Something is Missing: On the Value of Displaying Missing and Nuanced Data

Tamara Friedenberger
Julius-Maximilians-Universität
Würzburg, Germany
tamara.friedenberger@uni-wuerzburg.de

Jörn Hurtienne
Julius-Maximilians-Universität
Würzburg, Germany
joern.hurtienne@uni-wuerzburg.de

ABSTRACT
Missing values in data sets are usually omitted or replaced in further analysis, and hard-to-measure nuances are often simplified for analysis and visualization. This is sensible for standard statistical analyses but should be interrogated for visual and physical data representations: Including missing or nuanced data and keeping it visible may lead to deeper data insight and invoke more audience engagement. This workshop contribution aims to prompt designers to reflect on their handling of missing data when designing visual and physical data representations.

Author Keywords
Missing data; nuanced data; inconsistent data; data visualization; data physicalization.

CSS Concepts
• Human-centered computing~Visualization ~ Visualization design and evaluation methods

INTRODUCTION
Often, the possibility of missing data is ignored, and a complete data set is assumed for statistical analyses [1]. If not ignored, cases with missing data are often excluded, or missing values are replaced with estimations [1,7]. This is necessary to retain the integrity and validity of statistical analyses, but should only be used after having exhausted all methods to ensure a more robust data collection [7].

On another note, sometimes observed variables do not fit neatly into categorizations and contain difficult-to-measure subjectivity or uncertainty [2], e.g., people feeling equally represented by two options in a single-choice question. Classic statistical inference methods for this kind of data are limited and are usually not equipped to include these nuances [2].

When visualizing or physicalizing data, missing or nuanced data points are commonly handled and cleaned just like for statistical evaluations. I propose that this should be interrogated and evaluated for each data set in question, because this practice stands to lose the possibility of deeper insight. Previously, the importance of recognizing missing data and its depiction has been highlighted multiple times, in different situations, relating to different kinds of missing data. Some examples include:

1) When data could be collected but is not: explorations of data sets that are missing completely can spark conversations on power imbalance. Examples of this can be found in chapter one of [3].
2) When data can't be collected and is therefore missing, like trying to self-observe instances where oneself completely forgets something, an example of which is available in [4].

Contrary to these two examples, I will focus on the display of missing data points (as well as nuances in categorical data) in an otherwise collectable and collected data set and examine if they can provide more insight when displayed than when excluded.

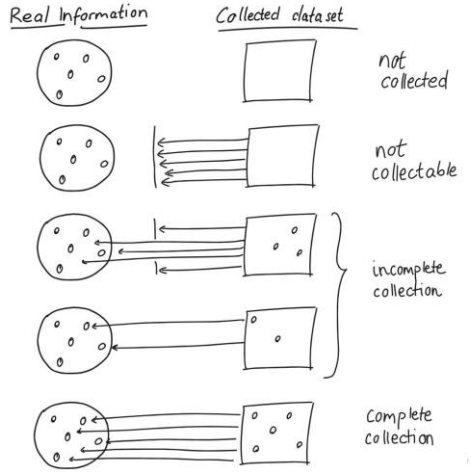

Figure 1: Reasons for missing data are varied: from simply not collected data, over not collectable data, to only partially collected data sets. Ideally, all real-world information is translated into the collected data.

To illustrate my point I will use two examples: The show concept “100% City” by the theater collective Rimini Protokoll [5, 8], and my data physicalization “Mercury 20”, which was inspired by key facets of the show. In the following, I will introduce both examples and then examine them side by side, with regard to their handling of missing or nuanced data.

100% CITY

In 100% City, 100 representative inhabitants of the city where the performance takes place, answer questions on stage about their demographic information, beliefs, values and past experiences. Depending on the question, participants answer by, for example, grouping themselves on stage under the labels “me” or “not me”, raising colored signs for multiple choice questions, or pantomiming activities of their everyday life depending on the specified time of day.

The show thrives on messiness and the personhood of the participants on stage, giving statistics "a face" and providing a deeper experience and richer data insight than classic data displays. For an analysis see [5].

Figure 2: 100% City. 100 participants display their own data for questions live on stage, e.g. grouping themselves under “me” and “not me” or pantomiming activities of their everyday life depending on the specified time of day. Retrieved on 28/6/22 from https://www.rimini-protokoll.de/website/en/projects/100-stadt-7-1

MERCURY 20

Building on these themes, I have designed a data physicalization displaying data on NASA’s candidates for their first astronauts in 1960, called subsequently “Mercury 20”. The data of 20 people is encoded (7 of which were chosen to become astronauts, known as the Mercury 7), inspired by NASA’s official selection criteria, as well as some further soft facts, that might have influenced the selection process. The data was collected from two documentaries [6, 9] and Wikipedia entries on each candidate [10, 11] with the explicit choice to not dig any further as to only display easily available data.

Starting from the top, each person is represented by a picture of themselves and their name. From this, a string weaves itself through all criteria (displayed as the white hoops). A selected number of candidates can be highlighted using string lights along their path, making it easier to compare and reference certain individuals. To guide the viewer’s interaction, a tablet provides some either-or questions: two candidates are highlighted, and the viewer is prompted to select the person they would send to space. Afterwards, they gain feedback on NASA’s actual choice and can read up on the importance of each selection criteria for NASA’s decision.

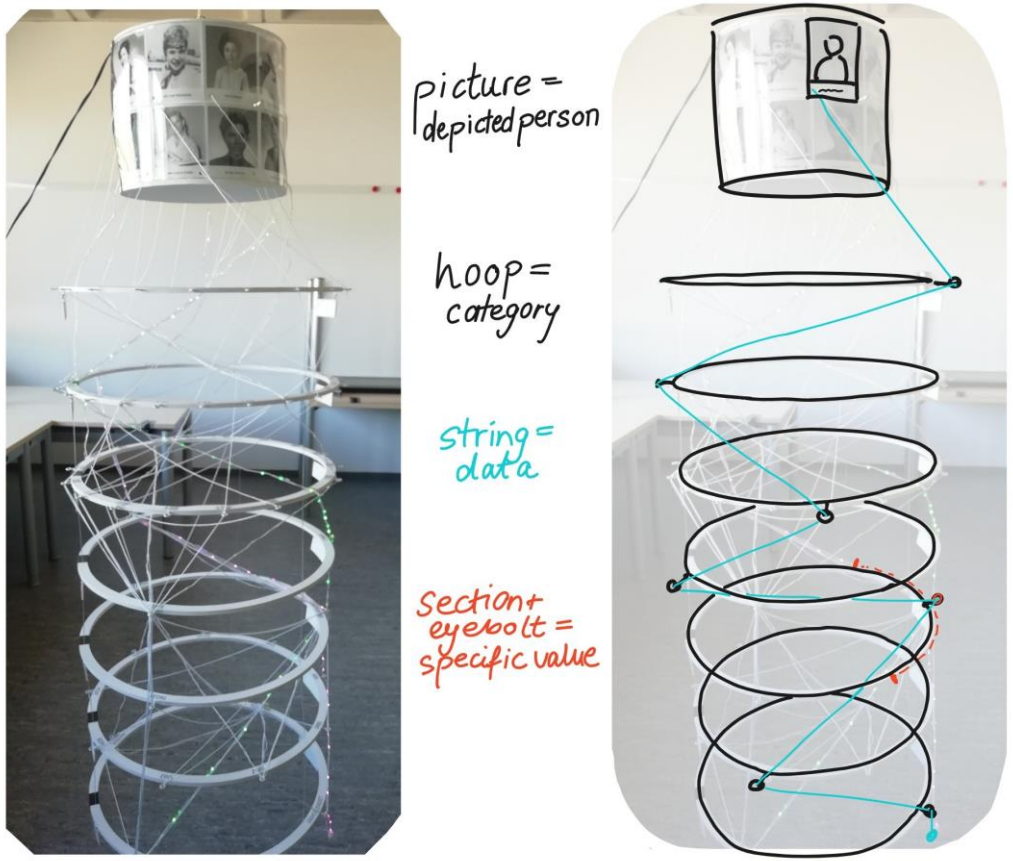

Figure 3: Mercury 20. A data physicalization based on data design themes of 100% City. 20 astronaut candidates are encoded by a string running from their picture to each hoop (representing a selection criterion category). A specific value is assigned to each person by running their string through the respective eye bolt attached to the hoop.

DEPICTING MISSING OR NUANCED DATA

Now I will discuss how 100% City and Mercury 20 depict missing data points, as well as nuanced data.

Missing Data

In 100% City, people stay on stage for the complete duration of the show. This means that even if they do not answer a question, this missing data point remains visible, as illustrated in Figure 4.

Mercury 20 keeps data visible by encoding data using a string, traveling through different layers of hoops (Figure 5). To more clearly distinguish actual data from missing data points, all existing data points travel through eye bolts fixed to the hoop, with the string remaining outside eye bolts for missing data points. The string remains visible, and viewers may stumble across missing values, for example when exploring the data of a single person. At the same time, it does not overpower the display and is only foregrounded upon closer inspection. In the case of the Mercury 20 data, keeping missing data points can invoke discussions about what data gets collected or stays available even after a number of years.

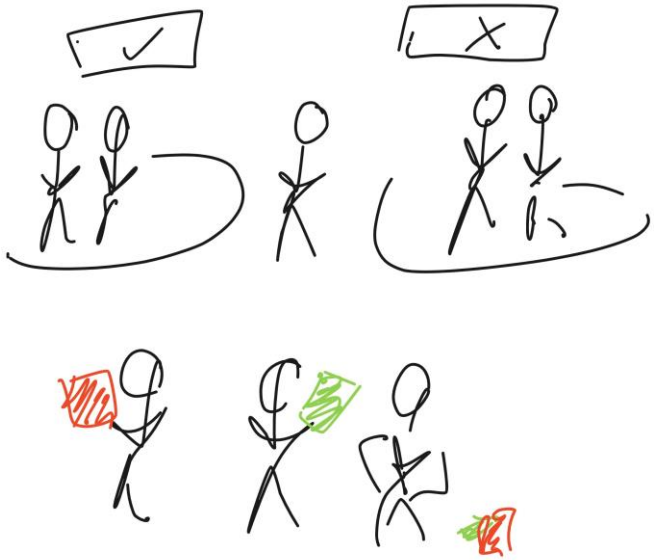

Figure 4: In 100% City missing data is depicted by keeping all participants on stage, even when they choose not to answer (e.g., by not entering an answer area or not raising a colored sign when prompted).

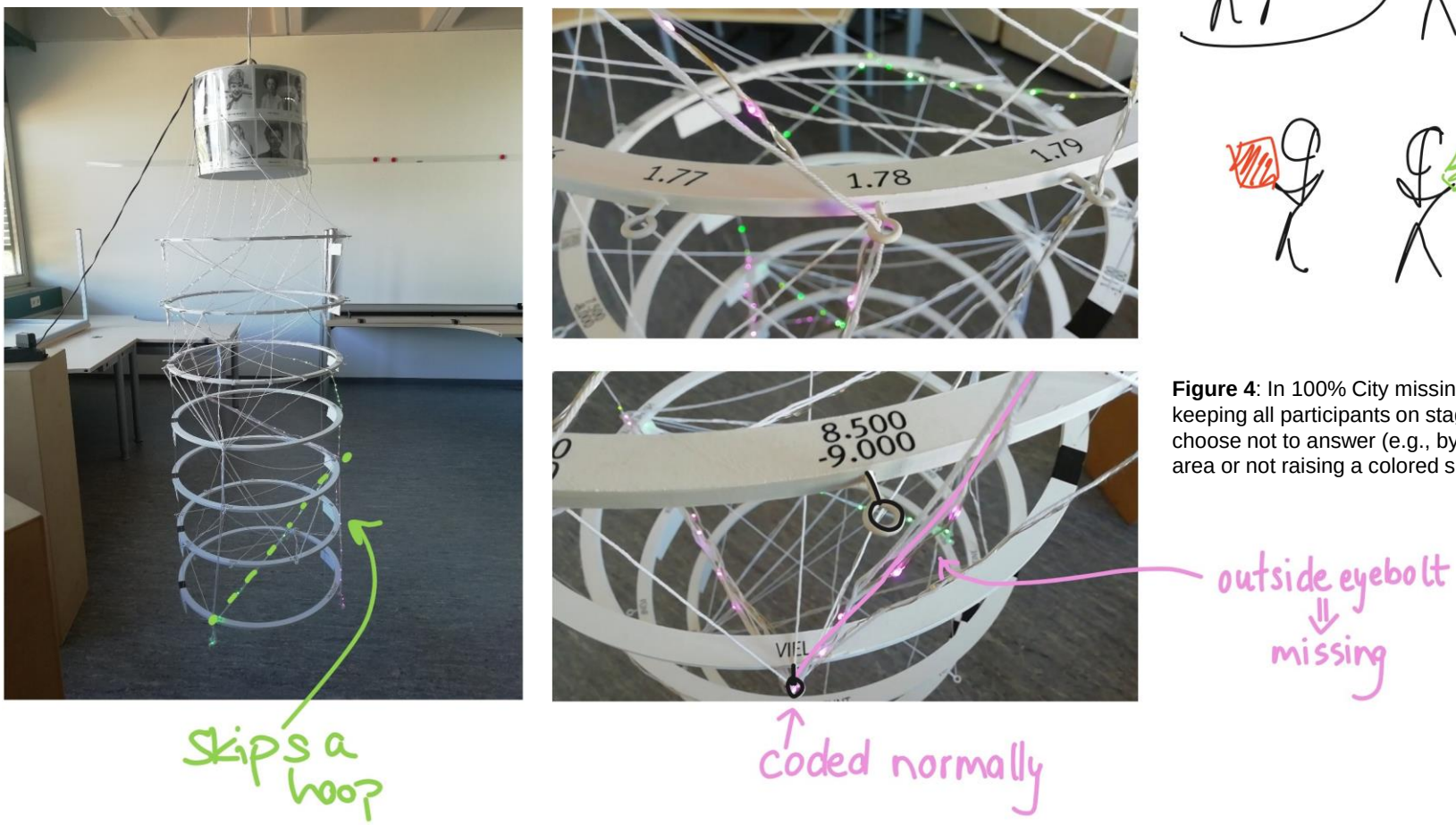

Figure 5: In Mercury 20 missing data stays visible, because the data from each person is encoded using a continuous string, which skips hoops where the respective data points are missing.

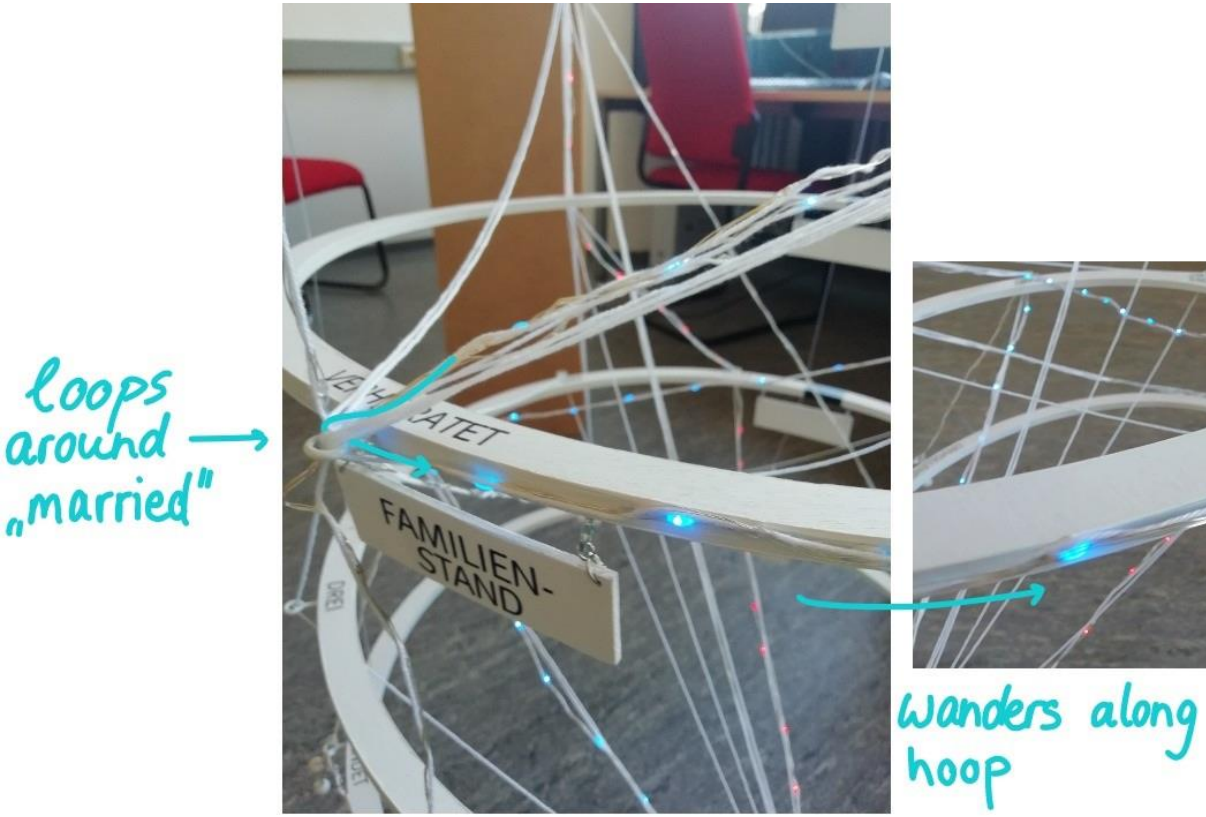

Figure 6: Mercury 20 - Nuanced Data. One data point in the category “marital status” breaks the standard encoding: The candidate at the point of selection was separated from his wife, but still posed as married to the public. His string loops around the eye bolt for “married”, travels along the hoop to the section “separated” and continues from there.

Nuanced Data

In 100% City, participants hesitating or switching answers mid-way through can be followed live by the audience. Also, participants can choose to break the encoding scheme and give an unexpected answer. Some examples can be seen in Figure 7. This makes the show more personal and prompts thoughts in the audience for what the reason for this could be, or how to interpret the unusual data encoding.

Similarly, Mercury 20 retains a nuanced data point: one candidate had separated from his wife at the time of the selection process, but still posed as married to the public, as seen in Figure 6.

I have displayed this in my design by guiding his string towards the “married” portion of the “marital status” hoop, but then have the string travel on to the “separated” category and continue to the next category from there.

Breaking the format may lead to a deeper data insight: what does it mean when it looks like the string is heading for “married”, but on closer inspection actually just loops around and travels to “separated”? Playing with and subverting the original encoding scheme provides a richer data display and can communicate more nuances than originally possible in the design.

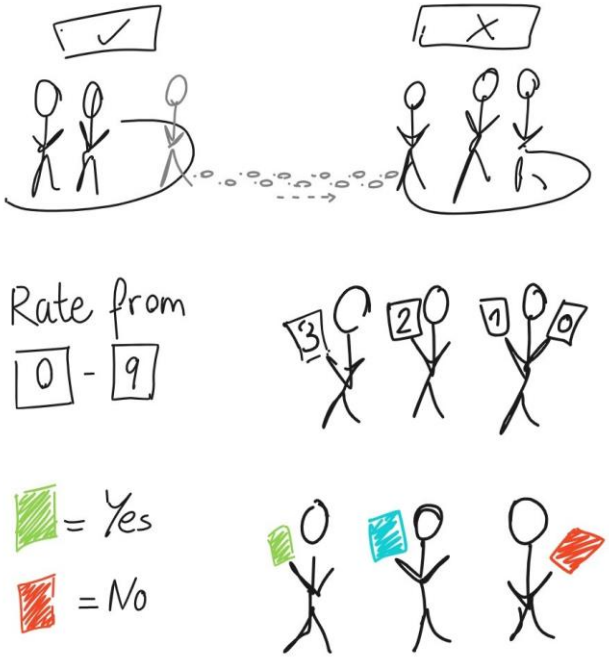

Figure 7: In 100% City, nuanced data is, for example, visible when participants change their mind and switch categories. Other possible deviations from the encoding scheme are building a 2-digit number from two single digit cards, or using a third color in a two-color question, which has no designated meaning.

CONCLUSION

Data sets with missing or nuanced data points may provide a chance for deeper insight when displayed in their entirety instead of defaulting to replacing or omitting unusual values. Designers should take a moment to ask themselves:

1) Do missing values contain relevant information on the data?
2) What story do the missing or nuanced values tell?
3) What happens when the data is allowed to break out of its encoding scheme?
4) What is the goal of the visual or physical data representation, and may it be supported through displaying data that is usually left out?

These unusual data designs can provoke thought in and discussions among its viewers. One promising way to keep missing data visible is to work with a continuous data display, like the string in my design or the physical presence of people in 100% City.

REFERENCES

[1] Douglas G. Altman and J. Martin Bland. 2007. Missing data. *BMJ* 334, 7590 (Feb. 2007), 424–424. https://doi.org/10.1136/bmj.38977.682025.2C

[2] Norberto Corral, Maria Angeles Gil, and Pedro Gil. 2011. Interval and Fuzzy-Valued Approaches to the Statistical Management of Imprecise Data. *Understanding Complex Systems* (Jan. 2011). https://doi.org/10.1007/978-3-642-20853-9_31

[3] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data feminism.* The MIT Press, Cambridge, Massachusetts.

[4] Mikhaila Friske, Jordan Wirfs-Brock, and Laura Devendorf. 2020. Entangling the Roles of Maker and Interpreter in Interpersonal Data Narratives: Explorations in Yarn and Sound. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. ACM, Eindhoven Netherlands, 297–310. https://doi.org/10.1145/3357236.3395442

[5] Jörn Hurtienne. 2018. Possibilities of Human Data Embodiment: 100% City *Position Paper for the Workshop: Towards a design language for Data Physicalization (IEEE VIS '18),* Berlin, Germany. IEEE. http://dataphys.org/workshops/vis18/wp-content/uploads/sites/6/2018/10/Rimini_IEEE_180903small.pdf

[6] Tom Jennings. 2020. The Real Right Stuff. Video. Retrieved 1 June 2022 from https://www.disneyplus.com/movies/the-real-right-stuff/7iE06CyNZFcq

[7] Hyun Kang. 2013. The prevention and handling of the missing data. *Korean Journal of Anesthesiology* 64, 5 (May 2013), 402–406. https://doi.org/10.4097/kjae.2013.64.5.402

[8] Rimini Protokoll. 2015. 100% Amsterdam (in Dutch with English subtitles). Video. Retrieved 1 June 2022 from https://www.youtube.com/watch?v=lHyTCBnqTbc&t=3s&ab_channel= EuropeanCulturalFoundation

[9] David Sington and Heather Walsh. 2018. Mercury 13. Video. Retrieved 1 June 2022 from https://www.netflix.com/title/80174436

[10] Wikipedia. 2022. Mercury 13 — Wikipedia, The Free Encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Mercury%2013&oldid=1101532100

[11] Wikipedia. 2022. Mercury Seven — Wikipedia, The Free Encyclopedia. Retrieved from http://en.wikipedia.org/w/index.php?title=Mercury%20Seven&oldid=1098699832