Investigating time series visualisations to improve the user experience

Citation for published version:
Adnan, M, Just, M & Baillie, L 2016, Investigating time series visualisations to improve the user experience. in *CHI'16: 2016 CHI Conference on Human Factors in Computing Systems Proceedings.* Association for Computing Machinery, pp. 5444-5455 , 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, United States, 7/05/16. https://doi.org/10.1145/2858036.2858300

Digital Object Identifier (DOI):
10.1145/2858036.2858300

Document Version:
Peer reviewed version

Published In:
CHI'16: 2016 CHI Conference on Human Factors in Computing Systems Proceedings

Investigating time series visualisations to improve the user experience

Muhammad Adnan
University of Leeds
Leeds, UK
m.adnan1@hotmail.com

Mike Just
Heriot-Watt University
Edinburgh, UK
m.just@hw.ac.uk

Lynne Baillie
Heriot-Watt University
Edinburgh, UK
l.baillie@hw.ac.uk

ABSTRACT
Research on graphical perception of time series visualisations has focused on visual representation, and not on interaction. Even for visual representation, there has been limited study of the impact on users of visual encodings and the strengths and weaknesses of Cartesian and Polar coordinate systems. In order to address this research gap, we performed a comprehensive graphical perception study that measured the effectiveness of time series visualisations with different interactions, visual encodings and coordinate systems for several tasks. Our results show that, while positional and colour visual encodings were better for most tasks, area visual encoding performed better for data comparison. Most importantly, we identified that introducing interactivity within time series visualisations considerably enhances the user experience, without any loss of efficiency or accuracy. We believe that our findings can greatly improve the development of visual analytics tools using time series visualisations in a variety of domains.

Author Keywords
Visualisation; graphical perception; evaluation; interaction technique; visual encoding; coordinate system; time series.

ACM Classification Keywords
H.5.2. [Information interfaces and presentation (e.g., HCI)]: User Interfaces---Evaluation/methodology, Interaction styles (e.g., commands, menus, forms, direct manipulation).

INTRODUCTION
The staggering volume of continuously generated data today creates extraordinary opportunities for businesses to improve decision making with deeper insights into their data. However, most organisations collect more data than they can analyse and present in a meaningful way [19,28]. This has led to an increasing interest and effort from both industry [40] and academia [15] in developing usable and scalable visual analytics (VA) solutions. VA is commonly defined as: "the science of analytical reasoning facilitated by interactive visual interfaces" [37, p.4]. Whether simple or complex, a VA system is essentially composed of two main components: *visual representation* and *interaction* [39]. The former deals with the mapping of underlying data to alternative visual representations, while the latter facilitates the dialog between the user and the VA system. Despite consistently stating the value and importance of interaction for visual data analysis [12,37], the focus of the research by much of the VA community has been on the visual representation of data rather than on the interaction, as has been noted by others [2,24,30].

Graphical perception is defined as the ability of users to interpret the visual encoding and thereby understand the information presented in a graph [8]. The aforementioned trend of focusing more on visual representation than on interaction is also predominant in graphical perception of time series visualisations, which is a well-studied area in visual analytics. A common practice in quantitative graphical perception studies of time series visualisations [4,10,14,16,21] is to conduct them in a static setting, i.e., not allowing the users to interact with the time series visualisations, thereby limiting our knowledge of the user experience. In addition, the overall coverage of different visual representations of time series data is lacking. Several previous studies [4,10,14,16,21] have compared the effectiveness of time series visualisations within and across the categories of positional (e.g., Figure 1(a&b)) and colour visual encodings (e.g., Figure 1(c&d)). However, to the best of our knowledge, there has been no study that compares the effectiveness of time series visualisations that use area visual encoding (e.g., Figure 1(e&f)) with their positional and colour counterparts. We use the term 'area' throughout even though we encode area-based visualisations using only segment (or arc) length, i.e., the width is fixed. Area does consist of length and width, however due to a relatively large segment/arc width, we believe it is more accurate to describe our testing as the stimulus of area. Similar to visual encodings, limited empirical evidence is available for the strengths and weaknesses of Cartesian (e.g., Figure 1(a,c&e)) and Polar (e.g., Figure 1(b,d&f)) coordinate systems for time series visualisations, using different visual encodings.

To address this research gap, we performed a comprehensive graphical perception study that measured

CHI'16, May 07-12, 2016, San Jose, CA, USA
© 2016 ACM. ISBN 978-1-4503-3362-7/16/05…$15.00
DOI: http://dx.doi.org/10.1145/2858036.2858300

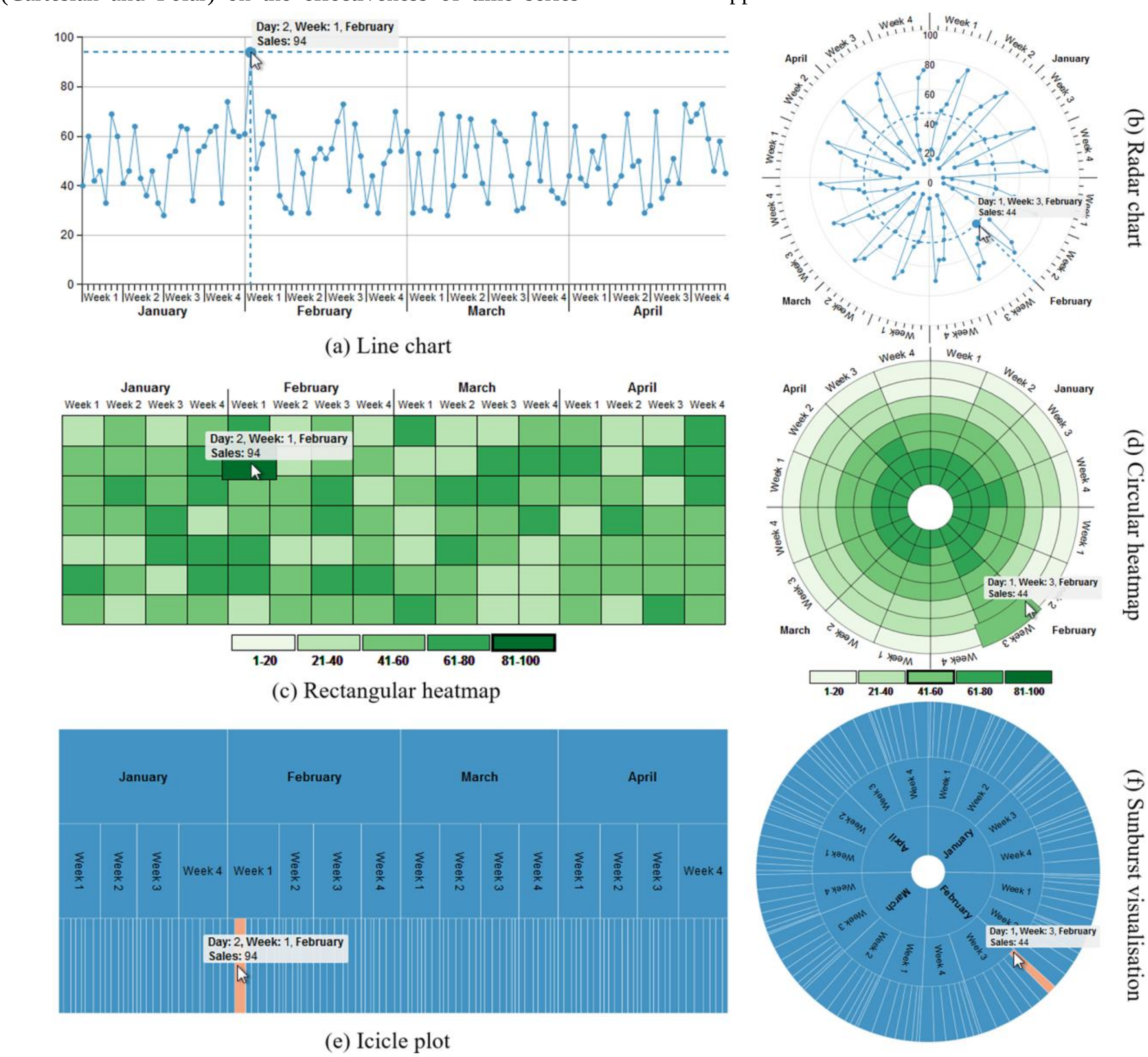
the effectiveness (i.e., completion time, accuracy, confidence and ease of use) of time series visualisations with different interactions, visual encodings and coordinate systems for several tasks. Our work makes three primary contributions:

C1: We systematically examine the effects of two commonly used interaction techniques (i.e., highlighting and tooltips) on the effectiveness of different time series visualisations.

C2: We compare the effectiveness of time series visualisations that use three different visual encoding techniques: position, colour and area.

C3: We investigate the impact of two coordinate systems (Cartesian and Polar) on the effectiveness of time series visualisations within each of the aforementioned visual encoding techniques.

Through these contributions, we hope to contribute to some very practical problems that rely on effective data visualisation of large time series data sets, such as analysing network security data to counter cyber-threats. We also believe that our work will motivate the visual analytics research community to focus equally on interaction and visual representation of data. In the next section we review related work. Then, we present our quantitative graphical perception study and explain the choice of time series visualisations, interaction techniques, study tasks and our time series data generation method. Finally, we present the results, discuss their implications and highlight further research opportunities.



(a) Line chart (b) Radar chart (c) Rectangular heatmap (d) Circular heatmap (e) Icicle plot (f) Sunburst visualisation

Figure 1. Selected time series visualisations with highlighting and tooltip interaction techniques. Visual encodings: positional (a) & (b); colour (c) & (d); area (e) & (f). Coordinate systems: Cartesian (a), (c) & (e); Polar (b), (d) & (f). Tasks: maxima (a), (c) & (e); trend detection (b), (d) & (f).

RELATED WORK

We review the related work for three areas that are key to our research, and the impact of each on the graphical perception of time series visualisations, namely the effects of interaction techniques, and then the impact of different visual representations, first with visual encodings and then with coordinate systems.

Interaction Techniques

The visual analytics (VA) research community has focused more on the visual representation of data rather than on different aspects of interaction, unlike the closely related research field of Human-Computer Interaction (HCI) that places equal emphasis on both the design and user experience. The small set of VA papers that discuss interaction largely concentrate on the classification and characterisation of interaction techniques [12,18,34,35,39], capturing interaction processes [31], designing and operationalising interactions [20,25,38] and identifying challenges and opportunities for further research [13,30]. However, limited attention has been paid to measuring the exact effects of commonly used interaction techniques on visual representations, both in general and specific to quantitative graphical perception studies of time series visualisations. There are two exceptions in this regard. Lam et al. [23] examined the effects of interactive low- and high-resolution visual representations of time series data on search and comparison tasks. Also, Perin et al. [29] compared interactive horizon graphs with static horizon graphs and small multiples of line charts for multiple time series datasets. The focus of these two quantitative studies was on comparing the effectiveness of alternative visual representations of time series data, rather than on assessing the effects of individual interaction techniques. A more quantitative approach to assess the impact of alternative interaction techniques for different time series visualisations remains an under-researched area.

Therefore, one of the main contributions (C1) of this work is to systematically examine the effects of commonly used interaction techniques (i.e., highlighting and tooltips) on the effectiveness of different time series visualisations.

Visual Encodings

Prior work on graphical perception found that the choice of visual encoding matters a great deal to users. In their seminal work in 1984, Cleveland and McGill [8] presented various charts to participants and asked them to compare the values of two marked objects by estimating the percentage of a smaller value versus a larger. This accuracy measure was then used to rank the following visual encodings (ordered from most to least accurate) for quantitative data: (1) position along a common scale, (2) position along non-aligned scales, (3) length, direction and angle, (4) area, (5) volume and curvature, and (6) shading and colour saturation. More recently, Heer and Bostock [17] in 2010, successfully replicated these results by conducting a crowdsourced experiment using Amazon's Mechanical Turk. Based on existing psychophysical results and analysis of different perceptual tasks, Mackinlay [27] extended the ranking of the effectiveness of visual encodings to ordinal and nominal data.

Time series data is among the most common type of data explored in quantitative graphical perception studies. The focus of such studies has been on evaluating the effectiveness of time series visualisations within and across the categories of positional (e.g., Figure 1(a&b)) and colour visual encodings (e.g., Figure 1(c&d)). Within the category of positional visual encoding, Heer et al. [16] compared the effectiveness of line charts with horizon graphs. They investigated the effects of chart size and layering on user performance for a single task of value comparison. Similarly, Javed et al. [21] evaluated the effectiveness of four different line chart techniques, involving multiple time series datasets. This study involved three tasks of comparison, slope and discrimination. On the other hand, Correl et al. [10] investigated how time series visualisations, using both positional and colour visual encodings, can be specifically designed to support aggregate comparisons. Albers et al. [4] also compared the effectiveness of eight different time series visualisations, using both positional and colour visual encodings, for a number of point and summary comparison tasks. The results from these studies [4,10,16,21] highlight the strengths and weaknesses of both positional and colour visual encodings for representing time series data and suggest that no individual visual encoding dominates in every task and data density. In particular, these findings indicate limited generalisability of the results of generic, non-time series studies conducted by Cleveland and McGill [8] and Heer and Bostock [17] to time series visualisations.

This encouraged us to compare the effectiveness of time series visualisations using area visual encoding with visualisations using positional and colour visual encodings. To the best of our knowledge, ours is the first study that compares the effectiveness of time series visualisations using three different visual encodings of position, colour and area. This relates to the second contribution (C2) of our work.

Coordinate Systems

The strengths and weaknesses of visualisations based on Cartesian and Polar coordinate systems have been discussed by developers but empirical evidence is limited. Some studies exist that compared the Polar visualisations with competing Cartesian counterparts [7,36]. However, the non-time series nature and limited control of confounding factors within these studies constrains the generalisability of their findings.

Cleveland and McGill [9] compared the effectiveness of bar charts with their polar counterparts, pie charts. They found bar charts to be more effective than pie charts, because comparing lengths is more accurate than comparing angles. However, Schonlau and Peters [33] found the effectiveness of pie charts closely comparable to that of bar charts. More recently, Diehl et al. [11] compared the effectiveness of a matrix-based Cartesian visualisation with its Polar counterpart. They employed the generic study task of memorising positions of visual elements. The findings from their user study suggested using a Cartesian coordinate system unless there are clear reasons in favour of a Polar coordinate system. While conducted with better control over confounding factors than [7] and [36], the findings of these generic, non-time series studies are not easily generalisable to Cartesian and Polar time series visualisations, mainly due to the large number of tasks with varying complexity, different data densities and the variety of available visual representations of time series data.

With regards to the impact of coordinate systems on time series visualisations, the study conducted by Fuchs et al. [14] is closest to our work. They compared the effectiveness of two Cartesian and two Polar time series visualisations in a small multiple setting (called "glyphs"). One set of glyphs (one Cartesian and one Polar) used positional visual encoding, while the other set used the colour visual encoding. They found that line glyphs generally performed better for peak and trend detection tasks, while the Polar glyphs were better suited for reading values at specific temporal locations. In contrast to Fuchs et al. [14], we investigated the impact of coordinate systems on normal-sized, temporally labelled time series visualisations, not only based on positional and colour visual encodings but also using an area visual encoding (i.e., Figure 1(e & f)). This relates to the third contribution of our work (C3).

USER STUDY

To better understand the impact on the user experience, we conducted a lab-based user study to evaluate the effectiveness of different time series visualisations that use varied interaction techniques, visual encodings and coordinate systems for four tasks.

Time Series Visualisations

Time series data is a set of quantitative values changing over time, and it is predominant in a variety of domains ranging from banking and finance to computing to climate measurements. It is also extremely conducive to visualisation. Aigner et al. [3] provides an extensive review of time series visualisations.

We selected two time series visualisations (one Cartesian and one Polar) from each of three visual encoding categories of position, colour and area. Firstly, a Cartesian line chart (Figure 1(a)) was used as a well-known baseline for the positional visual encoding; the radar chart (Figure 1(b)) is the Polar counterpart. Secondly, following similar studies [4,10,14], rectangular (Figure 1(c)) and circular heatmaps (Figure 1(d)) were chosen to serve as the Cartesian and Polar counterparts from the category of colour visual encoding. We used ColorBrewer 2.0 [6] to choose a colour-blind safe sequential colour scheme for heatmaps. Lastly, considering the hierarchal nature of time series data, the Cartesian icicle plot (Figure 1(e)) and Polar sunburst visualisation (Figure 1(f)) were selected to represent the area visual encoding. While we could have used Cartesian and Polar treemaps [22] for our area encoding, icicle plots and sunbursts offer a more logical and space-efficient way of representing time series data. Further, sunburst provides an accurate Polar representation of the Cartesian icicle plot, unlike treemaps, thereby providing some consistency with the pairings for positional and colour encodings.

Interaction Techniques

In this study, we investigated the effects of two commonly used interaction techniques: highlighting and tooltips. In its simplest form, highlighting acts as the viewing control to attract a user's attention to a subset of data within a visualisation [26]. On the other hand, a tooltip is a text message that appears when a cursor is positioned over an element in a graphical user interface. In our case, the purpose of both of these techniques was to facilitate the identification of the temporal location and quantitative value of a particular data point as accurately as possible. The primary difference was that highlighting tried to achieve this visually, whereas the tooltip explicitly provided the required information in textual form.

The implementation of both of these interaction techniques was achieved through mouseover. The implementation of highlighting varied for each visualisation pair, mainly due to the inherent structure of the visual encodings. However, the design principle remained the same in each case, i.e., to have maximum visual impact, while avoiding visual occlusion. In the case of line chart, the highlighting of a particular data point increased its radius and introduced two dashed lines, each stretching from the selected data point to the X and Y axis respectively (Figure 1(a)). The same highlighting technique was used for radar chart, except following the Polar coordinate system (Figure 1(b)). In the case of the colour-encoded heatmaps, highlighting a data point increased its size and added an outline around the corresponding legend (Figure 1(c&d)). Lastly, in the case of the area-encoded icicle plot and sunburst visualisations, the highlighting of a data point changed its colour (Figure 1(e&f)). Note that the visualisations that were using the colour visual encoding limited our choice of highlighting to alteration in the size of data points, as opposed to changing their colours. The opposite was true for time series visualisations that use the area visual encoding.

The implementation of tooltips was the same for all visualisations, regardless of the type of visual encoding and coordinate system being used. We derived the following four interaction scenarios from these two commonly used interaction techniques: no interaction (where neither highlighting nor tooltips were used), only highlighting, only tooltips, and both highlighting and tooltips (highlighting-tooltips).

Tasks

Following Fuchs et al. [14] our selection of tasks was based on two criteria: (1) their ecological validity and (2) their heterogeneity in terms of the elementary perceptual tasks. To meet the first criterion, our tasks were informed by investigations into the work practices of network security professionals, particularly the activity of data correlation to find new and unexplained patterns for further analysis [1]. However, we believe that the study tasks are also widely applicable to other domains, e.g., climate measurement and banking and finance. Heterogeneity was ensured by selecting tasks that were composed of distinct elementary perceptual tasks. The time series data was presented to participants in the form of sales data of a fictitious company, spanning over 16 weeks (with 7x16=112 data points) to provide an easy-to-understand context. Based on these criteria, we selected the following four tasks for this study:

1. Maxima: To identify the highest absolute value in a dataset (e.g., the highlighted day in Figure 1(a,c&e)). Specifically, we asked participants to “discover the day when the sales were at their highest”.
2. Minima: To identify the lowest absolute value in a dataset. Specifically, we asked participants to “discover the day when the sales were at their lowest”. This was functionally identical to the maxima task, except that “highest” was changed to “lowest”. Despite their similarities, prior work [4,32] suggests that there are differences in the performance of these two tasks and that different visualisations may be more suitable.
3. Comparison: To compare two sets of data points to find out which set has the highest aggregated value. Specifically, we asked participants to “discover which of the following weeks has the highest aggregated sales: <week X> or <week Y>”.
4. Trend detection: To identify a subset of data (i.e., a week) with the lowest value increase (or upward trend) within the dataset (e.g., 3rd week of February in Figure 1(b,d&f)). Specifically, we asked participants to “discover the week with the smallest difference in sales between the first and last day”.

In terms of ecological validity, maxima and minima tasks aim to identify the highest and lowest network traffic respectively, when detecting bandwidth depletion denial of service (DoS) attacks using data correlation. On the other hand, the comparison task aims to compare two subsets of data to identify unexplained patterns requiring further analysis. Lastly, the task of trend detection aims to identify increasing or decreasing trends in network traffic when performing the activity of data correlation.

Experiment Design

We followed a within-subject design with independent variables of 4 interaction scenarios (no interaction, highlighting, tooltips and highlighting-tooltips), 3 visual encodings (position, colour and area), 2 coordinate systems (Cartesian and Polar) and 4 tasks. This within-subject factorial design yielded a total of 96 (4x3x2x4) experimental conditions for each participant. The dependent variables used to measure the impact on users were the completion time of each experimental condition, answer correctness, confidence regarding the given answer, and the ease of use of the visualisation. The experiment was approved by our University's ethics board.

Following Javed et al. [21], the order of the tasks was not counterbalanced, but rather presented in the order of simple to complex, to better prepare the participants for the more difficult tasks at the end. We believed this would not have a significant effect on the validity of the study results since our purpose was not to compare the effectiveness of one task against another, but rather to compare the effectiveness of alternative time series visualisations within the selected tasks. We counterbalanced the order of time series visualisations and interaction scenarios between subjects using a balanced Latin square to minimise the systematic effects of practice.

Time Series Data

Following previous graphical perception studies [4,10,14,16,21], we used synthetic time series data in order to have greater control on the data values and their corresponding visual representation. We generated 96 distinct time series datasets, one for each experiment condition. These were assigned randomly to minimise the learning effect of a correct answer repeatedly appearing at a particular temporal location or having a particular value. Each dataset had 112 data points (1 per day) and spanned over a period of 16 weeks. Based on our pilot testing and the data density of datasets used by similar studies [4,10,14], we estimated that this scale of data would be substantial enough to achieve the appropriate level of difficulty for the different tasks of our study.

We adopted the methods of Fuchs et al. [14] and Correl et al. [10] to generate synthetic data for this experiment. All data values were drawn from a normalised range 0 to 100 to make the data generalisable across different display conditions. Below we explain the data generation method for each of the four study tasks:

Maxima: We selected a data point (p) at random in terms of its temporal placement and value. The value of p was drawn from the range 65 to 100. Following Albers et al. [4] and Fuchs et al. [14], we set the offset d=20, which is the difference between the maxima (p) and randomly generated noise. The remaining data points within the time series dataset were randomly drawn from the range 1 to (p - d).

Minima: Similar to maxima except that p was drawn from the range 1 to 35 and the remaining data points within the time series dataset were randomly drawn from the range (p + d) to 100, with offset d=20.

Comparison: We selected two weeks (w1 and w2) at random in terms of their temporal placement. Data points within w1 were drawn at random from the range 20 to 80. The sum (s1) of values within week 1 was then calculated. The values of data points within w2 were randomly drawn with the constraint that the sum of values (s2) was either 20% (offset d) lower than s1 to simulate the aggregated value of w1 > w2, or s2 was 20% higher than s1 to simulate the aggregated value of w2 > w1. The remaining data points within the time series dataset were randomly drawn from the range 1 to 100.

Trend Detection: First, we created 15 distractor weeks. In each week, the value of the 1st data point was drawn from the range 1 to 20 and the value of the 7th data point was drawn from the range 70 to 100. From these distractor weeks, we identified the week with the smallest difference between the first and last day (sDif). Then, we created the winner week and added it to the dataset at a random temporal place. In this week, the value of the 1st data point (p1) was drawn from the range 1 to 50 and the value of the 7th data point was calculated by using the formula 'p1 + sDif - offset (d)', where d=20. The middle 5 data points within each week were drawn along the trend line with a random variation of 1 to 5.

Participants

We recruited 24 participants (14 male, 10 female). All participants had normal or corrected-to-normal vision and did not report colour blindness. The age of participants ranged from 18-44 years. 11 participants (46%) were students, 1 participant (4%) was part-time employed and 12 participants (50%) were in full-time employment. The education level of participants was diverse with 5 participants (21%) enrolled as undergraduate or graduate students, 8 (33%) had a Bachelor's degree, 8 (33%) received a Master's degree and 3 participants (13%) had a Ph.D. degree. Following Javed et al. [21], we also screened participants to have reasonable computer experience, which was defined as using a computer more than 20 hours per week. All the participants were given a £10 voucher upon completion of the study.

Apparatus

The experiment was conducted on a 15.6 inch Samsung laptop with a 2.5GHz Intel i5 processor, 6 GB of RAM and screen resolution set to 1366x768. A standard, wired mouse was also connected with the laptop to facilitate the interaction with the study software.

The study was implemented as a web application using HTML5, CSS and JavaScript. Time series visualisations were developed using D3.js [5]. Based on our pilot studies, for the Cartesian time series visualisations, a width of 700 pixels and a height of 300 pixels was found to be of appropriate size on the 15.6 inch laptop with the screen resolution set to 1366x768. The size of Polar visualisations was set to 458x458 pixels to assign an approximately equal number of pixels for both the Cartesian and Polar visualisations.

Study Procedure

The experiment took place in a quiet room. The participants sat in front of a table at a distance of approximately 50cm from the laptop and interacted with the study software using only a mouse.

The experiment began by briefly explaining the purpose and procedure of the study. The participant was asked to sign a consent form and to complete a short demographic survey. The experimenter then explained (using a script) the time series visualisations, interaction techniques, data, and the first task. After completing the training for the first task, the participant completed 24 (4 interaction scenarios x 3 visual encodings x 2 coordinate systems) experimental conditions for the task in which he/she selected the correct answer using drop-down menus of week and month. After each condition, the participant answered two 5-point Likert scale questions, ranging from strongly agree (5) to strongly disagree (1): (i) I feel confident about the given answer and (ii) I think this visualisation is easy to use for this task. The participant completed the training and all experimental conditions for the task of maxima, then minima, comparison and trend detection respectively. Thus, training was completed at the start of each of the four tasks. Participants were advised that they should only move on to the actual study task once they were comfortable with the study software and fully understood the task.

Participants were advised to perform each experimental condition as quickly as possible to keep it in line with our 1st criterion of task selection (i.e., ecological validity); however, no strict time limits were imposed. The study duration ranged from 50 to 75 minutes.

RESULTS

In this study, completion time was the only continuous dependent variable and it was not normally distributed. Data normality was tested using the Shapiro-Wilk test, which identified 73.95% (71 out of 96) experimental conditions that were not normally distributed ($p<0.05$). The remaining 3 dependent variables in this study were either binary (accuracy) or ordinal (confidence and ease of use), which by definition could not be normally distributed. Therefore, we analysed the results using a non-parametric Friedman test. Pairwise comparisons were conducted with Wilcoxon signed-rank test. Significance levels were adjusted with the Holm-Bonferroni correction for multiple testing. We categorise the results into three sub-sections, corresponding to the three contributions of this paper. We only report on the statistically significant results at $p<0.05$.

Effects of Interaction Techniques

We compared the effectiveness (i.e., completion time (seconds), accuracy (percentage), confidence (Likert) and ease of use (Likert)) of four selected interaction scenarios (i.e., no interaction, highlighting, tooltips, and highlighting-tooltips) for three visual encodings (i.e., position, colour and area) and four tasks. Figure 2 provides an overview of statistically significant results.

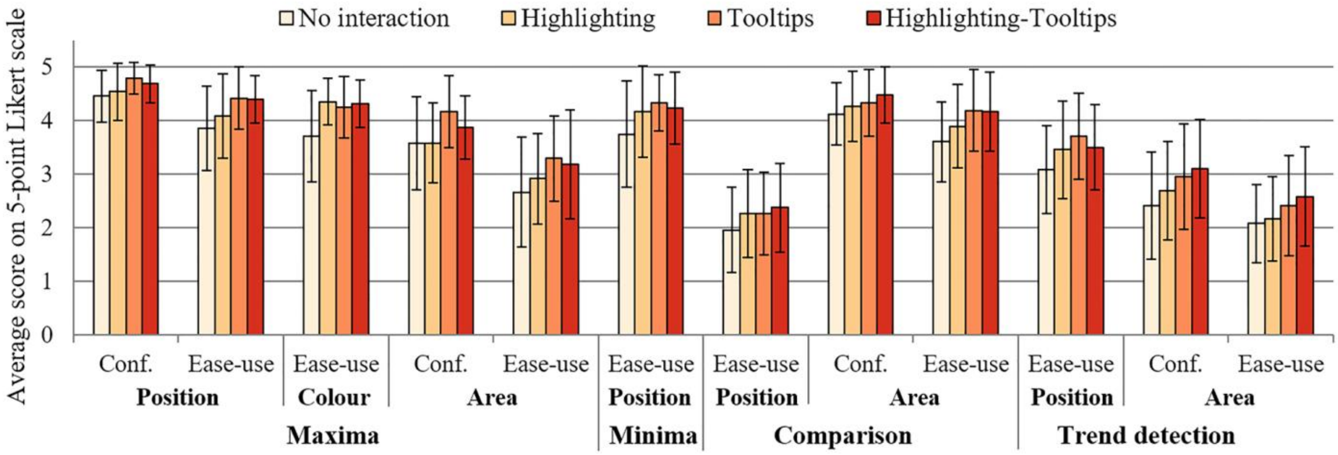


Figure 2. Overview of results for interaction scenarios. Conf. = confidence and ease-use = ease of use.

Maxima Results

For visualisations using positional visual encoding, there was an overall effect of interaction scenarios on confidence ($\chi^2(3)=11.53$, $p=0.009$) and ease of use ($\chi^2(3)=15.26$, $p=0.002$). Pairwise comparisons showed that participants had more confidence regarding their answers when using visualisations with tooltips than no interaction ($Z=-2.84$, $p=0.005$) and highlighting visualisations ($Z=-2.81$, $p=0.005$). Similarly, visualisations with tooltips were easier to use than no interaction ($Z=-3.03$, $p=0.002$) and highlighting visualisations ($Z=-2.91$, $p=0.004$). Visualisations with highlighting-tooltips were also found to be easier to use than no interaction visualisations ($Z=-2.89$, $p=0.004$).

For visualisations using colour visual encoding, there was an overall effect of interaction scenarios on ease of use ($\chi^2(3)=15.70$, $p=0.001$). Pairwise comparisons showed that visualisations with highlighting ($Z=-3.12$, $p=0.002$), tooltips ($Z=-2.80$, $p=0.005$) and highlighting-tooltips ($Z=-3.22$, $p=0.001$) were easier to use than no interaction visualisations.

For visualisations using area visual encoding, there was an overall effect of interaction scenarios on confidence ($\chi^2(3)=14.30$, $p=0.003$) and ease of use ($\chi^2(3)=17.31$, $p<0.001$). Pairwise comparisons showed that participants had more confidence regarding their answers when using visualisations with tooltips than no interaction visualisations ($Z=-2.74$, $p=0.006$). Moreover, visualisations with tooltips ($Z=-3.14$, $p=0.002$) and highlighting-tooltips ($Z=-3.06$, $p=0.002$) were easier to use than no interaction visualisations.

Minima Results

For visualisations using positional visual encoding, there was an overall effect of interaction scenarios on ease of use ($\chi^2(3)=10.35$, $p=0.016$). Pairwise comparisons showed that visualisations with highlighting ($Z=-2.99$, $p=0.003$) and tooltips ($Z=-2.76$, $p=0.006$) were easier to use than no interaction visualisations.

Comparison Results

For visualisations using positional visual encoding, there was an overall effect of interaction scenarios on ease of use ($\chi^2(3)=10.36$, $p=0.016$). Pairwise comparisons showed visualisations with highlighting ($Z=-2.68$, $p=0.007$) and highlighting-tooltips ($Z=-2.55$, $p=0.010$) were easier to use than no interaction visualisations.

For visualisations using area visual encoding, there was an overall effect of interaction scenarios on confidence ($\chi^2(3)=10.32$, $p=0.016$) and ease of use ($\chi^2(3)=17.01$, $p<0.001$). Pairwise comparisons showed that participants had more confidence regarding their answers when using visualisations with highlighting-tooltips than no interaction visualisations ($Z=-2.80$, $p=0.005$). Moreover, visualisations with tooltips were easier to use than no interaction ($Z=-3.12$, $p=0.002$) and highlighting visualisations ($Z=-2.66$, $p=0.008$). Visualisations with highlighting-tooltips were also found to be easier to use than no interaction visualisations ($Z=-3.12$, $p=0.002$).

Trend Detection Results

For visualisations using positional visual encoding, there was an overall effect of interaction scenarios on ease of use ($\chi^2(3)=24.05$, $p<0.001$). Pairwise comparisons showed that visualisations with tooltips ($Z=-3.40$, $p=0.001$) and highlighting-tooltips ($Z=-2.69$, $p=0.007$) were easier to use than no interaction visualisations.

For visualisations using area visual encoding, there was an overall effect of interaction scenarios on confidence ($\chi^2(3)=18.02$, $p<0.001$) and ease of use ($\chi^2(3)=13.04$, $p<0.005$). Pairwise comparisons showed that participants had more confidence regarding their answers when using visualisations with tooltips ($Z=-3.10$, $p=0.002$) and highlighting-tooltips ($Z=-3.17$, $p=0.002$) than no interaction visualisations. In addition, visualisations with highlighting-tooltips were easier to use than no interaction ($Z=-2.89$, $p=0.004$) and highlighting visualisations ($Z=-2.73$, $p=0.006$).

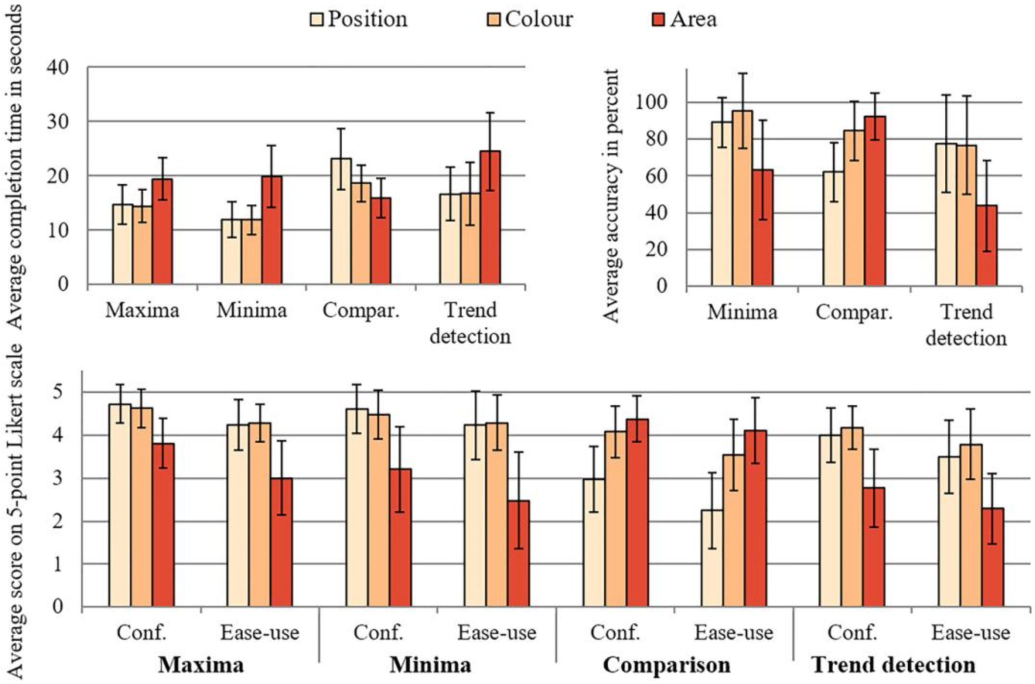


Figure 3. Overview of results for visual encodings. Conf. = confidence, ease-use = ease of use, compar. = comparison.

Impact of Visual Encodings

We compared the effectiveness (i.e., completion time, accuracy, confidence and ease of use) of visualisations using three different visual encoding techniques (i.e., position, colour and area) for the tasks of maxima, minima, comparison and trend detection. Figure 3 provides an overview of statistically significant results.

Minima Results

There was an overall effect of visual encodings on accuracy ($\chi^2(2)=33.74$, $p<0.001$). Pairwise comparisons showed that the participants gave more correct answers when using visualisations with colour visual encoding, compared to position (Z=-2.36, p=0.018) and area (Z=-4.18, p<0.001). Participants also gave more correct answers when using visualisations with positional visual encoding than area (Z=-3.89, p<0.001).

Comparison Results

There was an overall effect of visual encodings on completion time ($\chi^2(2)=37.00$, $p<0.001$), accuracy ($\chi^2(2)=33.09$, $p<0.001$), confidence ($\chi^2(2)=31.08$, $p<0.001$) and ease of use ($\chi^2(2)=31.63$, $p<0.001$). Pairwise comparisons showed that the completion time was lower for visualisations with colour (Z=-4.11, p<0.001) and area visual encodings (Z=-4.29, p<0.001) than position. The completion time was also lower for visualisations with area visual encoding than colour (Z=-3.46, p=0.001). In addition, participants gave more correct answers when using visualisations with colour (Z=-3.67, p<0.001) and area visual encodings (Z=-3.96, p<0.001) than position. Pairwise comparisons also showed that participants had more confidence in their answers when using visualisations with colour (Z=-3.79, p<0.001) and area visual encodings (Z=-3.95, p<0.001) than position. Participants also had more confidence in their answers when using visualisations with area visual encoding than colour (Z=-2.12, p=0.034). Lastly, visualisations with colour (Z=-3.64, p<0.001) and area visual encodings (Z=-3.94, p<0.001) were easier to use than position. Visualisations using area visual encoding were also easier to use than colour visual encoding (Z=-2.72, p=0.007).

For all other cases, there were no statistically significant differences between positional and colour visual encodings, however both were more effective than area visual encoding. Also, no statistically significant differences were found for the dependent variable of accuracy within the task of maxima.

Impact of Coordinate Systems

We compared the effectiveness (i.e., completion time, accuracy, confidence and ease of use) of Cartesian and Polar coordinate systems for three visual encodings (i.e., position, colour and area) and four tasks (i.e., maxima, minima, comparison and trend detection). Figure 4 provides an overview of statistically significant results.

Maxima Results

For visualisations using positional visual encoding, the participants had more confidence regarding their answers for the Cartesian coordinate system than Polar (Z=-2.08, p=0.038). Similarly, the Cartesian visualisation was easier to use than Polar (Z=-2.75, p=0.006). For visualisations using area visual encoding, the completion time was lower for the Cartesian coordinate system than Polar (Z=-2.46, p=0.014). In addition, the Cartesian visualisation was easier to use than Polar (Z=-2.49, p=0.013).

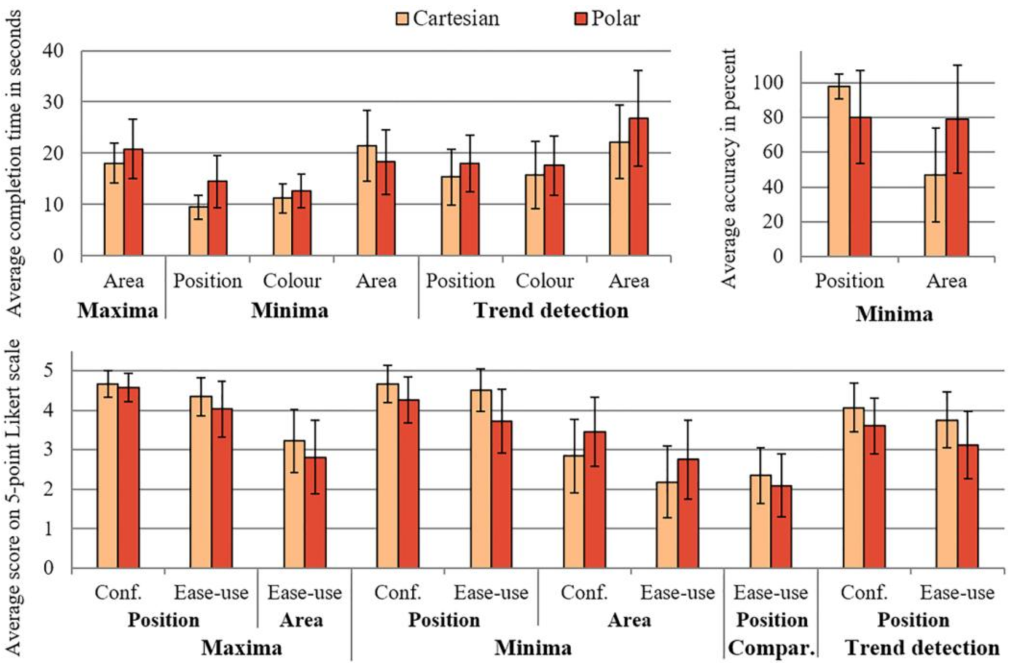


Figure 4. Overview of results for coordinate systems. Conf. = confidence, ease-use = ease of use and compar. = comparison.

Minima Results

For visualisations using positional visual encoding, the completion time was lower for the Cartesian coordinate system than Polar (Z=-4.00, p<0.001). The participants gave more correct answers when using the Cartesian coordinate system compared to Polar (Z=-2.80, p=0.005) In addition, the participants had more confidence in their answers for the Cartesian coordinate system than Polar (Z=-3.33, p=0.001) and found it easier to use, compared to Polar (Z=-3.92, p<0.001).

On the other hand, for visualisations using colour visual encoding, only completion time was lower for the Cartesian coordinate system than Polar (Z=-2.11, p=0.034).

In contrast to positional and colour visual encodings, the Polar visualisation performed better than its Cartesian counterpart for area visual encoding. The completion time was lower for the Polar coordinate system than Cartesian (Z=-2.14, p=0.032). The participants gave more correct answers when using the Polar coordinate system than Cartesian (Z=-4.00, p<0.001). Also, participants had more confidence regarding their answers for the Polar coordinate system than Cartesian (Z=-3.60, p<0.001) and found it easier to use, compared to Cartesian (Z=-3.11, p=0.002).

Comparison Results

For the task of comparison, the only statistically significant result was for the visualisations using positional visual encoding. The participants found the Cartesian visualisation to be easier to use than its Polar counterpart (Z=-2.59, p=0.010).

Trend Detection Results

For visualisations using positional visual encoding, the completion time was lower for the Cartesian coordinate system than Polar (Z=-2.46, p=0.014). In addition, the participants had more confidence regarding their answers for the Cartesian coordinate system than Polar (Z=-2.98, p=0.003). They also found it easier to use compared to Polar (Z=-3.47, p=0.001).

On the other hand, for visualisations using colour visual encoding, only completion time was lower for the Cartesian coordinate system than Polar (Z=-2.07, p=0.038). Similarly, for visualisations using area visual encoding, only completion time was lower for the Cartesian coordinate system than Polar (Z=-2.49, p=0.013).

DISCUSSION

We divide the findings into three sub-sections, corresponding to the three contributions (C1-C3) of this paper.

Effects of Interaction Techniques

We found that introducing interactivity for time series visualisations could considerably enhance the user experience (i.e., confidence and ease of use), without the loss of any efficiency or accuracy. Further, there are clear differences in the effectiveness of selected interaction scenarios (i.e., highlighting, tooltips, and highlighting-tooltips). Our findings indicate that visualisations with tooltips were not only considerably more effective than the visualisations with highlighting, but they were also slightly better than visualisations with both the highlighting and the tooltips (highlighting-tooltips). In terms of design implications for time series visualisations, this finding suggests that users prefer clear textual instructions (i.e., tooltips) than highlighting, when it comes to facilitating the identification of temporal location and quantitative value of a particular data point.

The trend of interactive time series visualisations resulting in better user experience was observed considerably less for the task of minima and visualisations that use colour visual encoding. This confirms the findings of prior work [4,32] that indicate that despite the similarity between minima and maxima, there are differences in the performance of these two tasks and that different visualisations may be suitable. This finding has two design implications for time series visualisations. Firstly, introducing interactivity may add little value for time series visualisations using colour visual encoding. Secondly, the effect of interactivity on the performance of minima task is very negligible.

Impact of Visual Encodings

Our findings suggest that for maxima, minima and trend detection tasks, time series visualisations that use positional and colour visual encodings were more effective than area visual encodings. However, for the task of comparison, visualisations that use area visual encodings were more effective than their positional and colour counterparts.

This confirms that the findings of generic, non-time series graphical perception studies [8,17] are not fully generalisable to time series visualisations. In other words, while visualisations using positional visual encodings are quite a strong option to visualise time series data, there are many scenarios where visualisations that use colour visual encoding are comparable or more effective. In addition, we have a new finding that there are certain scenarios (e.g., comparison task) where visualisations that use area visual encoding are better suited, compared to visualisations from both the visual encodings of position and colour.

In terms of design implications, our findings suggest that the choice of a time series visualisation should be based on the type of tasks and the metric used to measure the effectiveness, since there is no *"one-size-fits-all"* solution.

Impact of Coordinate Systems

We found that visualisations that are based on a Cartesian coordinate system are generally comparable or more effective than Polar, except for visualisations that use area visual encoding for the task of minima.

This confirms the findings of two generic studies conducted by Diehl et al. [11], which suggest using a Cartesian coordinate system unless there are clear reasons in favour of Polar. In our case, the only scenario where a visualisation based on the Polar coordinate system (i.e., sunburst visualisation) were more effective than its Cartesian counterpart (i.e., icicle plot) was from the category of area visual encoding, for the task of minima. There could be several possible reasons for this finding. We believe that this was due to the increased space to encode data values of days, provided by the outermost ring of the sunburst visualisation compared to the lowermost segment of the icicle plot.

Also note that, the effect of coordinate systems is negligible for time series visualisations that use colour visual encoding. In terms of design implications, this finding suggests that time series visualisations using colour visual encoding may be an appropriate choice when the coordinate system is not known beforehand.

LIMITATIONS AND FUTURE WORK

While comprehensive enough to explore three different aspects (C1-C3) of time series visualisations, this approach was not without limitations. To begin with, we investigated the effect of two commonly used interaction techniques (i.e., highlighting and tooltips) on the effectiveness of different time series visualisations. In the future, we would like to extend this study to other commonly used interaction techniques, e.g., selection, filtering and zoom in/out.

In addition, our experiment considers a small but diverse set of time series visualisations and tasks. We believe that our findings would be generalisable to a wider range of situations, but we have not confirmed this empirically. In our future endeavours, we plan to extend these findings by including additional time series visualisations (e.g., scatter plots and horizon graphs) and tasks (e.g., range and outliers). The same is true for the selected size and aspect ratio of time series visualisations, density of time series datasets and the offset between the correct answers and randomly generated noise.

CONCLUSION

In this paper, we presented the results of a comprehensive graphical perception study. The purpose of this study was to examine the effects of two commonly used interaction techniques on time series visualisations (C1), to compare the effectiveness of time series visualisations using the positional, colour and area visual encodings (C2) and to investigate the impact of Cartesian and Polar coordinate systems on the effectiveness of time series visualisations (C3).

Our results show that the time series visualisations that use positional and colour visual encodings were more effective than area visual encodings for maxima, minima and trend detection tasks. However, for the task of comparison, visualisations that use area visual encodings were more effective than their positional and colour counterparts. We also found that the time series visualisations that are based on the Cartesian coordinate system are generally comparable or more effective than Polar. Most importantly, we identified that introducing interactivity within time series visualisations considerably enhances the user experience, without any loss of efficiency or accuracy. We believe that our findings could vastly assist visual analytics tool developers when choosing time series visualisations for different tasks in a variety of domains. We also believe that our work will motivate the visual analytics research community to focus equally on interaction and visual representation of data.

REFERENCES

1. Muhammad Adnan, Mike Just, Lynne Baillie, and Hilmi G. Kayacik. 2015. Investigating the work practices of network security professionals. *Information & Computer Security* 23, 3: 347-367.
2. Wolfgang Aigner. 2011. Understanding the role and value of interaction: First steps. In *Proceedings of the International Workshop on Visual Analytics* (EuroVA '11), 17-20. http://dx.doi.org/10.2312/PE/EuroVAST/EuroVA11/017-020
3. Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. 2011. *Visualization of time-oriented data*. Springer Science & Business Media.
4. Danielle Albers, Michael Correll, and Michael Gleicher. 2014. Task-driven evaluation of aggregation in time series visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 551-560. http://dx.doi.org/10.1145/2556288.2557200
5. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D³ data-driven documents. *IEEE Trans Vis Comput Graph* 17, 12: 2301-2309.
6. Cynthia A. Brewer, Mark Harrower, and Pennsylvania State University. 2015. ColorBrewer 2.0 – color advice for cartography. Retrieved July 10, 2015 from http://colorbrewer2.org/
7. Michael Burch, Felix Bott, Fabian Beck, and Stephan Diehl. 2008. Cartesian vs. radial – a comparative evaluation of two visualization tools. In *Proceedings of 4th International Symposium on Visual Computing* (ISVC '08), 151-160. http://dx.doi.org/10.1007/978-3-540-89639-5_15
8. William S. Cleveland and Robert McGill. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387: 531-554.
9. William S. Cleveland and Robert McGill. 1986. An experiment in graphical perception. *International Journal of Man-Machine Studies* 25, 5: 491-500.
10. Michael Correll, Danielle Albers, Steven Franconeri, and Michael Gleicher. 2012. Comparing averages in time series data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12), 1095-1104. http://dx.doi.org/10.1145/2207676.2208556
11. Stephan Diehl, Fabian Beck, and Michael Burch. 2010. Uncovering strengths and weaknesses of radial visualizations---an empirical approach. *IEEE Trans Vis Comput Graph* 16, 6: 935-942.
12. Alan Dix and Geoffrey Ellis. 1998. Starting simple: adding value to static visualisation through simple interaction. In *Proceedings of the working conference on Advanced visual interfaces* (AVI '98), 124-134. http://dx.doi.org/10.1145/948496.948514
13. Niklas Elmqvist, Andrew V. Moere, Hans-Christian Jetter, Daniel Cernea, Harald Reiterer, and T. J. Jankun-Kelly. 2011. Fluid interaction for information visualization. *Information Visualization* 10, 4: 327-340.
14. Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. 2013. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 3237-3246. http://dx.doi.org/10.1145/2470654.2466443
15. John R. Harger, and Patricia J. Crossno. 2012. Comparison of open-source visual analytics toolkits. In *Proceedings of the SPIE 8294, Visualization and Data Analysis*, 82940E. http://dx.doi.org/10.1117/12.911901
16. Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. 2009. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 1303-1312. http://dx.doi.org/10.1145/1518701.1518897
17. Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '10), 203-212. http://dx.doi.org/10.1145/1753326.1753357
18. Jeffrey Heer and Ben Shneiderman. 2012. Interactive dynamics for visual analysis. *Communications of the ACM* 55, 4: 45-54.
19. Intel IT Center. 2013. Big Data Visualization: Turning Big Data Into Big Insights. Retrieved July 5, 2015 from http://www.intel.co.uk/content/dam/www/public/us/en/documents/white-papers/big-data-visualization-turning-big-data-into-big-insights.pdf
20. Yvonne Jansen and Pierre Dragicevic. 2013. An interaction model for visualizations beyond the desktop. *IEEE Trans Vis Comput Graph* 19, 12: 2396-2405.
21. Waqas Javed, Bryan McDonnel, and Niklas Elmqvist. 2010. Graphical perception of multiple time series. *IEEE Trans Vis Comput Graph* 16, 6: 927-934.
22. Brian Johnson and Ben Shneiderman. 1991. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of the IEEE Conference on Visualization* (Visualization '91), 284-291. http://dx.doi.org/10.1109/VISUAL.1991.175815

23. Heidi Lam, Tamara Munzner, and Robert Kincaid. 2007. Overview use in multiple visual information resolution interfaces. *IEEE Trans Vis Comput Graph* 13, 6: 1278-1285.

24. Heidi Lam. 2008. A framework of interaction costs in information visualization. *IEEE Trans Vis Comput Graph* 14, 6: 1149-1156.

25. Hai-Ning Liang, Paul Parsons, Hsien-Chi Wu, and Kamran Sedig. 2010. An exploratory study of interactivity in visualization tools: 'Flow' of interaction. *Journal of Interactive Learning Research* 21, 1: 5-45.

26. Jie Liang and Mao L. Huang. 2010. Highlighting in information visualization: A survey. In *Proceedings of the 14th International Conference on Information Visualisation* (IV '10), 79-85. http://dx.doi.org/10.1109/IV.2010.21

27. Jock D. Mackinlay. 1986. Automating the design of graphical presentations of relational information. *Acm Transactions On Graphics* 5, 2: 110-141.

28. Andrew McAfee and Erik Brynjolfsson. 2012. Big data: The management revolution. *Harvard Bus Rev* 90, 10: 61-67.

29. Charles Perin, Frédéric Vernier, and Jean-Daniel Fekete. 2013. Interactive horizon graphs: improving the compact visualization of multiple time series. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 3217-3226. http://dx.doi.org/10.1145/2470654.2466441

30. William A. Pike, John Stasko, Remco Chang, and Theresa A. O'connell. 2009. The science of interaction. *Information Visualization* 8, 4: 263-274.

31. Margit Pohl, Sylvia Wiltner, Silvia Miksch, Wolfgang Aigner, and Alexander Rind. 2012. Analysing interactivity in information visualisation. *KI-Künstliche Intelligenz* 26, 2: 151-159.

32. Jibonananda Sanyal, Song Zhang, Gargi Bhattacharya, Phil Amburn, and Robert J. Moorhead. 2009. A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE Trans Vis Comput Graph* 15, 6: 1209-1218.

33. Matthias Schonlau and Ellen Peters. 2008. Graph comprehension: An experiment in displaying data as bar charts, pie charts and tables with and without the gratuitous 3rd dimension. *Social Science Research Network Working Paper Series*.

34. Kamran Sedig, Paul Parsons, and Alex Babanski. 2012. Towards a Characterization of Interactivity in Visual Analytics. *Journal of Multimedia Processing and Technologies* 3, 1: 12-28.

35. Kamran Sedig, Paul Parsons, Mark Dittmer, and Robert Haworth. 2014. Human-centered interactivity of visualization tools: Micro-and macro-level considerations. In *Handbook of Human Centric Visualization*, Weidong Huang (eds). Springer, New York, 717-743.

36. John Stasko, Richard Catrambone, Mark Guzdial, and Kevin McDonald. 2000. An evaluation of space-filling information visualizations for depicting hierarchical structures. *International Journal of Human-Computer Studies* 53, 5: 663-694.

37. James J. Thomas and Kristin A. Cook (eds.). 2005. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press.

38. Christian Tominski, Camilla Forsell, and Jimmy Johansson. 2012. Interaction support for visual comparison inspired by natural behavior. *IEEE Trans Vis Comput Graph* 18, 12: 2719-2728.

39. Soo J. Yi, Youn ah Kang, John T. Stasko, and Julie A. Jacko. 2007. Toward a deeper understanding of the role of interaction in information visualization. *IEEE Trans Vis Comput Graph* 13, 6: 1224-1231.

40. Leishi, Zhang, Andreas Stoffel, Michael Behrisch, Sebastian Mittelstadt, Tobias Schreck, René Pompl, Simon Weber, Holger Last, and Daniel Keim. 2012. Visual analytics for the big data era - A comparative review of state-of-the-art commercial systems. In *Proceedings of IEEE Conference on Visual Analytics Science and Technology* (VAST '12), 173-182. http://dx.doi.org/10.1109/VAST.2012.6400554