6. **Between automation and interpretation: Using data visualization in social media analytics companies**

*Salla-Maaria Laaksonen and Juho Pääkkönen*

**Abstract**

This chapter explores the use of data visualizations in social media analytics companies. Drawing on a dataset of ethnographic field notes and thematic interviews in four Finnish social media analytics companies, we argue that data visualizations are crucially involved in how analytics-based knowledge claims become accepted by companies and their clients. Basing on previous research on visualizations in organizations and as a representational practice, we explore their role in social media analytics. We identify three practices of using visualizations, which we have named have simple-boxing, flatter-boxing, and pretty-boxing. We argue that these practices enable analysts to achieve the simultaneous aims of producing credible and valuable analytics in a context marked by high business promises.

**Keywords:** Visualization in data analytics; Analytics as business; Automated analytics; Interpretation; Visual analytics; Epistemology of data analytics

**Introduction**

The abundance of social media data and the development of computational methods have led to the birth of new business opportunities, including the growing field of data analytics. One nascent area inside this field is social media analytics, or refining and processing data generated by human behaviour on social media, with the aim of transferring them to
business knowledge. David Beer (2017) describes data analytics companies as new powerful *data intermediaries* who build infrastructure to shape the circulation of social data and construct narratives of omnipotent and intelligent calculative technologies to reach objective, correct solutions and decisions in organizations. Beer (2017, p. 466) posits that the ‘data analytics industry is powerful in shaping what is said, made visible or known through data’, and calls for more studies explicating the ways in which this industry cultivates visions of data-led thinking. At present, we lack an adequate account of knowledge production in data analytics, although the field’s societal implications are proliferating (cf. Beer & Burrows, 2013). This issue is particularly pressing for novel contexts of analytics, such as social media analytics, which are currently becoming established (Kennedy, 2016).

An essential part of making sense of digital datasets is the use of visualizations—such as regular line, bar, or pie charts, or more sophisticated algorithmic visualizations, such as clustering diagrams or network graphs (Kennedy & Hill, 2018). Visualization techniques also play an increasingly important role in organizations, which utilize visual representations of data in both their internal and outward communication and negotiations (Quattrone et al., 2012; Halpern, 2014). While the most obvious aim of data visualizations is to communicate information, they can also shape the actions and understandings of their users (cf. Beer & Burrows, 2013; Kennedy, Hill, Aiello, & Allen, 2016). This means that they are devices which bear material agency in their immediate purposes and contexts of use (cf. Leonardi, 2011) and work to construct conceptions of both data and analytics.

In this chapter, we explore processes of knowledge production in data analytics, by investigating the use of data visualization in the business of social media analytics. We approach visualizations as visual representations of data, the use of which is intertwined with epistemic conceptions that guide how analytics are conducted, both as business and as knowledge production. Thus we are interested in the conceptions of what constitutes knowledge in analytics, and the role visualizations play in the process. Drawing on ethnographic field notes and thematic interviews in four Finnish social media analytics companies, we argue that data visualizations are crucially involved in how analytics-based knowledge claims become accepted by companies and their clients. We explore their role in this process, and the associated epistemic conceptions concerning social media data and its analysis.

To begin, and to formulate our research question, we first introduce social media analytics as a research context, and discuss previous literature on the uses of visualization beyond social media analytics. Then, in the subsequent sections, we present our empirical material and findings.
Social media analytics as a research context

Visuals have achieved a dominant position in contemporary organizations (e.g. Sørensen, 2014). Organizational, economic, and political life is represented and enacted by visualizations such as flowcharts and diagrams (Locke & Lowe, 2012), infographics (Amit-Danhi & Shifman, 2018), or network visualizations (Venturini, Bounegru, Jacomy, & Gray, 2017), which have reached an almost paradigmatic status in social media visualizations. In the business context, visualizations are strongly marked with practical value evaluations: they make sense of complex data, translating them to usable or valuable information (Halpern, 2014). Being valuable means giving insights to the current state of affairs but also includes a prophetic dimension by enabling a sight to the future (Beer, 2017). Visualizations, hence, are technologies that find value which it is impossible for an unaided human eye to locate.

Visualization in social media analytics is marked by two specificities: consulting business and the nature of data. First, while statistical methods and visualization practices are old conventions, the data on which the analyses are conducted are of a novel nature: the validity and generalizability of social media data have not been demonstrated or commonly constructed. Many hype-generating narratives evolve around the business insights that can be extracted from social media analytics, but practices of using these novel data sources are yet to be firmly established.

Second, social media analytics are situated in a business-to-business (B2B) context, where companies’ business proposals are coined around the notion of selling data, which also means they need to construct value for those data and repeat the hype narrative highlighting its importance, omnipotence and visionary capabilities (Beer, 2017). This places visualizations in a new context, where their role and function is defined in B2B negotiations. Analytics companies are simultaneously required to produce credible analytics and justified knowledge claims in a nascent area, while meeting hype-inflated expectations of business value. These two aims are potentially in conflict, for instance when selling analytics demands that results are presented in an appealing form, which might not represent the analytics process accurately.

Data visualizations as revelation and persuasion

As noted above, in organizations, visualizations are an established form of knowledge production. They are representations that have been established
as useful and credible means of depicting information, making them intertwined with practices and structures of institutional politics and legitimacy (Scott & Orlikowski, 2012; Lynch & Woolgar, 1988). The role of visualizations is further emphasized through the prevalent process of datafication, whereby increasingly many aspects of human life are quantified and captured in a form amenable for computational analysis (e.g. van Dijck, 2014). Data visualization is essentially a way to access and make sense of data, and thus the increasing pace of datafication leads to a growing importance of visualization methods (Kennedy & Hill, 2018).

Previous research has discussed visualization in data analytics as a representational practice that enables one to reveal hitherto hidden, or otherwise inaccessible, information in data (Coopmans, 2014; Halpern, 2014). As such, the commercial attraction of visualization stems from the promise of yielding insight into data, portrayed as a vital strategic resource for business (Coopmans, 2014). However, data analysis promises unforeseen increases in business value only for those who are in possession of the required visualization tools and skills to use them—a conception dubbed as artful revelation by Catelijne Coopmans (2014). Underlying the notion of artful revelation is the idea that visual representations of data enable viewers to see for themselves, or witness patterns in data first-hand, thus lending credibility to the produced knowledge. The idea of first-hand witnessing as a source of credibility has its roots in the history of scientific experiments and publishing (Shapin, 1984). Further, in the context of visualizations, credibility through witnessing is linked to the ideal of ‘mechanical objectivity’ (Daston & Galison, 1992), according to which visual representations should strive to depict patterns in data truthfully, free from the biasing influence of subjective interests and aesthetic judgements (Frow, 2014; see also Kennedy et al., 2016).

However, digital visualization techniques can also allow images to be manipulated without being restricted by the underlying data, to make them aesthetically pleasing, or to highlight certain selected aspects and downplay others (Frow, 2014). Consequently, as Emma Frow (2014) has argued in the context of scientific publishing, the issue of credibility in digital visualization concerns the skills of those producing the visualizations. Visualizations that look carefully prepared and aesthetically pleasing are easily ‘interpreted by readers as reflecting skill and expertise on the part of the author’ (p. 258). Thus, credible revelation in digital data visualization is as much a matter of skilfully selecting and portraying the right aspects in data, as one of depicting objectively existing patterns.

In this vein, critical accounts portray visualizations as doing persuasive work; that is, presenting particular viewpoints as more acceptable or valuable.
than others (see Kennedy et al., 2016 for a summary). Kennedy et al. (2016) have shown that the persuasive powers of visualization can be linked to prevalent visualization conventions, which imbue visualizations with an ‘aura of objectivity’ (p. 723), making them seem transparent and factual. However, as Frow (2014) argues, perceptions of the objectivity of digital visualizations—and thus the issue of their credibility—are intertwined with assessments of the skill of the visualizers on the part of the visualizations’ viewers. Consequently, skilful use and aesthetic aspects are central to the persuasive work of visualizations.

Acknowledging the specificities of social media analytics as a research context, as well as the previous research on the purposes of visualizations in society, we formulate our research question as follows: How are data visualizations used in social media analytics, and how are the credibility and value of analytics constructed through visualization processes?

**Data and method**

Our empirical study is based on ethnographic field notes and thematic interviews collected in four companies that analyse social media data as a part of their business endeavours. In total six persons across the four companies were interviewed. The interviewees were all in positions of management or middle-management, although, in the smaller companies these are in practice also operative employees. The companies vary in size and stage: Three are mid or early-stage start-ups, with less than fifteen employees. The fourth company is an established firm, the main business of which is surveys, but which aims to expand to social media data. The main product of one of the smaller companies is network analyses and visualizations. Another smaller company focuses mainly on distributing and cleaning social media datasets instead of doing in-depth analyses themselves, but they offer some basic visualizations and summaries.

We approached the interview transcripts and ethnographic notes qualitatively, focusing on the parts where the material concerns visualization. Using Atlas.TI and an inductive approach, we trace the different functions and meanings associated with visualizations, and the ways in which they are used to reveal information and to persuade. Both authors first coded the material independently using a grounded approach (cf. Glaser & Strauss, 1967), after which we met to discuss and refine the coding, and to group our codings into larger categories and themes. After reclassifying the material, we ended up with six themes: interestingness, objectivity, credibility,
communication to lay people, aesthetics, and customer apprehension. Next, the passages under these codes were reread and reorganized to chart out, first, the central role visualizations play in imbuing social media analytics with credibility and value, and second, the practices of generating visualizations, through which this role is enacted.

Credibility and value in social media analytics

In data analytics, the belief that data constitute a baseline of truth is prevalent (Ruppert, Isin, & Bigo, 2017). However, in our material, we observed that social media analytics are characterized by the conception that social media data are messy and heterogeneous; it is a form of data that is unreliable compared to traditional, institutionalized data forms which have been accepted as solid evidence for phenomena. Despite this starting point, our case companies were preoccupied with the idea that there is something essential hidden in the data: it is a matter of revelation and presentation to find it. The analysts expressed that the credibility of analysis is based, first, on the data containing patterns, which can be uncovered and utilized in a reliable manner, and second, the practice of using automated means to produce results that are free from subjective bias. Consequently, the results of analysis and measurements were deemed credible when they were grounded in, or driven by, data and analysed using objective means. This notion relies on the idea of quantification as a source of trust in contexts where expertise has not been demonstrated (Porter, 1995; Halpern, 2014; Elish & boyd, 2018).

In a context marked by the idea of messy data, visualizations are used for imbuing analytics with credibility and value. They are used to build coherence and to reveal hidden structures and patterns of that which is in the data. Thus, epistemically visualizations not only describe, but bring out the essential, show patterns in data in a manner that does not distort their depiction, letting the data drive the visualization. Automation plays a crucial role in the process, as interpretation takes place as an interactive process between the machine and the human:

We would need something programmatic to browse through that stuff and condense it, and to categorize, classify and present it. And then the researcher could sort of see that infographic, in quotation marks, and detect the point they want to take up, what I want to bring out and make a point, that this information could be valuable.
Credibility through automation, then, is seen as a prerequisite for valuable, marketable analyses. However, although automation and data-drivenness are associated with credible analytics, visualizations are a means through which patterns in data are ultimately revealed, in the sense of making them visible, that is, amenable for assessment and examination. Without visualization one has to trust the algorithm, but visualizations let the interpreter do the interpretations. This is a form of credibility produced through witnessing, which is specific to visual depiction of patterns.

Despite being based on algorithmic data-driven analytics, data visualization is also an acquired skill when it comes to discovering patterns in data. In Halpern’s (2014, p. 22) words, visualizing is ‘making something that is out of sensory recognition relatable to the human being’. Thus, visualization expertise is relevant for assessing the credibility of analytics, as the results of algorithmic analysis can be hard to decipher without the aid of visual representations. Further, demonstrating expertise in visualization to clients also lends value to analytics, as will become apparent below. Next, we will scrutinize three practices through which human skills are involved in the process of creating credibility and value in social media data visualization.

**Simple-boxing to conceal complexity**

The consulting context is marked with an unequal distribution of knowledge: the customers are buying expertise from the consultant. This generates a need to communicate potentially complex information to the customers in an understandable delivery format, which implies simplification. The necessity of simplicity also affects visualization choices and preferences; in practice, it leads good visualization practice to be a customer-driven concept within the companies. Good visualizations are easily comprehended and effectively work as descriptions of the situation or, preferably, decision-making devices for the customers. This means that visualizations need to show, in one glimpse, understandable depictions of reality, stripped of its messiness and complexities. This is a process referred to as ‘simple-boxing’ by one of our informants, during which details of the complex reality become hidden.

So we will probably put some colour symbols there like red and green or something according to the sentiment. So that in a glimpse you could see if there’s a lot of red or blue or green.
In this vein, as a part of its simple-boxing process, the network analysis company has developed heuristics considering the amount of nodes that can be reasonably presented in a network graph so that it remains understandable to the customer. The company has learned by practice to filter out information that is deemed unnecessary and irrelevant based on expertise accumulated in business negotiations. This process is also affected by the presentation technologies and conventions institutionalized in business practices. One of our case companies explained how they select the dimensions of a statistical model so that they can be fitted and beautifully visualized on a single PowerPoint slide. They have developed an understanding that neither a salesperson nor the customer can handle more dimensions.

As a way of communicating complex information, simple-boxing also serves to enable assessments of the credibility and value of analytics. The following quote from the firm specializing in survey research demonstrates how communicating information effectively means revealing it by means of data visualization.

> We have in fact thought about the issue of how to get to raw data in such a way [...] that we could somehow define it more strictly, standardize in a way the output [...] to visualize it to the researcher so that it would be easier for them to [...] recognize whether there is anything meaningful there, versus just getting 50,000 lines of some text.

Simple-boxing thus is a way of crafting visualizations that enables the analysts to comprehend complex information and communicate it effectively to different parties, including clients, consultants selling the analytics, and colleagues. Nevertheless, simple, comprehensible visualizations are regarded as representing patterns in data truthfully, provided they are based on analyses that are judged to be reliable. For instance, one analyst suggested that a visualization of clustering results truthfully depicts the underlying patterns in data. This was despite a somewhat subjective choice of the number of clustering dimensions displayed in the results. Thus, simple-boxing with visualizations need not imply unrealistic representation, provided that the depicted results are produced through credible analytics.

**Flatter-boxing to highlight the interesting**

Apart from working as devices that reveal essential information, our informants also identify a need for visualizations to point out what is *interesting* in the data.
Revealing the interesting is a practice thought to be dependent on the skills of the analyst, which involves finding the potential structure and patterns, and picking correct ways to visualize them. For example, both generating and interpreting visualizations are referred to as craftsmanship in the network analyses consulting company—an expert can immediately see if there is a structure in a network visualization. As the following quote highlights, the represented patterns might not constitute an ‘absolute truth’, and thus their status is dependent on the conceptions and craftsmanship of the analyst.

Absolute truths are hard to formulate about [network visualizations]. When you have, over the years, done quite many of those [...] you start to form some sort of a picture about what the network might tell, kind of a carpenter’s feel, that you learn to recognize different kinds of wood merely by smell or touch.

One popular way of building structure and pointing out the interesting from social media data is by making lists and rankings. This is a practice postulated by the business context: in a situation of constantly striving for profit and competition any information that creates order becomes valuable (cf. Halpern, 2014). For instance, one of our case companies publishes lists of actors’ performance on social media and uses various visualizations to track their relative positions over time. These processes require visualizations to depict patterns in data that are deemed to be interesting, such as displaying differences or orderings among the measured units, ‘so that they can constantly observe on one screen if they are going down or up’.

Thus, visualizations generate value for analytics by highlighting differences. What is crucial for this kind of revelation is that patterns are discovered that match the analyst’s conceptions of valuable information, with less emphasis being based on reliable and data-driven analysis processes. Hence, the skills of the human analyst are essential for finding and depicting the interesting. We have named this practice flatter-boxing. In line with the notion of artful revelation (Coopmans, 2014) the use of visualization skills, or craftsmanship learned through experience, is a condition for the production of valuable analytics that reveal interesting patterns in data.

Pretty-boxing to induce marvel and convey expertise

Apart from being simple and interesting, visualizations also need to be attractive (cf. Brinch, this volume). Visualizations are representational
devices that both communicate expertise and build trust. This effect is largely gained through aesthetics; visualizations need to look beautiful but also difficult to create (cf. Coopmans, 2014; Frow, 2014). This is particularly the case with network analyses, which are easy to sell merely because they generate what the consultant calls a ‘wow effect’:

It’s always like wow, because you get fancy, pretty visualizations, so that it’s an essential part of the network analysis that you generate the visualization. People often equate [the analysis and the visualization], [...] people are satisfied since they get a view of what is really happening, that it’s not just a stream of messages, but there is a structure.

Compared to topic detection methods, another computational text analysis method with less clear visualization practices, the interviewee thought that networks are easier to interpret for people with no technical background. At the same time, the interviewee acknowledges that the customer can concoct a story based on network diagrams; they are like Rorschach tests that support many versions of the truth, when sometimes no truth actually exists.

Thus, aesthetically pleasing visualizations can give an illusory feeling of understanding. As Kennedy et al. (2016) have argued, the persuasive work done by visualizations is in part due to prevalent visualization conventions. Our evidence conforms to the notion that aesthetic considerations are involved in assessing analytics (cf. Frow, 2014). Through pretty-boxing analytics with beautiful visual renderings, analysts can convey expertise, which simultaneously builds credibility and value. However, this persuasiveness of visual presentation is considered dangerous in cases where the reliability of analytics has not been demonstrated. As the analysts of the survey analytics firm expressed in the interview, the results of even epistemically dubious analytics can be successfully marketed, given appropriate presentation.

Discussion: When and how visualizations work

The evaluation of visualizations among our informants is functionally oriented: visualizations should work, or manage to do what is expected in a business context. When it comes to social media data, the notion of the data being messy, heterogeneous, and complex is a narrative which is part of the business offering of these companies. In our case companies,
data visualizations are used to translate knowledge extracted by various algorithms to applicable business knowledge. In order to be applicable, the produced analytics are required to be simultaneously credible and valuable. These aims cannot be straightforwardly accomplished by applying seemingly objective algorithmic procedures to extract information from data. In addition to appearances of objectivity, successful consulting also requires that the product is relevant and interesting for the customer, and that the customer is able to understand and benefit from the results.

In our research, the translating role that visualizations play between humans and technology in the context of social media analytics becomes evident. As explained above, visualizations are regarded as tools which can hide complexity and represent data beautifully, potentially glossing over multifaceted decisions made during data analysis. As Coopmans (2014) argues, data visualization is a representational practice which enables analysts to discover valuable information in data and convey their expertise to customers. Our investigation highlighted how in social media analytics the status of visualizations is defined by the extent to which they communicate interesting and essential information to the customer in a simple and beautiful form. This status is constructed in the interplay between automated analysis processes and application of the analysts’ visualization expertise.

Whether a visualization works, or is able to fulfil its business purpose, is hence not straightforwardly related to the perceived objectivity of analysis. Rather, the status of a visualization is evaluated on the basis of how well it manages to display interesting and useful results, which depend on customer needs, and the abilities of both the customer and the analyst. Despite the continuous contemplation of data quality and the acknowledgement of craftsmanship, the buyers’ needs emerge as an important factor affecting the visualization practices. This makes the selection and interpretation of visualization models an example of an interested reading of reality (Rieder, 2017), where the actors are reading the data and visualizations in ways that show patterns or differences, following their business models’ predications. Hence, the aim of generating interesting patterns that further business objectives can override the epistemic aim of realistic representation. Here, we witness an interplay between assessments of credibility and value in evaluating social media analytics visualizations. Visualizations emerge as devices which enable analysts to retain analysis credibility while constructing a valuable representation of patterns in the data.
Conclusion

Our investigation identified three intertwined practices that help analysts to achieve the aim of producing credible and valuable analytics: simple-boxing, flatter-boxing, and pretty-boxing. In addition to methodological norms guiding data analysis, social media analytics are shaped by the need to communicate intelligible and attractive results to customers. This criterion potentially stands in tension with the aim of realistic representation. When the two aims conflict, the analysts are faced with the challenge of conveying their results in a simple enough form to sell their product, while retaining the customers’ interest and credibility of analysis. Using visualizations to simplify and beautify information provides a way to accomplish this goal, reconciling the pursuit of interesting results with credibility grounded in data-driven revelation.

Various features exhibited in our material influence how visualizations are constructed in social media analytics to serve the purposes of business. First, they are used to communicate exactness and objectivity of the analysis while hiding the complex processes of collecting, cleaning, and analysing the data. Visualizations are not necessarily shown in their original, valid form, but they are tweaked, simple-boxed, flatter-boxed, and pretty-boxed until they better communicate the desired message; for example, until they highlight the differences between measured units. As an end result, however, they still communicate objectivity based on numbers and impersonated analysis (cf. Porter, 1995). Second, after being tweaked, visualizations provide more clues to lay people to interpret the information in comparison to word lists or numerical representations, which again increases the perceived credibility of the analysis. Third, their credibility is intertwined with notions of usability, interestingness, and business value (cf. Frow, 2014). These are issues connected to the conventions accumulated in the business context over time.

Finally, the entire social media analytics business builds on the assumption that social media data can provide valuable insights to organizations. Visualization is what makes algorithmic output a form of human knowledge, but the interpretation and skills of the analysts are an essential part of the process. The algorithmic analysis process, human skills, and visualization all work together to build structure in and discern value from messy data. As Halpern (2014, p. 30) eloquently formulates, ‘[visualization] is a set of techniques by which to manage, calculate, and act on a world of incomplete information’. In business context, data visualizations act to increase the business value of data in concrete ways; they transform the data to an
authoritative object, an artefact that yields for effects and outcomes, facilitating not only understanding but also business. By doing this, they play an important role in establishing the omnipotent nature of data analytics (cf. Beer, 2017). As more and more organizations are tempted by the possibility of taking advantage of automated analysis to understand their shareholders’ discussions on social media, and to make better informed decisions, the practices of using visualizations within this field can have far-reaching effects in the society.

References


About the authors

Salla-Maaria Laaksonen (D.Soc.Sc.) is a postdoctoral researcher in the Centre for Consumer Society Research, University of Helsinki. Her research areas are technology, organizations, and new media, including organizational reputation in the hybrid media system, the organization of online social movements, and the use of data and algorithms in organizations.

Juho Pääkkönen (M.Soc.Sc, social and moral philosophy) is a PhD Candidate in Sociology at the University of Helsinki, Faculty of Social Sciences. His research focuses on epistemic conceptions and methodological issues pertaining to computational methods and big data across social sciences and data analytics.