Star Plots and Fundamental Analysis: Evaluation of Cognitive Limitations of Shape-Based Financial Benchmarking

Sujay Shalawadi*
Bauhaus Universität Weimar
sujay.shalawadi@uni-weimar.de

Anton Frolov†
Bauhaus Universität Weimar
anton.frolov@uni-weimar.de

ABSTRACT

This paper describes a user study conducted with a goal to evaluate the empirical usability of star plots applied to the task of financial benchmarking. The amount of error in user evaluations for direct and inverted mapping was assessed, operationalizing the influence of the star plot shape. The study concluded that star plots are favourable for specific tasks in financial benchmarking, with future research directions outlined on the basis of evaluated results.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in visualization**;

KEYWORDS

star plot, financial data, benchmarking, visualization

ACM Reference Format:
Sujay Shalawadi and Anton Frolov. 2018. Star Plots and Fundamental Analysis: Evaluation of Cognitive Limitations of Shape-Based Financial Benchmarking

1 INTRODUCTION

The use of advanced information visualization techniques representing multivariate financial datasets has been prevalent in the recent decade, with a growing body of research devoted to improving the usability of said visualizations. In this paper we wanted to focus on a particular application of a technique called star plot to the task of financial benchmarking.

Star plot is a method of displaying multivariate data. Each star represents a single observation. Star plots are two-dimensional charts of three or more quantitative variables represented on axes starting from the same point. The use of star plots could be beneficial for financial benchmarking.

Financial benchmarking is the process of comparing the business practices and performance standards of a company to that of other firms within the same industry. The data typically considered are related to the growth and profit, and the valuation of the company.

The information gained from financial benchmarking allows firms to determine how well they perform in comparison with other companies in their domain. This information also helps investors to make decisions regarding their investments in terms of short term and long term gains.

The process of financial benchmarking is a tedious process because of the necessity to keep the data up to date from financial data repositories. The sheer volume and speed of data coming in from companies is a challenge for investors to keep up. When information floods in, it becomes difficult to focus on important data. Finding the right resource due to the sheer volume, selecting the right information to be presented and using advertisements in a beneficial manner to support decisions of investors are main challenges in presenting the data.

The number of dimensions to financial data could be mundane for the human eye to focus simultaneously on several data values. The use of star plots to represent several disparate data values on a single plot could highly reduce the stress to the frontal cortex of the eye.

2 RELATED WORK

We wanted to get familiar with existing methods and concepts that could be beneficial for the design of our study.

In most interface design involving plots, quantitative metrics can easily be measured. They can include data density, occlusion rate, number of dimensions and number of identifiable points [2]. However, the real qualities to be evaluated are :

(1) visual representation usability, referring to the expressiveness and semantic quality of the resulting image;
(2) interface usability, related to interaction mechanisms to allow users interact with the data through the image;
(3) data usability, devoted mainly to the quality of data that support users' tasks.

To address this challenge, four criteria of testing usability of visual representations and three classes of testing interaction mechanism were identified by Carle et.al [2]. The four criteria are completeness, spatial organization, information coding and state transition. *Completeness* is associated with the concept of representing all the semantic contents of the data to be displayed. *Spatial organization* is related to the overall layout of a visual representation, which comprises analyzing how easy is to locate an information element on the display and to be aware of the overall distribution of information elements in the representation. *Information coding* is the use of additional symbols to aid perception of information. *State transition* is the smoothness with change of states of the system during the interaction.

Some of the popular evaluation methods listed in the work of Plaisant et.al [4] are:

*Immatrikulationsnummer. 117698
†Immatrikulationsnummer. 117747

(1) controlled experiments to compare specific widgets;
(2) tackling problems of users given as feedback;
(3) controlled experiments to compare two or more tools;
(4) case studies of users in their natural environment.

However, there are certain challenges that were also addressed in [4], e.g. matching tools to the users, tasks with real problems; improving user testing through realistic tasks; addressing universal usability to support users from diverse backgrounds. The inclination towards novelty must be complimented by integrating visualization tools into solutions of real problems.

The research conducted by Klippel et.al [3] explains the cognitive behaviour of users while analyzing star plots. The experiments conducted showed that certain shape features influences user's reaction time to respond. The shapes of the star plots were permuted by changing the axis labels to produce different shapes to analyze user response. The area of the star plot does not affect the response's correctness however the reaction time is faster for concave plots compare to convex plots.

The research conducted by Eklund et.al [1] develop a self organizing map (SOM) that is used to visualize multidimensional data on a 2 dimensional plot. The SOM is a 2 dimensional plot from a n-dimensional data. The method involves choosing a data material and then prepossessing it by standardization. Then, the network topology, learning rate and neighbourhood radius are defined. The use of SOM is preferable over other tools such as spreadsheets and graphs that cannot handle multidimensional data. However, for a dimension space of $6 \dots 10$ the use of SOMs might be considered excessive.

Star plots are used to display multivariate data and in combination with principle component analysis (PCA) more than two principle components can be displayed at a time. The research conducted by Wu et.al [6] on food and drugs data showed the better visualization of the multivariate dataset with star plots against PCA. It further showed that for larger datasets, star plots are faster to compute in comparison to PCA.

Numerous patents have been filed for financial benchmarking systems, such as one named **Financial visualization and analysis systems** by Caranama et.al. The patent describes a system which presents financial data and related non-financial data in a visual analog format for discussion and analysis by one or more users. The income statement, balance sheet and statement of cash flows are analyzed as a whole rather than separately as individualized statements. Additionally, it provides direct interfaces with underlying financial and related data. The analysis steps created by a user may be captured and recorded for later play back, communication and analysis.

3 RESEARCH QUESTION AND HYPOTHESIS

In our empirical study, we wanted to evaluate visual representation usability of star plots in representing *financial benchmarking dataset*(FBD) against the criteria of information coding in terminology of [4]. The goal of the study was to examine the prevalent tasks done to make analysis. Since the study is conducted in university setup, we managed to get a focus group of students aged between 20 to 30 with varying familiarity to the domain.

Table 1: Dependent and independent variables

Dependent	Independent
Error in shape area evaluation	Domain knowledge
Error in between-pairs ratio evaluation	Inquisitive interactions
Time of arrangement	Skewness of the shape

The research question that motivated our study was: *Are star plots a good visualization technique for representing FBD?*

With the research question defined, we concluded the hypotheses to be tested in the study. The hypotheses were:

H_0 : *The use of star plots to encode FBD does not introduce prohibitively large error in subjective user evaluation (not more than 6% … 15%, as per [5]).*

H_1 : *The skewness (assymetry) of the star plot shape affects the participants aptness to evaluate the area of the star plot.*

H_2 : *Between-pairs comparison of star plots enables an effective underlying ratio comparison.*

4 DATA COLLECTION

Following from the defined hypotheses the variables were identified along with the testing procedure and its conditions. The dependent and independent variables were defined as could be found in table 1.

The notion of error in $H0$ bears simultaneously both purely geometrical, i.e. difference between estimated and factual area of the shape, and domain-specific meaning and measure, wherein arbitrary weights are assigned to axes of the star plot by participants familiar to the domain (thus creating an implied confounding factor to what could be seen as a purely geometrical error assessment).

Considering this aspect, together with the absence of an ideal, peer-reviewed weighting scheme for financial benchmarking, we collected only the measured error in its geometrical sense, while simultaneously noting down presence of such factors as prior domain knowledge and inquisitive interactions with parameter descriptions and disambiguations, ubiquitously available in the interface.

To measure the skewness of the shape in $H1$ we had to derive a conservative metric which would measure the asymmetry of axis values. We devised such a metric:

$$Skew(A) = \frac{Var(A)\,\sigma(A)}{\mu(A)}, \tag{1}$$

where A is a set of normalized star plot axis values, $Var(A)$ is the variance of A and $\mu(A), \sigma(A)$ denote respectively the mean and standard deviation of the values in A. The values of $Skew$ are lower bounded with 0, where 0 is equivalent to the highest degree of rotational symmetry of the star plot. The value of $Skew$ changes conservatively along with the perceived skewness of the shape, as could be seen in figure 1.

To assess the participants ability to compare pairs of star plots in $H2$, we asked them to arrange a set of four pairs of star plots in the descending ranking order. To measure the amount of permutation error, we calculated the transpositional distance of the collected set permutation both to the ideal geometry-based permutation and to the ideal permutation based on the participants own evaluations. Second measurement was used to operationalize the factor of domain-specific axis weight assignment.

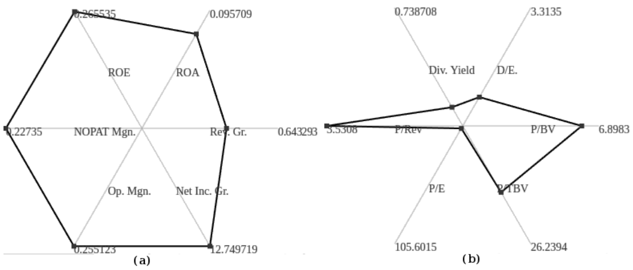

Figure 1: Skewness of the star plot shape, as measured by the Skew metric: (a) $4,397 \cdot 10^{-3}$**; (b)** $134,847 \cdot 10^{-3}$**.**

As a measure of transpositional distance, Jaro-Winkler Similarity (JWS) was calculated, given by:

$$sim_w = sim_j + (\ell p(1 - sim_j)), \tag{2}$$

where ℓ is the length of common prefix at the start of the string up to a maximum of four characters; p is a constant scaling factor for how much the score is adjusted upwards for having common prefixes, defined as $p = 0.1$; and sim_j is the Jaro Similarity, given by:

$$sim_j = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & \text{otw.} \end{cases} \tag{3}$$

where $|s_i|$ is the length of the string s_i, m is the number of matching characters and t is half the number of transpositions.

We saw the previously identified factor of domain-specific axis weight assignment as a major threat to internal validity of the study. To mitigate said threat, we operationalized the measured confounding factors (e.g., domain knowledge and inquisitive interactions). External validity was largely guaranteed by the use of commodity software and hardware, along with a high gender and cultural diversity among participants. As the procedure required time and concentration from the participants, the amount of available participants was limited (10), which was seen as an unavoidable bottleneck for external validity. Within subjects model of testing was chosen so as to simplify analysis and make testing with less people feasible.

To ensure the reliability of the study, we organized the data collection procedure so that the measurements were taken and reviewed by two people. Objectivity of the study was aimed to be reached through the detailed documentation of the procedure.

5 EXPERIMENT PREPARATION

The experiment was visioned to consist of a short introduction and the controlled evaluation procedure itself.

The purpose of the introduction was set to create a functional model of understanding of the evaluated star plot shapes, even if the participant was not familiar with the domain completely. Without going into specifics of financial benchmarking, we trained the participant, using the example of one company profile, to perform direct ("Growth & Profitability" benchmark) and inverted ("Valuation" benchmark) mapping of the star plot area to a value on a Likert-type scale from 1 to 10. Second goal of the introduction was to identify the presence of any domain knowledge in a participant through a verbal request.

Figure 2: Layout of a company profile: (a) "Growth & Profitability" benchmark; (b) "Valuation" benchmark; (c) the employed approach to profile arrangement.

The evaluation procedure consisted of two stages, starting from sequential evaluation of four company profiles (for both benchmarks) and finishing with an arrangement task, where participants were required to organize the previously evaluated companies in a descending ranking order through visual comparison. The layout and the approach to profile organization could be seen in figure 2.

As a software platform, we used a custom financial visualization framework, developed outside of the scope of the study. In this framework we collect and store, using a MySQL database, actual financial data on publicly traded companies in the United States. Through a RESTful API, implemented with Node.js, the data is served to a programmable front-end (AngularJS), which in turn allows for a variety of graphical data representations.

The structure of the data to be collected in a Google Sheets table was as follows.

Introduction input

Domain knowledge	Binary independent variable denoting presence of any existing domain knowledge (familiarity)

Per-company inputs (x4)

"Growth & Profitability"	Evaluation of a benchmark on a 10-position Likert-type scale, direct mapping
(Inspections)	Binary independent variable denoting presence of inquisitive interactions with interface
"Valuation"	Evaluation of a benchmark on a 10-position Likert-type scale, inverted mapping
(Inspections)	Binary independent variable denoting presence of inquisitive interactions with interface

Between-pairs comparison input

Arrangement order	Encoded in the form of a 4-character string, e.g. **"MHAC"** for an arrangement **M***SFT* → **H***PQ* → **A***APL* → **C***CUR*
Arrangement time	Time of performed arrangement, in seconds

6 EXPERIMENT CONDUCTION

The active study period lasted over the course of three weeks, with participants appointments arranged by email.

The study was conducted at the same location in the LINT-pool of Bauhaus Universität Weimar. As a means of confounding factor retainment, we maintained the same set of input (keyboard and mouse) and output (monitor) devices, together with same environmental conditions (i.e. lighting and temperature).

After meeting a participant, the short introduction was made. The concepts of a publicly traded company, share and financial benchmarking were explained, making sure that participant had a clear understanding of real-world purpose of the visualization tool. We saw it as a necessity to summarize every explanation with an instruction on the use of every benchmark (direct and inversed mapping). The conversation took different amounts of time, until a said degree of understanding was reached by the participant. The flow of the study was revealed to the participant at this point, without revealing the research question.

In the next step of the study, the participant was requested to search for a specific company and perform benchmark evaluation as intended. On multiple occasions a participant would confuse the mapping for the benchmark (direct instead of inverted). These cases were handled with discarding the participants evaluation and repetition of introduction summary. At this point, we expected to have potential incidents, where participants would completely misunderstand the goal of the evaluation, however there were in fact no such cases. The evaluation was repeated for four companies.

After the evaluation was over, the arrangement task was explained and performed. No direct or indirect time pressure was applied to the participant, while the assisting person for the study stated that the arrangement should continue for as long as the participant wants, i.e. until the participant feels confident about the final arrangement.

After the arrangement was over, we would engage in a retrospective conversation with the participant, revealing the research question and asking for feedback on their experience. The expressed ideas and concerns were noted carefully and stored in a link with the participants evaluations.

7 DATA EVALUATION

After the completion of data collection phase of the study, review and evaluation took place. We chose Google Sheets and IBM SPSS as our main tools for storage and statistical analysis of the values respectively.

7.1 Evaluation Errors Calculation

Having direct access to precise axis values presented in star plots we were able to calculate actual area and skewness of the star plot shapes.

Comparison of normalized actual and evaluated on a Likert-type scale areas gave us the absolute error of direct and inverted area evaluation. In the evaluation of this error we chose to keep the sign to be able to interpret data as either positive or negative bias. Over all samples (80) with values ranging from $-0,270$ to $0,217$ the mean value of error was $0,009$, with a 95% confidence interval being $\pm 0,046$.

Table 2: Results of Kolmogorov-Smirnov and Shapiro-Wilk normality tests, given for errors while evaluating plots with different shape asymmetry

Skew $\cdot 10^{-3}$	Kolmogorov-Smirnov Sig.	Shapiro-Wilk Sig.
$3,925$	0,245	0,820
$4,397$	0,370	0,752
$10,287$	0,329	0,655
$18,820$	0,302	0,829
$19,572$	0,300	0,815
$39,754$	0,308	0,756
$61,902$	0,231	0,924
$134,847$	0,289	0,731

Table 3: ANOVA for 8 groups (8 plots) of 10 samples of absolute error in evaluation

Source of Var.	Sum of Sq.	d.f.	Var.	F	p
Between Groups:	0,227	7	0,033	3,515	**0,003**
Within Groups:	0,665	72	0,009		
Total:	0,8925	79			

Error in between-pairs comparison was calculated as a JWS between an ideal arrangement order of company profiles and the one given by participant. The error was corrected by subtracting the value of operationalized factor of domain-specific axis weight assignment (JWS between given arrangement and the one implied in previous evaluations by participant). The resulting error ranged from 0 to $0,113$, with mean value $0,047$ and a 95% confidence interval being $\pm 0,079$.

As a particularly interesting yet unexpected find we uncovered a positive correlation between the arrangement time and the between-pairs comparison error. After a simple t-test we were able to conclude that longer time of arrangement predicts slightly higher error with a significance of $0,00093$.

7.2 Normality testing and ANOVA

To assess the influence of shape asymmetry on the evaluation aptness of the participant we decided to apply analysis of variance (ANOVA). Because ANOVA operates on the presumption of normal distribution in the dataset, we performed Kolmogorov-Smirnov and Shapiro-Wilk normality tests (table 2).

After applying ANOVA to detect difference between the distributions of error values, we found out evidence that the assymetry of the shape does in fact influence the evaluation aptness with a significance of $p = 0,003$ (table 3).

8 INTERPRETATION

In the analysis of collected evaluation errors it has been shown that the value of error is not prohibitively large $(0,9\% \pm 4,6\%)$. As the confidence interval for error spans $9,12\%$, the H_0 can be accepted for the most of financial benchmarking tasks.

Since the mean was leaning towards zero, one can not define the prevalence of positive or negative bias in benchmark evaluation. However, it has been noticed that participants behavior during the study follows in one of the two distinct patterns. The first pattern involves people spending more time to give evaluations and inspecting the meaning of every axis parameter. Participants who had behaved in this way eventually talked about mental models of visualization that they tried to create ("symmetry indicates balanced enterprise", "one must really understand every detail"). The second pattern dictates rapid answering, which paradoxically leads to better performance in between-pairs comparison. Coarse look into associated values provokes an association in the described patterns and the notion of positive and negative bias for first and second pattern respectively. As we did not measure the time of evaluation for every single plot and rather focused on between-pair comparison, we can not make any statements in this regard, however this might be an interesting direction for future research.

With shape asymmetry it has been shown that there is a clear difference somewhere between the distribution of values in evaluations of shapes with different values of Skew. This alone allows us to accept H_1, which is also corroborated by the research done in [3], however it remains an opened question if only certain extents of asymmetry cause significant change in evaluation error.

In future research, one may look into post-hoc analysis, e.g. Tukey's honest significant difference test, while varying the shape asymmetry parametrically. This could yield an insight into the dynamics of evaluation error in relation with the value of Skew.

Finally, the between-pairs comparison results show that the error in underlying ratio comparison is significantly higher than the single plot evaluation error ($4,7\% \pm 7,9\%$), yet still is acceptable in terms of [5]. Since the ratio comparison could also be a matter for subjective weighting (based on the employed investing strategy or the experience of the user), one has to be very hesitant about accepting H_3. We conclude that a future research is required to support H_3, specifically one directed to have higher external validity (larger number of participants, higher involvement of financial professionals).

Based on the objective observations we can conclude that the between-pairs comparison mechanic is optimal for coarse, time-intensive benchmarking. The unanswered question in the study was if the error grows with the amount of companies (pairs of star plots) compared sequentially. The learning effect was also not addressed in this context, which may constitute future work.

9 CONCLUSION

In classic studies of visualization techniques it is typical to isolate the values from their effective meaning, focusing in this way on the immediate properties of the visualization technique.

In this study we embarked on an attempt to produce a research that fuses together the meaning and the value in the context of star plots as a visualization of a small financial benchmarking dataset. Our participants used both direct and indirect mapping of a star plot shape area evaluating underlying questions of whether the company represented in the plot was a well-managed enterprise ("Growth Profitability" benchmark) and how the market was valuating it ("Valuation" benchmark) at the time of data acquisition.

Despite the challenges that appeared from the definition of the research question itself, e.g. confounding factors, we were able to confidently identify that star plot is in fact a great instrument for the tasks of financial benchmarking, which does not increase the error of evaluation more than is considered to be generally accepted in financial predictions and estimations.

In the same time, we identified and proved to be significant a problematic tendency for the evaluations to be affected by the asymmetry of the shape. The extent and the dynamics of said tendency are yet to be explored, however the devised environment and approach makes it a relatively low-hanging fruit.

Understanding that the most interesting use case for the end user of such system is between-pairs comparison, we scrutinized the usefulness of the visualization in this context, discovering that the produced errors in underlying ratio comparison are not prohibitively large.

While interpreting the results, we outlined the most promising directions for future work that we found reasonable, considering limited scope and external validity of the study.

REFERENCES

[1] Tomas Eklund, Barbro Back, Hannu Vanharanta, and Ari Visa. 2003. Using the Self-Organizing Map as a Visualization Tool in Financial Benchmarking. *Information Visualization* 2, 3 (2003), 171–181. https://doi.org/10.1057/palgrave.ivs.9500048 arXiv:https://doi.org/10.1057/palgrave.ivs.9500048

[2] Carla M. D. S. Freitas, Paulo R. G. Luzzardi, Ricardo A. Cava, Marco Winckler, Marcelo S. Pimenta, and Luciana P. Nedel. 2002. On Evaluating Information Visualization Techniques. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '02)*. ACM, New York, NY, USA, 373–374. https://doi.org/10.1145/1556262.1556326

[3] Alexander Klippel, Frank Hardisty, and Chris Weaver. 2009. Star Plots: How Shape Characteristics Influence Classification Tasks. *Cartography and Geographic Information Science* 36, 2 (2009), 149–163. https://doi.org/10.1559/152304009788188808 arXiv:https://doi.org/10.1559/152304009788188808

[4] Catherine Plaisant. 2004. The Challenge of Information Visualization Evaluation. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)*. ACM, New York, NY, USA, 109–116. https://doi.org/10.1145/989863.989880

[5] V. M. Silva and J. F. Coimbra Leite Costa. 2016. Selecting the maximum acceptable error in data minimising financial losses. *Applied Earth Science* 125, 4 (2016), 214–220. https://doi.org/10.1080/03717453.2016.1230972 arXiv:https://doi.org/10.1080/03717453.2016.1230972

[6] W Wu, Q Guo, PF De Aguiar, and DL Massart. 1998. The star plot: an alternative display method for multivariate data in the analysis of food and drugs. *Journal of pharmaceutical and biomedical analysis* 17, 6-7 (1998), 1001–1013.