Possibilities of Human Data Embodiment: *100% City*

Jörn Hurtienne
University of Würzburg

Abstract— In this paper we explore the use of living humans to represent statistical data as a new form of data physicalisation. In a case study, we analyse data representation formats and consider strategies for how human data embodiment can be used to "give statistics a face". This is done at the example of *100% City*, a performance directed by Rimini Protokoll, a German documentary theatre group. In this, 100 people statistically represent the population of a whole city and answer questions about their demographic status, politics, work, family, health and more. The results encourage further research into what we might call Computer Augmented Human Data Embodiment (CAHDE). It may also inspire traditional data physicalisations in how data about humans can be made more personally relevant to their audiences.

Index Terms— data physicalisation, visualisation, information visualization, statistical representation

Introduction

Data Physicalisation as a recent offspring of information visualization is "a research area that examines how computer-supported physical representations of data (i.e., physicalisations) can support cognition, communication, learning, problem solving, and decision making." ([10], p. 3230). Thus, data physicalisations, unlike most visualisations, have more dimensions available that can be used to code data, to deliver a more direct experience, and to add interpretational spice. These dimensions include the materials used, their haptic properties like texture, temperature and firmness, their permanent availability, their inherent three-dimensionality, and, most interesting, the implicit meaning learned through prior contextualised encounters with the material.

A specific material for data physicalisations with a rich potential for implicit and explicit meaning is the living human body. In this paper we analyse the expressive capabilities of representing data with people. This is done on the basis of a case study of a data performance by the theatre group Rimini Protokoll. In their performance *100% City* they bring 100 people of one city on stage that are selected to be statistically representative of the city's population. The central proposition of Rimini Protokoll is "to give statistics a face" to counteract notions of official statistics where families become 'households', children 'expenditures', and incomes 'average' [2]. Of interest here is the style of how *100% City* is performed: how can demographic statistics be shown in a way that is informative about a city's population and at the same time made personally relevant to the audience?

1 People as Data Points

The *100% City* series of performances started 2008 with *100% Berlin*, a statistical documentary involving 100 Berliners that were selected to represent the then 3.4 million population of Berlin in Germany. Each person on stage stood for 34,000 Berliners. They were sought out according to Berlin's composition of age bands, gender, living quarters and national origin. Since then *100% City* has been travelling around the world with adaptations made to each of the cities it has been performed, for example, in Vienna (2010), Melbourne (2012), Tokyo (2013), Philadelphia (2014), Yogyakarta (2015), and Marseille (2017). The recruitment is done through a chain system where one person recruits another within 24 hours. Recruitment follows specific criteria of age, gender, marital status, city neighbourhood and ethnicity in an attempt to reflect the demographics of the last census of the respective city. In contrast to the usual demographic surveys, participants are thus not a random sample of disconnected individuals but ideally form a representative chain of people that each at least know their recruiter and the person they recruited. This brings in a first element of personalisation and is made explicit in the subtitle, e.g. *100% Berlin: A Statistical Chain Reaction*. In the accompanying programme, each person is introduced with a picture, some personal details and how they are connected to the previous or next person in the chain.

2 Formats of Data (Re)Presentation

Like in a normal theatre performance the audience faces the stage. With 100 persons on stage, however, only the first row(s) of them are clearly visible so that an additional (over)view becomes necessary. Above the stage is a circular screen that provides a bird's-eye-view of all actors on stage (Figure 1). The circular screen corresponds to the circular form of the revolving stage on which most of the performance takes place. It conveys a 'Petri-dish' feeling that serves to enhance the analytic dimension of the performance. The demographic sample (much like a biological sample in a Petri dish) can be scrutinised by the audience and subjected to "chemical reactions" triggered by different questions and response formats the performers are exposed to [1,13]. The views of the sample in the Petri dish are sometimes enhanced by digitally superimposing quantitative scales or a city map (Figures 2, 3 and 8). Like under a microscope, this display also shows close-ups of single performers as they tell individual stories or ask a question (see [13] for a detailed analysis). The display is also used to display questions and the legend for Multiple Choice questions (see below and Figures 5 and 7).

At the beginning of the performance, people line up, one after the other, on the slowly revolving stage. Everyone introduces themselves within 1/100 of a revolution of the stage. One by one a circular diagram develops that is only visible in the Petri dish (Figure 2) while the new arrivals are better visible from the front. Here, the Petri dish representation acts as a circular progress indicator. Like a pie chart it shows how many people already have introduced themselves and, roughly, how many are still to come. The circumference of the circle is closed when 100 people are on the stage.

The next format represents the age distribution of the participants (Target display, Figure 3). For each year they lived, participant move a step towards the centre of the stage. At the top screen the resulting age distribution is overlaid by a ring diagram resembling a target

• *Jörn Hurtienne is head of the chair of Psychological Ergonomics at the University of Würzburg. E-mail: hurtienne@acm.org.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.
Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx/.

where the bullseye represents 100 years. The Target display is also used to represent other forms of scalable quantitative data, for example, household size.

Altogether, the performers respond to about 100 questions in different formats. In one format, performers gather under either of two surtitles: “me” and “not me”, according to their response to diverse questions like: Who was born in our city? Who is a member of a political party? Who believes in a god? (Figure 4). For responding to the next question, people would stay or move sides according to their responses. For some political questions people are filtered according whether they vote or not. Then non-voters have to leave the stage. In a different response format, performers stay in a fixed position and just raise their hands when they want to answer questions in the affirmative (“Show of Hands”).

Another format are multiple choice questions. The performers are seated and hold up large colour plates were the colours represent different answer categories. The top screen shows the questions and the accompanying legend to their answers (Figure 5). Holding up large plates is also conducive to short open answer formats in which performers can indicate, for example, the “country of my origin” or their “favourite Australian” (Figure 6). For the audience to be able to digest the variety of answers, performers need to move about, sometimes leaving the stage and walking into the audience.

In another format, performers move “Into the Spotlight” when a shown statement applies to them (Figure 7). Example statements include “We have children.” “We are in favour of the death penalty” or “We were raised without a father.” These statements are projected onto the top screen and people gather at the centre of a stair to be better visible by the audience

Other forms of data representation include switching on torches on a dark stage (“Light in the Dark”) or distributing people in space, for example, on a superimposed “City Map” (Figure 8). Data can be also represented sequentially. The “Day-in the-Life” format, for example, shows a sped-up 24-hour cycle in which performers act out those activities that they would normally do at the different times of the day.

At first sight, these data representation formats appear to be simple. Data is mainly represented as frequencies. Only a small number of answer categories is available and different variables are rarely combined. However, these seemingly simple representations quickly turn out to be rather complex considering that the data is embodied by real people. The performers’ bodies “naturally” encode demographic variables like gender and age that are easy to discover and always co-present.

The saliency of other co-present variables depends on their discoverability. Social status, for example, may be judged from the clothing of performers, their equipment or the way they tell their stories during the performance. Over time, their history of previous responses conveys an impression of, for example, liberal or conservative political attitudes. While these other variables are always co-present, it is difficult for the audience to actually make use of these. Analysing gender distribution across answers, for example, is difficult, because constellations change very quickly and the whole group can only be seen from the top view where identification is difficult. Tracking political attitudes from previous answers is difficult, too, because this depends on being able to identify many performers individually and memorise their previous answers.

3 Strategies for Giving Statistics a Face

Each series of performances is meticulously planned to make the experience relevant for the audience. Rimini Protokoll include questions that refer to the history of the place or its current political and economic development. For example, the fall of the wall in Germany was highlighted in Berlin and issues of indigenous people were highlighted in Melbourne. For the audience, recruiting performers to be representative of their own city creates authenticity and enhances the personal significance of the show. These effects may be stronger in smaller cities, where it is more likely that audience members know someone on stage.

Fig. 1. Performers and the Petri dish display.

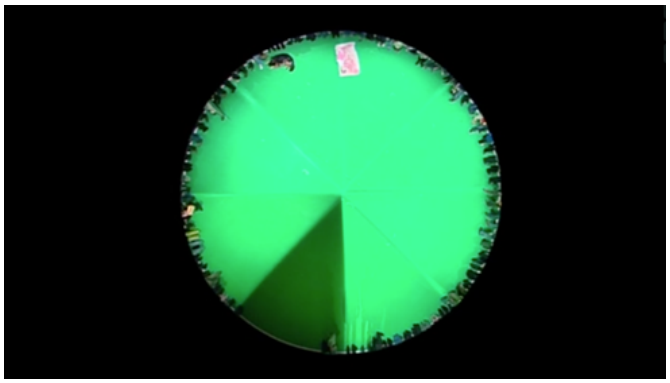

Fig. 1. The Petri dish as a Circular Progress Indicator.

Fig. 2. The Target Display.

Fig. 3. The Me/Not Me Display.

Fig. 4. Multiple Choice.

Fig. 5. Open Answers.

Fig. 6. Into the Spotlight.


Fig. 7. City Map.

The performance starts with 10-second introductions of each of the performers. They say their name, tell a fact about themselves, introduce a personal item they brought, and then introduce their relationship to the next person: "Hi, I am Cathy. I have 14 piercings and 3 tattoos. I am aspiring to be a body piercer. And this is my Chucky [puppet] and I brought my best friend and her daughter along." There is much room for creating personal significance: audience members form an impression based on performers' looks, what they say and do, and the personal items they bring. Very often, these introductions use humour for creating sympathy.

Between the questions, there are several opportunities for some performers to tell their personal stories in more detail: a war veteran talks about the difficult re-integration into civil society after returning home; a black woman tells a story about daily racism when working as a croupier at a casino; another man tells the story of his object, a 30-year-old piece of soap that has seen many places travelling with him around the world. In these stories performers often elaborate on their answer to the prior question or they prepare the ground for the next.

While most questions are pre-determined, the performance also includes interactive elements that allow for new and spontaneous questions. During the Me/Not Me format there is an "open mike" phase where performers can spontaneously ask questions to other performers. In another phase all performers line up at the edge of the stage and start asking questions to the audience. Then the audience asks questions to the performers. Several questions explicitly address the relation between audience and performers and encourage play: Who knows someone in the audience? Who is afraid of us? Who likes to take a picture of us?

Giving statistics a face (and a body) is taken literally in *100% City*. At first *100% City* appears to be an exercise in getting rough impressions of how the city population distributes between yes/no answers in a series of unrelated questions. Very soon, however, the observer can identify individual performers and starts to track them through the sequence of questions. The old lady from a more affluent neighbourhood confirms her conservative attitude in some answers, but in others surprises by being original and against-the-grain. The performance thus not only gives insights at the population level (e.g. whether more conservative or more liberal people would prefer the death penalty), it also reveals how demographic patterns break at the individual level: people are not what they seem. They cannot be put conveniently into generalised statistical categories. Of course, for the observer it is only possible to follow a few individuals. But this is enough for the effect to occur.

What makes people recognisable? Apart from gender and age, accessories and clothing are important. Some wear glasses; some their working clothes (a safety vest, a cooking hat, a school uniform); some wear a sporty, others a more formal outfit. People can be distinguished by their personal objects (photographs, skateboards, parachuting equipment). Again, the body plays a role in differentiating and recognising people: there are the tall and the small, the fat and the slim, the fit and the frail, the boring and the sexy.

Making the performance personally significant helps putting oneself into the shoes of those on stage. Some of the performers hesitate to answer; their faces are strained, decided, calm or sad when they answer certain questions. Performers can earn spontaneous bursts of applause from the audience when they showed courage and stood up for their values. For some of the more sensitive questions the lights are switched off and performers answered by lighting a torch without being visible themselves. With this simple trick, much more people admitted, e.g. to evade taxes and disclosed on other sensitive questions.

The dramaturgy of the event contributes to enhance emotional effects and makes the show more entertaining. Texts and stories of the performers are their own, but they have been selected for and revised by the directors to make them interesting, relevant and convincing [7]. The questions and their sequence appeal to the analytic mind at the beginning of the performance and strongly appeal to the emotions near the end when more personal and existential questions are asked.

Successive questions may contrast: "We have been victims of violence" followed by "We have committed acts of violence". They may increase in intensity: "We have suffered from depression" followed by "we have thought about taking our own life". The directors employ music that makes the movements on stage more entertaining to watch. Specific sounds are directly inspired from the city's soundscape, e.g. the underground railway in Berlin, pedestrian traffic lights in Melbourne. Some elements are about playfulness and entertainment. Performers move in a polonaise or perform local dances. They enter competitions: Who is the fastest runner? Who can do the most push-ups? In between the stories a roulette wheel rotates on the top screen to remind the audience that life is often determined by chance.

Altogether, events on stage are very lively, because of the planned moving of people between Me/Not Me questions, dance interludes, and small children and pets moving about on stage. In some performances people fool around, heckle, wave hands and pull faces. Commentators say: "here one has the impression of seeing democracy in action – a bit messy, unsynchronised, variable. It is even reassuring to see performers, whether out of ineptitude or resistance, consistently go off-program and use a fourth colour to answer a three colour question" ([5], p. 176). Thus, *100% City* makes the dropouts visible. On a higher level it also shows the limits of quantitative statistical surveys: we do not learn the reasons for the performers' answers, despite them telling their stories here and there.

4 Towards a Design Language of DataPhys

100% City can be seen as a recipe for what we might call Computer Augmented Human Data Embodiment (CAHDE) – a data physicalisation with live people. Several formats of data representations are possible. Classical forms of data visualisations are alluded to (e.g. the pie chart in Figure 2), but the analysis should have shown that CAHDE can find its own forms of data representations. These forms of representations are summarised in the Appendix to this paper (Table A.1).

One of the major benefits of CAHDE is "giving statistics a face", resulting in high emotional involvement of the audience. The analysis has shown that this goes beyond the mere mappings of variables and draws more on the performance aspects that relies on personal narratives, the use of music and the curation of the sequence in which the questions are presented. The role of the computer in these performances is subtle. It is used, for example, to prompt questions to the performers and to overlay special effects on the Petri dish display.

Human data embodiment, however, is a very costly method of data representation, and one that cannot be repeated at will. It would make sense for future research to not only explore human data embodiment further, but to also see whether the results can be transferred to classical forms of data physicalisation. The messiness, the playful, the engaging elements of human data embodiment may also appeal to other formats of data physicalisation. Some first ideas and possible classifications according to the effort required in turning from human data embodiment to data physicalisation with objects are also given in the Appendix of this paper.

Acknowledgments

Fig. 1 was taken from Rimini Protokoll's *100% Vienna* performance (https://www.rimini-protokoll.de/website/en/project/100-prozent-wien), Fig. 2 from *100% Berlin* (https://vimeo.com/40925638), Fig. 3-5, 7 and 8 from *100% Cork* (https://vimeo.com/72341849) and Fig. 6 from *100% Melbourne* (https://vimeo.com/155938916).

References

[1] Jennifer Bowen, Ulrike Garde, Anton Griffith, et al. 2015. Panel discussion 100% Melbourne: contemporary documentary performance that puts "real Melbournians" on the stage. In *Rimini Protokoll close-up: Lektüren*, Johannes Birgfeld, Ulrike Garde and Meg Mumford (eds.). Wehrhahn, Hannover, 302–317.

[2] Sebastian Breu. 2013. Statistics of Tokyo's 23 Wards. In *100% Tokyo*. Tokyo Metropolitan Theatre, Playhouse, Tokyo, 220–223.

[3] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann.

[4] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: Dynamic Physical Affordances and Constraints Through Shape and Object Actuation. *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, ACM, 417–426. http://doi.org/10.1145/2501988.2502032

[5] Ehren Fordyce. 2008. "We go live at 8 o'clock" Documentary theatre in the presence of performance art. In *Experts of the everyday: The theatre of Rimini Protokoll*, Miriam Dreysse and Florian Malzacher (eds.). Alexander Verlag, Berlin, 168–187.

[6] Gallup Europe. 2010. 5th European Working Conditions Survey, 2010 Technical Report.

[7] Ulrike Garde. 2015. Ein Spiel mit der Stadt : Die spielerische Durchbrechung von Schemata in 100% Melbourne. In *Rimini Protokoll close-up: Lektüren*, Johannes Birgfeld, Ulrike Garde and Meg Mumford (eds.). Wehrhahn, Hannover, 254–277.

[8] Trevor Hogan and Eva Hornecker. 2016. Towards a Design Space for Multisensory Data Representation. *Interacting with Computers*. http://doi.org/10.1093/iwc/iww015

[9] Yvonne Jansen, Pierre Dragicevic, and Jean-Daniel Fekete. 2013. Evaluating the Efficiency of Physical Visualizations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 2593–2602. http://doi.org/10.1145/2470654.2481359

[10] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, et al. 2015. Opportunities and Challenges for Data Physicalization. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ACM, 3227–3236. http://doi.org/10.1145/2702123.2702180

[11] Florian Malzacher. 2010. The Scripted Realities of Rimini Protokoll. In *Dramaturgy of the Real on the World Stage*, Carol Martin (ed.). Palgrave Macmillan UK, London, 80–87. Retrieved from http://dx.doi.org/10.1057/9780230251311_7

[12] Andrew Vande Moere. 2008. Beyond the Tyranny of the Pixel: Exploring the Physicality of Information Visualization. *2008 12th International Conference Information Visualisation*, 469–474. http://doi.org/10.1109/IV.2008.84

[13] Meg Mumford. 2015. 100% City and Popular Factual Television: A New Game Plan for Managing Proximity to People? In *Rimini Protokoll close-up: Lektüren*, Johannes Birgfeld, Ulrike Garde and Meg Mumford (eds.). Wehrhahn, Hannover, 278–301. Retrieved from http://hdl.handle.net/1959.14/1036735

APPENDIX: CAHDE DATA FORMATS AND POSSIBLE TRANSFER TO DATAPHYS WITH OBJECTS

Computer Augmented Human Data Embodiment (CAHDE) is a very costly method of data representation. Therefore, it makes sense to think about how the *100% City* format can inspire more traditional forms of data physicalisation, involving lifeless objects instead of living people. In the following we closely adhere to the *100% City* format. We assume a physicalisation with 100 objects that mimic the representation formats and strategies for personalisation found in the *100% City* performances. Trade-offs between strict adherence to the *100% City* formats and cost are discussed along the way. Designers might wish only to choose single elements for their own physicalisation projects.

Data Presentation Formats in DataPhys

First of all, *100% City* shows how it is possible to represent mass data with a relatively small number of data marks. Using a subsample of the data makes sense for data physicalisations that also need to be actuated. (For static representations using whole populations see *Of All the People in All the World* by Stan's Café: http://www.stanscafe.co.uk/project-of-all-the-people.html). The usual sample sizes of social surveys still tend to be too large for effective physicalisations *100% City* style. The *100%* format suggests to reduce the sample to just 100 data objects, each representing 1% of the population.

Second, representing statistics with objects instead of people can address the (few) shortcomings of the *100% City* format. Although an implicit co-representation of multiple variables is one of its strengths, the audience cannot really make use of it. The bird's-eye view is important for showing the distribution of participants, but only at the cost of making performers and their characteristics less easily identifiable. For the audience to combine distributions with the characteristics of individuals they would need more time to compare the front view on stage with the bird's-eye view above. Data physicalisations at a smaller scale can make both views accessible at the same time so that the distributions between genders across the answers, for example, are easily discernible by simple acts of active perception. Users could move their body or the physicalisation itself to bring the most interesting parts into view. In interactive physicalisations users can take time to compare and analyse, even to replay sequences of questions or jump between questions.

Third, how can 100 objects be made individually identifiable so that users can track these through a sequence of questions? One way is to look for ready-made objects that already differ in appearance. Various representations of people, for example, can be found in shops for toy train sets. More abstract objects are preferable when a more general applicability is needed. Jelly beans, for example, look more generic and differ in colour (combination), transparency, number of spots on the surface and taste. However, in many projects it will be necessary, that the objects are not just different, but that these differences are actually implicitly meaningful. Thus, most of the time, this will mean that designers prepare their own sets of data objects that differ in the required variables. These identifying variables preferably are "naturally" embodied (such as gender or age in toy train people) or they may be arbitrary signs (e.g. a triangle for "male", a circle for "female", different colours for different age groups). Such approaches may allow for very rich representations. In Figure A.1, for example, the changing colours along the length of each stick encode the importance of different variables across the lifespan (e.g. different job roles, qualifications, income, health status). A set of 100 such objects can be created from multivariate data where available and 3D-printed. These objects can be individually identified by their colour patterns and support rich multivariate comparisons.

Fourth, there are trade-offs between faithfully implementing the representation formats of *100% City* and the costs of doing so when using objects. Whether the objects need to be moved and whether users need to distinguish single objects are major drivers of cost. Table A.1 gives an overview. Comparatively easy to realise are physicalisations that do not need to move and need not be identifiable. This would be the format of *Of All the People in All the World* where one grain of rice represents one person. Only initial movement is necessary with the Circular Progress Bar. Next come physicalisations where un-identified objects do not need to move but can change, as in the "Lights in the Dark" representation. Fixed objects that require identification are a little more difficult to build when they also require local actuation. Formats like "Show of Hands" and "Multiple Choice" are in this category. The "City Map" format can be combined with "Lights in the Dark", "Show of Hands" or "Multiple Choice". When the objects remain fixed in one region, region can become an identifying variable, similar to gender and age in people.

All other representational formats require the objects to move, which may be costly to realise. How can individual objects be moved? The best solution is to use tiny robots (e.g. KiloBots, OzoBots), but orchestrating 100 of these quickly and simultaneously may still be a challenge. 100 objects could also be actuated mechanically, but again the effort is high. Another solution would be to use illusory movement across a number of fixed objects. Illusory movement can be achieved through apparent motion effects across LEDs or displays attached to these objects. A third solution would be to let users move the data objects. Although this reduces their efficiency and fluency, users may learn better and show a greater engagement with the statistical content. The instruction by the computer, however, needs to be designed carefully as not to reveal the answer before the user has finished arranging the objects. As long as individual objects do not need to be followed through a set of questions, any objects can be moved. As *100% City* needs identifiable objects this means to move specific objects, which may make this method too taxing for the user.

The most expensive representation formats are the "Day in the Life" and the "Open Answers" format, as these potentially include all three drivers of effort: moving objects through space, actuating objects in individual ways and providing identifying features. Such representations would be more easily created with a GUI than with physicalisations.

Fifth, even more effort is necessary when any spontaneity in answering questions should be retained. It might be possible to have a live panel of online participants, but it is very likely that representativeness will suffer and/or costs will increase disproportionately. Cost might be less when questions can be answered with delays of several hours or days. Another variant, when the data physicalisation is made for a large audience, is to ask these questions directly to other members of the audience (possibly at the cost of statistical representativeness).

Giving Statistics a Face

Many strategies in giving statistics a face include a combination of quantitative and qualitative data. In *100% City* performers give short introductions of themselves, they bring personal objects and some tell stories hinting at the reasons of their choices. Physicalisations with objects are often based on archived data. Qualitative data may be already available in the dataset when open-ended questions were asked or reasons for choices were obtained. In the physicalisations, qualitative data can be shown as text or read aloud by the system at the user's request. Qualitative data could also come from a different study, for example when complementing quantitative survey data with qualitative interview studies. Then the qualitative data cannot be tied to identifiable data objects in the physicalisation. Using photographs conveys a more personal touch – whether these are actual or substitute photographs that are shown to represent the variety in the sample (much as photos of personas in user-centred design do). Similarly, the objects performers brought to *100% City* ranged from toy pets, photographs, awards and heirlooms to specific items of clothing, jewellery, and musical instruments. Apart from helping to identify people, most of these had no further use in the show. In physicalisations it may be interesting to use topic-related objects that could be derived from the available database (e.g. when the profession of people is translated into physical symbols of their typical utensils).

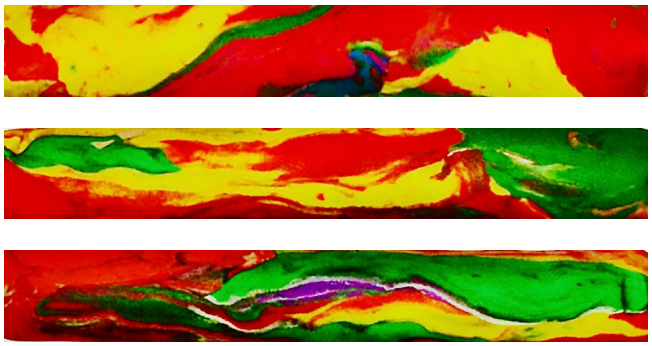

Figure A.1. Suggestion for multivariate data objects to be used in physicalisations with artefacts. Different colors show different variables. The length of the object codes progressing time (e.g. biographical time). Each cross-section at a specific time as well as the whole object are unique and individual.

Format	Does object need to move?	Need for indivi-dual objects?	Local actuation needed?	Live Data possible?
Stan's Café	No	No	No	No
Circular Progress Bar	Only initially	No	No	Yes
Lights in the Dark	No	No	Yes	Yes
Show of Hands	No	Yes	Yes	Yes
Multiple Choice	No	Yes	Yes	Yes
City Map	No	Yes	Yes	Yes
Target Display	Yes	No	No	Yes
Me/Not Me	Yes	Yes	No	Yes
Into the Spotlight	Yes	Yes	No	Yes
Open Answers	Yes	Yes	Yes	Yes
Day in the Life	Yes	Yes	Yes	No (sped up)

Table 1. *Stans Café* and *100% City* representation formats ordered with increasing effort of realisation when done with objects instead of people.

Another important feature in giving statistics a face is the possibility to identify individuals or groups and to follow these around through questions. As discussed above, data objects can encode different variables or display other features that make them identifiable. Identification is not only interesting for casual audiences, but may also prove valuable for professional users when doing exploratory data analysis and generating new hypotheses.

Other strategies of 100% City involve curating the questions to better achieve a fit between audience and data as well as to sequence and contrast questions for greater insight. This can also be done in physicalisations with objects. Users can encounter a fixed sequence of questions or may be able to select their own questions. More freedom of choice is probably needed for the professional user than for a casual audience.

Bidirectional questioning is another strategy: users can ask their questions spontaneously (or spontaneously select from pre-determined questions) and the system can also ask the user some of these questions. Thus, the audience asks and responds to questions about itself and immediately sees a summary of the results of the other audience members.

Much of the personal significance results from appreciating the courage of the performers when publicly voicing their opinions. Although this is a major factor to emotionally involve the audience, it is difficult to achieve with lifeless data objects. It is questionable, for example, whether publicly combining statements with personal information (e.g. name and address) would be able to mimic this effect or just lead to data misuse. The analysis has shown that un-anonymously asking people about their opinions in public can significantly decrease the reliability of the data. Therefore, our recommendation is rather to protect the anonymity of the data sources. Similarly, very high degrees of liveliness and of the messiness on stage are probably difficult to obtain in data physicalisations with lifeless objects. Introducing these artificially by including calculated randomness may not be useful and be rather distracting for the user.

100% City shows how the reliability of the data can be made more transparent to the viewer. In population surveys, answer options like "I don't know" or "I do not want to answer" are usually available and missing values are coded in the raw data. They, however, rarely enter formal analysis. Often only "valid" data is counted, the rest is classified as "missing". Thus, it can happen that 70% of the sample refuse to answer a question and data is reported from only 30% valid cases, thereby distorting the overall result. The 100% format shows how to handle "missing" data. People who want to abstain from voting do not indicate a colour in the "Multiple Choice" format or remain in the dark areas of the stage for "Into the Spotlight" questions, but they do not disappear entirely. By making "missing data" routinely available leads users to better understand their data and to make better decisions based on these data. Retaining missing data is also more conducive to data physicalisations where objects cannot simply disappear from the scene. The degree of cooperativeness of the interviewee is sometimes explicitly recorded by the interviewers. This variable can also be represented by subtle cues in the data physicalisation (e.g. slower onset of actuation, frequent abstentions or similar).

Conclusion

The transfer of the CAHDE analysis to data physicalisations with lifeless objects is not without trade-offs. Making objects individually identifiable and incorporating implicit variables allow the user to follow objects through a set of questions, but this is costlier to implement. Different forms of moving objects are envisioned: using small robots, employing effects of apparent motion, or computer-instructed actuation by the users. Combining quantitative with qualitative data, fostering bi-directionality, and representing "missing data" are other strategies that can be applied to physicalisations of social survey data.

Nevertheless, other and even better forms of data physicalisations may be possible than are realised in *100% City*. And, not all of the possibilities for "giving statistics a face" can be exhausted through data physicalisations, because many strategies rely on the presence of live bodies and people that dare (or struggle) to make public their opinions which creates empathy in the audience.