NBER WORKING PAPER SERIES

# SEPARATION OF CHURCH AND STATE CURRICULA? EXAMINING PUBLIC AND RELIGIOUS PRIVATE SCHOOL TEXTBOOKS

Anjali Adukia
Emileigh Harrison

Working Paper 33791
http://www.nber.org/papers/w33791

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2025

For helpful feedback, we thank Ben Arold, Elliott Ash, Sam Bazzi, Marianne Bertrand, Dan Black, Ethan Bueno de Mesquita, Judy Chevalier, Michael Dinerstein, Steven Durlauf, Alex Eble, Martin Fiszbein, Jeff Grogger, Sendhil Mullainathan, Heather Sarsons, Thomas Wollmann, David Yanagizawa-Drott, Hiro Yoshikawa, and seminar participants at AEFP, APPAM, Boston University, CESifo, Columbia, Harvard, London Business School, RAND, Stanford, SREE, UChicago, UCLA, Univ. of Michigan, UPenn, Warwick, BFI AI in Social Science conference, Yale, and the Zurich AI/Economics conference. For exemplary research support, we especially thank Matt Bonci. We also thank Sam Abers, Elisa Xi Chen, Paula Dastres Gallardo, Benjamin Leiva, Jake Nicoll, Celia Zhu, and other members of the MiiE Lab for excellent research assistance. For financial support, we thank the National Academy of Education, Spencer Foundation, and Becker Friedman Institute at UChicago. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200478 to the University of Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2025 by Anjali Adukia and Emileigh Harrison. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Separation of Church and State Curricula? Examining Public and Religious Private School Textbooks

Anjali Adukia and Emileigh Harrison

NBER Working Paper No. 33791
May 2025
JEL No. I0, I20, I21, I28, I29, J0, J10, J15, J16, J18, J19, Z10, Z11, Z12, Z13, Z18

## ABSTRACT

Curricula impart knowledge, instill values, and shape collective memory. Despite growing public funding for religious schools through U.S. school choice programs, little is known about what they teach. We examine textbooks from public schools, religious private schools, and home schools, applying computational methods—including AI tools—to measure the presence and portrayal of people, topics, and values over time. Despite narratives of political polarization, our findings reveal few meaningful differences between public school textbooks from Texas and California. However, religious school textbooks have less female representation, feature lighter-skinned individuals, and portray topics like evolution and religion differently. Over one-third of pages in each collection convey character values, with a higher proportion in religious school textbooks. Important similarities also emerge: all textbook collections rarely include LGBTQIA+ discussion, portray females in more positive but less active or powerful contexts than males, and depict the U.S. founding era and slavery in similar contexts.

Anjali Adukia
University of Chicago
Harris School of Public Policy
and NBER
adukia@uchicago.edu

Emileigh Harrison
University of Chicago
Harris School of Public Policy
harrisone@uchicago.edu

# I Introduction

Our collective memory and understanding of the world are shaped by a variety of forces including the stories we share (Michalopoulos and Xue, 2021; Adukia et al., 2023), the narratives propagated through popular culture (Riley, 2022; Michalopoulos and Rauh, 2024), as well as the monuments we build and the history they represent (Bazzi et al., 2023). These elements not only influence *what* is remembered but also determine *who* is remembered and *how*. Education plays a pivotal role in this process with curricular materials serving as critical sites of learning. These materials do more than convey academic knowledge; they also instill children with societal values such as responsibility, kindness, and citizenship. Consequently, changes in educational content can shape children's worldviews, influencing their perspectives both explicitly and implicitly through the discussion of certain topics and values, who are represented, and how they are portrayed (Cantoni et al., 2017; Arold et al., 2022; Arold, 2024; Riley, 2022).

The role of education in shaping belief systems has historically empowered religious organizations to engage in educational sectors around the world, shaping both curricular content and the educational landscape (Bazzi, Hilmy and Marx, 2020). In the United States, this dynamic is evident in the recent explosion of school choice programs, which allow parents to use public funds to enroll their children in schools that they feel better align with their values or religious beliefs. However, the lack of regulations regarding curricula means we know very little about what children are taught in religious educational settings and how it compares to what children are taught in public schools. Given the growing allocation of tax dollars to these alternatives through vouchers and other school choice mechanisms as well as the potential impact on children's development, it is important to understand the educational content to which children are exposed across educational settings so that we can make more informed decisions about where to direct scarce resources such that they align with the broader goals of our society.

Curriculum takes three primary forms: curriculum as written, curriculum as implemented, and curriculum as discussed. In this study, we focus on the former: curriculum as *written*, and specifically, the curriculum as it is presented through textbooks. In this study, we examine textbooks commonly used across a variety of educational settings in the United States. We assess state-adopted public school textbooks in California and Texas, states that serve the largest student populations in the U.S., which allows us to understand the content being provided to a large portion of children in the U.S. and how educational materials vary across different political landscapes. We also examine curricular materials often adopted in religious private school and home school contexts: those produced by conservative Christian

publishers Abeka, Accelerated Christian Education (ACE), and Bob Jones University Press (BJU Press), which allows us to understand how educational materials in religious settings compare to public school materials. We focus our analysis on textbooks for the foundational subjects of Reading, Science, and Social Studies and those intended for third and fifth grade students to understand the messages being sent to primary-school-age children.

We apply computational social science methods including artificial intelligence tools to examine three main categories of content analysis: First, we analyze *what* topics and values are being included. Second, we examine *who* are highlighted in these lessons, across race, gender, skin color, sexuality, birthplace, and occupation. Third, we explore *how* these groups of people, topics, and values are portrayed. We then examine how these content measures differ across educational settings and over time.

Overall, our analysis reveals meaningful parallels between the public school collections, while our collection of religious school textbooks differ notably, featuring less female representation, characters with lighter skin, more famous White individuals, and differential portrayal of topics such as evolution and religion, including in science textbooks. Important similarities, however, also emerge: for example, each collection portrays females in contexts that are more positive but less active and powerful than males, discusses Christianity more than other religions, rarely discusses historical or famous individuals who are Asian, Indigenous, or Latine, who were born outside of the North America and Europe, or who identify as lesbian, gay, or bisexual, and depicts the U.S. founding era and slavery in similar contexts.

To examine *what* is being communicated, we employ a large language model to summarize the topics and values present on pages in each collection. We find similarities in the top topics across public and religious school curricula which includes topics such as family, nature, friendship, biology, geography, agriculture, and astronomy. Each collection also discusses the U.S. founding era and slavery similarly. All three collections contain character values in more than a third of pages with a slightly larger proportion of pages classified as having these values in religious school textbooks, though there is meaningful heterogeneity in the specific character values espoused. In contrast to public school textbooks, religious school textbooks place a significantly greater emphasis on faith and religious values, including in Science textbooks. It is only in religious school textbooks that creationism is discussed, and that is often in conjunction with discussions related to evolution. In terms of geographical representation, children have the most exposure to the United States with little introduction to other parts of the world. The public school collections are more likely to have state-specific knowledge.

To understand *how* different topics are portrayed, we use both word and page embeddings to identify the top words associated with pages containing specific topics by collection. We find that the top words closest to topics relating to the Civil War, slavery, immigration, climate change, United States, and Native Americans are similar across our collections of public school and religious school textbooks. For example, each collection discusses slavery in the context of resistance, abolition, and colonies. However, differences emerge between public and religious school textbooks when examining religious topics and the discussion of evolution. In public school textbooks, religious topics are often discussed in the context of the founding era, whereas in our collection of religious school textbooks, religious topics are discussed in the context of god or Christianity. Similarly, the topic of evolution is often discussed in contexts related to creationism in our collection of religious school textbooks, but more in contexts related to adaption in public school textbooks. Though both California and Texas textbooks discuss evolutionary concepts such as adaptation and natural selection, Texas textbooks never explicitly mention the theory of evolution or the Big Bang theory.

We also see very little discussion of topics related to LGBTQIA+ issues. Indeed, it is only in two textbooks across all collections where we see any pages classified as talking about LGBTQIA+ topics. In the California collection, one passage highlights several notable figures from California, including Harvey Milk, a pioneering political leader who was assassinated, and Chaz Bono, a transgender activist, both discussed in the context of their advocacy for LGBTQIA+ rights. The second place where we see mention of LGBTQIA+ topics is in the Religious collection, in which LGBTQIA+ rights are presented as a threat to religious rights. This minimal discussion is in contrast to the growing narrative that children are being exposed to non-heterosexual or non-cisgender messages in schools.

When examining *who* is represented, we analyze the pictured individuals shown in images as well as names, pronouns, and identity-specific words mentioned in the text. For images, we first use a face-detection model described in Szasz et al. (2022) to identify the faces of all pictured characters (both in photographs as well as in illustrations). We then classify skin color by first segmenting the facial area of interest and then calculating a continuous measure of its darkness or lightness based on each pixel's color in the segmented skin. Separately, we predict the putative race and gender of each face using a feature classification model described in Szasz et al. (2022). From the text, we examine the presence of different groups of famous people by race, gender, sexuality, occupation (e.g. whether they were a religious figure or Founding Father), and birthplace. We also assess text-based gender representation by examining the proportion of gender-specific words used, including gendered pronouns, gendered terms, famous individual names, and other mentioned character names.

We see similar presence of females and males in Texas and California public school textbooks. However, in religious school textbooks, we see less female representation in both the images and text. In particular, we observe unequal gender representation in the Science and Social Studies textbooks across all collections, likely due to historical underrepresentation and lack of recognition of women in scientific fields and positions of power. Conversely, public school Reading textbooks show equal gender representation, presumably because these textbooks have greater flexibility in the stories they choose to present. However, this gender equality is not mirrored in the Reading textbooks from the Religious collection.

Across all three collections, most pictured characters and famous individuals mentioned are White, the second most pictured or mentioned group are Black individuals, with very low presence of Asian, Indigenous, and Latine individuals. We see that California and Texas textbooks picture characters with similar skin tones on average over time. Within religious school textbooks, however, we see that pictured faces are depicted with lighter skin colors on average. We also see that skin colors in religious school textbooks have gotten lighter in the last three decades, but they have remained relatively consistent in public school textbooks. A vast majority of the famous people mentioned in the public school textbooks were born in North America, in the religious school textbooks the vast majority were born in either North America and Europe, with minimal representation of individuals born in Africa, Asia, Australia, or South America across all collections. The most commonly included individuals in each collection are all identified as being heterosexual with very little inclusion of individuals who identify as lesbian, gay, or bisexual.

To understand *how* different identities are portrayed, we use word embeddings to characterize the context surrounding different groups of people. We match these words or domain vectors of words to a lexicon measuring sentiment as it pertains to valence (how positive or negative a word is), arousal (how active or passive a word is), and dominance (how powerful or weak a word is). In particular, we utilize word embedding association tests (WEAT) to gauge the proximity of these groups to various words or domains (groups of words related to specific topics including politics and family life). To facilitate this analysis, we introduce a novel placebo test to evaluate whether the word embedding associations we find are meaningfully large. Additionally, we apply semantic role labeling to discern and characterize the sentiment of actions attributed to different groups of people.

Despite differences between who is represented between public school and religious school textbooks, we see similar patterns across the three collections when examining how different marginalized groups are portrayed when they are present. All three collections describe females more often in the context of the family and the home compared to males, and

describe males more often in connection to politics and tools compared to females. The top words more associated with females and least associated with males predominantly revolve around appearance, family roles, or household duties such as sewing or groceries. By contrast, the top words associated with males and least associated with females revolve around military/fighting and leadership. In comparison to males, words and actions associated with females are more positive, but less active and less powerful.

Our study contributes to a vast literature focused on the content analysis of children's literature and educational materials (Wade, 1993; Allen, Allen and Sigler, 1993; Clark, 2007; Koss, 2015; Braden and Rodriguez, 2016; Bickford and Knoechel, 2017; Crisp, Gardner and Almeida, 2018; Stevenson and Zlotnik, 2018; Koss, Johnson and Martinez, 2018; Koss and Paciga, 2020; Lucy et al., 2020; Adukia et al., 2022a,b, 2023). The innovative uses of computer vision and natural language processing tools we develop and apply throughout our analysis contribute to the computational social science literature in a manner similar to Caliskan, Bryson and Narayanan (2017), Garg et al. (2018), and Kozlowski, Taddy and Evans (2019). These methodological advancements can have meaningful implications for expanding the broader field of computational social science and the scope of research inquiry that can be pursued.

The paper proceeds as follows: Section II provides background and context related to school choice in the United States. Section III discusses prior textbook content analyses. Section IV describes the textbooks in our sample. Section V explains our image and text processing tools. Section VI presents our findings. Section VII concludes. The appendices contain further information about the methods and additional exhibits presenting findings.

## II Public Education, Parental Values, and the Expansion of School Choice

Knowledge and values are transmitted to children through two primary pathways: vertical transmission from parents and family, and horizontal transmission from society and peers. Vertical transmission occurs from the beginning of a child's life in their home, and public schools have traditionally served as a key vehicle for horizontal transmission, imparting not only academic knowledge but also broader societal values. When parental preferences differ enough from those of mainstream society, they may choose to have greater control over their children's education and primarily transmit knowledge and values either at home through home schooling or in a private school whose values are perceived to be more closely aligned with their own.

Changes in public educational content and instruction have been shown to have a causal impact on children's worldviews and later life outcomes. In particular, changes in high

school politics curricula can affect students' views on democracy and free markets (Cantoni et al., 2017) and changes in public school content standards related to evolution can affect students' probability of choosing a STEM career (Arold, 2024). Similarly, messages about the abilities of others with similar racial and gender identities can impact children's academic performance (Riley, 2022). Consequently, parents' decisions about their children's education are driven by the values, cultural norms, and particular sets of knowledge or values they want to instill in their children.

Given the importance of educational content, debates about what should be taught—and what should be excluded—from public school curricula have been going on for centuries. In U.S. colonial times, schools were deeply intertwined with Christian instruction. However, as the United States developed a distinct national identity, tensions emerged between religious influence in public institutions and the ideal of a secular state. Figures like Thomas Jefferson championed religious liberty through the idea of a "wall of separation between church and state," arguing that public institutions should not promote religious doctrines. This principle has continued to shape legal battles in education, including debates over school-sponsored prayer (e.g. *Engel v. Vitale*) and the teaching of evolution (e.g. *Epperson v. Arkansas*, *Tennessee v. Scopes*). More recently, curricular debates have expanded beyond religion to include questions of how history and identity are taught in schools. Some states, like California, have embraced inclusive standards, such as the 2011 FAIR Education Act, which mandates broader representation of ethnic, gender, LGBTQIA+ groups, and persons with disabilities in history instruction. By contrast, other states have introduced "educational gag orders," seeking to restrict discussions of race, gender, and American history (Friedman and Tager, 2022). Notably, these restrictions apply only to public schools; private schools and homeschooling families face minimal or no curricular oversight.[1]

Faced with either dissatisfaction over academic standards or misalignment between their values and values prioritized in public education, many parents are increasingly turning to alternative school choice options, including private and home schools through the use of vouchers (EdChoice, 2023). However, evidence on the academic benefits of different school choice programs is mixed. Some research has found that academic improvements can occur, but only when students gain access to higher-quality schools (Deming et al., 2014). However, Abdulkadiroglu, Pathak and Walters (2015) find that parents in more rural settings are using private school vouchers to move their children to lower quality schools as measured by

[1]Most states require private school curricula to be "similar" to public standards, but 18 states either have no requirements or only require that private schools teach students the state constitution and state history. While states often have similar curricular requirements on paper for homeschoolers as they do for private schools, there is rarely any enforcement for homeschooling families (Bartholet, 2020).

students' academic performance. These findings suggest that many parents are motivated by factors beyond academics, particularly the desire for educational environments that are better aligned with their values. Survey data support this hypothesis: in a survey of parents with children in private schools in Indiana, 80% listed character or values instruction as a top desired school quality (Catt and Rhinesmith, 2017). A desire for religious values instruction is reflected in private school enrollment patterns: most students educated outside the public school system are educated in a religious school environment. Indeed, in 2017, among the 5.8 million children who were enrolled in a private school, 76% were in a religious school, with 49% enrolled in a conservative Christian school (PSS, 2015). 75% of parents who chose to educate their children at home listed a desire to provide moral instruction as a top reason why they choose to homeschool (National Center for Education Statistics, 2022). Pre-COVID estimates of the number of homeschoolers who identify as being religious, or for whom religion is a primary motivation for homeschooling, ranged from over 50% to 90% (Kunzman, 2010; Bartholet, 2020). Since the onset of the COVID-19 pandemic, the number of reported homeschoolers has increased (Musaddiq et al., 2022).

The content in school textbooks plays an important role in shaping collective memory and a shared narrative. While academic content certainly influences parents' school choice decisions, it is often not the sole or even most important factor. Given the role of educational content in shaping societal values and future citizens and the rising allocation of tax dollars to religious private schools or homeschooling families, it is critical to understand what children are being taught across different educational settings in order for families and taxpayers to make informed decisions.

## III Prior Textbook Analyses

Our study complements existing qualitative and quantitative work examining the content of curriculum in the United States and around the world. Much of this work focuses on the representation of gender and race in educational materials. For example, Crawfurd, Saintis-Miller and Todd (2024) analyze 1,255 English-language textbooks from 34 countries and find that women tend to be portrayed in stereotypical gendered roles. Lucy et al. (2020) also uses NLP tools to examine who is present in the text of a sample of 15 Texas high school textbooks used between 2015 and 2017 and find that Latine people are rarely discussed, the most common famous figures are nearly all White men, and that women tend to be discussed in the contexts of work and the home. Another related paper by Stevenson and Zlotnik (2018), focused on college-level introductory economics textbooks most commonly purchased in the U.S., finds that fictionalized women are more likely to be involved in food, fashion, or household tasks.

A smaller body of research examines the values transmitted through textbooks (c.f. Sallabaş 2013 which conducted a qualitative analysis of the character values present in Turkish textbooks). Even fewer studies examine textbooks commonly used in religious settings. The most relevant is a qualitative analysis by Scaramanga and Reiss (2018) on curriculum published by Accelerated Christian Education (ACE)–one of the textbook publishers we include in our study–which found that churches and schools are often depicted as racially segregated in their materials.

Finally, most of the previous work discussed above focuses solely on the text but largely ignores the images shown in textbooks. One exception is Deckman et al. (2018), who manually examined 1,468 images from U.S. elementary and middle school health textbooks to study the portrayal of race, gender, and sexuality. Despite finding substantial *presence* of gender and racial diversity, they saw that women and people of color were often portrayed in stereotypical roles, and no depictions of same-sex couples appeared in any of the images.

Our study expands on the textbook analysis literature by applying computational content analysis to both the images and the text of 261 primary school textbooks used in public and religious educational settings from 1980 to the present. Previous studies have typically limited their analyses to cross-sectional examinations of textbooks in a single setting. Our panel of textbooks affords a unique opportunity for longitudinal analyses across different educational settings and over multiple decades. Our broader set of tools enables new analyses, such as examining the representation of skin color in textbook images, an important but often overlooked dimension of human categorization (Keith and Monroe, 2016). Finally, we systematically measure the presence and portrayal of people, topics, and values over time in a single study. Our approach enables us to replicate qualitatively similar results found in previous work while also documenting new patterns. Our methods advance both image and text analysis in the computational social science literature along with recent papers by Adukia et al. (2023) which examines children's books, Ash et al. (2021) which examines newspapers, Yanagizawa-Drott and Voth (2023) which examines yearbooks, and Kamel and Woo-Mora (2023) which examines sports statistics websites.

## IV Data Sources

We draw data from three distinct sets of elementary school textbooks used in: (1) Texas public schools, (2) California public schools, and (3) religious private and home school settings. We digitize a novel data set of Reading, Science, and Social Studies textbooks written for the third and fifth grades. Our data set comprises over 75 thousand pages originating from 261 different textbooks, amounting to a corpus of approximately 16 million words and 60 thousand detected faces (Table 1). In Figure A.1, we show how many textbooks

in our sample were in use in a given year by subject, grade, and collection.[2]

### IV.A Texas State-Adopted Textbooks

We study Texas public school textbooks for four primary reasons. First, Texas adopts textbooks at the state level which means we can construct a panel of textbooks that children were likely to have been exposed to over time. Second, Texas serves one of the largest student populations in the U.S., so the messages that are sent in these books are likely to reach a broad audience. Third, because they are such a large share of the textbook publishing market, Texas has historically influenced the content of textbooks used across the United States. Finally, Texas is often portrayed as reflecting conservative values, both politically and in the media, which provides a view into how one side of the ideological spectrum presents education in the public sphere.

The state of Texas creates a list of state-adopted textbooks for all K-12 grades which is updated yearly. Before 2011, Texas public schools were only allowed to use funds to purchase books from this list. Starting in 2011, this law changed to allow schools to use funds to purchase any textbooks, but most schools continued to purchase textbooks off the state-adopted list. In light of this, we first digitized a convenience sample of approximately 50% of all textbooks on the yearly state-adopted textbook list in use between 1985-2010. Then, for the years 2011 through 2021, we focused our book digitization on the top three textbooks purchased by school districts each year for each relevant grade and subject (reliable textbook purchasing data were not collected before 2011).

### IV.B California State-Adopted Textbooks

We study California public school textbooks for various reasons. Similar to Texas, California also adopts elementary school textbooks at the state level and has a large influence on the textbook publishing market due to the size of its student population. It is typically associated with more progressive policies and is often posed as the political or substantive converse of Texas. Having these two political extremes in our sample will allow us to bound the most conservative and the most liberal ideals of traditional public school curriculum.

California curates a list of state-adopted textbooks for grades K-8 and updates it every couple of years for each subject.[3] We digitize the universe of books in the relevant grades and

[2]Note that our sample is not perfectly balanced across subjects, grades, and year in use across collections. Because of this, when comparing average measures across collections, we make sure to first calculate the measure for each collection, subject, grade, and year in use and then average these measures within a collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages. This approach allows us to attribute observed differences across collections to actual differences in content, rather than to shifts in the composition of the sample over time.

[3]Note that during the economic crisis of 2007, California did not make any changes to their state-adopted

subjects which were adopted in California for all years in which the list of adopted materials are available. This includes adopted textbooks beginning in 1999 for Social Studies, 2000 for Science, and 2002 for Reading and ending with the books that were in use as of 2022.

### IV.C Textbooks by Religious (Conservative Christian) Publishers

A third corpus that we examine comprises religious school textbooks published by three ideologically conservative Christian textbook companies: Abeka, Bob Jones University Press (BJU Press), and Accelerated Christian Education (ACE). We refer to this set of books as our religious school collection. These textbooks are commonly used by homeschooling families and adopted by over a third of non-Catholic Christian schools receiving tax-payer funding (Klein, 2017). We focus on textbooks published by these three Christian publishers both because of their wide adoption, but also because there are many taxpayer dollars funding schools which use these curricula. Notably, in Indiana, approximately 4,240 students received over $16 million in scholarships to attend schools that use the Abeka or BJU Press curricula, according to 2016-2017 figures from the Indiana Department of Education (Klein, 2017). We can then understand how the content in commonly used religious school curriculum compares to the public school content in both the politically conservative and liberal state contexts of Texas and California.

There are also few comprehensive curricular options that families can easily adopt when they choose to homeschool their children. These curricula conveniently provide such options for families. One can simply go online and purchase an entire grade's worth of curricular materials from one of these publishers or from second-hand resellers. Using Google Trends search data as a measure of interest in these topics (Figure 1), we see that there was steady interest in *Abeka* over time, but after the onset of the COVID-19 lockdowns in March 2020, there was a marked increase in searches for *Abeka*. We also see that there was a steady decline in search interest for *homeschool* over time, but then it spiked once schools shut down and then began to track the search interest for *Abeka*. This is consistent with evidence that since the onset of COVID, the number of parents choosing to homeschool their children has increased substantially (Musaddiq et al., 2022; Jamison et al., 2023).

We digitize each textbook published by these three publishers for the grades and subjects in our sample starting with books published in 1980 up until 2022, including updated editions of the same textbooks. Note that most of the curriculum published by ACE consists of workbooks called PACEs (self-guided Packets of Accelerated Christian Education) instead of textbooks, so we only include books from their Reading curriculum in our analysis.

textbook lists to save districts the expense of purchasing new textbooks.

## V Methods

In order to analyze large corpora of educational materials, we leverage tools from the computational social sciences, including artificial intelligence methods, which allow for a cost-effective and systematic approach to content analysis.

We focus on three main categories of content analysis: *What*, *Who*, and *How*. We describe the image and text analysis methods that we use for this exercise below and present our overall methods pipeline in Figure 2.

***What* is being communicated.** To measure *what* these textbooks cover, we use a large language model to identify topics and values conveyed in the text on a given page. We also use token (word) counts to measure whether and how often specific words are mentioned. Finally, we use named entity recognition (NER) to identify the geographies and historical events that are referenced.

***Who* is depicted.** We measure *who* is highlighted in the images and text. To examine the faces in photographs and illustrations in the textbooks, we use a face detection model to identify each pictured individual. We then classify skin color and predict features to measure how often people of different putative races, genders, and skin colors are represented in images. To measure who is highlighted in the text, we use token counts and named entity recognition to measure how often gendered words (including names, pronouns, and terms) are mentioned as well as how often famous people of different races, genders, sexualities, occupations, and nationalities are represented.

***How* content is portrayed.** To measure *how* different people, topics, and values are discussed, we use word embeddings to identify words found in similar contexts to them and semantic role labeling to identify the actions that people of different identities are described as taking. We then analyze the sentiment associated with these associated words and actions.

### V.A Images As Data

The process that we use to convert images into data is illustrated in Figure 2a. In this section, we describe how we use computer vision tools to detect faces, classify skin color, and predict features such as putative race and gender.

#### V.A.1 *Images as Data: Face Detection*

Our first step when measuring representation in images is to run face detection on every digitized page in our sample of textbooks to identify all depicted faces. While off-the-shelf face detection models are the state-of-the-art way to detect faces in photographs, they often underperform when applied to illustrations. Since many of the images found in

content targeted towards children contain illustrations, we use a face detection model trained to better detect faces in both photographs and illustrations as described in Szasz et al. (2022) and Adukia et al. (2023). This model increased the number of detected faces by 250% in their sample of books. We apply this model to detect faces in the textbooks in our sample.

### V.A.2 *Images as Data: Skin Color*

Skin color is one of the first ways that children learn to distinguish humans, and is distinct from measures of putative race but is highly related to racial classification (Adukia et al., 2025). To extract and classify skin color, we begin by identifying the facial area of interest in the pictured face by "segmenting" the skin. To do this, we use a face landmarker model,[4] which uses convolutional neural networks to predict and detect facial landmarks and then to isolate the parts of the face which contain skin in order to remove hair, clothing, and other accessories (known as "skin segmentation"). Then, we extract the dominant colors by grouping pixels of the segmented skin (based on their position in RGB color space) into clusters using $k$-means clustering. This groups each pixel into $k$ clusters that share the most similar RGB color space, and each cluster is represented by the RGB value of its center. We remove the smallest two clusters, which tend to contain irrelevant information such as shadows or highlights. We then select the centers (or centroids) of the remaining clusters–which represent the most common skin-like colors–are selected. These centroids are then treated as the dominant skin tones of the detected face. Finally, we collapse the remaining segmented skin into a single "representative color" by taking a weighted average of the dominant skin colors in CIE L*a*b* color space (where the weights are determined by the size of the cluster from which each dominant color was drawn). To construct a continuous measure of how dark or light the skin color of a detected face is, we use the L* value of this representative skin color which ranges from 0 (darkest) to 100 (lightest) as our continuous measure of skin tint that serves as our measure of skin color. This process follows a similar method as described in Adukia et al. (2023) and is illustrated in Figure 2a.

### V.A.3 *Images as Data: Feature Classification of Putative Race, Gender*

We then wish to classify different features of pictured characters, specifically putative race and gender. We use a feature classification model that we trained using a manually curated data set of both illustrations and photographs which can more accurately predict features of illustrated faces such as the race and gender, as described in Szasz et al. (2022). We use this improved feature classification model for our analysis.

[4]Specifically, we use Google's MediaPipe Face Landmarker.

### V.B Text As Data

We also examine the printed text, which is another important source of content in addition to images in textbooks. The process that we use to convert text into data is presented in Figure 2b.

#### V.B.1 *Text as Data: Large Language Models*

One way we measure *what* topics and values are included in the text is through the use of large language models (LLMs). LLMs are deep learning models that recognize, summarize, translate, and generate text through predictions. LLMs acquire these abilities by using substantial quantities of data and computational resources during their training to achieve general-purpose language understanding and generation.

For our analysis, we use OpenAI's GPT-4 model (GPT), which is an autoregressive pre-trained LLM that applies deep learning algorithms to generate what would be a human-like response to user input. This is done through the use of a transformer (i.e., neural networks) architecture, comprising encoder and decoder models. The encoder part takes the input given by the user, extracts its relevant features, "understands it" through its context, and creates a high-level representation of it. This representation is then passed as an input to the decoder model, where the final response is generated using information learned in the pretraining process. In essence, it uses past data to predict future information.

We use GPT to analyze single pages of textbooks and identify topics and values present in them.[5] It should be noted that LLMs use the text on pages to generate *predictions* using past information to predict future values. It never represents *ground truth*, which then may lead to measurement error. Moreover, because LLMs provide predictions as opposed to ground truth, results are not always exactly the same each time the model is deployed, which makes it more difficult to precisely replicate results. One alternative way to measure substantive topics is to measure the instance of topic-related words on a given page.

We validate the model's topic classification by counting how often a set of topic-related words appear on pages that were (or were not) classified as containing that topic by GPT. We find that pages which were classified as having a specific topic were much more

[5]We set the temperature of the model to be 1 to strike a balance between allowing for broader range of relevant topics and interpretations while still ensuring that output was grounded in the content of the textbook pages. The parameter *temperature* defines the freedom or creativity we allow the model to have for its responses. The optimal temperature value depends on the type of task we ask GPT to do, and to the best of our knowledge there is not a benchmark or standard for this. While higher temperature values give the model more room to hallucinate (create fake or made-up responses), lower ones make responses more deterministic, potentially limiting its flexibility to provide reasonable responses. We assign a system role to guide the model's behavior, instructing it to act as "a knowledgeable and unbiased judge."

likely to contain words related to that topic (Figure C.1). For example, pages that were not classified as containing topics related to "Slavery" have on average 0.0 words related to slavery such as captivity, enslavement, slave, plantation, Underground Railroad, Emancipation Proclamation, and Thirteenth Amendment. On pages that GPT classified as having the topic "Slavery," however, we see 3-4 words per page on average. A notable exception is the "religious/religion" topic. There seem to be instances of words related to religion on pages that are not identified as having religious topics by GPT. Manual inspection of these pages yield examples such as a passage which discusses Martin Luther King, Junior and mentions that he was a pastor. On the corresponding page, there is a religious word, but religion is not a topic.[6]

The above validation method presents another approach to measure the prevalence of specific topics, i.e., measuring counts of specific keywords instead of relying on a large language model to summarize the text. This would involve providing a list of topics as well as a comprehensive list of keywords for each subject. We use the GPT classifications to help identify topics, which helps us identify prevalent topics that we might not have thought to look for and systematically test for differences between the prevalence of topics across collections. This application of GPT is one example of using machine learning as a tool for hypothesis generation as discussed by Ludwig and Mullainathan (2024).

#### V.B.2 *Text as Data: Named Entity Recognition*

Named Entity Recognition (NER) identifies "named entities" (nouns predicted to be proper names such as locations, events, or people), which then allows us to count the number of times these named entities appear in the source text. This provides measures of both what topics and which identities are included in the text, including geography, historical events, and the race, gender, occupation, birthplace, and sexuality of famous people.

To understand the geography to which children may be exposed, we use both the named "GPE," or geopolitical entities, and the "LOC," or location tags. To measure what historical events may be included, we use the "EVENT" tag. Finally, to identify names of famous people, we merge the named entities tagged as a "PERSON" with a pre-existing data set, Pantheon 2.0, containing information from over 70,000 Wikipedia biographies to identify famous people mentioned in the text such as Martin Luther King, Jr. or Amelia Earhart (Yu et al., 2016). Using these data, we can examine the number of famous people mentioned by gender, occupation, sexuality,[7] and birthplace, and we separately manually label their

[6]Note that lexicon-based approaches may miss relevant terms and separately may include terms with multiple meanings that do not reflect the domain of interest.

[7]If sexuality is missing for a famous person, we impute "Heterosexual".

putative race. We aggregate certain occupations into larger categories of interest.[8]

We use the remaining "PERSON" named entities that were not matched to a famous person's name to measure the number of remaining female and male names mentioned in the text, which could be first names of famous individuals or names of other non-famous characters. To do this, we extract the first name of each remaining "PERSON" named entity and estimate the probability that the name is female (or male) using data from the Social Security Administration on the frequency of names by gender in the U.S. population.

#### V.B.3 *Text as Data: Token Counts*

Token counts (often referred to as word counts) are one of the simplest natural language processing tools which allow us to count the number of times specific groups or topics are mentioned. We can use these counts to see how often words related to females (such as *she*, *queen*, and *sister*) are mentioned relative to words related to males (such as *he*, *king*, and *brother*. These words include pronouns and other gendered terms.

#### V.B.4 *Text as Data: Semantic Role Labeling*

In order to understand *how* topics and identities are portrayed textually, we first want to identify the different parts of a sentence so we can separately analyze subjects and actions. We first create a data set which allows us to conduct sentence-level analysis. We use Semantic Role Labeling (SRL) models to identify the predicate-argument structure of a sentence. These models label the agent (active subject), main verb (predicate/action), and the patient (object or passive subject) in a sentence and the various elements that accompany it (e.g. "who" (agent) "did" (verb) "what" to "whom" (patient)).[9] For example, in the sentence "In the North, President Abraham Lincoln asked men to fight to preserve the union," the *agent* in the sentence is "President Abraham Lincoln," the *action* is "ask," and the *patient* is the statement, "to fight to preserve the union."

We use SRL to help us characterize the actions taken by agents of different identities in our text. To do this, we first apply SRL to all the sentences in each textbook to identify the agent and the action. Then we classify the agents' gender as well as the valence, arousal, and dominance scores associated with each action.

[8]For example, we separately examine presence and portrayal of Founding Fathers. We follow Morris (1973) in his identification of key founders, which included criteria related to leadership, longevity, and statesmanship. Thus, the Founding Fathers we include in this analysis are John Adams, Benjamin Franklin, Alexander Hamilton, John Jay, Thomas Jefferson, James Madison, and George Washington. Others take a more expanded view, considering any signer of at least one of three foundational documents: the Declaration of Independence, Articles of Confederation, and/or the U.S. Constitution. Our estimates will therefore reflect a lower bound on the representation of Founding Fathers.

[9]We implement Semantic Role Labeling using AllenNLP, a library for applying deep learning methods to NLP research (Gardner et al., 2018).

To classify the gender of each agent, we start by using named entity recognition to extract any names in the text labeled as the agent. If any names are detected, we classify their associated gender following the process in Section V.B.2. If the name belongs to a famous person, we can classify the agent by the famous person's known gender. If the name does not belong to a famous person, we classify the gender of the agent based off their first name. If no names are found, we apply part-of-speech tagging to identify if the agent of the sentence contains a gendered pronoun or term. We then follow the same process as described in Section V.B.3 to classify gender. If the agent of the sentence contains a gendered pronoun (e.g. she), we use the gender of the pronoun to classify the gender of the agent. Similarly, if the agent of the sentence contains a gendered noun/term (such as brother or queen), we use it to classify the gender of the agent.

#### V.B.5 *Text as Data: Sentiment Analysis using VAD*

To characterize the sentiment of words related to and the actions taken by agents of different identities, we measure three attributes of the words: their valence (V), arousal (A), and dominance (D). Valence measures range from happiness to unhappiness, or positive to negative. Arousal measures indicate a level of "affective activation," or active to passive. Dominance measures reflect the level of control of the emotional state, or dominant/powerful to submissive/weak. We use the NRC VAD Lexicon which includes a list of more than 20,000 English words and their valence, arousal, and dominance scores (Mohammad, 2018). Scores range from zero to one. For example, the words "teamwork" and "leader" have {V, A, D} scores of {0.92, 0.49, 0.77} and {0.83, 0.58, 0.93}, respectively. Both have relatively positive valence, but "teamwork" has slightly more positive association to it. However, "leader" has a higher arousal score and a meaningfully larger dominance score than "teamwork."

#### V.B.6 *Text as Data: Embeddings*

Embeddings at different levels – word, topic, page, document – are tools that can facilitate measurement of *how* topics and identities are depicted.

**Word Embeddings.** Word embeddings are vector representations of words which encode semantic meaning by incorporating information from the nearest neighbors (context) of that word. As a result, word embeddings can be used to examine the context in which the word is discussed. Specifically, the closer two words are positioned together in high-dimensional space (as determined by the cosine similarity of two word vectors), the more often they are used within the same context.[10] We train one Word2Vec model for each collection using the Skip-gram model architecture.

[10]Cosine similarity is the cosine of the angle between the vectors calculated using the dot product of the vectors divided by the product of their lengths.

We use word embeddings to measure the contexts in which different identity groups (in particular males vs. females) are represented within a given textbook collection in two different ways, as described below.

**Top Relative Words.** First, we identify the top relative words most closely associated with a target group (such as females) and least associated with a comparison group (such as males). This approach follows a similar methodology outlined in Manzini et al. (2019) and Charlesworth, Caliskan and Banaji (2022), where the "top relative words" are found by computing the group-to-word association between a single word and the target (or comparison) group, repeating for a given list of target words, and finding the top $N$ words for which the difference in group-to-word associations between the target and comparison group is the greatest. More formally, group-to-word association is calculated using Equation 1 where the association between group $G$ and word $w$ can be calculated as $a(G, w)$:

$$a(G, w) = \frac{1}{|G|} \sum_{g \in G} cos(g, w) \tag{1}$$

where $cos(g, w)$ indicates the cosine similarity (e.g. distance in embedding space) between the two words embeddings $g$ and $w$. We take the difference in word associations between the target and comparison groups to test whether a word $w$ is "skewed" towards the target group (e.g. is more often found in the same context as the target group) as shown in Equation 2:

$$skew(w) = a(G_{target}, w) - a(G_{compare}, w) \tag{2}$$

Finally, the top $N$ relative words most closely associated with a target group and least associated with the comparison group are the $N$ words $w \in W$ with the largest values of $skew(w)$.

We examine the Valence, Arousal, and Dominance (VAD) scores of the top relative words related to females vs. males using the same VAD lexicon we use in our Semantic Role Labeling analysis (Mohammad, 2018) as described in Section V.B.5. Following Charlesworth, Caliskan and Banaji (2022), we do this by finding the average VAD score of words in the VAD lexicon that are most closely associated with female words relative to male words.

**Group-to-Domain Similarity.** Second, we apply a textual analysis variant of the Implicit Association Test (IAT) called the Word Embedding Association Test (WEAT) introduced by Caliskan, Bryson and Narayanan (2017). This word embedding analogue of the IAT involves taking the estimated vectors for words related to a given gender group (e.g., "she" and "woman" belong to the "female" group) and calculating their average distance to words related

to a given domain (e.g., "baby" and "marriage" belong to the "family" domain).We then estimate a domain-specific parameter of "gender-skew" by comparing the average distance to a domain between females and males. We test whether different groups have a greater association with different domains (such as politics or the home). To do this, we extend Equation 1 to calculate group-to-domain similarity between a group $G$ and a domain $D$ as shown in Equation 3 below:

$$A(G, D) = \frac{1}{|G| \times |D|} \sum_{g \in G} \sum_{d \in D} cos(g, d) \tag{3}$$

Similarly, we extend Equation 2 to test whether a domain $D$ is "skewed" towards the target group (e.g. if words in domain $D$ are more often found in the same context as the target group $G$) as shown in Equation 4:

$$Skew(D) = A(G_{target}, D) - A(G_{compare}, D) \tag{4}$$

To determine where differences in domain association between groups are meaningfully significant from each other, we use a novel application of placebo tests as described in Section VI.C.

**Document Embeddings (Book and Page Embeddings).** In order to characterize larger sections of text such as at the page or book level, we apply an analogous approach to word embeddings called "document embeddings." Similar to word embeddings, document embeddings are high-dimensional vector representations of documents. Unlike word embeddings, document embeddings can capture the representations of input texts of variable length such as sentences, paragraphs, pages, or entire books as opposed to single words. We define a document to be either an entire textbook or a single page in a textbook. We train separate Doc2Vec models using the distributed memory (DM) architecture from which we estimate a set of textbook embeddings (one embedding per *textbook*), and separately, a set of page embeddings (one embedding per textbook *page*). In a similar manner to how we analyze word embeddings, we use cosine similarity to calculate how "close," or similar, textbooks in our sample are to each other across collections, subjects, and time.

To visually assess similarities between textbooks, we need to apply techniques that reduce the dimensionality of these vectors to two-dimensional space without losing the detailed information that larger vector spaces provide (such as 300-dimensional space). To do so, we implement t-SNE to reduce these 300-dimensional vectors into a two-dimensional space. There are other approaches one could use (e.g. PCA), but t-SNE is better suited for

this task, since it is more robust to outliers, less sensitive to the ordering of the data points, and better at preserving the clusters in the data.

**Topic Embeddings.** Finally, we integrate the output from GPT, which classifies topics on each page, with the page embeddings. We use this information in two analyses. First, for each topic, we calculate and average the pairwise cosine similarity of pages belonging to that topic for each pair of collections. This allows us to see, in aggregate, the semantic similarity of topics between collections. In the second analysis, we create a topic embedding for each topic and collection. To do so, we compute a single embedding by averaging the page embeddings of all topic-related pages in a given collection. Hence, there will be 3 topic embeddings for each topic, one for each collection. Then, using the word embeddings which were estimated in the same embedding space, we apply cosine similarity to find the closest words to each topic embedding. This allows us to describe, with greater granularity, the context in which topics are discussed in each collection based on a visual examination of proximity between words.

## VI Results

In this section, we discuss the content patterns that we uncover in our collections of textbooks. We begin by examining the *overall similarity* across the collections, by subject and over time. We then document *what* topics and values are included and *how* these topics are presented across collections. We follow this with a discussion of *who* is represented and *how* they are portrayed across collections.

### VI.A Overall Comparison

First, we analyze the overall textual similarity between textbooks by using document embeddings to map distances between textbooks in embedding space. The closer two textbooks are positioned in embedding space, the more similar their textual content is, both in terms of the words used and the context in which they are used. Figure 3 plots our textbook embeddings in a two-dimensional space, with collections represented by colors and subjects by shapes.[11] As expected, textbooks cluster together by subject: Reading and Social Studies textbooks are much closer to each other in the embedding space than to Science textbooks. This is because the vocabulary and topics found in Reading and Social Studies textbooks are more similar to each other than to those found in Science textbooks, which tend to be more technical.

We also see meaningful similarities between California and Texas public school text-

[11] For ease of graphing, each textbook embedding is reduced to two-dimensional space without loss of significant information using t-distributed stochastic neighbor embedding (t-SNE).

books (Figure 3).[12] There is distinct separation between these public school textbook clusters and the clusters with the religious school textbooks in our sample. We also see subjects cluster together across collections, which reinforces our confidence that the model is capturing meaningful contextual relationships in the text within subjects and across collections. The similarities between California and Texas textbooks are particularly notable, given the widely believed narrative of educational polarization between these states.

We quantify the similarities between textbooks across different collections by calculating the cosine similarities for every pair of textbook embeddings from distinct collections (e.g., California vs. Religious) that correspond to the same subject, grade, and year in use. Figure 4a displays histograms of these cosine similarities for pairs of textbooks from different collections (e.g. the blue histogram represents the distribution of cosine similarities between textbook pairs, one from the California collection and the other from the Texas collection, but both sharing the same subject, grade, and year in use). This figure shows that the distribution of cosine similarities between Texas and California textbooks is skewed toward higher similarity, compared to the distributions for California versus religious school textbooks or Texas versus religious school textbooks. Notably, some textbook pairs have near-perfect similarity scores. This is expected, as Texas and California often adopt textbooks from the same publishers, who likely reuse content across different editions of the same textbook tailored to each state.

In Figure 4b, we compute the average cosine similarity between textbooks from two different collections over time. The patterns in these plots are consistent patterns from our t-SNE plot in Figure 3. We see that the public school textbook collections (California and Texas) have higher cosine similarity with each other than either of them have with the Religious collection. When we disaggregate by subject (Figure 4c), the patterns largely hold, with textbooks from Texas and California consistently showing the highest similarities across all subjects. However, in the most recent decade, the similarity between Reading textbooks from Texas and California has decreased, reaching levels of similarity (or dissimilarity) comparable to those with textbooks from our Religious collection. Initially, Science textbooks across all collections exhibited the highest similarity, but this has declined over time. Texas Science textbooks are generally more similar to California Science textbooks than to those from religious schools; over the past three decades, however, Texas Science textbooks have become less similar to those from California and more similar to religious school textbooks. In contrast, California and Texas Social Studies textbooks have grown substantially more similar to each other, while religious school Social Studies textbooks have become slightly

[12] We visualize book embeddings with labeled grade-level clusters in Figure A.2.

more aligned with public school textbooks in the most recent decade.

In the following sections, we unpack specific dimensions of content to understand the factors driving these patterns of difference and similarity across textbook collections.

### VI.B What: Textbook Topics, Geography, History, and Values

Parents choose schooling options, in part, because they want their children to be exposed (or not) to certain knowledge and values. In this section, we document what topics and values are being imparted to children, either implicitly or explicitly.

**Topics.** We first examine what topics are shared with children in each subject (Figure 5). We see that similar topics are discussed across collections within subjects with some notable exceptions unique to each collection. In Reading textbooks, the top three topics identified were the same for each collection: family, nature, and education. The Religious collection is slightly more likely to include narratives related to family than the collections of public school textbooks (8% or 9% vs. 15%). It is also more likely to have narratives related to fear, survival, and perseverance. In Science textbooks, astronomy, biology, physics, observation, and science are among the most common topics across all three collections. In Social Studies textbooks, geography, history, and agriculture are among the most common topics (Figure 5). Note that we see religion as a common topic in Social Studies textbooks for our collection of religious school textbooks but not in the public school textbooks.

This suggests meaningful overlap in the most prevalent topics transmitted to children, which suggests that children are being exposed generally to similar topics across educational settings. Much of the public debate, however, does not center on these identified topics. Rather, discussions often focus on topics that are considered to be more contentious such as those that address issues such as evolution, enslavement, racism, and climate change. We first examine the prevalence of these topics as identified by a large language model, both overall and over time (Figures 7, A.3, and A.4). These figures show that religious topics are present on over 12% of pages in our religious school textbooks on average as opposed to 4.4% and 2.8% in California and Texas, respectively, and when a specific religion is mentioned it is overwhelmingly Christianity. These figures also show that California and Texas textbooks are much more likely to address topics related to climate change than textbooks in our religious school collection. By contrast, our religious school collection is the only set of textbooks that discusses creationism and is much more likely to address topics related to evolution (typically as a way of contrasting it to creationism). California textbooks are the most likely to include slavery-related topics overall, though patterns are mixed over time; whereas the Religious collection is more likely to include topics related to the Civil War, and increasingly

so over time. Both California and Texas discuss immigration more often than religious school textbooks, though in Texas textbooks we see an increase in the proportion of pages which discuss immigration over our sample period and in California we see a decrease. Racism and LGBTQIA+ related topics are rarely, if ever, mentioned across all three collections.

We have described how *often* certain topics are mentioned across textbook collections. However, just as important as *whether* certain topics are discussed is *how* they are discussed. By integrating the information on topics with our embeddings as described in Section V.B.6, we can examine differences in the context or sub-topics surrounding a given topic across collections.

To systematically examine how similar these collections are regarding the context in which they discuss different topics, we conduct pairwise comparisons of the cosine similarity of topics between each collection (Figure A.5a). This exercise reveals higher similarities between topics discussed in Texas textbooks and topics discussed in California textbooks than between these public school textbooks and those from our collection of religious school textbooks, indicating that public school textbooks across these states discuss topics more similarly to each other than they do compared to religious school textbooks.

We also plot topic embeddings for each collection along with the top word embeddings closest to each topic embedding in two-dimensional space to visually present and examine the context surrounding each topic embedding. We then cluster the word embeddings into subtopics.[13] These figures confirm that many of the topics of interest are discussed in similar contexts across collections which means the collection-specific topic embeddings are close to each other in embedding space, have similar top words, and are a similar distance to each of their top words. For example, the topic embeddings for *Slavery* are all similarly close to each other and to words related to abolitionists, resisted, and plantations (Figure D.5).

However, for some topics, differences emerge between collections. In Figure 6, for example, we see that *Religious* topics are discussed in different contexts across the public school and religious school textbooks. In the Religious collection, the *Religious* topic embedding is closer to word embeddings related to god and Christianity such as prayer, creator, eternal, Jesus, Moses, and heaven. In both the California and Texas collections, the *Religious* topic embedding is closer to word embeddings related to the U.S. founding era such as religious freedom, Puritans, Quakers, colony, William Penn, and Roger Williams. Finally, we see differences in the discussion of *Evolution* across these three collections. As shown in the topic embeddings figures, the topic of evolution is often discussed in the context of

[13] We show further topic embedding analysis and validation in Appendix C and Appendix D.

words such as creation, theories, evidence, and Genesis in our collection of religious school textbooks but this is not the case in public school textbooks. Unlike many of our other topics which appear on hundreds to thousands of pages, the topic of evolution only appears on a small number of pages in our sample, fewer than 100 pages in each collection (Figure 7). We manually examined these pages and extracted representative passages to illustrate qualitative similarities and differences in the treatment of evolution across collections.

Pages in our collection of religious school textbooks that contain the topic of evolution fall under three main categories: (1) concepts related to evolution are discussed, but evolution is not named or mentioned (for example from a 3rd grade Science religious school textbook classified as discussing natural selection: "A twig caterpillar is the same color as a twig. In fact it looks just like a twig. Birds which eat insects ... think the caterpillar is part of the tree."; (2) evolution is explicitly brought up and depicted as not being true; and (3) evolution is discussed in the context of Social Darwinism: "The idea that only the fittest people can survive in society is known as Social Darwinism. Because people are naturally sinful and selfish this belief was dangerous." Examples of (2) are common. One such passage discusses the human and dinosaur footprint found next to each other in Texas as evidence against evolution. A different example passage from a religious school 5th grade Science textbook reads as follows:

> "Evolutionists interpret the fact that dogs and cats are similar as being an evidence for evolution. Creationists on the other hand explain that God created living things with similar structures because they would need the structures to do similar jobs. Cats and dogs for example have similar teeth because they enjoy similar eating habits. If a job is to be done with greatest efficiency there is usually a best method for doing the job. God used the best most efficient design in his creatures so they would be well suited for living on the earth."

In our collections of public school textbooks, pages predicted to contain the topic of evolution discuss a wide variety of concepts including Neanderthals, fossils, natural selection, prehistoric animals, the Ice Age, and behavioral or structural adaptations. One difference between Texas and California textbooks is that in the Texas collection, even though there are discussions of natural selection, adaptation, and the Earth being millions of years old, there is never any explicit mention of evolution, the Big Bang theory, or Charles Darwin. An example of a passage from a Texas 3rd grade Reading textbook reads as follows:

> "If eohippus does not look like a horse of today, how do we know that eohippus was a horse? How do we know that eohippus was not a member of the dog family? We know that eohippus is a member of the horse family because we have found skeletons of other horses that lived long ago. When we put the skeletons in a row we can see small changes. Picture 2 shows just how horses changed over millions of years. It shows that

eohippus did not change into a dog. It changed into the horses of today."

In our California collection, not only are there explicit mentions of evolution, there is also one unique passage in a California 5th grade Social Science textbook called "Stories of How the World Began" which describes many different creation stories from different cultures. An excerpt from that passage reads:

> "How did the world begin? Scientists say there was a big bang and dust and particles of matter came together to form stars and planets. In the Western nations the Bible told of Adam and Eve and how God created the world in seven days. The Popol Vuh, the sacred book of the Mayan Indians, told this tale of creation: This is the story of how all was still, all quiet, in silence; there was no movement, no sound, and the whole of the sky was empty. This is the first story, the first speech..."

In addition to the topics which are commonly discussed across collections, it is useful to understand what topics are exclusively identified in specific collections and not in other collections (Figure A.6a). We see that only California textbooks are identified as having topics related to California history and geography, including discussion related to the Kumeyaay tribe, an Indigenous people from what is now Southern California. Only Texas includes topics related to the practice and knowledge of Texas Essential Knowledge and Skills (TEKS) state standards. Only the religious school textbooks have topics related to God, evangelism, and creationism, consistent with what we see in Figure 7.

Much of the debate surrounding what children learn involves the knowledge they are given about history and the world around them. Indeed, in each of the collections, geography and history were among the top Social Studies topics identified (Figure 5). When examining specific mentions of places, we see that children are most likely to be exposed to Earth generally and the United States overall with very little relative exposure to other parts of the world. California is most likely to include mentions of California-specific geography, and Texas is most likely to include mentions of Texas-specific geography. Half of the top ten places mentioned outside of the United States are located in outer space, including Mars, Jupiter, Venus, and the Moon. The Sun only only appears in the public school collections, and Israel is mostly only mentioned in the Religious collection (Figure 8a).

Figure 8b shows the most commonly mentioned historical events in our sample of textbooks. We see varying patterns across the collections, but in particular, we see an increase in mentions of the Civil War over time in California and the Religious collections, but a sharp decrease over the last two decades in Texas. By contrast, we see that Texas and California textbooks are more likely than the religious school collection to mention the American Revolution.

There were only two books in our sample that we identified as discussing LGBTQIA+ topics. These books were identified using pages predicted as discussing LGBTQIA+ topics by GPT and verified using token counts of the words lesbian, gay, bisexual, transgender, queer, intersex, and asexual. The first book is a 5th grade Social Studies textbook from our Religious collection. The relevant passage is taken from a section entitled "Gay Rights":

> "Another issue that troubled Americans was gay rights. During his time as president, Obama encouraged the growth of gay rights. He said that men should be able to marry men and that women should be able to marry women. When Obama first took office, many Americans disagreed with this view. By 2015, many agreed. Americans who disagreed faced problems. Many believed that God designed marriage to be between a man and a woman. They trusted that His plan benefited humanity. These Americans did not support gay marriage. They thought that it would harm society. For example, children in families where there is a same-sex marriage are denied having both a father and a mother. But in 2015, the Supreme Court ruled in favor of gay marriage. It had redefined marriage. Five justices voted in favor of gay marriage, but the other four justices voted against it and pointed out that this change threatened religious liberties. Some Christians who owned bakeries, flower shops, or photo studios were sued for not participating in a same-sex wedding. These businesses served all other customers, but refused to participate in same-sex weddings. Many of these business owners lost their court cases and were fined. This change in the United States raised other questions. Could religious schools continue to teach their views of marriage? Would schools and business owners come under more pressure to violate their beliefs on these issues? These questions remained unanswered."

The second book is a 3rd grade Social Studies textbook from our California collection. The relevant passage is from a section entitled "Supporting Many Different Communities":

> "Many other Californians worked hard to make changes in their communities...
>
> Harvey Milk was a political leader and a gay activist in the 1970s. Gay men are attracted to other men. Milk was ... one of the first elected officials in history to tell the public that he was gay. Milk worked hard to protect gay and lesbian rights. Lesbians are women who are attracted to other women. Milk's work also made it easier for gay, lesbian, bisexual (those attracted to both men and women), and transgender groups to stand up for their rights.
>
> Chaz Bono is a transgender activist. The word gender refers to whether someone is male or female. Transgender people feel that their gender is the opposite of what it was at birth. Bono was born female, but always felt male. He changed his gender and today lives as a man. Bono and other activists fight against discrimination and stereotypes of their community."

The passage in the Religious collection highlights the tension between LGBTQIA+ rights and religious freedoms. Given that this is the only representation of LGBTQIA+ individuals in our sample, children exposed to this curriculum are only seeing people in this community posed as a threat to the freedoms of others. The passage in the collection of California

textbooks gives definitions of what it means to be a gay, lesbian, bisexual, and/or transgender individual and highlights two LGBTQIA+ activists. While the sentiment of this passage is positive overall, by not discussing any LGBTQIA+ individuals of color or who identify as female, this passage fails to take an intersectional lens. We did not identify any passages in the Texas state-adopted curriculum that discuss LGBTQIA+ topics, rights, or individuals.

**Values.** The majority of parents who choose private or home schooling options report a desire for "morals," "values," or "character" education as among the top most important qualities when choosing a school (Catt and Rhinesmith, 2017). We accordingly examine the values that are being espoused on the pages of the textbooks.

Consistent with our analysis of top topics, the top values we identify are similar across collections with the top three values being centered around learning. We see this overlap especially between California and Texas textbooks (Figure 9). The primary outlier is that of *faith*, which shows up as a top value in the Religious collection (Figure 9). We explore this further by combining values related to religion such as *religious faith, religious freedom, religious beliefs, religious diversity, religiosity, religious tolerance*, and *religious devotion*. Using this group of values we find that religious values are present on 2%-16% percent of pages on average in a given subject in our collection of religious school textbooks as compared to 0%-5% of pages in our collections of public school textbooks. Two notable patterns emerge: First, in public school textbooks, religious values only appear in Reading and Social Studies textbooks; and second, they are present on approximately 2.1% of pages on average from Science textbooks in the Religious collection (Figure 10a).

Despite many parents emphasizing character and values education as a key factor when choosing a private school (Catt and Rhinesmith, 2017), we see that both Texas and California public school textbooks include a meaningful proportion of pages that espouse character values to children (Figure 10b). The religious school collection has a higher proportion of pages overall that include character values, but all three collections include character values on over 30% of all textbook pages. When examining the specific character values that are espoused, we see that the public school textbooks are more likely to have values related to creativity, curiosity, and teamwork, whereas the Religious collection is more likely to have values related to perseverance, compassion, courage, and gratitude (Figure A.7a). All three collections put a similar amount of emphasis on respect and responsibility. Surprisingly, there is a smaller proportion of pages in the religious school collection that are classified as espousing moral values than one may expect and than the public school collections have character or family values (Figure A.7b). Despite these differences in terms of prevalence, the contexts in which character and family values arise are similar across all collections when

conducting pairwise comparisons between collections (Figure A.5b).

Some values are exclusively identified with one collection and the others (Figure A.6b). We see that only California textbooks are connected with values related to learning from nature or housing affordability. Only Texas emphasizes values related to the recognition of historical figures and correctness. Only the Religious collection has values related to spiritual growth, trust in God, and obedience to God.

### VI.C Who: Historic Figures, Birthplace, Race, Skin Color, Sexuality, Gender

In addition to understanding *what* topics and values are included in textbooks, it is important to understand *who* is highlighted and *how*. In this section, we discuss the messages related to historic figures and their occupations, birthplaces, race, skin color, sexuality, and gender that are portrayed to children.

**Historic Figures and Birthplaces** Political and media discussions of curricula are commonly centered around which historical figures and events should be highlighted and how they should be portrayed. This discourse reflects concerns expressed by parents as highlighted in a survey conducted by the Heritage Foundation which found that 59.5% of teachers and 56% of parents surveyed noted the importance of schools including content related to the Constitution and the founding era of the United States (Sailor et al., 2021).

We explore the inclusion of historical figures across the collections in Figure 11a and compare how often they are mentioned across educational settings.[14] The most frequently mentioned famous people in each collection are shown in Table 2. George Washington, Thomas Jefferson, and Benjamin Franklin – White males who are all Founding Fathers – are among the most commonly included famous individuals across all collections. The Religious collection only has White males in the top twenty list of historical figures included. The most frequently mentioned individuals in each collection are all identified as being heterosexual, with the one exception of Jane Addams, a civic leader who identified as a lesbian, mentioned in the Texas collection (though her sexuality is never mentioned in the books themselves).

Textbooks in our public school collection are slightly more likely to discuss civic leaders (e.g. Martin Luther King Junior, Susan B. Anthony, Frederick Douglass, Rosa Parks) than textbooks in our religious school collection, and this is increasingly true over time. By

[14]We show these patterns by subject in Figure A.8. In Social Studies, politicians are the most commonly mentioned, followed by Founding Fathers. In Science, scientists are the most frequently mentioned famous people, but only half of the mentions of famous people are scientists. 4% of the famous individuals in religious school Science textbooks are religious figures. Science books across all collections are more likely to mention artists or athletes than Social Studies books are likely to mention scientists. Public school reading textbooks are more likely to include mentions of artists and athletes, followed by politicians.

contrast, textbooks in our religious school collection are more likely to discuss politicians and military personnel (e.g. Abraham Lincoln, John F. Kennedy, Robert E. Lee). In recent years, scientists, explorers, artists, founding fathers, athletes and religious figures are all discussed a similar amount across collections.

We also examine the contexts in which Founding Fathers are discussed across collections by calculating the average VAD scores of the top words closest to the word embeddings of the Founding Fathers' names. Figure 11b shows that these individuals are depicted in the context of words with similar valence, arousal, and dominance across the three collections.

We also see that the top words identified by creating a topic embedding in a similar manner as described in Section V.B.6 using all pages that mention at least one of the Founding Fathers depict Founding Fathers in similar contexts for all three collections (Figure D.8). Religious leaders, explorers, and astronauts present in the texts also have similar valence, arousal, and dominance scores across the collections.

When we examine the birthplaces of the famous individuals to whom children are exposed (Figure 12), we see relatively similar patterns over time and overall across public school textbooks, with most inclusion of individuals born in North America, followed by individuals born in Europe. Most of these individuals are male, but when females are present, they are mostly from North America (Figures A.9a and A.9b). The Religious collection has similar inclusion of individuals born in North America and Europe. For all of the collections, there is little to no representation of individuals from Africa, Asia (including East Asia and South Asia), the Middle East, and South America.

**Race.** A salient part of curricula are the famous individuals who are highlighted for children. History and science came from individuals, so in the conveyance of information, textbooks often include discussion of these individuals. They can serve as real-life characters in our histories, examples of leadership, and as role models to whom children may aspire to be similar. To measure how often famous individuals of different races/ethnicities are discussed, we apply Named Entity Recognition (NER) tools as discussed in Section V.

We examine the incidence of famous individuals across the corpora and find that the vast majority of famous individuals mentioned are White individuals (between 81%-92% across all collections and subjects as shown in Table 1). This pattern is consistent over time with a slight decrease in famous White individuals represented in recent decades (Figure 12). Over time, the Religious collection is more likely to feature famous individuals who are White and less likely to feature famous individuals who are Black relative to the public school collections. Each of the collections have very little inclusion of Asian, Latine, or Indigenous

famous or historical individuals. Generally, within racial groups, males are often more likely to be included than females (Figure A.10a).

We also examine the contexts in which Native Americans are raised by creating a topic embedding in a similar manner as described in Section V.B.6 and identifying the associated top words (Figure D.7). California and Texas textbooks are more likely to address Native Americans in the context of tribes and settlements, while religious school textbooks discuss them more in the context of Pilgrims and religious leaders such as Joseph Smith.

One caveat of focusing on famous individuals only to measure the representation of race is that we are unable to capture the race of non-famous individuals in the text. However, we can look at how often both famous and non-famous individuals of different races are pictured in textbooks by examining the race classifications of detected faces in our sample. We find that in all collections, most of faces depicted are White individuals. However, the difference between the proportion of White faces and the proportion of Black faces is much closer to parity in faces than it is when examining the race of famous individuals. In public school textbooks, California textbooks have the largest percentage of Asian and Latine faces, though this number is still very small; only about 5% each compared to 3% each in Texas public school textbooks and 2% each in the Religious collection (Table 1 and Figure A.10b).

**Skin Color.** In Figure 13a, we show the distribution of detected faces in our textbook by skin tint (e.g. how light or dark their skin color is on a scale from 0 to 100). We plot these distributions over a scatterplot showing the skin color of every face in our sample. The x-axis for both the density and scatter plot is skin tint, but the y-axis for the scatterplot is the vibrancy of each skin color. Figure 13b shows that the patterns are relatively stable over time, with a slight trend towards increased lightening of character skin color in the Religious collection over time.

Overall, we see that the pictured individuals in the public school collections from Texas and California are more likely to be depicted with darker skin on average than the pictured individuals in the Religious collection. This pattern holds even when conditioning on race; in the Religious collection the skin color of pictured faces is lighter than pictured faces in California and Texas textbooks, both on average and over time (Table B.1). This is true even when we condition on the predicted race of the face. We also see that older children (children in grade 5 as opposed to grade 3) are more likely to be shown characters with darker skin. Table B.2 further separates faces by predicted race and finds that faces in the Religious collection are lighter for all races/ethnicities except for Latine. However, we do not see a similar clear pattern when comparing Texas and California. In California, for example, characters classified as being Asian or more likely to have lighter skin than

characters classified as Asian in Texas; whereas this pattern is flipped for characters classified as being Black: they are shown with lighter skin in Texas compared to California.

**Sexuality.** We also examined the number of mentions of famous individuals who identified as Lesbian, Gay, or Bisexual and found fewer than 20 individuals mentioned in each of our public school textbook collections and fewer than 10 individuals mentioned within our religious collection of textbooks. In Table B.3, we list each of these individuals and the number of times they are mentioned and the number of books they are mentioned in. In all three collections, the total mentions of LGB famous individuals make up less than 1% of all mentions and this has been consistent over time (Figure A.11). Note that only Harvey Milk was mentioned as identifying as gay, while none of the other famous individuals were discussed in the context of their sexuality.

**Gender.** In looking at gender presence over time, we see the public school collections have higher female representation in both the text and the images of textbooks than the Religious collection (Figure 14a). These patterns are consistent over time; however, across all collections females are under-represented in relative to their share of the population (Table B.5). The disparity is much larger in the text than it is in the images. While females are depicted less often than males, female representation is close to parity in images (Table 1); this is not the case for female representation in the text. The total proportion of female words over time ranges from approximately 20% in our collection of religious school textbooks to approximately 40% in our collections of public school textbooks, with these proportions being relatively stable over time (Figure A.12). Out of the top ten famous people and top ten first names in a collection, at most one are female (Tables 2 and B.4). Interestingly, the top name in each collection by a meaningful margin is *John*, a male name that is of Biblical origin, originally appearing in the Old Testament.

One could hypothesize that females are less represented in textbooks due to their historical marginalization in social studies and science fields. We examine this by disaggregating gender representation by subject. We see a distinct pattern emerge. Science and Social Studies textbooks have less female representation. Notably, however, in public school Reading textbooks, females are mentioned as frequently as males (Figure 14b). This pattern may be attributed to the greater flexibility publishers have in selecting the gender of characters for Reading textbooks. Unlike history textbooks, where the narrative often revolves around historical events and figures from which women were historically marginalized—such as the consistent male dominance in the U.S. presidency—Reading textbooks are not bound by these historical constraints. Consequently, discussions in Reading materials can feature a more balanced gender representation. We do not see this same pattern in the Religious

collection (Figure 14b). Indeed, females are consistently less represented in the text across all subjects within the religious school books. The closest to gender parity is found in Reading textbooks, where female-related words constitute 41% of all gendered words.

***Gendered Actions.*** We next explore *how* females and males are portrayed. We first characterize the sentiment associated with actions attributed to females and males. We do this by calculating the average valence, arousal, and dominance scores for actions associated with female agents and male agents in respective sentences as described in Sections V.B.4 and V.B.5. We then measure the differences in these average scores to derive a gender skew metric for each sentiment dimension.

We see that there are consistent inequalities in gender portrayal across each of the collections: each collection portrays females with similar directions of inequality. In Figure 15a, we see that across all collections of textbooks, the actions taken by females are more positive, but less active and powerful than actions taken by males. This pattern remains even when we focus only on sentences where the gender of agents is identified using only gendered pronouns and not names and gendered terms (Figure A.13a) and is much more stark when we focus on actions that are only taken by female agents as opposed to actions that are only taken by male agents (Figure A.13b).

***Gendered Context.*** We then measure the contexts in which females and males are described in the text in in two different ways: (1) through the top relative words most closely associated with females vs. males, and separately, males vs. females and (2) by testing whether females or males are more likely to be found in the context of specific domains (such as in the context of family or politics). We describe this approach in Section V.B.6.

Table 3 shows the top 10 relative words most closely associated with females vs. males and, separately, with males vs. females for each collection, along with the average VAD scores for the top words. We see that females are more associated with words related to appearance and males are more associated with words relating to the military and horses. Consistent with the patterns we see for the actions related to each gender, the sentiment of the top ten relative words for females are more positive on average, but less active and powerful on average than the top relative male words. This pattern remains present when we extend our analysis to include the top 50 relative words by gender, as shown in Figure 15b.

In Figure 16 and as described in Section V.B.6, we show the gender centeredness of several domains: family, home, performance arts, politics, sports, and tools. For example, the *Home* domain (which includes words such as marriage, children, relatives, and household) is more skewed towards females and the Politics domain (which includes words such as

government, democracy, and political) is more skewed towards males.

A natural question is whether the estimated gender skew is meaningfully different from zero gender skew. We test this using a placebo test. In this test, we randomly draw a set of words from the vocabulary of the textbook collection that is the same size as the list of words in the domain for which we are testing for gender skew and then calculate the skew for this "random" domain. We repeat this 1,000 times and plot the distribution of these random domain associations over the top of the bar graphs using a barplot where the edges of the barplot show the 10th and 90th percentile values of the possible gender skewness in a domain of random words.

We find that in all three textbook collections, the *Family* and *Home* domains are skewed towards female, and this measure of skewness is statistically different than the skewness we would find in a random set of words. We find a similar (though smaller) gender skew in the *Politics* domain towards males for the Texas and Religious textbook collections. Lastly, we see no meaningful differences by gender when examining the associations in the *Sports* and *Performance Arts* domains in the Religious and Texas collections, but a slight skew towards females in these domains in California textbooks.

## VII Conclusion

What we teach our children matters. Curricula play a vital role in transmitting knowledge, shaping values, and constructing collective memory. Growth in U.S. school choice programs has increased public funds directed to religious schools, but little is known about what is taught and how it compares to public school curriculum.

Our study analyzes textbooks from public schools, religious private schools, and home schools using computational methods and advances from artificial intelligence to assess how people, topics, and values are represented over time. Contrary to widespread narratives of political polarization, we find few substantial differences between public school textbooks from Texas and California. By contrast, religious school textbooks show less representation of females, more depictions of light-skinned individuals, and differing treatments of topics such as evolution and religion, including in science content. Over one-third of each of the textbook collections communicate character values, with religious school textbooks doing so at higher rates. Notable similarities also appear: both public and religious textbooks rarely address LGBTQIA+ topics and individuals, have minimal representation of Asian, Indigenous, and Latine individuals, depict females more positively but in less active or powerful roles than males, and present the U.S. founding era and slavery in comparable ways.

It is important to note these methods perform better when assessing overall portrayals

rather than examining the nuances of individual representations, suggesting a potential gap that could be addressed through manual content analysis as evidenced by the differences in portrayal uncovered in our examination of the passages discussing topics related to evolution and LGTBQIA+ individuals.

An important consideration for the implications of our work is that curricula can take differing forms: curricula as written, curricula as implemented (by teachers), and curricula as discussed (among students). In our setting, we focus our analysis on the first of these, which is curricula as written. Similarities in textbooks across different educational settings do not necessarily guarantee that children receive similar educational experiences in the classroom. Pedagogy plays a crucial role in shaping how material is taught, and differences in teacher training, beliefs, and classroom practices can lead to substantially different lessons even when the same textbook is used. However, while standardizing classroom instruction requires a highly prescriptive and often impractical implementation strategy, textbook adoption remains a more feasible lever for policy intervention. Importantly, prior research has shown that changes in curriculum standards alone can influence student outcomes. This suggests that the written curriculum matters and the importance of analyzing textbook book content remains despite potential difference in pedagogy.

Our paper demonstrates how advances in computational social science and artificial intelligence tools can be used to measure differences in educational materials and other sources of text and image data. Prior textbook content analysis studies have focus on cross-sectional analyses of gender and racial representation in text alone. We build on and extend prior research on textbook content analysis by analyzing both the text and images of a novel panel of elementary school textbooks used in public and religious settings from 1980 to the present. Our longitudinal approach allows us to measure changes in the presence and portrayal of people, topics, and values over time and across educational settings.

In the midst of an increase in educational gag orders and statewide changes in the content standards of public school curriculum, our work shines a light on what messages children are receiving from informal and non-traditional curricula outside the public school system which could have long-term implications for children's beliefs and decision-making.

## References


**Abdulkadiroglu, Atila, Parag A Pathak, and Christopher R Walters.** 2015. "School vouchers and student achievement: Evidence from the Louisiana Scholarship Program."

**Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz.** 2023. "What we teach about race and gender: Representation in images and text of children's books." The Quarterly Journal of Economics, 138(4): 2225–2285.

**Adukia, Anjali, Callista Christ, Anjali Das, and Ayush Raj.** 2022a. "Portrayals of race and gender: Sentiment in 100 years of children's literature." Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, 20–28.

**Adukia, Anjali, Patricia Chiril, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha.** 2022b. "Tales and tropes: Gender roles from word embeddings in a century of children's books." Proceedings of the 29th International Conference on Computational Linguistics, 3086–3097.

**Adukia, Anjali, Richard Hornbeck, Daniel Keniston, and Benjamin Lualdi.** 2025. "The social construction of race during Reconstruction."

**Allen, Ann M, Daniel N Allen, and Gary Sigler.** 1993. "Changes in sex-role stereotyping in caldecott medal award picture books 1938—1988." Journal of Research in Childhood Education, 7(2): 67–73.

**Arold, Benjamin W.** 2024. "Evolution vs. creationism in the classroom: The lasting effects of science education." The Quarterly Journal of Economics.

**Arold, Benjamin W, Ludger Woessmann, Larissa Zierow, et al.** 2022. "Religious education in school affects students' lives in the long run." Vol. 23, 40–43.

**Ash, Elliott, Ruben Durante, Maria Grebenschikova, and Carlo Schwarz.** 2021. "Visual representation and stereotypes in news media."

**Bartholet, Elizabeth.** 2020. "Homeschooling: Parent rights absolutism vs. child rights to education & protection." Ariz. L. Rev., 62: 1.

**Bazzi, Samuel, Andreas Ferrara, Martin Fiszbein, Thomas P Pearson, and Patrick A Testa.** 2023. "The Confederate diaspora."

**Bazzi, Samuel, Masyhur Hilmy, and Benjamin Marx.** 2020. "Religion, education, and the state."

**Bickford, John Holden, and Lori A Knoechel.** 2017. "The historical representation of Native Americans within primary-and intermediate-level trade books." The Councilor: A Journal of the Social Studies, 78(2): 4.

**Braden, Eliza G, and Sanjuana C Rodriguez.** 2016. "Beyond mirrors and windows: A critical content analysis of Latinx children's books." Journal of Language and Literacy Education, 12(2): 56–83.

**Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan.** 2017. "Semantics derived automatically from language corpora contain human-like biases." Science, 356(6334): 183–186. Publisher: American Association for the Advancement of Science.

**Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang.** 2017. "Curriculum and ideology." Journal of Political Economy, 125(2): 338–392.

**Catt, Andrew D, and Evan Rhinesmith.** 2017. "Why Indiana parents choose: A cross-sector survey of parents' views in a robust school choice environment." EdChoice.

**Charlesworth, Tessa ES, Aylin Caliskan, and Mahzarin R Banaji.** 2022. "Historical representations of social groups across 200 years of word embeddings from Google Books." Proceedings of the National Academy of Sciences, 119(28): e2121798119.

**Clark, Roger.** 2007. "From margin to margin? Females and minorities in Newbery and Caldecott Medal-winning and Honor books for children." International Journal of Sociology of the Family, 263–283.

**Crawfurd, Lee, Christelle Saintis-Miller, and Rory Todd.** 2024. "Sexist textbooks: Automated analysis of gender bias in 1,255 books from 34 countries." PloS one, 19(10): e0310366.

**Crisp, Thomas, Roberta Gardner, and Matheus Almeida.** 2018. "The all-heterosexual world of children's nonfiction: A critical content analysis of LGBTQ identities in Orbis Pictus Award books, 1990–2017." Children's Literature in Education, 49(3): 246–263.

**Deckman, Sherry L, Ellie Fitts Fulmer, Keely Kirby, Katharine Hoover, and Abena Subira Mackall.** 2018. "Numbers are just not enough: A critical analysis of race, gender, and sexuality in elementary and middle school health textbooks." Educational Studies, 54(3): 285–302.

**Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger.** 2014. "School choice, school quality, and postsecondary attainment." American Economic Review, 104(3): 991–1013.

**EdChoice.** 2023. "Fast facts on school choice." Accessed on March 9, 2024.

**Friedman, Jonathan, and James Tager.** 2022. "Educational gag orders: Legislative restrictions on the freedom to read, learn, and teach."

**Gardner, Matt, Joel Grus, Mark Neumann, Oyvind Tafjord, Pradeep Dasigi, Nelson Liu, Matthew Peters, Michael Schmitz, and Luke Zettlemoyer.** 2018. "AllenNLP: A deep semantic natural language processing platform."

**Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou.** 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." Proceedings of the National Academy of Sciences, 115(16): E3635–E3644.

**Jamison, Peter, Laura Meckler, Prayag Gordy, C Ence Morse, and C Alcantara.** 2023. "Home schooling's rise from fringe to fastest-growing form of education." Washington Post.

**Kamel, Donia, and L Guillermo Woo-Mora.** 2023. "Skin tone penalties: Bottom-up discrimination in football."

**Keith, Verna M, and Carla R Monroe.** 2016. "Histories of colorism and implications for education." Theory Into Practice, 55(1): 4–10.

**Klein, Rebecca.** 2017. "Voucher schools championed by Betsy DeVos can teach whatever they want. turns out they teach lies."

**Koss, Melanie D.** 2015. "Diversity in contemporary picturebooks: A content analysis." Journal of Children's Literature, 41(1): 32.

**Koss, Melanie D, and Kathleen A Paciga.** 2020. "Diversity in Newbery Medal-winning titles: A content analysis." Journal of Language and Literacy Education, 16(2): n2.

**Koss, Melanie D, Nancy J Johnson, and Miriam Martinez.** 2018. "Mapping the diversity in Caldecott books from 1938 to 2017: The changing topography." Journal of Children's Literature, 44(1): 4–20.

**Kozlowski, Austin C, Matt Taddy, and James A Evans.** 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." American Sociological Review, 84(5).

**Kunzman, Robert.** 2010. "Homeschooling and religious fundamentalism." International Electronic Journal of Elementary Education, 3(1): 17–28.

**Lucy, Li, Dorottya Demszky, Patricia Bromley, and Dan Jurafsky.** 2020. "Content analysis of textbooks via natural language processing: Findings on gender, race, and ethnicity in Texas US history textbooks." AERA Open, 6(3).

**Ludwig, Jens, and Sendhil Mullainathan.** 2024. "Machine learning as a tool for hypothesis generation." The Quarterly Journal of Economics, 139(2): 751–827.

**Manzini, Thomas, Yao Chong Lim, Yulia Tsvetkov, and Alan W Black.** 2019. "Black is to criminal as Caucasian is to police: Detecting and removing multiclass bias in word embeddings."

**Michalopoulos, Stelios, and Christopher Rauh.** 2024. "Movies."

**Michalopoulos, Stelios, and Melanie Meng Xue.** 2021. "Folklore." The Quarterly Journal of Economics, 136(4): 1993–2046.

**Mohammad, Saif M.** 2018. "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words." Melbourne, Australia.

**Morris, Richard.** 1973. Seven Who Shaped our Destiny: The Founding Fathers as Revolutionaries. Harper & Row.

**Musaddiq, Tareena, Kevin Stange, Andrew Bacher-Hicks, and Joshua Goodman.** 2022. "The pandemic's effect on demand for public schools, homeschooling, and private schools." Journal of Public Economics, 212.

**National Center for Education Statistics.** 2022. "Homeschooled children and reasons for homeschooling." Condition of Education, Retrieved December 18th, 2023.

**Riley, Emma.** 2022. "Role models in movies: The impact of Queen of Katwe on students' educational attainment." The Review of Economics and Statistics, 1–48.

**Sailor, Angela, Lindsey M Burke, Anne Segal, and Adam Kissel.** 2021. "Civics studies: Why they matter, what parents and teachers think, and how they can reclaim truth." Heritage Foundation.

**Sallabaş, Muhammed Eyyüp.** 2013. "Analysis of narrative texts in secondary school textbooks in terms of values education." Educational Research and Reviews, 8(8): 361.

**Scaramanga, Jonny, and Michael J Reiss.** 2018. "Accelerated Christian Education: A case study of the use of race in voucher-funded private Christian schools." Journal of Curriculum Studies, 50(3): 333–351.

**Stevenson, Betsey, and Hanna Zlotnik.** 2018. "Representations of men and women in introductory economics textbooks." Vol. 108, 180–85.

**Szasz, Teodora, Emileigh Harrison, Ping-Jung Liu, Ping-Chang Lin, Hakizumwami Birali Runesha, and Anjali Adukia.** 2022. "Measuring representation of race, gender, and age in children's books: Face detection and feature classification in illustrated images." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.

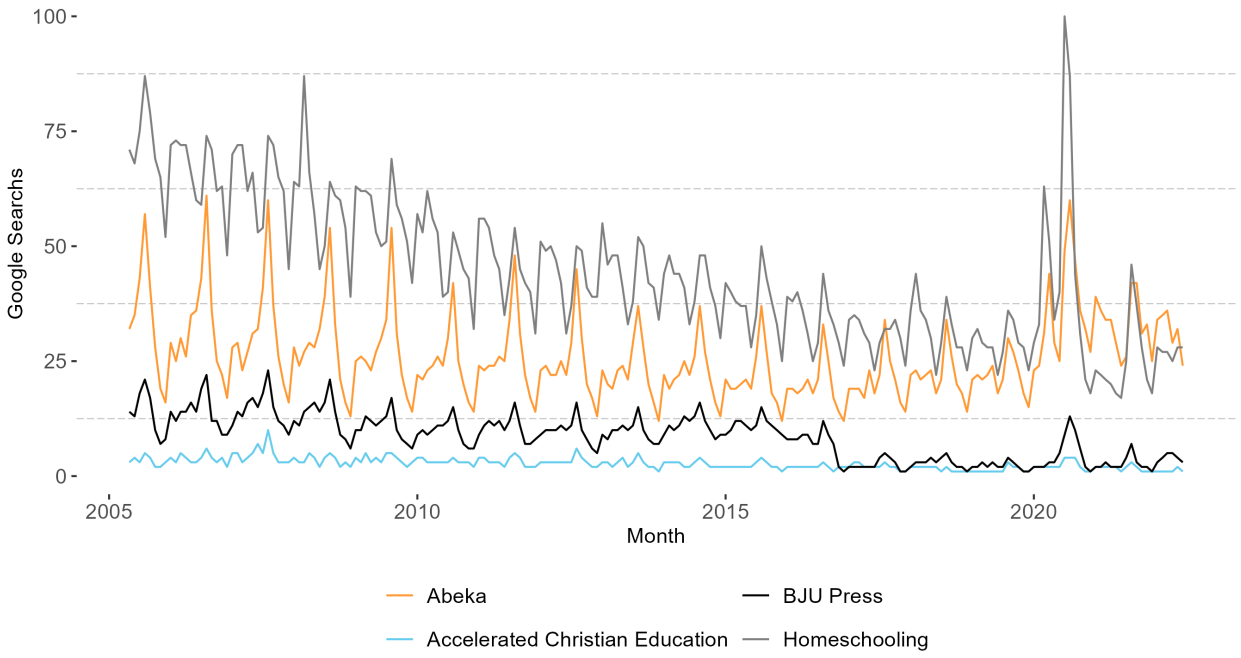
**Wade, Rahima C.** 1993. "Content analysis of social studies textbooks: A review of ten years of research." Theory & Research in Social Education, 21(3): 232–256.

**Yanagizawa-Drott, David, and Hans-Joachim Voth.** 2023. "Image(s)."

**Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A Hidalgo.** 2016. "Pantheon 1.0, a manually verified dataset of globally famous biographies." Scientific Data, 3(1): 1–16.

## VIII Main Figures

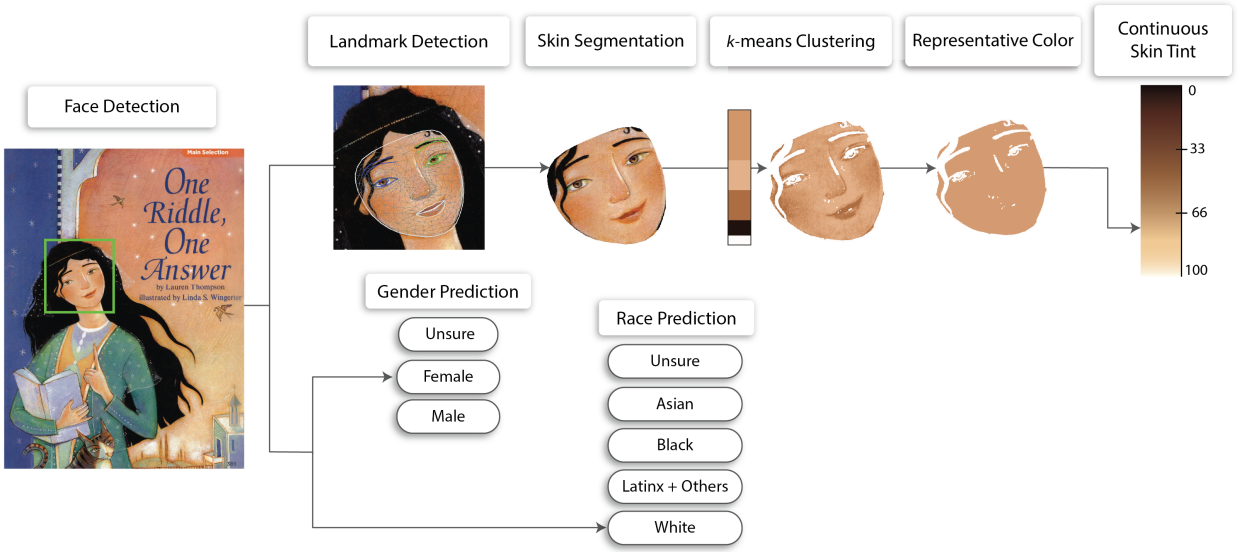
Figure 1: Google Searches Over Time: Homeschooling and Religious Collection



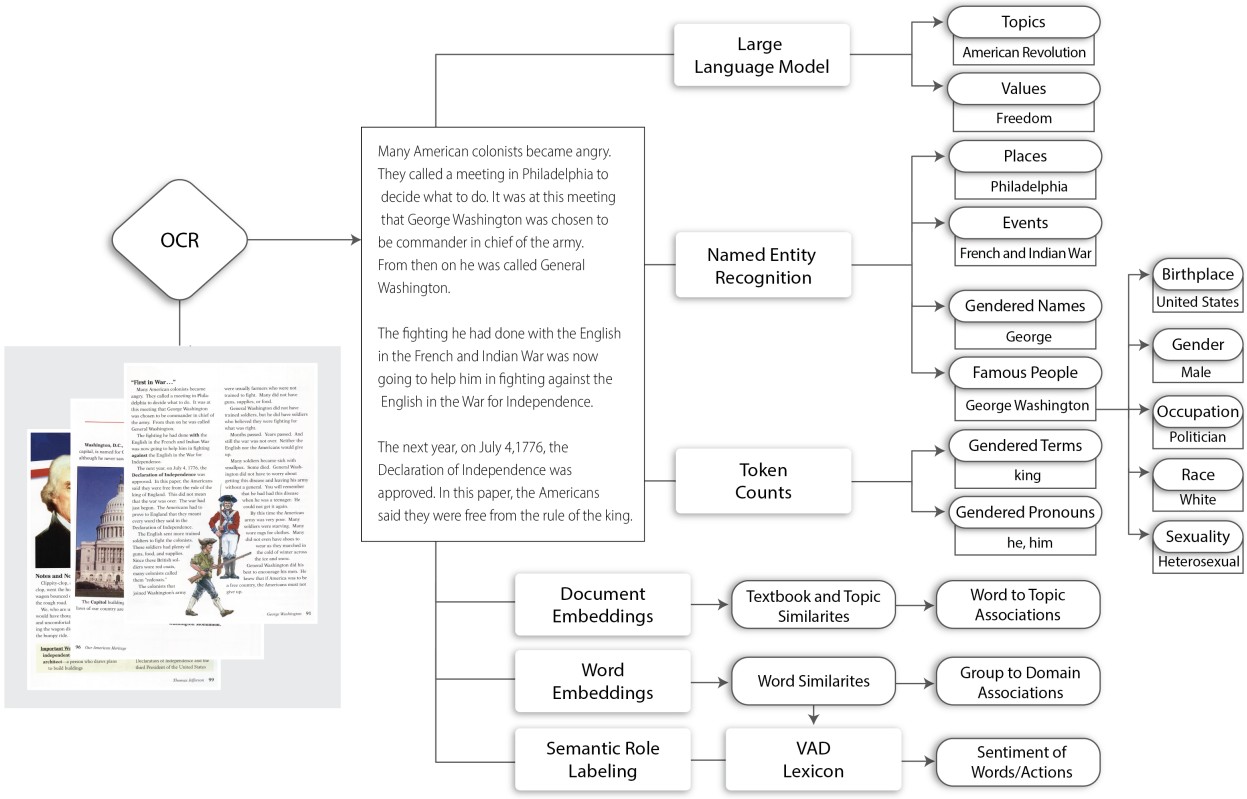

*Notes*: This figure illustrates the Google search trends for the search term homeschool and each of the education publishing companies in the religious school textbook collections; Abeka, Acceleration Christian Education (ACE), and BJU Press (BJU). Note that we use the company topics created by Google Trends instead of the search terms when looking at these three publishers. This figure shows the increase in search interest for homeschooling and Abeka curriculum after the start of the COVID-19 crisis.

Figure 2: Data Extraction Pipelines: Image and Text Analysis

(a) Image Analysis Pipeline

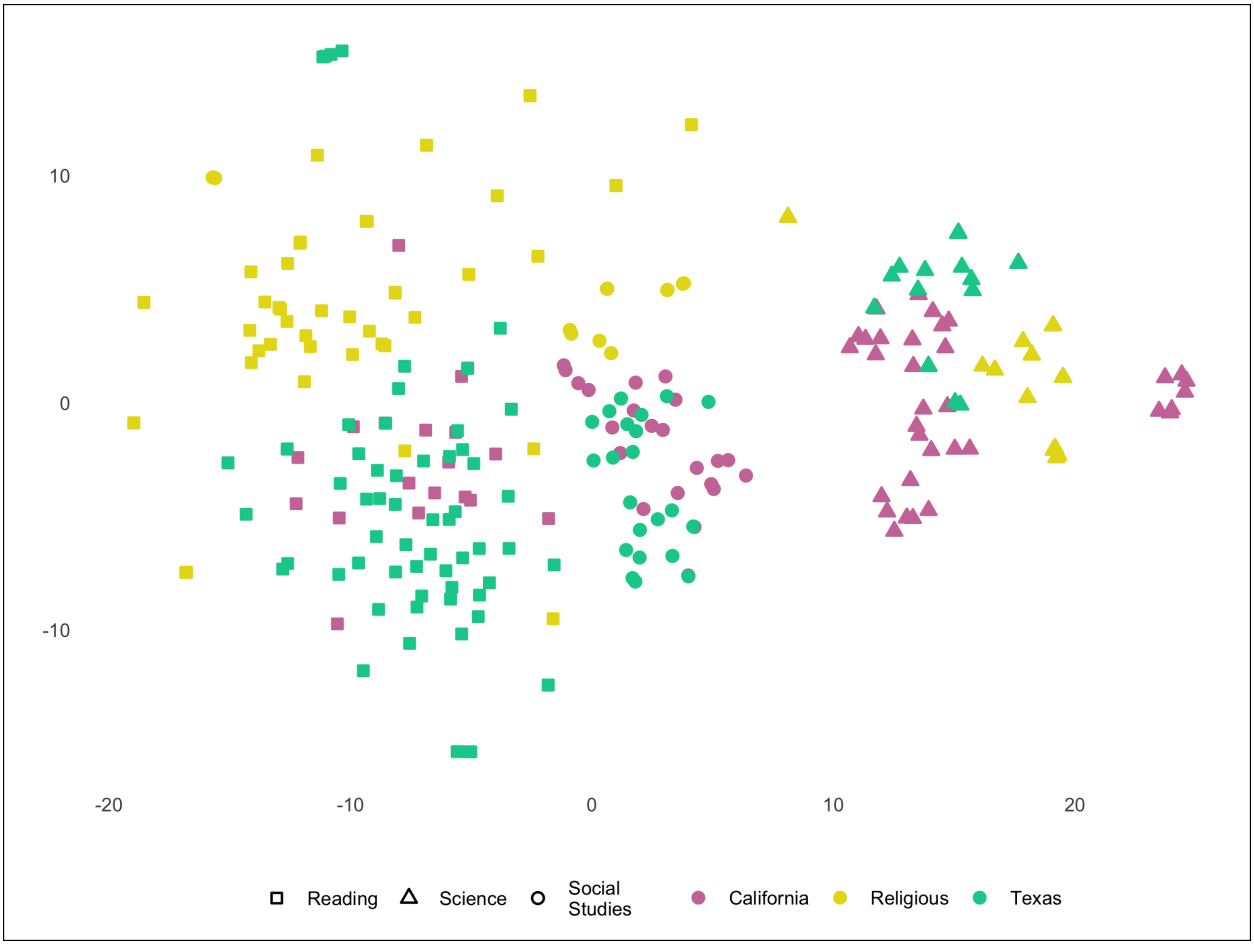



(b) Text Analysis Pipeline




*Notes*: This figure illustrates the steps in our image analysis (top) and text analysis (bottom) pipelines beginning with the raw scan of each page in our sample of books.

Figure 3: Book Embeddings: Similarities between Books across Collections and Subjects

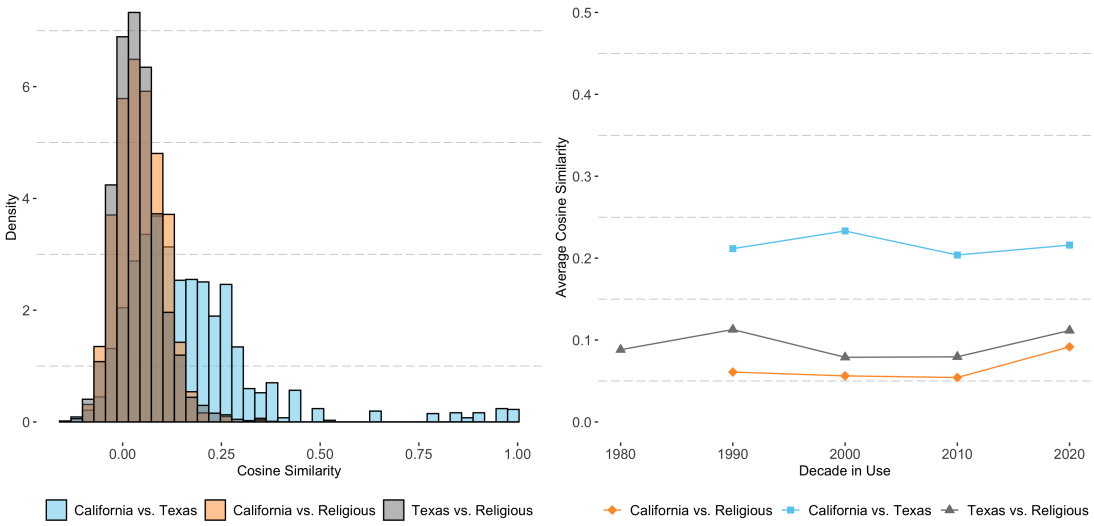



*Notes*: Each observation represents a book from a specific collection (color) and subject (shape). Their content, originally represented in a 300-dimensional space, is here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations. Among collections, pink signifies California, yellow signifies Religious, and green signifies Texas. Among subjects, the squares (□) signify Reading, the triangles (△) signify Science, and the circles (○) signify Social Studies.

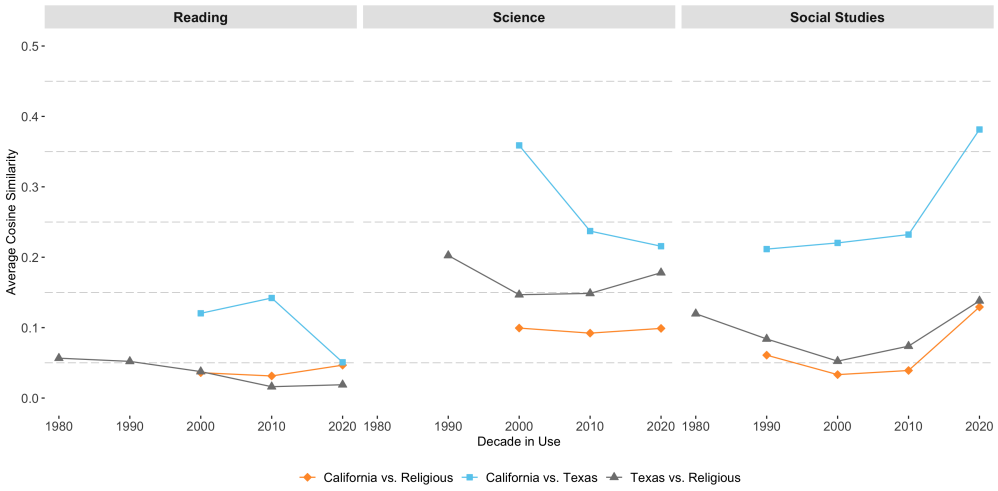
Figure 4: Average Cosine Similarity Between Book Embeddings by Collection Pairs

(a) Histogram of Cosine Similarities Between Book Embeddings by Collection Pairs

(b) Average Similarity Between Book Embeddings by Collection Pairs over Time



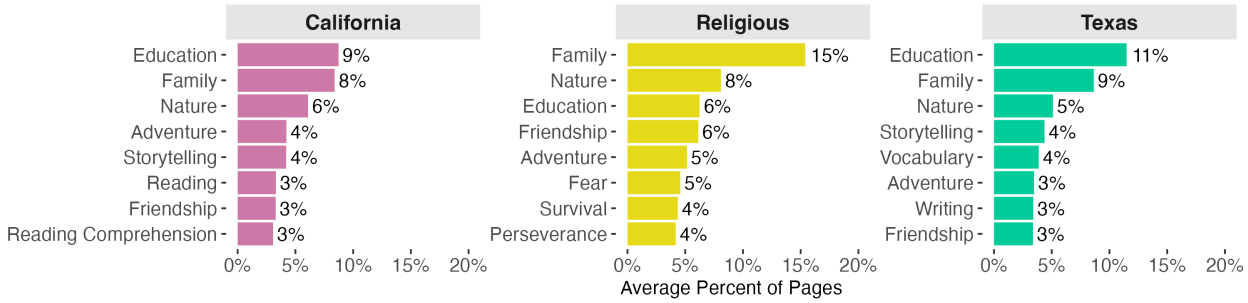

(c) Average Similarity Between Book Embeddings by Collection Pairs and Subject over Time



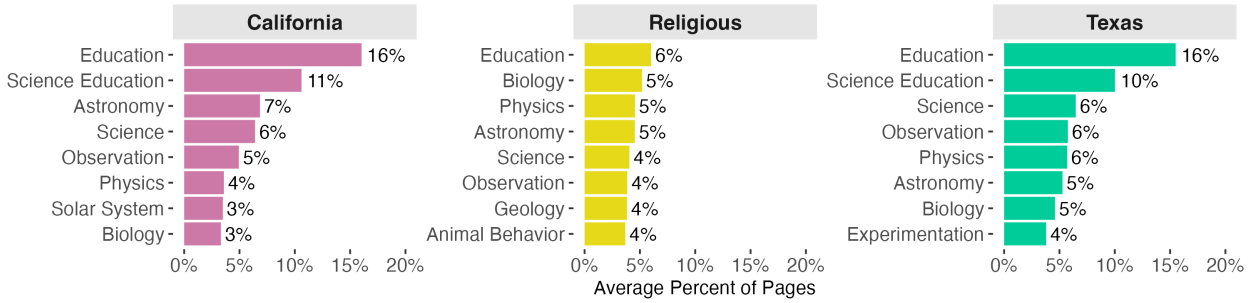

*Notes*: In Figure (a), we plot a histogram of the cosine similarities between any two books in different collections (e.g., California vs. Religious) that share the same subject, grade, and year in use. We separate the densities by collection pairs (e.g. one California textbook and one religious school textbook). Note that each book pair is repeated for every year those two books are both in use. In Figure (b), we average the cosine similarities of book pairs for each decade and collection pair; this average is weighted equally across subject, grade, and year in use. Figure (c) follows the same approach as Figure (b) except that we calculate the averages separately by subject.

Figure 5: Top Topics by Collection

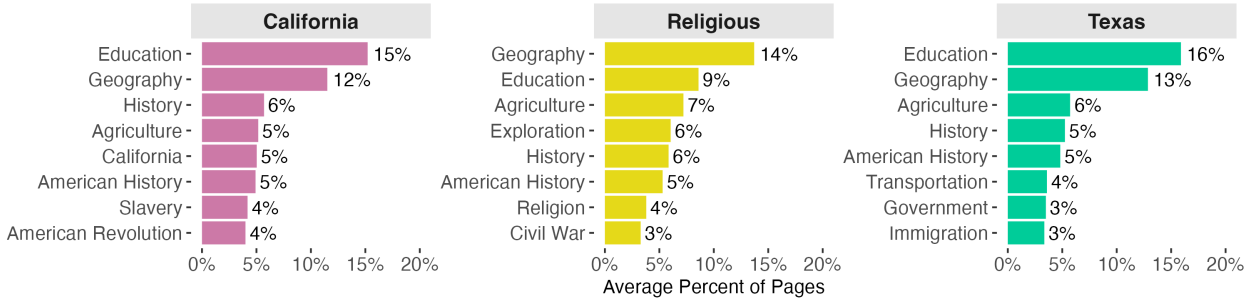
(a) Reading




(b) Science

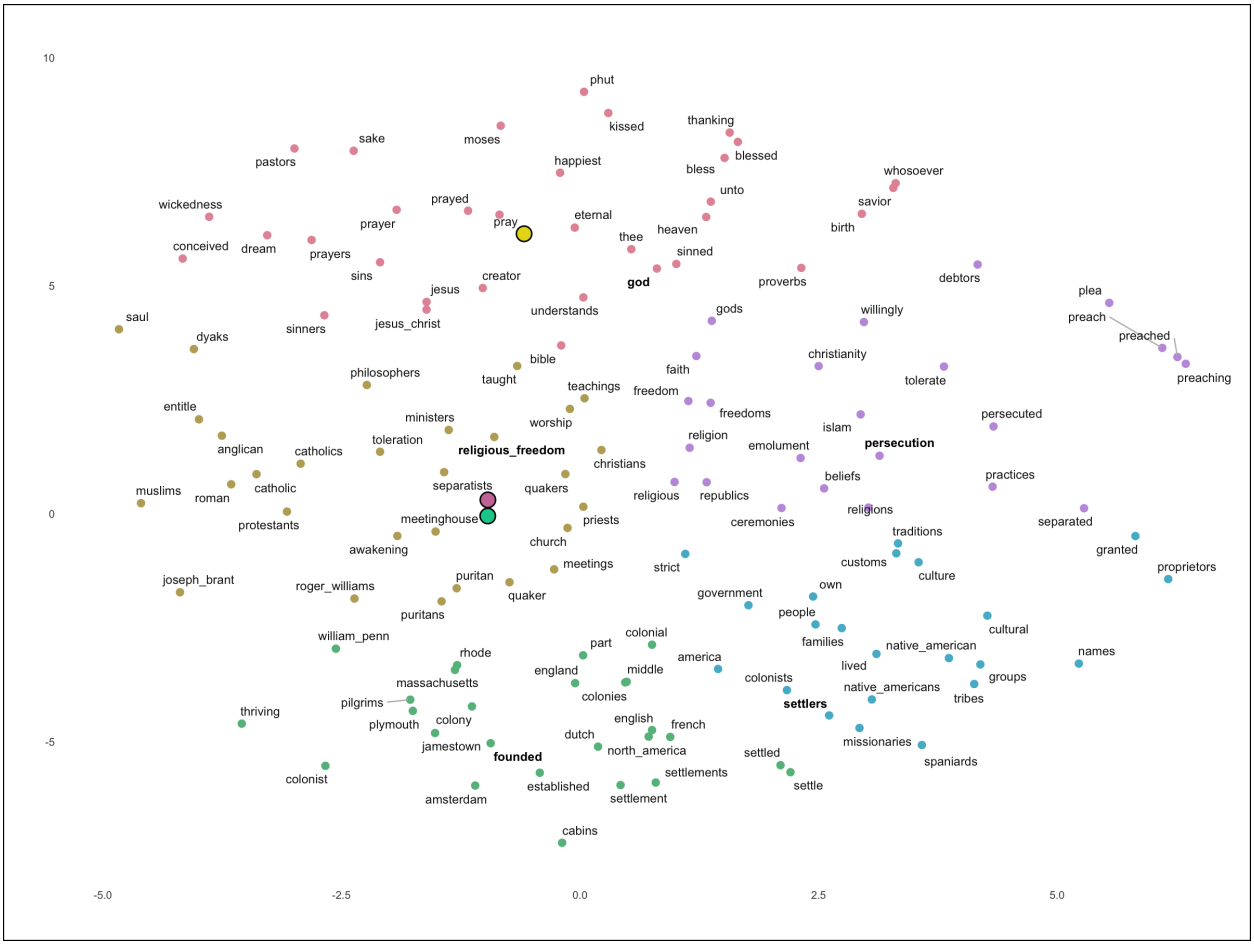



(c) Social Studies



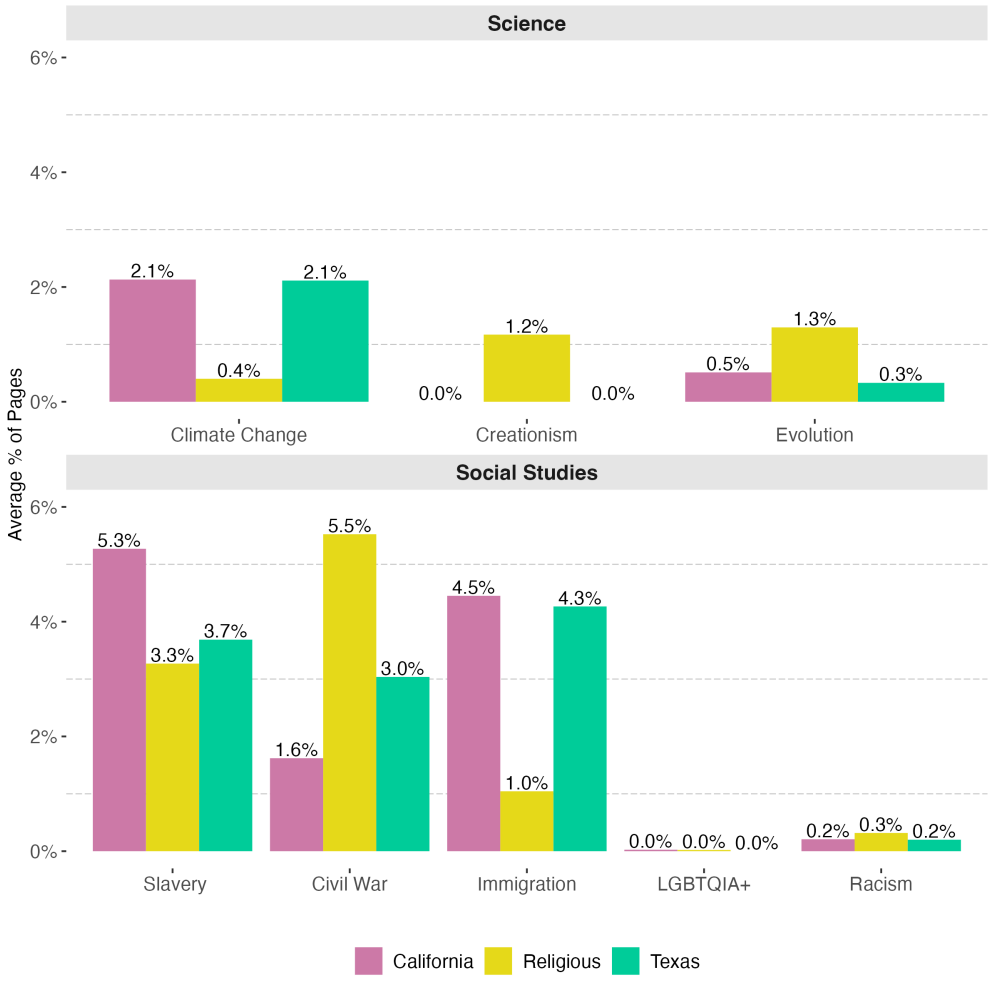

*Notes:* This figure shows the top topics (e.g. topics that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure 6: Religious Topics: Topic and Word Embeddings




*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

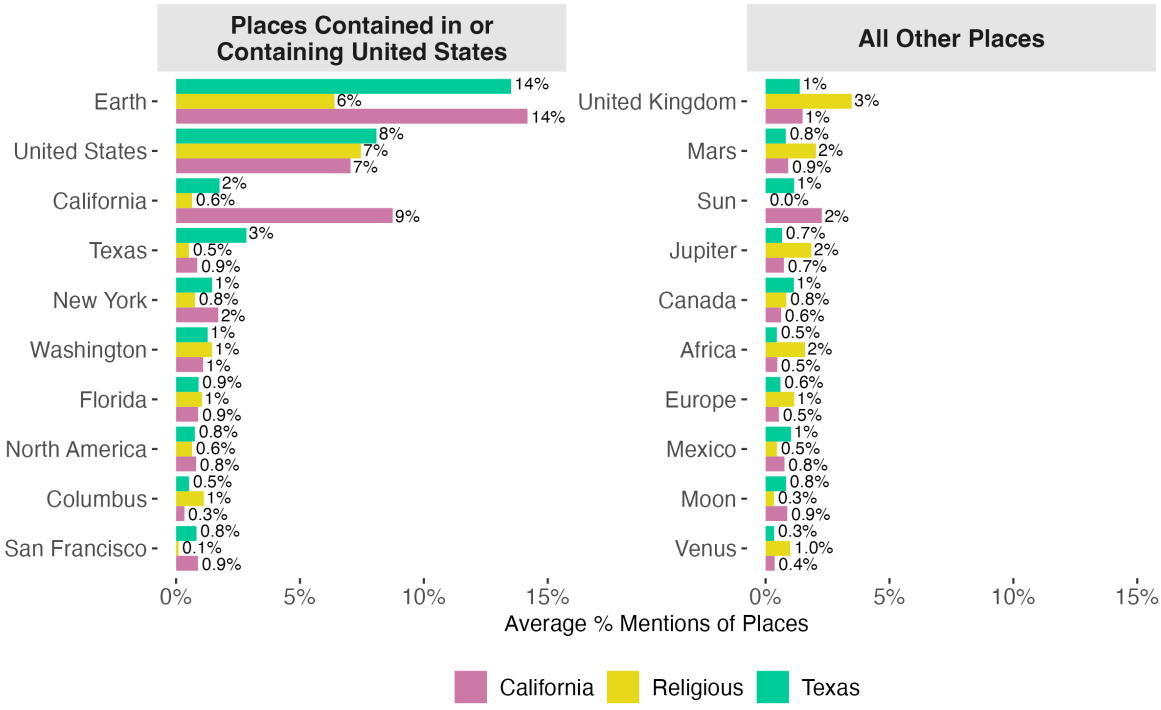
Figure 7: Average Percent of Pages Containing Contentious Topics



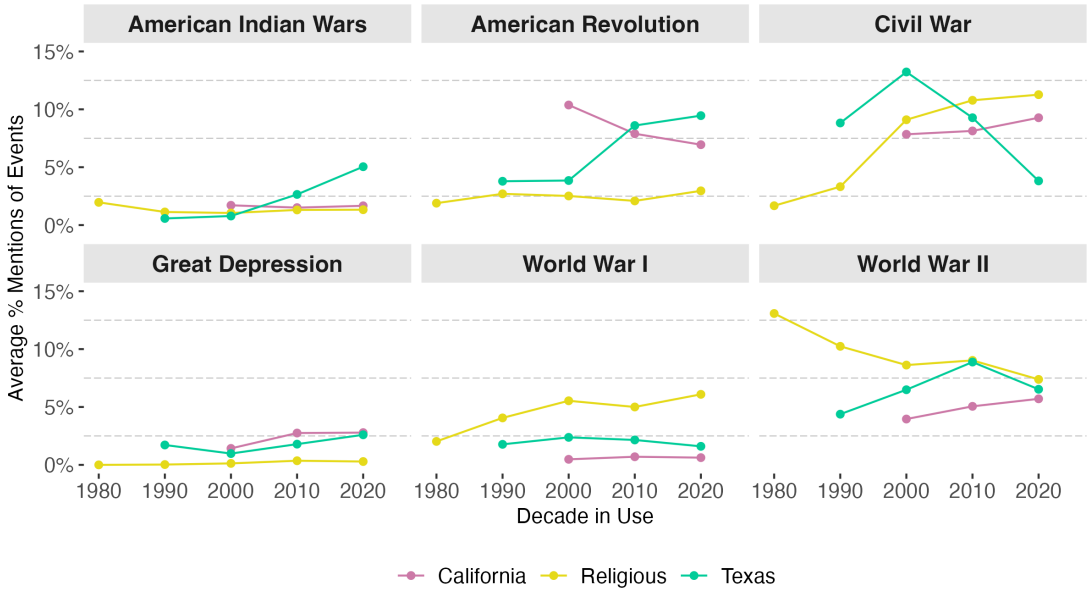

*Notes*: This figure illustrates the average percentage of pages within a specific subject that contain a specific topic. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure 8: Geography and History

(a) Geography



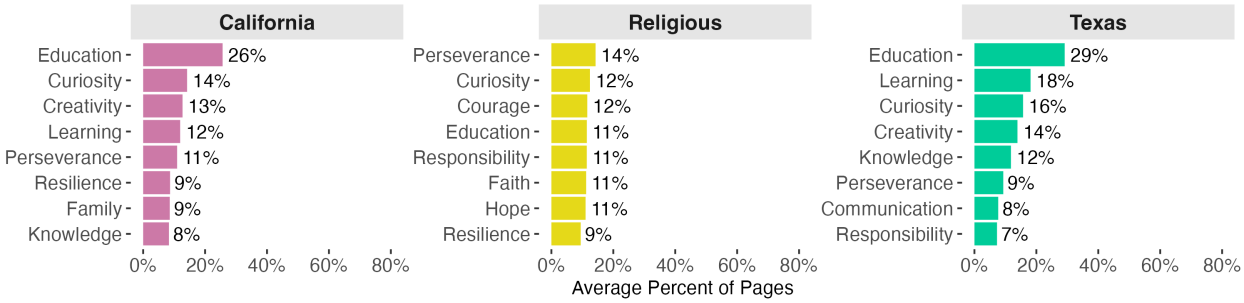

(b) Historical Events



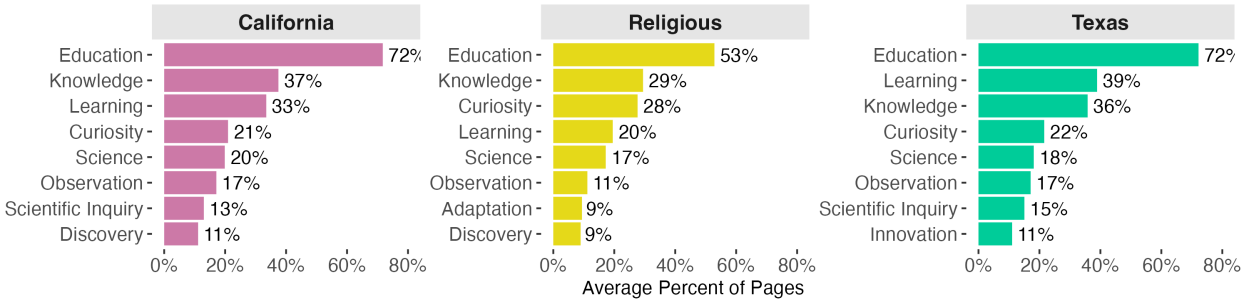

*Notes*: Panel (a) illustrates the most frequently mentioned geographic locations contained in or containing the United States and the most frequently mentioned geographic locations for all other places by collection. Panel (b) illustrates the most frequently mentioned events by collection over time. We determine both location and events using identified by Named Entity Recognition. To determine the average percentage of mentions for each location or event, we first calculate the percentage of mentions in each collection, subject, grade, and year in use. We then average these percentages at the collection and decade level to ensure that pages from each grade and subject are weighted equally in the final averages.

Figure 9: Top Values by Collection

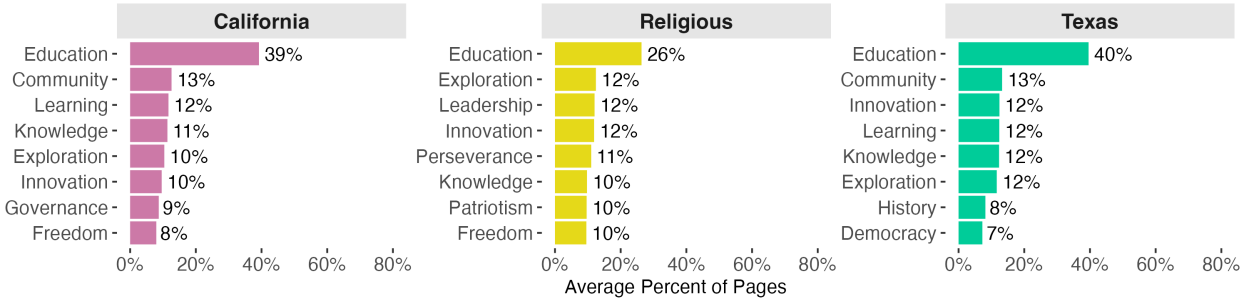
(a) Reading




(b) Science



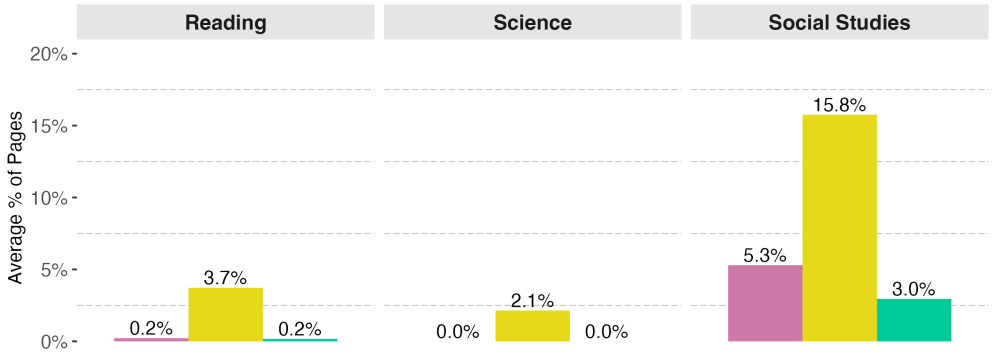

(c) Social Studies



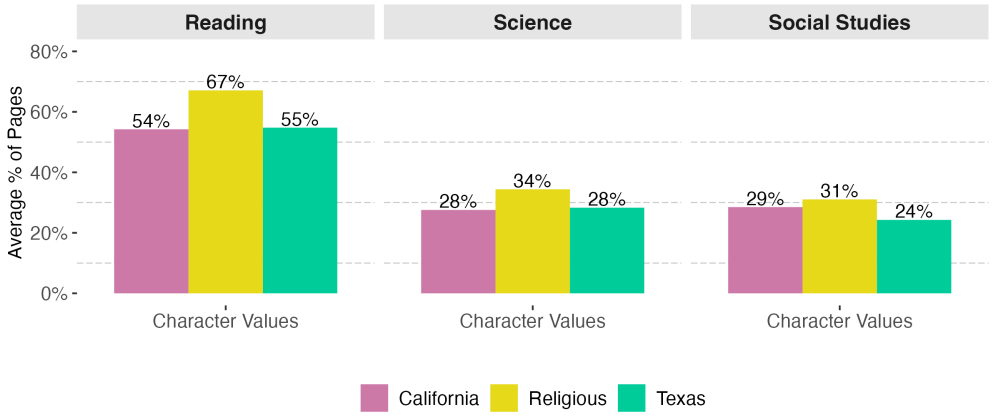

*Notes:* This figure shows the top values (e.g. values that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each value, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages by collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure 10: Specific Categories of Values

(a) Average Percent of Pages Containing Religious Values by Subject



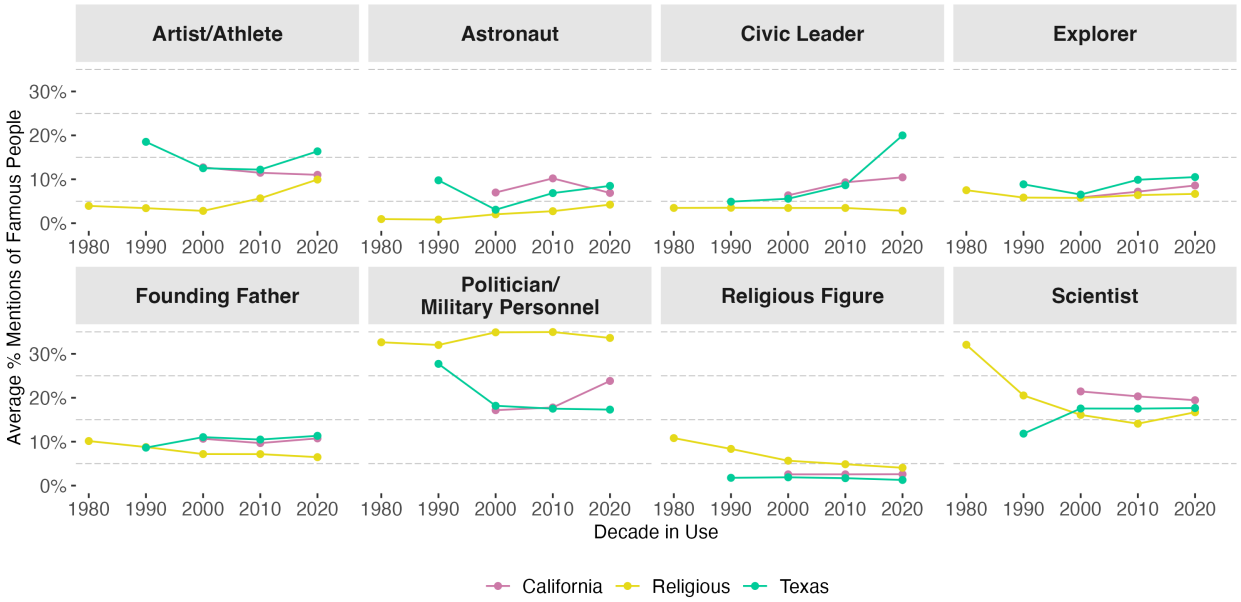

(b) Average Percent of Pages Containing Character Values by Subject




*Notes:* This figure shows the average percentage of pages within a specific subject that contain either (a) a religious or (b) a character value. To calculate the average percentage of pages containing each group of values, we first calculate the percentage of pages for each subject, grade, and year in use. We then average these percentages by collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure 11: Historical Figures

(a) Mentions



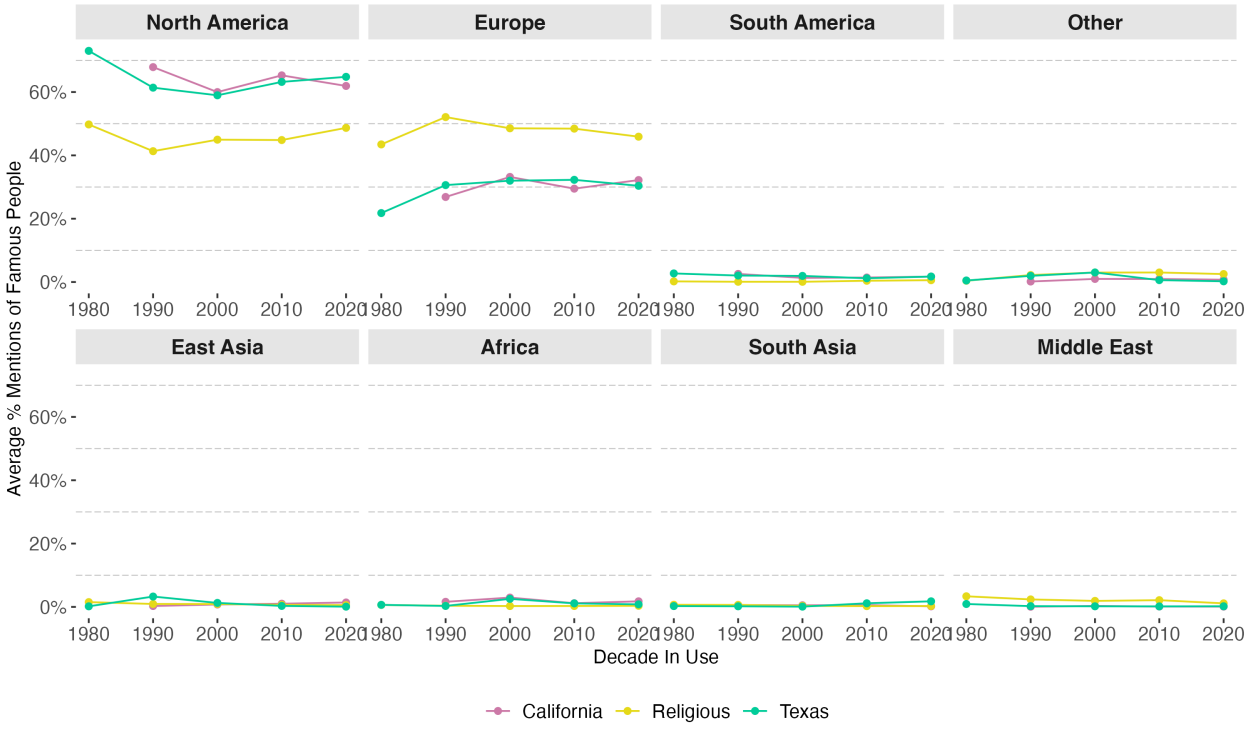

(b) Context



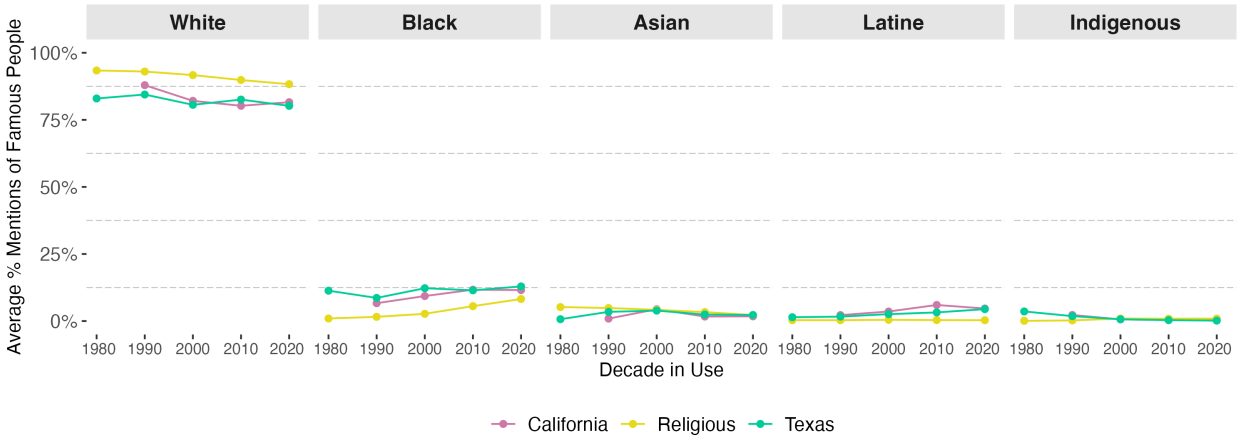

*Notes*: This figure shows our analysis of famous people by occupation. Panel (a) figure shows the number of times a famous person with a specific occupation is mentioned in a collection over time. To determine the average percentage of mentions for each occupation, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection and decade in use level to ensure that pages from each grade and subject are weighted equally in the final averages. Panel (b) shows the average Valence, Arousal, and Dominance scores of the top 50 words that are closest to the word embeddings of the names of famous people that have a given occupation.

Figure 12: Birthplaces and Races of Named Famous Figures

(a) Birthplaces



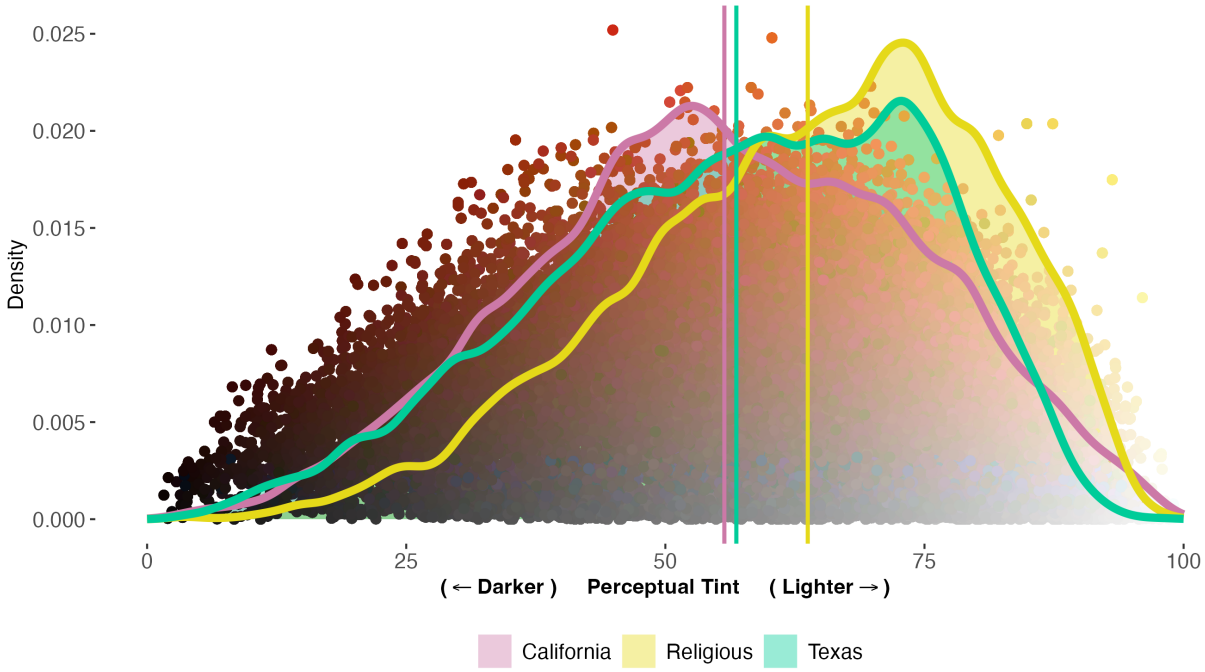

(b) Race



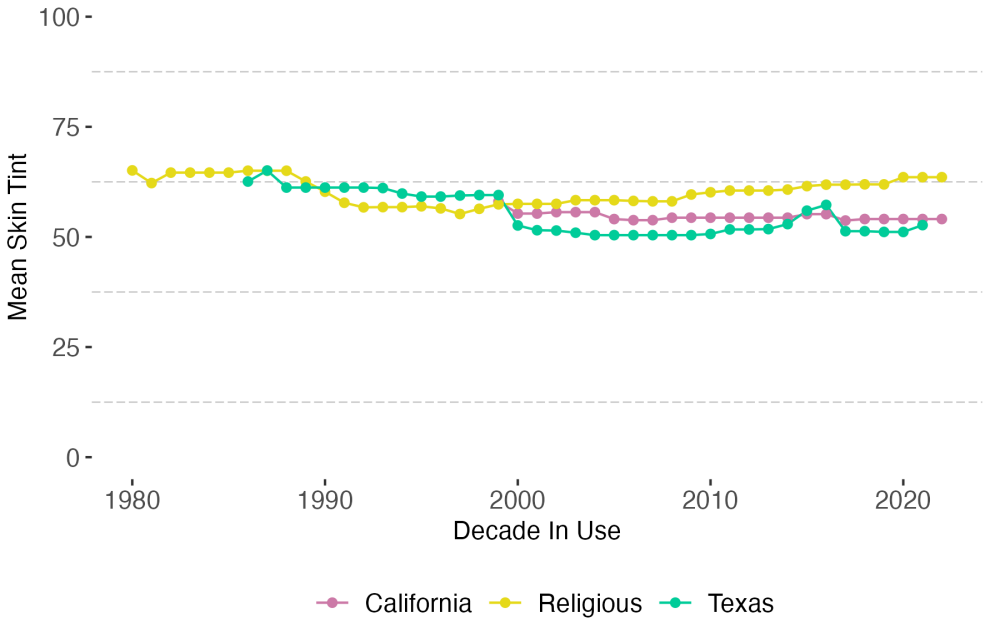

*Notes*: Panel (a) of figure shows the birthplaces of all famous people mentioned in a collection by decade in use. Panel (b) of this figure shows the average percent of mentions of famous people that are of a specific race in each collection by decade in use. To determine the average percentage of mentions for each race, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection and decade in use level to ensure that mentions of famous people from each grade, subject, and year in use are weighted equally in the final averages.

Figure 13: Skin Color of Pictured Faces by Collection

(a) Skin Tint Distributions



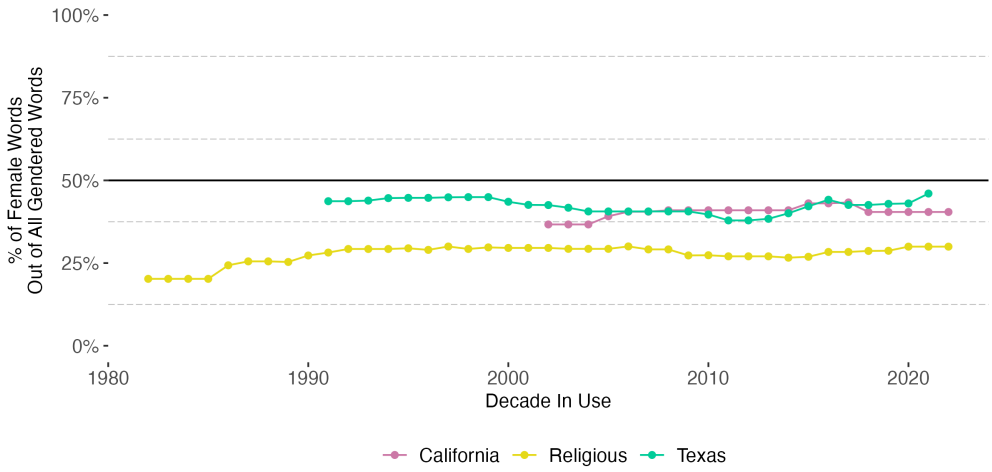

(b) Skin Tint Over Time



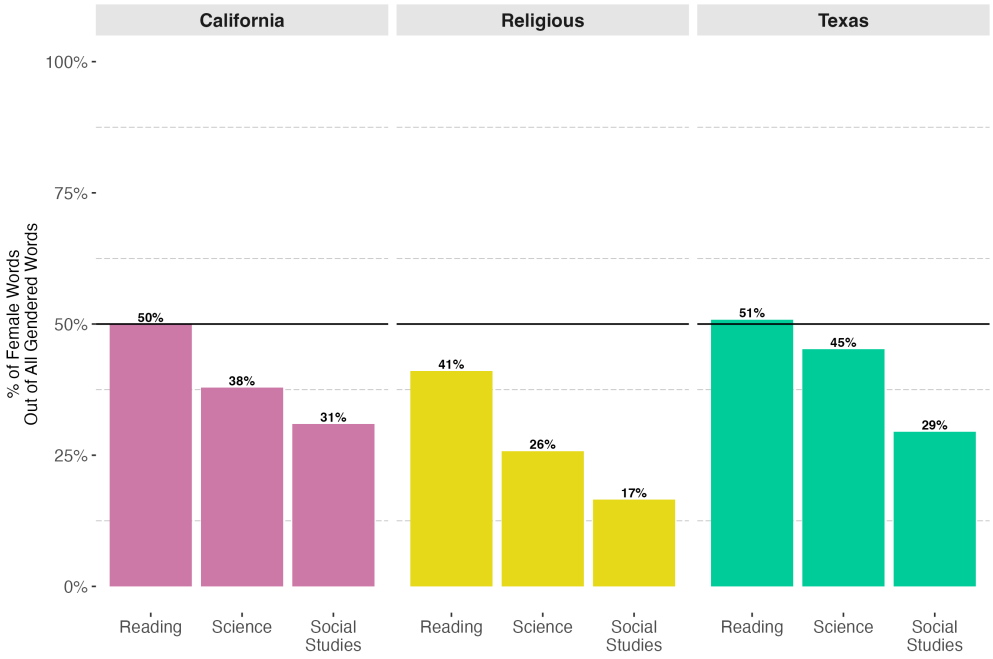

*Notes*: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze, focusing on faces considered to be human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B). Panel (a) shows the distribution of skin color tint for faces detected. The mean for each distribution is denoted with a dashed line. Panel (b) shows these patterns over time.

Figure 14: Female Representation in Text

(a) Over Time



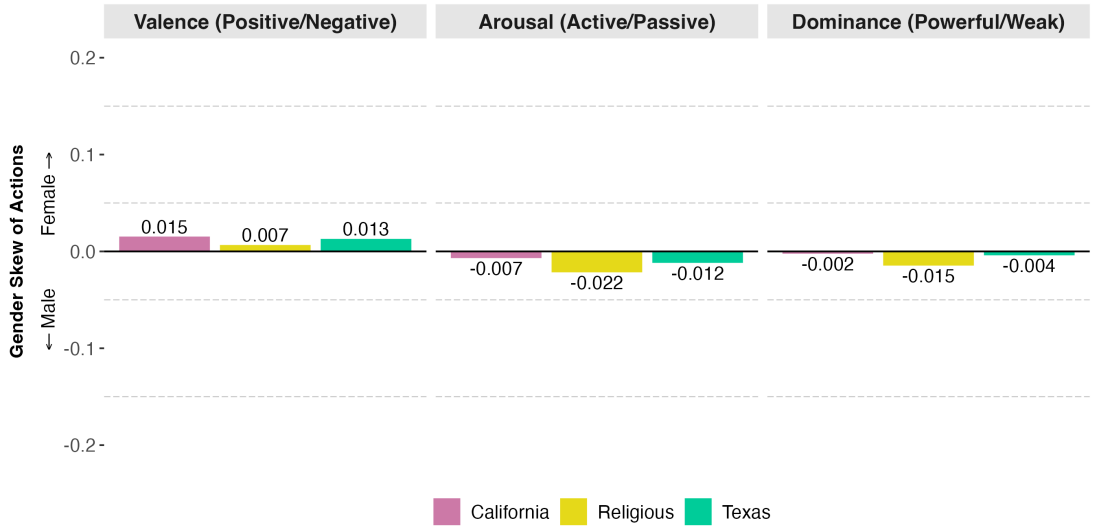

(b) By Subject



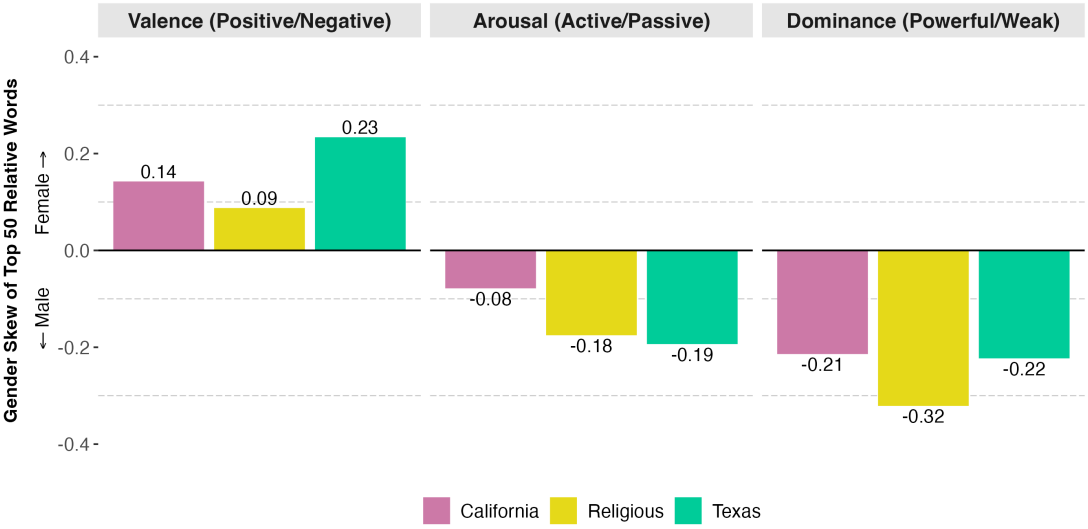

*Notes*: This figure shows the average percent of female words out of all gendered words (a) over time by collection and decade in use and (b) by collection and subject. Note that female words includes female names, female pronouns, and female terms such as queen or brother. To calculate the average percentage of female words by collection and decade in use, we first calculate the percentage each collection, subject, grade, and year in use. We then average these percentages at the collection and decade in use level, to ensure that textbooks in each subject, grade and year in use are weighted equally in the final averages. To calculate the average percentage of female words by collection and subject, we first calculate the percentage each collection, subject, grade, and year in use. We then average these percentages at the collection and subject level, to ensure that textbooks in each grade and year in use are weighted equally in the final averages.

Figure 15: Sentiment (Valence, Arousal, Dominance) by Gender

(a) VAD Scores of Female vs. Male Actions

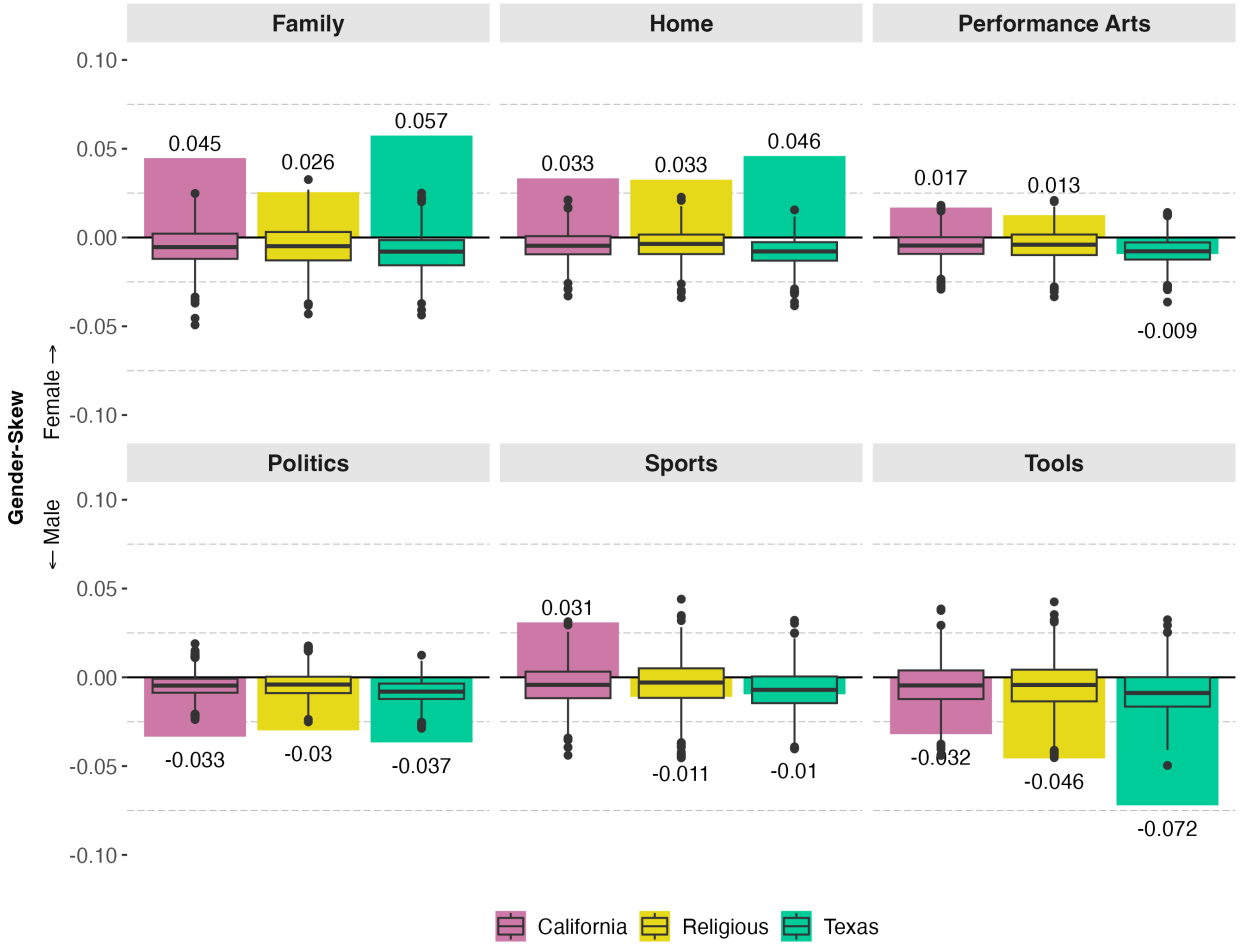



(b) VAD Scores of Top 50 Relative Words by Gender




*Notes*: This figure shows the Valence, Arousal, and Dominance (VAD) scores of words and actions associated with males and females and then takes the differences of these score to find the gender skew of that dimension. Panel (a) shows the difference in the VAD scores of actions associated with female words and the actions associated with male words. If the skew is positive, it means that context in which females are discussed uses words that are higher on that Valence, Arousal, or Dominance dimension than the context that males are discussed in. We use Semantic Role Labeling to find the actions associated with males versus females. Panel (b) shows the difference in the VAD scores of the top 50 relative words closest to female words and relative words closest to male words. We call this difference in scores the gender skew of scores. If the skew is positive, it means that context in which females are discussed uses words that have higher Valence, Arousal, or Dominance scores than the context that males are discussed in. To find the top relative words associated with females vs. males, we find the words that are closest to females in word embedding space and farthest from the males. We do the a similar exercise to find the top relative words associated with males vs. females. We estimate gender skew in this manner across 50 different word embedding models, estimated on the same source text and find the average gender skew across models. Note that we are limiting our analysis to words that are included in the VAD lexicon.

Figure 16: Gender Skew By Collection and Domain




*Notes*: This figure shows the gender skew of certain domains by calculating the difference in the association between female pronouns and a given domain with male pronouns and given domain. If the gender skew is positive, it means that words belonging to a specific domain are closer to female pronouns on average (e.g. meaning those words are more likely to be discussed in the same context as female pronouns). The boxplots show the distribution of gender skew in a random set of words the same size as the domain of interest. If the estimated gender skew is larger than the whiskers of the boxplot, this means that the gender skew for a collection is larger than that of a random set of words and therefore, likely to not have arisen by chance. Note that we are limiting our analysis to words that are included in the VAD lexicon.

## IX Main Tables

Table 1: Summary Statistics

| | California (1) | Religious (2) | Texas (3) |
|---|---|---|---|
| *Collection Totals* | | | |
| Total number of textbooks | 80 | 85 | 96 |
| Number of Reading textbooks | 22 | 55 | 60 |
| Number of Science textbooks | 36 | 17 | 14 |
| Number of Social Science textbooks | 22 | 13 | 22 |
| Number of Grade 3 textbooks | 42 | 50 | 55 |
| Number of Grade 5 textbooks | 38 | 35 | 41 |
| Total number of pages | 29,262 | 19,814 | 39,235 |
| Total number of detected faces | 23,664 | 14,065 | 22,040 |
| *Sample Period* | | | |
| Earliest year in use | 1999 | 1980 | 1986 |
| Latest year in use | 2022 | 2022 | 2021 |
| Average number of years textbooks were in use | 7 | 13 | 8 |
| *Weighted Average: Topics and Values* | | | |
| Average number of classified topics per page | 8 | 7 | 8 |
| Average number of classified values per page | 6 | 6 | 6 |
| *Weighted Average: Skin Color* | | | |
| Average skin tint of faces | 55 | 60 | 53 |
| *Weighted Average Percent: Putative Race of Faces* | | | |
| Average percent of faces classified as Asian | 5% | 2% | 3% |
| Average percent of faces classified as Black | 31% | 25% | 35% |
| Average percent of faces classified as Latine | 5% | 2% | 3% |
| Average percent of faces classified as White | 48% | 57% | 43% |
| Average percent of faces classified as other race | 1% | 1% | 1% |
| Average percent of faces classified as unsure race | 11% | 13% | 15% |
| *Weighted Average Percent: Putative Race of Famous People* | | | |
| Average percent of famous people classified as Asian | 3% | 4% | 3% |
| Average percent of famous people classified as Black | 11% | 3% | 11% |
| Average percent of famous people classified as Indigenous | 1% | 1% | 1% |
| Average percent of famous people classified as Latine | 5% | 0% | 3% |
| Average percent of famous people classified as White | 81% | 92% | 82% |
| *Weighted Average Percent: Gender of Faces* | | | |
| Average percent of faces classified as Female | 48% | 41% | 44% |
| Average percent of faces classified as Male | 50% | 55% | 53% |
| Average percent of faces classified as unsure gender | 3% | 4% | 4% |
| *Weighted Average Percent: Gendered Words* | | | |
| Average percent of female gendered words | 40% | 27% | 43% |
| Average percent of famous people classified as female | 19% | 7% | 19% |

*Notes:* For all averages, we first compute the average for all textbooks within the same subject, grade, and year in use. We then average across these initial averages for each subject, grade, and year in use. This methodology ensures that textbooks used over longer periods are given greater weight and that equal emphasis is placed on each subject and grade. Average number of topics and values per page on conditional on having at least one topic/value identified on a page.

Table 2: Top 20 Famous People by Collection

| Collection | Name | Race | Gender | Occupation | Sexuality | Rank | Mentions | Textbooks |
|---|---|---|---|---|---|---|---|---|
| California | Benjamin Franklin | White | Male | Founding Father | Hetero | 1 | 383 | 32 |
| California | Thomas Jefferson | White | Male | Founding Father | Hetero | 2 | 322 | 31 |
| California | George Washington | White | Male | Founding Father | Hetero | 3 | 417 | 30 |
| California | Abraham Lincoln | White | Male | Politician | Hetero | 4 | 197 | 25 |
| California | Martin Luther King Jr. | Black | Male | Civic Leader | Hetero | 5 | 133 | 24 |
| California | John Adams | White | Male | Founding Father | Hetero | 6 | 233 | 18 |
| California | Abigail Adams | White | Female | Companion | Hetero | 7 | 81 | 18 |
| California | Frederick Douglass | Black | Male | Civic Leader | Hetero | 8 | 95 | 17 |
| California | Paul Revere | White | Male | Military | Hetero | 9 | 194 | 16 |
| California | James Madison | White | Male | Founding Father | Hetero | 10 | 139 | 16 |
| California | Anne Hutchinson | White | Female | Religious Figure | Hetero | 11 | 88 | 16 |
| California | Susan B. Anthony | White | Female | Civic Leader | Hetero | 12 | 83 | 16 |
| California | John Hancock | White | Male | Politician | Hetero | 13 | 67 | 16 |
| California | Harriet Tubman | Black | Female | Writer | Hetero | 14 | 64 | 16 |
| California | William Clark | White | Male | Explorer | Hetero | 15 | 57 | 16 |
| California | Theodore Roosevelt | White | Male | Politician | Hetero | 16 | 32 | 16 |
| California | Alexander Hamilton | White | Male | Founding Father | Hetero | 17 | 217 | 15 |
| California | Neil Armstrong | White | Male | Astronaut | Hetero | 18 | 145 | 15 |
| California | George III | White | Male | Politician | Hetero | 19 | 124 | 15 |
| California | Patrick Henry | White | Male | Politician | Hetero | 20 | 112 | 15 |
| | | | | | | | | |
| Religious | George Washington | White | Male | Founding Father | Hetero | 1 | 310 | 33 |
| Religious | Abraham Lincoln | White | Male | Politician | Hetero | 2 | 369 | 27 |
| Religious | Benjamin Franklin | White | Male | Founding Father | Hetero | 3 | 158 | 25 |
| Religious | Paul Revere | White | Male | Military | Hetero | 4 | 185 | 19 |
| Religious | Thomas Jefferson | White | Male | Founding Father | Hetero | 5 | 146 | 18 |
| Religious | Noah Webster | White | Male | Linguist | Hetero | 6 | 280 | 17 |
| Religious | John F. Kennedy | White | Male | Politician | Hetero | 7 | 102 | 17 |
| Religious | James Madison | White | Male | Founding Father | Hetero | 8 | 50 | 16 |
| Religious | Robert E. Lee | White | Male | Military | Hetero | 9 | 132 | 15 |
| Religious | Daniel Boone | White | Male | Explorer | Hetero | 10 | 103 | 14 |
| Religious | David Livingstone | White | Male | Explorer | Hetero | 11 | 59 | 13 |
| Religious | Ulysses S. Grant | White | Male | Politician | Hetero | 12 | 56 | 12 |
| Religious | Charles The Bald | White | Male | Politician | Hetero | 13 | 45 | 12 |
| Religious | William Bradford | White | Male | Politician | Hetero | 14 | 31 | 12 |
| Religious | Rudyard Kipling | White | Male | Writer | Hetero | 15 | 20 | 12 |
| Religious | John Ball | White | Male | Religious Figure | Hetero | 16 | 12 | 12 |
| Religious | John Smith | White | Male | Explorer | Hetero | 17 | 123 | 11 |
| Religious | Woodrow Wilson | White | Male | Politician | Hetero | 18 | 31 | 11 |
| Religious | Francis Scott Key | White | Male | Writer | Hetero | 19 | 20 | 11 |
| Religious | Hans C. Andersen | White | Male | Writer | Hetero | 20 | 17 | 11 |
| | | | | | | | | |
| Texas | George Washington | White | Male | Founding Father | Hetero | 1 | 531 | 45 |
| Texas | Benjamin Franklin | White | Male | Founding Father | Hetero | 2 | 377 | 33 |
| Texas | Abraham Lincoln | White | Male | Politician | Hetero | 3 | 422 | 31 |
| Texas | Thomas Jefferson | White | Male | Founding Father | Hetero | 4 | 245 | 31 |
| Texas | Alexander Hamilton | White | Male | Founding Father | Hetero | 5 | 245 | 27 |
| Texas | Paul Revere | White | Male | Military | Hetero | 6 | 268 | 25 |
| Texas | Martin Luther King Jr. | Black | Male | Civic Leader | Hetero | 7 | 219 | 24 |
| Texas | Neil Armstrong | White | Male | Astronaut | Hetero | 8 | 198 | 24 |
| Texas | Harriet Tubman | Black | Female | Writer | Hetero | 9 | 127 | 22 |
| Texas | John F. Kennedy | White | Male | Politician | Hetero | 10 | 214 | 21 |
| Texas | John Adams | White | Male | Founding Father | Hetero | 11 | 175 | 21 |
| Texas | Theodore Roosevelt | White | Male | Politician | Hetero | 12 | 157 | 21 |
| Texas | Thomas Edison | White | Male | Inventor | Hetero | 13 | 101 | 21 |
| Texas | William Clark | White | Male | Explorer | Hetero | 14 | 63 | 21 |
| Texas | Francis Scott Key | White | Male | Writer | Hetero | 15 | 58 | 21 |
| Texas | Franklin D. Roosevelt | White | Male | Politician | Hetero | 16 | 454 | 20 |
| Texas | Langston Hughes | Black | Male | Writer | Hetero | 17 | 68 | 20 |
| Texas | Meriwether Lewis | White | Male | Explorer | Hetero | 18 | 59 | 19 |
| Texas | Lyndon B. Johnson | White | Male | Politician | Hetero | 19 | 58 | 19 |
| Texas | Jane Addams | White | Female | Civic Leader | Lesbian | 20 | 92 | 18 |

*Notes*: This figure illustrates the most frequently famous people mentioned within a collection. We find the number of times a famous person is mentioned and the number of textbooks they appear in and then sort first by number of textbooks within a collection and then by number of mentions to find the top individuals.

Table 3: Top 5 Relative Words by Gender

| | | *Mean* | | |
|---|---|---|---|---|
| *Top Words Associated with Female (vs. Male) Pronouns* | | V | A | D |
| **California** | husband, dress, sew, skater, cooking | 0.67 | 0.48 | 0.49 |
| **Religious** | doll, skirt, mama, shawl, grandmother | 0.74 | 0.37 | 0.39 |
| **Texas** | doll, dress, fabric, holly, skirt | 0.65 | 0.31 | 0.34 |
| *Top Words Associated with Male (vs. Female) Pronouns* | | V | A | D |
| **California** | master, king, admiral, horse, lieutenant | 0.69 | 0.56 | 0.81 |
| **Religious** | harness, inventor, return, architect, horseback | 0.68 | 0.53 | 0.64 |
| **Texas** | sword, king, spear, stallion, believing | 0.62 | 0.59 | 0.74 |

*Notes*: This figure shows the top 5 relative words associated with a specific gender as well as the average Valence, Arousal, and Dominance (VAD) scores of those words. To find the top relative words associated with females vs. males, we find the words that are closest to females in word embedding space and farthest from the males. We do the a similar exercise to find the top relative words associated with males vs. females. Note that we are limiting are analysis to words that are included in the VAD lexicon.

# Online Appendix

**Separation of Church and State Curricula?**
**Examining Public and Religious Private School Textbooks***

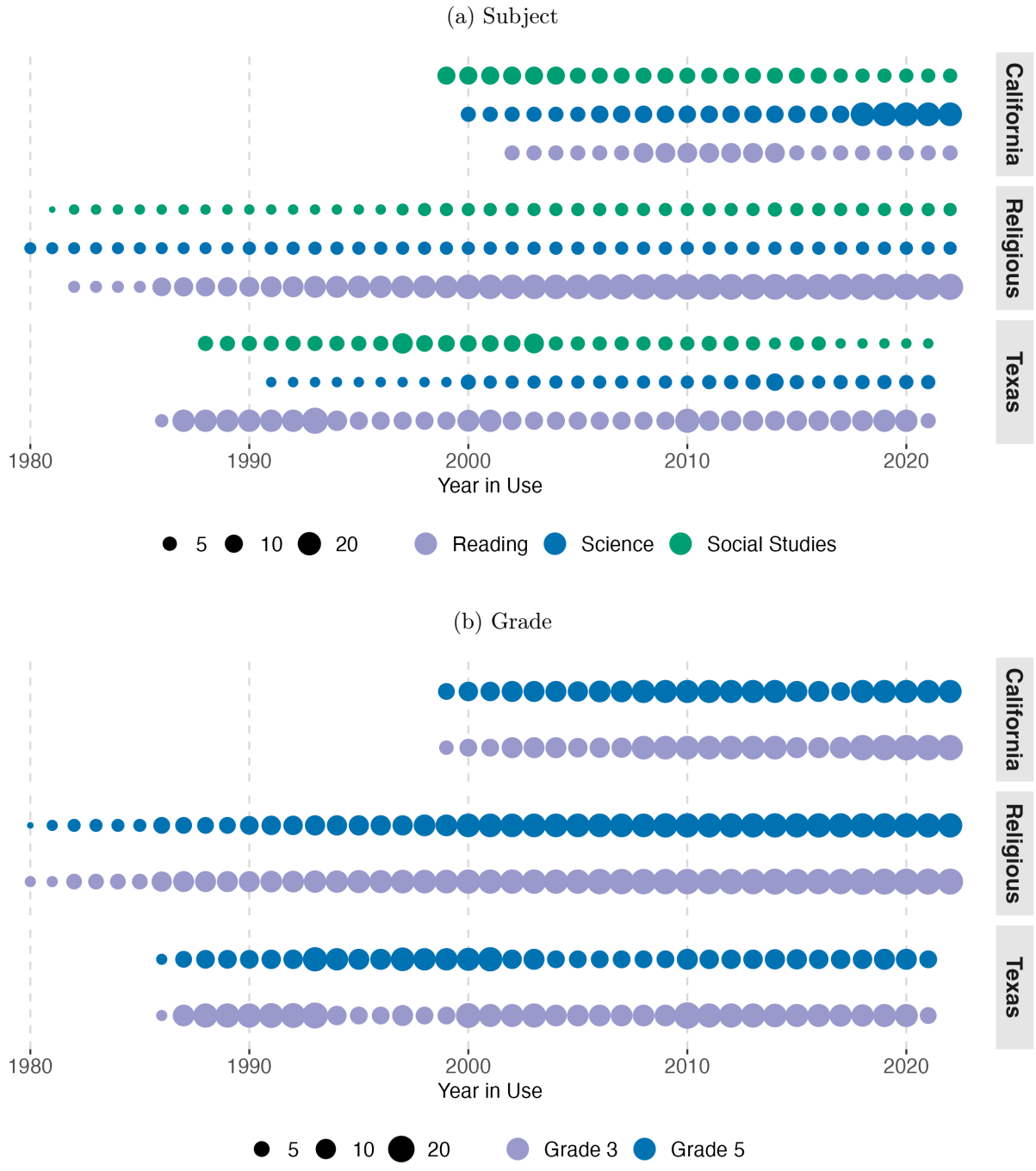
**April 2025**

**Anjali Adukia, University of Chicago, MiiE Lab, IZA, NBER**
**Emileigh Harrison, University of Chicago, MiiE Lab, Inclusive Economy Lab**

*Contact: adukia@uchicago.edu, harrisone@uchicago.edu

## A Appendix Figures

Figure A.1: Number of Textbooks in Sample by Grade, Subject, Collection, and Year in Use



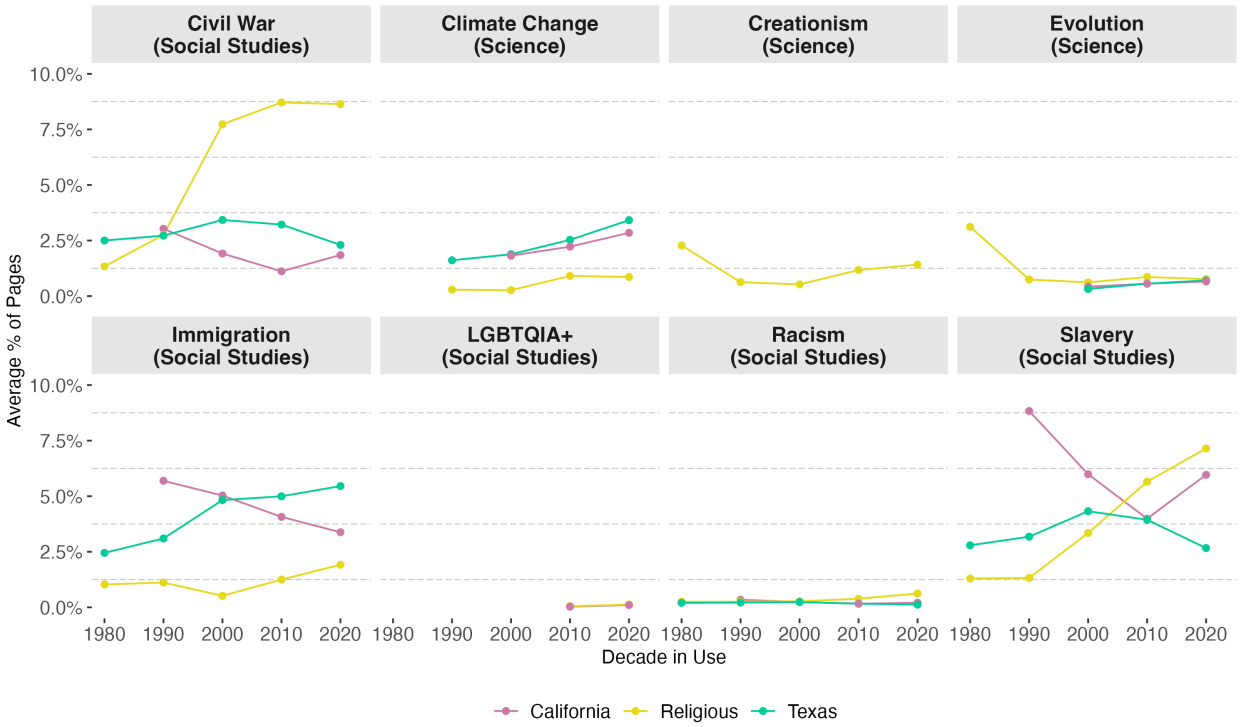

*Notes*: This figures shows the number of books in our sample that were in use within a given year by grade, subject, and collection. Note that most textbooks were in use for 7-13 years on average depending on the collection.

Figure A.2: Book Embeddings: by Collection, Subject, and Grade



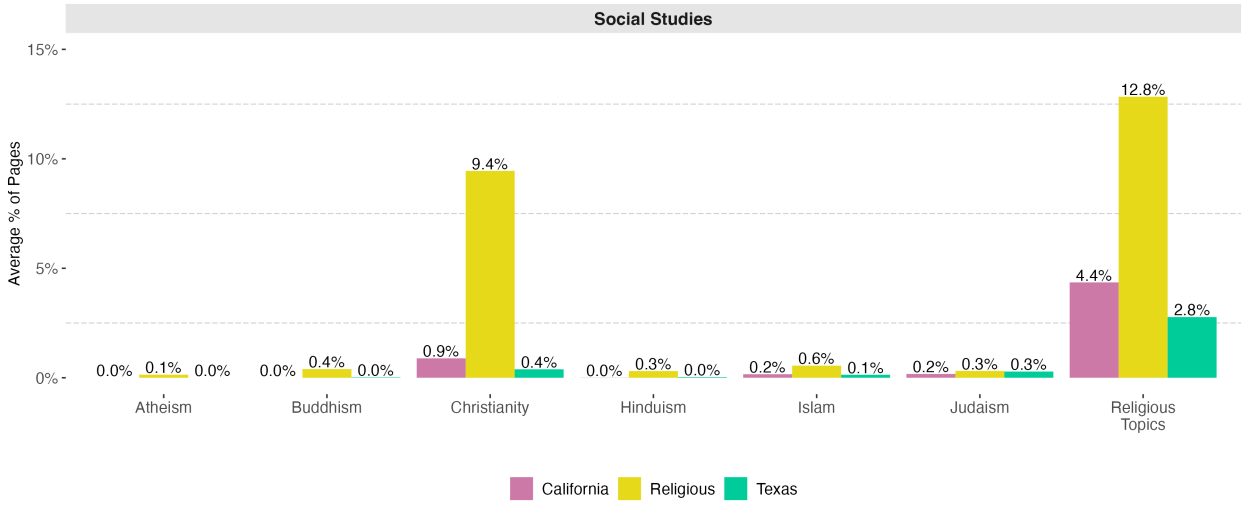

*Note*: Each observation represents a book from a specific collection (color), subject (shape), and grade (outline). The filled shapes indicate 5th grade textbooks and hollow shapes indicate 3rd grade textbooks. Their content, originally represented in a 300-dimensional space, is here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations. Among collections, pink signifies California, yellow signifies Religious, and green signifies Texas. Among subjects, the squares (□) signify Reading, the triangles (△) signify Science, and the circles (○) signify Social Studies.

Figure A.3: Average Percent of Pages Containing Contentious Topics Over Time




*Notes*: This figure illustrates the average percentage of pages within a specific subject as denoted in the parentheses that contain a specific topic over time. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, grade, and year in use. We then average these percentages at the collection level and decade level, to ensure that pages in each grade and year in use are weighted equally in the subject and collection specific final averages.

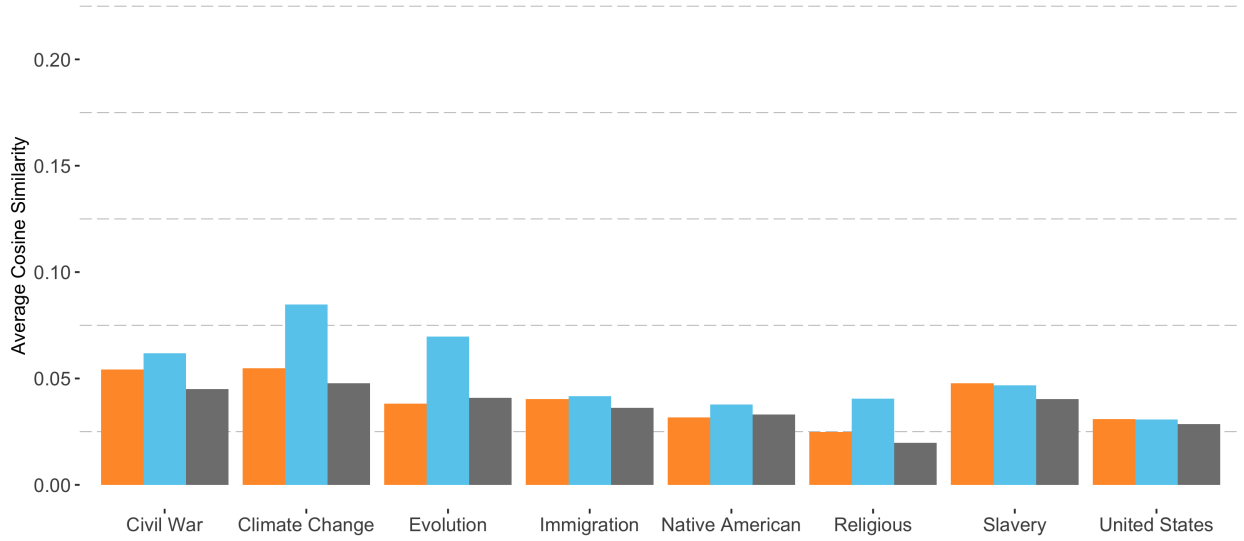
Figure A.4: Average Percent of Pages in Social Studies Textbooks Containing Religious Topics or Topics Related to a Specific Religion




*Notes*: This figure illustrates the average percentage of pages within Social Studies textbooks that contain a religious topic or topics related to a specific religion. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade and year in use are weighted equally in the final averages.

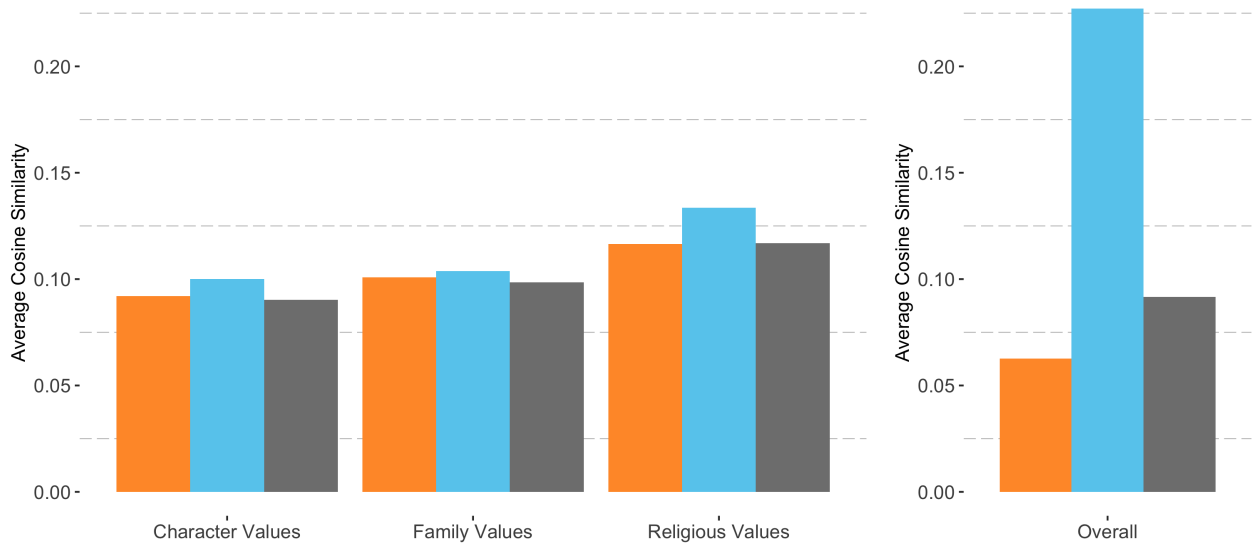
Figure A.5: Average Similarity of Page Embeddings Between Collections by Topics and Values Present

(a) Topics




(b) Values

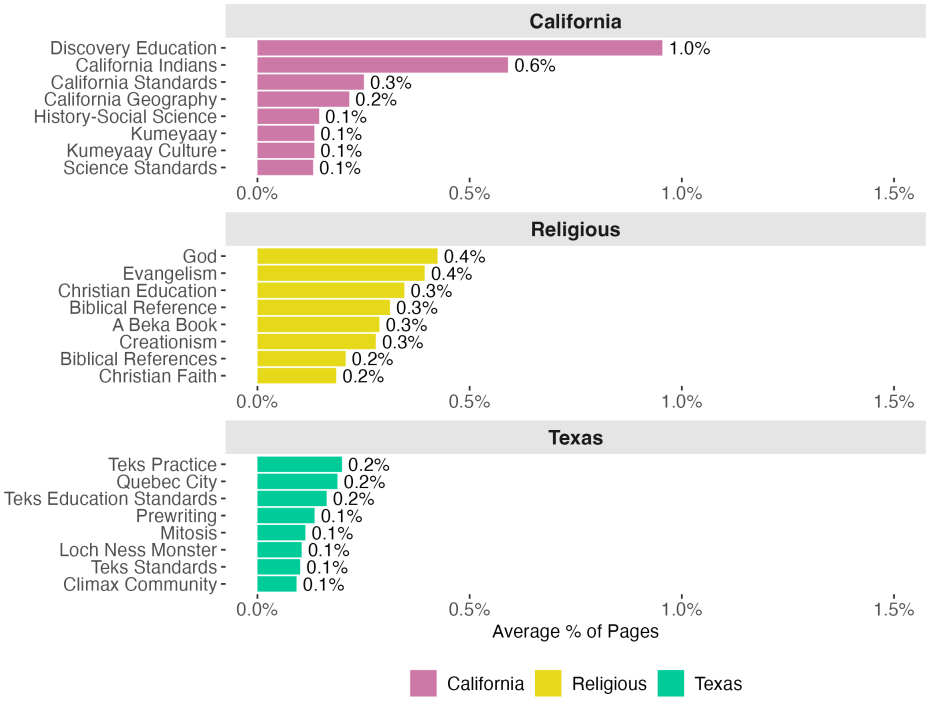
(c) Overall



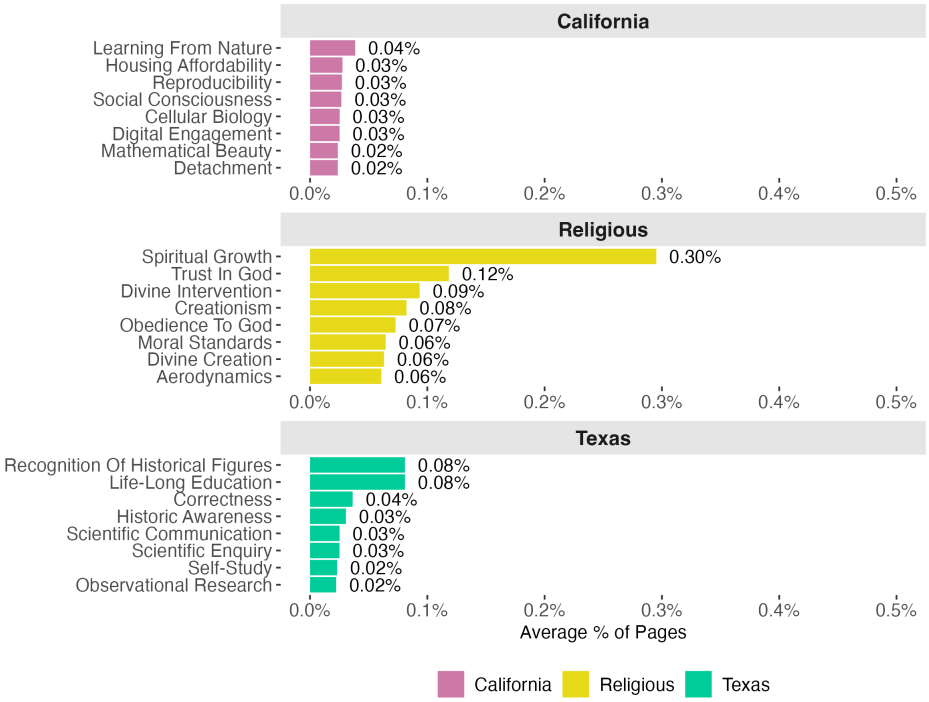

*Notes*: In this figure we calculate the average cosine similarities between any two pages in different collections (e.g., California vs. Religious) that share the same subject, grade, and year in use. In Figure (a), we average the cosine similarities of page pairs for all pages that contain a specific topic as predicted by GPT. In Figure (b) we do the same thing but for all pages that contain a specific value as predicted by GPT. In Panel (c) we show the average cosine similarity of all pages weighted equally across subject, grade, and year in use for comparison.

Figure A.6: Top Topics and Values Exclusive to Each Collection

(a) Topics



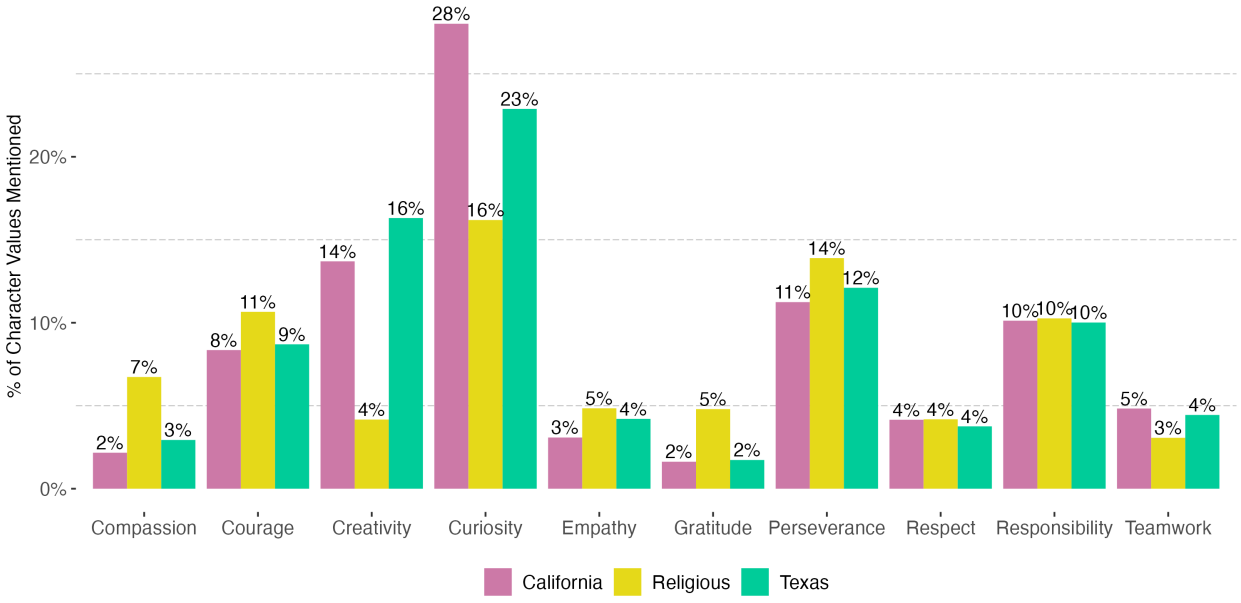

(b) Values



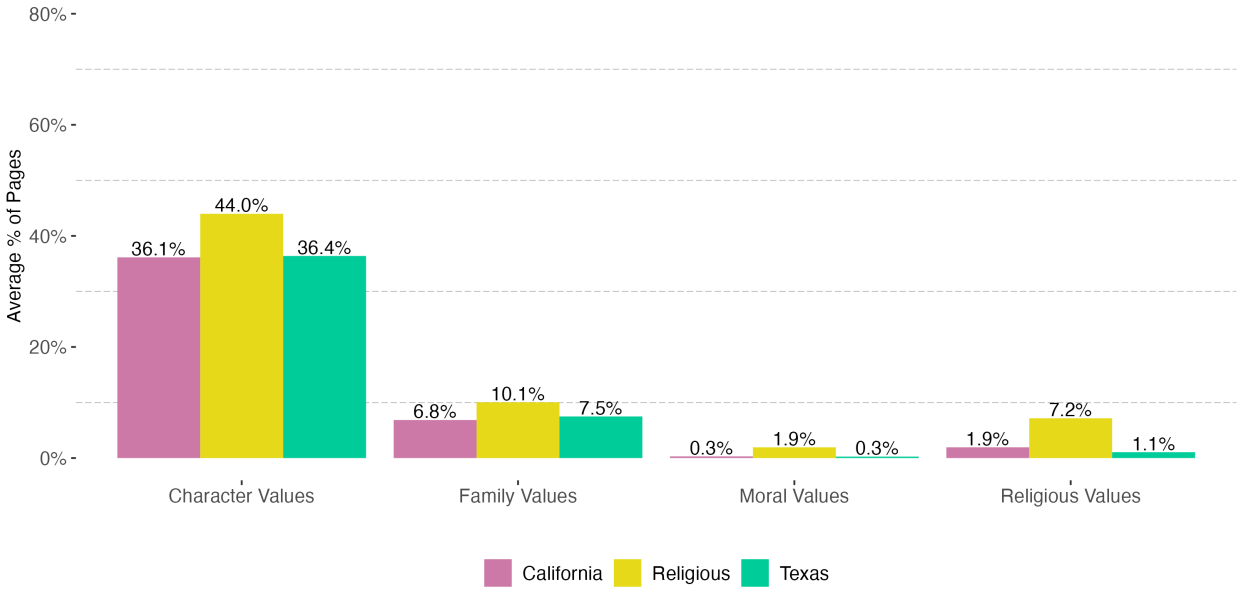

*Notes*: This figure shows the top topics (a) and values (b) which are exclusive to a single collection (e.g. only appear in one collection). To calculate the average percentage of pages containing each topic or value, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure A.7: Character, Family, Moral, Religious Values

(a) Specific Character Values Mentioned by Collection

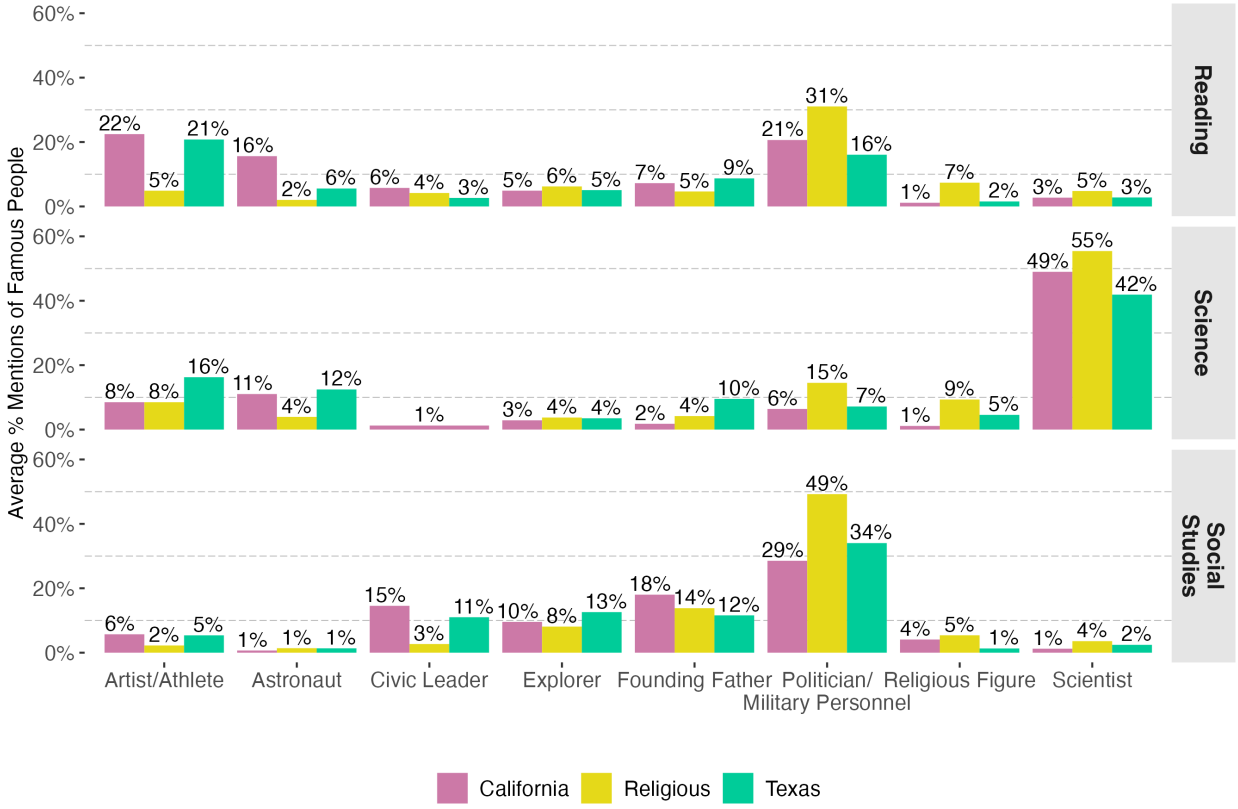



(b) Character, Family, Moral, and Religious Values by Collection




*Notes*: This figure shows (a) the proportion of all character value mentions that is accounted for by the most frequently mentioned values in each collection and (b) the average percentage of all pages in a collection that contain character, family, moral, or religious values. To determine the average percentage of pages, we first calculate the percentage of pages within each collection, subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally. The topics and values in both panels were as classified by a large language model.

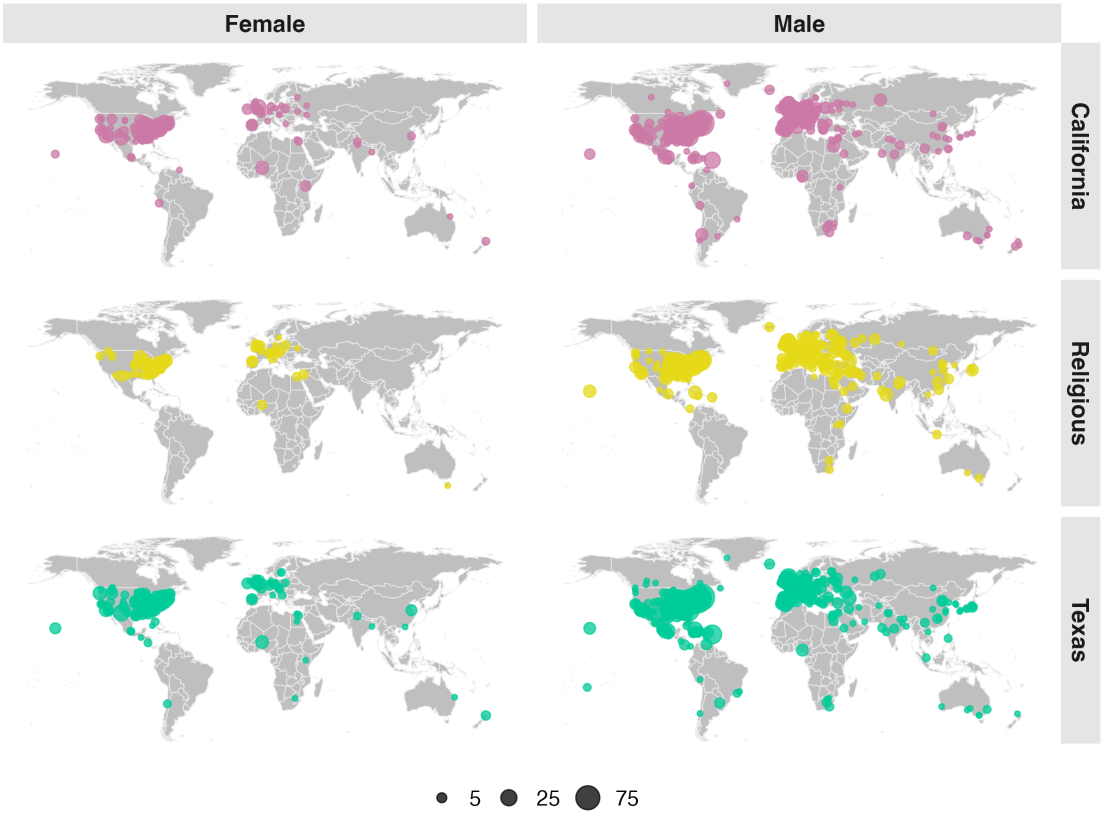
Figure A.8: Historical Figures, by Subject and Occupation



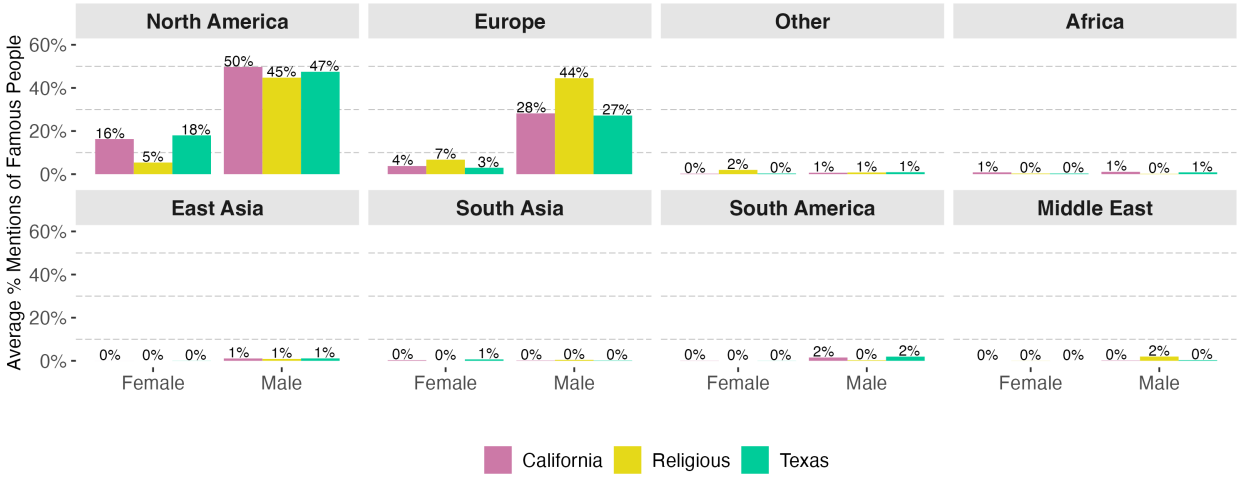

*Notes*: This figure shows the average percentage mentions famous people by their occupation, disaggregated by subject. To determine the average percentage of mentions for each occupation, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally in the final averages.

Figure A.9: Birthplaces of Historical Figures by Gender

(a) Map



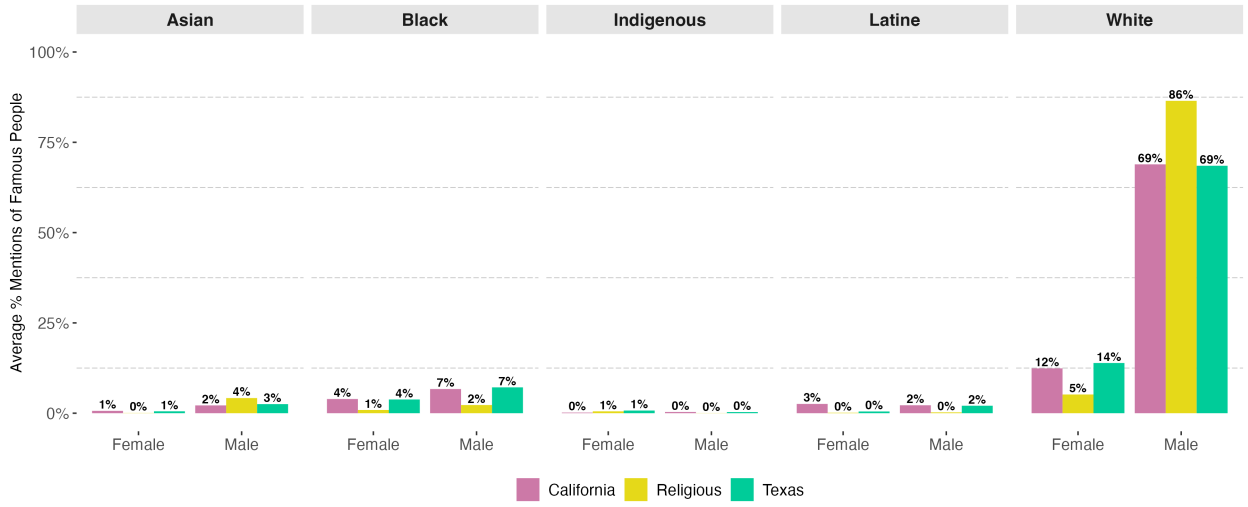

(b) Averages



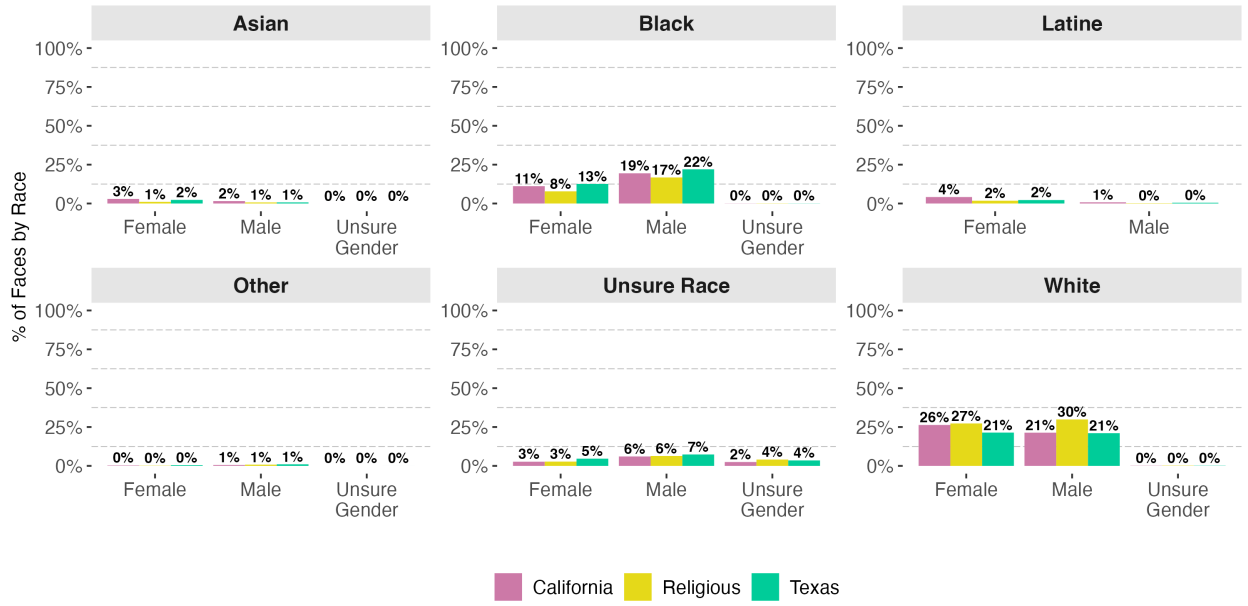

*Notes*: This figure shows the birthplaces of all famous people mentioned in a collection by gender. In Panel (a), the size of the point represents the number of mentions. In Panel (b) to determine the average percentage of mentions for each race, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection level to ensure that mentions of famous people from each grade, subject, and year in use are weighted equally in the final averages.

Figure A.10: Race/Ethnicity and Gender Representation in Text and Images by Collection

(a) Text (Famous Individuals)

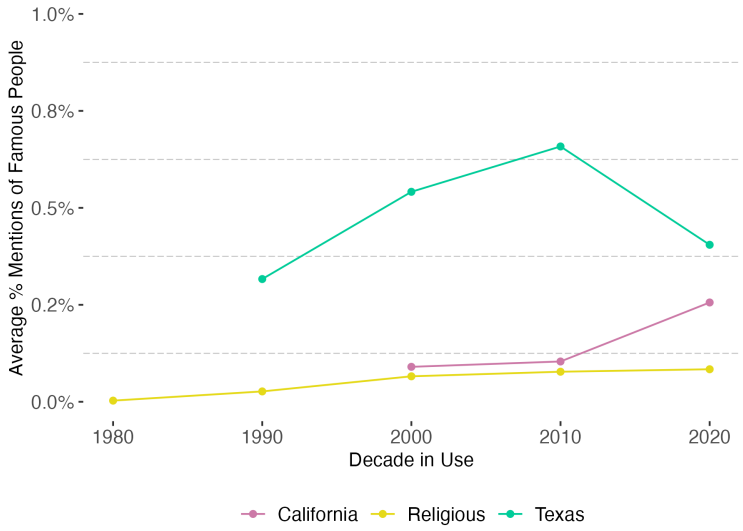



(b) Images



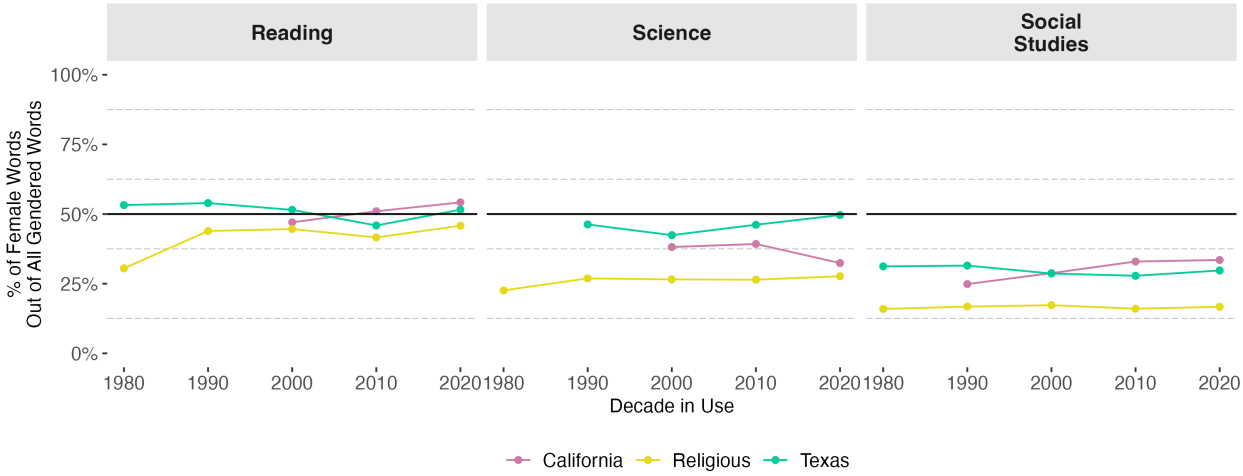

*Notes*: Panel (a) shows the average percentage of famous people that identify as a specific race and gender out of all mentions of famous people by collection. To determine the average percentage of mentions by collection, we first calculate the percentage of mentions of famous people of a specific gender by race across each subject, grade, and year in use. We then average these percentages at the decade level to ensure that pages from each grade, subject, and year in use are weighted equally in the final averages. Panel (b) shows the average percentage of faces classified as a specific race and gender by our feature classification model described in Section V.A.3 by collection and subject. To determine the average percentage, we first calculate the percentage of faces by race within each collection, subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally.

Figure A.11: Percentage of Famous Individuals Who Identify as Lesbian, Gay, or Bisexual by Collection Over Time




*Notes*: This figure shows the within-collection average percent of mentions of famous people who have identified as lesbian, gay or bisexual over time. To determine the average percentage of mentions, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection and decade in use level to ensure that pages from each grade and subject are weighted equally in the final averages.

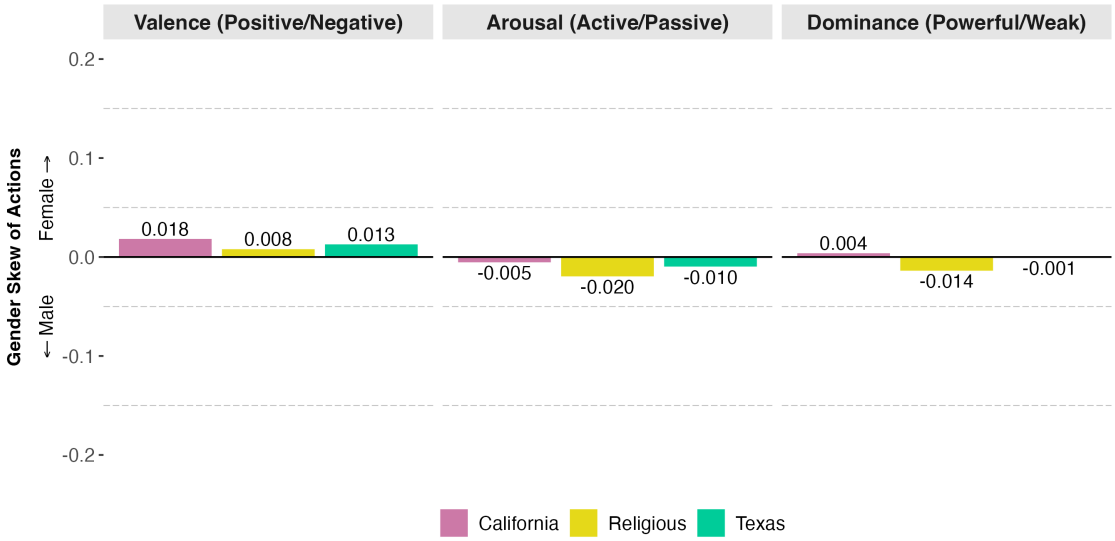
Figure A.12: Percent of Female Words by Subject Over Time



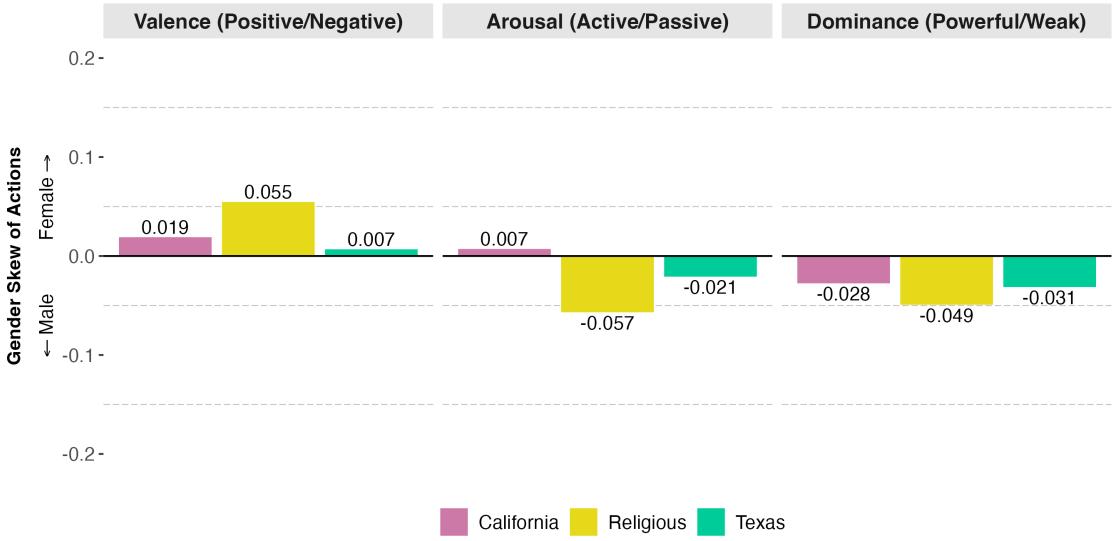

*Notes*: This figure shows the average percent of female words out of all gendered words by collection and subject over time. Note that female words includes female names, female pronouns, and female terms such as queen or brother. To calculate the average percentage of female words, we first calculate the percentage each collection, subject, grade, and year in use. We then average these percentages at the collection, subject and decade in use level, to ensure that textbooks in each grade and year in use are weighted equally in the final averages.

Figure A.13: Difference in Average Sentiment (VAD Scores) of Gendered Actions

(a) Actions Associated with Male Pronouns vs. Actions Associated with Female Pronouns




(b) Actions Associated Only with Male Words vs. Actions Associated Only with Female Words




*Notes*: This figure shows the average Valence, Arousal, and Dominance (VAD) scores of all actions associated with males and females and then takes the differences of these score to find the gender skew of that dimension. If the skew is positive, it means that context in which females are discussed uses words that are higher on that Valence, Arousal, or Dominance dimension than the context that males are discussed in. We use Semantic Role Labeling to find the actions associated with males versus females. Panel (a) shows the difference in the VAD scores of actions associated with female pronouns and the actions associated with male pronouns. Panel (b) shows the difference in VAD scores of actions associated only with female words vs actions only associated with male words.

## B Appendix Tables

Table B.1: Average Skin Tint by Collection

| | *Dependent variable:* | |
|---|---|---|
| | Skin Tint | |
| | (1) | (2) |
| Religious | 6.191*** | 5.013*** |
| | (1.343) | (1.329) |
| Texas | −0.356 | 0.270 |
| | (1.675) | (1.709) |
| Asian | | −3.711*** |
| | | (1.078) |
| Black | | −10.612*** |
| | | (0.500) |
| Latine | | −14.037*** |
| | | (0.801) |
| Other/Unsure | | −6.586*** |
| | | (0.830) |
| Grade 5 | −4.915*** | −4.681*** |
| | (1.290) | (1.303) |
| Science | −9.127*** | −9.208*** |
| | (2.120) | (2.182) |
| Social Studies | −2.687* | −2.367* |
| | (1.442) | (1.419) |
| Year in Use | −0.108* | −0.095 |
| | (0.059) | (0.058) |
| Constant | 64.022*** | 67.999*** |
| | (2.305) | (2.298) |
| Observations | 347,294 | 347,294 |
| Adjusted $R^2$ | 0.084 | 0.159 |

*Notes*: In this table, we show a regression of the skin tint of a face detected in a given textbook on indicator variables indicating which collection, subject and grade the textbook belongs to. We also include the year a textbook was in use as a continuous variable centered around the first year a textbook was in use in our sample. In column 2, we also control for the predicted race of faces. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

Table B.2: Average Skin Tint by Collection and Race

| | *Dependent variable:* | | | | |
|---|---|---|---|---|---|
| | Skin Tint | | | | |
| | Asian | Black | Latine | White | Other/Unsure |
| | (1) | (2) | (3) | (4) | (5) |
| Religious | 1.908 | 8.529*** | 1.750 | 4.476*** | 1.738 |
| | (1.696) | (1.708) | (1.354) | (1.328) | (1.911) |
| Texas | −7.699*** | 3.403** | −2.177 | 0.457 | −5.348** |
| | (2.617) | (1.544) | (2.296) | (1.770) | (2.559) |
| Grade 5 | −7.929*** | −2.786** | −2.147 | −5.215*** | −6.800*** |
| | (2.300) | (1.177) | (1.442) | (1.287) | (2.424) |
| Science | −8.043*** | −8.583*** | −3.613 | −9.289*** | −12.744*** |
| | (2.945) | (2.061) | (2.378) | (2.029) | (3.416) |
| Social Studies | −3.519 | −2.609* | −0.947 | −2.112 | −4.436** |
| | (2.241) | (1.341) | (1.557) | (1.477) | (2.115) |
| Year in Use | −0.159* | −0.139** | 0.160** | −0.080 | −0.078 |
| | (0.094) | (0.061) | (0.069) | (0.057) | (0.088) |
| Constant | 70.976*** | 55.471*** | 44.536*** | 67.901*** | 66.186*** |
| | (3.058) | (2.571) | (2.229) | (2.290) | (3.454) |
| Observations | 12,574 | 99,455 | 13,084 | 182,639 | 39,542 |
| Adjusted $R^2$ | 0.161 | 0.066 | 0.033 | 0.088 | 0.106 |

*Notes*: In this table, we show a regression of the skin tint of a face detected in a given textbook on indicator variables indicating which collection, subject and grade the textbook belongs to. In each column, we condition on a different predicted race of each face. Standard errors clustered at the textbook level.
*p<0.1; **p<0.05; ***p<0.01

Table B.3: Famous Lesbian, Gay, Bisexual Individuals Mentioned by Collection

| Collection | Name | Race | Gender | Occupation | Rank | Mentions | Textbooks |
|---|---|---|---|---|---|---|---|
| California | Jane Addams | White | Female | Civic Leader | 1 | 24 | 4 |
| California | Friedrich Wilhelm Von Steuben | White | Male | Military | 2 | 4 | 4 |
| California | Bessie Smith | Black | Female | Singer | 3 | 5 | 3 |
| California | Andy Warhol | White | Male | Artist | 4 | 2 | 2 |
| California | Michel Foucault | White | Male | Philosopher | 5 | 5 | 1 |
| California | Keith Haring | White | Male | Artist | 6 | 3 | 1 |
| California | Billie Jean King | White | Female | Tennis Player | 7 | 2 | 1 |
| California | Ellen Degeneres | White | Female | Comedian | 8 | 2 | 1 |
| California | Brian Boitano | White | Male | Skater | 9 | 1 | 1 |
| California | Claude Mckay | Black | Male | Writer | 10 | 1 | 1 |
| California | Frida Kahlo | Latinx | Female | Painter | 11 | 1 | 1 |
| California | Harvey Milk | White | Male | Politician | 12 | 1 | 1 |
| California | James Baldwin | Black | Male | Writer | 13 | 1 | 1 |
| California | Ma Rainey | Black | Female | Musician | 14 | 1 | 1 |
| California | Tom Daley | White | Male | Athlete | 15 | 1 | 1 |
| | | | | | | | |
| Religious | James Baldwin | Black | Male | Writer | 1 | 21 | 8 |
| Religious | Elagabalus | Asian | Male | Politician | 2 | 2 | 2 |
| Religious | Pyotr Ilyich Tchaikovsky | White | Male | Composer | 3 | 2 | 2 |
| Religious | Simone De Beauvoir | White | Female | Writer | 4 | 2 | 2 |
| Religious | Jane Addams | White | Female | Civic Leader | 5 | 2 | 1 |
| Religious | Rupert Brooke | White | Male | Writer | 6 | 1 | 1 |
| | | | | | | | |
| Texas | Jane Addams | White | Female | Civic Leader | 1 | 92 | 18 |
| Texas | Barbara Jordan | Black | Female | Politician | 2 | 13 | 3 |
| Texas | Claude Mckay | Black | Male | Writer | 3 | 9 | 3 |
| Texas | Bessie Smith | Black | Female | Singer | 4 | 5 | 3 |
| Texas | Frida Kahlo | Latinx | Female | Painter | 5 | 4 | 2 |
| Texas | Maurice Sendak | White | Male | Writer | 6 | 2 | 2 |
| Texas | Gertrude Stein | White | Female | Writer | 7 | 3 | 1 |
| Texas | Berenice Abbott | White | Female | Photographer | 8 | 2 | 1 |
| Texas | Michel Foucault | White | Male | Philosopher | 9 | 2 | 1 |
| Texas | Oscar Wilde | White | Male | Writer | 10 | 2 | 1 |
| Texas | Andrew Scott | White | Male | Actor | 11 | 1 | 1 |
| Texas | Billie Holiday | Black | Female | Singer | 12 | 1 | 1 |
| Texas | Greta Garbo | White | Female | Actor | 13 | 1 | 1 |
| Texas | Jason Brown | White | Male | Skater | 14 | 1 | 1 |
| Texas | Little Richard | Black | Male | Singer | 15 | 1 | 1 |
| Texas | Ma Rainey | Black | Female | Musician | 16 | 1 | 1 |
| Texas | Tom Daley | White | Male | Athlete | 17 | 1 | 1 |

*Notes*: This figure illustrates the number of times a famous person identified as lesbian, gay, or bisexual (LGB) is mentioned within a collection. We calculate both the number of times a person as mentioned and the number of textbooks they appear in. We then sort first by number of textbooks within a collection and then by number of mentions. Our data on LGB famous people comes from the Pantheon 2.0 dataset. Note that this data does not contain information on transgender individuals.

Table B.4: Top 20 First Names

| Collection | Name | Gender | Rank | Mentions | Textbooks |
|---|---|---|---|---|---|
| California | John | Male | 1 | 817 | 54 |
| California | David | Male | 2 | 376 | 47 |
| California | Robert | Male | 3 | 301 | 45 |
| California | Paul | Male | 4 | 308 | 44 |
| California | James | Male | 5 | 486 | 43 |
| California | Michael | Male | 6 | 289 | 43 |
| California | George | Male | 7 | 399 | 39 |
| California | Tom | Male | 8 | 305 | 39 |
| California | Richard | Male | 9 | 221 | 39 |
| California | Peter | Male | 10 | 255 | 38 |
| California | Charles | Male | 11 | 252 | 37 |
| California | William | Male | 12 | 355 | 36 |
| California | Mary | Female | 13 | 280 | 35 |
| California | Jim | Male | 14 | 173 | 35 |
| California | Gary | Male | 15 | 147 | 35 |
| California | Mark | Male | 16 | 119 | 35 |
| California | Steve | Male | 17 | 103 | 35 |
| California | Joseph | Male | 18 | 180 | 34 |
| California | Thomas | Male | 19 | 334 | 33 |
| California | Stephen | Male | 20 | 138 | 33 |
| | | | | | |
| Religious | John | Male | 1 | 729 | 60 |
| Religious | William | Male | 2 | 496 | 49 |
| Religious | Robert | Male | 3 | 380 | 45 |
| Religious | Joseph | Male | 4 | 264 | 45 |
| Religious | James | Male | 5 | 271 | 44 |
| Religious | George | Male | 6 | 454 | 41 |
| Religious | Mary | Female | 7 | 425 | 41 |
| Religious | Charles | Male | 8 | 352 | 39 |
| Religious | Matthew | Male | 9 | 104 | 38 |
| Religious | Jack | Male | 10 | 318 | 36 |
| Religious | David | Male | 11 | 273 | 36 |
| Religious | Henry | Male | 12 | 284 | 34 |
| Religious | Brian | Male | 13 | 50 | 34 |
| Religious | Thomas | Male | 14 | 203 | 33 |
| Religious | Richard | Male | 15 | 110 | 33 |
| Religious | Elizabeth | Female | 16 | 218 | 32 |
| Religious | Susan | Female | 17 | 106 | 32 |
| Religious | Tom | Male | 18 | 328 | 30 |
| Religious | Rachel | Female | 19 | 208 | 30 |
| Religious | Peter | Male | 20 | 131 | 30 |
| | | | | | |
| Texas | John | Male | 1 | 509 | 65 |
| Texas | George | Male | 2 | 411 | 58 |
| Texas | David | Male | 3 | 331 | 58 |
| Texas | Mary | Female | 4 | 433 | 56 |
| Texas | James | Male | 5 | 392 | 54 |
| Texas | Robert | Male | 6 | 319 | 54 |
| Texas | Tom | Male | 7 | 745 | 52 |
| Texas | William | Male | 8 | 303 | 52 |
| Texas | Michael | Male | 9 | 493 | 51 |
| Texas | Jim | Male | 10 | 248 | 48 |
| Texas | Charles | Male | 11 | 200 | 48 |
| Texas | Mike | Male | 12 | 277 | 47 |
| Texas | Susan | Female | 13 | 235 | 47 |
| Texas | Jean | Female | 14 | 240 | 45 |
| Texas | Peter | Male | 15 | 191 | 45 |
| Texas | Thomas | Male | 16 | 333 | 44 |
| Texas | Joe | Male | 17 | 388 | 43 |
| Texas | Richard | Male | 18 | 172 | 43 |
| Texas | Jack | Male | 19 | 292 | 42 |
| Texas | Karen | Female | 20 | 111 | 42 |

*Notes*: This figure illustrates the most frequently first names mentioned within a collection. We find the number of times a first name is mentioned and the number of textbooks they appear in and then sort first by number of textbooks within a collection and then by number of mentions to find the top individuals. Note that we remove mentions of the full names of famous individuals in this analysis.

Table B.5: Female Representation in Text and Images

| | *Dependent variable:* | |
|---|---|---|
| | % Female Words | % Female Faces |
| | (1) | (2) |
| Religious | $-0.103^{***}$ | $-0.068^{***}$ |
| | (0.023) | (0.023) |
| Texas | 0.022 | $-0.026$ |
| | (0.020) | (0.020) |
| Grade 5 | $-0.058^{***}$ | $-0.008$ |
| | (0.022) | (0.024) |
| Science | $-0.112^{***}$ | $0.041^{*}$ |
| | (0.020) | (0.022) |
| Social Studies | $-0.199^{***}$ | 0.002 |
| | (0.019) | (0.022) |
| Year in Use | 0.0002 | $-0.0003$ |
| | (0.001) | (0.001) |
| Constant | $0.523^{***}$ | $0.449^{***}$ |
| | (0.033) | (0.033) |
| Observations | 2,521 | 2,521 |
| Adjusted $R^2$ | 0.322 | 0.054 |

*Notes:* Regression of two different measures of female representation in textbooks on indicator variables indicating which collection, subject, and grade the textbook belongs to. We also include the year a textbook was in use as a continuous variable centered around the first year a textbook was in use in our sample. Standard errors clustered at the textbook level. $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01

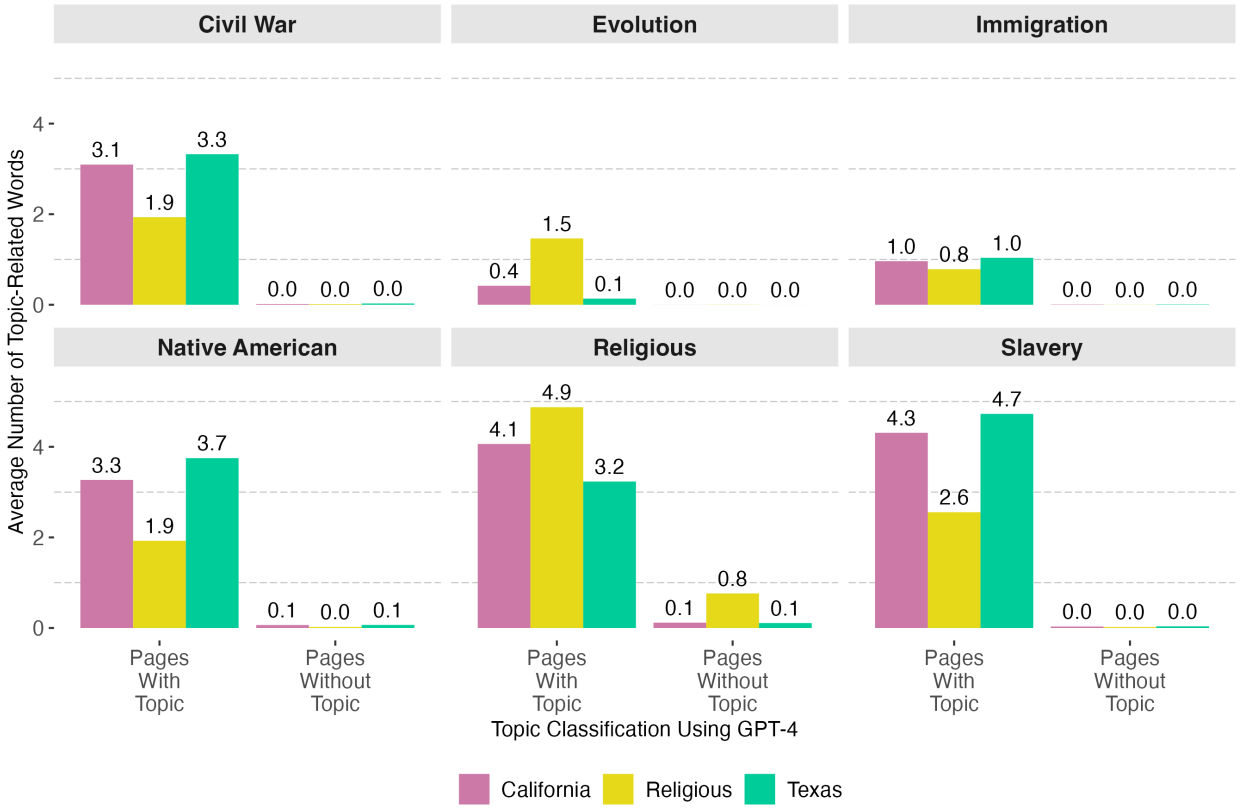
## C Methods Validation Appendix

In this appendix, we show the variety of ways in which we used manual validation to measure the effectiveness of our computational methods in measuring the people and topics present in our sample of educational curriculum.

### C.A Topic Classification

First we manually validate the topic classification performed by GPT using token counts of topic related words. In Figure C.1, we show that on average, pages that were not classified as containing specific topic, also do not have any mentions of those topic related words on the page.

Figure C.1: GPT-Identified Topics and Number of Topic-Related Tokens




*Notes*: This figure shows the average number of words on a page that are related to a given topic that are then also assigned to a predicted GPT topic classification. "TRUE" indicates that a given page has a topic-related word is related to a topic, and GPT also classified it as connected to that topic. "FALSE" indicates that the word is related a topic, but GPT did not classify it as such.

## C.B Feature Classification

In this section, we gauge the effectiveness of our feature classification model by comparing its performance with classifications completed by human annotators. The set of faces used in this validation procedure were randomly sampled from our textbooks and stratified by collection, subject and grade. Each randomly sampled face was randomly assigned 3 times to one of eight annotators. The annotators then determined whether the face was humanoid and, if so, labeled features of the face such as gender and race. Finally, the image was assigned the modal value out of the 3 labels, and any three way ties were labeled as "unsure." The annotators also had the option to mark their answers as "unsure" along with their response. For some features, we incorporate this uncertainty in the modal aggregation. This aligns with the feature classification model's structure, which could also predict "unsure" labels for certain features.

There were 1,080 images chosen for this exercise which were each labeled 3 times, resulting in 3,240 image-by-labeler observations. After subsetting these observations to those labeled as containing a face, and then keeping only the images with 3 labelers, our final validation set consists of 900 images. We compare the human and model classifications for these images in Figure C.2.

Regarding the gender of faces, in Figure C.2a, the model assigns male and female labels in almost equal proportion in the public school collection, whereas the human labels skew slightly towards male. In the religious school collection, both the model and the human annotators labeled more females than males, though the manual labels skewed even more male. This means, we may be slighting underestimating the gender imbalance in images within our textbooks. Turning now to race, Figure C.2b shows that the annotators more often assigned faces to non-white and non-black labels relative to the model's predictions. This may be indicative of the annotators' ability to differentiate racial nuances better than the model.

For both race and gender labels, we see higher uncertainty among human annotators than we see for the model. We see more uncertainly for both human and models among race labels than gender labels, highlighting that classifying race is a difficult task. This is confirmed in our measures of accuracy and inter-labeler reliability shown in Table C.1. To calculate the accuracy, we take the modal human response for each image as the true value and subsequently find the percentage of model predictions equal to said value. To calculate the inter-labeler reliability, we use the Fleiss' kappa measure commonly used to find the agreement between categorical labels given by a fixed number of labelers. This statistic

ranges from 0 to 1 where 0 means complete disagreement and 1 means perfect agreement between labelers. In particular, we apply the following formula:

$$\hat{\kappa} = \frac{I_o - I_e}{1 - I_e} \tag{C.1}$$

where $I_o$ is the observed agreement and $I_e$ is the expected value based on chance. The table shows that the model has higher accuracy (i.e., agrees more often with the human annotation) than random choice, performing best when predicting age and gender while struggling more with emotion and race. This aligns with the kappa measure, which demonstrates greater disagreement among human annotators when assigning emotion and gender to faces.

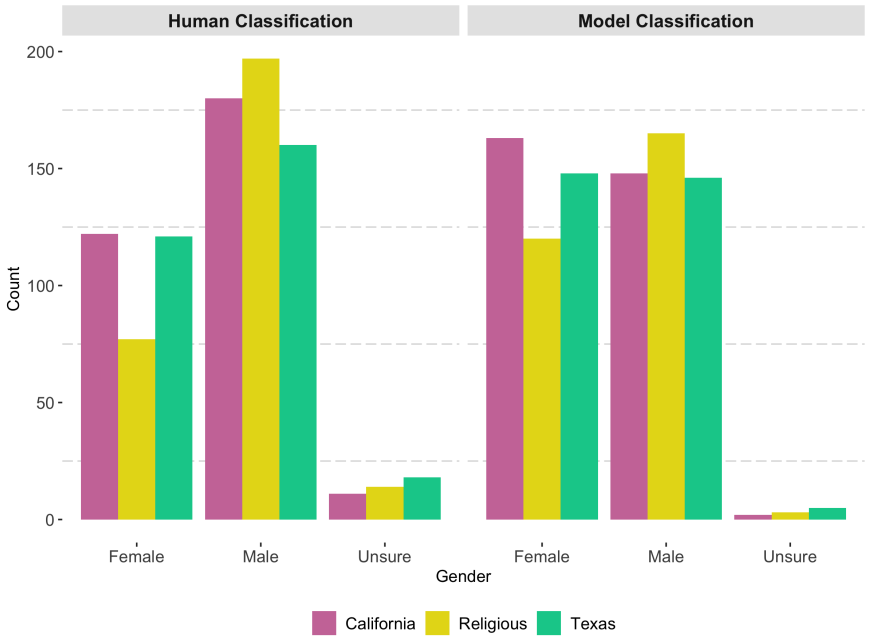
Table C.1: Accuracy and Inter-Labeler Reliability

| | Without Unsure | | With Unsure | |
|---|---|---|---|---|
| | Accuracy | $\hat{\kappa}$ | Accuracy | $\hat{\kappa}$ |
| gender | 64.944 | 0.805 | 62.000 | 0.723 |
| race | 57.349 | 0.501 | 48.333 | 0.425 |

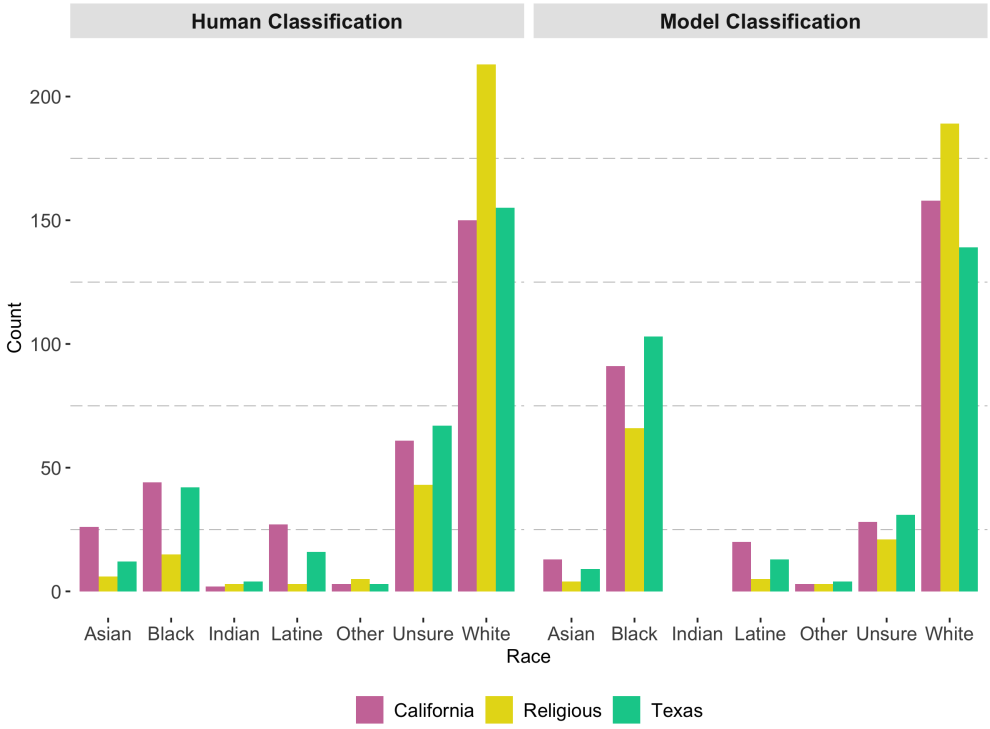
*Notes:* Accuracy is calculated as the percentage correct between the modal value of human labels and the predicted value for a given feature. Inter-labeler reliability was calculated using the Fleiss' kappa measure commonly used to find the agreement between categorical labels given by a fixed number of labelers, in our case 3 labelers. We repeat calculation with and without unsure labels.

Figure C.2: Manual vs. Model Classifications of Race and Gender in Images

(a) Gender



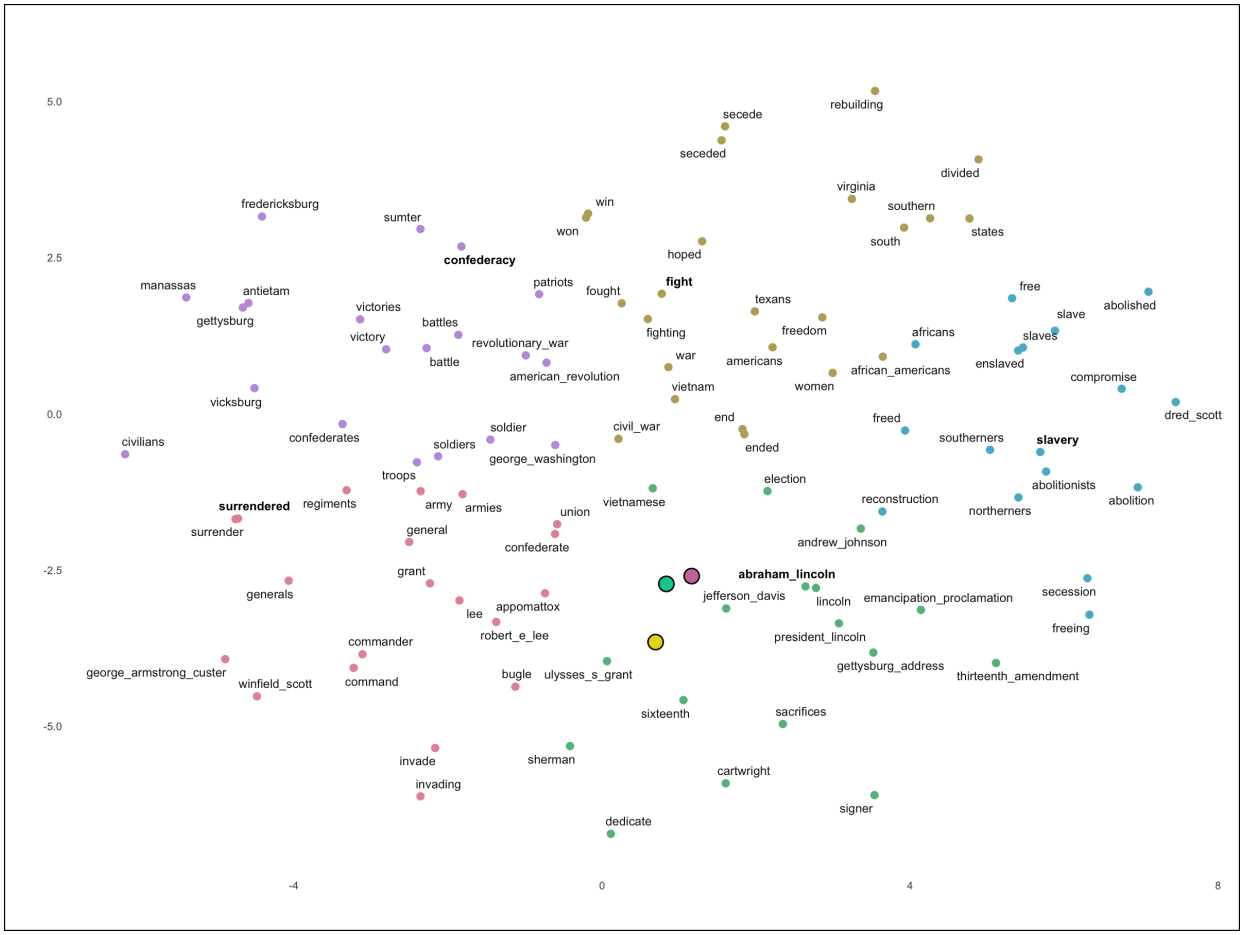

(b) Race




*Notes*: This figure compares the count of values between the human labels and the model predicted labels for a given feature. Each observation is an image where the human label is chosen based on the modal value among 3 independent labelers.

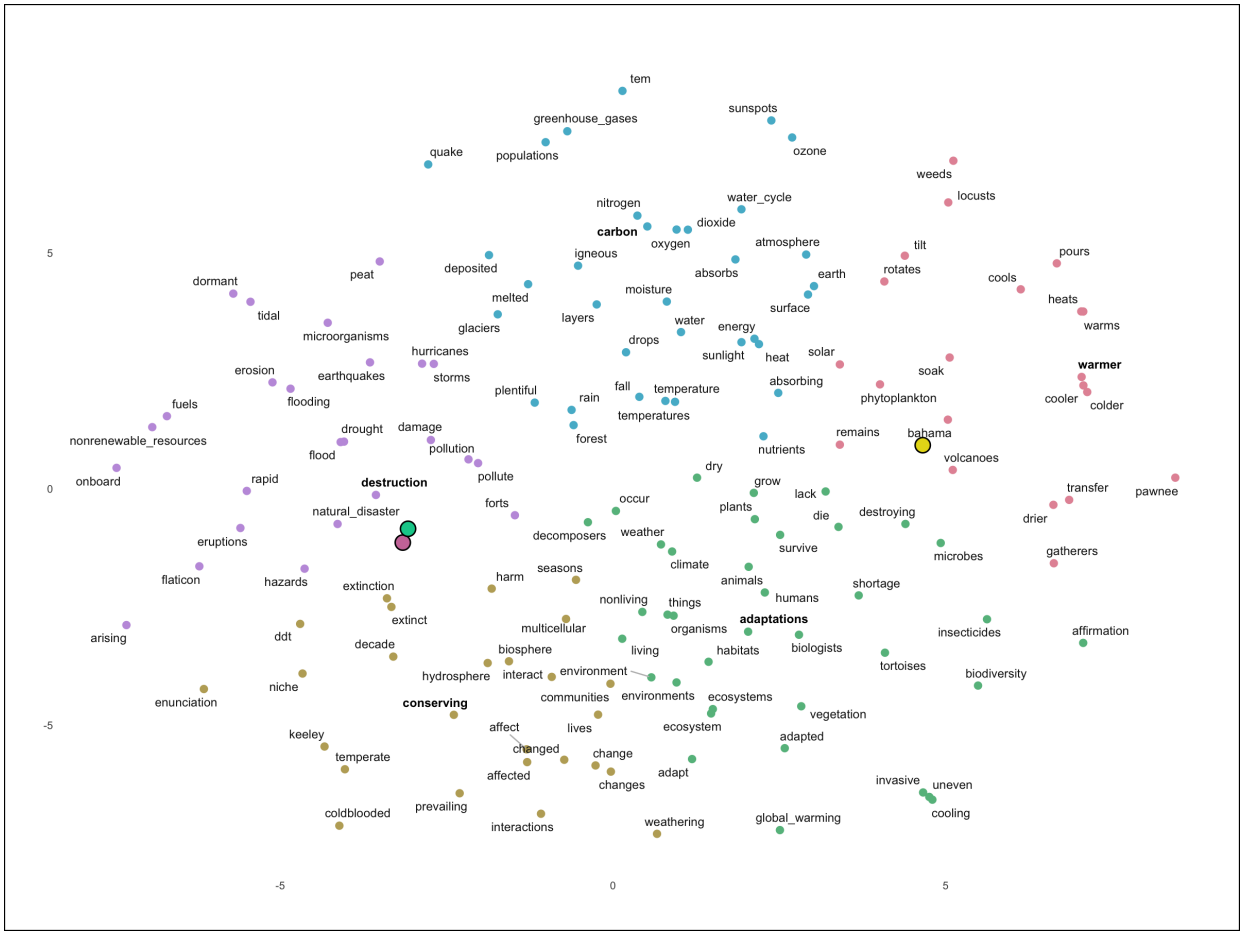
## D Topic Embedding Appendix Figures

Figure D.1: Civil War: Topic Embeddings (Doc2Vec)



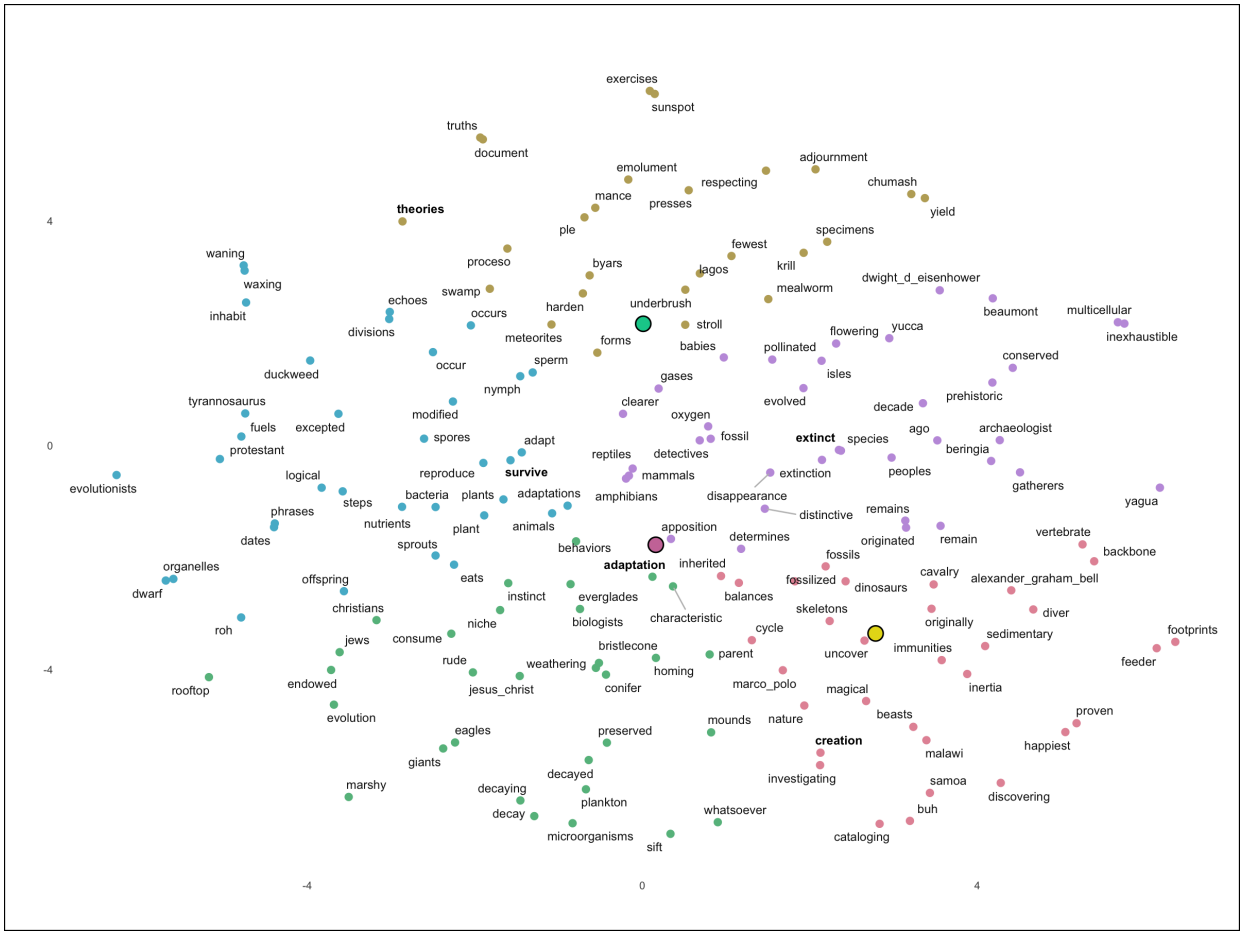

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. The name of the subtopic is bolded and is chosen by a human annotator. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.2: Climate Change: Topic Embeddings (Doc2Vec)



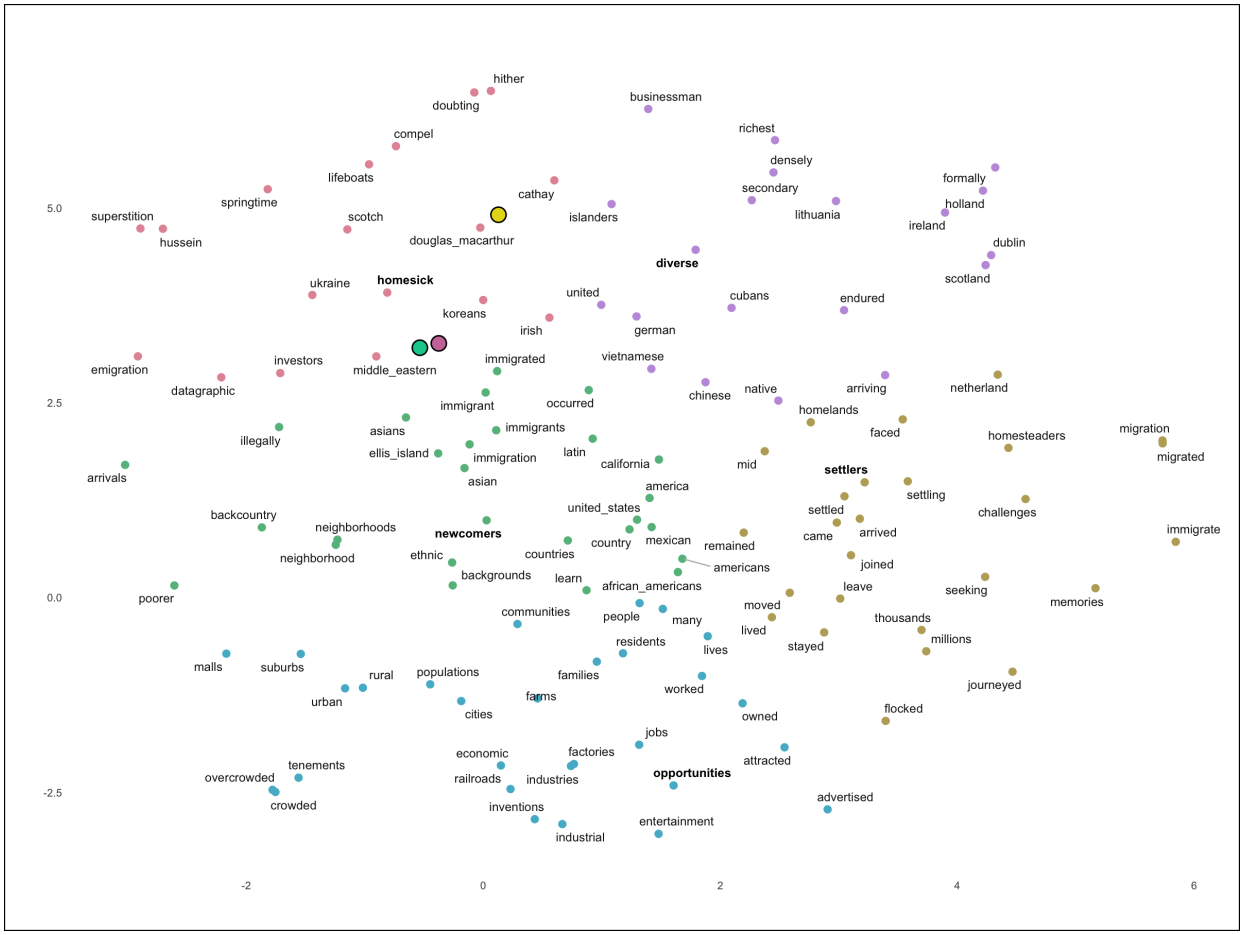

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.3: Evolution: Topic Embeddings (Doc2Vec)



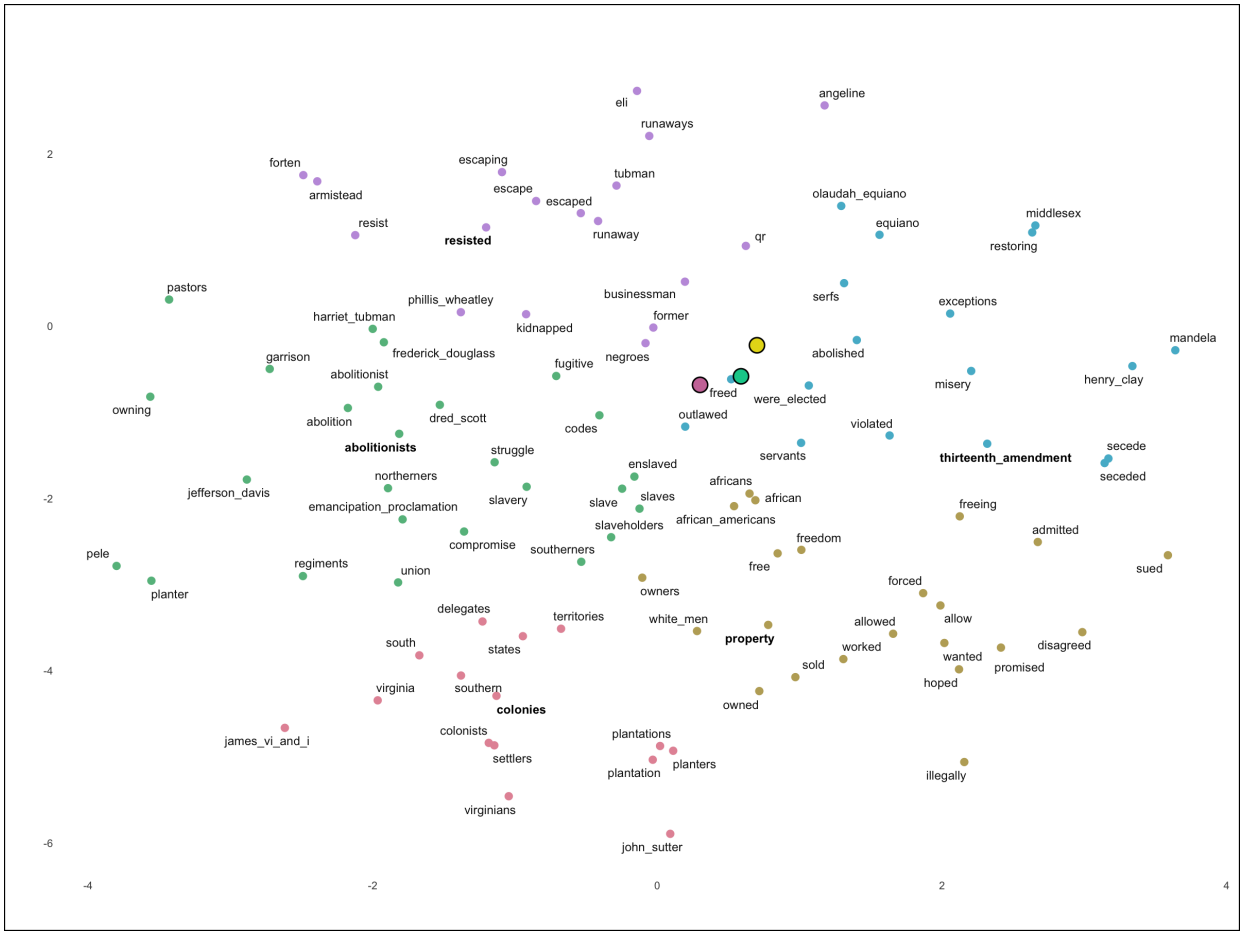

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.4: Immigration: Topic Embeddings (Doc2Vec)



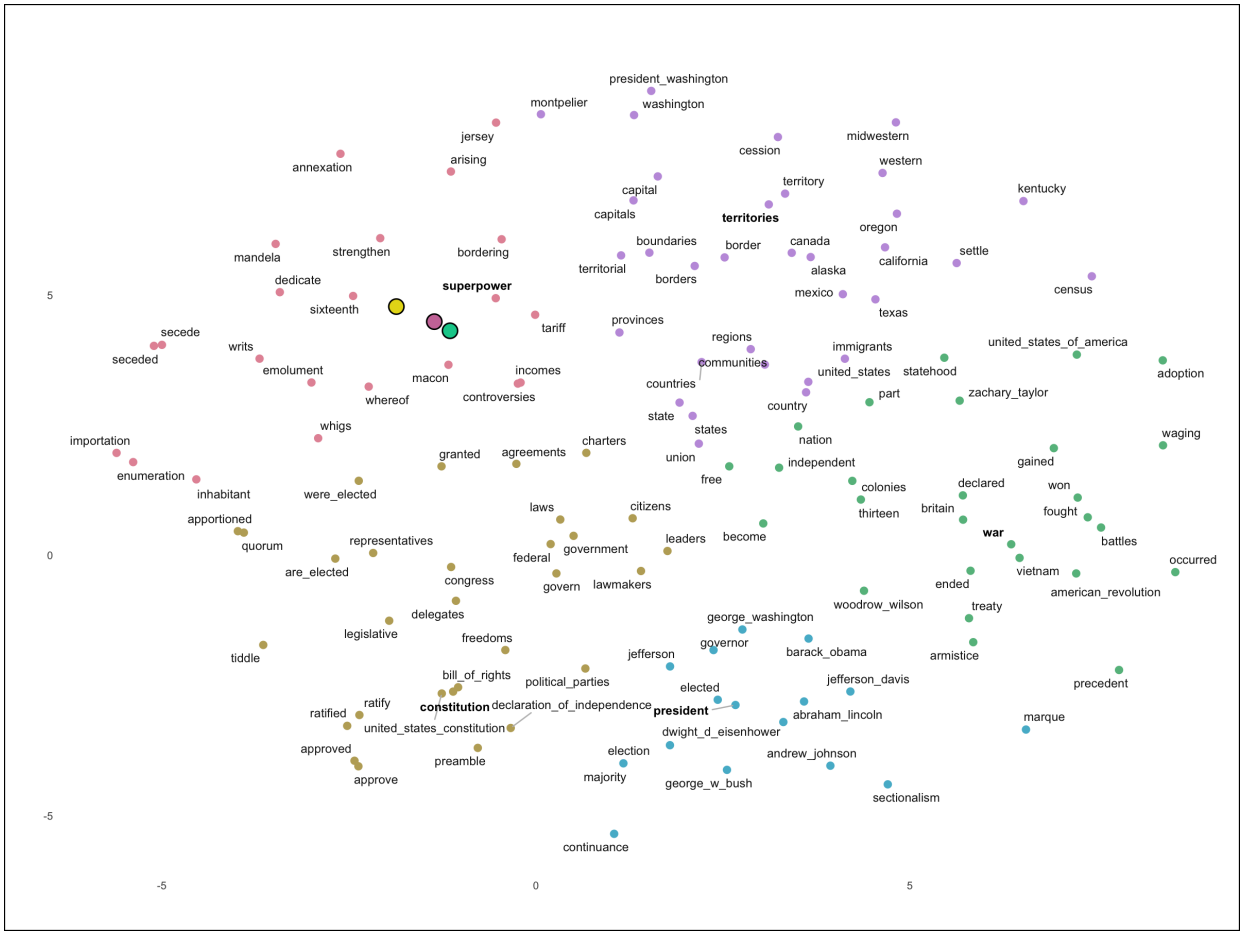

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.5: Slavery: Topic Embeddings (Doc2Vec)



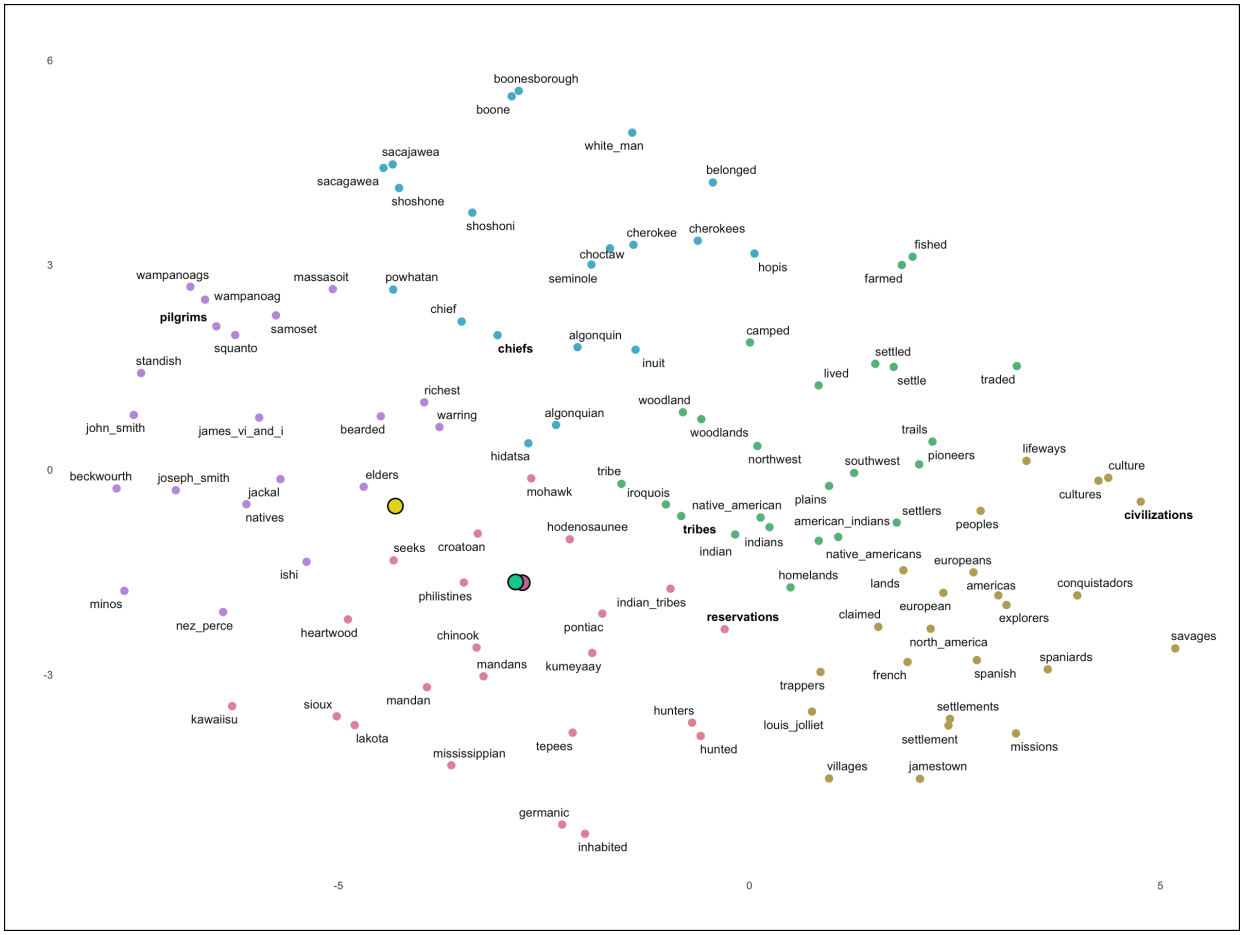

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.6: United States: Topic Embeddings (Doc2Vec)



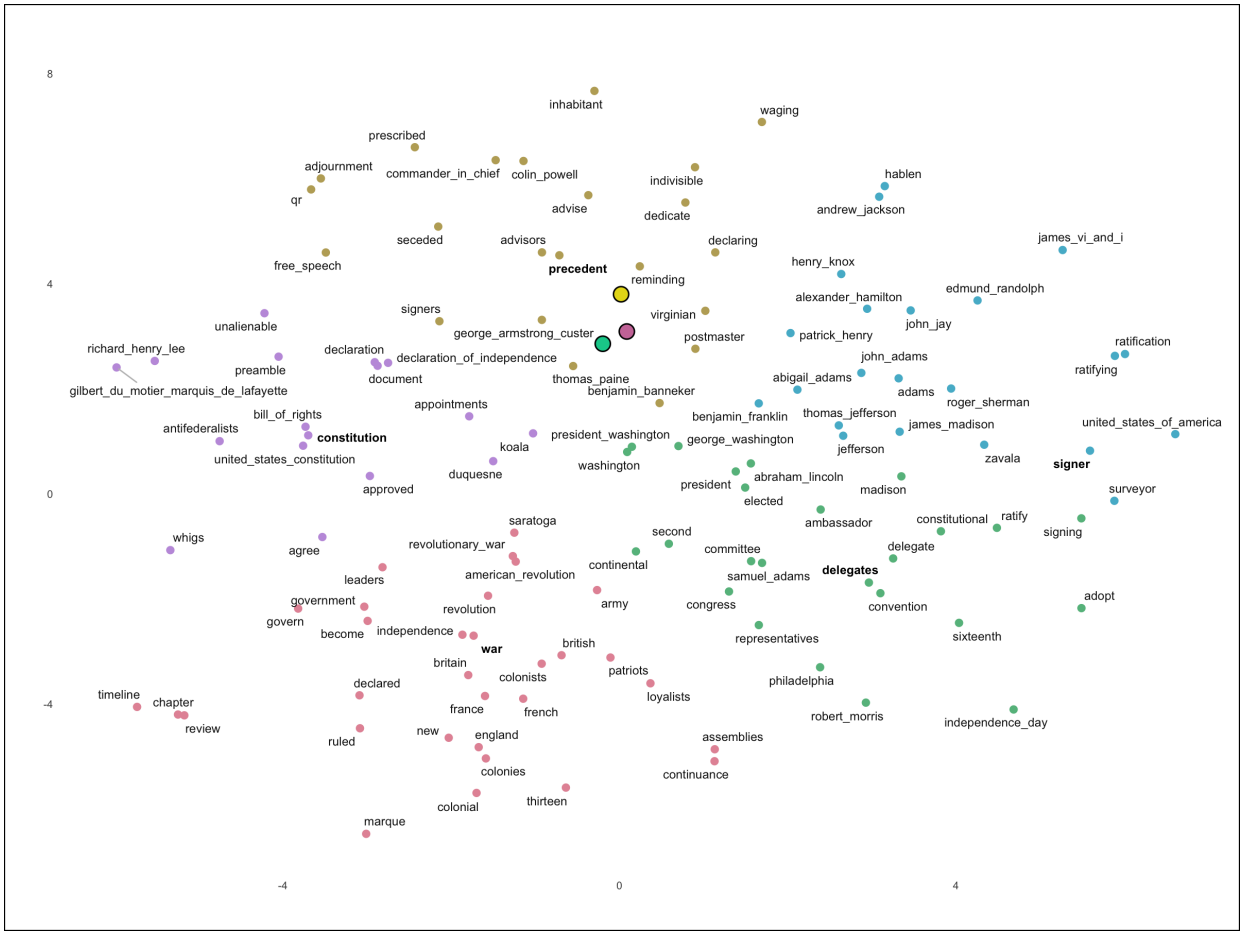

*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.7: Native Americans: Topic Embeddings (Doc2Vec)




*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.

Figure D.8: Founding Fathers: Topic Embeddings (Doc2Vec)




*Notes:* Each small dot represents one of the 60 words closest to a collection topic embedding from a specific collection. We cluster the words into subtopics and denote these subtopic by color. Each large dot represents a topic embedding; pink signifies California, yellow signifies Religious, and green signifies Texas. To create these embeddings, we use Doc2Vec, which transforms a collection's pages into fixed-length vectors and embeds them in the same semantic space as the words. We run this model on the combined pages of all 3 collections, creating one combined training corpus. A topic embedding is a normalized average of the resulting page embeddings for all pages designated with a specific topic. These embeddings, originally represented in a 300-dimensional space, are here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations.