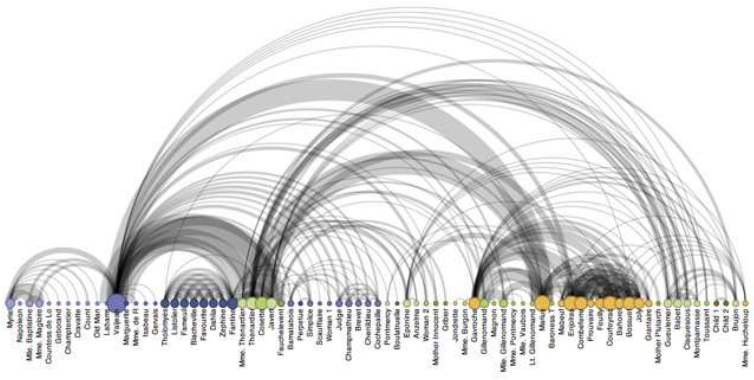


Figure 2-6: Knuth's arc diagram of *Les Misérables* characters [14].

An alternative for force-directed layout is an arc diagram. First coined by Wattenberg [24], arc diagrams were used by Knuth to illustrate interaction of characters in Victor Hugo's novel *Les Misérables*, seen in Figure 2-6 [14]. Each character is represented with a circular node, where size indicates the number of appearances. The nodes are arranged linearly, colored and ordered according to clusters of characters that appear together frequently. Semi-transparent arcs are drawn between the characters which appear in the same chapter, with the thickness of the arc representing the number of such appearances. While the arc diagram may fail to properly depict the structure of a network, Heer et al. point out it is advantageous because the one-dimensionality allows for other features to be easily displayed

near the nodes [11].

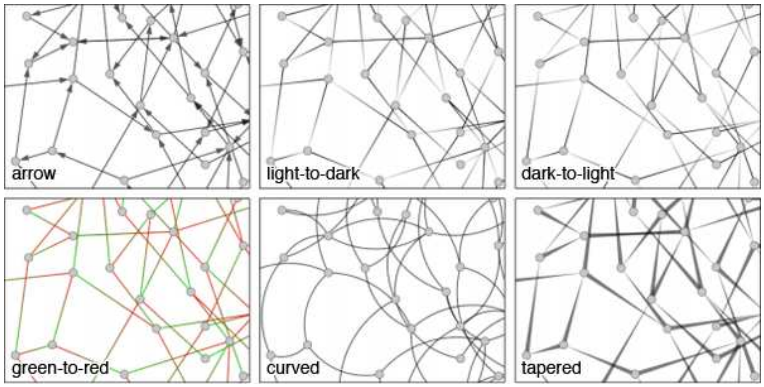



Figure 2-7: Different types of directed edges [12].

Relationships in a network may be directional, such as the predator-prey relationship. In a visualization of such a network, the direction of the edges must be encoded so these relationships can be understood. Holten and van Wijk studied the effectiveness of different techniques for indicating directionality of edges in a graph, seen in Figure 2-7 [12]. The traditional arrowhead was found to perform poorly, while a tapered edges performed best. As for an intensity-based direction cue, a dark-to-light representation was found to be clearer than light-to-dark.

Node-link diagrams can have occlusion problems when they are highly-connected, so a matrix-based representation of a network is a possible alternative [11]. In many cases, networks are stored as an adjacency matrix, so all that needs to be done is visualize that matrix as a grid, where the cell at the $i$th row and the $j$th column represents the relationship from entity $i$ to entity $j$. Figure 2-8 shows Knuth's visualization of *Misérables* characters in matrix-form [14]. The color of the cell indicates the presence or type of a relationship, with some neutral color indicating the lack of a relationship. Ghoniem et al. showed that a matrix-based view is suitable for large or dense networks for tasks that involve finding or counting links or nodes [8]. With proper ordering of the rows and columns, the structure of the network can be effectively displayed, however path-finding tasks may be difficult.

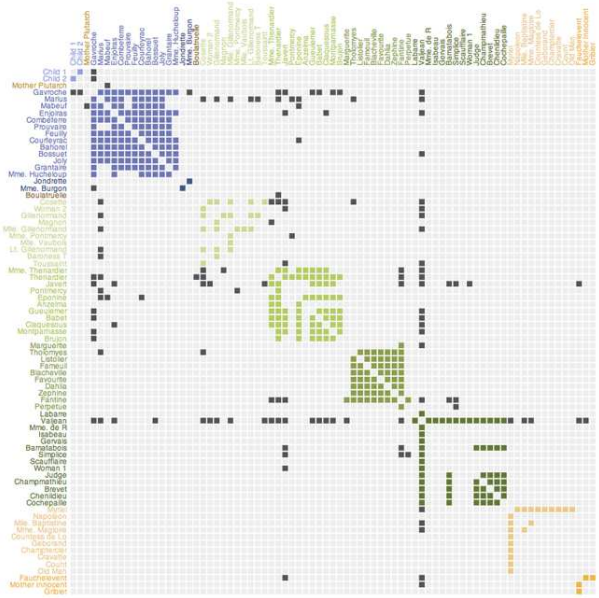
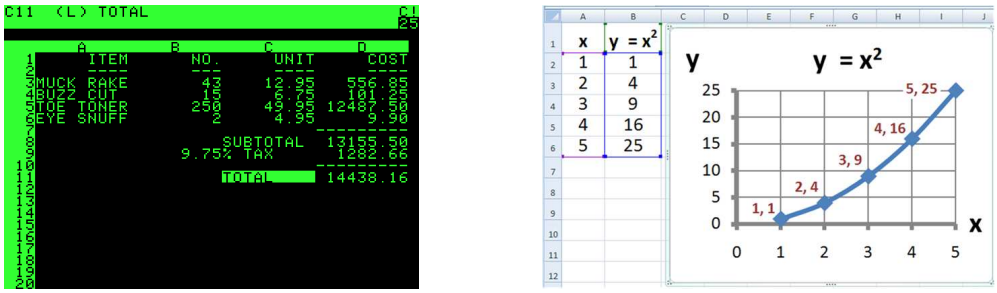


Figure 2-8: A matrix-based visualization of an adjacency matrix [14].

## 2.3 Understanding Models

Models of all kinds—ecological, engineering, statistical, etc.—can benefit from the aid of a visualization. With even a simple model, patterns and trends may be difficult, if not impossible, to discern from only a table of numerical values. A simple chart may be all that is necessary to elucidate the mathematics behind a model. The learning process can be even further enhanced through interaction with the model. Users can adjust parameter values, perceive a change (or perhaps no change) in the results, and begin to understand the degree of influence different parameters possess.

VisiCalc was a very early example of software assisting the understanding of models [9]. As a business student, Bricklin wished there was a faster way to change the input or fix mistakes when working out financial models by hand [2]. To address this, he worked with Frankston to develop VisiCalc, seen in Figure 2-9a. As the world's first electronic spreadsheet, VisiCalc consisted of rows and columns containing either text, numerical values, or formulas. Result cells were instantly updated according to changed inputs or adjusted formulas, allowing a user to work with models in a more efficient and dynamic manner.




(a) A screenshot of VisiCalc (GNU General Public License).

(b) A chart made using Microsoft Excel (public domain).

Figure 2-9: Two examples of spreadsheet applications.

VisiCalc was superseded by Lotus 1-2-3, which was in turn supplanted by Microsoft Excel. Microsoft Excel remains popular and features graphing tools which can generate charts, such as in Figure 2-9b.

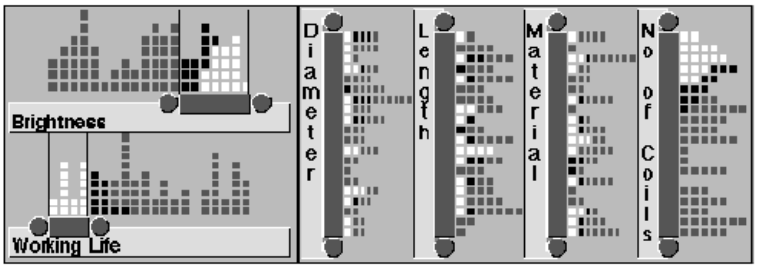



Figure 2-10: A screenshot of the Influence Explorer [22].

The Influence Explorer by Tweedie et al. is a good example of an interactive visualization [22]. They developed an interface for understanding the relationships between different attributes in a design process. Parameter values of the Influence Explorer are initially randomly selected to represent different possible items. For each attribute, there is a histogram including each of the items. The attribute ranges are controlled by sliders. When the user adjusts the slider of a given attribute, all items that are within that range are highlighted on all of the histograms. Figure 2-10 is a screenshot of the Influence Explorer being used to test the performance of different light bulb designs; white indicates the design passed, black

it failed one specification, and grey it failed two specifications. Industrial designers found the ability to interactively explore the effects of different parameter ranges to be valuable.

# Bibliography


[1] Wolfgang Aigner, Silvia Miksch, Wolfgang Müller, Heidrun Schumann, and Christian Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14:47–60, 2008.

[2] Dan Bricklin and Bob Frankston. VisiCalc: Information from its creators, Dan Bricklin and Bob Frankston. `http://bricklin.com/visicalc.htm`, 1999. [Online; accessed 2013-10-07].

[3] Norman L. Christensen, Ann M. Bartuska, James H. Brown, Stephen Carpenter, Carla D'Antonio, Rober Francis, Jerry F. Franklin, James A. MacMahon, Reed F. Noss, David J. Parsons, Charles H. Peterson, Monica G. Turner, and Robert G. Woodmansee. The Report of the Ecological Society of America Committee on the Scientific Basis for Ecosystem Management. *Ecological Applications*, 6(3):665–691, August 1996.

[4] William S. Cleveland, Marylyn E. McGill, and Robert McGill. The shape parameter of a two-variable graph. *Journal of the American Statistical Association*, 83(402):pp. 289–300, 1988.

[5] Andrew U Frank. Different types of "times" in GIS. *Spatial and Temporal Reasoning in GIS*, pages 40–62, 1998.

[6] Sarah K. Gaichas and Robert C. Francis. Network models for ecosystem-based fishery analysis: a review of concepts and application to the Gulf of Alaska marine food web. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(9):1965–1982, 2008-09-01T00:00:00.

[7] Robert J. Gamble and Jason S. Link. Analyzing the tradeoffs among ecological and fishing effects on an example fish community: A multispecies (fisheries) production model. *Ecological Modelling*, 220(19):2570 – 2582, 2009.

[8] Mohammad Ghoniem, Jean-Daniel Fekete, and Philippe Castagliola. A comparison of the readability of graphs using node-link and matrix-based representations. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '04, pages 17–24, Washington, DC, USA, 2004. IEEE Computer Society.

[9] B. Grad. The creation and the demise of VisiCalc. *Annals of the History of Computing, IEEE*, 29(3):20–31, 2007.

[10] S. Havre, B. Hetzler, and L. Nowell. ThemeRiver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization*, pages 115–123, 2000.

[11] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. A tour through the visualization zoo. *Commun. ACM*, 53(6):59–67, June 2010.

[12] Danny Holten and Jarke J. van Wijk. A user study on visualizing directed edges in graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2299–2308, New York, NY, USA, 2009. ACM.

[13] Waqas Javed, Bryan McDonnel, and Niklas Elmqvist. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):927–934, November 2010.

[14] Donald E. Knuth. *The Stanford GraphBase: a platform for combinatorial computing.* ACM, New York, NY, USA, 1993.

[15] P. H. Leslie and J. C. Gower. The properties of a stochastic model for the predator-prey type of interaction between two species. *Biometrika*, 47:219–301, 1960.

[16] Alfred J. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16(12):317–323, 1926.

[17] United States. National Marine Fisheries Service. Ecosystem Principles Advisory Panel. *Ecosystem-based fishery management: a report to Congress*. U.S. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, 1999.

[18] William Playfair. *The Commerical and Political Atlas: Representing, by means of stained copper-plate charts, the progress of the commerce, revenues, expenditure and debts of England during the whole of the Eighteenth Century*. 1786.

[19] M. L. Rosenzweig and R. H. Macarthur. Graphical Representation and Stability Conditions of Predator-Prey Interactions. *The American Naturalist*, 97(895):209+, January 1963.

[20] Takafumi Saito, Hiroko Nakamura Miyamura, Mitsuyoshi Yamamoto, Hiroki Saito, Yuka Hoshiya, and Takumi Kaseda. Two-tone pseudo coloring: Compact visualization for one-dimensional data. *IEEE Symposium on Information Visualization*, 0:23, 2005.

[21] Milner B. Schaefer. A study of the dynamics of the fishery for yellowfin tuna in the eastern tropical pacific ocean. *Inter-American Tropical Tuna Commission Bulletin*, 2(6):243–285, 1957.

[22] Lisa Tweedie, Robert Spence, Huw Dawkes, and Hua Su. The influence explorer. In Jim Miller, Irvin R. Katz, Robert L. Mack, and Linn Marks, editors, *CHI 95 Conference Companion*, pages 129–130. ACM, 1995.

[23] Vito Volterra. Fluctuations in the abundance of a species considered mathematically. *Nature*, 118:558–560, 1926.

[24] Martin Wattenberg. Arc diagrams: Visualizing structure in strings. In *Proceedings of the IEEE Symposium on Information Visualization*, INFOVIS '02, pages 110–116, Washington, DC, USA, 2002. IEEE Computer Society.

[25] M. Weber, M. Alexa, and W. Muller. Visualizing time-series on spirals. In *IEEE Symposium on Information Visualization*, pages 7–13, 2001.