Using Concrete and Realistic Data in Evaluating Initial Visualization Designs

Søren Knudsen
University of Copenhagen
Copenhagen, Denmark
sknudsen@di.ku.dk

Jeppe Gerner Pedersen,
Thor Herdal
University of Copenhagen
Copenhagen, Denmark

Jakob Eg Larsen
Technical Univ. of Denmark
Kgs Lyngby, Denmark
jaeg@dtu.dk

ABSTRACT

We explore means of designing and evaluating initial visualization ideas, with concrete and realistic data in cases where data is not readily available. Our approach is useful in exploring new domains and avenues for visualization, and contrasts other visualization work, which typically operate under the assumption that data has already been collected, and is ready to be visualized. We argue that it is sensible to understand data requirements and evaluate the potential value of visualization before devising means of automatic data collection. We base our exploration on three cases selected to span a range of factors, such as the role of the person doing the data collection and the type of instrumentation used. The three cases relate visualizing sports, construction, and cooking domain data, and use primarily time-domain data and visualizations. For each case, we briefly describe the design case and problem, the manner in which we collected data, and the findings obtained from evaluations. Afterwards, we discuss the potential means to data collection, and which outcomes we expect this might provide.

CCS Concepts

•Human-centered computing → Visualization design and evaluation methods; *HCI design and evaluation methods; Interaction design process and methods; User centered design;*

Keywords

Methodology; InfoVis; Personal Visualization; Evaluation; Pre-design Empiricism

1. INTRODUCTION

Visualization work typically operate under the assumption that data has already been collected, and is ready to be visualized. However, when starting visualization projects in new domains, this assumption does not always hold true. This seems particularly apparent in emerging visualization domains such as casual [20] and personal [11] visualizations. At the same time, previous work has emphasized the need to work with real data [24, 17]. However, it is broadly observed that gaining access to real data can be challenging in early phases of design studies [24]. Manual and semi-manual means of data collection might be an answer to the problem of evaluating early designs. Likewise, piggybacking on off-the-shelf tools, might be an interesting alternative. In personal visualizations, it seems particularly important to use real data that participants can relate to in order to derive meaningful results from evaluations. Being able to collect initial data would help the visualization community to identify new opportunities for visualizations, and can work as proof-of-concepts, to evaluate the value of visualizations for an unexplored domain. More broadly, we believe it is time to consider novel methods inspired by, and catering to casual [20] and personal [11] visualizations.

In this paper, we explore means of collecting data for the sake of evaluating initial visualization ideas and designs (i.e. formative evaluations). We aim to illuminate potentials for conducting pre-design empiricism [3]. We outline three cases in which we collected initial data to enable us to create prototype designs and evaluate the value of visualizations for the concrete cases. Each case represents a point along several factors that relate to manual data collection. The factors are for example, the duration of the data collection period, the role of the person doing the data collection, the instrumentation used in the process, and the persons involved in the process. In design studies [24], our techniques could be used as basis for *winnow*, *cast*, and *discover* phases, as well as early parts of the *design* phase.

Our approach is inspired by previous work in human centered design [25, 13], and work in information visualization [17, 12, 4]. The premise of quickly conducting and analysing usability evaluations presents an alternative to common evaluation methods (e.g., [15, 2]). Although long, structured, and thorough evaluations might often provide valuable results, they are not always the most effective for the problem at hand. Instead, we can pick "low-hanging fruits" with more rudimentary approaches. This way, we can go quickly from getting an idea or set of ideas, to evaluating these. Considering evaluation from practitioners' point of view, this seems to be particularly useful.

Additionally, evaluating multiple ideas might lead evaluation participants to provide more critique [27]. This presents a trade-off between number of ideas, evaluation perfection, and use of resources. Most importantly, given the same amount of time, we might choose between evaluating one

BELIV '16 Baltimore, Maryland, USA

DOI: 10.1145/1235

idea thoroughly, or several ideas less thoroughly. In early phases of design studies, many parallel ideas exist. In these situations, it is beneficial to evaluate several alternatives rather than a single design. In this paper, we use paper prototyping methodology [25] to quickly construct interactive prototypes that we can evaluate. Further, we produce visualization sketches by hand and use of off-the-shelf software tools such as Google Maps and Tableau where appropriate.

Our contributions are three-fold: First, we contribute a set of methods that help researchers and practitioners go from idea to first evaluation results for ideas for which no data existed, in short time (e.g., less than a day). The described techniques imposes limited work on designers and participants. They are both intended to be used concretely and to inspire similar approaches and techniques that might be applied in design. Second, we provide concrete insights from three cases, that show novel uses of visualizations. This showcases the usefulness of our data collection approach. Third, we identify a number of challenges related to collecting data and conducting evaluation in early design work, and outline limitations of the approach. These both illuminate potential problems and solutions in data collection.

2. RELATED WORK

Here, we discuss existing design methodologies which we can build on, and which complements our methods and techniques. We consider both the concrete methods (e.g., Roberts et al. [22]), and broader discussions of methodology in design studies (e.g., Sedlmair et al. [24]).

2.1 Design methodology

While evaluation methods have become more commonplace in visualization work, we know comparatively little about potential approaches for designing visualizations, particularly from a user-centered point of view. Brehmer et al. [3] highlighted the need for pre-design empiricism. We are particularly driven by the need for quick prototyping techniques, which we believe is necessary to construct methods that practitioners might adopt. We are not aware of any visualization design method with this particular focus, although several methods might be applicable (e.g, [17, 12, 4]). Greenberg et al. [10] provides a wealth of useful sketching techniques for design. While, not addressing visualization designing specifically, many of the techniques presented in this book are applicable to visualization design. Additionally, their unusual and creative use of tools is somewhat similar to some of the techniques we describe.

Pretorius and Wijk [21] argue to "let the data speak", while acknowledging that user-centered design approaches are fundamental. However, they also stress that a core part of visualization depend on the type of data being visualized. They argue that data-driven approaches are particularly important in uncharted visualization territory, in exploratory and complex analyses, with domain experts that have deep domain knowledge. We also believe that obtaining simple data for simple and casual applications could provide similar value in uncharted territories. To be able to start working with data, the existence of data is a crucial factor.

Within the quantified self community[1] data visualization of self collected data is a well known and widely used approach to reflect on and learn from self tracking data. Furthermore there is a tradition in the community for sharing knowledge on methods and experiences from personal self tracking, so that newcomers to self tracking can learn from and get inspired by other self trackers. While the phenomena that self trackers collect data about represent a wide range of activities [6] it is by definition always about self collected data that represents phenomena that relates to the individual self tracker. However, the process and possible methods of more directly assisting others in the process of self tracking and in particular visualization of the self collected data has received less attention. Additionally, focus in this community seems to — for a large part — be on fully automatic approaches to data collection.

Gaver et al. [9] suggested a method to collect experiences from participants by facilitating capture of everyday activities. While their method relates much to our methodology, the goals are different. Where Gaver et al. sought to portray everyday experiences to design new technology, we aim to contribute methods and techniques which helps to collect data to use as the basis for designing and prototyping visualizations. Similarly, diary studies (e.g., [8]) and experience sampling (e.g., [5]) methods provide ample ways of eliciting data from participants. While previously these methods have been used as the foundation for design, they might be useful in the context of understanding how visualizations might play a role in peoples' lifes. Specifically, Carter et al. [5], suggests that *"audio is a lightweight media appropriate for annotation"*. We also consider audio as annotation media.

In two of the decribed cases, we specifically aim to evaluate multiple ideas. This choice is driven by the suggestion that presenting participants with alternatives might lead participants to provide more critique [27]. To be able to quickly construct evaluate ideas, we base evaluations on paper prototyping methodology [25], produce visualization sketches by hand, and use off-the-shelf software tools such as Google Maps and Tableau where appropriate.

Within quantified self and personal informatics, the Stage-Based Model from Li et al. [16] suggests to model personal informatics systems with an iterative process of self tracking consisting of five different stages: preparation, collection, integration, reflection, and action. The method applied in the reflection stage is typically visualizations of self-collected data.

2.2 Evaluation

Our work also relates to evaluation methodology. Most importantly, we use the ability to quickly collect data and construct a prototype visualization (construed broadly), to evaluate this with the people for which the collected data makes most sense; the people that the data describes. Similar to our focus on generating quick insights, Kjeldskov et al. [13] considered this from the perspective of conducting usability evaluations in a day. However, our focus here is not on the specific act of formative or summative evaluation of visualizations. This has been the focus of much other work (e.g., [15, 4]). However, in our design cases, we make use of common evaluation methods such as the think-aloud protocol [2], interview techniques [14], and empirical data analysis methodology [14, 26]. We do so to illustrate how our methods fit into the broader narrative of going from idea, through data collection and prototype, to evaluation results.

[1] http://quantifiedself.com/

3. METHODOLOGICAL APPROACH

We have made the case that understanding ways to collect concrete and realistic data to evaluate initial visualization designs is valuable. To increase our understanding of how to approach this process, we chose to conduct three case studies that attempted to do follow such a process. We selected the three cases primarily from the perspective of varying 1) how the task of collecting data is distributed over time and people, 2) the sophistication of used technology, and 3) the degree of personal information. At the same time, we looked to collect data for cases where data is not collected presently, is not possible to collect fully automatically, and where the duration of a useful collection period does not span more than a few days. Lastly, we aimed to select cases that relate to personal visualizations. However, in looking for cases, we excluded domains that include very personal data (e.g., health), since we considered this to be too sensitive to study, considering our methodological focus.

In each case, we followed the following steps: **Identify** and characterise the domain and analysis problems. **Understand** the requirements for collecting data in the domain, and the nature of the data (e.g., what does the data say [21]). **Consider** the possibilities for collecting data and potential problems for each possibility. **Choose** a data collection technique. **Collect** the data. As part of this process, the chosen collection technique is prepared, the actual collection is performed, and data is extracted. **Produce** visualization sketches and prototypes, based on sketches drawn by hand and off-the-shelf visualization tools. **Evaluate** visualization sketches and prototypes, including preparing a study protocol. **Analyse** evaluation results. **Methodological** analysis and evaluation of data collection method.

In the next section, we describe three case studies in which we manually or semi-manually collected data for evaluating initial visualization designs. The cases study 1) how visualizations might help elite soccer children evaluate and improve their own performance, 2) how visualizations might help construction workers understand their use of time on construction cases, and 3) how visualizations of cooking activities might help people structure such activities.

4. CASES

In the following, we describe three case studies which follow the steps described above to collect data for initial design and evaluation of visualizations. The cases study 1) how visualizations might help elite soccer children evaluate and improve their own performance, 2) how visualizations of cooking activities might help people structure cooking activities, and 3) how visualizations might help construction workers explore and understand their use of time on construction projects.

4.1 Case 1: Sports Domain

Case 1 studies the potential of visualizing elite soccer children's performance data, to support them in understanding their performance. The case is derived from a larger study conducted with an elite soccer club, which is reported in Pedersen et al. [19]. In the study, we conducted interviews with team players, coaches, and members of club management, workshops with the team players, designed lo-fi visualizations prototypes, and finally evaluated these designs with three individual players on the team. This helped us *identify* and understand the domain, and inspired us to explore the concept of collecting data more broadly in this paper. In this paper, we focus on the evaluation of the lo-fi visualization prototypes.

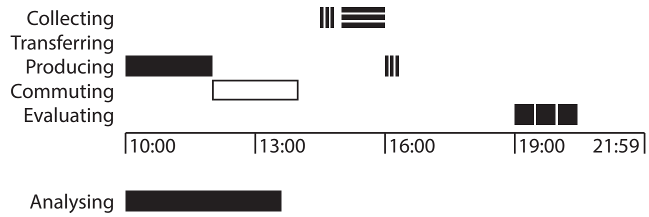

Figure 1: Case 1 design and evaluation process. All activities caried out within a single day.

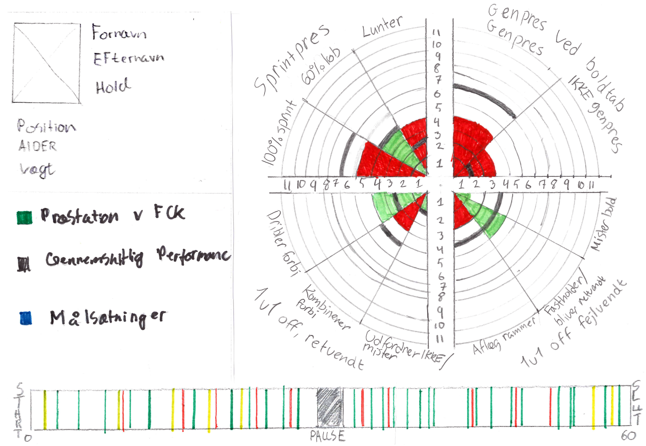

Figure 2: An example of an interactive paper prototype used in the evaluation in Case 1. Depending on participants' choices, we revised or replaced the top right part of the user interface with alternative visualizations. Here, the prototype shows performance data collected during a match played a few hours prior to the evaluation. For further information, see Pedersen et al. [19].

The premise of the study was an existing smartphone app that showed a tabular representation of each soccer player's individual key performance parameters for a match. This representation was accessible on the player's own smartphones, and allowed us to *understand* the requirements for data collection, and the nature of the data. After obtaining an understanding of the domain, we designed and evaluated lo-fi visualization prototypes, based on data collected up to, and during a soccer match. Since our prototype design was developed prior to the evaluation, we were able to construct a prototype frame before collecting data.

We *considered* data collection options, and *chose* to collect data based on the existing smartphone application, assisted by other participants in our larger study, who did the actual data collection. On the day of the evaluation, and thus *collection* of the data, we started by asking the three evaluation participants to provide their performance expectations. We asked them to fill in a form of performance expectations (see Figure 3). Afterwards, the match was played. During the match, team members on the bench were asked to tag performance data of one evaluation participant each. They

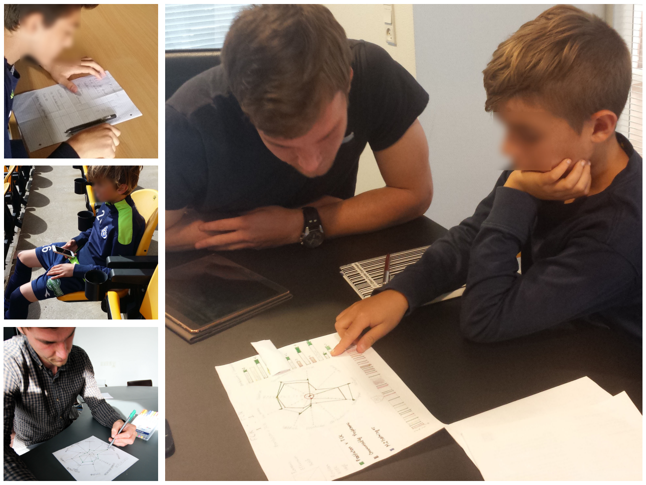

Figure 3: In Case 1, we asked study participants to tag soccer performance of other participants. We used this way of data collection to be able to conduct evaluations of lo-fi prototypes on the same day as collecting the data. The four situations show: A participant logging performance expectations prior to a match (top), a participant tagging another participants' performance during a soccer match (middle), a co-author producing a lo-fi prototype (bottom), and a co-author and participant in an evaluation situation (right).

used the existing smartphone app to do this (see Figure 3). After the match, we extracted the collected performance data from the smartphone app in raw format and imported it into a spreadsheet. From there, we augmented the paper prototypes that were created prior to the match with the collected data (see Figure 3). This process enabled us to rapidly *produce* paper prototypes based on real performance data immediately after the match had ended. We conducted the *evaluation* with the players for which the data had been tagged (see Figure 3). The paper prototype thus visualized the participants' performance data for a match they played earlier in the day. The players used the paper prototype as they would use a real interactive system. For example, when pointing with their fingers to indicate tapping, we would adapt the visualizations to their interactions (see Figure 3).

After the evaluation, we *analyzed* how the players used the visualizations to gain insights and how the visualizations supported their understanding of their performance. We observed how the visualizations helped the players' to understand their own performance data, and increase their ability to evaluate their own performance. We observed a player that evaluated his performance and contextualized it, in relation to previous and future matches. For example, he structured his actions by color: *"I can see that I have a surprising high amount of red actions in the end of the match. This is because I am getting tired. I should have played more non-risky passes"*, clearly considering how he might change behavior. Thus, we believe the improved visualizations helped him understand his performance data. In contrast, the other players received only marginal benefits from the paper prototype. A player that had a limited understanding of his own performance, only slightly increased his understanding of the data. This might stem from confusion caused by the visualization, and the player's inability to integrate the visualizations with his memory of the match. On the other hand, a player that had a particularly good grasp of his performance had few problems understanding the visualizations. However, he obtained few of his insights from the visualizations, and used most of the evaluation to draw connections between the visualizations and his memory of the match. His insights and his ability to compare his performance to the match resulted in a marginally enhanced level of insights.

Considering the *methodology*, the paper prototype method gave us opportunities to quickly and easily create designs based on the participants' real data. We expected the participants to contribute design improvements. However, they were far more interested in their performance, than the design.

4.2 Case 2: Construction Domain

Case 2 studies the potential for using visualizations to support self-employed construction workers (i.e., small businesses) in understanding and planning their use of time. We *identified* this as an interesting problem, as we expected this domain to pose an interesting mix of business visualizations and personal visualizations, due to the domain and size of the business. We used our previous knowledge, as well as informal discussions with people in our network that work as craftsmen or have done home improvements themselves, in order to gain a better *understanding* of the domain and data. The entire process in shown in Figure 4.

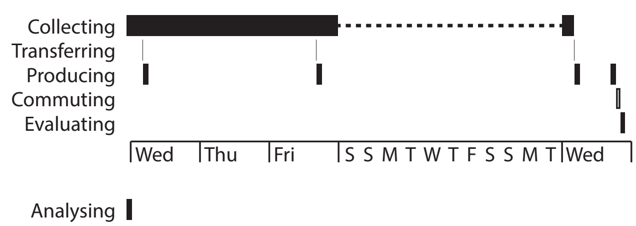

Figure 4: Case 2 design and evaluation process. We collected time+space data for the participant over a two-weeks period. However, the remaining phases in the case amounted to a days' worth of work, most of which was spent during the evaluation day.

To consider this problem from the point of view of collecting data, we made arrangements with a construction worker to track the time he spent on different cases during two normal work weeks. To *consider* data collection techniques, we considered the duration of the collection period, as well as potential instrumentation. We also considered more manual possibilities. We *chose* to ask the participant to use the smartphone app Moves[2] to track time and position throughout the work-day. Our choice provided additional benefits. Aside from collecting the time he spent on construction projects, the app collected time spent on transport, shopping materials and equipment, and in his workshop.

The participant *collected* data over two weeks. We first examined the raw files. The collected data provided names for positions, as well as latlongs and start and end times. During the period, the participant visited were located at his

[2] see http://moves-app.com

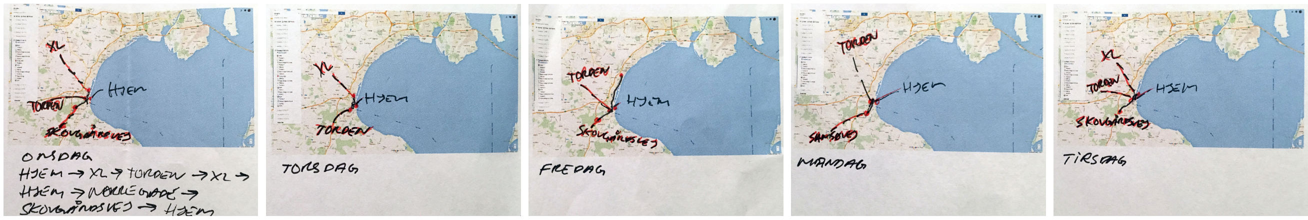

Figure 5: We used small multiple map visualizations to ground the evaluation in concrete data in Case 2. The figure shows the 5 days leading up to the day of the evaluation. In the evaluation, we presented all 10 days of data in this way, in three rows corresponding to weeks. We sketched routes on top of the Google Maps representation. We also added information about locations (e.g., street names of cases, names of material suppliers, home, etc.).

home, his workshop, an auto service, construction material suppliers, and at 7 construction sites.

To *produce* visualizations to evaluate, we first loaded the data into Tableau and the "My Maps" feature of Google Maps. Google Maps was helpful in obtaining an overview of the collected data. We chose to use a small multiples approach, and used Google Maps to create maps that showed the route that the participant had taken each day. To make this work in the context of small multiples, we printed the 10 maps on an large sheet of paper (A3 format), and sketched the route on top of the printed representation. We augmented these with information about locations. This is shown Figure 5. Additionally, we used Tableau to construct visualizations. One representation showed locations during the entire collection period as circles where size encoded time spent at the location. Another showed each day in a stacked bar, where the individual parts of bars represented locations.

To *evaluate* the produced visualizations, we brought the printouts and interactive visualizations with us to the evaluation. First, we showed the small multiples representation, which we used to talk about the context of his work. For example, we talked about the reasons for the frequent visits to the material suppliers, and the location of his workshop and the cases. Afterwards, we turned to the interactive version of Google Maps and explained how we could aggregate them. Then, we showed the visualizations we had prepared in Tableau, which served to illustrate approaches to visualize use of time for different activities. Finally, we used pen and paper to sketch these ideas together with the participant. This allowed us to talk through the different ways that activities could be shown. We focused on the choice of whether to show activities stacked, which worked well to represent total time spent per day, and to show a representation similar to a week calendar.

After the evaluation, we *analysed* the collected evaluation notes. While the small multiples seemed to help the participant recollect his activity, the aggregated time-visualization in Tableau seemed more useful to him. He suggested that these would be useful to keep track of time spent on, as well as plan construction projects. The participant for example suggested this could be used to decide between which construction construction jobs to accept based on location. Additionally, he imagined using such a tool planning in collaboration with the other construction workers that he sometimes collaborated with. From this, we sketched such a design in collaboration (see Figure 6).

Finally, we considered our *methodological* approach. The

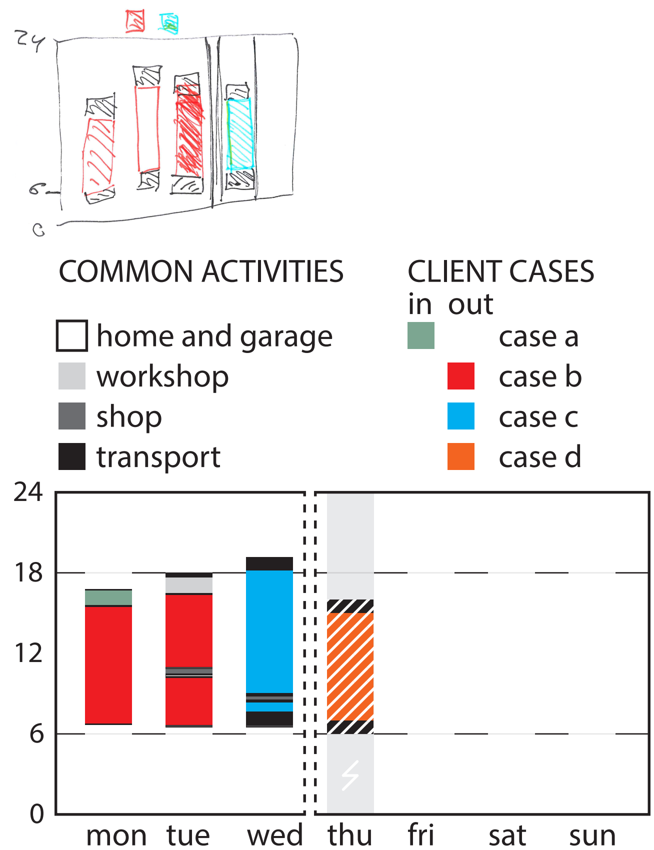

Figure 6: Sketches and drawings from Case 2. The top sketch was produced during the evaluation. We produced the bottom illustration, to explain the idea. The participant considered the need for planning the near future in the face of traffic and weather conditions. We provided a design suggestion which we explained would enable him to drag individual cases, workshop or supplier visits to the next days to schedule them. Traffic and weather predictions would be shown around the planned activities, to enable the participant to plan accordingly. Additionally, cases would be colored according to whether they were inside or outside cases.

evaluation gave us an improved understanding of the domain, supported by concrete data collected using an off-the-shelf smartphone app. We concluded the evaluation with novel ideas for visualizing activity data in the concrete domain, derived from the needs of a construction worker. Considering the participants' response to questions that related to more hypothetical data, we believe that the concrete data showed particularly value.

Finally, going from collected data to visualizations proved time-consuming. We used a broad selection of techniques to produce visualizations. However, we spent an inordinate amount of time preparing visualizations for the evaluation, which we do not believe provided value in the evaluation.

4.3 Case 3: Cooking Domain

Case 3 studies the potential of visualizing allocation of time for cooking activities in a household kitchen. We *identified* this as an interesting problem, due to its potential relation to personal and casual visualization and to Internet of things research (e.g., intelligent fridges). Monroe recently approached this domain from a different perspective [18]. In our case and in contrast to the previous cases, we used the data collection and design process, to increase our *understanding* of the domain and data. Like Case 1, this activity is relatively short, perhaps spanning 30 to 90 minutes. However, in contrast to soccer, it is possible for the data collection to occur parallel to the primary activity and be handled by the same person, without adversely interfering with the activity. The entire process, which comprised five hours, is shown in Figure 7.

Our choice enabled us to *collect* data, construct visualization sketches, and conduct an evaluation of this in five hours on a week day evening (see Figure 7).

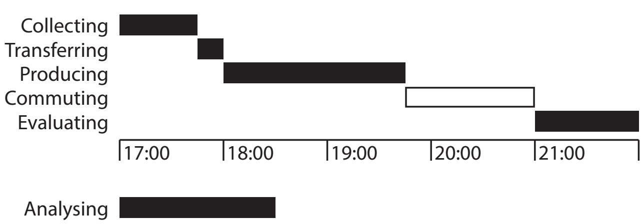

Figure 7: Case 3 design and evaluation process. All activities caried out within a single day.

To consider this problem from the point of view of collecting data, we made arrangements with a female participant to track her use of time during cooking activities. We *considered* alternative possibilities to enable her to track time spent on cooking activities. We considered the duration and limited interference with the activity, as important factors in this decision. Additionally, the potentially wet and messy environment limited the possibilities for data collection.

We *chose* to ask the participant to use her smartphone to record audio while cooking, and to indicate activities by clapping and verbally announcing them. We expected the claps would enable us to manually identify activity switches based on the sound signature. Further, we expected this compromise to only slightly influence the primary activity, while at the same time providing an easy way to collect, transfer, and extract the data, for which the participant used her smartphone.

The participant *collected* data during a late Tuesday afternoon while cooking food for her and her son. In the end, we received a Dropbox link to a 43-minutes long audio recording of our participant cooking pasta bolognese, while occasionally talking with her son, who happened to be in the room watching television. We identified claps by the sound wave, although we chose to also use a time-frequency plot. We identified 11 claps in total which were all succeeded by verbal activity announcements. The participant had chosen to announce activities related to vegetables as only two activities (handle vegetables and add vegetables to pot). We chose to manually divide this in more detailed steps by manually going through the recording and adding tags. From these, we created a comma-separated file describing the activities and timestamps. Afterwards, we explored possibilities for visualizing the data in Tableau and chose to print a few visualizations. From the raw data and print outs, we *produced* visualization sketches which we brought to the evaluation.

We subsequently conducted an informal *evaluation* with the participant, based on a prepared question sheet that we brought to the evaluation. The questions were written in black, and responses were added in red. We followed a similar style in some visualization sketches (see Figure 8, right).

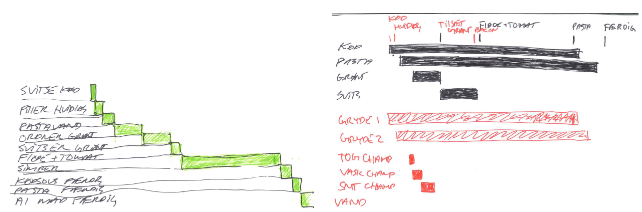

Figure 8: Two sketches used during the evaluation in Case 3. The sketch on the left uses a simple approach to dividing time into activities. The sketch on the right shows overlapping activities.

In the evaluation, we showed the visualization sketches to the participant one at a time. The first sketch was a simple and naive representation that divided time into the different activities (see Figure 8, left). The next sketch presented some activities as overlapping others (see Figure 8, right). For example, it showed the two cooking pots as activities that spanned a long time, from the moment they were found in the cupboard, and to the meal was done. We also showed a range of other time-line based visualizations.

We subsequently *analysed* the notes collected during the evaluation. In relation to the first sketch, the participant explained that it represented the data incorrectly, and that it didn't reflect that some activities were parts of longer activities. However, it also helped to form a mutual understanding of the collected data in a somewhat raw visual form. The second sketch we presented to the participant aligned much better with the her understanding of the process, and helped her to consider more broadly, how such visualizations could be useful.

The participant considered that cooking activities could be useful in a range of situations, such as following recipes while cooking, optimising cooking activities, and computing energy contents of meals. Interestingly, she also considered the usefulness of such visualizations in planning how to involve children in cooking activities. She expressed surprise about the potential value of visualizations for following recipes: *"It was interesting to see timelines. [I was surprised by] how easy it would be to obtain an overview of a recipe, compared to reading a list of ingredients and a procedure."*

We were specifically interested the *methodological* approach and its benefits and limitations. The prepared questions related to the experience of collecting data, to the type of collected data, and to the validity of the method.

The participant suggested that the data collection technique itself might be useful to author recipes, by using the audio recording as *"a kind of note-taking device"*, perhaps in combination with visualizations. While she claimed that cooking visualization were relevant, she also considered how data collection techniques and visualizations based on timelines could be useful in a professional context (e.g., project management). The participant stated that the method of collecting data felt easy, somewhat odd, but not difficult or interfering with the cooking activity.

As explained above, the participant had chosen to only mark the most important activities, since she was in doubt about how many activities to mark. From this, it is also clear that we observed issues in collecting sub-activities, which the participant also articulated. While we were able to manually infer the subdivisions, this was only possible because we had domain experience. In studies of novel, specialised domains, this might not be possible. However, we also used this issue to consider how we might collect information about subactivities in the course of conducting an evaluation itself. We asked the participant to explain the subdivisions. Based on this, we created a crude visualization based on the participant's input. We believe this approach might be fruitful, and consider this to be related to participatory design methods. Alternatively, participants could be asked to verbalise subdivisions during the data collection phase. However, we worry this might interfere more with the primary activity.

We also asked about the ease of marking activities by clapping. We had considered other techniques and wanted the participant to comment on these (e.g., using physical buttons). The participant argued against these alternatives, and commented that these were inferior to the used technique. While we believe that alternative techniques might be usable, we view the participants' comments to suggest that the chosen technique worked sufficiently well.

Finally, we were interested in the validity of the collected data. During the evaluation, we noticed that the participant had forgotten to mark the start of one activity (unwrapping and adding bacon). While we were able to infer some activities from the sound recording, we could not infer what it was. However, we handled this during the evaluation. Additionally, from the discussions inspired by the visualizations, we asked about other information that might be visualized. From the participant's suggestions, we figured that information which she described as activities, might relate to the used equipment (e.g., kitchen cooker and cookware) or kitchen locations (e.g., cooker, table, and sink).

5. SWIFT DATA COLLECTION

From the three cases described in the previous section, we describe the Swift Data Collection techniques for realistic quick'n'dirty visualization design and evaluation. As we show in the following, the techniques span a space of data collections approaches rather than a fixed set of techniques. In presenting them as such, it is our hope that they inspire researchers and practioners to adapt these ideas to their own particular circumstances.

The techniques span factors relating to Duration, Roles, Instrumentation, and Persons (DRIP). We describe these factors in the following.

Duration: We consider the duration of data collection. On one end of this scale, we find activities that span minutes. On the other end, we find activities that span days or even months. The duration of data collection relates closely to the goal of the visualization design.

Roles: We indentify three roles of people: The activity role (**R1**), which is fullfilled by a person carying out the logged activity. The data collection role (**R2**), which is fullfilled by a person doing the actual task of collecting data using the selected instrumentation. The facilitating role (**R3**), which is fullfilled by a person designing and evaluating visualizations in a potentially novel domain. The third role might be divided in two or more roles. However, for the purpose of our work, this division is an unnecessary complication.

Instrumentation: We consider the technological sophistication of the instrumentation used by **R1** for collecting data on activities performed by **R1**. On one end of this scale, we find techniques based on pen and paper. On the other end, we find fully automatic tools that work with no interaction from the data collector.

Persons: We identify three categories of people: Researchers, practitioners, or designers (**P1**). The study participants performing an activity (**P2**). Other study participants (**P3**). The last group might simply be part of the study, in order to facilitate data collection (see **R2**). However, they might also perform a subsequent activity which needs to be logged (see **R1**), thus changing their role.

These factors can and have been varied previously. For example, the choices of distributing *roles* over *persons*, has been varied in previous work. Within quantified self, it is common that one *person* maintains all three *roles*. In HCI, self-logging is typically used for longer durations of logging, which in our descriptions combines **R1** and **R2**. Whichever division, the choice of *roles* and *persons* influences other factors and decisions in collecting data. Additionally, the collection duration impacts the practicality of the other factors. For example, if the design is intended to visualize activities occuring over several days, it is probably impractical to assign different persons to **R1** and **R2**.

Obviously, **pen and paper** is the most simple *instrumentation* form. For many situations, this is a perfectly useful *instrument* (and perhaps the most optimal one). To log time stamps, a watch might augment these. In some cases, location might also be relevant to collect. However, we expect this rarely requires more than can be observed from the surroundings. Even using pen and paper, it is necessary to have a plan for what to collect (i.e. a schema), in order to be able to use the collected data immediately after the data collection process. Defining such as schema appears counter-intuitive to the free-form nature of pen and paper, and suggests that less flexible *instrumentation*, such as simple smartphone applications, might be as useful. Additionally, depending on the environment in which the activity is performed, pen and paper might be suboptimal. Finally, if **R1** and **R2** is combined, pen and paper might adversely interfere with the activity.

Both audio and video recording are well-known alternatives to pen and paper approaches. These make it possible to revisit collected material after the fact, to identify and record new observations (i.e. to adjust the collection schema). Additionally, if **R1** and **R2** is combined, these might interfere less with the activity. This is commonly used in observational studies. However, for the purpose of swift data collection for visualization, this is less useful, since extracting activity data from audio or video is time-consuming. To simplify extraction of data from audio or video, the data collector can manually create audible or visible marks. Audible marks might be indicated by clapping, whereas visible marks might be indicated with a strong light source. Both techniques should create a strong and short peak which is possible to identify from scanning the material (it might even be possible to automate this detection). Additionally, these marks might be narrated by the data collector. All smartphone systems we know offer dictation functionality either from factory settings, or through free third-party apps.

Additionally, it is possible to use a smartphone as a data collection *instrument*. It can be used to collect common data as described above, or more specific and custom data as described in Case 2. However, using off-the-shelf data collection tools might also be problematic. In Case 2, the participant could use the app to explore the collected data during the collection period. This might result in preconceptions about the potential in using this data. On the other hand, enabling participants to explore the collected data also brings awareness about the collected data.

Finally, we considered using physical buttons as a data collection *instrument* (e.g., in Case 3, using physical buttons to register events for each type of activity). From our own experience, instrumentation can be useful for tracking events over longer periods of time (i.e., months). We have for example experimented with tracking allergies with a simple smartwatch application. We believe that considering instrumentation options more broadly might be interesting.

If **R2** and **R3** is allocated to different *persons*, it is not necessary for them to meet. Instead, the collection *instrument* can be sent to the data collector and later returned. Additionally, in many parts of the world, the majority of the population have access to a smartphone, which in itself is a useful data collection *instrument*. In such situations, we might simply ask the data collector to send the data to the facilitator over the Internet. Such data might both come as audio, video, photographs, or structured data formats. However, we are concerned that long video recordings might be problematic to send over the Internet, and suggest that alternative options should be considered in this case.

6. DISCUSSION AND FUTURE WORK

To evaluate new domains for visualization, we have described three cases that outlined data collection possibilities and described factors relating to these.

Memory We believe that by using real data, the participants were both able to better recollect previous activities, and were more engaged and motivated to be part of the study, and the evaluations in particular. Thus, we might use real data as a motivational aid when exposing participants to prototypes, while their memory is still fresh, and any insights they can obtain, still useful. The small multiple map visualizations in Case 2 seemed to support the participants' recollection of events from the past weeks. However, we also observed in Case 1, that participants could short-circuit the evaluation. Instead of relying on visualizations for understanding their performance, they used their own memory. While this implies that a factor in the evaluation provided greater external validity, this might be problematic if the goal of an evaluation is to obtain insights on the use of the visualization. Additionally, this can serve both as an argument for an insights-based study approach [23, 28] and an argument for the use of real data, which might bridge the gap between data and peoples memory and understanding [1]. In contrast to Case 1 and 2, the benefit of the memory in Case 3 was not apparent. We figure this might be due to the relatively imprecise collection method.

Future Research Directions An element of our data collection techniques revolves around the idea of asking participants to do more, so designers need to do less. For example, in Case 1, we asked fellow soccer team members to use their idle time to tag performance data for other team members. In Case 2 and 3, we relied less on participants' time in collecting data. We believe it is necessary to strike a balance in the process. On the one hand, we believe it is acceptable to ask participants to make an effort in participating in studies. However, we see two important principles in this: First, participants benefit of participation should match their effort. Second, the amount of effort they put into a task, should not greatly exceed the effort the researcher or another professional would use to perform the same task, unless solving the task is the object of study (which for data collection purposes, is not the case).

While quantified self is typically driven by a particular motivation and goal, such as understanding a personal phenomena or learning about causes of such phenomena, the process in which to arrive at answers is often characterized by being exploratory. Self tracking spans a wide number of topics [6] and data visualization is the typical mean to reflect on self collected data [16]. However, self collected data often result in novel data sets for which no standard visualization is readily available, thus leading self trackers to develop custom visualizations that fit their particular purpose. While earlier work has proposed high level heuristics for visualization of self collected data [7], we suggest that our proposed method could assist early in the self-tracking process by informing self trackers how possible self collected data and visualization of those could assist the self tracker in reaching answers to the questions posed.

7. CONCLUSION

We have argued for the need and use of methods and techniques to collect data, to allow early design work to be based on real data, for design ideas for which no data existed. We have contributed methods that help researchers and practitioners go from idea, over prototype design, to first evaluation results in a short time (e.g., less than a day), and have illustrated this in three concrete cases. The described methods imposes limited work on designers and participants. Further, they are both intended to be used concretely and to inspire similar methods and techniques. We believe the methods and techniques are particularly relevant for personal visualizations researchers. Finally, we identified a number of challenges related to collecting data and conducting evaluation in early design work, and outline limitations of the approach, both illuminating potential problems in collecting data, and potential solutions.

8. REFERENCES

[1] R. Amar and J. Stasko. A knowledge task-based framework for design and evaluation of information visualizations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 143–150. IEEE, Oct. 2004.

[2] T. Boren and J. Ramey. Thinking aloud: reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3):261–278, Sept. 2000.

[3] M. Brehmer, S. Carpendale, B. Lee, and M. Tory. Pre-design Empiricism for Information Visualization: Scenarios, Methods, and Challenges. In *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation Methods for Visualization*, BELIV '14, pages 147–151, New York, NY, USA, 2014. ACM.

[4] S. Carpendale. Evaluating Information Visualizations. In A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North, editors, *Information Visualization*, number 4950 in Lecture Notes in Computer Science, pages 19–45. Springer Berlin Heidelberg, 2008.

[5] S. Carter and J. Mankoff. When Participants Do the Capturing: The Role of Media in Diary Studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 899–908, New York, NY, USA, 2005. ACM.

[6] E. K. Choe, N. B. Lee, B. Lee, W. Pratt, and J. A. Kientz. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1143–1152. ACM, 2014.

[7] A. Cuttone, M. K. Petersen, and J. E. Larsen. Four Data Visualization Heuristics to Facilitate Reflection in Personal Informatics. In *Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice*, pages 541–552. Springer, 2014.

[8] M. Czerwinski, E. Horvitz, and S. Wilhite. A Diary Study of Task Switching and Interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 175–182, New York, NY, USA, 2004. ACM.

[9] B. Gaver, T. Dunne, and E. Pacenti. Design: Cultural Probes. *interactions*, 6(1):21–29, Jan. 1999.

[10] S. Greenberg, S. Carpendale, N. Marquardt, and B. Buxton. *Sketching user experiences: The workbook*. Elsevier, 2011.

[11] D. Huang, M. Tory, B. A. Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. Personal Visualization and Personal Visual Analytics. *IEEE TVCG*, 21(3):420–433, 2015.

[12] B. Jackson, D. Coffey, L. Thorson, D. Schroeder, A. M. Ellingson, D. J. Nuckley, and D. F. Keefe. Toward Mixed Method Evaluations of Scientific Visualizations and Design Process As an Evaluation Tool. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization*, BELIV '12, pages 4:1–4:6, New York, NY, USA, 2012. ACM.

[13] J. Kjeldskov, M. B. Skov, and J. Stage. Instant Data Analysis: Conducting Usability Evaluations in a Day. In *Proceedings of the Third Nordic Conference on Human-computer Interaction*, NordiCHI '04, pages 233–240, New York, NY, USA, 2004. ACM.

[14] S. Kvale. *Doing Interviews*. Qualitative Research Kit. SAGE Publications, Thousand Oaks, CA, 2008.

[15] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, Sept. 2012.

[16] I. Li, A. Dey, and J. Forlizzi. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 557–566, New York, NY, USA, 2010. ACM.

[17] D. Lloyd and J. Dykes. Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2498–2507, Dec. 2011.

[18] M. Monroe. Classic Techniques in New Domains: An Alternative Recipe. In E. Bertini, N. Elmqvist, and T. Wischgoll, editors, *EuroVis 2016 - Short Papers*. The Eurographics Association, 2016.

[19] J. G. Pedersen, T. Herdal, and S. Knudsen. Designing Information Visualizations for Different Levels of Comprehension. In *Proc. NordiCHI '16*, 2016. Accepted for publication.

[20] Z. Pousman, J. Stasko, and M. Mateas. Casual Information Visualization: Depictions of Data in Everyday Life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1145–1152, Nov. 2007.

[21] A. J. Pretorius and J. J. V. Wijk. What Does the User Want to See? What do the Data Want to Be? *Information Visualization*, 8(3):153–166, Sept. 2009.

[22] J. C. Roberts, C. Headleand, and P. D. Ritsos. Sketching Designs Using the Five Design-Sheet Methodology. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):419–428, Jan. 2016.

[23] P. Saraiya, C. North, and K. Duca. An insight-based methodology for evaluating bioinformatics visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):443–456, July 2005.

[24] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2431–40, 2012.

[25] C. Snyder. *Paper Prototyping*. Morgan Kaufmann, 1st edition, 2003.

[26] A. Strauss and J. Corbin. *Basics of qualitative research. Techniques and procedures for developing grounded theory*. SAGE Publications, Thousand Oaks, CA, 2. ed. edition, 1998.

[27] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen. Getting the Right Design and the Design Right: Testing Many Is Better Than One. In *Proc. CHI '06*, pages 1243–1252, 2006.

[28] J. S. Yi, Y.-a. Kang, J. T. Stasko, and J. A. Jacko. Understanding and characterizing insights: how do people gain insights using information visualization? In *Proc BELIV' 08*, page 4. ACM, 2008.