Visualizing Dimensionally-Reduced Data: Interviews with Analysts and a Characterization of Task Sequences

Matthew Brehmer
Univ. British Columbia
Vancouver, Canada
brehmer@cs.ubc.ca

Michael Sedlmair
Univ. Vienna
Vienna, Austria
michael.sedlmair@univie.ac.at

Stephen Ingram
Univ. British Columbia
Vancouver, Canada
sfingram@cs.ubc.ca

Tamara Munzner
Univ. British Columbia
Vancouver, Canada
tmm@cs.ubc.ca

ABSTRACT

We characterize five task sequences related to visualizing dimensionally-reduced data, drawing from data collected from interviews with ten data analysts spanning six application domains, and from our understanding of the technique literature. Our characterization of visualization task sequences for dimensionally-reduced data fills a gap created by the abundance of proposed techniques and tools that combine high-dimensional data analysis, dimensionality reduction, and visualization, and is intended to be used in the design and evaluation of future techniques and tools. We discuss implications for the evaluation of existing work practices, for the design of controlled experiments, and for the analysis of post-deployment field observations.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Evaluation/Methodology*

Keywords

Tasks, dimensionally-reduced data, interview study.

1. INTRODUCTION

Dimensionality reduction is the process of reducing a high-dimensional dataset to a lower-dimensional representation that retains most of its important structure. It has been an active research area throughout several decades and across many domains, from its origins in psychology [46, 57] through statistics [8] to machine learning [18, 45, 49] and visualization [15, 16, 19, 55].

While many techniques and tools combining dimensionality reduction with visualization have been proposed, there is still no perfect automated solution that will generate the most effective visualization for every situation. Analysts are faced with complex choices between alternative dimensionality reduction techniques and between different visualization techniques for analyzing the resulting data. These choices are strongly dictated by the analysts' data and tasks [29, 44]. The statistics and machine learning communities have provided extensive classifications of dimensionality reduction techniques based on data and technique characteristics [11, 12, 13, 18, 50, 54]. In contrast, there is very little that is explicitly stated about the characteristics of tasks that analysts engage in when visually analyzing dimensionally-reduced data. To guide designers, analysts, and those who conduct evaluations of techniques and tools, a better understanding of these tasks is essential.

BELIV '14, November 10 2014, Paris, France.

http://dx.doi.org/10.1145/2669557.2669559

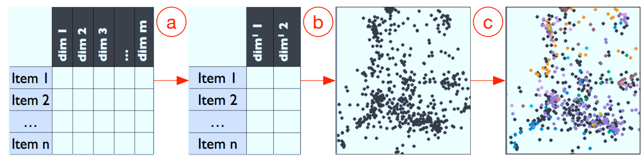


Figure 1: (a) Data is reduced to 2D; (b) encoded in a scatterplot to *verify* visible clusters; and (c) colour-coded according to preexisting class labels to *match* clusters and classes.

Our contribution is a characterization of five task sequences related to the visualization of dimensionally-reduced data: *naming synthesized dimensions*, *mapping a synthesized dimension to original dimensions*, *verifying clusters*, *naming clusters*, and *matching clusters and classes*. In the last of these sequences, illustrated in Figure 1, an analyst uses dimensionality reduction and scatterplots to verify clusters, and then match them with existing classes. Our characterization is based on an in-depth analysis of ten interviews with analysts who use dimensionality reduction for visualizing their data, as well as on a literature review of papers that apply dimensionality reduction for the purpose of data visualization. Our analysis framework is a recently proposed typology of abstract tasks [7], and allows practitioners to identify task sequences based on observed work practices, occurring in requirements gathering activities and in field evaluations of deployed tools.

2. RELATED WORK

Characterizing tasks. The systematic analysis of worker activities and tasks is a critical process in the design and evaluation of technology, and task analysis frameworks appear in many different fields, including human factors and ergonomics [51], human-computer interaction [32], and visualization [7, 38].

While many characterizations of visualization tasks are agnostic to data type, some address specific types of data [42], such as network-based data [25], time-oriented data [24], and tabular data [14]. Data-specific task characterizations are important because they provide a less generic description of tasks. They consider a specific set of *data abstractions*, facilitating a mapping to appropriate visual encoding and interaction techniques [7, 29]. A less generic description of tasks is also critical for evaluation, such as when specifying tasks to be performed by participants in controlled experiments. In this paper, we propose a data-type specific characterization of task sequences for *dimensionally-reduced* data.

Characterizations of tasks are often based on their authors' own experience in conjunction with a thorough consideration of the literature [3, 42], while others are based on observations of human behaviour in controlled laboratory settings [2]. In contrast, our characterization of task sequences is primarily based on an interview study with analysts working with their own data [28], allowing us to ground our findings in real data analysis practices.

Mapping tasks to techniques for high-dim data analysis. There are many techniques and tools that combine analysis of high-dimensional data, dimensionality reduction, and visualization, including some developed by our research group [15, 16, 53]. While there are helpful characterizations of high-dimensional data analysis techniques [4] and of dimensionally-reduced data [41], the mapping between data, tasks, and appropriate techniques remains unclear [44]. This problem is particularly apparent when designing *workflows*, or instantiations of task sequences within software tools for high-dimensional data analysis [15, 19].

One task for dimensionally-reduced data is that of matching clusters and categorical classes given with the data, discussed below in Section 4.2.3. Based on findings from an empirical data study, Sedlmair et al. identified effective techniques for visualizing this data that support this task [40], and called for similar work to be done for other tasks relating to dimensionally-reduced data. Our characterization of task sequences moves us closer to this goal.

Expert judgments and dimensionally-reduced data. We are aware of one other study involving expert analysts' interpretations of visualizations of dimensionally-reduced data, though they do not share our explicit examination of analysts' domain problems and tasks: Lewis et al. [26] asked expert and novice analysts in a controlled lab setting to subjectively rate the value of 2D scatterplot projections of seven dimensionally-reduced datasets, generated using nine different dimensionality reduction techniques. Their findings showed that experts were more consistent in their positive and negative ratings. Judging the value or quality of a visualization of dimensionally-reduced data should occur regardless of task, and analysts can additionally leverage automated quality metrics based on human perception [1, 4]. We did not seek out novice analysts, though the domain experts we interviewed varied in terms of their perceived understanding of dimensionality reduction; furthermore, we sought to identify and characterize experts' tasks and activities in naturalistic settings, rather than in a controlled lab study.

3. METHODS

Our methodological choice was motivated by a vibrant thread of work in the visualization community using qualitative methods in general [9, 17, 47], and interview studies in particular [21, 22].

Data collection. Between 2010 and 2012, we interviewed 24 data analysts working in academic and industry settings, representing over a dozen domains, spanning the natural sciences, computer science, policy analysis, and investigative journalism. These analysts were recruited from our extended personal and professional networks, and were known to work with high-dimensional data. These interviews were semi-structured, lasting in duration from one to four hours; some of these interviews were more akin to contextual inquiries, occurring at the analyst's workplace, while others were performed in our department or via teleconference.

We discussed the analysts' domain context, their data analysis goals, and their data; we also asked more specific questions about how they transformed their data and their use of dimensionality reduction and visualization techniques. We also collected artifacts from these analysts, including their published papers and theses, their unpublished manuscripts, screenshots of visualizations they had created, and in some cases, even their data.

Data analysis. We alternated between data collection and analysis, progressing from *initial* to *focused* coding of the data [10]. In this paper we concentrate our attention on the ten analysts who (a) specifically used dimensionality reduction algorithms in analyzing their high-dimensional data, and who (b) also visualized their dimensionally-reduced data.

To analyze data collected from these ten interviews, we use a recently proposed typology of abstract visualization tasks [7]. The typology distinguishes *why* tasks are undertaken at multiple levels of abstraction, *what inputs* and *outputs* a task may have, as well as *how* a task is supported by visual encoding and interaction techniques. Using this lens allowed us to better interpret our results from the standpoint of visualization design and evaluation, culminating in the task sequences presented in Section 4, where we use a `fixed-width font` to specifically highlight vocabulary from this typology. In Section 5, we revisit the typology and illustrate how it can summarize our five task sequences.

Finally, we enriched our analysis with further examples from the literature. We specifically sought papers that report on applications where dimensionality reduction and visualization were used in conjunction for analysis, and we consider these applications with respect to the task sequences we identified.

4. TASK SEQUENCES

We have identified five task sequences related to dimensionally-reduced data. In this section, we summarize each task sequence and illustrate the sequence with accompanying diagrams. Each is named after the terminal task appearing in the sequence. We also comment on how these task sequences arose in our interviews, and which visualization techniques were used to address these sequences. These task sequences are not exclusive: some analysts performed multiple task sequences in the course of their work. This descriptive survey of analysts' data, task sequences, and visualization is summarized in Table 1. The dataset sizes being investigated by these analysts ranged from dozens to over a million dimensions, and from hundreds to hundreds of thousands of items.

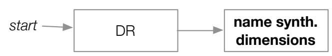
Dimensionality reduction. All the task sequences we characterized begin with dimensionality reduction (DR). In our context, we define DR as a means of dimensional synthesis: a set of m synthesized dimensions is `derived` from n original dimensions, where $m < n$. Dimensional synthesis techniques are commonly differentiated between *linear* and *non-linear* [18]. Linear techniques such as principal component analysis (PCA) [20] or classical multidimensional scaling (MDS) [46, 57] produce synthetic dimensions from linear projections of the original data. However, many datasets have an intrinsic structure that can only be revealed using non-linear techniques, such as Isomap [45], t-SNE [49], or Glimmer MDS [16]. Further distinction between linear and non-linear dimensional synthesis is outside of the scope of this paper, though we note that some techniques are more appropriate for verifying the existence of local cluster structure while others are more appropriate for identifying global intrinsic dimensions (or *manifolds*) [26]. In Table 1, we note who used linear and non-linear DR.

It is not our intent to catalog and differentiate the large body of DR techniques; we will concentrate our analysis on their output, asking `why` do analysts visualize these synthesized dimensions.

4.1 Dimension-Oriented Task Sequences

We describe two task sequences that specifically relate to *synthesized* dimensions as generated by dimensional synthesis DR techniques: *naming synthesized dimensions* and *mapping synthesized to original dimensions*.

4.1.1 *Name Synthesized Dimensions*



Given a set of synthesized dimensions, an analyst may want to `discover` what these dimensions mean, to `generate hypotheses` about the semantics of these synthesized dimensions. An analyst will `browse` the set of synthesized dimensions, and for each dimension of interest, she will `browse` items and their corresponding values; as a result, she may be able to `identify` the name of a synthesized dimension.

This task sequence was attempted by two of the analysts we interviewed (Analysts A and B in Table 1). Both worked in the field of human-computer interaction and attempted to identify the intrinsic dimensions related to usage data collected about online search behaviour and music listening behaviour, respectively.

A common approach, employed by both analysts, is to inspect data points plotted according to two synthesized dimensions in a 2D scatterplot, in which the analyst may be able to discern an interesting semantic relationship along the axes. In some cases, these scatterplots are augmented with text labels containing categorical information, such as item name, annotated adjacent to a subset of the plotted points [8, 27, 45] or available through interaction. Tenenbaum et al.'s paper describing the Isomap algorithm [45] contains a particularly compelling example (reproduced in Figure 2), in which each data point in a scatterplot corresponds to an image of a face; a random sample of these images are displayed directly in the scatterplot as thumbnails adjacent to their corresponding points. Given this display, it is possible to discern names for the three synthesized dimensions resulting from dimensional synthesis.

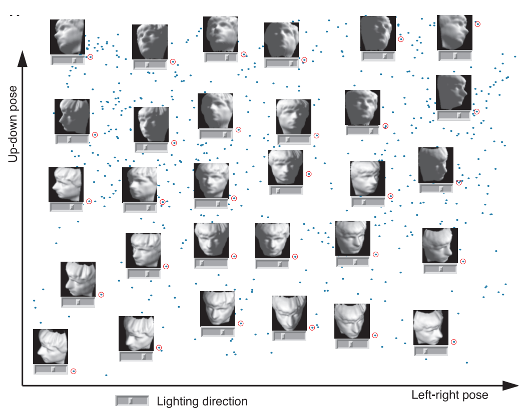
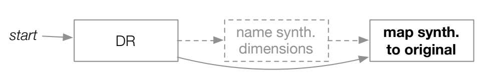


Figure 2: A figure from Tenenbaum et al. (2000) [45] (© 2000 AAAS), in which three synthesized dimensions have been identified: *up-down pose* along the y-axis, *left-right pose* along the x-axis, and *lighting direction* indicated below each image.

4.1.2 *Map Synthesized to Original Dimensions*



Regardless of whether an analyst is interested in naming synthesized dimensions, another possible task sequence involves mapping synthesized dimensions back to original dimensions. In the context of principal component analysis, this mapping is often referred to as the *loading* of the synthesized dimensions by the original dimensions [20]. Given a synthesized dimension, an analyst may want to `discover` this mapping. More specifically, the analyst may either `verify` a hypothesis that this mapping exists, or `generate` a new hypothesis about it. The analyst will `browse` items and their values along this synthesized dimension and `compare` these values to those along the set of original dimensions, looking for similarities and correlations. This mapping could allow analysts to `identify` groups of correlated original dimensions.

Four of the analysts we interviewed attempted to perform this sequence of tasks; two of these analysts had previously attempted to name some of their synthesized dimensions. Analyst A mapped her synthesized dimensions to a set of original dimensions in aggregated usage logs from an online music streaming service, while Analyst B attempted the same task sequence with aggregate search engine metrics but was unable to confidently map any of her synthesized dimensions to her original dimensions. Both used 2D scatterplots to carry out this task sequence. The other two analysts were explicitly interested in grouping original dimensions based on this mapping: a policy analyst (C) investigating survey data pertaining to recreational boating practices used 2D scatterplots to compare synthesized dimensions and original dimensions, while a bioinformatician (D) investigating protein regions used scatterplot matrices (SPLOMs), heatmaps, and density plots.

4.2 Cluster-Oriented Task Sequences

There exists another set of task sequences where the semantics of the synthesized dimensions are not a central interest; instead, analysts are interested in clusters of items that might be revealed in the dimensionally-reduced data. We characterize three task sequences: *verify clusters*, *name clusters*, and *match clusters and classes*.

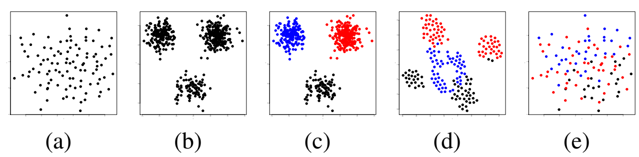
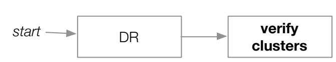


Figure 3: Example scatterplots of dimensionally-reduced data illustrating tasks related to item clusters: Verifying the existence of clusters, naming clusters, and matching clusters and classes. (a) no discernible clusters (b) three discernible clusters (c) a match between clusters and class labels (d) a partial match between clusters and class labels (e) no discernible class separation.

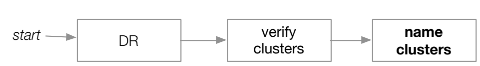
4.2.1 *Verify Clusters*



Analysts might seek to `verify` the general hypothesis that clusters of items will be revealed in the dimensionally-reduced data, or to `verify` hypotheses about specific conjectured clusters. In order to `discover` clusters, analysts must `locate` and `identify` item clusters in the low-dimensional representation of the data; in the example of Figure 3b, we can identify three clusters.

All ten of the analysts we spoke to were interested in verifying that clusters exist in their data. This task sequence is also captured by Buja and Swayne's discussion of visualizing data following multidimensional scaling [8]. The analysts we interviewed used a variety of visualization techniques when performing this task sequence, including 2D monochrome scatterplots, such as those depicted in Figures 3a-b, as well as 3D scatterplots, SPLOMs, dendrograms, heatmaps, and density plots.

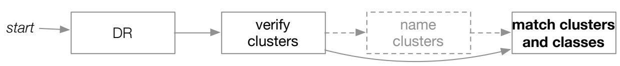
4.2.2 *Name Clusters*



Once the existence of clusters has been verified, such as in the example of Figure 3b, the next task is often one of `generating` hypotheses regarding the meaning of these clusters in the form of a name. In this `discover` task, an analyst will `browse` items within a cluster and attempt to `summarize` the cluster with a meaningful name. In some cases, this name is made explicit, as the analyst will `annotate` the cluster, thereby using the visualization to `produce` new information about their data.

Eight of the analysts who had previously *verified clusters* also attempted to *name clusters* in the course of their work, using the same visualization techniques mentioned above. For instance, Analyst F examined bibliometric data from a corpus of life sciences research literature, who attempted to identify and name clusters of related research concepts, such as *"cancer"* or *"RNA"*.

4.2.3 Match Clusters and Classes



The final task sequence we characterize is matching clusters with classes. The input to this *match* task is not only a set of item clusters, identified in the earlier *verify clusters* task, but also a set of categorical class labels. These classes might come directly with the data, be assigned using a clustering algorithm run by the analyst, or be the result of manual labeling. The analyst must `verify` a hypothesis that a cluster of items matches the class for those items. To `discover` a match, the analyst performs a `lookup` for the class and cluster membership of an item in order to `compare` them, resulting in a match (as in Figure 3c), otherwise referred to as a *true positive*, or a mismatch (as in Figures 3d-e), which could either be a *true negative* or a *false negative*. This task was examined in a recent paper on guidance for selecting appropriate visualization techniques for dimensionally-reduced data [40].

Naming the clusters is not a pre-requisite for this *match* task, though we did encounter four analysts who reported performing both tasks in succession (A, B, I, J); two other analysts performed this task without previously naming the clusters they identified (G, H). Typically, this task was performed using 2D scatterplots, wherein the points were coloured using the class labels; SPLOMs, interactive and non-interactive 3D scatterplots, and node-link graphs were also used. Note that the visual separability of colour-coded clusters differs perceptually from the separability of monochrome clusters, as summarized in a recent taxonomy of cluster separation factors [41]. These perceptual differences should be taken into account particularly when selecting experimental stimuli for use in controlled experiments.

A possible outcome of this task sequence is a partial match between classes and clusters: there may be more clusters than classes, or vice versa. In cases where there are more clusters than class labels, illustrated in Figure 3d, this outcome suggests that the class labels may not capture a finer-grained cluster structure in the data, as was the case for the investigative journalist we interviewed (J). In cases where there are more classes than clusters, illustrated in Figure 3e, this result may either be a *true negative*, in which perfect class separation is not possible, or a *false positive* [40]. If this mismatch is suspected to be a *false negative*, Sedlmair et al. recommend selecting other dimensions to visualize, using other visualization techniques such as a SPLOM, or revisiting the choice of dimensionality reduction technique.

5. A TASK TYPOLOGY REVISITED

The analysts we interviewed hailed from very different domains, each using a different terminology to describe their work processes. For instance, we needed a way to compare how *diagnosing cancer patients based on their genomic data* (Analyst H) was like *classifying types of human motion through the use of sensors attached to the body* (Analyst G). We required an abstract vocabulary for describing and comparing the work processes of these analysts.

For this reason, we use a recent typology of abstract visualization tasks [7] that provides a domain-agnostic vocabulary and framework for describing visualization tasks in terms of *why*, *what*, and *how*. By describing a task in this manner, we can link outputs and inputs to describe sequences of interdependent tasks, which Norman would refer to as *activities* [35]. This typology has already been used to characterize the tasks of journalists [6] and bioinformaticians [30]. We use it here to describe task sequences relating to visualizing dimensionally-reduced data across multiple domains.

Our analysis concentrated on the *why* and *what* aspects of the tasks pertaining to dimensionally-reduced data, as summarized in Figure 4. We chose not to be prescriptive about *how* these task sequences should best be supported by visualization techniques; instead, we described the variety of techniques used by the analysts we interviewed for each task sequence.

Table 1: Top: A summary of task sequences performed by the ten analysts we interviewed, along with the visualization techniques(s) used to perform these tasks sequences. Bottom: examples of task sequences in papers discussing DR and visualization.

Analyst or Paper		Data			DR		Task Sequence					Visualization Techniques						
ID	**Domain**	**Description**	**# Dims**	**# Items**	**Linear**	**Non-Linear**	*Name Dims*	*Map Dims*	*Verify Clusters*	*Name Clusters*	*Match Clusters*	**2D SPs**	**3D SPs**	**SPLOMs**	**Scree plots**	**Graph / Tree**	**Correl. matr.**	**Heatmaps**
A	Human-computer interaction	usage logs from online music service	48	310	✓		✓	✓	✓	✓	✓	✓		✓				
B	Human-computer interaction	aggregated search engine metrics	12–31	1,463	✓	✓	✓	✓	✓	✓	✓	✓						
C	Policy analysis	recreational boating survey data	39	543	✓	✓		✓	✓	✓		✓	✓	✓		✓		
D	Bioinformatics	protein region data	160	10–100K	✓	✓		✓	✓	✓				✓				✓
E	Computational chemistry	polymer molecule feature vectors	1K	10K	✓	✓			✓	✓		✓					✓	✓
F	Social network analysis	bibliometric co-occurrence matrix	20K	20K	✓	✓			✓	✓		✓		✓	✓		✓	
G	Human computer interaction	human motions from multiple sensors	1,170	9,120	✓				✓		✓	✓	✓					
H	Bioinformatics	genomic, clinical data from patients	1.4M	600	✓	✓			✓		✓	✓	✓					
I	Bioinformatics	distance matrix of genome sequences	100K	100K	✓	✓			✓	✓	✓	✓	✓	✓		✓		
J	Investigative journalism	distance matrix of text documents	10K	10K		✓			✓	✓	✓	✓				✓		
Ref.	**Citation**																	
[8]	Buja & Swayne (2002)	distance matrix of Morse codes	36	36		✓	✓		✓	✓		✓						
[27]	Matusik et al. (2003)	BRDF reflectance model	4.36M	104	✓	✓	✓					✓			✓			
[37]	Reveret et al. (2005)	quadruped skeleton models	348–406	9	✓		✓								✓			
[45]	Tenenbaum et al. (2000)	64 x 64 px images	4,096	698–1K		✓	✓					✓			✓			

The analysts we interviewed were all interested in `discovery`, which involves the `generation` and `verification` of hypotheses. Figures 4b-f show which tasks relate to `hypothesis generation` and which relate to `hypothesis verification`. The graphical depiction also shows which task can be associated with pure `consumption` of information and which task can additionally lead to the `production` of new information. When consuming information, an analyst will `search` for targets within a visualization. Whether the location and identity of these targets is known a priori will determine the type of `search`. In tasks related to visualizing dimensionally-reduced data, we found that `search` strategies used by analysts were either `browse`, `locate`, or `lookup`, as indicated in Figures 4b-f. Once targets are found, an analyst will execute some form of query: they might `identify` a single target, such as an item cluster, `compare` multiple targets, such as values along a synthesized dimension to values along an original dimension, or `summarize` all the targets, such as when *naming a cluster*.

Dependencies. The task sequences described in Section 4 contain dependencies. For example, in order to *match clusters and classes*, an analyst must first *verify* that clusters exist. Each of the sequences also depend on the output of dimensionality reduction techniques, the `derived` synthetic dimensions. The application of dimensionality reduction to a set of original dimensions is itself a task, as shown in Figure 4a. However, unlike the other tasks described above, it is about neither hypothesis generation nor verification, but rather about `producing` new information intended to support subsequent tasks.

While the distinctions between these tasks and task sequences may seem obvious in hindsight, we initially struggled to find a vocabulary and framework that would allow us to distinguish between these task sequences and their interdependencies. The task typology [7] allows us to describe these task sequences explicitly, whereas they were implicit in previous work combining dimensionality reduction and visualization.

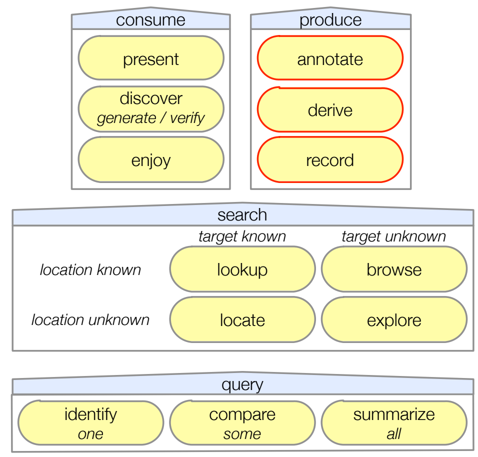


Figure 5: The *why* part of the abstract task typology [7], with the refinement (emphasized in red) that the actions of *annotate*, *record*, and *derive* are forms of *produce* [34].

Extended typology. Figure 5 reproduces the *why* part of an extended task typology [34]. The changes relevant to our analysis in this paper pertain to three actions: an analyst may `annotate` information, `derive` new information from existing, or `record` their use of a visualization so as to provide analytical provenance or to facilitate subsequent presentations of the visualization. The

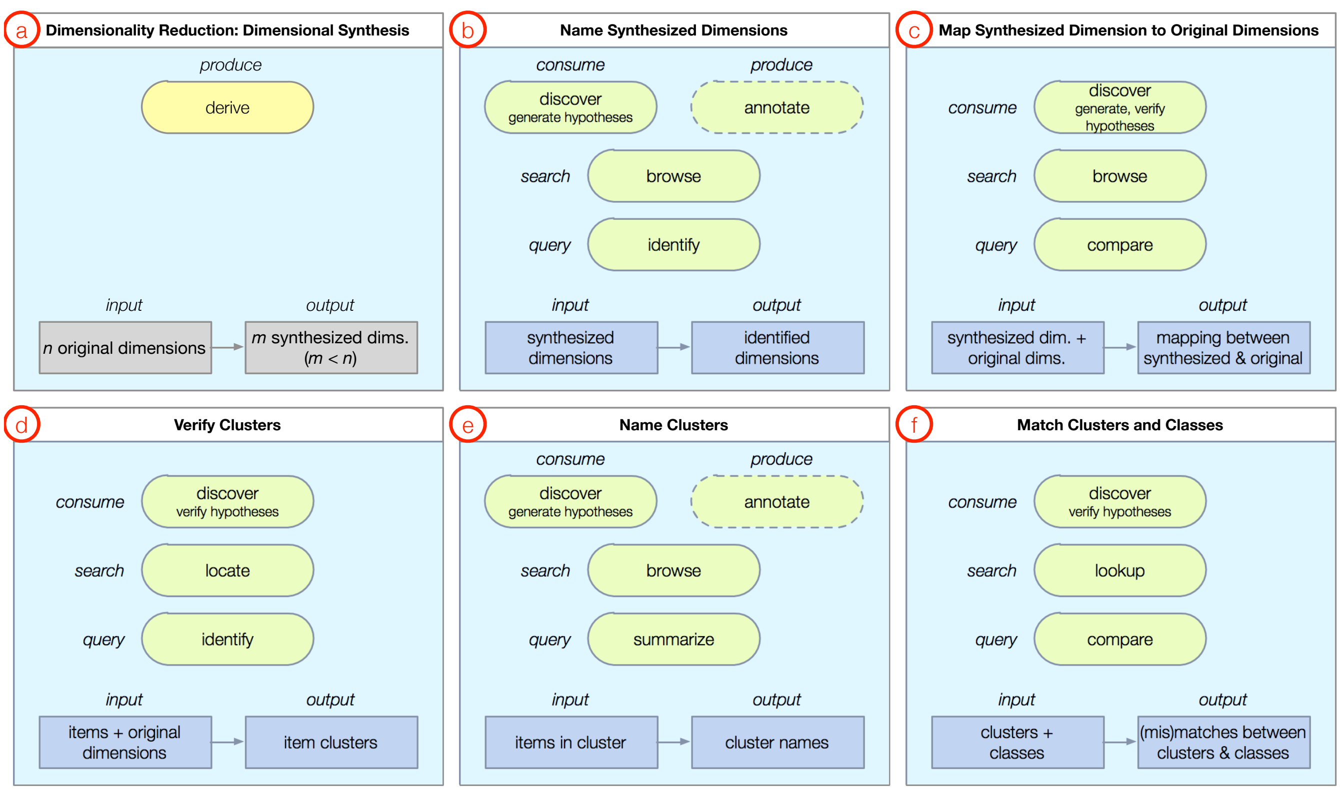


Figure 4: Six tasks related to dimensionally-reduced data, characterized using an abstract task typology [7, 34], which describes *why* a task is perform at multiple levels of abstraction (yellow) and *what inputs* and *outputs* a task has (grey). These tasks are combined to form the task sequences described in Section 4.

terms annotate, derive, and record were previously attributed to families of interaction techniques in the how part of the typology [7]; the extended typology classifies them as ends rather than means and thus situates them as forms of produce. Both versions of the typology distinguish whether a person will use a visualization either to consume or produce information. The remaining aspects of the typology describing lower levels of abstraction are unchanged.

6. DISCUSSION

We discuss the utility of our characterization of task sequences with regard to several visualization evaluation scenarios, the limitations of our current findings, and our planned future work.

6.1 Implications for Evaluation

Task characterization and evaluation are closely linked. An understanding of visualization tasks and activities informs how an evaluation is conducted, from the justification of experimental procedures to the collection and analysis of field observations.

Our current work adds to previous task characterizations proposed in the visualization evaluation literature [14, 25, 48]. As evaluation takes on many forms, we frame our discussion around four of Lam et al.'s scenarios for empirical studies [23].

Understanding work practices. Work practice evaluations provide a richer understanding of the perspective of people who might benefit from visualization, reflecting real work practices and activities. While their immense importance has been outlined several times [5, 29, 33], only a few dedicated examples exist in the visualization literature [21, 22, 47].

More commonly, however, such work practice evaluations occur in design studies, an increasingly popular form of problem-driven visualization research [39]. In particular, a design study's early *discover stage* [39] involves the analysis of work practices within a very specific usage context in a particular domain. These concrete work practices are then translated into abstract visualization tasks and design requirements.

Our current work goes beyond task characterization in design studies by conducting interviews with analysts across different application domains. We then abstracted our findings into more general task sequences, or activities [35], across these domains. In doing so, we intend to support researchers when conducting and analyzing future *work practice* evaluations, specifically when dimension reduction techniques are to be employed. We encourage practitioners to adopt our characterization of task sequences into a lexicon for coding observations of work practices and for translating domain-specific descriptions of these practices. We believe that using our task sequences will make the analysis process more efficient and, furthermore, will allow for better transferability between design studies from different application domains [39].

Evaluating user performance. Our characterization of task sequences can inform the design of experimental procedures and participant instructions in controlled laboratory studies, where the aim might be to quantitatively assess human performance on a newly proposed visualization technique. Many previous characterizations of tasks have informed experimental design, such as the adoption of Zhou and Feiner's task characterization [58] in a laboratory evaluation of an information retrieval tool [31]. We expect that our characterization of task sequences will play a similar role in the evaluation of techniques or tools that visualize dimensionally-reduced data. For instance, an experiment might compare multiple visualization techniques for *verifying clusters* and subsequently *matching clusters and classes*, where performance might be measured in terms of speed and accuracy.

Munzner refers to such studies as a form of downstream validation, in which a design has been implemented for its investigation in a study [33]. In contrast, upstream validation in this case refers to the justification of visual encoding and interaction design choices before its implementation. We deem our task sequences to be similarly helpful for such upstream evaluations. Researchers presenting new visualization or interaction techniques can refer to our task sequences to concisely state assumptions about which abstract tasks are supported, rather than leaving this description implicit in a way that places a burden on a potential adopter of the technique.

Evaluating user experience. In either lab or field settings, a researcher can evaluate the user experience of a tool or technique by dictating the tasks without specifying *how* to execute them, asking study participants to verbalize their actions while they attempt to execute a sequence of tasks. Such a protocol might allow the researcher to understand if features of the tool are learnable, useful, or in need of further usability improvements. Questionnaires and interview questions relating to user experience could also be framed around our characterization of task sequences.

We note that expertise has many facets; the distinction between novices and experts is a particularly nuanced question for studies considering dimensionality reduction. Several of the high-dimensional data analysts that we interviewed might be described as *middle-ground users* [15]: they had significant domain expertise but only partial understanding of the available dimensionality reduction tools and the mathematics underlying these techniques. This characteristic of users is important to keep in mind when recruiting participants for evaluations of user performance or user experience, as some evidence exists that participants with an understanding of dimensionality reduction will interpret visualizations differently than those who do not [26].

Evaluating visual data analysis and reasoning. While a researcher must dictate the tasks in a controlled laboratory experiment, another scenario is the observation of tasks in an open-ended qualitative evaluation of a visualization tool or technique. Here, the researcher must recognize when these task sequences appear in naturalistic settings, in order to better understand how visual data analysis and reasoning are supported following the introduction of a new visualization system. This form of evaluation is typical in design studies [39, 43], particularly after a tool is deployed.

As with evaluations of work practices, our characterization of task sequences could become part of a lexicon for coding observed behaviour after a tool is deployed. In cases where direct observation of tool use is not possible, our characterization of task sequences might be used to analyze interaction log files, or used as a basis for diary or interview questions, suggesting a consistent vocabulary for participant responses. Precedents for the use of task characterization in evaluation of deployed tools include the adoption of Yi et al.'s characterization [56] in a longitudinal field study of a social network analysis tool [36], or the use of Brehmer and Munzner's task typology [7] to evaluate why and how journalists used a tool for analyzing large document collections [6].

Finally, if we consider the task sequences *name synthesized dimensions* and *name clusters* in particular, one conceivable evaluation of visual data analysis and reasoning would involve collecting participant annotations and explanations of synthesized dimensions or clusters in visualizations of dimensionally-reduced data. Such a study might adopt a protocol similar to one used by Willett et al. to elicit participant annotations and explanations of time-series visualizations in an application deployed online [52]. This evaluation could help to identify the features of a visualization tool that facilitate or inhibit visual data analysis and reasoning.

6.2 Limitations and Future Work

Our interview findings are certainly not exhaustive, and despite conducting interviews with twenty-four analysts, only ten of these analysts contributed to our characterization of task sequences. This selection was based on our goal of studying task sequences relating to visualizing data reduced with *dimensional synthesis* techniques. There are many other interesting areas of high-dimensional data analysis that we did not address. Specifically, we found that many of our excluded interviewees used *dimensional filtering* techniques, in which a subset of the original dimensions are retained [19, 55]. Alternatively, other analysts applied dimensionality reduction to their data without visually analyzing it. In these cases, dimensionality reduction was used to reduce the data for algorithmic input, such as for classification and other machine learning applications.

We consider our findings to be existence proofs of the task sequences as performed by analysts as part of their ongoing work. We do not make claims about the prevalence of these task sequences in high-dimensional data analysis. Neither we do not make claims about completeness: our characterization of task sequences might be incomplete due to sampling or observer bias.

Future work. We intend to use our characterization of task sequences in future tool development, particularly in the design and evaluation of *workflows*: software instantiations of features to support these sequences. We plan to conduct further analysis of our collected data, characterizing the visualization techniques and tools used by the analysts we interviewed to perform these task sequences, situating these techniques among those proposed in the literature, and considering further implications for design. Finally, we hope to expand upon this set of task sequences to characterize additional high-dimensional data tasks, including those relating to data resulting from dimensional filtering.

7. CONCLUSION

We presented a characterization of five task sequences related to visualizing dimensionally-reduced data. Our abstract characterization of these task sequences fills a gap between the large body of technique-driven literature and analysts' domain problems in this area. We encourage other researchers to consider these task abstractions in the evaluation of existing work practices, in the *discover* phase of future design studies involving high-dimensional data and dimensionality reduction, in the design of controlled experiments, and in field evaluations of deployed systems.

Acknowledgments

We thank the data analysts who participated in this study for their time and energy. We thank NSERC for funding this project. We thank S. Bergner, J. Dawson, J. Ferstay, M. Meyer, T. Möller, T. Torsney-Weir, M. Tory, H. Younesy for assisting with interviews, ongoing discussions, and/or feedback on paper drafts.

References

[1] G. Albuquerque, M. Eisemann, D. J. Lehmann, H. Theisel, and M. Magnor. Improving the visual analysis of high-dimensional datasets using quality measures. In *Proc. IEEE VAST*, pages 19–26, 2010.

[2] R. Amar, J. Eagan, and J. T. Stasko. Low-level components of analytic activity in information visualization. In *Proc. IEEE InfoVis*, pages 111–117, 2005.

[3] R. Amar and J. T. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Proc. IEEE InfoVis*, pages 143–150, 2004.

[4] E. Bertini, A. Tatu, and D. A. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE TVCG (Proc. InfoVis)*, 17(12):2203–2212, 2011.

[5] M. Brehmer, S. Carpendale, B. Lee, and M. Tory. Pre-design empiricism for information visualization: Scenarios, methods, and challenges. In *Proc. ACM BELIV Workshop (to appear)*, 2014.

[6] M. Brehmer, S. Ingram, J. Stray, and T. Munzner. Overview: The design, adoption, and analysis of a visual document mining tool for investigative journalists. *IEEE TVCG (Proc. InfoVis, to appear)*, 2014.

[7] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE TVCG (Proc. InfoVis)*, 19(12):2376–2385, 2013.

[8] A. Buja and D. Swayne. Visualization methodology for multidimensional scaling. *J. Classification*, 19:7–43, 2002.

[9] S. Carpendale. Evaluating information visualizations. *Information Visualization: Human-Centered Issues and Perspectives (LNCS)*, 4950:19–45, 2008.

[10] K. Charmaz. *Constructing Grounded Theory. A Practical Guide Through Qualitative Analysis*. Sage, 2006.

[11] P. Cunningham. Dimension reduction. In *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval*, pages 91–112. Springer, 2008.

[12] S. France and J. Carroll. Two-way multidimensional scaling: A review. *IEEE T. Systems, Man, and Cybernetics*, 41(5):644 –661, 2011.

[13] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *J. Machine Learning Research*, 3:1157–1182, 2003.

[14] N. Henry and J. D. Fekete. Evaluating visual table data understanding. In *Proc. ACM BELIV Workshop*, 2006.

[15] S. Ingram, T. Munzner, V. Irvine, M. Tory, S. Bergner, and T. Möller. DimStiller: Workflows for dimensional analysis and reduction. In *Proc. IEEE VAST*, pages 3–10, 2010.

[16] S. Ingram, T. Munzner, and M. Olano. Glimmer: Multilevel MDS on the GPU. *IEEE TVCG*, 15(2):249–261, 2009.

[17] P. Isenberg, T. Zuk, C. Collins, and S. Carpendale. Grounded evaluation of information visualizations. In *Proc. ACM BELIV Workshop*. ACM, 2008.

[18] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE T. Pattern Analysis and Machine Intelligence*, 22(1):4–37, 2000.

[19] S. Johansson and J. Johansson. Interactive dimensionality reduction through user-defined combinations of quality metrics. *IEEE TVCG (Proc. InfoVis)*, 15(6):993–1000, 2009.

[20] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2nd edition, 2002.

[21] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. Enterprise data analysis and visualization: An interview study. *IEEE TVCG (Proc. VAST)*, 18(12):2917–2926, 2012.

[22] Y.-A. Kang and J. T. Stasko. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *Proc. IEEE VAST*, pages 19–28, 2011.

[23] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information

visualization: Seven scenarios. *IEEE TVCG*, 18(9):1520–1536, 2012.

[24] T. Lammarsch, A. Rind, W. Aigner, and S. Miksch. Developing an extended task framework for exploratory data analysis along the structure of time. In *Proc. EuroVA Workshop*, 2012.

[25] B. Lee, C. Plaisant, C. S. Parr, J. D. Fekete, and N. Henry. Task taxonomy for graph visualization. In *Proc. ACM BELIV Workshop*, 2006.

[26] J. M. Lewis, L. van der Maaten, and V. R. de Sa. A behavioral investigation of dimensionality reduction. In *Proc. Cognitive Science Society*, pages 671–676, 2012.

[27] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM T. Graphics*, 22(3):759–769, 2003.

[28] J. E. McGrath. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction: Towards the Year 2000*, pages 152–169. Morgan Kaufmann, 2nd edition, 1995.

[29] M. Meyer, M. Sedlmair, P. S. Quinan, and T. Munzner. The nested blocks and guidelines model. *Information Visualization*, In press, 2014. doi.org/10.1177/1473871613510429.

[30] B. Mirel and C. Görg. Scientists' sense making when hypothesizing about disease mechanisms from expression data and their needs for visualization support. *BMC Bioinformatics*, 15(1):117–128, 2014.

[31] E. Morse, M. Lewis, and K. A. Olsen. Evaluating visualizations: using a taxonomic guide. *Intl. J. Human-Computer Studies*, 53(5):637–662, 2000.

[32] P. M. Mullins and S. Treu. A task-based cognitive model for user-network interaction: Defining a task taxonomy to guide the interface designer. *Interacting with Computers*, 5(2):139–166, 1993.

[33] T. Munzner. A nested model for visualization design and validation. *IEEE TVCG (Proc. InfoVis)*, 15(6):921–928, 2009.

[34] T. Munzner. *Visualization Analysis and Design*. A K Peters Visualization Series, CRC Press, 2014.

[35] D. A. Norman. Human-centered design considered harmful. *interactions*, 12(4):14, 2005.

[36] A. Perer and B. Shneiderman. Integrating statistics and visualization for exploratory power: From long-term case studies to design guidelines. *IEEE T. CG&A*, 29(3):39–51, 2009.

[37] L. Reveret, L. Favreau, C. Depraz, and M.-P. Cani. Morphable model of quadrupeds skeletons for animating 3D animals. In *Proc. Eurographics Symp. Computer Animation*, pages 135–142. ACM, 2005.

[38] H. J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE TVCG (Proc. InfoVis)*, 19(12):2366–2375, 2013.

[39] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE TVCG (Proc. InfoVis)*, 18(12):2431–2440, 2012.

[40] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE TVCG (Proc. InfoVis)*, 19(12):2634–2643, 2013.

[41] M. Sedlmair, A. Tatu, M. Tory, and T. Munzner. A taxonomy of visual cluster separation factors. *Computer Graphics Forum (Proc. EuroVis)*, 31(3):1335–1344, 2012.

[42] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symp. Visual Languages*, pages 336–343, 1996.

[43] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proc. BELIV Workshop*, 2006.

[44] A. Tatu, G. Albuquerque, M. Eisemann, P. Bak, H. Theisel, M. Magnor, and D. A. Keim. Automated analytical methods to support visual exploration of high-dimensional data. *IEEE TVCG*, 17(5):584–597, 2010.

[45] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[46] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17:401–419, 1952.

[47] M. Tory and S. Staub-French. Qualitative analysis of visualization: A building design field study. In *Proc. ACM BELIV Workshop*, pages 7:1–7:8. ACM, 2008.

[48] E. R. A. Valiati, M. S. Pimenta, and C. M. D. S. Freitas. A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proc. ACM BELIV Workshop*, 2006.

[49] L. van der Maaten and G. Hinton. Visualizing data using t-SNE. *J. Machine Learning Research*, 9:2579–2605, 2008.

[50] L. van der Maaten, E. Postma, and H. van den Herik. Dimensionality reduction: A comparative review. Technical report, Tilburg University, TiCC-TR 2009-005, 2009.

[51] K. J. Vicente. *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. CRC Press, 1999.

[52] W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proc. CHI*, pages 227–236, 2012.

[53] M. Williams and T. Munzner. Steerable, progressive multidimensional scaling. In *Proc. IEEE InfoVis*, pages 57–64, 2004.

[54] I. Witten, E. Frank, and M. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2011.

[55] 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 InfoVis*, pages 105–112, 2003.

[56] J. S. Yi, Y. A. Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *IEEE TVCG (Proc. InfoVis)*, 13(6):1224–1231, 2007.

[57] G. Young and A. Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 3:19–22, 1938.

[58] M. X. Zhou and S. K. Feiner. Visual task characterization for automated visual discourse synthesis. In *Proc. CHI*, pages 392–399, 1998.