

Nathanael D Nihart. Digital Humanities Techniques for Asylum Studies in the Archive. A Master's Paper for the M.S. in I.S. degree. December 2022. 25 pages. Advisor: Ryan Shaw

This project explores the utilization of digital humanities techniques and tools for the purposes of transcription, data analysis, visualization and presentation of archival data pertaining to asylums in the 19<sup>th</sup> and 20<sup>th</sup> centuries. Since 2017, the Community Histories Workshop's Asylum in the Archive initiative has been digitizing and transcribing records relating to the admission and the assessment of people in the Dorothea Dix asylum from 1856 through 1922. My project builds off this work by developing machine learning models to increase transcription efficiency for future records as they become available, and working with the dataset to analyze, visualize and present the data to tell the history of the asylum and psychiatry. The result demonstrates the value of the work already done, leaves examples and models for future scholarship utilizing this dataset and contributes to broader understanding of the emergence of psychiatry and the asylum in the South.

Headings:

Digital Humanities

Data Analysis

Data Visualization

DIGITAL HUMANITIES TECHNIQUES FOR ASYLUM STUDIES IN THE  
ARCHIVE

by  
Nathanael D Nihart

A Master's paper submitted to the faculty  
of the School of Information and Library Science  
of the University of North Carolina at Chapel Hill  
in partial fulfillment of the requirements  
for the degree of Master of Science in  
Information Science.

Chapel Hill, North Carolina

December 2022

Approved by

---

Ryan Shaw

## Table of Contents

<b>INTRODUCTION .....</b>	<b>2</b>
<b>RELATED WORK .....</b>	<b>5</b>
<b>METHODS.....</b>	<b>9</b>
IMPROVING TRANSCRIPTION EFFICIENCY .....	9
ANALYZING AND VISUALIZING THE DATA .....	13
<b>PROJECT EVALUATION .....</b>	<b>17</b>
<b>CONCLUSION .....</b>	<b>18</b>
<b>REFERENCES .....</b>	<b>19</b>

## INTRODUCTION

The acquisition of data usable for quantitative analysis in the domain of asylum studies and the history of psychiatry is typically a labor-intensive process involving locating the relevant primary sources in an archive, arranging for a visit to review these materials, digitizing the materials, transcribing what is needed for the research question(s), cleaning this data, and running analyses on the resulting dataset. The Community Histories Workshop (CHW), a research entity affiliated with both the University of North Carolina at Chapel Hill and Auburn University, developed its Asylum in the Archive initiative in 2017 with the goal of utilizing archival records to facilitate research into the history of the asylum in the South, the emergence of psychiatry and the experience of those deemed mentally ill. As part of this initiative, they located primary sources related to the Dorothea Dix Hospital, the first asylum in North Carolina which operated from 1856 until 2012, in the North Carolina State Archives; these sources included admission ledgers and general case books detailing patient admissions and assessments from the opening of the asylum in 1856 through 1922 (state law restricts public access to these records with a 100 year sequester). These records were then digitized and transcribed by graduate students over the course of five years, resulting in two spreadsheets of data representing the admission ledgers and general case books. From 2021 to 2022, a relational database was built to house the consolidated data from these two spreadsheets, and significant work was done to clean and normalize the transcribed data. This database, thought to be the first relational database housing patient records from 19<sup>th</sup> century asylums, creates new

opportunities for research while reducing the data preparation burden on researchers wishing to use the data for data analysis and visualization.

My project engages with the Asylum in the Archives initiative by investigating the use of digital humanities techniques and tools for both the acquisition and use of data. As digitization and transcription of pages from the admission ledgers will continue indefinitely as they become accessible on a rolling basis under the public records law, I investigated the use of machine learning for automating handwriting transcription, ultimately training a model for the semi-automated transcription of admissions ledger pages. The application of this model on new pages significantly reduces the amount of time needed to transcribe the data. While the utility of the new relational database in allowing for efficient and low-burden research may seem apparent, it has not yet been utilized since being released in Spring 2022; my project serves to demonstrate this utility, by providing presenting descriptive analysis and visualization of what the data shows about the early days of psychiatry, the asylum, and its patients. In exploring tools, techniques and platforms for data analysis and visualization, some of my choices were informed by the need to create an example of engagement with the data that was as accessible as possible for researchers with a variety of skill sets, including students working with the asylum records and data as part of practicums and field experiences. After generating new examples of data visualizations in Tableau, I switched to using R for more robust analysis and visualization of the data. I created a shared workspace in RStudio Cloud for the Community Histories Workshop, storing the code I wrote for analysis and visualization in this space so that it can potentially be of use to future students and researchers. In publishing the data visualizations to the web, I worked to

make them as interactive as possible to allow for some dynamic exploration of the data through the visualizations.

While the focus of my project was on the exploration and application of digital humanities techniques and tools for engagement with archival data in the context of asylum studies, the resulting data analysis and published visualizations potentially contribute (albeit minimally) to a better understanding of the asylum in the south (much of the research around asylum studies has focused on the northeast United States) and illuminates conditions warranting more study. Of particular importance in contextualizing this research is the race of the patients in the data being used: there were 47 Black patients in the early years of the asylum (The state built a separate asylum - Cherry - for Black patients in 1880), with the remaining 8,297 presumably white. While some analysis and visualization has been done around race in the early years of Dix, the exclusion of non-whites from the asylum for the vast majority of the time span of our data necessitates the explicit recognition that the results are broadly contributing to our understanding of the white asylum and psychiatry as it applied to white people.

Key Terms:

**Descriptive analysis:** “Identifying patterns in the data to answer questions about who, what, where, when, and to what extent” (Loeb et al., v.)

**Digital Humanities:** The use of digital tools and computing techniques to further humanities research and the presentation of one’s findings. Notoriously hard to define: for a series of definitions provided by a diverse set of digital humanists, see Gold (2012).

## **Related Work**

The interdisciplinary nature of this project means that there is significant peripheral work in the primary domains surrounding the research – asylum studies, psychiatry and the history of medicine, archival studies and information studies, and digital humanities. As my project is primarily concerned with the demonstration of how to leverage digital humanities and information systems to enhance research using archival records, I will focus this section on general domain research in asylum studies as well as digital humanities while also touching on notable similar projects.

There have been a number of shifts in how researchers have approached asylum studies Brown (1994), with arguments over what allows for a meaningful representation of the experience of the asylum. We can find qualitative research describing the day-to-day experiences of patients in the asylum using case studies (Dwyer, 1987), as well as quantitative studies that will be discussed below. Specific to the asylum my project revolves around, O'Rourke (2010), provides a broad comprehensive overview of the politics, administration, history, development, and changes in the Dix asylum. Research looking at the parallel rise of psychiatry as a profession alongside the asylum bemoans the lack of accessible asylum data for research around the beginning of the 1900s (Scull, 2022). Only two quantitative analyses look at the demography and epidemiology of asylum populations: the first based on admissions ledgers from the asylum in Colorado (Magennis & Lacy, 2014), and the second looking broadly at hospitals across the Northeast (Maeda, 2016). Both studies mention the absence of this type of data in

historical analyses of asylums, and Maeda outlines the inaccuracies that the use of aggregated data has caused with analyses. Whether or not it is ethical for patient records (unredacted and including names) from a psychiatric asylum to be accessible to the public, and the duty of care for researchers utilizing these records, is a source of significant academic focus. A broad overview of the ethics of using old medical records has been written by Wright & Saucier (2013), and a SILS master's thesis addresses this (Almond, 2019).

Exploration of digital humanities techniques and tools in working with archival data on interdisciplinary projects can be seen in theory and practice. There is significant research around transformation of records in the archives and improvements in the transcription process of humanities data through crowdsourcing the transcription process (Causer et al., 2018; Blickhan et al., 2019). Because of the sensitivity of the data we're working with, however, we are unable to take advantage of public-facing crowdsourcing transcription platforms and it must be done in-house. A novel approach to this problem was developed by a graduate student at the University of Texas while working on the Colored Insane Asylum project; they developed an application to segment images into anonymized chunks of text (individual words) that can then be fed into a transcription crowdsourcing platform (Smith, 2014). Of the most significance for this project, however, are developments in the realm of handwritten text recognition. While earlier research used the Computer-Assisted Transcription of Text Images (CATTI) framework (Toselli et al., 2017), a case study (Muehlberger et al., 2019) and systematic review (Nockels et al., 2022) of the Transkribus application has demonstrated that this technology is both mature and highly useful to researchers for working with archival



collections and humanities data. As I learned the R programming language to conduct data analysis and visualization, I relied heavily on online textbooks and guides, particularly R for Data Science (Grolemund & Wickham, 2017) and The Epidemiologist R Handbook (Batra, 2021). Throughout the process of viewing, analyzing, and presenting data, I drew from ideas presented in “Data Feminism” (D’Ignazio & Klein, 2020). More specifically targeted to methods and tech for humanities research data analysis and presentation is Hai-Jew’s “Data analytics in digital humanities” (2017). Research into humanities scholarship and research practices reveals trends in challenges in the application of technology for digital humanities (Liu et al., 2021). There is a broad set of technologies used by humanities scholars, and a large part of the research process can be consumed by data preparation and acquiring the technological knowledge needed for this process (Given & Willson, 2018). There is also significant research looking at the role that information professionals can play in engaging with digital humanities tools and techniques to support and promote scholarship (Burns, 2016), suggesting that there is a natural fit for providing support in content and data, technologies and tools (Zhang et al., 2015).

Similar research projects include Davis (n.d.)’s Colored Insane Asylum project. This project digitized and transcribed records from Virginia’s Central State Hospital, storing some records in a spreadsheet for general analysis. This project has fostered significant research including dissertations and theses. The analysis of records from Colorado’s insane asylum by Magennis & Lacy (2014) utilized admission ledger data that had been transcribed and stored in an unknown “machine-readable” format at Colorado College, but additional details from other uses of this dataset are unknown. Related works

specific to the Asylum in the Archive project and the underlying dataset include a StoryMap that provides an overview of the Asylum in the Archive initiative (Allen & Tams, 2022), a website in which Huber (2019) cleaned and visualized a subset of the Dix records from the admission ledger showing general characteristics and trends in the data using R, and a website by Dasgupta (2020) using Stata MP in a more robust analysis of a subset of the data (a portion of the admissions ledger) to both describe the data and also do statistical analysis. There is additionally a current research project by a psychiatry resident at UNC that focuses on the relationship between race and diagnosis of Dementia Praecox/schizophrenia using both the dataset held by CHW for the Dix asylum and data from Cherry asylum (North Carolina's Black insane asylum during this time period).

## Methods

In this section, I discuss the investigation and application of digital humanities tools and techniques to the Community History Workshop's Asylum in the Archive initiative. I split this section into two parts: Improving Transcription Efficiency, and Analyzing, Visualizing, and Presenting the data.

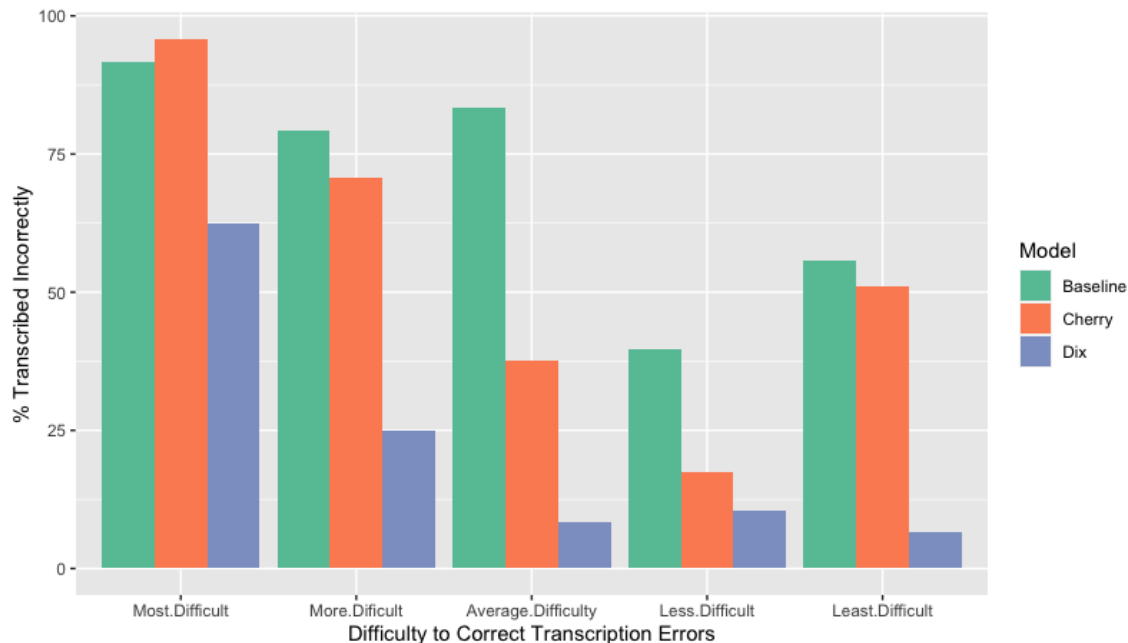
### Improving Transcription Efficiency

A significant bottleneck in the analysis of archival records is the labor-intensive nature of transcribing digitized records. At the onset of this thesis project, approximately 300 pages (representing approximately 8000 admissions) from the Dix admission ledger had already been transcribed, with 9 pages (representing 215 admissions) still needing transcription; additional pages will be digitized and require transcription as they become accessible on a rolling basis under the public records law. To increase the efficiency of this process, I investigated the use of Transkribus, an application for transcribing handwritten historical documents, for automating part of the transcription process. While Transkribus comes with models for text recognition, it also allows for the training of machine learning models for text recognition using your own data. In addition to the base model that comes with Transkribus, there were also pertinent models for recognizing text in admission ledgers from the Cherry asylum that were created by me in my role as a graduate research fellow for the CHW. Both the base and Cherry models provided underwhelming performance in the transcription of the 9 pages of Dix asylum admission records; after correcting the transcriptions provided by these models, I used the resulting 4689 words in

2906 lines to train a new model (Dix) that additionally utilized the existing data from the existing available models. To evaluate the effectiveness of the newly trained model, I compared character (CER) and word error rates (WER) between the transcriptions generated by the different models. Transkribus stores a versioned history including each stage of a document being processed and includes a tool to calculate differences and error rates between a reference ground truth version and the versions of the document that exist after a model has been applied to generate a transcription of the document. While I had utilized the 9 remaining pages of the Dix admission ledger that still needed transcription to train the model, I selected an older, previously transcribed, image from the admission ledger for evaluation of the models; this was necessary to evaluate accuracy correctly as the new model would overperform if given one of the pages that it was trained on. For this page, I ran the base model provided by Transkribus, the model trained on the Cherry records, and the new model that I trained on the Dix records. Word error rate, or the number of words that differ from the ground truth, was 43.2% using the base model, improving to 31.69% with the Cherry model, and 9.72% using the Dix model. Character error rate, or the number of characters that differ from the ground truth, was 22.36% using the base model, improving to 16.92% with the Cherry model, and 5.88% using the Dix model. There were 500 words in the test document, so calculating the number of words needing manual correction per model shows an improvement from 216 with the base model to 158 with the Cherry model to 49 with the Dix model. While this suggests a marked improvement in the model accuracy, it was also important to understand how these changes affect transcription efficiency regarding time spent correcting transcription errors. Certain columns are incredibly easy to correct, such as Admission ID (a four-digit

number) while others are time consuming (names). I analyzed which columns contained transcription errors for each model, assigning each field a difficulty rating from 1 (easiest) to 5 (hardest) to correspond to how much time and knowledge it requires to correct errors in these fields. The three most difficult fields to correct are name (difficulty: 5), form (difficulty: 4), and supposed cause (difficulty: 3), with the remaining fields having a 2 (n=6) or 1 (n=5) difficulty. Fields with a 1 difficulty can be corrected en masse, while fields with a 2 can be corrected quickly but at an individual level. Fields with difficulty 3 or 4 are based on somewhat standardized vocabularies that might need to be consulted by looking in a superintendent report, and the name field has such great variety that it requires either a comprehensive visual understanding of variations in cursive handwriting and/or a strong knowledge of historical names in North Carolina. To analyze improvements, I reviewed values at the cell level – a cell might contain several words, so I recalculated model accuracy at this level, finding that with the baseline model there were 55.06% with errors, 36.61% with the Cherry model, and 13.69% with the Dix model. I then analyzed how many cells contained errors and grouped them based on column correction difficulty.

Percentage of cells containing errors per model, by column difficulty:



As the supposed cause and form fields contain a limited controlled vocabulary, the newly trained model improved recognition of these values substantially. Fields with a 1 or 2 difficulty rating similarly were composed of either numeric values or semi-repetitive values (i.e., county names, textual numbers), and saw a similar rapid improvement in detection accuracy. The name field, due to massive variability, saw marked improvement due to increased character recognition, but retained a high error rate. While additional training data for the model would continue to improve its accuracy, this suggests that it already provides significant time savings.

The length of time for a seasoned worker familiar with the admissions ledgers to transcribe a page from the ledger is around 60 minutes; for a new worker becoming acquainted with the contents of the ledger or the style of handwriting from the time, it can take significantly longer. Using Transkribus, a subset of the data (Sex and Social Relation

must be manually transcribed as they use multi-column tick mark boxes that Transkribus cannot adequately recognize) with relatively high accuracy can be produced in under 30 minutes. I have provided conservative estimates for the steps of the process:

- Image layout setup: 5 minutes
- Transcription corrections: 8 minutes
- Copy/Paste to excel: 3 minutes
- Normalize data using OpenRefine: 5 minutes
- Fill in remaining fields (sex, social relations): 2 minutes
- Merge data into main transcription spreadsheet: 2 minutes
- Total: 25 minutes

It is important to note that similar to the process of manually transcribing the records, this time reflects how long it might take someone familiar with both Transkribus and the records to process a single page from the admission ledger. As there is a learning curve for using Transkribus, I created a training video outlining the steps for going through this process as well as various methods I found to increase efficiency in the process.

Due to the ongoing nature of the project and turnover in the students who work on transcribing the ledger pages, this transcription model should significantly reduce the time required to transcribe new ledger pages as they become available.

### Analyzing and visualizing the data

While the dataset we are working with is derived from records accessible to the public, we acknowledge the sensitivity of the data and, as such, made decisions regarding the ethics of its presentation – first, any data that is presented in a publicly accessible way

(i.e. visualizations) will be aggregate or otherwise deidentified data; second, it will only be published if it lends itself to pushing forward the understanding of the asylum, the patient within, and/psychiatry.

During the early stages of this project, I utilized Tableau, a proprietary tool for data visualization that provides a free license for students, for exploration of the dataset. The graphical interface of this tool lends itself to ease of use, and while completing a field experience for the CHW in the spring of 2022 I had already created a shareable Tableau file that contained a preconfigured connection to the data and sample visualizations. For publishing a presentation of the data on the web, I began investigating the use of StoryMaps, a proprietary tool for storytelling with images and maps. However, as we began asking more complex questions of the data, and it became clearer that the resulting product would be heavy in complex visualizations, I began exploring other options. For more complex and advanced analysis and visualizations, I began exploring the data using the R programming language and RStudio, an open-source tool for using R to analyze and visualize data. There are many packages available for R that enhance the ability to analyze, visualize and present data. One of the more challenging aspects of the visualization related to geographical representation of the data. The geographical data in the dataset is limited to county names (residence) and states (nativity); to map these, I utilized the *tigris* package to download shapefiles for NC counties and join them to dataframes being used for choropleth mapping. Further, to be able to show admissions or patients as a proportion of the population, I acquired census population data from the IPUMS National Historical Geographic Information System (Manson, et al., 2022). The county population data provided exists at decade intervals. Initial analysis was done



against total population counts; upon review, we switched to using county population counts of white residents, as Dix only admitted white people (with 47 exceptions in the first years of the asylum) during the timeframe of our dataset. Future work could incorporate census age and sex data for additional proportional analyses. As the data spans 66 years, some visualizations can feel difficult to interpret, or subsets of the data by time span are desirable to zoom in on certain periods. To facilitate this, I utilized the DYGraphs package, which transforms time-based visualizations by adding a slider at the bottom to select a subsetted time span. To improve readability and usability of visualizations, I also incorporated the Plotly package, which adds interactivity to visualizations – by hovering over the viz, you can see specific counts instead of having to rely on labels (that can make complex charts crowded) or make approximations using the axes and a visual assessment.

I used RStudio Cloud to create a shared workspace for the Community Histories Workshop, so that the R code I wrote could be utilized by future researchers.

To produce the resulting analysis and visualization in a meaningful way, I explored several publishing options for and landed on Quarto, a flexible system for publishing R (or Python) code and dynamic visualizations to a website alongside markup text. While it runs the underlying R scripts using live connections to the data source, it obfuscates the underlying data, making it inaccessible to the viewer of the website and satisfying our requirements for maintaining the sensitivity of the data. The product of this work is published online at <https://ndnihart.quarto.pub/dorothea-dix-asylum-ledgers-and-casebooks/>. The visualizations and context provided in this site provide a way for future students and researchers to become acquainted with the contents of the dataset and see

general trends they may wish to explore. It also provides a tool for marketing the possibilities of this dataset in seeking funding for further development of the datasets and continued exploration of the data.

Future work beyond what has been done could include greater textual and descriptive contextualization of the data being presented; by storing the underlying code in a space accessible to the CHW, visualizations can be easily modified or duplicated to present closer looks at subsets of the data as needed.

## **Project Evaluation**

To evaluate the success of the machine learning model to increase transcription efficiency, I calculated the accuracy of the new model in relation to the base model as well as situating the accuracy within fields of variable difficulty for human transcribers. The results of this analysis, along with an estimate of time savings, was provided in the section on improving transcription ability. Additional evaluation, including feedback about the ease of the process and time spent, should be conducted once new pages become available for transcription and a new person is brought in for that transcription. Development of the data analysis and visualization has been ongoing for the semester, with weekly check-ins with Dr. Allen to review ideas around visualizations and output. This guidance helped shape what was needed: visualizations that could be understood by, and meaningful to, a wide variety of users - from clinical staff to humanities researchers; further, the data needed to be presented in a way that allowed for dynamic interaction and which stimulated questions that could be asked of the data. Upon publishing the resulting visualizations to a Quarto notebook, I met with Dr. Allen again to review the product, and made changes to the visualizations based on feedback. There are many more strands to follow in the visualization of the data, lending itself to further development. Use of the notebook, and the R code left in the repository for future researchers, would be necessary for further evaluation.

## **Conclusion**

There is much knowledge to be gleaned from the patient records dataset from the Dix Asylum, and digital humanities techniques are well suited to explore, analyze, visualize, and present this data to a wide variety of audiences. My project highlighted the possibilities for utilizing digital humanities techniques and tools when working with archival asylum records; it contributed to the process of developing this dataset through a machine learning model for improved transcription efficiency of new records, and the possibilities for utilization of the dataset for research by sharing the code use for conducting analyses and visualizations of the data. The resulting published visualizations provide a resource for understanding some of what the data can contribute to broader knowledge around both the asylum in the south and the emergence of psychiatry.

## References

- Allen, R., & Tams, L. (2022). *The Asylum in the Archive*. Community Histories Workshop. <https://communityhistories.org/our-initiatives/asylum-archive/>
- Almond, S. (2019, May). *A Study of the Use of Unrestricted Health Information in Digitized Medical Archival Collections* (Master's paper). University of North Carolina School of Information and Library Science.  
<https://doi.org/10.17615/bchm-dh13>
- Batra, N., et al. (2021). *The Epidemiologist R Handbook*.  
<https://doi.org/10.5281/zenodo.4752646>
- Blickhan, S., Krawczyk, C., Hanson, D., Boyer, A., Simenstad, A., Hynning, V., & Van Hynning, V. (2019). Individual vs. Collaborative Methods of Crowdsourced Transcription. *Journal of Data Mining & Digital Humanities, Special Issue on Collecting*, . . . <https://doi.org/10.46298/jdmdh.5759>
- Brown T. E. (1994). Dance of the dialectic? some reflections (polemic and otherwise) on the present state of nineteenth-century asylum studies. *Canadian bulletin of medical history = Bulletin canadien d'histoire de la medecine*, 11(2), 267–295.  
<https://doi.org/10.3138/cbmh.11.2.267>
- Bryan, C. S. (2014). *Asylum doctor: James Woods Babcock and the red plague of pellagra*.
- Burns, J. A. (2016). Role of the Information Professional in the Development and Promotion of Digital Humanities Content for Research, Teaching, and Learning in

the Modern Academic Library: An Irish Case Study. *New Review of Academic Librarianship*, 22(2–3), 238–248.

<https://doi.org/10.1080/13614533.2016.1191520>

Causser, T., Grint, K., Sichani, A. M., & Terras, M. (2018). ‘Making such bargain’: Transcribe Bentham and the quality and cost-effectiveness of crowdsourced transcription1. *Digital Scholarship in the Humanities*, 33(3), 467–487.

<https://doi.org/10.1093/lc/fqx064>

Dasgupta, N. (2020). *Dix Hospital Ledger - data cleaning*. Opioid Data Lab.

<https://opioiddatalab.github.io/public/DixAsylum/docs/Dix%20Hospital%20Ledger%20-%20data%20cleaning.html>

Davis, K., Galloway, P. K., & Karadkar, U. (n.d.). *Central State Hospital Digital Library & Archives Project*. The Central State Hospital Digital Library & Archives Project. Retrieved 2022, from <https://www.coloredinsaneasylums.org/>

D’Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. Amsterdam University Press.

Dwyer, E. (1987). *Homes for the mad: Life inside two nineteenth-century asylums*. New Brunswick, N.J: Rutgers University Press.

esri. (2022). *Story Maps and the Digital Humanities*.

<https://collections.storymaps.esri.com/humanities/>

Hai-Jew, S. (2017). *Data analytics in digital humanities*.

Given, L. M., & Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6), 807–819.

<https://doi.org/10.1002/asi.24008>

- Gold, M. K. (2012). *Debates in the Digital Humanities* (1st ed.). Univ Of Minnesota Press. <https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/550ab4e6-ca58-4840-acba-ea555be32601>
- Grolemund, G., & Wickham, H. (2017). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (1st ed.). O'Reilly Media.
- Huber, K. M. (2019). *Visualizing the Dorothea Dix Hospital Ledger*. Visualizing the Dorothea Dix Hospital Ledger. Retrieved 2022, from <https://kmhuber.github.io/index.html>
- Liu, R., McKay, D., & Buchanan, G. (2021). Humanities Scholars and Digital Humanities Projects: Practice Barriers in Tools Usage. *Linking Theory and Practice of Digital Libraries*, 215–226. [https://doi.org/10.1007/978-3-030-86324-1\\_25](https://doi.org/10.1007/978-3-030-86324-1_25)
- Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive analysis in education: A guide for researchers. (NCEE 2017–4023). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Maeda, H. (2016). The Discovery of Mental Hospital Patients: *A Historical Epidemiology of Institutionalization in the American North, 1880–1920*. *Social Science History*, 40(3), 463–490. <https://doi.org/10.1017/ssh.2016.14>
- Magennis, A., & Lacy, M. (2014). Demography and Social Epidemiology of Admissions to the Colorado Insane Asylum, 1879–1899. *Social Science History*, 38(1-2), 251–271. doi:10.1017/ssh.2015.18

- Manson, S., Schroeder, J., Van Riper, D., Kugler, T., & Ruggles, S. (2022). *IPUMS National Historical Geographic Information System* (Version 17.0) [Dataset]. IPUMS. <http://doi.org/10.18128/D050.V17.0>
- Muehlberger, G., Seaward, L., Terras, M., Ares Oliveira, S., Bosch, V., Bryan, M., Colutto, S., Déjean, H., Diem, M., Fiel, S., Gatos, B., Greinoecker, A., Grüning, T., Hackl, G., Haukkovaara, V., Heyer, G., Hirvonen, L., Hodel, T., Jokinen, M., . . . Zagoris, K. (2019). Transforming scholarship in the archives through handwritten text recognition. *Journal of Documentation*, 75(5), 954–976. <https://doi.org/10.1108/jd-07-2018-0114>
- Nockels, J., Gooding, P., Ames, S., & Terras, M. (2022). Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research. *Archival Science*, 22(3), 367–392. <https://doi.org/10.1007/s10502-022-09397-0>
- Noll, R. (2011). *American Madness: The Rise and Fall of Dementia Praecox*. Harvard.
- O'Rourke, M. L., North Carolina., & North Carolina. (2010). *Haven on the hill: A history of North Carolina's Dorothea Dix Hospital*. Raleigh, N.C: Office of Archives and History, NC Dept. of Cultural Resources.
- Scull, A. (2022). *Desperate Remedies: Psychiatry's Turbulent Quest to Cure Mental Illness*. Belknap Press: An Imprint of Harvard University Press.
- Smith, D. (2014). *Steganoscription: exploring techniques for privacy-preserving crowdsourced transcription of handwritten documents* [MA Thesis]. School of Information, The University of Texas at Austin.



- Toselli, A. H., Leiva, L. A., Bordes-Cabrera, I., Hernández-Tornero, C., Bosch, V., & Vidal, E. (2017). Transcribing a 17th-century botanical manuscript: Longitudinal evaluation of document layout detection and interactive transcription. *Digital Scholarship in the Humanities*, 33(1), 173–202. <https://doi.org/10.1093/llc/fqw064>
- UNC University Libraries. (n.d.). *On The Books – Jim Crow and Algorithms of Resistance*. On The Books. Retrieved 2022, from <https://onthebooks.lib.unc.edu/>
- Wright, D., & Saucier, R. (2013). Madness in the Archives: Anonymity, Ethics, and Mental Health History Research. *Journal of the Canadian Historical Association*, 23(2), 65–90. <https://doi.org/10.7202/1015789ar>
- Zhang, Y., Liu, S., & Mathews, E. (2015). Convergence of digital humanities and digital libraries. *Library Management*, 36(4/5), 362–377. <https://doi.org/10.1108/lm-09-2014-0116>