Visually Exploring Multivariate Trends in Patient Cohorts using Animated Scatter Plots

Alexander Rind[1], Wolfgang Aigner[1 2], Silvia Miksch[1 2], Sylvia Wiltner[2], Margit Pohl[2], Felix Drexler MD[3], Barbara Neubauer[2], and Nikolaus Suchy[2]

[1] Danube University Krems, Austria
[2] Vienna University of Technology, Austria
[3] Landesklinikum Krems, Austria
aigner@ifs.tuwien.ac.at, margit@igw.tuwien.ac.at

Abstract. The effectiveness of animation in visualization is an interesting research topic that led to contradicting results in the past. On top of that, we are facing three additional challenges when exploring patient cohorts: irregular sampling, data wear, and data sets covering different portions of time. We present TimeRider, an improved animated scatter plot for cohorts of diabetes patients that tackles these challenges along with its evaluation with physicians. Results show that animation does support physicians in their work and provide further domain-specific evidence in the discussion on the effectiveness of animation.

Keywords: Information Visualization, animation, time, medical data

1 Introduction

In health care, analyzing and exploring data of patient cohorts are important tasks, especially in quality control for healthcare providers or clinical research. Here, it is not only important to show the developments of one or more variables over time individually but also to explore relationships of variables and the dynamics thereof over time. This substantially increases the complexity of the task and demands sophisticated methods, such as correlation analysis. However, following the concept of exploratory data analysis, visual methods can be a valuable aid for getting an overview of the data and its relationships as well as for generating hypotheses or discovering surprising insights that would have possibly been overlooked by using statistical methods only. Furthermore, visualization can help in presenting findings to non-experts that might have a hard time interpreting computational results. For example, a physician might be interested in an overall view of how the conditions of patients develop over time. This could reveal that a certain group of patients is behaving differently from the rest. More specifically, a possible question would be to examine whether this is caused by a certain treatment method, the patients' lifestyle, or a combination of both.

Probably the most popular visual method to explore relationships between variables is the *scatter plot*. However, each individual scatter plot is a static snapshot of the relationship of two variables only and developments over time cannot be seen. There are basically two possibilities for visualizing dynamics over time: *static* representations that use spatial features such as a time axis, or *dynamic* representations that map time to time (i.e., animation or slideshows) [1]. For scatter plots, both axes are already used for data variables. Therefore, we use animation for showing developments over time. Animated scatter plots were prominently used by Rosling [16] in his famous presentations where he effectively communicates how complex socio-economic relationships develop over time. However, this approach is not applicable straightforwardly in other situations. Particularly, three challenges arise in handling the dynamics of patient cohort data: 1) irregularly sampled data, 2) data wear, 3) comparing datasets that cover different portions of time. First, data of different patients are usually sampled independently of each other, which is in contrast to, e.g., Rosling who shows data gathered on a yearly basis. Second, the validity of a medical parameter is usually decreasing over time and simply interpolating between two readings would hide this fact. Third, also patients that have been treated sequentially should be visually comparable in parallel.

Next, we discuss related work followed by an introduction to the application domain of diabetes care. In Sect. 4, we present our improved animated scatter plot and how the mentioned challenges are met. To examine our method, a usability evaluation with domain users has been conducted and is presented in Sect. 5. Evaluation feedback has been applied, so that we can present an updated software version in Sect. 4. Finally, we provide a conclusion and pointers for future work in Sect. 6.

2 Related Work

Considering animation in visualization, contradicting results can be found in research. Bartram [2] elaborates about potential advantages and uses of animation whereas a main motivation is that animation is an additional display dimension easing the representation of large amounts of multivariate dynamic data. She discusses five main factors that are potential benefits of using motion: 1) perceptually efficient, 2) rich interpretative scope, 3) still relatively underused, 4) more potential coding bandwidth, and 5) it gets technologically easier to implement. Nakakoji et al. [11] conducted two user studies to investigate cognitive effects as well as interactivity of animated visualizations for exploratory analysis. Their results support the arguments of Bartram and find animation a powerful instrument. In a more recent study Griffin et al. [8] investigated the effectiveness of animation for detecting moving clusters on a 2D representation considering the effects of timing (i.e., frame rate). Both error rate and task completion time consistently show advantages for the animated conditions in comparison to a small multiples display. The authors refer to the perceptually efficient Gestalt principle of *common fate* as possible explanation.

However, there are also critical voices on using animation. One argument against animation is the perceptual effect of *change blindness*, which can lead to an unnoticed miss of important changes. Nowell et al. [13] take up this argument and investigate potential causes as well as solutions for two visualization systems they developed. Tversky et al. [17] conducted a critical survey of evaluations that compared static and animated representations, many of which saw animation as beneficial. The authors argue that these experiments suffered from several flaws, most importantly that the static and animated cases were not comparable (e.g., less or more information presented or different representations). Their main concerns on animation are threefold: 1) may be hard to perceive, 2) may be comprehended discretely, and 3) there is a lack of interactivity. Especially, concerning the last issue the authors point out that interactivity may be the key to overcoming the drawbacks of animation. Most importantly, users should be able to control the speed, view, review, start, stop, zoom in and out, and change orientation of parts or the whole animation.

Animation has been applied to scatter plots most prominently by Rosling [16] and his publicly available Gapminder Trendalyzer. Robertson et al. [15] conducted a study on the effectiveness of trend analysis comparing animated scatter plot, static representation of traces for all variables in one display, and static small multiples representations with traces individually for each variable. Their results show that the static variants were better suited for analysis tasks and the animated scatter plot was better suited for presentation tasks.

Overall, it can be noted, that no clear view on animation in visualization exists. Contradictory views have been presented in prior research suggesting for example that animation is well suited for identifying moving clusters [8] whereas others argue that static views are better suited for analysis tasks [15]. This suggests that further research is needed to investigate animation in visualization.

3 Medical Scenario

Our research on exploration of patient cohorts was conducted in cooperation with a diabetes outpatient clinic. Diabetes mellitus is a widespread chronic condition in which the human body is no longer capable of managing its glucose (blood sugar). Patients need to change their lifestyle, take oral medication, and/or inject insulin. Otherwise, they are at risk of many complications, e.g., diabetic coma or cardiovascular disease. The choice of treatment depends on many factors including diabetes type, comorbid diseases, and the patient's experience.

The data set was collected during checkup examinations at the clinic, which were scheduled in intervals between six weeks to three months, depending on the patient's condition. Consequently, the data set is sampled irregularly. It encompasses 35 patients (anonymized), ten quantitative variables (e.g., fasting blood glucose level), and 22 binary variables (e.g., insulin therapy).

Physicians plan to use TimeRider for quality control and clinical research in connection to their work in the clinic, for example to find out whether some therapies are more effective than others. They are interested in whether their patients' conditions improve or worsen. They want to relate an improvement of some variables to the development of other variables, for example, lower glucose and gaining weight. They also need to compare and filter the patient cohorts.

4 Visual Encoding and Interaction Design

In this section, we describe how TimeRider meets the challenges of visualizing the dynamics of patient cohort data.[4] It has been argued that rich interaction is essential to take full advantage of the benefits of animation [17]. Fig. 1 shows an annotated screenshot of our highly interactive prototype. The scatter plot is its central component. It represents patients as marks in a Cartesian coordinate system and maps two variables to the two axes. Since position is the most accurate visual variable, scatter plots allow for an expressive and effective processing by the human visual system [9]. Users can select variables from two combo boxes next to the respective axes. Further variables may be mapped to color, shape, and size of the marks. Thus, users may encode up to five variables visually.

Animation is used to explore the dynamics of cohorts. For tackling irregularly sampled data, we establish a common time unit (e.g., one day) that we use for frames in the animation. At each frame, we draw the patient mark on a linear trajectory between its previous and the next known position. To account for data wear and maintaining temporal context, we propose two techniques of enriching the visual encoding of time: transparency and traces. In *transparency* mode, marks fade out more and more as they move away from a known location. They have their full opacity only in a frame for which data exists. After that, their opacity linearly degrades. This makes patients with current data clearly stand out. In *trace* mode, a line starts from the previous known location and iteratively grows to the next location. The marks and traces of all past observations stay visible with trace diameter increasing over time. Thus, at the end of the animation complete patient histories can be seen (Fig. 2). However, this static view only tells about the direction of change or the existence of local extrema. Animation is needed to understand the timing and co-occurrence of these developments. To *navigate in time*, users may hit the play button and watch the events unfold. Alternatively, they can jump from frame to frame using the media player buttons or drag the time slider manually. In order to compare patient histories that cover different portions of time, our method allows users to *synchronize* the data set. Currently, it provides four synchronization options: 1) calendar date, 2) patient age, 3) start of treatment, and 4) end of treatment. For example, if users synchronize by calendar date, the animation will start with only one or a few patients and over time patients will appear and then disappear from the scatter plot. Each frame will show the probably interpolated patient state of the corresponding date. If they synchronize by start of treatment, the animation will start with all patients at once and each frame will show the patient state n time units after their first treatment.

Most variables in medicine have important *value ranges*. For example, glycated hemoglobin HbA1c has a normal range of 4 % to 6 %. If HbA1c is higher, the patient will be at risk of diabetes induced organ damages. Highlighting this information already in the visualization allows for a faster recognition of, for ex-

[4] Supplemental material and a Java Web Start application can be found at `http://ieg.ifs.tuwien.ac.at/research/timerider/`.

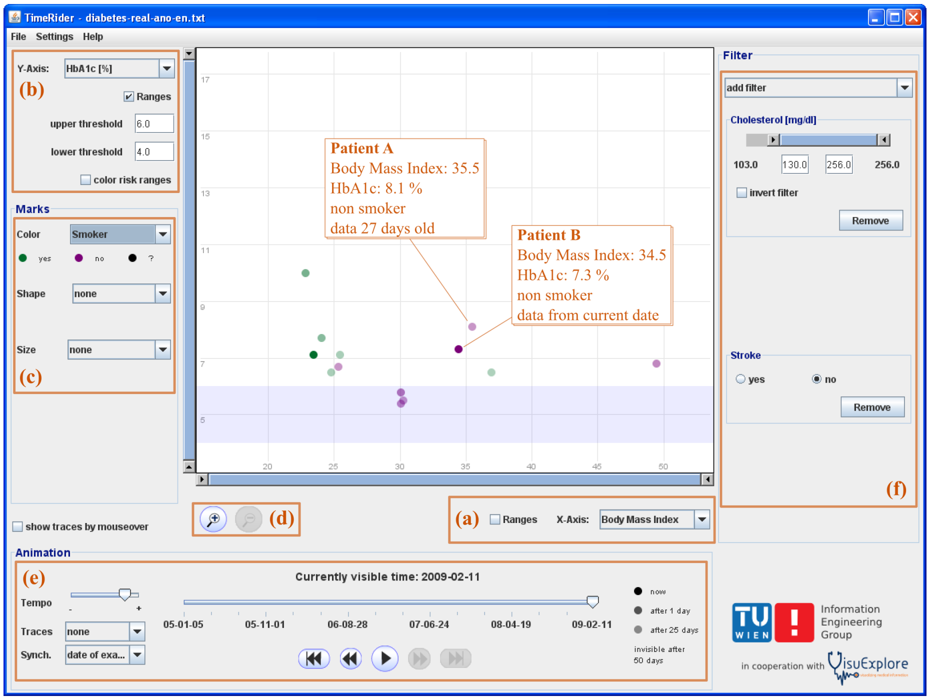

Fig. 1. TimeRider user interface: Patients are shown as marks in the scatter plot. Users may choose to map variables to (a) the x-axis, (b) the y-axis, (c) color, shape, and size of the marks. They may (d) zoom to show patients in more detail, (e) use animation to explore development over time, or (f) apply dynamic queries to filter patients. Here the normal range of HbA1c (y-axis) is represented by a light blue background. Two patients A and B demonstrate how transparency encodes the age of data values.

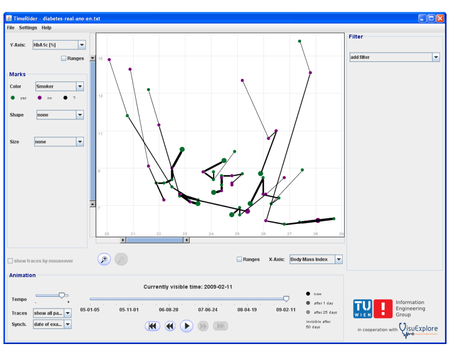

Fig. 2. TimeRider with traces shows the development over time for all patients.

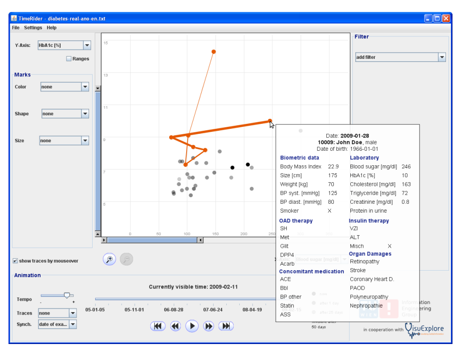

Fig. 3. On hovering the mouse over a patient mark, a trace shows its history and a pop-up window shows all current values.

ample, episodes with critical values. TimeRider represents these ranges using a light blue background in the scatter plot. Users can activate ranges for each axis (cp. Fig. 1 (b)) and choose whether they want to emphasize the normal range or the "risk" range, which is outside the normal range. They may also adapt range thresholds, though common thresholds are predefined via meta data.

TimeRider supports common interactions for select, pan, zoom, filter, and detail on demand (Fig. 3). Furthermore, it allows users to take a snapshot of the exploration state by *saving all settings* to a file. They can keep it for later reference, restore it with a different data set, or share it with colleagues. Alternatively, they may export a screenshot of the exploration state as an image file. A reset function will restore the initial program state, if they want to start over.

5 Evaluation

In the following evaluation we assessed the usability of TimeRider. Though we tested an earlier version than described above, results are still valid, because the visual mapping has not changed. Our subjects were physicians (in contrast to many studies using students). This investigation was guided by the following research questions: 1) Does animation, specifically in TimeRider, support physicians in getting insights from time-dependent data? 2) Is the mapping (e.g., color, traces) we developed appropriate for the task? 3) Are there any general usability/utility problems that might also occur in similar systems?

5.1 Research Methods

To answer our research questions we adopted the following methodologies:

Thinking aloud [6, 3] has been occasionally used to evaluate Information Visualizations (see e.g., [10], [4]). Despite its problems, thinking aloud is a valuable research methodology yielding interesting insights into human reasoning processes [5].

We combined the thinking aloud methodology with *screen capture*. Preece et al. [14] point out that any observation method based on video is rather time-consuming and should be based on some kind of criteria for the interpretation and categorization of the users' actions.

We used the *evaluation categories* developed by Forsell and Johannssen [7] to interpret the data. Their system of categories is based on other well-known heuristics (e.g., Nielsen [12]). They derived empirically the most important usability heuristics for Information Visualization: information coding (mapping), minimal actions, flexibility (number of possible ways to achieve a goal), orientation and help, spatial organization (e.g., distribution of elements on the screen), consistency, recognition rather than recall, prompting (all means to support the users to find alternative ways of doing things), remove the extraneous (distraction through unnecessary information), data set reduction (features for reducing data sets) [7, p. 203]. Even if these categories are mainly meant for heuristic evaluation, we think they can also form a valuable framework for the interpretation of user actions when working with Information Visualizations. We categorized self-contained events or activities as, e.g., that the users started the animation to go forward or backward in time or that they applied the filter mechanism. When users engaged in such activities several times in a row these activities were categorized as one activity, but on the categorization sheet a comment was made how often it was repeated.

5.2 Description of the investigation

The evaluation consisted of three parts: an introduction into the domain and the software (about 15 minutes), the solving of four tasks by the participants while thinking aloud, and, in the last part, a short interview was conducted. The dataset used was the same as described in Sect. 3. For the evaluation we used version 3.4 of the TimeRider software, which is a predecessor version of the prototype described in Sect. 4. Windows Media Encoder 9, a video and audio encoding software, was used to capture the participants' activities on the screen (not the participants themselves) and record their comments while working with the software. The following interview was recorded separately.

Ten physicians (four women and six men with the age ranging between 26 and 35 and a single person about 50) participated in the evaluation. The physicians had not previously been involved in the development and saw the software for the first time. The duration of the evaluation varied between one and two hours. This difference is due to the fact that tasks and interview were open ended, and some of the participants spent considerably more time on these activities. Four tasks, developed with the help of physicians, had to be solved independently by the participants. For tasks one to three concrete parameters for the x-axis and y-axis were given; task one was designed to let the participants familiarize themselves with the software, task two and three had concrete questions to answer. The fourth task allowed the participants to freely experiment and interact with the software.[5] In general, the tasks were exploratory in nature and no predefined solution existed. Below, as an example of such a task, is task three as it was given to the participants:

Parameters: x-axis: NBZ (fasting blood glucose level);
y-axis: RR diast [mmHG] (diastolic blood pressure)

Task description: Limit the data set to NBZ $\leq$ 100; RR diast $\leq$ 80. Choose a setting that gives a good overview over the trends of the patients. Which patients show a favorable trend? What is the general trend of the group? Experiment at will. Describe your findings.

In addition to the tasks participants were given a list of variables and abbreviations used in the software (e.g., RR diast = blood pressure diastolic). As all participants were physicians, they were aware of correlations between certain parameters (e.g., some types of insulin cause a weight gain). Participants were also asked to experiment with all of the interaction possibilities (traces, color/ size encoding), which they all did.

[5] Due to space restrictions, the full list of tasks cannot be listed here but can be found at `http://ieg.ifs.tuwien.ac.at/research/timerider/`

5.3 Results

All participants were able to solve the tasks and were able to predict trends. Solving the tasks seemed to be quite easy, but the participants were slightly hesitant about predicting trends. An example for an insight a participant got is: "When people started treatment they had a high HbA1c (indication for plasma glucose concentration) value at the beginning, and then this value decreased until the next measurement rather quickly. This indicates that at the beginning patients were diagnosed with diabetes, then the therapy started and the HbA1c went down."

More than 50 usability problems were detected, and additionally, a number of interesting remarks regarding the software were made. The most serious problems are listed below. The number in brackets indicates the number of participants having that specific problem once or several times during the evaluation. Also added is the categorization according to the heuristic of Forsell and Johanssen [7]:

dropdown lists: Participants could not find entries because the order of variables was neither consistent nor clearly communicated (9, prompting).

filter function: The range sliders were unfamiliar and it was not possible to edit the thresholds by keyboard (7, data set reduction).

traces: It was often impossible to follow traces, even a single one (4, information coding/mapping), and the traces were confusing when all of them were activated (5, information coding/mapping).

overlapping values: There was no way to tell if and how many values (marks on the screen) are overlapping (5, information coding/mapping).

risk levels: Value ranges were not correctly identified because the normal range was highlighted but the user interface referred to "risk range" (4, information coding/mapping). When both risk levels (x-axis, y-axis) were active, participants could not match them to the corresponding x-axis or y-axis (4, information coding/mapping).

zoom bar: While using the zoom bar the scatter plot was empty which confused the participants (3, spatial organization).

colors: Color changes of the values over time confused the participants and drew their attention to these points (4, information coding/mapping). The yellow color used to distinguish values was difficult to discern (2, spatial organization).

5.4 Discussion

In summary, the user study identified many ways to improve TimeRider's ease of use. We tackled many of the problems and iteratively developed a new version, which we present in Sect. 4 of this work. The main improvements are: 1) Value range controls are located next to the respective axes and wording has been improved, also users may now explicitly "color risk ranges". 2) Filter controls allow keyboard editing. 3) During zoom the scatter plot is continuously updated and buttons were added to make the zoom feature more visible. Furthermore, we added features to synchronize patient histories, select patients, and save settings.

However, the methods to represent irregularly sampled data and data wear were kept unchanged.

Some of the innovative approaches presented obstacles to the users. E.g., it took minutes for the participants to understand the control of filter settings because range sliders were unknown to them. After getting familiar, they were able to use them without problems. All participants were able to detect trends, clusters, and correlations in the data, though it should be noted that prior to this learning all participants had problems with the navigation or control of TimeRider. At the end of the evaluation some participants worked easily with it while others were slightly frustrated.

As far as our research questions are concerned, the following results can be derived from the study: Previous research [17, 15] indicated that subjects were often confused by animations and could not derive any insights. In addition, it is often difficult to understand trends from static visualizations. This was not the case in our study. Therefore, we assume that the animation we developed supports physicians in their work. More research is necessary to clarify why animation in our case was more successful than in other contexts. With the modifications indicated by our study, the system can be learned fairly easily. Some of the mappings posed problems to the subjects, especially the risk levels, overlapping values and traces. These problems will also occur in other, similar visualizations. They can be described as generic problems. Specific usability problems found in this visualization were problems with filters and dropdown lists. Most of the usability/utility problems belong to the category of information coding/mapping. This is probably to be expected in Information Visualizations.

6 Conclusions and Future Work

TimeRider is an improved animated scatter plot that provides solutions to three challenges arising when visually exploring patient cohort data: irregular sampling, data wear, and data sets covering different portions of time. For this, it enriches the visual encoding of time with transparency or traces. The evaluation showed that physicians using our method successfully gained insights. More generally speaking, our work provides evidence for the effectiveness of animation in visualization acquired in a domain-driven study.

Additionally, the evaluation yielded valuable input for the iterative improvement of TimeRider and animated scatter plots in general. The version presented in Sect. 4 accounts for much of this input. Furthermore, there are several directions for future research: Overlap in general is a persistent problem in visualization. Overlapping marks may be avoided through jittering, though that needs to play together with animation. Traces seem to be most effective when they do not overlap, but even then, they are sometimes difficult to interpret. No easy solutions seem to be possible for more than a few traces. The mapping of value ranges could also account for the patient heterogeneity, e.g., different thresholds depending on demographics or therapy. We plan to examine the strategies that physicians chose to solve the tasks in our study. Finally, more user studies are needed to investigate animation and specifically the enriched visual encodings we provided for patient cohort data.

Acknowledgments. This work was supported by the Bridge program of the Austrian Research Promotion Agency (No. 814316). We wish to thank our partners: NÖGUS, NÖ Landeskliniken-Holding, Landesklinikum Krems, Systema.

References

1. Aigner, W., Miksch, S., Müller, W., Schumann, H., Tominski, C.: Visualizing time-oriented data—a systematic view. Computers & Graphics 31(3), 401–409 (2007)
2. Bartram, L.: Perceptual and interpretative properties of motion for information visualization. In: Proc. Workshop New Paradigms in Information Visualization and Manipulation. pp. 3–7. ACM (1997)
3. Boren, T.M., Ramey, J.: Thinking aloud: Reconciling theory and practice. IEEE Trans. Professional Communication 43(3), 261–278 (2000)
4. Burmester, M., Mast, M., Tille, R., Weber, W.: How users perceive and use interactive information graphics: An exploratory study. In: Proc. IEEE Conf. Information Visualization (IV). pp. 361–368 (2010)
5. Carpendale, S.: Evaluating information visualizations. In: Kerren, A., Stasko, J.T., Fekete, J.D., North, C. (eds.) Information Visualization: Human-Centered Issues and Perspectives, LNCS, vol. 4950, pp. 19–45. Springer, Berlin (2008)
6. Ericsson, K.A., Simon, H.A.: Protocol Analysis. MIT Press, Cambridge, MA, London, UK, 2nd edn. (1993)
7. Forsell, C., Johansson, J.: An heuristic set for evaluation in information visualization. In: Proc. Int. Conf. Advanced Visual Interfaces. pp. 199–206. ACM (2010)
8. Griffin, A.L., MacEachren, A.M., Hardisty, F., Steiner, E., Li, B.: A comparison of animated with static small multiple maps for visually identifying space-time clusters. Annals of the Association of American Geographers 96(4), 740–753 (2006)
9. Mackinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graphics 5(2), 110–141 (1986)
10. Marsh, S.L., Dykes, J., Attilakou, F.: Evaluating a geovisualization prototype with two approaches: Remote instructional vs. face-to-face exploratory. In: Proc. IEEE Conf. Information Visualization (IV). pp. 310–315 (2006)
11. Nakakoji, K., Takashima, A., Yamamoto, Y.: Cognitive effects of animated visualization in exploratory visual data analysis. In: Proc. IEEE Conf. Information Visualization (IV). pp. 77–84 (2001)
12. Nielsen, J.: Usability Engineering. Morgan Kaufmann, San Francisco, CA (1993)
13. Nowell, L., Hetzler, E., Tanasse, T.: Change blindness in information visualization: A case study. In: Proc. IEEE Symp. Information Visualization. pp. 15–22 (2001)
14. Preece, J., Rogers, Y., Sharp, H.: Interaction Design: Beyond Human-Computer Interaction. John Wiley & Sons, New York, NY (2002)
15. Robertson, G., Fernandez, R., Fisher, D., Lee, B., Stasko, J.: Effectiveness of animation in trend visualization. IEEE Trans. Visualization and Computer Graphics 14(6), 1325–1332 (2008)
16. Rosling, H.: Visual technology unveils the beauty of statistics and swaps policy from dissemination to access. IAOS Statistical Journal 24(1–2), 103–104 (2007)
17. Tversky, B., Morrison, J.B., Betrancourt, M.: Animation: can it facilitate? Int. J. Human-Computer Studies 57, 247–262 (2002)