

**Exploring Data in Digital Editions
Lesson Plan for Twelfth Grade**

Teacher Copy

Standards/Objectives:

| <u>Mississippi Minorities Studies Curriculum Standards</u> | |
|-------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| MIN.2.1 | Describe the various causes of prejudice including social distance, economic competition, scapegoating, conflicting social and religious norms, and stereotypes. |
| MIN.2.2 | Describe political and cultural factors that may serve to maintain inequalities experienced by minority groups |
| <u>College, Career, and Civic Life C3 Framework</u> | |
| AAS.9 | Debate the issues confronting contemporary African Americans in the continuing struggle for equality. |
| <u>College, Career, and Civic Life C3 Framework</u> | |
| IC.3A.1 | Evaluate the ways computing impacts personal, ethical, social, economic, and cultural practices. |
| IC.3A.1a | Students should be able to evaluate the accessibility of a product to a broad group of end users, such as people who lack access to broadband or who have various disabilities. |
| IC.3A.2b | Students should also begin to identify potential bias during the design process to maximize accessibility in product design. |

Introduction: This lesson plan is designed for students to critically engage with digital primary sources and the methods of digital humanities, specifically digital editions. In particular, it pushes students to analyze whose voices are represented in these collections and how? Whose voices are absent and why? Using the subject tags, metadata, transcriptions, exportable data, and annotations available at the Civil War & Reconstruction Governors of Mississippi Project, students will critically evaluate the materials in the collection and how they are made available and accessible, and the significance of this process. In the process, this discussion will demonstrate how the humanities can inform ethical data practices and conversely, how data can be a useful but potentially problematic tool within the humanities.

The [Civil War & Reconstruction Governors of Mississippi Project](#) (hyperlinked) or CWRGM is a digital edition that is digitizing over 20,000 documents sent to the governors of Mississippi during the American Civil War and Reconstruction (1859–1882). Americans from all backgrounds wrote to the state’s governors with their concerns, and after 1865, this includes African American authors. With NEH/NHPRC-funding, these records are made freely-available online with high-quality images of the original documents alongside transcriptions, which also feature annotated subject tags. Document metadata, in-text subject tags, and annotations are

designed to increase the discoverability of the people within the documents, but they also generate massive amounts of data for research and classroom use. This [data](#) (hyperlinked) is made freely available at the website and is updated every six months. For greater context about these materials and the data included in this lesson, please read "[How to Evaluate Online Resources using a Digital Edition](#)," a blog post authored by Lindsey R. Peterson.

The documents housed at CWRGM and the methods their research team employs to increase discoverability within these collections creates fruitful opportunities for students to think critically about how data is produced, the ethical implications of that data, and the data's limitations. In this lesson plan, students will explore the website and reflect on data production and its ethical and scholarly consequences by exploring the data associated with three subject tags: African Americans, Mental illness, and Impoverished people.

Materials:

CWRGM Datasets:

- Students can access cooccurrences for any subject tag in the collection by contacting the CWRGM editors. Therefore, we suggest downloading the cooccurrences for the following subject tags, which CWRGM editors have made available here:
 - African Americans [subject tag](#) (hyperlinked) / [cooccurrences file](#) (hyperlinked)
 - Mental illness [subject tag](#) (hyperlinked) / [cooccurrences file](#) (hyperlinked)
 - Impoverished people [subject tag](#) (hyperlinked) / [cooccurrences file](#) (hyperlinked)
- "CWRGM Subject Index" file on the [Data page](#) (hyperlinked)

Other Materials:

- Humanities & Data Literacy Exploration worksheet (included below)
- Computer for online research

Procedures:

1. Depending on your classroom, you may have students explore all three files independently, in groups, or assign individual students or groups a single file to explore. Based on your work division, instruct students to download the cooccurrences files for the following subject tags:
 - a. African Americans [cooccurrences file](#) (hyperlinked)
 - b. Mental illness [cooccurrences file](#) (hyperlinked)
 - c. Impoverished people [cooccurrences file](#) (hyperlinked)
 - d. Other subject tags can be used, as these are just a few of the thousands of subject tags featured at the website. These three, however, offer large quantities of data that reveal problematic trends that offer opportunities for student learning and are readily available. To obtain other subject tag cooccurrences files, you must [email](#) (hyperlinked) the CWRGM research team ahead of time due to technological constraints on their system.

- i. Issue for students to consider: This data was provided here by CWRGM editors, but emailing the CWRGM editorial team for occurrence data creates a barrier to access. Discuss.
2. Introduce students to the methods CWRGM uses to create that data:
 - a. Depending on the level of students, the instructor may explain this process, but instructors are encouraged to let students explore the data and website to try and piece together what a cooccurrence is and how its data is created.
 - i. Prompt: Discuss how easy or challenging it was to find the data and methods on the project website and why this matters.
 - b. Using an internally controlled vocabulary list, when project editors identify language that falls under one of their nine subject tag categories, they link the original language in the document to their controlled vocabulary term (or tags them with a subject). For example, various references to enslaved people (i.e. servants, colored, servile population, slave, etc.) all are tagged with the controlled subject term "[African Americans–Enslaved people](#)" (hyperlinked). A term is added to a document no more than once.
 - i. Prompt:
 1. What benefits are created by subject tagging in a collection? How can they aid users?
 2. How can they make records pertaining to historically marginalized people more discoverable?
 3. What are some issues that may arise?
 4. Potential issues for students to consider:
 - a. While every document is drafted and goes through two more stages of review by independent researchers per CWRGM's [tagging protocols](#) (hyperlinked), subject tags are applied by humans and can be missed and misapplied in the collection.
 - b. The boundaries of every subject tag are arbitrary and authors' meaning is not always clear. Sometimes tags can be added that don't belong there or should have been added when they do apply. For example, how do you apply the African Americans tag when the subject is biracial?
 - c. Subject tagging creates an index of all of the documents in the collection that refer to [enslaved people](#) (hyperlinked). A cooccurrences file, then, exports all of the other subject tags that were also tagged in documents that received the [African Americans–Enslaved people tag](#) (hyperlinked). This file allows users to see what other terms are most commonly discussed alongside that subject tag.
 - i. Issue for students to consider: Subject tags are not nested. So documents that are tagged with the "African Americans–Enslaved people" tag, will not show up under the "African Americans" subject tag.
 1. Discuss solutions: This data can, however, be collated together.
 2. Prompt: What is the value of separating the "African Americans" and "African Americans–Enslaved People" subject tags in the

collection? Students may discuss the value of analyzing the distinct experiences of African Americans *within* slavery and in general, for example.

3. Using the cooccurrence data files for the African Americans, Mental illness, and Impoverished people subject tags, have students complete the Humanities & Data Literacy Exploration worksheet questions (provided below). The instructions listed here can be copied into the worksheet for independent work, or the instructor can collectively guide the class through the steps orally and using a projector system.
 - a. Instruct students to sort the data in a spreadsheet application according to different columns and discuss what jumps out at them in the process.
 - b. Instruct students to sort each data file according to the “Subject Title” field (column A). Then browse through the Subject Title (column A) field and identify the subject tags that most commonly appear alongside the respective subject tag (such as the Mental Illness tag).
 - c. Prompt: Brainstorm the potential opportunities and issues created by the data association.
 - i. Prompt: What terms appear most commonly? What opportunities and possible issues does this present?
 - ii. For example, students may discuss that the African Americans subject tag shows a high correlation with tags associated with criminality. This presents an issue as it highlights and can even reinforce long standing stereotyping of Blackness with deviance and criminality. Conversely, it shows that criminal records and criminal procedure may be a fruitful avenue for exploring Black experiences during the war and Reconstruction, leading students to learn more about convict labor systems and racially based adjudication of the law.
 - d. Next have students download the “CWRGM Subject Index” file on the [Data page](#) (hyperlinked) and sort according to the “Subject” field (column H).
 - i. Students should then find whichever of the three subject tags they are working with (African Americans, Impoverished people, and Mental illness) in the Subject field. They can do this by highlighting the Subject field column and then use a find command for the name of the subject tag.
 - ii. Next, copy and paste the data associated with all instances of their subject tag/s to individual tabs.
 - iii. In each tab, have students sort the data based on the “Text” field (column I) and browse what type of language was used to describe the group of people represented by that subject tag.
 1. Prompt: What kinds of language gets linked with the subject tag? How are the people in the documents represented here? (By name, group identity, slur, etc.)
 2. Prompt: Have students discuss why this representation matters and how it makes them feel.

3. Issue for students to consider: People are often reduced to a group identity in the collection. Where men, especially white men, are readily identifiable by name, women, people of color, and other historically disempowered people are less likely to be identified by their entire name (or a name of their choosing in the case of enslaved people) or are only identified by external factors such as their race, ethnicity, ability, marital status, etc.
- e. Direct students to the CWRGM website's pages for the same subject tags:
 - i. African Americans [subject tag](#) (hyperlinked)
 - ii. Mental illness [subject tag](#) (hyperlinked)
 - iii. Impoverished people [subject tag](#) (hyperlinked)
 - f. Then have students explore the documents connected to those subject tags.
 - i. Prompt: Can the letters help flesh out the stories, lived experiences, and humanity of the people in the collection where the data could not?
 - ii. Prompt: What are the limits of this and what resources can we turn to to address those limits?
 - iii. Prompt: How does shape what we know about history? And how does this make them feel?
 - iv. Issue for students to consider: Data can help point us in fruitful directions but it contains bias and can be used in harmful ways. Turning to the documents allows researchers to discover more about the people in the collection: their names, feelings, experiences, etc. But there are also limits to what the documents can uncover as well, and some groups (like those linked with the Mental Illness subject tag) are still not recognizable by name.

Humanities & Data Literacy Exploration Worksheet

Data Production & Methods:

1. What kinds of data does the project produce?
2. How is this data created?
3. How open and accessible is the data to the public?
 - a. How does this make you feel?

Data's Humanity:

1. How are people represented in the data? (By name, group identity, map marker, something else?)
 - a. How does this make you feel?
2. How does this negatively shape what we can learn about and how we view the people in the collection?
 - a. How does this make you feel?
3. How does this positively shape what we can learn about and how we view the people in the collection?
 - a. How does this make you feel?

Comparing Quantitative and Qualitative Data:

4. What are the differences between the kinds of information found in the spreadsheets created by subject tag indexes and cooccurrences and the letters in the collection?
5. How can we use the documents in the collection to address some of the collections' data limits?
 - a. Identify one document for each subject tag that demonstrates these possibilities.