# Interactive Non-Expert Information Visualizations and their Evaluation Beyond Time and Error

Johanna Fulda

**Abstract**— In this paper we look at the current state of interactive information visualizations and how they can be evaluated beyond time and error. We focus on non-expert visualizations that contain interactivity for the user and can be found in the environments of news websites, museums and public spaces. For each environment particularities and examples are illustrated and the different categories of interaction described, including for what purposes they can or should be used. For evaluating those graphics we look at the advantages and different use cases for "Lab Studies" and "Into-the-Wild Studies", and explain why qualitative and quantitative methods both are equally valuable. Furthermore we point out three aspects that should be considered when evaluating those visualizations: 1) The inclusion of the building process into evaluation, with user testings as well as expert analyses, 2) Rating the discoveries a user makes during exploration, and 3) The advantages of long-term studies.

**Index Terms**—Infovis, Evaluation, Visualization, Interactive Information Visualization, Evaluation methods, Non-Expert Information Visualization, Lab Studies, "Into-the-Wild" Studies

## 1 Introduction

If we consider previous research in the evaluation of information visualization (Infovis), there are two main components being considered, which are representation and interaction [39]. There is still more focus on evaluating the representational part [11], but with the increased dissemination and the facilitated possibilities to create interactive visualizations it is becoming more important to also focus on the interaction part and try to understand how people use it and how they gain insight. Of course these two parts go hand in hand, because also the interactive visualization needs to have a clear and understandable representation to make it usable. The interaction part however is still underrepresented in scientific research papers so far [39].
To first understand the advantage of the opportunity to interact with infographics, Endert et al. explained how it helps understanding complex and extensive datasets [12]. That is because the user can visually explore data, make decisions independently and navigate through the available dataset corresponding to his interests. Especially for the interactive "non-expert" graphics considered here, also the playful aspect is important, because the user has to be engaged and motivated to deal with the graphic. So the question arises how we can evaluate the quality of those interactive visualizations. For example they can either contain a lot of useful information but being ignored by the users because they don't catch their attention, or they are just not recognized as something to interact with and offering information - that can particularly happen with installations in public spaces [16]. Another possibility is, that it invites the users to play, but misses the informational part because the information isn't easily accessible or it just doesn't provide a lot of it. To regard all these circumstances for the evaluation, it isn't enough to only look at the time and error rate in a user study, but we also have to consider how likely it is that the graphic catches the users' attention and prevails upon them to interact with it, how much insight a user gains and how he understands using the provided possibilities of interaction. Therefore we look at established evaluation methods and their benefits. We also point out some aspects that should additionally be considered when evaluating interactive non-expert visualizations. Before that we define the addressed visualizations more precisely, present some examples and show different categories of interaction.

- *Johanna Fulda is studying Media Informatics at the University of Munich, Germany, E-mail: johanna.fulda@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Special Aspects of Usability', 2014*

## 2 Considered Visualizations

Information visualizations can be used in a lot of different areas. For this paper we narrowed down the target group and decided to focus on interactive "non-expert" visualizations. Those who can be found in three different environments. The following explains the details and shows some examples.

### 2.1 For "non-experts"

Many information visualization systems are considered to visualize a very specific research field and thus help experts doing analyses of their highly specialised work. But there are also use cases for Infovis systems that aim for explaining a general topic, where the target group is people with various backgrounds and different fields of expertise and interest [28]. Most of them are thus no experts in the specific field of the visualization. These "non-expert" visualizations are therefore rather for giving an overview of something and do not intend to dive into the topic too deeply. Even though the interactivity enables further exploration, it has to be assumed that the user is new to the topic and explain everything from scratch.
In general the use of those visualizations is voluntary and they often are just supplemental offerings. That means it first has to get the users attention and look interesting enough to motivate further investigation. There are a lot of factors influencing that decision, for example time constraints, familiarity, mood, age [18]. If the visualization attracted the user, the possibilities of interaction should be easy to discover. They should either be recognizable from previous experience and correspond to a "usual" way of interaction - as known from web applications or native apps - or provide the user with a challenging or creative task [18]. But all interactive elements have to be identifiable, more advanced interaction methods should be explained shortly. If it looks too complicated it might scare off the users, so there has to be a trade-off between looking interesting and being manageable [6].

### 2.2 Environment

These non-experts visualizations occur in many different environments. To narrow down the field of application, we want to look at three different cases, as there are news websites, museums/exhibitions and public spaces. Below, all three environments are explained in more detail:

#### 2.2.1 Interactive graphics at online news websites

Since Internet technologies, such as browsers or native apps offer many possibilities for interaction, it challenges the traditional "one-way directional flow of news" [9]. The readers can choose themselves what they are interested in und don't have to follow the guidelines

from the news producers [8, 17, 35]. They can choose from endless offers and even get the chance to participate interactively in different ways. One kind of interactivity is interactive graphics, which can be associated with an article, provide additional information or give an overview over a certain topic. There are several flagship examples from big online newspapers, like *The Guardian* or the *New York Times*. Together with the new species of data journalists they try to realize big interactive data projects more and more often.
In contrast to the traditional newspaper, that people often read calmly for example during their breakfast or while riding a train, the purpose of reading online news is mainly to get informed or updated very quickly, for example during a short break at work - but it can happen several times a day [17]. When online news users enter the landing page of the website, they find an overview with a huge selection of articles. The goal of the users is to find something that is interesting for them. Again in contrast to the traditional newspaper, where the reader at least scans the content and layout of most articles by flicking through the pages, the online news only show a small teaser of the article or even just a headline to invite to further reading. That means that in the end most texts there will never be opened [17]. So the first challenge for the interactive graphics is to be realized at all. They have to catch the readers attention and motivate them to follow the link. The second one then is the understanding of the graphic. The readers have to understand what the graphic shows, how they can interact with it and what advantages they get by doing so. And all this should be visible without the need for an extensive explanation [8].

Figure 1 shows an example of an interactive graphic at *sueddeutsche.de* (the online presence of the *Süddeutsche Zeitung*), which was developed together with *OpenDataCity* - the ”Zugmonitor”. It visualizes data on a map from the German Railway Company in real time and gives an overview of all trains and indicates those who have a delay. The view can also be predated, and show data from a former day. The data is obtained through data scraping from the train company's website, thus it unfortunately is dependent on their data structure and won't work anymore when the company decides to change it (as it happened in autumn 2013). It still is a good example for the possibilities of interaction a news website visitor can have.

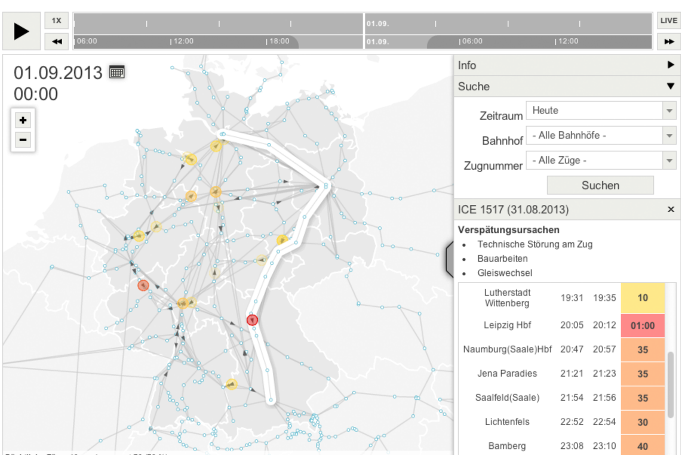


Fig. 1. Screenshot of the ”Zugmonitor” from sueddeutsche.de [33].

### 2.2.2 Museums and exhibitions

Increasingly augmented with digital technology [18] museums often offer interactive Infovis systems to their visitors. Those provide deeper insight or background information to the exhibited pieces. Care has to be taken that people visit museums for very different reasons, as for example to ”add to their specialized knowledge”, or for ”an entertaining and educational experience” [15]. That means on the one hand we cannot assume that visitors have prior knowledge in a certain topic and thus the installation should provide introductory information. On the other hand it should also be interesting for those who already have specialized knowledge and offer more profound information. Furthermore it is not very likely that a visitor will use the installation more than once [18] so it has to evoke their curiosity immediately as well as be understandable intuitively. But at least we can ascribe the visitors a certain interest for the exhibition.

As examples the *Austrian Technical Museum* in Vienna had an installation at their exhibition medien.welten (in summer 2013), where visitors could do calculations like on a real Abacus. The system evaluated the input and offered feedback and instructions if necessary [18]. Figure 2 shows another example, the ”The Virtual Fraunhofer Spektralapparat” at the *Deutsches Museum* in Munich. An interactive 3D model is displayed next to the original piece. The visitor can turn it around in every angle and gets additional information about the functionality of the Spektralapparat, like which part does what and how. This approach brings the visitor way more insight than if he was just looking at the model [21].

Fig. 2. The Virtual Fraunhofer Spektralapparat at Deutsches Museum Munich [21].

### 2.2.3 Public spaces

Public spaces are cluttered with displays nowadays. The vast majority contains passive content like advertisement, directions or restraints. Digital screens sometimes also display location-dependent information, such as the time one has to wait for the next bus or train, the local news in an underground station or redirections because of traffic jams on the road [10]. There are a few examples of visualizations that include the urban environment and shows information that is generated by the people around [23]. In Copenhagen and Aarhus for example, there are some ”Cykelbarometers” placed next to much-used cycle ways. They count and display the number of bikes passing by through the day. In New York the ”National Debt Clock” shows the American government's debt on a huge display and also indicates how much that is for each American family [23]. (In 2008 it ran out of digits [4].)
Only very few attempts have been made for installations where people have the possibility of ”hands-on” interaction in public spaces, partly because the displays still are a case of cost and maintenance, but above all because it is expensive and time-consuming to offer interesting and interactive content that matches all the requirements these installations have [2].

2009 the MIT SENSEable City Lab developed a design for a futuristic bus stop as seen in figure 3. The ”EyeStop” that is partially covered with touch screens and sensors, offering people to plan their trip (through entering their destination and getting a map with the shortest way), to see where the bus they are waiting for is right now or to write or read local announcements (and much more). Although it is not implemented outside in a real public space yet, it surely is a showcase project for ”the potential of next-generation urban transportation design” [25].

Fig. 3. EyeStop, developed by MIT’s SENSEable City Lab [25].

One less futuristic example is the installation ”Vote With Your Feet” from Steinberger et al.. A ”hyperlocal public polling tool for urban screens” [13] that is meant to engage people in thinking about certain topics and to show the community’s attitude towards it. With two buttons on the ground, the participant can answer questions with ”Yes” or ”No” and thus be part of a survey.

But the few interactive installations in public spaces seem to have additional problems with getting noticed and accepted. Especially in cluttered public spaces it is at first challenging to get people’s attention and to make them pause for something they don’t know and don’t see the immediate benefit of [24]. Many objects in public spaces may not be noticed at all or at least not with the main focus of a passing person. Sometimes they are seen unconsciously or as secondary task, while doing something else, which means people are paying only limited attention to these objects [32]. Furthermore the common human-computer interaction (HCI) findings for how people interact with interactive displays seem to differ a little for public spaces, as there is the factor of being in public and seen by others [24, 13, 34] and the factor that the system has to be realized before the interaction can take place [24] - HCI mostly assumes that the user already knows that there is something to interact with.

## 3 Categories of Interaction

In his book ”Visual Explanations: Images and Quantities, Evidence and Narrative” Tufte lists 58 different methods of interaction to ”winnow the wheat from the chaff and separate the sheep from the goats”, those are for example ”pair, merge, harmonize, synthesize, focus, organize” [37]. To narrow those down a little, Yi et al. made one persuasive categorization with seven methods. Those are: Select, Explore, Reconfigure, Encode, Abstract/Elaborate, Filter, Connect [39].
For the visualizations we observe here, not all of those seven methods seem equally important, because for the non-expert user (as defined above) the interaction should primarily be simple to lower the barrier to get started. If there is too much preparation necessary, it is very likely that people abandon the interaction.

As Table 1 shows exploring seems to be the most prevalent method of interaction for the non-expert interactivity. It enables the user to change to a different subset of data. Sometimes exploring reminds more of flipping through a book than being in a modern digital environment, where more fancy methods could be applied. But to also encourage inexperienced users to interact with digital devices it’s reasonable to use that method [19].

To show the applications for all the other interaction techniques, the seven methods are defined shortly and explained how they are used in the examples from Table 1:

### 3.1 Select

Enables the user to mark one item as interesting and thus lets him keep track of a transformation or get additional information for that item. In the example of the ”Zugmonitor” the user can click on one train to highlight the route it is on. It simplifies tracking one specific train without having to stare at it to not get lost in the clutter. Furthermore it lists all the stations that particular train is stopping, the scheduled time and the potential delay.

### 3.2 Explore

As already mentioned Explore lets the user change the displayed subset of data. It doesn’t necessarily change the whole view but can also just let some additional data appear and some previous data vanish. In the mentioned examples in Table 1 the interaction mostly changes the whole view.

### 3.3 Reconfigure

The spatial arrangement can be changed through reconfiguring the view, for example to be able to see more details and less clutter at a specific area of interest. Hidden information can thus be revealed and connections or rankings seen more clearly. For listings reconfiguration is used very often, to sort lists after specific criteria. In the example of the New York Times graphic, where the World Cup players are shown in different sizes depending on the number of mentions in Facebook posts during a day, the user can also reconfigure the view and either order them by the country they are playing for, by their name or by the visualized number of mentions. In the examples for the environment of a museum the reconfiguration is often used for 3D models, to enable the user to virtually turn around an object and observe it from any angle.

### 3.4 Encode

Encoding can be used to change the fundamental visual representation, like colours, shapes, the kind of representation - for example change a bar chart to a scatterplot. This technique doesn’t appear too often in the considered Infovis systems, therefore there is also no example for it in Table 1. The reason may be that encoding requires the user to already have a deeper understanding of the represented dataset, to be able to purposefully change the representation to a more suitable one for that specific dataset. It could surely also be used for playing around and thus accidentally discovering new features, but it probably is more useful for expert visualizations.

### 3.5 Abstract/Elaborate

Especially for big datasets it is impossible to show all information in one screen and very often it is also not necessary, because the user only wants an overview without going deeper into detail or he is only interested in a single aspect. Abstracting and Elaborating the infographic allows the user to adjust the degree of abstraction - to either go deeper into one aspect or go further away to see more of the overall image. At *The Guardian*’s ”Woman’s Rights” graphic, the reader first gets an overview of all the continents and their averaged data. If he then selects one continent, the graphic goes deeper into the data of that specific continent and shows details for the single countries inside. Here the colours and their saturation indicate the positioning of one country in the seven represented categories. If the user again selects one of those sections, a listing for the legislation behind that category is shown next to the graphic, displaying a tick or a cross for being in the country’s law or not. In that graphic there are three layers of abstraction. Many more can for example be found on geographical maps where data can either be seen for the whole world or alternatively for one street inside a city (and many layers in between).

| Environment | Name / Place | Underlying data | Tasks | Explanation |
|---|---|---|---|---|
| Websites | **Zugmonitor** Süddeutsche Zeitung (zugmonitor.sz.de) | data scraping from Deutsche Bahn website | Select, Explore, Filter, Connect | Visualizes long-distance trains in Germany and their delays. Live and historic. (Currently out of order since the German train company changed their data structure in autumn 2013) |
| | **Top World Cup Players on Facebook, Day by Day** New York Times (nytimes.com/interactive) | Number of mentions of players through Facebook API | Explore, Reconfigure, Encode | Pictures of the most discussed World Cup players are sized corresponding to how often they were mentioned on facebook posts (for each day of the World Cup 2010) |
| | **Women's rights, Country by country** The Guardian (theguardian.com) | Excel Sheet with data collected by World Bank and UN | Explore, Reconfigure, Abstract/Elaborate | Illustrates women's rights across the world. Details for single countries in seven categories and the legislation behind them (February 2014) |
| Museums/ Exhibitions | **Fraunhofer Spektralapparat** Deutsches Museum | 3D model of Spektralapparat and associated explanations | Explore, Reconfigure | Digital 3D model next to the original artefact to let visitors discover the piece more detailed |
| | **Abacus** Technical Museum Vienna | Exercises generated by museum | Explore | Digitally-augmented Abacus that lets the visitors do calculations on it and offers feedback and instructions |
| | **360grad Electric** BMW World in Munich | 3D model plus additional information | Explore, Reconfigure, Abstract/Elaborate | Shows a BMWi model and background information with slick animations |
| Public Spaces | **EyeStop** MIT SENSEable City Lab | APIs to public transport, news, etc., plus data from own sensors | *Everything possible* | Solar powered future bus stop that should also be used as a community gathering space |
| | **Cykelbarometer** Copenhagen, Denmark | Generated by people | Other (more participation than interaction) | Displays the number of how many cyclists already past that point that day |
| | **Vote With Your Feet** Brisbane, Australia | Generated by people | Select | Public polling tool installed at bus stop asking people Yes/No questions. Answers are given through two buttons on the ground |

Table 1. Examples for different environments

### 3.6 Filter

Filtering means the user can add or remove criteria for the displayed data and thus only see the matching information. He can for example select a range of time, because it is assumed that in that range something interesting happened. All the other information is not interesting at that moment and thus doesn't have to be displayed. In the "Zugmonitor" the user can filter the data for specific dates or time frames or he can choose one specific train station or one specific train number and get information about delays only for trains that match that criteria.

### 3.7 Connect

Connecting representation can either be used when one dataset is visualized in different ways in more than one view - to highlight one selected item also in all the other views. Or it can be used for highlighting similar items to show connections that otherwise would have remained unseen. Again in the "Zugmonitor", if the user clicks on one train its route is highlighted and thus shows the connected stations. That gives the user additional information because with the many routes overlapping and cutting across each other, it is impossible to recognize one specific route if it is not highlighted.

As table 1 shows not every kind of interaction fits into those definitions. Especially in the environment of Public Spaces it's not necessarily only touchscreens the user can interact with, but different objects that for example augment the visualization, make it more special or artistic.

## 4 Common Evaluation Methods

There are many well-established and valuable methods for the evaluation of information visualizations. They are mostly well documented and the creators of the study can use guidelines and checklists to prepare their tests. The types of studies reach from simple task solving exercises to more complex observations of unknowing users. There are quantitative and qualitative methods that can either take place in a lab or outside in the settings where the applications will actually be used when finished. Advantages and disadvantages of both environments will be explained and why assertors of qualitative and those of quantitative methods don't have to fight any longer. But first we discuss why we need more elaborate approaches than "the time and error approach" at all.

### 4.1 Problems with the time and error approach

If there is a classical way of evaluating information visualizations, then it probably is the one where a test user is asked to deal with a certain task and then the observer watches how long it takes him and how many errors are made in the meantime [30]. This approach is sufficient for simple visualizations, but with more and complex data and many different possibilities of interaction it becomes less suitable. Furthermore tools that visualize big data are mainly made for "innovation and discovery" [31] so it doesn't make sense to let the user do one specific exercise, but rather observe how he is trying to gain insight and discover the data. To "quantify the quality of a visualization system" [1] we thus need methods beyond time and error.

### 4.2 Lab Studies

It is very common to evaluate new things in the environment of a lab, because there the researcher has the necessary equipment and all the test users are in the same situation when they do the tests. Also ethically it is flawless because the users know that they are part of a study and can thus not feel spied on afterwards.

#### 4.2.1 Usability testing

The classical usability test requires three to ten test users that are given a set of typical tasks. The observers watch them to see where they run into trouble and write a report afterwards with all the identified problems, possibly ranked by importance. Then the product is being improved and a new usability test created. Through these many iterations it is possible to improve the understanding and the operationality of a product. And it can respond to the needs and wishes of the end user in a very effective way. Disadvantages are that the environment is different than the one the system would usually be used in. Also the devices may be different from the ones used normally. Improvements can be to relocate the tests into the work environment (including all the distractions and interruptions one normally has there) and to allow the test users to utilize their own devices [31].

#### 4.2.2 Questionnaires, Talking Aloud and Thinking Aloud

Additional to only looking for obstacles the user is facing, one can ask the test person a lot of questions afterwards, why he decided to do it this or that way. More live and unfiltered feedback one can get by asking the user to say everything he is doing out loud, that also reveals the consciousness about decisions and if everything is understood as it was supposed to. The "Thinking Aloud" method requests the user to talk even more. He has to speak out everything he is thinking while finishing the exercise. That can reveal indeterminations of what to do next or show how certain elements are received and if they are recognized as what they are supposed to be. But because the talking task demands a lot of concentration and users also tend to pay more attention to navigation problems than they normally would, those tests can not at the same time be used for testing the time needed for completion and thus the efficiency [38]. They are also highly subjective and can't be evaluated with an algorithm or some mathematical formula but have to be assessed individually (see 4.4).

#### 4.2.3 Cognitive methods

To not only rely on what the test users verbalize, there are methods where the accurate behaviour can be tracked and thus even unconscious activities can be noticed. Methods are for example "Eye Tracking", measuring the brain activity or even more physiological responses like muscle activity. All these measurements need quite expensive hardware, so it can't be expected to have them available for every project. If they are available, again the final evaluation is another big challenge, because all the recorded signals have to be interpreted. Therefore the researchers need special knowledge in neuroscience or have to collaborate with experts. It also makes sense to already involve experts during the design phase of the study, the results may be less cryptical then for non-neuroscientists - but still need some extra knowledge [3].

### 4.3 "Into-the-Wild" Studies

For the interactive non-expert visualizations we defined above, the previous described methods are not sufficient, because they only consider the last of the three phases. The first important phases of the evaluation of those visualizations is, that they have to be discovered and approached, before the interaction finally takes place. So we can not just invite test persons into a lab and place them in front of the system, but we want to know if untaught people do realize the existence of the systems at all and if so, if it makes them curious enough to actually approach them. That means we actually have to do tests outside "in the wild" to explore people's behaviour.

#### 4.3.1 Awareness and Motivation

There are at least three phases that have to be gone through for a successful interaction [24, 16] - some literature even mentions more phases (for example Michelis et al. name six in their "Audience Funnel" [22]), but those are the ones that seem to be most important for us:

- **Attention**: The first step is to be realised by passing people
- **Motivation**: The phase of understanding what it offers and deciding if it is worth trying
- **Interaction**: The actual phase of interaction and gaining insight

As mentioned above, we have to leave the lab environment to be able to evaluate the first two phases. For news websites it may even work to explore the user's behaviour in a lab, if the user doesn't know about what is there actually to discover, but the testing environment can bias the behaviour in that way, that the user is more attentive than he would normally be, when browsing a news site. So we want to find out, if the visualization is being discovered without the user having any expectations. That means we let unknowing people approach and interact with the installation and can start asking questions afterwards.

#### 4.3.2 Observation and Interviews

In the environment of a museum Hinrichs et al. chose a "qualitative ethnographically oriented study method" as described by Blomberg et al. in the book "Participatory Design - Principles and Practices". People were informed through a sign that a study was being conducted in the room to harm nobody's privacy. Even though they didn't use video or audio recordings and only took notes of their observations. The observer was also sitting quite far away from the installation, so they hoped there was no interference with the people's behaviour. After interacting with the object, the visitors were asked to fill out a questionnaire concerning their experience with the installation and if and how they gained insight through the visualization. Afterwards the researchers had to analyze all the notes and discoveries. Therefore they used the "open coding method" (the term was characterized in 1989 in the context of Social Sciences [5]) which means creating a category system based on the results of the observation. The categories, as many as necessary, are generated freely and afterwards grouped together and given more abstract labels, then grouped together again and so on [7]. The results then showed different types of visitors, different behaviour if they were alone or in a group and how long and intense they interacted with it. They concluded that the motivation for approaching the installation was "the display technology, the visual appearance of the visualizations, and seeing other people interact with it" [15].

#### 4.3.3 Hidden cameras

It often makes sense to capture the situation with a camera, to enable looking at it again afterwards. There are some rules the researchers have to stick to, to not infringe the test person's rights but if they keep those conditions it is a useful tool. It makes it even possible that the observers are at another place during the test phase, to not influence the behaviour of the passing people. For example Steinberger et al. used that method to evaluate their interactive bus stop. Like in the previous example they also observed the people (the ones using the object as well as the people watching) and tried to interview them after they finished their interaction with the object. Here the motivation for approaching the installation was either an interest in the displayed topic or it was seen as a nice occupation while waiting for the bus. Due to the fact that it offers a very simple task (one Yes, one No button) it was perceived as very easy and understandable [13].

### 4.4 Qualitative vs. Quantitative

All these observation, interviewing people or letting them fill out questionnaires entail a long post processing phase. They involve a big interpretive part and collection of empirical data, sometimes even philosophical questions. These qualitative methods "seek to make sense of personal stories and the ways in which they interact" [27, pp. 1]. In contrast to that there are the quantitative methods, which use numbers and statistical methods to produce results that are generalizable. "In quantitative research, the researcher's role is to observe and measure, and care is taken to keep the researchers from 'contaminating' the data through personal involvement with the research subject" [27, p. 6]. In many older literature about research methods, scientists were almost fighting about which of those methods is better. Positivists "regard the world as made up of observable, measurable facts", whereas the interpretivist "portrays a world in which reality is socially constructed, complex and ever changing" [27, p. 8-9]. Nowadays scientists agreed largely on both methods being useful for specific needs. And it is also common to use them both, for example enriching collected numbers with user interviews or to also make a survey with predefined answers during a qualitative study [36, p. 3-6].

For the considered Infovis systems the qualitative approach often is more convenient. The creators want to know how people perceive the objects and what exactly motivates them to actually approach them or rather not. Also because it is a not too established field of Infovis, they want to get new insights, suggestions and ideas, to keep improving the systems. And those insights may not be caught by quantitative methods. An exception can be the evaluation of online graphics. Because here the persons responsible often don't want to spend too much time for evaluation, so they draw back to the data that is generated for their website anyway. That is number of visitors, number of clicks, time they remained on a page, and if available some demographic values (age, gender, field of interest), where they come from (geographically and from which previous website), if they are new or returning visitors and so on. With that data they can see if a visualization is very popular or if it seems to remain rather undiscovered. The data about how long someone spends with it reveals insight too, but the reasons, the exact behaviour, and understanding cannot be caught through that data. A tool for getting qualitative feedback on websites (without having to prepare a study) is the comment function they often offer. People who are particularly excited about the article, or feel the need to contribute something get the chance there - however if that feedback is used to improve future graphics cannot be answered explicitly.

Many of the portrayed methods are established and well documented and can deliver qualitative as well as quantitative data. But especially for the considered interactive non-expert visualizations there are some ideas for additions to common evaluation methods that still could improve the results. In the following chapter we take a closer look at those ideas.

## 5 Required features

The previously portrayed methods are used very often and there are also many detailed instructions available. "A method of analysing interview transcripts in qualitative research" [7] for example guides through all the steps of a qualitative evaluation, the book "Blending qualitative and quantitative research methods in theses and dissertations" breaks it all down and offers a whole "Catalogue of Methods" [36]. Still there are some features missing in current evaluations, three suggestions are described in the following. The inclusion of the building process and the measuring of discoveries can be performed in lab study environments, Long-term studies would then take place in the final environment of the system.

### 5.1 Include building process

There are two aspects that should be evaluated already and repeatedly during the creation process of the interactive visualization. That is the interaction (usability and understandability) and the appeal (that includes aesthetics and the user's experience). Especially usability and experience overlap very often and the effects of aesthetics are sometimes only perceived unconsciously by the user. That is why next to testing the prototypes with users it also makes sense to include interaction design experts in the building process [14]. Through their experience and knowledge of guidelines they can help to avoid making mistakes that have been made before. Both suggestions are very time consuming and expensive but can avoid a lot of post-treatment.

#### 5.1.1 User testing

As mentioned earlier the non-expert visualizations have to be very easy to understand. Some projects have been portrayed that have the simplest possible interaction methods (for example only 2 buttons). But the bigger and more complex the data gets, the more possibilities of interaction could be used. Therefore a balance has to be found between using as many handy methods as possible and still being usable intuitively. But very often developers and designers lose their neutral view during a project and tend to overlook potential problems, because it seems just very clear to them. As idea for evaluating the balance between interesting and simple, it would make sense to already include the building process into the evaluation [20]. As described for the usability-tests, one could regularly invite few test users to see if they understand the current possibilities of interaction and based on that continue the development. That would avoid having something totally magnificent in the end that people won't use because they don't understand it. As described by Hinrichs et al. "a visually appealing information visualization can be experienced negatively if it is hard to explore due to awkward interaction techniques" [15].

#### 5.1.2 Expert testing

Additionally to evaluating the user's behaviour, experts can be asked to evaluate the overall usability, user experience and fluidity [14]. That also should happen during the building process to avoid too many adjustments afterwards. There are a lot of rules and usability design guidelines that help creating a system. Experts in the field of interaction design can identify potential problems that users might face during the usage of the system. Afterwards they can help improving them due to their experience and knowledge. Nielsen suggests to include three to five experts to identify the most relevant problems [26]. As this paper was considered looking more at the interaction part of Infovis, not the representational one, we mainly ignored the aesthetics part of our installations. But when it comes to interactivity, the user experiences the behaviour of the whole installation, and that goes closely together with the aesthetics. Our systems involve the "first-approach-moment" and they are only used voluntarily. So for attracting people to approach them at all they have to have a certain appeal. Interaction designers will also consider the effects of aesthetics and fluidity [11]. Besides getting information out of it, the user also wants to be entertained and enjoy the newest possibilities of technology. "Users should get a feeling of immersion, first-personness and direct engagement with the objects and the visualizations" [11, p. 4]. If an animation doesn't react immediately, gets stuck all the time or isn't visually appealing the chances of people abandoning the installation raise.

### 5.2 Measuring discoveries

Even though we focused a lot on the approaching process, gaining insight should still be the crucial part of information visualizations. But how can we measure the insight when we actually want the users to independently discover the provided data, rather than asking them to do one specific task? Saraiya et al. defined it in the following way: "More, valuable, faster, and deeper data findings correspond to more effective visualizations as it suggests users can gain more insight from the data" [29]. Even though their research concerned Bioinformatics Visualizations, they claim that definition to be domain independent. To measure the value and deepness of the findings, they suggest that the experimenters could either use self-reporting and ask the users to rate their own findings or to ask domain experts to rate them [29]. In our cases we think self-reporting could be sufficient, because the data of the installations is mostly preselected and thus it is less likely to discover something completely new (exceptions are possible). So the users could be asked to rate the importance of their findings on a scale, or the evaluators could, due to their expertise in that topic, define the value of a certain finding. The test user thus would achieves a score in the end. Both attempts would yield quantitative results and thus make the evaluation process quicker and less subjective. But until now it is only a suggestion, because there haven't been scientific studies for our specific use cases yet, to prove that method valuable.

### 5.3 Long-Term studies

The qualitative methods performed during "Into-the-Wild" studies (as mentioned above) seem to yield quite vague results sometimes. The results are strongly dependent on the persons being interviewed and finally the interpretation of the researcher. Also prototype installations often look more like art objects or alien elements especially in public spaces. But as the goal of the visualizations is to inform people, we want to know how much insight they gain through them [29]. And to get an as less distorted result as possible, the user has to be familiar with the system first [30]. Therefore it seems reasonable to observe them over a longer period of time. People then can get used to it and approach them less sceptical. So especially for the evaluation of installations in public spaces that approach should be considered more often. People passing a place regularly, might be in a hurry or just sceptical in the first place, but would maybe approach it another time. However for museums, where people rarely pass by more than once and for news sites, where the topics change really quickly long-term studies are rather inconvenient.

## 6 Conclusion

There are many great examples for bringing information to people in a digital and interactive way, but still there is a lot of potential for that number to grow. Creating them is connected with a lot of work and the benefits are controversial. For example generating an interactive information visualization for a news website involves editors, designers, programmers and often also a data expert. So it is only profitable for bigger topics that attract many readers and in general also only for bigger companies who can afford experimenting with those tools. But we can see a trend in more and more tools appearing that help building visualizations of data, and are not exclusively accessible for programmers. There is for example "InstantAtlas", "Leaflet", Visual.ly", "Dygraphs", "Google Charts" and many more. Because nowadays also the general "non-expert" is used to computers, the interaction on websites is the most convenient one for everybody and thus creators of interactive graphics can experiment more with new techniques than would be advisable for a museum for example.

For the evaluation there are a lot of established methods available. For the non-expert visualizations considered here it turned out that qualitative methods can bring the most benefit and they can also be combined with quantitative data. But still some improvements could be made when it comes to "Into-the-Wild" studies, which are a useful extension to the isolated lab studies. Even though the lab studies are also important and valuable in many cases, especially for the public installations it is necessary to test them in their actual environment and with untaught people. That again raises some ethical questions, like if it is acceptable to film random people in the public and to invade their privacy by approaching them with a lot of questions. There are some rules that for example the "Human Research Ethics Committee" established and if the researchers are decent and unobtrusive it can be acceptable for all participants.

I intentionally left out the discussion about all the different devices for visualizations and focussed on tabletop displays for museums and desktop computers for news sites. News graphics should ideally also work on mobile devices, but those are mainly still limited in speed and performance so the visualizations can't always be expected to have the same behaviour there as on a desktop computer - also the screen size limits the possibilities of interaction. For public spaces there doesn't seem to be one consistent kind of device yet, as the whole field still is in the phase of development and the rare installations try to use inventive and extraordinary input devices to increase the adventurous experience.

I also didn't go into much detail about how exactly the analysis of the evaluation results is being made. There is a lot of literature for that and especially for the qualitative methods there are some philosophical questions that have to be answered and it is dependent on the interpretation of the particular observer. It may be said that it makes sense to include some more people into the process of analysis to avoid being too influenced by one opinion.

The research also indicated that visualizations in public spaces are not too common yet and differ in many points to those in a museum or on a website. The differences are for example: the target group, the displays, the input devices, the robustness. On a public space one cannot tell a target group at all. Even though museums and websites also are accessible for everybody, it mostly attracts people who are interested in the specific field the museum or website shows. And there is also the problem of maintenance and vandalism but there might be possible alternatives found to build robust and immune displays that can stand all the resistance outside, it surely would be nice to see more public interactive installations in the future.

## References


[1] About BELIV 2014 (workshop series focusing on the challenges of evaluation in visualization). `http://beliv.cs.univie.ac.at/about.php`, Accessed: 2014-05-27.

[2] F. Alt, T. Kubitza, D. Bial, F. Zaidan, M. Ortel, B. Zurmaar, T. Lewen, A. S. Shirazi, and A. Schmidt. Digifieds: insights into deploying digital public notice areas in the wild. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*, pages 165–174. ACM, 2011.

[3] E. W. Anderson. Evaluating scientific visualization using cognitive measures. 2012.

[4] BBC News. Us debt clock runs out of digits. `http://news.bbc.co.uk/2/hi/7660409.stm`, October 2008, Accessed: 2014-06-19.

[5] B. L. Berg and H. Lune. *Qualitative research methods for the social sciences*, volume 5. Pearson Boston, 2004.

[6] J. O. Borchers. A pattern approach to interaction design. *AI & SOCIETY*, 15(4):359–376, 2001.

[7] P. Burnard. A method of analysing interview transcripts in qualitative research. *Nurse education today*, 11(6):461–466, 1991.

[8] D. S. Chung. Profits and perils online news producers' perceptions of interactivity and uses of interactive features. *Convergence: The International Journal of Research into New Media Technologies*, 13(1):43–61, 2007.

[9] D. S. Chung. Interactive features of online newspapers: Identifying patterns and predicting use of engaged readers. *Journal of Computer-Mediated Communication*, 13(3):658–679, 2008.

[10] S. Claes and A. Vande Moere. Street infographics: raising awareness of local issues through a situated urban visualization. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*, pages 133–138. ACM, 2013.

[11] N. Elmqvist, A. V. Moere, H.-C. Jetter, D. Cernea, H. Reiterer, and T. Jankun-Kelly. Fluid interaction for information visualization. *Information Visualization*, 10(4):327–340, 2011.

[12] A. Endert and C. North. Interaction junk: user interaction-based evaluation of visual analytic systems. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors-Novel Evaluation Methods for Visualization*, page 7. ACM, 2012.

[13] F. A. Fabius Steinberger, Marcus Foth. Vote with your feet: Local community polling on urban screens. In *Proceedings of the 3rd ACM International Symposium on Pervasive Displays. Copenhagen, Denmark*. ACM, 2014.

[14] J. L. Gabbard, D. Hix, and J. E. Swan. User-centered design and evaluation of virtual environments. *Computer Graphics and Applications, IEEE*, 19(6):51–59, 1999.

[15] U. Hinrichs, H. Schmidt, and S. Carpendale. Emdialog: Bringing information visualization into the museum. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1181–1188, 2008.

[16] L. E. Holmquist. Evaluating the comprehension of ambient displays. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1545–1545. ACM, 2004.

[17] K. Holmqvist, J. Holsanova, M. Barthelson, and D. Lundqvist. Reading or scanning? a study of newspaper and net paper reading. *Mind*, 2(3):4, 2003.

[18] E. Hornecker and M. Stifter. Learning from interactive museum installations about interaction design for public settings. In *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*, pages 135–142. ACM, 2006.

[19] F. Jabr. The reading brain in the digital age: the science of paper versus screens. *Scientific American*, 11, 2013.

[20] 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*, page 4. ACM, 2012.

[21] M. Krauß and M. Bogen. Conveying cultural heritage and legacy with innovative AR-based solutions. `http://www.museumsandtheweb.com/mw2010/papers/krauss/krauss.html`, Accessed: 2014-05-27.

[22] D. Michelis and J. Müller. The audience funnel: Observations of gesture based interaction with multiple large displays in a city center. *Intl. Journal of Human–Computer Interaction*, 27(6):562–579, 2011.

[23] A. V. Moere and D. Hill. Designing for the situated and public visualization of urban data. *Journal of Urban Technology*, 19(2):25–46, 2012.

[24] J. Müller, F. Alt, D. Michelis, and A. Schmidt. Requirements and design space for interactive public displays. In *Proceedings of the international conference on Multimedia*, pages 1285–1294. ACM, 2010.

[25] G. Niederhausern, S. Zhou, E. R. Kang, A. Biderman, and C. Ratti. MIT SENSEable City Lab's EyeStop. `http://senseable.mit.edu/eyestop/`, Accessed: 2014-06-17.

[26] J. Nielsen. Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 373–380. ACM, 1992.

[27] A. Peshkin and C. Glesne. Becoming qualitative researchers: an introduction. *NY.: Longman Publising Group*, 1992.

[28] Z. Pousman, J. T. Stasko, and M. Mateas. Casual information visualization: Depictions of data in everyday life. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1145–1152, 2007.

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

[30] P. Saraiya, C. North, V. Lam, and K. A. Duca. An insight-based longitudinal study of visual analytics. *Visualization and Computer Graphics, IEEE Transactions on*, 12(6):1511–1522, 2006.

[31] B. Shneiderman and C. Plaisant. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pages 1–7. ACM, 2006.

[32] J. Somervell, D. S. McCrickard, C. North, and M. Shukla. An evaluation of information visualization in attention-limited environments. In *Proceedings of the symposium on Data Visualisation*, volume 2002. Citeseer, 2002.

[33] Süddeutsche.de and OpenDataCity. Bahn-Pünktlichkeit - Zugmonitor. `http://zugmonitor.sueddeutsche.de`, Accessed: 2014-06-17.

[34] D. S. Tan and M. Czerwinski. Information voyeurism: Social impact of physically large displays on information privacy. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pages 748–749. ACM, 2003.

[35] D. Tewksbury. What do americans really want to know? tracking the behavior of news readers on the internet. *Journal of Communication*, 53(4):694–710, 2003.

[36] R. M. Thomas. *Blending qualitative and quantitative research methods in theses and dissertations*. Corwin Press, 2003.

[37] E. R. Tufte and S. Rifkin. *Visual Explanations: Images and Quantities, Evidence and Narrative*, volume 88. Philadelphia [etc]: Publication and Editorial Office, Dept. of History and Sociology of Science, University of Pennsylvania [etc.], 1997.

[38] L. Van Waes. Thinking aloud as a method for testing the usability of websites: the influence of task variation on the evaluation of hypertext. *Professional Communication, IEEE Transactions on*, 43(3):279–291, 2000.

[39] J. S. Yi, Y. ah Kang, J. T. Stasko, and J. A. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231, 2007.