A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations

Robert Amar[*], John Stasko[†]
College of Computing/GVU Center
Georgia Institute of Technology
Atlanta, GA USA 30332-0280

ABSTRACT

The design and evaluation of most current information visualization systems descend from an emphasis on a user's ability to "unpack" the representations of data of interest and operate on them independently. Too often, successful decision-making and analysis are more a matter of serendipity and user experience than of intentional design and specific support for such tasks; although humans have considerable abilities in analyzing relationships from data, the utility of visualizations remains relatively variable across users, data sets, and domains. In this paper, we discuss the notion of *analytic gaps*, which represent obstacles faced by visualizations in facilitating higher-level analytic tasks, such as decision-making and learning. We discuss support for bridging the analytic gap, propose a framework for design and evaluation of information visualization systems, and demonstrate its use.

CR Categories: H.5.0 [Information Interfaces and Presentation]: General; J.0 [Computer Applications]: General

Keywords: Information visualization, analytic gap, theory, framework, evaluation, knowledge tasks.

1 INTRODUCTION

The modern line of thought on effective presentation of information, espoused strongly by Tufte and others, is that good data speak for themselves [23]. In this sense, Tufte is mainly discussing the creation of static presentations built to convey a message around a particular selected subset of data. Information visualization has grown up around this principle, with the added charge of exploring the benefits of interaction with such displays.

Shneiderman's mantra of "Overview first, zoom and filter, details-on-demand" [17] nicely summarizes the design philosophy of modern information visualization systems, including better-known commercial tools such as Spotfire (2D/3D scatterplots) [22], Eureka (tables with fisheye views and value bars, now the Inxight Table Lens) [7], and InfoZoom (tabular zooming and overview browser) [11]. Beginning with graphical and tabular constructs, these systems provide broad overviews of data sets, support selection and examination of individual data, and provide facilities for dynamic query.

While most recent work on the design and evaluation of information visualization systems typically centers on faithful correspondence of representation to data, there remains uncertainty about the ability of current systems to adequately support decision making, for three reasons we shall discuss separately: limited affordances, predetermined representations, and the decline of determinism in decision-making.

[*] e-mail: bob@cc.gatech.edu
[†] e-mail: stasko@cc.gatech.edu

IEEE Symposium on Information Visualization 2004
October 10-12, Austin, Texas, USA
0-7803-8779-1/04/$20.00 ©2004 IEEE

1.1 Limited Affordances

The operations afforded by many visualization systems are equivalent to very simple database queries. The operations at which these systems excel tend to be those which their default displays and dynamic query interactors afford: simple sorting; filtering; approximate two-dimensional correlation. A recent study by Kobsa finding that users achieved only 68%-75% accuracy on simple questions involving some common commercial systems indicates that even these operations have room for improvement [13]. While such operations can be useful for initial exploration of data sets, decision makers are beginning to rely more and more on macro-level, statistical properties of data sets, as we will discuss below.

1.2 Predetermined Representations

The representations employed by common visualizations are not particularly agile, supporting the formation of simplistic, static cognitive models from elementary queries on typically historical, cross-sectional data. If a user's visualization software supports scatterplots but a contour map is really desired or needed, then a different package must be used. Recently, a number of visualizations that address a specific domain or problem area have emerged ([9, 19, 24] being examples from the InfoVis '03 Symposium); while they can be very effective, they raise the question of whether each new domain requires a new visualization.

1.3 Decline Of Determinism In Decision-Making

Finally, and most importantly, we live in a world that is not only dominated by information, but uncertainty. A growing number of business schools are shying away from information-centric, deterministic management practices; the new managerial "science" is statistical process control [8], with philosophies such as Six Sigma marking an emphasis on managing risk, especially with respect to a growing trend in lowering variability [16].

There is a growing belief that organizations do not resemble mechanical systems so much as holistic organisms, constantly self-organizing and reorganizing to deal with change. According to Freedman:

> "In a sense, managers are in a position rather similar to that of pre-chaos natural scientists. They *think* they understand the relationships between cause and effect in their organizations. But in fact, the links between actions and results are infinitely more complicated than most managers suspect…. As a result, managers are prisoners of the very systems they are supposed to manage. They understand neither the underlying dynamics of

> these systems nor how to influence those dynamics to achieve organizational goals." [8]

Most information visualization systems do not deal with the notions of uncertainty in data and interlinked causes and effects very well. To be fair, a system can only be as good as the data one provides to it, and many systems are optimized for illustrating a few select relationships on a smaller scale. However, data analysts are often interested in complex relationships, especially ones that are not immediately apparent.

2 Identifying The Analytic Gaps

2.1 Representational Primacy

primacy (n.) – the state of being first or foremost

The status quo of information visualization is one concerned primarily with what is being visualized, letting designer intuition and user knowledge bridge the gap between the data and its use in higher-level knowledge tasks. As Tufte encourages, "above all else, show the data." [23] Studies such as Kobsa's [13] test more how well users can unpack the representation of individual data than how users actually discern any higher-level trends or implications of the data sets. This pursuit of faithful data replication and comprehension is what we call *representational primacy*.

What we argue here is that representational primacy can be a limiting notion, perhaps focusing on low-level tasks that do not map well to the true needs and goals of users. Of course, good collection and presentation of data are clear precursors to visualizations of any usefulness. Nor does representational primacy represent insensitivity to users or their needs; rather, it probably represents uncertainty as to how to best support those needs. Technologists have a long and fruitful history of sharing information and building tools useful to their communities of practice [3]. However, it is not clear that information visualization will be more than a "gee whiz" tool of occasional value to users in general if its use in more analytic thinking is not considered.

2.2 The Gaps Between Representation And Analysis

A desire to go beyond representationally primal systems has existed for decades, as early as Bertin's assertion in 1977 that "in decision-making the useful information is drawn from the overall relationships of the entire set" [2]. In 2002, Wesley Johnston even went so far as to say information visualization was the wrong primary tool where the formation of explanatory or correlative models was the desired outcome, and asserted a need for "model visualization" rather than "data visualization" [12].

One logical end to this line of thought is to build systems that are "black boxes," in which we input our data and out comes "the answer." However, it is widely viewed as irrational and unethical to trust an important decision to a "black box" system, as the rationale for such a decision are obscured and the responsibility for its consequences difficult to allocate. Therefore, we echo the recent arguments of Shneiderman for combining tools such as data mining with information visualization [18] to provide user control.

Shneiderman mainly argues for using data mining to identify time-series trends as well as possible correlations for users to explore. We wish to go one step further to what might be called a "white box" approach: systems that promote the generation of higher-level knowledge about a domain that results in justifiable actions. This is certainly a lofty goal which a single system or framework would find difficult to address; however, it is our hope to problematize some of the difficulties visualization systems encounter in such knowledge-making. We group these issues into two major categories; as these represent distances that must be bridged between current systems and more analytical systems, we call these *analytic gaps*.

2.2.1 The Rationale Gap: No "Black Boxes"

We define the *Rationale Gap* as the gap between perceiving a relationship and actually being able to explain confidence in that relationship and the usefulness of that relationship. Systems built under representational primacy assist in the perception of relationships, but very often fail to elucidate the strengths of these relationships and the confidence in these relationships. As a simple example, comparing averages in a visualization tool is misleading unless you know something about the populations from which the averages came, and thus your confidence in the actual difference in averages. As a tool incorporates a wider range of techniques, this problem compounds itself.

Indeed, typical implementations of business intelligence software have proven to be overly complex and require too much specialist intervention; the end result is not analytic clarity but an endless stream of reports [6]. Systems that bridge the Rationale Gap not only provide accurate, clear answers, but instill in users identifiable rationale about the kinds of decisions that can be made through their use.

2.2.2 The Worldview Gap: Show The Wider Perspective

We define the *Worldview Gap* as the gap between what is being shown and what actually needs to be shown to draw a straightforward representational conclusion for making a decision. Wesley Johnston's comments about "model visualization" fit directly into this.

Although extremely careful data collection and graphic design can indeed create situations where the data indeed speak for themselves, in practice, representation primacy often fails due to imperfect data collection and inexperienced presentation design. Tufte ranks some of the U.S.'s most revered journalistic information sources, such as *The New York Times* and *The Wall Street Journal*, as having low graphical sophistication, and provides a litany of examples of graphics that decorate numbers rather than actually elucidate relationships between variables [23]. While many information visualization systems are more sophisticated, providing graphical views of correlation and statistical summarization functions, they do not take full advantage of the powerful tools statistics has to offer. While correlation is a gateway to causation, the nature and usefulness of any visualized correlation is uncertain, as the true explanatory variable(s) may lie just outside the reach of the data; for example, do family income levels explain standardized test performance, or are the two merely found together?

Nor is it clear that one representation fits all; although scatter plots and graphs facilitate certain comparisons for certain kinds of data, effective representation design remains decided on a case-by-case, domain-by-domain basis. Contrast this with the well-traveled tension of the power of defaults. Kobsa found that Spotfire users tended to use the default scatterplot visualization in solving problems, even when using a bar chart or histogram representation would have been a better fit [13]. This indicates that representational affordances of a visualization (which, as we have argued, are usually limited) strongly influence what users do with it.

Systems that bridge the Worldview Gap not only indicate the useful relationships among data, but also indicate useful representations and the limits of those representations.

3 Example Analytic Gap Scenarios

In order to provide further grounding for these gaps and how existing systems can fall into them, we provide two example scenarios.

3.1 Example: Sports Analysis

Consider being the general manager of a sports team, with the responsibility of trading and acquiring personnel to produce the best results. (In fact, many people live this dream daily through fantasy sports competitions.) Analyzing a data set of sports statistics for some given year for leaders in statistical categories is fairly straightforwardly done using current visualization tools and dynamic query operations such as sorting. With a tool that provides aggregation functions, one can even relatively quickly compare the actual performance and payrolls of whole teams across the year, such as exists in individual leagues or in the sport as a whole.

All of this is useful in making some intuitive sense out of the data given; it can be especially useful in spotting anomalies such as extremely poor or good performers, or extremely high team payrolls. Still, there are two major problems.

First, any intuition we may develop about the data set is hard to transfer away from the tool; we may be able to see correlations for two or three variables at one time, but what we really desire is a plug-and-play "causal model," especially for predictive actions such as determining future performance of certain players. Unfortunately, information visualization systems provide little to no support for the formulation of predictive models, let alone a clear explanation as to how such a model might be constructed, running headlong into the Rationale Gap. Second, while most tools visualize correlations and simple relationships, they fail to provide indications as to which relationships or combination of relationships most strongly suggest the attainment of a certain performance metric, such as win percentage or offensive effectiveness, falling into the Worldview Gap and leaving users to use their own intuition as to what aspects of the data set are most useful. Confounds in correlation of variables are especially troubling when decisions involve a lot of money, such as those about sports personnel movement.

Possibly even more troubling is that we cannot really use a visualization tool to apply any real-world constraints, such as economic constraints; while we can dream about the sorts of teams we can put together and even get a superficial sense for how much such teams will cost, we cannot easily reason about how to achieve such an outcome in reality, such as managing money to be committed to players in the future and coping with effects on the existing organization and personnel. While such forward-looking prediction is arguably out of the domain of a representational visualization tool, we believe it is not implausible for at least some of the analytic processes involved to be translated into the perceptual domain, offering a viable and accessible complement to data mining tools and spreadsheets.

3.2 Example: Managerial Decision-Making

In his book *The Fifth Discipline*, Peter Senge describes a case study of a fictional company called WonderTech, which began by growing quickly but eventually collapsed under huge cycles of alternating high and low demand. The end result was due to a vicious circle of needing to improve sales but not having the capacity to keep up with sales when they did improve; as a result, the fixed investments in manufacturing increased but sales failed to stay consistently high enough to support an increasingly expensive infrastructure. [15]

Here is an instance when presumably the managers of WonderTech had a multitude of numbers available to them, and possibly even saw cyclic trends in sales and a growing fixed cost of manufacturing, but either failed to see the basic feedback process, failed to see a way out of the feedback process, or were too occupied with short-term solutions to get an accepted long-term solution in place, such as a commitment to rapid delivery [8]. Most visualization tools would support a time-series view of sales and financials, which would go far in elucidating that there was a problem. However, it would take a miracle in the data set to show growing order fulfillment times (if that was even a problem) and an as-of-yet nonexistent capability to show that reducing these fulfillment times could result in a better long-term ability to support sales, an example of the Worldview Gap.

4 Bridging The Analytic Gaps: Knowledge Tasks

Evaluating systems on how they meet the Rationale and Worldview Gaps is to some degree an operational approach for evaluating and designing analytic systems. However, we feel there is much to be gained from concrete identification of common tasks that fall in the gaps. Therefore, we propose a taxonomy of common subtasks that can provide better support for designers and evaluators of information visualization systems.

4.1 The Use Of Taxonomies

A recent branch of information visualization research concerns itself with the development of taxonomies for organizing low-level tasks that a visualization should facilitate, and automatically creating presentations that match these tasks to appropriate techniques. Wehrend and Lewis create a matrix of techniques that correspond to a particular combination of an object type, such as scalar or vector, and a cognitive operation, such as correlation or association [25]. In [27], Zhou and Feiner describe a taxonomy that refines the Wehrend and Lewis operations into visual tasks which are organized by their visual accomplishments (low-level user or presenter goals, such as "inform" and "enable") and visual implications (what visual capabilities are called upon in the attainment of the visual accomplishments).

While these low-level tasks are essential, they do not in and of themselves provide a basis for consistently bridging the analytic gaps. Even automatic generation of visualizations or visual discourse, such as that provided by IMPROVISE [27] and BOZ [4], rely on designer-provided rule sets and/or complex logical and perceptual operator definitions to create coherent presentations. Our overall goal is to describe complementary, higher-level knowledge tasks that appear in the real world that people must do to bridge the analytic gaps described above.

In particular, we are concerned with describing knowledge tasks regarding the application of visualization to two closely-related goals:

- **Complex decision-making, especially under uncertainty.** Recall the sports team management example from earlier. Consider the seemingly simple task of deciding whether to trade players with another team. It is far from straightforward to understand the expected collective performance of arbitrary subsets of players, the costs and benefits to teams of making personnel changes, and the prediction of future performance, both in terms of average performance and variability. Essentially, this is the Rationale Gap.

- **Learning a domain.** Exploration of particular data sets can reveal a lot about the general discipline or phenomena which the data sets describe. It can also ideally suggest elements outside the data set that further elucidate the domain. Essentially, this is the Worldview Gap.

We now turn to discussing our higher-level knowledge tasks that visualization systems should support for complex decision-making and learning. We classify these tasks according to which analytic gap primarily motivates it, although overlap is possible.

4.2 Rationale-Based Tasks

Users need to be able to relate data sets to the realms in which decisions are being made. For example, analysis of a computational chemistry data set may produce an encoding for a promising lead compound for the design of a drug [5]. Proper visualization of the data set communicates how to modify existing compounds to obtain the promising lead. Also, given a set of criteria, users need to be able to use salient features of data sets to create a description of the realm in general, to validate decisions.

4.2.1 Rationale Task 1: Expose Uncertainty

Some uncertainty is involved in any data set. Is the data set large enough to mitigate any associated sampling error? Are there figures in a data set involving uncertainties, such as population estimates with associated standard errors or statistically distributed phenomena? An understanding of where values are uncertain and how that uncertainty affects the degree to which a data set can be a source for reliable conclusions is key in statistical process control.

For example, when considering several vendors for a part whose width must be exactly within a specified range, it is important to understand not just the width of the average part produced, but the standard deviation as well (to understand the proportion of unusable parts). Also, when comparing poll results or estimated financial figures, having a measure of the standard error of the estimates is crucial to having confidence in the statistical significance of any differences observed, especially when sample sizes are small.

We consider this a Rationale Gap task as it relates directly to the confidence one can draw based on correlation or aggregation analysis done within a visualization tool. To summarize, *a system can help bridge the Rationale Gap by exposing uncertainty in data measures and aggregations, and showing the possible effect of this uncertainty on outcomes*.

4.2.2 Rationale Task 2: Concretize Relationships

In the case of correlation, especially when viewed on a scatterplot, perceived relationships are usually easy to describe and quantify. Other representations may suggest relationships or decisions without a clear concretization of the nature of the relationships. This can be particularly problematic in expected value analysis. When the expected payoff or outcome of a decision is a weighted average of the elements of a clearly identifiable discrete distribution (called a *risk profile* in business), then the outcomes are not so clear and are often surprising when people think in terms of expected values.

This is a Rationale Gap task in the spirit of being able to rationalize decisions and outcomes based on a cognitive alignment of a perceived relationship with its representational elements. To summarize, *a system can help bridge the Rationale Gap by clearly presenting what comprises the representation of a relationship, and present concrete outcomes where appropriate*.

4.2.3 Rationale Task 3: Formulate Cause And Effect

When investigating data, there is usually some causation data embedded directly in the data set, as well as effect data that can become clear through iterations of a simulation. Both the isolation of demonstrated causes as well as the discovery of possible effects are important in cognitive model formation. All of this must be done with an understanding of what assumptions have gone into creating the data, and thus affect the outcomes inferred. As an example, consider the story of WonderTech we recounted earlier. Some causation can be inferred from time series data of sales and manufacturing costs; a further step would be to be able to investigate the effects of changing certain variables on the outcomes depicted by the data set, such as sensitivity analyses (e.g. the value of an investment opportunity as depends on factors such as market interest rates or growth predictions).

This addresses the Rationale Gap primarily because it serves to distinguish between causation and covariance. To summarize, *a system can help bridge the Rationale Gap by clarifying possible sources of causation*.

4.3 Worldview-Based Tasks

Many tasks we will describe here indirectly support formulation of a strategy for browsing a visualization when they provide insights as to what data should be explored to clarify certain relationships or test certain hypotheses.

4.3.1 Worldview Task 1: Determination Of Domain Parameters

The attributes of data in a visualization, and thus the parameters by which data is organized in a visualization, communicate both standards of measure within a data set and key parameters for understanding a domain. The very fact that a collection of American baseball scores includes data such as home runs, runs batted in, and slugging percentage indicates that these are parameters considered important (at least by the data collector), and suggests domain-specific measures that require clarification. As well, the relative positive or negative connotations of parameters are not always clear; for example, in American baseball, a batter with a high number of career strikeouts may not be considered a good batter, nor a pitcher with a high number of walks and hits allowed, but these interpretations are not always inherent in the visualization.

We consider this a Worldview Gap task because it points the way to formation of more expressive representations. To summarize, *a system can help bridge the Worldview Gap by providing facilities for creating, acquiring and transferring knowledge or metadata about important domain parameters within a data set.*

4.3.2 Worldview Task 2: Multivariate Explanation

Most visualization systems support determination of correlation between two or three variables, in the limit of representational ability. However, some relationships involve more than three explanatory variables and/or simple transformation of single explanatory variables using logarithms or polynomial relationships [1]. Such correlations, often found in domains such as queuing theory, are not widely handled by typical visualization tools. Also, when correlations expected by theory do not exist, correct interpretation and action usually involves user guidance. In general, while statistics offers methods such as stepwise regression to help automatically determine good explanatory models, mindlessly employing such tools generally yields bad results [1]. Combining these methods with user guidance could result in a very useful facility for data analysts.

In 1990, La Quinta Motor Inns retained the services of academic statisticians who derived a successful mathematical model for the selection of sites for La Quinta inns [1]. The model directly related site profitability to the room rate and inversely related profitability to the population of the state of the site, which both seem reasonable. However, the analysts also found a strong direct relationship between profitability and the number of college students within four miles (possibly surprising) and an inverse relationship between profitability and the *square root* of the median income of the area. The model explained 51% of the variation in profitability, which is respectable in practice; however, this possibility does need to be raised to a user of the model, who may experience deviations from the results.

This task is in the spirit of the Worldview Gap, as it can help elucidate useful representational transformations. To summarize, *a system can help bridge the Worldview Gap by providing support for discovery (whether automated or manual) of useful correlative models and constraints*.

4.3.3 Worldview Task 3: Confirm Hypotheses

Users need to test the accuracy of their deductions about a data set. Tools must help users define hypotheses, simulate possible outcomes, and verify the truth of such hypotheses. While we might include statistical hypothesis tests such as confirmation of expectation (e.g. statistical distribution of results, expected limits of data values) and comparison of averages with certain confidence intervals, this task includes higher-level hypotheses. If a particular region or outcome of interest is found, then hypothesis tests can also become a question of how far and how easily users can operate on that outcome. This analytic process is clearly difficult to support in a general manner across interfaces and representations, but may be useful for specific design decisions.

We consider confirmation of hypotheses a Worldview Gap task because it points to the expressiveness and completeness of cognitive or mathematical models derived from use of a visualization. To summarize, *a system can help bridge the Worldview Gap by providing support for the formulation and verification of user hypotheses*.

5 Employing The Knowledge Tasks

Now that we have described the analytic gaps and some common knowledge tasks, we would like to propose a design and evaluation framework. In essence, all one need do is apply the knowledge tasks (plus any other higher-level knowledge tasks one wishes to employ) to a given situation.

5.1 Using The Tasks For Design

When designing a visualization for a new domain or scenario, one can use the knowledge tasks to systematically:

- Generate new subtasks for a visualization to support or perform.
- Identify possible shortcomings in representation or data.
- Discover possible relationships to highlight or use as the basis for a visualization.

The general idea is to apply each knowledge task in turn as a user would to each scenario. For example, "Where might I be interested in multivariate relationships?" or "Exactly what is uncertain about this data and how will it affect the outcomes I show?" or even "How will I show the concrete outcomes from this process?"

5.2 Using The Tasks For Evaluation

One can also use these tasks as a form of heuristic evaluation [14] of the pragmatic value of a given visualization simply by evaluating how well the visualization supports the knowledge tasks. The Rationale Gap tasks provide particularly rich opportunities to ask questions both about how actual relationships and outcomes are shown to a user (e.g. must the user infer an outcome from the context of a representation, or can a user perform a direct action to see an outcome, such as in a brushing histogram), as well as how confident the user should be in these outcomes relative to any uncertainty inherent in the data set being visualized.

6 Design Example: The Infovis 2004 Contest

While the knowledge tasks and scenarios have their roots in quantitative domains, such as financial and scientific domains, the six knowledge tasks here provide a very fruitful way of thinking about visualizations for a decidedly less quantitative scenario: the InfoVis 2004 Contest [10]. The contest, which is to provide visualizations to support questions about the evolution of information visualization as a research area, is based on a dataset containing metadata (titles, abstracts, keywords, dates, and references) about articles from the InfoVis conference from 1995 to 2002. Although it is hoped that applying the knowledge tasks sheds new light on possible solutions to contest tasks, we wish to show more that the knowledge tasks provide a systematic basis for thinking about and identifying issues in the data set.

6.1 Rationale Task 1: Expose Uncertainty

For this dataset dominated primarily by nominal data, at first glance it seems there is no uncertainty to speak of. However, uncertainty can appear in more forms than standard deviations and measurement errors. If one examines the metadata for completeness, one notices a number of possible sources of uncertainty. For example, author names are sometimes spelled or formatted differently. Paper dates are sometimes exact, and sometimes involve a large range of dates. References may be missing or their formats may differ, requiring significant effort for tagging or cleaning.

In other words, being sure of who is who, when is when, and sometimes even what is what is difficult. If any uncertainties cannot be resolved in the process of data cleaning, they must be shown to the user. For example, if it is unclear whether or not "J. Smith" and "J. T. Smith" are the same person, this is an uncertainty, especially given the higher-level tasks contest entrants are asked to support.

6.2 Rationale Task 2: Concretize Relationships

If we are asked to relate two researchers' work in the field of information visualization, how will we do it? Ideally, a visualization should provide perceptual triggers [21] to show these outcomes. One possible approach is to use a concept map such as a themescape [26] and show two or more researchers' work as regions on that themescape, highlighting areas of overlap with brighter colors to indicate the degree of overlap of the researchers involved. But does that overlap imply or represent frequent co-authorship, common mutual referencing, unity in research subject matter, or something else entirely? If there is significant overlap in fringe areas, then does that represent the formation of new research areas, or just a coincidence? All of these items could be indicated to the user.

6.3 Rationale Task 3: Formulate Cause And Effect

Here, we can think about possible causes and effects in the field to generate interesting ideas for relationships to highlight. Did one paper spawn off a generation of related papers? Can we identify opposing schools of thought on a topic and their point evolutions in time? Do user studies (tagged externally by other participants) promote new and interesting ideas in the field? Most importantly, what data must we employ to validate this cause and effect? How can a user feel he/she is exploring the data set and knows where the relationships come from, rather than interacting with a "black box"?

6.4 Worldview Task 1: Determine Domain Parameters

Clearly, the attributes of the metadata dominate our thinking about the dataset. We have already discussed the notion of considering other factors that may come to bear on the dataset that might not currently be reflected. Another possibility is to consider how deeply the metadata allow us to make conclusions. Are abstracts enough to relate articles, or do we need more text to do the appropriate comparisons? Are references enough, or do we need more metadata on what kinds of papers (conference full papers, extended abstracts, technical notes, journal papers, etc.) are citing other papers and being cited?

6.5 Worldview Task 2: Multivariate Explanation

Returning to the themescape example, the outcomes highlighted for the user are a two-dimensional projection of a potentially multivariate trend. The important questions for design of a relevant visualization include generating possible multivariate explanations as well as how to communicate the variables' contribution to the overall analysis. For example, one may determine that the trajectory of a researcher on a themescape is determined by a particular correlation with the subject matter of other researchers, dates of publication, and keywords (possibly both author-provided and contest entrant-generated).

6.6 Worldview Task 3: Confirm Hypotheses

Even though the contest tasks are mainly qualitative, users may wish to experiment with different classifications or evolution along different dimensions: for example, using research money allocated to areas or number of people working in an area to show evolution rather than a size-agnostic time-based evolution. Considering the themescape example once more, if overlaps are identified in fringe areas, a user may wish to see if that fringe area eventually panned out into anything larger. One may even wish to ask higher-level questions, such as whether or not the development of a particular research area was hindered by or depended on the development of a different area. Ultimately, for the purposes of the contest, this form of experimentation may be limited, but considering the types and degree of utility of such experimentation may help decide the feature set available to a user.

7 Evaluation Example: Commercial Tools

We can also use the knowledge tasks to reflect upon how commercial tools might or might not be meeting the challenges posed by the analytic gap. Here, we consider the same tools considered by Kobsa in his evaluation [13]: Spotfire, Eureka, and InfoZoom.

7.1 Rationale Task 1: Expose Uncertainty

Again, most statistical facilities in these information visualization systems are limited to aggregation and correlation. Spotfire can bin data according to standard deviation and can indirectly show some variations around points, but the explicit treatment of uncertainty is otherwise limited. Eureka and InfoZoom generally display the data as given. None of the programs allow easy comparison of averages within a certain confidence, although InfoZoom's "derived attributes" functionality is programmatically expressive for those who can write programs. Granted, the data provided do not always show uncertainty well; still, uncertainty is not generally part of the data import facilities of these programs, and even if explicit measures of uncertainty were integrated into the data, the data importing facilities would require them to be treated as members of the data set rather than metadata.

7.2 Rationale Task 2: Concretize Relationships

All of these commercial systems can show details-on-demand for a particular item or set of items. As well, when filtering relationships are applied, single items or sets of items can be easily shown and isolated for individual examination. However, close but inexact matches, as well as relationships based on probabilistic links, are harder, if not impossible, to show and isolate. An approach such as the Attribute Explorer [20] can help increase the flexibility of such queries.

7.3 Rationale Task 3: Formulate Cause And Effect

Spotfire provides the "View Tips" functionality, which highlights interesting correlations for users to examine. Otherwise, users of these systems are left on their own to explore possible correlations. As well, no facilities for sensitivity analysis are provided.

7.4 Worldview Task 1: Determine Domain Parameters

Since these systems are largely data-driven, the tools communicate the domain parameters that are in the data set. Most of the issues here revolve around presentation; for example, Spotfire relegates some data to a smaller window for details-on-demand, and Eureka occasionally has problems displaying large labels. The ability to attach annotations or other metadata to certain domain parameters and present such metadata to the viewer would be advantageous.

7.5 Worldview Task 2: Multivariate Explanation

Spotfire offers explicit three-dimensional correlation; while Eureka and InfoZoom do not offer explicit correlation, they do use filtering, brushing, and sorting on many different attributes at once. For flexibility and ease of analysis, these systems could provide more tools for correlation, such as non-linear correlation and correlation to logarithms or polynomial functions of data.

7.6 Worldview Task 3: Confirm Hypotheses

In these systems, when items of interest are isolated, their context in the data set as a whole is visible. As mentioned before, InfoZoom does provide powerful derived attributes; in fact, all the tools provide some way of creating at least simple derived attributes, usually based on aggregation functions. However, the tools are not as well suited to time series analysis, which is a common basis for higher-level data analysis and hypothesis tests.

8 Conclusion

In this article, we have identified the focus of current information visualization systems on representational primacy, or the overriding pursuit of faithful data replication and comprehension. We have argued that to become even more useful, a parallel focus on analytic primacy must emerge. Limitations in current systems were classified into one of two analytic gaps: the Rationale Gap, representing the gap between perceiving a relationship and expressing confidence in the correctness and utility of that relationship; and the Worldview Gap, representing the gap between what is shown to a user and what actually needs to be shown to draw a representational conclusion for making a decision. For each gap, we proposed three task forms that serve to narrow or diminish these gaps, and then demonstrated how these tasks might be used for systematic design and heuristic evaluation.

While we have primarily concentrated on information visualization, similar challenges and problems exist in other visualization realms such as scientific visualization. In providing a knowledge task framework and a set of subtasks that are useful to consider, our intention is to check the status quo of visualization tools with the decision-making processes of the real world. In short, we are asking what more these systems could do to be more useful for decision makers. If, as Tufte asserts, we lack graphical sophistication as a population, then perhaps we need all the help we can get to make sense of the rapidly burgeoning mounds of information that we must deal with on a daily basis in our work and personal lives.

References

[1] Albright, S.C., Winston, W.L., and Zappe, C. (2003) *Data Analysis and Decision Making with Microsoft Excel, Second Edition*. Thomson Learning, Pacific Grove, CA, 2003.

[2] Bertin, J. (1981) *Graphics and Graphic Information Processing*, Berlin, Walter de Gruyter, 1981, being a translation of Bertin, J. *La Graphique et le Traitement Graphique de l'Information*, Paris, Flammarion, 1977.

[3] Brown, J.S. and Duguid, P. (1991) Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. In *Organizational Learning*, Cohen, M.D. and Sproull, L.S. (eds.), pp. 58-81, SAGE Publications, 1991.

[4] Casner, S.M. (1990) A task-analytic approach to the automated design of graphic presentations. *ACM Transactions on Graphics,* 10(2), pp. 111-151.

[5] Dietterich, T.G., Lathrop, R.H., and Lozano-Perez, T. (1997) Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence,* 89, pp. 31-71.

[6] Eick, S.G. (2001) Visual discovery and analysis. *IEEE Transactions on Visualization and Computer Graphics,* 6(1), pp. 44-58.

[7] Inxight Table Lens, http://www.inxight.com/products/oem /table_lens/ (April 2004).

[8] Freedman, D.H. (1992) Is management still a science? *Harvard Business Review* (Nov.-Dec. 1992), pp. 26-38.

[9] Glaser, D.C., Tan, R., Canny, J. and Do, E.Y. (2003) Developing architectural lighting representations. In *Proceedings of InfoVis 2003*, IEEE, pp. 241-248.

[10] InfoVis 2004 Contest, http://www.cs.umd.edu/hcil/iv04contest/, April 2004.

[11] InfoZoom, http://www.humanit.de/en/ products_solutions/products/iz/, April 2004.

[12] Johnston, W. (2001) Model Visualization. In *Information Visualization in Data Mining and Knowledge Discovery*, Fayyad, U., Grinstein, G., and Wierse, A. (eds.), pp. 223-228, Morgan Kauffman, 2001.

[13] Kobsa, A. (2001) An empirical comparison of three commercial information visualization systems. In *Proceedings of InfoVis 2001,* IEEE, pp. 123-130.

[14] Nielsen, J., and Molich, R. (1990) Heuristic evaluation of user interfaces. In *Proc. CHI 1990*, ACM, pp. 249-256.

[15] Senge, P.M. (1994) *The Fifth Discipline.* Currency, 1994.

[16] Sharit, J. (1997) Allocation of Functions. In *Handbook of Human Factors and Ergonomics, Second Edition*, Salvendy, G. (ed.), Wiley Interscience Publications, New York, NY, 1997, pp. 301-339.

[17] Shneiderman, B. (1996) The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. 1996 IEEE Conference on Visual Languages*, pp. 336-343.

[18] Shneiderman, B. (2002) Inventing discovery tools: Combining information visualization with data mining. *Information Visualization,* 1(1), pp. 5-12.

[19] Spell, R., Brady, R., and Dietrich, F. (2003) BARD: A visualization tool for biological sequence analysis. In *Proceedings of InfoVis 2003*, IEEE, pp. 219-225.

[20] Spence, R. and Tweedie, L. (1998) The Attribute Explorer: information synthesis via exploration. *Interacting with Computers*, 11, pp. 137-146.

[21] Spence, R. (2001) *Information Visualization*. ACM Press.

[22] Spotfire, http://www.spotfire.com (April 2004).

[23] Tufte, E.R. (2001) *The Visual Display of Quantitative Information, Second Edition*. Graphics Press, Cheshire, CT.

[24] van Ham, F. (2003) Using multilevel call matrices in large software projects. In *Proc. InfoVis 2003,* pp. 227-233.

[25] Wehrend, S. and Lewis, C. (1990) A problem-oriented classification of visualization techniques. In *Proceedings of InfoVis 1990,* IEEE, pp. 139-143.

[26] Wise, J.A., Thomas, J.J, Pennock, K., Lantrip, D., Pottier, M., Schur, A. and Crow, V. (1995) Visualizing the non-visual: spatial analysis and interaction with information from text documents. In *Proc. InfoVis 1995*, IEEE, pp. 51-58.

[27] Zhou, M.X. and Feiner, S.K. (1998) Visual task characterization for automated visual discourse synthesis. In *Proceedings of CHI 1998*, ACM, pp. 392-399.