Data Mapping: Interactive Big Data Visualization in Landscape Architecture

Pia FRICKER[1] and Georg MUNKEL[2]

[1]ETH, Zurich, Switzerland · fricker@arch.ethz.ch
[2]ETH, Zurich, Switzerland · munkel@arch.ethz.ch

Abstract

"Everyone spoke of an information overload, but what there was in fact was a non-information overload." (Richard S. Wurman, *What-If, Could-Be: An Historic Fable of the Future*, 1976)

Since the invention of statistical analyses of a scientific nature over 250 years ago (William PLAYFAIR, 1759–1800), we find ourselves in a situation today of almost unlimited data access. However, which data are relevant for design purposes, e.g. how can this data not only be represented understandably, but also be inserted directly into the planning process as design relevant criteria? This paper discusses the results of our current research in the area of abstract representation of data as a design support tool. The findings were tested on their methodological implementation and application in practice and subsequently presented in the one-year study program of the Master of Advanced Studies in Landscape Architecture (MAS LA).

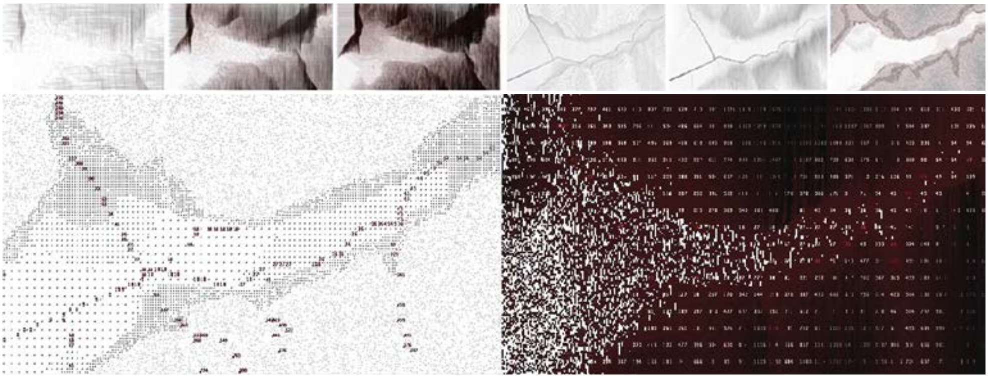

Fig. 1: Student project of Ana Krstulovic and Effrosyni Laskari, Title: "Natural Elements: Extraction, Interpretation". The program depicted three 'actors': the mountain, the valley and the river. All have their individual, unique form language. Thus, in the mountains, the rotation of lines was regulated by the respective heights; the valley was drawn using a rough point raster and the edges were indicated by a denser pattern; and the rivers were visualized using a very close point raster. In their work, the students placed the focus on the interplay of various graphical elements: text, lines, and surfaces were applied equally for drawing. The software played the visualizations back as animations.

Buhmann, E., Ervin, S. M. & Pietsch, M. (Eds.) (2015): Peer Reviewed Proceedings of Digital Landscape Architecture 2015 at Anhalt University of Applied Sciences. © Herbert Wichmann Verlag, VDE VERLAG GMBH, Berlin/Offenbach. ISBN 978-3-87907-555-3.

1 Introduction

The reality of 'Big Data' and direct accessibility to an incredible range of datasets already had a great influence on architecture in the 1990s. In their book *Metacity/Datatown* (ed. by MAAS 1999), MVRDV postulated a design strategy that is mainly generated through the interpretation of datasets. If one observes the current development critically, then we are already imperilled by the danger that we no longer want to exemplify or even understand the vast quantity of data available, especially if its meaning has become difficult to codify. Historically, statistical charts have aimed to communicate complex ideas in a clear, exact and efficient manner, however, information visualization today mainly serves to give meaning to a large amount of data in order to obtain a finding from it (KLANTEN 2008, 2010).

With the simple technical access of sensor technology, for example in combination with Arduino (an open-source electronics platform based on easy-to-use hardware and software), one can generate a plethora of environmental data regarding place-specific conditions in real time through the use of UAVs (Unmanned Aerial Vehicles). Data that can create datascapes, especially in landscape architecture as a supplement to conventional GIS information, can be used at various levels as design tools. In the MAS LA Program of Professor Christophe Girot (Institute of Landscape Architecture, ETH Zurich), we are trying to question the kind of data thematically before the integration of datasets in order to determine which data are relevant and how they can be integrated into the design process.

The spatial maps generated by Nadia AMOROSO (2010, 2012) using DataAppeal software, reveal a GIS web-based application that allows impressive new presentation methods of environmental data. This application could present hidden data in an understandable manner, in the form of animated maps, which could be used for participative processes, for example. Experience with students, however, has shown that the datasets are often not entirely understood and therefore false parameters can influence the design. The designs quickly become extremely complex and uncontrollable. Through the simple integration of visual programming languages, such as Grasshopper, into the design process, a strong basic idea can often be transformed into a completely new design through simple changes of sliders, which are only regulated for changes, however, not for the sensitive integration of essential location-specific parameters. This is why we started our research with the responsibilities of the designer. The representation of datasets can be easily, and often unaware, incorrectly interpreted, which means the consciousness must be trained. Students should therefore be trained to solve these problems. The challenge is to provide and understand a relevant set of data, moreover, it requires the ability to code the applications and fully control them. The graphic form does not always indicate the information. Likewise, the selection of the right graphic form is decisive for further application and understanding of the data compiled.

2 From Data Capturing to Data Programming

James Melsom, a specialist in the area of sensor technology, has conducted projects in this area for the LVML (Landscape Visualization & Modeling Lab) that built the foundation for the data survey. The modular structure of the study program allows us to lay down a methodical theme focus that can illuminate various technical aspects within the general problem. A central element is presented by overcoming the inhibition threshold to programming.

Within the curriculum of the MAS LA program, for the past three years we have sounded out the potential of programming relative to the visualization of real data (sensor data) as a design tool. Our previous experience was based on the use of processing (www.processing.org), an open source programming language with direct visual output. The use of this language made a relatively quick learning success possible as well as individual solutions, which could be integrated directly in the design chain. In addition to the pure visualization of data, the subsequent analyses of data represented a great potential for the direct interaction with the data. In particular, a mature understanding of the site and its hidden information emerged through this interaction, which could be immediately applied to the design process (Fig. 2).

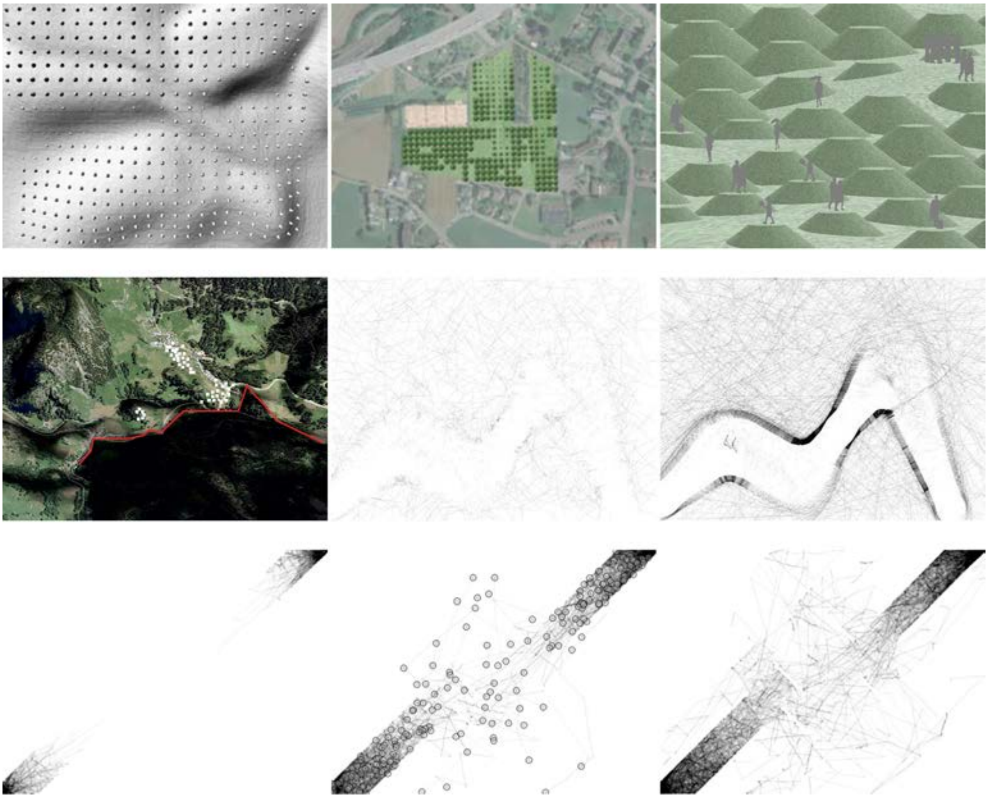

Fig. 2: The overview represents three different results of the modules from the last three years. Christine Baumgartner's project, in the top row, shows the development of a design tool. Through the script, students can control the amount of excravation material and its organization in the form of cones during the design process. In the second row, the tool from Pascal Werner simulates the lighting in a narrow valley dependent on urban design development. Through the results of the simulation, the various planning phases can be verified. The project in the last row, carried out by Vaso Stavrotheodorou, simulates the flow of visitors. The agent-based simulation serves as an analysis tool in order to recognize the spatial quality in the 2D drawing.

During the past two years, various experiments were conducted in the MAS LA Module Programming Landscape on the automated production of maps. The location was always the design site, however, the maps, which were programmed, depicted an individual level of the design problem that was created through the integration of the data. The various results were produced through the weighting of the data and opened new ways of interpretation (HUANG 2013). Finally, various interpretation and evaluation methods served to generate a foundation for the creation of operative maps that directly influenced the design. In this case, we wanted to close the gap between the trend towards "parameter-based design" and the subjectivity of the landscape architect and create an artistically significant unit.

3 MAS LA Module: Programming Landscape − How to Program a Map from Code?

We structured the course for students who didn't have any programming experience and thus taught the basics of a standardized, object-oriented language. The focus lies less on the rigid communication of a language's syntax and more on typical language elements of a higher programming language, such as loops, conditions, functions, and inheritance.

In one of the final projects of a module, students write a program that draws a map of the familiar design areas of previous modules. To do this, the programming teams, each made up of two students, follow the goal to crystallize decisive aspects of the site for their design. These aspects are handled using the smallest possible variables and, as a result, are separated into the most precise qualifable parameters possible. For example, these might include the height of the site, its distance to rivers, the visibility of other defined places or the slope of the land. These values can be found for any place on the map and are translated through drawing style and method. In doing this, drawings are generated that mirror the interplay of the investigated parameters for the entire planning area.

The script written by the students (Fig. 3) divides the design location into cells and starting from the local geographical lowest points, seeks the nearest cell that shows the smallest increase in elevation and the shortest distance to a side canal. From this cell, the procedure is repeated, until the canal is reached. The path suggested by the computer program surprised the students; they had expected that the horizontal proximity to the next canal would be decisive.

The main goal of the course is to teach creative handling of complex relationships relative to a design. First, an elementary understanding of data is required in order to subsequently investigate new strategies to visualize the data, which is needed in order to be able to derive design-relevant decisions in the next step. When one considers the commonly used software packages for visualizing data, for example ArcGIS, or direct interventions in the design through Grasshopper, the process of reflection is often forgotten.

A nearly perfect software package, such as ArcGIS, offers a vast selection of highly developed methods to visualize data. However, the working methods of landscape architects means that they must always adapt these tools, because the possibilities are indeed diverse, but nevertheless preassigned.

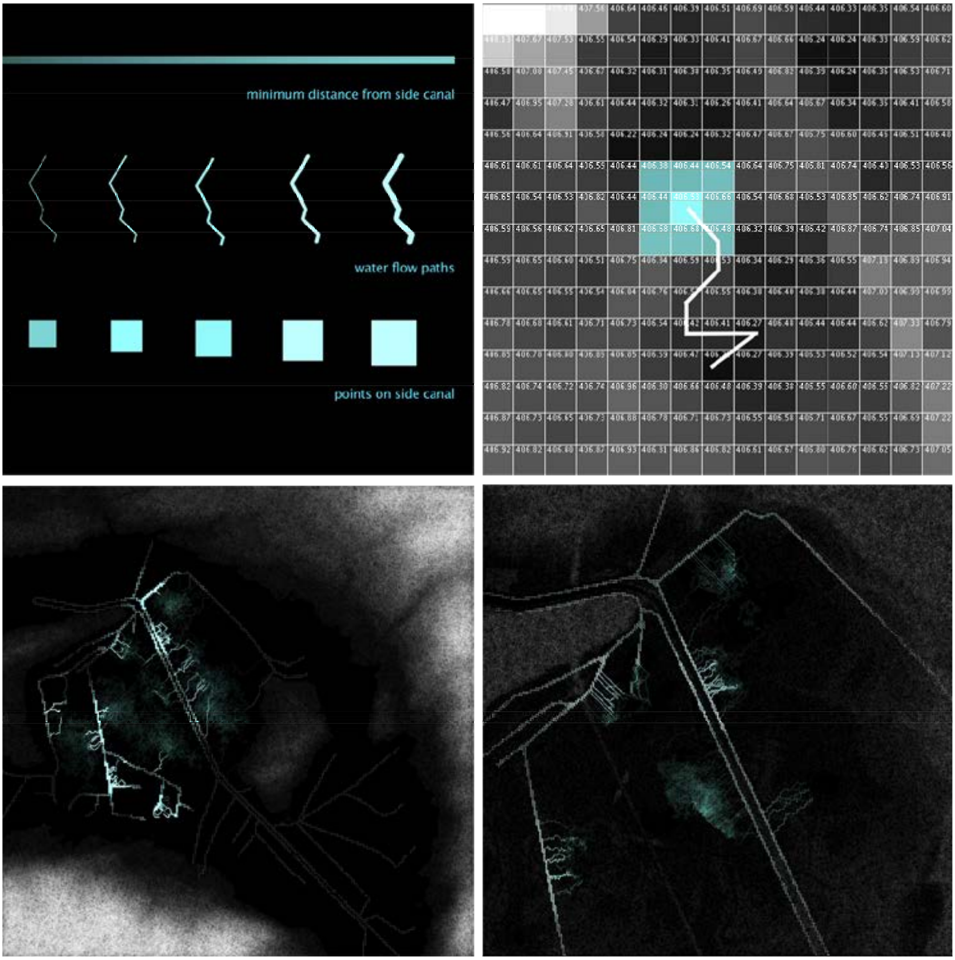

Fig. 3: Student project of Sofia Prifti and Jacqueline Frizi: "A Way Out": The program draws a map of possible solutions for drainage paths that direct the water from all the lowest points of the site to the nearest sewer-side canal. Terrain height data were used in order to specify the lowest points of the area as well as the height difference between them and the side canals. The parameters, which determine the optimum paths, were 1) the smallest height difference and 2) the minimum distance possible.

The more we trust the tools offered, the stronger is the danger that our view of things adjusts to that. "I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail." (MASLOW 1966)

3.1 Data Visualization in the Design Phase

Visual representation, whether as a model or a drawing, is often very similar in its spatial conditions to a design solution and, as a result, runs the danger of being mistaken for one (TUFTE 2001).

We convey an open, playful approach to the visualization algorithm to our students. They should recognize that even small changes can already produce completely different visual results. We intentionally transpose measurements exponentially or anti-proportionally into the sizes of drawn objects. This distortion of the figure, which can be pushed as far as a false interpretation (TUFTE 2001), demonstrates the effect of a row of data on the entire graphic, as well as the influence of the quantity on the algorithm. Another method is to reflect the physical size in completely different parameters of the drawn object, e.g. size, color, transparency, form, etc.

In addition, we have experienced that students hold too rigidly to a single graphic rendering and do not notice that geometric characteristics of the graphic are flowing into their design. For example, zones often emerge in places where a clear color contrast shows an apparent border in a graphic.

The students who prepared the project "I Explorer" (Fig. 4) expected, for example, that the autobahn would be a critical visual disturbance in their design. This was clearly emphasized in all the drawings. The representation on the left is the result from the software and reveals that a view of the autobahn is only possible in a few places.

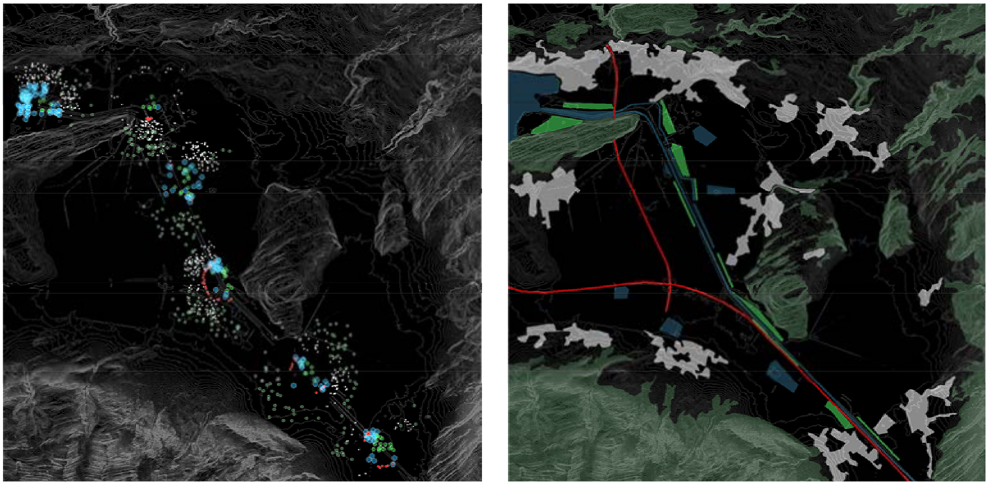

Fig. 4: Student project of Angelos Komninos and Argyro Theodoropoulous, "I Explorer": The program shows a map of views along a path. The question is the level of intensity that impacts the view of the different areas along a path: water, woods, cities, and highways. The algorithm developed for this can determine the visibility of all areas for any area along the path and finally draw a colored symbol, depending on the frequency of their occurrence in the displayed view. The different areas are drawn around a place, each within their own radius. This project had an instrument programmed in for capturing the real spatial relationship in an intuitive manner. This presentation technique makes it possible for students to test and adjust their design parameters.

3.2 Evaluation of Measurements Based on their Graphic Representation

When the students see that they can create an entire spectrum of output with the help of algorithms, the knowledge sinks in that there is more than just one single solution to the given design situation.

At the same time, it increases the probability of discovering correlations. When a connection exists between different values, this must also be perceivable in different presentations. Even though in some instances, it is clear, in others, it is more subtle, and, most important, it does not prove the causality of the correlation.

In addition, not only the logic can be changed: just reducing the underlying data set massively changes the information illustrated in the graphics. If one imagines, for example, a data set in which the values in the center vary between 0 and 10, but on the margins, a maximum von XX up to 100 appears. This makes it clear that completely different impressions emerge depending on whether the margins are included or trimmed off.

We promote discussions on exactly the issue of the difference between the two because we believe that it is important to understand how much a person can be influenced and lose a clear view. In the just mentioned example, the question of whether it is right to include the margins or to trim them off, offers no right or wrong answer. It lies in the discretion and responsibility of the author.

4 Integration of Large Software Libraries in Teaching Design

In order to sharpen the meta-level, the abstract interactive visualization of data, we are currently looking into the use of software used by data journalists. In using these tools, we expect the capacity to be able to operate much more analytically and efficiently with large amounts of data, for example, through cluster analyses, which will allow us to easily discover the locations of similarities and coherences within the special datasets.

At the moment, we are keeping an eye on these software packages: Javascript Libraries p5.js (http://p5js.org/) and d3js (http://d3js.org) as well as the database mongodb (http://www.mongodb.org), and the search and analysis engine, elasticsearch, (http://www.elasticsearch.org).

If we want to make these powerful tools available to our students, you can't get around the fact that you need to abstract them into a manageable framework: A software framework packages all libraries into a single programming interface. All functionalities can use the same principles, such as naming or return types. In addition, any feature you plan not to use can stay hidden behind the interface. The result is a lean yet extensible and readily comprehensible starting point for each student's software project.

Our current framework allows the students to manage their sites in multiple grids. These grids can hold values of two different types: numeric, e.g. height, distance, noise level, or Boolean, e.g. region markers or visibility. The grids can be easily generated from pixel images and in addition through a combination of grids. For example, if you have a grid containing the terrain height and a grid where all streets are marked, the framework can generate a grid containing all areas, in which the streets are visible.

To date, we have had very positive experiences with this strategy, because we still could extend the framework's capabilities during the preparation of the module's final project. The necessity of these extensions was first noticed in the work with the students.

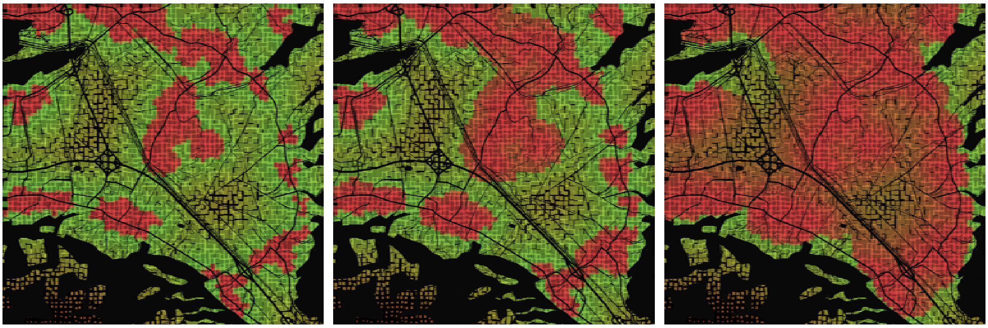

Fig. 5: Student Project of Maki Hasegawa and Gebhard Merk, "Programming Landscape": The program is based on the input of five layers: urban zones, buildings, street network, agricultural zones and water zones. The algorithm compares the distances of the layers depending on the importance of the given design. By changing the importance of the input parameters, the design proposal can be tested and adapted.

5 Conclusions and Outlook

Big Data visualization in landscape architecture requires, in addition to an understanding of the data, a certain basic understanding of programming in order to be in a position to independently control the use of data. Our teaching methods show great promise in disentangling the didactic-methodological complexity that arises. The students are able to collect data responsibly and independently and integrate it as relevant tools in the design process. With our applied research, we want to test a method and approach that would demonstrate the integration of data as design tools. To this end, we encourage students to cross-disciplinary and theoretical boundaries and critically analyze the dataset information. At the moment, rapid technical development allows almost the unrestricted use of arbitrary and complex datasets, which they either collect themselves using UAVs in combination with sensor technology, or from open source repositories. It also encourages the continued methodological development of design tools by universities.

Through the application of software packages from the field of data journalism (DDJ), we hope to achieve new levels in the visualization of data, which would then go beyond the portfolio of graphic data information on maps and be able to generate direct changes in the design (GRAY 2012). In contrast to data journalism, our goal is not about interpreting data subjectively in order to generate new journalistic presentation forms, rather ours is about testing the technical aspects used in collecting, preparing and analyzing the data for its application in the design process.

A further aspect is the integration of additional users and their needs, in which we see the possibility to be able to evaluate data directly in the program as an exciting use of the measurements and to be able to point out and integrate a further level of meaning and to be able to integrate the overlap and interaction of various dimensions. This understanding of the parameters triggers a strong fascination and excitement and goes beyond the standardized spatial description of information structures.

Acknowledgements

We would like to thank all the students who participated in the MAS LA program 2012–2014 for their hard work: Claudia Aracci, Silke Hartmann, Enise Karacizmeli, Karolina Katsambi, Tania Kaushal, Ana Krstulovic, Karol Kruk, Effrosyni Laskari, Wolfgang Novak, Georgios Orfanopoulos, Anastasios Roidis, Loulitta Stavridi, Marina Tsintzeli, Anna Walewska, Mohamed Abdel Wahab, Jacqueline Frizi, Maki Hasegawa, Angelos Komninos, Gebhard Merk, Sofia Prifti, Alexandre Roulin, Gaganjit Singh, Georgios Sarmaniotis, and Argyro Theodoropoulous, We also want to thank Prof. Christophe Girot, Institute of Landscape Architecture at ETH Zurich. Special thanks to all the assistants of the Chair who have been involved in the courses and to James Melsom and Karsten Droste for their great support.

References

AMOROSO, N. (2010), The Exposed Citiy, Mapping the Urban Invisibles. Routledge, London & New York.

AMOROSO, N. (2012), Digital Landscape Architecture. Thames & Hudson, U.K.

AMOROSO, N. & SECHTER, H. (2012), 3D Geo-Design Mapping Using DataAppeal. In: Peer Reviewed Proceedings Digital Landscape Architecture 2012. Wichmann, 346-355.

CONTIN, A., PAOLINI, P. & SALERNO, R. (2013), Innovative Technologies in Urban Mapping: Built Space and Mental Space. Springer.

GRAY, J. et al. (2013), The Data Journalism Handbook. O'Reilly, U.S.A.

HUANG, M. & WEIDONG, L. (Eds.) (2013), Innovative Approaches of Data Visualization and Visual Analytics (Premier Reference Source). IGI Publisher.

KLANTEN, R. (2008), Data flow: Visualizing Information in Graphic Design. Gestalten, Berlin.

KLANTEN, R., EHMANN, S., BOURQUIN, S. & TISSOT, T. (2010), Data flow 2: Visualizing Information in Graphic Design. Gestalten, Berlin.

MASLOW, A. H. (1966), The Psychology of Science. Maurice Bassett Publishing, U.S.A.

MVRDV, (1999), MVRDV − Metacity Datatown, 010 Uitgeverij, Rotterdam.

TIERNEY, T. (2007), abstract space − beneath the media surface, Taylor and Francis, New York.

TUFTE, E. (1990), Envisioning Information, Cheshire, Connecticut, U.S.A.

TUFTE, E. (1997), Visual Explanations, Graphics Press, Cheshire, Connecticut, U.S.A.

TUFTE, E. (2001), The Visual Display of Quantitative Information, Graphics Press LLC, U.S.A.

http://www.arduino.cc/
http://p5js.org/
http://d3js.org/

http://vimeo.com/29862153 − a quick d3js show reel
http://www.elasticsearch.org
http://www.lvml.net
http://www.mongodb.org
https://www.processing.org/
http://www.fastcodesign.com/