

DATA IN CONTEXT: CONCEPTUALIZING SITE-SPECIFIC VISUALIZATION PROJECTS

George Legrady, University of California, Santa Barbara, legrady@mat.ucsb.edu.

Angus Graeme Forbes, University of Illinois at Chicago, aforbes@uic.edu.

Abstract

Site-specific data visualization installations have distinct conditions of data collection, data analysis, audience interaction, and data archiving. This paper describes features of five data visualization projects related to their successful staging within different contexts.

Introduction

The field of data visualization explores the creation and analysis of visual representations of information; data visualization projects translate abstract data into a visual form that makes it easier for viewers to perceive relationships within the data [1]. Representing data in visual form is a complex practice that can be thought of as similar in approach to the structuring of language through rules of syntax and grammar, where discrete units such as characters and words are combined to create meaning. Through the ordering of visual primitives, such as form, space, color, line, dimensions, scale, balance, texture, direction, and motion, visual representation similarly involves a *construction* of meaning [2]. Visualization is both context and goal dependent; authorial intention determines a project's emphasis on the clarity of communication and visual impact [3]. Data visualization, which began as a critical tool in scientific inquiry that necessitated the statistical mapping of data, has in the past decade crossed-over into the arts, becoming a new hybridized artform. This new genre investigates visual form through computational means, leading to more abstracted outcomes that emphasize aesthetic exploration [4, 5].

Data visualization results are typically delivered in print or on screens; user interaction is usually limited to the use of a conventional keyboard and mouse set-up, and mediated by a graphical user interface. In most cases, the visualization is independent of the context within which it is presented. Public presentation of data visualization can take either of two approaches. In the first approach, the content has no connection to the site where it is presented. For example, the location of a billboard is predetermined and the lifespan of each advertisement it displays is usually brief. In the second approach, the visual display has a specific functionality that creates an engagement with the site where it is located, perhaps featuring visualized content that originates from or reflects back onto the site. Signage that overtly announces a building's function and that has a direct indexical relation to its location is one example of this. A site-specific location, such as Times Square, may impose visual elements that, by virtue of its location, function symbolically to signify spectacle.

In the early 1970s, the architects Robert Venturi and Denise Scott Brown published "Learning from Las Vegas," a challenging, seminal book about the role of spectacle signage, commercial billboards, and electronic signs that illuminate the Las Vegas urban landscape [6]. Their analysis acknowledged that billboards, generally considered as an aesthetic detriment to the landscape, were in fact significant, integrated architectural elements, and that it was necessary to acknowledge their impact in the scenery. During the same time period, the art-world expanded its activities beyond exhibiting artworks in the enclosed spaces characteristic of galleries and museums and

began to present work in new locations outside traditional contexts of artistic and visual experiences [7]. This direction in art, to some degree stimulated by public commissions, set in motion a research-style approach to analyzing the requirements and possibilities of site-specificity. Artists became more cognizant of how their artworks were situated within a location, and of how artistic intervention would transform that site. Such an approach is well articulated by the artist John Roloff when he explains how his planning methodology involves "strategies employing inversions, intrusions, displacements, assemblages and extended analogies/metaphors, often in geologic parlance, of existing, often predictable, ecological beliefs and systems in order to disrupt, re-cast and extrapolate their epistemological, ontological and associative potential" [8].

The conceptual development of temporary and permanent site-specific data visualization projects involves extensive study of the contexts of the site for the work. This is similar to the preliminary studies necessary for the planning of an architectural work. These include, among other considerations, a review of: the function of the space; the information generated within the space; the human circulation patterns within the space at different time periods; lighting and acoustic conditions; and appropriate locations for particular kinds of activity and organization within the space.

In this article we present five data visualization installations, each of which was created principally by the first author. These projects are introduced as a means to examine the site-specific conditions of public installations and issues of staging data visualization within these locations [9]. Specifically, we describe the nature of the data that was used in each project and discuss the contexts in which it was collected, analyzed, and visualized.

Pockets Full of Memories (2001-2007)

Pockets Full of Memories is an installation commissioned by the Centre Pompidou with the intent to activate the general public to create an archive of cultural data. The audience contributes a digitized image of an object in their possession and inputs textual descriptions about that object through an interactive questionnaire. The collection of objects is continuously re-organized as new images are dynamically contributed to the archive via a self-learning mapping algorithm. The algorithm arranges objects that have similar descriptions so that they have an optimal proximity to each other. In the process of building the collection at each site, each installation results in a unique historical record and expression of the venues' specific contextual conditions.

Each contribution consists of a scanned object with associated semantic data. The object is classified via an Osgood semantic differential slider that requires contributors to provide a value for properties of the object, such as its age and to what degree it is natural, personal, fashionable, useful, functional or symbolic [11]. Each object is further described by three keywords and a longer textual description. Additionally, demographic information is requested, including the name, age, gender, occupation, and origin of the user, all of which are used to calculate the object's semantic value in relation to every other object in the collection. The contributions can then be statistically analyzed to compare subject matter and explore the variations in the semantic descriptions. During each exhibition, the collection on display is accessible both in the museum and also on the Internet, where the public can further contribute by attaching comments to any of the objects.



Fig. 1: A photo showing a gallery visitor contributing data to the *Pockets Full of Memories* installation

The visual component of *Pockets Full of Memories* is featured on a large cinematic projection that displays a 2D map divided into a 12x32 matrix of 384 cells. Four animated visualizations feature the data in different ways, with the most recent contributions highlighted by bright yellow frames. The first visualization shows the objects and their relocation to new positions as the algorithm recalculates. This is followed by a visualization that shows lines indicating displacement for all objects; the greater the number of objects in the archive, the more active the visualization becomes due to the increased displacement of objects. A third visualization presents a textual description of each object along with the name of its contributor. Finally, a fourth visualization colors each cell with a shade of grey to indicate its relative similarity to neighborhood cells. In this animation, light-colored cells represent clusters with similar values, whereas darker cells act as cluster separators that denote semantic distance, resembling mountain ranges that separate valleys.

The staging of the installation follows the traditional cinematic layout with a large screen at the front of the installation. A data input kiosk is positioned in the middle of the space, and two internet accessible computers are placed near the rear of the space to allow viewers to browse the previously input objects' metadata and to send e-messages to each object. This format emphasizes an effective spatial distribution that allows contributors to gather around the kiosk while other viewers can study the large visualization in clusters at both ends of the space. The staging is further enhanced by the way in which lighting was used to organize the installation. The front space, where the projection is displayed, is darkened to better perceive the image, while the back space is lit to allow individuals to easily position themselves. Fig. 1 shows an image of the *Pockets Full of Memories* installation, including the interactive kiosk for contributing objects and the internet-enabled computers.

The exhibition took two years to plan and produce. Production required an international team made up of multimedia engineers, graphic and interface designers, a computer scientist, and an industrial psychologist [10]. It opened at the Centre Pompidou, Paris in the summer of 2001 to a large international general audience. Approximately 20,000 people visited the installation, and approximately 3,300 of them contributed objects. Since then it has been re-staged at a number of international venues, ultimately concluding at the Museum of Contemporary Art, Taipei in 2007. These various locations

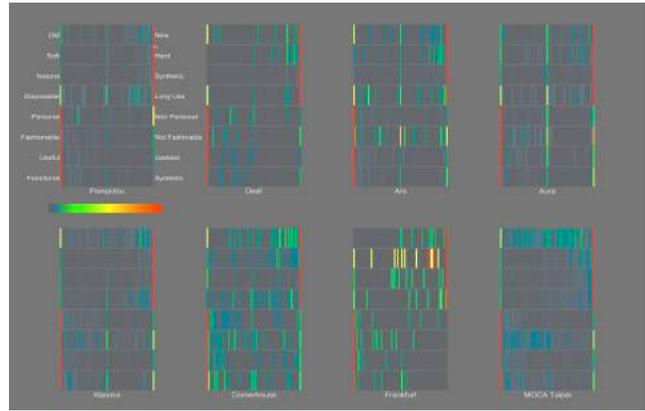


Fig. 2: A comparison of how cell phones have been classified in all the exhibition venues of the *Pockets Full of Memories* installation.

provided a broad international sampling of over 11,288 contributions [12].

Making Visible the Invisible (2005-2019)

Making Visible the Invisible is an installation commissioned by the Seattle Arts Council for the Rem Koolhaas designed Seattle Central Library. It consists of 6 large LCD screens, positioned side-by-side and located on a glass wall in an area labeled "Mixing Chamber," behind the librarians' main information desk on the 2nd floor of the library [13]. The screens feature animated algorithmic visualizations generated via custom statistical analyses of real-time data. Prototyping and production took over two years to complete and included the efforts of a team of hybrid media artists-engineers.

This artwork receives its data from the library's IT server, which provides a direct list of what items patrons check out at any given time. The item may be a book, a DVD, a CD, or a VHS tape. The database is refreshed every hour as new information is provided. This multivariate data is organized in chronological order, and a rich set of metadata is also sent to the installation, including check-out and check-in times, titles, keywords, call numbers, and unique IDs (such as barcodes or collection codes). The Seattle Central Library organizes its non-fiction items by the Dewey classification system, an ideal system for visualization due to its precise numerical categorization. Items that are classified as fiction have a less precise labeling system, but nonetheless allow for analyses of media type, subject classification, and time-based activities. Approximately 2,500 items are checked-out per day, adding up to about 900 thousand per year and totaling over 7.5 million since the artwork's inauguration in September 2005.

Four animations repeatedly cycle every few minutes, each presenting the newly received library data. The sequence of animations begins first with numeric statistics, and then proceeds to a linear chronological sequence of the items checked-out, followed by two-dimensional mapping to give an overview of how all the data in the hour have performed. The fourth animation consists of an animated, colorful visualization of the most popular keywords in the titles of check-outs; their color and position on the screen are determined by the averaged Dewey affiliations.

The placement of the visualization was finalized after multiple discussions with the architects. Ultimately it was decided that it would be best positioned above and behind the main information desk situated on the second floor. This location fit with the conceptual and aesthetic intention of the project: to



Fig. 3: A photo of the *Making Visible the Invisible* project, installed at the Seattle Public Library.

create a mood of “information exchange” somewhat like the check-in counters at airports or stock exchange information visualization, which incidentally fit right in with Koolhaas’ concept of the “Mixing Chamber” area as a “kind of trading floor” [14]. Fig. 3 shows a photo of the project installed in the Seattle Public Library.

A challenge in public data visualization projects involves determining the appropriate balance of sharing information while maintaining patrons’ privacy and ensuring the integrity of the data. Understanding the data, its functionality, and how it serves the institution that collects it are important considerations for any data visualization project that uses live data belonging to an organization. The longevity of the project presents another challenge. Initially intended to function as a ten-year project, the lifespan of *Making Visible the Invisible* has since been extended for another five years. Indeed, it may be the longest, continuously existing dynamic data visualization project in existence. Key issues have included planning the data storage and the maintenance of both the software and the hardware; a built-in challenge for a long-running project of this kind is the possibility of its technological obsolescence [15].

We Are Stardust (2008)

We Are Stardust is a two-screen projection installation commissioned by the Spitzer Space Center and the Art Center College of Design in Pasadena. The installation replays a sequence of 36,064 observations made by the NASA-launched Spitzer Satellite telescope over the course of five years (from the start of its mission in 2003 to the premier of the artwork in 2008). The Spitzer Space Telescope is an infrared sensing instrument that is orbiting the sun, trailing in the earth’s orbit. The intent of the work was to create a project based on data produced by the satellite. The initial plan included chronologically mapping the sequence of observations in order to visually reveal patterns in the data. It later became compelling to also integrate metadata about each observation: the name of the star, its distance and location, the date it was first observed, and information about the scientist or laboratory that observed it. The production team consisted of a small group of artists, computer graphics experts, and engineers. The installation was featured at the Art Center College of Design as part of a group exhibition titled *Observe* in 2008. It was also later reprised for the *CODE LIVE* exhibition at the 2010 Vancouver Olympics [16].

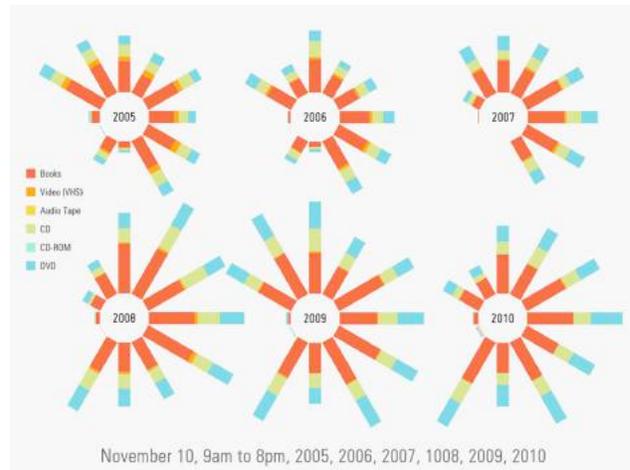


Fig. 4: Seattle Public Library checkout patterns of media based on hourly activity over a period of six years.

The *We Are Stardust* installation consists of two projections on opposite walls of a large exhibition space. Each of the projections chronologically sequences the history of the satellite’s observations. At one end of the gallery, full views of the universe’s constellations are overlaid with the traces of the Spitzer telescope’s observations. It would require five hours to cycle through the full sequence of these observations, thus the full sequence is subdivided into five-minute segments. Each segment begins with an animation of the “birth of the universe” where celestial objects are first placed at random, but then reposition themselves into an ordered cluster that accurately represents the constellations. At the end of each segment all of the 36,064 observations are briefly featured.

At the opposite end of the gallery space, another projection of the same sequences provides a radically different perspective. Using a military grade, heat-sensing, mobile infrared camera, the inside of the gallery is visually scanned according to the adjusted angle-of-view of the original Spitzer telescope coordinates. This visualization provides a colorful and engaging heat-based view of gallery spectators. The spectators can engage directly with the work as they position themselves under the gaze of this searching, moving optical instrument that maps their own thermal presence and actions onto the large screen. Fig. 5 shows the *We Are Stardust* installation.

The staging of this project was dictated by the necessity to provide two points of views: the macro, focused on mapping the observation data onto the universe; and the micro, engaging the local audience in the enclosed space of the gallery. This polarized differentiation dictated the positioning of the two projections at opposite ends with the heat-sensing camera mounted on the ceiling in the middle of the gallery space.

Cell Tango (2006-2009)

Cell Tango is a dynamic artwork presented as an interactive installation in fine arts museums and galleries. The public visiting the exhibition is invited to participate as a contributor to the project through the submission of tagged cell phone images sent from their phones to a database. These images, and the accompanying tags that categorize and describe them, are projected large-scale, continuously shifting as new contributions are added. The images are organized by a database which uses the images’ folksonomic tags to search and retrieve other associative images from the online Flickr photo-sharing website.

The installation features different visualizations of the images using four different animations. The animation ‘Cell_Bin’



Fig. 5: The *We Are Stardust* installation with visitors viewing the two different projections of the observed telescope data.

selects the most recent images in the database on the screen using an algorithm that scales the images. It then places large images first and gradually fills in blank spaces with smaller images until all of the empty screen spaces are filled. ‘Cell_Clusters’ consists of displaying thematic clusters of images based on contributors’ tags. The tags are placed around each of the incoming contributed images, which are differentiated with a yellow frame. The ‘Cell_Burst’ animation places images on the screen that then open like bursting fireworks. Tags are positioned once the image appears, followed by the retrieved Flickr images that are associated with each tag, creating a series of network structures indicating the software’s trace through the image database. ‘Cell_Finale’ concludes the visualization sequence by simulating a random throw of the full collection of submitted images on the screen.

Cell Tango was presented at multiple venues, each time additionally fine-tuned as different aesthetic ideas were more thoroughly developed. It premiered at the International Society of Electronic Arts conference in San Jose in 2006. It was then included in a broad range of venues, both nationally and internationally, over the next four years. In order to augment the movement of the images and to highlight the main activities of each visualization, sonification was added for an installation that premiered at the Lawrence Hall of Science at UC Berkeley in 2010, and algorithmic audio was incorporated again during the final presentation at the Poznan Biennale, Poland in 2010 [17].

The project’s unique strategy is to integrate a curated set of data (the cellphone images submitted by the public) with an open-ended selection from the public Flickr photography database (where the selection from Flickr is based on tags culled from contributors’ submissions). This conceptual premise explores the potential of unexpected juxtapositions where common semantic labels function to generate novel visual relationships. The site-specific conditions for this work have been very flexible, as the installation only requires a large screen surface on which to project the animations. Fig. 6 shows a photo of the installation at the Poznan Biennale.

Data Flow (2009)

Data Flow is an installation commissioned by Gensler Design for the Corporate Executive Board (CEB) Corporation in Arlington, Virginia. The project consists of three visualizations that map time-based communication exchanges between “members” (that is, users of the CEB website) and the CEB document repository. The artwork location was predetermined

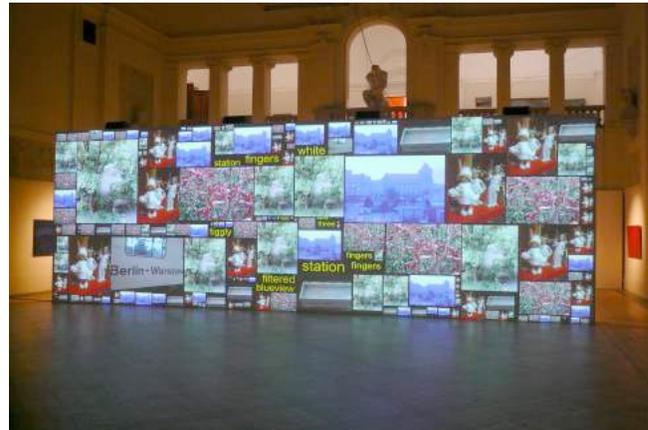


Fig. 6: A photo of the *Cell Tango* project, installed in Poznan, Poland, showing the ‘Cell Bin’ visualization.

in advance, and was situated on the corporation’s “Feature Wall” in the staircase areas extending from the 22nd to 24th floor of the CEB corporate headquarters. Following a study of the data and the architectural space, three locations were selected to position the three sets of screen; most locations on those floors allow for viewing at least two screens simultaneously. The project was installed for just over a year and took approximately six months to develop by a small team of artists and software engineers.

The flow of information from the CEB database to the installation consists of three main steps. Every ten minutes the *Data Flow* project server retrieves appropriately formatted and anonymized user data from the main CEB webservers, which is stored for 24 hours on a local database. The project server also retrieves the longitude and latitude of each user’s (approximate) location and discards any data that does not correlate with the requirements of the visualizations. This data is then forwarded to three visualization computers that each processes the received data according to their individual animation requirements.

Each of the three animations analyzes incoming members’ interaction from different perspectives. The top-level screen, titled ‘BIGRAMS,’ retrieves the most recent 20,000 search phrases and analyzes them to find the twelve most frequently used keywords. Differently colored links are used to indicate the industries and the “Programs” (subsections of the CEB website) that the members’ keywords are associated with. (Each member is classified according to their practice, which may be in the areas of finance, government, health, human resources, information technology, innovation strategy, legal, marketing, or operations and sales.) The middle-level screen, titled ‘GEOMAP,’ makes use of a novel dynamic rectangular cartogram algorithm to animate and deform a grid-based map of the world according to the location where members are checking-in from [18]. The animation receives the 500 most recent events and then sequentially expands cells that represent the active members’ locations. Each grid increases in size based on the length and number of actions. Every three minutes the distorted map collapses back to its original form. The animation indicates the sweep of activity as it moves westward over the course of each day. The bottom-level screen, titled ‘PROGRAM-TO-CONTENT,’ proportionately maps how various content choices and downloads are used by members from different membership categories. This visualization analyzes recent user activity and displays dynamic statistics relevant to these most active categories. This animation



Fig. 7: A detail of the three-story *Data Flow* project, installed at the Corporate Executive Board headquarters.

uses the language of spreadsheets and infographics to provide a creative yet functional overview of trending user activity.

Staging data visualization art within a corporate environment presents its own opportunities and challenges, including lengthy negotiations with administrators and in-house designers, and a heightened awareness of the importance of anonymizing and cleaning the data being used so as to respect customer privacy. Additionally, there was an ongoing concern regarding the possibility of inappropriate words appearing in the ‘BIGRAMS’ visualization. That is, a data visualization in this type of environment must, for better or worse, navigate a range of competing and potentially contradictory interests while retaining the intent of the project. Fig. 7 shows two of the data visualizations from the vantage point of the 24th floor of the CEB headquarters.

Conclusion

This paper has addressed the conditions by which large-scale site-specific data visualizations require extensive formulation in regards to their staging and presentation. Five artistic commissions engaged with visualizing data have been presented to address how each venue’s contextual circumstances, including location, positioning, lighting, scaling, and function, impact user perception and cognition. While some recent work has explored aesthetic issues, such as the effects of style on visualization and the ways in which artistic elements can augment data visualization techniques, the examination of substantial public data visualization installations could still provide insight into how visualization can most effectively be staged in site-specific locations [19, 20]. For most of the projects described above, an additional analysis step took place after the conclusion of the life of the installation. In these analyses, the collected data provides an insight into which stagings were most effective and that encouraged richer user interactions. In the case of *Pockets Full of Memories*, statistical studies were undertaken to explore how certain objects were described at the different locations where the installation was exhibited. Through these studies, one can see distinct trends in the categorization of objects at different locations, highlighting the diversity of visitors and the variations of each exhibition’s context and duration. For other projects, such as *Making Visible the Invisible*, the collected data is being used to create further data visualization projects. Figs. 2 and 4 each show an example of these post-installation data visualizations.

Data visualization projects created as site-specific installations additionally may integrate the sampling of data from the site itself within its process. This iterative approach is similar to design methodologies in architecture where the condition of a building’s location and function impacts its design [21]. This is further enhanced through technological sampling of data (such as demographics, noise, light levels, or traffic flow). The success of site-specific visualizations therefore requires extensive pre-production planning, in-depth negotiation with administrative and technical site managers, consideration for what the long-term impact of the work may be, as well as a clear understanding as to what degree efforts will be invested into the storage, archiving, and analysis of data collected from the installation.

References and Notes

1. C. Ware. *Information visualization: Perception for design*. Elsevier, 2013.
2. J. Bertin. *Semiology of graphics: Diagrams, networks, maps*. Esri Press, 2010.
3. R. Kosara. Visualization criticism—the missing link between information visualization and art. In *Proceedings of the 11th International Conference Information Visualization*, pages 631–636. IEEE, 2007.
4. A. R. Gaviria. When is information visualization art? Determining the critical criteria. *Leonardo*, 41(5):479–482, 2008.
5. E. A. Shanken. Art in the information age: Technology and conceptual art. *Leonardo*, 35(4):433–438, 2002.
6. R. Venturi, D. S. Brown, and S. Izenour. *Learning from Las Vegas*. MIT Press, Cambridge, MA, 1972.
7. G. Legrady. Intersecting the virtual and the real: Space in interactive media installations. *Wide Angle*, 21(1):105–113, 1999.
8. J. Roloff. Felsic/mafic/carbonate facies. <http://www.johnrolloff.com>, accessed March 2015.
9. G. Legrady and A. G. Forbes. Staging data visualization installations in site-specific situations. In *Proceedings of the IEEE VIS Arts Program (VISAP)*, pages 85–90. Paris, France, November 2014.
10. G. Legrady. Pockets Full of Memories: An interactive museum installation. *Visual Communication*, 1(2):163–169, 2002.
11. J. G. Snider and C. E. Osgood. *Semantic differential technique: A sourcebook*. Aldine Pub. Co., 1969.
12. G. Legrady. Pockets Full of Memories data archive. <http://tango.mat.ucsb.edu/pfom/databrowser.php>, accessed March 2015.
13. G. Legrady. Culture, data and algorithmic organization. *Leonardo*, 45(3):286–286, 2012.
14. R. Hackett. New piece for central library pushes art to the technical edge. *Seattle Post-Intelligencer*, September 12, 2005. <http://www.seattlepi.com/ae/article/New-piece-for-Central-Library-pushes-art-to-the-1182783.php>.
15. C. Paul. *New media in the white cube and beyond: Curatorial models for digital art*. University of California Press, 2008.
16. G. Legrady, J. Villegas, and A. Burbano. The We Are Stardust installation. In *Proceedings of the 17th ACM international conference on Multimedia*, pages 1087–1090. ACM, 2009.
17. A. G. Forbes and G. Legrady. Cell Tango: An evolving interactive archive of cellphone photography. In *Proceedings of the International Symposium on Electronic Art (ISEA)*, Istanbul, Turkey, September 2011.
18. A. G. Forbes, T. Hollerer, and G. Legrady. Behaviorism: A framework for dynamic data visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1164–1171, 2010.
19. A. Vande Moere, M. Tomitsch, C. Wimmer, B. Christoph, and T. Grechenig. Evaluating the effect of style in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2739–2748, 2012.
20. F. B. Viegas and M. Wattenberg. Artistic data visualization: Beyond visual analytics. In *Online Communities and Social Computing*, pages 182–191. Springer, 2007.
21. M. Kwon. *One place after another: Site-specific art and locational identity*. MIT press, 2004.