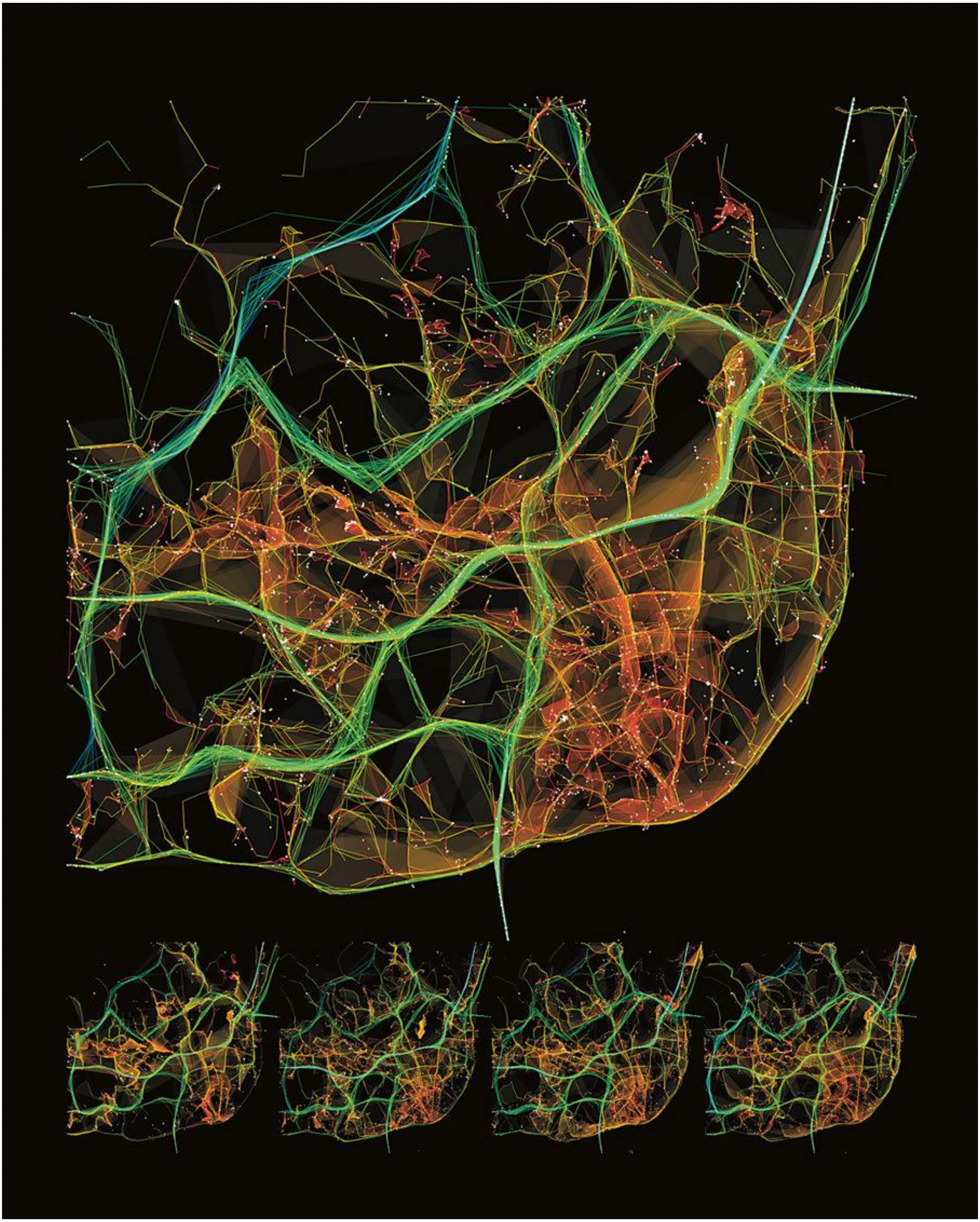
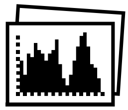


Fig. 1.
Top: **Photograph of the traffic of Lisbon at 9:00 a.m., rush hour**
Bottom: **Evolution of the city during the day: different areas are illuminated, and the color of the arteries changes. Visualization: Pedro Cruz, Penousal Machado**

Pedro Cruz, Penousal Machado

City Portraits and Caricatures



Introduction

Cities are a very rich source of data in order to understand how people organize themselves on different scales. The discourse on the visualization of cities offered different approaches, from more abstract to illustrative and figurative ones. In this chapter, we approach the visualization of cities in a less abstract, puristic way. Distancing ourselves from scientific visualization, we embrace what we call a *figurative approach to visualization*. With this approach we take a step back from the idea that visualization has to represent data only in the most direct and abstract way. The figurative approach extends the concept of portraits and caricatures into the realm of data, using visual metaphors and introducing visual distortions for emphasizing certain aspects of the data. This permissiveness enables us to build visualization artifacts that use playful analogies to better communicate the complex nature of cities to a nonacademic, general audience.

This article will describe the process that took two visualization cases from a direct to a figurative approach, not entering into detailed technicalities, but tying it with a cogent discourse that presents and applies portraits and caricatures in the context of data visualization.

A Figurative Approach to Visualization

Information visualization is an interdisciplinary area that intersects graphic design, human-computer interaction, computer graphics, and data mining. Its objective is to synthesize large amounts of data for broad audiences, creating and clarifying messages from data. The figurative approach described herein uses a taxonomy of photographs, portraits, caricatures, and disfigurements to describe distinct approaches to visualization. The main concept behind such taxonomy is degrees of authorship. We refer to authorship as the varying intent with which a particular perspective over the data is made visible by its creator. Authorship flourishes in contemporary information visualization, mostly under the term "information art" and is also embedded into scientific visualizations. It is important to notice that visualization artifacts that go beyond the sole purpose of analyzing patterns in data are often seen as distortions of the truth, especially in the context of academic visualization. While this is a legitimate argument under a puristic perspective, the caricature of data is conceptually inevitable, as Fernanda Viégas notes: "traditional analytic visualization tools have sought to minimize distortions, since these may interfere with dispassionate analysis. Is it possible that this focus on minimizing 'point of view' is misguided? For one thing, it is generally impossible to create a visualization that is truly neutral, just as it is impossible to create a flat map of the Earth's surface without distorting distances." (Viégas and Wattenberg 2007, 190)

Considering that authorship is always present to some degree in data visualization, our taxonomy identifies four major ways to undertake data visualization characterized by an increasing degree of authorship: photographs, portraits, caricatures, and disfigurements. But authorship is not the only characteristic to describe these concepts and tie them together in a continuous fashion. The figurative approach focuses on the idea that visualization does not have to invest mainly in abstract aesthetics, but can carry strong visual metaphors and exaggerate certain data features in its communication language. Since information visualization is part of the broader data visualization field, which also encompasses scientific visualization and, arguably, artistic visualization, this seemingly unscientific taxonomy describes this spectrum within visualization from the more scientific to the more artistic.

Data Photographs

In our taxonomy, photographs are the most direct reproduction of a data set – models that are the closest to a one-to-one mapping of data. This notion refers to Lev Manovich's concept (2011) of visualization without reduction or direct visualization. Instead of mapping data into a symbolic abstraction via Jacques Bertin's visual variables of position, size, value, texture, color, orientation and form (Bertin 1967), direct visualization preserves the nature of the data. What is text in a data set remains text on the representation space; what is an image remains an image; what is speech remains speech, and so forth. Although not deprived of authorship, because a photograph can be taken from different angles, they intend to faithfully reproduce the subject while maintaining every characteristic of it, in this case, data.

Data Portraits

In data portraits, the author has a more pronounced role than in photographs, by recurring to visual metaphors in the representation. Visual metaphors herein do not account for merely decorative visual elements, described as "chartjunk" by Edward Tufte (1983), but refers to graphical elaborations that have a close semantic meaning toward the data set and the author's message. These semantic metaphors are closely related to Donna Cox's visaphors (Cox 2006) in the sense that they are approximations of data, more dependent on subjective interpretations. A semantic visual metaphor is a figurative evidence of certain characteristics in data in addition to the ones directly mapped. Such evidences are elaborated on a graphical level and result in less abstract, often more familiar, natural and expressive artifacts.

In addition, visual metaphors carry a sense of novelty and uniqueness, that nevertheless can get lost once embedded into our culture. For example, the pie chart, first introduced by William Playfair (1801)[1] was at that time a novelty that has been lost with their pervasive use. Pie charts carry a tender visual metaphor that conveys a sense of unity of the parts in a whole.

[1] William Playfair, *The Statistical Breviary; Shewing the Resources of Every State and Kingdom in Europe* (London, 1801).

Data Caricatures

We regard data caricatures as an extension of data portraits since they also take advantage of semantic visual metaphors. They have, though, its distinctive features that we explain below.

A caricature is often regarded as a figurative representation of a subject that exaggerates its prominent features in order to improve recognizability (Redman 1984). This augmented recognizability is a consequence of what is called the "peak shift effect": an individual trained with a visual representation *r,* shows a response to a similar representation *b* that increases with its differences compared to *r* (Ramachandran and Hirstein 1999). Such a reference model *r* is an essential part of a caricature, which is always present, either physically or as a mental image.

The concept of caricature in a data visualization context comes with certain restrictions. As explained, caricatures depend on a mental image as a reference. Such a reference can generally not be taken for granted in data visualization since even the most direct visualizations introduce a new form for previously unvisualized data. Therefore, the application of caricatures in a visualization context depends on the familiarity of the reference model for its caricatured representation to allow for comparing the differences.

As suggested above, caricatures are tied with the concepts of exaggeration and recognizability. Exaggeration in the context of data means an overemphasis of one data dimension over the others. This can be graphically attained in several ways beyond the amplifying numerical differences in the data, as done in the work of Peter Rautek (Rautek et al. 2006), resulting in distortions of form, position, size, or color of a visual element toward the reference. Recognizability means, in the context of data, the clarification of a caricature's intent, highlighting messages derived from the data. While a data caricature is not necessarily superior in clarifying a message compared to a photograph or a portrait, we nevertheless believe that this is the case for many visualization solutions. To summarize, a data caricature is a visualization model that graphically distorts a reference representation of a data dimension in order to emphasize that same or other depicted dimension.

An obvious way to implement data caricatures in visualization is through the distortion of geographical positions. This principle has been applied since the nineteenth century with cartograms. Cartograms distort geographic maps in order to represent other data dimensions. For example, area cartograms resize countries in a world map proportionally to their population or GDP. This idea was simplified by Daniel Dorling, famous for the Dorling Cartograms (1996) – these cartograms do not maintain the shape of geographical objects, but usually replace them with circles of a size proportional to the mapped dimension. Although they oversimplify geographical shapes, Dorling Cartograms have proven to be very effective. The caricatural property of such methods lies on the emphasis given to a certain data dimension (population, for example) against a geographic map. This map is the reference model that is either fully represented or that can be mentally reconstructed.

Data Disfigurements

Data caricatures introduce distortions in the representation of data. When such distortions are exaggerated beyond certain levels, they can result in data disfigurements. Data disfigurements overemphasize certain dimensions to a point of communicating erroneous messages toward the data. They corrupt the semantic value of visual metaphors, undermining the clarification intent of visualization, and even producing unintelligible artifacts. Nevertheless, the exaggeration of distortions gives more room for authorship and potentially more memorable artifacts.

Figurative Visualization of Cities

The case studies in which we implement direct and figurative approaches to visualization refer to mobility systems of the cities of Lisbon and Singapore. Lisbon's data set[2] contains GPS traces of vehicles in the city during one month, with position and current speed indication. The Singapore data set includes information about people boarding and leaving buses (so-called tap-ins and tap-outs) at bus stations with its respective paid fares in the city of Singapore over one week.[3]

Photographing Lisbon

The spatial and temporal resolutions of Lisbon's data set did not allow a visualization that clearly depicted traffic patterns for each

[2] Data provided by the CityMotion project/MIT Portugal. Visualization project partially supported by the project PTDC/EIA-EIA/108785/2008 COSMO – COllaborative System for Mobility Optimization.

[3] Data provided by the Land Transport Authority in Singapore. Visualization developed for the LIVE Singapore! initiative, supervised by Kristian Kloeckl at MIT SENSEable City Lab and SMART (Singapore-MIT Alliance for Research and Technology).

day separately. Therefore, the information was condensed into a single virtual day, grouping the data by the second and displaying it as an animation. In order to enhance the temporal patterns in traffic even more, each vehicle, represented by a small white dot, leaves a trail that remained visible for 30 minutes in simulated time. The trail is almost transparent and colored accordingly with the speed of the vehicle. Using a restricted color scale of pure hues, the colors red and orange indicated slower velocities, pure green for 50 km/h, and the cyan tones for higher velocities. The trails tend to visually cluster into thicker lines for major arteries that emerge with mixed hues and opacities, representing traffic density (by thickness and opacity) and average speeds (by hue) at that point in time. For example, narrow streets tend to be thinner and red, while highways are thicker and green. The color of the highways that cross Lisbon changes to more yellow hues at rush hours.

Visualizing a data set usually involves looking for problems in the system being depicted: the most apparent salient features in Lisbon's data set are the congested areas. In order to emphasize those, another visual component was added that draws with a very low opacity the area covered by each vehicle in 30 minutes. The covered area is defined by closing the route, connecting its origin and destination points; its opacity is higher as the velocity lowers, and the hue is the same as the corresponding trail. This representation emphasizes the areas with slower traffic, being more opaque orange and red (speeds that get closer to green are almost transparent in this representation). Drawing such areas pinpoints the most problematic areas in Lisbon during the day, which is an easier perceptual task rather than trying to delineate those areas by only extracting colors of lines – it is notorious, for example how downtown keeps being illuminated during daytime, and how peripheral areas activate before the rest of the city. More than directing the audience's attention, it also adds visually to the artifact, making it more detailed and at the same time more compelling (figure 1). All the superimposition of semitransparent lines and shapes leaves little room for the visual extraction of concrete data points, such as the temporal velocity of a certain vehicle, but it is a direct representation, a photograph, of data and provides an overall picture of the traffic evolution during a day in Lisbon.

Due to the high level of detail, the artifact had to be rendered offline and assembled into a single animation afterward. However, it is sometimes of interest to have a real-time visualization, making it possible to pinpoint problematic issues as they unfold. The following project on Singapore addresses such intentions.

Photographing Singapore

The data set concerning Singapore's bus network describes stops at bus stations, but provides no information about the trajectories between stations. Nevertheless, we consider that the representation of such trajectories could provide a much more tangible artifact and an interesting visualization challenge: representing buses moving on a network, and not just buses at bus stations. To attain this, buses were implemented as programmed automatons that simply respond to inputs from a simulated environment (reactive agents), where data constitutes the environment. This also permitted the fulfillment of another objective: creating an interactive visualization that runs in execution time, enabling the rapid visualization of the outcomes of the applied techniques.

Each bus, as an agent, only knows its next stop, its arrival time, and the current simulation time. Data is injected on execution time into the simulation environment, redirecting each tap to the corresponding bus. Each bus has a buffer of the next stops, each in charge of retrieving the next bus stop when the time has come, as well as removing passed bus stops. Knowing only the arrival time, the next stop, the current position, and the current time, the bus travels to the next stop. Many approaches could be used to model the movement of each bus between bus stops: the first approach implemented a nonlinear movement, varying with the time that the bus has to reach the next stop. It was named "lazy buses" because the agents, in spite of knowing that they had to be at a certain stop at a certain time, start their travel slowly, but rush as the time urges (the distance traveled between two stations is inversely proportional to the squared time for the destination).

Beyond the portrayal of buses' travels, it was also interesting to depict which were the more congested. The density of bus stations in the visualization canvas did not allow for the labeling of each of them at the same time. A geographical clustering algorithm

Fig. 2.
Results of the clustering algorithm with a search radius of 44 meters and a minimum of 5 buses per cluster

was used to pick which stations were the most congested, with more than *N* buses stopped at the same time. Our approach to such an algorithm (Ester et al. 1996; Finkel and Bentley 1974) was capable of accommodating a good performance, running in execution time for the amounts of data being used. The first results of the clustering as well as the lazy buses' behavior implementation are depicted in **figure 2**. Small gray crosses represent all the bus stops, and triangles the actual buses, with the triangle indicating the current direction. Only the bus stations that correspond to clusters of buses at that instant are labeled.

Caricaturing Lisbon

In order to build a caricature of Lisbon's data set, we had to first work on its portrait and therefore a semantic visual metaphor. We see cities as systems with complex behaviors, but which have their own identity by shape, normal activity cycles, and abnormal activity bursts that should be diagnosed. On that account, and also on an authorial aesthetic choice, the traffic in Lisbon is portrayed exploring the metaphor of a living organism with circulatory problems by building a system of blood vessels.

A computational structure, a form-preserving "skeleton" subject to simulated elastic, spring-based forces, was generated for each major road in Lisbon. The skeletons are connected when roads intersect and also affect other neighboring areas when they move. This allowed the city to change its shape accordingly with the traffic on the arteries. Springs, as elastic devices, are able to change their shape with some resistance when reacting to a force. If changes on data values act as forces for those springs, then we are able to attain smooth transitions between shapes, as well as to naturally return to the initial shape configuration when the forces are null. Those shape mutations translate the actual perceived distances within a city, contrasting with the common geographical mapping (that is, for our caricature, the reference model). This type of mapping is a classic caricatural visualization since it is a type of distance cartogram – such cartograms are common in the form of isochronic cartograms, mapping all the geographic distances in function of travel time from a well-defined origin point. In our case we wanted to present how the travel times evolve for each artery at the same time and also how changing these distances alters the overall shape of the city.

We are dealing with a rather complex spring-based physics system – as previously mentioned, each spring distends or compresses according to the data; in this case, each spring is also connected to others, forming a skeleton. As a consequence, each spring cannot change its shape freely, being also influenced by connected springs and other data values. With so many frequently opposite forces being applied to the same elastic devices, there is a high risk of rupture – the skeletons collapsing in configurations that make unintelligible images of the city, and by extension, disfigurements. Therefore, we had to ensure temporal stability in order to attenuate abrupt variations, leaving time for the system to adapt to new variations and avoiding ruptures. That way, data was averaged to a single day, and only aggregations of one hour with intervals of

10 minutes were used to influence the system and produce the visualization (i.e., from 3:20 to 4:20, from 3:30 to 4:30, and so on). The system is then excited in the following ways: a greater number of vehicles on a vessel tends to make it thicker, and higher speeds tend to contract its length (and vice versa). The latter behavior was chosen in order to transmit a global impression of the perceived distances within the city. This behavior shrinks the city when the traffic velocities are higher, and distends it in the rush hours when the city faces congestion problems. The vessels were also colored accordingly: lower average speeds imply the darkening of a vessel, expressing slower circulation and stagnant blood **(figure 3)**. The vessels' visualization results in an artifact with crude aesthetics that is innate to our visual metaphor, pulsating for each rush hour and stressing which roads are more congested. The compression and distention of each artery and the entire city caricature shows the perceived distances and deviations from the average speeds in traffic.

Caricaturing Singapore

Another classical technique that exhibits caricaturist features is the fish-eye view, introduced by George Furnas (1986) and later applied to maps (Keahey and Robertson 1996). This technique augments a geographical region of interest, while distorting its periphery in order to maintain other parts of the map on the canvas. The main difference from cartograms is in how the data caricature works: cartograms distort one dimension (geographic positions) to emphasize another (e.g., GDP), while in fish-eye views the same dimension (geographic positions) is distorted to emphasize a restricted set of the same dimension (region of interest). In addition, fish-eye views also carry a visual metaphor that makes a device to build data portraits: the magnifying glass.

Caricaturing Singapore's data set followed this approach. We used a data lens to solve the problem of displaying the details of each bus stop on a street level while keeping a global context. Nevertheless, the magnification provided by a classic fish-eye view is not sufficient for the magnification requirements of this visualization. To attain it, a new type of lens was developed, one that transforms the behavior of a magnifying glass (typically the

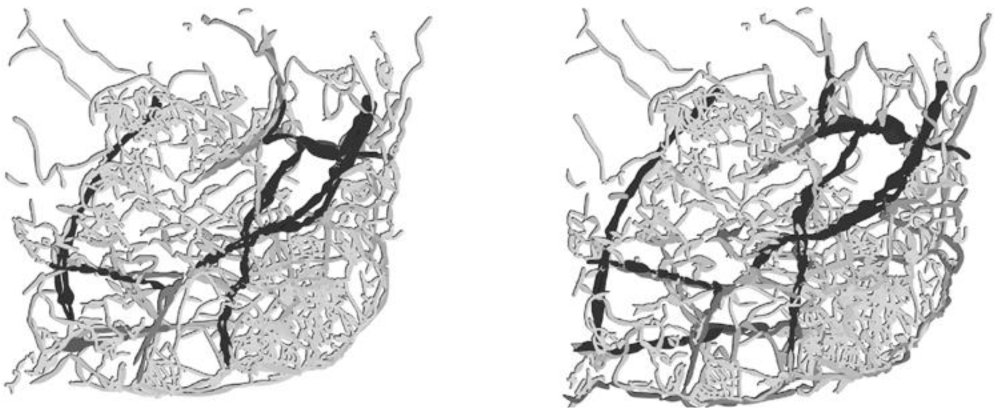


Fig. 3.
Blood vessels in Lisbon at 7:04 a.m. (*left*) and 8:44 a.m. (*right*), just before and after the morning rush hour. It can be observed that at the beginning of the rush hour, the main vessels of Lisbon carry a high number of vehicles, but without traffic congestion problems as the vessels are contracted. At 8:44 a.m. the average speed in the main vessels decreases, originating an expansion of vessels and of the entire city.

projection of the map on an hemisphere on top of the plane) into a projection on a very oblate ellipsoid on top of the plane (figure 4).

The implemented data lens is an interactive visualization tool that allows the user to uncover layers of information that reveal at greater detail the activity of Singapore's bus network. The lens can be dragged over the city's bus network or remain focused on a set of bus stops. The properties of the lens that a user can modify (the position, size and zoom level) help distinguish even very nearby bus stops. In addition, the lens provides access to every feature in the data set by filtering several types of information layers: the line numbers of buses currently stationary at each bus stop, the number of passengers on each bus, and the aggregate paid fares by passengers alighting at any one stop. Users can rapidly switch between these information layers to explore correlations between bus lines, stops, waiting times, passenger load, and paid fare prices.

Remarks

Data caricatures aim at pushing the boundaries in contemporary information visualization by balancing the distance from the dispassionate depictions of scientific visualization with the verge of artistic visualization. This balance is maintained by regarding data caricatures as instruments for design with a clarification intent – to communicate effectively and efficiently with an audience.

Data caricatures are framed in a figurative approach together with photographs, portraits, caricatures, and disfigurements for visualization. A photograph is mainly about direct mappings, or

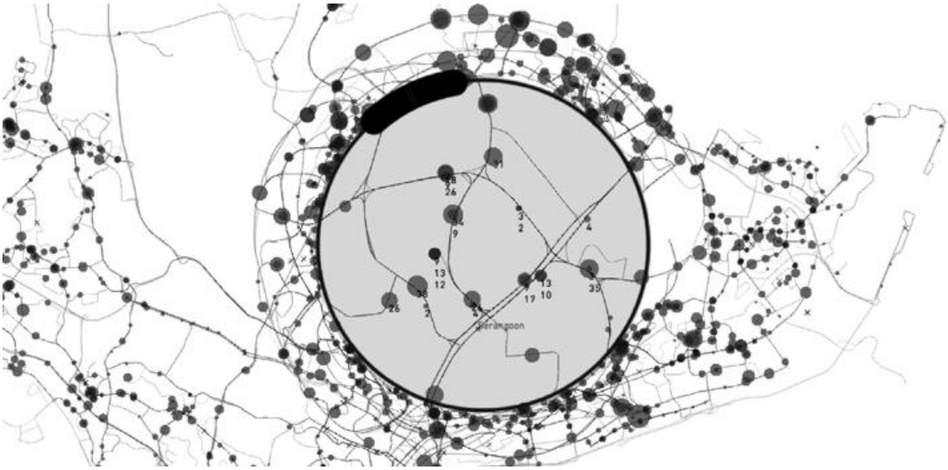


Fig. 4.
This implementation results in a much more diverting and interesting browsing of the space, where the user can direct the lens to a cluster of points, and fluidly unveil and understand each of its constituent parts.

visualization without reduction. Portraits differ from photographs by introducing strong visual metaphors. Visual metaphors have a semantic intent, constituting a figurative evidence of certain characteristics in data in addition to the ones directly mapped. Such evidence is elaborated on a graphical level and result in less abstract, but more expressive artifacts. A data caricature embraces this idea of semantic visual metaphors, extending it with exaggerations in order to emphasize aspects of data. Such exaggerations can culminate in pure distortions by introducing coarse inaccuracies – data disfigurements.

We implemented and applied the notion of caricatures to the visualization of cities. We believe that by using a caricatural approach to the portrayal of cities we can approximate audiences by using strong metaphorical semantics and delivering more concrete messages. For example, inhabitants of a city already have a mental image of certain aspects of it – if we compare such an image to the representation of an ideal model, then we have a distortion that is of interest to convey and where caricatures can adequately be used. Furthermore, such distortions can also be a natural consequence of visualization models that follow an approximation strategy to complicated mathematical problems. Such as in our work, agent- or physics-based models that smoothly adapt to data are much more generic than instantly reacting ones, enabling us to recycle them to other cities and even other types of data. Indeed, such models are ideal for a caricatural approach since they can be directly traduced to nature-based aesthetics, which, accordingly to Greg Judelman (2004), are often used in the context of visualization to decode complexity. Such as in nature, where complexity exists in various scales, the usage of nature-inspired systems seems to be a natural caricatural approach to cities, depicting information with different levels of granularity while maintaining the essence of the message. With this, we believe that caricatures are capable of playing a major role in information visualization, augmenting efficacy, efficiency, and memorability of communication, presenting, and approximating cities to people.

Bertin, Jacques. 1967. *Sémiologie graphique: les diagrammes, les réseaux, les cartes.* Paris: Mouton & Gauthier-Villars. Published in English as *Semiology of Graphics,* translated by William J. Berg. Madison: University of Wisconsin Press, 1983.

Cox, Donna. 2006. "Metaphoric Mappings: The Art of Visualization," in *Aesthetic Computing,* edited by Paul Fishwick. Cambridge, MA: The MIT Press, 89–114.

Dorling, Daniel. 1996. *Area Cartograms: Their Use and Creation.* Concepts and Techniques in Modern Geography 59. Norwich, England: School of Environmental Sciences, University of East Anglia.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, August 2–4, Portland, OR,* edited by Evangelos Simoudis, Jiawei Han, and Usama Fayyad. Menlo Park, CA: AAAI Press, 226–231.

Finkel, Raphael A., and Jon Louis Bentley. 1974. "Quad Trees a Data Structure for Retrieval on Composite Keys," *Acta Informatica* 4, no. 1: 1–9.

Furnas, George W. 1986. "Generalized Fish-eye Views," *ACM SIGCHI Bulletin* 17, no. 4 (April): 16–23.

Judelman, Greg. 2004. "Aesthetics and Inspiration for Visualization Design: Bridging the Gap Between Art and Science." *In Proceedings. Eighth International Conference on Information Visualisation, 2004, July 14–16, London, England.* Los Alamitos, CA: IEEE Computer Society Press, 245–250.

Keahey, T. Alan, and Edward L. Robertson. 1996. "Techniques for Non-linear Magnification Transformations," in *Proceedings IEEE Symposium on Information Visualization '96, October 28–29, San Francisco,* edited by Stuart Card, Stephen G. Eick, and Nahum Gershon. Los Alamitos, CA: IEEE Computer Society Press, 38–45.

Manovich, Lev. 2011. "What Is Visualisation?" *Visual Studies* 26, no. 1.: 36–49.

Playfair, William. 1801. *The Statistical Breviary; Shewing the Resources of Every State and Kingdom in Europe.* London: T. Bensley, for J. Wallis et al.

Ramachandran, Vilayanur S., and William Hirstein. 1999. "The Science of Art: A Neurological Theory of Aesthetic Experience," *Journal of Consciousness Studies* 6, nos. 6/7 (June/July): 15–51.

Rautek, Peter, Ivan Viola, and M. Eduard Groller. 2006. "Caricaturistic Visualization," *IEEE Transactions on Visualization and Computer Graphics,* 12, no. 5 (September): 1085–1092.

Redman, Lenn. 1984. *How to Draw Caricatures.* Chicago: Contemporary Books.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information.* Cheshire, CT: Graphics Press.

Viégas, Fernanda B., and Martin Wattenberg. 2007. "Artistic Data Visualization: Beyond Visual Analytics," in *Online Communities and Social Computing: Second International Conference, OCSC 2007, July 22–27, Beijing China,* edited by Douglas Schuler. Berlin: Springer, 182–191.