Off the Radar: Comparative Evaluation of Radial Visualization Solutions for Composite Indicators

Yael Albo, Joel Lanir, Peter Bak, *Member, IEEE*, and Sheizaf Rafaeli

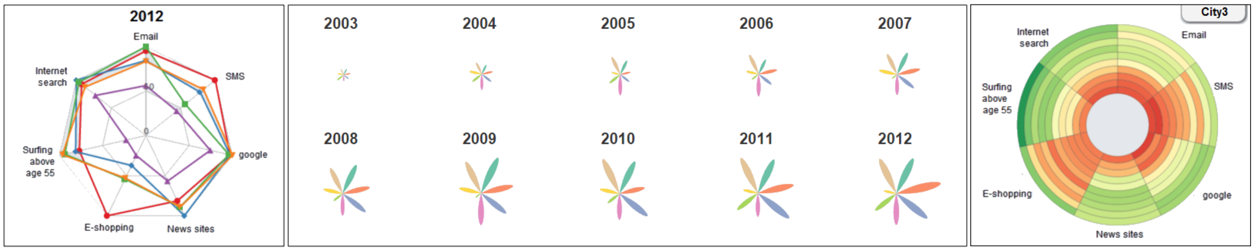


Fig. 1: Three radial solutions for composite indicator visualizations compared empirically for users' performance and preferences.

Abstract—A composite indicator (CI) is a measuring and benchmark tool used to capture multi-dimensional concepts, such as Information and Communication Technology (ICT) usage. Individual indicators are selected and combined to reflect a phenomena being measured. Visualization of a composite indicator is recommended as a tool to enable interested stakeholders, as well as the public audience, to better understand the indicator components and evolution over time. However, existing CI visualizations introduce a variety of solutions and there is a lack in CI's visualization guidelines. Radial visualizations are popular among these solutions because of CI's inherent multi-dimensionality. Although in dispute, Radar-charts are often used for CI presentation. However, no empirical evidence on Radar's effectiveness and efficiency for common CI tasks is available. In this paper, we aim to fill this gap by reporting on a controlled experiment that compares the Radar chart technique with two other radial visualization methods: Flower-charts as used in the well-known OECD Betterlife index, and Circle-charts which could be adopted for this purpose. Examples of these charts in the current context are shown in Figure 1. We evaluated these charts, showing the same data with each of the mentioned techniques applying small multiple views for different dimensions of the data. We compared users' performance and preference empirically under a formal task-taxonomy. Results indicate that the Radar chart was the least effective and least liked, while performance of the two other options were mixed and dependent on the task. Results also showed strong preference of participants toward the Flower chart. Summarizing our results, we provide specific design guidelines for composite indicator visualization.

Index Terms—Visualization evaluation, radial layout design, composite indicator visualization, experiment

1 Introduction

A famous quotation from Lord Kelvin says: "If you cannot measure it, you cannot improve it" [26]. Measuring and benchmarking (M&B) are necessary for understanding an entity's (e.g., a country, a region, a company) position and identifying growth opportunities. M&B is a continuous process of monitoring, benchmarking and improving [28].

A composite index (CI), as an M&B method, is a measure derived from a series of observed facts that can reveal relative positions on a regional basis in a given field (i.e. environment, economy, society or technological development). Multi-dimensional complex concepts are summarized into a single indicator. Sub-indicators construct a CI in a hierarchal way, as a weighted variables tree. When evaluated at regular temporal intervals, an indicator can point out the direction of change in different units over time [14]. A Composite Indicator might be useful in setting policy priorities and in benchmarking or monitoring growing year after year [9]. Simplicity and ease of interpretation is one of the advantages of CIs, as it produces a "bottom line" [31]. On the other hand, there is a need to support deeper insight acquisition (e.g., what is the best practice one should learn from), in order to improve performance. Domain experts, as well as non-experts, might be interested in making deeper sense of CIs.

Composite Indicators as a measurement and benchmarking tool are widely used in many areas including the social, political, management and technological domains. Communicating indicators can efficiently help citizens understand transparent data, and is thus perceived as vital for democracy [15]. For example, government authorities can advance e-government in order to improve their services to the citizens [22]. Educational authorities are interested in multi-dimensional measurements of institutions starting from the level of municipal schools and up to university level in order to compare educational and research achievements. In this paper, we specifically use a CI called Information and Communication Technology (ICT) index, aiming to measure the "information society" and pointing out digital divides. It is used by countries and municipalities to examine their development as information societies, and to compare performance in the area of ICT. The questions related to ICT indicators are general CI questions that might be linked with growth potential, as well as with gap risks: Which indicators and variables should be improved? Does a divide exist? What is the gap's size and trend - is it getting wider or narrower? What is the rate of change? [32, 10]

A useful way to examine these questions is by visualizing the components of the composite indicator. Indeed, it has been acknowledged that a clear visualization of a composite indicator is highly needed and recommended [14, 28]. Visualization of the benchmarking scores on a variety of measures and aspects can enable the evaluation of the effects

- *Yael Albo is with University of Haifa / Israel. E-mail: alboyael@gmail.com.*
- *Joel Lanir is with University of Haifa / Israel, E-Mail: ylanir@is.haifa.ac.il.*
- *Peter Bak is with IBM Research Haifa Lab, Haifa / Israel, E-mail: peter.bak@il.ibm.com.*
- *Sheizaf Rafaeli is with University of Haifa / Israel, E-mail: sheizaf@rafaeli.net.*

Manuscript received 31 Mar. 2015; accepted 1 Aug. 2015; date of publication 20 Aug. 2015; date of current version 25 Oct. 2015. For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.
Digital Object Identifier no. 10.1109/TVCG.2015.2467322

of policy-makers' actions, help the understanding of current problems and may help in selecting improvement measures [28]. Indeed, many CI websites and reports use visualizations to present the information to the public [1, 4, 7]. However, these websites use a variety of visualizations and it is not clear which one is best to use for different CIs and different tasks.

In this paper, we evaluate three radial visualization solutions for composite indicators with time-oriented data: the popular Radar chart, used in many existing CIs, and two possible alternatives. We explore the differences between these visualizations in terms of performance and subjective preferences and highlight the advantages and disadvantages of each visualization in various CI tasks. Summarizing our results, we present general guidelines for designing CI visualizations.

2 Composite Indicator data properties

CIs consists of a set of multidimensional properties and categories. A temporal dimension also exists, showing the evolution of the CI dimensions over time. The following list summarizes the data properties of a typical composite indicator:

Dimensions: Composite Indicators aim to represent a multidimensional reality. Since the phenomenon complexity is simplified by the CI's single value, unfolding dimensions is required for sense-making and taking actions as a result.

Items: A Composite Indicator is almost always used in a comparative manner. Many tasks have to do with comparing one's own score to some sort of norm or to other items. Common items can be cities, countries, schools, etc.

Time: Measurement and benchmarking is a continuous process. Supporting time-oriented data and tasks are one of the core challenges of Composite Indicator visualization. CI measurement is expressed in a manner that growth of values indicates improvement, and CI visualization almost always show growth over time.

In addition, a CI is often structured as a variable tree representing a hierarchy of indexes and subindexes, each composed of a different amount of dimensions. Since usually, each level is visualized separately (often using interaction to move between levels), in this paper we focus on a single-level visualization.

3 Composite Indicator Visualization Methods

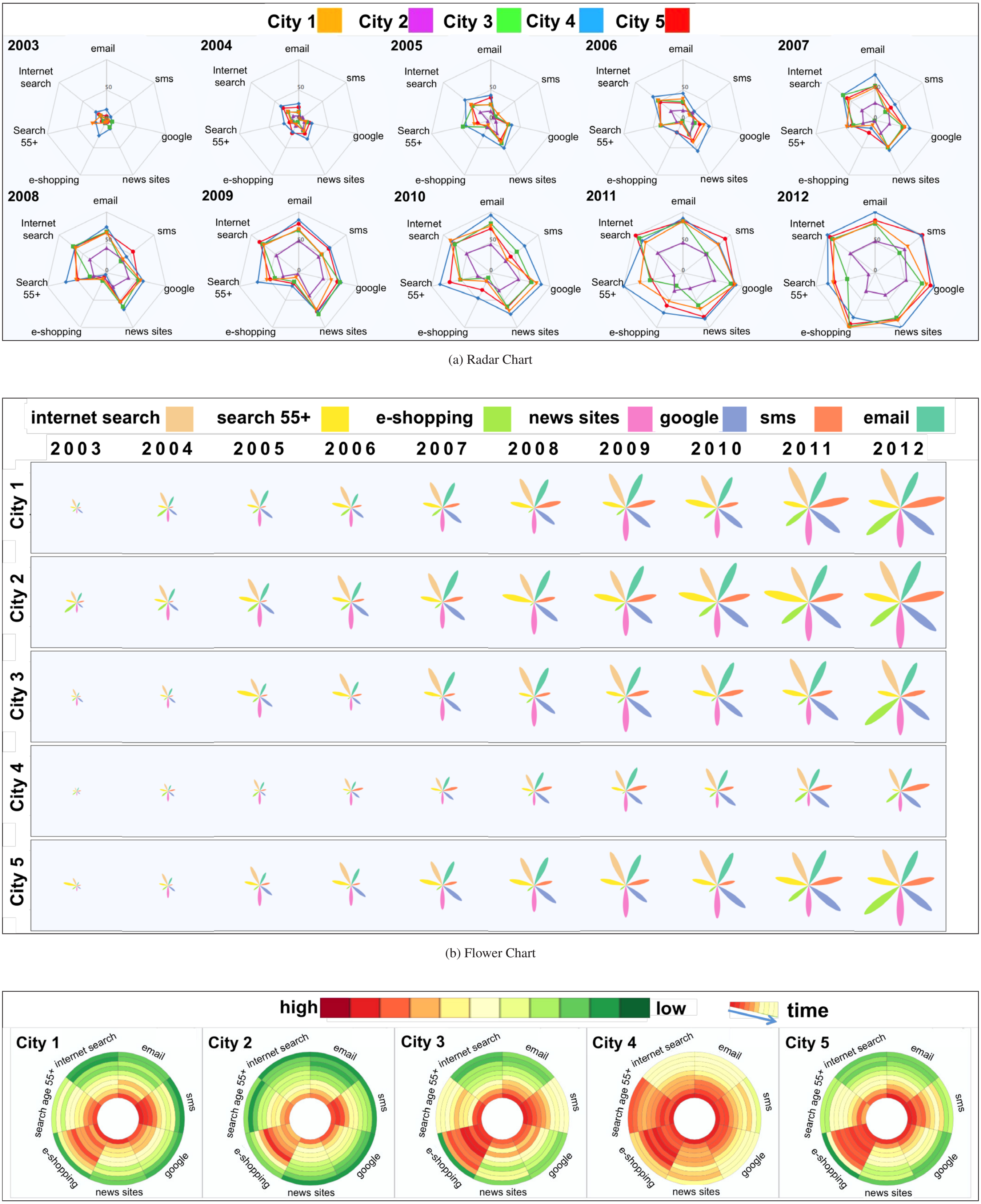
Composite indicators are often visualized with the aim of making the information accessible to the general public or to various stakeholders. Gapminder Trendalizer [2] is a well-known method that animates trends over time using a dynamic bubble chart and was used to show gaps and differences between countries among other things. This method was employed as a public data explorer tool in CIs like the UN Human Development index and ITU ICT Development index [3]. Robertson et al. [30] have evaluated its use, comparing it with two other methods. Their results show that while faster for presentation, the tool was less effective for analysis tasks. However, the dynamic bubble visualization used in Gapminder is not scalable, and can only work with up to four dimensions. Also, the focus of Gapminder is on comparing two dimensions, those being mapped to the x and y coordinates of the scatterplot. In a common CI, there are often 3-12 dimensions [14]. Thus, a multidimensional visualization is needed in order to equally convey all dimensions. When examining existing CI visualizations, we see that radial visualizations are very common in CI websites and reports [6, 4, 7, 1, 29, 18]. This is probably because of their ability to equally show multi-dimensional data in a single glyph [17]. In the CIs we examined, different dimensions were mapped to the radial sectors of the radial visualization. The use of radial visualizations can be related to their aesthetic appeal, compact layout, and ability to put selectable data within easy reach of the user (Fitts, 1954 in [17]). Easy focus on a particular data dimension is another advantage of radial presentation to be leveraged [16].

Among radial visualizations for CIs, Radar charts are undoubtedly the most commonly used. For example, Thomson Reuters use Radar charts to visualize and enable comparing academic institutions on several dimensions [6] and to create a university institutional profile [5]. In [33], a Radar chart was examined to measure the quality of clinical outcomes and services of health care institutions. In the area of ICT, the Global Information Technology report of the World Economic Forum uses a Radar chart to show performance of various countries in the Network Readiness Index [18]. Similarly, the Web Index, a popular index that ranks countries according to their progress and use of the Web, also used Radar charts for comparing countries [7]. An example of the Radar chart showing a CI is shown in Figure 2a. While popular, the use of Radar charts are in dispute. Criticism of Radar visualizations in the context of CI have emerged lately [19]. Because each item is represented by a polygon, the area of the polygon depends on the placement of axis dimensions, and thus may be inappropriate for comparing items. Furthermore, a Radar chart may turn out to be too cluttered if showing multiple items. Another problem with Radar charts is that we tend to prefer polygons with symmetrical shapes. Yet, we almost never care about something being more symmetrical than we care about the magnitude of ratings [20]. Thus, several questions arise in the context of using Radar charts for CI visualizations. Are they suitable for general CI tasks? for which CI task are they better and for which are they worse? Are there alternative visualizations that may be more effective?

The popular OECD Better Life index is one such alternative that uses a different kind of radial visualization. It uses a Flower-based visualization to show an 11-dimension CI that measures key factors (such as education, housing, etc.) that contribute toward well-being in countries [1]. Though this visualization was developed for a concrete purpose, Bertini suggests to use the Flowers technique "every time you want to represent multiple dimensions of a series of objects at the same time" [11]. Few pointed out this technique as "immediately engaging aesthetically and informatively [21]. In this radial-based Flower visualization (see Figure 2b), each flower glyph represents an item (i.e., a city). Each petal of the flower corresponds to a different dimension, with every dimension uniquely colored. The length of the petal is mapped to the value of that dimension. Multiple items are represented by multiple glyphs as small multiples. While not existing in the original website, we mapped the time dimensions as small multiples, with items shown as rows and time shown as columns.

As another alternative, we use an adaptation of Circle charts. We use a concentric Radial Space Filling (RSF) presentation based on Keim's circle view metaphor [25] (see Figure 2c). The main goal of the Circle View design is to represent temporal and multi-dimensional attributes in order to identify patterns, exceptions and similarities in the data. This approach was originally designed for fast changing visualizations caused by time related data streams [25]. We use the color coding traffic model (green for high values, red for low values) that was shown to be able to simplify M&B complexity [28], and use green to yellow and yellow to red, data mapped to 9 colors taken from ColorBrewer diverging colormap [23]. To keep consistent with the previous visualizations, the dimensions are mapped to the segments of the circle. However, the temporal dimension was mapped as the radius of the interleaving circles, as suggested in [25]. Also in our context, this seems to be more intuitive, than encoding this visual property to any other data property, such as the items or the dimensions. Mapping time to rings from inside out (internal rings represents the first year, outer ring represents latest year) results in a larger area dedicated to the current year. This focus on more recent performances is intuitive to the CI tool, since it highlights present status. Items were mapped as small multiples, resulting in a set of different glyphs of this type of chart. A possible advantage of this representation is that it may be easy to see changes and trends over time.

For example, a clear improvement over time can be seen in the "internet search" dimension in city1 in Figure 2c, from red in the inner circle (first year) to green in the outer circle (last year). Table 1 summarizes the visual mapping of the data properties to the visual properties of the three different charts. The way the visual mapping of the time, items and dimensions are designed, all resulting views consist of a set of small multiple glyphs. In these views, the number of ra-



(a) Radar Chart

(b) Flower Chart

(c) Circle Chart

Fig. 2: Three chart types encode CI data using different visual encoding for the data attributes. We compared users' performance and preference empirically for a number of elementary and synoptic tasks.

dial glyphs was fixed, and the only varying property was the number of segments. As a result, interaction such as scrolling, zooming or paging is not required - which could have biased or complicated our empirical study. Color has been used with great concern, following the advise and standards defined by real world usage. Visual choices were derived from our aim to compare the popular Radar charts with other radial solutions that also code dimensions to circle segments (see Table 1). The motivation was the need to have a satisfactory solution for varying number of dimensions (as expected in CIs' visualization), which led us to radial solutions on the first place. Other properties' coding was chosen in correspondence to this initial choice. It is important to emphasize that the study does not compare glyph vs. glyph (i.e. Radar chart vs. Flower glyph vs. Circle glyph). but rather about comparing visualization techniques as a whole (i.e. small multiples structure of radial icons). We believe that with this visual encoding we have created a basis for empirical validity and fair, unbiased comparability of visualization techniques.

Data property	**Radar Chart**	**Flower Chart**	**Circle Chart**
Time	Small Multiples	Small Multiples (columns)	Rings from Center
Items	Data Lines (color)	Small Multiples (rows)	Small Multiples
Dimensions	Axes	Segments (color)	Segments
Values	Distance from Center	Distance from Center	Color of the rings

Table 1: Summary of visual encoding of the data properties to different properties of the selected charts. Design decisions were made under the constrains of experimental validity.

4 Experimental Design

We performed a study to test the effectiveness of the three radial visualizations. We used a 3 (Vis: Radars vs. Flowers vs. Circles) x 2 (Dimensions: 3 vs. 7) x 10 (tasks) mixed study design. Vis and Tasks were within-subject variables while Dimensions was a between-subject variable. Since number of Dimensions may vary for each hierarchal level, we decided to explore two Dimensions amounts: of three and seven variables. Dependent variables that were measured were task completion time and task accuracy.

4.1 Visualizations

The three visualizations examined were Radars (R), Flowers (F) and Circles (C). The way they were presented to participants can be seen in Figure 2. Apart from the number of dimensions which we examined as an independent variable, we fixed the number of years to ten, showing a 10-year time frame, and the number of items to five. The number of measured items in a CI is usually rather big. CIs who aim to compare between countries performances might contain data of a few dozens of items (e.g., 85 countries in WebIndex [7], 138 countries in NRI index[4]). However, when directly comparing items, this number is usually much smaller. for example, in the WebIndex, a desired item comparison is "myself" vs. best practice, worst practice, mean and median practice (i.e. five items) [7]. Thus, we decided to fix the number of items to five in the controlled experiment.

To compare the performance of the three visualizations, it is important to describe the interaction techniques supported for each, and to ensure that differences in interaction techniques are understood with regards to the outcome of the study. Since basic characteristic of three visualizations are different from each other (e.g. small multiples of years in Radars vs. small multiples of items in Circles and Flowers), interactions were designed in order to improve comparison fairness. "Drag" interaction was available in Circles and Flowers enabling to bring one city closer another city in order to make item comparison more comfortable and similar to Radars. For Circles, drag is enabled by standing with the mouse on a city name's text, and then dragging and dropping it to desired position. For Flowers, same action enables moving the entire item's row. Tooltip interaction was available in all three visualizations, displaying concrete value titled by the related variable's name.

4.2 Hypotheses

We postulated the following main hypotheses:

H1. *Radars visualization will be less effective than the other techniques. This will be independent of task or number of dimensions*. That is, participants will be (a) slower and (b) have more errors when using the Radars visualization compared with the two other visualizations. This is based on the criticism of Radar charts as expressed in [19, 20]

H2. *Completion time with 7 dimensions will be slower than completion time with 3 dimensions.* Adding complexity to the visualization, in terms of more dimensions, should increase performance time.

H3. *The number of dimensions will not affect the choice of which visualization is more effective.* This is based on the fact that the number of dimensions in all three visualizations was mapped to radial segments. Thus, we anticipate that changing the number of segments will affect all visualizations in a similar way.

H4. *Participants will show a general preference for the Flower visualization over the two other visualizations*. This is based on Bertini's and Few's comments about the beauty and elegance of the Flower visualization [11, 21].

4.3 Dataset

Our data source is taken from real marketing and media surveys conducted during the years 2003-2012 twice a year asking about ICT use. The data contains information from a representative sample of 30 regions and about 10,000 adults annually. Data was collected and analyzed, and the indicators were calculated and displayed for the purposes of measuring, benchmarking and making decisions about digital gap and educational policy at the national and regional level. Data was normalized to 1-100 time dependent scores for each variable. We prepared two constructs of 3 and 7 dimensions, representing typical subindicators of ICT index. The first construct represents an "infrastructure" sub index with scores expressing ownership of Computer, Internet and Cellular. The second construct represents a "usage" sub index with scores evaluating usage of internet search, email, sms, google, news sites, E-shopping and internet use. We prepared three different data blocks of 3 dimensions and three different data blocks of 7 dimensions to accommodate the 3 (Vis) x 2 (number of dimensions) design. Each data block consisted of data of 5 regions (items).

4.4 Tasks

For task analysis, we use the formal task model by Andrienko and Andrienko [8]. On the upper level, tasks are divided into elementary and synoptic tasks. Elementary tasks address individual and separated data elements (values or groups of data) and include: lookup, comparison and relation seeking tasks. Synoptic tasks involve a general view (sets of values or groups of data in their entirely), and are divided into descriptive (including: lookup, comparison and relation seeking tasks) and connectional tasks (homogeneous and heterogeneous behavior). We developed 10 tasks, which we found to be relevant to ICT stakeholders. The last five questions, which are of synoptic nature, are especially of interest in relation to the digital divide and are typical CI questions. The tasks are presented in Table 2.

4.5 Participants

Forty eight participants took part in the study. Participants were students from our University from different faculties and departments. Half of the participants were male (24) and half were female. The average participant age was 25.7 (*SD*: 3.8). All participants had normal

Task No.	Question	Task Type
1	What was the score of *item a* in *dimension x* on *year y* ?	Elementary Lookup
2	When was the lowest/highest score of *item a*, in *dimension x* ?	Elementary Reversed Lookup
3	Was score of *dimension x* in *item a* higher than *item b* in *year y*?	Elementary Comparison
4	Was score of *dimension x* in *item a* higher than *v* before/after *year y*?	Elementary Comparison
5	Mark 2 years on which *dimension x1* was higher than *dimension x2* in *item a*.	Elementary Relation-Seeking
6	What was *dimension x1* trend in *item a* between years *y1-y2*?	Synoptic Pattern Identification
7	Which variable shows maximum divide in *year y*?	Synoptic Behavior Comparison
8	Does divide between *item a* and *item b* in *dimension x* getting narrower/wider?	Synoptic Behavior Comparison
9	Which item increased the most at any dimension from first to last year?	Synoptic Relation-Seeking
10	Growth rate of *dimension x* in *item a* was higher/lower than *item b*	Synoptic Behavior Comparison

Table 2: User tasks given in the experiment. Task type refers to the task taxonomy in [8]. Italic font is used as a placeholder for data attributes

or corrected-to-normal eye sight, and no participants were color blind (self-reported).

4.6 Procedure

Experiments were conducted in a quiet room, one participant at a time. Participants were seated in front of a 24” screen using a full HD resolution of 1920x1080. Participants were first given a short description of the purpose of the study, followed by a short questionnaire containing questions about personal information. After completing the questionnaire, participants were given three consecutive session blocks. Each block consisted of a different Vis (Radars, Flowers or Circles) that was presented to the participant. For each block, a training session was first conducted. The training session started with an introduction of the visualization technique and the corresponding interaction demonstrated by the experimenter. Next, four tutorial tasks were given in order to get the participant acquainted with the visualization and the tasks.

After the training part, the ten tasks were presented. Participants were asked to work as quickly and accurately as possible. The visualization was presented in full screen, with the task description written on the bottom of the screen. The task was completed by selecting an answer from the list or by entering a number in a text-box. After the participant chose an answer he or she would press a next button to prompt the next task. Completion time and the result for each task was recorded by the software and later analyzed. In addition, all user interactions such as tooltips and dragging of cities (when possible) were recorded.

Each participant saw all visualizations with either 3 variables or 7 variables (the between-subject variable of dimensions). After each block, a questionnaire was given to assess participant’s subjective opinion of that visualization. Visualization (block) order was counterbalanced using a Latin square design to avoid any order or learning effects. Within each block, the order of the tasks was fixed with the more difficult tasks at the end. This allowed participants to build their skills as they proceeded. At the end of the session, participants were asked to fill out a general questionnaire that asked comparative questions about all three visualizations. Each session lasted between 35-45 minutes.

5 Results

We present the study results in three parts: effectiveness (in task completion time and accuracy), behavior (in user interactions) and subjective preferences.

5.1 Quantitative measures

In order to analyze task completion time, we conducted a 3 (Vis: Radars vs. Flowers vs. Circles) x 10 (Tasks) x 2 (Dimensions: 3 vs. 7) Mixed Analysis of Variance (ANOVA) with dimensions being a between-subject variable and Tasks and Vis being within-subject variables. Results were analyzed in log time to control for the skewness in reaction time data, but are reported in seconds for clarity. In order to analyze accuracy, we conducted a mixed model for binary responses on correctness.

5.1.1 Main differences in completion times

We compared overall completion times of the three visualizations: Radars (R), Flowers (F) and Circles (C) (see Figure 3). In terms of task completion time we observed a highly significant main effect for Vis, $F_{(2,92)}$=29.56, $p<.001$. The average time when using the Radars visualization (M = 42.3 sec., SD = 1.78) was slower than when using Flowers (M = 35.6 sec., SD = 1.43) and Circles (M = 35.5 sec., SD = 1.50). Post-hoc analyses using the Bonferroni correction showed that the difference between the Radars and the other two visualizations was significant ($p<.001$), supporting H1a. There was almost no difference between Circles and Flowers.

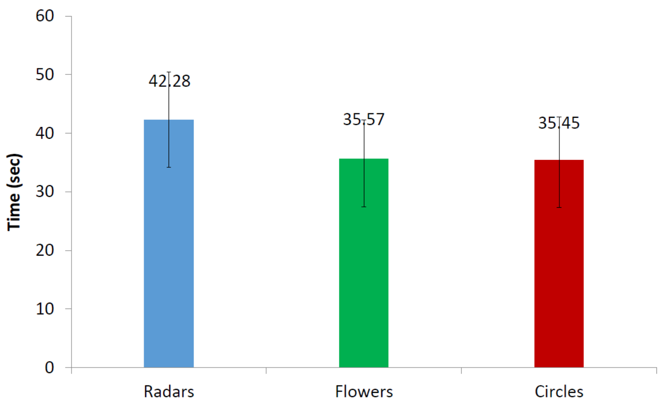


Fig. 3: Average task completion time of the three visualizations. Error bars denote 95% error interval. A significant main effect was found for Visualization, with Radars being significantly slower than Flowers and Circles

5.1.2 Number of Dimensions

Two dimension amounts were examined for each visualization (using a between-subject design): 3 and 7. In terms of task completion time, main effect of dimension amount was found significant, $F_{(1,34)}$=10.31, p=0.002. The average time when using 3 variables (M = 33.36 sec., SD = 1.99) was faster than when using 7 variables (M = 42.17 sec., SD = 1.99), supporting H2. This is not surprising since complexity of solving tasks involving 3 dimensions is naturally lower than when using 7 dimensions.

More interestingly, we did not observe a significant interaction effect between Vis and Dimension, as can be seen in figure 4. This finding supports H3. Practically, this is useful since it implies that the same visualization could be used independently on the amount of dimensions used.

5.1.3 Completion times within tasks

Looking at completion times, interaction effect of Visualization and Tasks was found highly significant ($F_{(18,828)}$=11.26, $p<.001$). Aver-

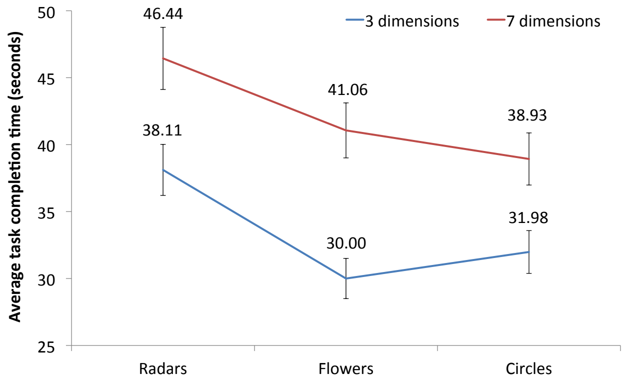


Fig. 4: Average task completion time per Vis and Number of Dimensions. No interaction effect was found between Vis and Number of Dimensions

age completion time in seconds per visualization for each task can be seen in Figure 5. Radar was slowest in all tasks, with the exception of task 7 ("find maximum divide between 2 items"), where radar was the fastest, and task 3 (comparing two items for a single year), where Radars was faster than Circles (and Flowers is the best). Tasks 3 and 7 are characterized by items comparison, which might explain why Radars performs better in those tasks, since divide between items can be clearly seen on the same Radar chart.

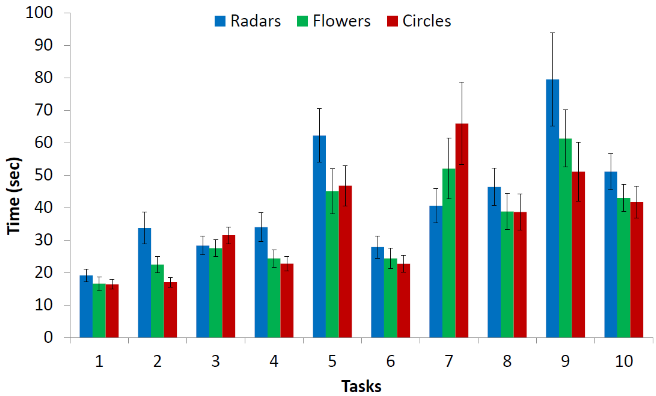


Fig. 5: Average task completion time per Task and Visualization. In most tasks, Circles performed fastest, and Flowers faster than Radars.

We found four main completion time patterns: Circles faster than Flowers faster than Radars (CFR pattern), Flowers faster than Radars faster than Circles (FRC pattern), Radars faster than Flowers faster than Circles (RFC pattern), and Flowers faster than Circles faster than Radars (FCR pattern) . Patterns and Pairwise comparison with Bonferroni correction are summarized in table 3. Prevalent pattern (in 7 out of 10 tasks) showed the CFR pattern, all but one with a significant difference between Circles and Radars (CR) and between Flowers and Radars (FR). This result shows that both Circles and Flowers outperform Radars for time completion in a broad set of tasks. However, difference between F and C was not found significant in most tasks (6 out of 10). Similarly, the only task showing an FCR pattern showed significant effect between Circles and Radars and between Flowers and Radars, and did not show an effect between Circles and Flowers.

Although in general Flowers and Circles task completion times averages were similar, in most tasks (no. 1, 2, 4, 6, 8, 9, 10) a CFR pattern was found. Significance in difference between C and F was found in tasks 2 (when was the lowest/highest score of a specific item and dimension) and 9 (Which item increased the most at any variable from first to last year). Advantage of Circles in tasks 2 and 9 might

Task No.	Task Type	Pattern	Pairwise comparison CF	RF	RC
1	Elementary Lookup	CFR		.027	.05
2	Elementary reversed Lookup	CFR	.001	.001	.001
3	Elementary Comparison	FRC	.015		.015
4	Elementary Comparison	CFR		.001	.001
5	Elementary Relation-Seeking	FCR		.001	.001
6	Synoptic Pattern Identification	CFR		.081	.006
7	Synoptic Behavior Comparison	RFC			.001
8	Synoptic Behavior Comparison	CFR		.015	.021
9	Synoptic Relation-Seeking	CFR	.028		.001
10	Synoptic Behavior Comparison	CFR		.041	.005

Table 3: Time pattern (from fastest to slowest) and p-values of pairwise comparisons per task between the three visualizations. C = Circles, F=Flowers, R = Radars. Empty cells in pairwise comparisons denote a non-significant effect

be explained by the fact that the time property in Circles is coded as rings on the same glyph rather than to Small Multiples in Radars and Flowers. Thus, scanning years in Circles is easier than scanning a row of 10 Flowers or 10 Radars. In addition, the color coding of the Circles' values assist in finding the "most green" or "most red" sector in a circle, which is the operation needed in both tasks 2 and 9.

Task 3 (elementary comparison between items) is the only task that shows an FRC pattern. Disadvantage of C is revealed in this task as a significant difference between C and the other visualizations was found. Radar's advantage in task 3 might be explained by the fact that comparing values between items were done on the same Radar, using position comparison. Comparing between Flowers required comparing a petal's size in 2 item rows.

Finally, a prominent advantage of Radars visualization in terms of task completion time was found in task 7 ("find maximum divide between 2 items"). An RFC pattern in this task shows that Radars was best for the maximum divide identification task.

5.1.4 Error rate

Error rates were relatively low and most participants managed to get the correct answers. Overall error rate was 14.79%. A distribution of correctness rate per task can be seen in Figure 6. Not surprisingly, the more complex synoptic tasks had higher error rates. To examine if error rates differed between visualizations, we conducted a mixed model for binary responses on correctness. Accuracy was analyzed on data without tasks 1, 3, 5 and 6, where correctness rates were higher than 95%. No significant difference was found between the Visualizations for correctness rates.

To analyze the accuracy of each task, we ran a mixed model for binary responses on error rates for each task separately. Results show a significant Vis effect only in task number 7 ("Find maximum divide between 2 items"), $F_{(2,94)} = 5.0$, p=.009. In task 7 the Radars method led to fewer errors (10.42% error rate) than Flowers (27.08%) or Circles (41.67%). This result is complementary with the result of task 7's completion time, which was found to be fastest with Radars.

5.2 Behavior

In order to understand the behavior patterns of the users when using the three visualizations, we analyzed their interaction with the interfaces using the system logs. To analyze tooltip use, we summarized all tooltip events for each task. To take into account arbitrary tooltip occurrences that result from quick movements of the mouse, we included only tooltip events that lasted more than two seconds. We conducted a 3 (Vis: Radars vs. Flowers vs. Circles) x 10 (Tasks) x 2 (Dimensions: 3 vs. 7) Mixed Analysis of Variance (ANOVA) with Dimensions being a between-subject variable and Tasks and Vis being within-subject variables on number of tooltip events. We observed a highly significant main effect for Vis, $F_{(2,92)}$=32.78, p<.001. The average number of tooltips per task when using the Flowers visualization (M =0.64, SD = 0.05) was lower than the number of tooltips per task when using Circles (M = 1.04, SD = 0.85), which in turn was lower than that of the Radars (M = 1.48, SD = 0.12). Post-hoc analyses using the Bonferroni correction showed that all differences between three visualizations were significant (p<0.001).

When looking at the number of dimensions, the average number of tooltips per task when using 3 dimensions (M = 0.97, SD = 0.1) was lower than when using 7 dimensions (M = 1.15, SD = 0.1). However, this difference was not significant. When looking at the interaction between dimensions and Vis, consistent with the results from Section 5.1.1, no interaction effect was found.

Figure 7 compares the number of tooltips in each task between the three visualizations. As can be seen in the figure, in all tasks, Radars had higher tooltips use than Circles and Flowers, with the exception of tasks 7 (Maximum divide Seeking) and 10 (Growth rate comparison). Flowers required the lowest tooltips use in most tasks (9 out of 10).

5.3 Subjective preferences

At the end of each block for a particular visualization, a feedback survey was given with nine questions. Questions were given on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). With ranked ordinal data and a relatively small sample size it is recommended to use a non-parametric statistical test [24]. We therefore used the Friedman non-parametric test to examine differences in ranking between the visualizations, followed by post hoc Wilcoxon tests with the Bonferroni correction for pair-wise comparisons.

Results of the Friedman tests ratings indicated significant difference between the three visualizations for all questions (p<.001). Table 4 shows the questions given and participants' average ratings, as well as which pair-wise comparisons were also significant. As can be seen in the table, results were uniform and consistent on all questions. In all nine questions Flowers was preferred overall, followed by Circles, with Radars receiving the lowers ratings.

At the end of the experimental session, 8 comparative questions were presented to participants, who were asked to choose their preferred visualization for each question. Results are presented in Table 5. We see that most participants stated overall preference towards Flowers (Q8) supporting H4. In addition, some further interesting observations can be made. First, we see that consistent with the previous results, most participants thought Flowers is generally most effective, most enjoyable and easiest to learn. This was again followed by Circles and then Radars. However, when asked about specific tasks, equal preference between Circles and Flowers for Q3 and Q4 was shown. Finally, when looking at common digital divide tasks, Circles was slightly preferred for trends identification (Q5) and Flowers was slightly preferred for divides (Q6). Radars was least preferred on all scales.

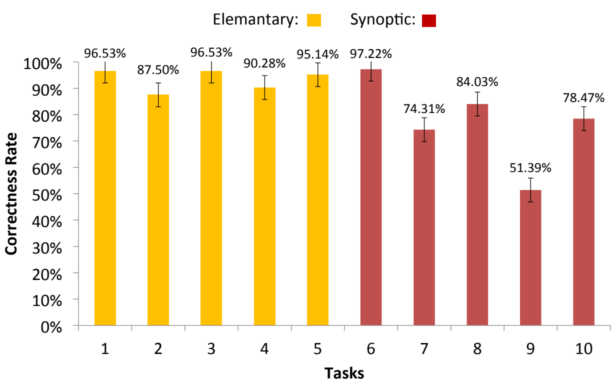


Fig. 6: Answer correctness rate per task revealed complimentary results to completion time.

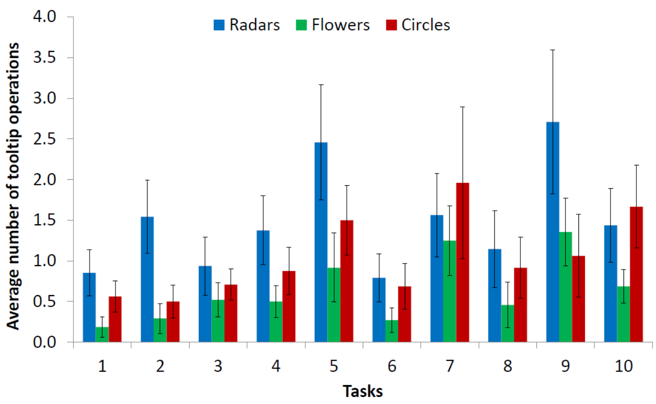


Fig. 7: Average number of tooltip operations per task shown for each task with each Visualization. In most tasks, Flowers required least interaction, and Circles less than Radars.

6 Discussion

Three radial visualization techniques for the display of multidimensional time-series CI were examined by a series of user tests with 48 subjects: Radars, Flowers glyphs and Concentric Radial Space Filling Circles. Two dimension amounts were tested (3 and 7) using elementary and synoptic tasks adjusted to tasks concerning the digital divide and related to CI. Performance was measured by task completion time and accuracy. User tooltip interactions behavior was studied, as well as user preferences.

A deeper look into the results and their interpretation reveal the following insights:

The Radars method both performed worst and was perceived as worst. Overall Radars performance in terms of time was significantly slower than Flowers and Circles supporting H1a. H1b, hypothesizing a difference in error rates was not supported. The salient inferiority of Radars in most tasks in terms of completion times is most likely related to the massive use of tooltips in the Radars condition. The average number of tooltips per task when using Radars was significantly higher than when using Circles and Flowers. This indicates that with Radars, participants needed to drill down more into the data rather than visually perceive the differences in the alternative visualization. These results are complementary with the subjective preferences. Preference analysis indicates Radars was judged to be the least helpful and least liked. It seems that the overlapping and cluttering of the Radars visualization made it difficult for participants to identify and compare values, as was commented by many participants: (e.g., *"It is very loaded which makes it confusing when the data has close values"*, *"it is difficult to isolate a value point"*. The connecting lines of the multidimensional values in the same item seemed to cause visual load, while not providing any actual contribution, in effect being chartjunk [34]).

Still, Radars are good for the task of finding the maximum divide. While Radars was slower in 8 out of the 10 tasks, in task 7 (maximum divide identification) we found a unique advantage of Radars in

	Question	Radars	Flowers	Circles	Sig.
1	The visualization was enjoyable	3.75 (2.09)	5.46 (1.54)	4.71 (1.73)	CR, CF, RF
2	Using the visualization aid was clear and understandable	4.52 (1.57)	5.85 (1.26)	5.48 (1.41)	CR, CF, RF
3	Overall, I believe that the visualization aid is easy to use	3.71 (1.83)	5.42 (1.53)	4.69 (1.64)	CR, RF
4	using the visualization aid was effective	4.08 (1.78)	4.08 (1.78)	4.65 (1.63)	CR, RF
5	Learning how to use the visualization aid was easy for me	5.27 (1.53)	6.19 (1.09)	5.71 (1.32)	CR, RF
6	Most people would learn to use the visualization aid very quickly	3.69 (1.85)	5.46 (1.44)	4.48 (1.56)	CR, CF, RF
7	**I did not find the usage of the visualization aid unnecessarily complex	2.94 (1.94)	5.00 (1.51)	4.04 (1.82)	CF, RF
8	**Using the visualization aid did not require a lot of mental effort	2.19 (1.68)	3.94 (1.68)	3.13 (1.48)	CR, RF
9	**Using the visualization aid was not frustrating	2.79 (1.98)	4.46 (1.48)	4.06 (1.45)	CR, RF

Table 4: Average ratings (and standard deviation) per visualization (Circles, Radars and Flowers) on a 7-point Likert scale. Significant pair-wise comparisons using the Wilcoxon text is listed in the significant column (N=48). ** denotes a reversed question

terms of completion times. In addition, a unique significant effect was found in this task's accuracy with Radars being most accurate. This result can be explained by the ease of item comparison in a single dimension in Radars, since all items are on same object (Radar), sharing axes. Task 7 is visually translated to finding the longest segment delimited by data points, all segments on the same axes. In Flowers on the other hand, Task 7 is translated to petal size comparison of different objects (Flower glyphs). In Circles, task 7 is translated to color comparison, again each color being on a different object (Circle). Furthermore, position is better than size and color as a differentiating a visual variable [35]. This result can help explain why Radars are a common way to illustrate divides.

For choosing which visualization to use, there was no difference when scaling from three to seven dimensions. Interaction effect of dimension amount and Vis was not found to be significant in terms of task completion times, in terms of accuracy, and in tooltip use. This result supports H3. Subjective preferences analysis also did not find a significant effect of amount of dimensions. This finding can be explained by the fact that in all three visualizations, the dimensions were mapped to the radial sectors. As such, while performance is slower with more dimensions, the difference is equal in all three visualizations. This finding is important, as it implies that the same visualization can be used independently on the amount of variables presented. Since a CI is often hierarchical with a different number of dimensions for each sub-index, this means that the same visualization can be used to present different sub-indexes.

People love flowers. Participants showed a strong preference toward the Flowers visualization, supporting H4. The Flowers method

	Question	Radars	Flowers	Circles
Q1	most effective	4	31	13
Q2	most enjoyable	5	32	11
Q3	best for checking concrete values	8	20	20
Q4	best for task 9	12	18	18
Q5	best for trends identification	7	17	24
Q6	best for divides identification	11	20	17
Q7	easiest to learn	1	35	12
Q8	do you prefer overall	5	27	16

Table 5: Overall comparison results between the three methods starting with: "which method was...". Each cell indicates how many participants preferred each method on that question. (N=48)

was found to be more enjoying, understandable, easy to use and easy to learn. They were also said to require the least mental effort and to be least frustrating. It is interesting to note that although there was no significant difference between Flowers and Circles in terms of actual effectiveness (and in two tasks, Circles was actually significantly faster), Flowers was perceived to be the most effective method. This clear cut result is surprising. It might be partly explained by the tooltip results - participants used the least amount of tooltips per task when using Flowers. It might be that participants prefer to see the information spread-out rather than use tooltips to extract information. Since Flowers use small multiple both for times and for items, participants perform less "drill downs" and value-specific extraction was less needed. Another explanation is the esthetic appeal of the Flowers which was commented on by several participants (e.g., *"A very pretty visualization. In contrast to the other two visualizations there is no visual load, and it is easy to see differences in leaf sizes"*). Blooming flowers visualization might serve as an attractive metaphor that may bring life to the "cold numbers" of statistics. This metaphor is cross-culture and age, as was expressed by one of the participants: *"Flowers are part of my world, I see them every day, they are nice for showing growth"*.

Circles Strengths. The Circles method was found significantly best in tasks 2 and 9 in terms of task completion time, and lead as fastest (although not significantly different from Flowers) in tasks 1, 4, 6, 7 and 10. Subjective preferences imply Circles was judged to be more helpful than Flowers and Radars for trend identification. This result might be explained by the advantage of red-to-green color code that emphasizes gradual changes or trends, as was expressed by participants: *"Red and green color coding contributed a lot to trend identification"*, *"it was especially comfortable for analyzing data of a specific city"*. Grouping together of all year values for one given item on same compact circle might also explain Circles' strengths in trend identification.

The Circles method was inferior for item comparison. In terms of task completion time, Circles was mostly faster than Flowers and Radars. Yet, in task 3 (elementary items comparison) an FRC pattern was found, and in task 7 (maximum divide identification) an RFC pattern was revealed. Both patterns show inferiority to Circles visualization comparing to Radars and Flowers. Tasks 3 and 7 are characterized by items comparison, which might explain why Circles show a disadvantage. Items comparison in the Circles visualization is visually translated to these actions: find the two circles that represent the items needed; possibly move the circles closer to each other, find the sector of the dimension; compare colors of that sector for the two circles. In task 3 (elementary items comparison) these sequence of actions should be made once, while task 7 (find maximum divide) requires these action made in a loop (in addition to divide estimation). As can be seen in Figure 7, Task 7 caused the most tooltips uses in the Circles condition. To summarize, our results show Circles design weakness for these items comparison tasks.

7 LIMITATIONS

In a controlled experiment, each experimental design decision brings with it some sort of tradeoff. In this section, we list the limitations of our study in light of the design decisions we made and the threats for validity in the methodology.

Choosing which charts to be compared bears some presumptions. There are many other techniques that could be applied for the data and tasks at hand. Especially star glyphs should be mentioned, which have been investigated empirically by Lee et al [27]. Also, parallel coordinates have the ability to efficiently encode multidimensional data, as evaluated for the purpose of information visualization by Caat et al. [13]. However, parallel coordinates are limited in the way they can encode time. Even with just the techniques investigated in this work, the visual possibilities of glyph design and alternations are almost endless [12]. Readers should be aware that results obtained here are limited to those techniques that were compared and alternative formations might help overcome shortcoming of the techniques examined.

The current design decisions pose some limitations on the empirical evaluation. Radar and Flower charts encode quantitative information on a continuous scale, whereas the same information is binned to 9 colors on the Circle chart. Consequently, color might give a faster impression about tendencies than sizes, while size may give a more accurate estimation of the value. However, in all cases participants had to use the hover feature to validate their hypotheses and perform the tasks. Consequently, we believe the differences in performance – if at all – is marginal. Nevertheless, the effects of using binary or tertiary colors, or limiting the tasks to pure tendency assessment, may have implications on performance and should be investigated further in the current context.

The screen-space occupied by charts have differed for the selected visualization types and visual encodings. Circle charts were most dense, one glyph for every city, 5 glyphs in total. Radars were second, having one glyph per year, 10 glyphs in total. Last, flowers were the most sparse having one glyph per year and city, 50 glyphs in total. The dimensions themselves were always part of the glyphs. Figure 2 provides a good impression of how much of the screen-space was occupied by each of the charts. This has implications on the scalability of the different graphs. A new year of data would hardly make any difference for the Circles, might have some for the Radars, but add a whole new column to the Flower chart. Similarly, adding another item adds a line to the Radars (yet increasing the clutter of the Radar chart, thus number of items shown together is limited), adds another glyph to the Circle chart and another row to the Flowers. Thus, while the Flower chart showed promising results, it should be noted that it is limited in its scalability to show multiple years and items.

Our results did not show performance differences when raising the number of dimensions from 3 to 7, thus we concluded that the same visualization can be used regardless of the number of dimensions. We chose 3 and 7 dimensions since 3 is a common amount for the top level hierarchy in many indices [3, 7, 4] and most CIs we examined did not go over 7 or 8 dimensions [7, 4] . Still, it might be that when raising the number of dimensions even further, there would be a difference in performance between visualizations. This does also not imply that the number of dimensions have no limitation. To the best of our knowledge, there was no systematic investigation on the maximum number of dimensions that can be encoded in radar charts. The number of items in our experiment was set to 5. This was necessary in order to conduct a controlled experiment. Future studies may investigate the effect of having different number of items.

8 DESIGN GUIDELINES AND CONCLUSIONS

When designing an interactive visualization for a CI, we aim to find one visualization that is both effective and engaging. One that would bring people's attention to and be memorable while enabling insight generation and deep understanding of the trends and gaps that might exist in the data. Multiple visualizations for a single CI is not recommended since it can confuse the user and require more learning time.

Our experimental results revealed a very clear understanding for devising design guidelines. Our results showed many recurring patterns through most of the tasks both in the effectiveness, the behavior and the preference. Many of these patterns were significant. Thus, we recommend the following guidelines for the design of CI visualizations:

Use one metaphor. The fact that no interaction effect was found between Visualization method and *number of dimensions* suggests that the same visualization can be used throughout a single CI website or report. This is important, because many CIs are built as an hierarchy, with several subindexes. Our results suggest that a single radial visualization can be used for all indexes and in all levels independent of the number of dimensions that each subindex might have.

Off with the Radar. In all tasks but one, Radars was outperformed by the two other visualizations. In addition, Radars was least liked. Thus, in general, we do not recommend using the Radar chart for an interactive CI visualization. However, in static reports that focus on gaps or divides (e.g., digital divide) and that focus on a single point in time (i.e., there is no temporal aspect) Radars can be useful.

When space is available, Flowers are recommended. For most purposes Flowers and Circles are equally suitable regarding performance. However, the strong subjective preference for Flowers makes Flowers the recommended option. It should be noted that in Flowers, because small multiples are used both for time and for items, it may take a lot of screen space, especially if many time intervals are examined.

Circles are recommended for detecting trends. Still, Circles was better than Flowers in detecting trends (and participants acknowledged this in their preference results). Thus, when there is an issue of screen space or when trend identification is most important, we recommend using Circles.

To conclude, although Radar visualization is most common, both its performance and popularity were lowest in the three-way competition. Although confirming our intuition, the extent of the results was still surprising. Future research would need to investigate more task types, which might leverage the advantages of Radar charts, e.g. detecting clustering and classifying behavior. We chose three radial visualizations that are applicable for CI and discussed their visual mapping and reflected on our decisions. We believe that our visual mapping is most applicable and led to a fair comparison. However, the options for these and other visualization types are basically limitless, and should be taken into consideration. We have not taken the mapping itself into the experiment as independent variables. Such an evaluation could be the focus of further research. Another direction for future work is to explore the option of animation for the temporal aspects of CI visualization. Dynamic visualization that animate CI's change over time might prove better than using static visualization as used in this study.

ACKNOWLEDGMENTS

This work is done as part of the National Israel ICT project by The Center of Internet Research (http://infosoc.haifa.ac.il) supported by the Israel Internet Association-ISOC-IK; Appleseeds Academy; the LINKS I-CORE Program of the Planning and Budgeting Committee and The Israel Science Foundation (1716/12).

REFERENCES

[1] Better life composite index. http://www.oecdbetterlifeindex.org/. Accessed: 2015-03-30.

[2] Gapminder trendalizer. http://www.gapminder.org/. Accessed: 2015-03-30.

[3] Itu ict development index (idi). http://www.itu.int/ITU-D/ict/statistics/explorer/index.html. Accessed: 2015-03-30.

[4] Network readiness index. http://www.weforum.org/issues/global-information-technology/the-great-transformation/network-readiness-index. Accessed: 2015-03-30.

[5] Semmelweis rankings. https://lib.semmelweis.hu/app/getFile&id=678. Accessed: 2015-03-30.

[6] Thomson & Reuters institution profiles. http://roundranking.com/materials/ThomsonReuters-GlobalInstitutionalProfilesProject.pdf. Accessed: 2015-03-30.

[7] Web index. https://thewebindex.org/wp-content/uploads/2012/09/2012-Web-Index-Key-Findings.pdf. Accessed: 2015-03-30.

[8] N. Andrienko and G. Andrienko. *Exploratory analysis of spatial and temporal data: a systematic approach*. Springer Science & Business Media, 2006.

[9] R. Bandura. A survey of composite indices measuring country performance: 2008 update. *New York: United Nations Development Programme, Office of Development Studies (UNDP/ODS Working Paper)*, 2008.

[10] K. Barzilai-Nahon, S. Rafaeli, and N. Ahituv. Measuring gaps in cyberspace: Constructing a comprehensive digital divide index. In *Workshop on Measuring the Information Society, the conference of Internet Research*, volume 5, 2004.

[11] E. Bertini. Oecds better life index. http://fellinlovewithdata.com/reviews/review-better-life-index, 2011.

[12] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen. Glyph-based visualization: Foundations, design guidelines, techniques and applications. *Eurographics State of the Art Reports*, pages 39–63, 2013.

[13] M. t. Caat, N. M. Maurits, and J. B. Roerdink. Design and evaluation of tiled parallel coordinate visualization of multichannel eeg data. *Visualization and Computer Graphics, IEEE Transactions on*, 13(1):70–79, 2007.

[14] J. R. C.-E. Commission et al. *Handbook on constructing composite indicators: methodology and user guide*. OECD Publishing, 2008.

[15] J. Cukier. Can data visualization help build democracy? *XRDS: Crossroads, The ACM Magazine for Students*, 18(2):26–30, 2011.

[16] S. Diehl, F. Beck, and M. Burch. Uncovering strengths and weaknesses of radial visualizations—an empirical approach. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):935–942, 2010.

[17] G. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 15(5):759–776, 2009.

[18] S. Dutta, B. Bilbao-Osorio, and T. Geiger. The global information technology report 2012. In *World Economic Forum*, pages 3–22. Citeseer, 2012.

[19] R. Feldman. Filled radar charts should not be used to compare social indicators. *Social indicators research*, 111(3):709–712, 2013.

[20] S. Few. Keep radar graphs below the radar–far below. *Perceptual Edge*, 2005.

[21] S. Few. Data blooms in beauty and truth. http://www.perceptualedge.com/blog/?p=1044, 2011.

[22] Å. Grönlund and T. A. Horan. Introducing e-gov: history, definitions, and issues. *Communications of the Association for Information Systems*, 15(1):39, 2005.

[23] M. Harrower and C. A. Brewer. Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

[24] S. W. Huck, W. H. Cormier, and W. G. Bounds. *Reading statistics and research*. Harper & Row New York, 1974.

[25] D. A. Keim, J. Schneidewind, and M. Sips. Circleview: a new approach for visualizing time-related multidimensional data sets. In *Proceedings of the working conference on Advanced visual interfaces*, pages 179–182. ACM, 2004.

[26] W. T. B. Kelvin. *Electrical units of measurement*. Institution, 1883.

[27] M. D. Lee, R. E. Reilly, and M. E. Butavicius. An empirical evaluation of chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24*, pages 1–10. Australian Computer Society, Inc., 2003.

[28] D. Maheshwari and M. Janssen. Measurement and benchmarking foundations: Providing support to organizations in their development and growth using dashboards. *Government Information Quarterly*, 30:S83–S93, 2013.

[29] H. Mosley and A. Mayer. Benchmarking national labour market performance: A radar chart approach. Technical report, WZB Discussion paper, 1999.

[30] G. Robertson, R. Fernandez, D. Fisher, B. Lee, and J. Stasko. Effectiveness of animation in trend visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1325–1332, 2008.

[31] M. Saisana and S. Tarantola. *State-of-the-art report on current methodologies and practices for composite indicator development*. Citeseer, 2002.

[32] G. Sciadas. International benchmarking for the information society. In *ITU-KADO digital bridges symposium*, 2004.

[33] M. Stafoggia, A. Lallo, D. Fusco, A. P. Barone, M. D'Ovidio, C. Sorge, and C. A. Perucci. Spie charts, target plots, and radar plots for displaying comparative outcomes of health care. *Journal of clinical epidemiology*, 64(7):770–778, 2011.

[34] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. Graphics press Cheshire, CT, 1983.

[35] M. O. Ward, G. Grinstein, and D. Keim. *Interactive data visualization: foundations, techniques, and applications*. CRC Press, 2010.