Representing the Unknown

A Critical Approach to Digital Data Visualizations in the Context of Feminist Film Historiography

Dr. Sarah-Mai Dang, Philipps-Universität Marburg, sarah-mai.dang@uni-marburg.de

Abstract

Considering the growing production and application of data in the era of digitalization, the objective of feminist historians to tell a story differently rather than telling a different story has acquired a new urgency. This article makes the case that in order to foster new perspectives and advance our understanding of women's influence in film culture, we need to further explore new forms of presenting historiographical research by taking advantage of digital tools and methods. In this context, it is suggested that data visualizations can offer a productive approach for telling women's achievements in early film industries while taking into account ambiguities, contingencies, and blind spots inherent to history.

Keywords: *digital film history, research data, data visualization, feminist theory, early cinema*

Chaplin, Griffith, and Smith

Charlie Chaplin is certainly one of the most famous actors of the silent era. Many people have seen one or more of his movies or at least have a clear image of him. Chaplin has become an iconic figure. In contrast, D.W. Griffith is not very well known beyond cinephiles and film professionals. At media departments, however, many students learn about the US-American director in one of the first classes on film history. Very early on, I was taught that Griffith' movie Intolerance (USA 1916) is considered a pioneering classic because of its extraordinary film sets and elaborate montage. Interestingly though, what I did not come to know was who was responsible for the acclaimed editing. It was only recently that I have learned form the collaborative online project *Women Film Pioneers Project* (WFPP) (https://wfpp.cdrs.columbia.edu/) that Rose Smith and her husband James Smith edited a number of Griffith' films, including Intolerance. Remarkably, however not surprisingly, notwithstanding the celebrated montage, James Smith' wife, Rose Smith, seemed to be forgotten in later sources – and thus in the course of history (Hatch 2013).

Smith was not the only woman effaced in film history. Since women's editing was "considered to be merely technical rather than creative", as film scholar Kristin Hatch explains, their work was not credited in the films. Film credits, as presented in the prints themselves, are one of

the sources historians would first go to in order to seek information. Thus, searching for evidence to tell the story of Rose Smith turns out to be quite an endeavor. In general, women's significance for Hollywood's visual style has been little documented (Hatch 2013a).

The reasons for this marginalization are manifold. To dismiss "women's work" as menial labor is probably the main reason, following the current research in feminist film history (see, for example, Gledhill and Knight 2015; Hill 2016; Gaines 2018). Furthermore, notwithstanding its significance for today's film theory, the focus on the audiovisual representation of women on screen in the 1970s and 1980s may have made the many women behind the scenes disappear from our sight, as feminist film scholars such as Jane Gaines or Heide Schlüpmann, as well as other colleagues, have argued, and as I have discussed elsewhere (Dang 2018). A third reason for the absence of women in film history, closely related to the first one (the disregard of labor), is that in the silent era, women's roles varied and were acknowledged differently. For instance, women commonly known as "cutters" were also referred to as "editors" or "scenario editors" if their work involved dramaturgical aspects even though there might have been only little manual cutting to it (Hatch 2013a). Job titles may change over time and differ from country to country. Therefore, it is difficult to get a comprehensive overview of the many women involved in film montage.

However, while this article focuses on women in film history, we need to be aware of that both women and men have been excluded from film history due to historic manufacturing conditions as well as specific conceptions of film and film history. For example, by focusing on the director of a film in the context of the *auteur theory* the many facets of film production including the numerous people that collaborated in various areas have been neglected for a considerable time. The development of film history is yet another example of how theoretical concepts, research interests and objects are closely interlinked and thus, as feminist theorist Donna Haraway has pointed out, that knowledges are always situated in a specifc context (Haraway 1988).

How we categorize and conceptualize tasks and professions such as "director", "authorship", or "editor" affects the (non-)representation of women in film history, as pointed out by a wide range of film scholars (see for example Bean und Negra 2002; Callahan 2010). Whether we identify a woman as "cutter", "editor", or "assistant director" matters because categories imply particular assumptions with regard to significance and status. Categorizations effect how we evaluate a woman's role in history. This is particularly important in the context of digital databases. Ascribing specific metadata to discrete elements is no neutral procedure but a deeply political act of interpretation, as digital humanities scholar Miriam Posner writes (Posner 2016). In other words, feminist film historians Shelly Cobb and Natalie Wreyford note, it is an authoritative process of power and authority that risks normalizing and essentializing meaning (2017, 115).

The example of Rose Smith is only one of many which demonstrate why it has been difficult to reconstruct women's work in early film industries and tell their stories to today's audiences. While the lack of evidence is certainly one major challenge, in this article, I would like to shift

the focus from archival research to the presentation of findings, that is to the presentation of research itself. How we provide access to sources and research today determines the way we will envision the past in the future. While it is crucial that historians carry on digging up treasures from the archives in order to tell more stories, we also need to further reflect upon *how* to tell more stories. How can we show research results in an engaging yet critical and self-reflective manner? How can we talk about past events we can only imagine (Dall'Asta and Gaines 2015)? How can we represent the unknown (Gaines 2018)?

In light of the increasing digitalization which is impacting both our research objects and methods these questions have become an even greater challenge. In the use of new technologies inclusion and exclusion mechanisms can be easily reproduced and exponentially amplified – or, and this is important to keep in mind, counteracted. As I have summarized in a previous article, there is the risk that due to the focus on 'big data' meta-history is favored over micro-history. Furthermore, while the implementation of standard metadata can foster interoperability and collaboration, at the same time it might reinforce blind spots and obscure specific details. Last but not least, mass digitization of objects allows for easy access and global circulation of artifacts, on the one hand. On the other, analogue sources will possibly be left out by students and scholars as well as users in general. At the same time, digital platforms can inform us of the many existing archives and their valuable collections and thereby encourage further research on site (Dang 2018).

Against this backdrop and considering the growing production and application of data in the era of digitalization, the objective of feminist historians to tell a story differently rather than telling a different story has acquired a new urgency. Today, there are numerous online projects that feature women's achievements in film history. In my view, however, platforms such as the *Women Film Pioneers Project* are still far from reaching their full potential (Dang 2020b). Without doubt the WFPP database offers a great deal of information on women workers in early cinema as well as additional references. Yet in order to foster new perspectives and advance our understanding of women's influence on audio-visual culture, I suggest to further explore new forms of presentation by taking much more advantage of digital tools and methods. Based on my studies on media aesthetics, research data and data bases and on what I shall present in this article, I assume that data visualizations in particular open a productive methodological path for telling women's significance in early cinema while taking into account ambiguities, contingencies, and blind spots inherent to film history.

Various data can be visualized in many different ways, for different purposes and in different contexts. For instance, visualizations provide access to research and cultural sources; they help us navigate archives and analyze data. They might demonstrate an idea and make us reflect on a particular subject. They can also invite us to ask further questions and explore new territories. Moreover, as I will show in this article, they enable us to rethink traditional approaches in the humanities and further develop film and media studies concepts and methods.

Before I lay out my arguments for further exploring data visualizations in the context of digital film historiography I want to sketch out some general challenges by drawing attention to the representation of the Corona crisis (as of July 2020).[1] The many familiar examples in this context can help us to better understand what is at stake when visualizing data and other types of research results.

Data visualizations and Covid-19

The Corona crisis has drawn particular attention to statistics and data visualizations in and beyond academia. In order to demonstrate the dimensions of the pandemic, news media have presented numerous graphics on its effects: for example, maps which display the development of the virus in specific regions, timelines which show a possible infection rate if no measures would be taken, and bar charts which compare the number of people infected, cured, and deceased in various countries (Lammar 2020). Other graphics elaborated on the goal of flattening the curve (Wiles 2020) or the rapid expansion of the outbreak (Stevens 2020). By now, we are all familiar with the mathematical term of exponential growth, the greater increase with passing time.

In most instances, the visualizations are based on data from the Johns Hopkins University (JHU), which aggregates data from various institutions such as the World Health Organization (WHO), the U.S. Centers for Disease Control and Prevention (CDC), the European Center for Disease Prevention and Control (ECDC), the National Health Commission of the People's Republic of China, as well as local reports and the international physicians online community DXY. Most readers probably recognize the university's COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) [Fig. 1].
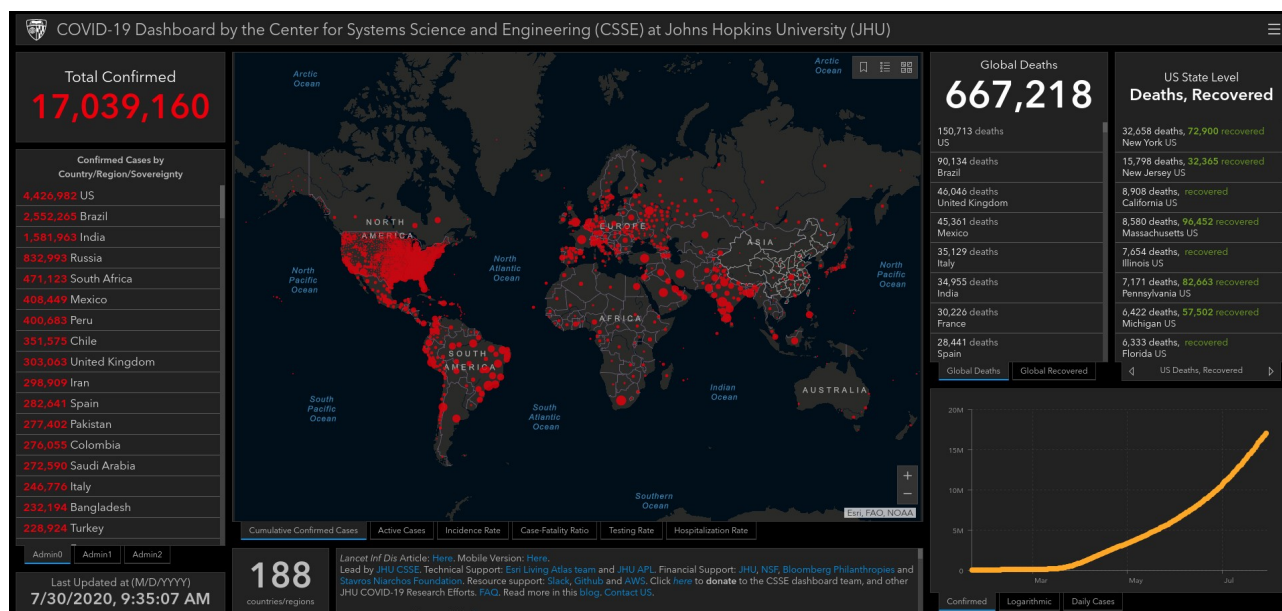

Figure 1: Screenshot of the COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. "COVID-19 Dashboard". *ArcGIS*, 2019,

---

1 This text was written in August 2020. Recent research on this topic could therefore not be taken into account. However, the examples are still helpful from my point of view.

As of July 2020, for example, the dashboard shows that the US is most affected by the virus, then comes Europe, India and parts of South America. Red circular areas indicate the regional gravity of the pandemic, confirmed by the stats of "Global Death" in the right column ranking US first (150,713), followed by Brazil (90.134) and the UK (46,046). The map clearly signals the fatal consequences of Covid-19, though it also displays the number of people who have recovered. From the dashboard we learn that the virus is to be taken seriously. This can be interpreted as the main message.

However, as we know from the media coverage, the pandemic and its data-based representation is more complicated. First, data on the developments of Covid-19 is captured in many different ways. Thus, the various data sets are only comparable to a limited extent. Second, in order to grasp the severeness of the virus we have to take into account the velocity of the spread, meaning the reproduction value (R), and not just the death accounts. Taking a closer look at the dashboard it becomes obvious that it gives us a very specific view on the virus' effects.

At first, by default, as pointed out by UX designers Dan Benoni and Louis-Xavier Lavalle in their animated case study of the COVID-19 Dashboard (Benoni and Lavalle 2020), we see an overview of cumulative cases. In this mode, almost the entire US seems to be infected. But in reality only a part – still a comparatively large number – of the population, has caught the virus. The fact is, as Benoni and Lavalle illustrate, the statistics and the visualization do not match. Readers might find this quite surprising since the Johns Hopkins University ranks as one of the top US American universities and thus the CSSE seems to be trustworthy – and it most likely is. Yet, as graphic designer and information scholar Edward Tufte has shown with his concept of the "lie factor", inadequacies are remarkably common in all kinds of institutions and areas.

In what has become a standard reference in the field of data visualization, *The Visual Display of Quantitative Information*, Tufte creates the "lie factor" in order to examine the proportional relation between data and its representation (Tufte 2001, 57–59). The "lie factor" can vary. In the case of the COVID-19 Dashboard it is relatively high. Benoni and Lavalle rightly argue, by completely 'infecting' an area visually, the map implies "it can't get worse, when in fact, it could" (Benoni and Lavalle 2020). In their view, a symbol map is unsuitable for representing the proportion of infected people. They suggest, "[w]hen data has a negative connotation, you should avoid showing cumulative cases" at all because it can "amplify/alter perceptions". If the dashboard's creators had chosen green or blue over red, besides a different type of representation, the numbers would have a very different effect on us (ibid.).

In addition to aesthetic questions of *how* data is presented – and consequently perceived and interpreted – it is, of course, also important to consider *what* is being represented. For example, the map tells us nothing about personal characteristics, age, gender, or health conditions (ibid.). We cannot trace how the virus has spread, though a map seems to be

particularly suitable for this information. Of course, a visualization can only focus on a limited range of factors without resulting in an information overload. However, we ought to try understand what these foci are. Visualizations do not simply represent what we assume is already there but also generate knowledge by relating to the world in a specific way. Data visualizations offer only a partial view; a view, however, that might appear natural in the act of re/presentation.

As stressed by scholars across disciplines, we have to closely look at the data a graph or diagram is based on. What data has been included and what data has been – deliberately or inevitably – excluded? Under which premises was the data generated? If we look at the dashboard, we see some substantial areas which are not red. According to news media coverage it is very much unlikely that this is because there are zero infections in these areas. As mentioned above, the tests are not equally performed across countries. For some regions, there hardly exists any data. Thus, media theorist Christoph Ernst concludes, also referring to Benoni's and Lavalle's analysis, that the freely accessible infographic does not just visualize the statistics of the global crisis but also the political agendas of nation states and how they seek to manage the curve (Ernst 2020).

Following this line of reasoning, we can state that data is always data politics, and so is data visualization. Despite the association with accuracy and evidence, data is neither self-explanatory nor neutral. All data "is *capta*, made, constructed, and produced, never given", media scholar Johanna Drucker asserts (Drucker 2016, 249). However, digital humanities scholar Charlotte Fillmore-Handlon explains, this does not mean that data is not objective. Referring to media historian Lisa Gitelman and literary scholar Virginia Jackson's introduction of *"Raw Data" Is an Oxymoron* (Gitelman and Jackson 2013), she writes, we need to:

> understand objectivity as 'situated and historically specific; it comes from somewhere and is the result of ongoing changes to the conditions of inquiry, conditions that are at once material, social, and ethical'. (Fillmore-Handlon 2016)

In this light, the terms "messy data" and "data cleaning" require careful scrutinization, as digital humanities researchers Katie Rawson and Trevor Muñoz argue in their plea "Against Cleaning" (Rawson and Muñoz 2019). They explain:

> The term 'cleaning' implies that a dataset begins as 'messy'. 'Messy' suggests an underlying order: it supposes things already have a rightful place, but they are not in it – like socks on the bedroom floor rather than in the bureau or the hamper. (ibid.)

Instead, the production and use of data is intertwined with human decisions and agency, statistician Nick Barrowman notes, arguing that data has no "mind of its own" (Barrowman 2018, 130). Therefore, like data visualizations, data itself is always already an interpretation (Fillmore-Handlon 2016). On this account, we need to analyze a) the source, b) the production process and c) the aesthetics of data visualizations as well as their perception in order to better understand the COVID-19 Dashboard and data visualizations in general.

Data visualizations and media studies

"Every discipline and disciplinary institution has its own norms and standards for the imagination of data, just as every field has its accepted methodologies and its evolved structures of practices", Gitelman and Jackson note. In their view, data is to be taken as a "matter of disciplines – rather than of computers" (2013, 3). I agree with their proposed perspective and like to add that, likewise, we need to consider how data visualizations – both as object of study and tool for investigation, or even method – are conceived in different disciplines.

Already before the Corona crisis, data visualizations have been broadly applied in various sectors and disciplines, mainly in fields known for their quantitative approaches, such as economics, demographics, or statistics. However, due to the growing production of research data in the (digital) humanities, data visualizations have been also – slowly but steadily – gaining in significance in the field of film and media studies.

In this article, I use the term data visualization in the broadest sense in order to retain an open mind for all sorts of digital representations of information and knowledge. A more general definition also accounts for the heterogeneity and complexity of data and data visualization in film and media studies (Dang 2020a). Nevertheless, for heuristic purposes it can be helpful to distinguish between visualizations of (meta)data *about* artifacts (information visualization) and visualizations *of* artifacts themselves (media visualization), as suggested by media theorist Lev Manovich (Manovich 2013). In the context of his much discussed Cultural Analytics approach, for more than ten years Manovich has been analyzing large amounts of images by applying various visualization techniques. For this purpose, Manovich's Software Studies Initiative (http://lab.culturalanalytics.info/) uses various tools that derive from different rather unfamiliar disciplines, media scholar Eef Masson points out (Masson 2017, 29). This is why we need to reflect how digital humanities are shaped by specific applications and thus specific intentions, assumptions, and epistemological definitions. For example, ImagePlot is based on the software ImageJ that was initially created for medical scans and later used for biological microscopy. To understand how digital tools work and how to interpret the results of data-based visualization is even for experts quite a challenge (ibid.).

ImageJ sum visualization of Sergio Leone's Once Upon a Time in the West
(Italy/US/Spain, 1968)

Figure 2: Screenshot of an example by media scholar Kevin L. Ferguson that Christian Olesen presents in his overview on image data visualization, cropped by the author. Olesen, Christian. "SEMIA and Moving Image Data Visualization: an overview and brief introduction". The Sensory Moving Image Archive, 20 May 2018, https://sensorymovingimagearchive.humanities.uva.nl/index.php/2018/05/20/semia-and-moving-image-dataviz-in-film-and-media-studies-an-overview-and-brief-introduction/. Accessed November 2020, cropped by the author.

Considering the many implications of humanistic digital research, we need to look at how new tools affect our approaches and also our objects of study. For instance, transforming artifacts such as films or paintings into data raises many fundamental methodological and epistemological questions. For example, how does this "recoding" alter our research object? What is the relationship between original artifact and data? How does this approach change film and media studies?

While the "translation" of images into data and then into visualizations provides new ways of comparison and analysis, as film historian Christian Olesen demonstrates (Olesen 2018), it is crucial to realize that visualizations present something different from the original object. We see references that might resemble the original object. But in contrast to what Manovich implies, I contend that we are not able to see the "objects themselves" (Manovich 2013). The "objects themselves", I would argue, is a misleading term, like "raw data", since it implies that each artifact can be defined by an ontological core aspect when in fact it is a matter of perception how we conceive an object. For instance, a "montage visualization", which accumulates all takes of one film in a mosaic-like overview, or a "summary visualization", which superimposes single images [Fig. 2], changes our perception of the visualized film and consequently our understanding of what film theorists usually define as a time-based medium.

As I have sketched out elsewhere, following Olesen, media visualizations in Manovich's sense allow us to see new aspects of artifacts, such as color schemes in genre film or image compositions in films, and thereby broaden our understanding of media research in terms of theories and methods (Dang 2020a, 124–127). However, regardless of the disciplinary potentials, I think that the translation process which according to Manovich characterizes

media visualizations is true for all kinds of visualization. For instance, visualization experts Katrin Glinka and Marian Dörk consider the translation of non-spatial data structures into geometrical forms and other visual arrangements (in order to show, for example, relations between philosophical concepts) a particular challenge for information visualizations in art history (Glinka and Dörk 2018, 238).

If we, in line of what I have addressed above with respect to the Covid-19 dashboard, understand both data and data visualization themselves as artifacts, and thus as "media data", the distinction between information visualization and media visualization becomes even more debatable. Since all data requires "material expression" (Gitelman and Jackson 2013, 6), a simple spreadsheet is already some form of visualization (ibid., 12). It is this "material expression" I will discuss in the following with regard to film historiography.

Data visualizations and film historiography

Despite the current discourses on 'big data' in the digital era, we must not forget that data has been essential in the humanities long before the "computational turn". Data has played a significant role in the context of historiographical studies, for example in stilometric film analysis developed in the 1970s when film studies was becoming a discipline (Olesen 2017) – in addition to all kinds of visual knowledge productions that can be traced way back in cultural history (Drucker 2014). Nonetheless, we can observe that the number of data-intensive projects in film and media studies has increased significantly in the past few years. Many different examples can be found in the context of film historiographical research.

Figure 3: Screenshot of the *Cinemetrics* website that shows a case study on Charlie Chaplin's CITY LIGHTS (USA 1931), slightly cropped by the author. This graph is also highlighted by Olesen to give an example of a statistical data visualization (2018). "CITY LIGHTS (1931, USA), directed by: Charles Chaplin". *Cinemetrics*, http://cinemetrics.lv/movie.php?movie_ID=5213#nogo517. Accessed August 2020, cropped by the author.[2]

One of the first data-driven projects that has become widely known is the online platform on film editing and shot length, *Cinemetrics* (http://www.cinemetrics.lv/) [Fig. 3]. It was created in

---

2 This graph is also highlighted by Olesen to give an example of a statistical data visualization (2018).

2005 by film scholar Yuri Tsivian and computer scientist Gunars Cijvans in order to provide statistical evidence for the transformation of film style. As Olesen explains in his comprehensive analysis of *Cinemetrics*, the crowd-sourced data uploads do not follow any research data management standards. Instead, the roughly 15,000 titles

> *constitute a heterogeneous data mass which facilitates comparison between primarily limited corpora with uniform, technical standards rather than providing evidence for a universal, evolutionary film history as in the 1970s.* (Olesen 2017, 44)

The central element of *Cinemetrics* consists of a standard format for statistical data visualization, a graph that can be annotated by users, Olesen observes. By providing various cutting parameters the website does not only allow for multilayered comparisons but reflects the variety of scholarly concepts of stilometrics such as the Average Shot Length (ASL) and the Median Shot Length (ibid., 45–46). While *Cinemetrics* seems to be indebted to positivist traditions that focus on accuracy and verifiability of patterns by means of statistical data, it allows for a detailed analysis of film editing. Thus, as Olesen argues, different than one might have initially assumed, in spite of the statistical focus, cinemetrics approaches qualify as exploratory, critically, and inductive, thus humanistic. Although they might appear as utterly scientistic, they possess the potential of bridging the hermeneutic and quantitative epistemic traditions (ibid., 40–41).
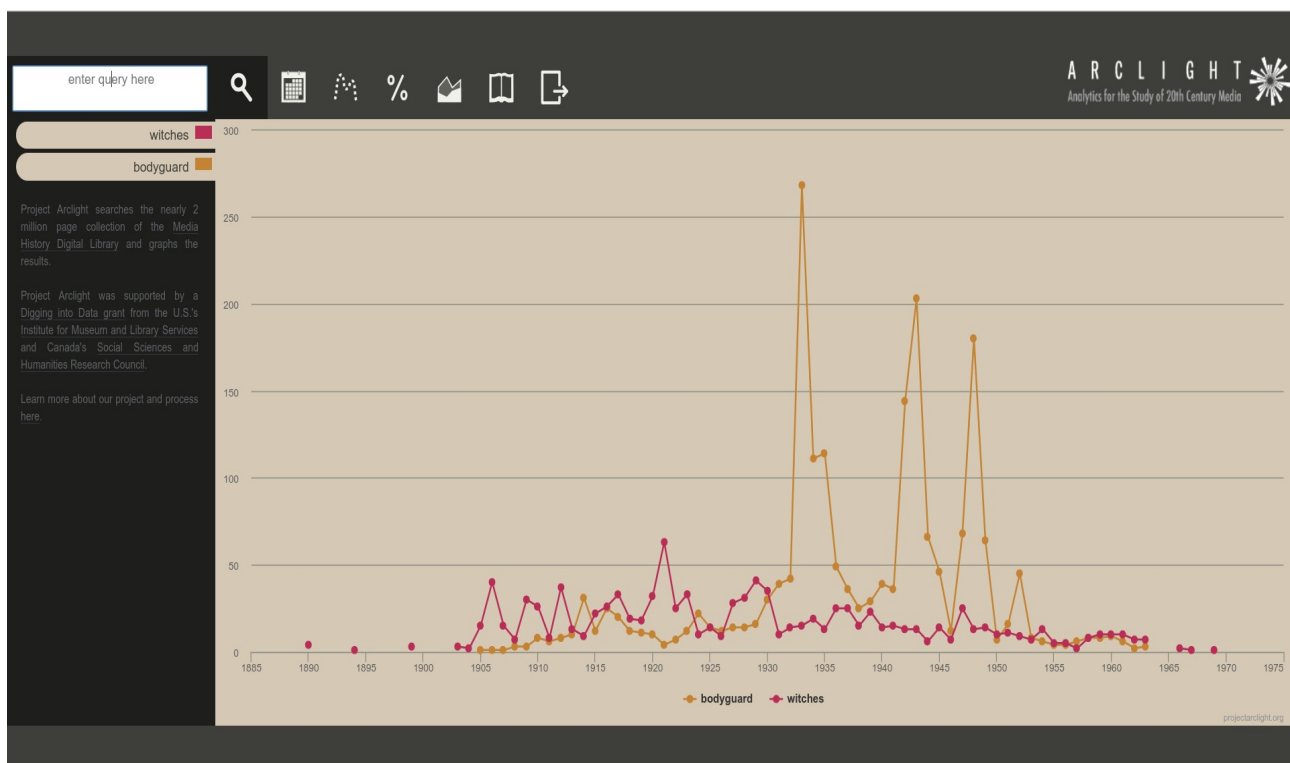

Figure 4: Screenshot of the *Project Arclight* website that shows how many times the terms "witches" and "bodyguard" are used in film magazines over several decades. *Project Arclight*, 2015, https://projectarclight.org/. Accessed December 2018, cropped by the author.

Besides *Cinemetrics*, a considerable number of additional film historiographical projects have been developed in recent years. When it comes to the question of how digital tools shape humanistic research it is also worthwhile to take a closer look at the *Project Arclight*

(https://projectarclight.org/) [Fig. 4]. This media history platform was initiated by media scholars Charles Acland and Eric Hoyt in 2014. Together with their teams they have created a software project that allows users to search for trending keywords in about two million film magazines and journals in the *Media History Digital Library* (MHDL) and the newspaper archive of the Library of Congress. The MHDL was founded by media historian David Pierce. It mainly contains sources up to 1964 since these are no longer protected by copyright but instead assigned to the public domain. The search results are visualized in diagrams and maps with direct access to the digitized artifacts aggregated by the Internet Archive (https://archive.org). *Project Arclight* demonstrates that data-based research can enable both a quantitative metadata analysis and a qualitative close reading approach. Due to the direct linking users can zoom in and zoom out while retaining the entities' integrity and historical context. Micro and macro histories are brought into a dialogue.

As for data bases and visualizations that explicitly focus on gender representation, besides the WFPP, the BFI Filmography (https://filmography.bfi.org.uk/) has gained international recognition [Fig. 5]. The project seeks to give a comprehensive overview of UK film industry from the beginning of film history. The website displays categories like the "most prolific actress" and "most prolific female director", as well as the gender balance in British feature films, among other aspects, for example, film subjects and international co-productions, in various graphs and diagrams. In order to address political issues such as diversity and inclusion, the project has added an extra layer that focuses on gender by drawing on additional data bases such as the Office for National Statistics and manual biographical research (ibid.). In doing so, the data curators are well aware that binary categories and external gender attributions leave out nuances. However, although this method is not perfect, as it is explained on the website, the focus on gender fosters further discussions about equality in film industries.

While the BFI Filmography does not tell individual stories, it serves as an impressive example of how quantitative data can make the absence of women in film history visible. Furthermore, it demonstrates that the absence of women is not a personal experience but a structural problem, as Wreyford and Cobb explain (2017). They both were engaged in the research project "Calling the Shots. Women and Contemporary Film Culture in the UK" led by Cobb and Ruth Linda Williams. Based on BFI data sets, their goal was to identify women's various roles in British film productions from 2000-2015 using both quantitative and qualitative methods in a feminist manner – that means, in their view, passionately, collaboratively, and critically (ibid., 114). Also taking into account the dilemma of labeling others (ibid., 115–117), in their article "Data and Responsibility", they argue that statistics and data have played a key role for feminist research and politics, for instance in understanding inequalities and consciousness raising (ibid., 108–109). Following these impetuses, they have created a research project that seeks to encourage a redistribution of financial means in a more equitable way by presenting various statistic findings to the public via news media, for example the low representation of women of color in the UK film industry (ibid., 116). Moreover, according to the BFI statistics, a general absence of women can be observed in contemporary British filmmaking (ibid., 124).

However, like other databases, the BFI database does not provide a complete overview of all the women who have been working in British film. The figures should be treated with caution, as Wreyford and Cobb emphasize. Ideally, they should be complemented with further investigations in order to provide a broader and more nuanced overview and to include women in the official history (ibid.).
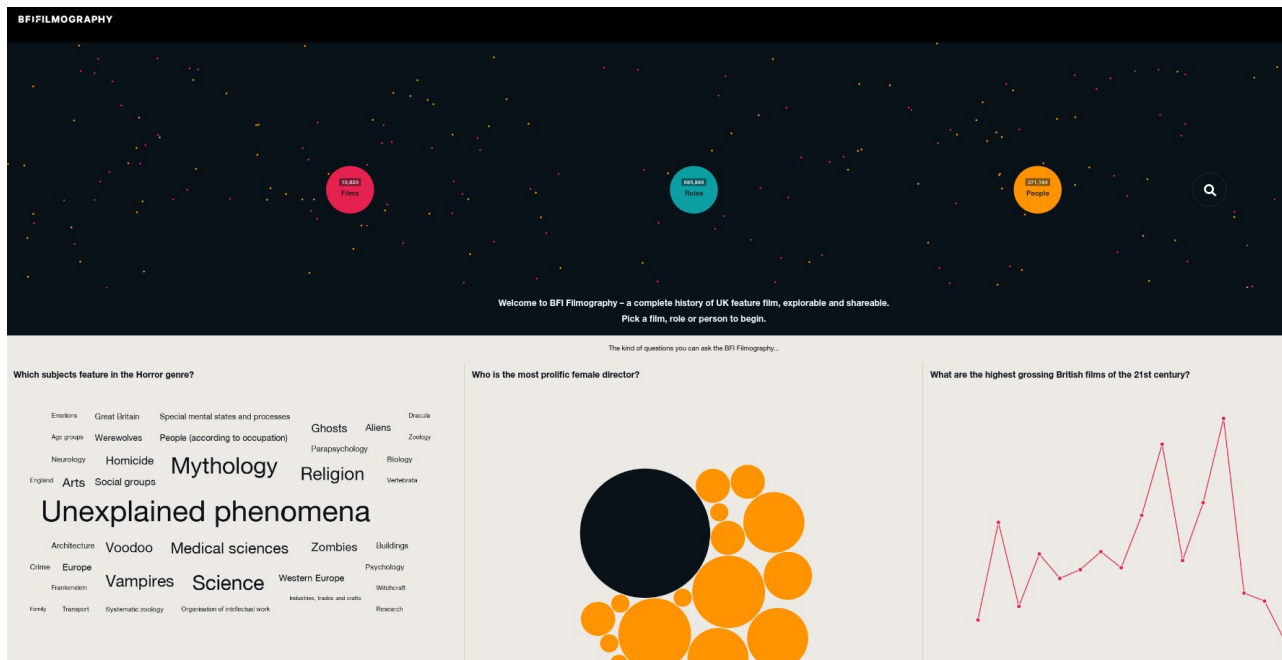

Figure 5: Screenshot of the *BFI Filmography* website that shows how different graphs are applied to different aspects in film history, slightly cropped by the author. *BFI Filmography,* 2017, https://filmography.bfi.org.uk/films. Accessed August 2020, cropped by the author.

These data visualization projects are just a few among many in the field of film and media historiography that have been created in the past two decades. This brief insight shows the variety of approaches. They all apply different data to different tools for various purposes and in various contexts. And they all look differently. How to classify visualization projects epistemologically concerns many digital humanities scholars at present. I also think that this issue needs to be further explored. So far, the analyses have mainly focused on the intended functions and pragmatic use of visualizations rather than their effects and possibilities.

It has been demonstrated that, generally speaking, each type of visualization serves specific functions. For instance, bar charts are appropriate for comparing values, pie charts show the percentages of values and network diagrams point out connections (Drucker 2016, 239–241). And, as shown in the case of the COVID-19 dashboard, symbol maps seem to be unsuitable for representing proportions. But, as Drucker reasonably argues, we need to further investigate the "intellectual implications of the use of graphical arguments built on tools borrowed from other disciplines" (Drucker 2016, 238). There is much more work to be done in order to better understand the epistemological conditions and effects of visualizations: a) in the humanities and b) from a humanities perspective (ibid.).

As indicated above, different from what one would expect, diagrams, maps and other forms of visualization can hardly serve as clear evidence that does not require any explanation or

interpretation. Due to representational conventions and epistemological premises visualizations appear ordered, comprehensive, and structured, when in fact they often obscure ambiguities, conflicts, and contradictions (Gitelman und Jackson 2013, 9). Therefore, if we consider visualizations themselves artifacts (Drucker 2016, 239) we have to closely examine each single case to grasp how an argument is made, what kind of knowledge is produced, and what underlying political structures are at play. As mentioned above, we ought to take into account that visualizations do not only represent information but at the same time also produce meaning. Or, in the words of Drucker: "The means by which a graphic produces meaning is an integral part of the meaning it produces" (ibid.). The challenge for media scholars is to not only distinguish various graphs but also grasp *how* the various visualizations are *creating* meaning. They do not just reveal or show something but they also "act" (Parry 2019). This argument needs be kept in mind, when we, as claimed by Glinka and Dörk, further educate ourselves in digital visualization literacy as a new facet of critical inquiry (Glinka and Dörk 2018).

Rethinking data visualizations

While it is necessary to thoroughly analyze how data visualizations re/produce – or perform and enact – specific values, ideologies, and politics, I would like to shift the focus to their critical potential for film historiography. Following current discourses in data feminism and other critical approaches in digital humanities, I, too, contend that data visualizations do not always obscure conflicts and contradictions but can, in contrast, help us reflect upon the situatedness of knowledge and epistemological uncertainties such as vague or ambivalent data or assumptions and probabilities. As a number of scholars have argued, in order to fully explore the potentials of digital knowledge production and representation we have to rethink our underlying premises of what data visualizations ought to accomplish. We need to look beyond the "lie factor" and recognize that they do not necessarily have to aim for the clearest and most comprehensive accurate picture. Graphics and other visual arrangements can be also unsettling and perhaps, in doing so, make us reconsider what is perceived as common knowledge and legitimate scholarly work. The following example shall outline how such a visualization can look like.
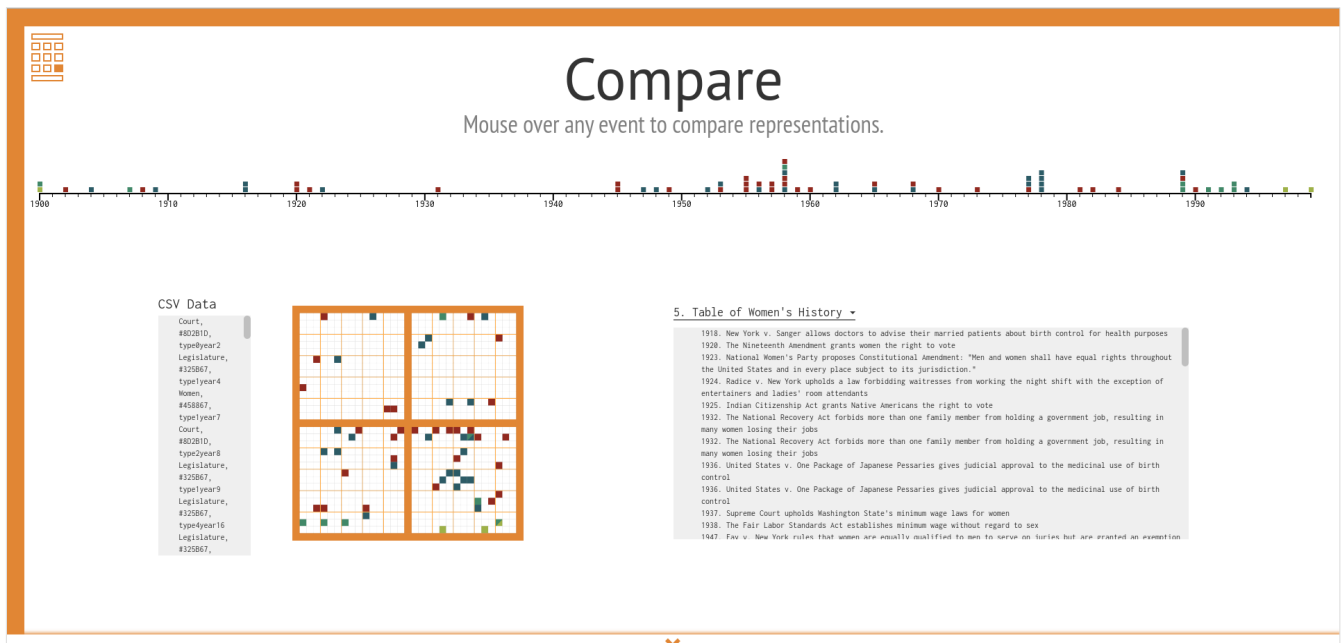
Figure 6: Screenshot of the *Shape of History Project* by Lauren Klein*, a multi-perspective, interactive data visualization project, slightly cropped by the author. "Compare. Mouse over any event to compare representations", *The Shape of History*, 2016, http://shapeofhistory.net/#compare. Accessed November 2020, cropped by the author.

In *Shaping History. Reimagining Elizabeth Palmer Peabody's Historical Visualization Work*, feminist data scholar Lauren Klein and her team introduce us to the grid as an alternative approach to history [Fig. 6]. In the introduction Klein wonders:

> *What would it mean if a visualization was designed to be difficult and abstract? If it was intended to send us back to the original source of the data in order to make sense of the image we encountered? What if the goal of visualization was to allow each person, individually, to interpret the image for herself? (Klein et al. 2016)*

The grid was designed by the nineteenth-century educator, writer, and publisher Elizabeth Palmer Peabody who believed in the active engagement of people. Instead of organizing historical data through a chronological timeline, which would be the standard choice for displaying past events, Palmer Peabody created a colorful grid with an interactive interface for users to implement data of their own and thereby create their personal report to history. The goal of this pedagogical approach was, as Klein explains, to reflect on the remediation process of data visualizations. It makes us think about the status of data and the importance of design that shapes history (ibid.). It encourages us to engage with historiography in an affective, playful, and self-reflective manner.

This is just one of many digital humanities projects that critically investigates knowledge production and that I think is inspiring for further exploring the potentials of data visualizations. Working in a larger research context with data feminism, Klein and data scholar Catherine D'Ignazio make the case that data visualization projects can productively draw on feminist theory (D'Ignazio and Klein 2016). They state:

When exploring the intersection of data visualization and the digital humanities, one must consider not only how the domain of digital humanities – and of the humanities more generally – can provide opportunities for the design and application of visualization tools and techniques, but also how theories from the humanities can themselves inform visualization design. (ibid., 1)

Based on four fields of critical inquiry (feminist science and technology studies, feminist human-computer interaction (HCI), feminist digital humanities, and cartography & geographic information system (GIS)), D'Ignazio and Klein outline six principles for feminist data visualization. I have slightly rephrased their claim to underline certain aspects I assume to be particularly relevant. They ask us to: 1) rethink binaries and categorizations, 2) in lieu of an universal objectivity embrace pluralism and allow for "multiple truths", 3) scrutinize power structures in the entire design process, also with regard to the production team and the users, 4) consider diverse contexts of knowledge production including data provenance and processing, 5) recognize aesthetic experience of data visualizations, and 6) credit the entire team's labor (ibid.).

As intended by D'Ignazio and Klein, these principles offer a fruitful starting point for further scrutinizing the complex framework of data visualizations. However, while I agree with their claim that a theoretically well informed approach to data visualization is much needed in order to understand and intervene in current methodological developments I want to emphasize that, as noted by various media scholars and mentioned earlier, vice versa, one must also consider how data visualization can enhance humanities approaches. As implicitly reflected by D'Ignazio and Klein in their paper cited above and further elaborated in their book *Data Feminism* (D'Ignazio and Klein 2020), I wish to stress that data visualization can also help us engage with feminist matters if we apply them as a "humanistic method" (Drucker 2014, 130–135). Thus, referring to what has been said earlier, in order to better understand digital forms of knowledge production, we need to further investigate data visualizations a) *in* the humanities and b) *from* a humanities perspective (Drucker 2016, 238), and, for the purpose of clarification I would like to add, echoing Drucker, c) *as* a humanities approach.

A humanistic method, or, humanities approach takes into account the constructed, subjective, and situated nature of scholarly knowledge. It shows that "phenomena and their observers are co-dependent" and that consequently data as well as data visualizations are always already a (performative) interpretation that is determined by particular historical, social, and political configurations (Drucker 2014, 130–135). Contrary to "realist approaches" which strive for transparency and equivalence as if the world to be presented was pre-existent (ibid.), humanistic data projects should re/present contingencies, partial views, and plural perspectives. In doing so, we need to distinguish between "the task of representing ambiguity and uncertainty" and "that of using ambiguity and uncertainty as the basis on which a representation is constructed" (Drucker 2014, 126–127).

Where do we go now

While digital film historiography is an emerging field where more and more scholars advance research by developing and applying new tools and methods, the many opportunities digital technologies provide are yet to be much more explored. In terms of critical inquiry, digital humanities have still a long way to go, or rather, should much more intervene in current developments and discourses (Posner 2016). In this respect, I hope to have shown why data visualizations play a particular important role that we need to further analyze both theoretically and application-oriented. Bearing in mind the foregoing, I conclude that the "means by which a graphic produces meaning" (Drucker 2016, 239) can best be understood in-depth by experimenting with data visualization itself – in addition to theoretical case studies – that means *as* a humanities approach. For instance, we should explore how to develop projects that "show us categories like race *as they have been experienced*, not as they have been captured and advanced by businesses and governments?" (Posner 2016) As Posner suggests, "a useful data model for race would have to be time- and place-dependent so that a person moved from Brazil to the United States, she might move from white to black." (ibid.)

In lieu of presenting information as if a priori reality exists that can be easily measured and grasped, as one might initially associate with data visualizations, we can take advantage of data visualizations to challenge absolute values, universalization and essentialization by foregrounding the particularity of knowledge. As Drucker reminds us, in the digital humanities we must not suddenly treat space and time as given as if philosophical discourses have never existed (Drucker 2014, 242). Therefore, we also have to scrutinize concepts of space and time as they are re/presented by powerful applications like Google Maps as well as by alternatives like OpenStreetMap (Posner 2016). How to model and show historical data that is vague and uncertain like "'for 6 months before the war,' 'around 1832', or 'during harvest season in her youth'" is still a desiderata, geographic researcher Karl Grossner and data visualization practitioner Elijah Meeks write (ibid., Grossner and Meeks 2014).

Another challenge lies in overcoming the "search-slot paradigm", a single query box that requires prior knowledge of a field, as Glinka, Dörk, and geovisualization scholar Sebastian Meier state (Glinka, Meier, and Dörk 2015, 109–110). Exploring how to visualize the "'Un-seen'" in cultural heritage collections they suggest to create a more flexible and open access to (data) collections by allowing users to search by a variety of interrelated metadata visualizations (ibid.). The BFI Filmography provides a good example for this approach. By facilitating different modes of access to the database, for example via a map that shows the distribution of regions, a histogram that focuses on dates and time ranges, and a tag cloud that illustrates the significance of a topic, experts and non-experts alike can benefit from the BFI platform. Thus, what might look like conventional statistics at first sight has much more to offer.

In addition to various access possibilities, interfaces could allow annotations and comments, or even the co-creation of data infrastructures in order to make users actively engage and perhaps help collect missing data (ibid., 111). Needless to say, this requires an inviting interface design so that the users will actually exploit such interactive opportunities. If we

consider power structures in the entire design process, it is crucial to assure, as Glinka, Dörk, and Meier emphasize, that also in collaborative environments diverse perspectives are included (ibid., 112). Diversity comprises recipients, producers, and approaches – in research and cultural heritage institutions alike. To change the point of view and look at collections beyond traditional logics such as metadata, similar to Manovich's Cultural Analytics approach they, too, suggest to use computer vision for relating objects, among others, by color, structure, or shape (ibid., 113). Data visualizations allow us to defamiliarize our research objects in order to recognize unexpected aspects and challenge traditional concepts (Olesen 2017, 50–52). If we, for example, conceive film not only as a moving image but as a color-intensive impression we might be able to ask new questions and develop new approaches.

The overarching argument here relies on carefully reconsidering the goal and thus the conceptualization and creation of data visualizations in order to present film historical research in a critical and self-reflecting manner. In this sense, it is essential to understand what visualizations do but also what they *could* do (Dörk et al. 2013). In terms of highlighting contingency, subjectivity, and serendipity, I see special promise in data visualization approaches to feminist film historiography.

By identifying artifacts as data and data as artifacts, data visualizations can extend film and media studies corpora as well as our repertoire of theories and methods. As pointed out by numerous scholars, data-intensive approaches do not replace established methods. Despite a change in perspective, when considering data visualization as a film historiographical approach for displaying research, essential humanities concerns – the critical investigation of knowledge production, the hermeneutic analysis of artifacts, or the contingency of history – still matter (Olesen 2017, 50–52). In the course of 'datafying' film historiography, fundamental historiographical methods such as finding, collecting, cataloging and interpreting remain relevant, but they change and new ones emerge. If we consider data visualizations in their versatility I argue that they offer a productive point of departure for actively intervene in current transitions in digital film historiography. I think that feminist film historiography can greatly benefit from digital data visualization and, vice versa, data visualization from feminist film historiography respectively, feminist theory. Considering the growing production and application of data in the era of digitalization as well as the debates on digital history, research data, and open access, it becomes obvious that new representation strategies are much needed to keep the memory of the countless women in early film industries alive and prevent them from getting lost again in the course of history. In my view, data visualizations can provide a critical response to the current challenges of digital film historiography and help us tell different stories differently.

References

Barrowman, Nick. 2018. "Why Data Is Never Raw." *The New Atlantis*, vol. 58, pp. 129–35. https://www.thenewatlantis.com/publications/why-data-is-never-raw.

Bean, Jennifer M., and Diane Negra, editors. *A Feminist Reader in Early Cinema*. Camera Obscura Book. Durham: Duke University Press, 2002.

Benoni, Dan, and Louis-Xavier Lavalle. "Coronavirus UX: How Dashboard Designs Can Impact Your Perception." *Growth.Design*, 2020. https://growth.design/case-studies/coronavirus-dashboard-ux/. Accessed August 2020.

Bishop, Catherine. "The Serendipity of Connectivity: Piecing Together Women's Lives in the Digital Archive." *Women's History Review*, vol. 26, no. 5, 2017, pp. 766–80. https://doi.org/10.1080/09612025.2016.1166883.

Callahan, Vicki, editor. *Reclaiming the Archive: Feminism and Film History*. Contemporary Approaches to Film and Television Series. Detroit: Wayne State University Press, 2010.

Dall'Asta, Monica, and Jane M. Gaines. "Prologue. Constellations. Past Meets Present in Feminist Film History." In *Doing Women's Film History: Reframing Cinemas, Past and Future*. Women and Film History International, edited by Christine Gledhill and Julia Knight. Urbana; Chicago; Springfield: University of Illinois Press, 2015, pp. 13–25.

Dang, Sarah-Mai. "Digital Tools & Big Data: Zu gegenwärtigen Herausforderungen für die Film- und Medienwissenschaft am Beispiel der feministischen Filmgeschichtsschreibung." *MEDIENwissenschaft: Rezensionen | Reviews*, no. 2–3, 2018, pp. 142–156. https://doi.org/10.17192/ep2018.2-3.7836.

———"Forschungsdatenmanagement in der Filmwissenschaft: Daten, Praktiken und Erkenntnisprozesse." *Montage/Av*, special issue *Praktiken des Digitalen*, vol. 29, no. 1, 2020a, pp. 120–40.

———"Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly*, special issue *Digital Humanities & Film Studies: Analyzing the Modalities of Moving Images*, vol. 14, no. 4, December 2020b, n.p.

D'Ignazio, Catherine, and Lauren F Klein. "Feminist Data Visualization." IEEE VIS Conference, 22-28 October 2016, Baltimore, Maryland, USA, n.p. https://www.semanticscholar.org/paper/Feminist-Data-Visualization-D%27Ignazio-Klein/2e3e2eb1bdc1cab5b0fab515266bb8849d416f33.

———*Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020.

Dörk, Marian, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. "Critical InfoVis: Exploring the Politics of Visualization." In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, 2189. Paris, France: ACM Press, 2013, pp. 2189–98. https://doi.org/10.1145/2468356.2468739.

Drucker, Johanna. *Graphesis: Visual Forms of Knowledge Production*. MetaLABprojects. Cambridge, Massachusetts: Harvard University Press, 2014.

———"Graphical Approaches to the Digital Humanities." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Raymond George Siemens, and John Unsworth. Chichester, West Sussex, UK: Wiley/Blackwell, 2016, pp. 238–50.

Ernst, Christoph. "Die Kurve Abflachen! – Über Informationsvisualisierung und die Corona-Pandemie – Universität Bonn." April 2, 2020. https://www.uni-bonn.de/neues/die-kurve-abflachen-2013-ueber-informationsvisualisierung-und-die-corona-pandemie. Accessed August 2020.

Fillmore-Handlon, Charlotte. "5 Things Data Cannot Do." July 8, 2016. http://projectarclight.org/arguments/5-things-data-cannot-do/. Accessed August 2020.

Gaines, Jane. *Pink-Slipped: What Happened to the Women in the Silent Film Industries?* Women and Film History International. Urbana: University of Illinois Press, 2018.

Gitelman, Lisa, and Virginia Jackson. "Introduction." In *"Raw Data" is an Oxymoron*, edited by Lisa Gitelman. Infrastructures Series. Cambridge, Massachusetts; London, England: The MIT Press, 2013, pp. 1–14.

Gledhill, Christine, and Julia Knight, editors. *Doing Women's Film History. Reframing Cinemas, Past and Future*. Women and Film History International, edited by Christine Gledhill and Julia Knight. Urbana; Chicago: University of Illinois Press, 2015.

Glinka, Katrin, and Marian Dörk. "Zwischen Repräsentation und Rezeption – Visualisierung als Facette von Analyse und Argumentation in der Kunstgeschichte." In *Computing Art Reader: Einführung in die digitale Kunstgeschichte*, edited by Piotr Kuroczyński, Peter Bell, and Lisa Dieckmann. arthistoricum.net, Universitätsbibliothek Heidelberg, 2018, pp. 235–50. https://doi.org/10.11588/arthistoricum.413.

Glinka, Katrin, Sebastian Meier, and Marian Dörk. "Visualising the »Un-seen«: Towards Critical Approaches and Strategies of Inclusion in Digital Cultural Heritage Interfaces." In *Kultur und Informatik. Cross Media*, edited by Carsten Busch and Jürgen Sieck, vol. XIII, Glückstadt: Werner Hülsbusch, 2015, pp. 105–18.

Grossner, Karl, and Elijah Meeks. "Topotime: Representing Historical Temporality." *Proceedings of DH2014 Conference,* 7–12 July 2014, Lausanne, Switzerland. Alliance of Digital Humanities Organizations, 2014, n.p. http://kgeographer.org/pubs/Grossner-Meeks_Topotime.pdf.

Haraway, Donna. "Situated Knowledges. The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist studies*, vol. 14, no. 3, 575–99.

Hatch, Kristen. "Cutting Women: Margaret Booth and Hollywood's Pioneering Female Film Editors." In *Women Film Pioneers Project*, edited by Jane Gaines, Radha Vatsal, and Monica Dall'Asta. New York NY: Columbia University Libraries, 2013a, n.p. https://doi.org/10.7916/d8-t0y9-hv61.

——— "Rose Smith." In *Women Film Pioneers Project*, edited by Jane Gaines, Radha Vatsal, and Monica Dall'Asta. New York NY: Columbia University Libraries, 2013b, n.p. https://doi.org/10.7916/d8-vqh1-0a87.

Hill, Erin. *Never Done: A History of Women's Work in Media Production*. New Brunswick, New Jersey: Rutgers University Press, 2016.

Klein, Lauren F., Caroline Foster, Erica Pramer, Adam Hayward, and Shivani Negi. "The Shape of History." 2016. http://shapeofhistory.net/. Accessed August 2020.

Lammar, Dominic. 2020. "COVID-19 - So veranschaulichen Datenjournalist:innen das Coronavirus." *netzpolitik.org*. March 11, 2020. https://netzpolitik.org/2020/so-veranschaulichen-datenjournalistinnen-das-coronavirus/. Accessed August 2020.

Manovich, Lev. "Museum Without Walls, Art History Without Names." In *The Oxford Handbook of Sound and Image in Digital Media*, edited by Carol Vernallis, Amy Herzog, and John Richardson. Oxford University Press, 2013, pp. 253–78. https://doi.org/10.1093/oxfordhb/9780199757640.001.0001.

Olesen, Christian Gosvig. "SEMIA and Moving Image Data Visualization: An Overview and Brief Introduction | The Sensory Moving Image Archive." May 20, 2018. https://sensorymovingimagearchive.humanities.uva.nl/index.php/2018/05/20/semia-and-moving-image-dataviz-in-film-and-media-studies-an-overview-and-brief-introduction/. Accessed August 2020.

Olesen, Christian Gosvig. 2017. "Towards a 'Humanistic Cinemetrics'?" In *The Datafied Society. Studying Culture through Data*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017, pp. 39–54. http://oapen.org/search?identifier=624771.

Masson, Eef. "Humanistic Data Research. An Encounter between Epistemic Traditions." In *The Datafied Society. Studying Culture through Data*, edited by Mirko Tobias Schäfer and Karin van Es. Amsterdam University Press, 2017, pp. 25–37. http://oapen.org/search?identifier=624771.

Parry, Kyle. "Reading for Enactment: A Performative Approach to Digital Scholarship and Data Visualization." In *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold and Lauren F. Klein. University of Minnesota Press, 2019, pp. 141–60. https://doi.org/10.5749/j.ctvg251hk.16.

Posner, Miriam. "What's Next: The Radical, Unrealized Potential of Digital Humanities." In *Debates in the Digital Humanities*, edited by Mathew K. Gold and Lauren F. Klein. University of Minnesota Press, 2016, n.p. https://dhdebates.gc.cuny.edu/read/untitled/section/a22aca14-0eb0-4cc6-a622-6fee9428a357.

Rawson, Katie, and Trevor Muñoz. "Against Cleaning." In *Debates in the Digital Humanities*, edited by Mathew K. Gold and Lauren F. Klein. University of Minnesota Press, 2019, n.p. https://dhdebates.gc.cuny.edu/read/untitled-f2acf72c-a469-49d8-be35-67f9ac1e3a60/section/07154de9-4903-428e-9c61-7a92a6f22e51.

Spongberg, Mary, Gina Luria Walker, and Koren Whipp. "Female Biography and the Digital Turn." *Women's History Review* vol. 26, no. 5, 2017, pp. 705–20. https://doi.org/10.1080/09612025.2016.1167342.

Stevens, Harry. "These Simulations Show How to Flatten the Coronavirus Growth Curve." *The Washington Post*. March 14, 2020. https://www.washingtonpost.com/graphics/2020/world/corona-simulator/. Accessed August 2020.

Tufte, Edward R. *The Visual Display of Quantitative Information*. Cheshire, Conn: Graphics Press, 2001.

Wiles, Siouxsie. "The Three Phases of Covid-19 – and How We Can Make It Manageable." *The Spinoff*. March 9, 2020. https://thespinoff.co.nz/society/09-03-2020/the-three-phases-of-covid-19-and-how-we-can-make-it-manageable/. Accessed August 2020.

Withers, Deborah M. *Feminism, Digital Culture and the Politics of Transmission: Theory, Practice and Cultural Heritage*. London; New York: Rowman & Littlefield International, 2015.

Wreyford, Natalie, and Shelley Cobb. "Data and Responsibility". *Feminist Media Histories*, vol. 3, no. 3, 2017, pp. 107–132. doi:10.1525/fmh.2017.3.3.107

Biography

Sarah-Mai Dang is Principal Investigator of the BMBF research group "Aesthetics of Access. Visualizing Research Data on Women in Film History" (DAVIF) (2021-2025) at the Institute of Media Studies at Philipps-Universität Marburg. Based on her current research and teaching focus on digital film historiography, data visualization, feminist theory, open science, and media aesthetics, she also founded the international DFG research network "New Directions in Film Historiography. Digital Tools and Methods in Film and Media Studies" (2019-2022). In addition, she initiated the collaborative Open Media Studies blog and the hybrid publishing project oa books, where she published her dissertation and blogs about academic publishing. She received a doctoral degree in Film Studies from Freie Universität Berlin and a Master of Arts from University of Michigan and has published widely on the effects of digitization in the humanities. (https://orcid.org/0000-0002-1960-247X, @SarahMaiDang, @SarahMaiDang@fedihum.org).