City Research Online

City St George's, University of London

Citation: Kachkaev, A., Wood, J. & Dykes, J. (2014). Glyphs for exploring crowd-sourced subjective survey classification. Computer Graphics Forum, 33(3), pp. 311-320. doi: 10.1111/cgf.12387

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/3393/

Link to published version: https://doi.org/10.1111/cgf.12387

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

Eurographics Conference on Visualization (EuroVis) 2014
H. Carr, P. Rheingans, and H. Schumann
(Guest Editors)

Volume 33 (2014), Number 3

Glyphs for Exploring Crowd-sourced Subjective Survey Classification

A. Kachkaev, J. Wood and J. Dykes, giCentre, City University London

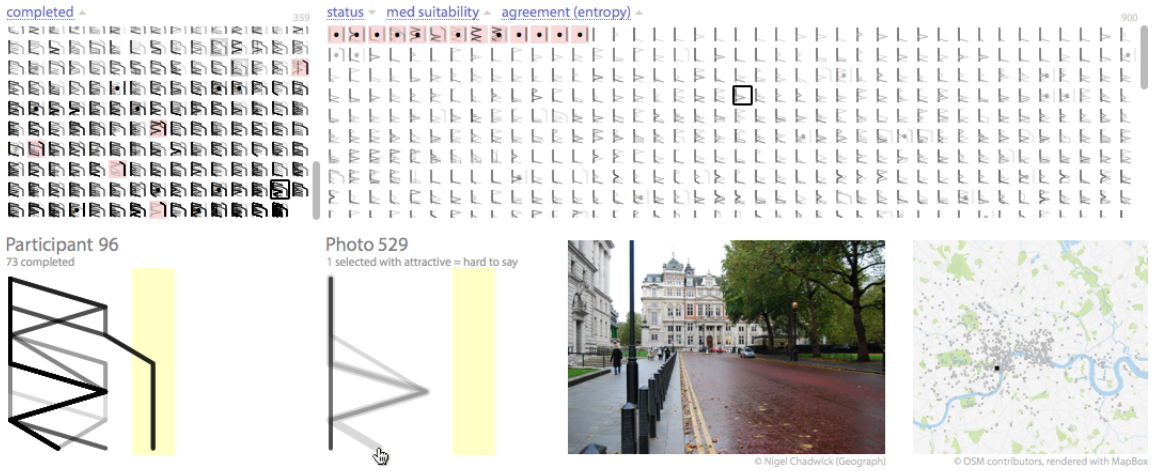


Figure 1: The interface of the survey analysis tool allowing navigation through 8,434 collected responses (49,285 answers) with use of glyphs, linked views and interaction. Available at photoassessment.org/results, demo video at vimeo.com/90299533.

Abstract

The findings drawn from opinion survey responses are usually made by producing summary charts or conducting statistical analysis. Both involve data aggregation and filtering as exploring the unaggregated data has traditionally been impractical or error-prone for large numbers of responses. We propose the use of glyphs with parallel coordinate plots to show all survey responses in a single view and design an interactive visual analytics tool around the representation to explore the data. We use this software for a 'photo content assessment' survey, where 359 participants classify 900 images by seven criteria. The proposed approach allows all 8,434 responses (49,285 answers to questions in total) to be represented in a single view and helps analysts to both clean the data and understand the nature of the survey responses. We describe the construction of the survey response glyphs and the interface to the interactive visual analytics software and generalise the design principles that arise from the approach. We apply the tool to two other datasets to evaluate the technique and to confirm its wider applicability for surveys with Likert scale responses.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation D.2.10 [Software]: Design Tools and Techniques—User interfaces

1. Introduction

In this paper we discuss the design of a visual analytics tool developed to explore subjective opinions of 900 geotagged photographs in central London. This particular task represents one instance of a more general process of categorical social survey analysis. While using statistical methods are commonly applied to find trends in such data, we make the case here for using visual analytical methods for both data processing and analysis. An approach that is well suited to handling imperfect and subjective responses in social survey data where it is not clear at the outset what types of patterns and structure should be identified. Central to our approach is in the concept of a *survey response glyph*, a glyph that em-

bodies all responses related to an individual subject or a participant. We evaluate the design by applying it to two additional datasets and propose that the insights we have gleaned as well as the techniques we have developed have wider applicability in survey response analysis.

2. Background and Related Work

The techniques of survey data representation differ depending on the purpose of the visualization, the nature of the data and the target audience. Commonly the results of the categorical surveys are aggregated into tables [Sap99] and are depicted in a tabular form. Despite the simplicity of the approach it is rather ineffective for exploring the relationships in the samples and extracting key trends. To facilitate comparison between values belonging to different categories, conditional formatting [Mic10, AS05], table-lens [RC94] or other related techniques have been applied. While tables are highly informative and precise, they are not very suitable for exploring multidimensional survey datasets or unaggregated data [Sap99].

The simplest and probably the most widely applied techniques of graphical representation of categorical survey data are bar charts and pie charts [Off11, O'B13]. While being familiar to a wide audience and easy to understand, they are characterised by a low data-ink ratio [Tuf83] and thus are only suitable for visualizing highly aggregated data. Besides, pie charts are often criticised for causing difficulties in comparing proportions between categories and problems with scaling underlying values [Few07, WW05, Tuf83].

Often the structure of categorical responses allows more sophisticated and data-rich visualization techniques. For instance, spatially tagged survey results are layered over conventional geographical maps, displayed on Choropleth maps or structured in a form of spatial treemaps [Uni13]. With the developing power of accessible and easy to use computer technologies different visualization techniques are more easily combined in an interactive environment, giving an opportunity to see collected survey results in multiple views and to interact with them [Off11, LSR10].

A special case of categorical responses is the Likert scale [TW06], used when the categories are ordinal. This type of scaling is widely applied for rating subjective opinion and is also used in this research. To visualize the spread between the categories, box plots or similar can be used instead of bar charts or in conjunction with them [TW06, WS12].

Survey data are commonly visualized to summarise findings, while cleaning and processing are performed using statistical methods [Sap99]. However, analysis by means of visual exploration can be more suitable in cases where the nature of trends to be discovered are unknown at the point of analysis. This leads to a need to effectively represent larger volumes of data with sufficient fidelity and clarity to support decision making. We could not find the evidence of use of high-resolution visualization techniques such as sparklines [Tuf83] or glyphs [BN11] for exploring individual results in larger categorical surveys and so propose an alternative described here.

3. The Survey Response Glyph

When a survey comprises a set of subjects that are classified by respondents using a number of criteria, a single response may be described as an array of natural numbers. This array is a subset of a matrix with questions (criteria) in rows and answers (categories) in columns. A group of responses aggregates individual answers and can be presented as a table with their frequencies. Another characteristic may be introduced to show relationships between classifications: it counts frequencies of all possible matching pairs of answers. This may help to reveal common patterns in behaviour, which is useful for detecting insincerity (deliberate insertion of erroneous responses) [Ame07] and understanding opinions of participants about subjects. The total number of parameters describing a group increases exponentially as the answer space grows and can be reduced if only relationships between answers to neighbouring questions are considered. This still allows patterns in individual responses to be detected while excluded parameters may be brought back into the set by changing the order of questions.

The challenge is to represent all these parameters compactly in order to both evaluate them and to be able to visualize multiple groups of responses using juxtaposition for the purpose of comparison [GAW*11].

Inspired by the successful use of glyphs in a wide range of fields [BN11, WHWC12, MRSS*12, LKH09], we apply this technique to survey data analysis. A glyph 'is a small visual object that can be used independently and constructively to depict attributes of a data record or the composition of a set of data records' [BKC*12]. A good overview of possible glyph designs can be found in [War02] and [War08].

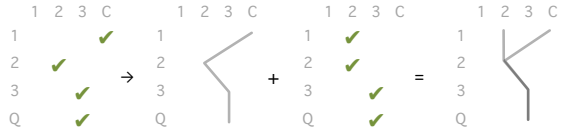


Figure 2: The concept of a survey response glyph. Rows correspond to survey questions and columns contain categories.

Each glyph is similar to a parallel coordinate plot (PCP) [ID90] rotated by 90°. Glyphs can then be arranged in a *survey response grid* having questions (dimensions) in rows and categories in columns. Each response is represented with a semitransparent polyline showing both individual answers and links between neighbouring questions within a response. With this approach we can show multiple response profiles in a single view by laying them on top

of each other. This provides a single view that can: (a) show each individual answer; (b) all responses by an individual; (c) allow estimates of the numbers of answers in a group; (d) show the most common answer profiles and (e) the degree of disagreement among answers. Because such design is sensitive to ordering of categories (answers), it better fits ordinal variables, but can also be applied for nominal values or ordinal scales mixed with a nominal value (e.g. rating + not applicable).

4. Survey and Tool Design

The motivation for proposing response glyphs was to be able to analyse results of a survey we conducted into subjective photograph assessment. Volunteers were asked to assess a sample of 900 photographs in London collected from three data sources: Flickr, Panoramio and Geograph. We wished to canvas views on what types of images depict attractive walkable areas (Figure 3) in order to build a pedestrian routing system.

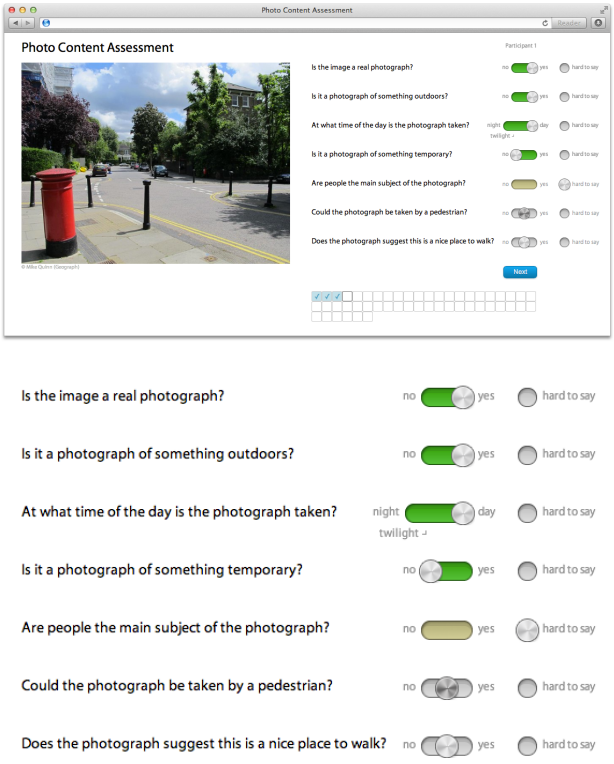


Figure 3: Photo content assessment survey: UI & questions.

Each participant was given a randomly ordered queue of images and was required to classify as many photographs as they wanted by seven criteria. In common with many survey designs, some questions were conditional on responses to previous questions, so the number of answers per photograph could vary between participants. Only responses with answers to all enabled questions were accepted; photographs could not be skipped. By the time the survey was closed in March 2013, it was viewed by 608 volunteers, 359 of whom classified at least one photograph excluding the test image used for learning purposes. We collected 8,434 responses enumerating 49,285 classifications; stratified sampling ensuring each photograph was rated by at least eight respondents.

There were two objectives in our processing of the responses. Firstly we wished to characterise the nature of the survey responses in order to remove erroneous or insincere entries. Secondly we wished to be able explore both respondents and the photographs being assessed in order to inform the process of automated selection of the photographs best suited to characterise urban attractiveness. In both cases, we did not know at the outset what types of survey response would characterise those we were most interested in so required a design that facilitated exploration.

4.1. Interface Design Principles

In constructing the exploration tool, we designed according to the following set of principles:

The data are the interface. We reinforce the link between data and interaction by making the elements that represent data also serve as the user interface. This reduces the need for separate buttons and menus which may cause clutter and decrease the data:ink ratio [Tuf83]. This was seen as particularly important for a design relying heavily on large numbers of glyphs with potentially complex structures.

Consistency of encoding. We map visual encodings of data and interaction consistently across all linked views. This was seen as important where a new 'grammar' of encoding had to be learned to make best use of the glyph design.

Use of states. Any change in the data view is considered as an action and can be stored, reverted or repeated. This was seen as important because quick and easy navigation between states eased comparison tasks as well as the process of early debugging and collaboration between analysts [WSD*13].

Maximise use of layout. We use ordering of glyphs to carry additional information. Ordering (position) as visual variable has the advantage of not introducing visual clutter – an important consideration for comparatively complex glyphs shape.

Use of keyboard shortcuts. In keeping with the objective of removing visual clutter we make use of keyboard in preference to mouse interaction. This was considered more appropriate for an expert-oriented visual analytics application [Fry08].

Transitions between views. We use smooth animated transitions between views where applicable in order to better re-

veal the differences in the data or to track amendments in their representation [HR07]. This had the secondary benefit of an aesthetic appeal to support engagement with the data [CM07].

Figure 1 shows the result of the design arising from following these principles. All survey responses are grouped by both participants and subjects (photographs) and are shown on a single screen in two sortable lists with glyphs of maximum possible size. The height of both lists is customisable while the widths are fixed to be proportional to group sizes. An enlarged interactive version for one glyph from each list can be opened on demand to explore the details. The layout is supplemented with a preview for a selected photograph (if any) and a map showing spatial locations of all images.

4.2. Representation of Participants' Responses

Following recommendations described in Section 3, we constructed two survey response grids (Figure 4).

Although the maximum number of categories per question was five (see *daytime*), we limited the grid to four columns by merging *twilight* and *hard to say* responses due to their semantic similarity in the context of our analysis. Although the answers to all questions are not numeric scores, they can be still considered as ordinal; 'n. a.' being an exception (nominal value) and may be optionally highlighted to support glyph interpretation.

Having one response grid should be usually enough for most surveys, however, in our case we have introduced *purpose-oriented ordering* (Figure 4b) of the grid in addition to *standard ordering* (Figure 4a). With this design the photographs that depict attractive walkable areas are represented with straight vertical lines on the left, which can be instantly decoded. We also change the order of questions in the *purpose-oriented grid mode* and move the one about people to be the second. This decision is determined by conditional question set within the survey: the question about the spatial environment of a photograph when the answer is 'indoors' disables remaining questions. Figure 5 shows examples of some response glyphs:

		1	2	3	4
real photo?	1	no	h. t. s.	yes	n. a.
outdoors?	2	no	h. t. s.	yes	n. a.
daytime?	3	night	h. t. s. / twilight	day	n. a.
temporal?	4	no	h. t. s.	yes	n. a.
people?	5	no	h. t. s.	yes	n. a.
by pedestrian?	6	no	h. t. s.	yes	n. a.
attractive?	7	no	h. t. s.	yes	n. a.

(a) Standard ordering of categories reflecting the original order of questions and response options in the survey.

		1	2	3	4
real photo?	1	**yes**	h. t. s.	**no**	n. a.
people?	**2**	no	h. t. s.	yes	n. a.
outdoors?	3	**yes**	h. t. s.	**no**	n. a.
daytime?	4	**day**	h. t. s. / twilight	**night**	n. a.
temporal?	5	no	h. t. s.	yes	n. a.
by pedestrian?	6	**yes**	h. t. s.	**no**	n. a.
attractive?	7	**yes**	h. t. s.	**no**	n. a.

(b) Purpose-oriented ordering of categories (alterations in bold). Answers ordered so that photographs showing attractive walkable areas are represented with a straight vertical line on the left.

Figure 4: Survey response grids.

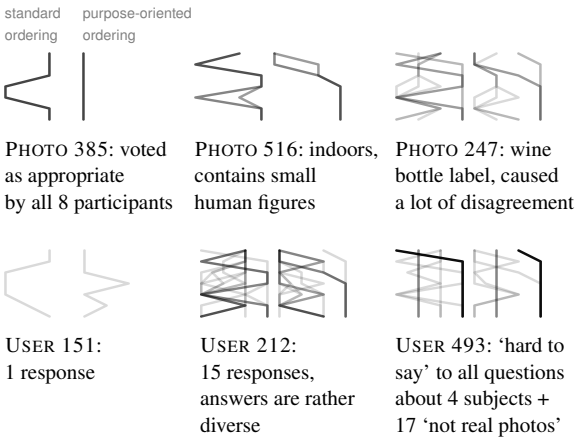


Figure 5: Glyphs representing responses grouped by photographs (samples) and users (participants). Opacity of a single line is 15%. Vivid patterns in groups denote a high level of agreement among responses.

Although there are fewer columns than rows, we choose to scale them to a square aspect ratio, resulting in a small horizontal stretching that eases discrimination between response patterns. This becomes less necessary when the semantic distance between categories is smaller. We set the opacity of the line representing a single response to be the minimum possible value that distinguished it from the background, 15% in our case. Because alpha composition is not linear [PD84], this value allows us to differentiate between up to 10–20 responses stacked on top of each other by the brightness of the composite lines. The need to detect line brightness placed a limitation on the minimum line width – the antialising of thin lines interfered with the line opacity. Equally a maximum line width was imposed by the need to minimise overlapping areas at line intersections. We therefore scaled line width between these bounds according to the size of the glyph area being displayed.

Where lines show repeated similar responses made by participants (*standard ordering*), they became a useful indicator of possible response insincerity. If a survey participant clicked on all available controls at the same position,

for example in an attempt to skip through the questionnaire rapidly, their actions are represented as vertical lines, which become darker for persistent behaviour. The fact that in our survey the answer 'no' to the first question disabled the remaining ones allowed participants to proceed to subsequent photographs even faster, in only two clicks. Such behaviour was also easily detectable by examining the glyph patterns. Because six of seven questions become 'n.a.' when an image is classified as not a photograph, the response becomes depicted as a ⅂ shape. A darker line with this shape is a clear indicator of unwanted behaviour that required removal from the sample. User 493 (Figure 5) shows such an example of insincerity: after giving a few possibly considered responses, he or she answered 'hard to say' to all questions as a means of proceeding to the next photograph quickly.

The main advantage of the purpose-oriented ordering over the standard one was to support selection of the photographs most suited to characterising urban attractiveness. With a universal and straightforward rule, which in our case is 'the further from the left the less suitable', this glyph type allows immediate detection of whether the photograph is good for our research purpose and estimating the degree of its unsuitability. This is clearly demonstrated in Figure 5 (photos 385 and 516).

After doing a number of experiments with the design of the glyphs we have not found their further simplification possible. For instance, replacement of lines with a grid of cells coloured by the numbers of answers, or use of micro bar charts, would cause a significant reduction of information shown, in particular loss of links connecting matching frequencies of answers to neighbouring questions. Such glyph designs would be less useful for the exploratory analysis of the categorical survey data.

Time Scaling When the amount of time it takes a participant to complete all answers about each subject is recorded, it is possible to merge this response attribute into a glyph. This revealed participants that are unusually quick, and a comparison of the overall difficulty respondents had in classification of different photographs. We propose two glyph designs incorporating the duration of the responses (Figure 6).

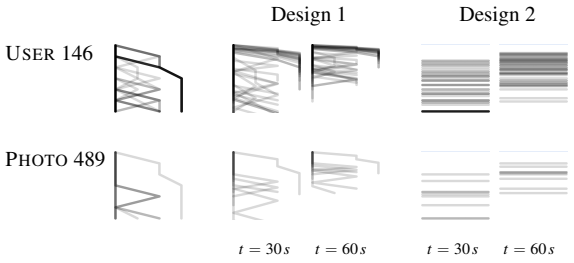
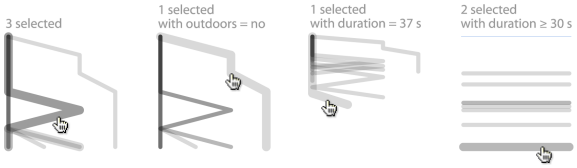


Figure 6: Time scaling of the response glyphs.



Mouse over an edge | Mouse over a node | Time mode 1 ($60\,s$) | Time mode 2 ($30\,s$)

Figure 7: Glyph interactions with mouse (Photo 489).

The first design was based on the idea of scaling lines vertically according to the response time of users to each question. The glyph in this case has similarities to the *DriftWeed* visual metaphor proposed by [RW00]. With a given time window t as a parameter the responses completed in t seconds have a height equal to a 100% of the glyph size, others are shorter or longer. When there are no data about how much time it takes a participant to move each control individually, the distortion of each shape is evenly distributed. This scaling reveals some expected patterns like the correlation between the number of available questions and the response duration. However the glyphs containing many responses become too cluttered and hard to interpret, so this does not provide a scalable solution.

The second design showed the durations of the responses with horizontal lines and thus displaying only one attribute per response in a glyph. In this case a baseline on top of the glyph may be added to serve as a zero reference point and being coloured differently to the responses. Despite the simplicity of the view it allows visual estimation of the complexity of the classification of the subjects and also the differences in the performance among the participants.

When an analyst is interested in the overall, mean, median, maximum or minimum response time for all responses in a group, a standard glyph may be combined with a coloured background encoding this attribute. As an alternative a coloured horizontal bar can be attached to the bottom of a glyph and encode the aggregated time parameter with length. Although the length is not preattentive as colour, it does allow more accurate comparison to be made [CM84].

Interaction When the glyphs are drawn in a software environment, they can be made responsive to mouse hovers and clicks (Figure 7). Mouse movements over the lines highlight the context and show additional information on demand. Clicking on a line in a glyph where the responses are grouped by subject may open the details of a corresponding person in the paired view; the opposite applies to the responses grouped by participants. If there is more than one participant or subject at a chosen edge or node, further clicking iterates through each of them allowing navigation to every entity linked to the glyph.

Animation To make the interchanges between glyphs smooth and clear, animated transitions between their states

[HR07] were added. This allowed tracking of individual responses when toggling between various glyph designs as well as whether different groups of responses shared the same subjects or participants.

4.3. Aggregated Lists of Responses

Because the glyphs are distinguishable even at small sizes, it is possible to display a large number of them on a single screen to aid comparison. Laying them out as a grid with cells of 20 by 20 pixels in row-prime order, similar neighbours may be compared while also scanning the list vertically for broader scale changes in response patterns. Expanding an item into a larger interactive glyph can be done by clicking on a glyph directly. Double clicking toggles between full and filtered grids; the filtered group excluding those responses previously rejected as insincere or though unavailability of a photograph. The maximum number of glyphs that can be displayed on a screen depends on the resolution, the amount of space left for bottom row with other data and also the proportion between counts of subjects and participants. For example, given a 17″/1,280×1,024 screen and 250 pixels in height for the bottom row, there is enough space for about 2,500 glyphs. Scroll bars are introduced if more space is required (Figure 1).

Placement plays an important role in analysis of data represented using glyphs [War02], so we pay extra attention to this feature. We suggest using customisable multilevel sorting (Figure 8) by both variables derived from the responses and attributes of entities related to the groups. For instance, we sort both participants and subjects by the amount of agreement between the responses and adopt two statistical measures for this purpose: Fleiss' Kappa [FC73] and entropy [TW06]. Entropy was found to be more robust due to some limitations of the Fleiss' Kappa [Pow12]. To improve the reliability of 'agreement' we weight the values based on numbers of available categories in each question. For instance, in our particular case 'n. a.' in six out of seven answers in a single response does not add 'strength' to the agreement if the same response pattern is repeated, due to the conditional nature of the remaining questions being contingent on the first response being other than 'n.a.'. The meaning of the agreement between the responses is different for groups with participants and subjects. If a participant has too many answers with the same profile so that the measure

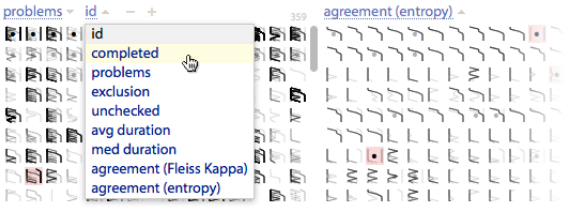


Figure 8: Multilevel sorting of the entity lists.

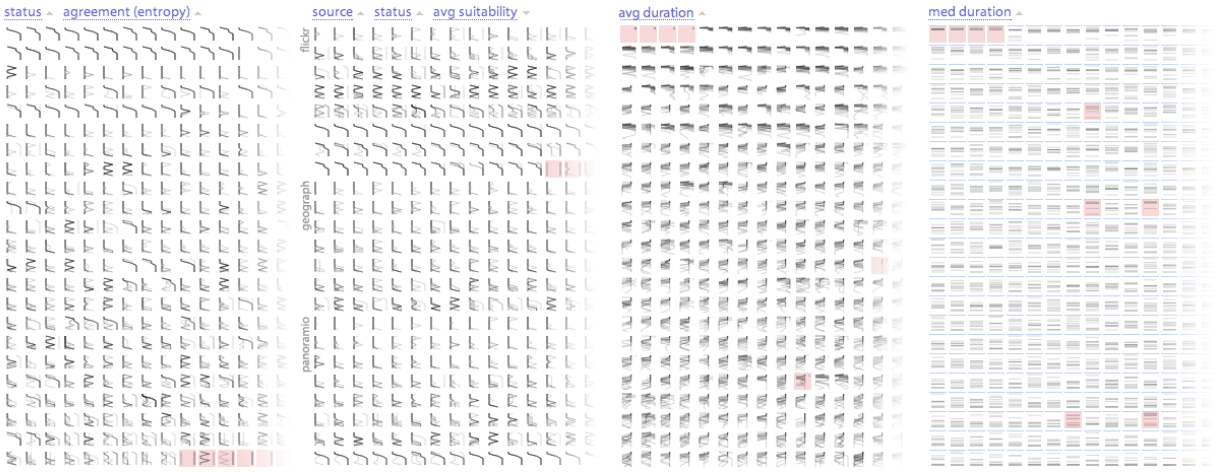
Figure 9: Cross-highlighting relationships between groups.

of agreement is high, this is a sign of possible insincerity. Equally, a high value of agreement between the responses in a context of subjects is what researchers may be expecting from the survey results meaning that entities are easy to classify.

A useful sorting method for subjects is by their overall score if numeric values can be assigned to all categories in all questions, from which totals or mean values may be derived. We can do this for our survey when using the purpose-oriented grid. With such sorting the photographs that describe attractive places for a leisure walking ( ) are gathered towards the beginning of the list, and the least appropriate ones ( ) are toward the end. Similar sorting can be done by scores for individual questions or groups of questions depending on the task. It is also possible to sort elements by frequencies of the same answers if sets of categories amongst all questions are equal. This may provide additional help in detecting insincerity. Placing similar elements next to each other simplifies the cognitive task. Searching for presence of differences between two subjects or estimating proportions of clusters of subjects becomes easier as identical glyphs form textures [War02], the boundaries of which are quick to detect.

The number of sorting modes can be increased significantly depending on the dataset that is currently loaded into the system. Our case counts 9 and 28 sorting modes for the participants and subjects, respectively. Some of the orderings are dedicated to specific tasks. For instance, we sort photographs by location using the Treemap algorithm [EvKSS13], greenness score, luminance derived from EXIF, timestamp, source, time from noon, etc.

The patterns of the responses are not the only characteristic of subjects and participants that can be visualized in the lists – a number of additional symbols may be used for other attributes of the data. If the chosen symbols do not interfere with the contents of the glyphs and are compatible with each other, they can be used simultaneously, increasing the richness of the visualization. For instance, we can combine glyphs with the colour of the item background to identify rejected participants and removed photographs (subjects). A bullet point can be added in middle of an item on demand if at least one response has not been given by a user due to a photo service API problem (i.e. when it was not possible to load a photograph, so it was skipped). If a problem

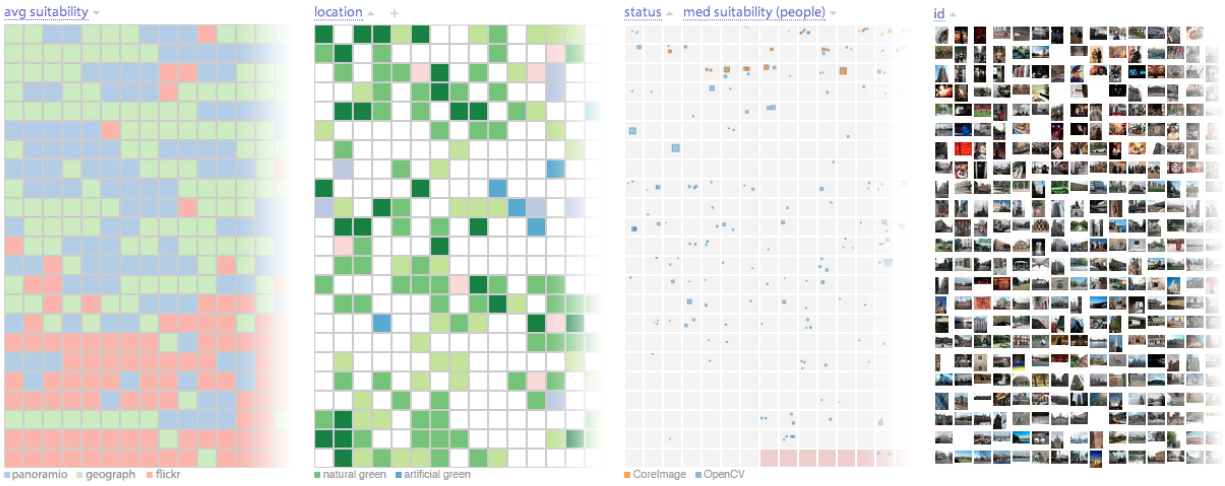


(a) Purpose-oriented grid glyphs sorted by the amount of agreement between responses.

(b) Same glyphs ordered by the source of photographs, their status and the average suitability.

(c) Glyphs drawn in time mode 1 (30 *s*, standard grid) and sorted by the average response duration.

(d) Glyphs drawn in time mode 2 (30 *s*) and sorted the median response duration.



(e) Colouring of items by source and sorting by the mean of suitability.

(f) Colouring items by the greenness score and ordering geographically.

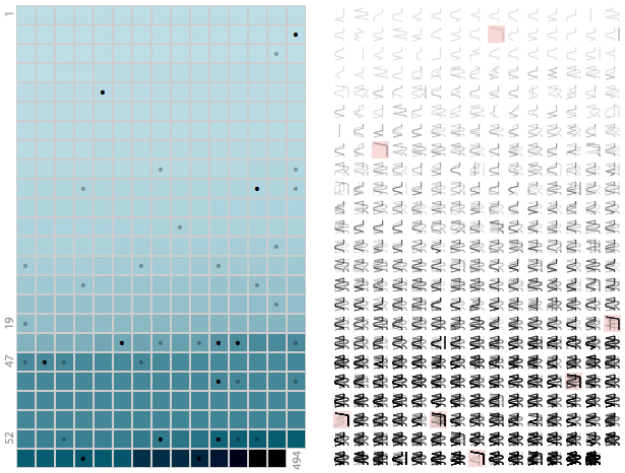
(g) Use of task-specific glyphs with face rectangles detected by CoreImage and OpenCV.

(h) Photo previews.

Figure 10: The list of subjects with different representations and orderings applied. Not all elements are shown for compactness.

takes place more than once, the mark becomes darker • . Grey border shading ☐ can be used to highlight unchecked participants – those who are new or who have given more responses since the last confirmation of their validity. Enabling this border helps in searching for new instances of insincerity; in our survey analysis tool it is removed automatically after a participant is selected for two seconds or their status is set to 'rejected'. A black border ☐ is used to highlight the currently chosen element, linking the list with an enlarged interactive glyph below it. A grey border ☐ appears around a number of elements when corresponding participants or subjects are selected in another view. For instance, hovering over a set of responses in a currently shown subject glyph highlights all participants these responses are made by. Similarly, hovering over a user in the list on the left highlights all subjects he or she has contributed to and vice versa (Figure 9).

We found replacing glyphs with a solid single-coloured fill useful for representing some values featuring each group



(a) Elements coloured by the number of completed responses; • – photoservice API faults.

(b) Standard grid glyphs; red cells represent excluded insincere participants.

Figure 11: The list of survey participants ordered by the number of completed responses.

(e.g. a number of completed responses). Having various encodings for different attributes, we can toggle between them and the glyphs to get more knowledge about the data.

Response glyphs can be replaced with other types of glyphs featuring the subjects of a survey. For instance, in our research it was necessary to perform automated and manual face detection for each photograph in the sample and then to compare the output with the subjective data obtained from the participants. Using another design for the glyphs we could show areas marked as faces on all photographs in a very compact from: . By ordering the subjects by the score on a question about faces and looking at these glyphs we could find the most appropriate face detection algorithm.

Finally, the glyphs can be converted into standard small multiples [Tuf83] and be used to, for instance, show the photographs themselves. This helped in navigating through the items when an analyst wished to return to the one recently viewed.

Some examples of visual representations and orderings for both participants and subjects are given in Figures 10 and 11. Although we cannot list all views and findings in the paper, we make the visualization available online for exploration: photoassessment.org/results.

5. Evaluation example: Measuring subject relevance

In order to evaluate our approach, we applied the idea of response glyphs to two other surveys with different structures. The first one [Zih13] is also related to photographs, but the images here were a bundled into groups of two to five, called *queries* that were examined by information retrieval domain experts. Participants were asked to classify each image as *irrelevant, marginally relevant, fairly relevant* or *highly relevant* to a defined subject. Each query contained a control photograph, which was not related to any subject by design.

The structure of this survey suggested using queries as subjects populating the rows of a response grid with photographs. Columns start with *don't know* (nominal value), followed by *irrelevant – highly relevant* (ordinal range). Thus, the size of the grid is from 2×5 to 5×5 (response lines are made shorter when needed). Figure 12 shows a screenshot of the adoption.

Visualising all 4,193 responses (19,931 answers to questions in total) allowed us to detect 34 individuals providing insincere responses among 332 participants; to evaluate the coverage of queries with answers and to explore diversity in opinion response. It confirmed the approach is still applicable to grids of five categories and ≈ 30 responses per subject. Different visual support for a glyph was also revealed: thin gridlines make it possible to decode positions of the bends as well as to distinguish between nominal and ordinal values. The implementation is available at ee1.photoassessment.org.

6. Evaluation example: Wikipedia Article Rankings

We also chose the anonymised dump of eleven million article ratings collected over one year (July 2011 – July 2012) from the English Wikipedia [Dat12]. This dataset was selected to examine the scalability of the approach.

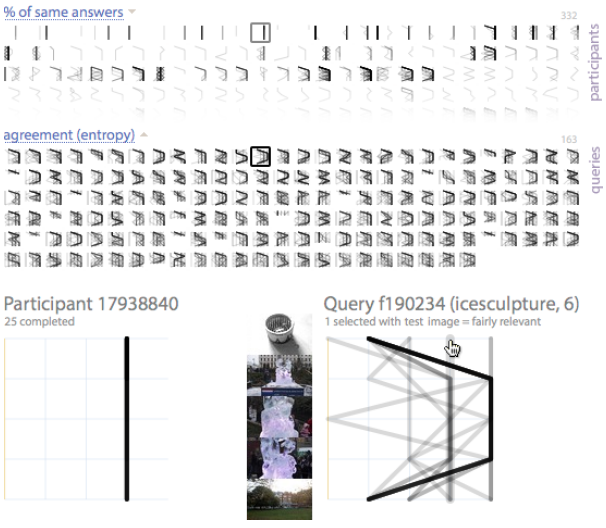


Figure 12: Most participants found the test image in *query f190234* irrelevant as expected. This also applies to the last photograph. Three other images are believed to be relevant, which is confirmed by the previews. Hovering over *test image* × *fairly relevant* reveals an insincere participant.

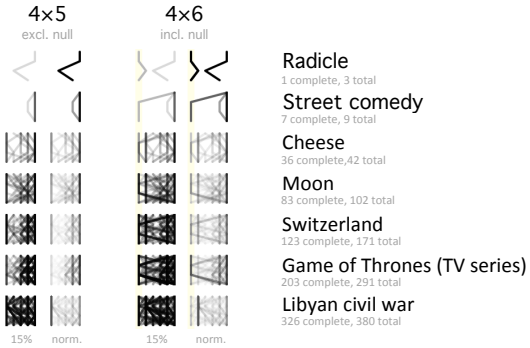


Figure 13: Examples of glyphs representing rankings of Wikipedia articles.

A corresponding form was deployed on 10% of articles, asking how a page was *trustworthy, objective, complete* and *well-written* [Med12]. Each parameter could be given a score from one to five (a typical Likert scale); it was allowed to skip some rankings. Such structure suggests a grid with five columns (the number of scores) and four rows (the number of dimensions); the sixth column may be prepended to show skipped rankings. The visualization of the responses for a random subset of 1,000 articles, where all four rankings a given, can be viewed at ee2.photoassessment.org. Figure 13 demonstrates some examples.

The purpose of this evaluation example was to identify the scalability of the glyph-based approach. When the distribution of responses by subjects is not balanced and the numbers differ significantly, glyphs with tens and hundreds of lines become overloaded, while the ones that contain only a few responses remain hardly seen. This may be partially compensated with normalisation of line opacity so that segments representing the same *proportion* of responses are depicted with similar brightness levels. Despite that normalisation still does not allow identification of individual response in glyphs with over 50 responses, common trends become salient. For example, a number of glyphs in this dataset can be characterised with two clear vertical lines on both sides. This suggests that responses are rather opposed: a significant proportion of readers find the same article extremely good or bad by all four parameters. This could not be observed if answers within responses were not linked – it would not be clear whether the majority of participants voting one star for objectiveness also vote one star for completeness. Adding incomplete responses to the glyphs demonstrates that it is not always possible to combine ordinal scale with a nominal value in the grid – too many line segments going to the edge of a glyph and then returning back to the middle of it clutter the view.

7. Conclusions

We found the combination of glyphs with parallel coordinate plots, linked views, interaction and animation a powerful instrument for navigating through the amounts of responses to some types of surveys. The visualization approach has helped us to conduct our own research, to clean the collected data and to test a number of task-specific hypotheses. A detailed look at all responses adds more confidence about the reliability of the survey results and also suggests possible solutions if there are any problems with some of them.

The same visualization tool can be applied for both looking at the collected data after the survey is complete and for supporting the process of its collection. When the visualization is supplemented with additional features and shows some parameters of the subjects of the research, it is possible to instantly compare them with the answers to the survey and to gain some insights (e.g. what spatial or temporal patterns exist in the data).

We identified a number of limitations with the approach. For instance, the glyphs inherit the scalability issues of parallel coordinate plots; it is not always possible to load the entire dataset into memory; the correct choice of the grid layout may be a challenge. We see a number of areas where future improvement can be made. For example, colour may be introduced into the glyphs to make it possible to distinguish between spikes related to different questions if such feature is needed (e.g. ⊢⊢ instead of ⊢⊢). The grid can be made more dynamic: columns and rows may become swappable on fly to make it easier to focus on answers to a particular question. Another depiction to the nominal value (usually meaning *n.a.*) may be applied – the edges sharing such answer may be represented with a thinner line rather than be diverted via an edge column on the grid. Varying widths of the lines can also be used in the larger versions of the glyphs to make nodes more distinguishable from links when multiple segments of one line often share the same angle. 'Gum' or 'weighted connections' techniques [Nag12] may be applied here. Additionally, responses within glyphs can be made filterable, for instance, to be able to reject only some of them when a participant starts behaving insincerely after giving a number of deliberate answers.

The source code of our implementation as well as the 'photo content assessment' survey dataset are open and free to use: github.com/kachkaev/survey-glyphs.

8. Acknowledgements

This work would not be possible without volunteers who agreed to spend their time on photograph assessment. We would like to thank all of them for participating in the studies. We also wish to acknowledge Ross Purves and Oliver Zihler (University of Zurich) who kindly provided their dataset on measuring subject relevance (Section 5). Finally, we express gratitude to Wikimedia Foundation for organising feedback collection and people who shared their opinion on the quality of Wikipedia articles (Section 6).

References

[Ame07] AMEGASHIE J. A.: *Intentions, Insincerity, and Prosocial Behavior*. Working Papers 0703, University of Guelph, Department of Economics and Finance, 2007. 2

[AS05] ABRAMOVICH S., SUGDEN S.: Spreadsheet conditional formatting: An untapped resource for mathematics education. *Spreadsheets in Education (eJSiE) 1*, 2 (2005), 3. 2

[BKC*12] BORGO R., KEHRER J., CHUNG D. H., MAGUIRE E., LARAMEE R. S., HAUSER H., WARD M., CHEN M.: Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics 2013-State of the Art Reports* (2012), The Eurographics Association, pp. 39–63. 2

[BN11] BRANDES U., NICK B.: Asymmetric relations in longitudinal social networks. *Visualization and Computer Graphics, IEEE Transactions on 17*, 12 (2011), 2283–2290. 2

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79*, 387 (1984), 531–554. 5

[CM07] CAWTHON N., MOERE A. V.: The effect of aesthetic on the usability of data visualization. In *Information Visualization, IV'07. 11th International Conference* (2007), p. 637–648. 4

[Dat12] DATAHUB: Wikipedia article ratings. http://bit.ly/wikipedia_article_ratings, 2012. 8

[EvKSS13] EPPSTEIN D., VAN KREVELD D., SPECKMANN B., STAALS F.: Improved grid map layout by point set matching. In *Visualization Symposium (PacificVis), 2013 IEEE Pacific* (2013), IEEE, pp. 25–32. 6

[FC73] FLEISS J. L., COHEN J.: The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement 33*, 3 (Oct. 1973), 613–619. 6

[Few07] FEW S.: Save the pies for dessert. Visual business intelligence newsletter. http://www.perceptualedge.com/articles/08-21-07.pdf, 2007. 2

[Fry08] FRY B.: *Visualizing data: exploring and explaining data with the Processing environment*. O'Reilly Media, Inc., 2008. 3

[GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization 10*, 4 (oct 2011), 289–309. 2

[HR07] HEER J., ROBERTSON G. G.: Animated transitions in statistical data graphics. *Visualization and Computer Graphics, IEEE Transactions on 13*, 6 (2007), 1240–1247. 4, 6

[ID90] INSELBERG A., DIMSDALE B.: Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *Proceedings of the 1st conference on Visualization '90* (Los Alamitos, CA, USA, 1990), VIS '90, IEEE Computer Society Press, p. 361–378. 2

[LKH09] LIE A. E., KEHRER J., HAUSER H.: Critical design and realization aspects of glyph-based 3d data visualization. In *Proceedings of the 25th Spring Conference on Computer Graphics* (2009), ACM, pp. 19–26. 2

[LSR10] LSR ONLINE: Place survey. http://www.lsr-online.org/placesurvey.html, 2010. 2

[Med12] MEDIAWIKI: Article feedback / version 4. http://www.mediawiki.org/wiki/Article_feedback/Version_4, 2012. 9

[Mic10] MICROSOFT: Quick start: Apply conditional formatting. http://office.microsoft.com/en-001/excel-help/quick-start-apply-conditional-formatting-HA010370614.aspx, 2010. 2

[MRSS*12] MAGUIRE E., ROCCA-SERRA P., SANSONE S.-A., DAVIES J., CHEN M.: Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments. *Visualization and Computer Graphics, IEEE Transactions on 18*, 12 (2012), 2603–2612. 2

[Nag12] NAGEL T.: Gum and weighted connections. http://tillnagel.com/2012/11/gum-and-weighted-connections/, 2012. 9

[O'B13] O'BRIEN, O.: This place, visualisation of 2011 census data for england and wales. http://casa.oobrien.com/thisplace/, 2013. 2

[Off11] OFFICE FOR NATIONAL STATISTICS: 2011 census interactive. http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-interactive-content/index.html, 2011. 2

[PD84] PORTER T., DUFF T.: Compositing digital images. *ACM Siggraph Computer Graphics 18*, 3 (1984), 253–259. 4

[Pow12] POWERS D. M.: The problem with kappa. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics* (2012), p. 345–355. 6

[RC94] RAO R., CARD S. K.: The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1994), ACM, pp. 318–322. 2

[RW00] ROSE S. J., WONG P. C.: Driftweed: a visual metaphor for interactive analysis of multivariate data. In *Electronic Imaging* (2000), International Society for Optics and Photonics, pp. 114–121. 5

[Sap99] SAPSFORD R.: *Survey research*. Sage Publications, London ; Thousand Oaks, Calif, 1999. 2

[Tuf83] TUFTE E.: *The visual display of quantitative information*. No. v. 914 in The Visual Display of Quantitative Information. Graphics Press, 1983. 2, 3, 8

[TW06] TASTLE W., WIERMAN M.: An information theoretic measure for the evaluation of ordinal scale data. *Behaviour Research Methods 38*, 3 (2006), 487–494. 2, 6

[Uni13] UNITED STATES CENSUS BUREAU: Data visualization gallery. http://www.census.gov/dataviz/, 2013. 2

[War02] WARD M. O.: A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization 1*, 3-4 (2002), 194–210. 2, 6

[War08] WARD M. O.: Multivariate data glyphs: Principles and practice. In *Handbook of data visualization*. Springer, 2008, pp. 179–198. 2

[WHWC12] WICKHAM H., HOFMANN H., WICKHAM C., COOK D.: Glyph-maps for visually exploring temporal patterns in climate data and models. *Environmetrics 23*, 5 (Aug. 2012), 382–393. 2

[WS12] WICKHAM H., STRYJEWSKI L.: *40 years of boxplots*. Tech. rep., had.co.nz, 2012. 2

[WSD*13] WALKER R., SLINGSBY A., DYKES J., XU K., WOOD J., NGUYEN P. H., STEPHENS D., WONG B., ZHENG Y.: An extensible framework for provenance in human terrain visual analytics. *Visualization and Computer Graphics, IEEE Transactions on 19*, 12 (2013), 2139–2148. 3

[WW05] WILKINSON L., WILLS G.: *The Grammar of Graphics*. Statistics and Computing. Springer, 2005. 2

[Zih13] ZIHLER O. F. M.: *Implementation of a Spatially-Aware Image Search Engine and Its Evaluation using Crowdsourced Relevance Judgements*. Master's thesis, Department of Geography, University of Zurich, 2013. 8