ASSESSING THE EFFECTIVENESS OF TEMPORAL LEGENDS IN ENVIRONMENTAL VISUALIZATION

R.M. Edsall  
Department of Geography, Penn State University  
M.-J. Kraak  
ITC Department of Geoinformatics Enschede The Netherlands  
A.M. MacEachren  
Department of Geography, Penn State University  
D.J. Peuquet  
Department of Geography, Penn State University

Abstract:

The increasing prevalence of interactive visualization applications for the representation of spatiotemporal environmental data has created a need for research into the cartographic design of these dynamic applications. The understanding of temporal animations requires that users not only perceive the spatial information presented, but also, just as importantly, be able to locate that information in time and understand changes through time. Spatial legends can support these needs by prompting appropriate mental views (schemata) for interpreting the temporal component of dynamic maps and by providing a method through which users can control what they see. Little empirical research has been directed toward discovering the relative advantages and disadvantages of the various different legend styles that have been designed to indicate the temporal location of an animation. This paper reports on an assessment of the relative effectiveness of three types of temporal legends to communicate information and facilitate visual thinking. These three types have been selected from a much larger set of legend styles currently in use for spatiotemporal data presentation and exploration. Our general hypothesis is that the effectiveness of a temporal legend depends to a significant extent on both the task (what questions are being answered) and the application (what phenomenon is being investigated) for which it is implemented. The results of this assessment will be important for the future design of well-reasoned cartographic animations intended to facilitate study of any type of spatiotemporal process.

INTRODUCTION

The integration of analysis and visualization of spatial data that has a significant temporal component has become a priority in environmental science. This synthesis is the focus of the Apoala project, a cooperative effort of the departments of Geography and Computer Science at Penn State. The application of multi-scale space-time database models, the effective querying of the data within those models, and the facilitation of traditional and exploratory analysis methods through visualization are the three primary objectives of the Apoala project. These realms of geographic information science are rich with research questions, including that which is addressed here: the design of the visual data representation to enable creative and successful investigations.

Geographic information scientists are recognizing the enhanced ability to identify and analyze processes and trends that is provided by systems that support temporal as well as spatial analysis and visualization. Langran (1992) introduced the potential applications of a system designed to handle spatiotemporal information, and Peuquet (1994) suggested a conceptual outline for the integration of time into GIS (also see Peuquet and Qian (1996)).

In order to visualize spatiotemporal data effectively, the utilization of interactive and dynamic cartographic techniques is required. The integration of time into geographic visualization applications is most often incorporated through animation, where the temporality of the data is displayed as a sequence of images within which changes, patterns, and processes can be identified visually. Animations, particularly those designed for computer display, afford a user many advantages over static maps (Monmonier, 1990; MacEachren, 1995), including an enhanced perception of change, or the ability to control, for example, the location in real-world time of the image.

As cartographers and analysts come to rely more and more heavily on cartographic animation to represent spatiotemporal information, a need has developed for empirical research focused on the design of cartographic animations as a visualization method integrated with GIS. Interface design (Edsall and Peuquet, 1997) and data representation techniques have a profound impact not only on what information is communicated by the representation but also on how that information is interpreted and explored. As an example, the method of representing and interacting with the temporal variable of the representation may influence the "meaning" of the map.
Kraak, Edsall, and MacEachren (1997) - hereafter referred to as KEM - present a conceptual approach to the design of temporal legends--tools for representing and interacting with the temporal variable--for animated maps. This paper follows on that theoretical framework and synthesizes it with the discussion on implementation by Edsall and Peuquet (1997). We report here on an experiment that addresses the possibility that the way in which the temporal variable is represented and modified influences the ability of a user to interpret data. In addition, we propose that the design of the temporal legend also has an impact on the strategies employed by the user to perform analysis tasks or explore a data set.

**APPROACH: DIMENSIONS OF TEMPORAL LEGEND DESIGN**

The general hypothesis - that the design of a temporal legend influences the way geographic tasks are conceptualized and executed - needs, of course, to be decomposed into more experimentally manageable units before an investigation can be designed. Among the many different ways that this problem might be considered is to classify the relevant aspects of visualization, including the design of the temporal legend, the nature of the phenomenon mapped, and the characteristics of the tasks and interpretations that are required of the user.

**Temporal Legends**

Legends in animated maps serve multiple purposes. Like traditional static maps, a legend must explain the meaning of the various visual symbols on the map. In a dynamic map, however, another role of the legend is to give information about the location in time of the image that is displayed. In one sense, the temporal legend is an indicator, not unlike a north arrow, which orients the user to a temporal framework. In interactive applications, the temporal legend can also serve as a tool to manipulate various aspects of time in the animation, by, for example, moving to a specific point in time, or specifying a certain temporal resolution or period for averaging or aggregation (see Monmonier's (1990) discussion of temporal brushing).

A wide variety of different methods of representing the dynamic variable (that variable, often time, which is the basis for the animation) currently exist. KEM (1997) classify them as follows: on one level, a temporal legend may exist either visually separated from the map (e.g. in a separate pop-up window) or embedded into the design of the map itself. The latter might involve a change in background color with change in real-world time or the addition of an audio message which indicates time (Krygier, 1994). However, visually separate graphic legends offer the greater opportunity for interaction and detailed investigation. Within this class of legends, three sub-categories are identified: (1) text, with simple letters and numbers that change with the animation, continually giving the precise time of the image displayed; (2) bar, a more graphic legend type, linear in form, which indicates the animation time through either an icon that moves along the length of the bar, or an area which "fills" with color, indicating time passed; and (3) clock, which depicts the passage of time as progress around a circle, not unlike a traditional analog clock.

**Space-time Phenomena**

The effectiveness of a legend type for data presentation and exploration may very well be application-specific. That is, it is possible that one style of legend prompts a particular perspective on time within a knowledge domain while another style prompts a different (but equally valid) perspective. As described by KEM (1997), classifying space-time phenomena can be accomplished in a variety of ways, no one of which is sufficient to describe all forms of data.

One such classification might be according to the temporal linearity of the phenomenon. Some phenomena are inherently periodic; others show no signs of periodicity over the time scale of interest. For example, El Nino cycles may be related to local drought patterns if their behavior is similarly periodic. Another distinction might be made according to the stability in space or feature identity of the phenomenon. Animations showing an object changing its location and those showing a location changing its attributes might be interpreted differently. A map showing a bird's migration path animates cartographic objects which change location over the course of the animation, but which themselves do not change identity, while a map of a community's land-use change over the last century shows attributes (land-use) changing, but no location change. A third classification of temporal data that might influence legend choice is the temporal regularity of the data observations. KEM (1997) point to the possible importance of making the user aware of the fact that recorded observations occurred or were collected at irregular time intervals. This may have implications on the graphic design of the legend; for example, the graphic size (say, the length of a "day" on a slider bar) of a given time period in the legend might need to vary if the data varies in regularity.
The use of the temporal legend not only to indicate but also to manipulate the place in time shown by the animation affords the user the ability to ask and answer a wide variety of different types of temporal questions. Of course, temporal queries represent a small subset of a whole host of analysis needs of a spatiotemporal geographic visualization system. Temporal queries are organized into six query types (KEM, 1997; MacEachren, 1995) which represent the most basic forms of simple and complex temporal investigative tasks. It is reasonable to expect that strategies for answering (and posing) these different queries will vary according to the legend type and that the design of the representation of the temporal variable should therefore be task-based in addition to being application-based. The six query types are related to the existence of an object or event (if?), its location in time (when?), its duration (how long?), its "temporal texture" (how often?), its rate of change (how fast?), and its sequence (what order?).

The classifications of legend types, phenomenon types, and query types described above are summarized in Figure 1. Each "box" in the cube represents one possible combination of legend, phenomenon, and query, and thus outlines one experimental unit to assess.

THE EXPERIMENT

Dependent Variables in the Experiment

To investigate temporal legend design, a preliminary human subjects experiment was developed. Choices must be made by a researcher to decide the variables by which the "effectiveness" of a legend type is to be determined. If a dynamic map is to communicate information, as many are created to do, the ability to accurately interpret that information should be a prime consideration of a map's design. The speed at which a map is (correctly) interpreted might also be of importance, particularly since, by their nature, animated maps exist in time as well as space and certain important data might literally be missed if the display is confusing or distracting. Some dynamic maps are created for data exploration as opposed to communication. In these cases, there is no particular "correct" interpretation of the map, and emphasis in the map's design might instead be placed on its facility in hypothesis generation or its flexibility in creative data manipulation tasks. In this respect, it might be valuable to collect information about how
The Scope of the Investigation

For this preliminary project, only one type of phenomenon - synoptic-scale weather over the United States - was selected for investigation. Such weather maps were chosen for several reasons. First, they are familiar to most potential subjects; many television weather reports today incorporate animated maps into their presentation, thus the weather phenomena presented to the subjects would be relatively comprehensible in a short amount of time. Second, weather maps can show multiple forms of phenomena. Some phenomena on weather maps, like storm centers, can be classified as features which change location but not attribute, while "cloudiness" is an attribute which can be conceived of as changing at a fixed location. From another perspective, weather maps show both temporally linear and cyclical phenomena, with the aperiodic motion of synoptic weather systems across the continent as well as the diurnal cycles of thunderstorm activity characteristic of late spring North American weather.

Legend types investigated. Three legend types with varying levels of interactivity were designed, each corresponding to one of the three "visually separate" legend classes described above. They were designed to resemble—at least loosely—existing temporal legends that may be familiar to subjects. The first of these was a timeline, consisting of seventy-six small boxes, each representing one hour of the animation (Fig 2). As the animation advances, the box representing the hour shown on the map is filled; the color of the fill indicates the day (27th, 28th, etc.) shown. This legend is fully interactive clicking on any one of the boxes in the line advances the animation to that hour. Advantages to this type of legend include the ability to instantly evaluate the position of the animation relative to its beginning and ending points, and the invocation of an intuitive concept of time as a linear phenomenon.

The second legend type investigated was a timewheel legend, designed to resemble a clock face, which advances hourly by filling in the appropriate portion of a circle that is divided into 24 wedges (Fig 3). A user may click on any of these wedges or on a
vertical timebar (with four boxes representing each of the four days in the animation) to advance to a desired time. Using this legend, cyclical, particularly diurnal, phenomena can be easily explored, and the clock-like design may also be user-friendly because of its familiarity.

![Figure 3. The cyclical legend type.](image)

Also in the experiment was a simple "text" legend, which indicated the time shown in the animation by numerals and letters, not unlike a digital clock (Fig 4). Advantages of this legend type include its ease of understanding, its familiarity, and its compact size.

![Figure 4. The text legend type.](image)

**Query types investigated.** Because one hypothesis of this project was that the effectiveness of the legends depends on the type of query made of the data represented, the experiment was also organized around three of the six query types mentioned earlier in this paper. Each subject thus answered nine questions: one (of the three) queries for each (of the three) legend types. The three query types — "when?", "how long?", and "how fast?" — were chosen because they are of varying complexity and require varying interaction and solutions strategies. In addition, questions based on these queries had objective and repeatable "correct" answers that were straightforward to test.

**Dependent variables recorded for analysis.** Each subject, as mentioned above, answered nine (multiple-choice) questions, each of which had one correct and at least two incorrect choices. The answer selected was recorded, as was the time necessary for the user to respond to the question. Also recorded was important information on the timing and nature of the interactions of the user with the display: which tools did the user utilize, and when? Data was also collected from subjects about their age, gender, and experience with weather forecasting and map interpretation. Subjects were also invited to give feedback and design suggestions in written form at the end of the test.
The survey instrument.

The map. An animated, interactive weather map was created for use in the preliminary human subjects experiment. The animation was developed within the Macromedia Director multimedia environment. Though the Apoala project's ultimate goal is the development of a UNIX-based application, for portability, the test environment for this experiment was developed on desktop PCs. Sixty-one satellite images were downloaded from the Purdue University Weather Processor Internet site. These images show, in grayscale, the relative amount of water vapor in the atmosphere over the continental U.S. for each hour, with several exceptions, for a 76-hour period from 00 GMT on April 27, 1997 to 03 GMT on April 30. Information on locations of lightning strikes -- both cloud-to-cloud and cloud-to-ground -- and on locations of tornadoes was included in the images, and the centers of the numerous cyclones (low-pressure systems) that formed and moved across the continent were also analyzed and marked on the images.

The methods of interaction. The animation was controllable by the user in a number of ways. As described above, both graphic legend types - the timeline and the timewheel - were "active" objects that would respond upon interaction. In addition, regardless of the legend types, the animation could be controlled at all times using a set of "VCR buttons," so named because they function similarly to those found on video cassette recorders. Using these buttons, a user can play, stop, and rewind the animation, and step it forward and backward. Before being asked to use a certain interface design to answer interpretation questions, subjects in the experiment were given the opportunity to practice each legend type to observe the function of each component.

The questions and the legends. Each legend type was applied to all three question types in succession. In other words, a subject might have first been asked to use a linear legend to answer "when?", "how fast?", and "how long?" questions, then been asked to use a cyclical legend to answer different "how fast?", "how long?", and "when?" questions. Subjects were assigned by the investigator into one of three experimental groups--A, B, or C. Group A saw the timeline (linear) legend first, then the timewheel (cyclical), then, for questions seven through nine, the text legend. Group B saw the cyclic legend first, then text, then linear, and Group C saw text, then cyclic, and finally linear. By evenly dividing subjects this way, order effects - the influence of practice allowing a user to perform better later in the experiment - were minimized. Questions varied in difficulty among the query types, but relative difficulty of the questions within a query type were kept as similar as possible. For example, the three questions of the "when?" type, questions 1, 5, and 9, were: "During these four days, at what time does lightning first occur in Pennsylvania?", "When is the tornado sighted in the state of Virginia?", and "When does a line of severe thunderstorms pass over the Florida peninsula?".

RESULTS AND DISCUSSION

Over two days in June 1997, 23 subjects were paid $5 each to participate in the study. They spent anywhere from 13 to 29 minutes in front of a Macintosh PowerBook. Of the 23, 20 subjects' responses were deemed useable for analysis; on two occasions, the computer crashed, requiring a reboot. The two subjects who suffered through this were compensated but their incomplete responses were not recorded or analyzed. In one other case, a subject's responses were discarded because the test was completed improperly. The remaining subject pool was balanced by gender (10 males, 10 females), and consisted mostly of university undergraduates. The subjects were provided with an atlas for placename reference, and they were free to ask the investigator present any question that would help them complete the tasks.

Hypotheses and data analyzed

Differential statistical tests were performed to assess whether subjects performed better and/or faster with any one legend type over another. Specifically, the following four hypotheses were assessed:

A1. Overall performance in correctly answering the nine questions depends on the legend type used to answer the question.
A2. Performance in correctly answering questions concerning a specific type of inquiry depends on the legend type used to answer the questions.
B1. Overall response times to answer all the questions vary according to the legend type shown.
B2. Response times to answer individual questions according to query type depend on the legend type shown.

The number of correct answers was counted and sorted according to legend type and query type. Response times were sorted and ranked according to legend type and query type.
Results

Overall, the results of this preliminary investigation showed no significant differences in performance or speed among the three legend types, or between any two of the three types. Two different tests, the $\chi^2$ and Kruskal-Wallis analysis of variance by ranks, were used for the analysis. Four hypotheses were tested, two concerning accuracy and two concerning response times.

**Hypothesis 1.** Overall performance in correctly answering the nine questions depended on the legend type used to answer the question. This hypothesis was not supported. A total of 180 questions was answered (9 questions, 20 subjects), 60 for each of the three legend types. The data show no significant difference in overall performance among the legend types. The cyclical legend was used 52 times with a correct result, and the linear and text legend both were used 43 times of the 60 possible to correctly answer the question (Fig 5, left). A $\chi^2$ test of these counts reveals that these are not statistically significant differences, even at the 90% probability level ($\chi^2 = 1.174, \chi_{crit}^2 = 4.60$ at $\alpha = 0.10, df = 2$).

**Hypothesis 2.** Performance in correctly answering questions concerning a specific type of inquiry depended on the legend type used to answer the questions. This hypothesis also failed to receive support statistically. The $\chi^2$ value for two of the three query types, "when?" and "how fast?", was below 1.0, indicating no statistical difference among the three legend types in facilitating the accurate answering of questions of those types (Fig 5, right). For the "how long?" query, a counterintuitive result, though not statistically significant: the linear legend was only used 8 of 20 times with success, whereas the cyclical legend was used 16 of 20, the text 13 ($\chi^2 = 2.655, \chi_{crit}^2 = 4.60$ at $\alpha = 0.10, df = 2$).

**Hypothesis 3.** Overall response times to answer all the questions vary according to the legend type shown. Again, the results of this preliminary investigation show that there is no significant statistical difference in response time among the three legend types, aggregated across query types. The mean response time for questions using the linear legend was 75.0 seconds, for the cyclical legend, 81.5 seconds, and for the text legend, 75.3 seconds (Fig. 6, left). The Kruskal-Wallace ANOVA by ranks reveals that this is a non-significant difference ($H = 0.912, H_{crit} = 4.60$ at $\alpha = 0.10, df = 2$). We considered the possibility that response times to correctly answered questions might vary according to legend type. However, after removing the 22 incorrect responses (out of 180), the mean response times of correct responses are still close enough to each other to resoundingly fail to support the hypothesis (linear: 73.3 sec., cyclic: 85.0 sec., text: 82.0 sec.; $H = 0.809, H_{crit} = 4.60$ at $\alpha = 0.10, df = 2$).

**Hypothesis 4.** Response times to answer individual questions according to query type depend on the legend type shown. Once again, the preliminary project failed to find conclusive evidence to support this hypothesis. Only the query type "when?" showed any substantial (but not significant) differences among the response times using the three different legend types: Subjects responded to the "when?" questions, on average, in 66.8 seconds, while the same questions took those who were using the cyclical legend 84.8 seconds (Fig 6, left-center). Those using the text legend required 71.3 seconds to accomplish the interpretation tasks ($H = 3.370, H_{crit} = 4.60$ at $\alpha = 0.10, df = 2$). No other query type showed a variation of over ten seconds among the means.
Forthcoming analysis

Though the above analysis showed no significant differences in performance among the three legend types at relatively simple information retrieval and interpretation tasks, potentially more valuable information about interface design for such data will come from the analysis of the "interaction" data. Each time any of the legend tools was used, the time and the interaction was recorded. From this data, a "history" of each subject's manipulation of the data can be extracted. This data is perhaps more useful to inform interface design, as trends and patterns of learning the interface and mentally organizing the query investigation might be identified. We are now in the process of analyzing these interaction logs.

FUTURE DIRECTIONS

The investigation of temporal legend design for animated maps of spatiotemporal data will continue with more subjects, more detailed questions, more (and more useful) legend types, and more applications. One significant change in future iterations of this investigation will be to ask subjects to generate their own hypotheses and conclusions from the information displayed. It is possible that different legend types will prompt different sorts of hypotheses; for example, if an animated map is shown with a wheel-like temporal legend, is a subject more likely to pose a question concerning a cyclical aspect of the data shown? What are the strengths and weaknesses of each of the legend styles? Ultimately, thus, we hope to understand the implications of legend form for interactive data analysis and information retrieval.

ACKNOWLEDGEMENT

This work was supported under Grant R825 195-01-0 of the U.S. Environmental Protection Agency and FAW NSF90-27 of the U.S. National Science Foundation.


REFERENCES


