

# Constructing and Evaluating Visualisation Task Classifications: Process and Considerations

N. Kerracher<sup>1</sup> and J. Kennedy<sup>1</sup>

<sup>1</sup>Edinburgh Napier University, United Kingdom

---

## Abstract

*Categorising tasks is a common pursuit in the visualisation research community, with a wide variety of taxonomies, typologies, design spaces, and frameworks having been developed over the last three decades. While these classifications are universally purported to be useful in both the design and evaluation processes and in guiding future research, remarkably little attention has been paid to how these frameworks have—and can be—constructed and evaluated. In this paper we review the task classification literature and report on current practices in construction and evaluation. We consider the stages of task classification construction and identify the associated threats to validity arising at each stage and in response to the different methods employed. We provide guidance on suitable validation approaches in order to mitigate these threats. We also consider the appropriateness of evaluation strategies according to the different aspects of the classification which they evaluate. In so doing, we seek to provide guidance for developers of classifications in determining appropriate construction and evaluation strategies when developing a classification, and also for those selecting between competing classifications for use in the design and evaluation processes.*

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology, Theory and methods

---

## 1. Introduction

Categorising tasks is an increasingly common pursuit in visualisation research, with a variety of taxonomies, typologies, design spaces, frameworks, and models having been developed over the last three decades. These classifications have been used to pre-empt and make sense of both peoples' aims and intentions, and the ever increasing body of work relating to visualisation tools, techniques, and systems. In particular, the usefulness of these classifications in the design and evaluation processes is widely accepted. However, while evaluation practices have also become a topic of increasing interest in the visualisation community e.g. [LBI\*11, IIC\*13], very little has been written regarding the evaluation of formal models such as task classifications.

Given the increasing number of task classifications being developed, recent work has been devoted to unifying the wide range of extant classifications (e.g. [BM13, SNHS13]) and teasing apart the terminology surrounding the use of the word 'task' ([RAW\*15]), a term agreed to be "deeply overloaded" [Mun09]. However, very little attention has so far been given to the construction and evaluation practices involved in developing task classifications. While we would expect a publication demonstrating a new visualisation technique or system to include some form of evaluation with respect

to its utility, performance, and limitations, this does not appear to be the case when newly developed classifications are reported. Further, while much work reflects on and provides guidance relating to appropriate design and evaluation practices when developing visualisation systems and techniques (e.g. [LBI\*11, IIC\*13, Mun09, MSM12, SMM12]), analogous guidance for developing task classifications does not exist. This is surprising, given that measuring the effectiveness of classifications has been recognised as a difficult problem [YKSJ07], and the benefits of evaluating classifications are comparable to those of evaluating systems, including identifying areas for improvement resulting in better classifications, convincing potential adopters of the validity and utility of the approach (particularly important for more complex classifications which may require significant effort to adopt), and helping adopters select between competing classifications.

This paper therefore seeks to act as a first step in filling this gap, by offering a review of the literature relating to construction and evaluation practices when developing task classifications. We begin with an outline of the method used in our literature review (Section 2) and a discussion of the range of terminology surrounding task classifications (Section 3). In Section 4 we consider the intended uses of task classifications in visualisation. In Section 5 we look at

the main stages of task classification construction and identify the associated threats to validity arising at each stage and in response to the different methods employed, along with strategies for their mitigation. In Section 6 we explore existing evaluation practices by distinguishing *what* is being evaluated (construction method, properties of the classification, intended use), and then consider evaluation methods in relation to these aspects.

## 2. Method

We used Rind et al.'s [RAW\*15] list of 31 abstract task categorisations as the basis for our review of the literature. Three of these publications [Mac95, TC05, SM13] were not available to us. We considered [Mun14] and [BSIM14] together, as the discussion relating to the construction and evaluation of this classification is presented in [BSIM14]. The literature was reviewed with regard to (1) terminology employed, (2) the intended uses of classifications, (3) the construction methods employed, (4) the aspects of classifications that were evaluated, and (5) the methods employed in evaluating these aspects. A table summarising the construction approaches and evaluation methods employed by each classification is included in the supplemental material. In our discussions we also draw on literature relating to the visualisation design process and evaluation practices, where appropriate.

## 3. Classification terminology

We use *classification* as an umbrella term to describe a construct in which items are collected together and grouped in some meaningful way. However, many terms are used in the visualisation literature to describe task classifications.

*Lists* of tasks are often offered in the visualisation literature. When referred to in this way, these are usually intended to be non-exhaustive illustrations of exemplar or common tasks, which may or may not be grouped into categories (e.g. [AS05]). The terms *typology* and *taxonomy* tend to indicate a more rigorous process of categorisation has been followed, and are often used interchangeably in the literature. However, Bailey [Bai94] distinguishes them on the grounds that a taxonomy is empirical (a set of extant entities are grouped according to their similarity to produce a classification), while a typology is conceptual (a classification is constructed a priori using multiple conceptual dimensions; the resulting categories represent concepts rather than empirical cases). Such classifications are used in the visualisation literature to describe and bring order to the range of tasks and visual techniques. Bailey [Bai94] also describes the case where the independent dimensions of a classification are combined to form a *property space*. In the visualisation literature, this idea—often termed *design space*—is becoming increasingly common [SNHS13]. The intended use of a design space is not simply as a means to classify existing items, but to map ‘*the space of the possible*’ [SNHS13], revealing potential items which may not yet exist. As such, it can be used as a generative method to specify novel visualisation techniques (e.g. [SHS11, JE12, KKC14]) or previously unconsidered tasks (e.g. [SNHS13, KKC15]). We discuss these uses further in Section 4.3.

Task classifications are often used to *characterise systems* ac-

ording to the tasks they support in order to help make comparisons between systems when selecting appropriate visual tools (e.g. [LPP\*06, APS14]). An extension of this idea are *task-technique mappings* or *catalogues* (e.g. [WL90, KKCG15]) which take a task classification and map to each category the visual techniques for their support. These are often intended to provide a useful inventory of appropriate techniques for use during the design process, and like design spaces, can help point to opportunities for research by identifying as yet unsupported tasks which could benefit from the development of appropriate visual techniques.

## 4. The Role of Classifications in Visualisation

Classifications of tasks and techniques play a number of useful roles in visualisation, including providing an overview of work in the field, and revealing potential research opportunities. Perhaps the most widely cited use of classifications in the visualisation literature is their employment in the design and evaluation processes. However, exactly how classifications can be utilised in these processes is not well documented. The following seeks to clarify the multiple ways in which classifications can be of use in visualisation.

### 4.1. Role in communication

Perhaps the most fundamental benefit of classifications is that they provide a common vocabulary to describe both analytical tasks and the means by which they can be achieved (objectives and actions, respectively, to use the language of Rind et al. [RAW\*15]). Having an agreed upon language allows researchers to communicate more effectively, reducing misunderstanding [SSS\*14, SNHS13]. As will be discussed further below, using classifications which present tasks in a consistent and abstract manner to describe the domain specific tasks of users, and the functionality of systems and tools from across domains and application scenarios, offers many benefits in the design and evaluation processes. Describing tasks in an abstract rather than domain specific manner also allows us to generalise when situating and communicating the results of research. For example, Lee et al. [LPP\*06] suggest the use of their classification in helping evaluators generalise results collected in a series of controlled experiments. Sedlmair et al. [SMM12] note the need to present clear abstractions of tasks when reporting on design studies, so that the bare minimum of domain knowledge is required to understand them. Similarly, Rind et al. note the use of abstract tasks when setting context in case studies. Using the recognised terminology of task classifications can be particularly beneficial in these circumstances.

### 4.2. Making sense of what's out there

Classifications can help us make sense of what already exists in our research area. They provide a useful means for bringing order to the range of user intentions and existing visualisation systems, tools, and techniques, often from across a wide range of domains. They can act as an entry point for researchers new to an area, in a manner similar to review and survey papers. All of the papers in this review can be said to be of use for this purpose.

### 4.3. Identifying what's not out there

Schulz et al. [SNHS13] describe how design spaces can be used to identify “the space of the possible”. A design space maps out the “universe of all possible design choices” [SHS11] and can be constructed by combining the independent dimensions of a taxonomy to produce all possible variants. Design spaces have been used to map the space of the possible for both visual techniques e.g. [SHS11, JE12, KKC14] and tasks e.g. [APS14, SNHS13, KKC15]. By mapping existing techniques to the possibilities identified in a technique design space, as-yet unexplored techniques may emerge, which could prove interesting opportunities for further research e.g. [SHS11, KKC14]. Further, visual techniques can be mapped to a task design space according to the categories of tasks which they support, to establish which tasks are currently addressed by existing techniques, and reveal areas which could benefit from further research e.g. [APS14, KKCG15]. In this way, mapping techniques to the ‘space of the possible’ can help guide the focus of future system development and encourage the pursuit of novel research questions [APS14]. Such a mapping could also help signpost opportunities for evaluation, as it identifies the case where multiple techniques support the same task: these techniques are potential candidates for use in controlled experiments to establish which techniques are the most effective in their support.

### 4.4. Use in the Design Process

Several authors note the role which task classifications can play in systematising the design process. Both Amar and Stasko [AS05] and Sedig and Parsons [SP13] assert that classifications can act as a systematic basis for thinking about the design process, while the use of classifications as a “checklist” of items to consider when designing visualisation tools is proposed by [AS05, HS12, LRAM12].

Classifications of tasks and techniques can be gainfully employed at multiple stages of the design process. One well known model of the visualisation design process is Munzner’s nested model [Mun09]. The following considers the use of classifications in the first three stages of this model.

#### 4.4.1. Task Understanding

Understanding which analytical tasks an analyst may wish to carry out is a non-trivial problem and a key component at the domain problem characterisation stage. In a typical design scenario, van Wijk [vW06] notes that visualisation researchers must spend time and effort bridging ‘the knowledge gap’ between themselves and the domain expert, in order to effectively understand the problem in what is potentially an unfamiliar domain with its own terminology. Generative methods [MMAM14] for eliciting possible tasks of interest can be roughly grouped into three strategies: deriving tasks in an analytical fashion, for example, by reviewing relevant literature; talking to domain experts, for example, through interviews or brainstorming sessions in focus groups; observing people at work, either using existing visualisation tools or the methods they currently employ.

As discussed further in Section 5.1.1, each of these strategies has limitations when eliciting tasks. One use for task classifications is in supporting the generative phase of task understanding in order

to mitigate some of the problems in the strategies outlined above. For example, they may act as a useful means upon which to base discussions with domain experts. By setting out the range of potential tasks of interest, they may overcome known problems associated with simply asking people to introspect. They may also help to keep the discussions focussed on tasks; one pitfall identified by Sedlmair et al. [SMM12] at this stage of the design process is allowing experts to focus on possible visualisation solutions, rather than explaining their problems. Potentially they may act as a useful bridge in the knowledge gap, presenting a collection of domain independent tasks from which concrete domain tasks can be derived. Finally, using a task classification in this way may help with task specification at a consistent level of granularity and abstraction.

Note that there are relatively few examples of task classifications being used at this stage in the design process. One example is Ahn et al. [APS14], who demonstrate how their task design space could be used as a generative method to help in the discovery of new tasks at the task understanding stage i.e. those tasks that analysts had not thought of during a requirements gathering process.

#### 4.4.2. Data/operation Abstraction Design

Once concrete, domain specific tasks have been captured, the data/operation abstraction stage requires that they be translated into the language of information visualisation. The resulting set of abstract tasks (operations) is used as the basis for selecting visual encodings at the encoding/interaction technique design stage. Task classifications can be utilised at this stage to describe domain tasks in appropriate abstract terms [RAW\*15], for example, Brehmer et al. propose using their classification as a “lexicon for coding” observed tasks. The process of abstraction reveals similarities between tasks that may initially appear to be rather different [Mun14]. This allows them to be meaningfully grouped together, thus categories of frequently occurring tasks can be identified. This can be useful when determining which tools to include when developing a system at the next stage of the design process.

#### 4.4.3. Encoding/interaction Technique Design

Visual technique classifications can be of assistance in revealing the range of potential design solutions at the encoding/interaction technique design stage. They can be divided into those that categorise the range of visual representations of the data, and those that deal with interaction [SNHS13]. They may classify techniques according to algorithms used [TM04], data structure e.g. [Shn96, LBAL09, HBO10], similarity of encodings e.g. [JE12, SHS11], or task support (either visual operations e.g. [YKSJ07, HS12], or user intention e.g. [WL90, MMC02]). However, where task support is not the basis on which the classification is made, a mapping between the technique categories and the tasks which they support is required for the classifications to be of assistance in directly helping designers to select appropriate tools for inclusion in their systems. Wehrend and Lewis [WL90], were among the first to propose a “problem-oriented approach” to tool classification, categorising scientific visualisation techniques according to the sub-problems (tasks) and objects supported. This results in a task-technique “catalogue”, which designers can use to look up potential visual solutions according to the problems for which they are trying to de-

sign. A particular advantage of the catalogue approach is that it provides a way to share visual solutions to similar problems across disparate application domains. Developing such task-technique mappings were also thought to be the first step towards automated system design [WL90, CR96].

Direct mappings between tasks and techniques for use in tool selection, however, may not always be possible or appropriate. Rind et al. [RAW\*15] note the use of guidelines translating between abstract objectives and abstract actions, citing Andrienko and Andrienko [AA06] who, given the intentionally generic nature of their task framework, derive a set of general principles which can be utilised when designing exploratory tools. Roth [Rot13] also notes the use of task classifications in the generation of design guidelines.

The importance of including process and provenance functionality in visual analysis systems, such as those described by [HS12], has recently been highlighted in the visualisation literature, and task classifications can play a role in identifying tasks for this purpose. Gotz and Zhou [GZ09] and von Landesberger et al. [vLFB\*14] develop task classifications with this purpose in mind. Rind et al. [RAW\*15] offer a more detailed discussion of the integration of tasks in visualisation artefacts.

#### 4.5. Use in Evaluation

An often-cited motivation for developing task classifications is their use in the evaluation process. They can be of use when selecting representative tasks for use in experiments, acting as a “check-list” covering the range of possible tasks for inclusion, in a manner similar to that of the design process (Section 4.4). Brehmer et al. [BSIM14] outline potential uses of their characterisation of task sequences in four of Lam et al.’s [LBI\*11] empirical evaluation scenarios: as a lexicon for coding observations when *understanding working practices* (as described in Section 4.4.2); to inform the design of experimental procedures when *evaluating user performance*; when specifying tasks for use when *evaluating user experience*, either as instructions in experiments or when constructing questionnaires and interview questions relating to user experience; and when coding observed behaviour when *evaluating visual data analysis and reasoning*.

The primary use of task classifications for evaluation purposes that we found in our review was their use in characterising systems in terms of task support. This allows evaluators to assess individual systems in terms of their capabilities and limitations e.g. [SSS\*14], or make comparisons across multiple systems e.g. [AS05, SNHS13, SSS\*14, AMA\*14, LPP\*06].

### 5. Classification Construction

There are a number of ways in which classifications can be constructed, although little reflection on the processes involved is to be found in the visualisation literature. Schulz et al. [SNHS13] consider the process of establishing recurring visualisation tasks and their description. They also discuss the consolidation of existing works. Building on this, Kerracher et al. [KKC15] identify three main steps in constructing a task classification: (1) generate the tasks, (2) collate and order them, (3) describe them. This

most closely describes a taxonomic approach to classification construction, where a set of existing items are gathered and grouped together based on their similarity. In contrast, what we refer to as *conceptual approaches* to classification construction—such as typologies and design spaces, as outlined in Section 3—begin with a set of important characteristics upon which tasks can be distinguished. In this case, rather than gathering a set of tasks, a set of conceptual dimensions are identified and used as the basis of classification construction. While these may need to be ordered in some way, the same process of rationalisation of tasks into categories required by step (2) of the taxonomic approach is not necessary. Some form of description of the resultant categories is still required, although for design spaces, the combination of choices along each dimension often serves to suitably define the category.

Different threats to validity arise from the different approaches that can be taken at each stage of classification construction, which consequently require different approaches to validation. We therefore reviewed the construction methods adopted by the task classifications surveyed, and comment on the potential threats to validity and appropriate validation approaches in each case. Our discussion here is largely based around the three steps in taxonomic construction. The discussion of each step is summarised in Tables 1, 2, and 3. We include a discussion specific to conceptual approaches towards the end of this section, which is summarised in Table 4.

#### 5.1. Task Generation

Task generation refers to the process of obtaining a set of tasks upon which a classification is based. Such a definition is most fitting when applied to taxonomies, where a set of items are collected and then organised. We expand this idea to include the process by which the dimensions of conceptual approaches are obtained. Schulz et al. [SNHS13] describe a number of common approaches to obtaining recurring visualisation tasks, including surveying individuals, observing visualisation users, and inferring from existing visualisation systems. In reviewing the literature, we found the most prevalent approach to task generation involved literature based methods (20 of 27 classifications): either involving extant classifications (16 of 27) or deriving tasks from the literature (9 of 27). Extant classifications may be unified [BM13, SNHS13, SSS\*14, vLFB\*14]; extended (e.g. to a specific data type, [AA06] by [LRAM12] and [KKC15], or for domain specific purposes, [YKSJ07] by [RWA\*13]); or—most commonly—used to derive, or cited as influencing, the task categories [LPP\*06, VPF06, YKSJ07, GZ09, Suo09, SNHS13, SP13, PPS14], with a small number making use of theoretical works from across a wider range of disciplines, such as HCI and cognitive science e.g. [GZ09, LS10, BM13, SP13]. Tasks can be derived from the literature by reviewing: existing systems and techniques for the tasks which they support [CR96, YKSJ07, GZ09, MMP09, SP13, APS14, AMA\*14]; problems addressed in the literature [WL90, YKSJ07]; tasks utilised in user evaluations [LPP\*06]; or studies examining users’ visual analytic behaviour [GZ09]. Far less common were the use of empirical methods to elicit tasks (8 of 27 classifications), including interviews with domain experts (either in a single domain [MMP09, Rot13], or across multiple domains [BSIM14]); surveys of those familiar with visualisation [AES05]; observational studies of people using visualisation sys-

Method	Threat	Mitigation
	<i>General threat -wrong or missing tasks, arising from:</i>	
Derive from literature (existing systems/techniques; existing problems; tasks in evaluations)	<ul style="list-style-type: none"> <li>- Understanding domain terminology</li> <li>- Requires similar problem having already been tackled</li> </ul>	Multi-strand approach to task gathering
Interviews with domain experts	<ul style="list-style-type: none"> <li>- Expert's availability</li> <li>- Skewing tasks towards a single domain</li> <li>- Difficulties with introspection/in articulating tasks</li> <li>- Maintaining focus on task discussions</li> </ul>	
Surveys of visualisation experts	Wrong people	
Observational strategies	<ul style="list-style-type: none"> <li>- Method of tackling problem must already exist</li> <li>- Lack of access to internal mental processes</li> </ul>	Downstream evaluation
System reviews	Method of tackling problem must already exist	
Author's own knowledge	<ul style="list-style-type: none"> <li>- Missing tasks</li> <li>- skew towards particular domain</li> </ul>	Downstream evaluation
Derive from extant frameworks	Validity of original classification used	

**Figure 1:** Task generation: summary of methods, threats to validity, and approaches for mitigating threats.

tems [GZ09, VPF06]; reviews of commercial systems [YKSJ07]; and expert reviews of the resultant classification to find missing tasks [APS14].

Finally, authors frequently draw on their existing knowledge of literature, systems, and practices when constructing classifications. Two of the papers we reviewed explicitly acknowledge this (drawing on “new thinking” [BM13], or the author’s experience [Suo09]), but many more likely do this implicitly, including the three papers which did not specify the means by which their tasks were generated [AS05, HS12, Shn96]: especially likely given their inclusion of extensive reference to the literature.

**5.1.1. Task Generation: threats to validity**

The two main threats at this stage are (1) gathering the wrong tasks (2) gathering an incomplete set of task. These threats arise in different ways depending on the method used.

The threats to validity when gathering tasks reported in the literature or via empirical methods such as interviews and observations are the same threats encountered at the task gathering stage of the design process, and thus the problems are well documented.

Relying on tasks reported in the literature requires a certain level of understanding of domain terminology on the classification constructor’s part (who is likely to be a visualisation researcher), and/or a similar problem having already been tackled in the visualisation literature (which preferably would include a clear characterisation of the problem; however, as noted by Munzner [Mun09] problem characterisation papers are somewhat lacking in the visualisation literature).

Talking to domain experts has a number of difficulties associated with it. In practical terms, access to domain experts may be limited in terms of the their availability [SA15]. Relying on experts from a single domain may also skew the set of tasks towards that of the represented domain [KKC15], a problem if the resultant classification is intended for more generic use. A more

general, well-known issue in HCI and psychology is that people find it difficult to accurately introspect about their needs and articulate them [Mun09, SMM12]. This difficulty is compounded when developing task classifications for Exploratory Data Analysis, where the purpose is exploration, and the potential analytical tasks involved in the exploration are not necessarily known at outset [AA06, FWSN08]. Keeping discussions focussed on tasks (rather than visual solutions) can be an issue [SMM12], while the gap in understanding of terminology used by domain experts, may also be a factor during such discussions.

Relying on tasks generated by those familiar with visualisation, as opposed to domain experts (as per the strategy of Amar et al. [AES05] who surveyed visualisation students to generate a set of analysis tasks) may also result in wrong or missing tasks. Indeed Amar et al. reflect on whether they would have obtained a different set of tasks had they surveyed professional analysts.

Observational strategies require that working methods be observed e.g. observing the domain expert using an existing visualisation system or some other tool; however, this requires that at least some method for tackling the problem already exists, which may not be the case for novel problems. Where systems do exist, researchers must still be careful to establish that the problem being tackled is indeed the right one. Moreover, the inherent lack of access to the internal mental processes of participants during fly-on-the-wall observation techniques makes observing the cognitive tasks which they are performing difficult, (although contextual enquiry, where the researcher interrupts to ask questions during the observation, may overcome some of this difficulty) [SMM12].

While we are unable to offer any new strategies for mitigating these known threats, adopting a multi-strand approach to task gathering may be one way to reduce the chances of gathering the wrong or an incomplete set of tasks upon which to base a task classification. Downstream evaluation of the resulting classification (using approaches such as those outlined in Section 6) may also highlight problems at this stage.

Finally, the principle of ‘garbage in, garbage out’ applies where the categories of extant classifications are used as the basis for constructing a classification: the validity of the resultant classification will be affected by the methods involved in constructing the original classifications. Those which have not been validated during construction or evaluated in a final form may contain errors which could be propagated to future classifications. Downstream evaluation of the resultant classification is therefore necessary where extant classifications are the basis for construction.

## 5.2. Categorisation

In the taxonomic approach, once tasks are gathered, some method of establishing meaningful categories is required. In 13 of the 27 classifications we reviewed, a set of tasks were gathered either from the literature or through empirical means. Of these, only five reported the method of categorisation employed when grouping the tasks. These included identifying common characteristics and uses of techniques and abstracting beyond the details of particular implementations [SP13]; an iterative coding process [BSIM14]; affinity diagramming [AES05, YKSJ07]; and card sorting with domain experts [Rot13].

### 5.2.1. Categorisation: threats to validity

The two main threats at the stage of classification construction are incorrect and missing task categories. These threats arise from two directions: upstream, from the set of items collected at the task generation stage, and at the current stage, from the method by which categorisation is performed.

In terms of upstream threats, the threat to validity may be propagated from the task generation stage, i.e. where the wrong or an incomplete set of tasks is collected, categories based on these items are likely to be flawed. Similarly, where classifications are constructed from the categories of extant classifications which have not been evaluated, any issues with the original classifications will potentially be propagated to the classification being developed. These threats may be mitigated by carrying out validation at the task generation stage, or identified during downstream evaluation of the final classification.

In terms of threats at the categorisation stage, when carrying out a taxonomic procedure, determining what constitutes ‘similarity’ between tasks can be a non-trivial problem, particularly where tasks are drawn from across a range of application domains and may be specified inconsistently (i.e. with respect to Rind et al.’s [RAW\*15] distinctions: at different levels of composition, abstraction and even in terms of actions vs objectives). Reasoning about similarities and differences between tasks often requires some level of abstraction. As Munzner [Mun14] notes when discussing abstraction in the visualisation design process, apparent differences between tasks are often misleading as “...there are a lot of similarities in what people want to do once you strip away the surface language differences” [Mun14, p 43–44]. The use of systematic methods such as iterative coding, affinity diagramming, and card sorting techniques (as outlined in the studies mentioned above) are one way to mitigate against producing the wrong categories; Bailey [Bai94] offers a taxonomy of clustering techniques along with useful guidance on as-

sessing the possible methods. Some of the evaluative methods identified by McKenna et al. [MMAM14] for the ‘understand’ activity of their design activity framework could also be of potential use at this stage in the construction process. However, consideration also needs to be given to who is carrying out these processes. In many cases, we found that categorisation was performed by the classification constructors (normally visualisation researchers). While this may be a valid approach (in many cases the intended users of the resultant classification are visualisation researchers), reasoning about similarities and differences amongst domain tasks may best be performed by domain experts.

One further threat to the potential usefulness of a classification is its structure, in terms of the granularity (size of categories), complexity, and depth (levels in a hierarchy) of categories. While we did not find any discussion of these aspects in our review of the literature, they are of interest in other disciplines which develop classifications (such as biology and information management). In terms of granularity, use of wide categories may have the advantage of producing a simpler classification with fewer categories, but may group together tasks with important distinctions (for example, where the classification is intended for use in a task-technique mapping, grouping tasks widely may result in difficulty in finding techniques supporting the full range of tasks). Kerracher et al. [KKC15] warn that during the task categorisation processes, subtle yet important distinctions between tasks may be lost, and less commonly occurring, but important, ‘corner case’ tasks may be discarded. Their use of matrix structures from which multiple categories can be derived is one attempt to overcome this problem. Meanwhile, narrow categories can result in the opposite problem—creating an overly complex structure by differentiating sets of tasks which could meaningfully be grouped together. Similarly, classifications which employ a hierarchical structure may wish to consider the depth and complexity of their structure. While other research areas have developed rules (such as ensuring consistency in depth to promote a ‘balance’ to the hierarchy, easing predictability when browsing and navigating the structure [Hla14], or limiting the depth of the hierarchy, as in the (now outdated) ‘3 click rule’ for designing web navigation), the potential effect of hierarchical structure has not been considered when developing task classifications. The optimal structure of a classification will likely depend very much on individual circumstances and intended use. Downstream validation of the resultant classification for the intended purpose and with the intended group of end users (such as that performed by Ahn et al., [APS14], who evaluated their classification via interviews with a number of experts from different domains) is therefore very important.

Finally, where extant categories are combined to either unite, or improve upon existing classifications, it is important to validate that this has been achieved. In the former case this may be done by demonstrating that all categories have been subsumed by the new classification (e.g. Brehmer and Munzner [BM13] map all extant categories to the categories of their classifications, while Sacha [SSS\*14] use discussion and illustration to demonstrate how the extant categories have been incorporated into their framework.) In the latter case, a discussion of the shortcomings of extant works and necessary additions helps validate the need for the new classification. Where additional categories are identified, validating the

Method	Threat	Mitigation
Ad hoc methods	Structural issues (granularity and depth)	
Systematic methods (iterative coding; affinity diagramming; card sorting with domain experts)	Wrong or missing categories arising from: <ul style="list-style-type: none"> <li>- Upstream threats (wrong or missing tasks at task generation stage; validity of extant classifications used)</li> <li>- Inconsistently specified tasks</li> <li>- Wrong people performing categorisation</li> </ul>	Upstream validation Downstream evaluation Use of systematic methods

**Figure 2:** Task categorisation: summary of methods, threats to validity, and approaches for mitigating threats.

Method	Threat	Mitigation
Verbal	<ul style="list-style-type: none"> <li>- Ambiguous/unclear descriptions</li> <li>- Descriptions too brief</li> <li>- Descriptions specified in an inconsistent manner</li> </ul>	Describe category in sufficient detail e.g. Yi et al.’s format Provide concrete examples
Formal notation	Difficult for those unfamiliar with notation	Accompany with verbal description

**Figure 3:** Description: summary of methods, threats to validity, and approaches for mitigating threats.

processes involved in their identification may require use of the methods discussed in this section and also at the task generation stage.

### 5.3. Description

Schulz et al. [SNHS13] identify four ways in which visualisation tasks can be described: verbal, functional, logical, or faceted. They also note that task descriptions may be hierarchical, allowing larger tasks to be represented as sequences of smaller subtasks. Almost all of the task descriptions in the classifications in our survey were verbal. Some, such as Brehmer and Munzner’s [BM13] typology and Schulz’s task design space [SNHS13], describe tasks in a faceted manner, in which case the task description is composed of a series of elementary components. Only the work of Andrienko and Andrienko [AA06] and the two works extending it ([KKC15] and [LRAM12]) offered a functional notation in the work we reviewed.

#### 5.3.1. Description: threats to validity

The main threats at this stage of classification construction are ambiguous or unclear descriptions, and descriptions specified in an inconsistent manner.

Ambiguous or unclear descriptions can make adoption of a classification by others more difficult. Kerracher et al. [KKC15] note that while the use of formal notation avoids ambiguity and allows highly nuanced distinctions between tasks to be made, it has the disadvantage that it may be difficult for those unfamiliar with the notation to read (they include verbal descriptions alongside their formal notation to overcome this limitation). In a minority of cases, we noted that some of the verbal descriptions found in our review were too brief for us to fully grasp the intended meaning of the

category, being only a few words long. We therefore highlight the format used by Yi et al. [YKSJ07] as an example of good practice (emphasis added): “*To each category, as a title, we assigned a short identifying name (e.g., Select) and also an illustrative phrase that captures the essence of the user’s intent in performing the interaction. We describe each category to provide a definition of what it means and we also include exemplary individual interaction techniques that fall within that category.*”

The problem of the overloading of the term “task” is evident when describing task categories, in that descriptions are not always specified in a consistent manner. While some of this may stem from earlier stages in the construction process (e.g. a number of the interaction classifications have been accused of conflating actions and objectives, which may arise at the task generation or categorisation stages), in order to describe each task category in a consistent manner, we suggest keeping in mind Rind et al.’s [RAW\*15] distinctions between actions and objectives, and the varying levels of abstraction and composition, when constructing task descriptions.

### 5.4. Conceptual approaches to classification construction

As described earlier, conceptual approaches begin with a set of conceptual dimensions upon which tasks can be distinguished, and result in a set of categories which represent concepts rather than empirical cases. Some means of establishing these dimensions is therefore required. As outlined in Section 5.1, dimensions are often gathered from extant classifications, for example, both Schulz et al.’s [SNHS13] design space and Brehmer and Munzner’s typology [BM13] draw on previous work to identify the dimensions of their classifications. An alternative approach is the formal modelling process used by Andrienko and Andrienko [AA06], who manipulate a metaphorical mathematical function in order to identify the types of tasks specified by their task typology. As the dimen-

Method	Threat	Mitigation
Gather and rationalise dimensions e.g. use extant categories	Wrong/missing dimensions Provenance of extant categories	Downstream evaluation
Functional modelling	Wrong/incomplete model	
Both approaches	Reification	Provide concrete examples

**Figure 4:** Conceptual approaches: summary of methods, threats to validity, and approaches for mitigating threats.

sions are established at outset, there is no need for the categorisation step of the taxonomic approach. However, where dimensions are gathered in a taxonomic fashion, some means of rationalising them and establishing which dimensions to include in the classification is required. Where categories are derived from extant classifications, these need to be combined to form the new system. We found that the process used to synthesise extant classifications is rarely reported, although how the resultant classification fits with those on which it is based is sometimes discussed and/or illustrated. Some form of description of the resultant categories is still required, therefore the discussion in Section 5.3 is relevant. For design spaces, the combination of choices along each dimension often serve to suitably define the category.

#### 5.4.1. Conceptual approaches: threats to validity

There are two main threats to validity for conceptual approaches: missing categories and reification.

While formal modelling approaches are able to claim completeness with respect to the model used [AA06], the classification is only as comprehensive as the model or dimensions upon which it is based. As discussed in Section 5.1.1, where extant categories are utilised, consideration needs to be given to their provenance. Downstream evaluation of the comprehensiveness of the resulting classification using approaches such as those outlined in Section 6 may highlight problems with missing or inappropriate categories.

Conceptual approaches also face a unique threat to validity: the question of whether the tasks are in fact ‘real world’, as opposed to constructs of the process employed. Bailey [Bai94] refers to this as the problem of “reification”, where theoretical constructs that do not exist empirically are ‘reified’ and treated as ‘real’ empirical entities. Providing concrete examples goes some way to mitigating this threat, however, validating the real world nature of tasks is a tricky problem, which we discuss further in Section 6.2.

## 6. Existing evaluation practices

In our review we found that evaluation of task classifications was in general rather lacking. Our definition of evaluation was broad, in that we considered any discussion regarding the limitations of the classification or its relation to other works to be a form of evaluation. Yet 9 of the 27 papers we reviewed offered no explicit evaluation. In many cases, where discussions were included, these reflections were brief and unsystematic (they were perhaps not intended to serve the purpose of evaluation). The vast majority of evaluations

in our review were discussion based (either discussions of limitations, discursive comparisons with extant classifications, or discussion in relation to some predefined evaluative criteria (see Section 6.2)). Only 8 papers used an empirical approach, such as use of the classification in the design or evaluation of systems, or testing it with domain experts. While discussions are a valid form of assessment, the brevity and lack of rigour in some indicate that this topic could benefit from more attention.

In the following discussion we divide evaluation strategies according to what they seek to evaluate: the method of construction, properties of the classification (descriptive powers, comprehensiveness, real world nature of tasks, syncretism, usability), and use of the classification (in the design and evaluation processes). These latter two strategies are of course not entirely distinct, as through evaluating the use of the classification, authors often seek to evaluate the fundamental properties of the classification. Finally, we discuss adoption rates as an additional method of evaluation. While reviewing the literature we also noted a distinction in evaluation practices between classifications which seek to improve upon extant classifications (and thus the need to evaluate in relation to other works in order to demonstrate some comparative advantage) and those which seek to unify extant works (which need to demonstrate that they have the ability to capture all aspects of extant classifications). We summarise this discussion in Table 5.

### 6.1. Evaluation of construction method

While we have identified the threats to validity and potential validation approaches at each stage of classification construction (Section 5), these were not widely cited in the literature we reviewed. One exception is Kerracher et al. [KKC15] who provide a detailed critique of the method employed in constructing their classification in comparison to alternative approaches as one means of evaluating their classification. Several papers also reflect on the construction method when discussing the limitations and advantages of their work. Roth et al. [Rot13] suggest that the empirically-derived nature of their framework makes it ecologically valid, and therefore offers advantages over other works. Amar et al. [AS05] reflect on the limitations of using student questions as the basis of their classification, while Brehmer et al. [BSIM14] acknowledge the limitations of using interviews to gather task sequences, noting that their classification may be incomplete due to sampling or observer bias. Wehrend and Lewis [WL90] consider the rigour with which their categories were selected, and whether an abstract mathematical approach would provide a “cleaner” analysis.

Aspect	Method
Construction method	Critique of method employed [KKC15, Rot13, AS05, BSIM14, WL90]
Property: descriptive power	<ul style="list-style-type: none"> <li>- Use classification to describe a known set of tasks: from existing systems [AS05, SSS*14]; common to a specific domain [SNHS13]; identified by domain experts [APS14]; carried out by users of visualisation systems [VPF06]; problems in the literature [WL90]</li> <li>- Compare with other classifications' descriptive powers [BM13]</li> </ul>
Property: comprehensiveness	<ul style="list-style-type: none"> <li>- Discussion of limitations [SNHS13, AES05, BSIM14]</li> <li>- Demonstrate able to cover more tasks than extant works [KKC15]</li> <li>- Describe a (large) known set of tasks (as per descriptive power)</li> <li>- Formal proof (formal modelling processes only) [AA06]</li> </ul>
Property: real world nature of tasks	<ul style="list-style-type: none"> <li>- Provide illustrative concrete examples [AA06, SP13] from across multiple domains.</li> <li>- Input from domain experts</li> </ul>
Property: syncretism	<ul style="list-style-type: none"> <li>- Discussion [SSS*14, vLFB*14]</li> <li>- Map categories of extant classifications to proposed classification [BM13].</li> </ul>
Property: usability	Assess use of classification by intended users; interviews [APS14, GZ09]
Usage: design process	<ul style="list-style-type: none"> <li>- Empirical evaluation using the classification in the design process: as a generative method/task organisation [APS14]; in task organisation, tool selection [SNHS13]</li> <li>- Demonstrate use via hypothetical design scenario [AS05]</li> <li>- Report on results of using classification in design process [WL90]</li> <li>- Implement classification (for analytical provenance) [GZ09, vLFB*14]; interviews to assess ease of implementing [GZ09]</li> </ul>
Usage: evaluation	Demonstrate use in evaluation process e.g. use of classification to characterise and compare task support in extant systems [AS05, SNHS13, SSS*14, AMA*14, LPP*06]; evaluate an individual system in terms of task support [SSS*14]
Adoption	Adoption rates as indicator of validity of classification.

Figure 5: Aspects of task classifications which can be evaluated and associated evaluation strategies

## 6.2. Evaluation of classification properties

Let us first consider the fundamental properties of a classification that we found to be subject to evaluation. Two papers explicitly validate their classification in relation to a set of pre-identified criteria. Yi et al. [YKSJ07] discuss their classification with respect to Beaudouin-Lafon's [BL04] dimensions to evaluate interaction models - descriptive power, evaluative power, and generative powers (which relates to use in the design process). Sedig and Parsons' [SP13] consider their classification in relation to four characteristics: syncretic (its ability to unify previously disconnected ideas), general (in its level of abstraction and applicability), comprehensive (in terms of coverage), and generative (in relation to use in design process and guiding future research). Ahn et al. [APS14] design an empirical evaluation in which they seek to validate comprehensiveness (task coverage); ease of use; precision (descriptive power); use in task organisation and clarification; and use in task discovery (as a generative method).

As mentioned, we consider evaluation with respect to use of the classification (e.g. in the design and evaluation processes) separately from the evaluation of properties. Beyond the papers mentioned above, descriptive powers and comprehensiveness were also the main properties evaluated more widely in the literature. We also found verifying the existence of tasks in the "real world", and usability of the classification to be of interest.

*Evaluating descriptive powers:* the fundamental purpose of a classification system is to use a common language to be able to describe the full range of tasks in a consistent way. A common method

to evaluate this ability is to use the classification to describe a set of known tasks and check that they can all be adequately captured. Examples include using the classification to describe the tasks which can be supported by an existing system [AS05, SSS\*14]; those common to a specific domain [SNHS13] or identified by domain experts [APS14]; or sequences of tasks carried out by users of visualisation systems [VPF06]. Brehmer and Munzner [BM13] go one step further by explicitly comparing their classification's ability to describe a sequence of tasks supported by an extant system, with task descriptions generated by other classifications. Thus they are also able to demonstrate how their classification overcomes the shortcomings of others in terms of its ability to describe.

*Evaluating comprehensiveness:* a classification's ability to capture all possible tasks is in many respects related to its descriptive powers. However, to evidence that a classification is complete is rather difficult; as per the problem of induction, we are always just one task away from finding a case which our classification cannot cover. It is particularly difficult to demonstrate when the taxonomy is intended to be useful across multiple domains, with a wide range of possible tasks. Additionally, classifications may not be able to capture tasks specified at multiple levels of composition (i.e. high or low level tasks). Discussions relating to limitations often seek to demarcate the limits of the claimed comprehensiveness of a framework e.g. with regards to tasks omitted or those that lie outwith the intended scope of the classification e.g. [SNHS13, AES05, BSIM14]. While not showing completeness, demonstrating that the developed classification is able to cover more tasks than extant frameworks is one form of valida-

tion e.g. [KKC15], (although [SNHS13] highlight the difficulties in comparing task classifications, including that categories which have been separated out in one work may be mixed in another, or authors may have fundamentally different ideas about what a task is). Similar to the evaluation strategy outlined with regard to descriptive powers, mapping a large set of tasks e.g. from problems identified in the literature to the task categories [WL90], may also go some way to demonstrate task coverage. Finally, where formal modelling approaches have been employed in the classification's constructions, a formal proof can be used to demonstrate comprehensiveness of the classification, at least with respect to the chosen model [AA06].

*Evaluating "real world" nature of tasks:* Classifications developed using formal modelling techniques e.g. [AA06, KKC15, LRAM12], or design spaces, where all possible permutations of tasks are generated by combining dimensions, e.g. [APS14, SNHS13], leave open the question of whether the generated categories are merely constructs of the formal process or are in fact representative of 'real world' tasks. Validation of such frameworks may therefore involve establishing that the tasks are indeed real world tasks. This type of validation is usually dealt with in the literature by providing illustrative concrete examples for each possible category of abstract task. The most comprehensive example is probably Andrienko and Andrienko [AA06] who include several data scenarios from different domains which they use to provide examples to illustrate most of the possible iterations of tasks generated by their modelling approach. Sedig and Parsons [SP13] offer examples of existing tools which implement each of their patterns in order to evidence the existence and necessity of each pattern in the real world. However, few frameworks offer an example task for every possible permutation of their model. Evaluating the real world nature of tasks can therefore prove tricky. Simply because we cannot readily think of a concrete example of a task category, it does not mean that it is not a real task, albeit perhaps exclusive to a particular domain or niche analysis scenario: Schulz et al. [SNHS13] note that "what looks like an inconsistency in the design space may actually be just a very creative and unusual combination of design choices". For the more extensive frameworks, examples may need to be drawn from multiple domains to cover all possible tasks, which may require input from multiple domain experts.

*Evaluating syncretism:* for classifications which seek to unite extant classifications, it is important to show that extant categories can be subsumed under the proposed system. Often the categories are compared through discursive methods e.g. [SSS\*14], sometimes highlighting the advantages of the proposed classification [vLFB\*14]. A more rigorous approach is to explicitly map the categories of extant frameworks to those of the proposed framework [BM13]. The resulting mapping not only shows where the categories sit, but reveals which categories are under- and over-represented in previous works.

*Evaluating usability:* Ahn et al. [APS14] is the only paper we reviewed which assesses the use of the classification by someone other than the classification developer, allowing them to evaluate its usability. They do so as part of their empirical study which involves interviews with domain experts.

### 6.3. Evaluation of usage

While a wide range of usage scenarios for task classifications have been identified (see Section 4), and envisaged uses of the developed classification are often outlined in detail e.g. [BSIM14, BM13], only a few papers evaluate their classifications directly with respect to their intended usage.

*Use in the design process:* Two papers include empirical evaluations of use in domain characterisation and abstraction. Ahn et al. [APS14] use interviews to explore the use of their classification as a generative method when establishing tasks of interest to domain experts, and its ability to help them organise, describe and clarify their tasks. Schulz et al. [SNHS13] report on a use case with domain experts. They use their classification to characterise and organise the tasks, establishing the most common and important tasks. Having also characterised existing tools according to the tasks that they support, they are able to select tools which are able to support the identified tasks. Other studies report more generally on the use of their classification in guiding the design process. Amar and Stasko [AS05] demonstrate the use of their framework in a hypothetical design scenario, in order to illustrate its use as a "systematic basis for thinking about and identifying issues in the data set". However, they do not develop (and therefore cannot evaluate) the system. Wehrend and Lewis [WL90] used their catalogue to develop a visualisation, presumably using it to select appropriate techniques for inclusion (they do not give a detailed report regarding its use in the design process). They do not formally evaluate the resulting representation, but conclude that it "appears to be an improvement over earlier representations designed in an ad hoc manner". This highlights one difficulty with this kind of validation, in that it can be difficult to say to what extent the classification was useful in guiding the design. For example, had the design process proceeded without the use of the classification, would the resulting system have been any different, or in some way less good?

Two papers, [GZ09] and [vLFB\*14] - both concerned with analytical provenance - directly implement their classifications in the design of a system in order to track users' analytical processes. Gotz and Zhou [GZ09] seek feedback from developers regarding the ease of implementing the model, and interview analysts who used the system with regard to how well it aligned with their mental models.

*Use in evaluation:* while we have identified many ways in which task classifications can be used in evaluations (indeed, this is often cited as the primary motivation/purpose for their development), we found only one type of evaluation scenario included in the body of work we reviewed: using the classification to characterise extant systems, then compare them according to task support [AS05, SNHS13, SSS\*14, AMA\*14, LPP\*06]. Sacha et al. [SSS\*14] also include a variation of this scenario, where they use their classification to assess a visual analytics system in terms of how it supports different aspects of their classification. This allows them to point to shortcomings and areas for improvement in the system's design.

#### 6.4. Evaluation with respect to adoption

One final evaluation strategy identified in the literature is that of adoption rates: Heer and Shneiderman [HS12] suggest validation via “community feedback, critique and refinement”. The adoption, evolution, and demise of task classifications ‘in the wild’, may provide significant information about their descriptive abilities, comprehensiveness, usefulness, and usability. Where limitations in task coverage are encountered, classifications are often adapted or extended (e.g. the extension of the Andrienko framework [AA06] to temporal data [LRAM12] or temporal graph data [KKC15]), unified, or new classifications are developed. Where task classifications are found not to be useful they are likely to be superseded. Even where a classifications may offer better descriptive abilities or more comprehensive task coverage, in the busy world of visualisation research, classifications which are easy to understand and require little learning overhead may be more likely to succeed.

#### 7. Conclusions and Future Work

In this paper we have reviewed the task classification literature in relation to classification construction methods and current evaluation practices. We identified the various approaches to construction at each of the stages and the associated threats to validity. For each of these threats we highlight potential strategies for their mitigation. We also discussed the evaluation strategies appearing in the literature, organising them according to what aspect of the classification they address: construction method, general properties of the classification, or usage scenarios.

We anticipate that this work will be of particular use to task classification developers when considering appropriate approaches to constructing and evaluating their classifications. It may also be of use as a reference to reviewers of papers reporting on task classification development and potential adopters of classifications, in making judgements about the validity and utility of the classification.

Based on the discussions reported in this paper, we make the following recommendations:

In order to develop a robust classification, during construction we recommend consideration of the relative advantages and disadvantages of the various strategies possible at each stage and employment of appropriate methods to validate them, in addition to downstream evaluation of the final classification. Importantly, we highlight the need for classification developers to report on the construction methods and corresponding validation approaches used when presenting their work, in order to assist readers and potential classification adopters make informed judgements with regard to the rigour and validity of the construction approach and resultant classification.

With regard to evaluation practices, one limitation of our work is its focus on extant evaluation practices, which we generally found to be rather lacking. Thus, while we have highlighted good practices where we found them, we are unable to offer fully prescriptive guidance based on our review of the extant task classification literature. While our primary recommendation is the need for inclusion of evaluation when developing task classifications (in the

same way that this would be expected when developing a visualisation system), we also recommend that further consideration to needs to be given to appropriate strategies for evaluating task classifications. Finally, we again highlight the need for better reporting practices with regard to evaluation. We recommend that classification developers report on both the evaluation strategy adopted and clearly identify which aspect(s) of the classification their evaluation approach intends to target. We anticipate that our identification of the different aspects of task classifications which can potentially be evaluated will be of benefit here.

While a large number of task classifications were included in our review, we are aware that the list on which it was based was not intended by the original authors to be exhaustive. It is therefore possible that we may have found additional use cases, construction methods, or evaluation practices by including additional works. However, a consistent set of practices and themes emerged from the work that we did include. We also limited our review to the papers specifically reporting on the *development* of a task classification. Many of the suggested use cases offered in the papers (particularly with respect to evaluation scenarios) were not actually demonstrated in these papers. By surveying the literature more widely to include papers which cite—and more specifically, adopt and use—these classifications, we may gain further information regarding the uses and usefulness of these works to the community. However, an initial review of citations relating to two of the publications (one older [AS05] and one more recent [APS14]), revealed very limited uptake with regard to the intended purposes of the classifications, such as employment in the design and evaluation processes. Instead, they were more frequently cited as examples of relevant extant classifications. Finally, we purposefully restricted our review to *task* classifications; however, as mentioned in our introduction, the penchant for classification in the visualisation community does not end with tasks; authors have sought to organise many other aspects relating to visualisation, including classifying visualisation tools and techniques, interaction approaches, data types, visual encodings, input methods, and evaluation strategies, amongst others. A wider review of the use of classification in visualisation, similar in structure to this paper and covering the benefits to the community, uses, construction methods, and evaluation practices, could offer a useful contribution to the movement towards formalising visualisation practices, and is planned for our future work.

#### References

- [AA06] ANDRIENKO N., ANDRIENKO G.: *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer, New York, 2006. doi:10.1007/3-540-31190-4. 4, 5, 7, 8, 10, 11
- [AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. (2005), 111–117. doi:10.1109/INFVIS.2005.1532136. 4, 5, 6, 9
- [AMA\*14] ALSALLAKH B., MICALLEF L., AIGNER W., HAUSER H., MIKSCHE S., RODGERS P.: Visualizing Sets and Set-typed Data: State-of-the-Art and Future Challenges. *Eurographics conference on Visualization (EuroVis)—State of The Art Reports* (2014), 1–21. doi:10.2312/eurovisstar.20141170. 4, 10
- [APS14] AHN J.-w., PLAISANT C., SHNEIDERMAN B.: A task taxonomy for network evolution analysis (supplemental material). *IEEE Transactions on Visualization and Computer Graphics* (2014). 2, 3, 4, 5, 6, 9, 10, 11

- [AS05] AMAR R. A., STASKO J. T.: Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics* 11, 4 (2005), 432–442. doi:10.1109/TVCG.2005.63. 2, 3, 4, 5, 8, 9, 10, 11
- [Bai94] BAILEY K. D.: *Typologies and taxonomies: an introduction to classification techniques*. No. 07. Sage, 1994. doi:10.1002/(SICI)1097-4571(199604)47:4<328::AID-ASI110>3.0.CO;2-Y. 2, 6, 8
- [BL04] BEAUDOUIN-LAFON M.: Designing interaction, not interfaces. In *Proceedings of the working conference on Advanced visual interfaces* (2004), ACM, pp. 15–22. 9
- [BM13] BREHMER M., MUNZNER T.: A multi-level typology of abstract visualization tasks. *IEEE transactions on visualization and computer graphics* 19, 12 (dec 2013), 2376–85. doi:10.1109/TVCG.2013.124. 1, 4, 5, 6, 7, 9, 10
- [BSIM14] BREHMER M., SEDLMAIR M., INGRAM S., MUNZNER T.: Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *In ACM Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*. (2014), Lam H., Isenberg P., Isenberg T., Sedlmair M., (Eds.), pp. 1–8. doi:10.1145/2669557.2669559. 2, 4, 6, 8, 9, 10
- [CR96] CHUAH M., ROTH S.: On the semantics of interactive visualizations. *Proceedings IEEE Symposium on Information Visualization '96* (1996), 29–36. doi:10.1109/INFVIS.1996.559213. 4
- [FWSN08] FEKETE J.-D., WIJK J. J. V., STASKO J. T., NORTH C.: The Value of Information Visualization. *Information Visualization* 4950, 2 (2008), 1–18. 5
- [GZ09] GOTZ D., ZHOU M. X.: Characterizing Users' Visual Analytic Activity for Insight Provenance. *Information Visualization*, 8, 1 (2009), 42–55. doi:http://dx.doi.org/10.1057/ivs.2008.31. 4, 5, 10
- [HBO10] HEER J., BOSTOCK M., OGIEVETSKY V.: A tour through the visualization zoo. *Commun. ACM* (2010), 59–67. doi:10.1145/1743546. 3
- [Hla14] HLAVA M. M.: The taxobook: Principles and practices of building taxonomies, part 2 of a 3-part series. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 6, 4 (2014), 1–164. 6
- [HS12] HEER J., SHNEIDERMAN B.: Interactive Dynamics for Visual Analysis. *Queue - Microprocessors* 10, 2 (2012), 30. doi:10.1145/2133416.2146416. 3, 4, 5, 11
- [IIC\*13] ISENBERG T., ISENBERG P., CHEN J., SEDLMAIR M., MOLLER T.: A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2818–2827. doi:10.1109/TVCG.2013.126. 1
- [JE12] JAVED W., ELMQVIST N.: Exploring the design space of composite visualization. *2012 IEEE Pacific Visualization Symposium* (feb 2012), 1–8. doi:10.1109/PacificVis.2012.6183556. 2, 3
- [KKC14] KERRACHER N., KENNEDY J., CHALMERS K.: The Design Space of Temporal Graph Visualisation. In *Proceedings of the Eurographics Conference on Visualization (EuroVis '14), Short Papers Track* (Swansea, 2014), Elmqvist N., Hlawitschka M., Kennedy J., (Eds.), The Eurographics Association, pp. 7–11. doi:10.2312/eurovisshort.20141149. 2, 3
- [KKC15] KERRACHER N., KENNEDY J., CHALMERS K.: A Task Taxonomy for Temporal Graph Visualisation. *Visualization and Computer Graphics, IEEE Transactions on* 21, 10 (oct 2015), 1160–1172. doi:10.1109/TVCG.2015.2424889. 2, 3, 4, 5, 6, 7, 8, 10, 11
- [KKCG15] KERRACHER N., KENNEDY J., CHALMERS K., GRAHAM M.: Visual Techniques to Support Exploratory Analysis of Temporal Graph Data. In *Eurographics Conference on Visualization (EuroVis) - Short Papers* (2015), Bertini E., Kennedy J., Puppo E., (Eds.), The Eurographics Association. doi:10.2312/eurovisshort.20151133. 2, 3
- [LBAL09] LTIIFI H., BEN AYED M., ALIMI A., LEPREUX S.: Survey of information visualization techniques for exploitation in KDD. In *IEEE/ACIS International Conference on Computer Systems and Applications, 2009. AICCSA 2009*. (2009), no. 1, pp. 218–225. doi:10.1109/AICCSA.2009.5069328. 3
- [LBI\*11] LAM H., BERTINI E., ISENBERG P., PLAISANT C., CARPENDALE S.: Empirical Studies in Information Visualization: Seven Scenarios. *IEEE transactions on visualization and computer graphics* (nov 2011), 1–18. doi:10.1109/TVCG.2011.279. 1, 4
- [LPP\*06] LEE B., PLAISANT C., PARR C. S., FEKETE J.-D., HENRY N.: Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization* (New York, NY, USA, 2006), BELIV '06, ACM, pp. 81–85. doi:http://doi.acm.org/10.1145/1168149.1168168. 2, 4, 10
- [LRAM12] LAMMARSCH T., RIND A., AIGNER W., MIKSCH S.: Developing an Extended Task Framework for Exploratory Data Analysis Along the Structure of Time. In *EuroVA 2012: International Workshop on Visual Analytics* (2012), Matkovic K., Santucci G., (Eds.), The Eurographics Association, pp. 31–35. doi:10.2312/PE/EuroVAST/EuroVA12/031-035. 3, 4, 7, 10, 11
- [LS10] LIU Z., STASKO J. T.: Mental Models, Visual Reasoning and Interaction in Information Visualization: A Top-down Perspective. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 999–1008. doi:10.1109/tvcg.2010.177. 4
- [Mac95] MACEACHREN A. M.: *How maps work: representation, visualization, and design*. Guilford Press, 1995. 2
- [MMAM14] MCKENNA S., MAZUR D., AGUTTER J., MEYER M.: Design Activity Framework for Visualization Design. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (dec 2014), 2191–2200. doi:10.1109/TVCG.2014.2346331. 3, 6
- [MMC02] MALETIC J., MARCUS A., COLLARD M.: A task oriented view of software visualization. *Proceedings First International Workshop on Visualizing Software for Understanding and Analysis* (2002), 32–40. doi:10.1109/VISSOF.2002.1019792. 3
- [MMP09] MEYER M., MUNZNER T., PFISTER H.: MizBee: A multi-scale synteny browsers. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 897–904. doi:10.1109/TVCG.2009.167. 4
- [MSM12] MEYER M., SEDLMAIR M., MUNZNER T.: The Four-Level Nested Model Revisited: Blocks and Guidelines. 1
- [Mun09] MUNZNER T.: A Nested Model for Visualization Design and Validation. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 921–928. doi:10.1109/TVCG.2009.111. 1, 3, 5
- [Mun14] MUNZNER T.: *Visualization Analysis and Design*. CRC Press, 2014. 2, 3, 6
- [PPS14] PRETORIUS A. J., PURCHASE H. C., STASKO J. T.: Tasks for multivariate network analysis. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8380 LNCS (2014), 77–95. doi:10.1007/978-3-319-06793-3\_5. 4
- [RAW\*15] RIND A., AIGNER W., WAGNER M., MIKSCH S., LAMMARSCH T.: Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization* (2015). 1, 2, 3, 4, 6, 7
- [Rot13] ROTH R. E.: An Empirically-Derived Taxonomy of Interaction Primitives for Interactive Cartography and Geovisualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2356–2365. 4, 6, 8
- [RWA\*13] RIND A., WANG T. D., AIGNER W., MIKSCH S., WONGSUPHASAWAT K., PLAISANT C., SHNEIDERMAN B.: Interactive Information Visualization to Explore and Query Electronic Health Records. *Foundations and Trends in Human-Computer Interaction* 5, 3 (2013), 207–298. doi:10.1561/11000000039. 4

- [SA15] SAKAI R., AERTS J.: Card Sorting Techniques for Domain Characterization in Problem-driven Visualization Research. *Eurographics Conference on Visualization (EuroVis)* (2015). doi:10.2312/eurovisshort.201511136. 5
- [Shn96] SHNEIDERMAN B.: The eyes have it: a task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996), pp. 336–343. 3, 5
- [SHS11] SCHULZ H.-J., HADLAK S., SCHUMANN H.: The Design Space of Implicit Hierarchy Visualization: A Survey. *IEEE transactions on visualization and computer graphics* 17, 4 (may 2011), 393–411. 2, 3
- [SM13] SCHUMANN H., MÜLLER W.: *Visualisierung: Grundlagen und allgemeine Methoden*. Springer-Verlag, 2013. 2
- [SMM12] SEDLMAIR M., MEYER M., MUNZNER T.: Design Study Methodology : Reflections from the Trenches and the Stacks. 1, 2, 3, 5
- [SNHS13] SCHULZ H. J., NOCKE T., HEITZLER M., SCHUMANN H.: A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2366–2375. doi:10.1109/TVCG.2013.120. 1, 2, 3, 4, 7, 9, 10
- [SP13] SEDIG K., PARSONS P.: Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Transactions on Human-Computer Interaction* 5, 1 (2013), 84–133. 3, 4, 6, 9, 10
- [SSS\*14] SACHA D., STOFFEL A., STOFFEL F., KWON B. C., ELLIS G., KEIM D. A.: Knowledge generation model for visual analytics. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1604–1613. 2, 4, 6, 9, 10
- [Suo09] SUO X.: *A Task-Centered Visualization Design Environment and a Method for Measuring the Complexity of Visualization Designs*. PhD thesis, Georgia State University, Atlanta, 2009. 4, 5
- [TC05] THOMAS J. J., COOK K. (Eds.): *Illuminating the path:[the research and development agenda for visual analytics]*. IEEE Computer Society, 2005. 2
- [TM04] TORY M., MOLLER T.: Rethinking Visualization: A High-Level Taxonomy. *IEEE Symposium on Information Visualization* (2004), 151–158. doi:10.1109/INFVIS.2004.59. 3
- [vLFB\*14] VON LANDESBERGER T., FIEBIG S., BREMM S., KUIJPER A., FELLNER D. W.: Interaction taxonomy for tracking of user actions in visual analytics applications. In *Handbook of Human Centric Visualization*, Huang W., Parsons P., Sedig K., (Eds.). Springer, 2014, pp. 653–670. doi:10.1007/978-1-4614-7485-2. 4, 10
- [VPF06] VALIATI E. R. A., PIMENTA M. S., FREITAS C. M. D. S.: A taxonomy of tasks for guiding the evaluation of multidimensional visualizations. In *Proceedings of the 2006 AVI workshop on BEyond time and errors novel evaluation methods for Information Visualization - BELIV '06* (New York, New York, USA, 2006), Bertini E., Plaisant C., Santucci G., (Eds.), ACM Press. doi:10.1145/1168149.1168169. 4, 5, 9
- [vW06] VAN WIJK J. J.: Bridging the Gaps. *Computer Graphics and Applications, IEEE* 26, 6 (2006), 6–9. 3
- [WL90] WEHREND S., LEWIS C.: A problem-oriented classification of visualization techniques. *Proceedings of the First IEEE Conference on Visualization: Visualization '90* (1990), 139–143. doi:10.1109/VISUAL.1990.146375. 2, 3, 4, 8, 10
- [YKSJ07] YI J. S., KANG Y. A., STASKO J., JACKO J.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE transactions on visualization and computer graphics* 13, 6 (2007), 1224–31. doi:10.1109/TVCG.2007.70515. 1, 3, 4, 5, 6, 7, 9