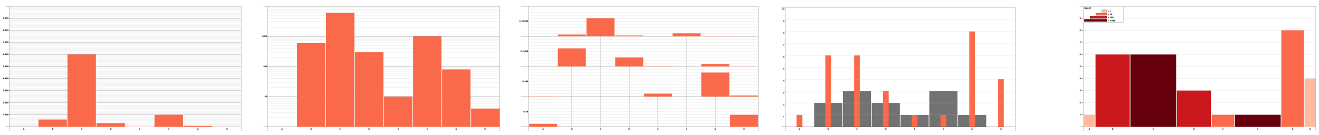
Width-Scale Bar Charts for Data with Large Value Range

M. Höhn[1,2], M. Wunderlich[1,2], K. Ballweg[1], T. von Landesberger[1,2,3]

[1]TU Darmstadt, Germany
[2]HiGHmed Use Case Infection Control Konsortium, Darmstadt, Germany
[3]Karlsruhe Institute of Technology, Germany



(a) *Linear (LIN)* **(b)** *Logarithmic (LOG)* **(c)** *Scale-stack (SSB)* **(d)** *Magnitude marker (OMM)* **(e)** *Width-scale (WSB)*

Figure 1: *A single data set with large value range using four related design and our* width-scale *bar chart.*

Abstract

Data sets with large value range are difficult to visualize with traditional linear bar charts. Usually, a logarithmic scale is used in these cases. However, the logarithmic scale suffers from non-linearity. Recently, scale-stack bar charts and magnitude markers, improve the readability of values. However, they have other disadvantages such as various scales or several objects for visualizing one value. We propose the width-scale bar chart that uses width, height and color to cover a large value range within one linear scale. A quantitative user study shows advantages of our design – especially for reading values.

1. Introduction

Large value ranges appear often in various data. Examples include population count of various countries [Eur20]. Each data set contains several values. Using scientific notation ($v = m \times 10^e$) this example contains values with differences in their exponents between four and eight. We consider this 'data with large value ranges'.

When data with large value ranges are visualized for their exploration, the major tasks are for reading values, comparing values, determining ratios of values, identifying extrema, sorting values or determining trends in the data [TM04, AES05]. The common visualization method is a bar chart with *linear* (LIN) (Fig. 1a) or *logarithmic* (LOG) (Fig. 1b) scale. LOG scale [Tuk77] can display larger value ranges more accurately, but increases difficulties in reading exact values due to its non-linearity [HSBW13]. Therefore, recently two special approaches, i.e., *scale-stack bar charts* (SSB) (Fig. 1c) and *order of magnitude markers* (OMM) (Fig. 1d), have been proposed. They improve the readability of values, but require several scales (SSB) or multiple encoding of values (OMM). Thus, reading values is potentially more difficult.

We present a novel approach to visualize data with large value ranges: the *width-scale bar chart* (WSB). With our new technique, bars can be arranged into one single scale (in contrast to SSB) by one bar object (in contrast to OMM). We compared our approach with LIN and LOG bar charts as well as the SSB and the OMM designs in an empirical user study. The results show, that our new design performs significantly better for reading values than all other designs and has comparable performance to the best design for determining ratios, sorting and find extrema.

2. Related Work

In addition to LIN and LOG design, Isenberg et al. [IBDF11] presented a technique where the x-axis can be locally transformed to show adequately dense data. Their evaluation showed that transformation of y-axis performs better than transformation of x-axis. One option is to use cut-off bars or scale break [CM86]. Recent techniques use the normalized scientific notation $m \times 10^e$ where $1 \leq m < 10$ and $e \in \mathbb{Z}$.

Scale-stack bar charts [HSBW13] use several scales to represent the values - one scale for every distinct exponent e. Within each scale the mantissa m is represented linearly (see Fig. 1c).

Order of magnitude markers [BDJ14] use different elements to visualize the mantissa m and the exponent e. The mantissa is displayed as a thin colored bar with height of m in front of a thicker gray bar represent the exponent e. Both elements are displayed using a linear scale from 0 to 10.

3. Visualization Method

Our new visual approach is the so-called *width-scale bar chart* (Fig. 2). This design is inspired by the scientific notation of num-

DOI: 10.2312/evs.20201056

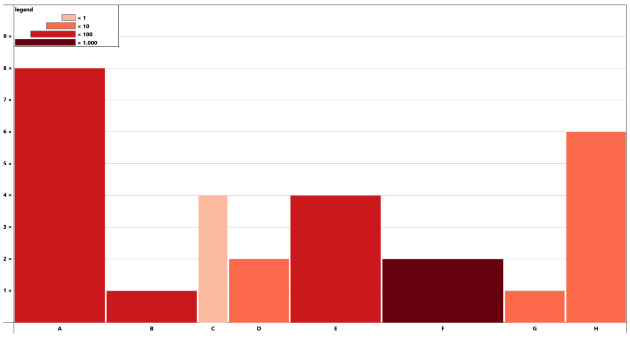


Figure 2: Width-scale bar chart. *Example data as used in the study.*

bers. Similar to the recent approaches [HSBW13, BDJ14], we split each value v into two parts to gain a tuple of mantissa m and exponent e so that $v = m \times 10^e$. As visual variables, we choose *width*, *height* and *color* due to the importance ranking by Bertin [Ber83, Car03]. The mantissa is linearly mapped to the height. The exponent is mapped to both the width of the bar and the color to facilitate readability. This assumption is based on [Rot16, HBF08], who propose to conjunct two visual variables to strengthen the graphic encoding of one attribute. Furthermore, the perception of the width is supported by the perceived brightness of the color [Few09, Bre15], i.e. a large value is darker than a small one. We chose an orange-yellow palette as it takes advantage of the fact that the human visual system has maximum sensitivity to luminance changes for the orange-yellow hue [LH92]. Additionaly, it is suitable for color blind people [FFM+13]. With this setting all values can be displayed with *one* scale from 0 to 10.

4. Evaluation

LIN and LOG designs are common designs for bar charts to visualize data [HSBW13] whereas SSB [HSBW13] and OMM [BDJ14] specifically address the visualization of data with large value range in bar charts. Therefore, we decided to make all these four and our new design – the WSB– subjects of our user study.

Data To cover the same range as in comparable studies [HSBW13, BDJ14], all data were created as $m \times 10^e$ with $1 \leq m < 10$ and $0 < e \leq 4$. The values within one data set were equally distributed within m and e so there were two values for every exponent in the set. In total we created 24 data sets.

Tasks In accordance with comparable studies concerning bar charts and data with large value range [HSBW13, BDJ14], we use these four tasks:

- *Task Value*: read the value
- *Task Sort*: sort all values in ascending order
- *Task Ratio*: determine the ratio of two values
- *Task Trend*: identify the trend in the data from the choice of linear, a logarithmic or an exponential, or none

Furthermore, these tasks are important tasks for exploration in visualization [TM04, AES05].

Task *Sort* is an extention of the exrtema taks used in previous studies [HSBW13, BDJ14], so the participants have to sort the whole data set. This has the advantage that not only the detection of smallest *or* largest value can be evaluated, but also creation of a sorted sequence. *Ratio* task differs from *Value* task. It asks for a multiplier of values rather than the value itself. Thus, the correct answer can be given in some cases without reading the values. For example, the ratio of 10 between 600 and 6.000 can be seen in OMM solely by difference in gray bar length (see Fig. 1d).

Experimental Setting A total of 136 participants (97 male, 29 female and 10 prefer not to say) took part in the study. The age distribution was from under 20 up to over 60, but the majority of the participants were between 20 and 30 years old (64%). We filtered out 21 participants, who did not answer the golden standard question correctly that checked for attention in the online study. In total, the distribution of the participants was as follows - filtered ones in parentheses: LIN 26 (2), LOG 25 (3), SSB 26 (4), OMM 22 (3), WSB 37 (9). A power analysis by Cohen [Coh13], performed with the R package [Hor20] *pwr*, supports the significance of these group sizes.

Procedure The study was set up with *SoSci Survey* [SoS03]. There was one task per page and the participants had to click the next button manually, so the given answer and the used time per task and participant could be stored. The visualizations had a resolution of 1000 × 520 px. Because the study was online, we are not able to control the display size, the brightness and contrast but we recommended a 13" or larger computer screen to take part in the study.

To achieve a processing time of 10–15 minutes the study itself was conducted as a between-subject study [CGK12]. This decision was based on our pilot study. Each participant had to solve each task for six times with different data sets so overall there were $1[design] \times 4[tasks] \times 6[repetitions] = 24$ tasks per participant.

The participants were recruited by advertising in lectures and online newsletters and word of mouth. In total, the recruiting phase was between Feb 2018 and Dec 2018. The participants were randomly assigned to one of the designs. We had a training phase introducing the study interface, explaining the visual design, to familiarize with the tasks. Each kind of task had to be solve once with direct feedback on the correctness. After all tasks the participants were asked to give feedback. The whole study is presented in supplementary material.

4.1. Analysis

For comparison with previous studies, we measure error (inaccuracy) and response times. These were logged in the system.

Log-Error In extension to the comparable studies of Hlawatsch et al. [HSBW13] and Borgo et al. [BDJ14], we decided to use the logarithmic error, as it indicates whether, if the given answer is wrong, it is too high or too low and if there is a mantissa error, an exponent error or both.
Hlawatsch et al. [HSBW13] use $e = |1 - \frac{user}{real}|$ as error. This tells us if a given answer is near to the correct answer or not. The drawback of this method is, that $e_1 = |1 - 1/10|$ and $e_2 = |1 - 100/1000|$ result in the same error value ($e_1 = e_2 = 0.9$) whereas the absolute error is quite different (9 resp. 900).

Borgo et al. [BDJ14] gave a 20% error tolerance to the correct value to decide on the answer's correctness. This is done because the tasks *Value* and *Ratio* involved estimation of an unknown value and therefore answers contain uncertainty. The drawback is an inaccurate error value, because an answer is only either correct or not.

We define the error $e_{log} = \log(\frac{\textit{respone value}}{\textit{encoded value}})$, where *response* is the answer given by the participant and *encoded* is the correct answer, and call it *Log-Error*. This *Log-Error* is used to evaluate task *Value* and task *Ratio*. For task *Sort* and task *Trend* we use a binary error with *true* and *false* value resulting in an accuracy for these tasks. Subsequently, the accuracy value was transformed by calculating $1 - \textit{accuracy}$ to get an error value and consistently maintain that smaller values represent better results.

4.2. Results

To perform our analysis we use a three-stage significant test for each task. Since we could not assume that the data is normally distributed in general, we first perfom a Shapiro-Wilk test on the error values. Due to the result of the Shapiro-Wilk test (not normally distributed for each individual task) as second stage of the analysis, we used a Kruskal-Wallis test to determine statistical significance between the design. On the third stage as post-hoc analysis we used a Wilcoxon signed-rank test for pairwise comparison of designs for tasks for which significance was found. All tests were performed with a standard significance level $\alpha = 0.05$, which was adjusted using a Bonferroni correction to $\alpha = 0.005$ for the post-hoc tests. Fig. 3 shows the mean error rates and mean response time for all tasks as well as p-values with pairwise significance.

Value Our WSB design performs best for both time and error. The Kruskal-Wallis test showed a significant mean effect in both error ($\chi^2 = 82.24, p < 2.2\text{e}{-16} \ll 0.05$) and time ($\chi^2 = 142.24, p < 2.2\text{e}{-16} \ll 0.05$).
Error values: The post-hoc Wilcoxon signed-rank test shows that WSB bar charts ($\mu = 0.02$) perform significantly better than all other designs. The LOG design ($\mu = 0.09$) performs second best. This corresponds to Hlawatsch et al. [HSBW13], but is in contrast to Borgo et al. [BDJ14], where the LOG design performed worst. This can be due to the differences in error measurement scale between Hlawatsch and Borgo. The SSB ($\mu = 0.17$) and LIN ($\mu = 0.18$) designs perform similar and better than the OMM design ($\mu = 0.19$).
Mean response time: the WSB design has the fastest mean response time ($\mu = 9.68s$), followed by OMM design ($\mu = 10.18s$), LIN ($\mu = 12.46s$) and SSB ($\mu = 14.89s$). The LOG design takes the longest ($\mu = 18.28s$) and also has the largest standard deviation.

Sort The SSB design has very low errors similar to WSB design. The Kruskal-Wallis test showed a significant main effect in both error ($\chi^2 = 31.53, p < 2.4\text{e}{-06} \ll 0.05$) and time ($\chi^2 = 52.12, p < 1.3\text{e}{-10} \ll 0.05$).
Error values: The post-hoc Wilcoxon signed-rank test shows no significance between the two best designs: WSB and SSB. LOG design has larger error ($\mu = 0.03$), however not significantly. LIN ($\mu = 0.11$) and OMM ($\mu = 0.38$) designs perform significantly worse.
Mean respone time: Our WSB design ($\mu = 27.38s$) is significantly faster than all other designs. The LOG design ($\mu = 30.61s$) performs second best similar to the LIN design ($\mu = 30.92s$) but with significance. Although the SSB design has low value error, it has the second longest response time ($\mu = 33.50s$). The OMM design ($\mu = 38.30s$) has the highest times.

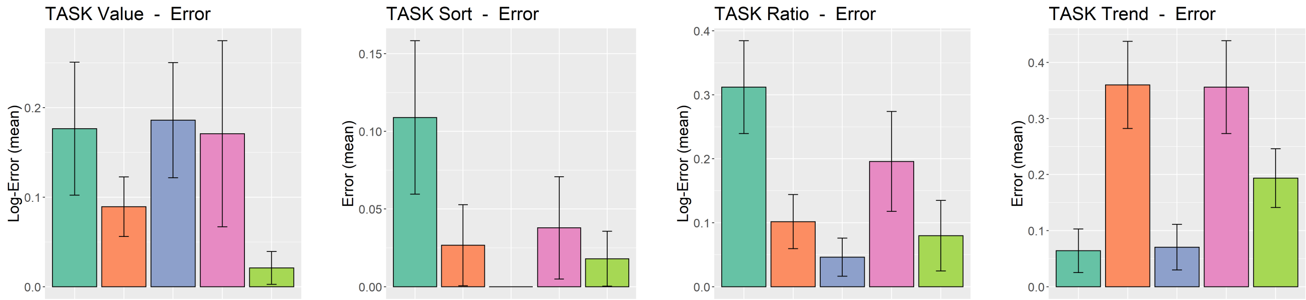
Ratio The SSB design performs best for this task similar to the WSB design. The Kruskal-Wallis test showed a significant main effect in both error ($\chi^2 = 199.57, p < 2.2\text{e}{-16} \ll 0.05$) and time ($\chi^2 = 81.76, p < 2.2\text{e}{-16} \ll 0.05$).
Error values: The post-hoc Wilcoxon signed-rank test shows that SSB and WSB have lowest error rates ($\mu = 0.05$ resp. 0.08) without significant differences. The LOG ($\mu = 0.10$) and OMM ($\mu = 0.20$) designs perfom significantly worse, but better than the LIN design ($\mu = 0.31$).
Mean response time: The LIN design ($\mu = 22.46s$) has best response time while having the worst error value. The WSB ($\mu = 26.89s$) and the OMM design ($\mu = 27.63s$) have a similar response time but with significant difference. The SSB ($\mu = 34.28s$) and the LOG design ($\mu = 34.65s$) perform worst.

Trend The LIN and the SSB designs perform best for this task, whereas the LOG and the OMM perfom worst. The WSB design has average performance. The Kruskal-Wallis test showed a significant main effect in both error ($\chi^2 = 199.57, p < 2.2\text{e}{-16} \ll 0.05$) and time ($\chi^2 = 81.76, p < 2.2\text{e}{-16} \ll 0.05$).
Error values: The post-hoc Wilcoxon signed-rank test shows that the LIN ($\mu = 0.06$) and the SSB ($\mu = 0.07$) design perfom best with significance to the WSB design ($\mu = 0.19$). The OMM ($\mu = 0.36$) and the LOG design ($\mu = 0.36$) preform significantly worse.
Mean response time: The LIN design ($\mu = 7.99s$) has the best response time followed by the SSB design ($\mu = 9.78s$). The WSB design ($\mu = 15.32s$) requires twice as much time as the LIN design whereas the LOG ($\mu = 20.95s$) resp. the OMM ($\mu = 21.12s$) design require three times as much as time as the LIN design.

4.3. Extended Analysis and Results

For the extended analysis we defined error classes and investigate the sign of the errors. This is possible due to the fine-graidend detail the *Log-Error* provides.
Error types: In addition to error size, we analyze the type of error occurring for large values: whether the error was only in mantissa, only in exponent or in both components. Table 1 shows that LIN and LOG designs have a larger amount of mantissa errors than SSB, OMM and WSB design. In the latter, the exponent error prevails. Interestingly, SSB design has only exponent errors, which indicated that participants had problems with the subdivision of the different scales.
Sign of errors The sign of errors shows, whether participants tend to over- or underestimate the correct value. Table 2 shows underestimation of values for LOG and overestimation for SSB design. There is no tendency for LIN, OMM and WSB, but more importantly the error rate is much lower for our design.

Free feedback from the participants confirms the numeric analysis. Four out of eight participants mentioned "problems to read



TASK Value - Error

p	LIN	LOG	SSB	OMM	WSB
LIN	◇	.008	.002	.000	.000
LOG	.001	◇	.639	.014	.000
SSB	.003	.526	◇	.114	.000
OMM	.000	.000	.000	◇	.001
WSB	.000	.000	.000	.650	◇

TASK Sort - Error

p	LIN	LOG	SSB	OMM	WSB
LIN	◇	.005	.000	.024	.000
LOG	.736	◇	.041	.595	.575
SSB	.021	.037	◇	.014	.093
OMM	.136	.230	.522	◇	.252
WSB	.000	.000	.000	.000	◇

TASK Ratio - Error

p	LIN	LOG	SSB	OMM	WSB
LIN	◇	.000	.000	.000	.000
LOG	.000	◇	.000	.814	.000
SSB	.000	.153	◇	.001	.742
OMM	.000	.009	.000	◇	.000
WSB	.000	.000	.000	.482	◇

TASK Trend - Error

p	LIN	LOG	SSB	OMM	WSB
LIN	◇	.000	.823	.000	.000
LOG	.000	◇	.000	.946	.000
SSB	.009	.000	◇	.001	.001
OMM	.000	.577	.000	◇	.001
WSB	.000	.378	.000	.120	◇

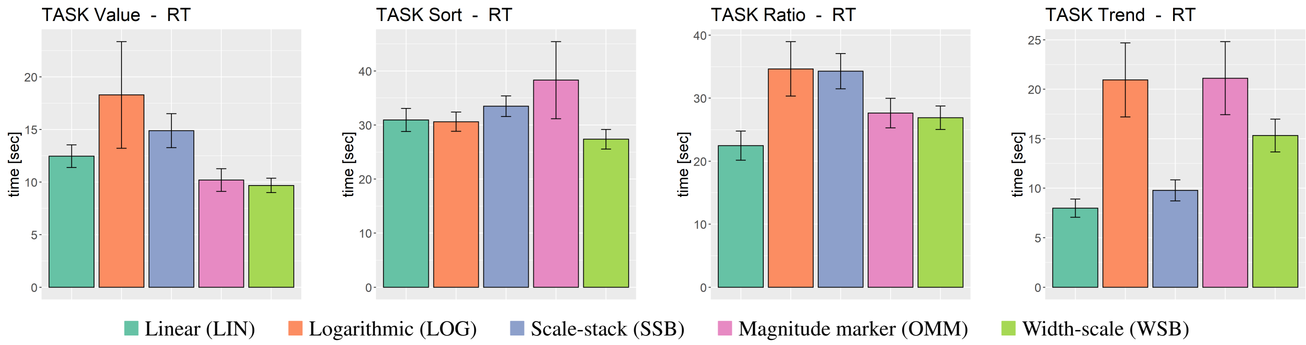


Figure 3: *The bar charts show the mean error rate (top) and the mean response times (bottom) for LIN, LOG, SSB, OMM and WSB for all tasks. Error bars show 95% confidence intervals, calculated with the R-Project [Hor20]. The color coding is done by using the RColorBrewer package [NB14]. Lower values are better. The tables show p-values. Pairwise significance is underlined. Upper triangle shows error, lower triangle response time.*

error type	lin	log	ssb	omm	wsb
Mantissa	35.29	22.67	0.00	1.54	**0.90**
Exponent	2.21	0.00	17.95	9.23	**1.35**
Both	2.94	1.33	0.00	0.00	**0.00**

Table 1: *Distribution of error types [%] for read value task.*

	lin	log	ssb	oom	wsb
$e > 0$	17.65	4.67	17.95	4.62	**1.35**
$e < 0$	22.79	19.33	0.00	6.15	**0.90**

Table 2: *Sign of error [%] for read value task.*

small values” concerning the LIN design. Several participants mentioned problems in the trend task with OMM, WSB and LOG design, e.g., “I had difficulties estimating trends on logarithmic scale”. The feedback indicated that the idea of the SSB design was liked. However, “it is really complicated to compare values”. WSB design is “interesting” and “sorting and reading values is very easy with this design. Even with strongly varying orders of magnitude” and “gets easy to use rather quickly”. Three out of 18 participants wondered about the double encoding by color and width. For all designs, participants suggested to show bars sorted.

In sum, this is our design ranking:

- *Task Value:* **WSB** $\prec$ LOG $\prec$ OMM $\preceq$ LIN $\prec$ SSB
- *Task Sort:* SSB $\preceq$ **WSB** $\preceq$ LOG $\preceq$ OMM $\prec$ LIN
- *Task Ratio:* SSB $\preceq$ **WSB** $\prec$ LOG $\preceq$ OMM $\prec$ LIN
- *Task Trend:* LIN $\preceq$ SSB $\prec$ **WSB** $\prec$ OMM $\preceq$ LOG

It means, our results are similar to the studies conducted during the development of the SSB [HSBW13] and the OMM [BDJ14].

5. Conclusion and Future Work

We presented a new design approach to visualize data with large value range in bar charts. The empirical study with a comparable methodology to previous studies, has shown that our WSB design improves the accuracy and time of value reading tasks. It has no significant difference to the best performing designs for ratio and sorting/extrema tasks and shows average performance for trend analysis. Our design can be used for data in economics, e.g., gross domestic product, or in medicine, e.g., number of infected persons across countries. Our design can easily be extended to show data with positive and negative values by using color, e.g., blue and red. In the future, we will address the participant’s feedback and investigate the influence of double encoding and value ordering in the visual design. Furthermore, we have to investigate the scalability of our WSB, both for the data sets, regarding the number of data and the value range, and for displaying the WSB on smaller screens, such as mobiles or smartwatches.

Acknowledgments The authors would like to thank all study participants. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the research and funding concepts of the Medical Informatics Initiative (01ZZ1802B/HiGHmed).

References

[AES05] Robert Amar, James Eagan, and John Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pages 111–117. IEEE, 2005. 1, 2

[BDJ14] R. Borgo, J. Dearden, and M. W. Jones. Order of magnitude markers: An empirical study on large magnitude number detection. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2261–2270, Dec 2014. 1, 2, 3, 4

[Ber83] Jacques Bertin. Semiology of graphics (wj berg, trans.). *Madison, WI: The University of Wisconsin Press.(Original work published 1967)*, 1983. 2

[Bre15] Cynthia A. Brewer. *Designing Better Maps: A Guide for GIS Users*. Esri Press, 2nd edition, 2015. 2

[Car03] M Sheelagh T Carpendale. Considering visual variables as a basis for information visualisation. 2003. 2

[CGK12] Gary Charness, Uri Gneezy, and Michael A Kuhn. Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1):1–8, 2012. 2

[CM86] William S. Cleveland and Robert McGill. An experiment in graphical perception. *International Journal of Man-Machine Studies*, 25(5):491–500, 1986. 1

[Coh13] Jacob Cohen. *Statistical power analysis for the behavioral sciences*. Academic press, 2013. 2

[Eur20] Eurostat. Population on 1st January by age, sex and type of projection. https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=proj_18np&lang=en, 2020. Accessed: 2020-02-17. 1

[Few09] Stephen Few. Introduction to geographical data visualization. *Visual Business Intelligence Newsletter*, March/April 2009. 2

[FFM+13] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 3237–3246, 2013. 2

[HBF08] Nathalie Henry, Anastasia Bezerianos, and Jean-Daniel Fekete. Improving the readability of clustered social networks using node duplication. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1317–1324, 2008. 2

[Hor20] Kurt Hornik. R FAQ, 2020. 2, 4

[HSBW13] Marcel Hlawatsch, Filip Sadlo, Michael Burch, and Daniel Weiskopf. Scale-stack bar charts. *Computer Graphics Forum*, 32(3.2):181–190, 6 2013. 1, 2, 3, 4

[IBDF11] Petra Isenberg, Anastasia Bezerianos, Pierre Dragicevic, and Jean-Daniel Fekete. A study on dual-scale data charts. *IEEE Trans Vis Comput Graph*, 12(2469):78, 2011. 1

[LH92] Haim Levkowitz and Gabor T Herman. Color scales for image data. *IEEE Computer Graphics and Applications*, (1):72–80, 1992. 2

[NB14] Erich Neuwirth and R Color Brewer. Colorbrewer palettes. *R package version*, pages 1–1, 2014. 4

[Rot16] Robert E Roth. Visual variables. *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology*, pages 1–11, 2016. 2

[SoS03] SoSci Survey GmbH. Sosci survey - die Lösung für eine professionelle Onlinebefragung. https://www.soscisurvey.de/, 2003. Accessed: 2020-02-20. 2

[TM04] Melanie Tory and Torsten Moller. Rethinking visualization: A high-level taxonomy. In *IEEE Symposium on Information Visualization*, pages 151–158. IEEE, 2004. 1, 2

[Tuk77] John W Tukey. *Exploratory data analysis*, volume 2. Reading, Mass., 1977. 1