Eurographics/ IEEE-VGTC Symposium on Visualization 2012
S. Bruckner, S. Miksch, and H. Pfister
(Guest Editors)

Volume 31 (2012), Number 3

A Qualitative Study on the Exploration of Temporal Changes in Flow Maps with Animation and Small-Multiples

Ilya Boyandin[1], Enrico Bertini[2] and Denis Lalanne[1]

[1]University of Fribourg, Switzerland
[2]University of Konstanz, Germany

Abstract

We present a qualitative user study analyzing findings made while exploring changes over time in spatial interactions. We analyzed findings made by the study participants with flow maps, one of the most popular representations of spatial interactions, using animation and small-multiples as two alternative ways of representing temporal changes. Our goal was not to measure the subjects' performance with the two views, but to find out whether there are qualitative differences between the types of findings users make with these two representations. To achieve this goal we performed a deep analysis of the collected findings, the interaction logs, and the subjective feedback from the users. We observed that with animation the subjects tended to make more findings concerning geographically local events and changes between subsequent years. With small-multiples more findings concerning longer time periods were made. Besides, our results suggest that switching from one view to the other might lead to an increase in the numbers of findings of specific types made by the subjects which can be beneficial for certain tasks.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [User Interfaces]: Evaluation/methodology—Qualitative evaluation

1. Introduction

Designing visualizations for the analysis of temporal changes in spatial interactions, that is, flows of entities (people, goods, money) moving between geographic locations, is a challenging task [BBBL11]. The complex nature of the data makes it difficult to find a suitable representation showing how the spatial relationships change over time while keeping the geographic metaphor intact.

A natural and popular representation of spatial interactions are flow maps [Tob87], which represent flows as lines connecting pairs of locations on a geographic map and their magnitudes by varying the widths or the color intensities of the lines. Adding the time component to this intuitive representation remains, however, a challenge.

There are two basic alternatives for this kind of extension which preserve the geographic metaphor and are, at the same time, easy to understand: animation and small-multiples.

Animation shows the changes of the flow magnitudes with interpolated transitions between the time periods. Small-multiples represent discrete time periods as static images arranged next to each other in a grid format. Animation allows for a higher resolution at each time step, but, given its transitory nature, puts a high load on the user's short memory. Small-multiples use space to represent time and, thus, provide only a limited resolution for each of the views.

Trying to better understand how these two solutions compare we studied the literature and found that:

- The comparison of animation and small-multiples is a classic problem studied in conjunction with a large number of visualization techniques [SSKY04, FRA*08, GMH*06];
- User studies carried out so far show that depending on the tasks and the exact settings of the experiment either animation or small-multiples is more efficient or leads to fewer errors [GMH*06, RFF*08, APP11, FQ11];
- No works comparing the use of animation and small-multiples with flow maps have been published so far.

Furthermore, most studies on dynamic graph visualization consider graphs in which the positions of the nodes can change, whereas the edge weights are constant. In flow maps, on the contrary, the nodes remain stationary and only

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd. Published by Blackwell Publishing, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

the edge weights, representing the magnitudes of the flows, change (flows may also disappear, when their magnitude becomes zero). Hence, flow maps represent an important special case of dynamic networks worth a separate investigation.

Our study takes a great deal of inspiration from the following research:

- Griffin et al's work on the perception of moving clusters [GMH*06], which shows that animated maps may reveal patterns that cannot be detected with static representations.
- Ellis and Dix's review of evaluation methods in visualization [ED06], in which they advocate for "explorative evaluation", an approach more focused on open-ended questions and with a higher chance of deriving new knowledge about a visualization.
- The broader line of research of insight-based evaluation [Nor06, SND05, YKSJ08] in which visualizations are evaluated in terms of the user-generated output rather than performance in the completion of benchmark tasks.

Inspired by these works, we decided to focus our research on the following question: "*Do animation and small-multiples lead to the detection of different kinds of information? And if yes, how do they differ?*" To this end, we used an open-ended exploratory protocol focusing on the collection of findings, that is, any kind of information users extracted by using the tool (we prefer to use the word "finding" instead of "insight" to allow for the inclusion of information at any level of complexity). We instructed the study participants to interact with the views and to document in the form of short sentences every piece of information they could find.

In the analysis phase we manually "coded" these findings using techniques drawn from grounded theory [Cha06] and captured the emergent categories. The distribution of the findings across these categories forms the basis for the comparison of the two views. The methodology we chose implies a qualitative nature of the study, therefore, our analysis is not based on statistical tests. However, we provided, where appropriate, statistical numbers to document our analysis (see Section 4.2).

In summary, the main contribution of this work is the observation that using animation or small-multiples may lead to different kinds of findings; without necessarily having one outperforming the other. Most notably, animation may promote findings of a smaller temporal and geographic scope than small-multiples. Our qualitative analysis of the collected data also leads to several useful guidelines for practitioners and open questions to pursue in future research. A secondary and minor contribution is the methodology itself. The coding approach we describe in the paper allows to compare interfaces in terms of the *types* of findings rather than just their *numbers*, allowing to gain a deeper understanding of the information people extract from visualizations.

2. Related work

Animated flow maps are discussed e.g. in [BEW95, TL96], but to our knowledge no user studies have been carried out yet which analyzed the effectiveness of animation and small-multiples for representing changes over time in flow maps.

MacEachren et al [MBHP98] discuss a user study of static and animated choropleth map representations of heart disease mortality rates. In this study a looping animation could reveal a specific pattern (a shift in location of high mortality rates) which was much more difficult to see with the use of discrete time stepping.

Slocum et al [SSKY04] present an evaluation comparing the use of animation and small-multiples in the software package MapTime for exploring temporal changes in geographic data. The evaluation which involved interviews and discussion groups showed that animations have a more important role for examining general trends, small-multiples for comparing arbitrary time periods, and change maps for explicitly depicting change.

Fabrikant et al [FRA*08] concluded based on the analysis of eye-movements of the participants of a controlled experiment that small-multiple displays are generally not informationally equivalent to non-interactive animations and that making an animation equivalent to a small-multiple display in order to achieve good experimental control for comparison may actually mean degrading its potential power.

In the user study by Robertson et al [RFF*08] the effectiveness of three alternative trend visualizations was evaluated by asking subjects to find answers to various analysis questions. Trend animation led to many participant errors and was the least effective form for analysis. The two static depictions of trends (small-multiples and traces) were significantly faster than animation, and the small-multiples display was more accurate. In this study each individual image in the small-multiples view represented the trace of changes over time of one single data element, not all the elements for a particular time slice like in our case.

Griffin et al [GMH*06] compared the effectiveness of animated vs static small-multiple maps for discovering space-time clusters and found that the subjects could more quickly and correctly identify clusters with animation than with small-multiples.

A lot of research has been carried out on the use of animation for representing changes in graphs [PS08, WB04, SP08, FQ11]. Archambault et al [APP11] evaluated the effectiveness of small-multiples and animation for the comprehension of graphs changing over time. In their study small-multiples gave significantly faster overall performance, whereas animation led to significantly fewer errors than small-multiples for the tasks of determining nodes and edges added to the graph. Moreover, the study showed that preserving the mental map (roughly, the node positions) while

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

showing changes both in animation and small-multiples had hardly any effect on the subjects' performance.

The effectiveness of animation and small-multiples is highly dependent on the information being visualized, the way it is presented to the users and the tasks they need to perform with it. It seems that very few general conclusions can be made about the usefulness of these two views for representing changes. Thus, their effectiveness has to be assessed explicitly for each particular situation.

	Group 1	**Group 2**	**Time limit**
	Training		-
Round 1 (main experiment)	SM	ANIM	20 min
	Rating findings		-
Round 2 (additional round)	ANIM	SM	10 min
	Rating findings		-
	Questionnaire		-

Table 1: *The protocol of our user study.*

3. Design of the study

The goal of our study is to find how animated (Fig. 1) and small-multiple (Fig. 3) flow maps compare in terms of the types of observations people make with them. Thus, we analyzed the findings made by the study participants in these two conditions. In the rest of the paper we will refer to the conditions as ANIM and SM respectively.

We performed an experiment with 16 subjects who were graduate and post-graduate students in computer science with no expert knowledge in migration flows or geographic visualization. The subjects were divided in two equally-sized groups. The first round of the study was designed as a between-subjects experiment. Each group was assigned one of the two views, either SM or ANIM (see Table 1). The participants were asked to explore the data by interacting with the view and to document their findings. The findings were collected in a database, and later, manually classified. We then performed a detailed comparison of the types of the first round findings between the views.

There was an additional second round in our experiment. The subjects were asked to continue making observations with the same dataset, but in the view which they *did not* use in the first round. Our goal was to see whether switching from one view to the other while still analyzing the same dataset would induce the subjects to make findings of different types compared to those which they made in the first view. Hence, we did not compare the findings made by the two subject groups in the second round. Instead, we compared the types of findings the subjects made in the second round with those which the same subjects made in the first round.

We preferred this approach over within-subject design because with a within-subject experiment, it would be much harder to analyze the effects of switching the view while still exploring the same dataset on the types of findings the subjects make. The main reason for designing the first round as a between-subjects experiment was also our desire to analyze the view change effects in the second round.

At the beginning of each session the subjects were trained. Both views and their interactive facilities were presented and explained to them. The subjects could interact with the views for a while and ask questions about them. The training was performed with a different dataset from the one used during the experiment itself.

After the training phase which took between 5 and 10 minutes, the first round started. The subjects were asked to make findings in the data using the view automatically selected for them depending on the group they were in. More precisely, the subjects were given the following task:

> *The views you will see represent refugee migration flows. Explore these views and type down the important findings you make.*

We also gave the subjects an idea of what an important finding is:

> *When deciding which findings are important, imagine, that you have to use them to present to somebody else what you have learned about these data.*

We decided not to give examples of findings to be sure the subjects are not biased by their particular types. The goal was to see what types of findings the subjects would come up with on their own.

In the view which was shown to the subjects beside the visualization there was a text field in which they had to type short one-sentence descriptions of the findings they make. After a subject had typed a finding and pressed "Enter", the finding was stored in the database and the text field was cleared, so that a new finding could be submitted. If a subject felt that no more important findings could be made in the view, the round was finished before 20 minutes were over.

After each round, the subjects rated the findings they just made by their importance on a Likert scale with four choices between "Not important" and "Very important". They also had the possibility to mark a finding as "Wrong" if they discovered an error, but they were not allowed to edit the findings.

In the second round the users were asked to continue the exploration of the same dataset during 10 additional minutes, but using the other visualization. In order to prevent too much fatigue, and taking into account the fact that the users would already be familiar with the data, we decided to make the second round shorter. During the analysis we did not compare the absolute numbers of findings made by the subjects, but the average percentages of the types of findings made in each of the rounds (see Section 4).

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

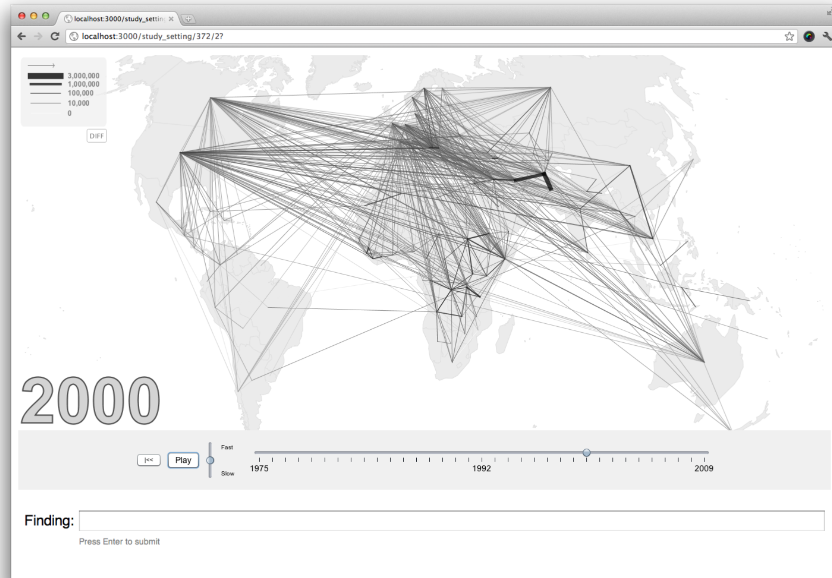

Figure 1: *ANIM condition (the animated view) representing migration flows between the world's countries. The subjects were asked to interact with the view and make findings about the dataset entering them in the text field below the view.*

Finally, after completing the second round and rating the findings the subjects were asked these multiple-choice questions on the computer:

- Which of the two interfaces did you prefer overall?
- Which of the two interfaces was easier to use?
- Which of the two views allows for a higher number of discoveries?
- With which of the two views it is more likely to miss relevant information?

The possible answers were: "animation", "small-multiples", "no difference", "I don't know". The participants were also asked to rate their overall impressions of the two conditions on a Likert scale with five choices and to describe the strengths and weaknesses of the conditions. The questionnaire ended with an open question for any general comments or suggestions.

The study was performed in a closed room in front of the same computer with a 24 inch monitor. One organizer was sitting next to the subject during the whole session.

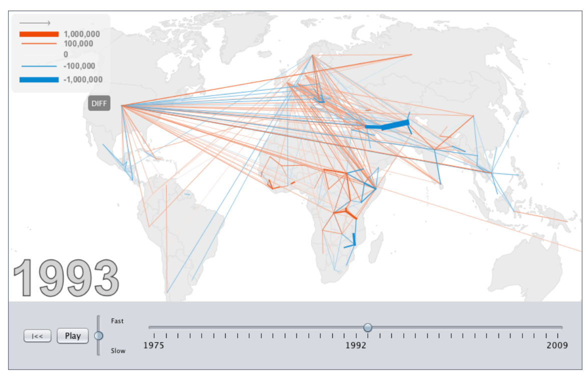

Figure 2: *The "difference view" of the ANIM condition which shows positive and negative changes of the flow magnitudes between the currently selected and the previous years. The study participants had the possibility to switch between the original view and the difference view at any time in both ANIM and SM.*

3.1. The conditions

The ANIM (Fig. 1) and SM (Fig. 3) conditions which we used in the experiment were based on the same flow map representation. In this representation flows of people migrating between the world's countries are shown with straight lines connecting the countries on a geographic map. The widths and the colors of each flow line represent the number of people migrating. We decided not to show the directions of the flows in order to simplify the views and avoid additional cluttering. For this experiment, we were more interested in analyzing findings concerning flows which changed their magnitudes over time, not in the flow directions.

The subjects had the possibility to highlight a flow by hov-

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

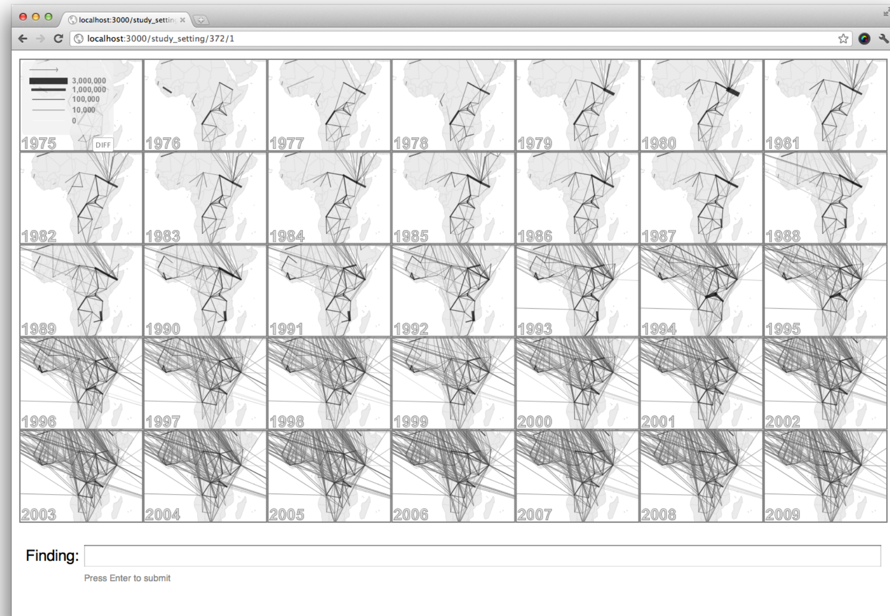

Figure 3: *SM condition (the small-multiples view). The condition supported zooming synchronously in all years' views (with the mouse wheel) and highlighting (by hovering mouse over a flow line). Here the user zoomed in to see Africa in more detail.*

ering over it with the mouse. When a flow was highlighted, detailed information about it was displayed, namely, the origin and the destination countries and the number of people migrating between them. In both conditions, ANIM and SM, it was possible to zoom and pan with the mouse. Zooming and panning in SM was applied simultaneously to each of the small-multiples. The animated view provided the users with animation controls: a play/stop button, a small slider for changing the speed of the animation, and a large slider which allowed to select any of the 35 years presented in the dataset. The animation smoothly interpolated the data between the years.

In both views, SM and ANIM, the participants had the possibility to switch to the "difference view" (see Fig. 2) which showed only the differences between the selected year and the previous one. Flows which had an increased number of people moving compared to the previous year were colored red and the ones which decreased blue. The width of these flows represented the absolute values of the differences. This way participants could see what exactly changed compared to the previous year. This made it easier for them to make and document findings concerning changes between subsequent years.

The main dataset we used for the experiment represented migration flows for 35 years (available from data.un.org). Each year's data contained about 200 nodes and a few hundred flows. We had to filter the flows, showing only the few hundred largest ones, in order to guarantee that the animation runs smoothly. The dataset we used for the tutorial was different: it represented commuters in Slovenia and contained data for 9 years. For our study we used a white-on-black color scheme, whereas for this paper we inverted the colors.

3.2. Data collected

During the experiment we collected the following data:

- short textual descriptions of the findings submitted by the subjects
- the importance of the findings as rated by the subjects
- screenshots of the views taken automatically when the findings were submitted
- interaction logs recorded during the sessions (all the users' actions supported by the views were logged, e.g. highlighting, zooming, starting animation)
- questionnaire submission
- videos with screen and audio recordings of the sessions.

Having such abundant data helped us during the analysis. Not only did it allow to discern various aspects of the process of making findings, it was also useful for clarifying the meanings of those findings, which were not clearly formulated.

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

4. Analysis

The main goal of the analysis of the collected findings was to find out whether there were qualitative differences between the types of findings made in the animated view and in the small-multiples. As our approach was based on grounded theory [Laz10], we did not have pre-formed hypotheses. Instead, we started from the analysis of the findings and the interaction logs developing a well-grounded theory from these data. To achieve this goal we identified coding categories, performed manual coding of the findings and carefully analyzed their distribution across the categories. In the rest of this section we discuss this process in detail.

4.1. Coding

As the findings concerned flows of people between geographic locations changing over time, we chose "geographic scope" and "temporal scope" as the main properties for the coding. Here are the definitions of the properties we used:

- **Temporal scope** - The time span the finding refers to.
- **Geographic scope** - The extent of the geographic entities mentioned in the finding.
- **Validity** - Whether the statement of the finding can be interpreted as a valid finding.

First, we applied a top-down approach and tried to identify the coding categories for each of the properties before the coding. Then, manual coding was performed with the predefined categories by the three authors of this paper. Each of the authors had to go through the whole list of findings and assign the property categories by choosing one of the predefined values listed for each of them. After that we calculated inter-annotator agreement rates to ensure the reliability of the coding. They were as follows:

Property	Initial agreement	Final agreement
Temporal scope	0.675	(1.0)
Geo scope	0.811	1.0
Validity	0.888	1.0

The initial agreement rates for the "temporal scope" was obviously way too low. Hence, we developed and used a simple web-based tool which helped us to find and resolve the disagreements by seeing the answers given by each of us and negotiating (see Fig. 4). It helped us to improve all the agreement rates, but still, the "temporal scope" agreement rate was not satisfactory.

The approach of choosing predefined categories and then categorizing the findings according to them led to ambiguous coding in many cases. Hence, after several failed efforts to improve the agreement rates we decided to apply a bottom-up approach instead and manually grouped the findings without predefined categories. The approach we used for this was based on card sorting [SW04]. We placed all the findings written on small cards on a virtual desktop and kept arranging and grouping them on the screen until they finally formed meaningful categories (see Fig. 5). We did not calculate the final agreement rates for the two properties we categorized this way, because we performed this final categorization collaboratively (therefore, the final agreement for "temporal scope" is put in parenthesis in the table above).

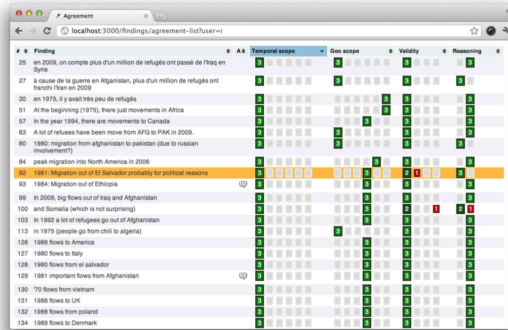

Figure 4: *Resolving disagreements during manual coding of the findings. Each row corresponds to a finding, each column with squares to a property. The square positions represent different classes which the could be chosen for each property. A green square shows an agreement indicating the number of those of us who agreed on a class, a red square is shown when there was a disagreement.*

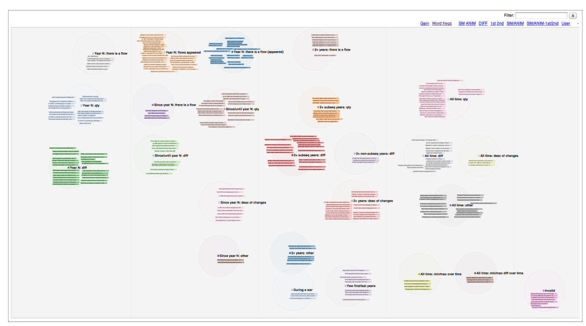

Figure 5: *Our web-based tool which we collaboratively used for establishing the coding categories by manually grouping the findings. Each finding is a small label which can be drag-and-dropped between categories. Categories are not predefined, but can be added and removed during the process. Here the categories are arranged in columns by their temporal scope, so that "one year" findings are placed in the leftmost column and "all time" findings in the rightmost column.*

The categories which we finally used for the "temporal scope" were as follows (the bottom-up approach):

- **One year** - Describes what was happening in one specific year (e.g. "1994 important flow from Rwanda to Congo").

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

- **Until or since** - Describes a pattern which was apparent for a time period before or after a specific year (e.g. "in 1979 movements from or to Vietnam started").
- **Interval** - Describes what was happening in a time span of several years.
- **All time** - Applies to the whole time period for which the data was available (e.g. "Migration involves increasingly more countries over time").

And for the "geographic scope" (the top-down approach):

- **Country** - Describes flows specifying only the country in which they originate or which they have as their destination, not both (e.g. "Large flows from Italy in 1992").
- **Country - Country** - Describes a flow between two specific countries (e.g. "Large flow from Russia to Italy in 1992").
- **Region** - Describes flows originating in a specific region or having the region as their destination.
- **Region - Country** - Describes flows between a country and a region.
- **Region - Region** - Describes flows between two regions.
- **Global** - Describes a global (geographically) pattern (e.g. "When going far, countries near the water are more popular destinations").

The values of the "reasoning" and "validity" properties were just "yes" or "no". In the end, the categories we came up with turned out to be useful for achieving our goal: pinpointing the differences between the types of findings made in the animated view and the small-multiples.

4.2. Results

In all the sessions with 16 users we collected 285 findings (17.8 findings per user on average with stdev of 4.65). There were 8 findings which were not formulated clearly enough, so that it would be possible to interpret them. They were marked as "invalid" and were not considered anymore. Out of the valid findings 173 were made in the first round (ANIM: 86, SM: 87), and 104 in the second round (ANIM: 55, SM: 49).

4.2.1. Main experiment

In the main experiment we only compared the types of the findings made in the first round. Concerning the temporal scope (Fig. 6), we observed that more findings of the types "one year" and "until or since" were made in ANIM, and more findings of the type "all time" were made in SM.

Notably, 93% of "one-year" findings in SM were made in the "difference view" (see 3.1). It was apparently too difficult to see the differences between subsequent years in the original view in SM. In contrast to that only 44% of "one-year" findings were made in the "difference view" in ANIM.

Concerning the geographic scope (Fig. 7), in the first round the subjects made more local observations ("country-country") in ANIM and more global observations ("region", or "global") in SM.

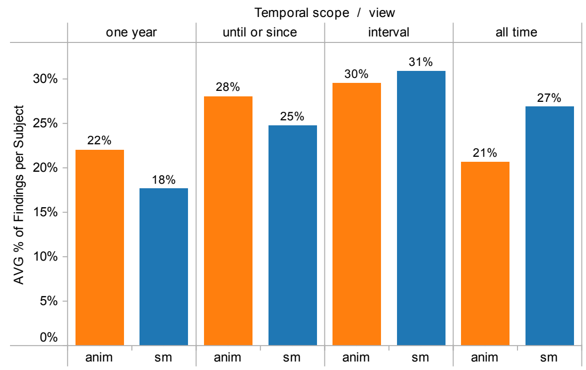

Figure 6: *Temporal scope of the findings made in the first round in each of the views.*

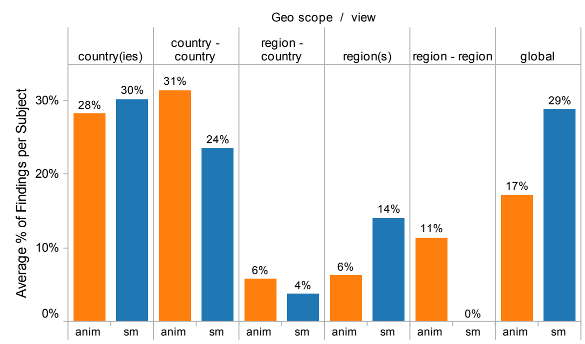

Figure 7: *Geographic scope of the findings made in the first round in each of the views.*

4.2.2. Additional round

In the additional round we explored how different were the types of observations the subjects made after switching from one view to the other.

Fig. 8 illustrates the distribution of the findings by their geographic scope when switching between the views. When switching from ANIM to SM the proportion of "country" findings decreased while the proportion of "global" findings increased. When switching from SM to ANIM the proportion of "country" findings increased, while the proportion of "global" stayed the same.

We observed a similar effect with the temporal scope (Fig. 9). After switching from ANIM to SM the proportion of "one year" and "until or since" findings decreased and "interval" and "all time" increased. Switching from SM to ANIM had mostly an opposite effect: proportion of the "until or since" findings greatly increased, whereas "interval" and "all time" decreased.

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

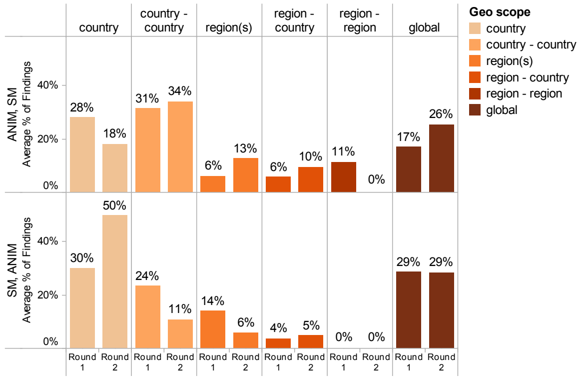

Figure 8: *Comparison of the geographic scopes of the findings made in the first and the second round depending on the subject group (i.e. the order in which the subjects were using the views).*

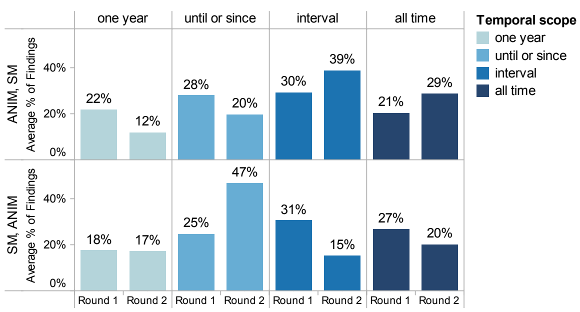

Figure 9: *Comparison of the temporal scopes of the findings made in the first and the second round depending on the subject group (i.e. the order in which the subjects were using the views).*

4.3. User feedback

In addition to the findings we analyzed the questionnaire submissions made by the study participants. Asked about their overall impressions of the two conditions, the participants favored ANIM: 15 of the 16 participants rated ANIM as "good" or "very good" and 10 of them rated SM in the same way. 44% of the subjects found ANIM easier to use, 13% SM, the rest had no preference. Further, we asked the participants to provide feedback concerning the strengths and weaknesses of each view. Several of them mentioned having one large view as the main advantage of ANIM over SM. It was also much easier for them to identify appearing or disappearing flows in ANIM and compare subsequent years. SM was credited with giving an overview over the whole dataset at once, providing better support for making quick comparisons and finding differences between non-subsequent years. The main weaknesses of SM mentioned by the study participants were that the views were too small and that there were too many of them. Thus, despite having a good overview in SM, it was difficult to focus on single elements and see how they were changing over time.

5. Discussion

5.1. On making findings

The results of the study show that alternative visualization techniques can generate or promote different types of findings and that switching from one view to the other might actually accentuate this effect. These results have a number of implications we discuss in the following.

- **Animation should be preferred for sudden change detection tasks.** Our results corroborate the outcome of the study on cluster detection by Griffin et al [GMH*06], which showed that certain changes might be better perceived when using animation. The higher level of small temporal scope findings and the users' feedback we received suggest that when the main task for a visualization is to be able to detect a sudden change, then animation is the preferable solution. This is also consistent with the results reported in [WB04] and in [APP11] for the node/edge appearance tasks.
- **Using only a single technique might lead to the loss of findings.** The complementarity of the two techniques is evident from the analysis of the findings and reinforced by the feedback we received from the participants. One corollary is that if only one technique is used important findings might be lost. Animation allows for higher resolution and easier detection of sudden changes, small-multiples reduces the load on short memory and allows for comparisons across many or arbitrary years. One possible solution is to provide both techniques in the same environment. Another would be to find a way of integrating them which would allow to overcome their limitations.
- **Switching between the views can have a beneficial effect on producing findings.** Our results suggest that switching from one view to another can lead to a boost in the number of produced findings of certain types. Thus, switching between the views can be explicitly used as a way to promote insights. Research on coordinated and multiple-view visualizations, e.g. [KERC09], does also show the usefulness of working with different representations. As well as the past work on investigative analysis which suggests that such an approach can help to avoid bias in judgments [RJH99]. A systematic analysis of the effect of switching between views can be an interesting line of research to pursue in the future.

5.2. Interaction patterns and the use of animation

People use different interaction patterns, i.e. they use different sequences of interaction techniques to produce a similar finding, and generally stick to their initial strategy and repeatedly use the same sequence to make new findings (see Fig. 10). We also noticed that the study participants tended

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

to rate findings requiring more time and interactions as more important and those which were easier to make as less important.

Figure 10: *Gantt chart showing a timeline of the findings (large bars) made by the 8 participants who used ANIM in the 1st round. Along with the findings we see the history of the animation-related actions (red and thin green bars) and the "change-year" slider action (thicker green bars) which the participants used to make these findings.*

In our analysis of the interaction logs we examined in particular how people use and perceive animation. First of all, we observed a disparity between what people like and what research suggests. Similarly to other studies on the use of animation, subjective user feedback is very favorable of animation; even when performance measures do not support it. Animation has definitely a special appeal on people which cannot be neglected and should be studied in more depth.

- People use the "change year" slider a lot (see Fig. 10), instead of playing the animation, and find it very helpful. Thus, providing animation controls without support for direct manipulation is clearly suboptimal.
- People use different strategies and might get stuck in a single one (see Fig. 10). Some use only the slider to control the animation manually, others only use the play/start button, some use it only at the beginning, and some combine the two. Users should be instructed on using more strategies.
- Additional research is needed to make interaction with animation smoother and more productive.

When comparing the interaction logs of ANIM and SM, we observed that people interact much more with the view in ANIM than in SM, not only because the ANIM view provides more interaction capabilities with the time slider. For instance, the highlight action was used about 12 thousand times with ANIM, and 6 thousand times for SM for about the same number of findings produced in both views in the first round (many of these highlight actions were triggered unintentionally, though, by just moving the mouse around). We made a similar observation for the panning action (542 times with ANIM, 293 times with SM). Zoom-in and zoom-out were used roughly equally numbers of times in both views. The total number of interactions is roughly two times larger with ANIM than with SM. Our interpretation of this is that ANIM favors interactions to observe local patterns and to detect sudden changes in time. On the other hand, SM favors reflection and requires less interactions to come up with findings concerning longer time periods.

6. Conclusion

In this study we analyzed the use of animation and small-multiples for exploring temporal changes in spatial interactions.

We observed that with animation the study participants made more findings concerning geographically local events or changes between subsequent years (especially, events in which flows appeared or disappeared in a specific year). With small-multiples more findings concerning longer time periods were made.

Besides, our results suggest that that switching from one view to another might have beneficial effects in terms of covering a larger spectrum of types of observations made. Thus, developing a smooth mechanism for integrating the two views in one exploration tool presents a great opportunity for future research.

Finally, we observed that different people used different sequences of interaction techniques to make similar findings, and that they often stuck to one strategy once they had learned how to make findings of a specific type. Visualization designers must take this observation into account to develop tools which not only support different types of findings, but support them equally well.

One important limitation is that, because of the qualitative nature of the study, the results were not statistically validated, and therefore, they have to be taken as suggestions for future research. Despite that, some of the results were confirmed by the feedback from the users or were consistent with previous research. In the future, a formal quantitative study must be performed for obtaining more generalizable results. The findings of the present study could be turned into questions which users would have to find answers to, thus, allowing to quantitatively compare different visualizations by measuring the users' performance.

Some of the tools which we presented in this paper are available at `http://bit.ly/flowmap-changes`.

7. Acknowledgments

The authors wish to thank all the people who participated in the study for their time and valuable feedback.

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.

References

[APP11] ARCHAMBAULT D., PURCHASE H., PINAUD B.: Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics 17* (Apr. 2011), 539–552. 1, 2, 8

[BBBL11] BOYANDIN I., BERTINI E., BAK P., LALANNE D.: Flowstrates: An approach for visual exploration of temporal Origin-Destination data. *Computer Graphics Forum 30* (June 2011), 971–980. 1

[BEW95] BECKER R., EICK S., WILKS A.: Visualizing network data. *IEEE Transactions on Visualization and Computer Graphics 1*, 1 (Mar. 1995), 16–28. 2

[Cha06] CHARMAZ K.: *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Sage Publications Ltd, 2006. 2

[ED06] ELLIS G., DIX A.: An explorative analysis of user evaluation studies in information visualisation. In *Proc. of the BELIV'06 workshop on BEyond time and errors: novel evaluation methods for information visualization* (2006), BELIV'06, pp. 1–7. 2

[FQ11] FARRUGIA M., QUIGLEY A.: Effective temporal graph layout: A comparative study of animation versus static display methods. *Information Visualization 10*, 1 (2011), 47–64. 1, 2

[FRA*08] FABRIKANT S. I., REBICH-HESPANHA S., ANDRIENKO N., ANDRIENKO G., MONTELLO D. R.: Novel method to measure inference affordance in static Small-Multiple map displays representing dynamic processes. *Cartographic Journal, The 45* (Aug. 2008), 201–215. 1, 2

[GMH*06] GRIFFIN A. L., MACEACHREN A. M., HARDISTY F., STEINER E., LI B.: A comparison of animated maps with static Small-Multiple maps for visually identifying Space-Time clusters. *Annals of the Association of American Geographers 96* (Dec. 2006), 740–753. 1, 2, 8

[KERC09] KEEFE D. F., EWERT M., RIBARSKY W., CHANG R.: Interactive coordinated multiple-view visualization of biomechanical motion data. *IEEE Transactions on Visualization and Computer Graphics (IEEE Visualization 2009) 15*, 6 (2009), 1383–1390. 8

[Laz10] LAZAR J.: *Research methods in human-computer interaction*. Wiley, Chichester West Sussex U.K., 2010. 6

[MBHP98] MACEACHREN A., BOSCOE F., HAUG D., PICKLE L.: Geographic visualization: designing manipulable maps for exploring temporally varying georeferenced statistics. IEEE Comput. Soc, pp. 87–94,. 2

[Nor06] NORTH C.: Toward measuring visualization insight. *IEEE Computer Graphics and Applications 26* (May 2006), 6–9. 2

[PS08] PURCHASE H., SAMRA A.: Extremes are better: Investigating mental map preservation in dynamic graphs. In *Diagrams 2008. Fifth International Conference on the Theory and Application of Diagrams* (2008), LNAI, Springer Verlag. Session 2.: Diagram Aesthetics and Layout (joint with VL/HCC). 2

[RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1325–1332. 1, 2

[RJH99] RICHARDS J. HEUER J.: *Psychology of Intelligence Analysis*. Central Intelligence Agency, 1999. 8

[SND05] SARAIYA P., NORTH C., DUCA K.: An Insight-Based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics 11* (July 2005), 443–456. 2

[SP08] SAFFREY P., PURCHASE H.: The "mental map" versus "static aesthetic" compromise in dynamic graphs: a user study. In *Proceedings of the ninth conference on Australasian user interface - Volume 76* (Darlinghurst, Australia, Australia, 2008), AUIC '08, Australian Computer Society, Inc., p. 85âĂŞ93. 2

[SSKY04] SLOCUM T., SLUTER R., KESSLER F., YODER S.: A qualitative evaluation of MapTime, a program for exploring spatiotemporal point data. *Cartographica: The International Journal for Geographic Information and Geovisualization 39* (Sept. 2004), 43–68. 1, 2

[SW04] SPENCER D., WARFEL T.: Card sorting: a definitive guide. *Document available online on 03.11.2011* (2004). 6

[TL96] THOMPSON W., LAVIN S.: Automatic generation of animated migration maps. *Cartographica: The International Journal for Geographic Information and Geovisualization 33*, 2 (June 1996), 17–28. 2

[Tob87] TOBLER W.: Experiments in migration mapping by computer. *The American Cartographer 14*, 2 (1987), 155–163. 1

[WB04] WARE C., BOBROW R.: Motion to support rapid interactive queries on node–link diagrams. *ACM Transactions on Applied Perception 1* (July 2004), 3–18. 2, 8

[YKSJ08] YI J. S., KANG Y.-A., STASKO J. T., JACKO J. A.: Understanding and characterizing insights. In *Proceedings of the 2008 conference on BEyond time and errors novel evaLuation methods for Information Visualization - BELIV '08* (Florence, Italy, 2008), p. 1. 2

© 2012 The Author(s)
Journal compilation © 2012 The Eurographics Association and Blackwell Publishing Ltd.