Workflow: Annotated pdfs, Tracked changes

PROOF COVER SHEET

Journal acronym: TCAG

Author(s): Tomasz Opach, S Popelka, J Dolezalova and J. K Rød

Article title: Star and polyline glyphs in a grid plot and on a map display: which perform better?

Article no: 1364169

Enclosures: 1) Query sheet
2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.

Please limit changes at this stage to the correction of errors. You should not make trivial changes, improve prose style, add new material, or delete existing material at this stage. You may be charged if your corrections are excessive (we would not expect corrections to exceed 30 changes).

For detailed guidance on how to check your proofs, please paste this address into a new browser window: http://journalauthors.tandf.co.uk/production/checkingproofs.asp

Your PDF proof file has been enabled so that you can comment on the proof directly using Adobe Acrobat. If you wish to do this, please save the file to your hard disk first. For further information on marking corrections using Acrobat, please paste this address into a new browser window: http://journalauthors.tandf.co.uk/production/acrobat.asp

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1		Tomasz	Opach	
2		S	Popelka	
3		J	Dolezalova	
4		J. K	Rød	

Queries are marked in the margins of the proofs, and you can also click the hyperlinks below.

Content changes made during copy-editing are shown as tracked changes. Inserted text is in red font and revisions have a red indicator ⅄. Changes can also be viewed using the list comments function. To correct the proofs, you should insert or delete text following the instructions below, but **do not add comments to the existing tracked changes.**

AUTHOR QUERIES

General points:

1. **Permissions:** You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. Please see http://journalauthors.tandf.co.uk/permissions/usingThirdPartyMaterial.asp.
2. **Third-party content:** If there is third-party content in your article, please check that the rightsholder details for re-use are shown correctly.
3. **Affiliation:** The corresponding author is responsible for ensuring that address and email details are correct for all the co-authors. Affiliations given in the article should be the affiliation at the time the research was conducted. Please see http://journalauthors.tandf.co.uk/preparation/writing.asp.
4. **Funding:** Was your research for this article funded by a funding agency? If so, please insert 'This work was supported by <insert the name of the funding agency in full>', followed by the grant number in square brackets '[grant number xxxx]'.
5. **Supplemental data and underlying research materials:** Do you wish to include the location of the underlying research materials (e.g. data, samples or models) for your article? If so, please insert this sentence before the reference section: 'The underlying research materials for this article can be accessed at <full link> / description of location [author to complete]'. If your article includes supplemental data, the link will also be provided in this paragraph. See <http://journalauthors.tandf.co.uk/preparation/multimedia.asp> for further explanation of supplemental data and underlying research materials.

AQ1 Please note that the ORCID for Popelka S has been created from information provided through CATS. Please correct if this is inaccurate.

AQ2 Please note that the ORCID for Dolezalova J has been created from information provided through CATS. Please correct if this is inaccurate.

AQ3 Please note that the ORCID for Rød J. K has been created from information provided through CATS. Please correct if this is inaccurate.

AQ4 Given name should be fullout as per journal style.

AQ5 The spelling of "Ho et al., 2015" has been changed to match the entry in the references list. Please provide revisions if this is incorrect.

AQ6 Please check if the edits made in the sentence "From the analysis…." retain its intended meaning and correct if necessary.

AQ7 The disclosure statement has been inserted. Please correct if this is inaccurate.

AQ8 The CrossRef database (www.crossref.org/) has been used to validate the references. Mismatches between the original manuscript and CrossRef are tracked in red font. Please provide a revision if the change is incorrect. Do not comment on correct changes.

AQ9 Please provide missing publisher location for reference "Dewey, 2007".

How to make corrections to your proofs using Adobe Acrobat/Reader

Taylor & Francis offers you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can mark up the proof directly using Adobe Acrobat/Reader. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

1. Save the file to your hard disk.

2. Check which version of Adobe Acrobat/Reader you have on your computer. You can do this by clicking on the "Help" tab, and then "About".

If Adobe Reader is not installed, you can get the latest version free from http://get.adobe.com/reader/.

3. If you have Adobe Acrobat/Reader 10 or a later version, click on the "Comment" link at the right-hand side to view the Comments pane.

4. You can then select any text and mark it up for deletion or replacement, or insert new text as needed. Please note that these will clearly be displayed in the Comments pane and secondary annotation is not needed to draw attention to your corrections. If you need to include new sections of text, it is also possible to add a comment to the proofs. To do this, use the Sticky Note tool in the task bar. Please also see our FAQs here: http://journalauthors.tandf.co.uk/production/index.asp.

5. Make sure that you save the file when you close the document before uploading it to CATS using the "Upload File" button on the online correction form. If you have more than one file, please zip them together and then upload the zip file.

If you prefer, you can make your corrections using the CATS online correction form.

Troubleshooting

Acrobat help: http://helpx.adobe.com/acrobat.html
Reader help: http://helpx.adobe.com/reader.html

Please note that full user guides for earlier versions of these programs are available from the Adobe Help pages by clicking on the link "Previous versions" under the "Help and tutorials" heading from the relevant link above. Commenting functionality is available from Adobe Reader 8.0 onwards and from Adobe Acrobat 7.0 onwards.

Firefox users: Firefox's inbuilt PDF Viewer is set to the default; please see the following for instructions on how to use this and download the PDF to your hard drive:
http://support.mozilla.org/en-US/kb/view-pdf-files-firefox-without-downloading-them#w_using-a-pdf-reader-plugin

CARTOGRAPHY AND GEOGRAPHIC INFORMATION SCIENCE, 2017
https://doi.org/10.1080/15230406.2017.1364169

Taylor & Francis
Taylor & Francis Group

ARTICLE

Star and polyline glyphs in a grid plot and on a map display: which perform better?

AQ1 AQ2 AQ3 AQ4

Tomasz Opach[a], S Popelka [b], J Dolezalova [b] and J. K Rød [a]

[a]Department of Geography, Norwegian University of Science and Technology, Trondheim, Norway; [b]Department of Geoinformatics, Palacký University, Olomouc, Czech Republic

ABSTRACT

Glyphs are small geometric shapes that in geovisualization are often used to represent multi-dimensional spatial data. The aim of this study is to investigate the effectiveness of their two types – star and polyline glyphs, as they can encode the same message and can provide similar functionality. Thus, if the two glyph types are similar and can be used for the same data, the question arises as to which of them better facilitates various user tasks. To address this question, an empirical study of 26 individual users is conducted to investigate differences in user performance for polyline and star glyphs shown either in a grid plot or on a map display. In this study, a task-based approach with eye-tracking is applied, as well as a subjective questionnaire and a psychological test of cognitive style. The finding is that polyline glyphs better facilitate tasks when datapoint values in glyphs are to be read, whereas star glyphs are better when a visual search among glyphs is to be done. Moreover, the results reveal that the map display works better than the grid plot. If star glyphs are to be used, the key (legend) needs to be better incorporated into a visual interface.

ARTICLE HISTORY
Received 9 February 2017
Accepted 2 August 2017

KEYWORDS
Eye-tracking; glyphs; polyline glyphs; star glyphs; user study

Introduction

Glyphs are a commonly used visualization technique that provides an overview of a dataset by showing its items as separate and simplified graphical entities in the form of juxtaposed small plots or charts without any labels (Borgo et al., 2013; Gribov, Unwin, & Hofmann, 2006; Ünlü & Malik, 2011). Hence, despite the limited explanations (labeling), juxtaposition enables access to particular data items, which often is unfeasible in the case of superimposition (Opach & Rød, 2017), such as when using parallel coordinates. Glyphs can differ greatly with regard to their form (Ward, 2002, 2008). One of the most commonly used glyph types is star glyphs. Although they have been implemented in many geovisualization environments (Gribov et al., 2006; Takatsuka & Gahegan, 2002), their use is impeded by polar coordinates in which visual scanning is more time-consuming and more error-prone than reading vertical and horizontal axes (Goldberg & Helfman, 2011). Therefore, Opach and Rød (2017) propose the use of polyline glyphs that resemble polylines from parallel coordinates as an alternative to star glyphs. The two glyph types can encode the same message and can provide similar functionality. Thus, if star and polyline glyphs are so similar and can be used for the same data, the question arises as to which of them better facilitates various user tasks.

In this paper, we aim to contribute to the body of previous work by investigating the performance of a data display consisting of either star or polyline glyphs (Figure 1). Moreover, as glyphs are frequently used on map displays and grid plots such as tables or matrices, we investigate these two layout arrangements to see whether there are differences regarding their performance. Finally, we examine whether there are differences regarding user behavior (undertaken actions when interacting with a graphical interface) between those who use star glyphs and those who use polyline glyphs. If such differences exist, what lessons can be learned from an empirical study in which user behavior is investigated? Can any findings be of value to map-makers and practitioners of information visualization?

The study consists of a theoretical part and an empirical part, in which a task-based approach with eye-tracking is employed. Additionally, we use a subjective questionnaire and a psychological test to gain deeper insights into the behaviors and opinions of users of polyline and star glyphs. The paper is organized as follows. After the background section, in

CONTACT Tomasz Opach tomasz.opach@ntnu.no

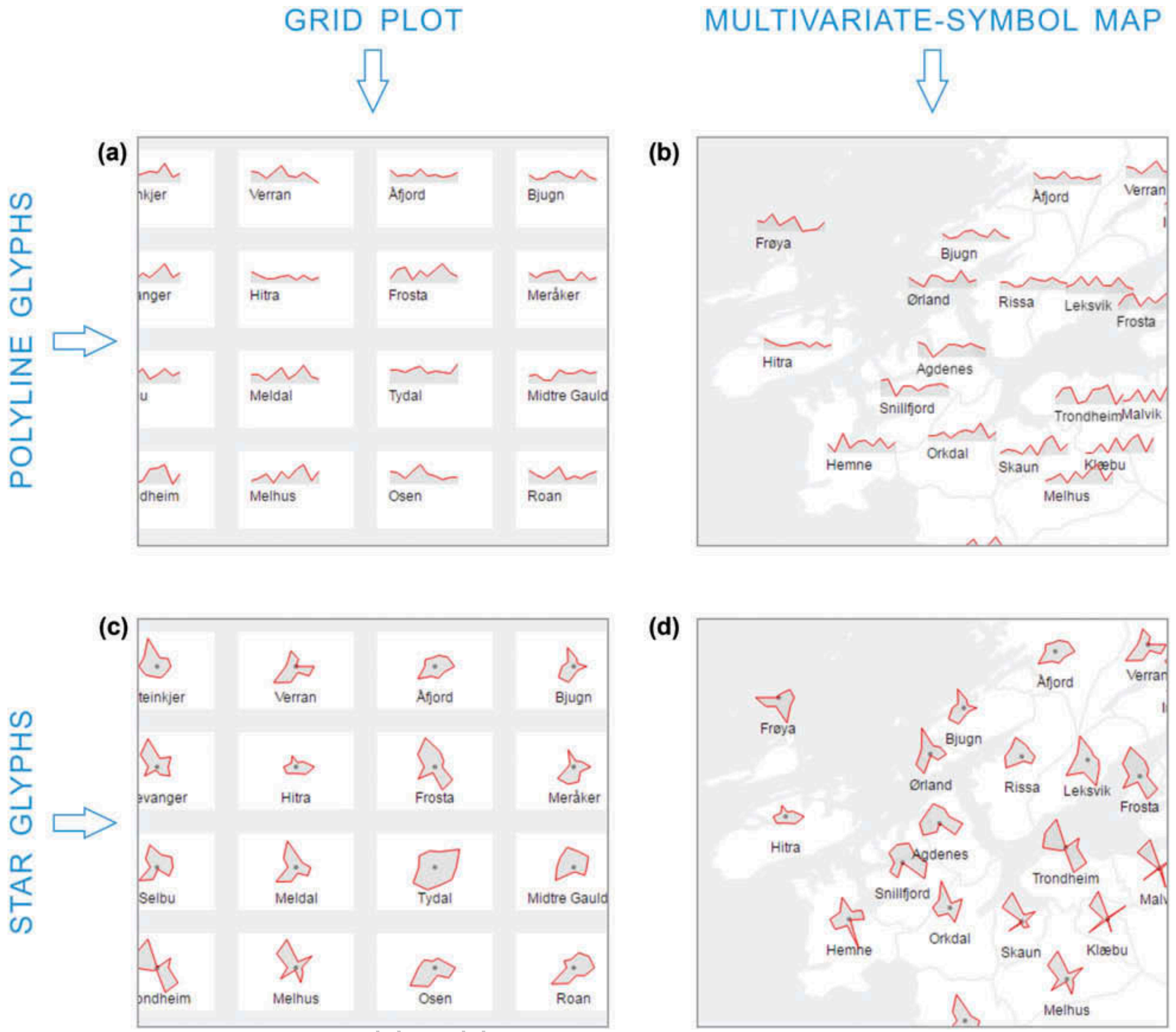

Figure 1. Polyline glyphs (a, b) and star glyphs (c, d) arranged as a grid plot (a, c) and as a multivariate symbol map (b, d).

which the state of the art in the glyph-based visualization is briefly discussed, we consider the advantages and weaknesses of star and polyline glyphs. Thereafter, we report the settings of our empirical study: its objectives, methods, and procedure. We then present and discuss the results, followed by our conclusions.

Background

Glyphs and multivariate data visualization

The growing role of information visualization in application areas such as information dashboards and business intelligence implies the need for a better understanding of various visualization techniques in general, and particularly techniques that, despite differing in form, can encode the same data. One such technique uses multivariate glyphs.

In order to encode hundreds of n-dimensional data items and show them in a limited space, the use of tiny graphical entities known as glyphs seems to be a sensible choice. Although there have been many studies of glyphs (Borgo et al., 2013), the areas in which they can be effectively used are still insufficiently studied (Ward, 2008). There are many ways in which glyphs can be used in information visualization in general and in geovisualization in particular. Glyphs seem to be suitable for visualization of multivariate vector fields (Forsberg, Chen, & Laidlaw, 2009). However, efforts needed to interpret such visualization may make glyph-based displays ineffective. Glyphs can also be embedded in a table view (Opach & Rød, 2013) or organized in a grid plot (Figure 1(a and c)) and, as part of coordinated and multiple views (CMVs), they can be dynamically linked with other visualization techniques, such as radar plots or parallel coordinates (Takatsuka & Gahegan, 2002). Additionally, glyphs can be superimposed on maps (Figure 1(b and d)) to form multivariate-symbol maps.

Glyphs in geovisualization: multivariate-symbol maps and grid plots

Glyphs have long been used within cartography and geovisualization. However, in these domains, glyphs are not a certain mapping technique to be studied. As Slocum, McMaster, Kessler, and Howard (2010, p. 337) claim, glyphs are “multivariate point symbols used to represent nonrelated attributes.” Glyphs can also appear in the form of uncommon shapes such as peculiar Chernoff faces (Chernoff, 1973); such faces can be arranged as cartograms (Dorling, 1995). In general, glyphs on maps can simply be called symbols or multivariate symbols in the case in which they encode more than one variable.

Glyphs can encode geographic objects directly onto maps, and then the mapping technique can be attributed a specific name, such as a bar-chart map (Figure 2(a)), a radar-plot map, or a multivariate-symbol map. The potential of such mapping techniques has been well known in cartography for decades, since cartographers have long been concerned with ways to visualize multivariate or time-series data (Arnberger, 1977; Bertin, 1967; Ratajski, 1989; Slocum et al., 2010), and simplified multivariate symbols have long been used in thematic maps. For instance, Ostrowski and Uhorczak (1972) introduced cartotypograms as a mapping technique in which n-dimensional (typically four-dimensional) star-plots (typograms) without coordinates are used to indicate types [typical multivariate signatures, see Figure 2(b)]. This approach enables the differences between plot shapes and sizes to be clearly visible. Therefore, users are able to distinguish types among data items. Currently, the technique is known as star map, and its implementations, can be found in many geovisualization tools, such as the GeoViz Toolkit, an application derived from GeoVISTA Studio (Takatsuka & Gahegan, 2002).

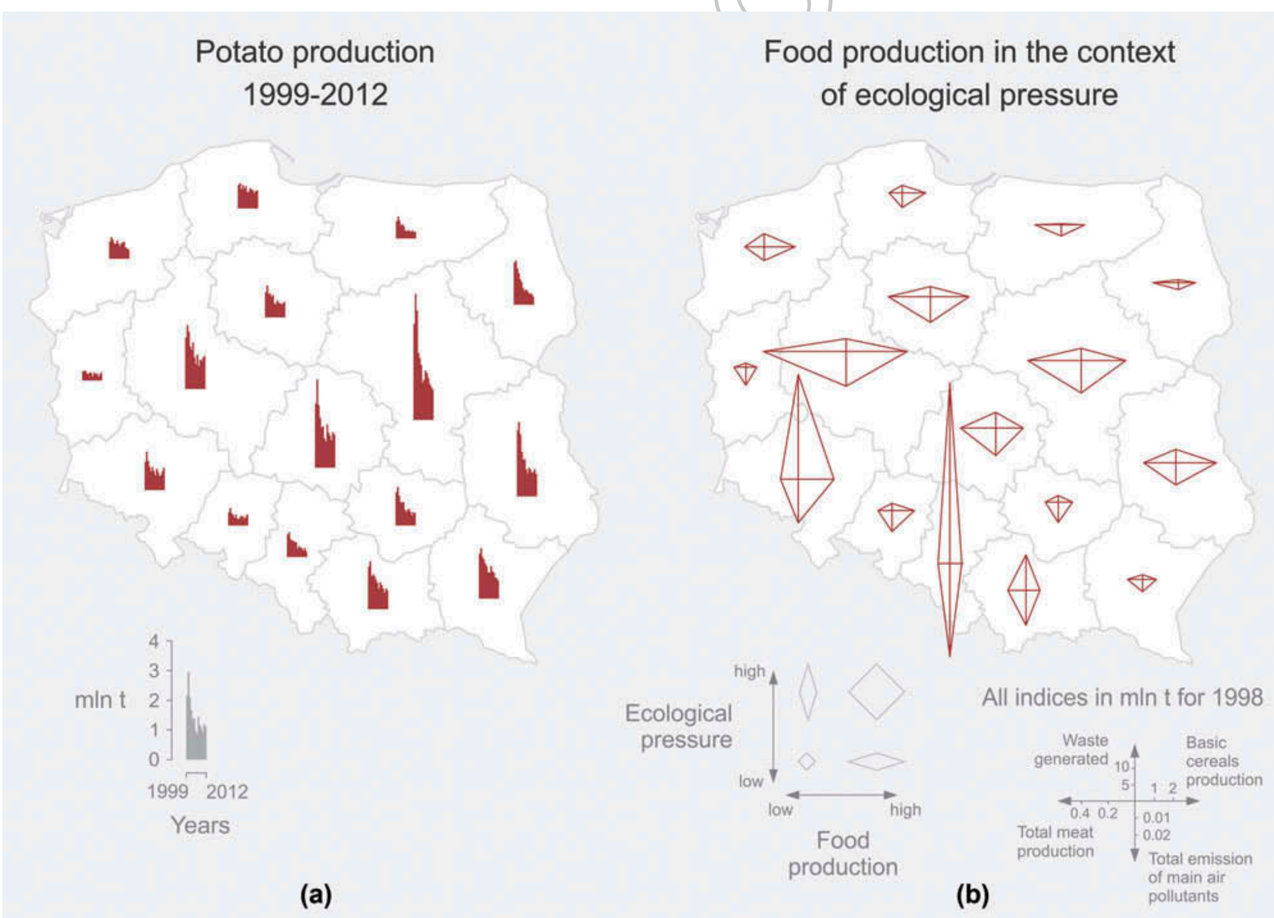

Figure 2. Two glyph maps of Poland: (a) the bar-chart map shows time-series data for potato production and (b) the cartotypogram map (star map) shows multivariate data on food-production types in the context of ecological pressure.

Recently, glyphs have been widely used in interactive map displays, but they are referred to differently, depending on the purpose of the displays [e.g., "utility symbols" (Andrienko, Andrienko, & Jankowski, 2003)]. However, in the broad geovisualization context, glyphs can be used more extensively than only as multivariate symbols on thematic map displays. They can be used differently, especially when they form part of CMV tools, whereby various visualization techniques are dynamically linked in order to facilitate information exploration and knowledge construction (Andrienko et al., 2002). In such cases, for more analytical purposes, glyphs can be grouped together and dynamically linked with other displays. They can be shown as scatterplot points (Chung et al., 2015; Gribov et al., 2006; Ünlü & Malik, 2011) or placed below or next to each other in a small multiple or a grid plot (matrix), in which "information slices are positioned within the eye span, so that viewers make comparisons at a glance–uninterrupted visual reasoning" (Tufte, 1990, p. 67). In this way, the similarities and differences between glyphs are likely to be identified more efficiently (Klippel, Hardisty, Li, & Weaver, 2009; Ward, 2008).

Star and polyline glyphs: advantages and weaknesses

Star glyphs are one of the most commonly used glyph types (Gribov et al., 2006; Ünlü & Malik, 2011). They show data items as graphic entities embedded in the polar coordinates context. Star glyphs can be thought of as a parallel coordinate plot in polar coordinates (Gribov et al., 2006; Ünlü & Malik, 2011). Therefore, in the CMV tools, in certain conditions, parallel coordinates can sometimes be replaced with star glyphs. It happens since visual attention can be shifted from parallel coordinates to star glyphs without adjusting visual reasoning (Klippel et al., 2009). Such adjusting is not needed if parallel coordinates are to be replaced with polyline glyphs (Opach & Rød, 2017) – graphical entities that resemble the polylines from parallel coordinates. Thus, glyphs can serve as an independent visualization component or they can support parallel coordinates.

While it can be assumed that polyline glyphs and star glyphs can be used interchangeably, since these two visualization techniques can encode the same data "payload," there is an essential difference in the way such encoded data are shown in the two glyph types. Since parallel coordinates are aligned in the polyline glyphs, users may find it easier to get datapoint values than with star glyphs, in which polar coordinates are used. Goldberg and Helfman's (2011) eye-tracking study revealed that visual scanning can be done more quickly along vertical and horizontal axes than circular scanning along rings, and the latter method is error-prone and not reliable. However, since star glyphs are more centered and compacted than polyline glyphs, they might perform better in tasks involving either similar or distinctive glyphs. However, such statements must be empirically tested. This raises the question as to whether star or polyline glyphs can be used for the same data, and if so, which glyph type performs better?

An example of a study in which line glyphs (similar to polyline glyphs) and star glyphs are compared has been published by Fuchs, Fischer, Mansmann, Bertini, and Isenberg (2013). Their study reveals that line glyphs are a good choice for tasks in which peak and trend detection is to be done when examining time-series data. By contrast, radial encoding of time in star glyphs works better if one has to find a particular temporal location. These findings contradict to some extent Goldberg and Helfman (2011) claim that linear graphs can better support the dimension-finding phase, since their linearly aligned dimensions support searches better than radial graphs. Lee, Reilly, and Butavicius (2003) compare four visualization techniques, including Chernoff faces and star glyphs, in terms of their usefulness in user tasks. Their study reveals that both types of glyph visualizations lead to slow, inaccurate answers being given, with a low degree confidence. In a more recent study, Chung et al. (2015) conclude that various interactive functions, such as glyph sorting, that support user exploration in glyph visualization can significantly enhance user performance.

Eye-tracking for evaluation of glyphs

Eye-tracking plays a particular role in empirical research on information visualization. According to Goldberg and Helfman (2011), to date, this technique has been underutilized as a method for understanding how individuals make use of information graphics. Although there has been a rapid increase in eye-tracking studies in geovisualization, studies of glyph-based visualization have been sparse and seldom. For example, Ho, Yey, Lai, Lin, and Cherng (2015) use the method to examine various 2D visualizations of flow, including a glyph-based technique, and Golebiowska, Opach, and Rød (2017) use eye-tracking to investigate a CMV interface consisting of a choropleth map, a parallel coordinate plot, and a table with polyline glyphs.

AQ5

Eye-tracking is not necessary to examine the performance of glyph-based displays. Nevertheless, it enables a better sense of the differences between the visual behavior of different users – what and how long they look at. Similar empirical data can be obtained with other methods, such as mouse-tracking or recording task execution time. However, without insight into eye-movement data, it is unfeasible to analyze, for example, how often participants use a key, how many times they look at particular map symbols (revisit them), or which glyphs attract their attention most. Therefore, to gain a comprehensive insight into how polyline and star glyphs work, and then consider their advantages and weaknesses, we conduct an empirical study with eye-tracking as our main empirical technique.

Empirical study

Objectives

Since there might be differences in user performance for polyline and star glyphs shown either in a grid plot or on a map display, we examine four layout modes (M1–4). These are presented in Figure 1. We address three research questions:

(1) RQ1: Are there specific user tasks for which one of the two investigated glyph types outperforms the other?
(2) RQ2: Are there specific user tasks in which glyphs shown by means of one of the two investigated display types (grid plot or map) work better than those arranged in the other display type?
(3) RQ3: Are there certain user skills that influence task-solving strategies with star or polyline glyphs?

Participants

Total of 26 individuals (15 males and 11 females, average age 23 years) attend the study voluntarily. They are not paid any compensation for the attendance. All of them are either bachelor or master's students taking the geoinformatics course at Palacký University. Their skills and knowledge are considered representative of target users. Students from the first year of the bachelor's study program are excluded because they have not had any training in GIScience, and therefore their performance is likely to be worse than those who have had this training.

Study material

As study material, we design a single-page web application where regular web browser can be used to run the tool. We use costless JavaScript APIs such as D3.js and jQuery Sparklines to develop the tool. Labels and comments in the tool's interface are in Czech, with the exception of the key (legend) which is in English. The interface features four layout modes (see Figure 1) in which glyphs are either polyline glyphs (M1-PolyGrid, M4-PolyMap) or star glyphs (M2-StarMap, M3-StarGrid), and in which glyphs are either regularly distributed in a grid (M1-PolyGrid, M3-StarGrid) or geographically distributed on a map (M2-StarMap, M4-PolyMap). Additionally, for the purpose of the empirical study, the tool has an opening dialog box in which the four layout modes are grouped as follows:

- Variant 1, in which a grid plot with polyline glyphs (M1-PolyGrid) is followed by a map with star glyphs (M2-StarMap);
- Variant 2, which displays the supplementary modes [i.e. a grid plot with star glyphs (M3-StarGrid) followed by a map with polyline glyphs (M4-PolyMap)].

In all modes, the main panel is accompanied by both a task panel and a key that explains how data are encoded in glyphs (see Figure 3).

For the tool's data content, we use 10 socioeconomic indicators (variables) describing 48 municipalities in the counties of Sør-Trøndelag and Nord-Trøndelag in central Norway. The participants are Czech or Slovak and such data content is unknown to them. Hence, prior knowledge of the visualized variables cannot influence the participants' answers.

Methods

We gain scientific evidence through individual user sessions in which we ask participants to:

- use the tool to solve six user tasks during an eye-tracking session,
- fill in a personal questionnaire and a subjective questionnaire on glyphs,
- perform a psychological test of the cognitive style of users.

We combine the methods above to get a comprehensive insight into participant choices and behavior.

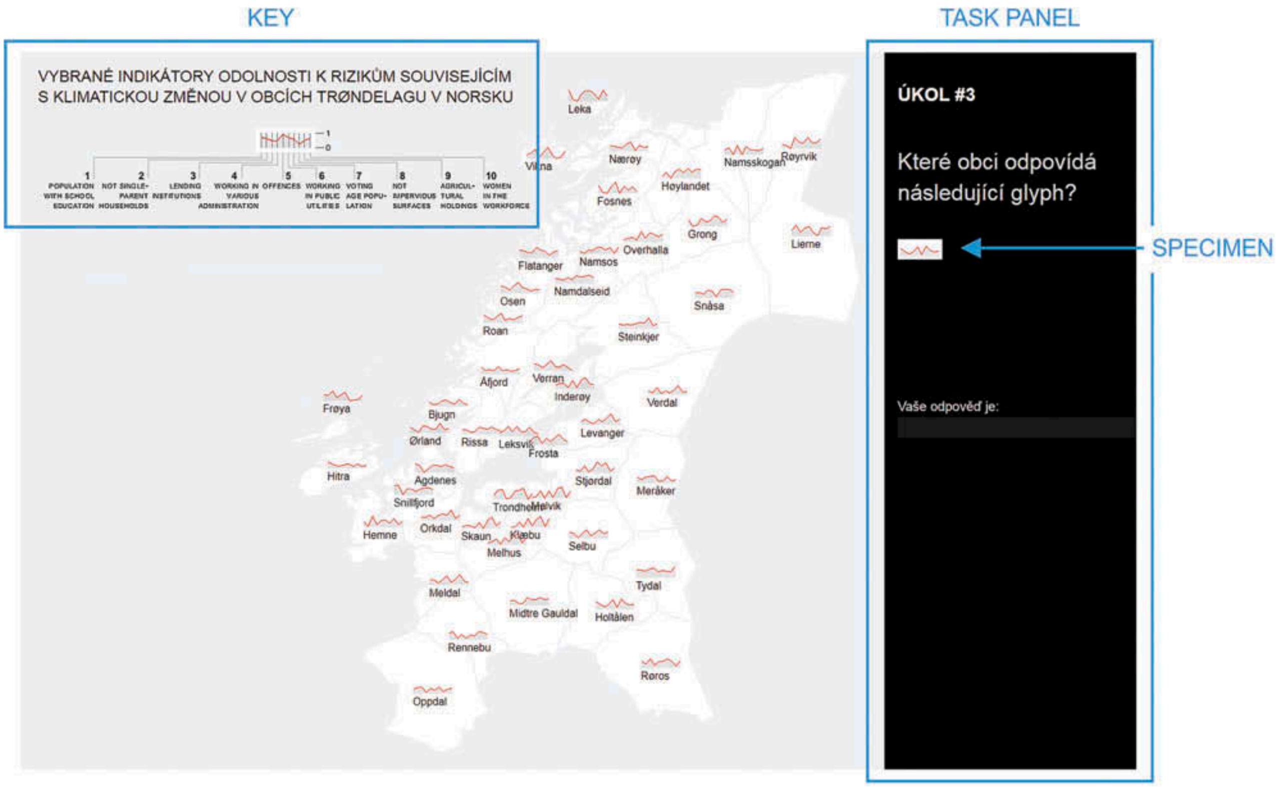

Figure 3. M4-PolyMap (polyline glyphs on a map display) – one of the four layout modes used in the empirical study. It is presented here with the task T3-FindGlyph in which participants search for the same glyph as the one (specimen) shown in the task panel.

User tasks

While participants execute the six user tasks, T1–6 (Table 1), we record their eye-movements, oral comments, the screen, and their task answers. Regarding the tasks, these are designed to reveal how users interact with polyline and star glyphs. T1-EstimVal is used to examine how users derive values from either polyline or star glyphs. T2-IdenGlyph is complementary to T1 as participants are expected to search for a glyph that encodes a certain variable score. In turn, in T3-FindGlyph, users search for the same glyph as the one shown in the task panel (see Figure 3). We use this task to verify which of the four layout modes performs better regarding participants' visual searches. In T4-SimilGlyphs, participants compare all glyphs between each other to find the two most similar glyphs. Regarding T5-DistinctGlyphs, a common task in visual analytics is to find the most distinctive cases among graphic entities, and we therefore request users to do the same. Again, we investigate whether this task can be more effectively accomplished with polyline glyphs or with star glyphs. In the final task, T6-CompArea, participants search for a compact area of three glyphs that are similar to each other.

Two questionnaires: personal and subjective about the glyph-based visualization

While the personal questionnaire will give information about participants' age and gender, the subjective questionnaire will inform about their subjective preferences regarding the two glyph types. The questionnaire consists of five questions (Q1–5) that participants answer on a 7-point scale. The questions are as follows:

- Q1 concerns overall feeling about the usability of glyphs.
- Q2 is about the aesthetics of glyphs.
- In Q3, participants are asked to specify a glyph type that is suitable for reading datapoint values; this question concerns T1-EstimVal and T2-IdenGlyph.
- Q4 concerns a comparison of glyphs between themselves and refers to T3-FindGlyph (search for a glyph), T4-SimilGlyphs (find similar glyphs), and T5-DistinctGlyphs (find distinctive glyphs).
- In Q5, participants state which glyph type works best for getting an overview of all glyphs; it refers to T6-CompArea, in which participants select a compact area of three similar glyphs.

Table 1. Six tasks used in the empirical study.

ID	Short name	Task question	Purpose of task
T1-EstimVal	Glyph value estimation	Use the key shown in the upper left corner and estimate the datapoint values of the selected glyph on variables 5 and 10	Examine how users derive values from polyline and star glyphs
T2-IdenGlyph	Search for a glyph featuring a concrete score on selected variable	Find a glyph with the score 1 on variable 5	Complementary to Task 1, examine how users search for a glyph that features a certain datapoint value
T3-FindGlyph	Search for a glyph	Search for the same glyph as the one shown in the task panel	Examine how effectively users search for either a polyline or star glyph
T4-SimilGlyphs	Point out the two most similar glyphs	Point out the two most similar glyphs	Examine how effectively users search for two similar glyphs (polyline glyphs or star glyphs)
T5-DistinctGlyphs	Find two glyphs with the most distinctive cases	Find two glyphs with the most distinctive cases	Examine how effectively users search for distinctive glyphs (polyline glyphs or star glyphs)
T6-CompArea	Find a compact area of three glyphs that are similar to each other	Indicate a compact area that consists of three glyphs of similar shape	Examine whether user performance regarding finding similarities is better for polyline or star glyphs

The scale used in the questionnaire is designed so that the values toward the left-hand end reflect the user's preference for polyline glyphs, whereas the values toward the right-hand end reflect their preference for star glyphs; the middle value means no preference (neutral choice).

Psychological test of the cognitive style of users

The last test – Navon's cognitive style test (Navon, 1977) – is one of the most frequently used tests for measurement of the global-analytic dimension of cognitive processing (Brand & Johnson, 2014). This test is designed to reveal whether an individual's preferred cognitive style is holistic or analytic – the distinction that is one of the most common among people's cognitive styles (Dewey, 2007). According to Dewey, in the analytic thinking, individuals comprehend a system, first, by recognizing its particular parts, and second, by understanding how they constitute a larger scale pattern. Whereas in the holistic thinking, of primary importance to individuals is to first recognize large-scale patterns of a system, not its particular elements. The fact that this cognitive style could affect work with a map, at least in the subprocesses, is well documented (Kubíček et al., 2016). We use a compound letter test, an adaptation of Navon's hierarchical figures test (Navon, 1977) developed as part of the GEOKRIMA project (Šašinka, 2013). In this test, big numbers composed of small numbers are displayed, and participants are requested to recognize either small or big numbers. We use the Hypothesis software (Morong & Šašinka, 2014) to perform the test. Its output will help to determine the cognitive style of participants and will be used to compare the affiliation to these cognitive styles with the strategy of solving the tasks.

Equipment

We use the eye-tracker SMI RED 250. The eye-tracker is arranged in the Eye-tracking Laboratory of the Department of Geoinformatics at Palacký University, Olomouc, in the Czech Republic. The stimulus is displayed on a 24-in screen with a resolution of 1920 × 1200 pixels. Eye positions are recorded at a frequency of 250 Hz. The eye-tracker is supplemented with a web camera that records participants during the sessions. We do this because audio and video recording can help to reveal the possible cause of missing data, participants' reactions to the stimuli, and their comments on the tasks.

Procedure

Individual user sessions are arranged as two-phase user testing (Figure 4) with a minimum of 3 days between each phase. This is done to avoid a learning effect, whereby, during the "later" stage of testing, participants may use their knowledge acquired in the "earlier" stage. Without two-phase user testing, the results might be influenced by already gathered experience. Moreover, a variant assigned to participants in each phase will shift: The first participant starts with variant 1, the next participant uses variant 2, and so on. In the second phase, participants use the complementary variant. For example, if a participant uses variant 1 in the first phase, they use variant 2 in the second phase.

The test sessions are performed using SMI Experiment Center. We collect participants' answers, eye-movement data complemented by audio and video recording of participants, screen recording, and task completion time. Each session is organized as follows (see Figure 4). In its first phase, after the participants have been welcomed, an eye-tracker is calibrated for each of them. The maximal allowed

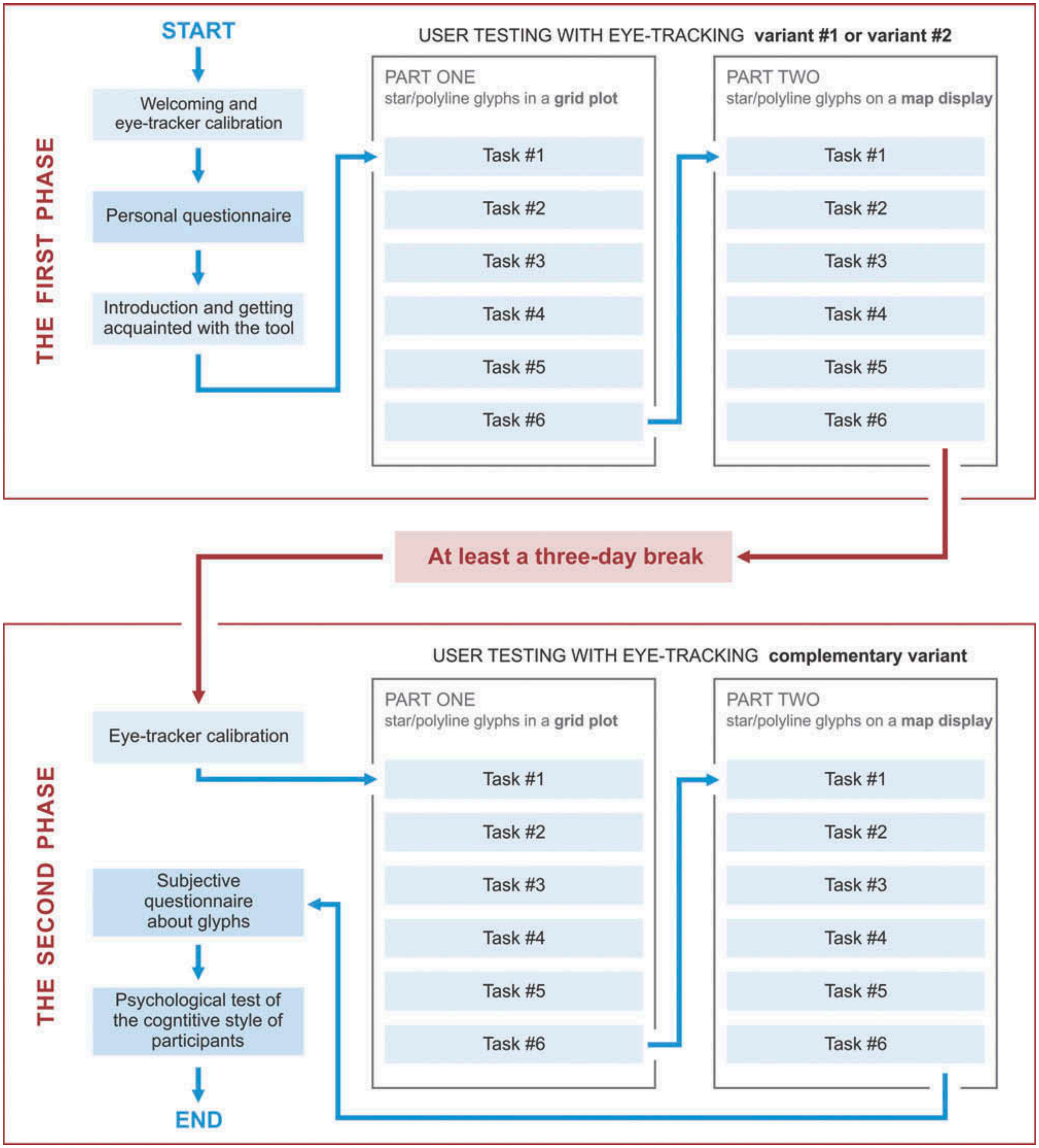

Figure 4. The design of the individual user session.

deviation is set as 1° of the visual angle. Next, each participant fills in the personal questionnaire. This is followed by a short instruction about glyph-based visualization and each participant is given 2 min to play around with the display. The testing consists of two parts. In both parts, participants solve six user tasks (see Table 1). However, in the first part, 48 glyphs constitute a grid plot, whereas in the second part, they constitute a map display with the coordinates of the municipalities they represent. In the second phase (a few days later), participants perform the same six tasks again, but with the complementary variant. Hence, within the entire session, each participant solves the six user tasks for each of the four layout modes. After the user testing in the second phase, participants fill in the subjective questionnaire and perform the psychological test on their cognitive style.

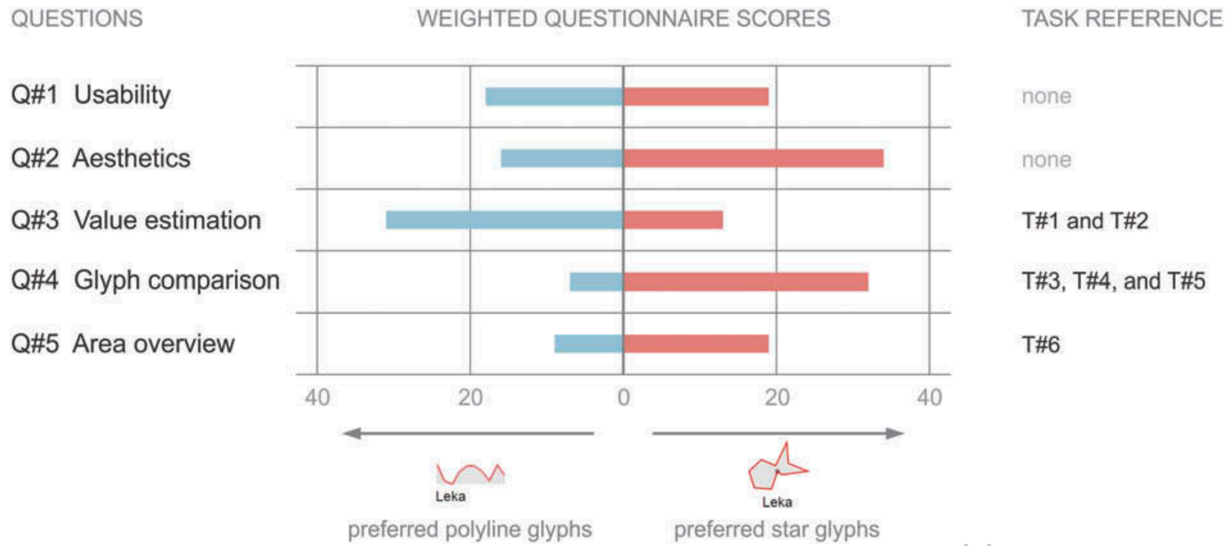

Figure 5. The outcomes of the subjective questionnaire about glyphs.

Data obtained

Eye-movement data are recorded as "screen recording" type of stimulus. Result of the recording is a separate video file obtained for each participant. Therefore, to analyze all recordings together, we use the Custom Trial Selector from the SMI BeGaze software to combine all videos by task. The custom trial is designed for each task as a screenshot, and then the corresponding part of each recording is assigned to it. In the analysis of eye-movement data, we use the I-DT algorithm for fixation detection (Salvucci & Goldberg, 2000). This algorithm is mostly used for low-frequency data (up to 250 Hz) and takes into account the close spatial proximity of the eye position points in the eye-movement trace. Threshold values in BeGaze are set to 80 ms for "duration threshold" and 50 pixels for "dispersion threshold," as these values are suggested as optimal (Popelka, 2014).

To analyze answer accuracy in T4–6, scores on the similarity measure for all possible pairs of glyphs need to be calculated. To do this, we use the SimUrb software (eyetracking.upol.cz/simurb) – a derivative of the ScanGraph software (Dolezalova & Popelka, 2016). It calculates the similarity measure for all pair of glyphs – the Euclidean distance in n-dimensional space. All statistical tests are executed in RStudio at 0.05 significance level.

Data analysis

The subjective questionnaire about glyphs

Since the questionnaire is presented at the end of the second phase (see Figure 4), all participants use all four layout modes (M1–4) ahead of the questionnaire. In its analysis, weighted scores are used to amplify higher ratings: The values leading to both ends of the scale have increasing weights, from 1 to 3 (the middle value, 0, is subtracted from the analysis). In all but Q3, star glyphs receive more points (Figure 5); it is not a surprise, because Q3 concerns value estimation (required in T1-EstimVal and T2-IdenGlyph), to which polyline glyphs are supposedly better tailored. Although the answers are almost balanced in Q1 about usability, from the aesthetics point of view (Q2), participants prefer star glyphs. Finally, more points are given to star glyphs in Q4 and Q5, about glyph comparison and area overview, respectively.

Eye-movement analysis: trial duration metric by task and layout mode

We analyze trial duration by task and layout mode. This metric shows how long it takes to solve a task. A quick look at the boxplots in Figure 6 will reveal that in the comparison of the grid plot and the map display (consisting of the same glyph types), all tasks are solved quicker if the map display is used. Statistically significant results of the Kruskal–Wallis test with the post hoc Nemenyi test are found for a number of configurations (marked with asterisks in Figure 6). However, statistically significant differences ($p = 0.02$) between polyline and star glyphs on the same display are found only for map display in T3-FindGlyph: It is faster to find a star glyph.

In turn, the Wilcoxon rank-sum test reveals statistically significant differences between the grid plot and the map display (regardless of the glyph type) for T1-EstimVal ($W = 1871$, $p < 0.001$), T4-SimilGlyphs ($W = 1766$, $p = 0.007$), and T6-CompArea ($W = 2195$,

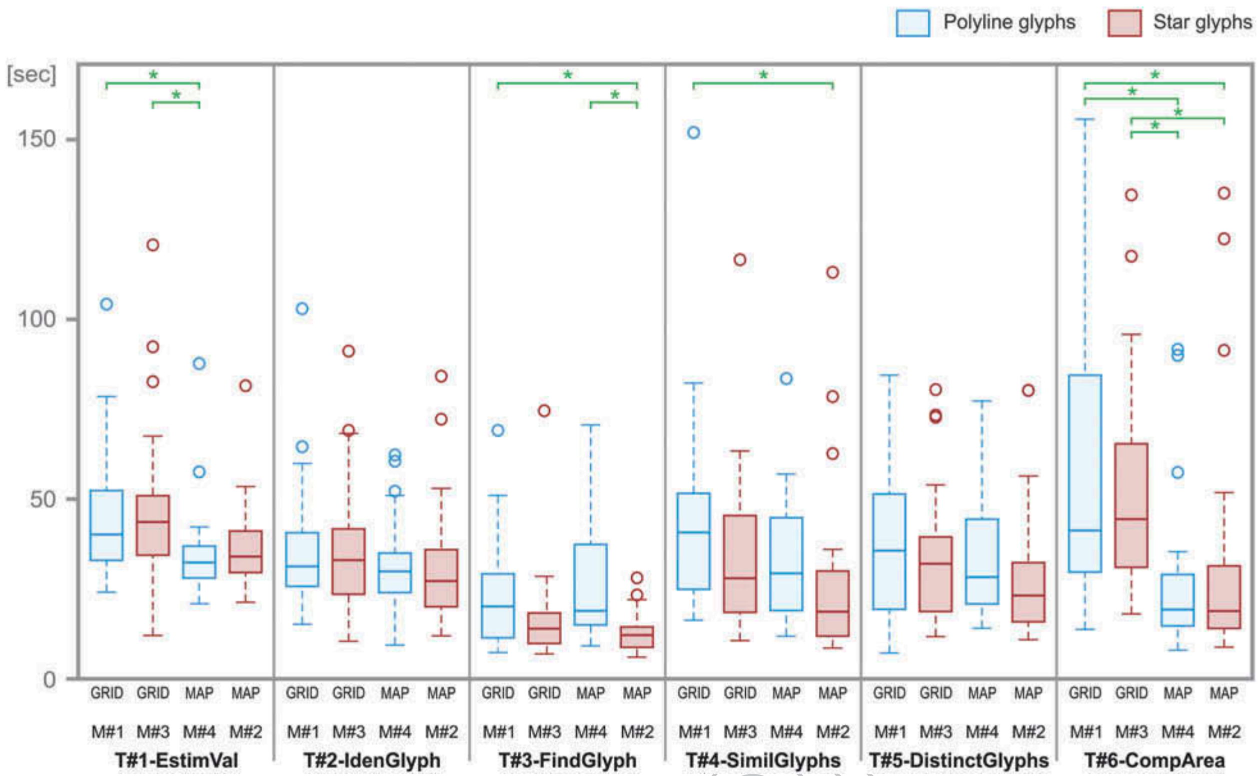

Figure 6. Trial duration metric by task and layout mode.

$p < 0.001$). In these three cases, the map display proves to be quicker. When it comes to the differences between polyline and star glyphs (regardless of the display type), the Wilcoxon test reveals statistically significant differences (star glyphs are quicker) for T3-FindGlyph ($W = 1927.5$, $p < 0.001$) and T4-SimilGlyphs ($W = 1770$, $p = 0.007$).

From the analysis above, it is apparent that trial duration metric is task dependent and in particular, tasks of higher values are observed for different configurations, for either polyline or star glyphs. From this reason, user behavior during task execution is further analyzed separately for subsequent tasks.

Eye-movement analysis: fixation counts in glyphs by task and layout mode

If fixation counts in glyphs are compared in various tasks and layout modes (see Figure 7), more numerous fixations occur in the last three tasks in general, and in T6-CompArea's grid plot in particular. These are caused by the intensive visual searches required for those tasks. The Kruskal–Wallis test with the post hoc Nemenyi test reveals statistically significant differences for a number of configurations (marked with asterisks in Figure 7). The significant differences are similar to those reported in the preceding section.

The glyphs that participants look at depend on their user tasks. For example, to read datapoint values encoded in a given glyph in T1-EstimVal, participants can either use the key ("absolute interpretation") or compare the glyph with other glyphs ("relative interpretation"). In T1-EstimVal, fixations are more numerous in the selected glyph and its neighborhood (the latter may be caused by the eye-tracker inaccuracy), because participants look at the nearest surroundings of the selected glyph and do not look at the bottom part of the grid plot (Figure 8). This may mean that participants do not tend to compare the glyph's shape with other glyphs. In more distant glyphs, only a few fixations occur and they probably accompany the gaze movements from the selected glyph to the key.

In the remaining tasks, fixations are scattered around the whole stimuli, as participants search for glyphs. In T2-IdenGlyph, such behavior can lead to glyph decoding executed through comparison with other glyphs ("relative interpretation"). Furthermore, in T2-IdenGlyph's grid plot, participants look mostly at the first two rows and finish solving the task just after localizing the first glyph fulfilling the requirement. In T3–6, the displays are also fully covered by fixations; hence, all glyphs attract attention. Nevertheless, in T4-SimilGlyphs, most fixations are recorded for the glyphs selected by participants, mostly for Klæbu and Malvik, particularly in the map display.

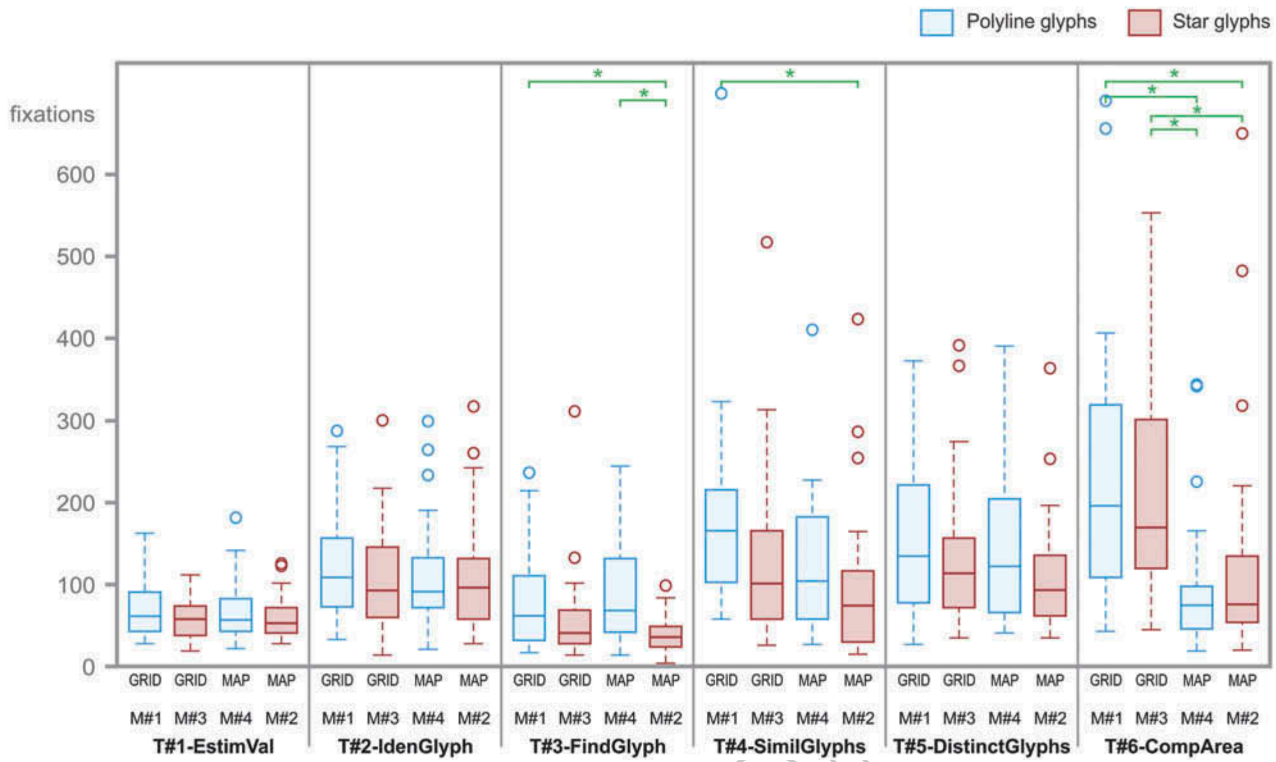

Figure 7. Fixation count in glyphs by task and layout mode.

Answer accuracy and visual behavior in particular tasks

T1: glyph value estimation

Answer accuracy. In T1-EstimVal, participants read datapoint values of two variables (5 and 10) in a marked glyph. To avoid a learning effect, there are two different glyphs marked in two subsequent layout modes used in the same test session: Leksvik in the grid plot and Selbu in the map display. To examine estimation accuracy, an average difference between estimated values and correct scores is calculated. The same as claimed in Q3 in the subjective questionnaire, better answer accuracy (lower average differences) is observed for polyline glyphs. However, the Kruskal–Wallis test with the post hoc Nemenyi test reveals no statistically significant differences in estimation accuracy between any pair of the four layout modes.

Areas of interest analysis. Users need a key (legend) to solve T1-EstimVal. We therefore examine how intensively the key is used in the four layout modes. The dwell time measure calculated for the areas of interest (AOIs) marked around various parts of the stimuli shows what portion (percentage) of the trial duration participants spend in particular AOIs. The key is used longer for star glyphs (medians of 9.5% and 12% for the grid plot in M3-StarGrid and the map display in M2-StarMap, respectively) than for polyline glyphs (8.5% for M1-PolyGrid and 8.6% for M4-PolyMap). However, the Kruskal–Wallis test with the post hoc Nemenyi test reveals no statistically significant difference. Slightly higher attention to the key when using star glyphs occurs in the revisits measure, which informs how many times participants revisit the AOI with the key during the trial duration. The numbers of revisits for star and polyline glyphs are, respectively, 4 and 3 (medians) in the grid plot, and 4 and 2 in the map display. Statistically significant difference is found between polyline and star glyphs (regardless the display type) using the Wilcoxon test ($p = 0.02$).

We also investigate the number of transitions between the AOI with the key and the AOI with either the grid plot or the map display. It turns out that participants look at the key more frequently if star glyphs are used: In the grid plot, participants move their visual attention from the main display to the key (total switches between the two AOIs) 146 times for star glyphs (M3-StarGrid), and 142 times for polyline glyphs (M1-PolyGrid). By contrast, in the map display, we again observe 146 such transitions for star glyphs (M2-StarMap), and only 127 for polyline glyphs.

T2: search for a glyph featuring a concrete score on selected variable

Answer accuracy. In T2-IdenGlyph, participants search for a glyph that leads to the value of 1 on

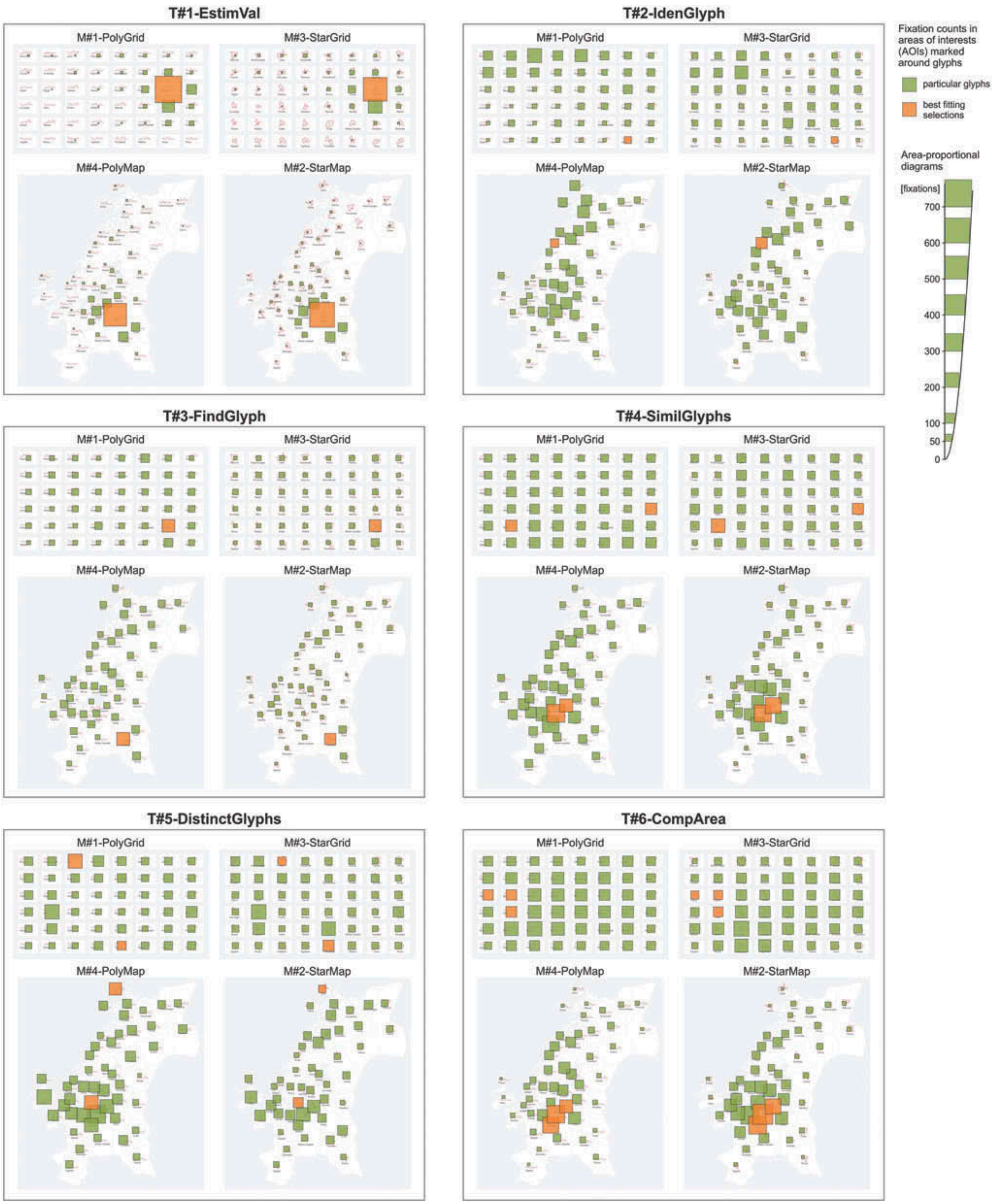

Figure 8. Fixations by task and layout mode.

variable 5. The difference between the chosen glyph's score on variable 5 and the value of 1 is calculated for each participant. The participants' answers for star glyphs are slightly better (this contradicts the responses to the subjective questionnaire (Q3) that polyline glyphs are more suitable than star glyphs for value estimation); however, the Kruskal–Wallis test reveals no statistically significant differences.

AOI analysis. Although participants also need the key to solve T2-IdenGlyph, they look at it a shorter time than in T1-EstimVal. They spend 9.05% of the trial duration looking at the key in T1, whereas in T2, the corresponding percentage is only 1.55% (3623 and 496 ms, respectively). This may be due to a learning effect. In T1, participants have already learned how to decode variables, and therefore they have less need for the key in T2.

We also analyze the visual behavior of participants with poor answer accuracy. In most cases of poor answers, participants use the map display with polyline glyphs. Furthermore, poor answers appear especially common for the glyphs for Stjørdal and Trondheim that score high on variable 6. This may be the reason why the participants mix it up with variable 5.

T3: search for a glyph

Answer accuracy. In T3-FindGlyph, participants search for the same glyph as the one – the specimen – shown in the task panel (see Figure 3). Although for both glyph types, the glyph for Holtålen municipality is used, the two glyph versions differ strongly with regard to shape.

The answer accuracy is high and all but three answers are correct. Three incorrect answers occur in the grid with polyline glyphs (M1-PolyGrid). This finding may confirm the results of the questionnaires, in which (Q3) star glyphs are claimed as more suitable for glyph comparison.

AOI analysis. Transitions between the display and the specimen reveal that participants move their attention to the specimen less frequently if a star glyph is found: 131 times (total switches between the two AOIs) when using the grid plot and 137 times for the map. For polyline glyphs, the corresponding number of times is 226 for the grid and 253 for the map. This may suggest that it is more difficult to remember the shape of a polyline glyph than of a star glyph since, in the latter case, participants do not need to bring back the shape of the specimen so often. This claim can be further backed up by the analysis of revisits of the specimen. The Wilcoxon rank-sum test reveals statistically significant differences ($W = 2000$, $p < 0.001$) between polyline glyphs and star glyphs for both the grid and the map (Figure 9). This means that the participants look at the specimen less frequently when searching for a star glyph. We thus interpret this that it is easier to remember a star glyph and that its specimen does not need to be checked so often.

We examine how quickly participants make the final decision after localizing the glyph in the display, and thus how certain they are. We also want to know what they do after localizing the glyph, how many times they revisit the specimen, and whether they check the remaining glyphs to be more confident. The analysis reveals that there is no difference between glyph and display types. Participants exhibit similar levels of confidence in all layout modes.

T4: point out two most similar glyphs

Answer accuracy. In T4-SimilGlyphs, participants point out the two most similar glyphs. To examine answer accuracy, the scores on the similarity measure need to be first calculated for all possible pairs of glyphs. To do this, we use the SimUrb tool. Then, we check the scores received for the pairs chosen by the participants. The highest similarity (0.89 in the range 0–1) features the pair Klæbu–Malvik, and this pair is selected 50 times in all 104 trials. The Kruskal–Wallis test reveals statistically significant differences ($p = 0.001$) between the accuracy of the answers given by those who use polyline and star glyphs in the map display. The test also returns statistically significant differences ($p = 0.001$) between polyline glyphs in the grid (M1-PolyGrid) and star glyphs on the map (M2-StarMap) (Figure 10(a)). In these cases, star glyphs perform better: Star glyphs selected by participants as pairs are more similar than pairs of polyline glyphs. This analysis correlates with the subjective questionnaire, in which star glyphs are claimed more suitable for comparisons of glyphs and hence for finding similar glyphs. Moreover, as expected from the trial duration analysis by layout mode (described earlier in this paper), the map gives better results than the grid plot for T4-SimilGlyphs. For these two displays (regardless of the glyph type), the Kruskal–Wallis test reveals a statistically significant difference ($p = 0.029$).

Scanpath length analysis. We use scanpath length (the length of the gaze trajectory) in the data analysis because, as Holmqvist et al. (2011) claim, it measures the efforts needed for visual search, and therefore it might reflect task complexity. In the comparison of the grid plot and the map display, the Kruskal–Wallis test reveals statistically significant differences ($p < 0.001$): Shorter scanpaths are recorded for the map display (Figure 10(b)). In turn, in the comparison of polyline and star glyphs, although shorter scanpaths are observed for the latter, the difference is not statistically significant ($W = 1569$, $p = 0.158$).

T5: find two glyphs with the most distinctive cases

Answer accuracy. T5-DistinctGlyphs is similar to T4-SimilGlyphs, except that participants in T5-DistinctGlyphs point out two the most

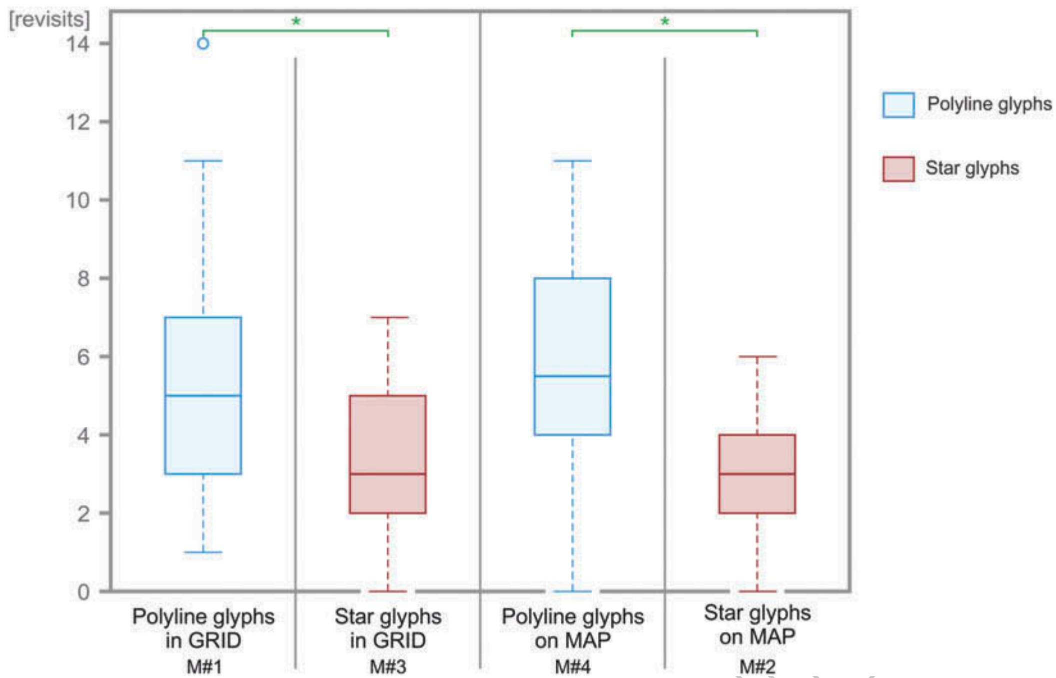

Figure 9. Revisits of the area of interest (AOI) marked around the specimen in the task panel.

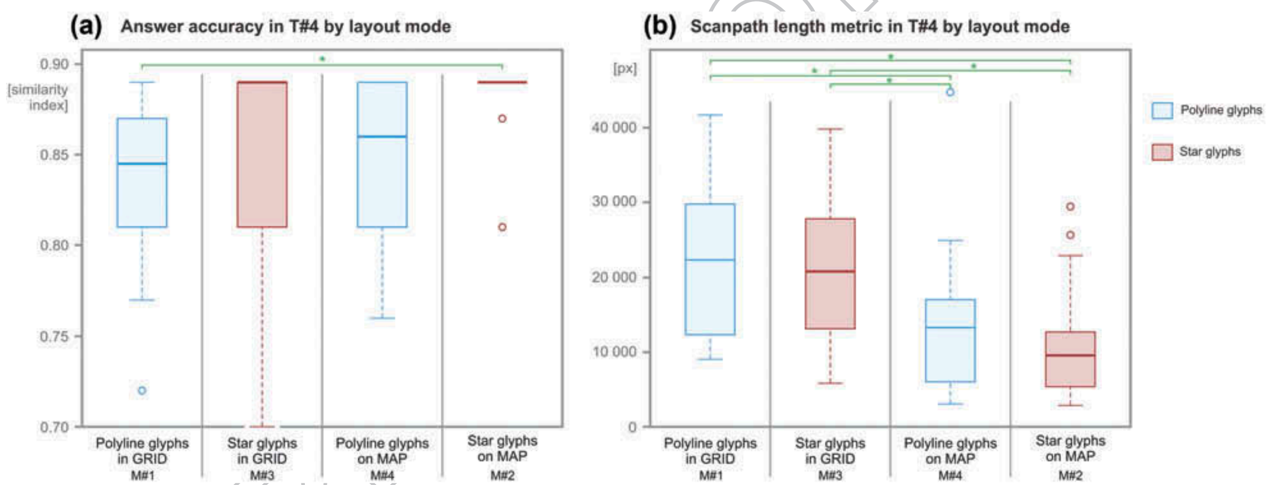

Figure 10. The outcomes of T4-SimilGlyphs (point out two most similar glyphs) by layout mode: (a) scores on the similarity measure calculated for the pair of glyphs selected by participants and (b) the scanpath length metric.

distinctive glyphs from all glyphs presented in the display. As in T4-SimilGlyphs, the SimUrb tool is used in T5-DistinctGlyphs for data analysis. However, we calculate an average similarity for each of two selected glyphs and all remaining glyphs. The lowest average similarity (i.e. most distinctive glyphs) is calculated for Leka (0.55) and Trondheim (0.59). These two glyphs are selected 30 and 8 times, respectively, in all 104 trials. However, a combination of both glyph types is selected only once, on the map display with polyline glyphs.

The Wilcoxon test reveals a statistically significant difference ($W = 993.5$, $p = 0.018$) between the answer accuracy of those who use star glyphs and those who use polyline glyphs (regardless of the display type). “Better” answers (lower average similarity of two selected glyphs) are given by those who use polyline glyphs. We also test the differences between polyline and star glyphs separately for the grid and the map. The Wilcoxon test reveals a statistically significant difference ($W = 398$, $p = 0.028$) between polyline and star glyphs only for the grid. In this analysis, too, lower average similarity is observed for polyline glyphs. The latter observation along with the previous one contradicts the observations from T4-SimilGlyphs in which star glyphs are found better for finding similar (and thus different) glyphs.

Scanpath length analysis. In the scanpath length analysis, the Kruskal–Wallis test shows a clear tendency toward significance ($p = 0.055$) only for the difference between polyline glyphs in the grid and star glyphs on the map. This may indicate that although in T5-DistinctGlyphs, polyline glyphs facilitate better answer accuracy, T5-DistinctGlyphs can be more "easily" solved if star glyphs are used, particularly on the map display. Further research is however needed to better elaborate this.

T6: find a compact area of three glyphs that are similar to each other

Answer accuracy. T6-CompArea is similar to T4-SimilGlyphs. However, participants search for a compact area of the three most similar glyphs. Although the purpose of this task is clear for the map display, since glyph positions are dependent upon the municipalities they represent, the purpose of this task may be questioned if the grid plot is to be used where glyph positions are random. We use T6-CompArea for the grid plot, as well to ensure consistency in the testing.

Although it is stated that selected glyphs must be adjacent, some participants select glyphs (5 times in 104 trials) that are far away. As we suppose, they do this because they do not read the task question sufficiently carefully. We calculate average similarity scores for all selected glyph triplets and use them as the answer accuracy measure. The highest similarity score for three adjacent glyphs in the map display is found for the triplet Klæbu–Malvik–Melhus (an average of 0.87). This triplet is selected 19 times of 104 trials. The Kruskal–Wallis test reveals statistically significant differences [shown marked with asterisks in Figure 11(a)] between four layout modes. Moreover, the Wilcoxon test reveals a statistically significant difference ($W = 371.5$, $p < 0.001$) between two display types: the map display gives better results. Regarding the difference between polyline and star glyphs, as visible in Figure 11(a), the average similarity of selected star glyphs seems to be higher than the similarity of polyline glyphs. However, the Wilcoxon test reveals no statistically significant difference ($W = 1136$, $p = 0.157$) between two glyph types.

Scanpath length analysis. For the scanpath length metric, the Kruskal–Wallis test gives statistically significant differences for the same combinations as in T4-SimilGlyphs. In this case, too, there are no statistically significant differences between two glyph types in the same display ($p = 0.976$ for the grid plot, $p = 0.795$ for the map display). Longer scanpaths are observed for the grid plot than for the map display and slightly longer scanpaths are observed for star glyphs than for polyline glyphs (Figure 11(b)).

The test of the participants' cognitive style: analytic versus holistic users

In the cognitive style test, 32 images are displayed that show big numbers (hereafter referred to as BNs) composed of small numbers (SNs). Participants are asked

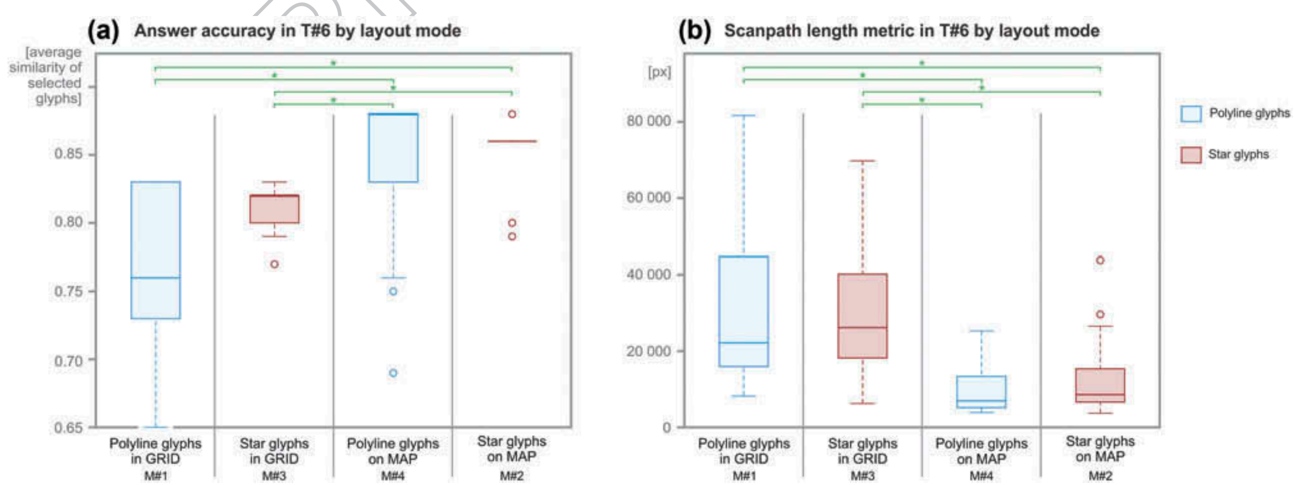

Figure 11. The outcomes of T6-CompArea (find a compact area of three glyphs that are similar to each other) by layout mode: (a) average scores on the similarity measures calculated for triplets of glyphs selected by participants and (b) the scanpath length metric.

Table 2. The distinction between analytic and holistic participants based on the outcomes of the test of their cognitive style.

ID	Ratio[a]	Difference	Cognitive style
P04	1.470	0.470	Analytic users
P10	1.470	0.470	
P09	1.234	0.234	
P17	1.161	0.161	
P11	1.158	0.158	
P14	1.107	0.107	
P24	1.079	0.079	Neutral users
P19	1.056	0.056	
P15	1.035	0.035	
P07	1.033	0.033	
P12	1.023	0.023	
P08	0.995	−0.005	
P03	0.947	−0.053	
P06	0.939	−0.061	
P22	0.928	−0.072	
P23	0.916	−0.084	
P05	0.914	−0.086	
P18	0.911	−0.089	
P02	0.902	−0.098	
P16	0.870	−0.130	Holistic users
P13	0.861	−0.139	
P20	0.841	−0.159	
P21	0.838	−0.162	
P01	0.837	−0.163	
P26	0.818	−0.182	
P25	0.769	−0.231	

[a]The quotient of the BL time and the SL time.

about either BNs or SNs, in 16 images for each. From the Hypothesis software used to run the test, we obtain the average times for SNs and BNs for each participant. Then, we calculate averages for SNs and BNs for the whole sample ($n = 26$). The times obtained for BNs are 13% shorter than those obtained for SNs. Therefore, to eliminate the global precedence effect (Navon, 1977), we equalize both data samples by multiplying the SN times by coefficient 0.87 (as it reduces them by 13%). Finally, we calculate the quotient of the BN time and the SN time; the value of 1 represents the most balanced (analytic vs. holistic) participants. Participants who feature the smallest deviation from 1 (who deviate to less than 10%) are labeled neutral (Table 2). Remaining participants are labeled as either analytics (six participants with better performance for SNs) or holistics (seven participants with better performance for BNs).

We examine which participants – analytic or holistic – perform better regarding answer accuracy, trial duration, and fixation frequency. For answer accuracy, holistics perform better only in T4-SimilGlyphs. In T2-IdenGlyph and T3-FindGlyph, the results are similar for both groups, whereas in T1-EstimVal, 5, and 6, analytics have better results. However, the Wilcoxon test does not reveal any statistically significant differences for any of these results. With regard to trial duration, holistics are faster in almost all tasks (Figure 12(a)), as expected. The only exception is T1-EstimVal, in which participants do not need to search for any glyph but only estimate two datapoint values in a selected glyph. Again, no differences between analytic and holistic participants are statistically significant.

The fixation frequency measure shows the number of fixations per second. In T1-EstimVal and T2-IdenGlyph, fixation frequency is higher for analytics, but in the remaining tasks, holistics have higher fixation frequency (Figure 12(b)). Furthermore, in T5-DistinctGlyphs and T6-CompArea, the Wilcoxon rank-sum test reveals statistically significant differences: $W = 201$, $p = 0.013$ and $W = 221.5$, $p = 0.035$, respectively.

Results and discussion

In this section, we relate the findings of the empirical part to the research questions (RQ1–3). We then summarize our assessment of which glyph and display type is most effective (answer accuracy) and most efficient (trial duration and other eye-tracking metrics). Moreover, we present our conclusions about whether users find star glyphs or polyline glyphs more satisfying. We refer to these three aspects – effectiveness (task completion by users), efficiency (task in time), and satisfaction (responded by users in terms of experience) – because according to ISO (9241-11), they constitute usability in a given context of use (users, tasks, equipment, and environments).

RQ1: polyline versus star glyphs

In general, star glyphs perform better than polyline glyphs. In most cases, their use leads to better answer accuracy and shorter task accomplishment time (Table 3). "Strong" results are obtained especially for T3–5, in which participants need to compare glyphs in a display. In these three tasks, star glyphs receive better scores in the subjective questionnaire (Q3). They also feature fewer revisits in T3-FindGlyph (i.e. better efficiency) in which the participant needs to compare glyphs with the specimen. The analysis reveals no statistically significant differences for any of the results from T2-IdenGlyph and T6-CompArea. Nevertheless, polyline glyphs receive better ratings in the subjective questionnaire in T2-IdenGlyph, whereas star glyphs receive better ratings in T6-CompArea.

Polyline glyphs receive better scores in T1-EstimVal, in which participants read the datapoint values of two variables. In this task, better answer accuracy and

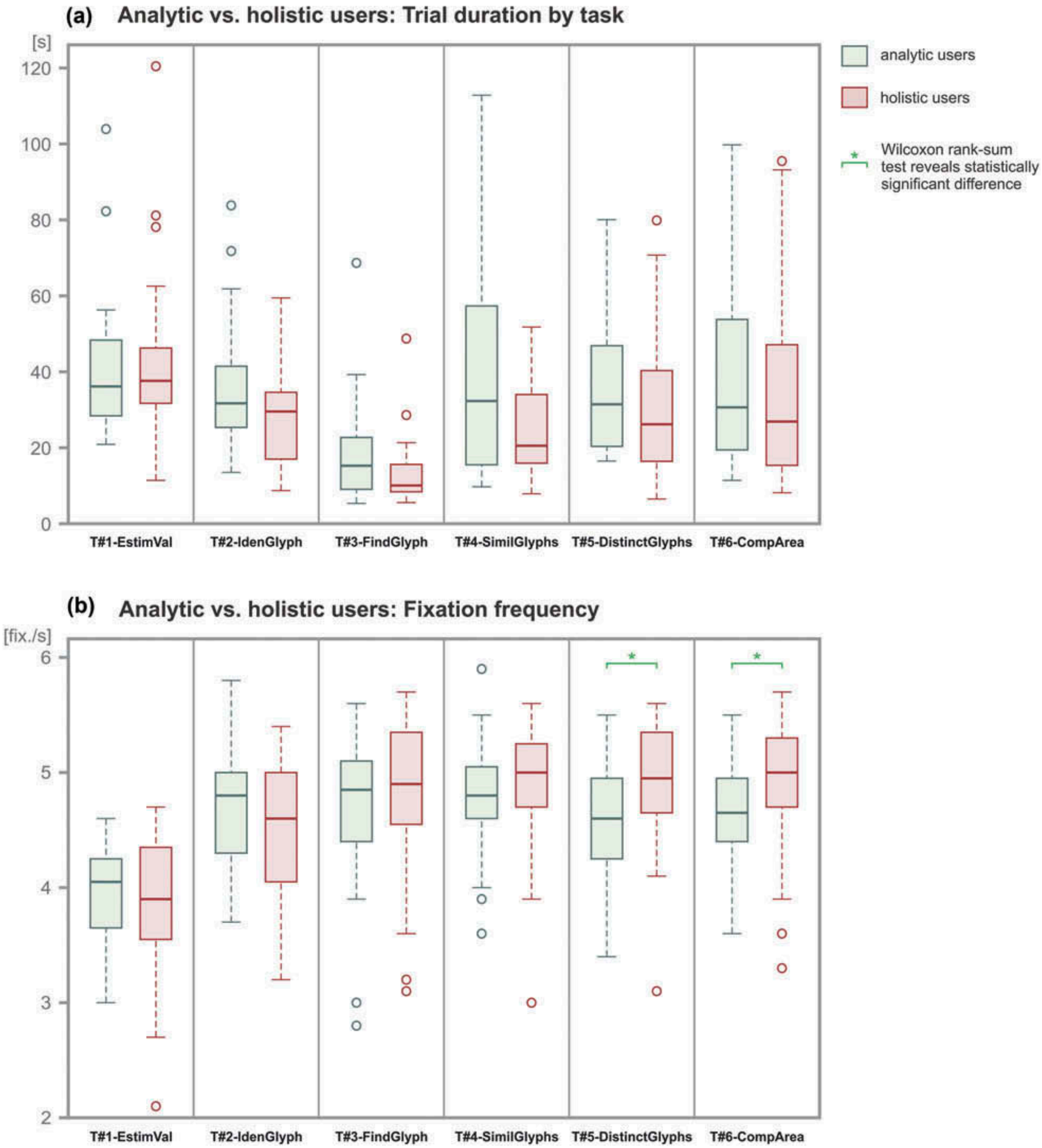

Figure 12. Analytic versus holistic participants by task: (a) the trial duration and (b) the fixation frequency measure.

shorter task accomplishment time are observed for those who use polyline glyphs; however, the analysis reveals no statistically significant differences. In T1-EstimVal, the analysis reveals significant result for the revisits to the key. Participants do not need to check the key as frequently in the case of polyline glyphs compared with star glyphs. We suppose that better performance of polyline glyphs might be caused by the linear order of variables encoded in a polyline glyph, thus resulting in their easier interpretation. In star glyphs, users must find variable positions, and this needs more effort (i.e. more time and more numerous revisits). Lastly, in the subjective questionnaire (Q1), participants' preferences for estimating values lean strongly toward polyline glyphs.

Polyline glyphs seem to perform better if they are used to read datapoint values encoded in glyphs. However, if comparisons are made among glyphs or

Table 3. Polyline versus star glyphs.

Task	Better answer accuracy	Shorter task accomplishment time	Eye-movement data analysis: revisits of specific AOIs (R) or SL		Subjective questionnaire
T1-EstimVal	Polyline glyphs	Polyline glyphs	R	Polyline glyphs[a]	Polyline glyphs
T2-IdenGlyph	Star glyphs	Star glyphs	SL	Polyline glyphs	Polyline glyphs
T3-FindGlyph	Star glyphs (no errors)[a]	Star glyphs[a]	R	Star glyphs[a]	Star glyphs
T4-SimilGlyphs	Star glyphs[a]	IStar glyphs[a]	SL	Star glyphs	Star glyphs
T5-DistinctGlyphs	Polyline glyphs[a]	Star glyphs	SL	Star glyphs	Star glyphs
T6-CompArea	Star glyphs	Polyline glyphs	SL	Polyline glyphs	Star glyphs

SL: Scanpath length. [a]Statistically significant or strong results.

if a compact area of glyphs is found with glyphs that are similar to each other, then star glyphs work better. These findings confirm, to some extent, those reported by Goldberg and Helfman (2011). Indeed, linear graphs support the dimension-finding task better than do radial graphs.

RQ2: grid plot versus map display

The outcomes are consistent when the grid plot is compared with the map display. Although the answer accuracy in T2-IdenGlyph and the revisits measure in T3-FindGlyph are inconclusive (Table 4) and median values of these metrics for the grid and the map are the same, in the majority of tasks, map works better than grid. The only exception is T1-EstimVal, in which a glyph is marked and participants estimate its datapoint values. However, given the nature of this task, display type makes no difference to users.

It can be assumed that maps' geographical background may function as noise in user tasks not related to the spatial context, and therefore arranging glyphs in a grid may facilitate their decoding and increase user performance in tasks such as T1-EstimVal and T3-FindGlyph. However, on maps, similar glyphs are more likely to be near to each other, since adjacent municipalities may feature similar variable scores. Therefore, the map's better results in T4-SimilGlyphs and T6-CompArea do not surprise us. In other cases, the map's better results are not explained. It can be speculated that this might be due to the participants' lack of familiarity with grid plots and due to the use of the grid plot as first in the empirical study. However, ahead of the testing, the participants are explained what grid plots are and how they work. The participants are also given 2 min to freely examine the display used in the empirical study.

RQ3: analytic versus holistic users

Apart from the differences between different glyph and display types, we examine whether there are certain features that influence user performance. We therefore verify whether different cognitive styles (analytic and holistic) influence user behavior. We take into account answer accuracy, task accomplishment time, and fixation frequency (Table 5). Although in T2-IdenGlyph and T3-FindGlyph, holistics and analytics show the same results, in other tasks analytics generally feature greater answer

Table 4. Grid plot versus map display.

Task	Better answer accuracy	Shorter task accomplishment time	Eye-movement data analysis: revisits of specific AOIs (R) or SL	
T1-EstimVal	Grid plot	Map display[a]	R	Map display
T2-IdenGlyph	Inconclusive	Map display	SL	Map display[a]
T3-FindGlyph	Map display (no errors)[a]	Map display	R	Inconclusive
T4-SimilGlyphs	Map display[a]	Map display[a]	SL	Map display[a]
T5-DistinctGlyphs	Map display	Map display	SL	Map display[a]
T6-CompArea	Map display[a]	Map display[a]	SL	Map display[a]

SL: Scanpath length. [a]Statistically significant or strong results.

Table 5. Analytic versus holistic users.

Task	Better answer accuracy	Shorter task accomplishment time	Higher fixation frequency
T1-EstimVal	Analytic users	Analytic users	Analytic users
T2-IdenGlyph	Inconclusive	Holistic users	Analytic users
T3-FindGlyph	Inconclusive	Holistic users	Holistic users
T4-SimilGlyphs	Holistic users	Holistic users	Holistic users
T5-DistinctGlyphs	Analytic users	Holistic users	Holistic users[a]
T6-CompArea	Analytic users	Holistic users	Holistic users[a]

[a]Statistically significant results.

accuracy, whereas holistics need less time to solve tasks and feature higher fixation frequency. However, in the comparison between analytic and holistic users, statistically significant differences are revealed only for the fixation frequency in T5-DistinctGlyphs and T6-CompArea. These results are in accordance with those reported by Tang (2010): For the majority of tasks, fixation frequency is higher for holistics, because they are more proficient at sensing a system's large-scale patterns and reacting to them instead of investigating the system's parts.

Although the comparison between analytic and holistic users provides inconclusive results, they can serve as suggestions for further research. In certain tasks (e.g. T1-EstimVal), analyzing glyph details is more important than sensing glyphs' large-scale patterns and reacting to them (Dewey, 2007). Therefore, it may explain the better answer accuracy observed for analytics. Furthermore, the shorter task accomplishment time and higher fixation frequency of holistic participants is not surprising, since this visual behavior is expected for such users, as they tend to act quicker and focus on general patterns (Kubíček et al., 2016; Navon, 1977). Nevertheless, it is necessary to take account of the fact that, in addition to cognitive style, the way glyphs are marked on a map may also be affected by cartographic knowledge and experience, the pursuit for innovative solutions, and certain personal aspects, which are not included in the analytical and holistic dimensions, such as care.

Conclusions

Glyphs facilitate visual analysis of multivariate geographical data. Star glyphs are particularly common in geovisualization; however, as they make use of polar coordinates, their decoding is impeded. To remedy this, in geovisualization tools – both in map displays and grid plots – star glyphs can be replaced by polyline glyphs. Our study reveals that if either polyline or star glyphs can be used, polyline glyphs are better for facilitating tasks in which datapoint values are to be read. By contrast, if the purpose is to facilitate visual search among glyphs (i.e. to find similar or distinctive glyphs), then star glyphs seem to be a better choice. Moreover, our study reveals that polyline and star glyphs arranged as a map display work generally better than glyphs grouped in a grid plot: Participants who use glyphs in the grid to solve user task perform worse than those who use glyphs shown on the map display. However, this finding needs more research in the future.

There are no particular differences in the visual behavior of participants who use polyline glyphs and participants who interact with star glyphs. One finding is that participants use the key (legend) more frequently if they read datapoint values from star glyphs than if they do so from polyline glyphs. Therefore, our research finding is that a key needs to be better incorporated in a visual interface if star glyphs are to be used to support such user tasks. Finally, glyphs are likely to be used more accurately by analytic users, although analytic users can take more time in comparison with holistic users.

Disclosure statement

No potential conflict of interest was reported by the authors. AQ7

Funding

This work was supported by the KLIMAFORSK program – the Norwegian Research Council's initiative for climate research [Grant Number 235490], climate change, and natural hazards: The geography of community resilience in Norway (ClimRes) and by the Internal Grant Agency of the Palacký University in Olomouc (project IGA_PrF_2017_024, Cloud-based Platform for Integration and Visualisation of Different Kind of Geodata).

ORCID

S Popelka http://orcid.org/0000-0003-2127-2614
J Dolezalova http://orcid.org/0000-0003-2694-5304
J. K Rød http://orcid.org/0000-0003-2935-6206

References

Andrienko, G., Andrienko, N., & Jankowski, P. (2003). Building spatial decision support tools for individuals and groups. *Journal of Decision Systems*, 12(2), 193–208. doi:10.3166/jds.12.193-208 AQ8

Andrienko, N., Andrienko, G., Voss, H., Bernardo, F., Hipolito, J., & Kretchmer, U. (2002). Testing the usability of interactive maps in commonGIS. *Cartography and Geographic Information Science*, 29(4), 325–342. doi:10.1559/152304002782008369

Arnberger, E. (1977). *Thematische Kartographie* [Thematic Cartography]. Brunswick: Westermann.

Bertin, J. (1967). *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes* [Semiology of graphics: Diagrams, networks, maps]. Paris: Gauthier-Villars.

Borgo, R., Kehrer, J., Chung, D. H. S., Maguire, E., Laramee, R. S., Hauser, H., . . . Chen, M. (2013). Glyph-based visualization: Foundations, design guidelines, techniques and applications. *Eurographics (STARs)* (pp. 39–63).

Brand, J., & Johnson, A. P. (2014). Attention to local and global levels of hierarchical Navon figures affects rapid scene categorization. *Frontiers in Psychology*, 5(Article 1274). doi:10.3389/fpsyg.2014.01274

Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American*

Statistical Association, 68(342), 361–368. doi:10.1080/01621459.1973.10482434

Chung, D. H. S., Legg, P. A., Parry, M. L., Bown, R., Griffiths, I. W., Laramee, R. S., & Chen, M. (2015). Glyph sorting: Interactive visualization for multi-dimensional data. *Information Visualization*, 14(1), 76–90. doi:10.1177/1473871613511959

Dewey, R. A. (2007). *Psychology: An introduction*. Russ Dewey.

Dolezalova, J., & Popelka, S. (2016). ScanGraph: A novel scanpath comparison method using graph cliques visualization. *Journal of Eye Movement Research*, 9(4), 5, 1–13. doi:10.16910/jemr.9.4.5

Dorling, D. (1995). The visualization of local urban change across Britain. *Environment and Planning B: Planning and Design*, 22(3), 269–290. doi:10.1068/b220269

Forsberg, A. S., Chen, J., & Laidlaw, D. H. (2009). Comparing 3D vector field visualization methods: A user study. *IEEE Transactions on Visualization and Computer Graphics*, 15 (6), 1219–1226. doi:10.1109/TVCG.2009.126

Fuchs, J., Fischer, F., Mansmann, F., Bertini, E., & Isenberg, P. (2013, April). Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3237–3246). ACM.

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information Visualization*, 10(3), 182–195. doi:10.1177/1473871611406623

Golebiowska, I., Opach, T., & Rød, J. K. (2017). For your eyes only? Evaluating a coordinated and multiple views tool with a map, a parallel coordinated plot and a table using an eye-tracking approach. *International Journal of Geographical Information Science*, 31(2), 237–252. doi:10.1080/13658816.2016.1191636

Gribov, A., Unwin, A., & Hofmann, H. (2006). About glyphs and small multiples: Gauguin and the expo. *Statistical Computing and Graphics Newsletter*, 17(2), 18–22.

Ho, H.-Y., Yeh, I.-C., Lai, Y.-C., Lin, W.-C., & Cherng, F.-Y. (2015, June). Evaluating 2D flow visualization using eye tracking. *Computer Graphics Forum*, 34(3), 501–510. doi:10.1111/cgf.12662

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.

Klippel, A., Hardisty, F., Li, R., & Weaver, C. (2009). Colour-enhanced star plot glyphs: Can salient shape characteristics be overcome? *Cartographica: The International Journal for Geographic Information and Geovisualization*, 44(3), 217–231. doi:10.3138/carto.44.3.217

Kubíček, P., Šašinka, Č., Stachoň, Z., Štěrba, Z., Apeltauer, J., & Urbánek, T. (2016). Cartographic design and usability of visual variables for linear features. *The Cartographic Journal*. Advance online publication. doi:10.1080/00087041.2016.1168141

Lee, M. D., Reilly, R. E., & Butavicius, M. E. (2003, January). An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualization – Volume 24* (pp. 1–10). Australian Computer Society.

Morong, K., & Šašinka, Č. (2014, May). Hypothesis – online software platform for objective experimental testing. Paper presented at the Applying principles of cognitive psychology in practice, Brno.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353–383. doi:10.1016/0010-0285(77)90012-3

Opach, T., & Rød, J. K. (2013). Cartographic visualization of vulnerability to natural hazards. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 48(2), 113–125. doi:10.3138/carto.48.2.1840

Opach, T., & Rød, J. K. (2017). Augmenting the usability of parallel coordinate plot: The polyline glyphs. *Information Visualization*. Advance online publication. doi:10.1177/1473871617693041

Ostrowski, J., & Uhorczak, F. (1972). Typogramy F. Uhorczaka jako środek graficznej prezentacji zjawisk wielocechowych [Typograms of F. Uhorczak as a means of graphical presentation of multi-feature phenomena]. *Polski Przegląd Kartograficzny*, 4(4), 145–151.

Popelka, S. (2014). Optimal eye fixation detection settings for cartographic purposes. *Proceedings of the 14th SGEM GeoConference on Informatics, Geoinformatics and Remote Sensing (SGEM'14)*, 1, 705–712.

Ratajski, L. (1989). *Metodyka kartografii społeczno-gospodarczej* [The Methods of Socio-Economic Cartography] (2nd ed.). Warszawa, Wrocław: PPWK.

Salvucci, D. D., & Goldberg, J. H. (2000, November). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71–78). ACM.

Šašinka, Č. (2013). *Interindividuální rozdíly v percepci prostoru a map* [Interindividual differences in perception of space and maps] (Doctoral dissertation). Faculty of Arts, Masaryk University, Brno.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2010). *Thematic cartography and geographic visualization* (3rd ed.). Upper Saddle River, NJ: Pearson.

Takatsuka, M., & Gahegan, M. (2002). GeoVISTA Studio: A codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, 28(10), 1131–1144. doi:10.1016/S0098-3004(02)00031-6

Tang, D. (2010). *Read between the eyes: Cognitive style difference between Americans and Chinese in a dual-display setting* (Unpublished master's thesis). Cornell University.

Tufte, E. R. (1990). *Envisioning Information*. Cheshire, CT: Graphics Press.

Ünlü, A., & Malik, W. A. (2011). Interactive glyph graphics of multivariate data in psychometrics: The software package gauguin. *Methodology. European Journal of Research Methods for the Behavioral and Social Sciences*, 7(4), 134–144. doi:10.1027/1614-2241/a000031

Ward, M. O. (2002). A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1(3–4), 194–210. doi:10.1057/palgrave.ivs.9500025

Ward, M. O. (2008). Multivariate data glyphs: Principles and practice. In C.-H. Chen, W. K. Hardle, & A. Unwin (Eds.), *Handbook of data visualization* (pp. 179–198). Heidelberg, Berlin: Springer.