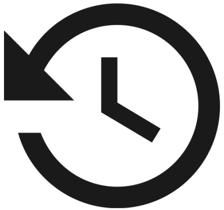
Survey on the Analysis of User Interactions and Visualization Provenance

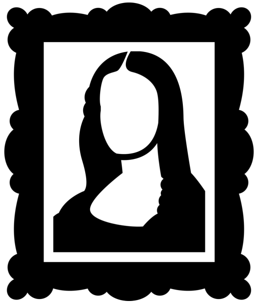
Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit,
Remco Chang and John Wenskovitch

Middlesex University London
Washington University in St. Louis
JOHANNES KEPLER UNIVERSITY LINZ
Tufts UNIVERSITY
VIRGINIA TECH

What is Provenance?

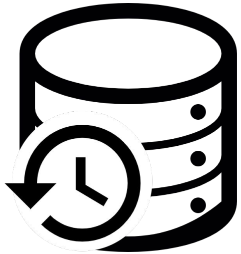


“The place of origin or earliest known history of something” (Oxford Dictionary)



Often used in the context of “a valued object or work of art" (Merriam-Webster)

In the field of Computer Science, the concept has been applied to data, computation, and user interactions.

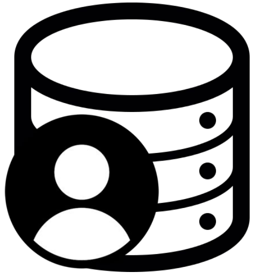


For example, “data provenance” includes the context information such as how/when/why data are collected/recorded/stored/processed.

Why do we need this survey?

Provenance is a fast growing area in the visualization research:

- Theories on visualisation and interaction provenance
- Provenance capture and visualization,
- Provenance analysis to
  - Understand users, for example for evaluating visual analytics tools, and
  - Support user sensemaking tasks such as providing personalization and helping collaboration.



Related Work

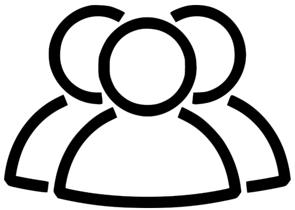
Provenance has been studied in many fields, often under different names

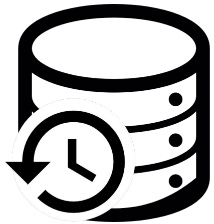
Human-Computer Interaction

- "Protocols", such as audio/video recording, computer logs, and user notebooks
- To understand user behaviors and intentions

Database, semantic web, and e-Science

- "Data lineage" and "Data provenance"
- For process debugging, data quality, and accountability





Reproducible science: make scientific experiments "repeatable" and "re-usable"

Scope of Survey

- Analysis of user interactions and provenance data in the field of visualization
  - Similar to meta-analysis as defined by *Ragan et al.*

✓ In Scope

User-generated (interaction) provenance with the goal of

- ✓ Improving
- ✓ Enhancing
- ✓ Understanding

... visual analysis system, visualization process, or visual artifact

✗ Out of Scope

- ✗ Only recorded information of interactive visual analysis session(s) without analysis
- ✗ Machine learning / Active learning based on binary decisions
- ✗ User studies without additional analysis of provenance information (beyond recording and sharing)

Ragan E. D., Endert A., Sanyal J., Chen J.: Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. IEEE Transactions on Visualization and Computer Graphics (2015).

Survey Methodology

Corpus

- Last eleven years
  2009 - 2019

- Issues and Processings
  4 Journals
  5 Conferences / Symposia

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
CG&A	3	1	-	-	-	1	2	-	-	-	3
CHI	-	-	1	2	-	-	2	-	-	-	4
EuroVis	-	-	-	-	1	-	2	2	1	3	5
IUI	1	-	-	-	1	1	-	1	3	2	3
TIIS	-	-	-	-		2	-	1	1	1	
TIST	-	1	-	-	-	-	-	-	-	-	-
TVCG	-	-	1	2	3	1	1	1	-	2	1
UIST	-	-	-	-	-	1	1	1	1	-	1
VIS	5	1	-	3	4	3	3	7	3	6	2

Sum ∑	**9**	**3**	**2**	**7**	**9**	**9**	**11**	**13**	**9**	**14**	**19**

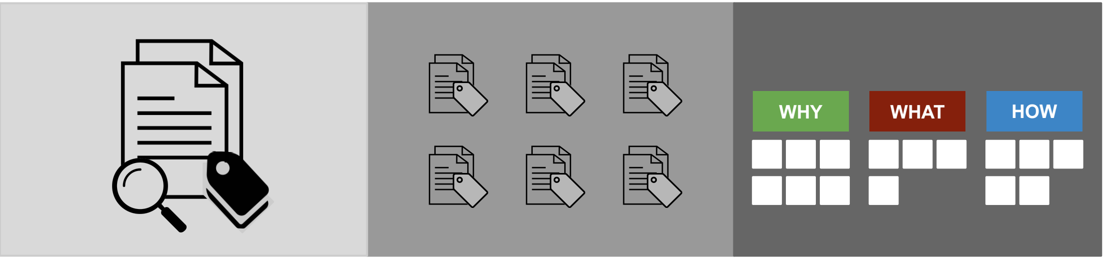
Survey Methodology

Coding Process

- Three stage approach for tagging
  1 - Explorative
  2 - Categorical
  3 - Supplementary

- Paper collection
  105 papers in total that are in-scope

1 Explorative	2 Categorical	3 Supplementary

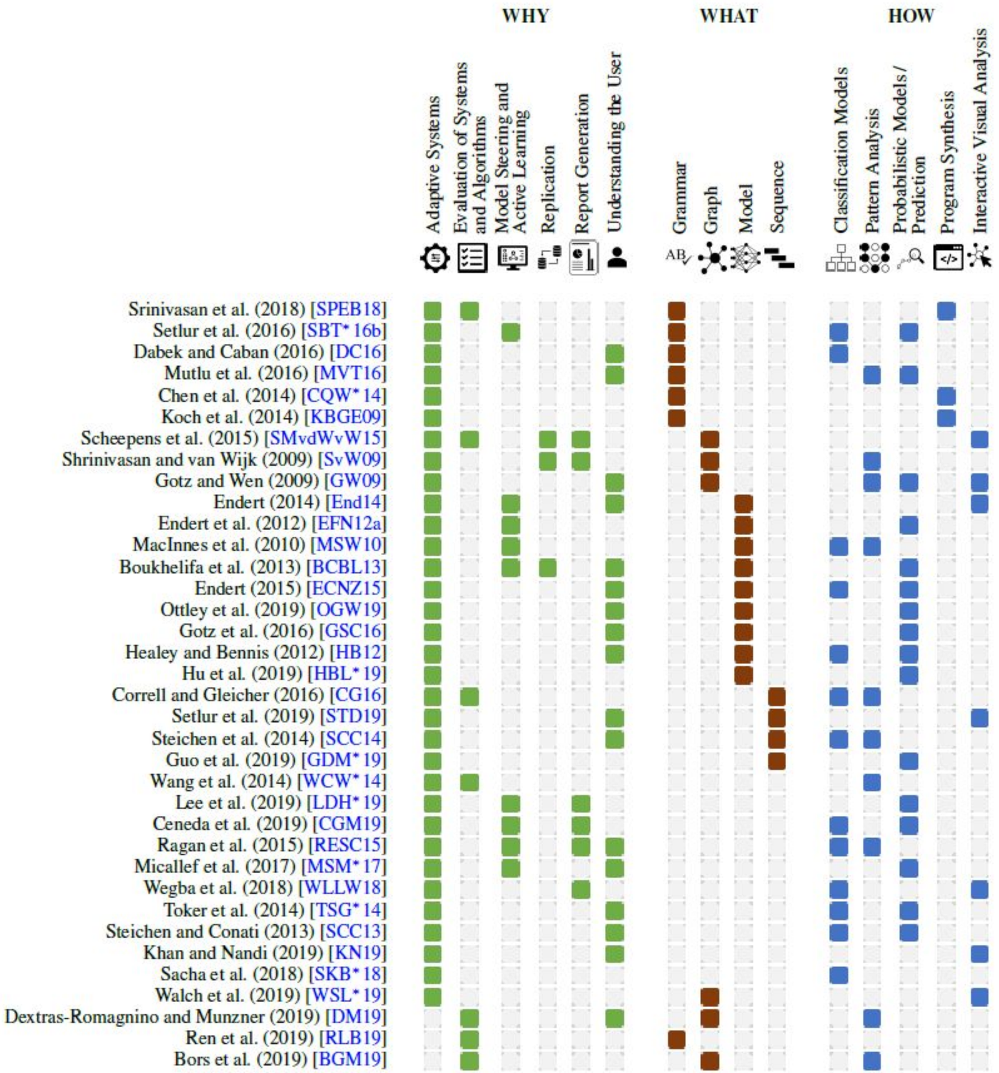


Goal Overview of relevant literature	**Goal** Unify categories and narrow down scope	**Goal** Distinct (sub-)categories
Example Keywords Provenance Analytics, Model Steering	**Example Keywords** WHY doing meta-analysis, HOW to analyze	**Example Keywords** WHY (6) WHAT (4) HOW (5)

Collection of Publications

Guiding Questions

- WHY analyze provenance data?
- WHAT types of provenance data and ways to encode it?
- HOW to analyze provenance data?



Analysis of User Interactions and Visualization Provenance

A companion website for the STAR Report on the Analysis of User Interactions and Visualization Provenance.

HOME | TECHNIQUES | OVERVIEW

About

This is a companion website for a review article on the analysis of user interactions and visualization provenance.

There is fast-growing literature on provenance-related research, covering aspects such as its theoretical framework, use cases, and techniques for capturing, visualizing, and analyzing provenance data. As a result, there is an increasing need to identify and taxonomize the existing scholarship. Such an organization of the research landscape will provide a complete picture of the current state of inquiry and identify knowledge gaps or possible avenues for further investigation. In this STAR, we aim to produce a comprehensive survey of work in the data visualization and visual analytics field that focus on the analysis of user interaction and provenance data. We structure our survey around three primary questions: (1) WHY analyze provenance data, (2) WHAT provenance data to encode and how to encode it, and (3) HOW to analyze provenance data. A concluding discussion provides evidence-based guidelines and highlights concrete opportunities for future development in this emerging area.

Browse through the techniques illustrated below, or use our wizard to find potential reasons for performing provenance analysis, the types of provenance data and encodings, and finally how to analyze them!

Get in touch if you have questions or comments.

Use the Interactive Table

Sort the paper collection according to your requirements!

Navigate to the wizard tab and select your WHY, WHAT, and HOW and receive a full list of publications on provenance analytics that are best suited to your selection.

Read the Article

Survey on the Analysis of User Interactions and Visualization Provenance

Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch

To appear in Computer Graphics Forum (EuroVis 2020)

Slides from the EuroVis 2020 Tutorial

Keynote Video

PDF Format

Learn about the WHY, WHAT, and HOW!

Click on a layout or operation to learn more!

https://provenance-survey.caleydo.org

Analysis of User Interactions and Visualization Provenance

A companion website for the STAR Report on the Analysis of User Interactions and Visualization Provenance.

HOME TECHNIQUES OVERVIEW

WHY analyze provenance data

The spectrum of possible reasons for conducting meta-analyses on provenance data is broad. Our goal is to provide a comprehensive overview of existing body of literature that analyze provenance datafor specific purposes. At a high-level, we can categorize the goals of the existing work as:

Understanding the User

"The goal of visualization is to create visual tools to support the user's reasoning and decision-making with data."

Evaluation of System and Algorithms

"Evaluation of a visual design or system with the primary purpose of non-trivial analysis of provenance data."

Adaptive Systems

"A better understanding of the system and the user's analytic process give rise to opportunities to create adaptive systems."

Model Steering

"Modeling steering leverages provenance data to improve the underlying data representations, machine learning models, or projection calculations in the case of high-dimensional datasets."

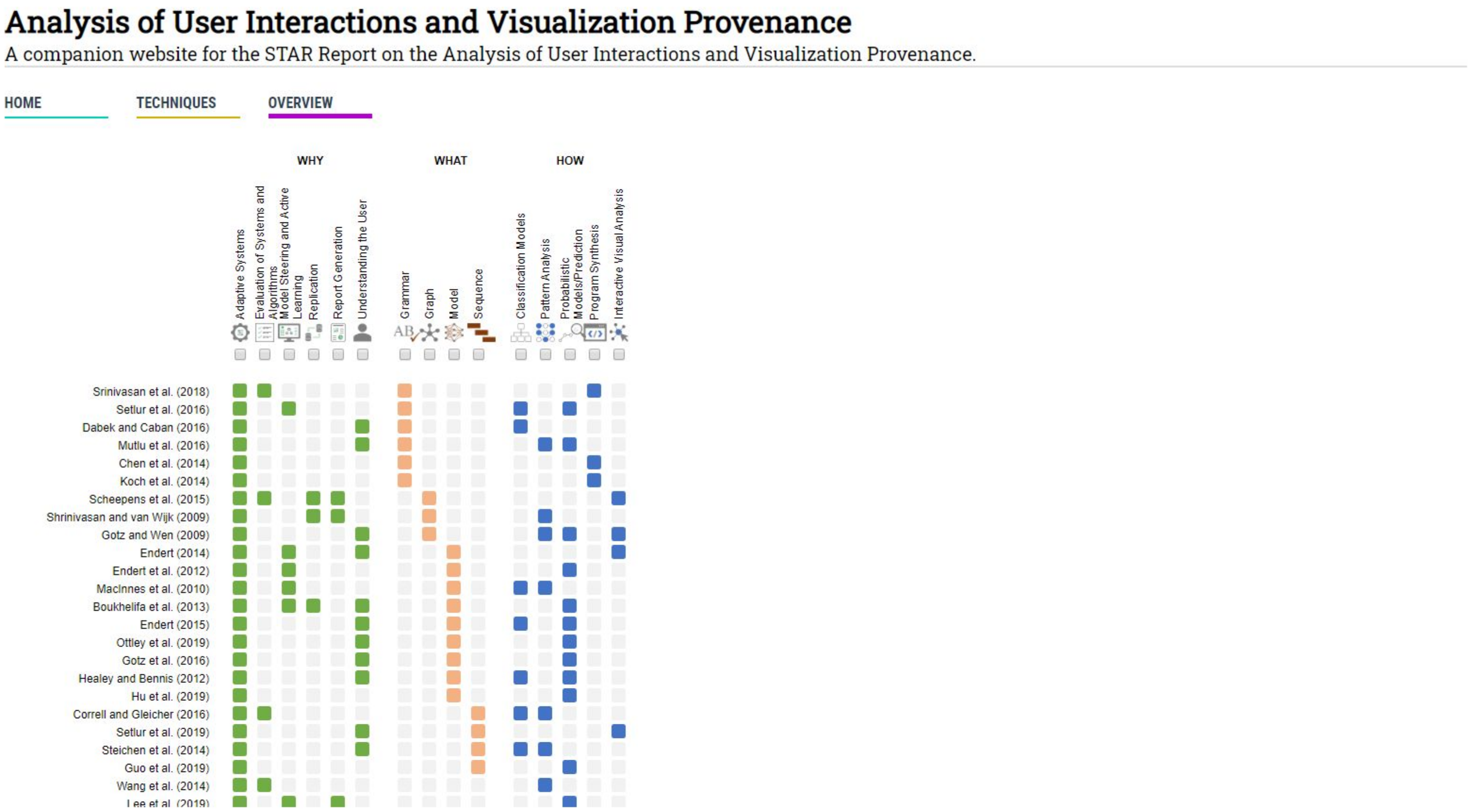
Replication, Verifiction, and Re-Application

"Using interaction logs to perform real-time or post-hoc quantification to validate the analysis results or to replicate the process when a similar problem arises."

Report Generation and Storytelling

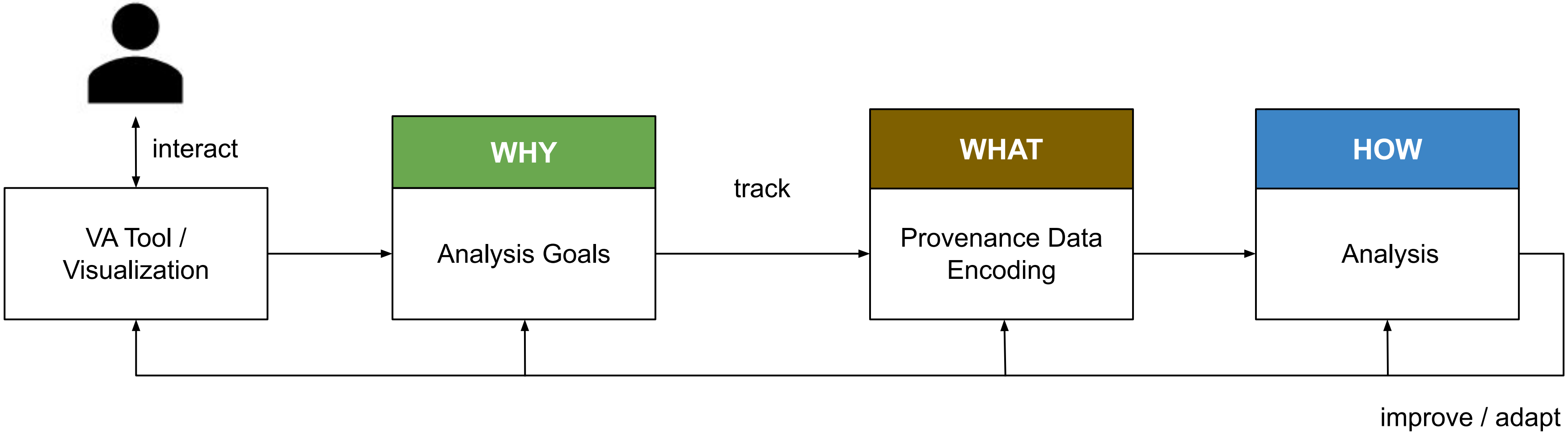
"Provenance data are used to automatically generate summary reports of an analysis session."

https://provenance-survey.caleydo.org



https://provenance-survey.caleydo.org

Structure of the Survey



The structure of our survey is based on a high-level provenance analysis model.

Goals: WHY Analyze Provenance Data (Alvitta)

Understanding the User

Evaluation of System and Algorithms

Adaptive Systems

Model Steering

Replication, Verification, and Re-Application

Report Generation and Storytelling

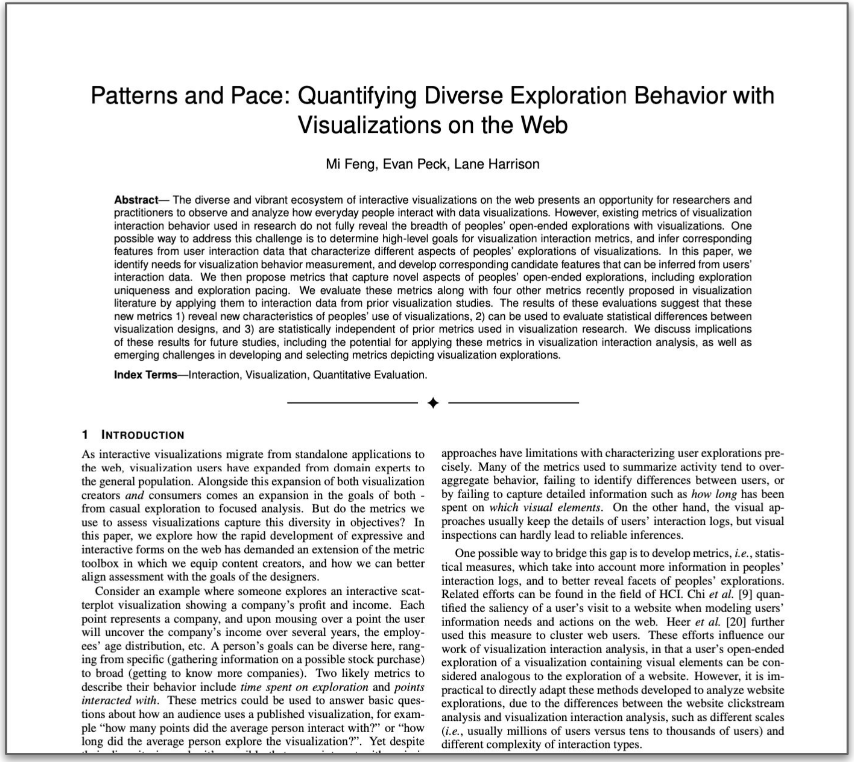
Understanding the User

Create theoretical and computational models that describe the reasoning process.
Examples include:

Understanding the User

Create theoretical and computational models that describe the reasoning process. Examples include:

- Quantifying hover and click patterns



Patterns and Pace: Quantifying Diverse Exploration Behavior with Visualizations on the Web

Mi Feng, Evan Peck, Lane Harrison

Abstract— The diverse and vibrant ecosystem of interactive visualizations on the web presents an opportunity for researchers and practitioners to observe and analyze how everyday people interact with data visualizations. However, existing metrics of visualization interaction behavior used in research do not fully reveal the breadth of peoples' open-ended explorations with visualizations. One possible way to address this challenge is to determine high-level goals for visualization interaction metrics, and infer corresponding features from user interaction data that characterize different aspects of peoples' explorations of visualizations. In this paper, we identify needs for visualization behavior measurement, and develop corresponding candidate features that can be inferred from users' interaction data. We then propose metrics that capture novel aspects of peoples' open-ended explorations, including exploration uniqueness and exploration pacing. We evaluate these metrics along with four other metrics recently proposed in visualization literature by applying them to interaction data from prior visualization studies. The results of these evaluations suggest that these new metrics 1) reveal new characteristics of peoples' use of visualizations, 2) can be used to evaluate statistical differences between visualization designs, and 3) are statistically independent of prior metrics used in visualization research. We discuss implications of these results for future studies, including the potential for applying these metrics in visualization interaction analysis, as well as emerging challenges in developing and selecting metrics depicting visualization explorations.

Index Terms—Interaction, Visualization, Quantitative Evaluation.

1 Introduction

As interactive visualizations migrate from standalone applications to the web, visualization users have expanded from domain experts to the general population. Alongside this expansion of both visualization creators *and* consumers comes an expansion in the goals of both - from casual exploration to focused analysis. But do the metrics we use to assess visualizations capture this diversity in objectives? In this paper, we explore how the rapid development of expressive and interactive forms on the web has demanded an extension of the metric toolbox in which we equip content creators, and how we can better align assessment with the goals of the designers.

Consider an example where someone explores an interactive scatterplot visualization showing a company's profit and income. Each point represents a company, and upon mousing over a point the user will uncover the company's income over several years, the employees' age distribution, etc. A person's goals can be diverse here, ranging from specific (gathering information on a possible stock purchase) to broad (getting to know more companies). Two likely metrics to describe their behavior include *time spent on exploration* and *points interacted with*. These metrics could be used to answer basic questions about how an audience uses a published visualization, for example "how many points did the average person interact with?" or "how long did the average person explore the visualization?". Yet despite

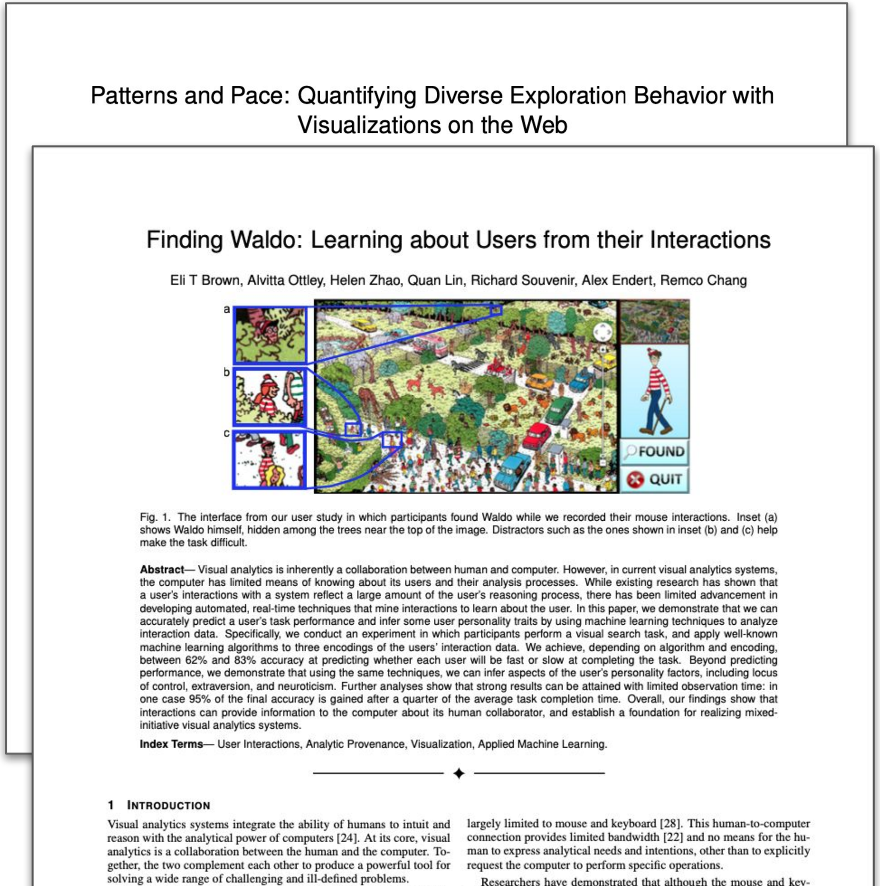
approaches have limitations with characterizing user explorations precisely. Many of the metrics used to summarize activity tend to over-aggregate behavior, failing to identify differences between users, or by failing to capture detailed information such as *how long* has been spent on *which visual elements*. On the other hand, the visual approaches usually keep the details of users' interaction logs, but visual inspections can hardly lead to reliable inferences.

One possible way to bridge this gap is to develop metrics, *i.e.*, statistical measures, which take into account more information in peoples' interaction logs, and to better reveal facets of peoples' explorations. Related efforts can be found in the field of HCI. Chi *et al.* [9] quantified the saliency of a user's visit to a website when modeling users' information needs and actions on the web. Heer *et al.* [20] further used this measure to cluster web users. These efforts influence our work of visualization interaction analysis, in that a user's open-ended exploration of a visualization containing visual elements can be considered analogous to the exploration of a website. However, it is impractical to directly adapt these methods developed to analyze website explorations, due to the differences between the website clickstream analysis and visualization interaction analysis, such as different scales (*i.e.*, usually millions of users versus tens to thousands of users) and different complexity of interaction types.

Understanding the User

Create theoretical and computational models that describe the reasoning process.
Examples include:

- Quantifying hover and click patterns
- Predicting personality traits



Patterns and Pace: Quantifying Diverse Exploration Behavior with Visualizations on the Web

Finding Waldo: Learning about Users from their Interactions

Eli T Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, Remco Chang

Fig. 1. The interface from our user study in which participants found Waldo while we recorded their mouse interactions. Inset (a) shows Waldo himself, hidden among the trees near the top of the image. Distractors such as the ones shown in inset (b) and (c) help make the task difficult.

Abstract— Visual analytics is inherently a collaboration between human and computer. However, in current visual analytics systems, the computer has limited means of knowing about its users and their analysis processes. While existing research has shown that a user's interactions with a system reflect a large amount of the user's reasoning process, there has been limited advancement in developing automated, real-time techniques that mine interactions to learn about the user. In this paper, we demonstrate that we can accurately predict a user's task performance and infer some user personality traits by using machine learning techniques to analyze interaction data. Specifically, we conduct an experiment in which participants perform a visual search task, and apply well-known machine learning algorithms to three encodings of the users' interaction data. We achieve, depending on algorithm and encoding, between 62% and 83% accuracy at predicting whether each user will be fast or slow at completing the task. Beyond predicting performance, we demonstrate that using the same techniques, we can infer aspects of the user's personality factors, including locus of control, extraversion, and neuroticism. Further analyses show that strong results can be attained with limited observation time: in one case 95% of the final accuracy is gained after a quarter of the average task completion time. Overall, our findings show that interactions can provide information to the computer about its human collaborator, and establish a foundation for realizing mixed-initiative visual analytics systems.

Index Terms— User Interactions, Analytic Provenance, Visualization, Applied Machine Learning.

1 INTRODUCTION

Visual analytics systems integrate the ability of humans to intuit and reason with the analytical power of computers [24]. At its core, visual analytics is a collaboration between the human and the computer. Together, the two complement each other to produce a powerful tool for solving a wide range of challenging and ill-defined problems.

largely limited to mouse and keyboard [28]. This human-to-computer connection provides limited bandwidth [22] and no means for the human to express analytical needs and intentions, other than to explicitly request the computer to perform specific operations.

Researchers have demonstrated that although the mouse and key-

Understanding the User

Create theoretical and computational models that describe the reasoning process. Examples include:

- Quantifying hover and click patterns
- Predicting personality traits
- Uncovering exploration bias

Patterns and Pace: Quantifying Diverse Exploration Behavior with Visualizations on the Web

Finding Waldo: Learning about Users from their Interactions

Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics

Emily Wall* Georgia Tech

Leslie M. Blaha† Pacific Northwest National Laboratory

Lyndsey Franklin‡ Pacific Northwest National Laboratory

Alex Endert§ Georgia Tech

ABSTRACT

Visual analytic tools combine the complementary strengths of humans and machines in human-in-the-loop systems. Humans provide invaluable domain expertise and sensemaking capabilities to this discourse with analytic models; however, little consideration has yet been given to the ways inherent human biases might shape the visual analytic process. In this paper, we establish a conceptual framework for considering bias assessment through human-in-the-loop systems and lay the theoretical foundations for bias measurement. We propose six preliminary metrics to systematically detect and quantify bias from user interactions and demonstrate how the metrics might be implemented in an existing visual analytic system, InterAxis. We discuss how our proposed metrics could be used by visual analytic systems to mitigate the negative effects of cognitive biases by making users aware of biased processes throughout their analyses.

Keywords: cognitive bias; visual analytics; human-in-the-loop; mixed initiative; user interaction;

Index Terms: H.5.0 [Information Systems]: Human-Computer Interaction—General

1 INTRODUCTION

Visual analytic systems gracefully blend sophisticated data analytics with interactive visualizations to provide usable interfaces through which people explore data [43, 71]. User interaction is central to the effectiveness of visual analytic systems [21, 56, 80]. It is the mechanism by which people and systems communicate about the data, representations, and to augment models with valuable subject matter expertise. These human-in-the-loop (HIL) approaches enable insights in many domains, especially where uncertainty is high and human reasoning is a valuable addition to data-intensive computation [20].

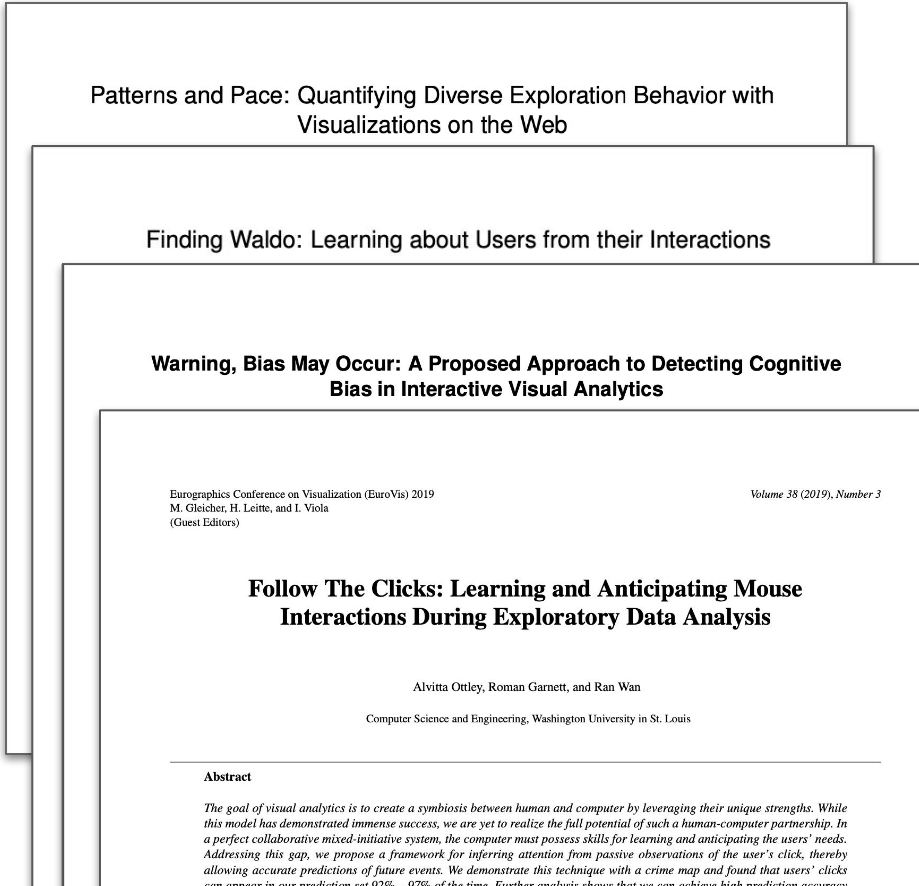
However, incorporating human reasoning and analysis into computational models may have unwanted side effects. Prior work in cognitive psychology informs us that there are inherent limitations to cognitive processes, such as working memory capacity limits [11, 51]. One limitation relevant to analytic processes and visual data analysis is cognitive bias, errors resulting from the use of fallible decision making heuristics [29, 42]. Evidence that cognitive biases impact users' decision making abounds; recent work has shown that information visualization users are not immune to cognitive biases [13]. While bias might exist and be propagated through a system via data collection (e.g., convenience sampling bias), data processing (e.g., algorithm bias), visual mappings (e.g., visual perception bias), etc. [27, 64], here we focus on cognitive bias injected by analysts.

Several cognitive biases have been previously identified as particularly relevant to data analysis and the intelligence process [38] (see Table 1). Such biases can have far-reaching effects, influencing the evidence upon which analysts rely and the hypotheses they form. Further, when user interaction in visual analytic tools is intended to guide analytic models, cognitive biases might be propagated to and amplified by the underlying computational models. The resulting biased analytic models may ultimately prompt analysts to make incorrect or inferior decisions, or simply echo the users' biases back

Understanding the User

Create theoretical and computational models that describe the reasoning process.
Examples include:

- Quantifying hover and click patterns
- Predicting personality traits
- Uncovering exploration bias
- Modeling attention



Patterns and Pace: Quantifying Diverse Exploration Behavior with Visualizations on the Web

Finding Waldo: Learning about Users from their Interactions

Warning, Bias May Occur: A Proposed Approach to Detecting Cognitive Bias in Interactive Visual Analytics

Eurographics Conference on Visualization (EuroVis) 2019
M. Gleicher, H. Leitte, and I. Viola
(Guest Editors)

Volume 38 (2019), Number 3

Follow The Clicks: Learning and Anticipating Mouse Interactions During Exploratory Data Analysis

Alvitta Ottley, Roman Garnett, and Ran Wan

Computer Science and Engineering, Washington University in St. Louis

Abstract

The goal of visual analytics is to create a symbiosis between human and computer by leveraging their unique strengths. While this model has demonstrated immense success, we are yet to realize the full potential of such a human-computer partnership. In a perfect collaborative mixed-initiative system, the computer must possess skills for learning and anticipating the users' needs. Addressing this gap, we propose a framework for inferring attention from passive observations of the user's click, thereby allowing accurate predictions of future events. We demonstrate this technique with a crime map and found that users' clicks

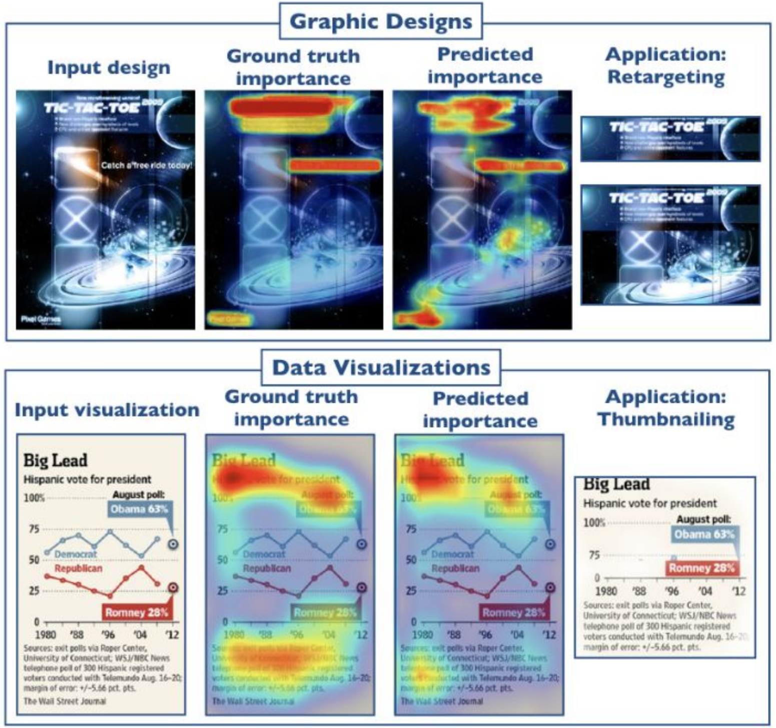
Evaluation of System and Algorithms

Prior work has used provenance data to understand the visualization system itself and to evaluate its usefulness. These include:

Evaluation of System and Algorithms

Prior work has used provenance data to understand the visualization system itself and to evaluate its usefulness. These include:

- Using mouse and eye data to learn the importance of visual elements

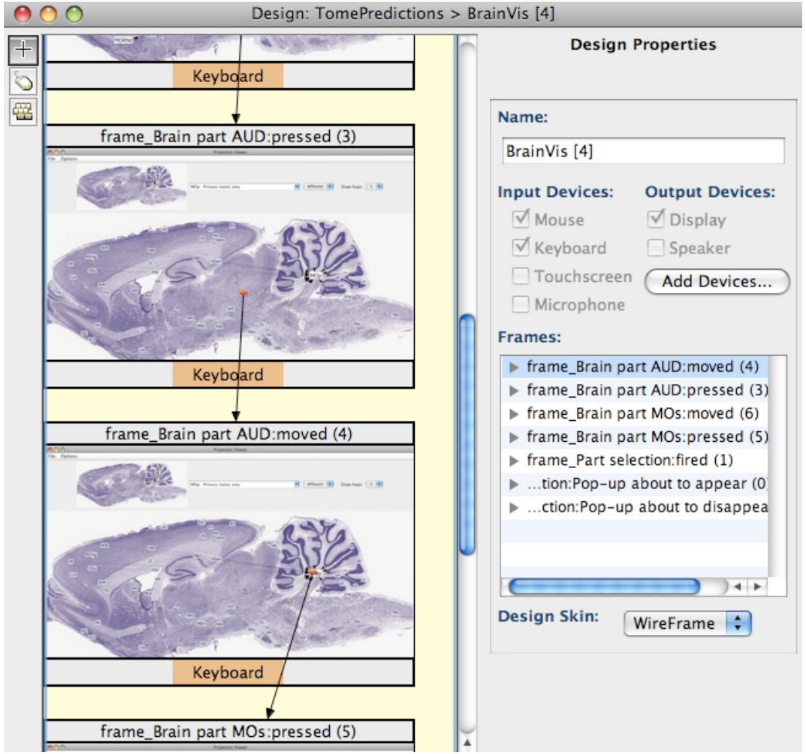


Bylinskii et al. 2017

Evaluation of System and Algorithms

Prior work has used provenance data to understand the visualization system itself and to evaluate its usefulness. These include:

- Using mouse and eye data to learn the importance of visual elements
- Modeling task performance to guide system designs

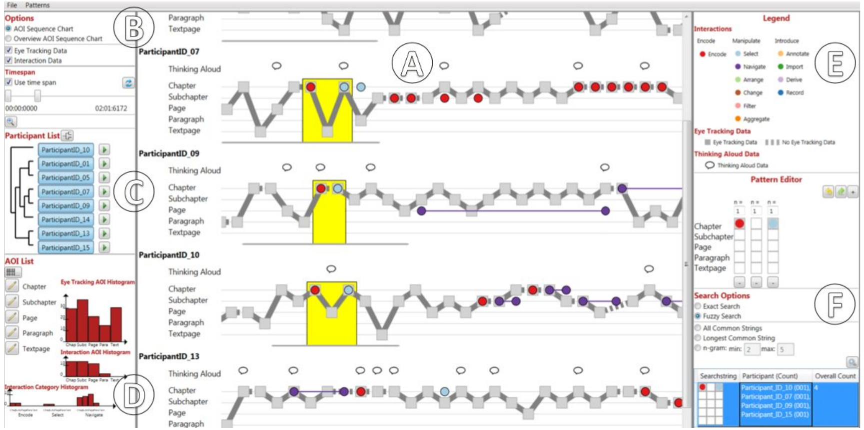


Gomez and Laidlaw 2012

Evaluation of System and Algorithms

Prior work has used provenance data to understand the visualization system itself and to evaluate its usefulness. These include:

- Using mouse and eye data to learn the importance of visual elements
- Modeling task performance to guide system designs
- VA systems for evaluating interactive visualizations



Blascheck et al. 2016

Adaptive Systems

Aim to improve the usability and performance of a visualization system. Topics ranged from:

Adaptive Systems

Aim to improve the usability and performance of a visualization system

Topics ranged from:

- Recommenders

Behavior-Driven Visualization Recommendation

David Gotz
IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10578 USA
dgotz@us.ibm.com

Zhen Wen
IBM T.J. Watson Research Center
19 Skyline Drive
Hawthorne, NY 10578 USA
zhenwen@us.ibm.com

ABSTRACT
We present a novel approach to visualization recommendation that monitors user behavior for implicit signals of user intent to provide more effective recommendation. This is in contrast to previous approaches which are either insensitive to user intent or require explicit, user specified task information. Our approach, called *Behavior-Driven Visualization Recommendation* (*BDVR*), consists of two distinct phases: (1) pattern detection, and (2) visualization recommendation. In the first phase, user behavior is analyzed dynamically to find semantically meaningful interaction patterns using a library of pattern definitions developed through observations of real-world visual analytic activity. In the second phase, our BDVR algorithm uses the detected patterns to infer a user's intended visual task. It then automatically suggests alternative visualizations that support the inferred visual task more directly than the user's current visualization. We present the details of BDVR and describe its implementation within our lab's prototype visual analysis system. We also present study results that demonstrate that our approach shortens task completion time and reduces error rates when compared to behavior-agnostic recommendation.

Author Keywords
Intelligent visualization, Information visualization, User behavior modeling, Visualization recommendation

ACM Classification Keywords
Algorithms, Human Factors

INTRODUCTION
Visualization has long been used to harness the power of human perception to uncover insights from large collections of data. However, it is impossible to create a "one-size-fits-all" technique for visualizing data because every task and data

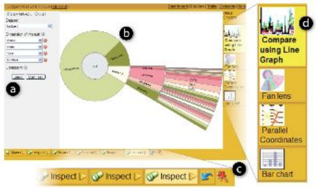


Figure 1. Behavior-driven visualization recommendation has been integrated into our lab's visualization system. Users can (a) issue queries and (b) interact with visualizations to analyze data. When a new recommendation is provided due to a user's behavior, he/she is notified via (c) a magic wand icon in the history panel and (d) a flashing segment on the recommendation sidebar. Users can accept the recommendation with a single click, or ignore it to continue uninterrupted.

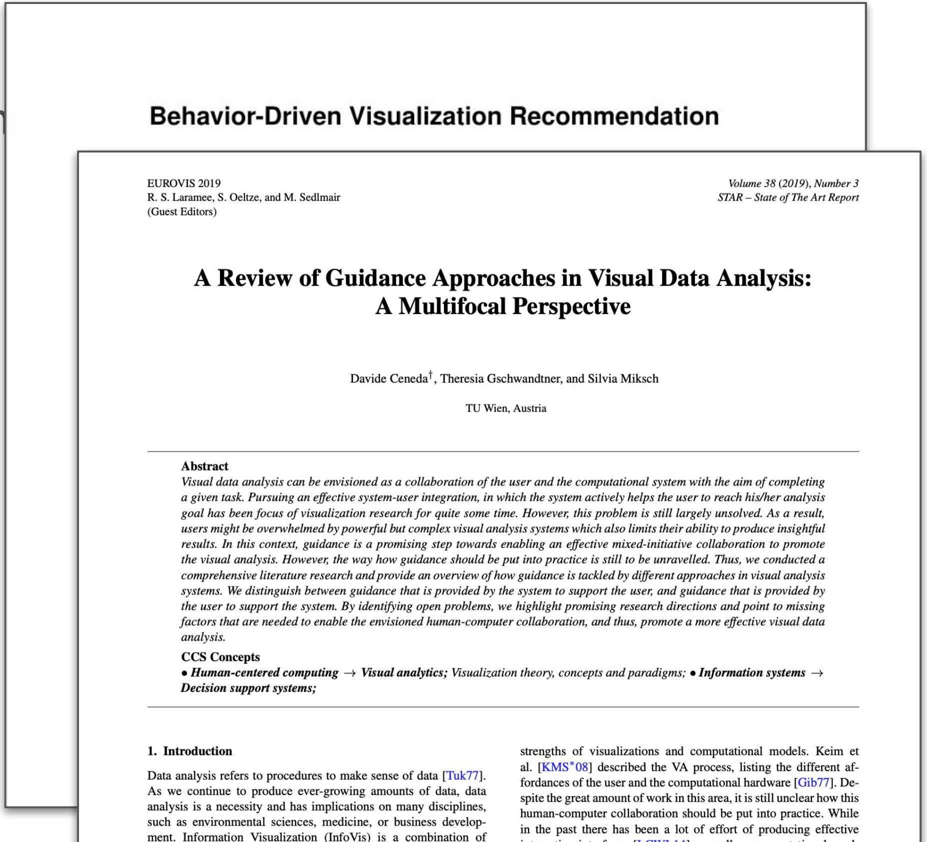
Given the variety of options, how and when to use a particular visual metaphor requires a significant level of visual literacy. Unfortunately, average business users don't typically posses these skills. While domain experts within their own area, they usually have little or no training in visualization. Companies must therefore hire professional analysts (with visualization and analysis skills, but little domain knowledge) to generate reports that are in turn used by business-line employees to make decisions. This dramatically increases the cost of visualization-based solutions and places them beyond the reach of the legions of business users who might otherwise benefit from their capabilities.

Recognizing the challenge of supporting average users, several visualization systems have integrated intelligent algorithms to automatically compose or recommend effective vis-

Adaptive Systems

Aim to improve the usability and performance of a visualization system. Topics ranged from:

- Recommenders
- Providing guidance to the user



Behavior-Driven Visualization Recommendation

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;***

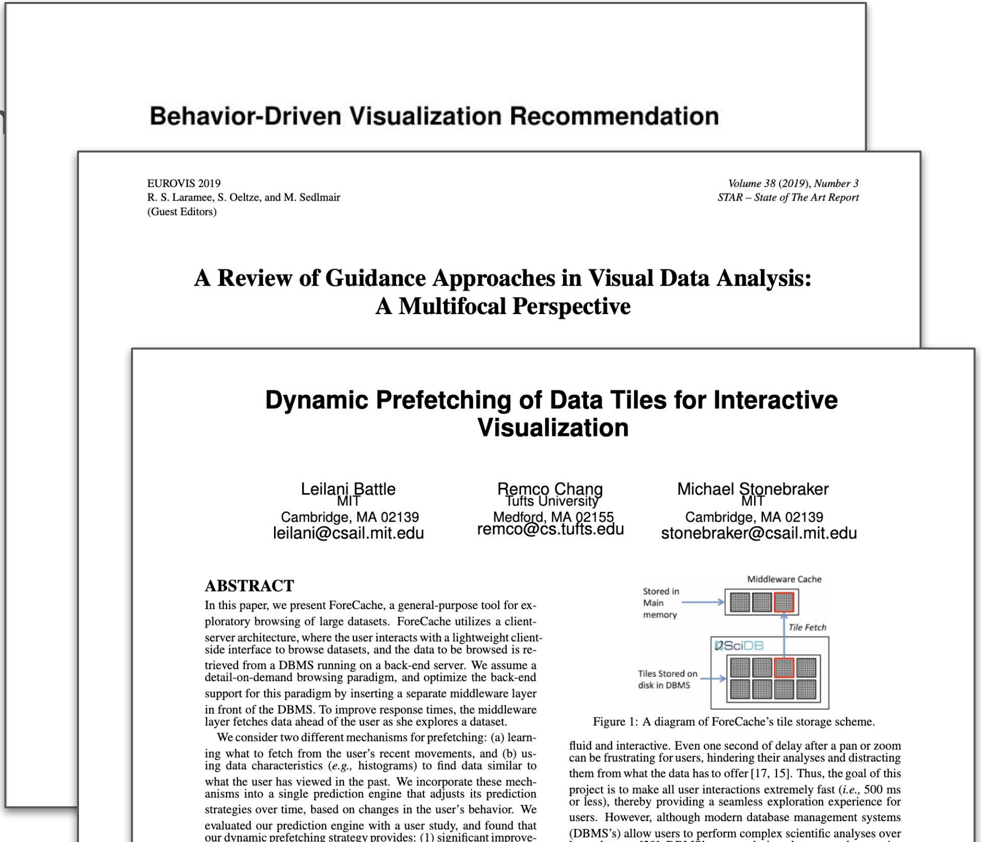
1. Introduction

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 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

Adaptive Systems

Aim to improve the usability and performance of a visualization system
Topics ranged from:

- Recommenders
- Providing guidance to the user
- Adaptive prefetching


Behavior-Driven Visualization Recommendation

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

Dynamic Prefetching of Data Tiles for Interactive Visualization

Leilani Battle
MIT
Cambridge, MA 02139
leilani@csail.mit.edu

Remco Chang
Tufts University
Medford, MA 02155
remco@cs.tufts.edu

Michael Stonebraker
MIT
Cambridge, MA 02139
stonebraker@csail.mit.edu

ABSTRACT

In this paper, we present ForeCache, a general-purpose tool for exploratory browsing of large datasets. ForeCache utilizes a client-server architecture, where the user interacts with a lightweight client-side interface to browse datasets, and the data to be browsed is retrieved from a DBMS running on a back-end server. We assume a detail-on-demand browsing paradigm, and optimize the back-end support for this paradigm by inserting a separate middleware layer in front of the DBMS. To improve response times, the middleware layer fetches data ahead of the user as she explores a dataset.

We consider two different mechanisms for prefetching: (a) learning what to fetch from the user's recent movements, and (b) using data characteristics (*e.g.,* histograms) to find data similar to what the user has viewed in the past. We incorporate these mechanisms into a single prediction engine that adjusts its prediction strategies over time, based on changes in the user's behavior. We evaluated our prediction engine with a user study, and found that our dynamic prefetching strategy provides: (1) significant improve-

Figure 1: A diagram of ForeCache's tile storage scheme.

fluid and interactive. Even one second of delay after a pan or zoom can be frustrating for users, hindering their analyses and distracting them from what the data has to offer [17, 15]. Thus, the goal of this project is to make all user interactions extremely fast (*i.e.,* 500 ms or less), thereby providing a seamless exploration experience for users. However, although modern database management systems (DBMS's) allow users to perform complex scientific analyses over

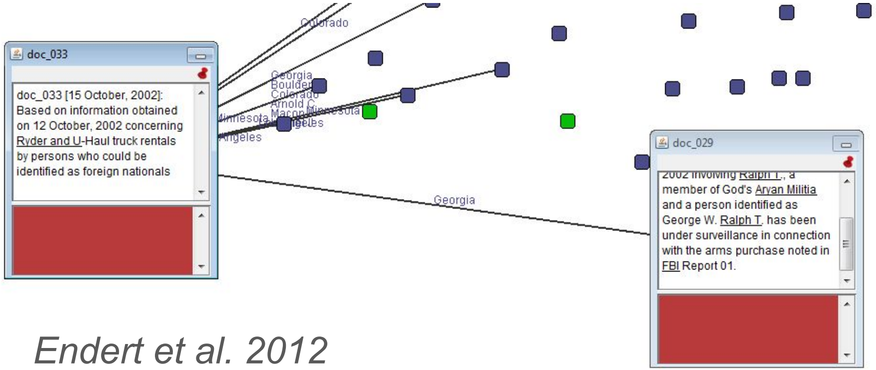
Model Steering

Modeling steering uses provenance data to improve the underlying data representations, machine learning models, or projection calculations. This category includes system such as:

Model Steering

Modeling steering uses provenance data to improve the underlying data representations, machine learning models, or projection calculations. This category includes system such as:

- ForceSpire

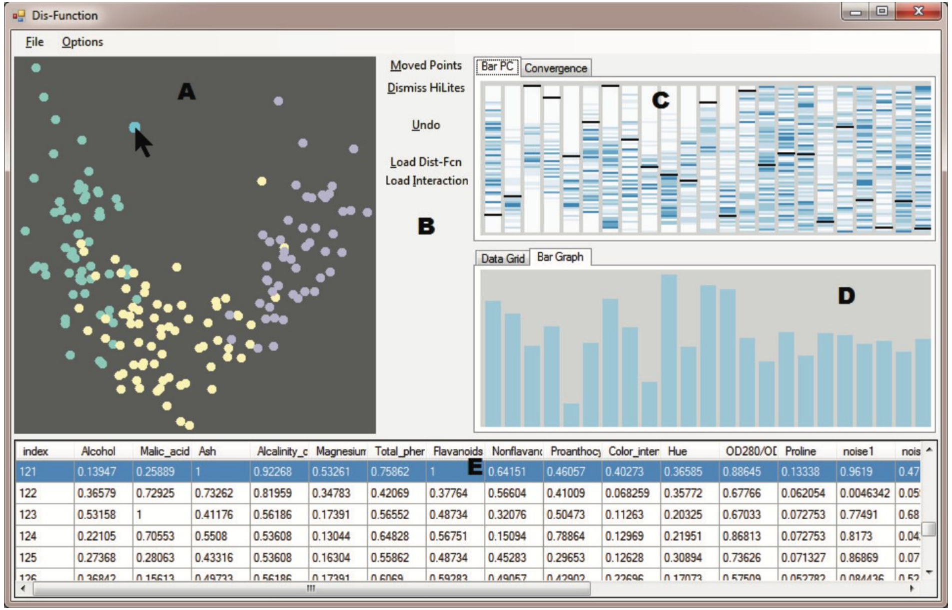


Endert et al. 2012

Model Steering

Modeling steering uses provenance data to improve the underlying data representations, machine learning models, or projection calculations. This category includes system such as:

- ForceSpire
- Dis-Function



Brown et al. 2012

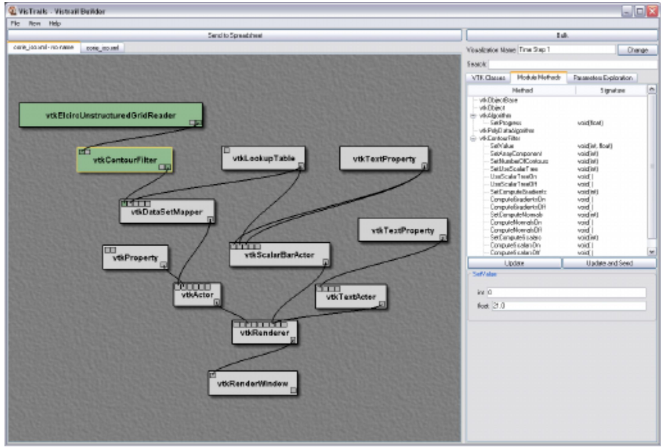
Replication, Verification, and Re-Application

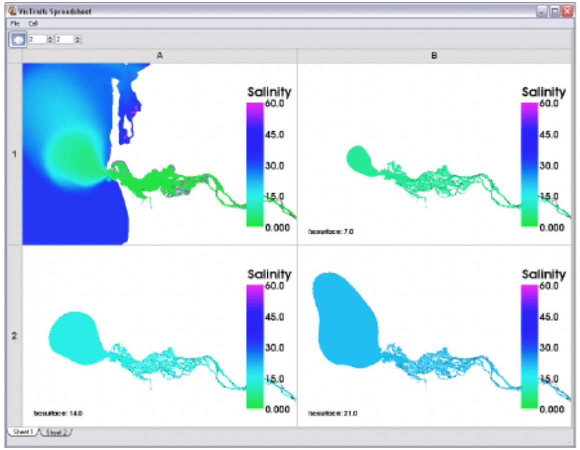
Another usage of provenance data is to verify, replicate or re-apply analysis sessions. For instance:

Replication, Verification, and Re-Application

Another usage of provenance data is to verify, replicate or re-apply analysis sessions. For instance:

- VisTrails



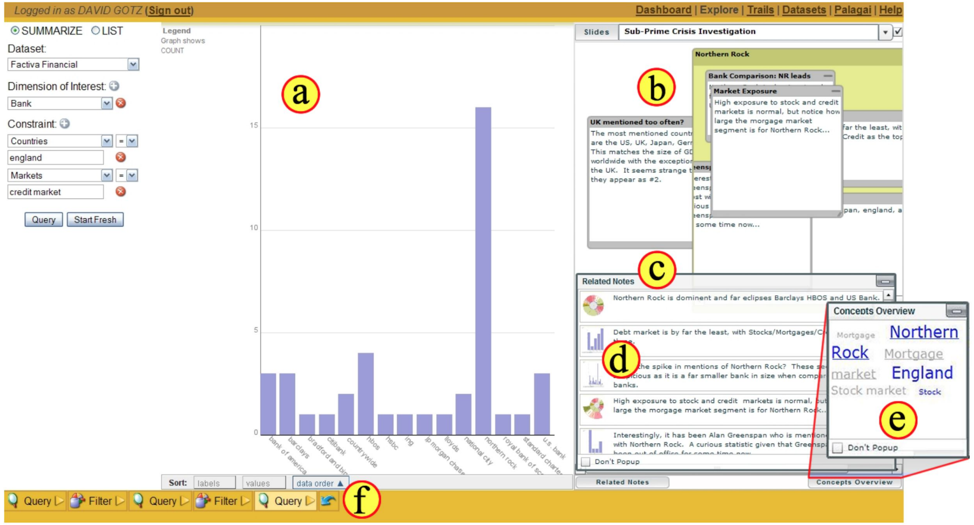


Callahan et al. 2012

Replication, Verification, and Re-Application

Another usage of provenance data is to verify, replicate or re-apply analysis sessions. For instance:

- VisTrails
- Harvest

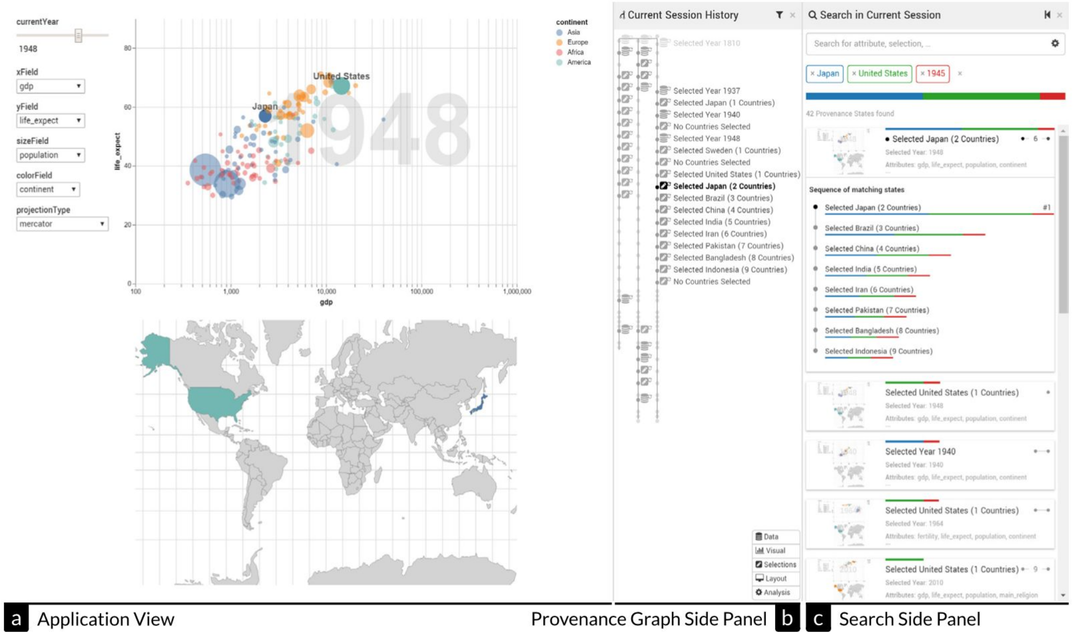


Shrinivasan et al. 2009

Replication, Verification, and Re-Application

Another usage of provenance data is to verify, replicate or re-apply analy sessions. For instance:

- VisTrails
- Harvest
- KnowledgePearls



Stitz et al. 2019

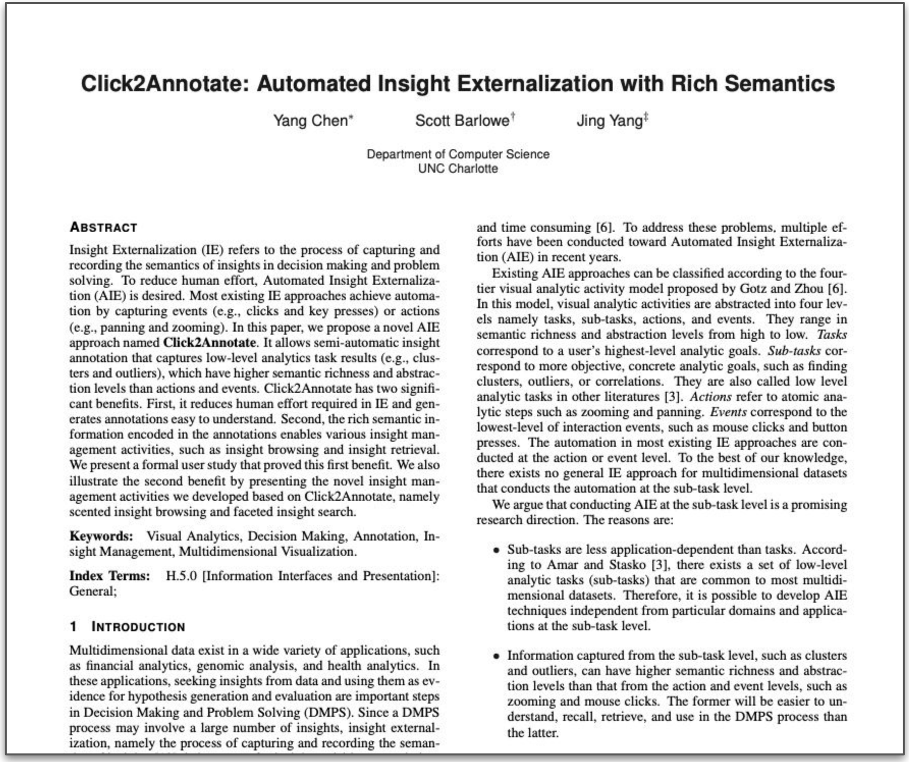
Report Generation and Storytelling

Finally, research has analyzed provenance data to automatically generate summary reports of an analysis session. We found papers on producing:

Report Generation and Storytelling

Finally, research has analyzed provenance data to automatically generate summary reports of an analysis session. We found papers on producing:

- Automated annotations



Click2Annotate: Automated Insight Externalization with Rich Semantics

Yang Chen* Scott Barlowe† Jing Yang‡

Department of Computer Science
UNC Charlotte

Abstract

Insight Externalization (IE) refers to the process of capturing and recording the semantics of insights in decision making and problem solving. To reduce human effort, Automated Insight Externalization (AIE) is desired. Most existing IE approaches achieve automation by capturing events (e.g., clicks and key presses) or actions (e.g., panning and zooming). In this paper, we propose a novel AIE approach named **Click2Annotate**. It allows semi-automatic insight annotation that captures low-level analytics task results (e.g., clusters and outliers), which have higher semantic richness and abstraction levels than actions and events. Click2Annotate has two significant benefits. First, it reduces human effort required in IE and generates annotations easy to understand. Second, the rich semantic information encoded in the annotations enables various insight management activities, such as insight browsing and insight retrieval. We present a formal user study that proved this first benefit. We also illustrate the second benefit by presenting the novel insight management activities we developed based on Click2Annotate, namely scented insight browsing and faceted insight search.

Keywords: Visual Analytics, Decision Making, Annotation, Insight Management, Multidimensional Visualization.

Index Terms: H.5.0 [Information Interfaces and Presentation]: General;

1 Introduction

Multidimensional data exist in a wide variety of applications, such as financial analytics, genomic analysis, and health analytics. In these applications, seeking insights from data and using them as evidence for hypothesis generation and evaluation are important steps in Decision Making and Problem Solving (DMPS). Since a DMPS process may involve a large number of insights, insight externalization, namely the process of capturing and recording the semantics of insights in decision making and problem solving, is tedious and time consuming [6]. To address these problems, multiple efforts have been conducted toward Automated Insight Externalization (AIE) in recent years.

Existing AIE approaches can be classified according to the four-tier visual analytic activity model proposed by Gotz and Zhou [6]. In this model, visual analytic activities are abstracted into four levels namely tasks, sub-tasks, actions, and events. They range in semantic richness and abstraction levels from high to low. *Tasks* correspond to a user's highest-level analytic goals. *Sub-tasks* correspond to more objective, concrete analytic goals, such as finding clusters, outliers, or correlations. They are also called low level analytic tasks in other literatures [3]. *Actions* refer to atomic analytic steps such as zooming and panning. *Events* correspond to the lowest-level of interaction events, such as mouse clicks and button presses. The automation in most existing IE approaches are conducted at the action or event level. To the best of our knowledge, there exists no general IE approach for multidimensional datasets that conducts the automation at the sub-task level.

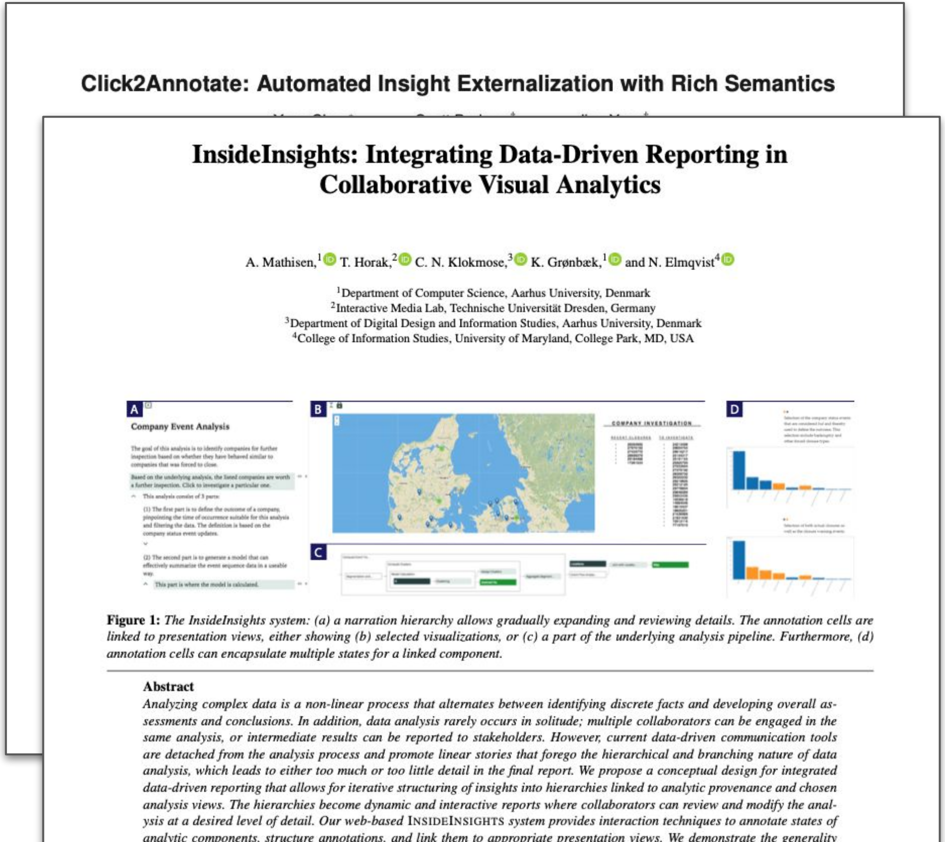
We argue that conducting AIE at the sub-task level is a promising research direction. The reasons are:

- Sub-tasks are less application-dependent than tasks. According to Amar and Stasko [3], there exists a set of low-level analytic tasks (sub-tasks) that are common to most multidimensional datasets. Therefore, it is possible to develop AIE techniques independent from particular domains and applications at the sub-task level.
- Information captured from the sub-task level, such as clusters and outliers, can have higher semantic richness and abstraction levels than that from the action and event levels, such as zooming and mouse clicks. The former will be easier to understand, recall, retrieve, and use in the DMPS process than the latter.

Report Generation and Storytelling

Finally, research has analyzed provenance data to automatically generate summary reports of an analysis session. We found papers on producing:

- Automated annotations
- Data-driven reports



Click2Annotate: Automated Insight Externalization with Rich Semantics

InsideInsights: Integrating Data-Driven Reporting in Collaborative Visual Analytics

A. Mathisen,[1] T. Horak,[2] C. N. Klokmose,[3] K. Grønbæk,[1] and N. Elmqvist[4]

[1]Department of Computer Science, Aarhus University, Denmark
[2]Interactive Media Lab, Technische Universität Dresden, Germany
[3]Department of Digital Design and Information Studies, Aarhus University, Denmark
[4]College of Information Studies, University of Maryland, College Park, MD, USA

Figure 1: *The InsideInsights system: (a) a narration hierarchy allows gradually expanding and reviewing details. The annotation cells are linked to presentation views, either showing (b) selected visualizations, or (c) a part of the underlying analysis pipeline. Furthermore, (d) annotation cells can encapsulate multiple states for a linked component.*

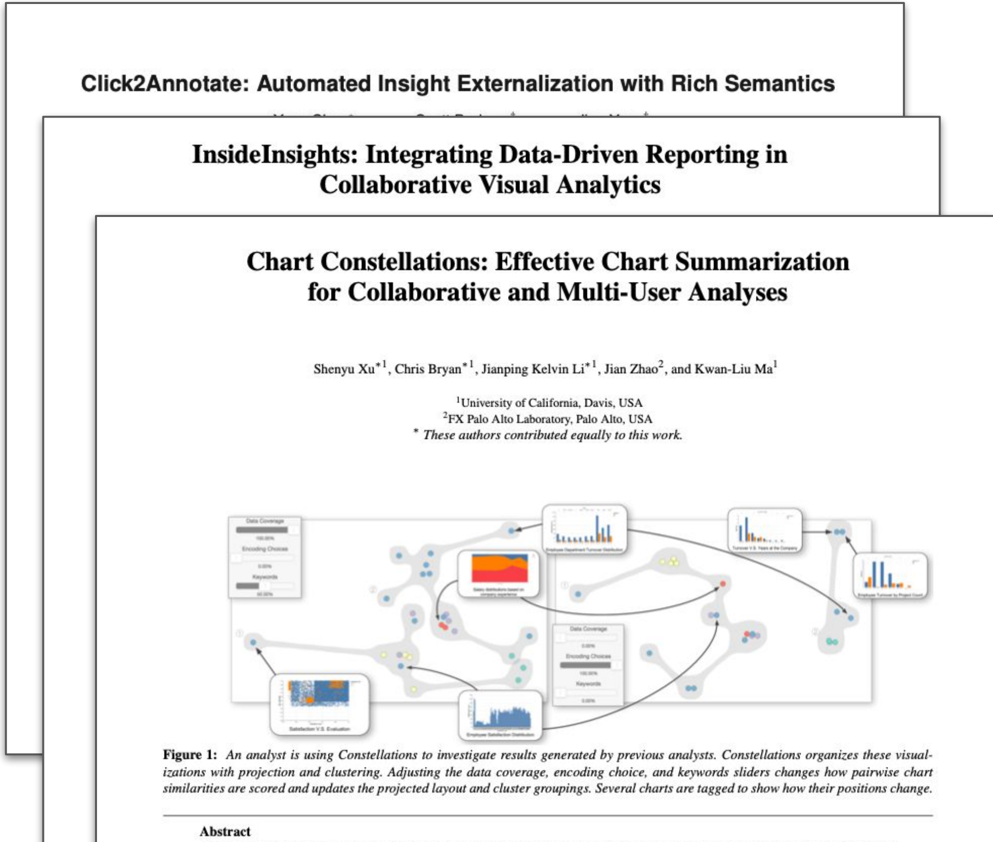
Abstract

Analyzing complex data is a non-linear process that alternates between identifying discrete facts and developing overall assessments and conclusions. In addition, data analysis rarely occurs in solitude; multiple collaborators can be engaged in the same analysis, or intermediate results can be reported to stakeholders. However, current data-driven communication tools are detached from the analysis process and promote linear stories that forego the hierarchical and branching nature of data analysis, which leads to either too much or too little detail in the final report. We propose a conceptual design for integrated data-driven reporting that allows for iterative structuring of insights into hierarchies linked to analytic provenance and chosen analysis views. The hierarchies become dynamic and interactive reports where collaborators can review and modify the analysis at a desired level of detail. Our web-based INSIDEINSIGHTS system provides interaction techniques to annotate states of analytic components, structure annotations, and link them to appropriate presentation views. We demonstrate the generality

Report Generation and Storytelling

Finally, research has analyzed provenance data to automatically generate summary reports of an analysis session. We found papers on producing:

- Automated annotations
- Data-driven reports
- Summary insights



Click2Annotate: Automated Insight Externalization with Rich Semantics

InsideInsights: Integrating Data-Driven Reporting in Collaborative Visual Analytics

Chart Constellations: Effective Chart Summarization for Collaborative and Multi-User Analyses

Shenyu Xu*[1], Chris Bryan*[1], Jianping Kelvin Li*[1], Jian Zhao[2], and Kwan-Liu Ma[1]

[1]University of California, Davis, USA
[2]FX Palo Alto Laboratory, Palo Alto, USA
* These authors contributed equally to this work.

Figure 1: *An analyst is using Constellations to investigate results generated by previous analysts. Constellations organizes these visualizations with projection and clustering. Adjusting the data coverage, encoding choice, and keywords sliders changes how pairwise chart similarities are scored and updates the projected layout and cluster groupings. Several charts are tagged to show how their positions change.*

Abstract

Encodings: WHAT Types of Provenance Data to Analyze

Sequence

Grammar

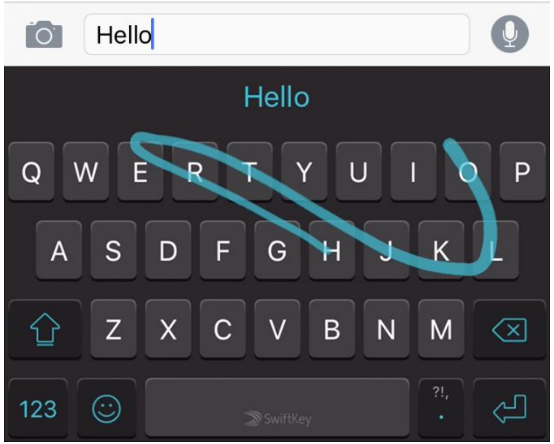
Model

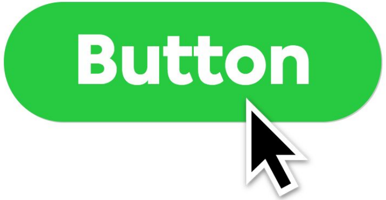
Graph

WHAT: Sequence

Sequence: Logs of Activities

- User Interaction





WHAT: Sequence

Sequence: Logs of Activities

- User Interaction
- Application State

escuela_access_log

```
200.90.177.197 - - [07/Nov/2004:04:14:17 -0300] "GET /novedades/rss/ HTTP/1.1" 200 3284
200.73.40.132 - - [07/Nov/2004:04:14:19 -0300] "GET / HTTP/1.1" 200 7550
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/escmovil1.jpg HTTP/1.1" 200 7748
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/novedades1.jpg HTTP/1.1" 200 7952
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/mapa1.jpg HTTP/1.1" 200 8035
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/lado1.jpg HTTP/1.1" 200 12964
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/lado2.jpg HTTP/1.1" 200 8340
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/escuela1.jpg HTTP/1.1" 200 7858
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/servicios1.jpg HTTP/1.1" 200 6835
200.73.40.132 - - [07/Nov/2004:04:14:20 -0300] "GET /img_index/departamentos1.jpg HTTP/1.1" 200 7900
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/instructivos1.jpg HTTP/1.1" 200 8597
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/calendarios1.jpg HTTP/1.1" 200 7217
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/organizaciones1.jpg HTTP/1.1" 200 7543
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /imagenes/fmellado.jpg HTTP/1.1" 200 2675
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/cabierta.png HTTP/1.1" 200 5464
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/wap_ing_uchile_cl.jpg HTTP/1.1" 200 5419
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /imagenes/logo_ucursos.jpg HTTP/1.1" 200 36799
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /img_index/novedades2.jpg HTTP/1.1" 200 8089
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /novedades.htm HTTP/1.1" 200 655
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /barraizquierda2.htm HTTP/1.1" 200 3258
200.73.40.132 - - [07/Nov/2004:04:14:21 -0300] "GET /main_novedades.htm HTTP/1.1" 200 514
200.73.40.132 - - [07/Nov/2004:04:14:22 -0300] "GET /img_index/lupa.gif HTTP/1.1" 200 1566
200.73.40.132 - - [07/Nov/2004:04:14:22 -0300] "GET /head_principal.htm HTTP/1.1" 200 3361
200.73.40.132 - - [07/Nov/2004:04:14:22 -0300] "GET /img_index/organizaciones1.jpg HTTP/1.1" 200 7543
200.73.40.132 - - [07/Nov/2004:04:14:22 -0300] "GET /img_index/novedades2.jpg HTTP/1.1" 200 8089
200.73.40.132 - - [07/Nov/2004:04:14:22 -0300] "GET /img_index/escmovil2.jpg HTTP/1.1" 200 8077
```

WHAT: Sequence

Sequence: Logs of Activities

- User Interaction
- Application State
- User State

WHAT: Sequence

Sequence: Logs of Activities

- User Interaction
- Application State
- User State
- Taxonomy-Based

***Select*::** mark something as interesting

***Explore*::** show me something else

***Reconfigure*::** show me a different arrangement

***Encode*::** show me a different representation

***Abstract/Elaborate*::** show me more or less detail

***Filter*::** show me something conditionally

***Connect*::** show me related items

WHAT: Sequence

Sequence: Logs of Activities

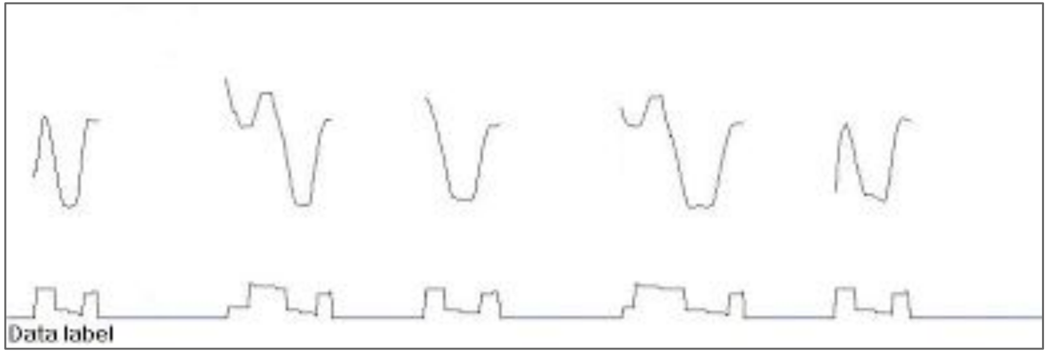
- User Interaction
- Application State
- User State
- Taxonomy-Based
- Image Space

Finding Waldo: Learning about Users from their Interactions, Brown et al. VAST 2014

WHAT: Sequence

Sequence: Logs of Activities

- User Interaction
- Application State
- User State
- Taxonomy-Based
- Image Space
- Temporal Signal



AB✓ WHAT: Grammar

Grammar: Generate Reusable Scripts

- Logic Rules

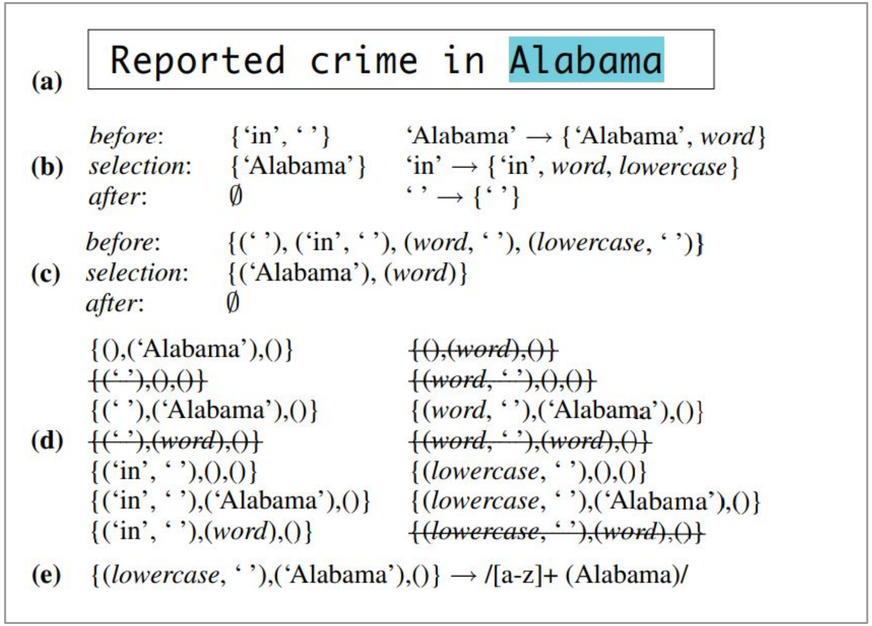
```
Portmap = first_dest_port_111(x1,x2,…) AND
          identical_src_ip(x1,x2,…) AND
          identical_dest_ip(x1,x2,…) AND
          time_sequence_0.5s(x1,x2,…)

Fast Scan = identical_src_ip(x1,x2,…) AND
          dest_ip_in_sequence(x1,x2,…) AND
          more_than_50_events(x1,x2,…) AND
          high_dest_ip_access_rate(x1,x2,…) AND
          (is_tcp(x1) AND is_tcp(x2) AND …) AND
          (low_duration(x1) AND low_duration(x2) AND
     …) AND
          (low_total_data(x1) AND low_total_data(x2) AND
     …)
```

Enhancing Visual Analysis of Network Traffic Using a Knowledge Representation, Xiao et al. VAST 2006

WHAT: Grammar

Grammar: Generate Reusable Scripts

- Logic Rules
- Languages and Scripts



Wrangler: interactive visual specification of data transformation scripts, Kandel et al. CHI 2011

WHAT: Grammar

Grammar: Generate Reusable Scripts

- Logic Rules
- Languages and Scripts
- Specifications

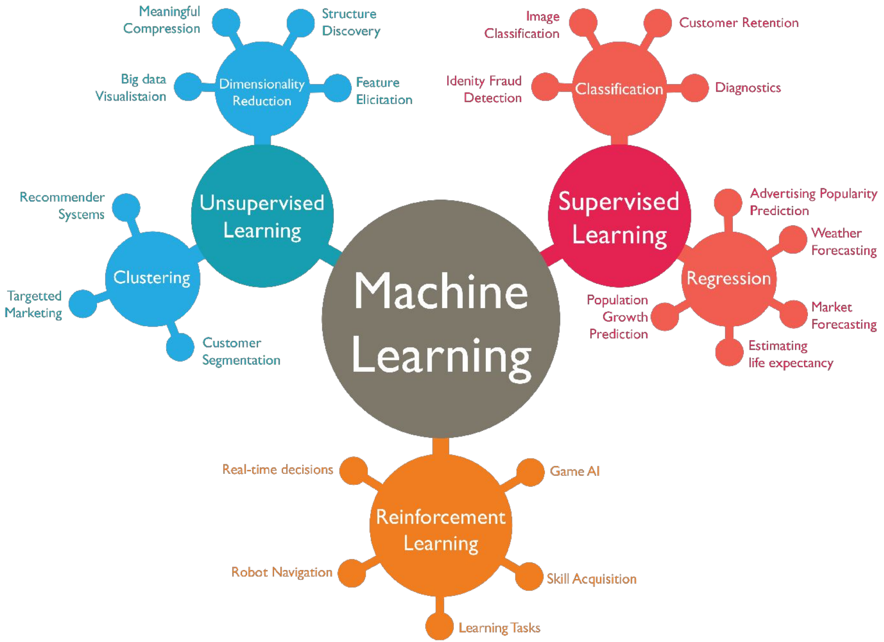
```
Mark := Rectangle | Symbol | Line | Text
GlyphElement := Mark | Guide | GuideCoordinator |
    DataDrivenGuide
Glyph := GlyphElement*, LayoutConstraint<GlyphElement>*
ChartElement := PlotSegment | Link | Mark | Legend | Guide |
    GuideCoordinator
PlotSegment := Glyph, (Scaffold | Axis){0..2}, Sublayout,
    CoordinateSystem
Attribute := "x1" | "y1" | "x2" | "y2" | ...
ElementAttribute<ElementType> := ElementType, Attribute
ParentAttribute := Attribute
ConstraintType := "equals"
LayoutConstraint<ElementType> := (ParentAttribute |
    ElementAttribute<ElementType>){2}, ConstraintType
Chart := ChartElement*, Scale*, LayoutConstraint<ChartElement>*

Notation  "*": zero or more; "{0..2}": zero to two; "|": or;
          "X<Type>": template with parameter "Type"
```

Charticulator: Interactive construction of bespoke chart layouts, Ren et al. TVCG, 2018

WHAT: Model

Model: Abstractions of Actions

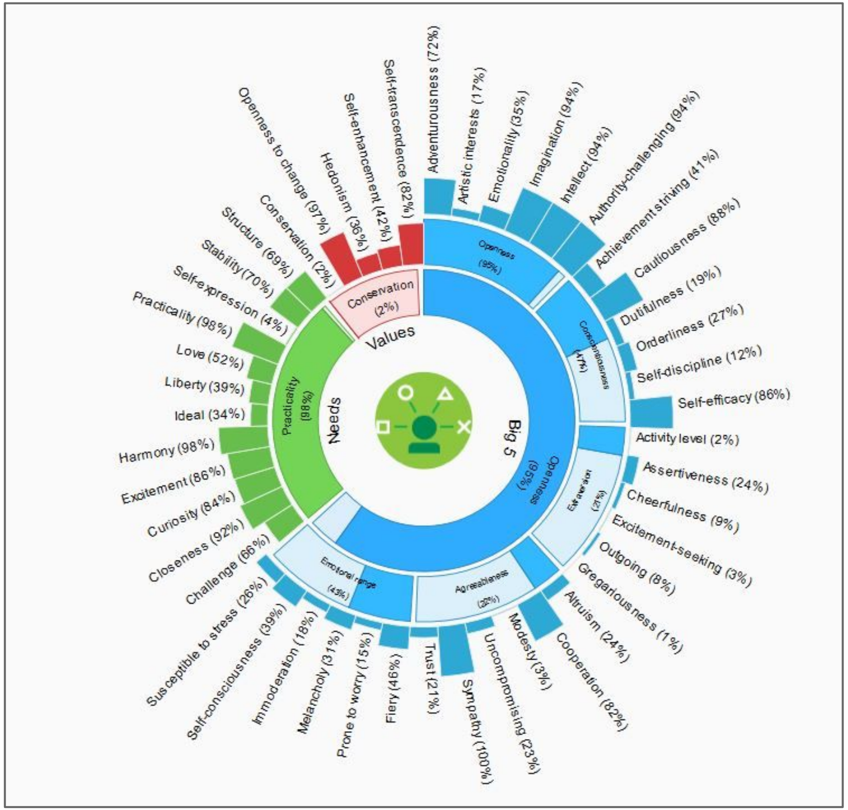
- Machine Learning Models



WHAT: Model

Model: Abstractions of Actions

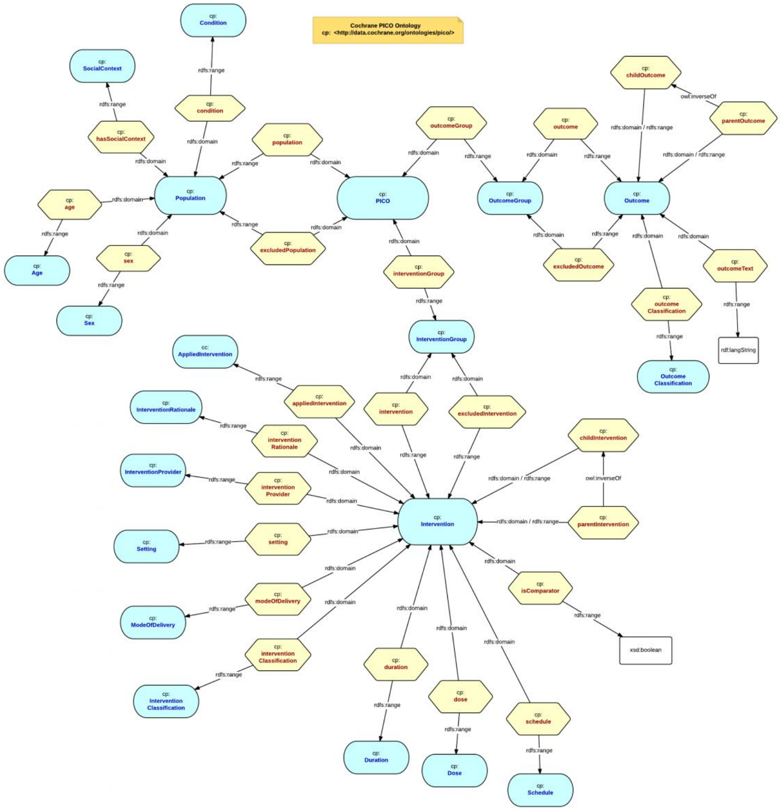
- Machine Learning Models
- User Models



WHAT: Graph

Graph: Actions and Relations

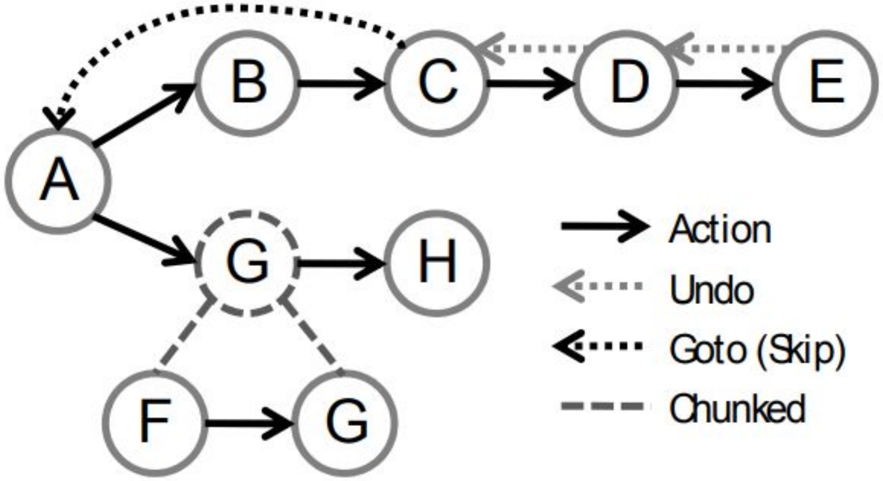
- Entity and Concept Graph



WHAT: Graph

Graph: Actions and Relations

- Entity and Concept Graph
- History Graph



Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation, Heer et al. InfoVis, 2008

Techniques: HOW to Analyze Provenance Data

Classification Models

Pattern Analysis

Probabilistic Models / Prediction

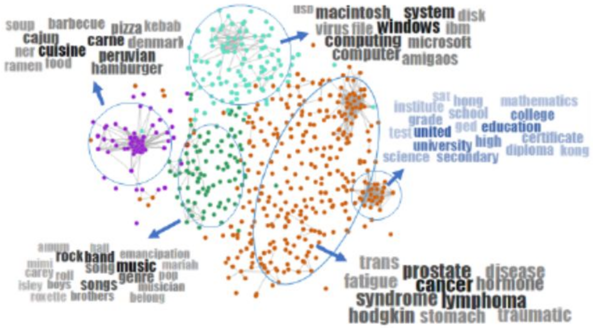
Program Synthesis

Interactive Visual Analysis

HOW: Classification Models

Classification: Differentiate sequences of interactions into meaningful groupings

- Clustering

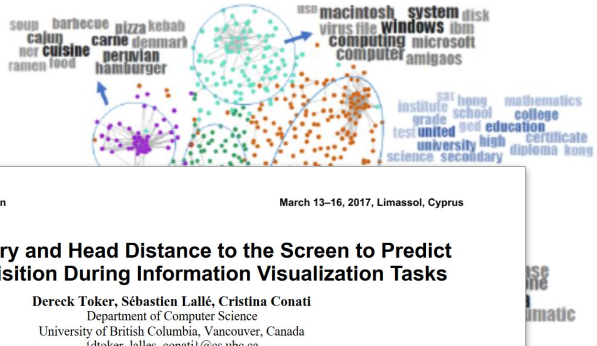


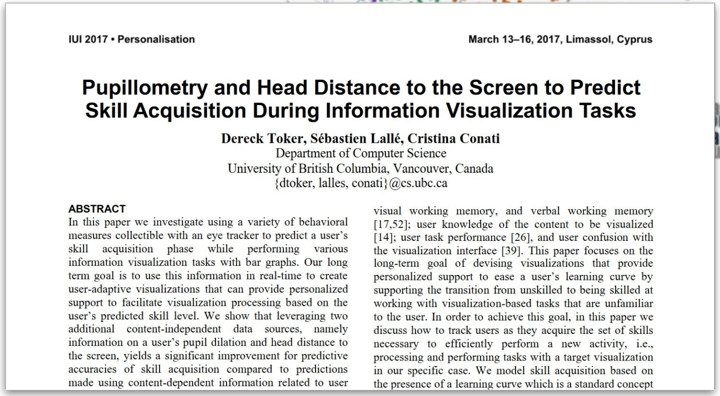
Sherkat et al., 2018

HOW: Classification Models

Classification: Differentiate sequences of interactions into meaningful groupings

- Clustering
- Regression





IUI 2017 • Personalisation March 13–16, 2017, Limassol, Cyprus

Pupillometry and Head Distance to the Screen to Predict Skill Acquisition During Information Visualization Tasks

Dereck Toker, Sébastien Lallé, Cristina Conati
Department of Computer Science
University of British Columbia, Vancouver, Canada
{dtoker, lalles, conati}@cs.ubc.ca

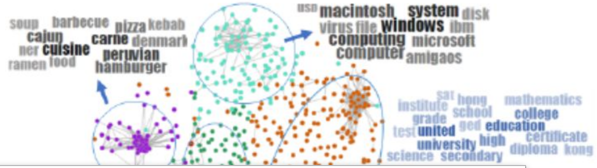
ABSTRACT
In this paper we investigate using a variety of behavioral measures collectible with an eye tracker to predict a user's skill acquisition phase while performing various information visualization tasks with bar graphs. Our long term goal is to use this information in real-time to create user-adaptive visualizations that can provide personalized support to facilitate visualization processing based on the user's predicted skill level. We show that leveraging two additional content-independent data sources, namely information on a user's pupil dilation and head distance to the screen, yields a significant improvement for predictive accuracies of skill acquisition compared to predictions made using content-dependent information related to user

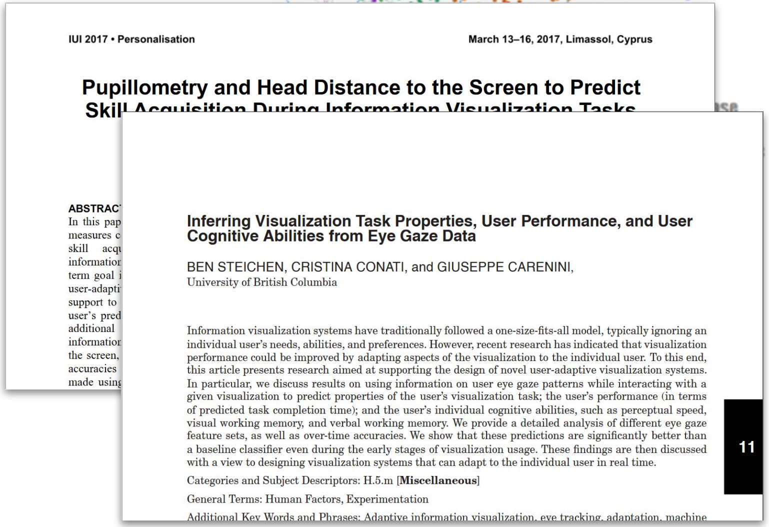
visual working memory, and verbal working memory [17,52]; user knowledge of the content to be visualized [14]; user task performance [26], and user confusion with the visualization interface [39]. This paper focuses on the long-term goal of devising visualizations that provide personalized support to ease a user's learning curve by supporting the transition from unskilled to being skilled at working with visualization-based tasks that are unfamiliar to the user. In order to achieve this goal, in this paper we discuss how to track users as they acquire the set of skills necessary to efficiently perform a new activity, i.e., processing and performing tasks with a target visualization in our specific case. We model skill acquisition based on the presence of a learning curve which is a standard concept

HOW: Classification Models

Classification: Differentiate sequences of interactions into meaningful groupings

- Clustering
- Regression
- SVMs




IUI 2017 • Personalisation March 13–16, 2017, Limassol, Cyprus

Pupillometry and Head Distance to the Screen to Predict Skill Acquisition During Information Visualization Tasks

ABSTRACT
In this pap... measures c... skill acqu... information... term goal i... user-adapti... support to... user's pred... additional... information... the screen,... accuracies... made using...

Inferring Visualization Task Properties, User Performance, and User Cognitive Abilities from Eye Gaze Data

BEN STEICHEN, CRISTINA CONATI, and GIUSEPPE CARENINI,
University of British Columbia

Information visualization systems have traditionally followed a one-size-fits-all model, typically ignoring an individual user's needs, abilities, and preferences. However, recent research has indicated that visualization performance could be improved by adapting aspects of the visualization to the individual user. To this end, this article presents research aimed at supporting the design of novel user-adaptive visualization systems. In particular, we discuss results on using information on user eye gaze patterns while interacting with a given visualization to predict properties of the user's visualization task; the user's performance (in terms of predicted task completion time); and the user's individual cognitive abilities, such as perceptual speed, visual working memory, and verbal working memory. We provide a detailed analysis of different eye gaze feature sets, as well as over-time accuracies. We show that these predictions are significantly better than a baseline classifier even during the early stages of visualization usage. These findings are then discussed with a view to designing visualization systems that can adapt to the individual user in real time.

Categories and Subject Descriptors: H.5.m [**Miscellaneous**]

General Terms: Human Factors, Experimentation

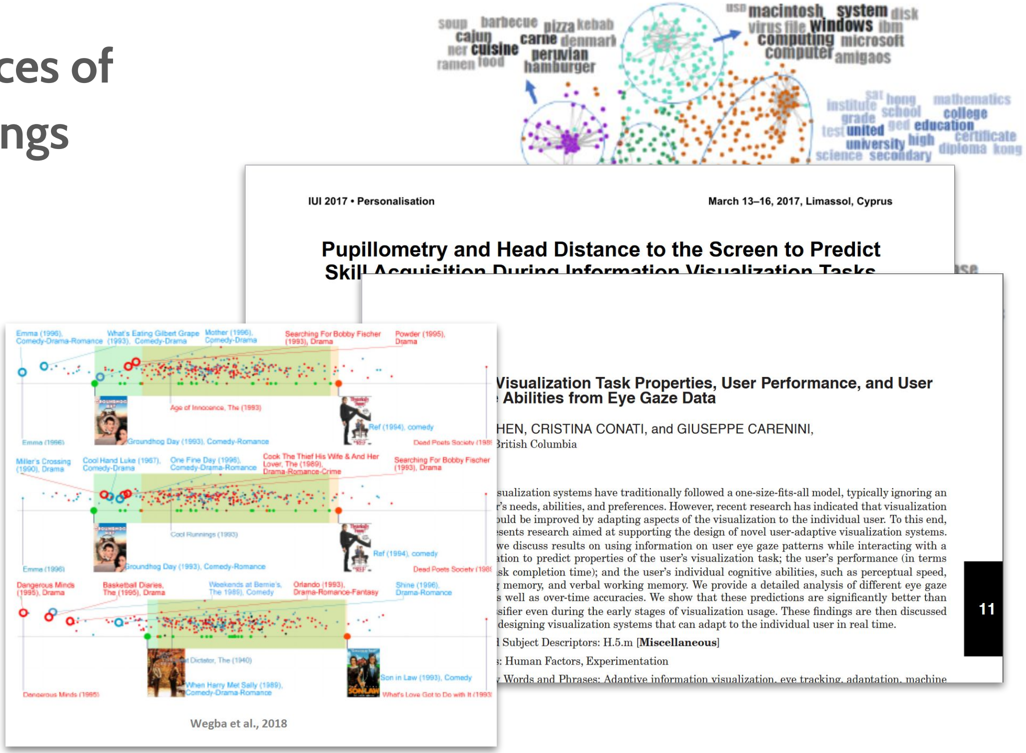
Additional Key Words and Phrases: Adaptive information visualization, eye tracking, adaptation, machine

11

HOW: Classification Models

Classification: Differentiate sequences of interactions into meaningful groupings

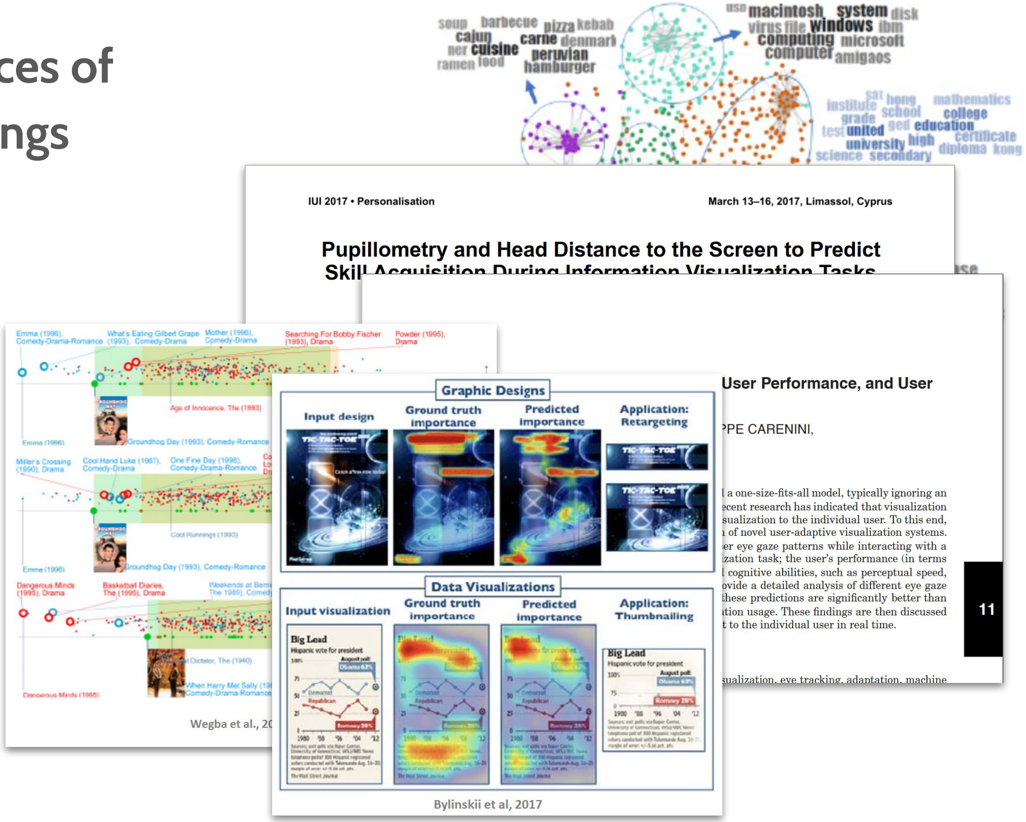
- Clustering
- Regression
- SVMs
- Topic modeling



HOW: Classification Models

Classification: Differentiate sequences of interactions into meaningful groupings

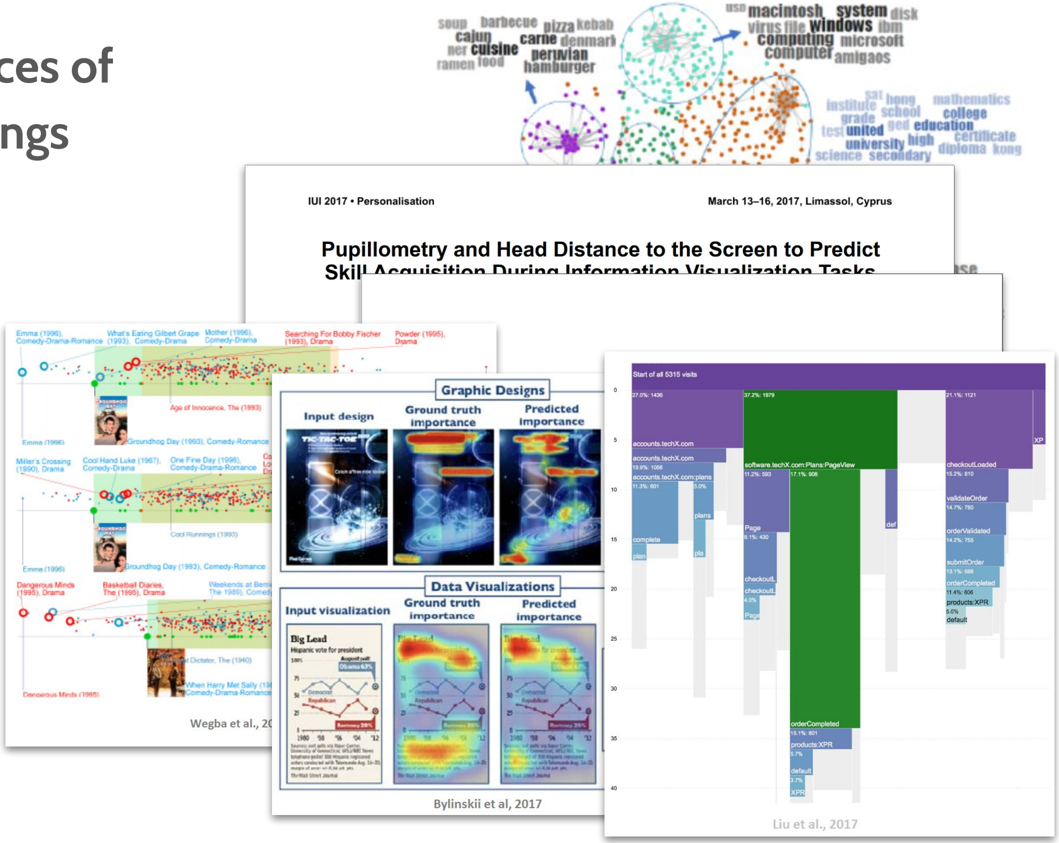
- Clustering
- Regression
- SVMs
- Topic modeling
- Neural networks



HOW: Classification Models

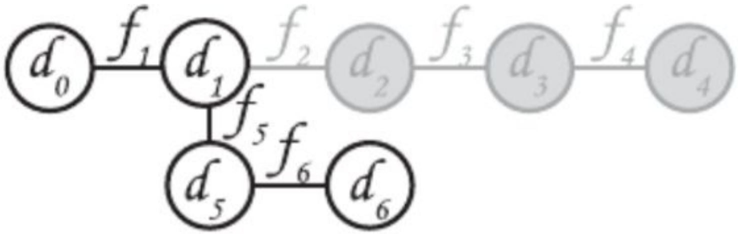
Classification: Differentiate sequences of interactions into meaningful groupings

- Clustering
- Regression
- SVMs
- Topic modeling
- Neural networks
- Hierarchical techniques



Wegba et al., 20
Bylinskii et al, 2017
Liu et al., 2017

HOW: Pattern Analysis



Gotz et al., 2017

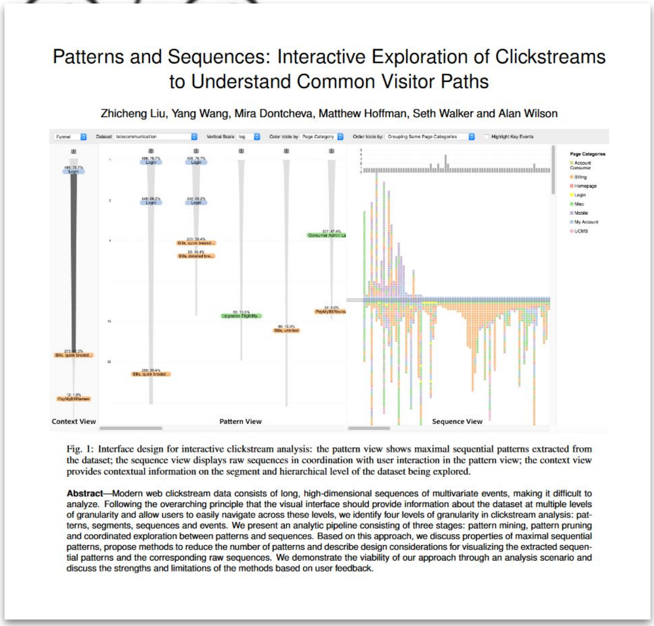
Pattern Analysis: Detect themes either via automation or by user-driven approaches

- Adaptive contextualization

HOW: Pattern Analysis

Pattern Analysis: Detect themes either via automation or by user-driven approaches

- Adaptive contextualization
- Extraction of branching patterns



Patterns and Sequences: Interactive Exploration of Clickstreams to Understand Common Visitor Paths

Zhicheng Liu, Yang Wang, Mira Dontcheva, Matthew Hoffman, Seth Walker and Alan Wilson

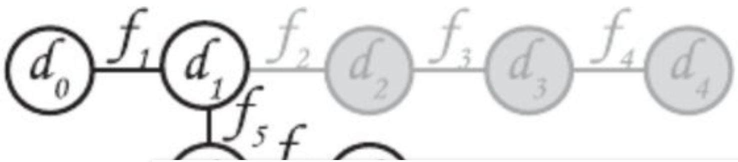
Fig. 1: Interface design for interactive clickstream analysis: the pattern view shows maximal sequential patterns extracted from the dataset; the sequence view displays raw sequences in coordination with user interaction in the pattern view; the context view provides contextual information on the segment and hierarchical level of the dataset being explored.

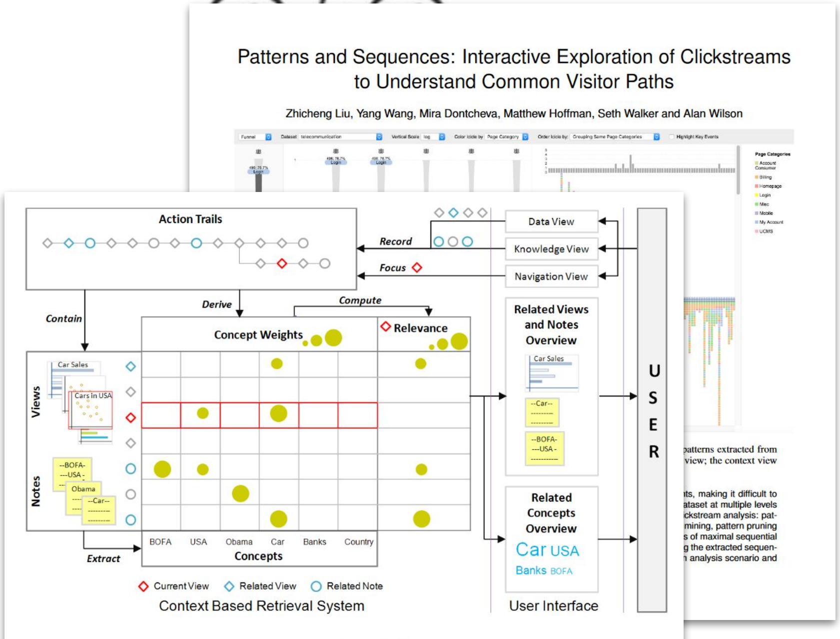
Abstract—Modern web clickstream data consists of long, high-dimensional sequences of multivariate events, making it difficult to analyze. Following the overarching principle that the visual interface should provide information about the dataset at multiple levels of granularity and allow users to easily navigate across these levels, we identify four levels of granularity in clickstream analysis: patterns, segments, sequences and events. We present an analytic pipeline consisting of three stages: pattern mining, pattern pruning and coordinated exploration between patterns and sequences. Based on this approach, we discuss properties of maximal sequential patterns, propose methods to reduce the number of patterns and describe design considerations for visualizing the extracted sequential patterns and the corresponding raw sequences. We demonstrate the viability of our approach through an analysis scenario and discuss the strengths and limitations of the methods based on user feedback.

HOW: Pattern Analysis

Pattern Analysis: Detect themes either via automation or by user-driven approaches

- Adaptive contextualization
- Extraction of branching patterns
- Retrieving notes from past analyses



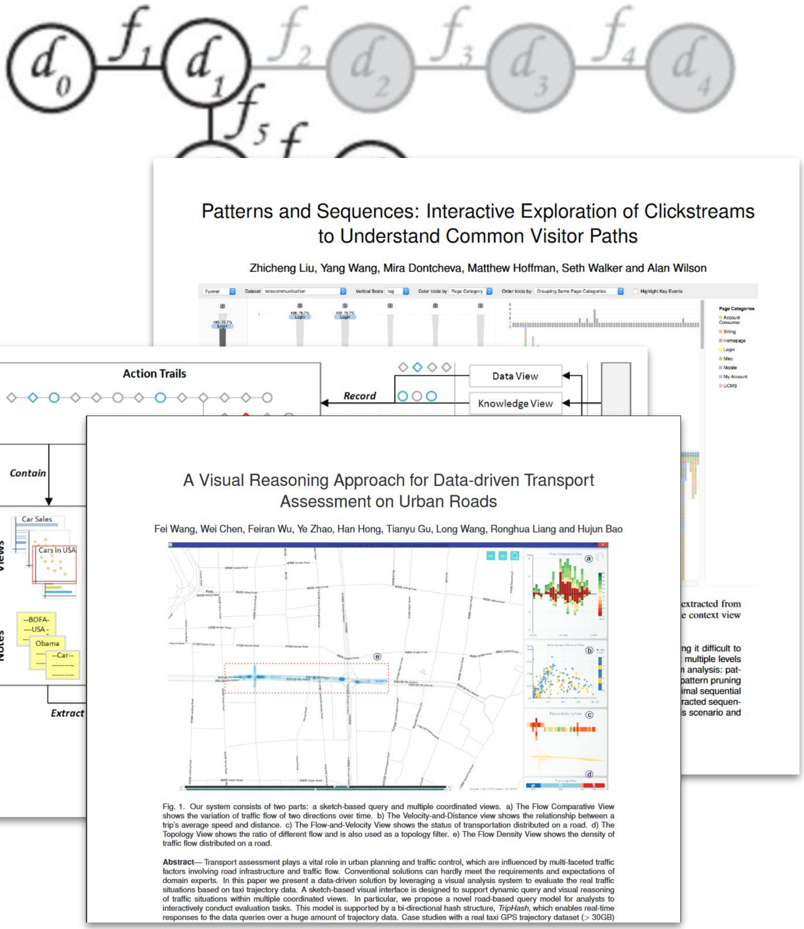


Shrinivasan et al., 2009

HOW: Pattern Analysis

Pattern Analysis: Detect themes either via automation or by user-driven approaches

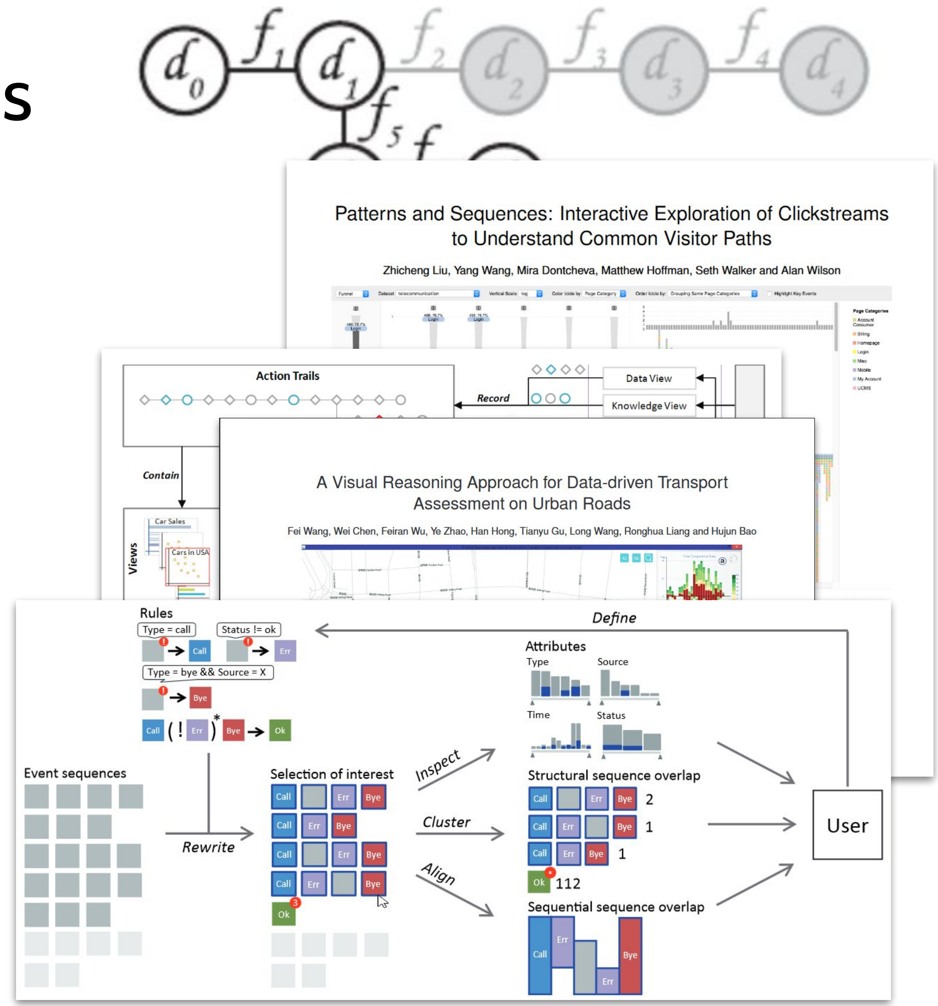
- Adaptive contextualization
- Extraction of branching patterns
- Retrieving notes from past analyses
- Sketches



HOW: Pattern Analysis

Pattern Analysis: Detect themes either via automation or by user-driven approaches

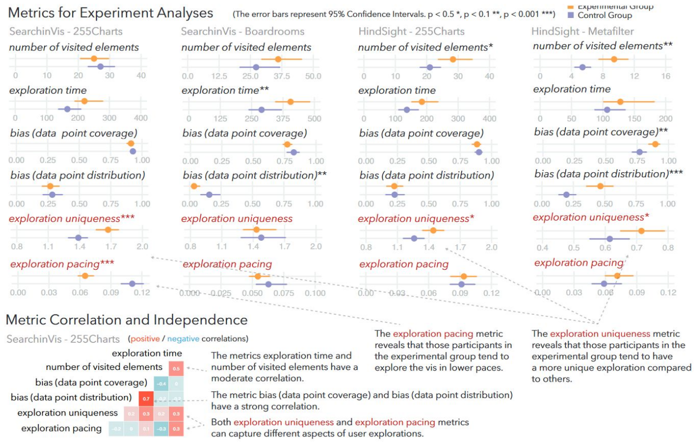
- Adaptive contextualization
- Extraction of branching patterns
- Retrieving notes from past analyses
- Sketches
- Rule-based systems



HOW: Probabilistic Models / Prediction

Probabilistic Models: Mitigating the inherent uncertainty in interpreting and inferring from provenance data

- Traditional statistical models

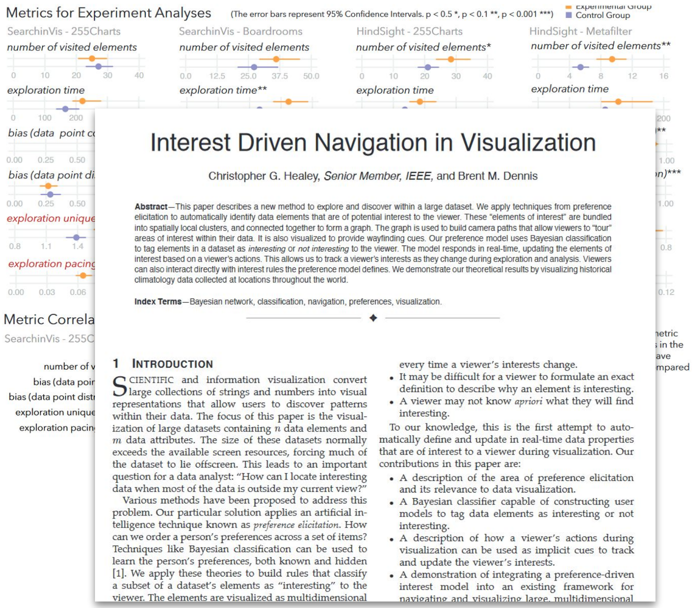


Feng et al., 2018

HOW: Probabilistic Models / Prediction

Probabilistic Models: Mitigating the inherent uncertainty in interpreting and inferring from provenance data

- Traditional statistical models
- Bayesian probability and inference



Interest Driven Navigation in Visualization

Christopher G. Healey, *Senior Member, IEEE*, and Brent M. Dennis

Abstract—This paper describes a new method to explore and discover within a large dataset. We apply techniques from preference elicitation to automatically identify data elements that are of potential interest to the viewer. These "elements of interest" are bundled into spatially local clusters, and connected together to form a graph. The graph is used to build camera paths that allow viewers to "tour" areas of interest within their data. It is also visualized to provide wayfinding cues. Our preference model uses Bayesian classification to tag elements in a dataset as *interesting* or *not interesting* to the viewer. The model responds in real-time, updating the elements of interest based on a viewer's actions. This allows us to track a viewer's interests as they change during exploration and analysis. Viewers can also interact directly with interest rules the preference model defines. We demonstrate our theoretical results by visualizing historical climatology data collected at locations throughout the world.

Index Terms—Bayesian network, classification, navigation, preferences, visualization.

1 INTRODUCTION

SCIENTIFIC and information visualization convert large collections of strings and numbers into visual representations that allow users to discover patterns within their data. The focus of this paper is the visualization of large datasets containing n data elements and m data attributes. The size of these datasets normally exceeds the available screen resources, forcing much of the dataset to lie offscreen. This leads to an important question for a data analyst: "How can I locate interesting data when most of the data is outside my current view?"

Various methods have been proposed to address this problem. Our particular solution applies an artificial intelligence technique known as *preference elicitation*. How can we order a person's preferences across a set of items? Techniques like Bayesian classification can be used to learn the person's preferences, both known and hidden [1]. We apply these theories to build rules that classify a subset of a dataset's elements as "interesting" to the viewer. The elements are visualized as multidimensional

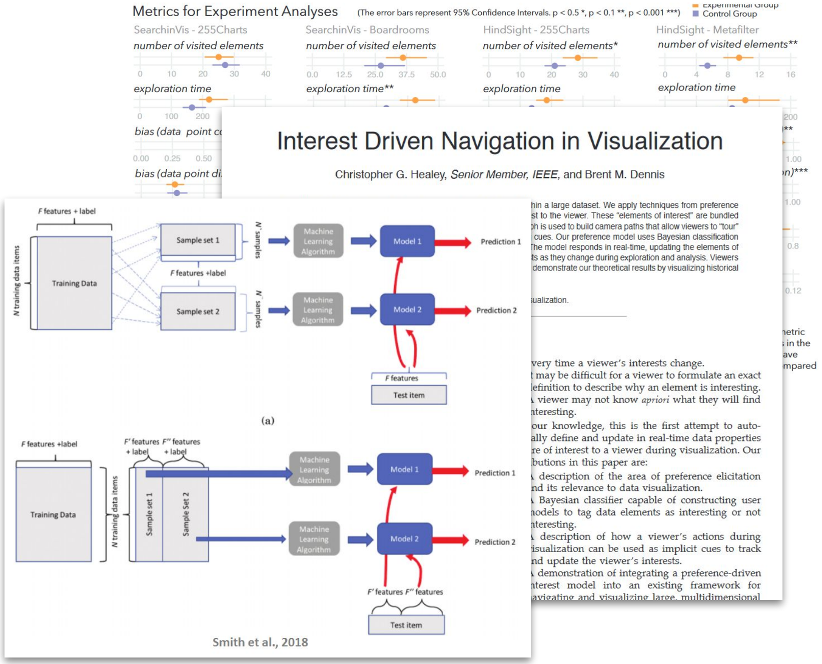
every time a viewer's interests change.
- It may be difficult for a viewer to formulate an exact definition to describe why an element is interesting.
- A viewer may not know *apriori* what they will find interesting.

To our knowledge, this is the first attempt to automatically define and update in real-time data properties that are of interest to a viewer during visualization. Our contributions in this paper are:
- A description of the area of preference elicitation and its relevance to data visualization.
- A Bayesian classifier capable of constructing user models to tag data elements as interesting or not interesting.
- A description of how a viewer's actions during visualization can be used as implicit cues to track and update the viewer's interests.
- A demonstration of integrating a preference-driven interest model into an existing framework for navigating and visualizing large, multidimensional

HOW: Probabilistic Models / Prediction

Probabilistic Models: Mitigating the inherent uncertainty in interpreting and inferring from provenance data

- Traditional statistical models
- Bayesian probability and inference
- Neural networks

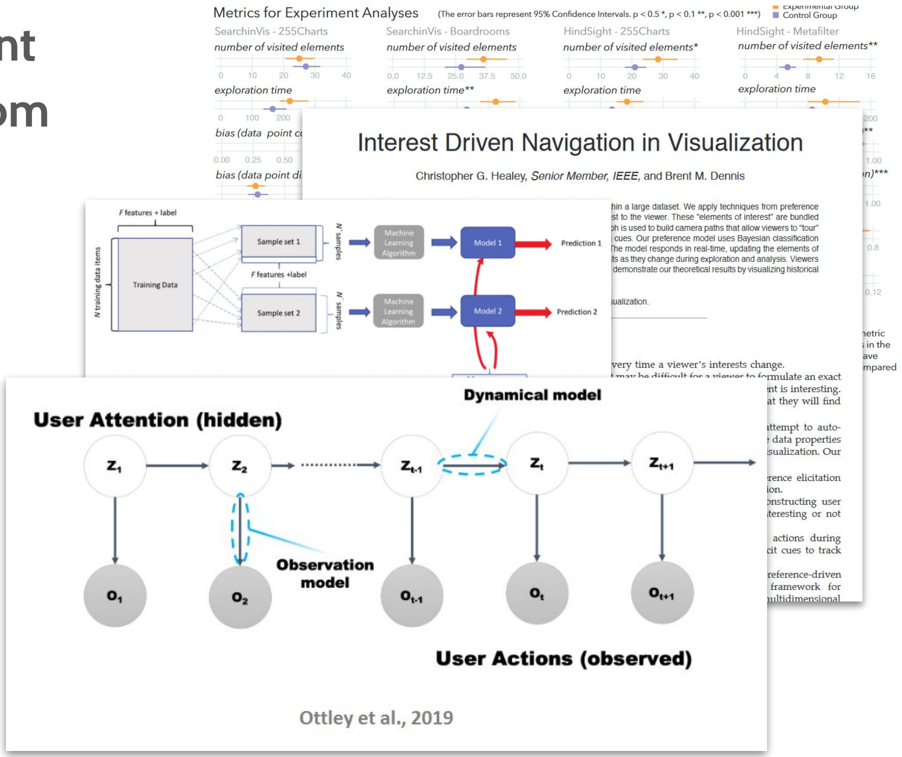


Smith et al., 2018

HOW: Probabilistic Models / Prediction

Probabilistic Models: Mitigating the inherent uncertainty in interpreting and inferring from provenance data

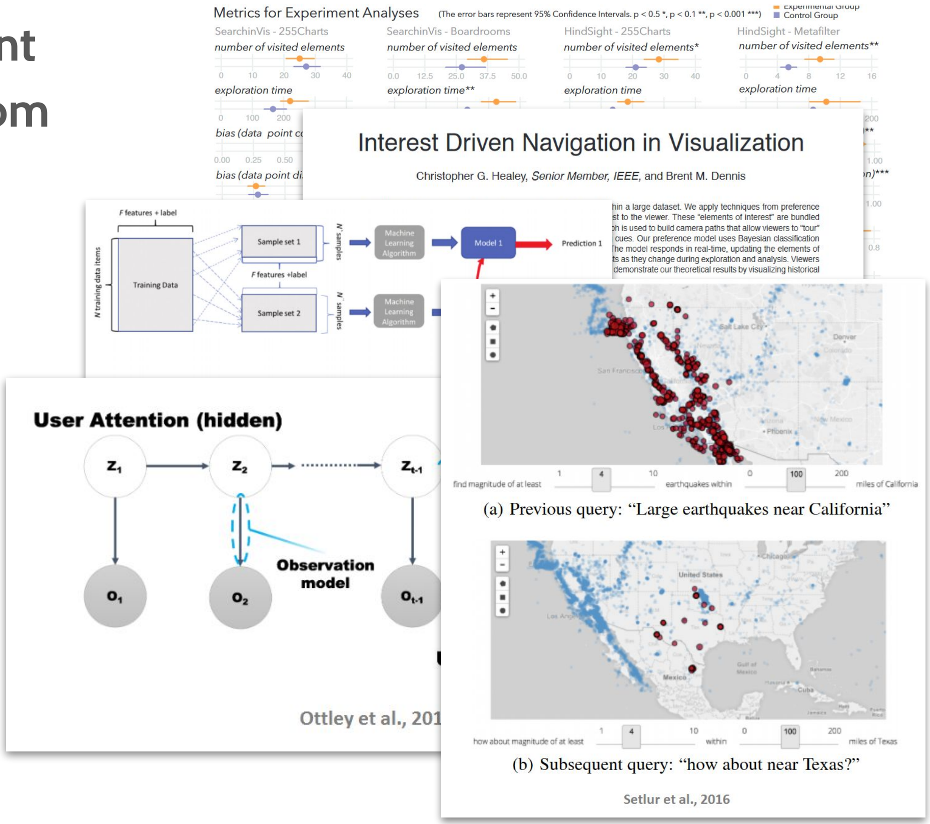
- Traditional statistical models
- Bayesian probability and inference
- Neural networks
- Markov models



HOW: Probabilistic Models / Prediction

Probabilistic Models: Mitigating the inherent uncertainty in interpreting and inferring from provenance data

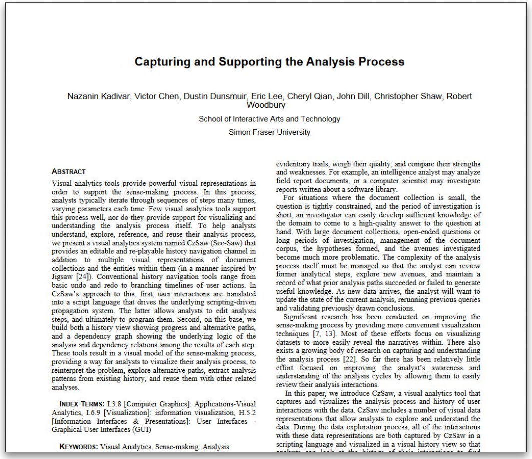
- Traditional statistical models
- Bayesian probability and inference
- Neural networks
- Markov models
- Natural language processing



HOW: Program Synthesis

Program Synthesis: Create executable scripts based on past user interactions

- Domain-specific languages



Capturing and Supporting the Analysis Process

Nazanin Kadivar, Victor Chen, Dustin Dunsmuir, Eric Lee, Cheryl Qian, John Dill, Christopher Shaw, Robert Woodbury

School of Interactive Arts and Technology

Simon Fraser University

ABSTRACT

Visual analytics tools provide powerful visual representations in order to support the sense-making process. In this process, analysts typically iterate through sequences of steps many times, varying parameters each time. Few visual analytics tools support this process well, nor do they provide support for visualizing and understanding the analysis process itself. To help analysts understand, explore, reference, and reuse their analysis process, we present a visual analytics system named CzSaw (See-Saw) that provides an editable and re-playable history navigation channel in addition to multiple visual representations of document collections and the entities within them (in a manner inspired by Jigsaw [24]). Conventional history navigation tools range from basic undo and redo to branching timelines of user actions. In CzSaw's approach to this, first, user interactions are translated into a script language that drives the underlying scripting-driven propagation system. The latter allows analysts to edit analysis steps, and ultimately to program them. Second, on this base, we build both a history view showing progress and alternative paths, and a dependency graph showing the underlying logic of the analysis and dependency relations among the results of each step. These tools result in a visual model of the sense-making process, providing a way for analysts to visualize their analysis process, to reinterpret the problem, explore alternative paths, extract analysis patterns from existing history, and reuse them with other related analyses.

INDEX TERMS: I.3.8 [Computer Graphics]: Applications-Visual Analytics, I.6.9 [Visualization]: information visualization, H.5.2 [Information Interfaces & Presentations]: User Interfaces - Graphical User Interfaces (GUI)

KEYWORDS: Visual Analytics, Sense-making, Analysis

evidentiary trails, weigh their quality, and compare their strengths and weaknesses. For example, an intelligence analyst may analyze field report documents, or a computer scientist may investigate reports written about a software library.

For situations where the document collection is small, the question is tightly constrained, and the period of investigation is short, an investigator can easily develop sufficient knowledge of the domain to come to a high-quality answer to the question at hand. With large document collections, open-ended questions or long periods of investigation, management of the document corpus, the hypotheses formed, and the avenues investigated become much more problematic. The complexity of the analysis process itself must be managed so that the analyst can review former analytical steps, explore new avenues, and maintain a record of what prior analysis paths succeeded or failed to generate useful knowledge. As new data arrives, the analyst will want to update the state of the current analysis, rerunning previous queries and validating previously drawn conclusions.

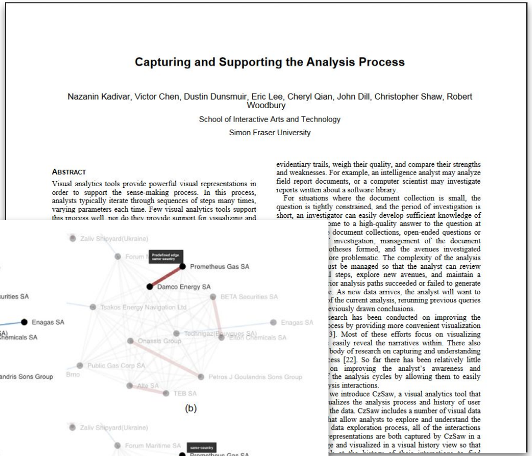
Significant research has been conducted on improving the sense-making process by providing more convenient visualization techniques [7, 13]. Most of these efforts focus on visualizing datasets to more easily reveal the narratives within. There also exists a growing body of research on capturing and understanding the analysis process [22]. So far there has been relatively little effort focused on improving the analyst's awareness and understanding of the analysis cycles by allowing them to easily review their analysis interactions.

In this paper, we introduce CzSaw, a visual analytics tool that captures and visualizes the analysis process and history of user interactions with the data. CzSaw includes a number of visual data representations that allow analysts to explore and understand the data. During the data exploration process, all of the interactions with these data representations are both captured by CzSaw in a scripting language and visualized in a visual history view so that

HOW: Program Synthesis

Program Synthesis: Create executable scripts based on past user interactions

- Domain-specific languages
- Graphs


Capturing and Supporting the Analysis Process

Nazanin Kadivar, Victor Chen, Dustin Dunsmuir, Eric Lee, Cheryl Qian, John Dill, Christopher Shaw, Robert Woodbury

School of Interactive Arts and Technology

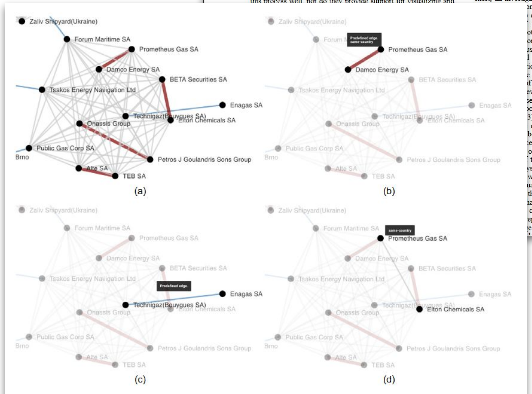
Simon Fraser University

ABSTRACT

Visual analytics tools provide powerful visual representations in order to support the sense-making process. In this process, analysts typically iterate through sequences of steps many times, varying parameters each time. Few visual analytics tools support this process well, nor do they provide support for visualizing and

evidentiary trails, weigh their quality, and compare their strengths and weaknesses. For example, an intelligence analyst may analyze field report documents, or a computer scientist may investigate reports written about a software library.

For situations where the document collection is small, the question is tightly constrained, and the period of investigation is short, an investigator can easily develop sufficient knowledge of

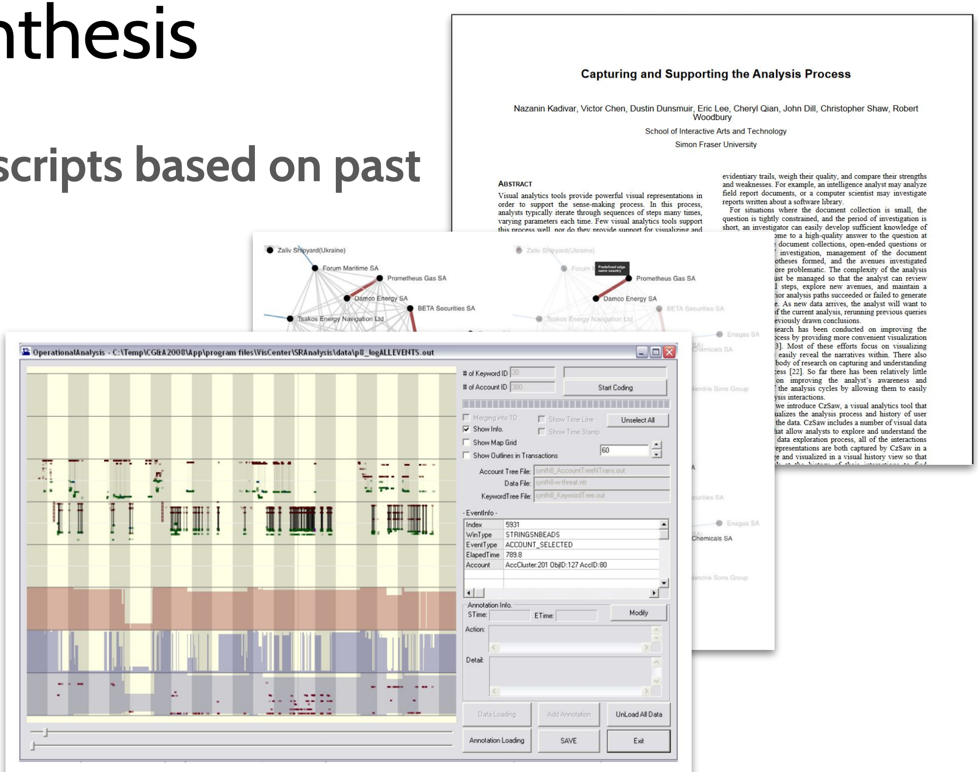


Srinivasan et al., 2018

HOW: Program Synthesis

Program Synthesis: Create executable scripts based on past user interactions

- Domain-specific languages
- Graphs
- Think-aloud studies

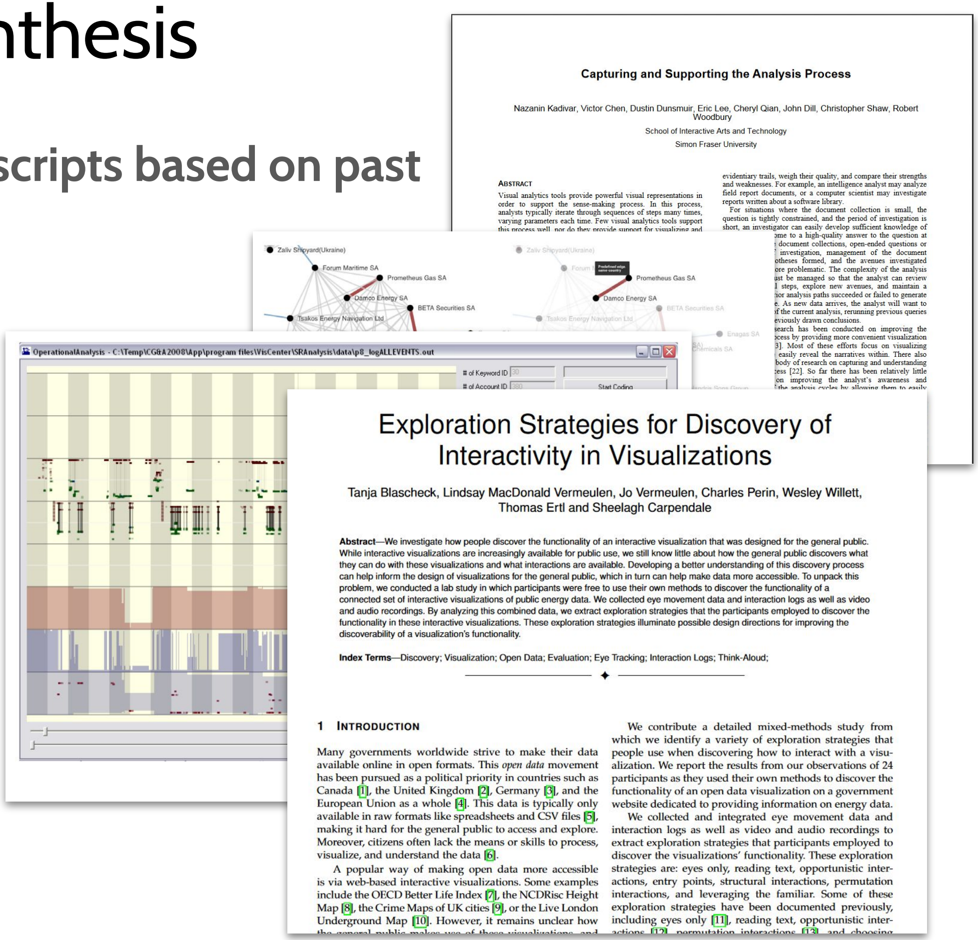


Dou et al., 2009

HOW: Program Synthesis

Program Synthesis: Create executable scripts based on past user interactions

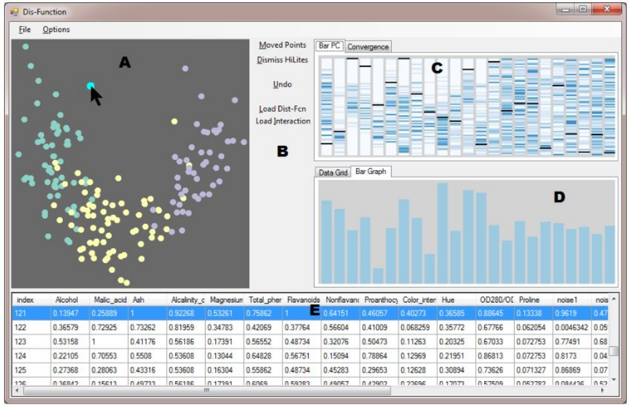
- Domain-specific languages
- Graphs
- Think-aloud studies
- Reviewing past provenance



HOW: Interactive Visual Analytics

Interactive Visual Analytics: User-driven approaches for provenance analysis

- Semantic interaction

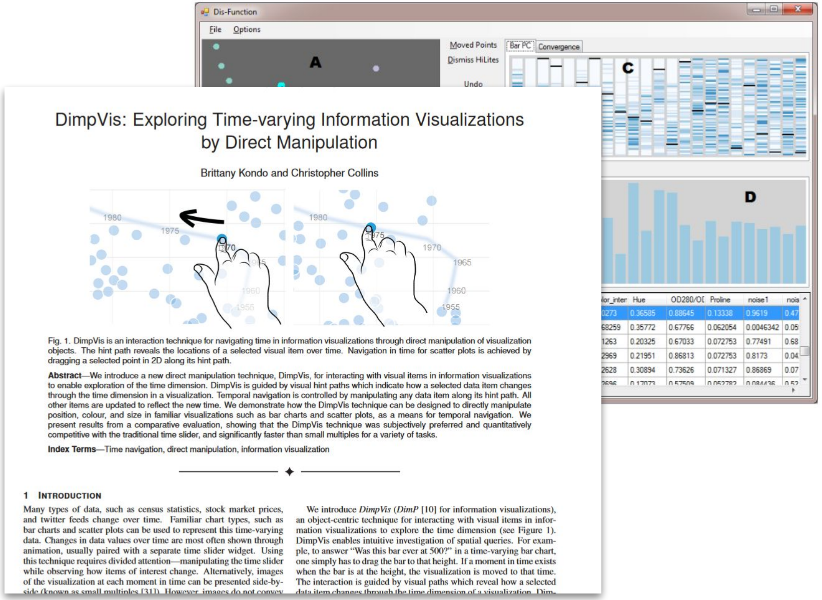


Brown et al., 2012

HOW: Interactive Visual Analytics

Interactive Visual Analytics: User-driven approaches for provenance analysis

- Semantic interaction



DimpVis: Exploring Time-varying Information Visualizations by Direct Manipulation

Brittany Kondo and Christopher Collins

Fig. 1. DimpVis is an interaction technique for navigating time in information visualizations through direct manipulation of visualization objects. The hint path reveals the locations of a selected visual item over time. Navigation in time for scatter plots is achieved by dragging a selected point in 2D along its hint path.

Abstract—We introduce a new direct manipulation technique, DimpVis, for interacting with visual items in information visualizations to enable exploration of the time dimension. DimpVis is guided by visual hint paths which indicate how a selected data item changes through the time dimension in a visualization. Temporal navigation is controlled by manipulating any data item along its hint path. All other items are updated to reflect the new time. We demonstrate how the DimpVis technique can be designed to directly manipulate position, colour, and size in familiar visualizations such as bar charts and scatter plots, as a means for temporal navigation. We present results from a comparative evaluation, showing that the DimpVis technique was subjectively preferred and quantitatively competitive with the traditional time slider, and significantly faster than small multiples for a variety of tasks.

Index Terms—Time navigation, direct manipulation, information visualization

1 INTRODUCTION

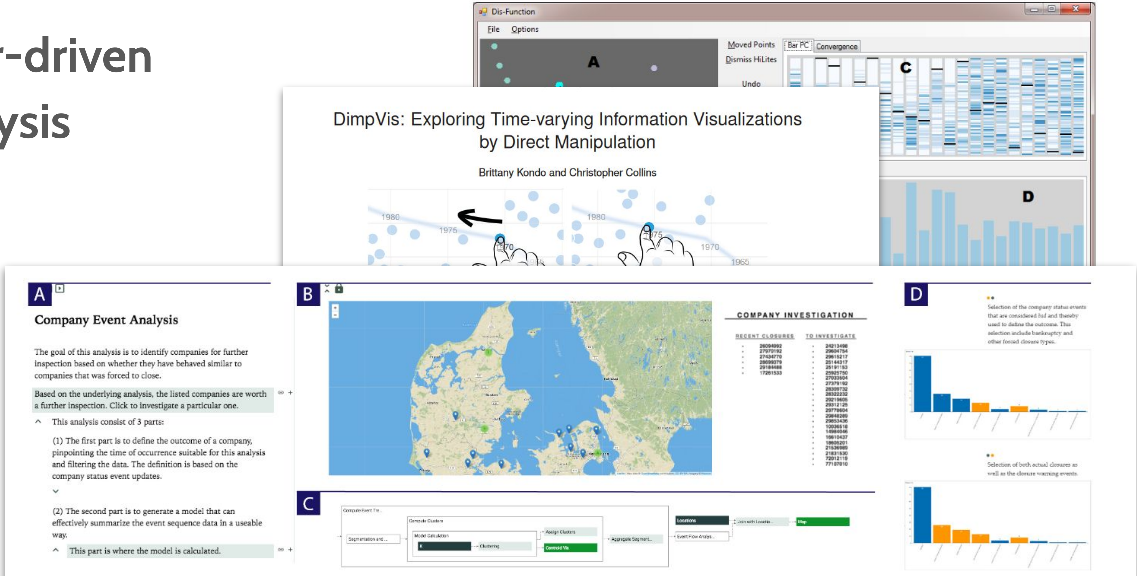
Many types of data, such as census statistics, stock market prices, and twitter feeds change over time. Familiar chart types, such as bar charts and scatter plots can be used to represent this time-varying data. Changes in data values over time are most often shown through animation, usually paired with a separate time slider widget. Using this technique requires divided attention—manipulating the time slider while observing how items of interest change. Alternatively, images of the visualization at each moment in time can be presented side-by-side (known as small multiples [31]). However, images do not convey

We introduce *DimpVis* (*DimP* [10] for information visualizations), an object-centric technique for interacting with visual items in information visualizations to explore the time dimension (see Figure 1). DimpVis enables intuitive investigation of spatial queries. For example, to answer "Was this bar ever at 500?" in a time-varying bar chart, one simply has to drag the bar to that height. If a moment in time exists when the bar is at the height, the visualization is moved to that time. The interaction is guided by visual paths which reveal how a selected data item changes through the time dimension of a visualization. Dim-

HOW: Interactive Visual Analytics

Interactive Visual Analytics: User-driven approaches for provenance analysis

- Semantic interaction
- Visual analytics tools

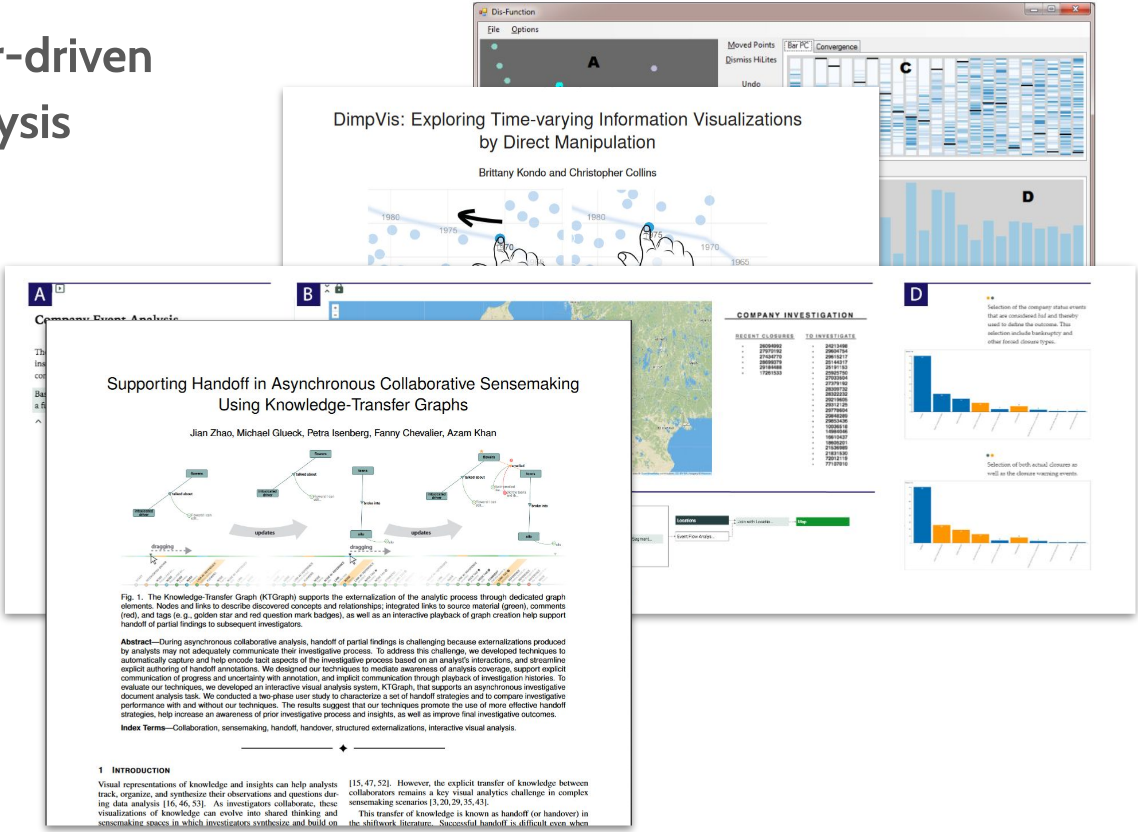


Mathisen et al., 2019

HOW: Interactive Visual Analytics

Interactive Visual Analytics: User-driven approaches for provenance analysis

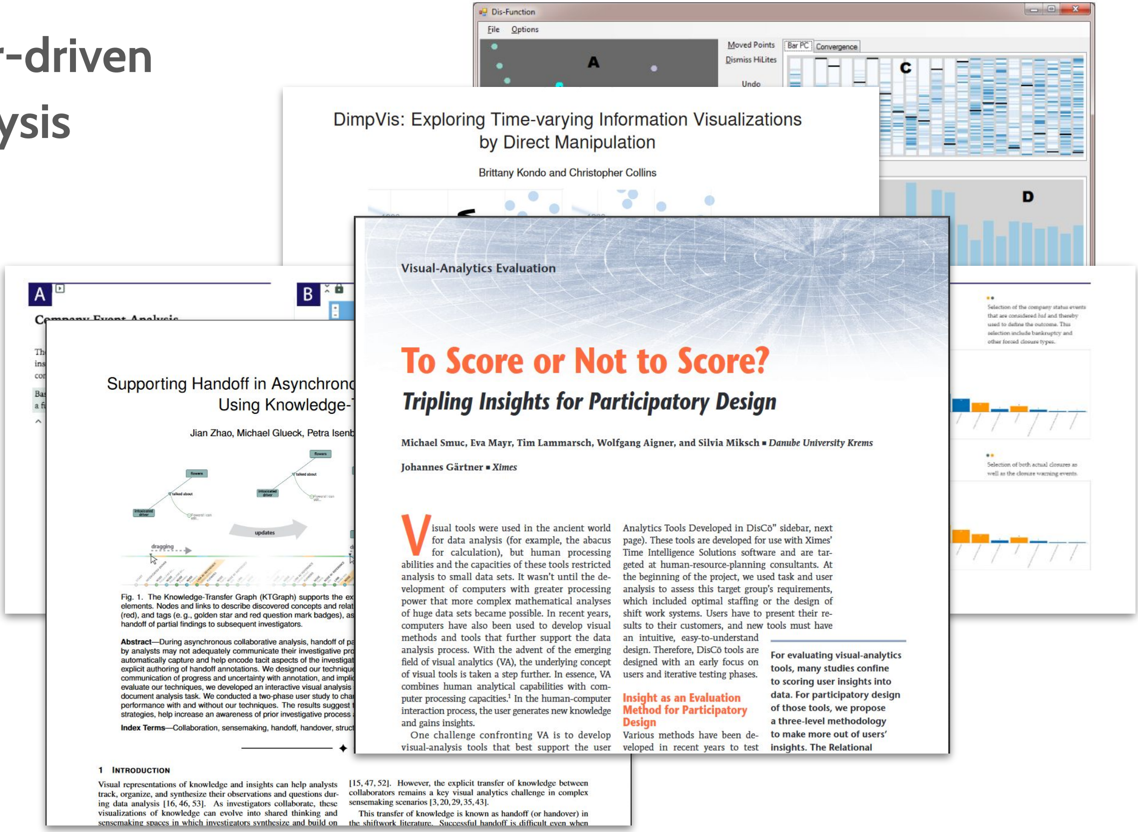
- Semantic interaction
- Visual analytics tools



HOW: Interactive Visual Analytics

Interactive Visual Analytics: User-driven approaches for provenance analysis

- Semantic interaction
- Visual analytics tools
- Analysis of visual design



Future Research

WHY

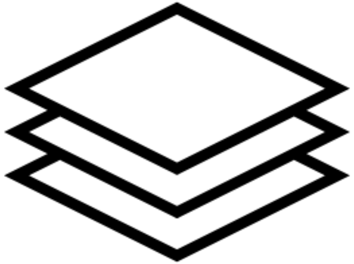
Still many opportunities such as user intent modelling.

Can support important questions in related field such as Explainable AI.

WHAT

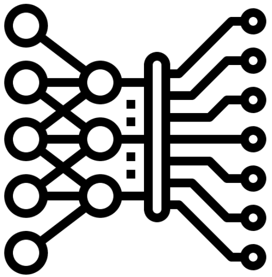
Multi-layer provenance model: from system log to user reasoning

Challenges in user reasoning capture



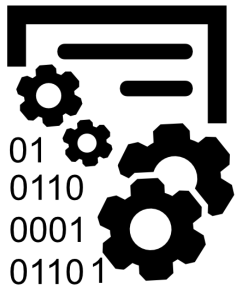
Future Research

HOW



- Utilization of more advanced Machine Learning methods
- Fundamental challenges such as “chunking”: grouping the steps in an interaction sequence

Standard and Integration



- Most provenance tools have their own formats, which makes data exchange and integration almost impossible
- A common standard will be beneficial to the all related fields