ARTICLE IN PRESS

Visual Informatics ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons

Indratmo [a,*], Lee Howorko [a], Joyce Maria Boedianto [a], Ben Daniel [b]

[a] *Department of Computer Science, MacEwan University, Canada*
[b] *Higher Education Development Centre, University of Otago, New Zealand*

HIGHLIGHTS

- Variants of stacked bar charts have different levels of effectiveness for supporting overall-attribute comparisons.
- Appropriate visual representations of data can reduce the complexity of the task at hand.
- Spatial separation of information can help users manage the perceived difficulty of the task at hand.

ARTICLE INFO

Article history:
Received 21 May 2018
Received in revised form 24 August 2018
Accepted 16 September 2018
Available online xxxx

Keywords:
Stacked bar chart
Comparison task
User study
Multi-attribute data
Information visualization

ABSTRACT

Stacked bar charts are a visualization method for presenting multiple attributes of data, and many visualization tools support these charts. To assess the efficacy of stacked bar charts in supporting attribute-comparison tasks, we conducted a user study to compare three types of stacked bar charts: classical, inverting, and diverging. Each chart type was used to visualize six attributes of data where half of the attributes have the characteristics of 'lower better' whereas the other half 'higher better.' Thirty participants were asked to perform two types of comparison tasks: single-attribute and overall-attribute comparisons. We measured the completion time, error rate, and perceived difficulty of the comparison tasks. The results of the study suggest that, for overall-attribute comparisons, the inverting stacked bar chart was the most effective with regards to the completion time. The results also show that performing overall-attribute comparisons using the classical and diverging stacked bar charts required more time than performing single-attribute comparisons using these charts. Participants perceived the inverting and diverging stacked bar charts as easier-to-use than the classical stacked bar chart for overall-attribute comparisons. However, for single-attribute comparisons, all chart types delivered similar performance. We discuss how these findings can inform the better design of interactive stacked bar charts and visualization tools.

1. Introduction

Stacked bar charts are a visualization method that is particularly useful for presenting the sums of data attributes while allowing users to see how the values of these attributes contribute to the totals (Streit and Gehlenborg, 2014). Stacked bar charts have many applications, including for exploring rankings of items based on multiple attributes (Gratzl et al., 2013), visualizing survey data collected using Likert-type scales (Heiberger and Robbins, 2014), and presenting probabilities of discrete events (Spiegelhalter et al., 2011). Many software tools, including statistical tools (e.g., R and SPSS), visualization tools (e.g., Tableau and VisComposer Mei et al., 2018), and visualization libraries (e.g., D3 Bostock et al., 2011), support stacked bar charts.

Despite their common usage, stacked bar charts can be ineffective if not designed correctly (Knaflic, 2015; Munzner, 2015). First, due to their limited scalability, stacked bar charts are visually inefficient when used to present data with many attributes. Second, it is common to assign varying colors to different segments of bars in a stacked bar chart. Such colors should be chosen suitably to allow users to differentiate categories in the chart quickly (Healey, 1996). Third, without thoughtful design, stacked bar charts cannot support attribute-comparison tasks effectively, especially when the

* Corresponding author.
E-mail addresses: indratmo@macewan.ca (Indratmo), lhoworko@gmail.com (L. Howorko), boediantoj@macewan.ca (J.M. Boedianto), ben.daniel@otago.ac.nz (B. Daniel).

Peer review under responsibility of Zhejiang University and Zhejiang University Press.

https://doi.org/10.1016/j.visinf.2018.09.002

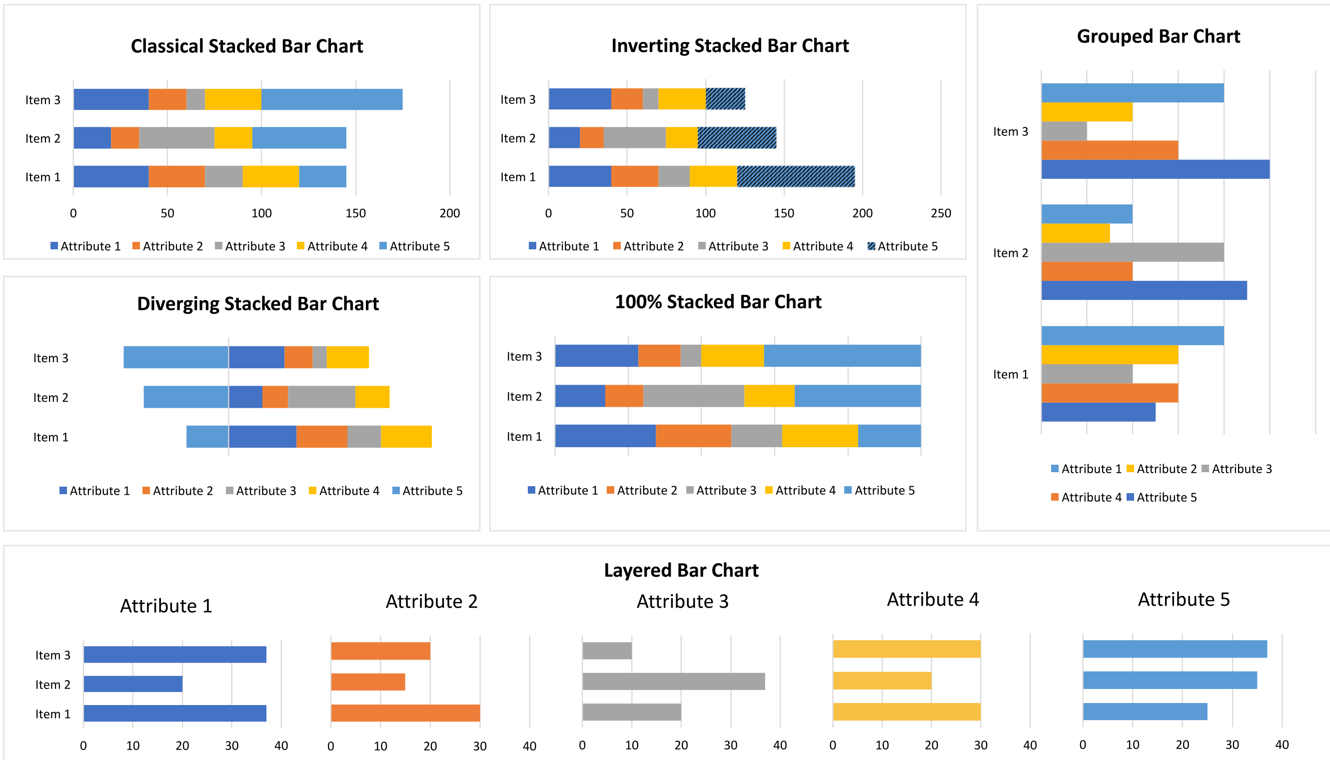

Fig. 1. Bar charts visualizing multiple attributes of data.

data attributes have diverging characteristics (e.g., 'lower better' vs. 'higher better').

Comparing data is one of the essential visualization tasks (Brehmer and Munzner, 2013). While analyzing multi-attribute data, users may need to focus both on a single attribute of data (e.g., finding a cheaper product) and on the overall attributes of the data (e.g., finding the best overall product based on the price, energy consumption, durability, and user rating). These attributes have the characteristics of lower better (i.e., price and energy consumption) and higher better (i.e., durability and user rating). Merely showing these attributes using a classical stacked bar chart may not help users perform product comparisons.

To understand the effectiveness of stacked bar charts in supporting both single-attribute and overall-attribute comparisons, we conducted a user study with 30 participants. We compared the performance of three types of stacked bar charts: classical, inverting, and diverging (see Fig. 1). The effectiveness of each chart type was assessed regarding the completion time, error rate, and perceived difficulty of the attribute-comparison tasks. The experimental results suggest that, for overall-attribute comparisons, the inverting stacked bar chart was the most effective with regards to the completion time. The results also show that performing overall-attribute comparisons using the classical and diverging charts required more time than performing single-attribute comparisons using these charts. Participants perceived the inverting and diverging charts as easier-to-use than the classical stacked bar chart for overall-attribute comparisons. However, for single-attribute comparisons, all chart types delivered similar performance.

Our study aims to contribute to the value of stacked bar charts in supporting single-attribute and overall-attribute comparisons by identifying the most effective type of stacked bar charts for facilitating attribute-comparison tasks. Identifying the relationship between chart type and task type has practical value, as the usefulness of charts varies across tasks (Saket et al., 2018). The outcome of this research can inform the better design of interactive stacked bar charts and guide visualization designers to provide the appropriate stacked bar chart to help users perform their tasks efficiently.

We organize the rest of this article as follows. Section 2 reviews relevant research on information visualization and bar charts. Section 3 illustrates the features of the classical, inverting, and diverging stacked bar charts used in the study. Section 4 describes the objective, design, procedure, and participants of the study. Section 5 reports on the experimental results and analysis. Section 6 discusses the main results of the study and their implications for the design of effective stacked bar charts and visualization tools. Section 7 summarizes the key ideas in this research and provides directions for future work.

2. Background and related work

2.1. The design space of bar charts

Bar charts typically use *length* to encode quantitative data and *color hue* to encode nominal data. Using these visual channels appropriately facilitates an accurate interpretation of data. Users can extract quantitative information from the length of bars and make length judgment accurately (Cleveland and McGill, 1984; Mackinlay, 1986), and utilize a small number of color hues to identify categories in charts (Ware, 2004). Further, length and color hue are visual features that the human visual system can recognize and process very quickly through parallel processing (Healey and Enns, 2012; Treisman and Gormican, 1988).

The use of bar charts and their variants has been discussed and documented by Brinton in 1939 (Brinton, 1939). Fig. 1 illustrates various kinds of bar charts that are commonly used to present multiple attributes of data. The orientation of the bars can be horizontal, as shown, or vertical. The style and name of these charts may vary over the years. For example, diverging bar charts

Please cite this article in press as: Indratmo, et al., The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics (2018), https://doi.org/10.1016/j.visinf.2018.09.002.

Table 1
The suitability of bar charts for single-attribute and overall-attribute comparisons.

Type	Single-attribute comparison	Overall-attribute comparison
Classical stacked bar chart	Yes	Yes
Inverting stacked bar chart	Yes	Yes
Diverging stacked bar chart	Yes	Yes
100% stacked bar chart	Yes	No
Grouped bar chart	Yes	No
Layered bar chart	Yes	No

were called bilateral bar charts by Brinton. However, the basic features and purposes of the charts remain the same. Below, we discuss each chart type (please refer to Fig. 1 when necessary) and summarize its suitability for single-attribute and overall-attribute comparisons in Table 1.

2.1.1. Classical stacked bar charts

In a classical stacked bar chart, the length of a bar indicates the sum of attribute values of an item while the segments of the bar show how each attribute value contributes to the total. Bar segments representing attributes typically have different visual features (e.g., different colors) to ease attribute comparisons and recognition. This visualization facilitates both single-attribute and overall-attribute comparisons. To make single-attribute comparisons, users can focus on the bar segment that represents the attribute of interest. When they need to perform overall-attribute comparisons, they can focus on the length of the bars.

Classical stacked bar charts can support overall-attribute comparisons well as long as all of the data attributes have the same characteristics. For example, when analyzing a dataset where larger values mean better, users can find the best overall item by locating the longest bar in a classical stacked bar chart. This approach, however, does not work when the data attributes have diverging characteristics. Consider the classical stacked bar chart in Fig. 1 and assume that attributes 1–4 are higher-better whereas attribute 5 is lower-better. From the chart, it is not apparent which one is the best overall item. Users must spend some cognitive effort before identifying that item 1 is the best item.

2.1.2. Inverting stacked bar charts

Inverting stacked bar charts have the potential to overcome the difficulty of overall-attribute comparisons involving diverging attribute characteristics. We first encounter the idea of inverting stacked bar charts in LineUp—a visualization application that uses interactive bar charts for exploring multi-attribute rankings (Gratzl et al., 2013). In LineUp, users may apply inversion to modify the mapping function so that the smaller the value of an attribute, the longer the bar that represents it. Visualizing the same data as shown in the classical stacked bar chart in Fig. 1, the inverting stacked bar chart inverts the visual representation of attribute 5 so that all bar segments now have a consistent meaning, i.e., longer means better. This inversion simplifies the task of finding the best overall item, as item 1 stands out to be the longest bar in the chart. Users of inverting stacked bar charts, however, may forget about the inverted bar segments and misinterpret the charts. Therefore, inverted bar segments should look different from regular bar segments. In our example, we fill the inverted bar segment (attribute 5) with a pattern.

2.1.3. Diverging stacked bar charts

In a diverging stacked bar chart, the bars use both areas of a common baseline (left and right or top and bottom) (Brinton, 1939). This chart type is suitable for presenting attributes with diverging characteristics (e.g., profit vs. loss) and supports both single-attribute and overall-attribute comparisons. Following our earlier example, a diverging stacked bar chart can be used to separate attribute 5 from the other attributes. More specifically, attribute 5 is shown in the left area of the baseline whereas the other attributes remain in the right area (see Fig. 1). This spatial separation may be able to help users deal with diverging attribute characteristics without using inversion, which may cause incorrect data interpretation.

2.1.4. 100% stacked bar charts

A 100% stacked bar chart is similar to the classical stacked bar chart except that the length of each bar is identical. In this chart, a bar represents 100%, rather than indicating an actual value, and the bar segments represent the percentages of the whole (Brinton, 1939). This chart can be considered equivalent to multiple pie charts.

A 100% stacked bar chart enables users to perform quick *relative* comparisons of individual attributes. Consider a dataset containing different kinds of languages spoken in a country and the numbers of people who speak these languages over the years. If we are interested in relative comparisons between these languages or finding trends in their use relative to the whole population, then using a 100% stacked bar chart to plot the dataset can help users see such patterns. For example, the graph may reveal that the percentages of English speakers in the country have increased rapidly in the last ten years. The 100% stacked bar charts, however, are not suitable for overall-attribute comparisons because every bar has the same length and consequently does not indicate which item is better.

2.1.5. Grouped bar charts

In stacked bar charts, typically only one attribute shares a common baseline. An exception is a diverging stacked bar chart where two attributes are sharing a baseline. All other attributes in stacked bar charts are nonaligned. Consequently, many attribute comparisons require nonaligned bar comparisons (length judgment), which prove to be less accurate than aligned bar comparisons (position judgment) (Cleveland and McGill, 1984; Heer and Bostock, 2010; Simkin and Hastie, 1987; Talbot et al., 2014).

As illustrated in Fig. 1, a grouped bar chart puts all the bars on a common baseline to support attribute comparisons especially *within* each item (Streit and Gehlenborg, 2014). To a certain degree, this chart type still allows users to compare attributes across items, but such comparisons are not as easy as comparisons within an item due to the presence of distractors (other bars) and the gap between items. Grouped bar charts support single-attribute comparisons well but not overall-attribute comparisons across items.

2.1.6. Layered bar charts

Similar to grouped bar charts, layered bar charts remove the need for nonaligned bar comparisons as in stacked bar charts (see Fig. 1). While grouped bar charts are ideal for comparing attributes within an item, layered bar charts are suitable for comparing individual attributes *across* items (Streit and Gehlenborg, 2014). This chart type, however, is not ideal for overall-attribute comparisons due to gaps between the attribute bars. It is not easy to get a sense of and compare the overall quality of items in layered bar charts.

Please cite this article in press as: Indratmo, et al., The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics (2018), https://doi.org/10.1016/j.visinf.2018.09.002.

2.2. The perceptual accuracy of bar charts

Since the primary purpose of visualization is to communicate data to users, many studies have focused on the perceptual accuracy of bar charts (Cleveland and McGill, 1984; Heer and Bostock, 2010; Simkin and Hastie, 1987; Talbot et al., 2014; Skau et al., 2015; Srinivasan et al., 2018; Zacks et al., 1998). These studies examined various issues that might affect perceptual accuracy and interpretation, such as the positions of bars (aligned vs. nonaligned), the presence of distractors (other bars not being judged), the addition of depth cues in bar charts, and the designs of multi-series bar charts.

In stacked bar charts, aligned bars can be compared more easily than nonaligned bars because bars on the same baseline allow viewers to make position comparisons instead of length comparisons (Cleveland and McGill, 1984; Heer and Bostock, 2010; Simkin and Hastie, 1987; Talbot et al., 2014). Moreover, these studies have shown that position judgment is perceptually more accurate than length judgment and therefore questioning the effectiveness of stacked bar charts. However, other research suggests that, for specific tasks (e.g., combining the total number of data in different categories), the sum of the bar segments in a stacked bar chart can help users complete the tasks effectively and efficiently (Böschen et al., 2017). Further, knowing how the sum is composed also provides contextual information to users and allows users to examine each attribute of data in detail.

Zacks et al. (1998) investigated the effects of depth cues and nearby graphical elements on the perceptual accuracy of bar charts. Adding depth cues to bar charts lowers perceptual accuracy, as these depth cues do not necessarily encode additional information and may be distracting to some viewers. Earlier studies also suggest that volume judgment is less accurate than length and area judgment (Cleveland and McGill, 1984; Mackinlay, 1986). Compared to depth cues, the height of the target bar and its relative height to nearby bars have more significant effects on accuracy, causing distortions on the perceived height of the target bar. This result is consistent with research on parallel-lines illusion, where lines of the same length may be perceived differently (under- or overestimated) depending on the length and location of adjacent lines (Jordan and Schiano, 1986).

2.3. The effects of embellishments in bar charts

With the availability of many visualization tools, people can create embellished bar charts easily, such as bar charts with rounded or pointy ends. While such modification might improve the aesthetics of bar charts, embellished bar charts, in most cases, lower perceptual accuracy compared to the standard bar chart (Skau et al., 2015). Skau's study shows that users are less accurate both in judging the absolute values of bars and in assessing relative difference (in percentage) between two bars in embellished bar charts. None of the embellished bar charts performs better than the standard bar chart.

However, when designed carefully, embellished charts may improve the memorability of the charts and information they convey, as the embellishments can provide additional recall cues to users (Bateman et al., 2010). For example, a chart depicting steep increases in government expenditure over a short period may use a monster illustration to convey a message that the increases are monstrous or shocking. The monster illustration may help users recall the central message of the chart (i.e., significant increases in government expenditure). Essentially, embellishments must be relevant to the content of charts; random illustrations attached to charts will not efficiently serve as recall cues.

2.4. Extensions to the traditional bar charts

Bar charts can only present a limited amount of information to users (Keim et al., 2002b, a, 2007). When a bar chart visualizes a complex dataset, it displays only the aggregate data. For example, a bar chart visualizing monthly sales data in a year shows twelve aggregate values across the months but does not show the individual transactions that contribute to these aggregate values. To improve business, data analysts need to identify the products or customers that have contributed most to the sales. However, this information is not visible in traditional bar charts.

To deal with this limitation, researchers have proposed extensions to the traditional bar charts (Keim et al., 2002b, a, 2007; Huang et al., 2009). These approaches use the area within bars in a bar chart to display individual items that make up the aggregate values. In this way, users can get both an overview of a dataset (as represented by the length of each bar) and a sense of how each aggregate value is composed of individual items (as shown in the area within each bar).

Pixel bar charts use the area within bars to visualize detailed items where each item is encoded using a single pixel (Keim et al., 2002b). Items within a bar are arranged based on one or two attributes that determine their x- and y-position within the bar. Then, the pixels representing these individual items are assigned colors based on a particular attribute of the data (e.g., price, quantity). Hierarchical pixel bar charts extend the concept of pixel bar charts further by allowing users to drill down data by selecting a specific bar/a subset of data (Keim et al., 2002a). When users select a particular bar (e.g., a specific year), the bar expands and displays more detailed information (e.g., monthly data of the selected year). This feature fits the hierarchical nature of business data. For example, sales data usually contain hierarchical information, such as time (year, month, day), location (city, neighborhood, store), and product (category, brand, item).

Business analysts need to identify critical items (e.g., large transactions) in their data so that they can focus their effort on these select entities. Pixel bar charts, however, treat individual items in a dataset uniformly so that no matter how critical an item is, each item is shown only as a single pixel. Value-cell bar charts proportionally map data values onto rectangular cells within bars to help analysts notice critical items in their data (Keim et al., 2007). For example, if a cell represents a $50 value, small transactions of less than $50 are combined into a single cell, while a $1000 transaction occupies 20 cells. The values of these transactions also determine the color of cells that represent them. In this way, users can notice the notable items in a dataset more efficiently, as these items are more visible than less critical items.

Treemap bar charts apply a similar approach used in pixel and value-cell bar charts (Huang et al., 2009). Treemap bar charts combine the features of traditional bar charts and the treemap visualization method (Johnson and Shneiderman, 1991). The treemap visualization method is a space-filling technique and used to place items in the area within a bar chart. Since this area is quite limited, treemap bar charts use a focus+context interaction technique (Rao and Card, 1994) to allow users to zoom in on a specific bar so that users can view an expanded treemap within the selected bar.

2.5. Summary of related work

Table 2 summarizes related work on information visualization and where our research fits into these categories. Our research falls into the effectiveness category, as it focuses on comparing the efficacy of three types of stacked bar charts to support attribute-comparison tasks. We use earlier research on information visualization to analyze and interpret our experimental results.

Please cite this article in press as: Indratmo, et al., The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics (2018), https://doi.org/10.1016/j.visinf.2018.09.002.

Table 2
Comparison of our research with related work on information visualization.

Topics	Description
Perceptual accuracy	Assessing the perceptual accuracy of visual channels (including bar charts) (Cleveland and McGill, 1984; Heer and Bostock, 2010; Simkin and Hastie, 1987; Talbot et al., 2014; Skau et al., 2015; Zacks et al., 1998).
Embellishment	Evaluating the effects of embellishment in bar charts (Skau et al., 2015; Bateman et al., 2010).
Extension	Extending the features of bar charts by using the area within bars to visualize detailed items (Keim et al., 2002b, a, 2007; Huang et al., 2009).
Effectiveness	Assessing the effectiveness of bar charts for specific tasks (Saket et al., 2018; Srinivasan et al., 2018; Böschen et al., 2017) (our research falls into this category).
Application	Using stacked bar charts for specific purposes (Gratzl et al., 2013; Heiberger and Robbins, 2014).

3. The classical, inverting, and diverging stacked bar charts

3.1. The rationale for the chart types

We studied the classical, inverting, and diverging stacked bar charts because they are suitable for both single-attribute and overall-attribute comparisons (see Table 1). The classical stacked bar chart is the most common type of stacked bar charts and therefore served as a baseline in the experiment. Both the inverting and diverging stacked bar charts have the potential to facilitate the attribute-comparison tasks. In particular, the inverting chart would work well for overall-attribute comparisons because it offers a consistent visual representation of data (i.e., longer bars mean better). However, there is a possibility that users might forget or misinterpret the inverted bar segments in an inverting stacked bar chart. The diverging chart does not pose such a risk and might improve performance by providing two separate spaces for the different data attributes (higher better vs. lower better). Such spatial separation might help users perform overall-attribute comparisons and lower their cognitive load. We explored these aspects in the study.

3.2. Construction of the charts

Data shown in the charts were a snapshot of stock market data with six attributes (Fish, 0000): NY% Growth (“the percentage change of next year’s earnings estimate compared with this year’s estimate”), ROE (“the trailing twelve months’ rate of return on shareholder equity”), Dividend Yield (“the dividend yield of the latest dividend rate on an annualized basis”), EPS% Payout (“the annual dividend as a percentage of trailing twelve months Earnings Per Share”), PE Ratio (“the price/earnings ratio using trailing twelve months earnings divided into current price”), and PEG Ratio (“the price/earnings ratio divided by 5-year estimates growth rate”). This study considered that each attribute was equally important and that NY% Growth, ROE, and Dividend Yield values were considered as higher better, whereas EPS% Payout, PE Ratio, and PEG Ratio values were assumed to be lower better.

We designed the charts for the experiment as follows. Each bar segment in the stacked bar charts had the same maximum length (i.e., normalized) to deal with different scales of stock attributes and because all of the attributes were considered equally important. For the experiment purpose, each chart showed only two stocks at any time. The classical and diverging stacked bar charts assigned the maximum length of a bar segment to the higher attribute value between two stocks, and the lower value took a proportion of it. For example, consider that stock A and B had PE Ratio values of 12 and 36 correspondingly and that the maximum length of a bar segment was 90 pixels. Since stock B’s PE Ratio value was higher than stock A’s, the bar segment representing stock B’s PE Ratio value would be 90-pixels long whereas stock A’s 30-pixels long. The inverting stacked bar chart used the same principle, except the bars for the inverted attributes were swapped between these stocks, e.g., stock A’s PE Ratio bar segment would be 90-pixels long whereas stock B’s 30-pixels long.

We implemented all charts using D3 (Bostock et al., 2011) and used one of D3’s categorical color palettes (category10) to assign colors to the stacked bar charts. The bar segments were stacked horizontally to ease reading (Heiberger and Robbins, 2014; Knaflic, 2015). Since this study aimed to assess the perceptual accuracy and interpretation of data, the exact values of the attributes were not essential and therefore not shown in the charts.

3.2.1. Classical stacked bar charts

In the classical stacked bar chart, each attribute value is shown without any modification and is stacked together as illustrated in Fig. 2. The chart places attributes that have the same characteristics (lower better or higher better) together to help users perform the comparison tasks. In this case, the lower-better attributes (PEG Ratio, PE Ratio, and EPS% Payout) are stacked together followed by the higher-better attributes (Dividend Yield, ROE, and NY% Growth).

3.2.2. Inverting stacked bar charts

The inverting stacked bar chart inverts the mapping function for the lower-better attributes (PEG Ratio, PE Ratio, and EPS% Payout) so that their visual representations have a consistent meaning with the other attributes, i.e., longer bars mean better. The inverted bar segments are filled with a pattern to give a visual cue to users about the inversion. For example, if we are looking for a lower PE Ratio stock and comparing CL and KO as shown in Fig. 3, the correct answer is KO because KO’s PE Ratio bar segment is more extended than CL’s. The mapping function for the higher-better attributes remain the same. For example, a longer ROE bar segment indicates a higher ROE value.

3.2.3. Diverging stacked bar charts

The diverging stacked bar chart uses the right area to display the lower-better attributes and the left area for the higher-better attributes (see Fig. 4). This chart uses the normal mapping function for all attributes. Thus, shorter bar segments indicate lower values and vice versa.

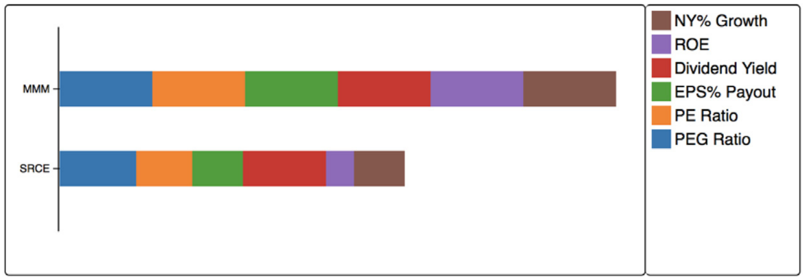

Fig. 2. Classical stacked bar chart.

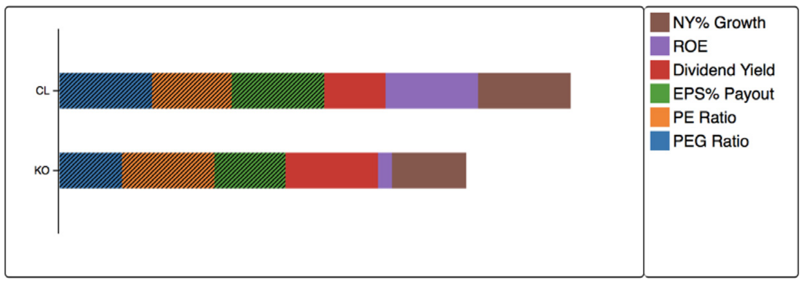

Fig. 3. Inverting stacked bar chart.

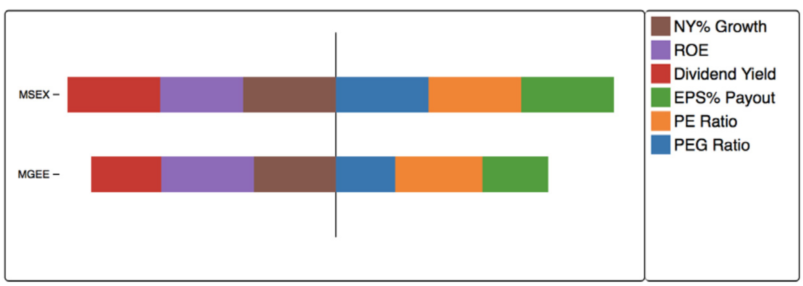

Fig. 4. Diverging stacked bar chart.

4. Methods

4.1. Motivation and objective

Our research was motivated by the application of interactive bar charts for exploring multi-attribute rankings in LineUp (Gratzl et al., 2013). LineUp enables users to compare rankings of items (e.g., universities) based on multiple attributes (e.g., academic reputation, faculty/student ratio, citations per faculty). The application supports various types of stacked bar charts and allows users to change the visual representation of data from one type to another (e.g., from classical to diverging stacked bar charts). However, little is known about which chart type would be best for supporting attribute comparisons especially when the data attributes have mixed characteristics. Our study aimed to provide empirical evidence to answer this research question.

The primary objective of this study was to assess the effectiveness of stacked bar charts for supporting comparison tasks. This study explored the effects of two factors – chart type and task type – on the completion time, error rate, and perceived difficulty of the comparison tasks. The chart types were classical, inverting, and diverging stacked bar charts. The comparison tasks consisted of single-attribute and overall-attribute comparisons. Each chart displayed six attributes of stock market data (Fish, 0000) where half of the attributes were assumed to be lower better, whereas the other half higher better (see Section 3). This assumption was stated explicitly as part of the instructions given to participants.

4.2. User tasks: single-attribute and overall-attribute comparisons

In the single-attribute comparison task, participants were asked to compare two stocks and select the one that had a lower price-to-earnings (PE) ratio. As shown in Fig. 2, PE ratio values were displayed as the second bar segment from the baseline so that they required nonaligned comparison. We made this arrangement to reflect a real-case scenario where users may focus on a particular attribute in a given stacked bar chart with the assumption that the attribute can be anywhere in the stack. Further, we designed this task so that participants needed to compare inverted bar segments (PE Ratio) in the inverting stacked bar chart to assess the effectiveness of the inverting chart. Below is the excerpt for the single-attribute comparison task description:

> Single Attribute Comparison Task. In this task, you only need to focus on a single attribute – PE Ratio – and select the stock that has the lower PE Ratio. Please try to complete the task carefully and quickly. Note that PE, PEG, and EPS% Payout are all lower is better, while Dividend Yield, ROE, and NY% Growth are all higher is better.

In the overall-attribute comparison task, participants were asked to compare two stocks and select the best overall stock. This

Please cite this article in press as: Indratmo, et al., The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics (2018), https://doi.org/10.1016/j.visinf.2018.09.002.

task assumed that each attribute had equal weight. Therefore, the best overall stock was a stock that, in total, had lower PEG ratio, PE ratio, and EPS% payout values and higher dividend yield, ROE, and NY% growth values. Below is the overall-attribute comparison task description given to participants:

> Overall Comparison Task. In this task, assume that each attribute has equal weight and select the best overall stock. Please try to complete the task carefully and quickly. Note that PE, PEG, and EPS% Payout are all lower is better, while Dividend Yield, ROE, and NY% Growth are all higher is better.

In the inverting stacked bar chart, the best overall stock between two stocks was merely the longer bar because all bar segments had a consistent meaning in the inverting chart. In the classical and diverging stacked bar charts, we could classify the six bar segments representing six attributes of the stock market data into two categories: shorter-better or longer-better bars. Selecting the best overall stock involved computing the total difference between these bars. For example, consider that stock A had a total length of 200 pixels shorter-better bar and 180 pixels longer-better bar and that stock B had a total length of 240 pixels shorter-better bar (worse than A's by 40 pixels) and 250 pixels longer-better bar (better than A's by 70 pixels). Compared to stock A, stock B was the better overall stock because, for stock B, the total difference between shorter-better bars (−40 pixels) and longer-better bars (+70 pixels) was positive (+30 pixels). Recall that every data attribute was assumed to have equal weight in this task, and the implementation of the charts followed this assumption.

4.3. Performance measures

The effectiveness of stacked bar charts was measured using three variables: the completion time (time to complete a task), error rate (the number of incorrect answers), and perceived difficulty completing a task using a particular chart. The completion time and error rate were collected automatically using the software used by participants to complete the tasks. The perceived difficulty was collected after each participant completed a set of tasks using the following questions:

- Finding a stock that has a lower price-to-earnings (PE) ratio was…(1: easy, 5: difficult)
- Finding the best overall stock was…(1: easy, 5: difficult)

4.4. Participants

We recruited 30 participants from undergraduate students who were taking a second-year computer science course in introduction to human–computer interaction. All participants used the same equipment and performed the given tasks in the same computer lab. They did not receive an honorarium, but we provided snack food after each data collection session as appreciation for their time and participation. Participants did not need to fully understand the meaning of the attributes of stock market data used in the study, as they received information that NY% Growth, ROE, and Dividend Yield values were considered as higher better, while EPS% Payout, PE Ratio, and PEG Ratio values were assumed to be lower better. This information was available as part of the instructions.

4.5. Experimental design

We used a 3 chart types × 2 task types, within-subjects design. Each participant used all variants of stacked bar charts to perform both single-attribute and overall-attribute comparisons. Each combination of chart type and task type was repeated five times. For example, each participant did five single-attribute comparisons using a classical stacked bar chart (i.e., comparing five randomly selected pairs of stocks). The completion time and error rate were recorded every time a participant compared a pair of stocks.

Table 3
The experimental matrix.

Group order	The order of chart types seen by participants		
1	Classical	Inverting	Diverging
2	Diverging	Classical	Inverting
3	Inverting	Diverging	Classical

We used a Latin Square to counterbalance the order of chart types seen by participants and to control learning effects associated with a within-subjects design (MacKenzie, 2013). Table 3 shows the 3 × 3 Latin Square used to determine the order of chart types used by participants in our study. We invited the whole class of students to our study and then scheduled the experiments based on the availability of participants at random. Each participant was assigned to group 1, 2, or 3 based on the order of their availability. That is, the first participant was assigned to group 1, the second one to group 2, the third one to group 3, the fourth one to group 1, and so on. For example, participants assigned to group 3 performed the tasks using variants of stacked bar charts in the following order: inverting, diverging, and classical.

4.6. Procedure

At the beginning of each experimental session, participants received information about the objective of the study, the experimental protocol, and the required tasks. Then, each participant was assigned to a group that determined the order of chart types they would see in the experiment (see Table 3). Based on this group order, the software displayed a classical, inverting, or diverging stacked bar chart. The features of the displayed chart were explained to the participants and then participants received a practice task. The practice task had to be completed correctly to ensure that participants were familiar with the chart and understood the task.

For each chart type, participants always performed single-attribute comparisons first before performing overall-attribute comparisons. After performing a comparison task using a specific chart type (e.g., single-attribute comparisons using a classical stacked bar chart), participants filled out a paper-based perceived difficulty questionnaire that asked how easy or difficult the task was. Then, the participants moved on to the next task or chart type.

5. Data analysis and results

This study used a within-subjects design to explore the effects of two factors (chart types and task types) on the completion time, error rate, and perceived difficulty. Given the settings of the experiment, it is appropriate to test the mean differences between the two conditions using two-way repeated measures analysis of variance (RM ANOVA). Mauchly's test showed that the data met the assumption of sphericity, which is essential to RM ANOVA. When the F values were significant, post hoc pairwise comparisons were performed with Bonferroni adjustment. The experimental data were analyzed using SPSS.

Table 4 provides the descriptive statistics of the results. It shows the means and standard deviations of the performance measures across two different task types (single-attribute and overall-attribute comparisons) and three different chart types (classical, inverting, and diverging). The completion time indicates how much time (in seconds) participants needed to complete a specific task using a particular chart. The error rate indicates the number of incorrect answers submitted by participants. The perceived difficulty indicates how easy or difficult participants felt about completing a particular task using a specific chart (1: easy, 5: difficult).

Table 4
Means (M) and standard deviations (SD) of completion time, error rate, and perceived difficulty ($N = 30$).

Task type	Chart type	Completion time		Error rate		Perceived difficulty	
		M	SD	M	SD	M	SD
Single-attribute comparisons	Classical	13.43	8.09	0.27	0.64	2.03	1.00
	Inverting	14.74	7.10	0.27	0.691	2.00	1.23
	Diverging	14.17	5.00	0.23	0.43	1.93	1.05
Overall-attribute comparisons	Classical	28.97	13.66	0.7	1.236	3.13	1.17
	Inverting	12.64	10.41	0.23	0.679	1.80	1.19
	Diverging	27.91	13.29	0.4	0.855	2.40	1.04

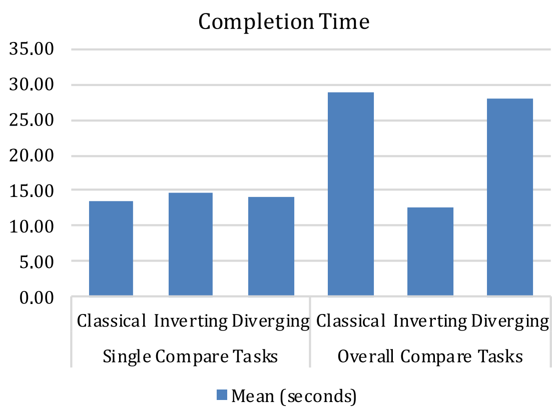

Fig. 5. The mean values of completion time for each chart type for the comparison tasks.

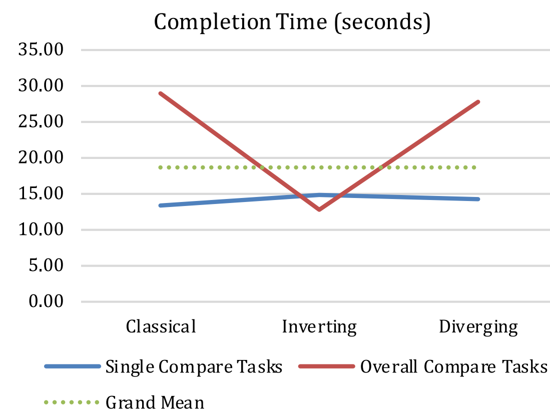

Fig. 6. The interaction effect between chart type and task type on completion time.

5.1. Effect of chart type and task type on completion time

Two-way RM ANOVA showed a main effect of chart type, $F(2, 58) = 14.642$ ($p < .001$), task type, $F(1, 29) = 41.935$ ($p < .001$), and an interaction between the chart type and task type, $F(2, 58) = 16.922$ ($p < .001$), on completion time.

The mean completion time for the inverting stacked bar chart was the lowest among the other means ($M = 12.64$, $SD = 10.41$), indicating that participants performed best in the overall-attribute comparison task when they used the inverting stacked bar chart (see Fig. 5). Participants took significantly less time to complete the overall-attribute comparison task while using the inverting chart type ($M = 12.64$) compared to the classical ($M = 28.97$) and diverging chart type ($M = 27.91$). Post hoc pairwise comparisons confirmed a significant difference between the mean completion time of the classical and inverting chart type ($p < .001$), and between the mean completion time of inverting and diverging chart type ($p < .001$).

No significant differences were found between the mean completion time among chart types for the single-attribute comparison task. This result suggests that, for single-attribute comparisons, all chart types delivered similar performance regarding completion time.

Fig. 6 illustrates the interaction between chart type and task type on completion time. The mean completion time for overall-attribute comparisons was significantly longer when participants used the classical stacked bar chart ($p < .001$) and the diverging stacked bar chart ($p < .001$) compared to the mean completion time for single-attribute comparisons. When they used the inverting stacked bar chart, however, completing single-attribute and overall-attribute comparisons required a similar amount of time.

5.2. Effect of chart type and task type on error rate

The results showed no significant differences between the mean error rates across different chart types and task types. Most of the comparison tasks were performed correctly as indicated by the low mean values of error rate. This finding suggests that, given enough time, participants were able to process information delivered by all three stacked bar charts correctly.

5.3. Effect of chart type and task type on perceived difficulty

Two-way RM ANOVA revealed a main effect of chart type, $F(2, 58) = 7.773$ ($p = .001$), task type, $F(1, 29) = 7.421$ ($p = .011$), and an interaction between chart type and task type, $F(2, 58) = 6.34$ ($p = .003$), on perceived difficulty.

For overall-attribute comparisons, the mean perceived difficulty of the inverting chart type was the lowest among the other means ($M = 1.80$, $SD = 1.19$). This result suggests that participants perceived the inverting chart type the easiest to use for completing overall-attribute comparisons (see Fig. 7). Post hoc pairwise comparisons revealed a significant difference between the mean perceived difficulty of the classical and inverting chart types ($p < .001$), and between the mean perceived difficulty of the classical and diverging chart types ($p = .049$). This implies that participants were least satisfied with the classical chart type ($M = 3.13$) compared to the inverting ($M = 1.80$) and diverging ($M = 2.40$) chart types. However, there was no significant difference between the mean perceived difficulty of the inverting and diverging chart types. For single-attribute comparisons, there were no significant differences in the mean perceived difficulty of all chart types.

Fig. 8 illustrates the interaction between chart type and task type on perceived difficulty. Post hoc pairwise comparisons showed a significant difference between the mean perceived difficulty of the classical chart type ($p < .001$) for single-attribute comparisons ($M = 2.03$) and overall-attribute comparisons ($M = 3.13$). However, there was no significant difference in the mean perceived difficulty of the inverting and diverging chart types for single-attribute and overall-attribute comparisons.

Please cite this article in press as: Indratmo, et al., The efficacy of stacked bar charts in supporting single-attribute and overall-attribute comparisons. Visual Informatics (2018), https://doi.org/10.1016/j.visinf.2018.09.002.

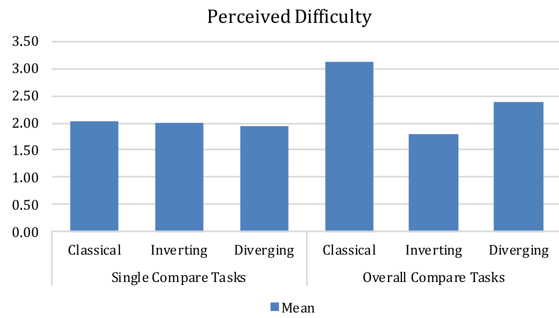

Fig. 7. The mean values of perceived difficulty (1: easy, 5: difficult).

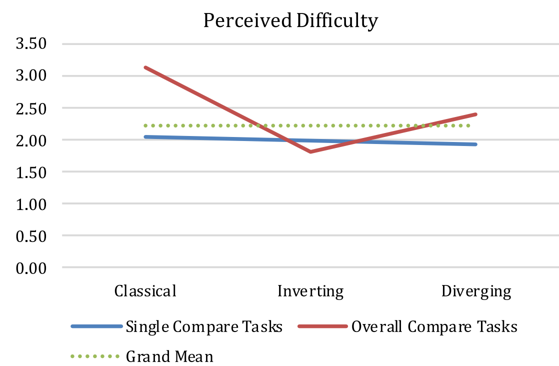

Fig. 8. The interaction effect between chart type and task type on perceived difficulty.

6. Discussion

Visualization is the hallmark of data literacy and knowledge utilization in a digital age with unprecedented growth in volumes of data. Visualization techniques evoke the user's visual perception, fast pattern detection, and recognition (Saket et al., 2018). Understanding various forms of charts and their appropriate uses can enhance understanding of the underlying data. Therefore, studying the efficacy and limitations of visualization methods, such as stacked bar charts, is fundamental research with critical implication to knowledge creation and utilization.

This study examined the effects of stacked bar chart type and task type on indicators such as the completion time, error rate, and perceived task difficulty. The aim is to contribute to the growing research on the value and utility of various forms of visualization charts and their variant contributions to efficacy on the user's tasks performance. This section highlights the main findings from the study and their implications for the design of useful stacked bar charts and visualization tools.

6.1. Single-attribute comparisons

For single-attribute comparisons, all charts delivered similar performance across the performance measures: the completion time was considered reasonable, and both the error rate and perceived difficulty were low. Participants thought it was easy to complete the single-attribute comparison task using any given charts. These results imply that comparing two nonaligned bars in a stacked bar chart can be performed quickly and accurately using the classical, inverting, and diverging stacked bar charts.

The single-attribute comparison task required participants to compare inverted bar segments while using the inverting chart to see if there were any significant disadvantages of using this chart. The experimental results indicate that the inverting chart seemed to have not caused any confusion or put more time pressure on users even though some of the bar segments have been inverted (i.e., longer bars represent lower values). It is worth noting, however, that the inverted bar segments were filled with a pattern so that they appeared different from the regular bar segments and that participants performed the task immediately after they learned about the inverting chart. These factors might have played a role in preventing potential confusion with the inverting stacked bar chart.

6.2. Overall-attribute comparisons

There were no significant differences with regards to the error rates of the three chart types. Most overall-attribute comparisons were performed correctly by participants. This result implies that participants were able to process information represented by the classical, inverting, and diverging charts correctly despite the different characteristics of the attributes (lower better vs. higher better).

Regarding the completion time, the inverting stacked bar chart was the most effective for supporting overall-attribute comparisons. In the inverting chart, all bar segments representing attributes of data have a consistent meaning with regards to the task at hand. That is, longer bars represent better values and vice versa. When participants compared the overall attributes of two stocks, they could choose the longer bar, as it represents the better overall stock. They needed only to focus on the sum of stacked bar segments and did not have to pay attention to individual attributes of data. Such comparisons require only position judgment, which is more accurate and faster than length judgment (Cleveland and McGill, 1984; Heer and Bostock, 2010; Simkin and Hastie, 1987; Talbot et al., 2014).

The other performance measures also indicated the simplicity of performing overall-attribute comparisons using the inverting stacked bar. The error rate of the inverting chart was low, and its perceived difficulty was significantly lower than the perceived difficulty of the classical chart. This result was expected for this task because participants had to consider the different characteristics of the bar segments in the classical stacked bar chart.

It is worth noting that both the classical and diverging charts required participants to account for the differences between the sum of shorter-better bar segments and the sum of longer-better bar segments to find the best overall stock. Consequently, it took more time to complete the overall-attribute comparison task using the classical and diverging charts compared to the time needed using the inverting chart.

Interestingly, for overall-attribute comparisons, the perceived difficulty in using the diverging and inverting charts was similar, and the scores indicated lower perceived difficulty in using these charts compared to the perceived difficulty in using the classical chart. This result was not surprising for the inverting chart, as the chart was the most effective regarding the completion time. While using the diverging chart, however, participants needed more time to perform overall-attribute comparisons, and the completion time for this task was similar to that with the classical chart. Nonetheless, participants thought it was easier to perform the task using the diverging chart than doing it using the classical chart. The spatial separation between the lower-better attributes (right area) and the higher-better attributes (left area) in the diverging chart might have helped lower the cognitive load of participants (see Fig. 4).

Overall-attribute comparisons took more time than single-attribute comparisons when participants performed the tasks using the classical and diverging charts. This observation suggests that overall-attribute comparisons are a more complicated endeavor than single-attribute comparisons. In particular, participants must account for the differences between the sum of shorter-better bar segments and the sum of longer-better bar segments when working with the classical and diverging charts. However, the inverting chart was an exception in this case: the completion time for single-attribute and overall-attribute comparisons was not significantly different when participants used the inverting chart. The inverting chart made a complex task simpler by transforming the visual representation of the attributes so that they all had a consistent meaning. This consistency has transformed the overall-attribute comparison task into a more straightforward task of finding a longer bar in the chart.

6.3. Implications for design

This study demonstrated that different variants of stacked bar charts had different levels of effectiveness for supporting overall-attribute comparisons. Although all of the charts contained the same information, users found that some of the charts were perceived easier to use, and the inverting chart, in particular, stood out from the other charts. This finding suggests that, depending on the task at hand, individual charts can be more effective than others to support users to complete the task. More generally, different representations of the same information can change the difficulty level of tasks significantly (Larkin and Simon, 1987). Therefore, visualization tools should allow users, possibly guide them, to change visual representations of data that they explore.

Visual representations of data may reduce the complexity of tasks when they have a consistent meaning to the tasks. In our study, the inverting chart was the easiest to use and most useful for overall-attribute comparisons because its visual representation of data had a consistent meaning to the user's task. That is, longer bars represent better overall stocks and vice versa. This principle applies to other visual channels, such as position, area, and color hue and saturation. Since we cannot anticipate all user needs in advance, what we can do is to allow users to specify how they want to use the available visual channels and provide logical options to change the visual representation of data (e.g., using normalization or inversion). Our finding suggests that, if we can transform a task into a more straightforward perceptual task (e.g., length judgment into position judgment as in the case of the inverting chart in our study), it will reduce the user's cognitive load and may improve the completion time and lower the perceived difficulty of the task.

Finally, spatial separation of contrasting information might have an essential role in the perceived difficulty of a task. In our study, participants spent a similar amount of time to complete overall-attribute comparisons while using the classical and diverging stacked bar charts. However, they perceived the diverging chart as easier to use than the classical chart. These charts differed only in where the lower-better and higher-better attributes were visualized. Having a sense of spatial separation in the diverging chart has lowered the perceived difficulty of the overall-attribute comparison task. Visualization tools can apply this finding by allowing users to create boundaries of their workspace (e.g., arranging where to display charts on their dashboard).

6.4. Limitations of the study

Our participants did not have any difficulty in completing the given tasks using the inverting stacked bar chart. They performed the tasks right after they learned about the features of the inverting stacked bar chart. Therefore, they were fully aware of the fact that some of the bars had been inverted for the experimental purposes and were able to interpret the data correctly. This ease of use, however, is not necessarily guaranteed when inverting stacked bar charts were accessed a few days or weeks later or used by new users. Users might miss the inverted values/bars and consequently misinterpret the charts. Visualization designers should be aware of this risk and take the necessary steps to prevent potential confusion with the inverting stacked bar chart.

Our study focused on the quantitative approach to assessing the effectiveness of stacked bar charts. While this approach is useful for showing how chart types and task types affect the performance measures, the experimental data did not capture users' experiences and possible challenges when performing the tasks. Also, the potential problem with the inverting chart (as discussed above) could not be observed in such short experiment sessions. It would have been useful to triangulate the experimental outcomes with qualitative techniques, such as interviews.

Our participants had a high level of computer literacy, as they were taking a second-year computer science course. Prior knowledge might influence the outcomes of the experiment. Therefore, findings in the study may not be readily generalized to non-computer users.

7. Conclusion

This study assessed the effectiveness of three variants of stacked bar charts – classical, inverting, and diverging – for supporting single-attribute and overall-attribute comparisons. We used completion time, error rate, and perceived difficulty as performance measures. The experimental results showed that the inverting stacked bar chart allowed participants to complete the overall-attribute comparisons with the least amount of time and that the inverting and diverging charts were perceived to be easier to use than the classical chart for overall-attribute comparisons. Further, the error rates for all charts for both tasks were low and not significantly different.

We attributed the effectiveness of the inverting stacked bar chart to two factors. First, the visual representation of data had a consistent meaning with regards to the user's task at hand: longer bars meant better overall stocks and vice versa. Second, this consistency reduced the complexity of assessing the overall quality of stocks into a simple perceptual task of position judgment, which users can perform accurately and quickly.

We highlighted how our findings could contribute to the design of useful visualization tools. First, visualization tools should allow users to modify visual representations of data to support user tasks. Second, visual representations of data that have a consistent meaning to the user's task have the potential to reduce the complexity of the task. Finally, having spatial separation of contrasting information may lower the perceived difficulty of the task at hand.

There are several directions to follow up on this research and address its current limitations. For instance, the efficacy of stacked bar charts can be assessed in a longitudinal study where participants use the charts to support their day-to-day tasks in the real world. The chart usage can be measured and compared over an extended period to understand the usability of the charts. Since the user's experiences and taste affect the usability of an object, a study that combines qualitative and quantitative methods would help gain broader and in-depth insights into the use of stacked bar charts. Data from interviews and think-aloud protocols, for example, may shed better insights into the reason why a specific type of stacked bar charts works better than others for supporting specific tasks for a particular group of users. Further, in a longitudinal study, we will be able to observe how participants use charts to compare and analyze data, switch between one chart type to another type, and other practice that is not observable in a lab setting.

Acknowledgments

We thank Dr. Sharon Bratt for allowing her students to participate in this study.

Funding

Lee Howorko received funding from MacEwan University, Canada through the Undergraduate Student Research Initiative Grant.

Authors' statement

We have reviewed and approved this article for publication.

Conflict of interest

None.

References

Bateman, S., Mandryk, R.L., Gutwin, C., Genest, A., McDine, D., Brooks, C., 2010. Useful junk? the effects of visual embellishment on comprehension and memorability of charts. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 2573–2582. http://dx.doi.org/10.1145/1753326.1753716.

Böschen, F., Strobel, B., Goos, S., Liebers, C., Rathje, B., Scherp, A., 2017. Evaluation of the comprehensiveness of bar charts with and without stacking functionality using eye-tracking. In: Proceedings of the Conference on Human Information Interaction and Retrieval. pp. 337–340. http://dx.doi.org/10.1145/3020165.3022147.

Bostock, M., Ogievetsky, V., Heer, J., 2011. D3: data-driven documents. IEEE Trans. Vis. Comput. Graphics 17 (12), 2301–2309. http://dx.doi.org/10.1109/TVCG.2011.185.

Brehmer, M., Munzner, T., 2013. A multi-level typology of abstract visualization tasks. IEEE Trans. Vis. Comput. Graphics 19 (12), 2376–2385. http://dx.doi.org/10.1109/TVCG.2013.124.

Brinton, W.C., 1939. Graphic Presentation. Brinton Associates, New York City, URL https://archive.org/details/graphicpresentat00brinrich.

Cleveland, W., McGill, R., 1984. Graphical perception: theory, experimentation, and application to the development of graphical methods. J. Amer. Statist. Assoc. 79 (387), 531–554. http://dx.doi.org/10.2307/2288400.

Fish, D., 2015. U.S. dividend champions [online]. Accessed 03.04.15.

Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., Streit, M., 2013. LineUp: visual analysis of multi-attribute rankings. IEEE Trans. Vis. Comput. Graphics 19 (12), 2277–2286. http://dx.doi.org/10.1109/TVCG.2013.173.

Healey, C.G., 1996. Choosing effective colours for data visualization. In: Proceedings of the 7th Conference on Visualization. pp. 263–270. http://dx.doi.org/10.1109/VISUAL.1996.568118.

Healey, C.G., Enns, J., 2012. Attention and visual memory in visualization and computer graphics. IEEE Trans. Vis. Comput. Graphics 18 (7), 1170–1188. http://dx.doi.org/10.1109/TVCG.2011.127.

Heer, J., Bostock, M., 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 203–212. http://dx.doi.org/10.1145/1753326.1753357.

Heiberger, R., Robbins, N., 2014. Design of diverging stacked bar charts for Likert scales and other applications. J. Stat. Softw. 57 (5), 1–32. http://dx.doi.org/10.18637/jss.v057.i05.

Huang, M.L., Huang, T.H., Zhang, J., 2009. Treemapbar: Visualizing additional dimensions of data in bar chart. In: Proceedings of the 13th International Conference on Information Visualisation. pp. 98–103. http://dx.doi.org/10.1109/IV.2009.22.

Johnson, B., Shneiderman, B., 1991. Tree-maps: A space-filling approach to the visualization of hierarchical information structures. In: Proceedings of the 2nd Conference on Visualization. pp. 284–291.

Jordan, K., Schiano, D.J., 1986. Serial processing and the parallel-lines illusion: length contrast through relative spatial separation of contours. Percept. Psychophys. 40 (6), 384–390. http://dx.doi.org/10.3758/BF03208197.

Keim, D.A., Hao, M.C., Dayal, U., 2002a. Hierarchical pixel bar charts. IEEE Trans. Vis. Comput. Graphics 8 (3), 255–269. http://dx.doi.org/10.1109/TVCG.2002.1021578.

Keim, D.A., Hao, M.C., Dayal, U., Hsu, M., 2002b. Pixel bar charts: a visualization technique for very large multi-attribute data sets. Inf. Vis. 1 (1), 20–34. http://dx.doi.org/10.1057/palgrave.ivs.9500003.

Keim, D.A., Hao, M.C., Dayal, U., Lyons, M., 2007. Value-cell bar charts for visualizing large transaction data sets. IEEE Trans. Vis. Comput. Graphics 13 (4), 822–833. http://dx.doi.org/10.1109/TVCG.2007.1023.

Knaflic, C.N., 2015. Storytelling with Data: A Data Visualization Guide for Business Professionals. John Wiley & Sons, Hoboken.

Larkin, J.H., Simon, H.A., 1987. Why a diagram is (sometimes) worth ten thousand words. Cogn. Sci. 11 (1), 65–100. http://dx.doi.org/10.1111/j.1551-6708.1987.tb00863.x.

MacKenzie, I.S., 2013. Human-computer Interaction: An Empirical Research Perspective. Elsevier, Waltham.

Mackinlay, J., 1986. Automating the design of graphical presentations of relational information. ACM Trans. Graph. 5 (2), 110–141. http://dx.doi.org/10.1145/22949.22950.

Mei, H., Chen, W., Ma, Y., Guan, H., Hu, W., 2018. VisComposer: A visual programmable composition environment for information visualization. Vis. Inf. 2 (1), 71–81. http://dx.doi.org/10.1016/j.visinf.2018.04.008.

Munzner, T., 2015. Visualization Analysis and Design. CRC Press, Boca Raton.

Rao, R., Card, S.K., 1994. The table lens: Merging graphical and symbolic representations in an interactive focus + context visualization for tabular information. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 318–322. http://dx.doi.org/10.1145/191666.191776.

Saket, B., Endert, A., Demiralp, C., 2018. Task-based effectiveness of basic visualizations. IEEE Trans. Vis. Comput. Graphics http://dx.doi.org/10.1109/TVCG.2018.2829750.

Simkin, D., Hastie, R., 1987. An information-processing analysis of graph perception. J. Amer. Statist. Assoc. 82 (398), 454–465.

Skau, D., Harrison, L., Kosara, R., 2015. An evaluation of the impact of visual embellishments in bar charts. Comput. Graph. Forum 34 (3), 221–230. http://dx.doi.org/10.1111/cgf.12634.

Spiegelhalter, D., Pearson, M., Short, I., 2011. Visualizing uncertainty about the future. Science 333 (6048), 1393–1400. http://dx.doi.org/10.1126/science.1191181.

Srinivasan, A., Brehmer, M., Lee, B., Drucker, S.M., 2018. What's the difference? Evaluating variations of multi-series bar charts for visual comparison tasks. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 304:1–304:12. http://dx.doi.org/10.1145/3173574.3173878.

Streit, M., Gehlenborg, N., 2014. Bar charts and box plots. Nat. Methods 11 (2), 117. http://dx.doi.org/10.1038/nmeth.2807.

Talbot, J., Setlur, V., Anand, A., 2014. Four experiments on the perception of bar charts. IEEE Trans. Vis. Comput. Graphics 20 (12), 2152–2160. http://dx.doi.org/10.1109/TVCG.2014.2346320.

Treisman, A., Gormican, S., 1988. Feature analysis in early vision: evidence from search asymmetries. Psychol. Rev. 95 (1), 15–48. http://dx.doi.org/10.1037/0033-295X.95.1.15.

Ware, C., 2004. Information Visualization: Perception for Design, second ed. Elsevier, San Francisco.

Zacks, J., Levy, E., Tversky, B., Schiano, D.J., 1998. Reading bar graphs: effects of extraneous depth cues and graphical context. J. Exp. Psychol. Appl. 4 (2), 119–138. http://dx.doi.org/10.1037/1076-898X.4.2.119.