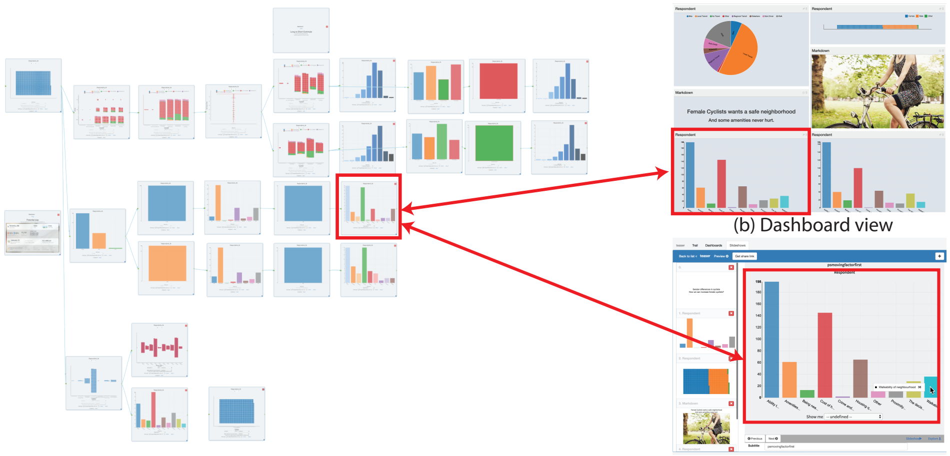
StoryFacets: Generating Multiple Representations of Exploratory Data Analysis for Communication

Anonymous Authors
Undisclosed Institutions
UndisclosedEmails



(a) Trail view (b) Dashboard view (c) Slideshow view

Figure 1: The StoryFacets system creates multiple formats—called *facets*—of the same analysis for both exploration and presentation. The trail facet (a) shows multiple analysis steps and visual provenance links between them. The result of one step (red box) is also used in a dashboard (b) and slideshow (c). Each visual element is linked to shared data storage.

ABSTRACT

We present a design study where we introduced visualization to urban transportation planning as part of a smart cities project. We learned early that a single form of data analysis was insufficient due to the diverse stakeholders involved in this process. This prompted us to design a visual analytics system integrating presentation functionality in a multi-stage design process. Our system exposes several different views of the same core data analysis session, each view being designed for a specific audience: (1) the *trail view* provides a data flow canvas supporting in-depth exploration (expert analysts); (2) the *dashboard view* organizes visualizations and other content into a space-filling layout to support high-level analysis (managers); and (3) the *slideshow view* supports linear storytelling by facilitating interactive step-by-step presentations (laypersons). Views are linked so that when one is changed, all others are automatically updated.

CHI'19, May 2019, Glasgow, UK
2018. ACM ISBN 123-4567-24-567/08/06…$15.00
https://doi.org/http://dx.doi.org/10.475/123_4

CCS CONCEPTS

• **Information systems** → **Multimedia content creation**; • **Human-centered computing** → **Empirical studies in collaborative and social computing**;

KEYWORDS

Exploratory analysis; visualization; data flow; communication; storytelling; narrative visualization; presentation.

ACM Reference format:
Anonymous Authors. 2018. StoryFacets: Generating Multiple Representations of Exploratory Data Analysis for Communication. In *Proceedings of ACM Conference on Human Factors in Computing Systems, Glasgow, UK, May 2019 (CHI'19),* 12 pages.
https://doi.org/http://dx.doi.org/10.475/123_4

1 INTRODUCTION

Urban transportation planning is a form of transportation planning concerned with establishing goals, policies, and investments to prepare for future means of moving people and goods from one place to another in an urban environment. Given the importance of transportation in modern cities, urban transportation planning is a key component of most *smart cities* initiatives that aim to take advantage of modern information and computing technology to optimize the efficiency, sustainability, and social well-being of a city. In such settings, transportation planning becomes a highly data-driven activity where multiple and heterogeneous data sources are collected and fused to enable elected and career officials to make informed decisions about highway networks, mass transit, street infrastructure, etc. Due to the vast scale of the data as well as the wide range of disciplines involved, transportation planning is a highly collaborative analysis process involving large teams of experts with different backgrounds. However, because these decisions directly concern individual citizens, decisions cannot be made only within specialized expert teams, but must also involve stakeholders such as managers, politicians, and even local residents. These stakeholders are mostly consumers of the findings from the expert teams, but by definition tend to lack the expertise to understand all of the complex analyses underlying a proposed solution.

In this work, we present results from a design study [32] on supporting data-driven urban transportation planning that we conducted as part of the iCity project in the City of Toronto, Canada. During the domain characterization and requirements gathering process of the project, our key finding was that while visual analytics is a prime candidate to aid experts in marshalling, analyzing, and synthesizing the heterogeneous data sources used in urban transportation planning, these analysis interfaces—even if they are visual—are often not suitable for more casual stakeholders [26]. More specifically, we found that each of the identified user groups (1) analysts, (2) data consumers/managers, and (3) laypersons such as politicians and individual citizens all have different needs and requirements for how to view, analyze, and understand this data. While expert analysts are well-versed in visual analytics and data science workflows and would benefit from a full-fledged data-flow analysis system [4, 9, 11], managers may prefer only a high-level interactive dashboard with the overall findings, and politicians and citizens may just want a slideshow or an infographic summarizing important outcomes. In current practice, we found that this meant that analysts or managers had to spend an inordinate amount of time building carefully crafted reports and slideshows in order to communicate their results to broader audiences, often using data storytelling methods [22, 33, 39]. Clearly, there is a need for quickly creating multiple storytelling formats derived from a single analysis session.

As part of our multi-phase design process, we designed, implemented, and evaluated StoryFacets, a data exploration system that addresses this gap by allowing users to generate multiple linked representations of the data and user analysis process from a single source (Figure 1). The StoryFacets workflow typically begins with an analyst exploring data in a full-fledged analysis view, called the *trail view*. In the background, StoryFacets automatically maintains multiple different views, or *facets*, of the same data-driven story:

- *Trail view:* a data-flow view [4, 9, 11] where the analyst can load datasets, apply data transformations, run statistical methods, and interact with visualizations.
- *Dashboard view:* a space-filling information dashboard that organizes all or selected parts of the analysis and allows direct interaction and exploration.
- *Slideshow view:* a traditional slideshow format where each visualization is shown as an interactive slide.

With these different presentation formats, StoryFacets supports a wide variety of communication scenarios—such as formal presentations, informal analysis reviews, and collaborative analysis—as well as a wide audience, from experts to beginners. The dashboard view can be used by data consumers or intermediate-level analysts that are knowledgeable about visualization but prefer to use explorations created by others, or turned into an infographic for sharing static results by adding qualitative content. The slideshow view is intended for formal presentations to a larger audience. Furthermore, not only are the additional views created in StoryFacets synchronized with the actual exploration, but the visualizations in each view are live and fully interactive. In fact, the analyst or data scientist can easily switch between views at any time, such as when transitioning from the slideshow view during a small group presentation to the more detailed exploration-centric trail view in response to an in-depth question from a viewer.

Along with our presentation of the StoryFacets system and a formative evaluation using expert participants, we also present the results of an experiment to evaluate how different presentation formats influence presentation style.

2 BACKGROUND

Here we survey the literature on communication, storytelling, and provenance for visualization.

Communication for Visualization

Presenting the outcome of a visual exploration process has always been a priority in visualization research; in fact, it can be argued that the accessibility and familiarity of visual communication is one of the primary reasons that visualization is useful. Viégas and Wattenberg unified these ideas into the concept of *communication-minded visualization* (CMV) [39]: the notion that useful visualizations are part of a greater ecosystem where viewers also participate in a collaborative data analysis process facilitated by the representation.

Unfortunately, most early visualization systems were designed for expert users and thus provided few visualizations suitable for novice stakeholders such as managers, policy makers, or the general public. In 2007, Pousman et al. captured the grassroots effort to democratize visualization for the masses by reducing barriers as the concept of *casual visualization* [26]. However, even today, many visualization tools still lack an easy path from analysis to presentation [35].

Take-away: The optimal data-driven storytelling method depends on the context and audience of the presentation [22, 33]. Thus, supporting a single presentation format is generally not sufficient. We are aware of no visual analytics tools that supports multiple presentation formats.

Storytelling in Visualization

Storytelling conveys sequences of events using plot, locations, actions, and characters, and visual storytelling is the use of visual communication for storytelling. Already in 2001, Gershon and Page [14] suggested that the combination of storytelling and visualization could become a powerful one, and drew on multiple media such as comics, film, and visual metaphors to argue this point. However, it was not until in 2010 that these ideas fully consolidated yielding two workshops at the IEEE VisWeek conference in quick succession (2010 [7] and 2011 [6], respectively), a survey by Segel and Heer [33], and a Dagstuhl scientific workshop in 2016.

Segel and Heer's work is particularly interesting because it identifies seven distinct genres for presenting data narratives: magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show, and film/video/animation. Taking data narratives a step further, Kosara and Mackinlay [22] argue that storytelling may in fact be the next grand challenge for visualization research, and they go on to survey the history of visual communication and its core mechanisms, such as annotations, highlights, textual descriptions, etc. Several specific narrative visualization techniques have since been proposed, such as the use of sequence [17, 18], geographic stories [13], spatiotemporal events [10], sketch-based presentations [23], and narrative annotations [31]. Some commercial tools provide support for this activity—e.g. Story Points in Tableau [21] and dashboards in Spotfire [1]. Still data analysis and storytelling processes are handled separately in those tools. In comparison, our focus is on connecting the storytelling process and the analysis process organically. The main difference is that a specific chart in a slide or dashboards is connected to the original exploration in StoryFacets, preserving the history for that chart.

Take-away: *Visualization* and *storytelling* is a powerful combination, but many current tools generally lack native storytelling support. The *StoryFacets* method automatically generates multiple storytelling representations during analysis.

Provenance for Visualization

Recording the history of an analysis process is referred to as maintaining its *analytic provenance*, and is important for overview as well as recall during the analysis itself [28]. However, it is also useful for communication to stakeholders; history and provenance can be used to construct stories about a visual exploration such as by serializing the exploration into a slideshow [19, 20]. For both of these reasons—improved analysis and improved communication—provenance for visualization has long been an important research topic, and multiple avenues have been explored [12, 28].

Graphical histories is one such avenue. Heer et al. [16] propose different types of graphical histories to save intermediate visualization states during the analysis process. Similarly, Dou et al. [8] demonstrate how interaction logs can help users understand the history of financial data analysis. Sarvghad and Tory [30] study several representations (including sequence diagrams, treemaps, and radial diagrams) to summarize the history of analysis coverage of different dimensions of data sets. In another example, Matejka et al. [24] represent interaction history by augmenting an interface with a heatmap of frequency of button clicks.

Facilitating the user's mental model for provenance is important for both overview and recall. The sensemaking loop proposed by Pirolli and Card [25] explains the iterative nature of the analysis process. For example, once an analyst finds an interesting insight, she might go back to the search and filter process to validate the insight by changing parameters in search of more examples of the same principle.

However, the iterative process changes when multiple analysts are involved. Information sharing on a team often takes place in the final stage of the loop: *presentation*. For this reason, many provenance-tracking visualizations use a spatial analysis workspace where elements are organized in a semi-structured manner. Maintaining spatial persistence in such representations promotes recall [27]; in fact, Ragan et al. [29] found that merely showing the final visual state of a spatial analysis workspace was a sufficient memory aid to significantly help analysts remember the analysis. Many representations—pioneered by the GRASPARC [4] system—are therefore based on branching *exploration trees* that can be deterministically arranged on a spatial workspace. Similarly, Derthick and Roth [5] show how this form of "branching time model" supports memory off-loading and comparison across time and exploration paths. One of the best-known provenance-tracking systems with a branching exploration tree is VisTrails [2, 12], which uses a tree diagram to represent sequences of actions, function calls, and resulting visualizations during computational data analysis. Similarly, Shrinivasan and Van Wijk [34] use a branching timeline to let the user navigate in time for a complex analysis.

Data flow systems replace a strict hierarchy with a directed acyclic graph representation of intermediate states wired together to form a flexible pipeline. The Sandbox [40] is one of the early examples of data flow systems; it provides a semi-structured analysis canvas for intelligence analysis. DataMeadow [11] allows the user to create branching chains of visualization glyphs for multidimensional analysis. Similarly, LARK [36] exposes the full visualization pipeline as a data flow chain on a collaborative space, allowing users to branch and modify the pipeline at different stages. GraphTrail [9] applies the data flow model to graph and network data, providing chains—or *trails*—of connected charts to visualize, filter, and drill into a dataset. ExPlates [20] automatically generates new nodes in an exploration graph in response to interaction in multidimensional data, such as filtering, selection, or brushing.

Take-away: *Visual provenance* facilitates overview and recall both during exploration and afterwards, and *data flow systems* provide a flexible method of explicitly representing provenance. The *StoryFacets* system uses a data flow model where each component can be automatically reformatted into multiple communication-oriented representations.

3 OVERVIEW: DATA FOR DIVERSE TEAMS

The goal of this study was to support urban transportation planners from the iCity smart cities project, a collaboration between the City of Toronto, University of Toronto, OCAD University, the University of Waterloo, Esri Canada, and IBM. Thus, we worked closely with the planners through a multi-phase design process that was inspired by the design study methodology proposed by Sedlmair et al. [32].

Our study quickly turned from focusing on the specifics of urban transportation planning to the more general challenge of supporting *data analysis for diverse teams*, i.e., where team members and stakeholders have varying levels of expertise, knowledge, and motivation. In the following sections, we describe the project phases in detail:

- I **Domain Characterization:** formative sketching, brainstorming, prototyping, and requirements gathering for the users, tasks, and design rationale for supporting diverse teams in urban transportation planning;
- II **Initial Tool Design:** designing our StoryFacets visual analytics tool based on expert feedback;
- III **Formative Evaluation:** qualitative evaluation of our tool with three visual analytics experts; and
- IV **Iterative Refinement:** adding needed features based on feedback from the formative evaluation.

4 PHASE I: DOMAIN CHARACTERIZATION

As part of the winnowing and discovery stages of a design study [32], we worked with our urban transportation planning users through interviews, discussions, brainstorming sessions, formative design sketching, and wireframing. We identified three generalizable roles—analyst, data consumer/manager, and client—and a variety of problems faced by urban transportation planners.

Subdomain: Complete Streets

Transportation planning is a broad and diverse domain. Here we focus on the design of individual city streets using the "Complete Streets" concept—a design and policy framework for designing city streets to support the needs of all users. Complete streets are "designed for all ages, abilities, and modes of travel. Safe and comfortable access for pedestrians, bicycles, transit users, and the mobility impaired is not an afterthought, but is an integral planning feature." [37]. So as to quantify as many of these factors as possible, our collaborators in the iCity project are developing tools to model the benefits and costs of alternative street designs. These models take into account mode and purpose demands as well as costs such as emissions, exposure, travel delay, access to facilities, physical activity, and conflicts between users.

User and Task Analysis

Our discussions with transportation planning and civil engineering experts revealed three main audiences for us to support. These categories are not disjoint; for example, if given easy access to the necessary tools and data, motivated clients can become analysts in their own right.

Analysts: These are the expert users who are conducting the data analysis, either individually or in collaboration with others. Analysts have motivation and capability to learn complex interfaces as well as invest in long analyses. *Unique tasks* include creating, modifying, and presenting a visual exploration. Specific domain users include city planners (municipal workers who designs streets or approves designs), transportation services engineers (municipal workers who works on overall transportation issues), and consultants contracted to design streetscapes. Analyses are triggered by specific events, such as the city deciding to modify the streetscape, and advocates calling for an evaluation, changes to adjacent land use. The goals of the analyst generally include:

- Determine the best allocation of space in a corridor;
- Evaluate consistency of street design w.r.t. demands;
- Identify deficient corridors per transportation mode;
- Compare multiple alternatives w.r.t. costs and benefits;
- Survey clients to determine priorities and feedback;
- Iteratively refine street designs; and
- Convey designs to stakeholders, collect feedback.

Data consumer/manager: While not directly involved in data analysis, these users are deeply invested in the outcome of a data exploration. *Unique tasks* include interpreting and presenting the outcome from a visual data exploration. They can apply analysis to new data or may be capable of rudimentary analysis, but generally do not have time or skills for this. Specific domain users include city planners and transportation services engineers (when consultants are the primary analysts), municipal boards, police and emergency service agencies, maintenance providers, transit agencies, and advocacy groups. Their goals include:

- Understand analysis outcomes;
- Evaluate the costs and benefits of alternatives;
- Inform deliberation and negotiations;
- Improve transparency between stakeholders; and
- Understand an analysis process on demand.

Client: The end-user or stakeholder of an analysis, the client is a visualization consumer mainly interested in understanding findings. *Unique tasks* include following narratives and validating results. Clients are non-experts in analysis, and generally do not have the resources to engage in analyses themselves. Specific domain users include city councilors, residents and businesses in the study area, people and services that use transportation within the study area, and advocacy groups. Goals of clients include:

- Understand a design proposal;
- Understand analysis outcome and its implications;
- Provide feedback on a design; and
- Understand an analysis process on demand.

Design Rationale

The user and task analysis gave rise to several requirements:

- *Multiple media:* Effective visualization design depends on the intended audience and the nature of the presentation; each use case requires its own design.
- *One source:* Separating presentation from analysis is time-consuming and error-prone; presentations should update as the analysis changes.
- *Data flow system:* Visually representing data provenance promotes overview and recall, which is important for iterative refinement by multiple users.

5 PHASE II: INITIAL TOOL DESIGN

StoryFacets is a web-based visual analytics system for network data where nodes and edges have attributes (i.e., multivariate), and there are multiple types of nodes and edges (i.e., heterogeneous or multimodal [15]). StoryFacets's unique characteristic is that it maintains a multi-format representation of the data exploration, each representation called a *facet.* Users can chain together *cards* to visualize data or show annotations. The system automatically updates the different facet views that serve as different formats for presentation. Below, we discuss the rationale behind our design choices, the data model, and the user interface.

For the rest of the paper, we will use the StudentMoveTo survey data[1] collected by four universities in Toronto. The survey collected detailed data about where students live and travel throughout the day, as well as what factors influence their work, studies, and daily activities. The survey received 15,226 complete responses where students recorded over 36,000 trips in single-day travel diaries.

Data Model

We need to link multiple datasets together in order to answer domain questions, thus creating a network. This network has multivariate node and edge attributes as well as multiple types of nodes and edges, which we store in our database.

The main operations needed are linking, filtering, and pivoting. Linking is the process by which users create links between data tables they have loaded, building the network. Filtering is reducing the number of elements displayed through interaction. Pivoting is when users select a set of nodes in the network and then traverse edges to select a connected set of nodes. For example, we can do a many-to-many pivot from several students to the households they are part of by pivoting on the "student-household" edge type.

We also use a directed acyclic graph to track the history of an exploration. Nodes represent either (1) a source data set, (2) a subset of a dataset resulting from a query along with an associated visualization and parameterization of that visualization, (3) a data transformation, or (4) markdown content. Edges represent specific user operations.

Visual and Interaction Design

The StoryFacets design centers around the concept of *cards*—visual representations of analysis artifacts—that can be arranged on *facet views*—visual canvases used for exploration or presentation. A key feature of a card is that they are persistent across all views in the system. This eliminates the need to manually update a card when switching to a new view; by definition, all cards are always up to date.

Cards. The basic visual element used in StoryFacets is the *card*, which contains content along with interactive widgets for editing. Each card has UI widgets that allow cards to be selected, resized, scrolled, zoomed, reset (zoom and scroll), deleted, or made full screen. *Visualization cards* (Figure 1) include a visualization of a subset of the data from the *data flow system* as well as parameterization widgets and a caption. Visualization cards include additional widgets for filtering, pivoting, selection, axis and color variable selectors, and choosing relative or absolute sizing. Filtering and pivoting create new *child* cards that are linked to their *parent* card in the exploration history. For example, see the labeled links

[1] http://www.studentmoveto.ca/about/

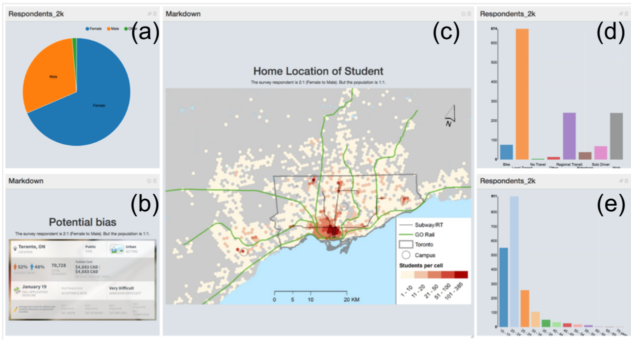


Figure 2: Dashboard view showing markdown and visualization cards in a customizable, space-filling orthogonal layout. (a) shows the gender composition of respondents. (b) warns of potential bias by comparing student enrollment statistics to survey results. (c) shows student home locations and transportation infrastructure. (d) shows the main commuting methods. (e) shows the age distribution of respondents.

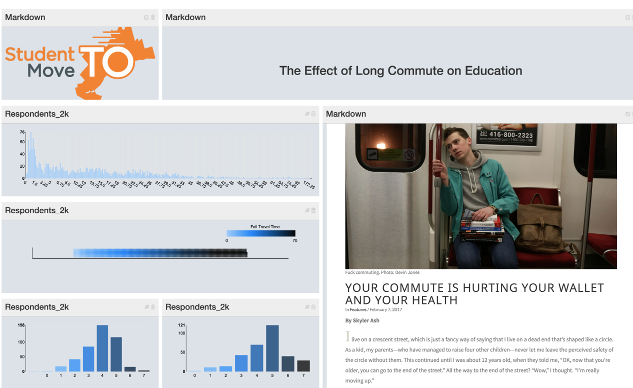


Figure 3: The dashboard view supports infographic elements such as text, images, videos, and mashups to tell a story.

between cards in Figure 4. Our modular architecture allows easy integration of new visualization types.

Facet Views. Cards can be combined into a *facet view* to present a subset of the exploration history and underlying data. A view is owned by a user and can be shared with others via hyperlink, either in live or read-only mode. Live views support *one source* and *media switching*, as edits are immediately updated in other views. Read-only views do not preserve edits so as to prevent unintentional alterations, but a read-only view can be forked as a new live view. To further support media switching, a link on each card takes the user to occurrences of that card in other views.

In order to support the *multiple media* design rationale, we provide different view designs for each of our target audiences: *trail view*, *dashboard view*, and *slideshow view*.

The *trail view* (Figure 4) is designed for *analysts* and provides a visual *data flow system* that renders the directed acyclic graph of the entire exploration history as a node-link diagram on a zoomable 2D canvas. Large rectangular nodes contain visualization cards, and directed edges show user operations that create new subsets of the data. New cards created by user interaction are automatically placed, and can then be rearranged. By placing visualization cards from parallel explorations beside each other, we support *direct comparison* through juxtaposition. Cards can be locked, preventing dependent cards to be affected.

The *dashboard view* (Figures 2 & 3) for *data consumers or managers* organizes cards in a compact layout. Users can choose which cards to include as well as move and resize cards to achieve the desired layout. Again, *direct comparison* between alternatives is achieved by juxtaposing cards.

The *slideshow view* (Figure 5) for *clients* shows each card as a slide. Users choose which cards to include as well as their order of appearance using an interface similar to Microsoft PowerPoint. Users can also include an existing dashboard view as a slide. Slides can be presented in full-screen mode with one card/dashboard per screen. Cards remain interactive both for editing and presentation.

Main Menu & Data Loading. Upon first logging in, the main menu provides a list of existing datasets and user-created views, which users can click to explore. Users can log in to fork a copy of an existing analysis and its associated views, or they can create a new session starting from a current dataset or upload their own CSV files. To help introduce users to a dataset, the menu shows a descriptive list of available node types (tables) which are clickable to display their attributes (columns). Users can create edges (many-to-many relationships) between node types by linking tables by key attributes. These edges can be directed or undirected, where the directionality affects pivoting.

Implementation Notes

We implemented StoryFacets as a web-based client/server framework. The client uses the AngularJS, D3 [3], and JQuery libraries. The server is based on Meteor.js and MongoDB.

6 PHASE III: FORMATIVE EVALUATION

We conducted a formative evaluation of the initial StoryFacets design to collect insights about appropriateness of different views for different purposes. Because we were looking for actionable guidelines on how to best improve the tool, we opted for an expert review [38] involving three experienced visual analytics researchers from industry.

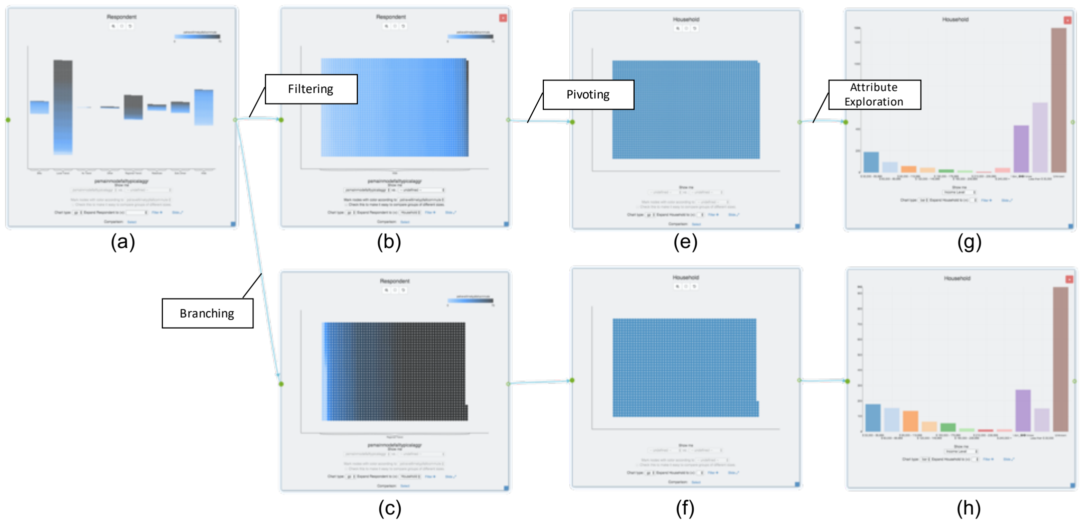


Figure 4: Visualization cards in an example trail view. (a) shows students by commute method. Node color represents individual commute times. To compare those groups, (b) shows the result of *filtering* to students who walk while (c) is filtered to students using regional transit. As (b) and (c) have the same parent card (a), we conceptually have a *branching* exploration. Through *pivots*, all households each student belongs to are shown in cards (e) and (f). We examine household income in (g) and (h). Students using regional transit tend to be from households with more income than students who walk. Perhaps this reflects the trend that students who live with wealthy parents tend to live in suburbs far from campus.

Figure 5: Slideshow view with markdown and visualization cards presented linearly. Each card is interactive and visualization parameters can be changed to explore the various attributes. Thumbnails for the other cards is seen on the left.

Method

We conducted study in order to better understand how experts would perceive and utilize the multiple views of the StoryFacets application. he expert users were recruited based on their experience in the field. Given the demands on their time, we only requested one hour for the study (actual times ranged between 59–65 minutes). Because our experts were geographically distributed, the expert reviews were conducted remotely using video conferencing with screen sharing. The experts were also briefed on the general problem statement that StoryFacets aims to address prior to beginning the study. We were most interested in their feedback about the use of multiple views for different purposes or audiences. Thus, we specifically told the experts we were interested in their feedback about the design.

Since our experts were all external to the project, we decided to use Star Wars movie data from the Star Wars API (https://swapi.co/) rather than specific urban transportation planning data, which would have required lengthy instructions. This dataset contains various types of information about characters, places, and vehicles in the fictional Star Wars universe. As with the types of data commonly used for urban transportation planning, the Star Wars data also involves multidimensional information and entity relationships, which makes it sufficient to demonstrate the true design focus—the multiple linked facets of the StoryFacets tool.

The procedure called for our experts to explore the data in a free-form fashion while speaking their observations and thoughts aloud. To guide their exploration, we seeded the experts with a list of questions.

Results

Our experts (E1, E2, and E3) provided qualitative feedback while answering questions as well as related ones. Here, we focus on qualitative feedback rather than task performance. The main finding is that the experts all saw different and specific usages for the various views.

All three experts stated that the trail view was better suited for a technical audience. E3 further thought this view would be the best for sharing results, at least with other analysts. However, E2 noted that a complicated analysis with many branching paths may cause the trail view to grow out of control. Nevertheless, the trail view was collectively lauded across all three experts; this was not surprising, as they were representatives of the intended users of this view.

The dashboard view was seen by E1 and E2 as most suitable for presentation to a less technical audience (e.g., management). E1 thought that it represents a good tradeoff between clarity and flexibility, and can even be used to explain complex data analysis with many branches. Simple branching can be shown as parallel rows or columns, especially with annotations. It could also support off-the-cuff presentations of an analysis currently in progress when the optimal order of presentation has not yet been established. E2 particularly enjoyed the animated data transitions in the slideshow view.

Of the three views, the slideshow view was the most controversial; all experts agreed that its utility was limited to presentations to novice stakeholders, but that it was highly useful for this specific purpose. E1 noted that creating a slideshow requires knowing the correct order of presentation, which is not always known in the midst of data analysis, but E2 stated this was the very aspect of the slideshow view that made it appealing once such an order is established. One compelling scenario E1 suggested was that when preparing a routine presentation for management, a traditional presentation can be quickly and easily created. However with StoryFacets, a presenter could switch to the trail view and retarget the presentation for an unengaged audience.

The experts all gave suggestions for future improvement. E1 noted a Prezi-style interactive tour of the workspace in the trail view could be a good alternative to present the state of a visual exploration to other analysts. Both E2 and E3 agreed that the trail view may also be useful for communication, and suggested adding the ability to add annotations directly. This was unexpected because we designed the trail view to be primarily an exploration space and the others as presentation formats. However, in many cases, this boundary is not strict, and the trail view can also function as an effective communication medium. This leads to a fundamental trade-off between related to provenance and presentation. For representing data provenance, a complex analysis trail should be preserved as a reminder of previous actions. But for the purpose of presentation to non-experts, complex exploration processes with branching analysis pathways and dead-end results may be irrelevant, redundant, or unrelated to the intended message. E2 noted a data flow system such as our trail view could easily become visually complex, and suggested simplifying the workspace with mechanisms such as editing exploration paths, collapsing or expanding branches, or eliminating fruitless paths.

E3 suggested version management for cases when the original author shares the exploration and colleagues build on the exploration in the original space. This use case raises important questions about how to facilitate modifications, notify the original author, and visualize differences across versions of the same exploration.

Outcome: Modifications Needed

Since the intention of this study was to guide the design of StoryFacets, a key aspect was identifying actionable modifications. Here are the changes needed based on the study:

- Adding annotation capabilities to each view;
- Adding standard visualization types;
- Making cards and views responsive and resizable;
- Rectangular, individual item, and modifier key interactions for selecting items and aggregates;
- Maintaining consistent color scales across cards;
- Using a natural sort order for labels that sorts string and numeric components separately; and
- Fixing label overplotting.

7 PHASE IV: ITERATIVE TOOL REFINEMENT

We refined StoryFacets iteratively based on the results of the formative expert review as well as informal usability tests.

To support annotations—as well as to integrate qualitative data into an exploration—we developed markdown cards. *Markdown cards* allow annotations and qualitative content to be added, such as text captions, bullet points, hyperlinks, images, video, and even interactive webpages. These can be used for integrating the results of analyses conducted in other tools, including embedded web pages, images, and video. Examples are shown in Figure 3 (a), (b), & (d), where they are used to create an infographic-style interactive dashboard suitable for novice users. Markdown cards can also take the place of non-visualization slides in the slideshow. This can help build a strong narrative about the data.

Several changes we made were targeted at increasing consistency and readability. This included using top-level color scales for each attribute for all views as well as an improved label ordering algorithm. We also implemented responsive resizing of cards in the trail and dashboard view to support exploring elements in more detail.

While our project was canceled at this stage, preventing us from running an in-depth followup study with our group of urban transportation planners, informal feedback on the new version of our tool was very positive.

8 STUDY: UTILITY OF STORYTELLING FORMATS

Given the refined StoryFacets tool, we use the application to address our initial hypothesis on the utility of different storytelling formats to support data analysis in diverse teams.

Method

To study how different presentation formats influence presentations, we designed a study in which participants prepared a presentation about a brief data analysis session and then presented to an experimenter. We compared the use of all three views in StoryFacets, varying the view between subjects so each participant would prepare and deliver one presentation with an assigned format.

For study participants, we recruited participants from the undergraduate and graduate student population at our university. We recruited 24 participants to complete the study (9 female) with a median age of 24. The participants' academic majors varied widely, but half of the participants focused on graphics and animation. Due to the experience of this participant population, we felt we could not use the highly technical urban transportation data, so we again used the Star Wars dataset for the study.

We created two sample analyses using the trail view. The analyses were video recorded along with think-aloud comments explaining the goals, intentions, actions, and findings of the "analyst" (in this case, a researcher). With this approach, our study ensures experimental control for the base content all participants were asked to present. Both analyses were designed to be similar in scope and complexity; both lasted approximately six minutes and involved the creation of multiple branches of visualization cards while investigating aspects of the Star Wars data. One analysis involved the investigating the relationship between vehicles and character genders, and the another analysis involved exploring differences in starships used in different movies.

Upon arrival and providing consent, participants completed a background questionnaire with questions about age, gender, education, academic discipline, and data visualization experience. The experimenter then gave a short demonstration of StoryFacets while explaining the core features of the trail view. Participants were next given instructions about the presentation they would be asked to make. They then watched and listened to one of the prepared analysis videos.

After viewing the analysis, the experimenter explained the assigned view format to be used for the presentation. The experimenter also explained the requirements and goals of the presentation: in a maximum of five minutes describe the analyst's findings as well as the steps taken throughout the analysis. The participant then had ten minutes to prepare a presentation using the assigned StoryFacets view, a mouse, a keyboard, and a 27-inch monitor. Every participant started with the default view of the presentation that was automatically generated for the assigned view mode from the assigned sample analysis (thus, the beginning state of the presentation materials were controlled for by the independent variables). Next, the participant moved to the opposite side of the room where they used a different computer attached to a 64.5-inch display to give the presentation with their prepared materials. The presentations were video and audio recorded for later analysis. The entire procedure took approximately one hour per participant.

Results

To analyze the results, we reviewed the prepared presentation material along with the recordings of the presentations themselves. We counted the cards in the presentation view and categorized the presentation style according to its structural organization and temporal order.

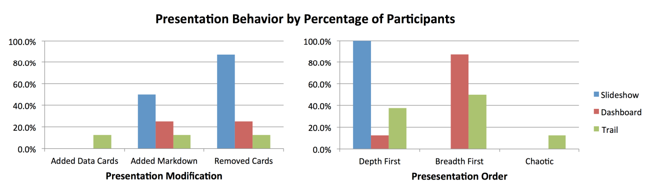


Figure 6: The results summary from shows how the different views affected presentation approaches.

A summary of the results is shown in Figure 6. The left side of Figure 6 shows different ways participants modified the presentation material during preparation. Of all the participants, only one opted to add any new visualization cards when preparing the presentation; this participant was in a *trail* condition. However, seven participants (of 24) added new markdown cards and content. Half the participants in the slideshow condition added supplemental markdown, while only two participants added markdown to the dashboard and one added to the trail view. This behavior suggests that participants were more likely to think the existing material was sufficient for presentation with the dashboard and trail views, but the slideshow-style presentations sometimes required some extra modification to assist in the explanation. The results for removal of the default material were similar. Ten participants removed cards or slides from the default set, and most of these (seven) were using the slideshow.

The right side of Figure 6 summarizes different styles of presentation based on the order content was presented. All participants in the slideshow condition presented using a depth-first approach, meaning that they presented one entire branch of the analysis at a time, detailing each card in order along the branch. It makes sense that the linear nature of the slideshow organization would encourage this style. In contrast, most participants (all but one) assigned the dashboard view used a breadth-first approach during their presentation. These participants attempted to explain multiple branches of the analysis in parallel, jumping back and forth between cards to draw comparisons and describe relationships between different data plots. The trail view was the most varied in terms of presentation style, as shown by the green bars in right side of Figure 6. Some participants used a depth-first approach while others used a breadth-first strategy. In addition, one participant used a "chaotic" strategy that could not be characterized by either depth-first or breadth-first labels. Interestingly, this participant did an excellent job providing a complete summary of the analysis content, though the presentation did not clearly describe the order the analyst used to arrive at the findings.

Overall, these results demonstrate the importance of presentation format for the nature and style of presentation delivery, which is the guiding principle behind the StoryFacets design. The familiar slideshow-style presentations were naturally better suited for linear presentations, while the dashboard view supported a more relational style of explanation to describe connections among items. The results support the notion that the more open workspace of the trail view provided more flexibility for use of different approaches, but participants also were less inclined to supplement the presentation space with additional material (i.e., images and annotations). When asked about this latter behavior, participants reported that they were comfortable and confident enough with being able to present the analysis without making many changes to the workspace. This further supports the hypothesis that the spatial layout of the trail view's workspace can be effective for communication.

9 DISCUSSION

The results from our evaluations support the hypothesis that different views have different uses. In our study, the trail view supported more flexible presentation styles, while the slideshow view caused participants to present in a more linear fashion. Additionally, the formative study in Phase III contributed evidence that the trail view was preferred for data analysis, whereas the slideshow view was favored for formal presentations. This confirms that the basic rationale behind the StoryFacets is sound: data exploration can be viewed through several radically different lenses, each with a unique and valuable *raison d'etre*. In other words, this is a validation of our "*one source, multiple media*" motto that arose in the early stages of this design study.

Designing appropriate views can be challenging, however. Many participants in the user studies expressed the notion that the slideshow view was best suited for formal presentations, or when an exploration was "finished" and should not be changed. We suspect that these opinions stem from familiarity; despite the ability to create interactive presentations that encourage additional exploration, people are simply not used to interacting during a standard slideshow presentation.

Finally, our current StoryFacets prototype aptly demonstrates the concept of multi-representation visual exploration and presentation, but it does not provide a polished and finished product. Unfortunately, our overall research project on urban transportation planning was canceled, and we never got far enough to actually deploy this work with our intended users. However, while much more work clearly remains to be done, the findings from this design study will be formative and points to the overall utility of the ideas we uncovered in the project. One specific goal for future improvement includes media and mechanisms for sharing exploratory analysis. Another idea may be to automatically organize visualizations and findings based on the content. For example, if two adjacent exploration branches show two related data subsets, the dashboards or slide layout algorithm should be able to position them intelligently side-by-side to enable easy comparison. Similarly, the animation for the filtering in the slide deck can be improved so the items maintain their identity over the slides. Animation support could show entities appearing and disappearing in response to filtering and pivoting operations to maintain object identity.

10 CONCLUSION AND FUTURE WORK

Exploratory data analysis involves much more than the initial data exploration that generates findings; the analysis and findings must often be shared with colleagues, discussed with managers, and eventually presented to stakeholders or the general public. In this paper, we reported on an in-depth design study with urban transportation planners, which yielded a common theme about the need for multiple stakeholders of varying expertise to have access to the outcomes of data analysis. As a result, we designed StoryFacets, a communication-minded visualization system that provides multiple, linked visual formats for analysis and presentation. A unified platform that supports all of these goals can enable a smooth and streamlined workflow for all stages of this process. Our evaluations confirm this design rationale and provide new insights for designing multi-faceted approaches.

StoryFacets is just the beginning of the new generation of one-source/multiple-media visualization systems. We hope to continue further investigation of such platforms in the future, looking specifically at methods to automatically generate other media types such as reports, animations, and guided tours directly from analysis artifacts.

ACKNOWLEDGMENTS

[Anonymized for double-blind review.]

REFERENCES

[1] Christopher Ahlberg. 1996. Spotfire: An Information Exploration Environment. *SIGMOD Record* 25, 4 (1996), 25–29. https://doi.org/10.1145/245882.245893

[2] Louis Bavoil, Steven P. Callahan, Carlos Eduardo Scheidegger, Huy T. Vo, Patricia Crossno, Cláudio T. Silva, and Juliana Freire. 2005. VisTrails: Enabling Interactive Multiple-View Visualizations. In *Proceedings of the IEEE Conference on Visualization*. IEEE, Piscataway, NJ, USA, 135–142. https://doi.org/10.1109/VISUAL.2005.1532788

[3] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3: Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 6 (2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[4] Ken Brodlie, Andrew Poon, Helen Wright, Lesley Brankin, Greg Banecki, and Alan Gay. 1993. GRASPARC: A Problem Solving Environment Integrating Computation and Visualization. In *Proceedings of the IEEE Conference on Visualization*. IEEE, Piscataway, NJ, USA, 102–109. https://doi.org/10.1109/VISUAL.1993.398857

[5] Mark Derthick and Steven F. Roth. 2001. Enhancing data exploration with a branching history of user operations. *Knowledge-Based Systems* 14, 1-2 (2001), 65–74.

[6] Nick Diakopoulos, Joan DiMicco, Jessica Hullman, Karrie Karahalios, and Adam Perer. 2011. Telling Stories with Data: The Next Chapter—A VisWeek 2011 Workshop. (2011).

[7] Joan DiMicco, Matt McKeon, and Karrie Karahalios. 2010. Telling Stories with Data—A VisWeek 2010 Workshop. (2010).

[8] Wenwen Dou, Dong Hyun Jeong, Felesia Stukes, William Ribarsky, Heather Richter Lipford, and Remco Chang. 2009. Recovering reasoning process from user interactions. *IEEE Computer Graphics & Applications* 29, 3 (2009), 52–61. https://doi.org/10.1109/MCG.2009.49

[9] Cody Dunne, Nathalie Henry Riche, Bongshin Lee, Ron Metoyer, and George Robertson. 2012. GraphTrail: Analyzing Large Multivariate, Heterogeneous Networks while Supporting Exploration History. In *Proceedings of the ACM Conference on Human Factors in Computer Systems*. ACM, New York, NY, USA, 1663–1672. https://doi.org/10.1145/2207676.2208293

[10] Ryan Eccles, Thomas Kapler, Robert Harper, and William Wright. 2008. Stories in GeoTime. *Information Visualization* 7, 1 (2008), 3–17. https://doi.org/10.1057/palgrave.ivs.9500173

[11] Niklas Elmqvist, John Stasko, and Philippas Tsigas. 2007. DataMeadow: a visual canvas for analysis of large-scale multivariate data. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology*. IEEE, Piscataway, NJ, USA, 187–194. https://doi.org/10.1109/VAST.2007.4389013

[12] Juliana Freire, David Koop, Emanuele Santos, and Cláudio T Silva. 2008. Provenance for computational tasks: A survey. *Computing in Science & Engineering* 10, 3 (2008), 11–21. https://doi.org/10.1109/MCSE.2008.79

[13] Tong Gao, Jessica Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. 2014. NewsViews: an automated pipeline for creating custom geovisualizations for news. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3005–3014. https://doi.org/10.1145/2556288.2557228

[14] Nahum D. Gershon and Ward Page. 2001. What storytelling can do for information visualization. *Commun. ACM* 44, 8 (2001), 31–37. https://doi.org/10.1145/381641.381653

[15] Sohaib Ghani, Bum Chul Kwon, Seungyoon Lee, Ji Soo Yi, and Niklas Elmqvist. 2013. Visual Analytics for Multimodal Social Network Analysis: A Design Study with Social Scientists. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2032–2041. https://doi.org/10.1109/TVCG.2013.223

[16] Jeffrey Heer, Jock D. Mackinlay, Chris Stolte, and Maneesh Agrawala. 2008. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1189–1196. https://doi.org/10.1109/TVCG.2008.137

[17] Jessica Hullman and Nicholas Diakopoulos. 2011. Visualization Rhetoric: Framing Effects in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2231–2240. https://doi.org/10.1109/TVCG.2011.255

[18] Jessica Hullman, Steven M. Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. 2013. A Deeper Understanding of Sequence in Narrative Visualization. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2406–2415. https://doi.org/10.1109/TVCG.2013.119

[19] Waqas Javed and Niklas Elmqvist. 2010. Stack Zooming for Multi-Focus Interaction in Time-Series Data Visualization. In *Proceedings of the IEEE Pacific Symposium on Visualization*. IEEE, Piscataway, NJ, USA, 33–40. https://doi.org/10.1109/PACIFICVIS.2010.5429613

[20] Waqas Javed and Niklas Elmqvist. 2013. ExPlates: spatializing interactive analysis to scaffold visual exploration. *Computer Graphics Forum* 32, 3pt4 (2013), 441–450. https://doi.org/10.1111/cgf.12131

[21] Robert Kosara. 2013. Story Points in Tableau Software. Keynote at Tableau Customer Conference. (Sept. 2013).

[22] Robert Kosara and Jock D. Mackinlay. 2013. Storytelling: The Next Step for Visualization. *IEEE Computer* 46, 5 (2013), 44–50. https://doi.org/10.1109/MC.2013.36

[23] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. 2013. SketchStory: Telling More Engaging Stories with Data through Freeform Sketching. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2416–2425. https://doi.org/10.1109/TVCG.2013.191

[24] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2013. Patina: Dynamic heatmaps for visualizing application usage. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3227–3236. https://doi.org/10.1145/2470654.2466442

[25] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of the International Conference on Intelligence Analysis*, Vol. 5. MITRE Corporation, Bedford, MA, USA, 2–4.

[26] Zachary Pousman, John T. Stasko, and Michael Mateas. 2007. Casual information visualization: Depictions of data in everyday life. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1145–1152. https://doi.org/10.1109/TVCG.2007.70541

[27] Eric D. Ragan, Alex Endert, Doug A. Bowman, and Francis Quek. 2012. How spatial layout, interactivity, and persistent visibility affect learning with large displays. In *Proceedings of the ACM Conference on Advanced Visual Interfaces*. ACM, New York, NY, USA, 91–98. https://doi.org/10.1145/2254556.2254576

[28] Eric D. Ragan, Alex Endert, Jibonananda Sanyal, and Jian Chen. 2015. Characterizing Provenance in Visualization and Data Analysis: An Organizational Framework of Provenance Types and Purposes. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 31–40. https://doi.org/10.1109/TVCG.2015.2467551

[29] Eric D. Ragan, John R. Goodall, and Albert Tung. 2015. Evaluating How Level of Detail of Visual History Affects Process Memory. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2711–2720. https://doi.org/10.1145/2702123.2702376

[30] Ali Sarvghad and Melanie Tory. 2015. Exploiting analysis history to support collaborative data analysis. In *Proceedings of the Graphics Interface Conference*. ACM, New York, NY, USA, 123–130.

[31] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring narrative visualizations with Ellipsis. *Computer Graphics Forum* 33, 3 (2014), 361–370. https://doi.org/10.1111/cgf.12392

[32] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. 2012. Design Study Methodology: Reflections from the Trenches and the Stacks. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2431–2440. https://doi.org/10.1109/TVCG.2012.213

[33] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148. https://doi.org/10.1109/TVCG.2010.179

[34] Yedendra Shrinivasan and Jarke van Wijk. 2008. Supporting the analytical reasoning process in information visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1237–1246. https://doi.org/10.1145/1357054.1357247

[35] James J. Thomas and Kristin A. Cook (Eds.). 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, Piscataway, NJ, USA.

[36] Matthew Tobiasz, Petra Isenberg, and M. Sheelagh T. Carpendale. 2009. Lark: Coordinating Co-located Collaboration with Information Visualization. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (2009), 1065–1072. https://doi.org/10.1109/TVCG.2009.162

[37] Toronto Centre for Active Transportation. 2012. Complete Streets for Canada. (2012). http://completestreetsforcanada.ca Retrieved September 6, 2018.

[38] Melanie Tory and Torsten Möller. 2005. Evaluating Visualizations: Do Expert Reviews Work? *IEEE Computer Graphics and Applications* 25, 5 (2005), 8–11. https://doi.org/10.1109/MCG.2005.102

[39] Fernanda B. Viégas and Martin Wattenberg. 2006. Communication-minded visualization: A call to action. *IBM Systems Journal* 45, 4 (April 2006), 801–812. https://doi.org/10.1147/sj.454.0801

[40] William Wright, David Schroh, Pascale Proulx, Alex Skaburskis, and Brian Cort. 2006. The Sandbox for Analysis: Concepts and Evaluation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 801–810. https://doi.org/10.1145/1124772.1124890