EUROVIS 2019
R. S. Laramee, S. Oeltze, and M. Sedlmair
(Guest Editors)

Volume 38 (2019), Number 3
STAR – State of The Art Report

A Review of Guidance Approaches in Visual Data Analysis: A Multifocal Perspective

Davide Ceneda†, Theresia Gschwandtner, and Silvia Miksch

TU Wien, Austria

Abstract

Visual data analysis can be envisioned as a collaboration of the user and the computational system with the aim of completing a given task. Pursuing an effective system-user integration, in which the system actively helps the user to reach his/her analysis goal has been focus of visualization research for quite some time. However, this problem is still largely unsolved. As a result, users might be overwhelmed by powerful but complex visual analysis systems which also limits their ability to produce insightful results. In this context, guidance is a promising step towards enabling an effective mixed-initiative collaboration to promote the visual analysis. However, the way how guidance should be put into practice is still to be unravelled. Thus, we conducted a comprehensive literature research and provide an overview of how guidance is tackled by different approaches in visual analysis systems. We distinguish between guidance that is provided by the system to support the user, and guidance that is provided by the user to support the system. By identifying open problems, we highlight promising research directions and point to missing factors that are needed to enable the envisioned human-computer collaboration, and thus, promote a more effective visual data analysis.

CCS Concepts

• ***Human-centered computing*** → ***Visual analytics;*** *Visualization theory, concepts and paradigms;* • ***Information systems*** → ***Decision support systems;***

1. Introduction

Data analysis refers to procedures to make sense of data [Tuk77]. As we continue to produce ever-growing amounts of data, data analysis is a necessity and has implications on many disciplines, such as environmental sciences, medicine, or business development. Information Visualization (InfoVis) is a combination of data analysis, human-computer interaction (HCI), psychology, visual design, and computer graphics [BS03]. InfoVis is commonly known as an effective *user-centered* way to make sense of large and complex data by means of external cognition. Non visual approaches to data analysis exist too. Data mining [FPSSU96] is probably one of the most known *machine-centered* methods. Its main focus is extracting patterns and models of the data, both for describing phenomena but also for predicting future events. While InfoVis solutions focus predominantly on providing visual means to interact and explore the data, mining methods are instead seen as black boxes, consuming data and producing models.

Visual Analytics (VA) stems from the goal of intertwining the strengths of visualizations and computational models. Keim et al. [KMS*08] described the VA process, listing the different affordances of the user and the computational hardware [Gib77]. Despite the great amount of work in this area, it is still unclear how this human-computer collaboration should be put into practice. While in the past there has been a lot of effort of producing effective interactive interfaces [LCWL14] as well as computational models [GG99, Ber06], just a few approaches were focused on bringing together the strengths of humans and computers in a *mixed-initiative* manner [Hor99]. A widely used term in this research direction is *human-computer collaboration*, which refers to the joint efforts of two or more agents (of which at least one is human and one is a computer) to achieve a common analysis goal [Ter95].

Achieving an effective system-user integration ($SYSTEM \rightleftharpoons USER$), however, is still a largely unresolved task, mostly because it requires to effectively combine multiple aspects of the data analysis process, like for instance the tasks to be supported, the computational methods needed, the knowledge of the users, and the visual means to be utilized, just to name a few. In past years, great efforts have been conducted into this direction. Scientists tried to tackle selected aspects of the knowledge generation model [SSS*14b]. Bertini and Lalanne [BL09] describe scientists' attempts to enhance visualizations with computational

† The authors are with TU Wien, Faculty of Informatics, Institute of Visual Computing & Human-Centered Technology, Favoritenstrasse 9-11/193, A-1040 Vienna, Austria. E-mail: *name.surname*@tuwien.ac.at

methods, or support mining with visual means. However, they also state that an effective integration, in which the affordances of human and computer are balanced and effectively combined, has not been accomplished yet. In particular, they claim that opening the black box of data mining techniques so to allow users to steer the process is far from being solved.

A promising attempt to enable a better collaboration of the human and the computer is *guidance*. Guidance is for its nature a mixed-initiative process: It comprises (1) the assistance a system gives to a user, and (2) to the feedback the user provides to the system in order to steer the analysis process and to achieve a common goal. In the context of mixed-initiative data analysis, Ceneda et al. [CGM*17] describe what input data source the guidance process can exploit, what are the possible outcomes of this human-computer collaboration and the different degrees of guidance the users may need [CAS*18]. Although all ingredients (*What?*) necessary for this collaboration exist, a description of the ways they should be intertwined to obtain effective results (*How?*) is still missing [BL09].

With the aim of providing an initial answer to this question, we survey the literature of mixed-initiative guidance approaches. We complement the original work on guidance with a description of *how* the system and the user interact during the data analysis process. We achieve this, by focusing on four dimensions:

1. the objective of guidance, i.e., the reason why guidance is provided in the first place;
2. the guidance degree, i.e., how much and which quality of guidance is provided;
3. from the user perspective, we will describe how users are enabled to provide feedback to steer the guidance process and
4. the types of feedback the user gives to the system. In particular, users' feedback can be tailored towards the guidance received in the past or by encouraging future guidance.

Considering the *guidance degree* and the *type of guidance*, will allow us to describe how the general analysis *objective* is reached from a *system perspective*. The focus on the *interaction modalities and feedback types*, will allow us to illustrate how the objective is achieved from a *user perspective*.

In the effort to identify the guidance objectives, we exploit the knowledge generation process described by Sacha et al. [SSS*14b]. More specifically, we mark the steps of the analysis process on which a computational system acts to provide guidance to the user ($SYSTEM \rightharpoonup USER$). We further describe the visual means used to convey such guidance. In this way, we want to highlight *how* guidance is provided to the user. For the two other dimensions, we exploit the guidance characterization by Ceneda et al. [CGM*17], and the well-known literature regarding interaction and feedback in visualization [Dow99]. In this way, we aim to describe *how* conversely the users are able to influence and steer the guidance process ($SYSTEM \leftharpoondown USER$). We formulate our goals in two research questions:

1. What approaches provide guidance to support the different tasks/steps of the visual analysis process, and how is this human-computer collaboration taking place?
2. What traits of the visual analysis process are nowadays uncovered by the majority of mixed-initiative guidance techniques?

By answering the aforementioned questions, we aim at providing an overview of guidance approaches and at the same time stimulate future research of guidance approaches in visual data analysis. Thus, our main contributions are:

- We provide a categorization of existing guidance approaches, in visual data analysis;
- We illustrate how guidance can be used to support the most common analysis tasks;
- We outline open challenges and emerging research directions, in the field of guidance and derive what are the missing factors to make the human-computer collaboration work effectively.

2. A Systematic Review of Literature

In this paper, we provide a systematic review of guidance approaches by describing how all the procedures and activities involved in the data analysis process are conducted and fostered by the system and the user in a shared and collaborative manner. In the following, we outline our research methodology and provide a clarification of the terms we use.

Literature Research and Methodology Our methodology was characterized by a sequence of refinement cycles, in which we alternated bottom-up to top-down phases. On the one hand, we performed a comprehensive literature research, looking for visual data analysis approaches containing elements of guidance. On the other hand, in a top-down fashion, we refined the literature research with respect to the categories we identified in our literature review.

Being a mixed-initiative process, guidance is strongly rooted in the HCI community. This is where we started our research. We looked into the proceedings of major conferences like the *Conference on Human Factors in Computing Systems (CHI)*, using keywords like: *"mixed-initiative"*, *"guidance"*, as well as synonyms and related words. We expanded the research following the references of the initial pool of papers we collected and using the functions of *Google Scholar* to retrieve pertinent related works. We further explored the proceedings of further conferences, like *IEEE VIS (InfoVis, VAST)*, and *IEEE International Conference on Knowledge Discovery and Data Mining (KDD)*. In this last group of conference proceedings, we looked for techniques and approaches focusing on specific aspects of the knowledge generation process (e.g., data transformation, model building).

A final phase of the literature research was then reserved to make an initial selection of the papers and skimming the non pertinent ones. Among all the works we collected, we kept just the approaches that constitute a mixed-initiative approach, i.e., containing elements of both, human and system guidance. In total, we collected more than three hundred papers. After the categorization we ended up with fifty-three papers, from the over three-hundred papers initially selected (see Table 1).

What we do cover In this report, we review guidance approaches. Guidance is envisioned as a dialogue between human and computer where the user brings his/her own problems and the system tries to support. At the same time, the user is also enabled to steer the course of the conversation, guiding the system. In terms of data analysis, the aim of guidance is to define and take actions to provide an answer to the analytical needs a user may develop during an interactive visual data analysis session, which hinder him/her from concluding the analysis. To understand if a given approach can be described as guidance, a simple question must be asked: is it the system or the user who makes decisions, takes actions and promotes the analysis? Being guidance a mixed-initiative process, the only answer to it is: Both of them, the user and the system [CGM*18]. Thus, just approaches complying to this definition and comprising both user and system guidance were collected.

What we do not cover It is important to specify further our focus, by describing what we do not cover. Guidance is defined as a mixed-initiative process. Since a collaboration consists of two parties working together, in our categorization we do not consider approaches in which only one agent is responsible for completing the whole task. Hence, we do not consider those approaches that are usually referred to as *human-computation* systems [CC12, QB11, YCK09] in this report. Conversely to guidance, where human and system efforts are intertwined, human-computation systems consider just one side of the collaboration. They mainly consist of methods to exploit human abilities (e.g., image/sound/pattern recognition) for all those tasks that a computer is not yet good at, sometimes using different entertaining expedients, such as gamification [VAMM*08, VALB06]. For the same reason, we do not consider pure *system-initiative* nor *user-initiative* approaches [CGM*18]. Those are approaches in which only the user (or the system) is in charge of taking decisions and advance the analysis, with the other agent taking part in the analysis as a passive executor. Since many InfoVis, VA, and also data mining papers may be ascribed to these categories we excluded them from the report. The reason is very easily explained. Conversely to guidance, only one agent is responsible for producing results and decision making. For this reason, we did not consider them in our work.

2.1. Concepts and terminology

Guidance. We build on the characterization of guidance by Ceneda et al. [CGM*18]. Guidance is defined as a mixed-initiative *"computer-assisted process that aims to actively resolve a knowledge-gap during an interactive visual analytics session"* [CGM*17, p.2]. Formally, guidance is characterized by an input, which consists of a description of the needs of the user (a knowledge-gap), plus a list of resources the process could use to generate assistance. The output consists of the computed answer to the user's knowledge gap and a set of visual means to communicate this answer. The output may be further characterized by a degree, indicating the amount of guidance provided by the answer. Three degrees of guidance are described: prescribing, directing, and orienting guidance. We will use them together with a description of the guidance answer and the visual means used to convey this answer, in order to specify how guidance is provided from the system-side to reach the given analysis objective:

- *Prescribing guidance* is the highest degree of guidance. The system establishes a set of mandatory actions, or specifies step-by-step instructions the user should take to proceed.
- *Directing guidance* is an intermediate degree. The system leads the user along multiple analysis paths and trails. The user has more freedom, but receives also less guidance. This is usually achieved by the system-indication of multiple alternative analysis options.
- *Orienting guidance* is the lowest degree. At this stage, the system aims only at keeping the user oriented, easing the exploration of the dataset, by providing, for instance, some suggestions or additional information, but without pointing to some concrete analysis path or to actual actions.

Knowledge generation model. The ultimate goal of the data analysis is making sense of data by exploration, verifying hypotheses, and generating new insights and knowledge. This process is illustrated by Sacha et al. [SSS*14b]. We build upon the dimensions of the knowledge generation process to characterize the analysis *objective*, the guidance process aims to support.

- *Data:* the user may need help to manipulate the data. This includes all the pre-processing procedures before the data is visualized or analyzed.
- *Visualization:* the user may need help to visualize the data or to refine an existing visualization. This includes finding a proper visual mapping of the data.
- *Model:* the user may need help to create a model of the data, or to refine an existing model. This includes finding a correct model of the data or appropriate parameter settings.
- *Exploration:* Although high level activities, such as exploration of a dataset, are mostly a human prerogative, the system may still be able to provide support. This includes supporting the interaction with the system itself, but also activities, like the definition of an analytical goal or the discovery of new *findings*.
- *Verification and Knowledge Generation:* Other high level activities a system might support are verification and knowledge generation. This includes the provision of means for collecting *findings* found during the exploration phase and connecting them with each other in order to foster complex *insights*, to prove or disprove *hypotheses*, and thus, to generate new knowledge.

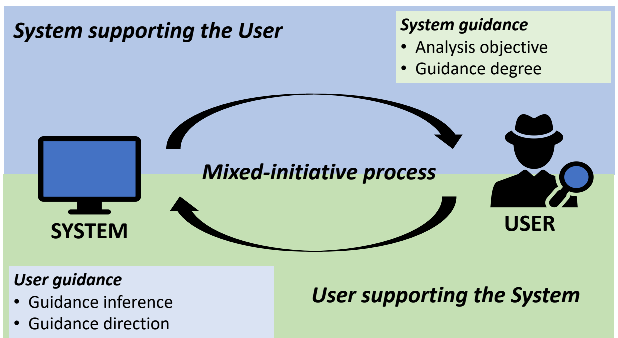


Figure 1: *Schematic representation of the mixed-initiative process and guidance. The whole process can be subdivided into two main phases: user and system guidance, which alternate until a complete analysis is achieved.*

3. Categorization Scheme

In the previous sections, we described the basic concepts of this survey. In the following, we explain the different dimensions we consider and how we compose those to form a categorization scheme. Note that a given paper may fall in more than one category, since user and system guidance coexist in the same approach. However, each category covers different aspects of the same approach.

3.1. Categories

The guidance process can be seen as a back and forth loop between the user and the system. The system guides the user and the user guides the system by providing feedback to the system guidance (see Figure 1). Thus, the system and the user are in a constant dialog until the analysis objective is reached. We describe these two different directions of guidance separately.

$SYSTEM \rightharpoonup USER$ When guidance is produced by the system and directed towards the user, the system takes the initiative to support the user. Notice that this is not necessarily the first step of a guidance process, since any of the two parties can initiate the process at any time depending on who is most qualified to solve the current step [Hor99]. To categorize approaches that are directed from the system towards the user, we ask three questions:

- *What* is the analysis objective?
- *How much* guidance is provided to reach the objective? What is the degree of the provided guidance?
- *What* type of suggestion is provided?

To answer the first question, we exploit the knowledge generation model of VA [SSS*14b]. This model gives an overview of the high-level procedures of a visual data analysis process. The user might need help at different points of this process. Thus, analysis objectives might be, for instance, supporting *data transformation*, *model visualization*, or *model building*. To answer the second and the third question, we use the guidance dimensions. In particular, we use the guidance degree [CGM*17] to describe how much assistance is provided. We further describe the visual dimensions used to communicate the guidance [CM84], to illustrate what type of guidance answer is provided and how it is communicated to the user. In this way, we aim to describe not only the *degree* of guidance, but also to the *type* of output the system produces to guide the analysis. These could be for instance simple suggestions, ordered options, or step-by-step instructions.

$SYSTEM \leftharpoondown USER$ The second type of guidance is directed from the user towards the system, and is characterized by the actions a user can take to influence the system guidance (i.e., the next suggestions or steps). This could also include asking for the system guidance in the first place.

- *How* can the user provide feedback and guidance to the system?

To answer this question, we discriminate how the user guidance can be derived from user's actions, and what is the direction of such guidance. There are two ways for the user to provide information that can be utilized as user guidance: 1) direct actions, and 2) indirect actions. The first group is the most common and includes approaches in which the user is directly responsible for modifying parameters of the guidance mechanism or selecting data elements. Usually, this is achieved by interaction with the user interface (i.e., widgets, sliders, handles, lists etc.). The second group includes approaches in which the user guidance is not directly communicated, but is indirectly and implicitly derived from the actions the user takes in the analysis process. For instance, a user moving data elements to distinct regions of the interface might indirectly signal the intention of grouping these elements, instead of directly choosing a clustering algorithm and modifying its parameters. Approaches that adjust their guidance mechanism in this way assign a semantic to each action (or sequence of actions) the user performs. This is closely related to the notion of *"the user is the loop"* forged by Endert et al. [EFN12, EHR*14]. We will refer to this category as guidance *inference*, as we describe how user guidance is inferred, either through direct or indirect actions.

Besides this, we consider the *direction* of this user guidance. According to cognitive science and psychology [Dow99], the actions a user performs on the data or on the visualization can be used for fine-tuning the guidance provided in the previous analysis loop by the system. In this case, we talk of *feedback* to the guidance provided in the *past*, which can be utilized to fine-tune future guidance. In the other case, the user actively asks for what she/he would like to see (e.g., with sketches), or what should be expected from future guidance loops. In this case, we refer to those actions as *feedforward* or actions performed to call for *future* guidance.

The guidance direction can be further grouped into: *positive and negative feedback* and *positive and negative feedforward*, as the user might evaluate (positively or negatively) the guidance provided by the system in a previous analysis loop, or she/he might use positive or negative examples of what the guidance results should look like or should not look like.

Thus, we categorize approaches according to the way the guidance is fine-tuned: either by feedback to previous guidance or by feedforward calls for guidance. In this manner, we are able to provide a description of how guidance is provided to the system.

3.2. Axes of the Categorization

In this section, we describe our categorization scheme, which also structures the remainder of this paper. For a visual representation of this categorization, we refer to Figure 2. From a high-level perspective, we distinguish between two general processes: user and system guidance. We do not describe these two processes in a completely symmetrical manner. This is due to the fact that some aspects, such as the guidance objective, are the same for user and system guidance within one data analysis cycle. Thus, we describe this objective only once for the system guidance. Although multiple goals can be pursued, existing approaches generally support one goal at the time.

Our categorization scheme is structured as follows: on the highest level we distinguish between system and user guidance. For each group we consider two categorization axes (i.e., four categorization axes in total): for system guidance we outline 1) The **analysis objective** to describe what is the final goal of the guidance process and 2) the **guidance degree** to assess how much assistance is provided to the user. This also includes a description of the **type of guidance**, and the means the guidance process utilizes to provide it. For user guidance we consider 3) the **guidance inference** which describes how user guidance is derived: this could be either from *direct* or *indirect* actions, as well as 4) the **guidance direction** which specifies if the user guidance is provided by evaluating the past or by giving directions for the future. We call these two directions *feedback* and *feedfoward*. In summary, with these four axes we aim to give a comprehensive overview of *how* guidance approaches support a data analysis process, and in general *how* the human-computer collaboration is pursued.

We will use these categories also in Table 1: at this regard we report in bold the suffixes of the categories as used in the table (e.g., **Map** for Mapping, **Trans** for data transformation etc.). The categorization scheme is defined by the following description:

- **System guidance**
  The system provides guidance to the user.
  - **Analysis Objective**. We distinguish between different analysis objective that are supported by the guidance system:
    - **Data**: approaches that provide guidance to data **trans**formation activities. We further stress if the guidance is aiming to support *single operations* (e.g., the imputation of missing values) or the *whole manipulation process* (e.g., compose a data preprocessing pipeline).
    - **Visualization**: approaches that guide the *visual **mapping** of data* or ***mod**el **v**isualization*.
    - **Model**: approaches that provide guidance for ***mod**el **building*** or ***par**ameter refinement* activities.
    - **Exploration**: approaches that guide data **explo**ration by supporting the *discovery of findings* or suggesting *user actions*.
    - **Verification and knowledge generation**: approaches that provide guidance to the generation of new **know**ledge from raw findings (i.e., *guidance for the management of findings and insights*).
  - **Guidance degree and guidance type**.
    We further categorize approaches according to what extent and what type of guidance they provide:
    - **Prescribing guidance**: approaches that support a given analysis task by ***pres**cribing step-by-step actions*.
    - **Directing guidance**: approaches that provide *ranked recommendations* to steer and **dir**ect the analysis process.
    - **Orienting guidance**: approaches that aim at improving the user's **ori**entation. This is achieved with the use of *visual hints* and the provision of *unranked suggestions*.
- **User Guidance**
  The user provides guidance to the system. We distinguish ways in which the user is able to influence and steer the guidance process or ask for guidance in the first place. We do this by focusing on how user guidance is derived from the user's actions and on the direction of such guidance.
  - **Guidance Inference**
    - **Direct actions**: approaches in which ***dir**ect actions* are the main way for the user to steer and provide guidance to the system. We further distinguish if *single actions* are taken into consideration or the whole *actions history* is taken into account.
    - **Indirect actions**: approaches that allow users to steer the analysis through ***indir**ect actions*. User intentions and needs are discerned from user actions such as manipulating data elements or rearranging the view.
  - **Guidance Direction**
    - **Feedback**: approaches that allow the user to steer the analysis by evaluating the guidance and the results provided by the system in the previous analysis loop (i.e., giving positive or negative feedback). Thus, this kind of guidance is directed **back**wards, towards the *past*.
    - **Feedforward:** approaches that allow the user to steer the analysis by proactively suggesting what he/she wants (or does not want) to see as a results in the future analysis steps. This guidance is directed **f**or**w**ar**d**, towards the *future*.

4. System Guidance to Human Activities

$SYSTEM \longrightarrow USER$ - At first, we describe the main mechanisms by which the a computational system provides guidance to the user.

4.1. Analysis Objective

We start by categorizing the reasons *why* a system provides guidance in the first place. We do so by investigating the analysis objective.

Data. Approaches falling in this category provide guidance to preprocessing operations that operate directly on the data, for instance, data wrangling and data cleansing [KHP*11]. Although the literature covering this first elaboration step is vast, just a few works contemplate guidance. This is usually achieved by means of recommendations and prediction of appropriate algorithms, parameters and visualizations [HHK15]. Most of these works stem from the initial ideas of Kandel et al. [KPHH11] and are nowadays pursued in the context of Trifacta [Tri].

Data Wrangling and Cleansing - On the subject of data wrangling, Kandel et al. proposed Wrangler [KPHH11], a visual interactive tool to support data transformation. Aside the visual design considerations, some aspects of their tool relate to guidance. In particular, Wrangler is able to guide the selection of appropriate data transformations, based on the data type and by matching the current data with a shared database of data transformations. Furthermore, in line with the guidance objectives, the provided transformations are not executed automatically, but the user is left the possibility to modify them according to the specific scenario. These modifications are then used to adapt the generation of future recommendations. On the other hand, Kandel et al. [KPP*12] support data cleansing. Data cleansing is often a semi-automatic activity, usually based on algorithms exploiting different metrics for determining data quality problems and user actions to consider the different quality issues in the right context. To support this task, Kandel et al. focus on providing suggestions of proper visualizations to compare quality metrics and the corresponding data values. Finally,

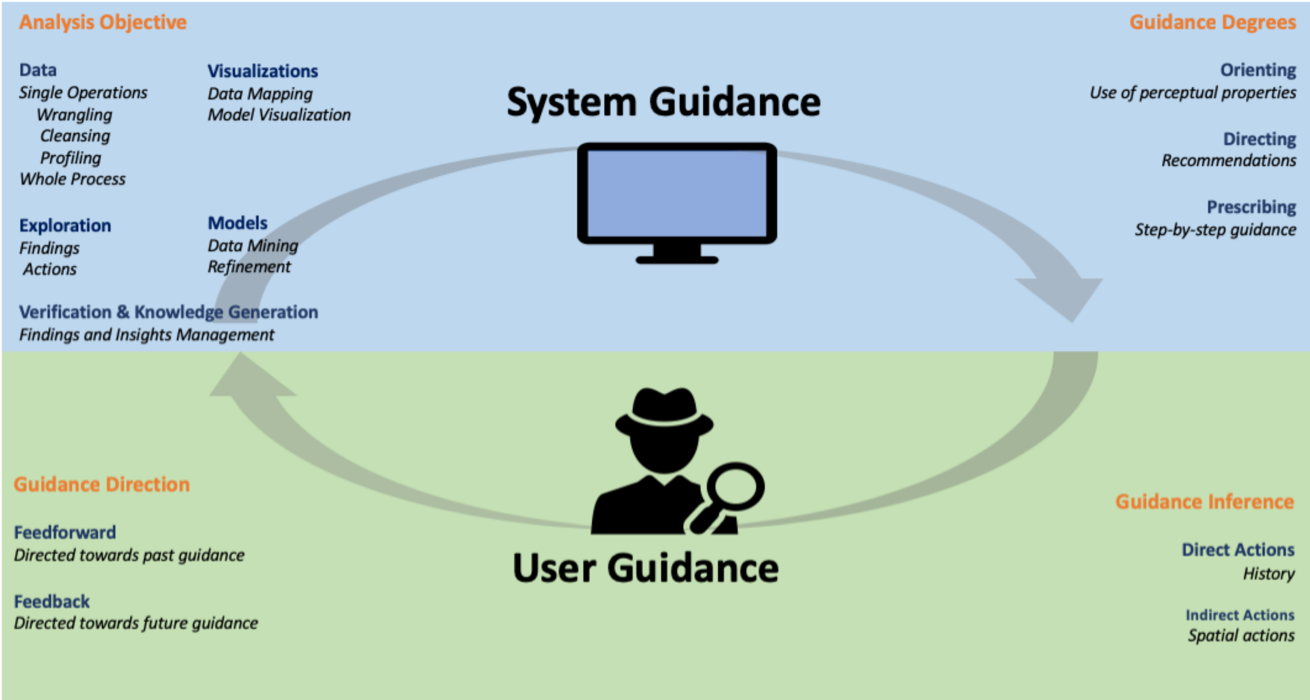


Figure 2: *Schematic representation of the categorization axes we used in this survey. Two main processes, user guidance and system guidance are represented. For each process we describe two categorization axes including multiple sub-categories.*

May et al. [MBD*11] present Smart-Stripes a tool that provides the users the possibility to steer the process of feature selection. Feature subset selection is a procedure that usually is done before data analysis, to extract the most important features from a large multidimensional dataset. In their work, automated methods and user-interaction are intertwined so to open the algorithmic black box and provide the user with an informative overview of the most interesting features. This is achieved, by decomposing the different measures characterizing the data features and relating them to precise data subsets, showing to the user the overall influence of precise portions of data on the overall measures, and thus on the resulting features.

Data Preprocessing - While the previous approaches focus on single tasks, Bernard et al. [BRG*12] instead focus on the whole data preprocessing process by providing guidance to compose the steps and procedures necessary to have usable data for analysis (see Figure 3). In particular, they aim at integrating domain knowledge and metrics to guide the imputation of parameters and the selection of appropriate values for the single processing steps. This is achieved by showing promising parameters, but also the possible effects of the single choices on the overall result. Focused on the overall preprocessing procedures is also the work by Heer et al. [HHK15], which constitutes a good starting point for a better human-computer collaboration in this area. In fact, they do not propose a solution to a specific problem in data transformation, but instead they propose a framework for supporting and guiding the user during the whole process. Their idea is to have a so-called *predictive interaction*, in which the system suggests next steps, and the user selects features (that will influence the generation of future suggestions) and chooses among the system suggestions.

Visualization. In this category we outline tools and approaches that aim to support either the mapping of data to visual forms or the visualization of data models. However, just a few guidance techniques are devoted to support the latter scenario.

Visual Mapping - Fujishiro et al. [FTIN97] developed GADGET, a system that presents the user either possible additions to existing visualizations and complete visual mappings for the user's convenience. The suggestions are based on the data and on task descriptors as well as on the similarity of the current visual mapping to a database of example mappings. The interaction and the choices of other users indirectly influence the provision of new suggestions. Bertini et al. [BS06] designed an approach to support the user in visualizing over-plotted areas and improve the overall image quality. This is achieved by algorithms and metrics that measure the degree of overlapping data. Through these metrics, the visualization is modified and the over-plotted areas are sampled preserving the most important data features. Koop et al. [KSC*08] present VisComplete, a system that aids users in the process of creating visualizations by using a database of previously created visualization pipelines. The system learns common design paths, and according to the current user input, it suggests visual additions. Gotz et al. [GLK*10] describe a behavior-driven system suggesting the user a set of visualizations that should be effective for a given inferred analytical task. This work is based on a previous study, in which the authors show the relation between tasks

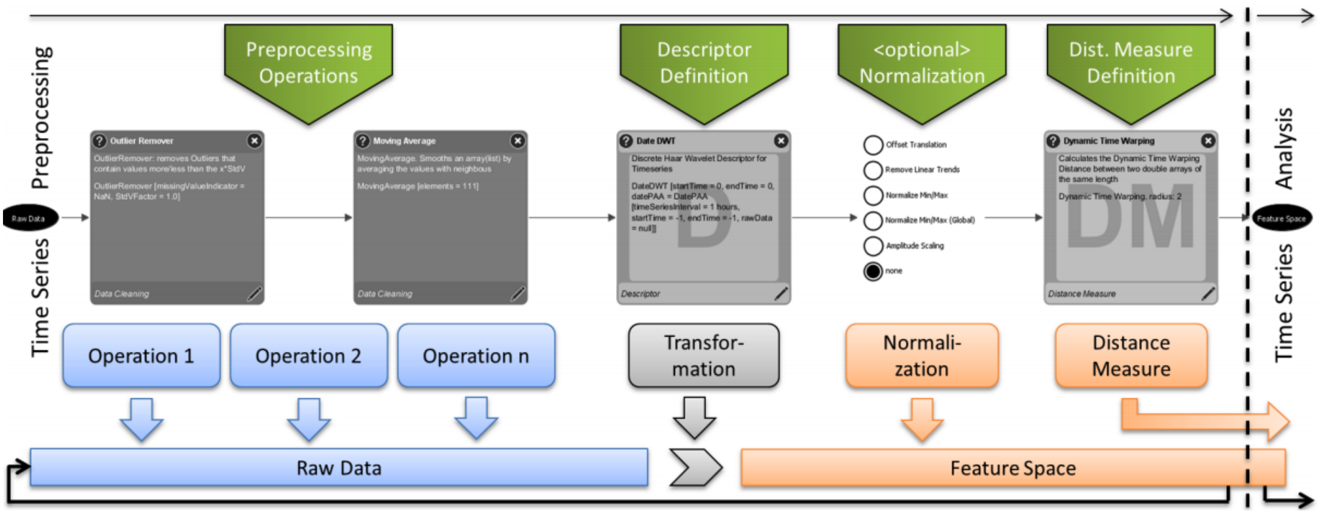


Figure 3: *Jürgen Bernard [Ber15] provide directing guidance to the data preprocessing step. The approach aims to support the user in composing and input suitable parameters for data manipulation processes. User guidance is derived from user's direct interaction with the tool. This feedback is then used to refine the system guidance.*

and user interactions [GW09]. In a similar manner, O'Donovan et al. [OAH15] present DesignScape, a system proposing a set of ordered suggestions to improve the current visual design. Two distinct types of suggestions are available: refinement suggestions, which improve the current design, and brainstorming suggestions, which change the style. Bouali et al. [BGV16] designed a system providing suggestions of proper visual mappings to the user. The user can choose and select the most promising one and provide weights of the most appropriate data columns to be included in the final visualization. On the base of these interactions, the guidance algorithm refines the recommendations. Finally, Wongsuphasawat et al. [WMA*16, WQM*17] describe Voyager, a system featuring a recommendation engine capable of suggesting effective visual mapping, considering both the current data selection, as well as expressiveness criteria (see Figure 4).

Model Visualization - Zheng et al. [ZAM11] present a tool that allows the user to visualize the model of a given 3D object. The system suggests informative views based on the results of a clustering algorithm. The views are sorted and the choice of a suggestion triggers the recalculation of the rendered scene and the calculation of new suggestions. Ankerst et al. [AEK00] in their work propose some useful hints for model visualization too. Although mainly focused on model building, the system they describe supports also the visualization of the model. Guidance is provided by proposing change of visualizations (e.g., expand a tree node), and by offering a look-ahead function which is useful for the user to understand how the model will look like in the future.

Models. Approaches in this category deal with supporting the creation and optimization of data models. This is usually achieved by providing the users with the most promising algorithms and by guiding the selection of proper parameters.

Data Mining - Most of the literature on this topic is built around classification and clustering algorithms augmented with visual means. They differ mainly for the kind of algorithms involved, the application scenario, and for the use of specific visual means to achieve their goal. Bernard et al. [BDV*17] devised a tool for supervising the labeling of human motion data. This is achieved by the provision of suggestions of viable candidates for labeling (see Figure 5).

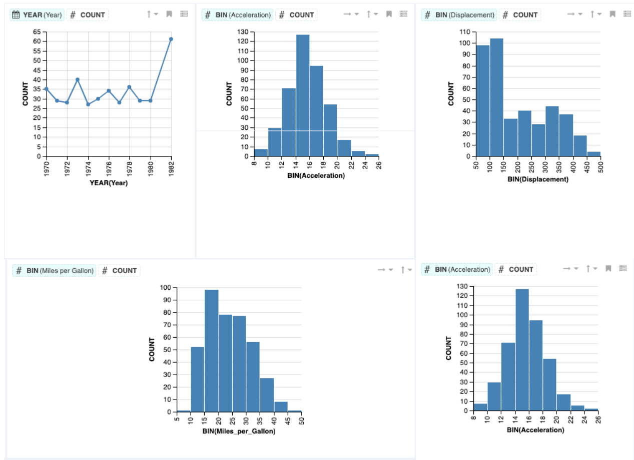


Figure 4: *Wongsuphasawat et al., [WMA*16] provide directing guidance to the visual mapping step. The approach aims to provides the user with recommendations of suitable visual mappings. Recommendations are ranked based on some objective visual criteria as well as thanks to user's direct feedback*

Choo et al. [CLKP10] describe a system that interactively helps the user to classify data. Through the use of dimension reduction algorithms, the computed clusters are visualized as scatter plots. Subsequently, the user is able to modify the initial classification thanks to similarity and distance metrics, showing how the initial clusters relate among each other. Through such metrics, the user is allowed to modify and steer the classification process. The results of this interaction cycle will be then incorporated and used for the calculation of the next clusters. In a similar way, Garg et al. [GRM10] use hidden markov models to segment data. By providing her with a semantic interpretation of the clusters, the user is then able to refine the initial segments. Migut et al. [MW10, MvGW11] apply a similar methodology to a risk-assessment scenario. Patients with psychiatric diseases are classified by intertwining algorithms and user feedback. Drucker et al. [DFB11] designed a system supporting the creation of a data model by proposing the most prominent elements to be added to the different clusters. The recommendations are based on a machine learning model that adapts over time, making the suggestions dynamic.

Ankerst et al. [AEK00] proposed a supervised tool for building decision trees combining the computational power of the system and the knowledge of the user. This is one of the few examples where the model building procedure (the black box) is opened and split into separated steps to allow fine tuning of operations. The user is also able to steer the process, and intervene at each elaboration step. On the same line, Endert et al. [EFN12] allow the user to fine tune and steer model building activities, by representing data on the 2D plane. The system supports the user by searching for similar related data entities, that are displayed and positioned together. The user can directly move data points around to alter the clusters and influence the discovery of similar data.

Parameter Refinement - Müller et al. [MAK*08] present Morpheus, a tool supporting the visualization and interactive exploration of subspace clusterings. In particular, it helps users to interactively choose the best parameter setting. By presenting different visualizations representing the results of different parameter choices, users may choose the best ones that generates the desired subspace clustering. Dörk et al. [DLB13] emphasize and visually highlight parameter options leading to relevant views based on their popularity, timeliness, and availability. To foster a better parameter selection, Jeong et al. [JZF*09] describe a tool that shows the relations between the data and the output of a principal-component analysis algorithm. This is achieved by highlighting the effects that each data column has on the final result.

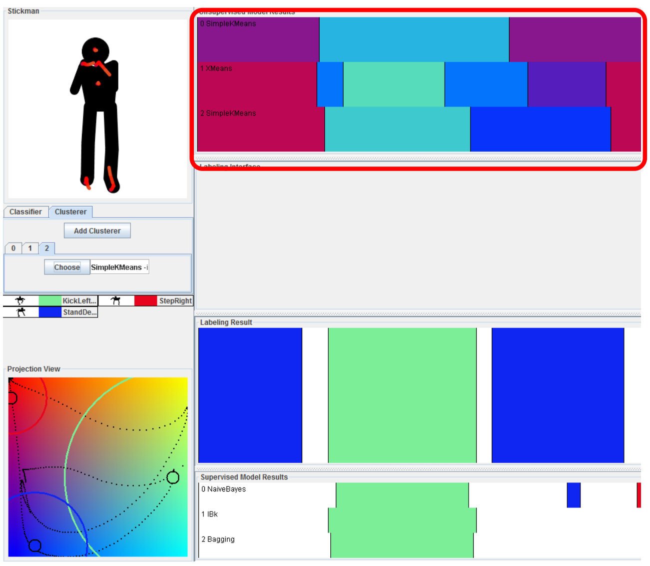


Figure 5: *Bernard et al. [BDV*17] provide orienting guidance to model building activities. Through the use of unsupervised and supervised methodologies, the approach is able to provide the user with suitable suggestions of labels (see red box in the figure) for different captures of human motion data.*

Exploration. Exploratory analysis combines the previous analysis steps (i.e., data preprocessing, data visualization, data models and algorithms) to achieve a higher goal, usually a complex task. The primary aim of providing guidance during data exploration is to support the user interaction and the discovery of new findings. A finding is an interesting observation made by an analyst [SSS*14b], and it usually refers to the disclosure of a pattern, or some interesting data subset.

Findings - The majority of the works providing exploration guidance, achieve their goal by pointing the user to interesting data and data structures. Heer and Boyd. [HB05] describe Vizster, a visualization system supporting the exploration and navigation of large social networks with the goal of finding communities. The system idientifies and highlights the such communities and provides the user with means means to browse them. Johansson et al. [JLJC05] show clusters using parallel coordinates and apply a number of visual techniques (i.e., highlighting, grouping, coloring, applying textures) that support an efficient analysis of the structure within these clusters.

To support exploration of weather phenomena, Steed et al. [SSJKF09] designed a system that displays the correlation among environmental variables and the underlying data, so that users can understand them better but also be able to use them more efficiently to predict weather events such as hurricanes. Adler et al. [ASM*10] support visual navigation in surgical operations by augmenting the visualization environment with patient-specific anatomical data. The user/surgeon is enabled to set and change the most appropriate visual target as the exploration evolves. By using a flexible degree of interest function, Alsakran et al. [ACZ*11] show the user interesting relations among a set of streaming textual data. The user is allowed, at anytime, to modify the interest measure and influence the layout of incoming nodes. Ip et al. [IV11] present a system that helps the user to identify salient patterns and interesting areas in very large images (e.g., landscapes). This is achieved by means of a saliency measure, that serves to identify interesting areas for user exploration. Domino [GGL*14] supports the exploration of multiple datasets. It supports the user by providing hints to arrange, combine, and extract subsets of different datasets. In contrast to that, Stratomex [LSS*12] is focused on the exploration of relations in cancer sub-type datasets. The system displays ribbons between data columns to highlight relations among data features. Scorpion [WM13] is a tool that supports the exploration of data outliers, by pointing users to the possible data tuples from which these outliers originated. Finally, Bernard et al. [BSW*14] developed a system that emphasizes the most interesting relations among data subsets, thus helping the user gaining an overview of the dataset. Gladish et al. [GST13] developed an approach that, by using a flexible degree-of-interest measure, is able to show interesting data regions to explore during the analysis of hierarchical data (see Figure 6).

Actions - A few works deal with guiding the exploration by supporting directly the interaction. This is usually translated into suggesting the next actions to take. One of the first attempts in this direction is Systematic-yet-flexible system by Perer and Schneiderman [PS08]. Unexplored states are shown to the user so he/she can systematically explore all the dataset. In the context of personalized learning, Krishnamoorthy et al. [KB06] present the user a set of personalized suggestions about the next documents to explore. Streit et al. [SSL*12] designed a model to steer exploratory analysis. Based on data features and task descriptors the system shows the users the data to explore and the next (alternative) steps to take in order to complete a given task. The approach developed by May et al. [MSDK12] presents signpost pointing at interesting regions of the graph, thus informing the user about the possible next steps to take. In a similar way, Crnovrsanin et al. [CLWM11], upon selection of a node in a graph, recommend the user a set of interesting actions to perform to reach interesting nodes.

Verification and Knowledge Generation. While the previous works focus on findings, the following ones deal with arranging those findings into valuable insights and new knowledge.

Yang et al. [YXRW07] designed an approach for managing discoveries in visual analysis. The system supports the organization of facts and findings by suggesting clustering of a given discovery based on semantic similarity. Shrinivasan et al. [SGL09] present an approach for helping the user in the activity of connecting the dots. Based on the current line of inquiry, the system suggests findings, notes and concepts and how to arrange them together. Chen and Scott [CBY10] developed an approach for semi-automated annotation to support insight externalization activities and the reconstruction of the process that lead to this insight. The work by Hossain et al. [HARN11] guides the user through the process of arranging facts from a collection of documents with the aim of creating a story. The system provides the user a set of paths (stories) connecting an initial and a final document. The user can navigate and explore the suggestions and also adjust the provided stories.

4.2. Guidance Degrees and Guidance Type

In the past section, we described the objectives, i.e., the reasons why a system provides guidance. In the following, we focus on how much guidance is provided as well as the type of assistance. Ceneda et al. [CGM*18] describe three guidance degrees: prescribing, directing, and orienting guidance. The first two degrees, prescribing and directing, provide a high level of assistance, with the former focused on the choice of the one most appropriate way to reach the results, and the latter describing the mechanisms to calculate a wide set of options to continue the analysis. In simple words, prescribing and directing guidance are focused on the provision of suggestions. Orienting guidance, on the other hand, plays its role at a lower subtle level, exploiting the user's perceptual abilities to provide him/her with a set of visual hints to foster the analysis. In summary, orienting guidance is more focused on providing means to understand the answer to his/her problem, instead of providing a ready-made answer.

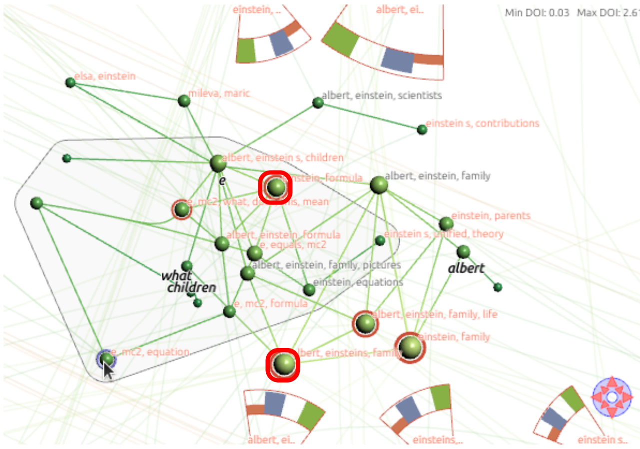


Figure 6: *Gladisch et al., [GST13] provide orienting guidance for the exploration of large graphs. The guidance solution, thanks to a flexible degree-of-interest measure, helps the user in identifying interesting areas of the graph to explore. For instance, in the image, interesting nodes (highlighted in red) can be expanded to reach the desired analysis target.*

Prescribing. One way of providing guidance is directing the user along a promising analysis path. The system, given the analysis context (i.e., the current data and tasks) computes the best way, also in terms of efficiency, to reach directly a satisfactory conclusion of the analysis [CGM*18].

Horvitz et al. [HBH*98] describe the design and the implementation of a system to provide assistance to software users. The system exploits bayesian user modeling to transform interaction into useful hints of user's intentions. The system is able to infer the different phases of the analysis, the tasks and user needs, and subsequently provide suggestions to continue the analysis and pursue an inferred task. Chen and Scott [CBY10] automatically calculate annotations of data snippets selected by the user. The user can directly modify the annotation which again affects the generation of future annotations. Ip et al. [IV11] guide the user through the visualization of large images by calculating and providing a step-by-step exploration of the most promising and interesting views.

Directing. The name suggests its main purpose: directing the analysis. In fact, this guidance degree aims at solving the user's knowledge gap by presenting a set of alternative options for further analysis. The given suggestions/options could differ in terms of quality and costs for different paths leading to the same result or in terms of interest for paths leading to similar or new targets. When compared to prescribing guidance, the options provided by this guidance degree are higher in number and differ in quality. For instance, an analytical system, based on an interestingness indicator, may suggest the user a set of alternative data cases that may be useful for the analysis, or provide a set of alternative interaction steps. Although it is clear that the freedom of the user is higher (given the higher number of options provided), this guidance degree may also introduce a certain level of uncertainty, and provide suggestions not directly related to the tasks in focus.

Recommendations. Directing guidance is tied strongly to recommender systems. Recommendations assume different forms according to the different goals they are created for, and to the analysis step they should support. To support data transformation, the system simply suggests to the user the most suitable functions to modify the data [KPHH11], clean and polish the data [KPP*12], or to support feature selection for data profiling [MBD*11]. The same happens for data visualization where steering guidance translates into systems that provide suggestions of different visualization alternatives, usually ordered on the basis of specific perceptual characteristics [WMA*16, BGV16, FTIN97]. To support data modeling, steering guidance usually provides the user with different algorithms and parameters [DFB11, AEK00]. However, this guidance degree assumes a particular interest when it allows the steering of the whole exploration process, pointing users to interesting findings [JLJC05, MSDK12, SSL*12].

Orienting. Maintaining user's mental map and orientation is a fundamental goal of any visualization tool. This importance has been recognized in various studies [PHG07, AP13]. With the term *orientation*, we refer to the structural cognitive information a user creates internally by observing an image, which represents the user's underlying understanding of the information. Hence, sustaining context comprehension and improving the user's orientation during the analysis influences the user's perception of the dataset and of the tasks. Orienting guidance can be provided to the user by exploiting low level information extracted from the dataset, in this case this information is mapped to basic perceptual properties to guide the user, or by exploiting users' interaction to support the analysis by providing suggestions (see Figure 7). Differently from directing guidance, the suggestions do not have a clear order, or priority. In the following, we distinguish the different orienting approaches according to the visual properties used to support guidance and the kind of suggestions they provide. The two groups are not mutually exclusive, but refer to different aspects of the same process.

Highlighting and removals. These approaches play with the preattentive skills of human perception to provide guidance. Contrasting the color hue and intensity of important elements, with those of the surroundings allow the users to quickly and preattentively identify them, without the need for a longer sequential search [Maz09, SKMW17].

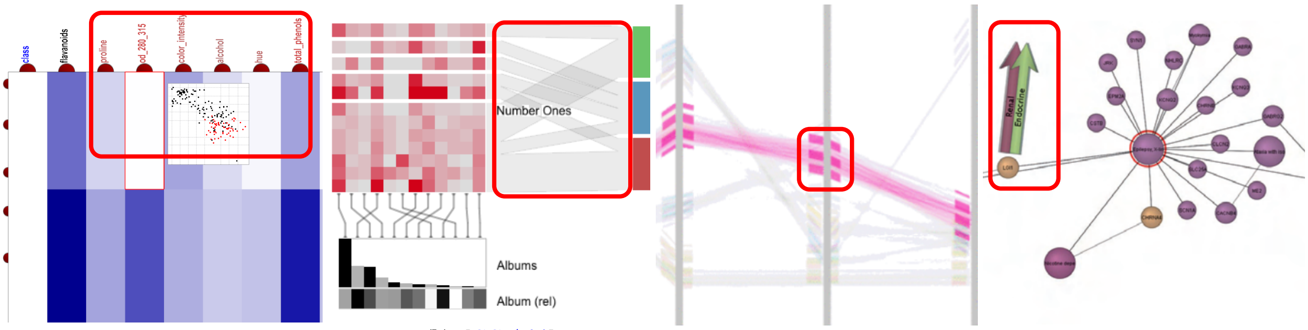
Vizster [HB05] signals the presence of communities in social networks by color change. Nodes and links of such communities are highlighted and color is used to encode distance. To guide feature subset selection, May et al. [MBD*11] color code interesting data columns. Color change is also used to relate selected data features to the overall quality measures, highlighting in this way causal relationships. Ankerst et al. [AEK00] present a visual technique for building decision trees. Possible split attributes and split points are highlighted, so to steer the building process. Similar techniques adopt highlighting for classification. Data points for which a given label has not yet been assigned, or for which the classification is uncertain, are presented in a different color [GRM10, MW10].

Layout and Form. 2D position, spatial grouping, and marks are properties that our eyes perceive faster and that attract our attention [CAS*18]. Guidance approaches use, for instance, links to signal relations or the positioning of elements to suggest the user that a (hopefully) better layout can be obtained. Usually this is achieved by means of (user-defined) metrics that express the user's intentions and goals. Closeness is often used to signal the belonging of a certain point to a cluster [MAK*08]. However, when uncertainty is involved, it might still not be obvious to the user which cluster to choose. To support this task, Choo et al. [CLKP10] visualize links among the data element and the most appropriate cluster. Similarly, but at a higher level of abstraction, Stratomex and Domino [LSS*12, GGL*14] are two approaches that present the users relations between different datasets using glyphs (e.g., ribbons). Building a story is important to flawlessly compose connections among facts and events. Thus, guiding story building activities is the aim of the approach developed by Hossain et al. [HARN11]. In particular, relations among documents are shown for the user's convenience.

Motion. Flicker and motion are also important preattentive visual features. They are very useful to quickly attract the user's attention. This is why they are frequently used in our daily life, for instance in commercials and in traffic lights. However, we could find just one approach exploiting such visual properties for guidance. Johansson et al. [JLJC05] use animation and textures to bring the attention of the user to important characteristics of the data. Aiming to show clusters in a high density parallel coordinates plot, they animate different lines with differing phase velocities to emphasize the skewness or the variance of data clusters.

Suggestions. In general, providing orienting guidance is achieved by presenting suggestions. The system utilizes a complex combination of the expedients described above (i.e., highlighting, glyphs, etc.) to provide analytical options, so that the user can proceed towards her/his goal. Differently from directing guidance, these suggestions have equal weights (i.e., are not sorted according to importance) and therefore the resulting guidance degree is considered lower.

Using a flexible degree-of-interest function the system developed by May et al. [MSDK12] is able to produce recommendations regarding data subsets that are worth investigating. In particular, the system supports orientation by pointing the user to graph regions outside the active exploration area by means of visual glyphs, as well as a possible shortest path to reach that region. Luboschik et al. [LMS*12] ease the exploration of multiscale data. The approach points the user to scales and regions within the data that exhibit behavior of interest without the need for an exhaustive search. The system aggregates the data of the finest granularity into more coarse-grained data. Consecutive data scales are compared, the important data characteristics of the fine-grained data are preserved and visualized in the overview, so that the user can have a view of the most important characteristics without the necessity of panning and zooming operations. Jiang et al. [JN15] developed a method to support the creation of queries. The system calculates the relevance of the query parameters with respect to what a user is doing (interaction) and highlights the most prominent values for those parameters.



(a) *[MBD*11]* **(b)** *[GGL*14]* **(c)** *[JLJC05]* **(d)** *[MSDK12]*

Figure 7: *Orienting guidance can be provided by encoding the information the user needs by using different perceptual properties. Highlighting (a) can be used to make visible information that could have some interest for the user, for instance data columns (highlighted in red). Changing the layout and using forms for different data subsets may stimulate the user to explore them. For instance, in Figure (b), different data subsets are connected with visual ribbons (highlighted in red in figure b) to signal interesting relationships. Motion can be a way to convey guidance too. In Figure (c), motion is used to signal that the data under analysis has a specific characteristic. The reader can imagine the pink line moving up and down inside the rectangle highlighted in red. In general, the end-goal of orienting guidance is provide suitable suggestions to proceed the analysis. In figure (d), the system suggests interesting graph regions to explore next (the arrow highlighted in red).*

Minor Guidance. We dedicate a separate paragraph to the description of those borderline approaches that comprise very small elements of guidance. Usually, these approaches cannot be considered regular guidance techniques, but still they utilize a design rational that is interesting from a guidance perspective. These approaches are signaled by an empty circle in Table 1.

Undo/redo of actions and history of visualizations are common practice [MYIM98,DR01,KNS04]. Usually, those approaches cannot be considered as guidance approaches, since they present almost static visualizations of the interaction history. Guidance, on the other hand, is a dynamic process focused on the future of the analysis. In this context, the approach proposed by Sarvghad and Tory [ST15] differs from a standard history of actions in that it relates the exploration history with data dimensions to enable users to see which data dimensions have been explored in the past and in which combinations. Hence, it encourages the user to proceed in a way that promotes a more complete data exploration.

Usually, a static visualization of a model is not considered guidance. The approach by Krause et al. [KPN16], however, is different in that it provides guidance to explore the output of predicting algorithms by showing relations between the output model and the data features that influenced it. Users cannot only understand why certain results are predicted, but also see how the predictive model responds to modifications of the data itself, which also facilitates parameter refinement.

5. User Guidance

SYSTEM $\longleftarrow$ *USER* - In this section, we review the ways guidance is provided by the user to the system. User guidance may serve, for instance, to close the guidance loop after system guidance has been provided, or to initiate guidance in the first place. Theoretically, as a consequence of user guidance, a system should provide further additional visual cues to the user, so to acknowledge a change in the analysis course (due to the received input), and initiate (once again) the guidance process. In the following, we describe the direction of user guidance and how it is derived from the user's actions.

5.1. Guidance Direction: Feedback and Feedforward

We discern approaches that allow the user to foster the analysis by either evaluating the results and the guidance the system has produced in the *past* and those that allow the user to input directly what results or what kind of guidance suggestions he/she expects to see in the *future*. Following the terminology used in the cognitive science, we call this *feedback*, in the first case, and *feedforward* actions, in the latter case. We further specify and the quality of such actions: they can be *positive* and *negative*. Thus, we end up with four combinations: **Positive and negative feedback** and **positive and negative feedforward**. Ideally, feedback and feedforward action sequences should alternate: A feedforward action may stimulate the system to produce specific results that are subsequently evaluated by user feedback and so on.

Positive and Negative Feedback. Most of the works we analyzed enable the user to provide feedback (either positive and negative), so in the following we will report just a few example approaches to show how those approaches implement the feedback loop. For a complete overview, refer to Table 1.

To guide view selection in volume visualization, Zheng et al. [ZAM11] propose an approach that suggests optimal viewpoints. The user is enabled to provide positive feedback to these suggestions by selecting the most promising ones. As a reaction, the system updates the suggestions, pointing to new and promising but so far unseen view directions. Fujishiro et al. [FTIN97] designed a system supporting the design of appropriate visualizations. As the user interacts with the tool, the system proposes and suggests additions to the actual design. The user, by selecting the most appropriate additions, guides and provides feedback to the guidance mechanism, influencing future suggestions. A similar approach is proposed by Müller et al. [MAK*08]. The Morpheus system supports the interactive exploration of subspace clustering, by presenting suitable results. Based on the discovered knowledge, the user can give feedback to the system for improving the suggestions. In this case, the feedback loop enables the user to set parameters and thus to discover meaningful subspace clusters. Andrienko et al. [AAR*09] provide guidance to support the visual clustering of trajectories. In this context, users are enabled to modify the cluster result computed by the system, by excluding one or several subclusters from the cluster itself. Other aspects of this approach relate to the *feedforward* concept, thus we will discuss it in the next paragraph. Stein et al. [SHJ*15, SSS*14a] propose a visual analytics approach supporting the analysis of soccer matches. By extracting the most interesting features from the data, the system is able to propose to the user interesting events that characterize the match. The user can steer the exploration process by confirming or rejecting previously unlabelled events and use them as additional training data for the classifier. Finally, by capturing low-level analytical task results, Click2Annotate [CBY10] supports semi-automated insights annotation. If the user is not satisfied with the annotations generated by the system, he/she can modify and affect the outlook of future annotations by dragging and dropping statistical measures into the annotation itself.

Positive and Negative Feedforward. Another group of approaches enable the user to directly input what he/she wants to obtain from the analysis. Therefore, the following techniques differentiate themselves from the previous in that the user is focused on the future of the analysis instead of the past. This is usually achieved by sketching examples. Sketch-based information retrieval is a very popular field. However, in the context of this review, we just consider the literature related to guidance techniques.

Chegini et al. [CSG*18] developed a system supporting the visual exploration of patterns in large scatter-plot matrices. Usually, the analysis space is huge, so to reduce users' effort, the system recommends suitable patterns for close-up investigations. On the other hand, the user is enabled to actively input what he/she is currently looking for in the data: the user can directly draw or select patterns representing the searched output. Andrienko et al. [AAR*09] in their work support clustering of trajectories. They allow users to provide feedforward actions. Users are in fact enabled to split, combine, and create their own clusters, thus suggesting directly to the system how the clustering algorithm should categorize the data. The work by Janetzko et al. [SSS*14a] comprises elements that can be related to feedforward actions. In fact they allow the user to steer the exploration of semantically meaningful soccer events by integrating the possibility for the user to sketch and describe dangerous situations that should be taken into consideration. Migut et al. [MvGW11] guide the classification of psychiatric patients. The user can steer the model building process by indicating to the system prototypes of patients they are interested in. The iCluster system [DFB11] helps the user to cluster large document collections by providing recommendations. The system learns and subsequenty adjusts the suggestions as the user interacts with the tool showing to the system how he/she would like to organize the documents. The approach conceived by O' Donovan et al. [OAH15] aims to guide the design of visual layouts. Users can specify their own intents in form of constraints, and by sketching partial layouts, hence, steering the guidance process.

5.2. Guidance Inference

While in the previous section, we analyzed the direction of user guidance, in this section we describe in what way the user can convey such guidance to the system. The most common way to infer user guidance, is by taking the user's *direct actions* with the interface widgets into consideration, for instance, using drop-down lists, buttons, check-boxes, etc. This comprises, for instance, considering the direct input of weights for mining algorithms or the selection of visual parameters for a visual mapping, but also annotating data for insight generation.

Different taxonomies are available to discern how a user can provide input [AES05, YaKS*07, Shn96, AAB*11]. Existing taxonomies on this topic are mainly focused on assigning a meaning to certain actions, with the aim to understand the user's intents. However, in the context of this survey, we analyze how this interaction affects the guidance process. The other way to derive guidance is by considering *indirect actions*. We chose this focus in line with the notion of *"user is the loop"* forged by Endert et al. [EHR*14]. They move the focus from approaches exploiting direct actions towards the creation of new approaches in which the user does not simply take part in the analysis process, but is part of it. This new concept fosters the creation of more immersive tools, that directly learn from user interactions, instead of waiting for direct user input.

Direct Actions. This is the most common way of providing input and providing guidance. We analyze tools and techniques that offer direct user input through interface widgets. We also take into consideration the temporal aspects of such interaction, for instance, when exploiting a history of actions to derive the user's intent.

Direct manipulation of parameters - To support data transformation, May et al. [MBD*11] propose a method that highlights interesting data features. On the user side, the direct selection of those interesting features causes the recalculation and updates those measures and metrics, closing the guidance loop. Bernard et al. [BRG*12] support the design of a preprocessing pipelines for time-series data. While the system points to the most promising parameters for each processing step, the user is enabled to steer the process by selecting appropriate weights directly.

Bouali et al. [BGV16] provide guidance for the generation of visualizations. The system proposes a set of suggestions, that the user can choose from. On the other hand, the selection of the most appropriate visual mapping provides an input for the creation of the next visualization generation. Similarly, in DesignScape [OAH15] previews of design suggestions are shown to the user, who can select the most promising one. Many other approaches exploit direct user feedback to generate visual mappings [WMA*16, KSC*08, GLK*10].

Ankerst et al. [AEK00] support the generation of decision trees. The direct selection of the next split points steers the tree construction process. Choo et al. [CLKP10] propose a system to guide the data classification process. Users can directly select an area in the data view containing uncategorized items, and subsequently re-run clustering algorithms to optimize the process. Similarly, Garg et al. [GRM10] help the user to associate elements to proper clusters. The user is directly involved in the manipulation of the clustering parameters, to split and join clusters. Other approaches follow a similar strategy to model building [MvGW11, AAR*09, MW10, JZF*09, DFB11, PSHD96].

Similar concepts exist for data exploration and annotation. Domino and Stratomex [GGL*14, LSS*12] support this task by suggesting how different parts of data are related. Subsequently, the user can interact, connect, and arrange data chunks, based on those suggestions. Streit et al. [SSL*12] allow exploration steering by presenting next analysis steps. Based on the given task, the user can subsequently choose the most appropriate analysis direction. Ip et al. [IV11] guide the exploration of large images. By means of direct actions performed on the interface, the user can modify the selection of interesting views and image areas. Shrinivasan et al. [SGL09] help the construction of data stories by structuring the data according to a given start and end document. The user can directly choose among the suggested structures and affect the composition of the story, as well as of the story-line. Many other approaches are present in literature [YXRW07, HB05, JLJC05, SSJKF09, CBY10, HARN11, Fia12].

History - This category comprises those approaches that go beyond capturing single actions. A temporal component is also taken into consideration. Hence, complex action patterns and sequences of interactions are compared within each other to understand the user's analytical intents. The system can exploit such findings to fine tune the guidance, provide better suggestions, and steer the analysis process accordingly.

Gotz et al. [GW09] support visualization creation by taking into consideration complex interaction patterns. The captured patterns are compared with a knowledge base to understand the visual task the user is performing. This, in turn, influences the suggestions of visualizations that best fit the inferred task. Horvitz et al. [HBH*98] support data exploration by modeling the time-varying needs of the users by means of bayesian networks. The suggestions proposed by the system to pursue an analytical goal are influenced by the user actions. Temporal series of actions are interpreted and based on the inferred goal a next step is proposed. Yang et al. [YXRW07] provide guidance to extract valuable information nuggets hidden in the data based on the user's preferences. Also in this case, such interests are inferred directly from the user's action history, and are the base for the retrieval of new information nuggets.

Indirect Actions. Indirect actions involve providing feedback by acting on the data, rather than explicitly stating intentions through the interface widgets.

Spatial Actions - Strictly connected with implicit feedback is the concept of *spatialization*. For instance, a user provides guidance to the system by acting directly on the data. The system, thus, learns and infers weights, parameters, and preferences from the user's actions. In particular, spatialization is derived from how the user modifies the spatial properties of the data (e.g., moving and grouping data).

Endert et al. [EFN12] designed ForceSpire to guide the visual exploration of text documents. To achieve that, this tool modifies the spatialization of data items on a canvas, in such a way that the rendered layout reflects the user notion of similarity among documents. A decisive aspect of this approach is that distance metrics are implicitly suggested by the user through what they call *semantic interactions*. The movement, the juxtaposition of data points (documents) is interpreted, and the weights determining the similarity metrics are implicitly changed.

To the best of our knowledge, ForceFire is the only approach using semantic interaction and the implicit feedback paradigm through spatialization. However, although not directly connected with Endert's ForceSpire [EFN12], other approaches (marked with an empty black circle in Table 1) utilize similar expedients to allow users to signal their intents, and implicitly provide feedback to the system. Garg et al. [GRM10] support model building activities allowing the user to resolve inconsistently categorized data points by moving them to different clusters. Jeong et al. [JZF*09] allow the user to move data items to different clusters to influence the dimension reduction algorithm. Finally, Vizster [HB05] allow layout modification by moving the nodes of a social network graph.

6. Discussion, Future Research and Limitations

In the previous sections, we described the categorization of guidance approaches (Table 1). We illustrate how the system provides guidance to the user to reach a given objective, and how much guidance is provided (Section 4). We gave details of how the user is enabled to guide the system, closing the guidance loop (Section 5). In this section, we discuss and analyze the outcome of our categorization, listing opportunities for future research.

A Comprehensive View: When looking at the summary Table 1, one immediately notices that it presents several white areas. In the table, the different papers (the rows) are ordered according to the supported guidance objective. One result of our review is that guidance approaches are not new to the field of visual data analysis. However, as we can see, most of the approaches we reviewed offer simple solutions to specific contextualized problems, overlooking the more general problem of providing guidance in data analysis. Just recently, the situation has started to change, with approaches that try to offer more comprehensive guidance solutions. However, the analysis process is not yet covered in its entirety. To sum up, although the interest is growing, guidance is still a young research topic in the vast area of data visualization, data analysis, and visual analytics, and substantial research is required to make guidance a widely used and effective technique.

Guidance Objective: If we look at the guidance objective, the trend is represented by approaches providing guidance to single objectives and single tasks. Just a few approaches deal with multiple objectives at the same time, and moreover, some problems are not even sufficiently tackled yet. The least represented category is the one covering approaches that provide guidance to *insight verification and knowledge generation* as well as *model visualization* activities. For the latter objective, we think that this is simply due to the fact that such a task is usually covered by the general data visualization and visual mapping step. For the former (i.e., guidance to support insight verification and knowledge generation), the

	System Guidance										**User Guidance**			
Papers	Guidance Objective							Guidance Deg.			Inference		Direction	
	Transf	Map	Par	ModV	ModB	Expl	Know	Or	Dir	Pre	Dir	Ind	Bck	Fwd
Total: 53	**5**	**7**	**5**	**2**	**12**	**28**	**4**	**35**	**18**	**3**	**51**	**5**	**48**	**5**
[MBD*11]	●							○	●		●		●	
[BRG*12]	●								●		●		●	
[KPHH11]	●								●		●		●	
[KPP*12]	●								●		●		●	
[HHK15]	●								●		●		●	
[BGV16]		●						●			●		●	
[WMA*16]		●							●		●		●	
[FTIN97]		●							●		●		●	
[GLK*10]		●							●		●		●	
[GW09]		●							●		●		●	
[KSC*08]		●						●			●		●	
[OAH15]		●							●		●			●
[MAK*08]			●			●		●			●		●	
[GRM10]			●		●			●			●	○	●	
[MW10]			●		●			●			●		●	
[AAR*09]			●		●			●			●			●
[DLB13]			●			●			●		●		●	
[AEK00]				○	●			●			●		●	
[ZAM11]				●		●			●		●		●	
[CLKP10]					●			●			●		●	
[MvGW11]					●	●		●			●			●
[JZF*09]					●			●			●	○	●	
[EFN12]					●	●			●			●	●	
[DFB11]					●				●		●		●	
[KPN16]					●			●			●		●	
[BDV*17]					●			●			●		●	
[EFN12]					○	●		●				●	●	
[ASM*10]						●			●		●			●
[HB05]						●		●			●	○	●	
[JLJC05]						●		●			●		●	
[YXRW07]						○	●		●	○	●		●	
[IV11]						●		●		●	●		●	
[SSJKF09]						●		●			●		●	
[ACZ*11]						●		●			●		●	
[LSS*12]						●		●			●		●	
[GGL*14]						●		●			●		●	
[HBH*98]						●				●	●		●	
[SSL*12]						●		●	●		●		●	
[PS08]						●		●			●		●	
[BSW*14]						●		●			●		●	
[MSDK12]						●		●			●		●	
[CLWM11]						●		●			●		●	
[GST13]						●		●			●		●	
[TSTR12]						●		●			●		●	
[ST15]						●		●			●		●	
[JN15]						●		●			●		●	
[LMS*12]						●		●			●		●	
[KB06]						●		●			●		●	
[WM13]						●		●			●		●	
[CBY10]							●	●			●		●	
[SGL09]							●		●		●		●	
[HARN11]							●	●			●			●

Table 1: *Table summarizing the classification of guidance papers. Columns represent the different aspects we took into consideration, while papers are listed as rows. The rows are sorted according to the guidance objective they support. We considered approaches providing guidance for different* ***objectives****: approaches supporting data* ***Transf****ormation, visual* ***Map****ping,* ***Par****ameter setting,* ***Mod****el* ***V****isualization,* ***Mod****el* ***B****uilding,* ***Expl****oration and* ***Know****ledge generation. We considered three* ***Guidance deg****rees:* ***Or****ienting,* ***Dir****ecting, and* ***Pre****scribing guidance. We also describe the user side of guidance, in particular the guidance* ***Inference****:* ***Dir****ect and* ***Ind****irect actions. Finally, we provide details of the guidance* ***Directio****ne: whether it is Feed****back*** *or Feedforward. Empty black circles (○) show approaches that offer this functionality but to a minor extent.*

lack of approaches is due to the fact that those tasks are historically a human prerogative, requiring high-level reasoning, where the system can just provide minor support. Still, with the recent advancements in machine learning, there is much room for further improvements. Users could benefit greatly from more guidance tailored towards supporting high-level tasks. On the other extreme, the most supported task is data exploration (see Table 2). This is where there is more space for creativity and research, especially if we consider the large amount of scenarios in which data exploration is usually needed. We can further see that just a small group of techniques offers guidance for more than one objective. One exception is model building activities and parameter setting tasks that are usually supported at the same time, since they are naturally interconnected. However these objectives represent rather an exception. We see a promising research direction in this lack. A comprehensive guidance approach supporting the user along the whole analysis process, i.e., from data manipulation, to knowledge and insights generation, would be a decisive step towards the goal of obtaining effective guidance, however, it is far from being realized.

Deg.	**Guidance Objective**						
	Transf	Map	Par	ModV	ModB	Expl	Know
Or.	1	2	4	1	9	21	2
Dir.	5	5	1	1	2	4	2
Pres.	-	-	-	-	-	3	1

Table 2: *Summary of the papers offering a specific guidance degree in relation to the guidance objective. The numbers in the cells represent the amount of approaches providing a certain degree of guidance (i.e., orienting, directing and prescribing) for a particular task (i.e., data transformation, visual mapping, parameter refinement, model visualization and building, exploration and knowledge generation).*

Guidance Degree: Looking at the guidance degrees, we see a similar situation. The majority of approaches provides basic guidance (i.e., orienting guidance) and uses simple expedients to support the analysis (see Table 2). In this scenario, orienting guidance is mostly provided during data exploration tasks. Directing guidance (i.e., providing alternative options to continue the analysis), is mostly provided during data transformation tasks (i.e., options to manipulate the data are offered) and visual mapping tasks (i.e., alternative visual encodings are provided). It is also easy to see the scarcity of approaches providing prescribing guidance, which may not be a bad thing, since a certain degree of freedom is usually not only required but also recommended during data analysis. We see a chance for future research in the development of further directing guidance approaches, since they provide the user with support but at the same time leaving him/her sufficient freedom to steer the analysis.

Finally, also the number of approaches providing multiple degrees of guidance is limited. We believe that the support of multiple guidance degrees is a fundamental step towards the provision of effective and dynamic guidance solutions, since they would allow for adapting the guidance degree as the as the user becomes more experienced.

User Guidance: Although some approaches consider aspects that can be referred to as implicit interaction means, direct interaction is offered by the vast majority of the considered approaches. This is no surprise, since it is the most straight-forward form of interaction. Just one among the approaches we considered, offers indirect interaction. To make a short recall, this means that information that can be an input to the guidance mechanism (i.e., algorithms and parameters) is derived indirectly, while the user interacts with the data (in contrast to a direct input via sliders or buttons). For instance, a user may signal that a certain data-point belongs to a given cluster just by changing its position in the visualization, instead of changing the clustering parameters. This represents an open area for further research. It is clear that changing interaction strategy may be difficult, however indirect interaction represents also a more natural way of dealing with the data, especially from a guidance point of view, where this may lead to a better user-system *dialogue* and thus to guidance solutions that are better accepted by the users. On a similar note, another open challenge is considering not only a single user action as means to derive the user's intent, but also multiple sequences of actions, thus exploiting the temporal information and context that is provided.

Feedback: Finally, we analyze the types of feedback users can to provide to the guidance system. Most of the techniques provide traditional feedback, which means that users evaluate the results provided by the system in the previous analysis loop. A few approaches allow for feedforward, i.e., what the user would like to see from future analysis steps (see Table 3). In this case, a combination of both types would give users the possibility to better fine-tune the guidance offered by the system, and would add to the communication possibilities between user and system.

Visualization of Guidance: One limitation of our work is that we did not capture completely the way the guidance suggestions are communicated to the user. We attempted to do so providing categories for the different guidance degrees, instantiating an initial categorization, and distinguishing among the different visual means used to communicate the guidance, like highlighting, change of colors, movement, and so on. However, since the visual means used, the tasks to be supported and the guidance degree are strongly interdependent, we could not provide a finer-grain description of this aspect. This aspect should be further elaborated in future research.

Deg.	**Guidance Direction**	
	Feedback	Feedforward
Or.	31	3
Dir.	16	1
Pres.	3	-

Table 3: *Summary of the papers offering a specific guidance degree in relation to the guidance direction. The numbers in the cells represent the amount of approaches providing a certain degree of guidance (i.e., orienting, directing and prescribing) for a particular direction (i.e., feedback and feedforward).*

Black box Approaches: One of the most challenging aspects of the research on guidance is the possibility that it offers to open the so-called black box of mining algorithms. This means that through guidance, users are supported in the process of transforming the algorithmic black box into a white box, allowing the user to steer the analysis at each step of the model building process, fine tuning parameters on-the-fly, thus allowing a better control and in the end obtain better results. In the past, Bertini and Lalanne [BL09] already pointed out that just a few approaches allow the user to steer the analysis process in such a way. At the time, they cited Ankerst et al. [AEK00] as the only approach allowing the user to control the model building process. Unfortunately, we did not find a much better situation today. Aside Ankerst, a few authors published works that try to open the black-box of data analytics [MBD*11, CLKP10, BRG*12]. We think that guidance could be helpful also in this respect. Having the possibility to fine tune at such fine-grain the analysis process will result in improved visual analysis results. However it also presents challenges: not all the operations should be allowed, nor all the possible permitted actions will lead to the desired results. At this regard, the study and development of guidance approaches could ease and facilitate the process.

6.1. An agenda for guidance in visual data analysis

We summarize the findings and contributions of our work by listing a set of qualities and research directions that we consider decisive for the development of comprehensive guidance mechanisms in visual analysis. As mentioned earlier, the expected goal of each guidance approach is to be *effective* [CAS*18, CGM*17]. This means that user and system guidance should concur to reach the established analysis objective. To realize such an end-goal, guidance approaches must be *accepted and trustworthy*. We see a timely provision of guidance as necessary not to interrupt the analysis flow. All the approaches we considered offer guidance tailored towards single objectives. Therefore, providing guidance at the right time is relatively easy since only objective is in focus. If future approaches aim to support multiple objectives, the timing of guidance will be even more important. It may be the case that some tasks can be solved easily by the user without any help, in such situations the provision of guidance may be rather harmful than advisable. In the same way, providing guidance at the wrong time might be also counter productive.

In line with the previous point, guidance should be designed to be *context-aware*. This means that the system must know what is the actual state of the analysis. Moreover, appropriate visual means should be used not to confuse the user. The majority of approaches we analyzed offers a single guidance degree, and with that, simple visualization means. However existing solutions might not be sufficient for more complex tools. Also in this case, the risk is to confuse and distract the user from the what is important, if those visual means are wrongly used in the analysis environment. We do not call for the development of completely new visualization techniques for guidance, but instead for the conscious use of the one already existing, for a seamless integration of the guidance suggestions in the analysis loop. At this regard, the collection of guidelines to design and integrate guidance into visualization solutions should be on the research agenda.

We further argue for making sure that guidance is *controllable and predictable*. Users should be enabled to steer the analysis, turn off the guidance if not needed, and ask themselves for assistance in other cases. In the same way, guidance must be predictable and non-disruptive, in the sense that it should not alter unexpectedly the course of the analysis, and thus preserving the user's mental map. Some of the guidance approaches we describe offer the possibility to steer the guidance process and the course of the analysis at a fine-grained level. However, as already observed black box approaches are still predominant.

We see guidance as a *dynamic* process. It must adapt to different situations, recognize different knowledge-gaps, and consequently adapt the level of intervention. Thus, correctly identifying the user's intent is fundamental for well-suited system guidance. Considering common signals to initiate guidance, like for instance long stall times or absence of interaction, may not be sufficient. A tight collaboration with cognitive and psychology sciences might be key to adequately capture and understand the user's behavior at a deeper lever. One way to formalize tasks and support dynamic guidance is by using taxonomies. Because of their generality, they may be exploited to frame general scenarios in which guidance may be useful. However, due to this generality, it is also not straightforward to apply such taxonomies to specific practical contexts. This limits the design of guidance for a wide set of tasks. Thus, we see potential for further improvement of guidance in the research of multilevel taxonomies.

Finally, especially because this was not considered by any of the approaches we identified, we call for more research to support guidance in *collaborative* scenarios. At this regard, cloud intelligence could be used as an input to the guidance process and facilitate analysis tasks that single users may find hard to solve by themselves. From our point of view, this is an important step towards effective and widely established guidance in modern real-life scenarios.

7. Conclusions

In this paper, we selected, analyzed, and reviewed fifty-three papers dealing with guidance to support visual analysis tasks. We chose to look at such approaches from two complementary perspectives: user and system guidance. We did this in accordance with the characterization of guidance [CGM*18] and to emphasize its *mixed-initiative* nature [Hor99]. Analyzing existing work in this respect, we identified details of how users and systems are able to join their efforts to reach an analysis objective and support the visual data analysis process. We summarized our findings in Table 1 which shows that guidance, although it has its roots in the visual data analysis, is still a young and promising research field, offering many unresolved challenges. In conclusion, users are usually left alone with powerful, yet overwhelming visual analysis techniques. We believe that augmenting these techniques with efficient guidance is a crucial step towards unburdening the user, and thus, leveraging the real potential of such systems. While existing work already tackles some aspects of guidance in visual analysis, much more research is needed to realize truly comprehensive guidance approaches, that exploit state-of-the-art methods of artificial intelligence to constantly monitor the user's intents and knowledge gaps, and efficiently support the analysis goals.

Acknowledegments

This work was supported and funded by the Austrian Science Fund (FWF), grant P31419-N31.

References

[AAB*11] ANDRIENKO G., ANDRIENKO N., BAK P., KEIM D., KISILEVICH S., WROBEL S.: A conceptual framework and taxonomy of techniques for analyzing movement. *Journal of Visual Languages & Computing 22*, 3 (2011), 213–232. 12

[AAR*09] ANDRIENKO G., ANDRIENKO N., RINZIVILLO S., NANNI M., PEDRESCHI D., GIANNOTTI F.: Interactive visual clustering of large collections of trajectories. In *Visual Analytics Science and Technology (VAST). IEEE Symposium on* (2009), IEEE, pp. 3–10. 12, 13, 14

[ACZ*11] ALSAKRAN J., CHEN Y., ZHAO Y., YANG J., LUO D.: Streamit: Dynamic visualization and interactive exploration of text streams. In *Visualization Symposium (PacificVis), IEEE Pacific* (2011), IEEE, pp. 131–138. 8, 14

[AEK00] ANKERST M., ESTER M., KRIEGEL H.-P.: Towards an effective cooperation of the user and the computer for classification. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (2000), ACM, pp. 179–188. 7, 8, 10, 12, 14, 16

[AES05] AMAR R., EAGAN J., STASKO J.: Low-level components of analytic activity in information visualization. In *Information Visualization. INFOVIS. IEEE Symposium on* (2005), IEEE, pp. 111–117. 12

[AP13] ARCHAMBAULT D., PURCHASE H. C.: Mental map preservation helps user orientation in dynamic graphs. In *Graph Drawing* (2013), Springer, pp. 475–486. 10

[ASM*10] ADLER S., SALAH Z., MECKE R., ROSE G., PREIM B.: Overlay of patient-specific anatomical data for advanced navigation in surgery simulation. In *Proceedings of the First International Workshop on Digital Engineering* (2010), ACM, pp. 52–58. 8, 14

[BDV*17] BERNARD J., DOBERMANN E., VÖGELE A., KRÜGER B., KOHLHAMMER J., FELLNER D.: Visual-interactive semi-supervised labeling of human motion capture data. *Electronic Imaging 2017*, 1 (2017), 34–45. 7, 8, 14

[Ber06] BERKHIN P.: A survey of clustering data mining techniques. In *Grouping multidimensional data*. Springer, 2006, pp. 25–71. 1

[Ber15] BERNARD J.: *Exploratory search in time-oriented primary data*. PhD thesis, Technische Universität, Darmstadt, December 2015. 7

[BGV16] BOUALI F., GUETTALA A., VENTURINI G.: Vizassist: an interactive user assistant for visual data mining. *The Visual Computer 32*, 11 (2016), 1447–1463. 7, 10, 12, 14

[BL09] BERTINI E., LALANNE D.: Surveying the complementary role of automatic data analysis and visualization in knowledge discovery. In *Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration* (2009), ACM, pp. 12–20. 1, 2, 16

[BRG*12] BERNARD J., RUPPERT T., GOROLL O., MAY T., KOHLHAMMER J.: Visual-interactive preprocessing of time series data. In *Proceedings of SIGRAD; Interactive Visual Analysis of Data; November 29-30; Växjö; Sweden* (2012), no. 081, Linköping University Electronic Press, pp. 39–48. 6, 12, 14, 16

[BS03] BEDERSON B., SHNEIDERMAN B.: *The craft of information visualization: readings and reflections*. Morgan Kaufmann, 2003. 1

[BS06] BERTINI E., SANTUCCI G.: Give chance a chance: modeling density to enhance scatter plot quality through random data sampling. *Information Visualization 5*, 2 (2006), 95–110. 6

[BSW*14] BERNARD J., STEIGER M., WIDMER S., LÜCKE-TIEKE H., MAY T., KOHLHAMMER J.: Visual-interactive exploration of interesting multivariate relations in mixed research data sets. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 291–300. 8, 14

[CAS*18] COLLINS C., ANDRIENKO N., SCHRECK T., YANG J., CHOO J., ENGELKE U., JENA A., DWYER T.: Guidance in the human–machine analytics process. *Visual Informatics 2*, 3 (2018), 166–180. 2, 10, 16

[CBY10] CHEN Y., BARLOWE S., YANG J.: Click2annotate: Automated insight externalization with rich semantics. In *Visual Analytics Science and Technology (VAST), IEEE Symposium on* (2010), IEEE, pp. 155–162. 9, 12, 13, 14

[CC12] CROUSER R. J., CHANG R.: An affordance-based framework for human computation and human-computer collaboration. *IEEE Transactions on Visualization and Computer Graphics 18*, 12 (2012), 2859–2868. 3

[CGM*17] CENEDA D., GSCHWANDTNER T., MAY T., MIKSCH S., SCHULZ H.-J., STREIT M., TOMINSKI C.: Characterizing guidance in visual analytics. *IEEE Transactions on Visualization & Computer Graphics*, 1 (2017), 111–120. 2, 3, 4, 16

[CGM*18] CENEDA D., GSCHWANDTNER T., MAY T., MIKSCH S., STREIT M., TOMINSKI C.: Guidance or no guidance? a decision tree can help. In *EuroVA: International Workshop on Visual Analytics* (2018), Eurographics Digital Library, p. 19–23. 3, 9, 16

[CLKP10] CHOO J., LEE H., KIHM J., PARK H.: ivisclassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *Visual Analytics Science and Technology (VAST), IEEE Symposium on* (2010), IEEE, pp. 27–34. 7, 10, 12, 14, 16

[CLWM11] CRNOVRSANIN T., LIAO I., WU Y., MA K.-L.: Visual recommendations for network navigation. In *Computer Graphics Forum* (2011), vol. 30, Wiley Online Library, pp. 1081–1090. 9, 14

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association 79*, 387 (1984), 531–554. 4

[CSG*18] CHEGINI M., SHAO L., GREGOR R., LEHMANN D. J., ANDREWS K., SCHRECK T.: Interactive visual exploration of local patterns in large scatterplot spaces. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 99–109. 12

[DFB11] DRUCKER S. M., FISHER D., BASU S.: Helping users sort faster with adaptive machine learning recommendations. In *IFIP Conference on Human-Computer Interaction* (2011), Springer, pp. 187–203. 8, 10, 12, 13, 14

[DLB13] DÖRK M., LAM H., BENJELLOUN O.: Accentuating visualization parameters to guide exploration. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (2013), ACM, pp. 1755–1760. 8, 14

[Dow99] DOWRICK P. W.: A review of self modeling and related interventions. *Applied and preventive psychology 8*, 1 (1999), 23–39. 2, 4

[DR01] DERTHICK M., ROTH S. F.: Enhancing data exploration with a branching history of user operations. *Knowledge-Based Systems 14*, 1-2 (2001), 65–74. 11

[EFN12] ENDERT A., FIAUX P., NORTH C.: Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2012), ACM, pp. 473–482. 4, 8, 13, 14

[EHR*14] ENDERT A., HOSSAIN M. S., RAMAKRISHNAN N., NORTH C., FIAUX P., ANDREWS C.: The human is the loop: new directions for visual analytics. *Journal of intelligent information systems 43*, 3 (2014), 411–435. 4, 12

[Fia12] FIAUX P. O.: *Solving intelligence analysis problems using biclusters*. PhD thesis, Virginia Tech, 2012. 13

[FPSSU96] FAYYAD U. M., PIATETSKY-SHAPIRO G., SMYTH P., UTHURUSAMY R.: Advances in knowledge discovery and data mining. 1

[FTIN97] FUJISHIRO I., TAKESHIMA Y., ICHIKAWA Y., NAKAMURA K.: Gadget: Goal-oriented application design guidance for modular visualization environments. In *Proceedings of the 8th conference on Visualization* (1997), IEEE Computer Society Press, pp. 245–ff. 6, 10, 11, 14

[GG99] GOEBEL M., GRUENWALD L.: A survey of data mining and knowledge discovery software tools. *ACM SIGKDD explorations newsletter 1*, 1 (1999), 20–33. 1

[GGL*14] GRATZL S., GEHLENBORG N., LEX A., PFISTER H., STREIT M.: Domino: Extracting, comparing, and manipulating subsets across multiple tabular datasets. *IEEE Transactions on Visualization & Computer Graphics*, 1 (2014), 1–1. 8, 10, 11, 13, 14

[Gib77] GIBSON J.: The theory of affordances: perceiving, acting, and knowing. 67–82. 1

[GLK*10] GOTZ D., LU J., KISSA P., CAO N., QIAN W. H., LIU S. X., ZHOU M. X.: Harvest: an intelligent visual analytic tool for the masses. In *Proceedings of the first international workshop on Intelligent visual interfaces for text analysis* (2010), ACM, pp. 1–4. 6, 12, 14

[GRM10] GARG S., RAMAKRISHNAN I., MUELLER K.: A visual analytics approach to model learning. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (2010), IEEE, pp. 67–74. 8, 10, 13, 14

[GST13] GLADISCH S., SCHUMANN H., TOMINSKI C.: Navigation recommendations for exploring hierarchical graphs. In *International Symposium on Visual Computing* (2013), Springer, pp. 36–47. 9, 14

[GW09] GOTZ D., WEN Z.: Behavior-driven visualization recommendation. In *Proceedings of the 14th international conference on Intelligent user interfaces* (2009), ACM, pp. 315–324. 7, 13, 14

[HARN11] HOSSAIN M. S., ANDREWS C., RAMAKRISHNAN N., NORTH C.: Helping intelligence analysts make connections. *Scalable Integration of Analytics and Visualization 11* (2011), 17. 9, 10, 13, 14

[HB05] HEER J., BOYD D.: Vizster: Visualizing online social networks. 8, 10, 13, 14

[HBH*98] HORVITZ E., BREESE J., HECKERMAN D., HOVEL D., ROMMELSE K.: The lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence* (1998), Morgan Kaufmann Publishers Inc., pp. 256–265. 9, 13, 14

[HHK15] HEER J., HELLERSTEIN J. M., KANDEL S.: Predictive interaction for data transformation. In *CIDR* (2015). 5, 6, 14

[Hor99] HORVITZ E.: Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (1999), ACM, pp. 159–166. 1, 4, 16

[IV11] IP C. Y., VARSHNEY A.: Saliency-assisted navigation of very large landscape images. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (2011), 1737–1746. 8, 9, 13, 14

[JLJC05] JOHANSSON J., LJUNG P., JERN M., COOPER M.: Revealing structure within clustered parallel coordinates displays. In *Information Visualization. INFOVIS. IEEE Symposium on* (2005), IEEE, pp. 125–132. 8, 10, 11, 13, 14

[JN15] JIANG L., NANDI A.: Snaptoquery: providing interactive feedback during exploratory query specification. *Proceedings of the VLDB Endowment 8*, 11 (2015), 1250–1261. 10, 14

[JZF*09] JEONG D. H., ZIEMKIEWICZ C., FISHER B., RIBARSKY W., CHANG R.: ipca: An interactive system for pca-based visual analytics. In *Computer Graphics Forum* (2009), vol. 28, Wiley Online Library, pp. 767–774. 8, 13, 14

[KB06] KRISHNAMOORTHY G., BRUSILOVSKY P.: Personalized guidance for example selection in an explanatory visualization system. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (2006), Association for the Advancement of Computing in Education (AACE), pp. 2122–2127. 9, 14

[KHP*11] KANDEL S., HEER J., PLAISANT C., KENNEDY J., VAN HAM F., RICHE N. H., WEAVER C., LEE B., BRODBECK D., BUONO P.: Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization 10*, 4 (2011), 271–288. 5

[KMS*08] KEIM D. A., MANSMANN F., SCHNEIDEWIND J., THOMAS J., ZIEGLER H.: Visual analytics: Scope and challenges. In *Visual data mining*. Springer, 2008, pp. 76–90. 1

[KNS04] KREUSELER M., NOCKE T., SCHUMANN H.: A history mechanism for visual data mining. In *Information Visualization. INFOVIS. IEEE Symposium on* (2004), IEEE, pp. 49–56. 11

[KPHH11] KANDEL S., PAEPCKE A., HELLERSTEIN J., HEER J.: Wrangler: Interactive visual specification of data transformation scripts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011), ACM, pp. 3363–3372. 5, 10, 14

[KPN16] KRAUSE J., PERER A., NG K.: Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (2016), ACM, pp. 5686–5697. 11, 14

[KPP*12] KANDEL S., PARIKH R., PAEPCKE A., HELLERSTEIN J. M., HEER J.: Profiler: Integrated statistical analysis and visualization for data quality assessment. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (2012), ACM, pp. 547–554. 5, 10, 14

[KSC*08] KOOP D., SCHEIDEGGER C. E., CALLAHAN S. P., FREIRE J., SILVA C. T.: Viscomplete: Automating suggestions for visualization pipelines. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (2008), 1691–1698. 6, 12, 14

[LCWL14] LIU S., CUI W., WU Y., LIU M.: A survey on information visualization: recent advances and challenges. *The Visual Computer 30*, 12 (Dec 2014), 1373–1393. 1

[LMS*12] LUBOSCHIK M., MAUS C., SCHULZ H.-J., SCHUMANN H., UHRMACHER A.: Heterogeneity-based guidance for exploring multiscale data in systems biology. In *BioVis* (2012), IEEE, pp. 33–40. 10, 14

[LSS*12] LEX A., STREIT M., SCHULZ H.-J., PARTL C., SCHMALSTIEG D., PARK P. J., GEHLENBORG N.: Stratomex: visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization. In *Computer graphics forum* (2012), vol. 31, Wiley Online Library, pp. 1175–1184. 8, 10, 13, 14

[MAK*08] MÜLLER E., ASSENT I., KRIEGER R., JANSEN T., SEIDL T.: Morpheus: interactive exploration of subspace clustering. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (2008), ACM, pp. 1089–1092. 8, 10, 12, 14

[Maz09] MAZZA R.: *Introduction to information visualization*. Springer Science & Business Media, 2009. 10

[MBD*11] MAY T., BANNACH A., DAVEY J., RUPPERT T., KOHLHAMMER J.: Guiding feature subset selection with an interactive visualization. In *Visual Analytics Science and Technology (VAST), IEEE Conference on* (2011), IEEE, pp. 111–120. 6, 10, 11, 12, 14, 16

[MSDK12] MAY T., STEIGER M., DAVEY J., KOHLHAMMER J.: Using signposts for navigation in large graphs. In *Computer Graphics Forum* (2012), vol. 31, Wiley Online Library, pp. 985–994. 9, 10, 11, 14

[MvGW11] MIGUT M., VAN GEMERT J. C., WORRING M.: Interactive decision making using dissimilarity to visually represented prototypes. In *Visual Analytics Science and Technology (VAST), IEEE Conference on* (2011), IEEE, pp. 141–149. 8, 12, 13, 14

[MW10] MIGUT M., WORRING M.: Visual exploration of classification models for risk assessment. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (2010), IEEE, pp. 11–18. 8, 10, 13, 14

[MYIM98] MENG C., YASUE M., IMAMIYA A., MAO X.: Visualizing histories for selective undo and redo. In *Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific* (1998), IEEE, pp. 459–464. 11

[OAH15] O'DONOVAN P., AGARWALA A., HERTZMANN A.: Designscape: Design with interactive layout suggestions. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems* (2015), ACM, pp. 1221–1224. 7, 12, 14

[PHG07] PURCHASE H. C., HOGGAN E., GÖRG C.: How important is the mental map–an empirical investigation of a dynamic graph layout algorithm. In *Graph drawing* (2007), Springer, pp. 184–195. 10

[PS08] PERER A., SHNEIDERMAN B.: Systematic yet flexible discovery: guiding domain experts through exploratory data analysis. In *Proceedings of the 13th international conference on Intelligent user interfaces* (2008), ACM, pp. 109–118. 9, 14

[PSHD96] PIROLLI P., SCHANK P., HEARST M., DIEHL C.: Scatter/gather browsing communicates the topic structure of a very large text collection. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1996), ACM, pp. 213–220. 13

[QB11] QUINN A. J., BEDERSON B. B.: Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems* (2011), ACM, pp. 1403–1412. 3

[SGL09] SHRINIVASAN Y. B., GOTZY D., LU J.: Connecting the dots in visual analysis. In *Visual Analytics Science and Technology. VAST. IEEE Symposium on* (2009), IEEE, pp. 123–130. 9, 13, 14

[SHJ*15] STEIN M., HÄUSSLER J., JÄCKLE D., JANETZKO H., SCHRECK T., KEIM D. A.: Visual soccer analytics: Understanding the characteristics of collective team movement based on feature-driven analysis and abstraction. *ISPRS International Journal of Geo-Information 4*, 4 (2015), 2159–2184. 12

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages. Proceedings., IEEE Symposium on* (1996), pp. 336–343. 12

[SKMW17] SCHWÄRZLER M., KELLNER L.-M., MAIERHOFER S., WIMMER M.: Sketch-based guided modeling of 3d buildings from oriented photos. In *Proceedings of the 21st ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (2017), ACM, p. 9. 10

[SSJKF09] STEED C. A., SWAN J. E., JANKUN-KELLY T., FITZPATRICK P. J.: Guided analysis of hurricane trends using statistical processes integrated with interactive parallel coordinates. In *Visual Analytics Science and Technology. VAST. IEEE Symposium on* (2009), IEEE, pp. 19–26. 8, 13, 14

[SSL*12] STREIT M., SCHULZ H.-J., LEX A., SCHMALSTIEG D., SCHUMANN H.: Model-driven design for the visual analysis of heterogeneous data. *IEEE Transactions on Visualization and Computer Graphics 18*, 6 (2012), 998–1010. 9, 10, 13, 14

[SSS*14a] SACHA D., STEIN M., SCHRECK T., KEIM D. A., DEUSSEN O., ET AL.: Feature-driven visual analytics of soccer data. In *Visual Analytics Science and Technology (VAST), IEEE Conference on* (2014), IEEE, pp. 13–22. 12

[SSS*14b] 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. 1, 2, 3, 4, 8

[ST15] SARVGHAD A., TORY M.: Exploiting analysis history to support collaborative data analysis. In *Proceedings of the 41st Graphics Interface Conference* (2015), Canadian Information Processing Society, pp. 123–130. 11, 14

[Ter95] TERVEEN L. G.: Overview of human-computer collaboration. *Knowledge Based Systems 8*, 2 (1995), 67–81. 1

[Tri] trifacta.com. Accessed: October 20th, 2018. 5

[TSTR12] TOLEDO A., SOOKHANAPHIBARN K., THAWONMAS R., RINALDO F.: Personalized recommendation in interactive visual analysis of stacked graphs. *ISRN Artificial Intelligence* (2012). 14

[Tuk77] TUKEY J. W.: *Exploratory data analysis*, vol. 2. Reading, Mass., 1977. 1

[VALB06] VON AHN L., LIU R., BLUM M.: Peekaboom: a game for locating objects in images. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2006), ACM, pp. 55–64. 3

[VAMM*08] VON AHN L., MAURER B., MCMILLEN C., ABRAHAM D., BLUM M.: recaptcha: Human-based character recognition via web security measures. *Science 321*, 5895 (2008), 1465–1468. 3

[WM13] WU E., MADDEN S.: Scorpion: Explaining away outliers in aggregate queries. *Proceedings of the VLDB Endowment 6*, 8 (2013), 553–564. 8, 14

[WMA*16] WONGSUPHASAWAT K., MORITZ D., ANAND A., MACKINLAY J., HOWE B., HEER J.: Voyager: Exploratory analysis via faceted browsing of visualization recommendations. *IEEE Transactions on Visualization & Computer Graphics*, 1 (2016), 1–1. 7, 10, 12, 14

[WQM*17] WONGSUPHASAWAT K., QU Z., MORITZ D., CHANG R., OUK F., ANAND A., MACKINLAY J., HOWE B., HEER J.: Voyager 2: Augmenting visual analysis with partial view specifications. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (2017), ACM, pp. 2648–2659. 7

[YaKS*07] YI J. S., AH KANG Y., STASKO J. T., JACKO J. A., ET AL.: Toward a deeper understanding of the role of interaction in information visualization. *IEEE Transactions on Visualization & Computer Graphics*, 6 (2007). 12

[YCK09] YUEN M.-C., CHEN L.-J., KING I.: A survey of human computation systems. In *Computational Science and Engineering, 2009. CSE'09. International Conference on* (2009), vol. 4, IEEE, pp. 723–728. 3

[YXRW07] YANG D., XIE Z., RUNDENSTEINER E. A., WARD M. O.: Managing discoveries in the visual analytics process. *ACM SIGKDD Explorations Newsletter 9*, 2 (2007), 22–29. 9, 13, 14

[ZAM11] ZHENG Z., AHMED N., MUELLER K.: iview: A feature clustering framework for suggesting informative views in volume visualization. *IEEE transactions on visualization and computer graphics 17*, 12 (2011), 1959–1968. 7, 11, 14