Quality Metrics in High-Dimensional Data Visualization: An Overview and Systematization

Enrico Bertini, *Member, IEEE*, Andrada Tatu, and Daniel Keim, *Member, IEEE*

Abstract—In this paper, we present a systematization of techniques that use quality metrics to help in the visual exploration of meaningful patterns in high-dimensional data. In a number of recent papers, different quality metrics are proposed to automate the demanding search through large spaces of alternative visualizations (e.g., alternative projections or ordering), allowing the user to concentrate on the most promising visualizations suggested by the quality metrics. Over the last decade, this approach has witnessed a remarkable development but few reflections exist on how these methods are related to each other and how the approach can be developed further. For this purpose, we provide an overview of approaches that use quality metrics in high-dimensional data visualization and propose a systematization based on a thorough literature review. We carefully analyze the papers and derive a set of factors for discriminating the quality metrics, visualization techniques, and the process itself. The process is described through a reworked version of the well-known information visualization pipeline. We demonstrate the usefulness of our model by applying it to several existing approaches that use quality metrics, and we provide reflections on implications of our model for future research.

Index Terms—Quality Metrics, High-Dimensional Data Visualization.

1 Introduction

The extraction of relevant and meaningful information out of high-dimensional data is notoriously complex and cumbersome. The *curse of dimensionality* is a popular way of stigmatizing the whole set of troubles encountered in high-dimensional data analysis; finding relevant projections, selecting meaningful dimensions, and getting rid of noise, being only a few of them. Multi-dimensional data visualization also carries its own set of challenges like, above all, the limited capability of any technique to scale to more than an handful of data dimensions.

Researchers have been trying to solve these problems through a number of automatic data analysis and visualization approaches that cover the whole spectrum of possibilities: from fully automatic to fully interactive. Visualization researchers have discovered early on that searching for interesting patterns in this kind of data can be done through a mixed approach, where the machine based on quality metrics automatically searches through a large number of potentially interesting projections, and the user interactively steers the process and explores the output through visualization.

The pioneering work of Friedman and Tukey in 1974 with their *projection pursuits* method [21] introduced the idea. They recognized the limit of human beings in exploring the exponential set of projections and tackled the high-dimensionality issue by letting an algorithm discover interesting linear projections in 1D (histograms) and 2D (scatter plots) and letting the user evaluate the corresponding output.

During the last few years the use of this paradigm has witnessed a growing interest, and an increasing number of techniques has been published in key data visualization conferences and journals. Quality metrics have been used for very disparate goals such as: searching for interesting projections, reducing clutter, and finding meaningful abstractions. However, the initial idea of quality metrics has been elaborated and expanded so much further and into so many different directions that it is hard to come up with a coherent and unified picture for them. A reader of one of these papers may well appreciate the value of a single technique without having a way to place it into a larger context. Also, researchers who might want to approach this area of investigation for the first time and develop new techniques may have a hard time appreciating the whole spectrum of possibilities and directions related to the use of quality metrics.

In this paper we move first steps towards filling this gap. We provide a systematization of using quality metrics in high-dimensional data analysis through a literature review. We analyzed numerous papers containing quality metrics and went through an iterative process that led to the definition of a number of *factors* and a *quality metrics pipeline*, which is inspired to the traditional information visualization pipeline [12].

The extracted factors and the pipeline have the following interrelated goals: (1) putting the existing methods into a common framework, (2) easing the generation of new research in the field, (3) spotting relevant gaps to bridge with future research.

In the paper, we provide an extensive explanation of the methodology we followed, the results we obtained, and their practical use. In particular, we demonstrate by going through a number of selected examples how we are able to describe existing approaches through the proposed models. Also, we spot a number of interesting gaps and give guidelines on how to carry out new research in this area. To the best of our knowledge, despite the numerous techniques that can be categorized under the umbrella of quality-metrics-driven visualization, this is the first attempt in this direction.

1.1 Definitions

In order to make the goal and scope of our work clear, we provide some initial definitions.

Information Visualization Pipeline: a reference model that describes how to transforms data into visualizations through a series of processing steps, as defined in [12].

Quality Metric: a metric calculated at any stage of the information visualization pipeline that captures properties useful to the extraction of meaningful information about the data.

High-Dimensional Data: any data set with a dimensionality that is too high to easily extract meaningful relations across the whole set of dimensions. In the context of this paper, any dimensionality higher than 10 is considered high-dimensional.

Our focus is on the analysis of methods that apply *quality metrics* at any stage of the *information visualization pipeline* as a way to facilitate the detection and presentation of interesting patterns in *high-dimensional data*.

- *Enrico Bertini is with University of Konstanz, Germany, E-mail: enrico.bertini@uni-konstanz.de.*
- *Andrada Tatu is with University of Konstanz, Germany E-mail: tatu@inf.uni-konstanz.de.*
- *Daniel Keim is with University of Konstanz, Germany, E-mail: keim@inf.uni-konstanz.de.*

Manuscript received 31 March 2011; accepted 1 August 2011; posted online 23 October 2011; mailed on 14 October 2011.
For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

1.2 Examples

We first discuss a few short examples of the approaches covered in our review to familiarize the reader with the concepts exposed in the paper and get the feeling of their heterogeneity. They cover a broad selection of the factors, denoted with italics, which will be presented in detail in Section 5.1.

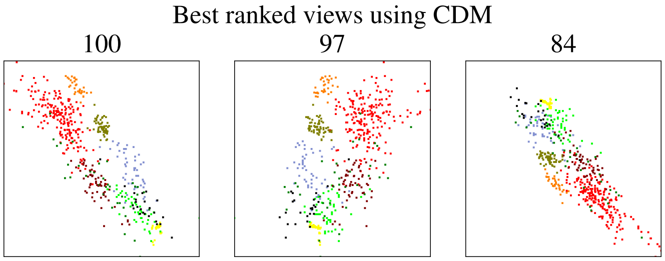


Fig. 1. Ranking projections according to their *class density measure*, favoring projections with minimal overlap between predefined classes (i.e., the colors) [48].

Example 1. Tatu et al. in [48] analyze high-dimensional data sets by computing an interestingness score for every scatter plot generated with all the possible combinations of axis pairs from the original data. The score is calculated by running *image* processing algorithms on top of each scatter plot in order to detect images with *clusters* in the visualization. The system returns a list of scatter plots as those presented in Figure 1 sorted in order of relevance according to the chosen quality measure.

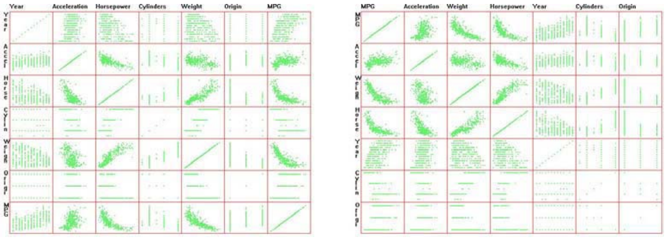


Fig. 2. Clutter reduction achieved through axes reordering in a scatter plot matrix (initial visualization on the left, reordered on the right) [39].

Example 2. Peng et al. in [39] provide algorithms to reorder the axes of multi-dimensional data visualizations (parallel coordinates, scatter plot matrices, glyphs, recursive patterns) in order to *reduce clutter* and make interesting patterns more clearly visible. For each visualization a specific quality metric calculated in the *data space* is used to find the best ordering. In Figure 2, we present an example on scatter plot matrix reordering.

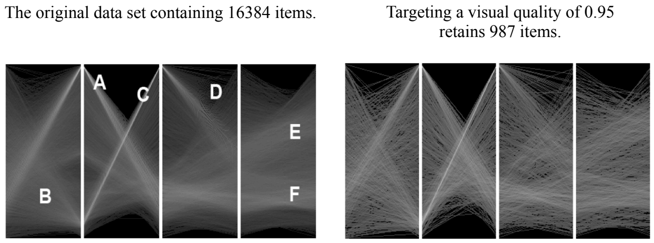


Fig. 3. Data abstraction algorithm based on sampling, aiming at reducing data size while preserving relevant patterns. Original visualization on the left with 16384 data items. Sampled visualization on the right with 987 items and a visual quality of 0.95 [28].

Example 3. Johansson et al. in [28] study the abstraction obtained by applying sampling or aggregation algorithms on top of parallel coordinates and provide quality metrics to judge when the *abstraction* disrupts relevant patterns in the data. In Figure 3 we show an example from their work, where on the left the data set containing 16384 items is displayed with parallel coordinates. On the right side they display an image targeting a visual quality of 0.95 (on a scale from [0,1]) by displaying only 987 items. The image quality is calculated by a *screen* metric using distance transforms.

All the approaches have in common that they use quality metrics in the context of high-dimensional data visualization; nonetheless they can differ on a variety of aspects. For instance, in Example 1 the purpose is to find interesting projections, in Example 2 the purpose is to reduce clutter, whereas the purpose in Example 3 is to find the right abstraction level. The approaches can as well differ in a number of other aspects such as: the visualization techniques employed, the space in which the quality metrics are calculated, or the level of interaction they provide.

Therefore the questions are: *How we can put all the approaches into a common framework which is able to highlight commonalities and differences? What are the main factors through which we can describe them? How can we learn from the approaches and build on top of them to systematically move the idea of quality-metrics driven visualization forward?*

These are the main questions that motivate our work and in the following sections we will provide the results of our investigation.

2 Background

Quality metrics in visualization have a long history. While in our work we focus only on their specific use in high-dimensional data analysis, they have a broader scope than we can describe here. Early attempts to calculate quality metrics can be traced back to the work of Tufte [51], where he proposed metrics such as the *data to ink ratio* and the *lie factor*, which respectively optimize the use of the visualization space and reduce the distortions that visualization may introduce. Later in 1997 Richard Brath proposed a rich set of metrics to characterize the quality of business visualizations [11] and, around the same period Miller et al. advocated the use of visualization metrics as a way to compare visualizations [37]. The graph drawing community developed its own set of metrics, most notable aesthetic metrics such as those found in the foundational work of Ware et al. on *cognitive measurements of graph aesthetics* [52]. Later, the word quality metrics assumed a more specific meaning; in particular it appeared in the context of a number of papers related to clutter reduction and scalability [9, 10, 28, 30, 39].

While all these works are related to our goal, early in our project we decided to focus on the use of quality metrics in high-dimensional data exploration only. Our initial data gathering process included a broader class of papers, including those cited above. However we soon realized there is no all encompassing model able to synthesize the relevant aspects and, at the same time, is useful in practice. For this reason the paper focuses only on the use of quality metrics in high-dimensional data.

There exist a number of research papers which try to categorize existing work in the visualization area. Here we briefly mention some recent ones to put our work in a larger context. In Rethinking Visualization [50] Tory and Möller provide a taxonomy to describe scientific and information visualization under the same structure. Ellis and Dix organize a large number of existing clutter reduction techniques into a clutter reduction taxonomy [18]. Yi et al. review a large number of visualization systems to better understand the role of interaction in visualization [60]. Segel and Heer analyze a large body of story telling visualizations to identify common design patterns [43]. All these papers share with ours the need of putting some order into a complex aspect of data visualization by starting from a detailed analysis of what researchers and practitioners have proposed in the past.

Since our proposed systematization uses a data visualization pipeline as the basis for the analysis of quality metrics, we deem important to briefly discuss existing data processing pipelines. The information visualization pipeline has been presented by Card et al. [12] and is widely accepted as the standard processing model for information visualization. The Data State Reference model [13] is largely based on the information visualization pipeline and classifies visualizations according to how they use the operators in the pipeline. In this regard it is similar to our work in that we also use elements of the pipeline to classify the papers we have analyzed. The KDD pipeline [19] has been developed in the early nineties to describe the data processing stages involved in knowledge discovery. While we took inspiration from this model, as quality metrics involve automatic computation and visualization, we decided not to use it as a basis for our work because visualization does not explicitly appear in the intermediary steps of the process. Keim et al. [31] and Bertini et al. [8] present alternative pipelines that show how automated data analysis algorithms can be included in the data visualization process. These papers are also sources of inspiration for our work as they focus on the integration of automated algorithms and data visualization.

3 Methodology

We followed an iterative data gathering, coding, and modeling approach inspired to the methods used in grounded theory analysis [47]. We started from a small set of papers about quality metrics we knew from our own experience and used this initial list to derive a first set of descriptive factors. After that, we expanded the list by analyzing the references contained in the first set of papers and by searching in relevant visualization venues. In particular, we used Google Scholar [1] to search for references to and from the collected papers. We also expanded our list by targeted keyword search.

At this stage we decided to narrow down the scope of our study and focus on quality metrics for high-dimensional data analysis. We discarded the papers that (1) did not explicitly address high-dimensional data, (2) did not propose quality metrics systems or algorithms. For instance we discarded a number of interesting papers on the use of quality metrics for generic data visualizations [27], for graph drawing [16], or the discussions on generic aspects of quality metrics [10].

Two of the authors went independently through the current list of papers and completed a table with the current version of the classification and took notes on necessary modifications/additions to accommodate new aspects discovered during the analysis. After this first phase the two lists and the notes where confronted in order to reach a consensus on table factors and paper coding. The third author played the devil's advocate role at this stage to confirm the factors were *explicative*, *understandable* and *relevant*. A third set of additional papers were gathered and coded at this point to test the classification further.

We proceeded then to the definition of a visualization pipeline able to capture the data visualization processes described in the papers. We started from the traditional information visualization pipeline [12] because it is widely known and helps capturing key elements of quality-metrics-driven visualizations (details in Section 4).

We generated the quality metrics pipeline iteratively using the set of gathered papers and the descriptive table with quality metrics factors as reference. In particular, (1) we built a first draft of the new pipeline; (2) we went through the whole list of papers and checked whether the pipeline was able to describe every aspect involved in the process; (3) where discrepancies were found, we refined the pipeline accordingly. As a final step, we double-checked that every paper in the list could be described by a specific instance of the pipeline. Similarly to the procedure followed in the first phase we let one of the authors, not involved in the model generation phase, again play devil's advocate and refine the model at intermediary steps. The work on the pipeline generated also small adjustments that led to the final version of the quality metrics table (Table 2).

It is important to note that while we followed a systematic approach there is no guarantee that this is the only way to describe quality metrics and their use. Many of the elements introduced in the proposed models are the result of our own experience and are thus necessarily subjective. Nonetheless, the usefulness of the proposed model is demonstrated by its ability to describe the whole set of papers and to identify relevant gaps interesting for future research.

4 Quality Metrics Pipeline

We briefly recall the main elements of the Card et al.ś pipeline [12] and then we move forward to the description of our extensions.

The original purpose of the infovis pipeline was to model the main steps required to transform data into interactive visualizations. The quality metrics pipeline in Figure 4 preserves its main elements: processing steps (horizontal arrows), stages (boxes), and user feedback (with few naming differences we will explain soon). *Data transformation* transforms data into the desired format. *Visual mapping* maps data structures into visual structures (visualization axes, marks, graphical properties). *View transformation* creates rendered views out of the visual structures. The whole set of transformations is influenced by the user who can decide at any time to transform the data (e.g., filter), use different visual structures and, navigate the visualization through different view points.

The infovis pipeline captures extremely well the key elements of interactive visualization across a variety of domains and visual techniques. However, when we focus on the visualization of high-dimensional data patterns a practical problem arises. While the whole set of processes is still valid, the number of possible combinations at each step is so high that it is impractical to find interactively the most effective ones. An example in the spirit of Mackinlay's seminal analysis [36] helps to clarify the problem: if the original data has dimensionality $n = 10$ (still a quite low number) and the number of available visual parameters is $k = 4$ (e.g., a scatter plot with the following visual primitives: x-axis, y-axis, size, and color), the number of alternative mappings at the *visual mapping* stage is already more than 5000 (k-permutations, i.e., the number of sequences without repetition: $\frac{n!}{(n-k)!}$).

The main function of quality metrics algorithms is to aid the user in the selection of promising combinations. Typically, the algorithms search through large sets of possibilities and suggest one or more solutions to be evaluated by the user. To describe these steps we created an additional layer in Figure 4 that we call *quality-metrics-driven automation*, which depicts how quality metrics fit into the process. The metrics draw information from the stages of the pipeline (green upwards arrows) and influence the processing steps (blue downwards arrows) with their computation. The user remains in control of the whole process letting the machine perform the computationally hard tasks. We named the new pipeline the *quality metrics pipeline*.

The concept of generation of *alternatives* and their evaluation is at the core of the method. Regardless the purpose, all the systems we have encountered follow a common general pattern:

1. Create alternatives (projections, mappings, etc.)
2. Evaluate alternatives (rank views, orderings, etc)
3. Produce a final representation (ranked list of views, small multiples, etc.)

As we will show in Section 6, systems with disparate purposes can be described by this same model.

Processing. In the following we provide details about specific features of the processing steps of the quality metrics pipeline.

1. *Data Transformation* (source data → transformed data). In the original pipeline this step has the main role to put the data in a tabular format, hence the original name *tabular data* of its output. Since here we focus on high-dimensional data we assume the source data to be already in a tabular format and we rename it into *transformed data*. At this stage data transformation is responsible for the generation of alternative data subsets or derivations. Common operations include: *feature selection*, *projection*, *aggregation*, and *sampling*.

2. *Visual Mapping* (transformed data → visual structures). Visual mapping is the core stage of the pipeline where data dimensions are mapped to visual features to form visual structures. Distinct mappings of data features to visual features provide alternatives

[1] http://scholar.google.com/

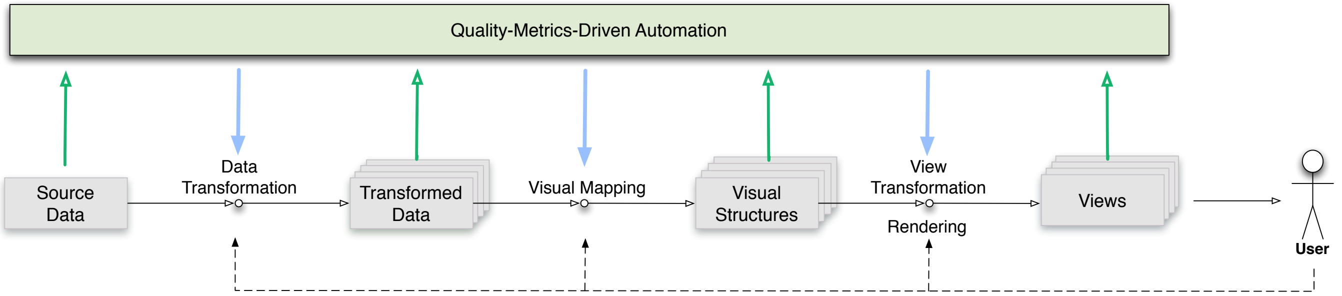


Fig. 4. Quality metrics pipeline. The pipeline provides an additional layer named quality metrics base automation on top of the traditional information visualization pipeline [12]. The layer obtains information from the stages of the pipeline (the boxes) and influences the processes of the pipeline through the metrics it calculates. The user is always in control.

that can again be evaluated in terms of quality metrics. The most common type of operation at this stage is the generation of orderings; by assigning data dimensions to visualization axes in different orders. In general, alternatives can be generated by considering the full set of visual features (e.g., color, size, shape).

3. *Rendering/View Transformation (visual structures → views).* Rendering transforms visual structures into views by specifying graphical properties that turn these structures into pixels. We added the word Rendering to the pipeline to emphasize the role of the image space; many quality metrics are thus calculated directly in the image space considering the pixels generated in the visualization process. At this stage alternatives views of the same structures can be generated automatically. Surprisingly, as we discuss in Section 7, this stage is, in the context of our inquiry, rarely used.

Quality metrics computation. Quality metrics can draw information from any of the stages of the pipeline. As we describe later in Subsection 5.1 quality metrics can be calculated in the data space, image space or a combination of the two. Metrics calculated at the *View* stage draw information from the rendered image, whereas the others draw information from the data space (and elements of the visual structures in some few cases). Many different kind of metrics are possible. Our analysis of quality metrics features in Subsection 5.1 provides numerous additional details.

Quality metrics influence. As described above, quality metrics algorithms generate alternatives and organize them into a final representation. At the data processing stage they can for instance generate 1D, 2D, or nD projections (e.g., [20, 22, 44]), data samples (e.g., [9, 28]), or alternative aggregates (e.g., [14]). At the visual mapping stage the layer generates alternative orderings or mappings between data and visual properties (e.g., [39, 42]). At the view stage the layer can generate modifications of the current view like changing the point of view, highlighting specific items, or distorting the visual space (e.g., [4]).

User influence. The quality metrics layer does not want to substitute the user in favor of the machine. While the users can always influence all the stages of the pipeline, their main responsibility becomes to steer the process, e.g., by setting quality metrics parameters, and to explore the resulting views. It is worth noting that the process is not necessarily a linear flow through the steps. As will be evident from the examples in Section 6 in many cases complex iteration takes place.

5 Systematic Analysis

Through our paper review we identified two main areas of investigation. First, we classify the papers according to quality metrics criteria that help explaining their key features. Second, we provide a more detailed categorization of the visualization techniques we have come across.

5.1 Quality Metrics

Through the literature review we identified a number of factors that describe the methods encountered. Each factor has a number of possible values and each paper can assume one or more of these values (see Table 2).

What is measured. This factor describes what is measured by the quality metric. In our analysis we have grouped the metrics in the following categories: *Clustering* metrics measure the extent to which the visualization or the data contain groupings, that is, well-separated clusters that can be easily identified. Clustering is loosely defined because we have encountered many alternative approaches. It is worth to keep in mind that with clustering here we intend any measure in the data or image space which is able to capture groupings. *Correlation* relates to two or more data dimensions and captures the extent to which systematic changes to one dimension are accompanied by changes in other dimensions. Simple Pearson correlation between two variables is one of the most commonly used measure in this category but global correlation among multiple data dimensions are also used [30]. *Outlier* metrics capture the extent to which the data segment under inspection contains elements that behave differently from the large majority of the data, i.e., outliers. *Complex patterns* metrics capture aspects that cannot be easily categorized as any of the classes described above. We detected a number of papers with such measures and grouped all of them in this class. An example is Graph-Theoretic Scagnostics [54] a technique where it is possible to characterize scatter plots with features like “stringy” or “skinny”. *Image quality* refers to metrics where the purpose is not necessarily to find specific patterns but more to identify the degree of organization of a visualization or, as some of the papers call it, the amount of clutter. *Feature preservation* metrics focus on the comparison between a reference state and the representation in the visualization, or between the features in the data and the visualization, with the intent to preserve the features of interest as much as possible. A subset of these papers focus on classified data, searching for projections where the original classes are well separated [46, 48]. In the same category we can find papers that measure the information loss due to data abstraction techniques such as sampling and aggregation [9, 14, 28]. It is worth noticing that in this categorization we classified the techniques according to their main target. This however does not hinder a metric of one type to also detect patterns of another type. For instance, clustering and correlation, as well as complex patterns and image quality, may have such an overlap.

Where it is measured (data/image space). In our review we have found a completely mixed set of approaches with respect to where the metrics are calculated: *data space* or *image space*. Metrics calculated in data space detect data features directly in the data without using information from the view that will be used to display the results. For instance, the Rank-by-Feature technique ranks 1D and 2D projections according to a number of statistical properties calculated only in data space. Metrics calculated in image space bypass the analysis of the data and work directly on the rendered image. Often these methods employ sophisticated image processing techniques like in the work of

Tatu et al. where interesting scatter plots are ranked using a Hough Transformation [48]. A *mixed-space* approach, where both data and and image space are used at the same time, is also possible. We found two distinct cases. Bertini and Santucci [9] present a measure to compare features in the data space to features in the image space; with the intent of preserving as much as possible data features in the final image. Peng et al. [39] measure clutter in relation to the ordering of visualization axes: these calculations need data features (outliers, correlations) and visualization features (e.g., axes adjacency) at the same time. Please note that the entries in Table 2, where both data and image space are present, do not necessarily imply the use of the aforementioned mixed approach. More often, they simply mean that alternative approaches co-exist in the context of the same paper.

Purpose. Purpose describes the main reason for using quality metrics, that is, what is the goal to be achieved with the metric. We identified the following purposes. *Projection* aims at finding subsets of the original dimensions in which interesting patterns reside, e.g., analyzing all the possible 2D projections of a multidimensional data set by checking whether interesting groupings exist in a scatter plot. *Ordering* aims at finding, where possible, an ordering of the visualization axes that eases the visual detection of interesting patterns. Parallel coordinates is a classical example where the order of the axes greatly influences the chances of detecting interesting patterns in the data. *Abstraction* aims at maintaining or controlling a certain degree of data representation quality when data reduction techniques are used to increase the scalability of a visualization. Sampling and aggregation are the two main types of abstraction techniques we encountered. For instance, in [14] the authors propose a data abstraction technique that permits to measure the information loss due to abstraction and to find a trade-off between data loss and data reduction. *Visual mapping* aims at finding interesting mappings between the original data features and the visual features of the visualization technique. Features such as color, size or shape fall into this category. *View optimization* aims at modifying parameters of the view with the intent to produce better visualizations, in which, for example, data segments with a high degree of interest are highlighted.

Interaction. The last column of the table indicates which papers offer the possibility to interact with the quality-metrics-based automation. We extracted two main classes of interaction: *threshold selection* and *metrics selection*. With threshold selection we mean the possibility to set thresholds in the quality metrics computation mechanism (e.g., the data abstraction level in [14] or the density estimation smoothing parameter in [20]). With metrics selection we mean systems in which the user can either switch from one metrics to another or combine them into an integrated one (e.g., [15, 30]). Please note that some of the papers may contain interaction capabilities and still be marked as not interactive because they do not provide direct interaction with the quality metrics mechanisms.

5.2 Visualization

The original table we have designed to classify the full set of papers (see Table 2 below) contains a rough categorization of visualization techniques into three main classes: *scatter plots (SP)*, *parallel coordinates (PC)* and *others* (which include a fairly large number of different techniques). While this categorization helps understanding how these techniques distribute over the whole set of papers (SP and PC accounts for 80% of the total) it does not say anything about key features of visualization techniques; especially those closely related to the usage of quality metrics.

We define **layout dimensionality** as the number of data axes a visualization has. A *data axis* is the visualization feature that establishes what *position* a single visual mark takes in the visualization. For instance, scatter plots have dimensionality two because they can accommodate two spatial dimensions.

The visualization techniques are classified into 1D, 2D, 3D, 4D and nD, where nD stands for techniques that can accommodate an arbitrary number of dimensions (with obvious scalability limits when the number of dimensions grows too big).

It is worth noticing that in general every visualization has an additional number of visual features to which data features can be mapped, e.g., color and size, but here we focus on the layout because it is the variable that most characterizes every visualization technique and that has the biggest impact on the use of quality metrics. Table 1 shows the dimensionality of all the techniques we have identified in the review.

The visualization techniques that are not in the nD class necessarily need an additional mechanism for the analysis of high-dimensional data. Typically, as discussed below, they are organized in a higher level structure that accommodates several projections. Those which can accommodate an arbitrary number of dimensions (nD) all need some kind of ordering mechanisms.

Table 1. Visualization techniques categorized by their layout dimensionality (i.e., the number of axes of the visualization).

Visualization	Layout Dimensionality
histogram	1D
jigsaw map [53]	1D
scatter plot	2D
pixel bar charts [32]	4D
dimensional stacking [33]	nD
matrix [7]	nD
parallel coordinates [26]	nD
radvis [24]	nD
scatter plot matrix [56]	nD
star glyphs [45]	nD
table lens [40]	nD

While not explicitly discussed in any of the reviewed papers, we have noticed that often a quality-metrics-driven approach needs some kind of (implicit or explicit) **meta-visualization**. With meta-visualization we mean a visualization of visualizations. More specifically, a visualization layout strategy that organizes single visualizations into an organized form. For instance, when a quality-metrics-driven technique produces a number of interesting scatter plots as an output, there is the need to organize them into a schema that facilitates their comprehension and analysis (e.g., organized into a list sorted by interestingness). From our analysis we have identified the following main meta-visualization strategies. *List:* a layout strategy that organizes visualizations in an ordered linear fashion (often sorted to reflect quality metrics rankings). *Matrix:* a layout strategy that organizes visualizations in a grid format, where grid entries are organized according to some data features (e.g., column and rows represent data dimensions) (often called also Small Multiples, Trellis, Lattice, Facets).

It is worth noticing that some basic visualization techniques can be considered meta-visualizations themselves. A notable example is the *scatter plot matrix* which shows a set of scatter plots organized in a matrix layout.

In general there is a strong interplay between visualizations and meta-visualizations. As mentioned above, techniques with a fixed dimensionality need to be organized in a meta-visualization. The meta-visualization influences the ordering of the visualizations and in some cases also the content. For instance, the matrix layout requires that the visualization within a grid cell corresponds to the data values it represents.

Finally, meta-visualizations can themselves be influenced by quality metrics. All the layout strategies have some degree of freedom in terms of reordering, and an optimal reordering (according to some given goal) can only be achieved by searching in the space of solutions (e.g., as presented in [39]).

6 Examples

In this section we provide three selected examples from our review as a way to show how our proposed model can describe existing approaches in this area. We selected the examples in a way to cover as many interesting aspects as possible. In particular, we picked papers

Table 2. Quality metrics papers classified according to quality metrics factors (sorted by purpose).

Paper Title	Visualization technique			What is measured						Where it is measured		Purpose					Inter-act-ion
	SP	PC	other	clustering	correlation	outliers	complex patterns	image quality	feature pres.	data	image	projection	ordering	abstraction	visual mapping	view optimization	
										space							
A Projection Pursuit Algorithm for Exploratory Data Analysis - Friedman & Tukey [21]	SP			clustering						data		projection					
A Rank-by-Feature Framework for Unsupervised Multidimensional Data Exploration Using Low Dimensional Projections-Seo & Shneiderman[44]	SP		histogram, matrix, list	clustering	correlation	outliers	complex patterns			data		projection					S
Finding and Visualizing Relevant Subspaces for Clustering High-Dimensional Astronomical Data Using Connected Morphological Operators**[20]	SP		histogram	clustering							image	projection					T
Graph-Theoretic Scagnostics - Wilkinson et al. [54]	SP			clustering		outliers	complex patterns				image	projection					
Selecting good views of high-dimensional data using class consistency - Sips et al. [46]	SP								class pres.	data		projection					T
Coordinating computational and visual approaches for interactive feature selection and multivariate clustering - Guo [22]			matrix		correlation					data		projection	ordering				
Exploring High-D Spaces with Multiform Matrices and Small Multiples - MacEachern et al. [35]			pixel based vis., matrix, small multiples		correlation					data		projection	ordering				
Improving the Visual Analysis of High-Dimensional Datasets Using Quality Measures - Albuquerque et al. [4]			jigsaw map, radvis, table lens	clustering	correlation	outliers				data	image	projection	ordering		visual mapping		
Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration of High Dimensional Datasets - Yang et al. [58]		PC	histogram, star glyphs		correlation					data		projection	ordering			view optimization	S, T
Interactive Dimensionality Reduction Through User-defined Combinations of Quality Metrics - Johansson & Johansson [30]		PC		clustering	correlation	outliers				data		projection	ordering				S, T
Pargnostics: Image-Space Metrics for Parallel Coordinates - Dasgupta & Kosara [15]		PC		clustering	correlation			image quality			image	projection	ordering				S
Combining automated analysis and visualization techniques for effective exploration of high-dimensional data - Tatu et al. [48]	SP	PC		clustering	correlation		complex patterns		class pres.	data	image	projection	ordering				
High-Dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distributions - Wilkinson et al. [55]	SP	PC		clustering		outliers	complex patterns				image	projection	ordering				
Clutter Reduction in Multi-Dimensional Data Visualization Using Dimension Reordering - Peng et al. [39]	SP	PC	star glyphs, dim. stacking		correlation	outliers		image quality		data	image		ordering				
Similarity Clustering of Dimensions for an Enhanced Visualization of Multidimensional Data - Ankerst et al. [5]		PC	recursive pattern, circle segments		correlation					data			ordering				
Measuring Data Abstraction Quality in Multiresolution Visualizations - Cui et al. [14]	SP	PC	histogram						feature pres.	data				abstraction			T
Quality Metrics for 2D Scatterplot Graphics: Automatically Reducing Visual Clutter - Bertini & Santucci [9]	SP			clustering					feature pres.	data	image			abstraction			
A Screen Space Quality Method for Data Abstraction - Johansson & Cooper [28]		PC							feature pres.		image			sampling			
Enabling Automatic Clutter Reduction in Parallel Coordinate Plots - Ellis & Dix [12]		PC						image quality			image			sampling			T
Pixnostics: Towards measuring the value of visualization - Schneidewind et al. [42]			jigsaw map, pixel bar chart		correlation		complex patterns			data	image				visual mapping		

** Ferdosi et al.

Legend: SP = scatter plot (& matrix), PC = parallel coordinates, feature/class pres. = feature/class preservation, S = select metric, T = set threshold.

with different purposes because they guarantee a larger variety of features.

The first example comes from the work of Tatu et al. [48]. The main goal of this paper is to find interesting projections of n-dimensional data using image processing techniques. The paper contains several measures and visualization techniques, here we focus only on the part dealing with parallel coordinates and one specific metric.

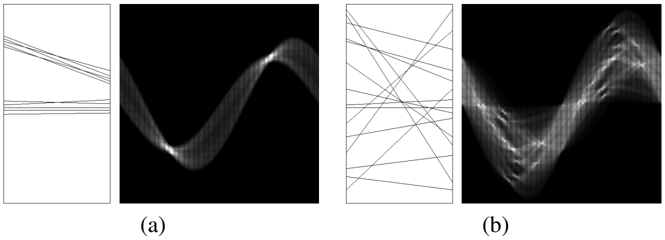


Fig. 5. Synthetic examples of parallel coordinates and their Hough transform: (a) two well defined clusters with bright areas in the hough plane, (b) no clear clusters visible, no bright pattern in the hough space [48].

The basic idea of the method is to generate all possible 2D combinations of the original dimensions and evaluate them in terms of their ability to form clusters in a 2-axis parallel coordinates representation (see Figure 5). Every pair of axis is evaluated individually using a standard image processing technique (the Hough transform), which permits to discriminate between uniform and chaotic distributions of line angles and positions (for details please refer to the original paper). Once interesting pairs have been extracted, they are joined together to form groups of parallel coordinates of a desired (user-defined) size (e.g., in Figure 6, groups of 4-dimensional parallel coordinates).

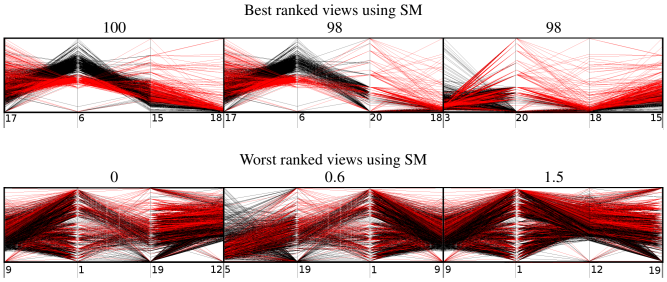


Fig. 6. Ranked list of four-dimensional parallel coordinates. Best ranked on top, worst ranked on the bottom [48].

Figure 7 presents the pipeline for this example. We can recognize three main elements: (A) all 2D parallel coordinates are generated in the *data transformation* phase; (B) all the alternatives are evaluated in the image space at the *view* stage; (C) the algorithm combines the interesting segments into a list of parallel coordinates (like those in Figure 6) using the *visual mapping* stage.

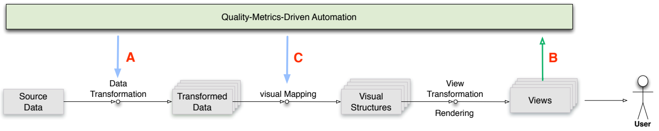
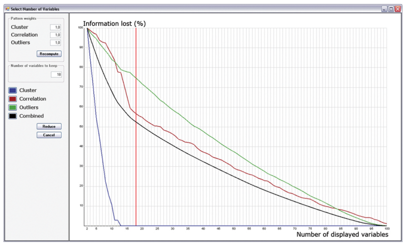


Fig. 7. Quality metrics pipeline for [48]: (A) generation of alternatives; (B) evaluation of alternatives (image space); (C) creation of the final representation.

The technique uses *parallel coordinates (PC)* as principal visualization technique and a *list* as a meta-visualization. It measures *clustering* properties, in the *image space*, and its main purpose is to find interesting *projections*. Interaction, in the way it is discussed in the paper, is very limited if not absent.

The second example comes from the work of Johansson and Johansson on interactive feature selection [30]. The technique ranks every single dimension for its importance using a combination of correlation, outlier, and clustering features calculated on the data. This ranking is used as the basis for an interactive threshold selection tool by which the user can decide how many dimensions to keep; weighting the choice with the corresponding information loss presented by the chart (see Figure 8). Once the user selects the desired number of dimensions the system presents the result with parallel coordinates and automatically finds a good ordering using the same data features calculated for ranking the dimensions. The user can also choose different weighting schemes to focus more on correlation, outliers or clusters. Figure 9 shows the results of clustering (top) and correlation (bottom).


Select Number of Variables

Pattern weights
Cluster 1,0
Correlation 1,0
Outliers 1,0

Number of variables to keep 18

Fig. 8. Interactive chart to select number of dimensions to keep vs. information loss [30].

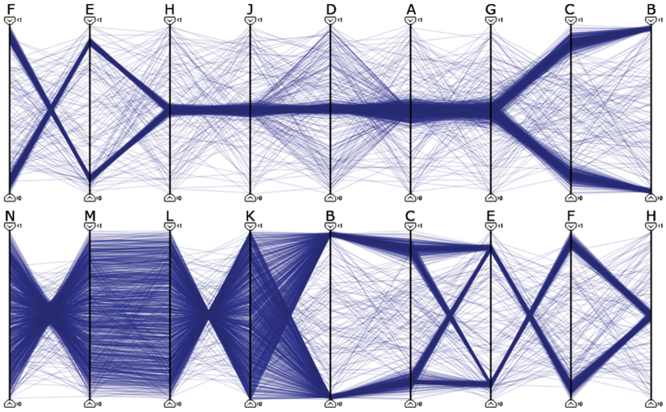


Fig. 9. Top: best ordering to enhance clustering. Bottom: best ordering to enhance correlation [30].

Figure 10 shows the pipeline for this example. Again we have three main elements: (A) every single dimension is ranked by the quality metrics directly from the *source data*. The reason why the source data is needed is because the importance measure of a single dimension is computed taking into account the full set of dimensions (see the paper for details); (B) the user selects the dimensions guided by the quality metrics, both the user and the quality metric influence the *data transformation* process; (C) the system finds the best ordering according to the weighting scheme proposed by the user producing one specific *visual mapping*. The view is presented to the user.

This technique uses *parallel coordinates* as principal visualization. There is no meta-visualization to organize alternative results in a schema but the interactive chart functions as a way to pilot the generation of alternatives. It measures *clustering*, *correlation* and *outliers* in the *data space* and its main purpose is to find interesting *projections* and *orderings*. Interaction plays a central role in the selection of the number of dimensions and in the weighting scheme.

The third example is taken from the work of Cui et al. on data abstraction quality [14]. This paper proposes a technique to create

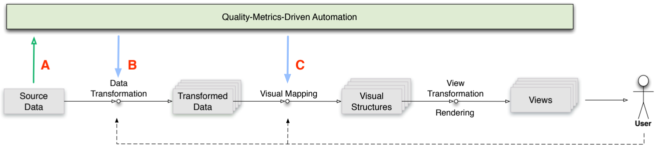


Fig. 10. Pipeline for [30]: (A) dimensions ranked by their importance; (B) selection of number of dimensions to retain vs. information loss; (C) creation of the final mapping with ordering.

abstracted visualizations in a user-controlled manner. The system features data abstraction metrics (Histogram Difference Measure and Nearest Neighbor Measure) and controllers to let the user find a tradeoff between abstraction level and information loss. In particular, the data abstraction quality is calculated by comparing features of the original data to features in the sampled or aggregated data.

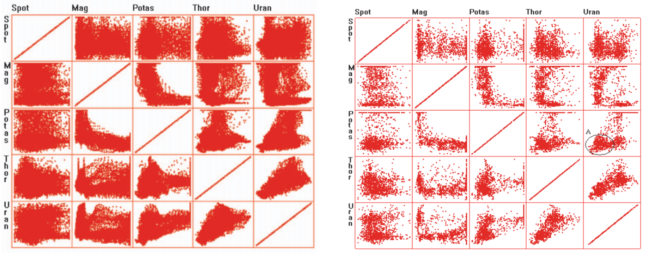


Fig. 11. Visual abstraction of a scatter plot matrix from [14].

Figure 12 shows the pipeline for this example. We have two main elements: (A) the data abstraction quality measures are calculated by comparing the *source data* to the *transformed data*; (B) the user selects the desired abstraction quality and receives feedback on its quality by steering the *data transformation process*.

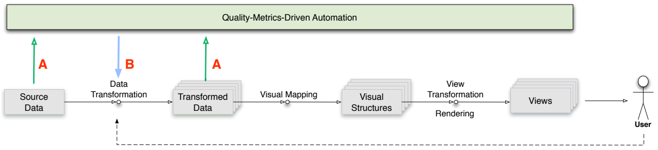


Fig. 12. Pipeline for [14]: (A) data features compared between the original data and the abstracted data; (B) instantiation of the desired abstraction level guided by quality metrics.

The paper applies the technique to *scatter plots* and *parallel coordinates* but it is generic enough to be applied to many other techniques. There is no meta-visualization to organize alternative results but similarly to the second example an interactive chart is used to set an abstraction threshold (see Figure 13). It measures *feature preservation*, and its main purpose is *abstraction*. Interaction plays a central role in the selection of the right abstraction level.

These three examples cover many aspects discussed in the paper, especially metrics calculated in the data vs. image space, different purposes, different measure types, different uses of the pipeline, and different interaction levels. Many of the papers we have reviewed have similar elements and functions, nonetheless there are others that deviate considerably from these ones. While we cannot provide the full set of examples in the scope of this paper we discuss in Section 7 some findings that stem from the analysis of the whole set, including those with uncommon approaches.

7 Findings

In the following we discuss some major trends we have observed during our analysis.

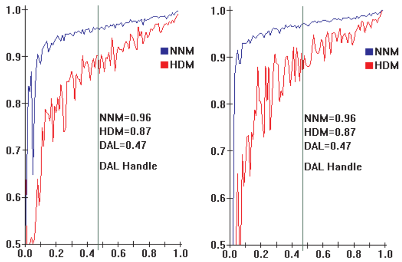


Fig. 13. Visual abstraction chart with threshold setting for the abstraction level and feedback on abstraction quality [14].

From the visualization point of view we already discussed the role of meta-visualizations, that is, visualizations with the purpose to accommodate other visualizations. During the paper review we found very limited explicit discussions of this aspect which we deem extremely relevant. Many of the papers we have analyzed seem to assume that providing a simple list of interesting visualizations will automatically solve the user's task. To the best of our knowledge, the only work that analyzes the issue explicitly and in great depth is the Trellis display [6], which organizes the display in a way to make patterns among views apparent. We believe a deeper investigation of this issue is needed.

Interestingly, some of the papers we reviewed do take care of the navigation issue, that is, how to explore configurations automatically found by the algorithm. These papers usually provide an additional visualization that permits to navigate from one configuration to another. For instance, Johansson et al. provide a line chart visualization to interactively show alternative projections in parallel coordinates [30]. Similarly, "hierarchical dimension ordering" [58] uses the InterRing visualization to the let the user navigate through alternative subsets of dimensions organized in a hierarchical fashion. Finally, the Rank-by-Feature framework [44] uses color-coded interactive lists and scatter plot matrices to provide a preview of the statistical properties of each views.

We also noticed a lack of systematic approaches to the ordering problem; every paper proposes its own method. The whole topic of *seriation*, introduced in the early work of Bertin [7] and discussed in depth by Hahsler et al. [23], deserves deeper investigation and acknowledgment. Also, innovative ways of ordering data dimensions may exist, like the *eulerian tours and hamiltonian decompositions* presented by Hurley et al. [25], which explores the possibility of repeating the axes in order to reduce dependency on a specific order.

In Subsection 5.2 we list a series of meta-visualizations that we have found, namely list, small multiples, and matrix. We believe this list can be expanded if novel solutions are developed. A promising one we have noticed in a few papers, but not included in the review (because they are not specifically using quality metrics) is the idea of arranging iconic versions of the visualizations generated in a scatter plot view (e.g., using MDS or similar techniques). Such a technique is for instance proposed in the work of Yang et al. where pixel-based icons are laid out with an MDS projection in a scatter plot [57].

Another issue we noticed from our analysis is the limited use of the *visual mapping* and *view transformation* functions in the pipeline. More specifically, visual mapping is almost exclusively used as a way to generate alternative orderings, taking into account exclusively the mapping between the original data dimensions and the visualization axes. But alternative mappings can also be generated by linking data dimensions to the whole spectrum of visual features like color, size, shape, etc., as is common in several systems based on visual languages like ggplot2[1], tableau[3], and protovis[2]). Pixnostics [42] is the only technique in our review presenting this kind of a process supported by quality metrics.

View transformation is also rarely used in the quality metrics pipeline. The only example we found is the use of quality metrics to automatically select focus area parameters in table lens [4]. The automatic selection of interesting point of views in 3D scatter plots, for example, is one clear case where the use of quality metrics at the view transformation stage would be beneficial. Another one is the automatic highlight of interesting items in a view (e.g., visual boosting in pixel-based visualizations [38]).

Finally, the purposes we have considered can be roughly classified into two broad higher level purposes: finding interesting visualizations and scaling visualizations to larger data sets. When considering these goals it is evident how *clustering*, *correlation*, *outliers*, and *complex patterns* support more the first goal, whereas *image quality* and *feature preservation* tend to support more the second one. One interesting pending issue is whether the use of quality metrics in high-dimensional data is confined to these two general purposes. One purpose which to the best of our knowledge is totally unexplored is the use of quality metrics to automatically or semi-automatically compare different visual techniques of the same data.

8 Filling the gaps: a research agenda

In the following we present a selected set of research issues we deem important for the advancement of quality-metrics-driven data visualization.

Evaluation and applications. Surprisingly, none of the papers we have analyzed reported on user evaluation. While we are convinced that quality metrics are useful and need to be further developed, we also realize that the whole idea has not yet been tested. Usefulness is therefore one of the most important aspect to consider, followed by usability issues. To the best of our knowledge there are no studies reporting on the use of the quality metrics approach in real-world settings. Observatory studies or even simple case studies would greatly improve the approach and most likely direct research to specific issues hard to anticipate without observation.

Perceptual tuning. All the metrics that work in the image space try to simulate the human pattern recognition machinery to some extend. They try to partially substitute human vision with image processing algorithms with the (implicit) assumption that algorithm rankings will match user rankings. This assumption needs a much deeper investigation. The study presented in [49], where quality metrics rankings of clusters in scatter plots are compared to human rankings, represents a first step in this direction. In addition, it is necessary to validate and tune the image space metrics in a way that the parameters take models of human perception into account. Excellent examples of initial steps in this direction are in the following papers [29, 34, 41], where the perception of visual patterns has been tuned according to user studies aimed at modeling the way humans perceive them.

Metrics systematization. During our review we collected a very large number of alternative quality metrics, some calculated in data space some in image space. While this proliferation of metrics is a sign of the richness of this approach, it is currently very hard to compare them and understand which one is suitable for a given task. Some authors provide a number of metrics in the same environment letting the user choose which one to use. Nonetheless we fear that this approach with limited guidance may not be effective for end users, especially, if there is a lack of understanding of the level of redundancy between one metric and another. Similarly, given the above mentioned dichotomy, it is hard if not impossible to state which approach yields the best results in which contexts. On a side note, the mixed approach of giving the user the possibility to combine several metrics into a composite one need much more investigation, validation, and guidance.

Scalability. Image space and data space quality metrics have different scalability issues. Quality metrics in image space have the advantage of being independent from the original data size, e.g., [15], that is, their computational complexity only depends on the screen dimensions. However, as data grows in size, virtually all visualizations experience some degrees of degradation that may influence the discriminatory power of the metric. For instance, visualizations with a lot of clutter might hinder the discovery of the desired patterns. Quality metrics in data space, on the other hand, are expected to be more robust in terms of pattern detection, but their computation is directly affected by data size. A thorough investigation of these issues and how to find a compromise between the two is clearly an interesting subject for future research.

9 Limitations

Our work has some important limitations to take into account; first of all its subjective nature. We are by no means suggesting this is the only way to describe the current state of quality metrics in high-dimensional visualization. There are no doubt a number of equally good alternative ways to describe it; this paper provides a much-needed starting point. We encourage the reader to use the paper as a way to get inspiration for further research and to understand its status.

Similarly, while we did our best to follow a thorough methodology (see Section 3), there might be relevant papers we overlooked. Even though we tried to be very broad and inclusive, the review is heavily influenced by our background. Especially, given our focus on Computer Science we might have missed relevant literature from Statistics. However, we feel confident that at this point of our review any additional paper would not change the structure or the elements of our model. In other terms, the real goal of our review was not to include every possible paper on the discussed matter but more to have enough coverage to build a coherent and useful picture.

10 Conclusion

We presented a systematic analysis of quality metrics as a way to support the exploration of high-dimensional data sets. Quality metrics have been used in a variety of contexts and purposes. With this work we started a collection of these disparate systems under one umbrella and provided a way to reason about their characteristic features. Specifically, we presented an analysis of the visualization techniques, the quality metrics, and the processing pipeline. The analysis has two main outcomes. First, it permits to describe the methods in details to capture their key components. Second, as shown in Section 7 and Section 8, it permits to spot interesting research gaps and promising directions for future research. While we consider this work just an initial step, we hope it will spur new ideas and support researchers and practitioners in the development of interesting new applications and novel techniques.

Acknowledgments

This work was partially supported by DFG Research Training Group GK-1042 ”Explorative Analysis and Visualization of Large Information Spaces”, University of Konstanz.

References

[1] ggplot2. `http://had.co.nz/ggplot2/`.

[2] Protovis. `http://vis.stanford.edu/protovis/`.

[3] Tableau. `http://www.tableausoftware.com/`.

[4] G. Albuquerque et al. Improving the visual analysis of high-dimensional datasets using quality measures. In *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, 2010.

[5] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 1998.

[6] R. A. Becker, W. S. Cleveland, and M.-J. Shyu. The visual design and control of trellis display. *Journal of Computational and Graphical Statistics*, 5(2):123–155, 1996.

[7] J. Bertin. *Semiology of graphics*. University of Wisconsin Press, 1983.

[8] E. Bertini and D. Lalanne. Investigating and reflecting on the integration of automatic data analysis and visualization in knowledge discovery. *SIGKDD Explor. Newsl.*, 11:9–18, 2010.

[9] E. Bertini and G. Santucci. Quality metrics for 2D scatterplot graphics: Automatically reducing visual clutter. In *Proc. Smart Graphics (SG)*, 2004.

[10] E. Bertini and G. Santucci. Visual quality metrics. In *Proc. AVI workshop on BEyond time and errors: noveL evaluation methods for Information Visualization (BELIV)*. ACM, 2006.

[11] R. Brath. Metrics for effective information visualization. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 1997.

[12] S. K. Card, J. D. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., 1999.

[13] E. H. Chi. A taxonomy of visualization techniques using the data state reference model. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2000.

[14] Q. Cui, M. Ward, E. Rundensteiner, and J. Yang. Measuring data abstraction quality in multiresolution visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 12:709–716, 2006.

[15] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE Trans. on Visualization and Computer Graphics*, 16:1017–1026, 2010.

[16] C. Dunne and B. Shneiderman. Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts. Technical Report HCIL-2009-13, University of Maryland, 2009.

[17] G. Ellis and A. Dix. Enabling automatic clutter reduction in parallel coordinate plots. *IEEE Trans. on Visualization and Computer Graphics*, 12:717–724, 2006.

[18] G. Ellis and A. Dix. A taxonomy of clutter reduction for information visualisation. *IEEE Trans. on Visualization and Computer Graphics*, 13:1216–1223, 2007.

[19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM*, 39:27–34, 1996.

[20] B. J. Ferdosi et al. Finding and visualizing relevant subspaces for clustering high-dimensional astronomical data using connected morphological operators. In *Proc. IEEE Conf. Visual Analytics Science and Technology (VAST)*, 2010.

[21] J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.*, 23:881–890, September 1974.

[22] D. Guo. Coordinating computational and visual approaches for interactive feature selection and multivariate clustering. *Information Visualization*, 2:232–246, 2003.

[23] M. Hahsler, K. Hornik, and C. Buchta. Getting things in order: An introduction to the R package seriation. *Journal of Statistical Software*, 25(3):p. 1–34, 3 2008.

[24] P. Hoffman, G. Grinstein, and D. Pinkney. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proc. Workshop on New Paradigms in Information Visualization and Manipulation (NPIVM)*. ACM, 1999.

[25] C. B. Hurley and R. W. Oldford. Pairwise display of high-dimensional information via eulerian tours and hamiltonian decompositions. *Journal of Computational and Graphical Statistics*, 19(4):861–886, 2010.

[26] A. Inselberg and B. Dimsdale. Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proc. IEEE Conf. on Visualization (VIS)*. IEEE Computer Society Press, 1990.

[27] H. Jänicke and M. Chen. A salience-based quality metric for visualization. *Computer Graphics Forum (Proc. EuroVis)*, 29(3):1183–1192, 2010.

[28] J. Johansson and M. Cooper. A screen space quality method for data abstraction. *Computer Graphics Forum (Proc. EuroVis)*, 27(3):1039–1046, 2008.

[29] J. Johansson, C. Forsell, M. Lind, and M. Cooper. Perceiving patterns in parallel coordinates: determining thresholds for identification of relationships. *Information Visualization*, 7:152–162, April 2008.

[30] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE Trans. on Visualization and Computer Graphics*, 15:993–1000, 2009.

[31] D. A. Keim et al. Visual analytics: Scope and challenges. In S. Simoff, M. H. Boehlen, and A. Mazeika, editors, *Visual Data Mining: Theory, Techniques and Tools for Visual Analytics*. Springer, 2008.

[32] D. A. Keim, M. C. Hao, U. Dayal, and M. Hsu. Pixel bar charts: A visualization technique for very large multi-attribute data sets. *Information Visualization*, 1(1):20–34, 2002.

[33] J. LeBlanc, M. O. Ward, and N. Wittels. Exploring N-dimensional databases. In *Proc. of the IEEE Conf. on Visualization (VIS)*. IEEE Computer Society Press, 1990.

[34] J. Li, J.-B. Martens, and J. J. van Wijk. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9:13–30, 2010.

[35] A. MacEachren et al. Exploring high-D spaces with multiform matrices and small multiples. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2003.

[36] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. on Graphics*, 5:110–141, 1986.

[37] N. Miller, B. Hetzler, G. Nakamura, and P. Whitney. The need for metrics in visual information analysis. In *Proc. Workshop on New Paradigms in Information Visualization and Manipulation*. ACM, 1997.

[38] D. Oelke et al. Visual boosting in pixel-based visualizations. *Computer Graphics Forum (Proc. EuroVis)*, 2011.

[39] W. Peng, M. O. Ward, and E. A. Rundensteiner. Clutter reduction in multi-dimensional data visualization using dimension reordering. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2004.

[40] R. Rao and S. K. Card. The table lens: merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In *Proc. SIGCHI Conf. on Human factors in Computing Systems (CHI)*. ACM, 1994.

[41] R. A. Rensink and G. Baldridge. The perception of correlation in scatterplots. *Computer Graphics Forum (Proc. EuroVis)*, 29(3):1203–1210, 2010.

[42] J. Schneidewind, M. Sips, and D. A. Keim. Pixnostics: Towards measuring the value of visualization. In *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, 2006.

[43] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Trans. on Visualization and Computer Graphics*, 16:1139–1148, 2010.

[44] J. Seo and B. Shneiderman. A rank-by-feature framework for unsupervised multidimensional data exploration using low dimensional projections. *Information Visualization*, 4:96–113, 2005.

[45] J. H. Siegel, E. J. Farrell, R. M. Goldwyn, and H. P. Friedman. The surgical implication of physiologic patterns in myocardial infarction shock. *Surgery*, 72:126–141, 1972.

[46] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum (Proc. EuroVis)*, 28(3), 2009.

[47] A. Strauss and J. M. Corbin. *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, 1998.

[48] A. Tatu et al. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *Proc. IEEE Symp. Visual Analytics Science and Technology (VAST)*, 2009.

[49] A. Tatu et al. Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data. In *Proc. International Conf. on Advanced Visual Interfaces (AVI)*. ACM, 2010.

[50] M. Tory and T. Moller. Rethinking visualization: A high-level taxonomy. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2004.

[51] E. R. Tufte. *The visual display of quantitative information*. Graphics Press, 1986.

[52] C. Ware, H. Purchase, L. Colpoys, and M. McGill. Cognitive measurements of graph aesthetics. *Information Visualization*, 1:103–110, June 2002.

[53] M. Wattenberg. A note on space-filling visualizations and space-filling curves. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2005.

[54] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2005.

[55] L. Wilkinson, A. Anand, and R. Grossman. High-dimensional visual analytics: Interactive exploration guided by pairwise views of point distributions. *IEEE Trans. on Visualization and Computer Graphics*, 12:1363–1372, 2006.

[56] C. William S. and M. E. McGill. *Dynamic Graphics for Statistics*. Wadsworth Inc., 1988.

[57] J. Yang, D. Hubball, M. O. Ward, E. A. Rundensteiner, and W. Ribarsky. Value and relation display: Interactive visual exploration of large data sets with hundreds of dimensions. *IEEE Trans. on Visualization and Computer Graphics*, 13:494–507, 2007.

[58] J. Yang, W. Peng, M. O. Ward, and E. A. Rundensteiner. Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets. In *Proc. IEEE Symp. Information Visualization (InfoVis)*, 2003.

[59] J. Yang, M. O. Ward, E. A. Rundensteiner, and A. Patro. Interring: a visual interface for navigating and manipulating hierarchies. *Information Visualization*, 2:16–30, March 2003.

[60] J. S. Yi, Y. a. Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans. on Visualization and Computer Graphics*, 13:1224–1231, 2007.