Visual Analysis of Historic Hotel Visitation Patterns

Chris Weaver* David Fyfe† Anthony Robinson‡ Deryck Holdsworth§ Donna Peuquet¶
Alan M. MacEachren‖

The GeoVISTA Center and Department of Geography
The Pennsylvania State University

ABSTRACT

Understanding the space and time characteristics of human interaction in complex social networks is a critical component of visual tools for intelligence analysis, consumer behavior analysis, and human geography. Visual identification and comparison of patterns of recurring events is an essential feature of such tools. In this paper, we describe a tool for exploring hotel visitation patterns in and around Rebersburg, Pennsylvania from 1898-1900. The tool uses a wrapping spreadsheet technique, called reruns, to display cyclic patterns of geographic events in multiple overlapping natural and artificial calendars. Implemented as an Improvise visualization, the tool is in active development through a iterative process of data collection, hypothesis, design, discovery, and evaluation in close collaboration with historical geographers. Several discoveries have inspired ongoing data collection and plans to expand exploration to include historic weather records and railroad schedules. Distributed online evaluations of usability and usefulness have resulted in numerous feature and design recommendations.

Keywords: Geovisualization, exploratory visualization, historical geography, coordinated multiple views, travel pattern analysis.

Index Terms: D.2.2 [Software Engineering]: Design Tools and Techniques—User Interfaces; H.5.2 [Information Systems]: Information Interfaces and Presentation—User Interfaces

1 INTRODUCTION

As a part of the research and development agenda for visual analytics [15], the development of integrated software for rapid construction of visual data analysis tools is a recent focus of research in both information and geographic visualization [11]. The 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 to use. In pursuit of this goal, Improvise [17] is an exploratory visualization software application that is written entirely in Java and is freely available on the web under an open source license. In Improvise, analysts rapidly build and browse multiple coordinated views of their data. What makes Improvise special is that it provides precise control over how interaction affects the presentation of space, time, and abstract data attributes 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 geovisual exploration and analysis.

*e-mail:cweaver@psu.edu
†e-mail:dfyfe@psu.edu
‡e-mail:arobinson@psu.edu
§e-mail:dwh6@ems.psu.edu
¶e-mail:peuquet@psu.edu
‖e-mail:maceachren@psu.edu

IEEE Symposium on Visual Analytics Science and Technology 2006
October 31 - November 2, Baltimore, MD, USA
1-4244-0591-2/06/$20.00 ©2006 IEEE

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 archival documents. Most registers contain a date, name, and hometown 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 a spouse, or whether they stabled a horse in the livery.

Starting with a data set transcribed from the guest register for the Rebersburg Hotel in Rebersburg, PA from 1898-1900, 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 set 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 guests and their residences, and an arc diagram showing sequences of visits by guests individually and in groups.

The key 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 cyclic temporal patterns across multiple natural and artificial calendars. Using glyphs rather than pixels (as in visualization of infant sleep cycles [18]) increases the number of timescales that can be encoded in a single view. Interactive scrubbing over time (by panning and zooming rows and columns of cells) offsets the corresponding decrease in the number of data items that can appear inside the view at any given time. Interactive adjustment of period length enables visual identification of periodic behavior of varying duration, much the same as in spiral-based visualizations of periodic data [2].

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 prompt a large increase in data collection and a desire to expand exploration to include weather and transportation information. Evaluation of the tool suggests strong support for continued efforts along these avenues as well.

2 DATA COLLECTION AND TRANSCRIPTION

The register for the Rebersburg Hotel [3] 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. With nearly a hundred guests each month signing the register from this

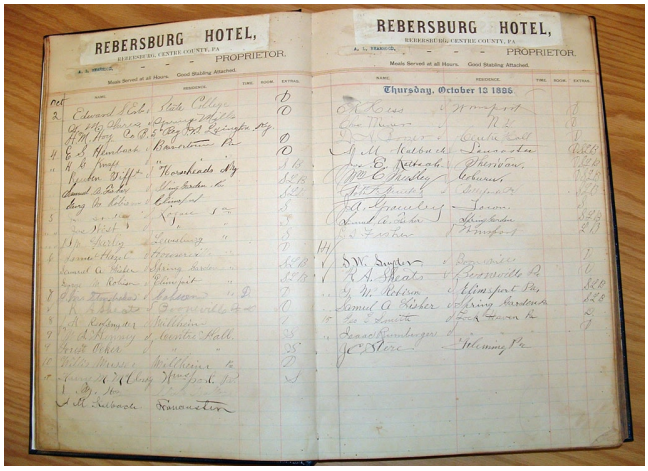

Figure 1: The Rebersburg Hotel Register.

wide variety of places, thousands of data points represent patterns of local and regional connectivity. However, when analyzing the actual archival document (figure 1), 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 set was transcribed into a spreadsheet 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 spreadsheet took approximately 80 hours. Once the data set was transcribed, the columns for guests and residences were sorted to check for errors and/or inconsistencies within the data set. 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 data set 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 visual representation of which visitors and residences appeared most often. Many guests appeared several times throughout the two-and-a-half-year period. Forty people registered at the hotel on at least four occasions; one came 78 times. Further research using other archival resources revealed that some of these repeat visitors were draymen (heavy cart drivers), hucksters (local peddlers), and traveling salesmen. Three of the hucksters, 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 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 the hotel data set with an existing spatial data set. The home location and number of visits were then mapped using commercial GIS software, revealing distance-decay gradients. More than 1800 visits 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.

The spatial and temporal patterns found by these methods are limited to aggregated representations of the data set. The texture of individual dates and guests was lost. Moreover, these search methods are 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 [8, 9].)

3 Exploratory Visualization in Improvise

Improvise is a self-contained, web-capable visualization builder and browser that shares many capabilities with other coordinated multiple view visualization systems [10, 13, 12] and toolkits [5, 7]. Users browse visualizations using mouse and keyboard interactions to navigate multidimensional space and select data items across multiple coordinated views.

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 data sets across multiple views. Query expressions specify how to map data attributes into graphical attributes in views. Multiple data sets 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 data sets interactively.

Improvise consists of a graphic user interface on top of a modular library of visualization components (figure 2). In the user interface, designers create, layout, parameterize, and coordinate views. Building occurs inside the same top-level window that contains views. Building is fully live, so that all design changes take effect immediately without the need for a separate compilation stage. This live, amodal interface design allows trained users to switch rapidly between building and browsing. 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, to support *improvisational visualization* on a timescale of hours or days.

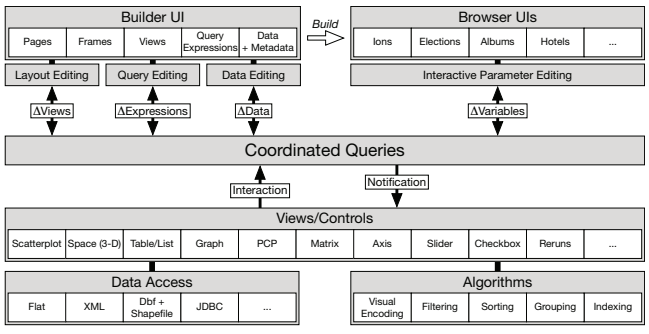

Figure 2: The Improvise software architecture.

Improvise has been used to build a wide variety of information visualizations, and has recently been extended for geovisualization of census, election, and ham radio communications involving spatial, temporal, and abstract data attributes. The ongoing evolution of the hotel register visualization is one of the major successes of this

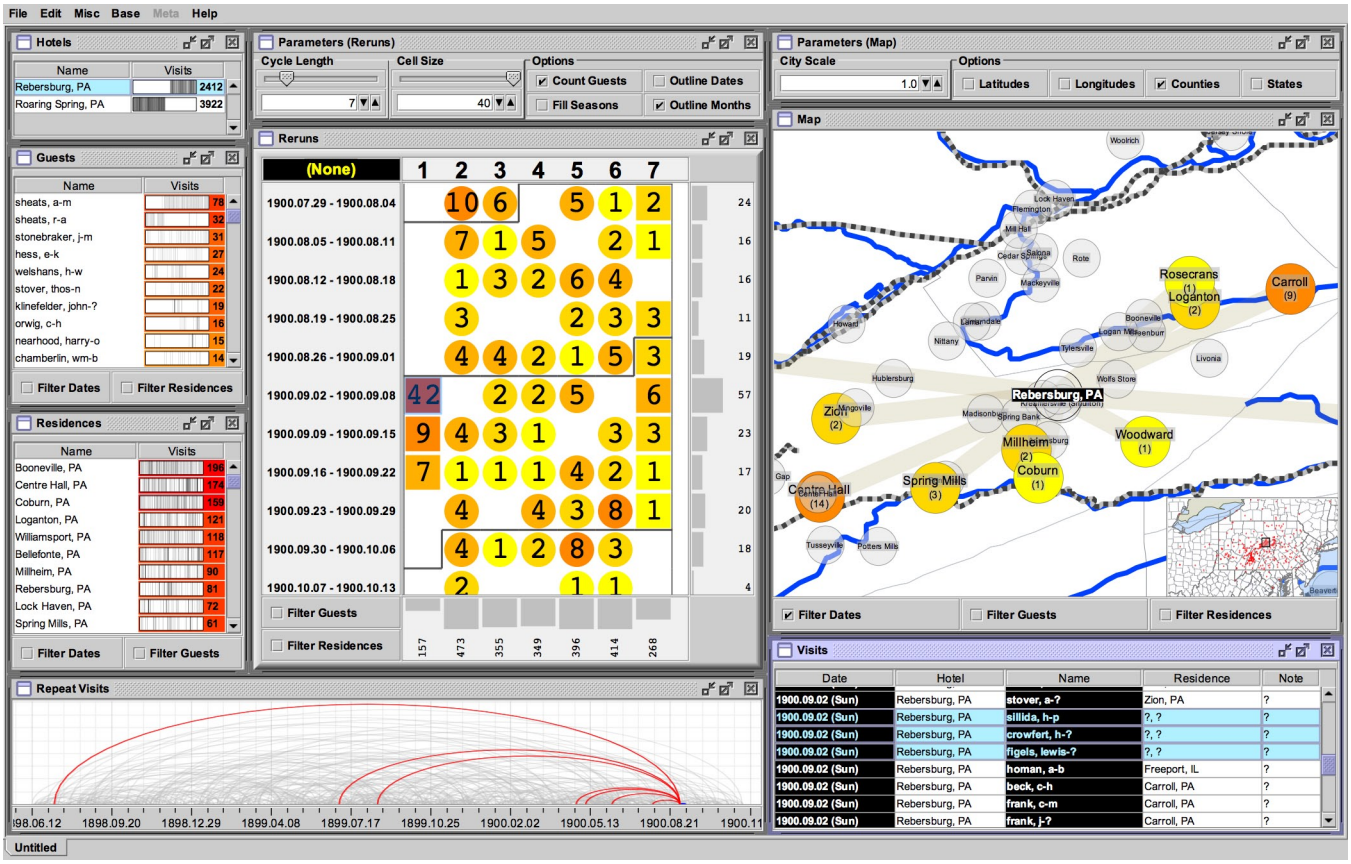

Figure 3: The hotels visualization. The map shows travel on Sunday, September 2, 1900 (Labor Day weekend) by 42 visitors to the Rebersburg Hotel, including three who came from places unknown. Although many guests may have come from places on rail lines or rivers, the location of the hotel in Rebersburg suggests that a portion of each trip may have been in coaches, by horse, on foot, or by similar means.

effort. (Because it is just one of many Improvise "documents"—all saved as regular, self-contained XML files—we have not adopted a formal name for the visualization as an independent tool. In fact, we refer to it internally simply as "the hotels vis.")

4 Hotels Visualization Interface

Using Improvise, we have developed a visualization of the guest register for a hotel in Rebersburg, PA from 1898-1900. 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, MacEachren] and historical geographers [Fyfe, Holdsworth]. This process has resulted in four major versions and dozens of minor versions of the visualization over the course of six months. Evolution of the visualization has been driven primarily by the exploratory and analytic needs of the historical geographers, who have adopted the visualization as a important tool in their research arsenal.

The visualization (figure 3) consists of the following views:

- A table view, showing the names, total visits, and pattern of visits over time for each guest.
- A table view, showing the same information for each residence (place of origin).
- 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 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.
- 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 the hotel, relative to railroads and rivers.
- An arc diagram [16], showing past (red) and future (blue) visits by guests visiting on selected dates.
- A drill-down table view, showing individual register entries.

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. Improvements to querying and rendering performance have resulted in navigation and selection coordination that is much faster than in earlier versions, enabling smooth interaction throughout the visualization.

Extensive cross-filtering between views enables exploration of complex interdependent groupings of people, dates, and places. Bidirectional filtering allows users to show subsets of 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 guests and residences table views. Filtering between the residences table view and the nested bar 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.

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), and preceding or following dates. The ability to visualize uncertainty is critical for dealing with noise in the original data and errors in the transcription process. In the visualization, uncertain guest (44%) and place (14%) 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.

5 Visual Analysis

Visual exploration has been a key factor in the discovery of several spatiotemporal 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 travelling merchants, the effects of weather and seasonal climate variations, circuitous routes taken by salesmen, and gatherings on holidays, as well as a variety of idiosyncratic travel patterns involving individuals, families, and other groups.

5.1 Organized Groups

When examining the visitation patterns of the hucksters from Booneville, relative consistency in weekday visits for each huckster is immediately identifiable (figure 4). 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.

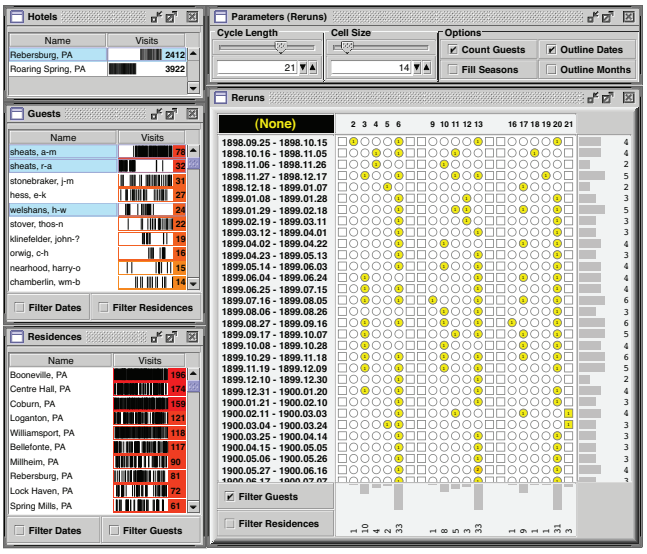

Figure 4: Alternating visits on Tuesdays and Fridays by an organized group of hucksters.

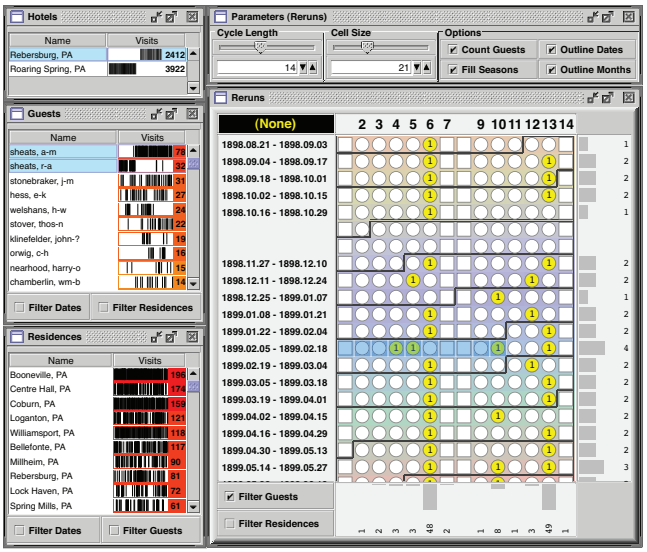

Figure 5: Variation in regular visits due to weather. The days highlighted in the reruns view involve recorded extreme winter conditions.

5.2 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 5). 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." [6]

The visualization also reveals how time of year strongly correlates with the overall number of visits, possibly due to seasonal

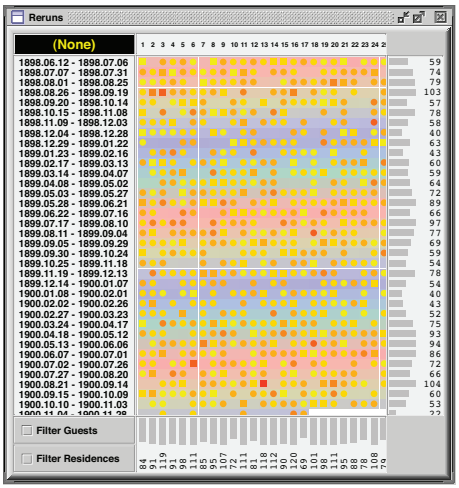

Figure 6: Seasonal variation in the overall number of visits. A pale color gradient from blue to green to red to gold indicates winter, spring, summer, and fall.

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.

5.3 Biweekly Circuits

While exploring the travel patterns of frequent visitors, selecting traveling salesman E. K. Hess of Williamsport led to the discovery of visitation patterns spanning multiple weeks (figure 7). 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 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 two week periods.

5.4 Gatherings

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 register (figure 8). Drill-down reveals that the double visit involves two entries under the same name (W. L. Smith) but from different hometowns (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, 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 9).

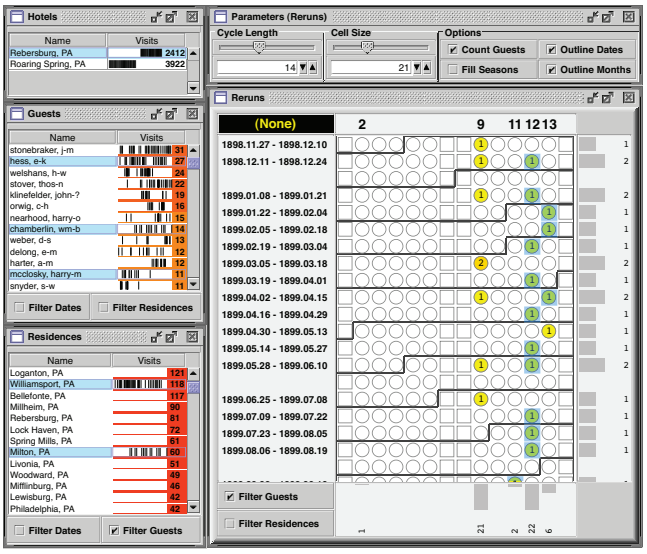

Figure 7: Biweekly travel patterns by traveling salesmen. Highlighted dates are visits by Hess. Chamberlain and McClosky both visited on Monday, March 13, 1899.

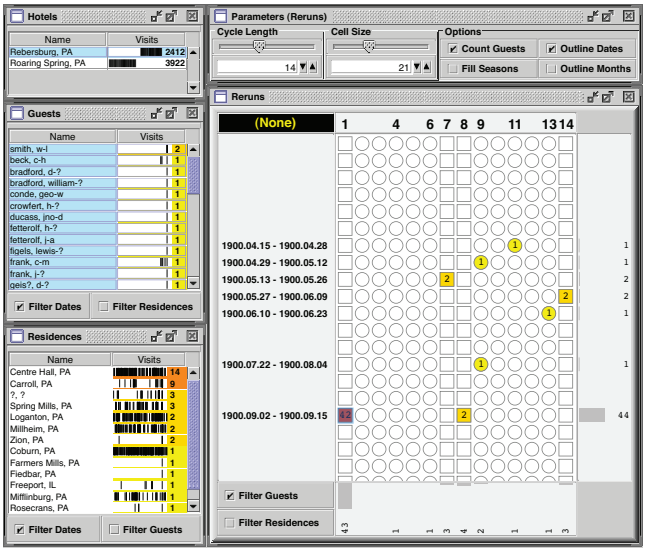

Figure 8: Visits on Sunday, September 2, 1900 (Labor Day weekend). Non-highlighted dates show other visits by the same guests.

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? Genealogical data-construction activity could be supported in the visualization by incorporating data sets for county birth, marriage and death registers, and local church or cemetery records. Showing multiple calendar data sets 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.

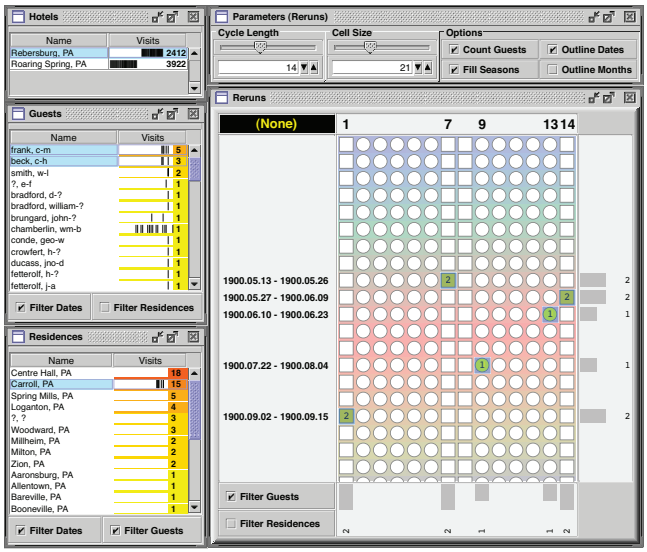

Figure 9: Repeat visits by two of the Labor Day weekend guests over summer months.

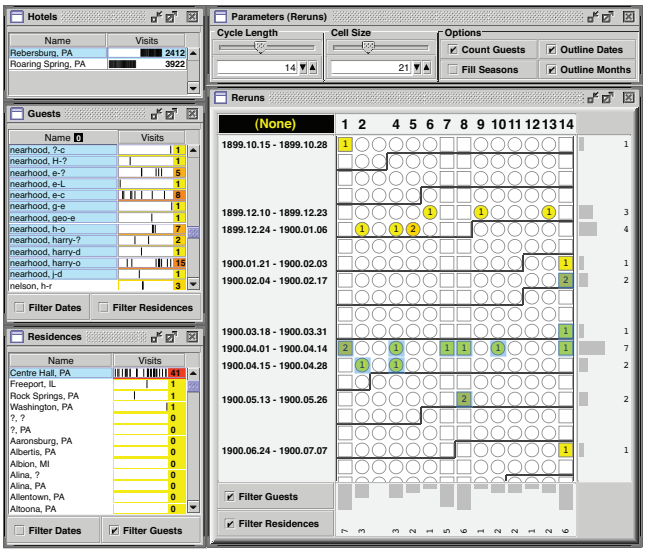

Figure 10: Travel from Centre Hall by possible members of the hotel proprietor's family. Nine of 15 visits by Harry O. Nearhood (highlighted) occurred sporadically over a three week period in April 1900.

5.5 Interesting General Relationships

In addition to enabling specific discoveries, the visualization reveals general relationships in the data set that uphold prior observations and hypotheses. For example, we knew from transcribing that the number of guests 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 hotel who shared the same surname, Nearhood, as the proprietor of the hotel. When querying all guests with this last name (figure 10), 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 ten 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.

5.6 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 eleven 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.

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 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 the Rebersburg Hotel? In other registers being analyzed, notations sometimes give instructions for forwarding mail, and invariably it is to a named hotel in another town. Toward answering these questions, we have recently added a second hotel register data set to the visualization and are currently transcribing several more.

6 Evaluation

We conducted a series of formative evaluations of the visualization. These evaluations targeted two goals: to solicit ideas for improvements and enhancements, and to evaluate it in terms of how well it satisfies a set of precepts for the design 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 fields in the school of Information Science & Technology. Members of the user group possess varied knowledge of GIS systems and are familiar with research priorities for the advancement of visual analytics.

6.1 Distributed Usability Assessment

Software evaluation activities often require significant investment of human and financial capital to develop and organize activities and procure technology to capture data. To facilitate assessment of the hotels visualization, we adapted e-Delphi, a web-based toolkit designed to support Delphi exercises. The Delphi method [4] was devised by the RAND Corporation shortly after World War II as a way for structuring group decisions in a manner that ensures personality conflicts and power relationships do not taint the presentation of ideas. The e-Delphi toolkit has been used in the past in consensus building exercises with climate change scientists [14]. 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.

We adapted the e-Delphi toolkit as a means to support group input to software evaluations. 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, which are the typical uses of the Delphi method. The toolkit allows us to conduct qualitative evaluations via the web with little investment in time and effort by either researchers or participants. The flexible design of the toolkit allows us to support a distributed, asynchronous, user centered design process.

6.2 Process

We conducted three rounds of evaluation using e-Delphi. 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.

6.2.1 Round 1

We asked users to complete two short tasks to become familiar with the capabilities of the visualization. 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 minutes (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.

6.2.2 Round 2

Amar and Stasko [1] 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 data set. [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 assessed the visualization in terms of these precepts. 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 as necessary to refresh their memory. The activity took 90 minutes and we again provided several days for revision.

6.2.3 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 minute session during class time and a few days afterward for revision. We extended response time an additional week to compensate for spring break.

6.3 Results

6.3.1 Round 1

Round 1 provided a great deal of valuable feedback on potential extensions to the visualization, additional data sources, and situations in which it might be particularly useful. One proposed extension was 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. Another user suggested the need for a means to help find groups of people who travelled together on the same day, perhaps from different places and for different lengths of stay. 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.

Users also requested sophisticated capture and annotation tools as part of the visualization so that they could develop stories about what they were seeing. 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.

6.3.2 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 noted by users was the inability to expose uncertainty in data and indicate its effect on outcomes. 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 set. 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 set.

6.3.3 Round 3

Round 3 provided further clarification of the survey responses received in Round 2. To better expose uncertainty and show its effects

on data, users suggested the creation of a visual summary for register entries that are incomplete or otherwise uncertain. Because the register data set is coded with question marks where names or places are uncertain, it will be relatively easy to implement a summary table that either contains all such records or tags each record with a more obvious indication of uncertainty.

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.

To clarify causation, the visualization could provide details on demand for individual dates, guests, and residences. Users expected to see ancillary information like holidays, special events, and weather 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.

7 Conclusions

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.

The e-Delphi evaluations resulted in overall positive feedback as well as concrete suggestions for design improvements and extensions to the visualization. As we implement these changes we will continue to solicit feedback from both domain experts and general users. The e-Delphi toolkit allows us to quickly conduct evaluations that can be repurposed for multiple user groups with little effort on our part. As the visualization matures, we will transition our evaluation efforts to techniques such as verbal protocol analysis and formal focus groups composed of professional analysts.

By creating innovative ways of examining spatiotemporal data, the 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 twentieth century.

Acknowledgements

Thanks to Bill Pike at the Pacific Northwest National Laboratory (PNNL) for creating e-Delphi and helping us adapt it for our evaluation work. Thanks also to 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] R. A. Amar and J. T. Stasko. Knowledge precepts for the design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):432–442, July 2005.

[2] J. V. Carlis and J. A. Konstan. Interactive visualization of serial periodic data. In *Proceedings of User Interface Software Technology (UIST)*, pages 29–38, San Francisco, CA, 1998. ACM Press.

[3] H. Collections and S. C. L. Labor Archives. *Rebersburg Hotel Guest Register, Rebersburg, PA, 1898-1900*. The Pennsylvania State University Libraries, University Park, PA.

[4] N. C. Dalkey. The Delphi method: An experimental study on group opinion. Technical Report RM-5888-PR, The Rand Corporation, 1969.

[5] J.-D. Fekete. The infovis toolkit. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 167–174, Austin, TX, October 2004. IEEE.

[6] B. Gelber. *Pocono Weather*. Uriel Publishing, December 1992.

[7] J. Heer, S. K. Card, and J. A. Landay. prefuse: A toolkit for interactive information visualization. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, pages 421–430, Portland, OR, April 2005. ACM Press.

[8] D. W. Holdsworth. Historical geography: New ways of imaging and seeing the past. *Progress in Human Geography*, 27(4):486–493, 2003.

[9] A. K. Knowles, editor. *Past Time, Past Place: GIS for History*. ESRI Press, Redlands, CA, 2002.

[10] M. Livny, R. Ramakrishnan, K. Beyer, G. Chen, D. Donjerkovic, S. Lawande, J. Myllymaki, and K. Wenger. DEVise: Integrated querying and visualization of large datasets. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 301–312, Tucson, AZ, 1997. ACM.

[11] A. M. MacEachren, M. Gahegan, W. Pike, I. Brewer, G. Cai, E. Lengerich, and F. Hardisty. Geovisualization for knowledge construction and decision support. *IEEE Computer Graphics and Applications*, 24(1):13–17, January 2004.

[12] C. North and B. Shneiderman. Snap-together visualization: A user interface for coordinating visualizations via relational schemata. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 128–135, New York, NY, USA, May 2000. ACM Press.

[13] C. Olston, A. Woodruff, A. Aiken, M. Chu, V. Ercegovac, M. Lin, M. Spalding, and M. Stonebraker. DataSplash. In *Proceedings of the International Conference on Management of Data (SIGMOD)*, pages 550–552, Seattle, WA, June 1998. ACM.

[14] W. Pike, B. Yarnal, A. M. MacEachren, M. Gahegan, and C. Yu. Infrastructure for human-environment collaboration: Building a prototype for the future of science. *Environment*, 47(2):8–21, 2005.

[15] J. J. Thomas and K. A. Cook, editors. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, August 2005.

[16] M. Wattenberg. Arc diagrams: Visualizing structure in strings. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 110–116, Boston, MA, October 2002. IEEE.

[17] C. Weaver. Building highly-coordinated visualizations in Improvise. In *Proceedings of the IEEE Symposium on Information Visualization*, pages 159–166, Austin, TX, October 2004. IEEE.

[18] A. T. Winfree. *The Timing of Biological Clocks*. Scientific American Library, December 1987.