Harvard Data Science Review • Issue 3.2, Spring 2021

Lessons From the ‘Data Exhaust’ of Plantation Slavery

Caitlin Rosenthal¹

¹Department of History, University of California Berkeley, Berkeley, California, United States of America

Published on: Apr 30, 2021

DOI: <https://doi.org/10.1162/99608f92.c7418199>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Column Editors' Note: *Historians, sociologists, and philosophers often struggle to trace the ways that the meaning of data shifts over time, and in particular how the 'same' data might be deployed to answer radically different questions. In this Mining the Past column, historian Caitlin Rosenthal introduces an important example from her work on the history of plantation slavery to emphasize the way the same data may be used not just to different ends, but also to reveal entirely new understandings of the past.*

Keywords: plantation slavery, data exhaust

Over the past decade, companies and data scientists have touted the strategic importance of 'data exhaust'—data that are a byproduct of other processes (Fourcade & Healy, 2016; Zuboff, 2019, pp. 68–70). So-called data exhaust usually refers to the trail of information that individuals leave behind as they move through the digital world.

In a sense, historians have always relied on data exhaust to do our work—archives contain trails of discarded data and documents, albeit on paper. In my own research, I rely on the thousands of plantation account books left behind by American slaveholders, many of them highly quantitative. The numerical remains of this brutal economic system offer a cautionary tale for modern researchers. Scholars who have relied on slaveholders' data have tended to answer slaveholders' questions. Modern data sets are no less laden with values, and revisiting the account books of American slavery can help us to recognize the risks of replicating the biases of our sources.

Consider this record of cotton picking. In August of 1859, overseer Henry L. Eggers opened a slim volume to record data about Elder Grove, a Louisiana cotton plantation ("Statement of Cotton," 1859–1866; *1860 United States Federal Census*). The picking season had just begun, and Eggers listed the names of each person who he forced into the fields to gather cotton. All of these individuals were enslaved, and Eggers was tasked with pushing them to pick as much cotton as they could. Beside each name, he entered the number of pounds picked each day, Monday through Saturday. At bottom he tallied up this data by day, and at right, by person.

In the first week of the harvest, Eggers's spreadsheet showed a variety of tasks. The cotton boles were just beginning to open, and not everyone was picking cotton. A month later, the pages had become neat grids of numbers, and the daily output had doubled (Figure 1). Almost every hand was forced to pick cotton every day, and the overseer carefully noted their daily output. In the last week of September, output reached between 7,510 lbs. and 8,930 lbs. per day. Individual productivity ranged from 780 lbs. for Julian up to 2,425 lbs. for Little Ellick. In total, the enslaved workers brought in 57,260 lbs. over the course of the week. Even Ellen, who had been "out sick for the past eight months," was deemed well enough to pick, harvesting 840 lbs. of cotton over the week. The only person working elsewhere, B. Henry, was "in the lint room" at the cotton gin—perhaps weighing the cotton so the data could be entered into the account book.

While the scale of this paper spreadsheet was, of course, much smaller than the data exhaust of today, data-minded Southern slaveholders fantasized about what they might learn from not just their own ledgers, but from hundreds or even thousands of similar records dispersed across the South. As slaveholder and accounting expert Thomas Affleck wrote in a letter, “Think of the advantage to both planters & overseers, of even 1,000 books written from day-to-day experience, scattered over the country!” (Affleck, 1855). Data from such books, kept in a standardized format, promised precise calculations and comparisons, turning Southern plantations into a vast laboratory for understanding labor productivity.

Such data promised an answer to a question that obsessed Southern slaveholders: How much cotton could they expect people to pick? This was, in a sense, the key performance indicator for increasing profits on cotton plantations. In a crop where output predicted profits and picking was the biggest bottleneck for production, knowing how much someone could pick was essential information. As South Carolina planter Plowden C. J. Weston put it, “In nothing does a good manager so much excel a bad, as in being able to discern what a hand is capable of doing and in never attempting to make him do more” (Weston, 18–, p. 8).

Yet, at the time, preformatted account books like the one used on Elder Grove plantation yielded only shadowy answers to these questions. These primitive spreadsheets enabled slaveholders to arrange and organize productivity data for each of the men, women, and children they enslaved. Planters and overseers tallied up the pounds, experimenting with incentives and brutal punishments. They made comparisons and drove people to work faster. However, without the tools for more complex analysis, definitive answers remained elusive.

Slavery’s Data, Slaveholder’s Questions

More than a century and a half later, we have the tools to analyze planter’s data easily and fully. Two economists—Alan Olmstead and Paul Rhode—have, in a sense, answered slaveholders’ questions. Toting up data from 142 plantations like Elder Grove, they have estimated changes in picking rates over time, finding that the amount of cotton each person picked increased dramatically under slavery (Olmstead & Rhode, 2008).

The increase in productivity was dramatic. By 1862, enslaved people were picking on average about four times as much cotton apiece as they had in 1801. Olmstead and Rhode have also analyzed the micro data to estimate productivity by age and by gender. They are finally answering the questions that drove Southern slaveholders to collect the data in the first place. The causes of the productivity gains remain a matter of intense debate and analysis.¹ Yet amid this debate, scholars concerned mainly with analyzing data have rarely paused to ask whether slaveholders’ questions are the most important ones to answer.

Presented with slaveholders’ data, we have continued to ask slaveholders’ questions. Indeed, the latest round of fighting over the productivity of slavery sometimes feels like a replay of a similar clash over economists Robert Fogel and Stanley Engerman’s *Time on the Cross* (1974/1995). Fogel and Engerman argued that slavery was highly profitable and that slaves even had a “Protestant ethic,” not only using planters’ data but also imbibing some of the lies they told about enslaved peoples’ motivations. Fogel and Engerman’s findings

around profitability have come to be widely accepted and have helped us to understand the extent to which slaveholders' practices reflected the growth of capitalism. Yet even as we acknowledge the ways slaveholders blended profit seeking with brutality, we are still gravitating toward slaveholders' numbers and prioritizing the kinds of questions those numbers can answer. Slaveholders' data—and the way scholars have used it—reflects this fundamental limitation: Data that is the exhaust of an economic system will be most useful for answering questions about the things that economic system values.

Data ethicists have already begun to document such biases in modern data practices, demonstrating that both the composition of data sets and the ways algorithms are trained can harm vulnerable groups, entrenching and perpetuating race and gender stereotypes (Benjamin, 2019; Criado-Perez, 2019; Noble, 2018). For example, training data from sources like newspapers has produced tools that are more useful for populations seen as newsworthy—disproportionately White men. Commercially available facial recognition algorithms value Whiteness and masculinity, performing best for lighter-skinned males and worst for darker-skinned females (Buolamwini & Gebru, 2018).

More generally, the data exhaust of the online economy will be most useful for answering a narrow set of questions about a specific set of people. Though there are surely unexpected insights to be gleaned from the digital breadcrumbs that trail behind us as we move through the online world, a huge proportion of this data is gathered as we buy and sell. The data exhaust of consumption will prompt us to tell stories about consumer capitalism that highlight our identities as consumers—narrow identities that leave many (perhaps most) important things out.

Slavery's data tells us how much cotton people could be forced to pick. Data gathered as we consume will tend to offer insights about how to sell us more.

Slow Data, Meaningful Data

What have economic historians missed by devoting their energy to answering slaveholders' questions? Another group of historians have also relied on plantation account books for their research, taking a different, slower path through the data and acknowledging the violence embedded in historical archives (Fuentes, 2016; Rosenthal, 2018; Smallwood, 2007). Data scientists can learn from historians as we all seek to develop critical approaches to biased data sets.

Returning to the Elder Grove account book, we can learn as much about the institution of slavery by questioning the data as we can by analyzing it. The grueling nature of the work—invisible in the productivity calculations—is immediately apparent.

Labor was not freely given. On the neat grid of numbers for the last week of September, there is one gap, for Charles who has “run away.” He escaped at the peak of the picking season—at the very moment when the planter most desired his labor. Following Charles's row forward, we see that he managed to stay away for 3 full

weeks. How did he survive away for so long? He stayed away on 2 days too rainy for anyone to work—who sheltered him? Did he negotiate the conditions of his return? Did he inspire (and perhaps even aid) the escape of Mose the following week?

We can see, too, the constant sicknesses that afflicted enslaved people. As the picking season began, Ellen, Emily, Big Watson, and Angeline were all too ill to go to the fields, yet as the need for pickers increased, they were sent to work anyway. Would Francis, out for 4 days and described, as “always sick” in the margins, have been so ill if he had not been constantly sent to the fields? On Residence plantation, also in Louisiana, the journal refers to a “measles gang,” a team of laborers who were still sick but had been deemed well enough to return to making sugar (Rosenthal, 2018, p. 115).

Looking around and between the numbers raises different questions—about power, struggle, and coercion. These questions are easier to ask with the benefit of hindsight, but even at the time, abolitionists sometimes remade data collected in the pursuit of profit into exhibits of the human costs of slavery. British writer James Stephen relied on data from a plantation account book to critique West Indian slavery. He used both the numbers and the horrifying descriptions of punishment interspersed between them to reconstruct the poor treatment of enslaved people, even depositing the original with his bookseller so that other individuals could reflect on it and see that his “extracts are correct” (Rosenthal, 2018, pp. 80–81).

Can we craft a slow data approach that opens up similar possibilities in modern data science? Though the scale of today’s data sets makes it more difficult to analyze them qualitatively, researchers have called for careful reflection and documentation of their biases. The hope is that through auditing and documentation—creating “Datasheets for Datasets”—the creators and distributors of databases can help users to see their biases and thus avoid misuse (Geburu et al., 2020; Raji et al., 2020). Citing lessons from historical archives, they have suggested that careful curation can help us to critique and counter underlying biases in data sets—for example, developing supplementary data sets that counter biases in representation (Jo & Geburu, 2019).

Developing adequate documentation and supplementary data sets will not be fast. Indeed, artificial intelligence ethicists have been clear that these processes are “not intended to be automated” (Geburu et al., 2020, p. 3). While automation may be fast and cheap, it compounds the risk of unthinkingly replicating data’s biases and overlooking what is left out (Eubanks, 2018).

Nor will such ‘slow data’ approaches make data exhaust useful for every question. More often, careful documentation and critique will expose what is missing, making clear what questions ‘data exhaust’ simply will not answer. Here, too, scholars of slavery can show the way. Even as historians have uncovered stories about enslaved people in the data exhaust of slavery, they have also recognized how little the records often tell us (Fuentes, 2016; Hartman, 2007; Morgan, 2016). No amount of reading between the numbers will wring full stories about enslaved people from records of their value as labor and capital. If historians have sometimes found the human in the data, more often we are left searching.

Yet here modern data scientists have the advantage over historians, who cannot go back in time to ask different questions. Today's data scientists can reckon with bias, crafting new research questions and recognizing that answering them fully will require multiple kinds of data, qualitative and quantitative, fast and slow.

Sold by W. H. Fla. Natchez

STATEMENT OF COTTON

Picked on *Robert Stewart's* Plantation,
for the Week ending *September 4th* 18*77*

NAMES.	Monday. <i>71</i>	Tuesday. <i>72</i>	Wednesday. <i>73</i>	Thursday. <i>74</i>	Friday. <i>75</i>	Saturday. <i>76</i>	Total.	REMARKS.
<i>Isaac</i>	345	360	380	375	390	400	2270	
<i>Isaac</i>	350	380	410	410	410	415	2320	
<i>2 Ellick</i>	370	405	410	420	415	405	2425	
<i>Shelley</i>	380	370	325	345	335	353	1965	
<i>B. Henry</i>	<i>in the Lint Room at the Lint</i>							
<i>1 Henry</i>	270	260	270	275	285	300	1680	
<i>B. Rung</i>	295	310	320	340	335	325	1925	
<i>B. Watson</i>	330	375	330	345	335	355	2070	
<i>B. Carter</i>	210	285	325	345	325	340	1830	
<i>Dave</i>	<i>sick</i>		230	350	335	355	1270	
<i>Mose</i>	250	275	285	300	295	290	1695	
<i>Gordon</i>	265	300	325	345	330	345	1910	
<i>Charles</i>	<i>Anna</i>							
<i>Bole</i>	215	230	230	260	250	245	1430	<i>B. Sleeping Run Away</i>
<i>Andrew</i>	380	390	405	405	410	410	2400	
<i>children</i>	270	305	315	350	350	370	1960	
<i>Evert</i>	280	280	290	300	305	315	1770	
<i>Aaron</i>	195	220	280	270	220	275	1240	
<i>Frank</i>	320	325	345	355	365	375	2085	
<i>Isaac</i>	275	235	235	255	255	255	1545	
<i>Mahala</i>	200	270	205	225	220	270	1510	
<i>Lige</i>	290	310	305	335	345	345	1925	
<i>Emby</i>	215	215	230	240	230	230	1350	
<i>Catherine</i>	230	220	255	250	240	240	1435	
<i>1 Mary</i>	210	300	300	270	200	200	1580	
<i>Ellen</i>	125	135	150	150	155	175	840	
<i>Mariah</i>	200	200	210	210	210	205	1235	
<i>Liddy</i>	275	210	225	235	230	230	1350	
<i>B. Mary</i>	225	230	250	250	260	265	1480	
<i>Martha</i>	180	180	190	200	200	195	1145	
<i>Angeline</i>	000	170	210	230	220	270	1000	
<i>Julian</i>	120	125	135	140	157	155	780	
<i>Francis</i>	175	185	195	210	195	195	1155	
<i>Doctor</i>	160	155	150	165	155	190	975	
<i>B. Morris</i>								

Figure 1. "Statement of Cotton," 1859–1866, Robert H. Stewart Account Books.

Disclosure Statement

Caitlin Rosenthal has no financial or non-financial disclosures to share for this article.

References

1860 United States Federal Census, Ward 3, Carroll, Louisiana (p. 368). Family History Library Film 803409. <https://www.ancestrylibrary.com/discoveryui-content/view/38477521:7667>

Affleck to Hammond, 3 January 1855, Box 32, Folder 10, Thomas Affleck Papers, Louisiana and Lower Mississippi Valley Collections, LSU Libraries, Baton Rouge, LA.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the New Jim Code*. Polity.

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of Machine Learning Research: Vol. 81. Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77–91). <http://proceedings.mlr.press/v81/buolamwini18a.html>

Criado-Perez, C. (2019). *Invisible women: Data bias in a world designed for men*. Abrams Press.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

Fogel, R. W., & Engerman, S. L. (1995). *Time on the cross*. W. W. Norton & Company. (Original work published 1974)

Fourcade, M., & Healy, K. (2016). Seeing like a market. *Socio-Economic Review*, 15(1), 9–29. <https://doi.org/10.1093/ser/mww033>

Fuentes, M. J. (2016). *Dispossessed lives: Enslaved women, violence, and the archive*. Early American Studies. University of Pennsylvania Press.

Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., & Crawford, K. (2020). Datasheets for datasets. *arXiv*. <https://doi.org/10.48550/arXiv.1803.09010>

Hartman, S. V. (2007). *Lose your mother: A journey along the Atlantic Slave Route*. Farrar, Straus and Giroux.

Jo, E. S., & Gebru, T. (2019). Lessons from archives: Strategies for collecting sociocultural data in machine learning. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 306–316). ACM. <https://doi.org/10.1145/3351095.3372829>

Morgan, J. L. (2016). Accounting for “The Most Excruciating Torment”: Gender, slavery, and Trans-Atlantic passages. *History of the Present*, 6(2), 184–207. <https://doi.org/10.5406/historypresent.6.2.0184>

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Olmstead, A. L., & Rhode, P. W. (2008). Biological innovation and productivity growth in the antebellum cotton economy. *The Journal of Economic History*, 68(4), 1123–1171. <https://doi.org/10.1017/S0022050708000831>

Weston, P. C. J. (18–). *Rules for the government and management of ____ plantation: To be observed by the overseer*. A. J. Burke.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *arXiv*. <https://doi.org/10.48550/arXiv.2001.00973>

Rosenthal, C. (2018). *Accounting for slavery: Masters and management*. Harvard University Press.

“Statement of Cotton,” 1859–1866, Robert H. Stewart Account Books, Mss. 404, 4732, Louisiana and Lower Mississippi Valley Collections, LSU Libraries, Baton Rouge, LA.

Smallwood, S. E. (2007). *Saltwater slavery: A middle passage from Africa to American diaspora*. Harvard University Press.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

©2021 Caitlin Rosenthal. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Rhode and Olmstead argue persuasively that the introduction of new strains of cotton was the decisive factor. Historians have countered that such increases would not have been so easily achieved without the violence that underpinned planters' control. These are the questions that slaveholders loved to debate. What was the best strategy for increasing yields? Which seeds should they plant? How should overseers maintain order? During the 19th century southern agricultural journals were full of articles debating seeds, manure, and labor discipline. ↩