This is a preprint[1], please cite: Schnürer, R., Ritzi, M., Çöltekin, A., & Sieber, R. (2020). An empirical evaluation of three-dimensional pie charts with individually extruded sectors in a geovisualization context. *Information Visualization*, https://doi.org/10.1177/1473871619896103

An empirical evaluation of three-dimensional pie charts with individually extruded sectors in a geovisualization context

Raimund Schnürer[1], Martin Ritzi[2], Arzu Çöltekin[2,3*] and René Sieber[1]
1 Institute of Cartography and Geoinformation, ETH Zurich, Zurich, Switzerland
2 Geographic Information Visualization & Analysis, University of Zurich, Zurich, Switzerland
3 Institute of Interactive Technologies (IIT), University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Switzerland

*** Corresponding author:** Arzu Çöltekin, Institute of Interactive Technologies (IIT), University of Applied Sciences and Arts Northwestern Switzerland (FHNW), Bahnhofstrasse 6 5210 Brugg-Windisch, Switzerland.
Email: arzu.coltekin@fhnw.ch

Abstract

This study compares participants' performance in extracting information from 3D pie charts with individually extruded sectors in a single map frame against 2D pie and bar charts in adjacent map frames. Specifically, we examine the response accuracy and response times of 181 adults who were asked to 1) identify the highest magnitude, 2) estimate a proportion, 3) perform both at the same time, and 4) solve a map-related task using the two chart setups. For each task, charts were shown on backgrounds with increasing visual complexity: a blank, a borders-only, and a choropleth map. Furthermore, we tested whether participants' performance improved through additional practice with the two chart types. We did not observe any differences in participants' aggregated response accuracy or response times between the tested 3D and 2D chart types for the highest magnitude task (1) and proportion task (2). However, participants solved the combination task (3) with 2D pie and bar charts on a blank background more accurately and were faster in fulfilling the spatial task (4) with 3D pie charts. The first difference, however, levelled for participants who gained more practice and who accomplished the combination task on maps with higher visual complexity, whereas the second difference persisted even for more trained subjects.

Keywords

2D, 3D, pie charts, bar charts, maps, performance measurement, visual complexity, learning

[1] This version is the reviewed and accepted version of the manuscript before it was copyedited by the publisher.

Introduction

Charts are a widespread means to represent multivariate data. Over the years, scientists have developed and formalized various chart types (Wilkinson, 1999; Schnabel, 2007; Heer, Bostock, & Ogievetsky, 2010; Schnürer, Eichenberger, Sieber, & Hurni, 2015), among which pie charts and bar charts are probably the most common. Users and applications of charts are manifold. For example, decision-makers frequently encounter charts in reports about financial data (Ervin, 2011) or when analyzing usage metrics in web dashboards (Elias & Bezerianos, 2011). If the underlying data is spatially referenced, charts can be placed on a map to depict the proportions of a thematic variable at different locations. This enables, for instance, marine biologists to explore oceanographic data (Kreuseler, 2000) or educators to display imports and exports of a country in school atlases (Hurni, 2017).

While chart use is rather commonplace, the question of which chart types should be used in which contexts is subject to debate. To determine this, user performance with pie charts and bar charts has been previously investigated for certain setups and scenarios. In a seminal study on the topic, Cleveland and McGill (1984) reported that nearly double as many subjects judged the change in percentage of the second largest value correctly with bar charts than with pie charts when the maximum value has been indicated beforehand. On the other hand, Simkin and Hastie (1987) demonstrated that while participants compared values more accurately with bar charts, they judged proportions of the whole more accurately with pie charts. Later, Spence and Lewandowsky (1991) confirmed that bar charts are favorable for comparing magnitudes, while pie charts are advantageous over bar charts when comparing combinations of proportions. In another study, further confirming the previous findings, Hollands and Spence (1992) stated that participants needed more time and made more errors when detecting small differences in changing values with pie charts than with bar charts, whereas the performance was better with pie charts in proportion judgement tasks.

Similar to the pie vs. bar chart debate, another research topic is whether and when to add a depth cue to charts (i.e. the third dimension or 3D). Arguably, adding a depth cue to 2D charts increases their attractiveness (Fausset, Rogers, & Fisk, 2008). It has been demonstrated in a study wherein a selection of 70 different diagram types were compared that participants favored 3D pie charts against all others in terms of aesthetic appeal (Burch, 2015). What people prefer is an important factor to consider in designing graphics, in order to ensure that the graphics are embraced by users; however, how depth cues on charts affect user performance while solving different tasks is critically important as well. Several experiments demonstrate that people perform better or at least similarly well with 2D charts than with their 3D equivalents. For instance, Siegrist (1996) reports that participants were slower and less accurate in estimating proportions in pie charts and magnitudes of bar charts when decorative depth was added. Participants in Hughes' (2001) experiment were also less accurate in comparing relative sizes with 3D bar charts than with 2D bar charts. Stewart et al. (2009) showed that participants performed a read-off task 'equally well' with 2D pie charts and 2D bar charts compared to their 3D counterparts. Participants performed better with 2D charts than with 3D charts when the task type was changed to a spatial transformation task or a trend prediction task. In another study examining different task types, Schonlau and Peters (2012) reported that depth cues reduce subjects' comprehension of pie charts, but not of bar charts. It is important to note that in all of these studies, the third dimension was *extraneously* used, that is, it was task-irrelevant and did not convey any information and thus essentially would be regarded as 'chartjunk' by Tufte (1983).

Influence of visual variables—such as size, orientation, and color (Bertin, 1983)—on user performance with charts has been examined in addition to depth cues. For instance, Zacks et al. (1998) stated that participants' judgements of absolute and relative bar heights were less accurate when depth cue was added. Zacks et al. (1998) altered the length of depth cues and changed their type to rectangles; these modifications led to similar results. Likewise, Fischer (2000) examined the effect of 2D and 3D frames as visual aids for 2D and 3D bars. In Fischer's (2000) study, participants needed the most time with 3D bars in 3D frames to compare bar sizes. Fischer (2000) interpreted the result as an indication that the working memory load was the highest in the 3D setup.

Studying yet another visual variable, Rangecroft (2003) reported that the orientation of pie chart sectors may influence performance. Participants in Rangecroft's (2003) experiment identified the smallest and largest segments more accurately in 2D than in 3D pie charts, whereas the accuracy was low (under 75%) or very low (50% or lower) for half of the 3D pie chart configurations differing in orientation. In a recent study, Kosara (2019) varied single pie chart sectors in terms of viewing angle, body height, value range, and rotation around the center. Study participants were asked to replicate the altered pie segments interactively in a 2D reference pie chart. Kosara (2019) observed a higher number of errors for viewing angles of 15° and 30° than those of 60° and 90°, as well as for sector orientations at the side than at the front or back of the chart; however, body heights did not affect the error rates. Note that newer studies, such as Schonlau and Peters' (2012), used colored chart depictions, whereas charts in older experiments, like Siegrist (1996), were mainly drawn in black and white. Even though color is a strong visual variable (e.g. Brychtová & Çöltekin, 2017), Stewart et al. (2009) stated that "graph comprehension accuracy was similar for color and black-and-white graphs" (p. 198) as a result of an experiment manipulating color, in which they compared charts both with and without depth cues. Depending on the colors used in the charts, this result could be challenged, but a comprehensive review of the effects of color is beyond the scope of this paper.

In this study, inspired by the use of 2D and 3D charts in the Atlas of Switzerland (Sieber, Serebryakova, Schnürer, & Hurni, 2016), an important objective is to understand how charts can be successfully combined with maps. One can question if it would be better to avoid adding charts altogether, and depict all attributes on side-by-side "small multiple" maps (Tufte, 1983); however, visualizing high-dimensional data has been a persistent challenge (Çöltekin, Bleisch, Andrienko, & Dykes, 2017) and empirical evidence on the effectiveness of particular representations or their combinations is often missing. Thus, we briefly review some of the previous studies analyzing user performance with charts on maps or similar geographic visualizations.

Cleveland, Harris, and McGill (1982) compared the performance of high-school students and scientists with proportional circles on statistical maps with 'map-like' stimuli (i.e. ticks, border, labels, scale bar) and 'non-map-like' stimuli (i.e. a blank background). Circles were randomly positioned in an invisible four by three grid in both setups. The authors noted that the difference between map-like and non-map-like stimuli is "sufficiently small" (p. 544), thus they did not include it in their analysis (Cleveland et al., 1982). As one of the earliest works in cartography examining the role of 3D symbols on maps, Kraak's (1988) dissertation included border maps for mono and stereo vision with three-dimensional point symbols (i.e. cubes, spheres) varying in visual variables (i.e. size, color). Three different types of questions involving the subjects to read single, neighbored, or all symbols were formulated. In Kraak's study (1988), geodesy and cartography students were significantly faster and provided more correct answers in stereo than in mono maps with differently sized 3D symbols. In another study, Lewandowsky et al. (1993) conducted two experiments, in which they compared color and grayscale choropleth maps, dot density maps, and maps with pie charts, where all maps included borders of Health Service Areas or US states. Participants were asked to identify and recall clusters of particularly high or low mortality for certain diseases. Lewandowsky et al. (1993) reported that people forgot more clusters on Health Service Areas than on US states boundaries, although the immediate perception of clusters was about the same. Participants identified fewer clusters on maps with pie charts than on monochrome choropleth maps on average. On the other hand, Seipel and Carvalho (2012) did not find any differences concerning accuracy and time when participants judged the relative heights of single bars visualized on a 2D and a tilted 3D base map. The latter results coincide with those of Fischer (2000), who left the background blank in his study.

Whether the background of a chart is blank or not, it is important to remember that most modern maps, atlases, and other geovisualization outputs are presented interactively. Thus, results from studies examining static charts and maps alone might not necessarily apply to more recent visualization work. Examining the role of interaction in participant performance with charts and maps, Bleisch, Dykes, and Nebiker (2008) stated that participants needed significantly more time in estimating differences of bars in an interactive 3D geovisualization environment than in a static 2D alternative. However, the authors suspected it was mainly due to the navigation in the 3D space. Participants also recognized bar heights faster with a reference grid than without in the 3D maps (Bleisch et al., 2008).

When people face new visualizations, they need to learn how to read and interpret them. This is broadly termed 'graphical literacy' and it is a part of the curriculum in schools in many countries. In connection to graphical literacy and learning, Ben-Chaim, Lappan, and Houang (1988) demonstrated that high school students could improve their spatial visualization abilities through instruction. Ben-Chaim, Lappan, and Houang's (1988) finding was based on a task to match 2D squares from different views of 3D cubes, which represented buildings on a map, and vice-versa. Two years later, in a comprehensive essay, Pinker (1990) noted that the ability to read graphs can be improved by explicit or implicit instructions with practice, and with experience in drawing quantitative relationships. Shah and Hoeffner (2002) added that it is helpful if readers understand the scientific context and if they are able to switch between chart representations, because they can connect visual features and meaning.

As learning is a lifelong process, professionals may also encounter and need to adapt to unfamiliar visualizations. Krishnamoorthy and North (2005), for instance, tested the learnability of a mapping application, including visualizations like a histogram and a scatter-plot in side-by-side frames. In particular, they examined the effect of exploratory learning by guiding users only by the application without providing any additional explanations—apart from the possibility of consulting the help menu. Krishnamoorthy and North (2005) observed that "[n]ovices tend to stick to initially learnt strategies" (p. 310), which may not be the most efficient ones. The authors suggested 'self-disclosure' as a method to overcome this problem, that is, giving context-sensitive help to users while performing a task for the first time. To measure the visualization literacy of laymen, Lee, Kim, and Kwon (2017) developed a test with eight tasks for 12 visualization types. Tasks for pie and bar charts, for example, involved retrieving a value, finding the extremum, and making a comparison. Domain experts judged the content of all tasks for pie charts as equally valid, while they assigned higher scores to the first two tasks for bar charts than to the comparison task. This comparison task was also the most difficult to answer for tested participants in the study of Lee et al. (2017).

We can conclude from the studies reviewed above that people can compare magnitudes well with bar charts, while pie charts might fit better for identifying a portion of a whole. Adding a decorative depth cue to either chart type has an overall negative impact on performance in a variety of tasks. To our knowledge, however, no studies have examined pie charts in depicting *meaningful* information in the third dimension on the pie sectors. Current off-the-shelf software like spreadsheet programs (e.g. Microsoft Excel) or geographic information systems (GIS) software (e.g. ESRI ArcGIS) cannot generate 3D pie charts with individually extruded sectors, which may explain the lack of studies. Likewise, it seems that measuring the influence of what is shown in the background has not been subject of research yet. Our literature review revealed that charts are often placed on blank or neutral backgrounds in information visualization experiments, whereas in a geovisualization context, a map in the background is often used. Lastly, researchers have offered observations on how people learn to work with new visualizations, and charts in particular, suggesting that design of a chart might have a different effect on first-time and returning users.

With this paper, we contribute towards a better understanding of chart use by answering the following questions through an empirical study: 1) How well do people perform in chart- and map-reading tasks with 3D pie charts in comparison to 2D charts when the third dimension represents meaningful information? 2) Do the performance differences among people using 2D vs. 3D charts persist when the background is no longer blank? 3) How does practice affect performance with 'novel' or less usual chart types? Below, we present more precise questions and the related hypotheses.

Hypotheses

This study primarily examines how well people can read 3D pie charts with their sectors extruded proportionally to individual data values: One attribute is mapped on the pie sector angle and another is depicted on the height of the extruded sector. A real-world example, where 3D pie charts represent the number and capacity of wood-fired heating systems, has been implemented in a national atlas (Sieber, Schnürer, Eichenberger, & Hurni, 2013). The topic of the current study pertains to traffic volumes and commuting times for different means of transportation. We compare 3D pie charts to 2D pie and bar charts in adjacent frames, where both the 2D and 3D setup encode identical information. The extrusion in the 3D pie chart is essentially a way of integrating the 2D bar chart onto a 2D pie chart.

Task type

We test the participants' performance (response accuracy and response times) with four different tasks varying in type and difficulty. We hypothesize that while participants should be more accurate in estimating the magnitude of the sectors with the 2D chart combination due to perceptual skew introduced by the perspective view in 3D pie charts (Wilkinson, 2001), they should be faster in using 3D pie charts than using 2D pie and bar charts in adjacent frames because the 2D setup would require more eye movements over a larger space (Brügger, Fabrikant, & Çöltekin, 2016).

Background complexity

Differences in performance were often observed in previous studies when the chart was presented on a blank background. As charts have been used on maps for nearly two centuries (e.g. Minard, 1858), we want to examine whether the presence of visual 'noise' (i.e. map elements and colors) influences participants' performance in reading chart values when the background map is not relevant for solving the task in that moment. In geovisualization contexts, the map in the background is typically relevant; however, it might also act as a distractor if one is trying to read information only from the charts. When the map is relevant, attention is usually switched between the background map and the charts.

In this study, we examine one task with a meaningful interaction with the background map, but also focus on the effect of 'noise' as one of the possible scenarios. We vary the backgrounds from a blank to two 'noisy' backgrounds in the first tasks, and finally present a task where the map is relevant. For the tasks with visual 'noise' in them, we hypothesize that the perceptual advantages offered by each chart type should be more pronounced in the presence of a non-blank background, because the visual complexity created by this 'noise' makes the tasks cognitively harder for the participants by forcing them to process more visual input than necessary (MacEachren, 1982; Neider & Zelinsky, 2011; Schnur, Bektaş, & Çöltekin, 2017).

Practice levels

An additional consideration in a study like ours is whether the proposed visualization type, that is, the 3D pie chart with individually extruded sectors—is 'learnable'. A strong motivation for this aspect is the frequent use of charts in school atlases or in other educational materials. We hypothesize that if participants had a chance to get some practice (e.g. by solving similar tasks more than once), their performance would overall improve, especially for a visualization type that is unknown or less familiar to them. More specifically, we expect more practiced participants to show level performance with the new visualization (i.e. the 3D pie chart) and the baseline visualization type (i.e. the 2D pie and bar chart combination).

Methods

Participants

We recruited participants from the second author's personal and university network by email and social media to participate in the experiment. A total of 181 adults (111 male, 70 female) aged 19–77 years (M = 32.28, SD = 11.10) completed our study. We excluded four participants from the statistical analysis as they exhibited color deficiency in our screening with the Ishihara (1917) test. Most participants (63%) indicated that they hold a university degree, whereas the remaining participants have completed an apprenticeship (15%) or secondary school education (22%). Participants reported that they were familiar with current information technologies: 6% have very poor/poor, 30% average, and 64% good/very good general computer and visualization knowledge and skills. Based on self-reports, participants' experience with maps and other cartographic products is balanced: 37% have very poor/poor, 36% average, and 27% good/very good cartographic knowledge and skills.

Participants were randomly assigned into one of two groups: 91 adults (53 male, 38 female) aged 22–77 years (M = 33.51, SD = 11.68) to Group A; and 86 adults (54 male, 32 female) aged 19–71 years (M = 31.14, SD = 10.51) to Group B. The differences between the groups are explained in detail in the Procedure section.

Materials

We created 36 maps varying *in chart type, number of frames and charts, and background type (Figure 1)*.

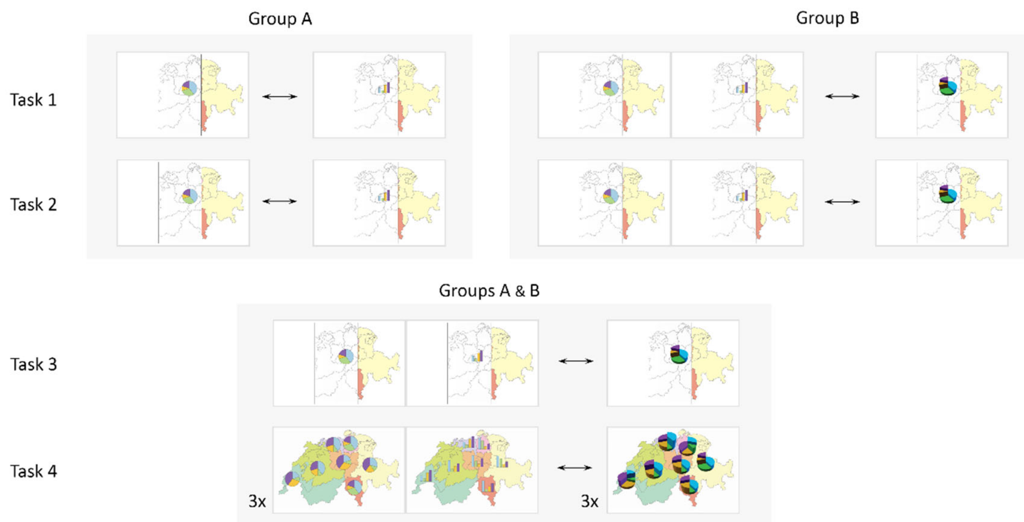

Figure 1: Chart types and background maps used in our study for different groups and tasks: Group A used 2D pie charts or 2D bar charts in a single map frame for Tasks 1 (highest magnitude) and 2 (proportion). For Tasks 3 (combination) and 4 (map-related), Group A used 2D pie and bar charts in adjacent map frames and 3D pie charts with individually extruded sectors in a single map frame. Group B used the latter chart setup for all tasks. A task consisted of three questions. Questions of Tasks 1 to 3 were performed on a blank map, a borders-only map, or a choropleth map. Questions of Task 4 were performed only on a choropleth map. Chart proportions varied between the questions, but were counterbalanced.

As shown in Figure 1, six maps containing one 2D pie chart and six other maps containing one 2D bar chart were presented to Group A in Tasks 1 and 2. These 12 maps were meant as 'fillers' to prevent Group A participants from practicing answering questions with 2D charts in adjacent (= juxtaposed or side-by-side) frames and 3D pie charts. The latter stimuli, which included six maps with one 2D pie chart and one 2D bar chart in adjacent frames as well as six other maps with one 3D pie chart in a single frame, were shown to Group B participants in Tasks 1 and 2 to comparatively assess their performance with these setups. Map sets with the same chart type in each task differed in background (i.e. blank, border, choropleth) to additionally examine the influence of visual 'noise' on participants' performance.

The same setup for Group B, comprising three maps each with adjacent 2D charts and single 3D pie charts, was provided to both groups in Task 3, but the background was still not meaningful for solving the task. In Task 4, however, all maps had the choropleth background consisting of seven colors, where each color represented a greater region. Three of the maps comprised seven 2D pie charts and seven 2D bar charts in adjacent frames, and three other maps had seven 3D pie charts in a single frame. The charts were centered in the greater regions in Task 4 and in the map frame in the other tasks. Charts and backgrounds are explained in detail in the following paragraphs, while tasks and groups are described in depth in the Procedure section.

We represented fictional values for the topic "Commuting in Switzerland" in all of the charts used in the experiment. The means of transportation was arbitrarily color-coded in the chart segments (i.e. blue = by foot/no commute, green = by bike/moped, yellow = by public transportation, violet = by car/motor bike). The share of people using a certain means of transportation was represented by proportions of 2D and 3D pie chart sectors (e.g. 10%, 20%, 30%, 40%). We chose differences of 10%, as according to Slocum (1981), differences in two-sectored pie charts smaller than 2.5% cannot be distinguished by participants. Since a recent study shows that people are more effective with basic pie charts (Kosara & Skau, 2016), we did not alter their shapes (e.g. to ellipsis, square) or sectors (e.g. individual radii, translated sectors). The average duration of travel to work (i.e. 0.5h, 1h, 1.5h, 2h) was proportional to heights of 2D bar charts and extruded sectors of 3D pie charts. We aligned the bars vertically, as it has been shown that people are more efficient in solving tasks with vertical than with horizontal bars (Fischer, Dewulf, & Hill, 2005).

The assignment of proportions to pie chart sectors as well as to heights of bar charts and 3D pie charts varied for each question in a counterbalanced manner to prevent an additional learning effect. The assignments are not meant to be representative for the topic, since we primarily aimed to control the influence of chart proportions on participants' performance by using the same proportions (i.e. 1 : 2 : 3 : 4) for all 2D and 3D chart segments. Chart size and order of colors of chart segments were also the same in all maps. 3D pie charts were arranged in a way that no occlusions occur, which we could not guarantee when we used real data. The viewing angle of 3D pie charts was 45° and extruded sectors were shaded in black. As shades are not common for 2D chart sectors, we chose a black outline in the corresponding depiction. We created 2D pie and bar charts with Microsoft Excel. As in the study by Schnürer et al. (2015), 3D pie charts were produced with OpenJSCAD, a JavaScript library for constructing solid geometries, and Blender, a 3D modeling application.

We systematically manipulated the visual complexity of the background on which the charts were displayed, except for the maps with seven charts, since here the background was essential for solving the task. In total, 10 maps had a blank background, 10 had black lines in the background representing borders of Swiss cantons, and 16 (i.e. choropleth maps) had borders of Swiss cantons as well as seven color-coded NUTS-2 regions (European Parliament and the Council, 2003) in the background (*Figure 1*). Borders of Swiss cantons originate from the "Vector 200" data set provided by the Federal Office of Topography swisstopo. The colors of the NUTS-2 regions were chosen close to a map of the Swiss Federal Statistical Office (SFSO)—dark green for Lake Geneva region (LG), light green for Espace Mittelland (EM), light violet for Northwestern Switzerland (NW), yellow for Eastern Switzerland (ES), orange for Central Switzerland (CS), pink for Zurich (ZH), and red for Ticino (TI).

We provided static maps to the participants so that they are exposed to the same information and cannot change the viewed information by panning, rotating, and zooming. The title was placed in all maps at the top center, the data source at the left bottom, the north arrow at right top, and the scale bar at the right bottom. Legends for angles of pie chart sectors were omitted as their relation to percentages was implied by the questions. Legends for chart segment colors and chart heights were put below data source and scale bar, while the legend for colored NUTS-2 Swiss regions was put at the left top the map. A short textual explanation was given below the legend (*Figure 2*). Backgrounds and map elements were created with ESRI ArcMap, apart from the legends, which we generated using Adobe Illustrator.

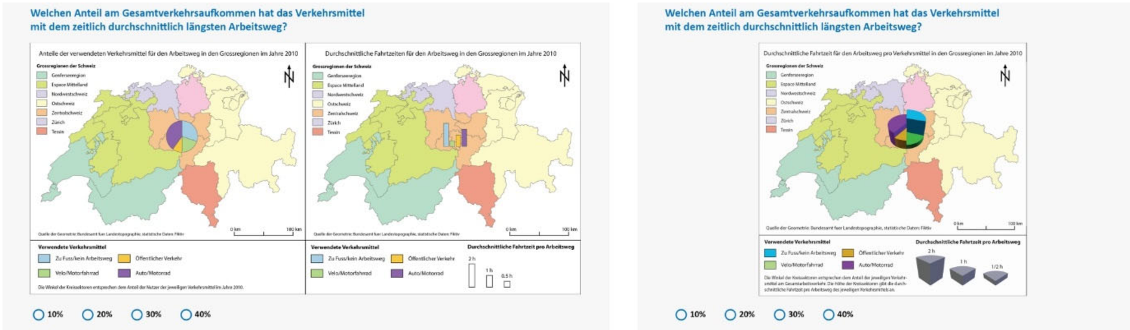

Figure 2: Two questions, used in our study for Task 3, containing 2D pie and bar charts in adjacent map frames (*left*) and 3D pie charts in one map frame (*right*), both on choropleth background. Participants were asked to select the share of the total traffic volume (= Anteil am Gesamtverkehrsaufkommen) for the the transportation means (= Verkehrsmittel) with the longest travel time (= Fahrtzeit). The correct answer is 40%.

Procedure

The experiment was conducted as an online study and the data collection period was three weeks (*Figure 3)*.

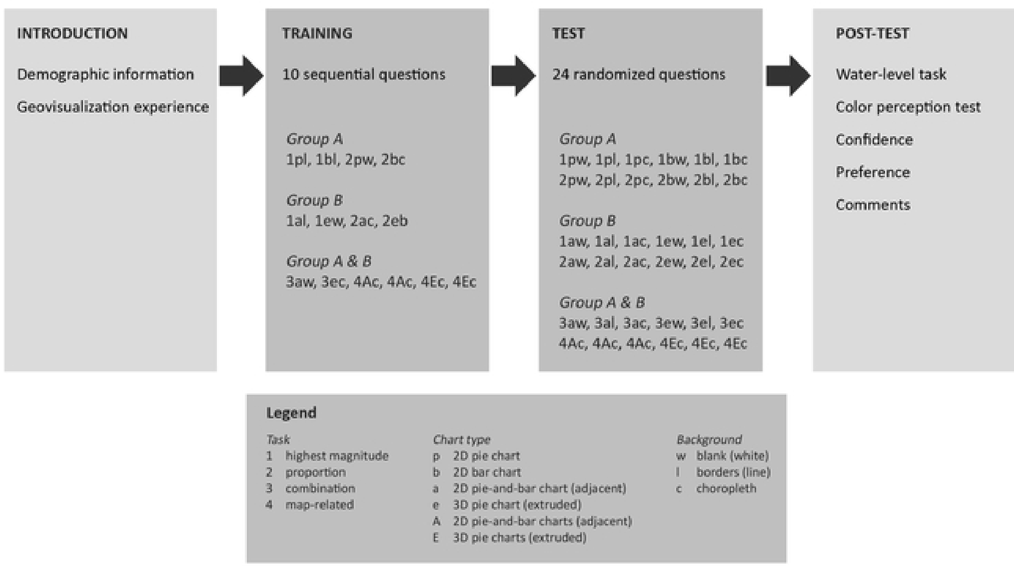

Figure 3: Study design

All textual descriptions were given in German. Participants were asked to use desktop and mobile devices with a sufficiently large screen so that they would not need to scroll on the page. Participants were ensured of anonymity and confidentiality through a consent form. In the introductory part of the survey, participants provided demographic information, such as their age, gender, and education level. We also asked about their experience, interest, and skills in using computers, maps, and visualizations using five-point Likert scales.

In the main experiment, participants were first familiarized with tasks and maps in a training session. Tasks in the training session were similar to those in the actual experiment; however, proportions of charts (and thus the answers) were different. One example question was given for each task and chart type, except for the last task, for which we gave the participants two example questions because it was the most difficult task. The correct answer was presented after each question to the participants to help them self-assess their accuracy.

With the training session, we aimed to acquaint participants with the tasks and visualizations as we assumed that the majority of them would be unfamiliar with 3D pie charts with individually extruded sectors. The actual experiment in the survey consisted of 24 mandatory questions, that is, one question for each combination of two chart types, three backgrounds, and four tasks. Before the experiment started, participants were asked to answer questions as quickly and accurately as possible. Questions were randomized to control for order effects due to learning, which would favor a certain chart type, background, or task. After every four questions, participants had the opportunity to relax in open-ended breaks. The page layout for each question was as follows: There was a map in the middle of the page, a question on top of the map, and a radio-button box with possible answers below the map (*Figure 2*). One of four possible answers was correct for the first three tasks, and one of seven possible answers (i.e. the NUTS-2 regions) for the last task. The selected answer of participants and the time needed to respond were recorded. There was no time limit on any question.

Tasks and example questions are listed in *Table 1*.

Table 1: Task types and exemplary questions in our study

Task type	**Example question**
Task 1 (Highest magnitude): Tick the color of the largest pie chart sector, highest bar, or the most extruded pie chart sector.	For 2D pie-and-bar charts: Which bar is the highest one? For 3D pie charts: Which extruded pie sector is the highest one?
Task 2 (Proportion): Tick the percentage of a pie chart sector or an extruded pie chart sector in a certain color.	Which share of the whole circle has the orange sector?
Task 3 (Combination): Tick the percentage of a pie chart sector from the highest bar or the most extruded pie chart sector in the same color.	Which share of the total traffic volume has the transportation means with the longest travel time?
Task 4 (Map-related): Tick the region from a color-coded means of transportation, a certain pie chart percentage, and the highest bar, or most extruded pie chart sector in the same color.	Which NUTS-2 region has the following properties: a\ the highest average travel time for "car/motor bike", and b\ a share of 40% on the total traffic volume for "car/motor bike"?

The readability of maximum values in Task 1 is similar to that in Meyer, Shinar, and Leiser's (1997) experiments, whereas proportion estimations—also known as part or portion of the whole judgments—in Task 2 are based on the study by Simkin and Hastie (1987). Tasks 3 and 4 are combinations of Tasks 1 and 2, with the difference that Task 4 additionally involves a geospatial component (i.e. identification of a region). A map-related task, such as Task 4, is representative of chart use in a geovisualization context (e.g. atlases), whereas Tasks 1 to 3 are also common outside geovisualization (e.g. reports). We designed the tasks to progress from 'simple' to 'complex' based on our own judgement. Task difficulty was controlled to examine if effects, if any, would persist over more complex tasks as well.

Each task consisted of six questions. For Group A, Tasks 1 and 2 had six questions each for regular pie vs. bar charts. For Group B, Tasks 1 and 2 again had six questions each for 2D vs. 3D chart setups as described earlier. Tasks 3 and 4 had identical questions for both groups. This design shows whether there is a practice effect for Tasks 3 and 4 because this way, participants in Group B would solve the questions with 2D pie and bar charts in two adjacent frames and 3D pie charts twice as often as participants in Group A during the study. Tasks 1, 2, and 3 were executed twice on a blank background, twice on a 'borders-only' background, and twice on a choropleth map in the background. Task 4 could not be solved without the map; thus, the charts were always shown on a choropleth map. A detailed overview of our study setup is given in *Table 2*.

After the main experiment, participants took a color perception test (Ishihara, 1917) for red-green and blue-yellow, which we used as an inclusion criterion. Participants also responded to a water-level task (Piaget & Inhelder, 1956), rated their confidence during the experiment, and stated their preferred chart type. In this paper, we focus on performance metrics, as described above. Water-level task, confidence, and preference results have been reported earlier by [Anonymized for reviewers] (2016). The experiment ended with an optional comment box, in case participants had any feedback about the study. We created the experiment using the online survey tool LamaPoll. Participants did not receive any compensation for their effort and participation was entirely voluntary. In our study, we followed Standard 8 of the Ethical Principles and Code of Conduct for Psychologists (American Psychological Association, 2017).

Statistical Analysis

As mentioned in the hypothesis about the task type, we regard performance as a combination of effectiveness (response accuracy) and efficiency (response time) in this study. To quantify response accuracy, we gave one point for correct answers and zero points for incorrect answers, making accuracy a binomial variable. Response time was measured as the time spent by participants to answer a question. We log-transformed response times in all tests to meet the assumption of normal distribution (Lawrence, 1988). We carried out the statistical tests with R (version 3.5.3) and assumed a significance level of $\alpha = .05$.

Results

In this section, we summarize the results for comparing participants' performance with *2D pie-and-bar charts* (= 2D pie and bar charts in adjacent frames) and *3D pie charts* (in a single frame). Results for 2D pie vs. 2D bar charts in single frames are not reported as they simply replicate the previous findings (i.e. pie charts are better for judging proportions of a whole, while bar charts are better for estimating magnitudes).

Overall performance

Figure 4 shows the main effects for participants' accuracy (*left*) and response times (*right*) with 2D pie-and-bar and 3D pie charts. The descriptive statistics suggests a slight advantage with 2D pie-and-bar charts in terms of accuracy ($M_{2D} = 91.32\%$, $SD_{2D} = 28.17\%$ vs. $M_{3D} = 89.48\%$, $SD_{3D} = 30.69\%$) and with 3D pie charts in terms of response time ($M_{2D} = 18.11s$, $SD_{2D} = 13.91s$ vs. $M_{3D} = 16.94s$, $SD_{3D} = 12.69s$). However, a weighted Welch's t-test for the main effects shows that the differences in neither accuracy ($t(3120.78) = 1.84$, $p = .07$) nor response time ($t(3148.21) = 1.67$, $p = .09$) are statistically significant at the aggregate level.

Table 2: Groups, tasks, questions, chart types, backgrounds, geographic regions, chart segment proportions, solutions, response accuracies, and response times in our study. The following chart types were used: 2D pie = 2D pie chart in a single frame; 2D bar = 2D bar chart in a single frame; 2D pie-and-bar = 2D pie and bar chart(s) in adjacent frames; 3D pie = 3D pie chart(s) in a single frame. These geographic regions were present in the choropleth map: LG = Lake Geneva Region; EM = Espace Mittelland; NW = Northwestern Switzerland; ES = Eastern Switzerland; CS = Central Switzerland; ZH = Zurich; TI = Ticino.

Group	**Task**	**Question**	**Chart type**	**Back-ground**	**Re-gi-ons**	**Proportions of pie chart sectors (2D / 3D)**				**Bar chart heights (2D) / Extrusion heights of pie chart sectors (3D)**				**Correct answer**	**Mean and SE of correct answers by participants [%]**	**Mean and SE of answer times by participants [s]**
						Blue	**Green**	**Orange**	**Violet**	**Blue**	**Green**	**Orange**	**Violet**			
A	1	Color of largest sector	2D pie	blank		40	30	10	20	-	-	-	-	blue	100	7.34 ± 0.74
		Color of largest sector	2D pie	borders		30	20	10	40	-	-	-	-	violet	100	7.89 ± 0.78
		Color of largest sector	2D pie	choropleth		30	10	20	40	-	-	-	-	violet	100	7.73 ± 0.83
		Color of highest bar	2D bar	blank		-	-	-	-	2	0.5	1	1.5	blue	100	5.40 ± 0.38
		Color of highest bar	2D bar	borders		-	-	-	-	1.5	1	0.5	2	violet	100	7.32 ± 0.86
		Color of highest bar	2D bar	choropleth		-	-	-	-	1	0.5	1.5	2	violet	100	7.31 ± 0.71
B	1	Color of highest bar	2D pie-and-bar	blank		25	25	25	25	2	0.5	1	1.5	blue	100	7.15 ± 0.63
		Color of highest bar	2D pie-and-bar	borders		25	25	25	25	1.5	1	0.5	2	violet	100	7.2 ± 1.13
		Color of highest bar	2D pie-and-bar	choropleth		25	25	25	25	1	0.5	1.5	2	violet	100	7.44 ± 0.88
		Color of highest sector	3D pie	blank		25	25	25	25	2	0.5	1	1.5	blue	100	6.84 ± 0.50
		Color of highest sector	3D pie	borders		25	25	25	25	1.5	1	0.5	2	violet	100	7.41 ± 0.97
		Color of highest sector	3D pie	choropleth		25	25	25	25	1	0.5	1.5	2	violet	100	7.09 ± 0.68
A	2	% of orange sector	2D pie	blank		30	10	20	40	-	-	-	-	20	97.80 ± 3.07	10.94 ± 2.53
		% of green sector	2D pie	borders		40	30	10	20	-	-	-	-	30	96.70 ± 3.74	13.38 ± 2.07 †
		% of orange sector	2D pie	choropleth		30	20	10	40	-	-	-	-	10	98.90 ± 2.18	11.11 ± 1.66
		% of orange bar	2D bar	blank		-	-	-	-	2	0.5	1	1.5	20	90.11 ± 6.25	15.53 ± 2.92
		% of green bar	2D bar	borders		-	-	-	-	1.5	1	0.5	2	20	89.01 ± 6.55	16.84 ± 3.14
		% of orange bar	2D bar	choropleth		-	-	-	-	1	0.5	1.5	2	30	89.01 ± 6.55	18.53 ± 3.12
B	2	% of orange sector	2D pie-and-bar	blank		30	10	20	40	2	1.5	1	0.5	20	94.19 ± 5.05	11.13 ± 1.52
		% of green sector	2D pie-and-bar	borders		40	30	10	20	2	1.5	1	0.5	30	95.35 ± 4.54	11.7 ± 1.11
		% of orange sector	2D pie-and-bar	choropleth		30	20	10	40	2	1.5	1	0.5	10	98.84 ± 2.31	9.57 ± 1.10
		% of orange sector	3D pie	blank		30	10	20	40	2	1.5	1	0.5	20	88.37 ± 6.91	11.09 ± 1.52
		% of green sector	3D pie	borders		40	30	10	20	2	1.5	1	0.5	30	94.19 ± 5.05	11.76 ± 1.53
		% of orange sector	3D pie	choropleth		30	20	10	40	2	1.5	1	0.5	10	96.51 ± 3.96	9.22 ± 1.03
A&B	3	% of longest travel time	2D pie-and-bar	blank		30	10	20	40	1	0.5	1.5	2	40	94.35 ± 3.43	16.43 ± 1.41
		% of longest travel time	2D pie-and-bar	borders		40	30	10	20	1	0.5	1.5	2	20	90.4 ± 4.38	16.98 ± 1.74
		% of longest travel time	2D pie-and-bar	choropleth		30	20	10	40	2	0.5	1	1.5	30	89.83 ± 4.50	16.36 ± 1.58
		% of longest travel time	3D pie	blank		30	10	20	40	1	0.5	1.5	2	40	84.18 ± 5.43	15.83 ± 1.66
		% of longest travel time	3D pie	borders		40	30	10	20	1	0.5	1.5	2	20	89.83 ± 4.5	17.64 ± 1.51
		% of longest travel time	3D pie	choropleth		30	20	10	40	2	0.5	1	1.5	30	87.01 ± 5	18.3 ± 2.09

(cont. next page)

(cont from previous page)

A&B	4	Region with 40% share and longest travel time for car/motor bike	2D pie-and-bar	choropleth	LG EM NW ES CS ZH TI	20 40 40 30 20 40 40	10 10 10 10 10 30 30	30 20 20 20 30 10 10	40 30 30 40 40 20 20	1 2 1 2 2 2 2	0.5 0.5 0.5 1.5 1 1.5 0.5	1.5 1 1.5 0.5 0.5 0.5 1	2 1.5 2 1 1.5 1 1.5	LG	86.44 ± 5.09	27.78 ± 2.23
		Region with 30% share and longest travel time for by foot/no commute	2D pie-and-bar	choropleth	LG EM NW ES CS ZH TI	20 40 20 30 30 40 20	10 10 10 10 10 30 10	30 20 30 20 20 10 30	40 30 40 40 40 20 40	1 2 1 1.5 2 2 1	0.5 0.5 0.5 0.5 1 1.5 0.5	1.5 1 1.5 1 0.5 0.5 1.5	2 1.5 2 2 1.5 1 2	CS	82.49 ± 5.65	27.84 ± 1.83
		Region with 40% share and longest travel time for by foot/no commute	2D pie-and-bar	choropleth	LG EM NW ES CS ZH TI	20 20 20 30 40 40 30	10 10 10 10 20 30 10	30 30 30 20 10 10 20	40 40 40 40 30 20 40	1 2 1 2 2 1 1.5	0.5 0.5 0.5 1.5 1 0.5 0.5	1.5 1 1.5 0.5 0.5 1.5 1	2 1.5 2 1 1.5 2 2	CS	84.75 ± 5.35	29.73 ± 2.78
		Region with 40% share and longest travel time for by foot/no commute	3D pie	choropleth	LG EM NW ES CS ZH TI	20 40 40 40 40 30 40	10 10 10 30 10 10 10	30 20 20 10 20 20 20	40 30 30 20 30 30 30	2 1.5 2 1 1 2 2	1 0.5 0.5 0.5 0.5 0.5 0.5	0.5 1 1 1.5 1.5 1 1	1.5 2 1.5 2 2 1.5 1.5	NW	90.4 ± 4.38	23.2 ± 1.44
		Region with 30% share and longest travel time for by car/motor bike	3D pie	choropleth	LG EM NW ES CS ZH TI	20 30 40 40 40 30 40	10 10 10 30 10 10 20	30 20 20 10 20 20 10	40 40 30 20 30 40 30	2 2 2 1 1 2 2	1 1.5 0.5 0.5 0.5 1.5 1	0.5 0.5 1 1.5 1.5 0.5 0.5	1.5 1 1.5 2 2 1 1.5	CS	84.18 ± 5.43	25.23 ± 2.11
		Region with 20% share and longest travel time for by car/motor bike	3D pie	choropleth	LG EM NW ES CS ZH TI	20 40 40 30 40 40 40	10 10 30 10 10 20 10	30 20 10 20 20 10 20	40 30 20 40 30 30 30	2 1.5 1 2 2 2 1	1 0.5 0.5 1.5 0.5 1 0.5	0.5 1 1.5 0.5 1 0.5 1.5	1.5 2 2 1 1.5 1.5 2	NW	80.79 ± 5.86	24.86 ± 2.42

† one outlier removed (the subject needed more than seven minutes to answer the question)

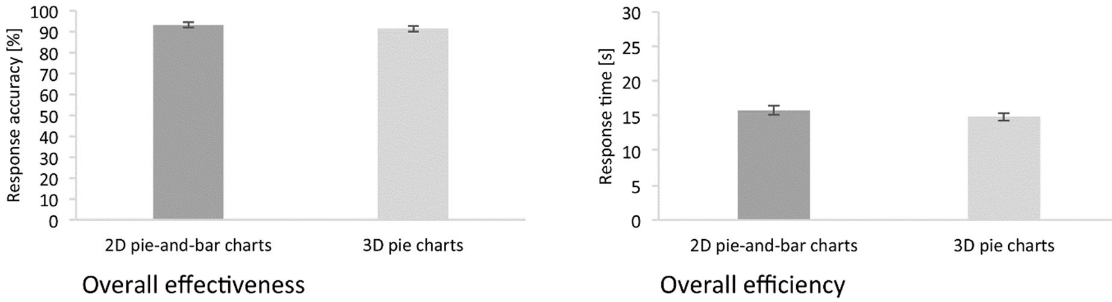

Figure 4: Mean response accuracies (*left*) and response times (*right*) of questions involving 2D pie and bar charts in adjacent frames and 3D pie charts in a single frame over all tasks. Results of tasks 1 and 2 (N = 6 x 86) got more weight in Welch's t-test than results of task 3 and 4 (N = 6 x 177) to balance the different response counts as only participants of group B solved Tasks 1 and 2 with 2D pie-and-bar charts as well as 3D pie charts. Error bars: ±SEM.

Effects of task type on performance

Figure 5 depicts the results at the task level for participants who solved all questions with 2D pie-and-bar and 3D pie charts (Group B). At the *left* of *Figure 5*, we see that, irrespective of chart type, the accuracy levels drop as we go from simpler tasks to more complex ones (see *Table 1* for their descriptions). Similarly, at the *right* of *Figure 5*, we observe that the response time increases in the same direction, confirming that task complexity is highest for the map-related task (Task 4). When we examine the plots for participants' performance differences based on chart type, a Welch's t-test shows that subjects are faster in the map-related task ($t(257) = 4.6$, $p < .001$, $d = 0.29$) with 3D pie charts ($M = 22.64$s, $SD = 11.38$s) than with the 2D pie-and-bar charts ($M = 25.62$s, $SD = 12.64$s). No other results show significant effects (*Table 3*). We considered only the results of the first three tasks in this analysis, because the background map is relevant for solving Task 4, therefore making it inherently different from the other tasks. Results did not appear to differ very much (*Table 4*) between background types or based on the chart types, except in one case: When the background is blank, McNemar's test confirmed that participants have a higher accuracy ($X^2(1, N = 349) = 12.41$, $p < .001$, $\varphi = 0.19$) with 2D pie-and-bar charts ($M = 95.70\%$, $SD = 20.31\%$) than 3D pie charts ($M = 89.11\%$, $SD = 31.19\%$). For Group A, we report the results in the section "Interactions between task type and practice level" because participants in this group solved only the last two tasks with 2D pie-and-bar and 3D pie charts to measure the effect of practice for this task type.

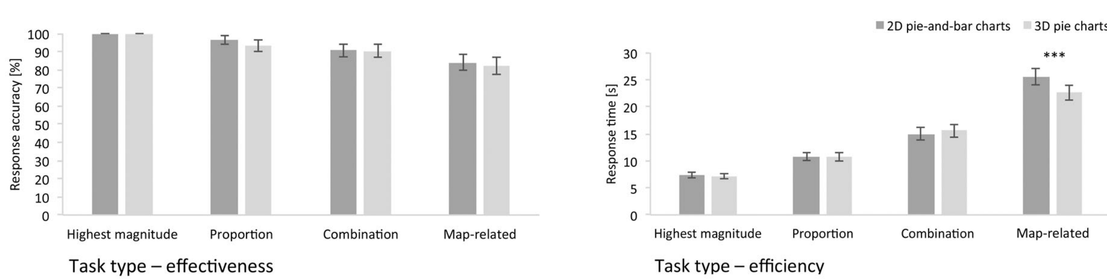

Figure 5: Mean response accuracies (*left*) and response times (*left*) for the four task types in the experiment. Results of participants with less practice are not included in this chart as they solved only the last two tasks with 2D pie and bar charts in adjacent frames and 3D pie charts in a single frame. Error bars: ±SEM. ***p < .001.

Table 3: Effects of task type on participants' performance of Group B using 2D pie-and-bar charts and 3D pie charts. Group A solved only Tasks 3 and 4 with these charts. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Task 1: Highest magnitude	100	0	100	0	0	1	0	258
Task 2: Proportion	96.12	19.34	93.02	25.52	2.72	.1	0.1	258
Task 3: Combination	90.7	29.1	90.31	29.64	0	1	0	258
Task 4: Map-related	84.11	36.63	82.17	38.35	0.31	.58	0.03	258
Response time [s]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**t**	**p**	**D**	**N**
Task 1: Highest magnitude	7.26	4.21	7.12	3.46	-0.19	.85	0.01	258
Task 2: Proportion	10.8	5.92	10.69	6.5	0.89	.37	0.06	258
Task 3: Combination	14.98	9.11	15.58	9.96	-0.36	.72	0.02	258
Task 4: Map-related	25.62	12.64	22.64	11.38	4.6	**< .001**	0.29	258

Effects of background complexity on performance

Figure 6 illustrates the participants' performance with 2D and 3D chart setups as the backgrounds vary.

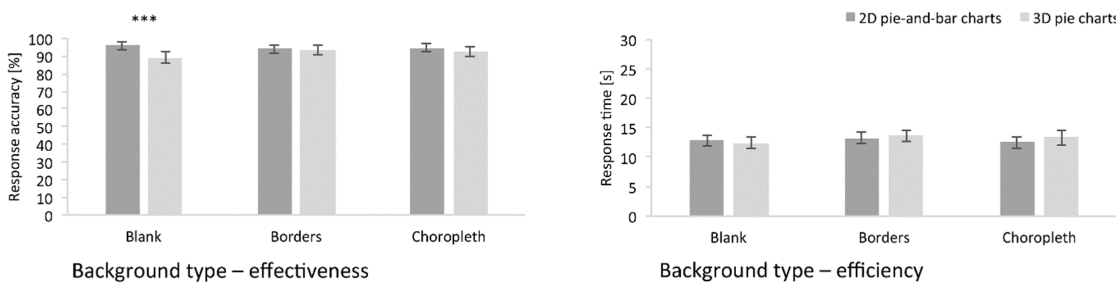

Figure 6: Mean response accuracies (*left*) and response times (*right*) for different backgrounds shown for each visualization type. Results of Task 4 (i.e. the map-related task) are not included in this chart, as the task could be performed only on a choropleth map and not on other backgrounds. Error bars: ±SEM. ***p < .001.

Table 4: Effects of background complexity on participants' performance using 2D pie-and-bar charts and 3D pie charts in Tasks 1 to 3. Task 4 was performed on a choropleth map only. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Blank map	95.7	20.31	89.11	31.19	12.41	**< .001**	0.19	349
Border map	93.98	23.81	93.41	24.85	0.04	.84	0.01	349
Choropleth map	94.56	22.72	92.55	26.3	1.16	.28	0.06	349
Response time [s]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**t**	**p**	**d**	**N**
Blank map	12.84	8.7	12.45	9.53	1.15	.25	0.06	349
Border map	13.27	10	13.67	9.41	-1.12	.27	0.06	349
Choropleth map	12.49	9.18	13.3	11.6	-0.52	.61	0.03	349

Effects of practice level on performance

To understand if the performance improves with practice or if the differences based on chart type might disappear once the participants have had a chance to practice, we analyzed the differences between the two groups of participants—one group with less practice (Group A), and another with more practice (Group B). Even though descriptive statistics suggest that participants' response accuracy is slightly better with the 2D chart combination and that this difference disappears with practice (*Figure 7, left*), the differences are statistically significant for neither the less practiced nor the more practiced participants (*Table 5*).

Table 5: Effects of practice level on participants' performance using 2D pie-and-bar charts and 3D pie charts in Tasks 3 and 4. Tasks 1 and 2 were solved with these chart types by Group B only. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Group A: Less practice	88.64	31.76	85.9	34.84	2.06	.15	0.06	546
Group B: More practice	87.4	33.21	86.24	34.48	0.31	.58	0.02	516
Response time [s]								
Group A: Less practice	24.62	16.31	22.48	14.88	3.55	**< .001**	0.15	546
Group B: More practice	20.3	12.22	19.11	11.25	2.78	**< .01**	0.12	516

Response time analysis shows a different pattern: Both the less practiced and more practiced participants were faster with the 3D than the 2D chart setup (*Figure 7, right*). These differences are statistically significant both for more practiced participants ($t(515) = 2.78$, $p < .01$, $d = 0.12$): 3D pie charts ($M = 19.11$s, $SD = 11.25$s) vs. 2D pie-and-bar charts ($M = 20.3$s, $SD = 12.22$s); and for less practiced participants ($t(545) = 3.55$, $p < .001$, $d = 0.15$): 3D pie charts ($M = 22.48$s, $SD = 14.88$s) vs. 2D pie-and-bar charts ($M = 24.62$s, $SD = 16.31$s).

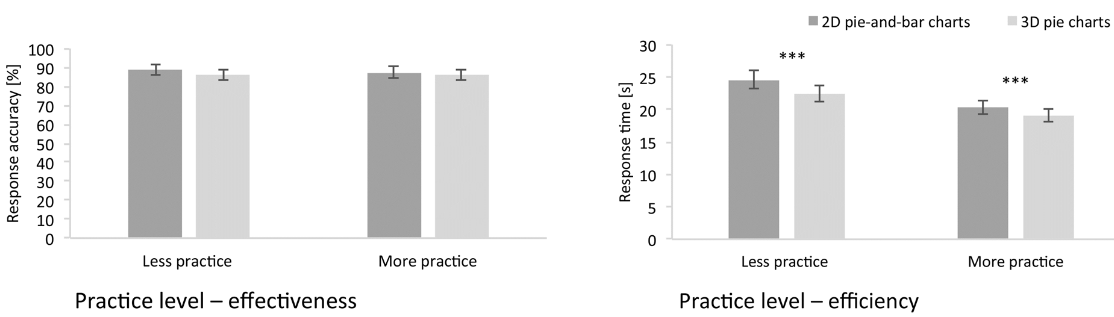

Figure 7: Mean response accuracies (*left*) and response times (*right*) of participants with less practice and more practice. The first two tasks are not included in this chart because less practiced participants solved these tasks with 2D charts in single map frames only. Error bars: ±SEM. **p < .01, ***p < .001.

Interactions between task type and background complexity

As mentioned earlier, we varied backgrounds for the first three tasks and compared the participants' performance with 2D and 3D charts against these backgrounds. In other words, the backgrounds themselves were introduced as 'noise' in the first three tasks. For Tasks 1 (highest magnitude) and 2 (proportion), we observed no statistically significant difference in participants' performance in terms of accuracy or response time with the 2D and 3D charts on individual background types (*Table 6*).

Table 6: Interactions between task type and background complexity for 2D pie-and-bar charts and 3D pie charts. Task 4 was performed on a choropleth map only. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Task 1: Highest magnitude + Blank map	100	0	100	0	0	1	0	86
Task 1: Highest magnitude + Border map	100	0	100	0	0	1	0	86
Task 1: Highest magnitude + Choropleth map	100	0	100	0	0	1	0	86
Task 2: Proportion + Blank map	94.19	23.54	88.37	32.24	1.45	.23	0.13	86
Task 2: Proportion + Border map	95.35	21.18	94.19	23.54	0	1	0	86
Task 2: Proportion + Choropleth map	98.84	10.78	96.51	18.46	0.5	.48	0.08	86
Task 3: Combination + Blank map	94.35	23.15	84.18	36.6	10.32	**< .001**	0.24	177
Task 3: Combination + Border map	90.4	29.55	89.83	30.31	0	1	0	177
Task 3: Combination + Choropleth map	89.83	30.31	87.01	33.72	0.55	.46	0.06	177
Task 4: Map-related + Choropleth map	84.56	36.17	85.12	35.62	0.04	.84	0.01	531
Response time [s]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**t**	**p**	**d**	**N**
Task 1: Highest magnitude + Blank map	7.15	2.94	6.84	2.31	0.48	.63	0.05	86
Task 1: Highest magnitude + Border map	7.2	5.28	7.41	4.54	-0.94	.35	0.1	86
Task 1: Highest magnitude + Choropleth map	7.44	4.12	7.09	3.18	0.3	.76	0.03	86
Task 2: Proportion + Blank map	11.13	7.09	11.09	7.07	0.23	.82	0.02	86
Task 2: Proportion + Border map	11.7	5.2	11.76	7.15	0.51	.61	0.06	86
Task 2: Proportion + Choropleth map	9.57	5.11	9.22	4.8	0.86	.39	0.09	86
Task 3: Combination + Blank map	16.43	9.54	15.83	11.18	1.2	.23	0.09	177
Task 3: Combination + Border map	16.98	11.75	17.64	10.18	-1.34	.18	0.1	177
Task 3: Combination + Choropleth map	16.36	10.68	18.3	14.07	-1.61	.11	0.12	177
Task 4: Map-related + Choropleth map	28.45	15.61	24.43	13.7	7.7	**< .001**	0.33	531

For Task 3 (the combination task), participants were more accurate in identifying the proportion of the 2D pie chart sector and the highest 2D bar (M = 94.35%, SD = 23.15%) than in specifying the proportion of the most extruded 3D pie sector (M = 84.18%, SD = 36.60%) on a blank map ($X^2(1, N = 177) = 10.32$, $p < .001$, φ = 0.24). No interactions were observed between chart types on border and choropleth maps for this task (*Table 6*). Also, as mentioned earlier, participants performed Task 4 (the map-related task) only with a choropleth background. We found no statistically significant differences in participants' response accuracy between the 2D and 3D chart setup with Task 4 (*Table 6*). However, participants solved Task 4 faster ($t(530) = 7.7$, $p < .001$, $d = 0.33$) with 3D pie charts (M = 24.43s, SD = 13.70s) than with 2D pie-and-bar charts (M = 28.45s, SD = 15.61s).

Interactions between task type and practice level

For the simpler tasks (Task 1: highest magnitude; Task 2: proportion), there were no interactions, because those tasks were meant to allow one group to practice more with 2D pie-and-bar and 3D pie charts than the other group. For Task 3 (the combination task), participants with less practice solved questions with 2D pie-and-charts (M = 92.31%, SD = 26.70%) more accurately ($X^2(1, N = 273) = 10.3$, $p < .001$, $\varphi = 0.19$) than with 3D pie charts (M = 83.88%, SD = 36.84%). Among the participants with more practice, we did not find any statistically significant differences concerning accuracy rates for the combination task (*Table 7*).

Table 7: Interactions between task type and practice level for 2D pie-and-bar charts and 3D pie charts. Tasks 1 and 2 were solved with these chart types by Group B only. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Task 3: Combination + Group A: Less practice	92.31	26.7	83.88	36.84	10.3	**< .001**	0.19	273
Task 3: Combination + Group B: More practice	90.7	29.1	90.31	29.64	0	1	0	258
Task 4: Map-related + Group A: Less practice	84.98	35.79	87.91	32.66	1.02	.31	0.06	273
Task 4: Map-related + Group B: More practice	84.11	36.63	82.17	38.35	0.31	.58	0.03	258
Response time [s]								
Task 3: Combination + Group A: Less practice	18.11	11.79	18.85	13.39	-1.1	.27	0.07	273
Task 3: Combination + Group B: More practice	14.98	9.11	15.58	9.96	-0.36	.72	0.02	258
Task 4: Map-related + Group A: Less practice	31.13	17.58	26.12	15.41	6.23	**< .001**	0.38	273
Task 4: Map-related + Group B: More practice	25.62	12.64	22.64	11.38	4.6	**< .001**	0.29	258

Response times appear to be similar among participants of both groups with 2D and 3D charts for Task 3. With Task 4 (the map-related task), participants' response accuracy did not differ statistically significantly between the charts (*Table 7*). However, participants solved the questions faster with 3D pie charts than with the 2D chart combination irrespective of their practice levels: Less practiced participants were faster ($t(272) = 5.13$, $p < .001$, $d = 0.31$) with 3D pie charts (M = 26.12s, SD = 15.41s) than with 2D pie-and-bar charts (M = 31.13s, SD = 17.58s); the same results were seen for the more practiced participants ($t(257) = 3.51$, $p < .001$, $d = 0.22$; $M_{3D} = 22.64$s, $SD_{3D} = 11.38$s; $M_{2D} = 25.62$s, $SD_{2D} = 12.64$s).

Interactions between task type, background complexity, and practice level

As backgrounds were varied in the first three tasks and participants with different practice levels had the same questions in the last two tasks, interactions between background visualizations and practice levels could have only occurred in Task 3, that is, the combination task. Here, only participants with less practice ($X^2(1, N = 91) = 13.47$, $p < .001$, $\varphi = 0.38$) were more accurate with 2D pie-and-bar charts (M = 95.60%, SD = 20.61%) than with 3D pie charts (M = 76.92%, SD = 42.37%) on the blank map, and not the participants with more practice

(*Table 8*). We did not find any interactions between the chart types on other background types (border and choropleth) among participants with different practice levels.

Table 8: Interactions between task type, background complexity, and practice level for 2D pie-and-bar charts and 3D pie charts. Tasks 1 and 2 were solved with these chart types by Group B, and Task 4 was performed by both groups on a choropleth map. Differences in response accuracy were calculated by McNemar's test and differences in response time were computed by Welch's t-test.

Response accuracy [%]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**X^2**	**p**	**φ**	**N**
Task 3: Combination + Blank map + Group A: Less practice	95.6	20.61	76.92	42.37	13.47	**< .001**	0.38	91
Task 3: Combination + Border map + Group A: Less practice	91.21	28.47	89.01	31.45	0.08	.77	0.03	91
Task 3: Combination + Choropleth map + Group A: Less practice	90.11	30.02	85.71	35.19	0.56	.45	0.08	91
Task 3: Combination + Blank map + Group B: More practice	93.02	25.62	91.86	27.5	0	1	0	86
Task 3: Combination + Border map + Group B: More practice	89.53	30.79	90.7	29.22	0	1	0	86
Task 3: Combination + Choropleth map + Group B: More practice	89.53	30.79	88.37	32.24	0	1	0	86
Response time [s]	**M_{2D}**	**SD_{2D}**	**M_{3D}**	**SD_{3D}**	**t**	**p**	**d**	**N**
Task 3: Combination + Blank map + Group A: Less practice	18.02	9.95	17.82	12.83	0.5	.62	0.05	86
Task 3: Combination + Border map + Group A: Less practice	18.35	12.1	19.03	11.41	-1.04	.3	0.11	86
Task 3: Combination + Choropleth map + Group A: Less practice	17.96	13.22	19.7	15.67	-1.38	.17	0.14	86
Task 3: Combination + Blank map + Group B: More practice	14.75	8.83	13.73	8.72	1.21	.23	0.13	91
Task 3: Combination + Border map + Group B: More practice	15.52	11.26	16.18	8.52	-0.85	.4	0.09	91
Task 3: Combination + Choropleth map + Group B: More practice	14.66	6.75	16.82	12.07	-0.89	.38	0.1	91

Participants' comments

Two participants commented in an optional text area for feedback that 3D pie charts are practical because everything is included in one chart and not split into two. Four participants remarked that it is difficult to read 3D pie sector proportions. One participant gave a negative comment about using 3D visualizations in general.

Discussion

Before we interpret the observations presented in this manuscript, it is worth noting that, overall, the success rates are quite high (80–100%) in our study, in which 3D pie charts with individually extruded sectors were compared against 2D pie-and-bar charts in adjacent frames. On the one hand, these high success rates indicate that the participants took the survey seriously (i.e. they did not guess but solved the tasks). On the other hand, we see a *ceiling effect*, especially for Task 1 (i.e. estimating the highest magnitude), where 100% of the participants were able to give the right answer. This may partly be explained by the way the task was designed: We gave our participants four options to choose from, which may have made the task too easy to detect any performance differences between 2D pie-and-bar and 3D pie charts, and which is in contrast to Cleveland and McGill's experiment (1984) where participants were asked to respond with exact values. We opted for this task design because, as Spence and Lewandowsky (1991) posited, the purpose of charts is not in telling exact values (tables should be used for such tasks) but to estimate approximate relative proportions. Our results thus provide evidence that the chart types in this study worked well for such a task.

The response times varied considerably over the entire experiment: Participants took 5.4 seconds for the easiest question and 29.7 seconds for the hardest question on average. We believe that these differences in response times are linked to task complexity—the easiest question required only one cognitive action (i.e. identification of the highest magnitude), whereas the hardest question required six mental steps (i.e. identification of the highest magnitude and proportion, color-coding and -encoding, spatial comparison, and referencing). We speculate that the number of operations involved in harder tasks must have led to considerably higher cognitive load, which, in turn, would explain this gap in response times. Even though we examined outliers and found only one extreme value for the proportion task with 2D pie charts in a single frame (*Table 2*), which was not relevant for the results we presented, it is still possible that our participants could have been distracted during the online survey. If they did, the distractions most likely occurred randomly, and thus would be distributed on different conditions. Hence, there is no reason to believe that such interference would have favored a particular variable that we examined. Similar to Fischer et al. (2005), we had breaks after every four questions, which we believe should have helped to counter such disturbance and also fatigue; however, overall response times may be lower in a controlled setting. Below, we provide a more fine-grain interpretation of our findings.

Task type and its interactions with other variables

Among the four task types examined, we formulated a two-level hypothesis regarding participants' performance: For the first two tasks types, we assumed that the 2D pie-and-bar charts should facilitate overall better (i.e. more accurate and faster) information extraction, because only one size variable (pie sector proportion *or* bar/extruded pie sector height) is tested in these tasks. For the last two tasks, however, participants interpreted two size variables (pie sector proportion *and* bar/extruded pie sector height). For these tasks (Tasks 3 and 4), 3D pie charts could offer advantages over 2D pie-and-bar charts, because the 3D setup should mitigate the potential split-attention effects by bringing the multivariate information closer together than the 2D setup, thus restricting the search space. Our results demonstrate that the differences in participants' performance are not as 'clear-cut' as we expected between the two chart setups for many tasks.

Reported as an interaction between the task types and background complexity, we see that in only one of the four tasks (Task 3, the combination task), participants solved one question (i.e. blank background) using the 2D pie-and-bar charts with a higher accuracy than with the 3D pie charts. We believe that this result is more relevant for the background complexity (elaborated more in the next section) than the task type. Nonetheless, if we look at this result from the perspective of tasks types, we see that this task required the reading of two size variables, which contradicts our hypothesis that 3D pie charts should be better suited for tasks because of the compact representation of the information. For the other three tasks, participants' success rates did not differ between 2D and 3D chart setups. While interpreting "no difference" is statistically not justified, the fact that we do not observe a difference in these cases seem to be in contradiction with some earlier findings (e.g. Siegrist, 1996), where participants were less successful in solving various chart reading tasks with 3D pie charts with equally extruded sectors. A more detailed examination of the results (*Table 2*) reveals that participants' accuracy in estimating proportions with single 2D pie charts (Group A) and 3D pie charts (Group B) on a blank background for Task 2 matches with results of previous studies regarding depth cues in charts, which makes sense because the earlier work also used blank backgrounds.

General change in task difficulty remains only a descriptive observation for response times in Tasks 1, 2, and 3, that is, there were no significant differences in participants' response times between 2D and 3D chart setups in these tasks. In Task 4, in accordance with our hypothesis, participants solved the more complex map-related task faster with 3D pie charts than with 2D pie-and-bar charts. A reason for this may be the distance between the pie chart and the bar chart in the side-by-side representation, which would require more eye movement to interrelate corresponding values of the two charts. The larger difference in response times for more difficult tasks is consistent with the results of Stewart, Cipolla, and Best (2009) with one important difference: participants are faster with 3D charts in our study–aside from the observations of 'no difference' which we cautiously interpret as "at least as fast". This might be because the information depicted by the extruded sectors of 3D pie charts is necessary to solve the task. Note that we do not observe a speed-accuracy tradeoff (i.e. participants did not rush to finish the experiment and took the time they needed to work out the answer); thus, we believe that the interpretation of the response time as a performance measure is meaningful.

Background complexity and its interactions with other variables

Previous studies comparing chart types have been typically conducted on blank backgrounds (e.g. Rangecroft, 2003), whereas this is rarely the case when the charts are used with maps (e.g. in atlases). Thus, we tested the effect of having some visual noise in the background. Note that we use the word 'noise' because the background maps in Tasks 1, 2, and 3 were irrelevant for solving the tasks, and these are the tasks in which we have a comparative measure of how background complexity affects the performance outcome. We expected any perceptual advantages offered by either chart setup—2D or 3D—to be amplified when there was noise in the background, because the participants would need all the assistance they could get in the presence of extraneous information. We did not expect differences in time taken by participants in locating the charts with different backgrounds, because the charts were quite salient and easy to detect in all conditions: 3D charts were shaded and 2D charts were outlined, and both chart types were placed in the center of the maps. Here, we will discuss the results only at an aggregate level and not between individual backgrounds per task type, since we also changed chart proportions in the questions to counter-balance for learning effects.

On the whole, varying the background complexity did not create a statistically significant effect in participants' response times with the 2D and 3D charts in this experiment. In an earlier study, Neider and Zelinsky (2006) reported similar results for a visual search task where neutral and 'camouflage' backgrounds caused no difference in efficiency (they did not report effectiveness). However, importantly, and as mentioned in the previous section, we observed that participants' accuracy was higher with the 2D pie-and-bar than with the 3D pie charts only on the blank background. This difference was no longer statistically significant when we introduced visual noise in the background. As opposed to our hypothesis, this finding suggests that increasing the background complexity reduces the differences between the 2D and 3D chart setups, seemingly impairing participants' performance in the 2D condition. Why should visual noise in the periphery weaken performance with 2D pie-and-bar charts? We suspect that it might be related to the size of the frame in which the 2D and 3D charts were displayed in this study: 2D pie and bar charts were presented in two adjacent frames, covering a considerably larger space than the 3D pie charts on the display. This possibly forced the participants to use more information from the peripheral vision as they needed to compare the pie chart with the bar chart, which was placed away. Thus, the presence of visual noise in this operation might have increased cognitive load, as one must actively ignore the incoming signal. The 3D pie chart, on the other hand, was compact in presentation and thus possibly did not require processing peripheral (or perhaps even parafoveal) visual information at all (Brychtová & Çöltekin, 2017). A future study including eye movement analyses would confirm this speculation.

Furthermore, we used a specific color configuration in this study. Following up on this work, it would be interesting to test other color configurations or distances. For example, as demonstrated by Brychtová and Çöltekin (2017) for choropleth maps, varying the color distance between chart sectors and background might lead to different results; further studies examining this aspect of the visual design would make our findings more robust.

Practice levels and its interactions with other variables

To understand if additional practice would remove some of the issues with the 'seemingly more difficult' chart types—the 2D pie-and-bar chart combination and 3D pie charts—we allowed Group B participants to work with these charts twice as often as Group A participants. Apart from improving overall performance, we speculated that any performance differences with 2D and 3D charts might disappear or at least become weaker, because as they practice, participants might develop strategies to counter the issues with the particular chart types. If participants were more experienced with one rather than the other chart type, this gap might become smaller through practice for either of the two chart types. Note that all participants had a training session at the start of the experiment; thus, Group B's exposure to twice as many tasks with 2D pie-and-bar charts and 3D pie charts was an additional chance to practice and improve. However, as opposed to the training session, Group B did not receive additional instructions or feedback for these questions.

Confirming our expectations, we observed that, overall, participants with more practice (Group B) responded faster than those with less practice (Group A), although the overall accuracy rates did not differ between the two groups (*Figure 7*). This finding confirms the common sense understanding that repeating a task is likely to increase efficiency. In some other setups, it may also lead to higher accuracy, but in our experiment, the accuracy rates were already very high, as discussed earlier.

Examining the interactions between the experimental factors revealed an interesting result related to our hypothesis on practical levels: Differences in accuracy between 2D and 3D chart setups vanished for more practiced participants compared to those with less practice for the combination task (Task 3, i.e. identifying the highest magnitude and estimating a proportion at the same time). We believe this is an important result in the 2D vs. 3D debate as it suggests that in some contexts, if people are able to practice, 3D charts might not hurt performance. Future studies examining the effectiveness of visualizations might benefit from bearing this in mind (i.e. by controlling for practice levels in the experiments). Following from this observation, it would be interesting to assess the steepness of the learning curve, the number of attempts after which the maximum performance converges, and how well participants can remember their acquired chart literacy skills.

Further remarks and future work

Below, we list some constraints to the study of which the readers should be aware when interpreting or attempting to reproduce the results. First of all, it is important to remember that participants in this study were highly educated (i.e. 63% with university degrees); therefore, our sample is likely not representative of the general public. Furthermore, our participants were possibly mostly Swiss German, meaning that our results may or may not apply to people with different cultural backgrounds. For example, results may be different if the study is performed in countries where science education or preferences for chart colors differ (Mackiewicz, 2007). As there is always an illumination model involved for 3D visualizations to produce shaded surfaces, colors in 3D pie charts were not exact matches of those in 2D charts. These differences were very subtle and arguably not perceptually relevant, and hence we believe they do not pose a limitation on our findings. Nevertheless, it would be interesting to examine whether the presence of shadow and outlines influences the performance of reading chart values. Moreover, we arranged 3D pie charts in a way that no occlusions occur between sectors. When mapping non-fictional values onto the 3D charts, it is likely that they will be less readable or even non-readable—e.g. when sectors in the foreground are larger than those in the background. Furthermore, occlusions would occur between extruded sectors in interactive 3D environments when moving and rotating the camera, whereas this is not the case for 2D charts. On the other hand, one can adjust the view in an interactive system to find the desired vantage point that enables maximum access to relevant information. Such additional action(s) would probably lower participants' performance with the 3D pie charts, which remains to be tested in a future study. Furthermore, occlusion occurs not only between the 3D pie charts sectors, but also between the charts and the underlying map. While all of the different map areas are still partly visible in our experiment, point generalization techniques (McMaster & Shea, 1992) could be applied to the charts in the cases where the charts fully cover areas. An example solution would be to displace the charts to an 'empty' space on the map and add a leader line to link the chart with the associated map area.

In many previous studies involving charts, such as summing up values in a pie chart (Hollands & Spence, 1998), tasks were only performed with one chart, but not for a chart combination (i.e. pie-and-bar charts), as in our experiment. We decided to position the 2D charts in adjacent map frames, since interactive geovisualization tools often offer a split-screen function by which different datasets can be compared. We always placed pie charts within the left frame and bar charts within the right frame in our experiment. Variations in position (e.g. top-bottom) may have led to different results. Moreover, the two static frames in this study were kept adjacent to each other, causing split attention. This effect would possibly be amplified if the information was spread to even more maps, which is the case for other high-dimensional representations such as small multiples (Tufte, 1983). Small multiples usually show only one variable per map and would probably facilitate comparison of values of the same category between map areas. However, the interrelation of different category values for a particular map area would be impeded. Other 2D chart types, e.g. wing charts (Schnabel, 2007), or interactive techniques such as data lenses or swipe tools (Lobo, Pietriga, & Appert, 2015), are possible alternatives to integrate all information in one map, thus to isolate the effect only to the change from 2D to 3D. Finding the best techniques to visualize multivariate data from a cognitive performance point of view may be a worthwhile subject to study in the future for similar configurations.

Further considering experimental control vs. ecological validity, Tasks 1 to 3 should not be regarded as 'realistic' use cases because they examine a single chart. Most real-world situations are similar to Task 4 where maps contain multiple charts. On the other hand, based on selection, filtering, or highlighting techniques that help to focus the map reader's attention on a single chart, one can argue that the first three tasks have fairly high ecological validity for digital maps. A different use case to our experiment, where participants were asked to identify the maximum value of bar charts or extruded pie chart sectors, would be to judge the values represented by height. As previous research only tested equally extruded pie chart sectors with meaningless heights, it would be interesting to verify whether the additional mental rotation to compensate the tilted view or summation of sector heights in the foreground and background of the chart decreases peoples' performance of 3D pie charts when reading height values. This question is motivated by the fact that mental rotation, especially in 3D, is known to be a difficult cognitive task for many people (Shepard & Metzler, 1988).

While questions for Tasks 1 and 2 have been constructed for theoretical purposes, questions for Tasks 3 and 4 tested participants' understanding of the context as well. In other words, in Tasks 3 and 4, participants needed to read values from the legend, which is a common task in geographic products such as atlases. The design of the questions in Tasks 3 and 4 is similar to that in Schonlau and Peters' (2012) experiment, and is recommended by Shah and Hoeffner (2002).

Conclusion

We depicted meaningful information on individually extruded pie chart sectors and compared this 3D chart type against a 2D alternative, in which pie and bar charts were placed in adjacent frames. In eight out of ten comparisons between the two chart types, participants had similar response accuracies and response times. Participants' performance did not differ between 2D and 3D chart setups for the easier tasks (i.e. Task 1, identifying the highest magnitude, and Task 2, estimating proportions). For the more difficult tasks, however, we observed nuanced differences in participants' performance with 2D and 3D chart types, which were as follows:

a) Participants were more accurate with 2D pie-and-bar charts than with 3D pie charts on a blank background when identifying the highest magnitude and estimating a proportion *at the same time* (i.e. Task 3, the combination task).
b) Participants were faster in responding with 3D pie charts than with 2D pie-and-bar charts when performing a spatial comparison task and when using the map legend *in addition to* the combination task (thus, when executing Task 4, the map-related task).

The difference in accuracy (a) vanished for participants who received more practice during the experiment, whereas the difference in efficiency (b) persisted even after gaining more practice.

Increasing the visual complexity of the background—either to border-only or to choropleth maps—also leveled out the accuracy differences, as described in (a).

Based on our results, we believe that information visualizers and map editors can use either chart setup successfully, bearing in mind the discussed experimental constraints (e.g. occlusion). The arguments and previous findings against the 3D pie charts appear to be rather about the extraneous (i.e. decorative) use of the depth cues, and not necessarily detrimental to participants' performance when the third dimension is used for representing a task-relevant variable. Overall, 3D pie charts with individually extruded sectors, as well as 2D pie and bar charts in adjacent frames, work well to represent multivariate data on static media.

Acknowledgements

We thank Lorenz Hurni for critically revising the paper.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

American Psychological Association. (2017). *Ethical Principles of Psychologists and Code of Conduct*. Retrieved from http://www.apa.org/ethics/code/

Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal*, *25*(1), 51–71. https://doi.org/10.2307/1163159

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press.

Bleisch, S., Dykes, J., & Nebiker, S. (2008). Evaluating the Effectiveness of Representing Numeric Information Through Abstract Graphics in 3D Desktop Virtual Environments. *The Cartographic Journal*, *45*(3), 216–226. https://doi.org/10.1179/000870408X311404

Brügger, A., Fabrikant, S. I., & Çöltekin, A. (2016). An empirical evaluation of three elevation change symbolization methods along routes in bicycle maps. *Cartography and Geographic Information Science*, *44*(5), 436–451. https://doi.org/10.1080/15230406.2016.1193766

Brychtová, A., & Çöltekin, A. (2017). The effect of spatial distance on the discriminability of colors in maps. *Cartography and Geographic Information Science*, *44*(3), 229–245. https://doi.org/10.1080/15230406.2016.1140074

Burch, M. (2015). The Aesthetics of Diagrams. *Proceedings of the 6th International Conference on Information Visualization Theory and Applications*, 262–267. https://doi.org/10.5220/0005357502620267

Cleveland, W. S., Harris, C. S., & McGill, R. (1982). Judgments of Circle Sizes on Statistical Maps. *Journal of the American Statistical Association*, *77*(379), 541–547. https://doi.org/10.1080/01621459.1982.10477844

Cleveland, W. S., & McGill, R. (1984). Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, *79*(387), 531–554. https://doi.org/10.1080/01621459.1984.10478080

Çöltekin, A., Bleisch, S., Andrienko, G., & Dykes, J. (2017). Persistent challenges in geovisualization–a community perspective. *International Journal of Cartography*, *3*(sup1), 115–139. https://doi.org/10.1080/23729333.2017.1302910

Elias, M., & Bezerianos, A. (2011). Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, & M. Winckler (Eds.), *Human-Computer Interaction – INTERACT 2011* (pp. 274–291). https://doi.org/10.1007/978-3-642-23768-3_23

Ervin, C. W. (2011). Pie charts in financial communications. *Information Design Journal*, *19*(3), 205–215. https://doi.org/10.1075/idj.19.3.02erv

European Parliament and the Council. (2003). *Regulation (EC) on the establishment of a common classification of territorial units for statistics (NUTS)* (No. 1059/2003). Retrieved from http://data.europa.eu/eli/reg/2003/1059/2018-01-18

Fausset, C. B., Rogers, W. A., & Fisk, A. D. (2008). *Visual Graph Display Guidelines* (Technical Report No. HFA-TR-0803). Retrieved from https://smartech.gatech.edu/bitstream/handle/1853/40573/HFA-TR-0803%20Graph_Guidelines_Final.pdf

Fischer, M. H. (2000). Do irrelevant depth cues affect the comprehension of bar graphs? *Applied Cognitive Psychology*, *14*(2), 151–162. https://doi.org/10.1002/(SICI)1099-0720(200003/04)14:2<151::AID-ACP629>3.0.CO;2-Z

Fischer, M. H., Dewulf, N., & Hill, R. L. (2005). Designing bar graphs: Orientation matters. *Applied Cognitive Psychology*, *19*(7), 953–962. https://doi.org/10.1002/acp.1105

Heer, J., Bostock, M., & Ogievetsky, V. (2010). A tour through the visualization zoo. *Magazine Communications of the ACM*, *53*(6), 59–67. http://doi.acm.org/10.1145/1743546.1743567

Hollands, J. G., & Spence, I. (1992). Judgments of Change and Proportion in Graphical Perception. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(3), 313–334. https://doi.org/10.1177/001872089203400306

Hollands, J. G., & Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology*, *12*(2), 173–190. https://doi.org/10.1002/(SICI)1099-0720(199804)12:2<173::AID-ACP499>3.0.CO;2-K

Hughes, B. M. (2001). Just Noticeable Differences in 2D and 3D Bar Charts: A Psychophysical Analysis of Chart Readability. *Perceptual and Motor Skills*, *92*(2), 495–503. https://doi.org/10.2466/pms.2001.92.2.495

Hurni, L. (2017). *Schweizer Weltatlas*. Zürich: Schweizerische Konferenz der kantonalen Erziehungsdirektoren (EDK).

Ishihara, S. (1917). *Tests for color-blindness*. Tokyo: Kanehara Shuppan.

Kosara, R. (2019). Evidence for Area as the Primary Visual Cue in Pie Charts. *IEEE VIS Short Paper Proceedings*. Presented at the IEEE VIS, Vancouver. https://doi.org/10.31219/osf.io/fcna4

Kosara, R., & Skau, D. (2016). Judgment Error in Pie Chart Variations. *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers*, 91–95. https://doi.org/10.2312/eurovisshort.20161167

Kraak, M.-J. (1988). *Computer-assisted cartographical three-dimensional imaging techniques* (Dissertation). Delft University of Technology, Delft.

Kreuseler, M. (2000). Visualization of geographically related multidimensional data in virtual 3D scenes. *Computers & Geosciences*, *26*(1), 101–108. https://doi.org/10.1016/S0098-3004(99)00036-9

Krishnamoorthy, S., & North, C. (2005). Learnability of interactive coordinated-view visualizations. *Proceedings of the Ninth International Conference on Information Visualisation*, 306–311. https://doi.org/10.1109/IV.2005.70

Lawrence, R. J. (1988). The Lognormal as Event-Time Distribution. In *Lognormal Distributions: Theory and Applications* (1st ed., p. 387). New York: Routledge.

Lee, S., Kim, S., & Kwon, B. C. (2017). VLAT: Development of a Visualization Literacy Assessment Test. *IEEE Transactions on Visualization and Computer Graphics*, *23*(1), 551–560. https://doi.org/10.1109/TVCG.2016.2598920

Lewandowsky, S., Herrmann, D. J., Behrens, J. T., Li, S.-C., Pickle, L., & Jobe, J. B. (1993). Perception of clusters in statistical maps. *Applied Cognitive Psychology*, *7*(6), 533–551. https://doi.org/10.1002/acp.2350070606

Lobo, M.-J., Pietriga, E., & Appert, C. (2015). An Evaluation of Interactive Map Comparison Techniques. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3573–3582. https://doi.org/10.1145/2702123.2702130

MacEachren, A. M. (1982). The Role of Complexity and Symbolization Method in Thematic Map Effectiveness. *Annals of the Association of American Geographers*, *72*(4), 495–513. https://doi.org/10.1111/j.1467-8306.1982.tb01841.x

Mackiewicz, J. (2007). Perceptions of Clarity and Attractiveness in PowerPoint Graph Slides. *Technical Communication*, *54*(2), 145–156. http://www.ingentaconnect.com/contentone/stc/tc/2007/00000054/00000002/art00002

McMaster, R. B., & Shea, K. S. (1992). *Generalization in Digital Cartography*. Washington, DC: Association of American Geographers.

Meyer, J., Shinar, D., & Leiser, D. (1997). Multiple Factors that Determine Performance with Tables and Graphs. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *39*(2), 268–286. https://doi.org/10.1518/001872097778543921

Minard, C. J. (1858). *Carte figurative et approximative des quantités de viandes de boucherie envoyées sur pied par les départements et consommateurs à Paris*. Retrieved from https://commons.wikimedia.org/wiki/File:Minard-carte-viande-1858.png

Neider, M. B., & Zelinsky, G. J. (2006). Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research*, *46*(14), 2217–2235. https://doi.org/10.1016/j.visres.2006.01.006

Neider, M. B., & Zelinsky, G. J. (2011). Cutting through the clutter: Searching for targets in evolving complex scenes. *Journal of Vision*, *11*(14), 1–16. https://doi.org/10.1167/11.14.7

Piaget, J., & Inhelder, B. (1956). *The Child's Conception of Space.* New York, NY: Humanities Press, Inc.

Pinker, S. (1990). A Theory of Graph Comprehension. In R. Freedle (Ed.), *Artificial Intelligence and the Future of Testing* (pp. 73–126). Hillsday, NJ: Lawrence Erlbaum Associates, Inc.

Rangecroft, M. (2003). As easy as pie. *Behaviour & Information Technology*, *22*(6), 421–426. https://doi.org/10.1080/01449290310001615437

Schnabel, O. (2007). *Benutzerdefinierte Diagrammsignaturen in Karten, Konzepte, Formalisierung und Implementationen* (Dissertation, ETH Zurich). Retrieved from https://doi.org/10.3929/ethz-a-005345413

Schnur, S., Bektaş, K., & Çöltekin, A. (2017). Measured and perceived visual complexity: A comparative study among three online map providers. *Cartography and Geographic Information Science*, *45*(3), 238–254. https://doi.org/10.1080/15230406.2017.1323676

Schnürer, R., Eichenberger, R., Sieber, R., & Hurni, L. (2015). 3D Charts – Taxonomy and Implementation in a Virtual Globe. *Revista Brasileira de Cartografia*, *67*(5). http://www.rbc.lsie.unb.br/index.php/rbc/article/view/1494

Schonlau, M., & Peters, E. (2012). Comprehension of Graphs and Tables Depend on the Task: Empirical Evidence from Two Web-Based Studies. *Statistics, Politics, and Policy*, *3*(2). https://doi.org/10.1515/2151-7509.1054

Seipel, S., & Carvalho, L. (2012). Solving Combined Geospatial Tasks Using 2D and 3D Bar Charts. *2012 16th International Conference on Information Visualisation*, 157–163. https://doi.org/10.1109/IV.2012.36

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review*, *14*(1), 47–69. https://doi.org/10.1023/A:1013180410169

Shepard, S., & Metzler, D. (1988). Mental Rotation: Effects of Dimensionality of Objects and Type of Task. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 3–11.

Sieber, R., Schnürer, R., Eichenberger, R., & Hurni, L. (2013). The Power of 3D Real-Time Visualization in Atlases – Concepts, Techniques and Implementation. *Proceedings of the 26th International Cartographic Conference*. Presented at the 26th International Cartographic Conference, Dresden (Germany). Retrieved from http://icaci.org/files/documents/ICC_proceedings/ICC2013/_extendedAbstract/178_proceeding.pdf

Sieber, R., Serebryakova, M., Schnürer, R., & Hurni, L. (2016). Atlas of Switzerland Goes Online and 3D—Concept, Architecture and Visualization Methods. In G. Gartner, M. Jobst, & H. Huang (Eds.), *Progress in Cartography* (pp. 171–184). Springer International Publishing.

Siegrist, M. (1996). The use or misuse of three-dimensional graphs to represent lower-dimensional data. *Behaviour & Information Technology*, *15*(2), 96–100. https://doi.org/10.1080/014492996120300

Simkin, D., & Hastie, R. (1987). An Information-Processing Analysis of Graph Perception. *Journal of the American Statistical Association*, *82*(398), 454–465. https://doi.org/10.1080/01621459.1987.10478448

Slocum, T. A. (1981). Analyzing The Communicative Efficiency Of Two-sectored Pie Graphs. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *18*(3), 53–65. https://doi.org/10.3138/462R-12GP-G723-8033

Spence, I. (2005). No Humble Pie: The Origins and Usage of a Statistical Chart. *Journal of Educational and Behavioral Statistics*, *30*(4), 353–368. https://doi.org/10.3102/10769986030004353

Spence, I., & Lewandowsky, S. (1991). Displaying proportions and percentages. *Applied Cognitive Psychology*, *5*(1), 61–77. https://doi.org/10.1002/acp.2350050106

Stewart, B. M., Cipolla, J. M., & Best, L. A. (2009). Extraneous information and graph comprehension: Implications for effective design choices. *Campus-Wide Information Systems*, *26*(3), 191–200. https://doi.org/10.1108/10650740910967375

Tufte, E. R. (1983). *The Visual Display of Quantitative Information* (1st ed.). Cheshire, CT: Graphics Press.

Wilkinson, L. (1999). *The Grammar of Graphics*. New York, NY: Springer Science+Business Media.

Wilkinson, L. (2001). Graphical Methods: Presentation. In P. B. Baltes (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 6368–6379). Oxford: Pergamon.

Zacks, J., Levy, E., Tversky, B., & Schiano, D. J. (1998). Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, *4*(2), 119–138. https://doi.org/10.1037/1076-898X.4.2.119