Chaomei Chen (2004) *Information Visualization: Beyond the Horizon.* Springer. pp. vii-ix

Foreword

It is with enthusiasm and excitement that I join the community of information visualization researchers and designers in celebrating our still fresh accomplishments of the past decade. However, even as we take pride in how far we have come, we should acknowledge that these are just the first steps of a much longer journey.

This book and the rich literature from conferences, journals, and a few pioneering books reveals a flourishing, but still emerging academic field, that fights for recognition every day. Similarly, the product announcements from new and mature companies, demonstrate the passionate commitment of venturesome entrepreneurs who struggle to cross the chasm to commercial success.

Readers of the academic literature and corporate press releases probably believe that the allure of information visualization is in finding appropriate representations of relationships, patterns, trends, clusters, and outliers. This belief is reinforced by browsing through conference titles that weave together technical topics such as trees, networks, time series, and parallel coordinates, with exotic verbs such as zoom, pan, filter, and brush. However, I believe that the essence of information visualization is more ambitious and more compelling; it is to accelerate human thinking with tools that amplify human intelligence.

Chaomei Chen captures the spirit of this emerging academic discipline in this second edition and cleverly uses knowledge domain visualization to trace the growth and spread of topics. His survey highlights the dramatic progress during the past five years in a way that celebrates and challenges researchers and developers. His numerous screen shots of research and commercial systems give a glimpse of what is possible, but readers will have to see the demos for themselves and view working products to get the full impact of the interaction dynamics.

Chen's book shows us how the rapidly maturing information visualization tools are becoming as potent as the telescope and microscope. A telescope enabled Galileo to see the moons of Jupiter, and a microscope made it possible for Pasteur to see bacteria that enabled him to understand disease processes. Similarly, remarkable technologies such as radar, sonar, and medical scanners extend human vision in powerful ways that facilitate understanding. The insights gained provide support for air traffic controllers, naval officers, physicians, and others in making timely and effective decisions.

The payoffs to users of information visualization tools will be in the significant insights that enable them to solve vital problems at the frontiers of their fields. By extending their vision to higher dimensional spaces, users of information visualization tools are making meaningful and sometimes surprising breakthroughs. These users, such as genomic researchers, financial analysts, or patent lawyers, are often struggling to understand the important relationships, clusters, or outliers hidden in their data sets. Their quest may last days or years as they seek identify surprising groupings hidden among naturally occurring combinations or distinguish novel trends from well-understood seasonal variations. The outcome may be to discover

secondary functions of known genes, or stocks that will outperform others in their industry group.

The users' goals are often noble, valuable, and influential. Which sets of genes limit cancer growth? Which stock movements are often precursors of a major market rise? Which companies are distinctively active in developing new patents in wireless applications for e-commerce? In other circumstances, the users of information visualization deal with difficult topics such as tracking epidemics, uncovering fraud, or detecting terrorists.

The process of information visualization is to take data available to many people and to enable users to gain insights that lead to significant discoveries. Chen appropriately focuses attention on how information visualization techniques "make the insights stand out from otherwise chaotic and noisy data." The often noisy data must be cleaned of anomalies, marked for missing values, and transformed in ways that are more conducive to insight and discovery. Then users can choose the representations that suit their tasks best. Next users can adjust their view by zooming in on relevant items and filtering out unnecessary items. Settings of control panels may have to be changed to present the items in appropriate colors, positions, shapes, orientation, etc.

Some parts of this process can be automated, and some data mining or statistical algorithms can be helpful, but often the insight comes to those who have a hypothesis to test or who suspect a novel relationship. Visualizations are especially potent in promoting the intuitions and insights that lead to breakthroughs in understanding the relevant connections and salient features.

Typically, the quest for understanding requires looking at the details of an outlier or a surprising correlation. At that point, the benefit of domain knowledge and the need for more data becomes strong. Chen's practical examples illustrate this process and the role of domain knowledge, especially in the case of detecting abrupt changes and emerging trends. Only the experienced geneticist can make the leap to recognize how a raised level of gene expression signals its participation in a meaningful biological pathway. Only the knowledgeable stock market analyst recognizes that the reason for a sudden rise in value is due to a successful marketing trial of a new product.

There are three implications of the situated nature of information visualization that will influence future research and the success of products: (1) input data usually needs to be cleansed and transformed to support appropriate exploration (2) related information is often needed to make meaningful judgments, and (3) effective presentation of results is critical to influence decisionmaking.

Sources of input data need to be trusted and possibly consulted to understand its meaning and resolve inconsistencies. Then these data can be cleansed of anomalies, transformed to appropriate units, and tagged for missing values. Sometimes data needs to be aggregated to an appropriate level of analysis, such as web log data that is grouped by session, by hour, or by domain name.

The source data may need to be supplemented by related information to provide context for decisions. For example, sales data that records customer zip codes, may only become meaningful when the zip code demographics, geographic location, or income distribution is accessible. It will be no surprise that ski equipment is sold

heavily in mountain states, but the surprising insight may be the high level of sales in wealthy southern cities. Similarly, genomic researchers need to know how a tight cluster of highly expressed genes relates to the categories of molecular function in the gene ontology. Stock market analysts will want to understand why a group of stocks rose and then fell rapidly by studying recent trading patterns and industry news reports.

Since effective presentation of results is critical to influence decisionmaking, designers must understand how users collaborate. The first step is simply recording the state of a visualization by allowing the saving of settings. Other important services are to support extraction of subsets, posting results to a web page, and producing high quality printed versions. Chen reports on the collaborative environments that allow simultaneous viewing of a shared display, accompanied by a synchronous chat window, voice conversation, or instant messaging, are increasingly common. Asynchronous environments with web-based discussion boards, are also important as they better support larger communities, where coordination for a synchronous discussion is difficult. Chen deals with this topic, as well as the visualization of group processes in online communities.

These three aspects of effective information visualization are in harmony with Geoffrey Moore's analysis in his insight-filled book *Crossing the Chasm* (1991). His formula for successful software products is that they are "whole product solutions" which solve a known problem with an end-to-end solution (no additional components needed). He cautions that training has to be integrated, benefits have to be measurable, and users have to be seen as heroes. Many early products failed to adhere to this formula, but newer offerings are in closer alignment.

Researchers can also learn from this formula, because it encourages a practical approach. Prof. Fred Brooks, long ago encouraged researchers to focus on a "driving problem." His advice remains potent, especially for those who are entranced with colorful animated displays and elaborate statistical manipulations. Explorers of the vastness multidimensional spaces are more likely to make important discoveries if they keep their mind's eye focused on solving their driving problem. They are also more likely to experience those wonderful Aha! moments of insight that are the thrill of discovery.

Then researchers and developers will need to get down to rigorous evaluations. Chaomei Chen places a strong emphasis on empirical studies to help researchers and developers get past their understandable infatuation with their innovations. Rapid progress will be made as more evaluations are done using benchmark tasks and standard data sets, coupled with carefully reported in-depth case studies of collaborations with problem solvers in many disciplines.

There's work to be done. Let's get on with it!

Ben Shneiderman
University of Maryland