Nineteenth-Century Art Worldwide

a journal of nineteenth-century visual culture

Kaylee P. Alexander

book review of

Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art by Diana Seave Greenwald

Nineteenth-Century Art Worldwide 20, no. 3 (Autumn 2021)

Citation: Kaylee P. Alexander, book review of *Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art* by Diana Seave Greenwald, *Nineteenth-Century Art Worldwide* 20, no. 3 (Autumn 2021), <https://doi.org/10.29411/ncaw.2021.20.3.8>.

Published by: [Association of Historians of Nineteenth-Century Art](#)

Notes:

This PDF is provided for reference purposes only and may not contain all the functionality or features of the original, online publication.

License:

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#) [Creative Commons License](#).

Diana Seave Greenwald,
Painting by Numbers: Data-Driven Histories of Nineteenth-Century Art.
Princeton, NJ: Princeton University Press, 2021.
256 pp.; 55 color illus.; 9 b&w illus.; notes, bibliography; index; 14 tables; appendices.
\$35.00 (hardcover)
ISBN: 978-0-691192-45-1

The culmination of nearly a decade of data work, Diana Seave Greenwald's *Painting by Numbers* brings together various sources of aggregate information on the production and exhibition of paintings in nineteenth-century France, Britain, and the United States. Greenwald, who holds a DPhil in History and an MPhil in Economic and Social History (both from the University of Oxford), utilizes methodologies more common in the social sciences in an attempt to circumvent biases in the history of art that tend to focus attention on paintings that have survived into the present rather than those that have been lost or are poorly documented. Specifically, Greenwald applies descriptive statistics and econometrics (the application of statistical analyses to economic data) to aggregates of data culled from historical indices and catalogues of paintings, such as Jon Whiteley's *Subject Index to Paintings Exhibited at the Paris Salon, 1673–1881*.^[1] The benefits of this approach are undeniable, as using aggregates of information in this way often helps us to identify patterns and trends otherwise undetectable through the examination of single works or a selection of them, or to identify expected and unexpected outliers that might help us refine our understanding of a given time and place. These benefits have been demonstrated in countless art markets studies dating from the 1980s to the present, although these works have tended to focus on the early modern period rather than the nineteenth century.^[2] Where Greenwald claims to depart from past research, however, is in her emphasis on subject matter and artists' gender over market factors such as price.

Though claiming to provide a novel methodology with which to approach the history of art, Greenwald's study is anything but groundbreaking. Indebted to the work of Harrison and Cynthia White, John Michael Montias, and other pioneers in the field of cultural economics, *Painting by Numbers* promises to be a guide for what Greenwald sees as an emerging field of data-driven humanities research.^[3] Part of the problem with this book, however, is the author's attempt to cover too many disparate, and very specific, subjects—from rural landscape painting in France to representations of empire in British painting—without a set of precise and overarching research questions. In fact, the only question that links

Greenwald's case studies together is the rather general one of what data analysis can bring to histories of nineteenth-century painting.

Greenwald's book is organized into two main parts: the opening two chapters constitute an introduction both to the book and its methodological roots (chapter 1: "What Is a Data-Driven History of Art?"), and to the specific datasets used in this study (chapter 2: "The Historical Data of the Art World"). This second chapter provides a detailed account of Greenwald's primary data sources—her own Historical American Art Exhibition Database, Jon Whiteley's *Subject Index to Paintings Exhibited at the Paris Salon*, and the catalogs of exhibitions that took place at the Royal Academy in London—and summarizes the information they provide. Together these sources record hundreds of thousands of paintings, but, as Greenwald importantly notes: "just because a dataset is large does not necessarily mean it is perfectly representative" (25). In focusing on the rationale of the present study and the historiography of comparable digital humanities work, Greenwald convincingly answers her overarching question and makes clear that the highest potential benefit of data for art historians comes from the combination of macroscopic and microscopic modes of analysis, drawn from the social sciences and humanities, respectively. Her attempt at implementing these methods is what follows.

The next three chapters are case studies that explore subsets of Greenwald's three primary datasets within their relevant chronological and geographic contexts. Here, Greenwald applies her knowledge of and experience with econometrics to the study of nineteenth-century painting to explore themes within French landscape painting, roughly from 1831 to 1881 (chapter 3: "Between City and Country: Industrialization and Images of Nature at the Paris Salon"); the labor economics of considering women painters in American art exhibitions (chapter 4: "Why Have There Been No Great Women Artists?: Artistic Labor and Time-Constraint in Nineteenth-Century America"); and representations of empire in paintings exhibited at the Royal Academy in London (chapter 5: "Implied But Not Shown: Empire at the Royal Academy").

"Between City and Country" specifically takes on the question of rural subject matter and scenes of peasant life in French landscape painting in the period following the establishment of the July Monarchy. Ultimately, despite the pages of statistical analysis, this chapter serves to reinforce established and accepted narratives about nineteenth-century French art, echoing the contributions of, for example, Albert Boime, Steven Adams, Richard R. and Caroline B. Bretell, and Patricia Mainardi.^[4] Although one of the benefits of taking a data-driven approach can be the confirmation of past assumptions through aggregate evidence, in this instance the sample size is so small—less than 1,600 landscape paintings out of a total 30,419 works present in the whole dataset (61)—and not placed within the context of *all* paintings shown at the Paris Salon within Greenwald's chosen chronology (1831 to 1881) that her potential contribution falls short. Additionally, the fact that we tend to see more landscape paintings exhibited at the Salon in the years following the introduction of the Prix de Rome for *paysage historique* (first awarded in 1816, and eliminated as a category after 1863) makes the choice to analyze landscape painting during this fifty-year period somewhat contrived.^[5] Rather than selecting landscape painting as the object of study from the outset, more interesting or concrete results might have emerged had Greenwald examined all exhibited works to question current scholarly assumptions about the role of landscape

painting in relation to other genres and subject categories included in Whiteley's *Index*. This alternate approach would have allowed the data to be used to help formulate or refine guiding research questions. Here we find that the questions we choose to ask can also represent a form of selection bias. Turning the process around to let the data speak first can potentially reduce partiality in humanistic research.

Although Greenwald's findings are unsurprising, key aspects of this book are valuable. She rightfully calls attention to art history's hesitancy to accept data-driven methodologies and makes a strong case for the methodological benefits of thinking in quantitative terms that straddle the delicate line between anonymous observations and datapoints as individual objects or persons with unique stories to tell.

In recent years, two camps have emerged with respect to how data-driven humanities are perceived: on the one side are those who have become increasingly tenacious in their rejection of data-driven art history and see such methods as a dehumanizing or anonymizing approach that precludes the in-depth study afforded by more traditional approaches in the humanities.^[6] This first camp has been most usefully described and discounted by Victoria Szabo in her article on database thinking in which she makes a strong and well-supported case for a judicious balance between aggregate study and carefully selected case studies.^[7] For Szabo, it is not a matter of qualitative vs. quantitative, but rather of cohesive analysis that does not need to rely on what have traditionally been viewed as competing, or mutually exclusive, methods. Unfortunately, the second camp often sees any humanities project that ventures into the world of quantitative analysis as, by definition, cutting-edge, whether or not that study in fact brings something new to existing scholarship.

What stands between the "shiny new toys" and the successful data-driven study is a critical understanding of what the data are showing, how they were gathered, and how they contribute to historical narratives; it is not sufficient to take raw data at face value. As Jessica M. Johnson has noted, for example, the sources of historical data are as biased as the individuals and institutions who originally recorded and preserved the information that comprise them. Johnson has made an important and compelling case for the mindful exploitation of data in her discussion of the ways in which scholars' uses of the Trans-Atlantic Slave Voyages Database have often perpetuated and restated the vantage point of white oppressors.^[8] A similar critique might be made of *Painting by Numbers*, because Greenwald attempts to tackle too many disparate topics in too few pages, and this prevents her from grappling robustly with the biases of her sources—the choice of terminologies, the contexts in which they were compiled, and their survival over other forms of documentation—and makes only a limited contribution to existing scholarship beyond her general call for more researchers to consider this methodology.^[9]

A more useful approach would have been to let the data speak, or, in other words, to develop focused research questions from the existing data. Beginning with the question of how data on nineteenth-century painting can help to revise preexisting histories of art is a necessary starting point. From there, with the choice to focus on three highly specific, yet unrelated, topics Greenwald limits the attention she is able to devote to each of these big ideas. Here, rather than confronting biases in the data head-on, she runs the risk of confirmation bias, or, the tendency to seek out information that supports preexisting assumptions while

(intentionally or not) ignoring data that may conflict. This is particularly the case in the chapter on women artists (chapter 4) and images of imperialism (chapter 5), topics that have traditionally been subjected to poor or heavily biased record-keeping. Chapter 4 especially runs the risk of confirming traditional narratives about “women’s” subjects (namely motherhood) prevailing in art produced by women, while ignoring the role of other subject matter (or the biased categorization of paintings produced by artists understood as female).

Still, Greenwald’s nod to Linda Nochlin in chapter 4, “Why Have There Been No Great Women Artists,” is without doubt the strongest section in this book. Here, the potential of data-driven histories of art is demonstrated clearly and usefully. Achieving a more nuanced balance between data and case study, she demonstrates how art historians can effectively use quantitative methods to highlight patterns and trends, coupled with relevant theoretical frames (here, those borrowed from labor economics) that demonstrate concrete, aggregate evidence for what have previously been assumptions based on a limited selection of artists’ work. Standing in sharp contrast to chapter 3, which relies on relatively little data, cursory associations between artistic production and social/political developments in nineteenth-century France (such as labor strikes), and a rather redundant case study of Jean-François Millet, chapter 4 makes clear the need for further data-driven studies in art history.

Greenwald is undoubtedly successful in making clear the limitations of her data sources, and in demonstrating that null results are to be embraced in the humanities just as they are in the natural and social sciences. In this regard, her scientific method is refreshingly sound. Yet, there are a number of methodological problems that are left conspicuously unaddressed despite the repercussions they have on her subject matter. For example, in her discussion of the program used to assign gender to the data on exhibitors at the National Academy of Design, it is surprising to see no allusion to the implications of relying on existing knowledge to assign gender manually in cases where the program reads a woman’s name as male, such as in the case of Claude Hirst (95). Although Greenwald makes clear that many women could have been missed here (and are likely to remain unidentified due to various biases in the historical record), her text does leave a critical loose end. What is more important: the ways in which names would have been perceived or the actualities of the artist’s biological sex? I would argue that they are of equal significance, and Greenwald could usefully explore the relationships between perceived gender and known gender by comparing the raw data assigned by the computer with the data she was able to clean to reflect historical knowledge of these artists.

The data preparation for this book demanded a herculean effort, though this represents the labor of the many hands that came before or were hired by the author rather than of Greenwald herself—a detail, buried in the book’s appendices, that has serious implications for understanding the quality of the author’s data. Although using third parties to process data is common in the social and natural sciences, historical data sources possess a number of unique challenges—damaged or unclear sources, variations in spelling, historical print or handwriting—that need to be considered carefully from the start. Here, I would like to stress that how and by whom data are prepared is not something to be taken lightly. In all data-driven studies, gathering and cleaning should be as methodical as any statistical analysis performed on those data. This is true in the humanities perhaps more than in any other field

as we are continually presented with imperfect and incomplete data. Transcription services, typically unaccustomed to working with historical documentation, can thus be problematic.

While Optical Character Recognition (OCR) software is advancing rapidly in its accuracy and work is in development for reading historical handwriting and print with the help of AI, current technologies are rarely as accurate and precise as we would like them to be. What is lost or misinterpreted when the computer is granted such responsibility? These sorts of errors are minimal when dealing with large aggregates of data, though even thousands of paintings do not constitute “big data,” and thus such errors do in fact alter outcomes. There is something to be said, too, for gaining an intimate knowledge of the document itself: its imperfections, idiosyncrasies, and biases. Though they constitute limitations for analyses, these details allow us to critique how knowledge has been transferred and preserved over time.

What is even more important than the gathering and cleaning processes, however, is understanding precisely what and who these data represent, and what and who they do not. Absences of data may also have interesting, contradictory, and surprising stories to tell. Although Greenwald begins to tackle the question of such absences in chapter 5 by addressing how descriptions of subject matter directly related to British imperialism tended to be more implied than explicitly evident in exhibition catalogues, more work on these gaps undoubtedly needs to be undertaken. Considerations of how gender was inaccurately perceived in exhibitions is just one example. These types of thorough discussions of omissions in the material record are regrettably missing from Greenwald’s exploration of painting beyond the canon, yet still within the scope of official exhibitions.

Painting by Numbers is an important text that should inspire future data-driven studies in art history, but data-curious art historians will need to consider it both as an example of the benefits of using data-driven approaches and as a study that displays the need to continue to think critically about how such methodologies are used.

Kaylee P. Alexander
Lecturer
Guilford College
kalexander[at]guilford.edu

Notes

[1] Jon Whiteley, *Subject Index to Paintings Exhibited at the Paris Salon, 1673–1881* (unpublished, 1993, deposited at the Sackler Library, University of Oxford).

[2] See, for example, Neil de Marchi and Hans J. van Miegroet, eds., *Mapping Markets for Paintings in Europe, 1450–1750* (Turnhout: Brepols, 2006); Hans van Miegroet and Neil De Marchi, “The Antwerp-Mechelen Production and Export Complex,” in *Album Amicorum J. Michael Montias*, ed. Mia Misozuki (Amsterdam: University of Amsterdam Press, 2007), 133–47; Marten Jan Bok and Tom van der Molen, “Productivity Levels of Rembrandt and His Main Competitors in the Amsterdam Art Market,” *Jahrbuch Der Berliner Museen* 51 (2009): 61–68, <http://www.jstor.org/stable/25674337> [login required]; Sandra van Ginhoven, *Connecting Art*

Markets: Guiliam Forchondt's Dealership in Antwerp (c. 1632–78) and the Overseas Paintings Trade (Leiden: Brill, 2017); Felipe Álvarez de Toledo López-Herrera. "Beyond Murillo: New Data-Driven Research on the Painting Market in Early Modern Seville," *Journal of Art Market Studies* 3, no. 2 (2019), <https://doi.org/10.23690/jams.v3i2.94>; Anne-Sophie V.E. Radermecker, "Artworks without Names: an Insight into the Market for Anonymous Paintings," *Journal of Cultural Economics* 43 (2019): 443–83, <https://doi.org/10.1007/s10824-019-09344-5>.

[3] See, for example: Harrison C. White and Cynthia A. White, *Canvases and Careers: Institutional Change in the French Painting World* (Chicago: University of Chicago Press, 1965); John Michael Montias, *Artists and Artisans in Delft: A Socio-Economic Study of the Seventeenth Century* (Princeton, NJ: Princeton University Press, 1982); and John Michael Montias, *Art at Auction in 17th Century Amsterdam* (Amsterdam: Amsterdam University Press, 2002).

[4] See, for example: Albert Boime, *The Academy and French Painting in the Nineteenth Century* (New Haven: Yale University Press, 1971); Steven Adams, *The Barbizon School & the Origins of Impressionism* (London: Phaidon Press, 1994); Richard R. Brettell and Caroline B. Brettell, *Painters and Peasants in the Nineteenth Century* (New York: Rizzoli, 1983); and Patricia Mainardi, *The End of the Salon: Art and the State in the Early Third Republic* (Cambridge: Cambridge University Press, 1993).

[5] Here I must emphasize that I am not suggesting that the coincidence of the Prix de Rome category for *paysage historique* and the rising number of such paintings exhibited is an indication of causality, as the establishment of this category equally represents market-driven factors such as a general upward trend in the preference for landscape painting. My point is that the interconnectedness of the Prix de Rome category, trends within the market for paintings, and exhibition history is a topic that needs further discussion within a larger context of the genres' relative popularity over time.

[6] See, for example: Nan Z. Da, "The Digital Humanities Debacle: Computational Methods Repeatedly Come Up Short," *The Chronicle of Higher Education* 65, no. 29 (2019), <https://www.chronicle.com/article/the-digital-humanities-debacle>.

[7] Victoria Szabo, "Transforming Art History Research with Database Analytics: Visualizing Art Markets," *Art Documentation: Journal of the Art Libraries Society of North America* 31, no. 2 (2012): 158–75, <https://doi.org/10.1086/668109>.

[8] Jessica Marie Johnson, "Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads," *Social Text* 36, no. 4 (2018): 57–79.

[9] For recent discussions on methodological considerations for data-driven approaches to art history, see: Paul B. Jaskot, "Digital Art History as the Social History of Art: Towards the Disciplinary Relevance of Digital Methods," *Visual Resources* 35, nos. 1–2 (2019), <https://doi.org/10.1080/01973762.2019.1553651>; and Anne Helmreich, Matthew Lincoln, and Charles van den Heuvel, "Data Ecosystems and Futures of Art History," *Histoire de l'art*, no. 87 (June 29, 2021), 45–54.