Information Visualization (2007) 6, 89 – 103

www.palgrave-journals.com/ivs

Visual exploration and analysis of historic hotel visits

Chris Weaver[1]
David Fyfe[1]
Anthony Robinson[1]
Deryck Holdsworth[1]
Donna Peuquet[1]
Alan M. MacEachren[1]

[1]*The Pennsylvania State University, University Park, PA, U.S.A.*

Correspondence:
Chris Weaver, 302 Walker Building, The Pennsylvania State University, University Park, PA 16802, U.S.A.
Tel: +1 814 865 3472;
Fax: +1 814 863 7943;
E-mail: cweaver@psu.edu

Abstract
Understanding the spatial and temporal characteristics of individual and group behavior in social networks is a critical component of visual tools for intelligence analysis, emergency management, consumer analysis, and human geography. Identification and analysis of patterns of recurring events is an essential feature of such tools. In this paper, we describe an interactive visual tool for exploring the visitation patterns of guests at two hotels in central Pennsylvania from 1894 to 1900. The centerpiece of the tool is a wrapping spreadsheet technique, called reruns, that reveals regular and irregular periodic patterns of events in multiple overlapping artificial and natural calendars. Implemented as a coordinated multiple view visualization in Improvise, the tool is in ongoing development through an iterative process of data collection, transcription, hypothesis, design, discovery, analysis, and evaluation in close collaboration with historical geographers. Numerous discoveries have driven additional data collection from archival newspaper and census sources, as well as plans to enhance analysis of spatial patterns using historic weather records and railroad schedules. Distributed online evaluations of usability and usefulness have resulted in feature and design recommendations that are being incorporated into the tool.
Information Visualization (2007) **6,** 89 – 103. doi:10.1057/palgrave.ivs.9500145

Keywords: Geovisual analytics; exploratory visualization; historical geography; coordinated multiple views; social behavior analysis; serial periodic data

Introduction

Promoting the development of integrated software for visual data analysis is a key facet of the research and development agenda for visual analytics.[1] In both information and geographic visualization, an important goal is to allow users to explore information that contains geospatial, temporal, and abstract components in a flexible, integrated, interactive graphical environment that requires minimal training and little or no programming to use.[2] In pursuit of this goal, Improvise[3] is an exploratory visualization application – written entirely in Java and freely available on the web under an open source license – in which analysts rapidly build and browse multiple coordinated views of their data. What Improvise brings to visual analytics is precise control over how interaction affects the presentation of spatial, temporal, and abstract information in and between multi-layer maps, scatter plots, parallel coordinate plots, tables, spreadsheets, and other views. Most importantly, Improvise visualizations can be rapidly modified and extended to develop hypotheses and exploit discoveries during ongoing exploration and analysis.

This paper describes the evolutionary process of developing a visual tool for exploring visitation patterns in hotel guest registers. Genealogists and autograph collectors have long valued information found in the registers of historic hotels. Academic researchers can also extract a wealth of information about commercial and cultural connectivity patterns from these

Received: 23 June 2006
Revised: 31 July 2006
Accepted: 23 October 2006
Online publication date: 1 February 2007

archival documents. Most registers contain a date, name, and home town for each visitor. They may also include information such as whether the guest represented a business, minstrel show, or baseball team, whether they were traveling with companions, or whether they stabled a horse in the livery.

Starting with data transcribed from guest registers for two hotels in central Pennsylvania from 1894 to 1900, the visualization evolved through an iterative process of creating views, populating them with data, and coordinating them in terms of user interaction. This tightly integrated method of building and browsing visualizations in Improvise makes it possible to realize and experiment with different displays of the hotel data in a matter of hours, including a multi-layer map of guest hometowns along railroads and rivers, table views containing nested bar plots of visits over time for individual hotels, guests, and residences, and an arc diagram showing sequences of visits by guests individually and in groups.

The central component of the visualization is a glyph-based technique called *reruns* that displays temporal information in the cells of a wrapping spreadsheet, thereby allowing exploration of periodic temporal patterns across multiple artificial and natural calendars. Using glyphs rather than pixels (as in visualization of infant sleep cycles[4]) allows encoding of more than two cyclic timescales in a single view. Interactive adjustment of period length using reruns enables visual identification of cyclic behavior of varying duration, much the same as in spiral-based visualizations of periodic data.[5]

Extensive cross-filtering between the reruns view, map, and other views allows analysts to follow chains of evidence involving complex groupings of people, dates, and places. Moreover, the data domain, rapid design approach, and analysis process appear to be directly applicable to current efforts to develop visual tools for use in intelligence analysis and emergency response management. Using the visualization, historical geographers have discovered numerous interesting patterns that are motivating the transcription of additional hotel guest registers as well as the collection of historic weather and transportation records. Evaluation of the tool suggests strong support for continued efforts along these avenues as well.

Data collection and transcription

The register for the Rebersburg Hotel[6] lists daily guests at a small hotel in Rebersburg, a village in the Brush Valley region of Centre County, Pennsylvania, between June 1898 and November 1900. All told there are 2411 guest entries. Some 219 different places of origin are recorded, including 162 towns and cities in Pennsylvania, 14 in New York, and seven in Ohio, as well as 36 places in 13 other states, plus Canada and a visitor from London, England.

A second hotel register, this one for the National Hotel (Figure 1) in Roaring Spring,[7] a small railroad town located in Morrison's Cove region of Blair County, covers the period from December 1894 through December 1900.

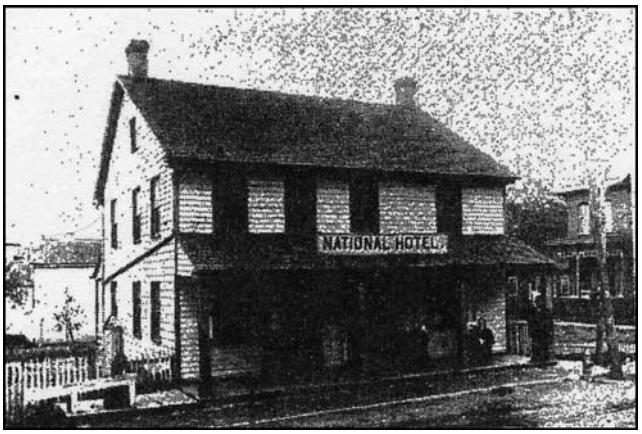

Figure 1 The National Hotel in Roaring Spring, PA.

Figure 2 The Rebersburg Hotel Register.

However, this register was transcribed in two phases, the first from December 1894 through December 1897 and second from June 1898 through December 1900. There were 3947 guest entries for the first 3 years and 4128 entries during the two and a half years that correspond to the same period as the Rebersburg Hotel register. Unlike Rebersburg, Roaring Spring is located on a line of the Pennsylvania Railroad and likely benefited by having more traffic. Indeed, the National Hotel had 1500 more guests than the Rebersburg Hotel during the same two-and-a-half year period. Visitors hailed from more than 230 different places in Pennsylvania, and 27 different states.

With more than a hundred guests each month signing each of the registers from this wide variety of places, thousands of data points represent patterns of local and regional connectivity for small-town communities in central Pennsylvania. However, when analyzing the actual archival documents (Figure 2), patterns were not immediately apparent. Barely legible signatures and cryptic notations were bewildering and the eye rested on

the bizarre, such as a sketch of a log cabin added by a traveling vaudeville company or the batting line-up of a visiting baseball team.

The data was transcribed into spreadsheets so that visits could be sorted and temporal and spatial patterns discerned using computational methods. The process of transcribing each entry into appropriate columns in the spreadsheets took approximately 80 h for the Rebersburg data and 300 h for the Roaring Spring data. Once the data was transcribed, the columns for guests and residences were sorted to check for overt errors and/or inconsistencies within the dataset. Place-name indexes were consulted to clarify settlement names. Some place names could not be deciphered or located, nor could some surnames (overall, 6% of the dataset was not legible). Summary tables were made for total number of visits per person and per declared home residence. Then, graphs of these tables were created to provide a crude visual representation of which visitors and residences appeared most often. Many guests appeared several times throughout the two-and-a-half-year period covered by both registers. Forty people registered at the Rebersburg Hotel on at least four occasions; one individual came 78 times. For the National Hotel these figures were much larger, with 122 individuals registering at the hotel on at least four different occasions during the first 3 years; one individual registered 83 times. The dramatic difference in repeat visitors could be attributed to the fact that Roaring Spring is easier to reach, as it is located on the railroad lines. Further research using other archival resources revealed that some of these repeat visitors were draymen (heavy cart drivers), hucksters (local peddlers), clerks, and traveling salesmen. Three of the hucksters visiting Rebersburg, A. M. Sheats, R. A. Sheats, and Harry Welshans, came to the hotel a combined 148 times from Booneville in neighboring Sugar Valley. For each of these men, the observed pattern of visits involves a single day of the week: Tuesdays for Welshans, Fridays for both Sheats. E. K. Hess, a traveling salesman from Williamsport, paid 40 visits to the Rebersburg Hotel. Like the Booneville hucksters, his visits were also predominantly on a single day, in his case Thursdays, with very few visits on any other day of the week. A graph of monthly visit totals by all guests revealed evidence of seasonal rhythms; some summer months had nearly twice the number of visitors as winter months.

The geographical location of each home residence was obtained by merging residence columns with an existing spatial dataset. The residence location and number of visits were then mapped using commercial GIS software, revealing decay in the number of visits as a function of distance. More than 1800 visits to the Rebersburg Hotel were by people whose claimed residence was less than 50 miles from Rebersburg, and more than half were visiting from within 20 miles of the hotel. Many of these places were accessible only by horse and wagon, as the nearest railroad stations were nine miles south in Coburn and 20 miles north in Mill Hall. Similar spatial patterns exist for Roaring Spring guests with the exception that most visitors came from places that were also on railroad lines. However, given the number of guests traveling with horses, it is evident that many residents throughout the valley still relied on horse and buggy for their travels 'into town'.

The spatial and temporal patterns found by these methods were limited to aggregated representations of the data. The texture of individual dates and guests was lost. Moreover, these search methods were relatively time consuming. In order to understand the intricacies of social connectivity patterns at the level of individual people and relationships, new methods are needed to bridge historical domain knowledge with visual analysis techniques. (For examples of historical geographers working in this area, see Holdsworth[8] and Knowles.[9])

Exploratory visualization in Improvise

Improvise is a self-contained, web-enabled visualization builder and browser with capabilities inspired by earlier coordinated multiple view visualization systems such as DEVise,[10] DataSplash,[11] and Snap-Together Visualization.[12] These systems support rich browsing using mouse and keyboard interactions to navigate multidimensional space and select data items across multiple coordinated views. Unfortunately, building occurs either through programming or in simple design interfaces that support small libraries of views, data sources, querying algorithms, and visual representation techniques. Conversely, sophisticated visualization toolkits such as VisAD,[13] the Infovis Toolkit,[14] and prefuse[15] provide extensive libraries of visualization components, but are primarily intended as programming interfaces and as such provide only rudimentary user interfaces for building and browsing visualizations interactively. Although GeoVISTA *Studio*[16] integrates an extensible JavaBeans-based (geo)visualization library with live, graph-based visual design of visualization tools, its coordination architecture limits the practical number of views and amount of coordination that can be built into its visualizations.

Improvise consists of a graphic user interface on top of a modular library of visualization components (Figure 3). The interface follows the multiple document model in which one or more visualizations run in parallel, each in its own top-level window. The builder interface is built around a set of editor dialogs, local to each visualization window, in which designers access data, specify data queries, and create, coordinate, and layout views. The browser interface consists of the set of views as they are laid out inside each window. All changes made in editing dialogs take place immediately and reversibly without the need for a separate compilation stage, resulting in a live, incremental, organic style of visualization design that has the added benefit of allowing trained users to switch rapidly between building and browsing. Several enhancements facilitate both tasks, including predefined macros that recreate common coordinated view 'chunks', built-in queries, screenshots, and integrated

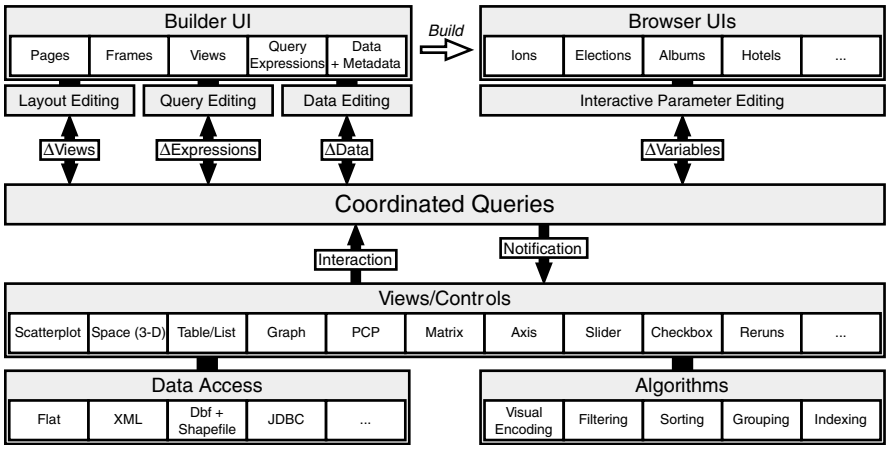

Figure 3 The Improvise software architecture.

metavisualization that allows users to visualize coordination structure *in situ* as they work.[17] The goal is to facilitate open-ended visual exploration and analysis of information by close-knit teams of collaborating domain experts and experienced visualization designers. In other words, the goal is to support *improvisational visualization* on much shorter timescales than previously possible.

Improvise visualizations are built around a coordination model called *Live Properties* that is coupled with a visual abstraction and data processing language called *Coordinated Queries*. Coordination takes place through shared interactive parameters that determine what, how, and where views display data. Multiple views are coordinated whenever they are connected through at least one such parameter. Coordinated Queries is a flexible, yet high-level visualization query language for coordinating access, processing, and rendering of multiple datasets across multiple views. Query expressions specify how to map data attributes into graphical attributes in views. Multiple datasets can be loaded, indexed, grouped, filtered, sorted, and visually encoded in terms of navigation and selection in and between multiple views. The combination of Live Properties and Coordinated Queries enables open-ended visual analysis by allowing users to design, construct, explore, and extend highly coordinated visualizations of multiple simultaneous datasets interactively.

Improvise has been used to build a wide variety of information and geographic visualizations involving spatial, temporal, and abstract data attributes. The ongoing evolution of the hotel register visualization is one of the major successes of this effort. (Because they are just a few of many Improvise 'documents' – all saved as regular, self-contained XML files – we have not adopted formal names for the various incarnations of the visualization as independent tools. Internally, we refer to them collectively simply as 'the hotels vis.')

Hotels visualization interface

Using Improvise, we have developed a visualization of guest registers for two hotels in central Pennsylvania from 1894 to 1900. The visualization is a result of an iterative process of design, construction, testing, exploration, and evaluation involving close collaboration between visualization researchers (Weaver, Robinson, Peuquet, and MacEachren) and historical geographers (Fyfe and Holdsworth). This process has resulted in five major versions and numerous minor versions of the visualization over the course of nine months. Continuing evolution of the visualization is motivated primarily by the exploratory and analytic needs of the historical geographers, who have adopted the visualization as an important tool in their research arsenal.

Figure 4 shows a snapshot of the visualization in the course of being used to detect events involving a vanguard for an organized group. The current visualization design consists of the following views:

- A table view, showing the location, total visits, and pattern of visits over time for each hotel.
- A table view, showing the name, total visits, and pattern of visits over time for each guest, including many with uncertain or completely illegible names.
- A table view, showing the same information for each residence (place of origin), with similar occurrences of uncertainty.
- A reruns view, showing total visits on each day using text and/or color. Squares distinguish weekends from circular weekdays. Months and seasons optionally appear as cell edges and a fill gradient, respectively. Cycle length, cycle phase, and cell size can be rapidly adjusted using sliders. Mousing over a cell shows its date in the top left corner.
- A vertical histogram along the right side of the reruns view, summarizing total visits for each period.

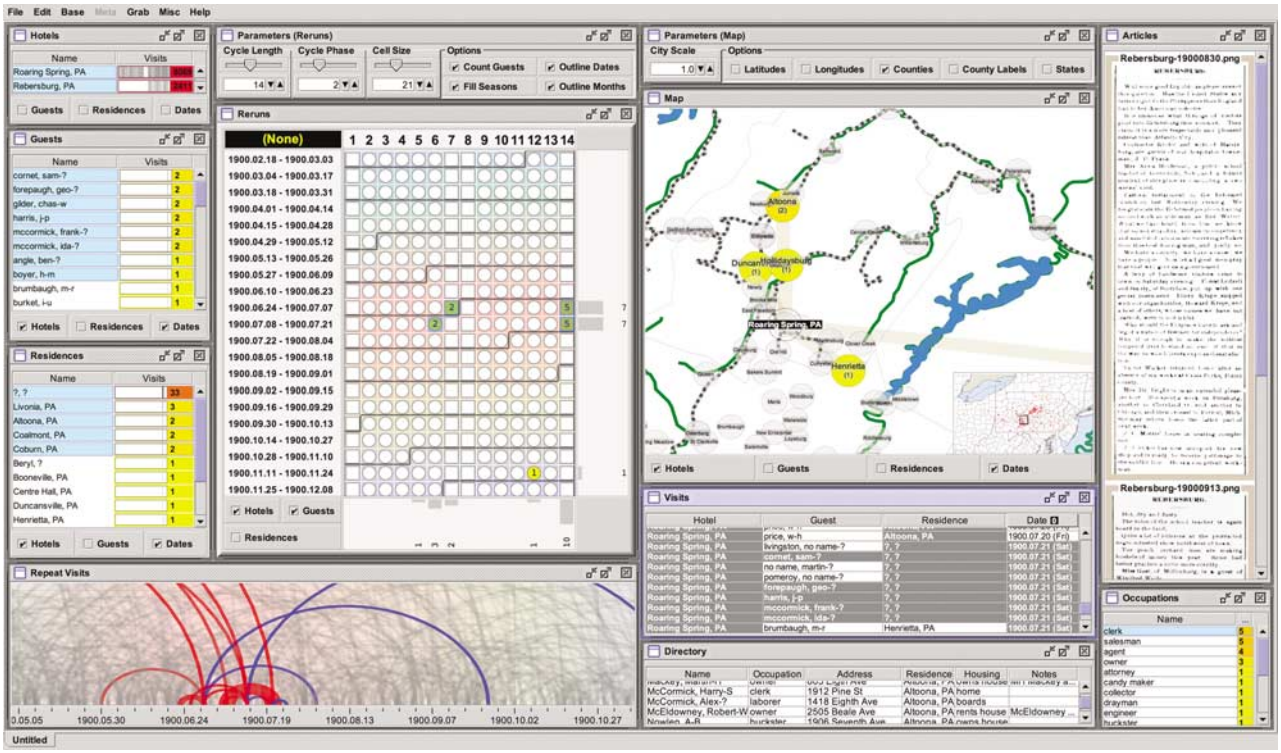

Figure 4 The hotels visualization.

- A horizontal histogram along the bottom of the reruns view, summarizing total visits for each day in the cycle.
- A multi-layer map, showing crow-flies paths from residences to hotels, relative to railroads and rivers. Continuous pan-and-zoom in the map allows exploration of potential travel routes to and between hotels. An overview + detail inset map suggests how geography affects the overall clientele of the hotels.
- An arc diagram,[18] showing past (red) and future (blue) visits by guests visiting on selected dates. Arc thickness indicates the number of guests. Seasons optionally appear in the background in the same way as in the reruns view.
- A drill-down table view, showing individual register entries. Compound brushing highlights cells in the table as a result of selecting hotels, guests, and residences in the three table views.
- A list view, showing scanned images of newspaper articles published around the time of selected dates.
- A table view, showing city directory information (including occupation and address) about guests.
- A table view, showing aggregated information about guest occupations.

In this case, the reruns views shows two visits to the National Hotel by a vaudeville troupe, the J. P. Harris Show, each preceded 8 days prior by their agent, Charles W. Gilder. Although the members of the troupe come from places unknown, the map reveals nearby towns from which other guests visited on the same days. The accessible location of the hotel in Roaring Spring suggests that the agent and troupe may have traveled there on rail lines, although a portion of each trip may have been by river, in coaches, by horse, on foot, or by similar means. The single thick arc in the arc diagram suggests that the troupe visited the hotel only twice; additional exploration might reveal whether members of the troupe visited individually or in small groups on other dates, possibly in relation to other visits by their agent.

All views support additive selection of multiple data items. In the reruns view, selection of individual dates or particular cycles and periods happens by clicking individual cells or rubberbanding around blocks of cells. The map allows selection of residences by clicking, rubberbanding, or lassoing regions. Extensive cross-filtering between views as a function of these selections enables exploration of complex interdependent groupings of people, dates, and places. Bidirectional filtering allows users to show subsets of hotels, guests, and/or home residences (as selected in the corresponding table views) in the reruns view, and conversely to show subsets of dates (as selected in the reruns view) in the hotels, guests, and residences table views. Filtering also occurs between these views and the nested time plots in other views. For instance, filtering between the residences table view and the time plots in the guests table view allows users to show temporal

visitation patterns for all guests, restricted to travel that involves selected residences. Similar filtering in the opposite direction shows temporal patterns of travel from all towns by selected guests only. Filtering on hotels makes it possible to focus analysis on one hotel, or look for travel patterns that involve both.

Using a process of successive selection and filtering, it is possible to ask specific questions and follow detailed chains of evidence. For example, an analyst might explore possible repeat meetings between guests prior to some critical event by asking the question, 'For visitors on a particular date, on what previous dates did two or more of them visit repeatedly?' To start to answer this question, the analyst would select the critical event date in the reruns view, filter the guests table on date, select all guests in the guests table, filter the reruns view on selected guests, select all dates in the reruns view that involve multiple visitors, then look for guests having multiple visits with similar temporal patterns in the guests view. Filtering the residences table and map view on dates and guests during this process would provide increasingly specific information about travel from particular cities, possibly revealing overlapping origins, destinations, or paths for guests under scrutiny.

Moreover, this process can involve unknown or uncertain people, places, and times by selecting similar names, nearby residences (as indicated in the map view), and preceding or following dates. The ability to visualize uncertainty is crucial for dealing with noise in the original data and errors in the transcription process. In the visualization, uncertain guest and place names are indicated with question marks. Uncertain portions of such names are similarly indicated (e.g '?, PA'). Because all views can display all register entries – including uncertain ones – it is possible to explore patterns involving unidentified persons across time and space. The exception is the map view, which does depict uncertain guests from certain places but does not depict any guests (certain or otherwise) from uncertain places.

Many instances of possible errors in the transcription process were discovered using the visualization itself. In particular, the ability to select and filter on multiple guests and residences gave us a means to explore whether similar names in fact refer to the same person or place. A counterintuitive consequence of this approach is that it is actually desirable to avoid correcting errors in data being visualized in order to maintain interactive access to uncertainty. Doing so allows analysts to conceptualize and interact with uncertainty due to deception, historical clerical error, document degradation, modern transcription error, or a combination of these factors.

Visual analysis

Visual exploration has been a key factor in the discovery of several spatial and temporal patterns of social relationships at a time when travel was by horse and buggy locally, or along specific railroad corridors over longer distances. In particular, the visualization has provided significant evidence in support of hypotheses regarding cooperation between traveling merchants, the effects of weather and seasonal climate variations, circuitous routes taken by salesmen, gatherings on holidays, and vanguard behavior, as well as a variety of idiosyncratic travel patterns involving individuals, families, and other groups. Table 1 summarizes some of the various periodic patterns that we have observed in the reruns view during analysis of the two hotel registers.

Weekly circuits

When examining the visitation patterns of the hucksters from Booneville, relative consistency in weekday visits to Rebersburg for each huckster is immediately identifiable (Figure 5A). Filtering the reruns view on the three selected guests reveals alternating visits on Tuesday and Friday. Rapidly switching which of the three is selected reveals that Welshans stays on Tuesday and that Friday visits by R. A. Sheats transition to Friday visits by A. M. Sheats in March 1898. The latter is also suggested by per-guest charts of visits over time in the guests table view.

Prior to the use of the visualization, each register entry for these visitors had to be manually added to a calendar. In this cumbersome process, the individuals' visitation patterns could not be compared in the context of all the other guests. In addition, the original calendar did not have the ability to change the cycle length for exploration of different cyclic temporal patterns.

Weather and climatic effects

Finding the regular pattern of Friday visits by A. M. Sheats prompted us to look for deviations from this routine. By setting the reruns view to a 14-day cycle then scrolling to earlier dates, we were able to determine that there were two periods when his scheduled visits were not on Fridays. It was only after the seasons option was turned on that we noticed that the only times he did not arrive on a Friday was during winter months (Figure 5B). This has led us to believe that the weather may have had something to do with these variations. While the exact reason for these deviations has not yet been determined, the tool has now guided us to examine historical climate data to test this hypothesis. Such records indicate that February 5–14, 1899 included 'the arrival of the greatest Arctic outbreak in United States meteorological records.'[19]

The visualization also reveals how time of year strongly correlates with the overall number of visits, possibly due to seasonal variations in climate that affect travel. In the vertical histogram (Figure 6), total number of visits is highest during the summer and lowest during the winter, with the exception of major holidays.

Biweekly circuits

While exploring the travel patterns of frequent visitors, selecting traveling salesman E. K. Hess of Williamsport led

Table 1 Periodic patterns in the visits to one or both hotels, as observed using the reruns view in coordination with other views

Periodicity	*Visits*	*Guest/Home Occupation*	*Hotel*	*Observations*
Weekly Same day	45	Harry Welshans Booneville Huckster	Rebersburg Hotel	Vertical pattern. Almost all visits on Tuesday.
Weekly Different days	62	S. D. Fessler Hollidaysburg Unknown	National Hotel	Vertical pattern. Most visits Monday, Tuesday or Wednesday.
Biweekly Same day	21	Harry McCloskey Coburn Unknown	Rebersburg Hotel	Vertical pattern. Almost all visits on Monday.
Biweekly Different days	24	R. W. McEldowney Altoona Wholesale Salesman	National Hotel	Vertical pattern. Two weeks apart but no single day prominent.
Biweekly Mixed days	36	John G. Felty Harrisburg Unknown	National Hotel	Vertical patterns. Several cases of two visits the same week (mostly Mondays, Thursdays).
Multiweekly Same day	40	E. K. Hess Williamsport Traveling Salesman	Rebersburg Hotel	Vertical pattern. Most visits on Thursday. Always even number of weeks (up to 8) apart.
Multiweekly Different days	27	Wm. R. Lylte Altoona Salesman	National Hotel	Vertical pattern. Almost always Monday–Wednesday, usually even number of weeks apart.
Bimonthly	48	R. L. Ehrenfeld Altoona Bar tender in 1900 census	National Hotel	Diagonal pattern. Visits during the middle and end/beginning of months.
Monthly Consecutive days	30	Samuel McCurdy Hollidaysburg Unknown	National Hotel	Diagonal pattern. Consecutive days most months.
Monthly	44	Wm. Smeltzer Woodbury Unknown	National Hotel	Diagonal pattern. Most visits occur from 17th to the 22nd each month.
Mixed	44	Thos. Weir Altoona Unknown	National Hotel	Begins with vertical pattern every other Monday. Switches to diagonal pattern mid-month (around the 15th).
Weekday/ Weekend	Various	Various	Both	Vertical patterns. Fewer on weekends than weekdays.
Seasonal	Various	Various	Both	Vertical histograms undulate with seasons. More visits during the summer than winter.
Seasonal variation	78	A. M. Sheats Booneville Huckster	Rebersburg Hotel	Deviations from vertical pattern during winter months.

to the discovery of visitation patterns spanning multiple weeks (Figure 5C). Changing the cycle length of the reruns view to 14 days reveals that Hess' visits strongly correlate with the second Thursday in this cycle, suggesting that he visited every other week. Perhaps more interesting is that there are several cases in which he did not return until the fourth or even sixth week, always maintaining an even number of weeks between visits. This pattern suggests that salesmen from larger cities such as Williamsport may have had larger circuits than those of the local hucksters, thus preventing them from coming every week. The visualization also shows that two other guests, William B. Chamberlain of Milton and Harry McClosky of Williamsport, also adhered to similar schedules, always visiting an even

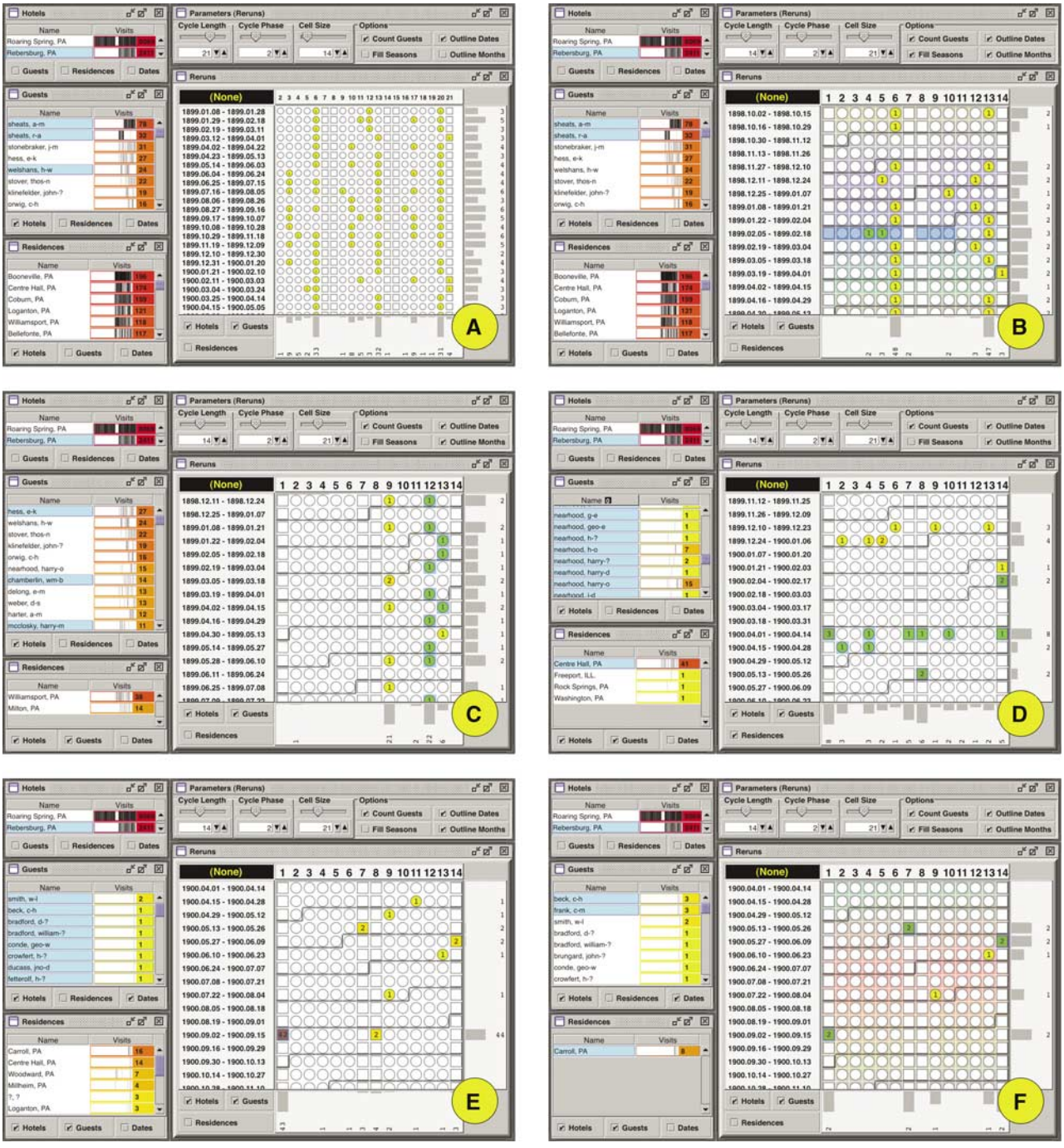

Figure 5 Patterns of visitation. (A) Alternating visits on Tuesdays and Fridays by an organized group of hucksters. (B) Variation in regular visits due to weather. The days highlighted in the reruns view involve recorded extreme winter conditions. (C) Biweekly travel patterns by traveling salesmen. Highlighted dates are visits by Hess. Chamberlain and McClosky both visited on Monday, March 13, 1899. (D) Travel from Centre Hall by possible members of the hotel proprietor's family. Nine of 15 visits by Harry O. Nearhood (highlighted) occurred sporadically over a 3-week period in April 1900. (E) Visits to Rebersburg on Sunday, September 2, 1900 (Labor Day weekend). Non-highlighted dates show other visits by the same guests. (F) Repeat visits by two of the Labor Day weekend guests over summer months.

number of weeks after their previous visit and on the same day of the week. This pattern prompts further research inquiries to determine the profession of these men and possibly identify additional commercial connectivities, especially as all three men always came in the second week of successive 2-week periods.

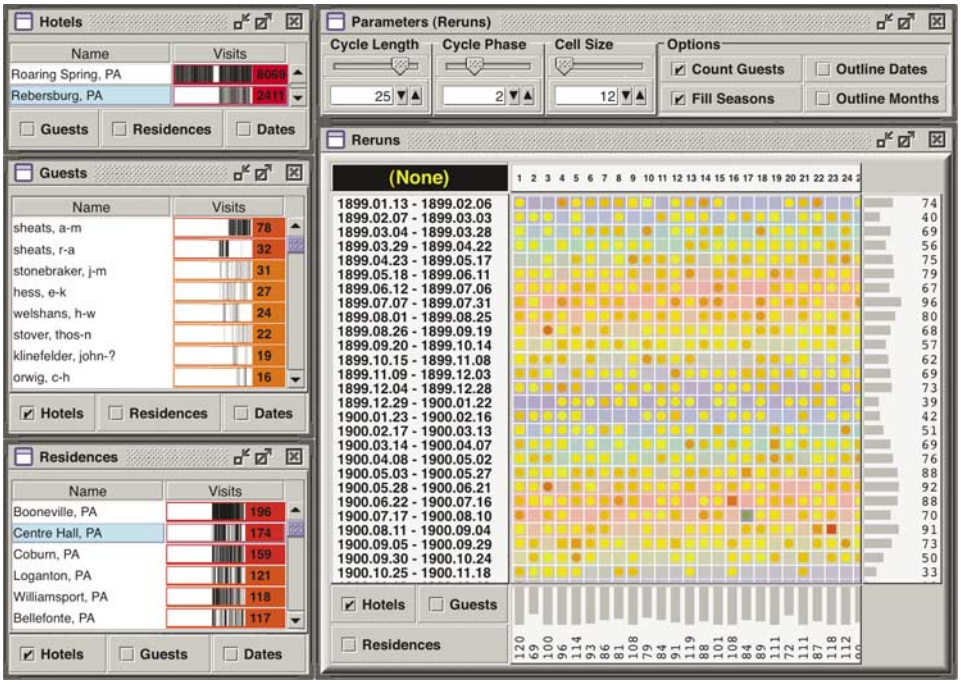

Figure 6 Seasonal variation in the overall number of visits. A pale color gradient from blue to green to red to gold indicates winter, spring, summer, and fall.

Hotel characteristics

In addition to enabling specific discoveries, the visualization reveals general relationships in the data that uphold prior observations and hypotheses about the nature of the hotels. For example, we knew from transcribing that the number of guests at the Rebersburg hotel on weekend days was much lower than during the week, a result reaffirmed in the horizontal histogram when showing all visits over cycle lengths that are multiples of seven.

In another case, several different guests were registered at the Rebersburg hotel who shared the same surname, Nearhood, as the proprietor of the hotel. When querying all guests with this last name (Figure 5D), it appears that, unlike most guests of the hotel, Nearhoods frequent the hotel mainly on weekends. All but three are from Centre Hall, a town 10 miles to the west, serving as another reminder that roads were unimproved and thus places were too far away to allow for a daytime visit without an overnight stay.

Gatherings/groups

Using the reruns view, it is possible to efficiently determine on which days the hotel had an unusually high number of guests, then query those days to see if there are large groups of people traveling together. For example, it can be seen that on Sunday September 2, 1900, there were 42 guests listed in the Rebersburg register (Figure 5E). Drill-down reveals that the double visit involves two entries under the same name (W. L. Smith) but from different home towns (Carroll and 'Fiedbar', an illegible place name). Selecting all dates for visits by the 42 guests reveals that 32 of them were one-time visitors.

Additional exploration reveals that two repeat visitors, C. M. Frank and C. H. Beck, both from Carroll, may have met twice at the beginning of the summer, perhaps to plan the fall event. Identifying this pattern of repeat visits involved filtering guests on date (Labor Day weekend Sunday), selecting all guests, filtering dates on selected guests, selecting the other two dates with more than one visitor, then noting that Frank and Beck had three visits total each (Figure 5F).

These results prompt further research inquiries as to whether there was Labor Day activity, perhaps with a dinner at the hotel, and a cascading set of questions: What other holidays attracted visitors? Were there other times when different people with the same surname congregated, perhaps for a family reunion? Do busy days involve weddings, baptisms, or funerals? Showing multiple calendar datasets in parallel might involve visual encoding of even more data dimensions in the existing reruns view, or addition of multiple reruns views that coordinate with each other using synchronized scrolling and brushing.

Genealogical data-construction activity could be supported in the visualization by incorporating datasets for newspaper articles, county birth, marriage and death registers, and local church or cemetery records. The images of nearby *Millheim Journal* newspaper articles could be augmented with data-driven highlighting of keywords and

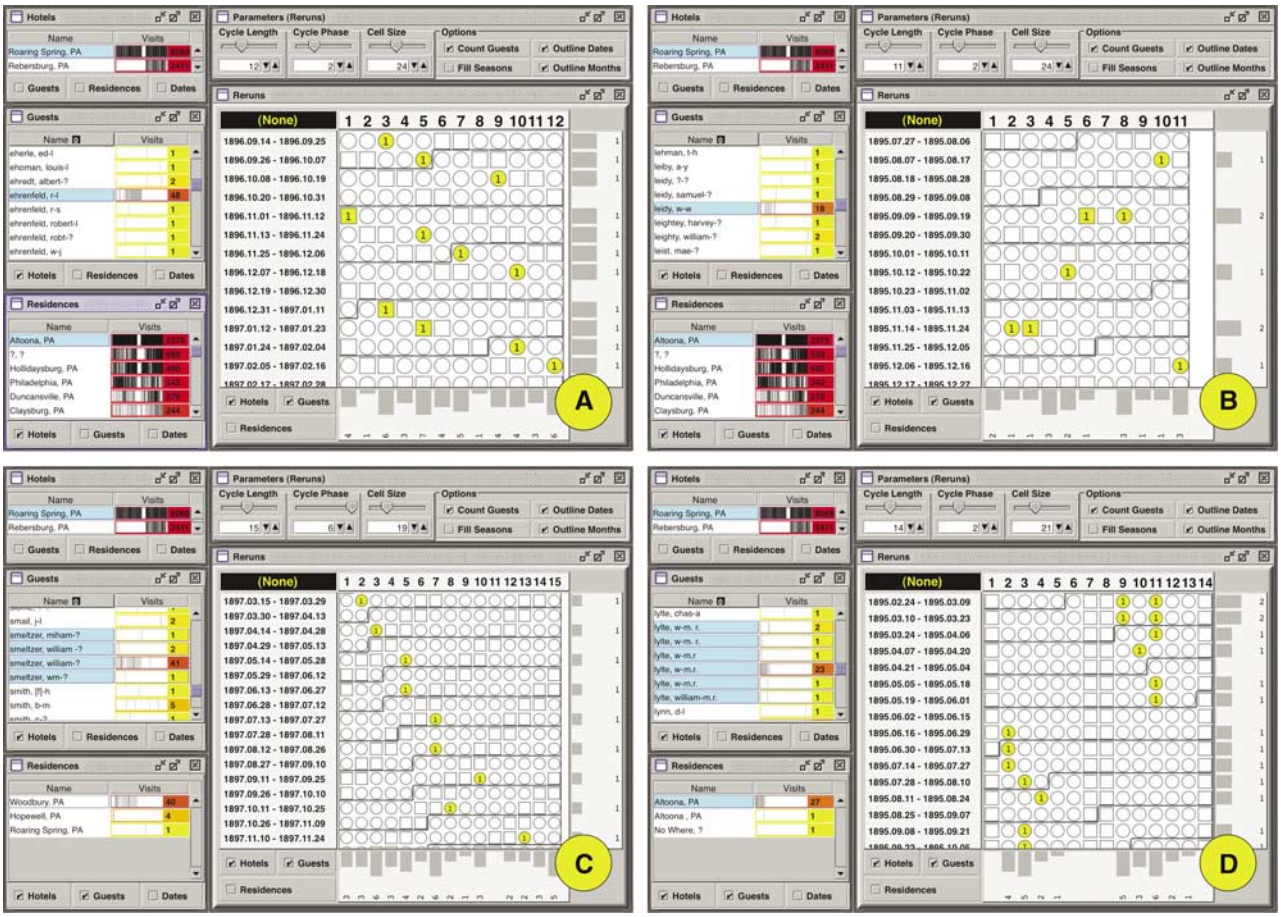

Figure 7 Extended patterns of visitation. (A) Bimonthly visits by R. L. Ehrenfeld on different days of the week. (B) Single and multiple day mid-monthly visits by W. W. Leidy. (C) Visits by Wm. Smeltzer on the 17th–22nd of each month, appearing as a drifting vertical pattern. (D) Biweekly visits on Wednesday through Saturday by Wm. R. Lylte, with a mysterious missing week in mid-June 1895.

phrases involving dates, locations, and persons. Temporally aware text analysis algorithms and views like those in CiteSpace II[20] could be used to extract word and phrase datasets for highlighting, as well as for coordination and cross-filtering with the other views in the visualization.

Non-periodic regular visits

While the reruns window is designed to explore periodic cycles of events, such as those in the previous patterns that involve one or more weeks, we have also found that it is possible to detect non-periodic cycles such as monthly or twice-monthly intervals. While exploring several individuals from the National Hotel register using various cycle lengths, diagonal patterns began to emerge that were not related to particular days of the week. For example, R. L. Ehrenfeld visits the hotel on 48 different occasions but there is no one day of the week that stands out from others (Figure 7A). By filtering by guests then perusing the visits drill-down table, it becomes clear that his visits are approximately twice per month, alternating from the middle of the month (from the 16th to the 19th) to the end/beginning of the month (from the 27th to the 2nd). Similarly, W. W. Leidy from Duncansville visits every month anywhere from the 14th to the 17th (Figure 7B), Wm. Smeltzer from Woodbury visits about two-thirds into each month (Figure 7C), and Wm. R. Lylte of Altoona skips a week in his normal biweekly visits (Figure 7D).

Vaudeville agents

During the transcription process, it was observed that for several of the traveling vaudeville shows there was an agent who came to the hotel a week prior to the actual show, presumably as an advance-man to make all of the local arrangements. What is interesting is that one of the shows, The J. P. Harris Show, stays at each of the hotels within a 2-week period. Indeed, the agent, Charles W. Gilder, arrived at the Rebersburg Hotel on Saturday, June 30, 1900 along with two other agents. The rest of the 19-person troupe followed 1 week later when they performed

'Uncle Tom's Cabin' on the 7th of July. Two weeks later the same scene played out at the National Hotel, this time with a single agent, Charles W. Gilder, arriving on Friday July 13th followed by a smaller troupe consisting of eight members of the same cast as Rebersburg who performed an unidentified show (it is not named in the hotel register).

In addition to providing clues about circuits and travel times during this period – Rebersburg and Roaring Spring are approximately 80 miles apart – this finding raises questions as to where the group performed between their visits. Did the company only perform on weekends? Where did they stay during the week? Why did the cast of 19 dwindle to eight? A newspaper advertisement for Millheim indicates that they performed there, just five miles south of Rebersburg, at some point during the prior week.

Connection to other sources

Using other sources to add information about hotel guests has illuminated commercial and social patterns at the individual level. The process of looking up a person's occupation using the census is tedious and usually involves several steps, even if their hometown is known. The census is also only conducted every 10 years, so finding an individual at a specific date between enumerations is not possible. City directories are also useful for gaining additional information about individuals. These directories are sometimes updated annually but often do not extend into rural regions or small towns.

Rather than finding every individual from the hotel registers in the census or directories, a truly daunting task, we have looked up several individuals for which interesting temporal patterns exist. It has been noted that 'hucksters' have weekly circuits while 'commercial travelers' or 'traveling salesmen' often came at larger intervals. Using this hypothesis, we looked up individual guests of the National Hotel having specific visitation patterns. Amos Nowlen, a weekly visitor at the National Hotel on 31 different occasions, continued to support this hypothesis when the census indicated that he was a huckster from Altoona. Similarly, R. W. McEldowney, a guest who visited the hotel 24 times roughly every 2 weeks, was found to be a wholesale salesman. (Further research indicates that he may have been involved in the soda industry.)

Since a large portion of the visitors to the National Hotel were from Altoona, a larger commercial center 17 miles north of Roaring Spring, a city directory for that place was examined to add occupational information about the hotel guests. Although the city directory was from 1893, 119 guests were listed with a wide variety of occupations that ranged from the expected salesmen, hucksters, clerks, and wholesalers, to teachers, butchers, tinners, jewelers and railroad employees. We are currently transcribing the rest of these entries for inclusion in the aggregated and drill-down city directory tables in the visualization.

Future queries

Research in historical geography is much like intelligence analysis in the sense that it involves careful detective work to coax answers to highly specific questions from sparse and often extremely noisy data. For instance, women appear very infrequently throughout the register. When they do appear, most often they are listed as coming with men under the terms 'and wife' or 'Mr. and Mrs.' In some cases, 'and lady' appears next to a guest's name – though we never found 'Mr. and Mrs. Smith'! Only 11 women registered by themselves under the titles 'Mrs.' or 'Miss'. This prompts further research inquiries concerning where these women came from, what means and patterns of travel were involved, and whether associated males registered by themselves on other occasions. Examination of newspaper columns indicates that women did indeed travel, but in most cases were staying with friends or family.

Some guests note when they sign the register that they are traveling with one or two horses (housed in the livery stable behind the hotel). This prompts further inquiries as to who these people were and where they came from. For guests not listing a horse, how did they arrive at the hotel? One means of transportation may have been the stagecoach – the block for which still sits in front of the building today – but this does not account for the lack of a horse annotation by our local hucksters (who by definition would have the goods they were selling with them, and hence likely a wagon). Does two horses signify a freight wagon, whereas one horse implies a buggy? A search of census records for all names of people traveling with a horse might reveal a specific occupational sector. The literature on hucksters suggests that they often had a weekly circuit, only returning to their 'home' place on weekends to refill their wagon at their store. In such cases, the obvious question is: What hotels were frequented the night prior and the night following their stay at these hotels? In other registers being analyzed, notations sometimes give instructions for forwarding mail, and invariably it is to a named hotel in another town.

Evaluation

We applied the e-Delphi toolkit developed at the GeoVISTA Center[21] to conduct a series of formative evaluations of the visualization. These evaluations targeted two primary goals: to solicit ideas for improvements and enhancements, and to evaluate the visualization toolkit in terms of how well it satisfies a set of precepts for the design and evaluation of visualization tools.

Twelve Penn State graduate students participated in the evaluation. All were enrolled in a graduate seminar on geovisual analytics. All but two were enrolled for graduate studies in Geography that focus on emerging issues in GIScience. The other two students came from related domains in our school of Information Science & Technology. Members of the user group possess varied knowledge of GIS systems and from their participation in the

graduate seminar they are familiar with research priorities for the advancement of visual analytics.

Distributed usability assessment

E-Delphi is a web-based toolkit designed to support Delphi exercises. The RAND Corporation devised the Delphi method[22] shortly after World War II as a way to structure group decisions in a manner that ensures personality conflicts and power relationships do not taint the presentation of ideas. E-Delphi has been used in the past in consensus building exercises with climate change scientists. We adapted the e-Delphi toolkit as a means to support group input to software evaluations. Evaluation sessions take place in an online collaboratory in which moderators can quickly initiate and lead multiple rounds of discussion and other activities. Round-based activities include surveys, metrics, free responses, voting, threaded discussions, or a mixture of these methods. It is important to note that while we used a tool designed initially to support Delphi exercises, we did not apply the Delphi Method *per se*; we were not focused specifically on building a consensus among our users, developing a forecast, or setting policy (typical uses of the Delphi method). Instead, we adapted the flexible e-Delphi toolkit to support distributed, asynchronous, group input to a user-centered design process.

Software evaluation activities often require significant investment of human and financial capital to develop and organize activities and procure technology to capture data. The e-Delphi toolkit allows us to conduct qualitative evaluations via the web with little investment in time and effort by either researchers or participants.

Process

We conducted three rounds of evaluation using the e-Delphi toolkit. In Round 1, we asked users to imagine extensions to the visualization tool. In Round 2, we had users indicate their opinions on the fit of the tool to a set of design precepts. Round 3 had users discuss the responses they provided in Round 2.

Round 1 We asked users to complete two short tasks to become familiar with the capabilities of the visualization. The tasks walked users through exploring the patterns of hucksters we had found in our collaborative work with historical geographers (Fyfe and Holdsworth). We then asked them to respond to four short-answer questions in e-Delphi:

(1) Briefly describe a scenario in which a historian might use a visualization tool like the one you have tested.
(2) What other types of data would be interesting to examine with the tools you used today? Why?
(3) Assume your task is to design one additional visualization tool to work with the set you have used today. Describe what you would add next and why.
(4) Imagine you work at a newspaper and you have been given this software to develop story ideas for future articles. What tools and interactions would you need to do your job?

These questions were designed to provoke suggestions for situations in which the visualization might be useful, additional kinds of data that might be required to tell a story using the visualization, and extensions that could be incorporated into the visualization to enhance exploration and analysis. We asked users to avoid critiquing basic appearance and behavior aspects of the visualization interface in favor of focusing attention on higher-level ideas and goals for future prototypes. Users had 90 min (as part of seminar class time) to complete the sample tasks and answer these prompts. We allowed users to return to their responses for three days afterward in case they wished to revise their answers.

Round 2 Amar and Stasko[23] identify six precepts that can be used to evaluate the design of information visualizations in terms of how well they bridge *Worldview* and *Rationale* gaps:

- Provide facilities for creating, acquiring, and transferring knowledge or metadata about important domain parameters within a dataset (Worldview).
- Support discovery (automated or manual) of useful correlative models and constraints (Worldview).
- Support the interactive formulation and verification of user hypotheses (Worldview).
- Expose uncertainty in data measures and aggregations and show possible effects of uncertainty on outcomes (Rationale).
- Clearly present what comprises the representation of a relationship and present concrete outcomes when appropriate (Rationale).
- Clarify possible sources of causation (Rationale).

In Round 2 we asked participants to assess the visualization in terms of these precepts. We chose to focus our attention on evaluating the toolkit based on these particular design goals because they appeared to be especially relevant to the goals of visual analytics. We asked users to indicate their level of agreement that the visualization satisfied these precepts on a five-point Likert scale. We also requested a short text response to provide justification for their choices. The instructions suggested that they re-examine the toolkit and the introductory tasks we had provided in the first round as necessary to refresh their memory. The activity took 90 min and we again provided several days for revision.

Round 3 In Round 3, we asked users to examine a major theme that we noticed in the results from Round 2: the Worldview precepts were more satisfied by the visualization than the Rationale precepts. The instructions were to discuss this dichotomy in terms of which things should

be implemented next to introduce a more appropriate balance to the design. Users had another 90-min session during class time and a few days afterward for revision. We extended response time an additional week to compensate for spring break.

Results

Results from our usability activities were collected via the moderator tools in the e-Delphi interface following the conclusion of each round. Discussion and short answer results are automatically sorted by participant and provided in a small window where they can be copied into a word processing application. Likert scale answers are automatically summarized to show the number of votes for each answer and a visual indication is provided to show the mean and mode of all the answers. We used the results of the first round to help shape the composition of the second round, and the results of the second to shape the third. In the following sections, we discuss the results from each round, and where appropriate we include quotes from our participants.

Round 1 Round 1 provided valuable feedback on potential extensions to the visualization, additional data sources, and situations in which users felt it might be particularly useful.

One user proposed the addition of a geographically based filtering capability, to allow the use of the map as the primary interface for selecting guest visits with a bounding box over cities of interest: 'I would add interaction capabilities to the map view that allowed the selection of certain origin cities to filter the date and names from the ledger. Similarly, it would be a help to select a group of guest names and have the map expand to bound the extent for the originating cities of that set of guests'.

Another user suggested a tool that would help find groups of people who traveled together on the same day, perhaps from different places and for different lengths of stay. As a complement to this, several users mentioned a desire to drill down to greater detail for each guest, to view a profile for each person. This profile would contain personal information such as their workplace, address, birthday, and other details.

Other suggestions for additional data sources include: transportation schedules, letters and other personal communication, bank transactions, hospital records, and the ability to integrate the results of Google searches for dates and names. It is worth noting that we observed users opening web-based searches alongside the visualization to see if there were interesting results based on the dates and names of people they were examining.

Multiple users requested sophisticated capture and annotation tools as part of the visualization so that they could develop stories about what they were seeing: 'I need a window to jot down notes as I go. I need to be able to save views, i.e, to be able to save the current state of all my windows'.

Suggestions for situations in which such tools might be useful included: tracking the impact of a major event (such as a flood or hurricane), analyzing the travel behavior of a specific individual or group of people, and characterizing hotels in terms of the types of people they typically attract. One user said: 'I could imagine historians or other social scientists using this tool to look at any phenomenon that has temporal data (with the reruns window) and if there is a spatial component the map window would also allow the researcher to look for spatial patterns'. Another suggested that it could be used to answer questions, 'To investigate the origins and spatial patterns of visits. Where do visitors come from? Do these patterns change over time? Is this hotel a choice for visitors from far away?'

Round 2 The agreement between responses we received in Round 2 indicated that our users generally agreed that the visualization works to bridge the three Worldview gaps to create knowledge, support discovery of correlative models, and support the interactive formulation of hypotheses. On the other hand, users generally disagreed that it satisfies the three Rationale gap precepts to expose uncertainty, present concrete outcomes, and clarify causation.

The biggest weakness that users noted was the visualization's inability to expose uncertainty in data and indicate its effect on outcomes, thus failing to satisfy the Rationale Gap precept. This was the only precept of the six that received all negative or neutral responses. Users pointed out that there was no explicit visual indication of uncertainty in the data they were viewing: 'Missing data seem to be indicated with '?', but users would probably love to have some quantitative measures of missing and unclear values'. A few noted that the only way for uncertainty to enter the picture in the current design was if the user could infer it from prior intimate knowledge of the data: 'This could be an issue of training and understanding more where the data comes from, but for the first time, it is hard to understand'.

Round 3 Round 3 suffered a bit in terms of participation due to the fact that spring break started shortly after we began the activity. Despite this, Round 3 provided further clarification of the survey responses we received in Round 2.

To better expose uncertainty and show its effects on data, users suggested the creation of a visual summary for the guest records that were incomplete or otherwise uncertain: 'There is no apparent indication that errors may exist in the datasets, so in theory, someone could take this at face value. There is no view indicating 'no data' or 'uncertain' or something to that effect'. Because the register data is coded with question marks where names or places were uncertain, it will be relatively easy for us to implement a summary table that either contains all such records or tags each record with a more obvious indication of uncertainty that can be visually encoded in the visualization.

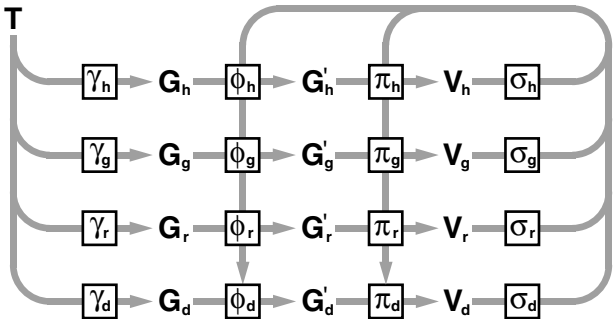

Figure 8 Filtering between views in the visualization.

In order to better satisfy the need to present what comprises a relationship and to suggest concrete outcomes, users suggested we work first on clarifying the tasks that might be accomplished using the visualization. This request stemmed from the fact that some users were unclear about the relationships they should be looking for, so they could not confidently define better ways to suggest outcomes: 'I keep trying to do simple things like figure out who came from where – which is kind of hard to figure out from this toolkit – but I don't think that would be a desired end task for someone visualizing this data'.

To clarify causation, it was suggested that the visualization should reveal bits of relevant metadata for each guest, hotel, or date on demand: 'Having a mouse over functionality in the Reruns view which shows names of people who visited a hotel on a given date would be useful'. Users also requested ancillary information like holidays, special events, and weather to appear in the reruns view. In the map, the number of guests and their names could appear as a popup when mousing over residences and paths of travel. The guests view could provide an option to further visually highlight individuals or groups using avatars to better break down patterns viewed in the reruns and map views.

Generalization and extension

We designed the hotels visualization using a 'group, select, and filter' approach in which multiple views each display the set of all unique values that occur for a particular raw or derived data attribute. Analysts select arbitrary subsets of these values in each view. Selections affect the filtering and visual appearance of all items across all views, allowing analysts to visually interrogate the data about who, what, where, and when by quickly drilling down into specific subsets of high-dimensional space–time information.

In the specific case of the hotels visualization (Figure 8), register entries in the dataset of hotel visits (T) are grouped (γ) by hotels (h), guests (g), residences (r), and dates (d). Each group is filtered (ϕ) and projected/visually encoded (π) in order to populate the hotels table (V_h), guests table (V_γ), residences table (V_r), and cyclic view of dates (V_d). Arbitrary selection (σ) of items in these views optionally affects the filtering and projection of items in other views, thus allowing the analyst to drill-down by selecting arbitrary subsets of people, places, and times and by toggling cross-filtering between pairs of views.

In general, this 'group, select, and filter' design approach incorporates query and display aspects of visualizations created in IVEE,[24] which automatically matches filtering sliders to data attributes, and Polaris,[25] which performs hierarchical categorization queries to generate multiscale displays. A key advantage of Improvise over these systems is the ability to calculate derived attributes and parameterize querying (including grouping/categorization), filtering, and visual encoding on these attributes in terms of interaction in any view or slider. The corresponding disadvantage is the challenge of capturing this flexibility in a fast, easy-to-use builder interface – like the attribute-drag-and-drop interface in Polaris/Tableau – that enables true exploratory visualization.

We are currently working to extend the Improvise builder interface to allow domain experts to drive rapid, semiautomatic construction of their own 'group, select, and filter' visualizations. In such an interface, users would choose raw data attributes or functionally specify derived data attributes of interest, then associate each attribute with an appropriate view type in the same way that visit dates are associated with the reruns view in the hotels visualization. Although initial design would occur in a form-based interface at a higher level of abstraction than the normal Improvise design dialogs, advanced users could then choose to continue working directly in the latter, editing query statements to customize their visualizations as needed for deeper or more specialized analysis.

Conclusion

Exploratory visualization provides a faster and more productive means of analysis than previous methods of examining historic hotel visits. By coordinating the reruns view with map and tabular displays, the hotels visualization allows detailed and contextualized queries regarding cyclic temporal patterns (in days, weeks, months, and years), seasonal temporal patterns with corresponding climate indications, social grouping patterns, spatial patterns of travel from particular places or regions, and complex combinations of all of these. Although the visualization focuses on historical travel patterns in a specific social context, application to current problems of interest in the intelligence analysis and emergency response communities – the behavior, movement, and communication of persons, groups, vehicles, and resources under the influence of terrain, climate, and weather – is an obvious next step. In fact, we anticipate that guest databases for multiple modern hotels (or other nexus points in travel networks) could be incorporated into the visualization in a short time. Future plans involve increasing the usefulness of the visualization by adding historical weather records and railroad schedules in order to analyze actual means and paths of travel by hotel visitors.

Distributed, structured evaluations have provided valuable critiques as well as concrete suggestions for improving and extending the visualization. The e-Delphi toolkit allows us to quickly conduct evaluations that can be repurposed for multiple user groups with little effort on our part. We have found it to be a method for getting useful feedback in situations in which it is difficult to assemble users in a single place at a single time. Due to the need for users to type out their thoughts, however, it is an imperfect substitute for face-to-face group discussion. As the visualization evolves, we will transition our evaluation efforts to techniques such as verbal protocol analysis and formal focus groups composed of professional analysts in order to elicit rich feedback to further refine our tools. Recent informal feedback from analysts reaffirms and expands upon many of the priorities we have identified as a result of the initial evaluation done with graduate students.

By creating innovative ways of examining spatiotemporal data, the hotels visualization aids significantly in exploration of historical hotel register data, providing an opportunity to see travel patterns that had previously gone unrecognized. Data for other hotels in other places, once transcribed, can be quickly incorporated, and spatiotemporal patterns and irregularities can be quickly discerned. As a result, hypothesis generation and testing has accelerated over the traditional techniques utilized in historical geography. Visual analysis has led to multiple discoveries that have triggered new thought processes and shaped new research questions about social connectivity relationships in rural areas at the turn of the 20th century.

Acknowledgements

We thank Bill Pike at the Pacific Northwest National Laboratory (PNNL) for creating e-Delphi and helping us adapt it for our evaluation work. We also thank Scott Pezanowski and Ritesh Agrawal for preparation of the geospatial layers used in the map display. This study was supported and monitored by the Advanced Research and Development Activity (ARDA) and the Department of Defense. The views and conclusions contained in this document are those of the author(s) and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Geospatial-Intelligence Agency or the U.S. Government.

References

1 Thomas J and Cook K (Eds). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press: Chicago, 2005; 186pp.
2 MacEachren A, Gahegan M, Pike W, Brewer I, Cai G, Lengerich E, Hardisty F. Geovisualization for knowledge construction and decision support. *IEEE Computer Graphics and Applications* 2004; **1**: 13–17.
3 Weaver C. Building highly-coordinated visualizations in Improvise. *IEEE Symposium on Information Visualization 2004* (Austin, TX), IEEE Computer Society Press: Chicago, 2004; 159–166.
4 Winfree A. *When Time Breaks Down: The Three-dimensional Dynamics of Electrochemical Waves and Cardiac Arrhythmias*. Princeton University Press: Princeton, NJ, 1987. 339pp.
5 Carlis J, Konstan J. Interactive visualization of serial periodic data. *ACM Symposium on User Interface Software and Technology 1998* (San Francisco, CA), ACM Press: New York, 1998; 29–38.
6 Rebersburg Hotel Guest Register, Rebersburg, PA, 1898–1900. *Historical Collections and Special Collections Library Labor Archives*. The Pennsylvania State University Libraries: University Park, PA.
7 Roaring Spring Hotel Guest Register, Roaring Spring, PA, 1894–1900. *Historical Collections and Special Collections Library Labor Archives*. The Pennsylvania State University Libraries: University Park, PA.
8 Holdsworth D. Historical geography: new ways of imaging and seeing the past. *Progress in Human Geography* 2003; **4**: 486–493.
9 Knowles A (Ed). *Past Time, Past Place: GIS For History*. ESRI Press: Redlands, CA, 2002.
10 Livny M, Ramakrishnan R, Beyer K, Chen G, Donjerkovic D, Lawande S, Myllymaki J, Wenger K. DEVise: integrated querying and visualization of large datasets. *International Conference on Management of Data 1997* (Tucson, AZ), ACM Press: New York, 1997; 301–312.
11 Olston C, Woodruff A, Aiken A, Chu M, Ercegovac V, Lin M, Spalding M, Stonebraker M. DataSplash. *International Conference on Management of Data 1998* (Seattle, WA), ACM Press: New York, 1998; 550–552.
12 North C, Shneiderman B. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. *Working Conference on Advanced Visual Interfaces 2000* (New York, NY), ACM Press: New York, 2000; 128–135.
13 Hibbard W, Dyer C, Paul B. Display of scientific data structures for algorithm visualization. *IEEE Conference on Visualization 1992* (Boston, MA), IEEE Computer Society Press: Chicago, 1992; 139–146.
14 Fekete J. The InfoVis Toolkit. *IEEE Symposium on Information Visualization 2004* (Austin, TX), IEEE Computer Society Press: Chicago, 2004 167–174.
15 Heer J, Card S, Landay J. Prefuse: a toolkit for interactive information visualization. *Conference on Human Factors in Computing Systems 2005* (Portland, OR), ACM Press: New York, 2005; 421–430.
16 Takatsuka M, Gahegan M. GeoVISTA *Studio*: a codeless visual programming environment for geoscientific data analysis and visualization. *Computational Geoscience* 2002; **10**: 1131–1144.
17 Weaver C. Visualizing coordination in situ. *IEEE Symposium on Information Visualization 2005* (Minneapolis, MN), IEEE Computer Society Press: Chicago, 2005; 165–172.
18 Wattenberg M. Arc diagrams: visualizing structure in strings. *IEEE Symposium on Information Visualization 2002* (Boston, MA), IEEE Computer Society Press: Chicago, 2002; 110–116.
19 Gelber B. *Pocono Weather*. Uriel Publishing: Stroudsburg, PA, 1992. 291pp.
20 Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 2006; **3**: 359–377.
21 Pike B, Yarnal B, MacEachren A, Gahegan M, Yu C. Infrastructure for human–environment collaboration: building a prototype for the future of science. *Environment* 2005; **2**: 8–21.
22 Dalkey N. The Delphi method: an experimental study on group opinion. Technical report RM-5888-PR. The Rand Corporation; 1969.
23 Amar R, Stasko J. Knowledge precepts for the design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics* 2005; **4**: 432–442.
24 Ahlberg C, Wistrand E. IVEE: an information visualization & exploration environment. *IEEE Symposium on Information Visualization 1995* (Atlanta, GA), IEEE Computer Society Press: Chicago, 1995; 66–73.
25 Stolte C, Tang D, Hanrahan P. Polaris: a system for query, analysis, and visualization of multi-dimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 2002; **1**: 52–65.