# Bridging From Goals to Tasks with Design Study Analysis Reports

Heidi Lam, Melanie Tory and Tamara Munzner, *Member, IEEE*

**Abstract**—Visualization researchers and practitioners engaged in generating or evaluating designs are faced with the difficult problem of transforming the questions asked and actions taken by target users from domain-specific language and context into more abstract forms. Existing abstract task classifications aim to provide support for this endeavour by providing a carefully delineated suite of actions. Our experience is that this bottom-up approach is part of the challenge: low-level actions are difficult to interpret without a higher-level context of analysis goals and the analysis process. To bridge this gap, we propose a framework based on analysis reports derived from open-coding 20 design study papers published at IEEE InfoVis 2009-2015, to build on the previous work of abstractions that collectively encompass a broad variety of domains. The framework is organized in two axes illustrated by nine analysis goals. It helps situate the analysis goals by placing each goal under axes of specificity (Explore, Describe, Explain, Confirm) and number of data populations (Single, Multiple). The single-population types are Discover Observation, Describe Observation, Identify Main Cause, and Collect Evidence. The multiple-population types are Compare Entities, Explain Differences, and Evaluate Hypothesis. Each analysis goal is scoped by an input and an output and is characterized by analysis steps reported in the design study papers. We provide examples of how we and others have used the framework in a top-down approach to abstracting domain problems: visualization designers or researchers first identify the analysis goals of each unit of analysis in an analysis stream, and then encode the individual steps using existing task classifications with the context of the goal, the level of specificity, and the number of populations involved in the analysis.

**Index Terms**—Framework, Data Analysis, Analysis Goals, Design Studies, Open Coding, Task Classifications

## 1 Introduction

Although the data analysis process is a central concern in visualization research and practice, we have found that significant challenges remain in supporting, understanding, and teaching visual data analysis because it is difficult to bridge from low-level analysis tasks to higher-level analysis goals. In visualization design and validation, we have advocated for translating domain-specific problems into abstractions describing the data and tasks of target users in domain-agnostic language [29]. As researchers, we have conducted numerous qualitative studies to understand analysis processes [15, 25, 26, 42] where we practiced this suggestion. In our experience, abstracting study results such as system logs and observational data using existing task classifications is time consuming, labour intensive, and difficult; similar frustrations have been noted by others (*e.g*., Reda *et al*. [39]). The low-level nature of most task classifications, which cover individual steps in analysis such as *identify outliers* or *sort* [2, 5, 30, 41, 43, 51, 53], may be an underlying cause of this frustration. In theory, a bottom-up approach should work: it should be possible to abstract each step in the analysis stream and iteratively group these steps to constitute an analysis goal. In practice, however, this bottom-up approach forces the coder to abstract low-level steps without the benefit of important context such as the analysis goal and how that goal is situated within a wider frame of analysis.

We have found that the alternative of explicitly chunking an analysis stream into manageable units, each with an analysis goal, makes the task abstraction process easier and faster than considering low-level steps without the framing context of what the analysis is actually about; i.e., it is easier than performing task abstraction in a vacuum. We have also found that this alternative process is not well supported by existing task classifications, despite previous efforts to bridge the gap between steps and goals. For example, our recent typologies of abstract tasks seek to broadly cover the why, how, and what of visualization design choices, but the proposed action-target pairs such as *discover outliers*, *compare distributions*, or *present trends* constitute low-level steps rather than high-level goals [5, 30]. Efforts to include user goals have either yielded goals such as *exploratory*, *confirmatory*, and *presentation* [43] that are too high-level to connect with tasks, or goals that are too domain specific to directly connect to the abstracted steps across a broad set of domains [14].

We therefore propose a framework that aims to bridge the gap between analysis goals and steps. We identify nine analysis goals that are situated within a larger frame of the analysis process, in order to support holistic reasoning about analysis steps. Central to the framework are two axes: *specificity* and *number of populations*. Specificity ranges from explore to describe to explain to confirm, indicating the degree of selectiveness of the analysis goal in pursuit of the expected outputs. Number of populations scopes the main focus of the analysis to either a single population or multiple populations, distinguishing between characterization and comparison.

We based the framework on a qualitative analysis of IEEE InfoVis design study papers that cover a diverse set of analysis problems and visualization solutions. We open-coded analysis reports in the 20 papers published between 2009 and 2015 that met our inclusion criteria. For each analysis goal, we identified the inputs and outputs of the analysis reported. We also identified the steps reported, to illustrate how the framework can be used in concert with existing classifications, which we intend to augment rather than replace. Note that even though the analysis reports do reference the specific visualizations used in the analysis, we did not explicitly code the visualization techniques; in this way, we focus on the *why* rather than the *how* of analysis.

We illustrate the need for this framework with a concrete analysis example: an analyst observed a spike in airline transaction failures in transaction logs and wanted to understand why the spike occurred. She examined distributions of attributes of the failed transactions such as airline, flight time, travel agent, and so on, and eventually identified one travel agent as a possible culprit for the failures. The actual goal of this analysis is to find the main reason behind the unexpected failure spike in the logs, which fits into our framework as the *Identify Main Cause* analysis goal. It is not possible to adequately capture this higher-level goal with typical task classifications because they aim to characterize the lower-level analysis steps. For example, while it is possible to capture the first step of "The analyst observed a spike" as *action: Analyze>Consume>Discover* and *target: All Data>Outliers* using Munzner's typology [30], the full analysis goal of finding the main reason for the failure cannot be encoded with it.

We envision that our framework can be used as a thinking aid to help

• *H. Lam is with Tableau Research. E-mail: hlam@tableau.com.*
• *M. Tory is with Tableau Research. E-mail: mtory@tableau.com.*
• *T. Munzner is with University of British Columbia. E-mail: tmm@cs.ubc.ca.*

*Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx*

visualization designers and researchers to characterize domain problems, from a starting point at a level that corresponds to how analysis is understood and described by most domain experts. Identifying the analysis goals allows the designer to chunk the analysis streams into more understandable units of abstraction, with each unit situated according to a level of specificity and the number of populations involved. Given this initial designation, the lower-level steps within it can be related to existing task classifications. Note, however, that our framework is only a first step and one interpretation of analysis reports; we envision that future work examining additional sources of data could lead to the addition of goals and perhaps even new axes.

Our contributions are: (1) A framework of nine analysis goals with two axes, each further characterized with an input, an output, and a series of steps; (2) Preliminary reports to illustrate how the framework can be used in concert with existing task classifications.

## 2 Related Work

We first relate our framework to a previous proposal that addresses the connection between goals and tasks, but is not specific to data analysis. We then discuss the extensive previous work on visualization task classifications, followed by previous work involving meta-analysis.

### 2.1 Linking Goals to Tasks

Cooper *et al.* described the interactions between humans and digital products in a hierarchy of *goal-activity-task-action-operation* [10]. The first level in this hierarchy corresponds closely to our own definition of *goal*, but their definition is more encompassing than ours; we focus more narrowly on data analysis aided by visualizations. Their level of *task* corresponds to both the individual analysis *steps* in our framework, and *tasks* in most existing classifications.

### 2.2 Visual Analysis Task Classifications

Rind *et al.* introduced a three-axis conceptual space to compare existing task classifications [40]. One of their axes is composition, as high-level tasks (*e.g.*, 'problem detection') can be broken down into lower-level subtasks (*e.g.*, 'find outliers in data'). The analysis goals in our framework sit closest to the *high* level of composition in their conceptual space. Rind *et al.* identified five previous classifications at this level: [3, 9, 41, 44, 48]. Some of these five identify analysis goals at a level comparable to ours (*e.g.*, Amar & Stasko's *identifying the nature of trends* [3], Suo's *problem detection*, *diagnosis* [48]). However, most are less specific (*e.g.*, Amar & Stasko's *complex decision making*, *learning a domain*, *predicting the future* [3], Roth's *procure*, *predict* [41], Schumann & Muller's *explore*, *confirm*, *present* [44], Thomas & Cook's *assess*, *forecast*, *develop options* [9]). Our framework extends these efforts by providing a consistent list of goals situated in two axes, with each goal illustrated with an example, scoped by an input and an output, and characterized by a series of reported steps.

Very few of the previous task classifications have an overarching analysis-based structure. Some simply consist of a list of identified tasks, either expressed as verbs such as *identify* and *compare* (*e.g.*, Wehrend & Lewis [51]), or verbs and nouns such as *identify outlier* and *compare trends* (*e.g.*, Amar *et al.* [2]). The structured classifications are either framed by data (*e.g.*, the type by task taxonomy [46]) or by the level of abstraction (*e.g.*, Gotz & Zhou's *task-subtask-action-event* [14], Andrienko & Andrienko's elementary and synoptic tasks [4]). In contrast, open-coding design study papers allowed us to examine analysis reports from multiple domains. We were therefore able to identify enough analysis goals to deduce an analysis-based structure: we identified axes of specificity (akin to the exploratory-confirmatory spectrum) and population (akin to the unit of analysis).

Two of the classifications in Rind *et al.* 's set are structured around the analysis process. Brehmer & Munzner [5, 30]; and Schultz *et al.* [43] express task-related questions as why, what, how, and so on. Our analysis goals can be viewed as the *why* question in their typology, but at the analysis level. Rather than asking "why is the task pursued?", we are asking "why is the analysis pursued?". In doing so, our framework

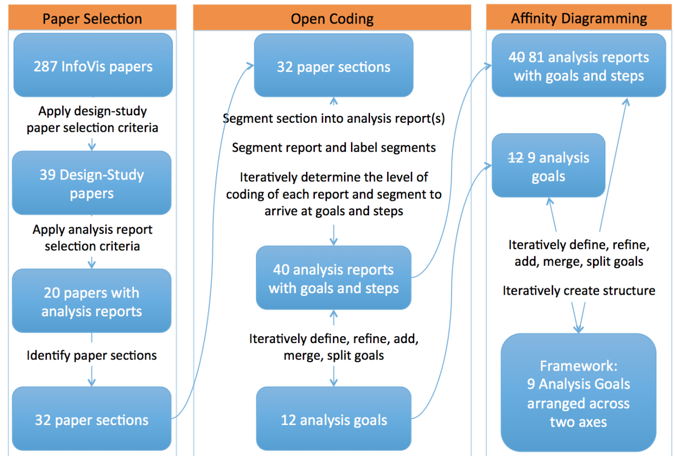


Fig. 1: Overview of framework creation process.

extends Munzner's typology [30] by identifying and delineating nine analysis goals, whose steps can be translated to abstract tasks.

In relation to Schultz *et al.*'s design space [43], our specificity axis is comparable to their goal dimension and our population axis is related to their cardinality dimension. Schultz *et al.* exemplified their design space with ten tasks in the domain of climate impact research. By considering a broader set of domains, we were able to add two more levels (Describe and Explain) to Schultz *et al.*'s goal dimension of *exploration-confirmation-presentation*.

### 2.3 Meta-analysis

Kerracher & Kennedy summarized the construction and evaluation of existing task classifications; our task generation method falls under the category of "derive from literature" [19]. Since our meta-analysis used design study papers as the objects of analysis, by definition we synthesize and re-use the intellectual products created by previous authors in contexts that diverge from the original goals of those authors. We re-use the methods in the 2012 Seven Scenarios paper [21], which has sparked further work in this vein by others [18]. Similarly, in 2014, Sedlmair *et al.* [45] qualitatively re-analyzed design study papers to build a model for visual parameter space analysis. Our work here continues with this trend to use publications as raw materials to derive our framework.

In terms of the data analysis method, we followed the Grounded Theory approach [8, 11], in a similar spirit to previous qualitative visualization research [7, 50].

## 3 Method

Figure 1 summarizes the process we used to create the framework, following the Grounded Theory appraoch [8, 11]. Our source materials were design study papers published at IEEE InfoVis from 2009 to 2015: the 20 papers that passed our selection criteria are listed in Table 1. After we open-coded relevant sections of these papers as analysis reports, we went on to an iterative affinity diagramming phase.

### 3.1 Paper Selection

We scoped our search space to InfoVis published between 2009 and 2015. Our choice of venue is due to the larger number of design-study papers published in InfoVis (43), when compared to VAST (24), and SciVis (7). From 287 papers published at InfoVis in these seven years, we did a first pass to identify 39 papers we deemed to be design study papers based on these criteria:

- addressed a real-world problem,
- used real (*i.e.*, non-synthetic) data,
- involved at least one target user in design or evaluation of the tool in a non-trivial way.

Table 1: Design study papers used to derive our framework. We derived the code names from system names and domains from paper keywords.

| Code Name | First Author | Domain | Ref. |
|---|---|---|---|
| ABySS-Explorer | Nielsen | Bioinformatics | [31] |
| BallotMaps | Wood | Governance | [52] |
| BirdVis | Ferreria | Ornithology | [12] |
| BoxFish | Landge | (Computer) performance | [22] |
| DAViewer | Zhao | Discourse structure | [53] |
| Entourage | Lex | Biological networks | [23] |
| MovExp | Palmas | Human-Computer Interaction | [32] |
| MulteeSum | Meyer | Gene expression | [28] |
| NeuroLines | Al-Awami | Neuroscience | [1] |
| Paramorama | Pretorius | Image analysis | [37] |
| Poemage | McCurdy | Humanities | [27] |
| Ravel | Isaacs | (Computer) performance | [17] |
| SellTrend | Liu | Investigative analysis | [24] |
| SignalLens | Kincaid | Signal processing | [20] |
| SnapShot | Pileggi | Sports | [34] |
| SoccerStories | Perin | Sports | [33] |
| TenniVis | Polk | Sports | [36] |
| Variant View | Ferstay | Bioinformatics | [13] |
| Vials | Strobelt | Biology | [47] |
| Weaver | Quinan | Weather | [38] |

Since we required realistic analysis reports as our input data, we further restricted our paper scope by requiring that the paper include at least one analysis report where a target user performed analysis on real data. This requirement reduced our input set to 20 papers.

Although the design study paper type was introduced in InfoVis 2003, it took several years for the field to develop to the point where authors routinely wrote papers that fully satisfied our selection criteria. We found fewer papers that satisfied our criteria as we surveyed backwards in time. We noted that in 2008 only two papers met our initial three criteria and none met the final requirement; we therefore closed our survey time window at 2009.

### 3.2 Open Coding

Two of the authors performed the open coding and the third author acted as the adjudicator in cases of conflicts. From the 20 papers, we identified 32 paper sections that contained sufficient information for us to delineate the goals and the steps of the analysis. These reports are typically found in the application or case study sections of the papers. As a summary, for each paper section, we

- split the section into segments that each encapsulated an analysis unit with a single goal; these segments were our unit of investigation and are referred to as *analysis reports*
- attached a tag that captured the goal of the analysis as described in the analysis report
- divided each analysis report into analysis steps
- attached a tag that captured the goal of each analysis step
- identified an input and an output of the analysis report

Note that even though we characterize the analysis goal with an input and an output, users may fail by not obtaining the expected output of the analysis. Due to the limited reporting of floundering in the source material of design study papers, our framework does not address the issue of failure.

Even though the above summary sounds like a linear process, open-coding was iterative and collaborative, as shown in Figure 1. Two coders met frequently to discuss how to meaningfully capture each analysis by discussing and reconciling the segmentation of paper sections into analysis reports, as well as definitions and levels of detail of the tags to conceptualize the analysis question and sub-questions pursued in the reports. These tags eventually evolved to be our analysis goals and steps, which also define the scope of our open-coding.

Table 2: Open-coded report from Section 6.2 of SoccerStories [33].

| Goal | Step | Paper Text |
|---|---|---|
| Discover Observation | Note observation | *The Offensive Defender: He began the first article during his initial exploration of all the phases when he was surprised to see that Real Madrids defender Varanne (number 2) was, despite his nominal role, active in many offensive phases of the first game.* |
| Describe Observation (Aggregate) | Identify attribute(s) to define/refine population(s) | *To illustrate this, he selected Varanne to highlight his actions and took the screenshot shown in Figure 14(a).* |
| Explain Differences | Identify attribute difference(s) between populations | *Proceeding with his analysis, he found out that this player was much less involved in offensive phases in the second game.* |
| | Identify attribute difference(s) between populations | *He also compared Varanne's statistics in both games, which showed that the player made much more passes (48) in the second game than in the first (33).* |
| | Relate finding(s) to domain | *Based on what he found out with SoccerStories and his previous knowledge, he deduced that Varanne (and to some extent the whole Real Madrid team) performed this way due to the location of the games: when not playing at home they preferred to wait for the other team to make a risky move and then counter-attack.* |

Table 2 is one example of a coded analysis report. The analysis goals and steps assigned to all the analysis reports are included in the supplementary materials.[1] By the end of the open coding phase, we had 12 analysis-goal tags applied to 40 analysis reports, each further segmented into labeled steps. These were the input data to our next step where the two coders collaboratively performed an affinity diagramming exercise.

### 3.3 Affinity Diagramming

In the affinity diagramming phase, we further clarified the definitions, merged, split, and added analysis goals. For example, we merged two of the goals with other existing goals as we recognized that our initial tagging was clouded by specific domains. One case was merging Parameter Optimization with Identify Main Cause, as we recognized that the process of parameter optimization was just a series of problem diagnosis and resolution steps. In the other case, we merged Compare Trajectories with Compare Entities as we recognized that time and space did not need to be distinguished from other kinds of attributes. We split two of the goals into Item and Aggregate variants, as we discovered that the steps taken by the papers' target users were different depending on the nature of the input data. We also explicitly called out two initial exploratory analysis goals (Discover Observation and Describe Observation) as a few reports convinced us of their importance, even though most reports did not emphasize them. Due to these changes in the goal tags, our total number of analysis reports increased from 40 to 81. Of these, 16 were due to splitting out Discover Observation and 20 were due to splitting out Describe Observation: these were the first steps in many analyses where the analysts then proceeded to other analysis goals. The final counts of the goals are listed in Table 4.

Another outcome of affinity-diagramming was to discover a structure between our analysis goals. We derived two axes that arrange the goals based on their specificity and the number of populations involved in the

[1] Supplementary materials posted at `http://tinyurl.com/gt27fau`.

Table 3: Analysis Goals Framework. Analysis goals (in bold) are organized by two axes: *Specificity* (horizontal) ranges from Explore to Confirm and denotes the selectiveness of the goal; *#Populations* (vertical) denotes the number of populations under consideration. Each analysis goal is characterized by an I=input and O=Output that are further described as Data, Obs=Observation, Pop=Population, Pop Defn=Population Definition, or Pop Contrasts. Observations can be found in a single record (Item) or across a population (Aggregate).

| Spec\#Pops | Explore | Describe | Explain | Confirm |
|---|---|---|---|---|
| Single | **Discover Observation**<br>I: Data only<br>O: Obs<br>(Item or Aggregate) | **Describe Observation (Item)**<br>I: Obs (Item)<br>O: Pop Defn (all attributes)<br>**Describe Observation (Aggregate)**<br>I: Obs (Aggregate)<br>O: Pop Defn (all attributes) | **Identify Main Cause (Item)**<br>I: Obs (Item)<br>O: Pop Defn (dominant attribute)<br>**Identify Main Cause (Aggregate)**<br>I: Obs (Aggregate)<br>O: Pop Defn (dominant attribute) | **Collect Evidence**<br>I: Hypothesis<br>O: Confirm / Reject |
| Multiple | | **Compare Entities**<br>I: Pop Defn<br>O: Pop Contrasts (similarities and differences) | **Explain Differences**<br>I: Pop Defn<br>O: Pop Contrasts (differences) | **Evaluate Hypothesis**<br>I: Pop Defn; Hypothesis<br>O: Confirm / Reject |

Table 4: Goals, reported steps, and their counts in (analysis reports, design study papers). Steps listed in order of counts. A=Aggregate.

| Single-Population Analyses | |
|---|---|
| Discover Observation (18, 14) | Note observation (18, 14)<br>Examine attributes for unusual/interesting observations (1, 1)<br>Examine finding(s) with other instances of observation (1, 1)<br>Calculate derived attributes (1, 1)<br>Relate finding(s) to domain (1, 1) |
| Describe Observation (Item) (2, 2) | Identify attribute(s) to define/refine population(s) (2, 2)<br>Examine finding(s) with other instances of observation (1, 1)<br>Verify observation externally (1, 1) |
| Describe Observation (A) (21, 13) | Identify attribute(s) to define/refine population(s) (21, 13)<br>Identify exception to observation (1, 1)<br>Describe population (1, 1) |
| Identify Main Cause (Item) (9, 7) | Identify likely dominant cause (9, 7)<br>Focus on instance (5, 4)<br>Examine finding(s) with other instances of observation (4, 4)<br>Assess hypothesis based on external information (3, 3)<br>Examine related data to understand observation (1, 1)<br>Relate finding(s) to domain (1, 1) |
| Identify Main Cause (A) (1, 1) | Identify likely dominant cause (1, 1)<br>Identify attribute(s) to define/refine population(s) (1, 1)<br>Assess finding(s) with data (1, 1) |
| Collect Evidence (9, 6) | Form hypothesis (9, 6)<br>Identify evidence to support hypothesis (9, 6)<br>Identify attribute(s) to define/refine population(s) (1, 1)<br>Assess hypothesis (1, 1) |
| **Multiple-Population Analyses** | |
| Compare Entities (5, 3) | Identify attribute difference(s) between populations (5, 3)<br>Identify attribute similarities between populations (4, 3)<br>Relate findings to domain (4, 2)<br>Identify attribute(s) to define/refine population(s) (3, 1)<br>Describe population (1, 1) |
| Explain Differences (9, 8) | Identify attribute difference(s) between populations (9, 8)<br>Relate findings to domain (4, 4)<br>Identify attribute(s) to define/refine population(s) (1, 1)<br>Overview data (1, 1) |
| Evaluate Hypothesis (7, 5) | Identify attribute differences between populations (7, 5)<br>Identify attribute(s) to define/refine population(s) (5, 3)<br>Form hypothesis (6, 4)<br>Assess hypothesis (6, 4)<br>Identify evidence to support hypothesis (2, 2)<br>Relate findings to domain (1, 1)<br>Broaden population scope (1, 1) |

analysis. The supplementary materials show how the coded analysis reports were distributed across the nine analysis goals.

Our framework is therefore the output of an iterative process where we collaboratively built, tested, and rebuilt the components to arrive at a stable structure. The next section details our framework.

## 4 Components of the Framework

Our framework consists of two axes and nine analysis goals, as shown in Table 3. We first introduce the two axes and the concept of an analysis goal, followed by the details of each goal. We use the SoccerStories report featured in Table 2 to illustrate these concepts.

### 4.1 The Two Axes

The first axis represents **specificity**, shown horizontally across Table 3. The specificity axis is akin to the exploratory-confirmatory spectrum and denotes the degree of selectiveness of the analysis goal in pursuit of the expected outputs. We identified four levels: Explore, Describe, Explain, and Confirm. In the SoccerStories analysis report in Table 2, Discover Observation is at the Explore level of specificity: the analyst was exploring the data set to see if there were interesting observations (*e.g.*, noted a defense player was surprisingly active in many offensive phases). Once identified, the analyst wanted to better characterize the observation in Describe (*e.g.*, studied the positions and types of actions). With a better understanding, the analyst then explained the observations, in this case by comparison in Explain Differences (*e.g.*, compared the defense player's actions in the first and the second game and found that he made more passes in the second game). If the analyst had had specific hypotheses, the goal could have become Confirm (*e.g.*, when not playing at home, do players prefer to wait for the other team to make a risky move first?).

The second axis, **population**, is shown vertically in Table 3 and denotes how many groups of records are involved in the analysis. We identified two levels: single and multiple. The SoccerStories analysis report involves multiple populations: the two games.

In addition to the two axes, we also identified a third trend in some single-population analysis goals: **Item** vs. **Aggregate**. The Item variant is a bottom-up approach where the analyst examines individual records one at at time to build up knowledge, *e.g.*, by studying individual actions in SoccerStories. In contrast, the Aggregate variant is a top-down approach where the analyst examines the entire population, *e.g.*, the entire game comprised a population of actions. We call out this variation as the steps taken by the analysts differed, but given that the Item/Aggregate distinction was only found in some analysis goals and only in single-population analyses, our data so far do not support making it into a full axis in our framework despite its importance.

### 4.2 Analysis Goal

An **analysis goal** captures what the analyst wants to achieve in the process, such as Compare Entities, Explain Differences, and Evaluate Hypothesis. In one of the analyses in the SoccerStories report, the analyst wanted to understand the differences in player Varanne's behaviours in the two games. This analysis belongs to the Explain Differences goal. Prior to this, the analyst wanted to describe Varanne's actions and the analysis was labeled as Describe Observation.

We further characterize the goal with an **input** and an **output**. For each goal, we also identified **analysis steps** conducted to achieve the goal given the input/output requirements.

To facilitate understanding, we strive to label these goals as succinctly and as closely to colloquial English as possible. Unfortunately, doing so means we cannot always use the level of specificity in the goal labels (e.g., Compare Entities belongs to the Describe level of specificity, as a shorthand for "Describe the similarities and differences between populations"). As a remedy, we ensure that the specificity level is included in the goal definitions.

#### 4.2.1 Inputs and Outputs

We identify an input and an output for each goal, described using the following terminology:

An **observation** is an interesting finding in the data, typically a trend, outlier, or feature. The observation can be made at the item (record) level or at the aggregate (population) level. A **record** denotes the smallest unit of analysis. In non-tabular data such as a network, a record represents a node or a link, either of which may have attributes. In tabular data, a record is typically stored in one table row, and is comprised of attributes that each hold one item of information. In the SoccerStories report, each record is an action such as a pass or a goal attempt. Each action has associated attributes such as the game, the player, and so on. An example of a Item-level observation is an outlier action that determined the outcome of a game.

A **population** is a set of records. Here the attribute values can be summarized with statistics such as averages and extrema. In the SoccerStories report, an example observation at the aggregate level is an outlier player behaviour (*e.g.*, Varanne was more active in the defense phase of a game). Observations are the output of Explore: Discover Observation and the input to most other single-population analyses, where the analyst further characterizes the unusual record or the population. In SoccerStories, the analyst may want to better describe a key goal attempt action (an Item) and locate other instances to describe the population of effective goal attempts. For the aggregate case, the analyst may want to identify the reason for a player's behaviour.

A **population definition** is a set of attributes and associated values that delineate a population. The definition may involve a single attribute (*e.g.*, game='first') or may involve multiple attributes (*e.g.*, game='first', location='home'). For example, if each action in a soccer game is a record, these records could be grouped into a population such as the "actions in the first games where the match was played at home". This is the output of many single-population types and input to all multiple-population types, as the latter require the populations to be defined before they can be compared (*e.g.*, home games vs. away games).

**Population contrasts** are similarities and/or differences between populations obtained from comparing the populations. The contrasts are expressed as data attributes (*e.g.*, location) and values (*e.g.*, home). This is the output of many multiple-population analyses. An example is the observation that "Varanne's strategy differs by location".

A **hypothesis** is a supposition based on the analyst's belief or limited evidence. This is the input to the two Confirm analysis goals, where the analyst may obtain enough evidence to **Confirm/Reject** the supposition. Note that in some cases, the analyst may not reach the output of Confirm/Reject due to insufficient evidence.

#### 4.2.2 Analysis Steps

In addition to identifying the analysis goals, we also open-coded the reported steps taken to achieve these goals. Similar to abstract tasks, these steps do not specify the tools or techniques used to achieve the goals and can therefore be translated to visualization tasks using existing task classifications. The intent of including these steps is to better characterize the goal rather than to produce a task classification. Table 4 summarizes steps we identified for each goal. Note that since analysis reports may not reflect the actual sequence of steps, the steps presented in Table 4 are ranked by frequency and do not correspond to the order within analysis sequences.

## 5 A Framework of Analysis Goals

We now describe our analysis goals grouped by the number of populations involved in the analysis. For each goal, we provide a definition, an input and an output, illustrated by steps described in design study analysis reports. For clarity, these steps are listed in tables, where the middle column contains generic descriptions of the steps and the right column has an example from a specific analysis report. We strive to be inclusive by selecting representative examples from the design study reports, but not all reported analyses of the same goal shared all the steps, as shown in Table 4.

### 5.1 Single Population Analyses

Single population analyses are summarized in the first row of Table 3 and documented in detail in the following subsections.

#### 5.1.1 Explore: Discover Observation

Discover Observation aims to explore the data to identify interesting trends, patterns, or anomalies. An observation can be found at the item (record) level or at the aggregate (population) level. For an analysis report that began in an open-ended way rather than being driven by a domain question, this analysis goal was the first in the process.

**Example: SignalLens [20]**. SignalLens is a tool to support analysis of electronic signals. The observation (a narrow outlier peak) was made at the item level by examining derived attributes of the signals.

| | Generic description | Example: SignalLens [20] |
|---|---|---|
| Input | Data: The data to analyze, either viewed at the aggregate or item level | A sample electrical signal output from a computer component viewed at the item level |
| Steps | Calculate derived attributes | Calculated rise and fall times to obtain signal pulse width |
| | Examine attributes for unusual or interesting observations | While most pulse widths corresponded to those in the standard specifications, the analyst found a very narrow outlier peak in a histogram of pulse widths |
| | Note observation | Noticed a slight whisker appears to protrude from the otherwise uniform shape |
| | Continue to another goal | Continued to describe the outlier as Describe Observation (Item) |
| Output | Obs: An observation | An outlier (narrow peak) |

#### 5.1.2 Describe: Describe Observation (Item)

The goal in Describe Observation (Item) is to define an observation in terms of data attributes and values. In this case, the observation is at a item level. The process can be iterative where the observation definition is refined over multiple steps.

**Example: SignalLens [20]**. In this report, the analyst knew and isolated a "glitch" within the signal and wanted to identify other instances in the data.

| | Generic description | Example: SignalLens [20] |
|---|---|---|
| Input | Obs (Item): An observation made at the item level | A "glitch" in a signal record |
| Steps | Examine finding(s) with other instances of observation | Used the motif finder to locate other instances of the glitch, but its occurrence was not uniform so the motif finder specification was imperfect |
| | Identify attribute(s) to define/refine observation | Noticed the repetition of the glitch and examined first and second derivatives to refine the filtering condition |
| Output | Pop Defn: A definition of the observation | Outputs of the motif finder and the manually derived filtering conditions |

### 5.1.3 Describe: Describe Observation (Aggregate)

The goal in Describe Observation (Aggregate) is to derive a definition for an observation that is noted at the aggregate level, with an implicit goal to better understand the observation in terms of data attributes. Like the Item version, the definition is normally refined over multiple steps. While the Item version of the goal strives to capture as many similar instances of the observation as possible, the aggregate version here aims to narrow down the definition of the population so that it minimally captures all instances of the observation.

**Example: BallotMaps [52]**. BallotMaps facilitated the analysis of election results. The goal of the main analysis was to understand the sources of name-order bias in elections. In other words, the analyst aimed to understand how positions of candidate names on the ballot influenced the number of votes received.

| | Generic description | Example: BallotMaps [52] |
|---|---|---|
| Input | Obs (Aggregate): An observation made at the aggregate level | The analysts saw evidence of name-order bias in the election result data |
| Steps | *Iteratively* identify attribute(s) to define/refine population(s) | *One example*: the bias was more evident in certain electoral boroughs |
| | Identify exception to observation | Identified borough with no name ordering effects |
| | Describe population | Summarized population with descriptive statistics |
| Output | Pop Defn: A definition of the observation | Identified that attributes borough and political party were associated with the name-bias observation. |

### 5.1.4 Explain: Identify Main Cause (Item)

The goal in Identify Main Cause (Item) is to explain an observation by finding the main contributor to an observation found at the item level. Since the diagnosis is deduced based on a single record, the analysts typically need to verify their hypotheses by extending the investigation to other similar instances.

In contrast to the Describe Observation types above, the goal here is not to obtain a complete set of attributes to fully describe the observation, but rather to identify the dominant attribute that contributed to the observation. That is, the goal is more specific as it is guided by the domain knowledge of the analysts.

**Example: DAViewer [53]**. DAViewer enabled computational linguistics researchers to explore, compare, and annotate the results of document parsers; they investigated the flaws of a document-parsing algorithm.

| | Generic description | Example: DAViewer [53] |
|---|---|---|
| Input | Pop Defn: A population that includes an observation noted at the item level | A set of documents with fragments that scored poorly with a document parser |
| Steps | *Iteratively* identify likely dominant cause | *One example loop* |
| | Focus on instance | Examined a specific document fragment with low score ("individual prosperity inevitably would result") |
| | Examine related data to understand observation | Examined the associated text for context |
| | Assess hypothesis based on external information/data | The parser did not classify the Cause relationship correctly |
| | Examine observation(s) with other instances | Looked for other fragments containing phrases "because" or "as a result" to see if the classification error also occurred |
| Output | Pop Defn: Dominant attribute explaining the observation | Concluded that the parser did not classify the Cause relationship correctly |

### 5.1.5 Explain: Identify Main Cause (Aggregate)

The goal in Identify Main Cause (Aggregate) is to explain an aggregate observation to find the main contributor to that observation. Analysts typically refined the population definition iteratively to identify the main cause. This strategy is different from the Item variant where the analysts took a bottom-up approach based on individual records. In contrast, the Aggregate variant involved top-down refinement.

**Example: SellTrend [24]**. SellTrend supported analysis of airline travel purchase requests. The analysts wanted to find the main attribute that accounted for an abnormally high count of failed transactions.

| | Generic description | Example: SellTrend [24] |
|---|---|---|
| Input | Pop Defn: A population with an observation found at the aggregate level | A spike in the time trend of daily failed airline transactions |
| Steps | *Iteratively* identify likely dominant cause | *One example loop* |
| | Identify attribute(s) to define/refine population(s) | Identified that Airline A80 had contributed the most failed transactions |
| | Assess finding(s) with data | Confirmed that the airline's failure count was worse than the historical average |
| | Identify attribute(s) to define/refine population(s) | Filtered to focus on airline A80's failed transactions |
| Output | Pop Defn: The dominant attribute that explains the observation | Travel agent Z7F contributed to most of airline A80's failed transactions |

### 5.1.6 Confirm: Collect Evidence

The goal of Collect Evidence is to confirm one's beliefs about a population. The process can be iterative, where the analyst triangulates the evidence before coming to a final verdict.

**Example: Ravel [17]**. Ravel helps software engineers optimize performance of parallel programs. Their analysis focused on confirming that poor performance of a system was due to load imbalance.

| | Generic description | Example: Ravel [17] |
|---|---|---|
| Input | Hypothesis: Prediction about the population based on previous analyses or domain expectations | Hypothesis that poor performance of the experiment runs were caused by load imbalance in the process |
| Steps | *Iteratively* identify evidence to support hypothesis | *One example*: the process that was assigned more input data exhibited poorer performance |
| Output | Confirm or reject hypothesis | Confirmed that poor performance was caused by load imbalance of the processes |

## 5.2 Multiple Population Analyses

Multiple population analyses are summarized in the second row of Table 3 and focus on comparing two or more populations. The outputs are expressed as *population contrasts*, or similarities and differences between the populations' attribute values. The populations are defined such that their contrasts can answer domain questions.

### 5.2.1 Describe: Compare Entities

The goal of Compare Entities is to describe two or more populations by comparing one to another, similar to the single-population case of Describe: Describe Observation. The process is usually iterative, where the analysis progressively compares the populations to collect similarities and differences of attribute values.

**Example: Entourage [23]**. Entourage supported analysis of relationships between biological pathways. The analyst wanted to compare different populations of cells that came from various tissues in terms of their responsiveness to two drugs. They also wanted to understand how drug responsiveness related to gene expression levels.

| | Generic description | Example: Entourage [23] |
|---|---|---|
| Input | Pop Defns: Two or more populations defined by attribute values | Cell populations denoted by tissue type, with drug responsiveness measurements and expression levels of a gene |
| Steps | *Iteratively* identify attribute similarities between populations | *One example*: Cell lines from lung, breast and three other tissues were responsive to the drug |
| | *Iteratively* identify attribute differences between populations | *One example*: Breast and lung cell lines exhibited over-expression; this was different from other cell lines that were also responsive to the drug. |
| Output | Pop Contrasts (similarities and differences): to understand domain questions | Cell lines that were responsive to the test drug did not all exhibit over-expression of a gene. |

### 5.2.2 Explain: Explain Differences

The goal of Explain Differences is to identify differences between the attribute values of multiple populations of records. The implicit goal is to deduce the cause of the differences, similar to Explain: Identify Main Cause. The process is iterative, where the analyst collects differences until they reach an understanding.

**Example: DAViewer [53]**. The following analysis report continues from §5.1.4, Identify Main Cause (Item), where the analyst hypothesized that the main reason behind the poor performance of a document parser was its inability to classify the Cause relationship in text fragments. Here, she wanted to further understand the problem by investigating if and how tuning different parameters affected the outputs of the parsers.

| | Generic description | Example: DAViewer [53] |
|---|---|---|
| Input | Pop Defns: Two or more populations defined by attribute values | A set of documents processed by three parsing algorithms and the differing scores between the algorithms. One of the algorithms had a feature deactivated. |
| Steps | *Iteratively* identify attribute differences between populations | *One example*: the algorithm with a feature deactivated could only identify Elaboration and Same Unit relationships, but missed Contrast relationships. |
| Output | Pop Contrasts (differences): to answer domain-level questions | The algorithm with a deactivated feature over-classified or missed certain relationships, which indicated that the deactivated feature was essential. |

### 5.2.3 Confirm: Evaluate Hypothesis

The goal of Evaluate Hypothesis is to confirm suspected similarities or differences between populations, similar to the single-population case of Confirm: Collect Evidence. The investigation is therefore very focused, as only specific attributes are examined.

**Example: SnapShot [34]**. SnapShot is a tool to aid analysis of hockey tactics. The analysts hypothesized that "animals defend their home turf", *i.e.*, players would play more defensively at home than away.

| | Generic description | Example: SnapShot [34] |
|---|---|---|
| Input | Hypothesis, Pop Defn: Two or more defined populations and a hypothesis about them | Players would be more defensive when at home; the populations were therefore games played by home teams and guest teams |
| Steps | *Iteratively* identify evidence to support hypothesis | *One example*: the analyst noticed that the home teams had more long shots and fewer short shots |
| Output | Confirm or reject hypothesis | More long shots from home teams implied a more defensive strategy, confirming hypothesis |

## 6 Illustrations of Use

We provide several examples illustrating the use of our framework.

## 6.1 Use in Connecting to Existing Task Classifications

We first illustrate how the framework can be used in concert with existing task classifications, using Munzner's Action-Target task typology [30] and the SellTrend [24] analysis report to illustrate the analysis goal "Explain: Identify Main Cause (Aggregate)" in Section 5.1.5. After we had extracted that analysis goal, we identified the main steps in that analysis and translated these steps into the terminology of Munzner's typology [30], as shown in Table 5. Every step in our framework corresponds to some step in that typology, showing that the gap between goals and steps is indeed bridged.

## 6.2 Use in Studying Interview Scripts

We captured an analysis in a contextual interview where two of the authors "tagged along" to watch data analysis in action. The analyst was a data blogger who explored a dataset about the participants of Burning Man to create a dashboard to showcase her findings. She explained her work as she progressed, and we asked clarification questions periodically. The one-hour session was video-recorded and one of us transcribed and open-coded the recording using the framework. The coded video transcripts are provided in the supplemental materials. Using our framework, we were able to summarize the analysis using the analysis goals of Explain Differences and Describe Observation.

Of note is that we initially tried to analyze this and another similar session by coding each line of the session transcript using Munzner's typology [30] and Schultz *et al.*'s design space parameters [43], also

Table 5: Translating an Identify Main Cause (Aggregate) report to Munzner's typology [30]

| Analysis Steps (Our framework) | Analysis Report (SellTrend [24]) | Munzner [2014] Actions | Munzner [2014] Targets |
|---|---|---|---|
| Note an observation at the aggregate level | Noted spike in the time trend of daily failed airline transactions | Analyze>Consume>Discover<br>Search>Explore<br>Query>Identify | All Data>Outliers |
| Identify the likely dominant attribute that caused the observation | Identified that airline A80 contributed the most failed transactions | Analyze>Consume>Discover<br>Search>Locate<br>Query>Identify | Attribute>One>Distribution>Extremes |
| Confirm the identified attribute | Confirmed that the airline's contribution is worse than historical average | Analyze>Consume>Discover<br>Search>Lookup<br>Query>Compare | Attribute>Many>Correlation |
| Refine the analysis population by applying the identified dominant attribute | Focused on airline A80s failed transactions | Analyze>Consume>Discover<br>Search>Browse<br>Query>Identify | Attribute>One>Distribution>Extremes |

in the supplemental materials. We abandoned that attempt as we frequently floundered in our coding of the transcript lines and lost our focus in the long discussions. Only after we identified the analysis goals did we succeed in finishing this coding.

On reflection, we realized that the other task classifications inadvertently pulled us into thinking at the low level of transcript actions. This level is arguably inappropriate, as interview data is very noisy and participants followed their stream of consciousness rather than presenting clear analysis goals and steps. In the re-analysis, we found that the framework let us look at the analysis through a higher level lens to quickly group a series of transcript activities and understand the analyst's higher level goals, even though they were rarely articulated verbally. The framework's elaboration of Munzner's Discover goal into nine specific analysis goals and their characteristic steps enabled us to recognize instances of those analysis goals in our data.

### 6.3 Use in Studying Analysis Logs

We captured an analysis performed with Tableau Desktop via computer logs (to collect user actions and visualization states) and a questionnaire (to solicit context and intent). The computer logs consist of time-stamped user commands (e.g., "drop-ui"), the associated visualization states (e.g., data dimensions on the various shelves such as row, column, filter), and screen captures of the visualizations.

We captured 341 user commands with 328 associated screen captures. From the questionnaire, the over-arching goal of the analysis was to explore a dataset of United States executions with the intent to tell a story. To ease our analysis, we built a tool to construct the visualization state and to associate each screen capture with a user command.

We first did a rough pass to identify the main intent of the user commands and chunk the commands that shared the same goal. With that as context, we described the command in English, then abstracted the description to remove domain-specific details.

The complete results are included in the supplemental materials and we offer one example to illustrate our process. The first ten user commands correspond to the user building a line chart of total number of executions over time, then grouping the records by race. The user then named the worksheet as "More white people than black are executed". We inferred the goal of these ten commands as Discover Observation, with the steps of "Examine attributes for unusual or interesting observations" followed by "Note observation".

We were able to capture all of the analysis with our proposed analysis goals. In terms of steps, we found one instance that was absent in the design study analysis reports: "Examine if populations are appropriate for comparison" in Compare Entities. In these cases, the analyst performed an extra screening to see if the populations were indeed interesting before doing a deeper comparison; oftentimes, the analysis tasks were "dead-ends" as the analyst quickly moved onto other sets of data slices in his investigation. Here the more detailed analysis steps collected via computer logs capture the floundering and quick examination steps that occur in exploratory data analysis; these are not described in highly curated design study papers. This example illustrates the limitations of using design study papers as source materials to identify steps, as discussed further in Section 7.1.

### 6.4 Use in Visualization Design

One of our former students, AS, was asked to improve an existing dashboard used by HIV researchers to explore incidence data, which includes geographic and demographic information of the patients. He started by collecting information about the needs of his target users from interviews, as well as from the reports generated by the HIV researchers. At first, he tried to abstract the domain-specific tasks using Brehmer and Munzner's typology [5] but without success: he found the terminology confusing (e.g., search vs. query) and did not understand the structure of the typology. AS then consulted his previous adviser and she suggested our unpublished framework, which he found helpful.

Specifically, AS appreciated the simple structure of the two axes of the framework. While the number of populations did not resonate with him, the specificity axis helped him place the analysis needs of his users in the framework as Explore, Describe, Explain, or Confirm. Doing so not only helped him abstract the domain problems but also helped him better understand and differentiate seemingly similar levels such as Describe and Explain. He then moved from one of these specificity levels in the framework to the specific goals. At times, he was uncertain as to where to put a user need on the specificity axis. In those cases, the input and the output associated with each goal helped clarify his understanding of the user need.

For example, the HIV researchers wanted to understand why there were more new cases of HIV in a geographic region than expected. AS could characterize that domain problem as Explain, where the researchers filtered the data to focus and narrow their investigations to identify the cause; for example, movement of people into the region may have brought in new cases and new viruses that were rapidly spreading. As another example, the researchers wanted to understand the similarities and differences between the two fastest growing clusters in a specific geographic region based on factors such as age of the patients, their sexual orientations, and ethnicity. The analysis input was the clusters along with patient demographics. AS recognized this user analysis as Compare Entities, and subsequently explicitly supported this goal in the dashboard to address a previously unmet user need.

With a clearer understanding of his users' needs, AS was able to build an interactive prototype and communicate its value to the HIV researchers, who included the prototype as part of a grant proposal.

## 7 Reflections on our Framework

We envision our framework as a thinking aid to assist visualization designers and researchers in translating domain problems into abstract tasks. Here, we reflect on the methods we used for our framework's derivation and the reliability of its results, as well as future directions.

### 7.1 Using Design Studies as Sources

We chose to use design study analysis reports as our source material so that we could survey a broad range of domain problems and visualization approaches and benefit from the intellectual work of the authors in distilling their observations into a concise and coherent form. We believe that the high level of curation for these reports may have precisely enabled to us to develop this framework, given the failure of our initial efforts to analyze analysis streams collected from observation studies (Section 6.2). However, despite the stringent paper-selection criteria designed to ensure a certain level of realism (real users with real-world data), design study papers are not intended to be complete reports of the analysis process and thus our framework is also not complete.

First, the curtailment of the curated reports may explain the terseness we found in the Explore: Discover Observation analysis goal and the limited reports of failures and floundering. One exception is SnapShot, which reported an analysis where the analyst repeatedly failed to verify his hypotheses and the experience inspired the analyst to think more creatively. This exception gives us a glimpse of what might be missing and we believe more faithful recordings of analyses are required to more comprehensively cover initial stages of analysis (such as exploring relationships between attributes) as well as failures and floundering.

A second and perhaps more limiting aspect of our source materials is that one cannot assume the reports to faithfully record actual analysis processes as they happened, even though we believe the essence of the analyses are preserved in these analysis reports. For this reason, our identified analysis steps serve to illustrate the goals but not fully characterize them, and should not be considered as a sole source from which to build a task classification of the kind identified by Brehmer *et al.* for dimensionally reduced data [6]. For that purpose, we believe actual analysis records, obtained via logs or recordings, are required.

Another shortcoming of the design study reports are potential biases such as the use of domain experts as evaluators, uneven distribution of domains (*e.g.*, biology-related domains seemed over-represented, as seen from Table 1), and the use of a single tool built by the design study authors for each analysis. Also, because we only surveyed InfoVis publications, we naturally missed domain problems that are more prominent in other venues such as VAST, SciVis, and CSCW, as well as other sources such as VAST challenge entries. For example, we did not observe explicit examples of collaboration or presentation, nor extensive use of computations such as machine learning and statistics.

Finally, we merged goal tags during the affinity diagramming phase as we did not have enough materials to confidently call out tags as different goals. Examples include Parameter Optimization (merged with Identify Main Cause) and Compare Trajectories (merged with Compare Entities). More analysis goals are likely to be identified with more source materials.

In short, while using design study analysis reports enabled us to broadly survey domains and systems, we recognize the need to include other venues for a more complete set of analysis goals, as recommended by Kerracher & Kennedy as the "multi-strand approach to task gathering" [19]. In addition, we also need more faithful recordings of the analysis process to truly capture the steps.

### 7.2 Framework Axes

We derived two axes to organize the nine analysis goals: specificity and number of populations. The specificity axis, ranging from exploratory to confirmatory, is not a new concept. Tukey stated that "exploratory and confirmatory can–and should–proceed side by side" [49]. Pirolli and Card's sensemaking model depicts a continuous process for intelligence analysis, which starts with exploration and proceeds to problem structuring, evidentiary reasoning, and decision making [35]. Here, we characterize the exploratory-confirmatory axis more generally as Explore, Describe, Explain, and Confirm, based on analysis goals taken from a diverse set of domain problems. Comparing our analysis goals with cognitive models of analysis such as that of Grolemund and Wickham's [16], we notice that less specific levels (Explore and Describe) are less schema-driven than the more specific levels (Explain and Confirm). Further work is needed to better relate these levels to cognitive models. We also anticipate further enrichment of this specificity axis as more analysis goals are collected.

The second axis concerns the number of populations. We call out this attribute of analysis as we observe that the steps in multiple-population analyses are different from their single-population counterparts. Namely, only multiple-population analyses focus on comparison. To us, the unique task of comparison calls for different visualization techniques (such as small multiples) and is therefore worth highlighting.

Despite using design-study analysis reports as our source materials, we did not specifically encode the visualization techniques used in the analyses. It is possible to support the same analysis goal with different techniques and thus we consider these questions to be separable, along the same lines as previous authors [4, 30, 43]. Our goal was to create a useful framework for goals that balances generality and simplicity.

That said, our background in visualization did influence the organization of our framework through the choice of axes. Number of populations was considered important due to the importance of comparison tasks in visualization design, and the notion of Item/Aggregate stood out because of the different techniques used to support the two variants, as indicated by our preliminary foray into open-coding the techniques. We observed that while multiple-population analysis goals are typically supported by small-multiple techniques, single-population analysis goals with Item observations are typically supported by details-on-demand. This concept was also identified by Sedlmair *et al.* as local-to-global and global-to-local strategies [45] and is similar in spirit with Andrienko & Andrienko's elementary and synoptic tasks [4]. Due to the limited analysis reports we used as our source material, we only identified Item and Aggregate variants of two analysis goals (Describe Observation and Identify Main Cause), which is insufficient to call it out as a separate axis. We believe further addition of analysis goals, as well as integrating visualization techniques into our framework, will help clarify the Item/Aggregate dimension.

### 7.3 Analysis Goals and Steps

The nature of our source materials impacted the fidelity of our steps, which we cautiously use to characterize goals rather than as observations of actual strategies. Nonetheless, it is instructive to see whether reports attributed to the same goals share common steps. Table 4 shows the counts of reports associated with each goal and step. There are indeed common steps for the analysis goals. For example, "identify attribute difference(s) between populations" is present in all Explain Differences reports, while all but one report of Compare Entities have a step of identifying attribute similarities. In the case of Evaluate Hypothesis, all analysis reports contain the form hypothesis step and/or the assess hypothesis step. The presence of dominant steps in the analysis reports indicates that we are consistent in characterizing analysis goals.

## 8 Conclusions

From 81 analysis reports extracted from 20 design study papers, we derived a framework that aims to bridge the gap from high-level domain goals to specific low-level tasks that are typically the focus of existing task classifications. The framework consists of two axes with nine analysis goals. Each goal is scoped by an input and an output, and illustrated with steps identified from analysis reports.

Our framework consists of two axes of specificity and number of populations. We identified four levels of analysis specificity, enriching the exploratory-confirmatory spectrum by adding Describe and Explain between them. We also refined the situations in which Item and Aggregate analysis strategies are applied.

Preliminary uses show that the current framework has value, although it is necessarily incomplete due to the limitations of our methods. We invite the visualization community to augment this initial effort to form a more complete understanding of the data analysis process.

## Acknowledgments

We thank Michael Sedlmair, Matthew Brehmer, the UBC InfoVis group, and Tableau Research for their constructive feedback. We also thank AS for sharing his experience in using the framework.

## REFERENCES


[1] A. K. Al-Awami, J. Beyer, H. Strobelt, N. Kasthuri, J. W. Lichtman, H. Pfister, and M. Hadwiger. NeuroLines: a subway map metaphor for visualizing nanoscale neuronal connectivity. *IEEE Trans. Visualization and Computer Graphics*, 20(12):2369–2378, 2014.

[2] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symp. Information Visualization*, pp. 111–117, 2005.

[3] R. A. Amar and J. T. Stasko. Knowledge precepts for design and evaluation of information visualizations. *IEEE Trans. Visualization and Computer Graphics*, 11(4):432–442, 2005.

[4] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer, 2006.

[5] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2376–2385, 2013.

[6] M. Brehmer, M. Sedlmair, S. Ingram, and T. Munzner. Visualizing dimensionally-reduced data: Interviews with analysts and a characterization of task sequences. In *Proc. BELIV*, pp. 1–8. ACM, 2014.

[7] S. Carpendale. Evaluating information visualizations. In J.-D. F. Andreas Kerren, John T. Stasko and C. North, eds., *Information Visualization: Human-Centered Issues and Perspectives*, vol. 4950, pp. 19–45. Springer LNCS, 2008.

[8] K. Charmaz. *Constructing grounded theory. A practical guide through qualitative analysis*. Sage Publications, Inc, 2006.

[9] K. A. Cook and J. J. Thomas. *Illuminating the path: The research and development agenda for visual analytics*. National Visualization and Analytics Center), 2005.

[10] A. Cooper, R. Reimann, D. Cronin, and C. Noessel. *About face: The essentials of interaction design*. John Wiley & Sons, 2014.

[11] J. W. Creswell. *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Prentice Hall, 2002.

[12] N. Ferreira, L. Lins, D. Fink, S. Kelling, C. Wood, J. Freire, and C. Silva. BirdVis: Visualizing and understanding bird populations. *IEEE Trans. Visualization and Computer Graphics*, 17(12):2374–2383, 2011.

[13] J. A. Ferstay, C. B. Nielsen, and T. Munzner. Variant View: Visualizing sequence variants in their gene context. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2546–2555, 2013.

[14] D. Gotz and M. X. Zhou. Characterizing users' visual analytic activity for insight provenance. *Information Visualization*, 8(1):42–55, 2009.

[15] L. Grammel, M. Tory, and M.-A. Storey. How information visualization novices construct visualizations. *IEEE Trans. Visualization and Computer Graphics*, 16(6):943–952, Nov. 2010.

[16] G. Grolemund and H. Wickham. A cognitive interpretation of data analysis. *International Statistical Review*, 82(2):184–204, 2014.

[17] K. E. Isaacs, P.-T. Bremer, I. Jusufi, T. Gamblin, A. Bhatele, M. Schulz, and B. Hamann. Combing the communication hairball: Visualizing parallel execution traces using logical time. *IEEE Trans. Visualization and Computer Graphics*, 20(12):2349–2358, 2014.

[18] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. A systematic review on the practice of evaluating visualization. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2818–2827, 2013.

[19] N. Kerracher and J. Kennedy. Constructing and evaluating visualisation task classifications: Process and considerations. In *Computer Graphics Forum*, vol. 36, pp. 47–59. Wiley Online Library, 2017.

[20] R. Kincaid. SignalLens: Focus+ context applied to electronic time series. *IEEE Trans. Visualization and Computer Graphics*, 16(6):900–907, 2010.

[21] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *IEEE Trans. Visualization and Computer Graphics*, 18(9):1520–1536, 2012.

[22] A. G. Landge, J. A. Levine, A. Bhatele, K. E. Isaacs, T. Gamblin, M. Schulz, S. H. Langer, P.-T. Bremer, and V. Pascucci. Visualizing network traffic to understand the performance of massively parallel simulations. *IEEE Trans. Visualization and Computer Graphics*, 18(12):2467–2476, 2012.

[23] A. Lex, C. Partl, D. Kalkofen, M. Streit, S. Gratzl, A. M. Wassermann, D. Schmalstieg, and H. Pfister. Entourage: Visualizing relationships between biological pathways using contextual subsets. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2536–2545, 2013.

[24] Z. Liu, J. Stasko, and T. Sullivan. SellTrend: Inter-attribute visual analysis of temporal transaction data. *IEEE Trans. Visualization and Computer Graphics*, 15(6):1025–1032, 2009.

[25] N. Mahyar, A. Sarvghad, and M. Tory. Note-taking in co-located collaborative visual analytics: Analysis of an observational study. *Information Visualization*, 11(3):190–204, 2012.

[26] N. Mahyar and M. Tory. Supporting communication and coordination in collaborative sensemaking. *IEEE Trans. Visualization and Computer Graphics*, 20(12):1633–1642, 2014.

[27] N. McCurdy, J. Lein, K. Coles, and M. Meyer. Poemage: Visualizing the sonic topology of a poem. *IEEE Trans. Visualization and Computer Graphics*, 22(1):439–448, 2016.

[28] M. Meyer, T. Munzner, A. DePace, and H. Pfister. MulteeSum: a tool for comparative spatial and temporal gene expression data. *IEEE Trans. Visualization and Computer Graphics*, 16(6):908–917, 2010.

[29] T. Munzner. A nested model for visualization design and validation. *IEEE Trans. Visualization and Computer Graphics*, 15(6):921–928, 2009.

[30] T. Munzner. *Visualization Analysis and Design*. CRC Press, 2014.

[31] C. B. Nielsen, S. D. Jackman, I. Birol, and S. J. Jones. ABySS-Explorer: visualizing genome sequence assemblies. *IEEE Trans. Visualization and Computer Graphics*, 15(6):881–888, 2009.

[32] G. Palmas, M. Bachynskyi, A. Oulasvirta, H.-P. Seidel, and T. Weinkauf. MovExp: A versatile visualization tool for human-computer interaction studies with 3D performance and biomechanical data. *IEEE Trans. Visualization and Computer Graphics*, 20(12):2359–2368, 2014.

[33] C. Perin, R. Vuillemot, and J.-D. Fekete. SoccerStories: A kick-off for visual soccer analysis. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2506–2515, 2013.

[34] H. Pileggi, C. D. Stolper, J. M. Boyle, and J. T. Stasko. SnapShot: Visualization to propel ice hockey analytics. *IEEE Trans. Visualization and Computer Graphics*, 18(12):2819–2828, 2012.

[35] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc. Intl. Conf. Intelligence Analysis*, vol. 5, pp. 2–4, 2005.

[36] T. Polk, J. Yang, Y. Hu, and Y. Zhao. TenniVis: Visualization for tennis match analysis. *IEEE Trans. Visualization and Computer Graphics*, 20(12):2339–2348, 2014.

[37] A. J. Pretorius, M.-A. P. Bray, A. E. Carpenter, and R. A. Ruddle. Visualization of parameter space for image analysis. *IEEE Trans. Visualization and Computer Graphics*, 17(12):2402–2411, 2011.

[38] P. S. Quinan and M. Meyer. Visually comparing weather features in forecasts. *IEEE Trans. Visualization and Computer Graphics*, 22(1):389–398, 2016.

[39] K. Reda, A. Johnson, M. Papka, and J. Leigh. Modeling and evaluating user behavior in exploratory visual analysis. *Information Visualization*, 15(4):325–339, 2016.

[40] A. Rind, W. Aigner, M. Wagner, S. Miksch, and T. Lammarsch. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization*, 15(4):288–300, 2015.

[41] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2356–2365, 2013.

[42] A. Sarvghad and M. Tory. Exploiting analysis history to support collaborative data analysis. In *Graphics Interface*, pp. 123–130, 2015.

[43] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Trans. Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[44] H. Schumann and W. Müller. *Visualisierung: Grundlagen und allgemeine Methoden*. Springer-Verlag, 2013.

[45] M. Sedlmair, C. Heinzl, S. Bruckner, H. Piringer, and T. Möller. Visual parameter space analysis: A conceptual framework. *IEEE Trans. Visualization and Computer Graphics*, 20(12):2161–2170, 2014.

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

[47] H. Strobelt, B. Alsallakh, J. Botros, B. Peterson, M. Borowsky, H. Pfister, and A. Lex. Vials: Visualizing alternative splicing of genes. *IEEE Trans. Visualization and Computer Graphics*, 22(1):399–408, 2016.

[48] X. Suo. *A task-centered visualization design environment and a method for measuring the complexity of visualization designs*. PhD thesis, Georgia State University, Atlanta, GA, 2009.

[49] J. W. Tukey. *Exploratory data analysis*. Addison-Wesley, 1977.

[50] J. Walny, S. Huron, and S. Carpendale. An exploratory study of data sketching for visual representation. In *Computer Graphics Forum*, vol. 34, pp. 231–240. Wiley Online Library, 2015.

[51] S. Wehrend and C. Lewis. A problem-oriented classification of visualiza-

tion techniques. In *Proc. IEEE Conf. Visualization*, pp. 139–143. IEEE, 1990.
[52] J. Wood, D. Badawood, J. Dykes, and A. Slingsby. BallotMaps: Detecting name bias in alphabetically ordered ballot papers. *IEEE Trans. Visualization and Computer Graphics*, 17(12):2384–2391, 2011.
[53] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating discourse analysis with interactive visualization. *IEEE Trans. Visualization and Computer Graphics*, 18(12):2639–2648, 2012.