Appendices

A Appendix Tables

Table A1. Top Five Most Mentioned Famous People, by Collection

Collection	Rank	Name	Race	Gender	Mentions	Books
Mainstream	1	George Washington	White	Male	152	32
Mainstream	2	Abraham Lincoln	White	Male	270	25
Mainstream	3	Thomas Jefferson	White	Male	71	15
Mainstream	4	John Adams	White	Male	60	14
Mainstream	5	Benjamin Franklin	White	Male	23	12
Diversity	1	Martin Luther King Junior	Black	Male	282	51
Diversity	2	Abraham Lincoln	White	Male	72	41
Diversity	3	George Washington	White	Male	62	40
Diversity	4	Frederick Douglass	Black	Male	131	30
Diversity	5	Langston Hughes	Black	Male	109	30
People of Color	1	Martin Luther King Junior	Black	Male	263	48
People of Color	2	Abraham Lincoln	White	Male	70	39
People of Color	3	George Washington	White	Male	58	37
People of Color	4	Frederick Douglass	Black	Male	131	30
People of Color	5	Langston Hughes	Black	Male	108	29
African American	1	Langston Hughes	Black	Male	53	17
African American	2	Martin Luther King Junior	Black	Male	130	16
African American	3	Malcolm X	Black	Male	69	12
African American	4	Frederick Douglass	Black	Male	43	12
African American	5	Duke Ellington	Black	Male	25	12
Ability	1	Harold Pinter	White	Male	78	2
Ability	2	Andy Warhol	White	Male	4	2
Ability	3	Marco Polo	White	Male	3	2
Ability	4	Duke Ellington	Black	Male	2	2
Ability	5	Judy Blume	White	Female	2	2
Female	1	John F. Kennedy	White	Male	8	4
Female	2	Martin Luther King Junior	Black	Male	19	3
Female	3	Jimmy Carter	White	Male	15	3
Female	4	Betty Friedan	White	Female	10	3
Female	5	Richard Nixon	White	Male	9	3
LGBTQIA+	1	Alicia Keys	Multiracial	Female	3	3
LGBTQIA+	2	Britney Spears	White	Female	3	3
LGBTQIA+	3	Marilyn Monroe	White	Female	3	3
LGBTQIA+	4	Julia Roberts	White	Female	5	2
LGBTQIA+	5	Alexander Hamilton	White	Male	4	2

Note: This table shows the five most frequently mentioned famous people in each collection, along with their race, their gender, the number of times they were mentioned, and the number of books in which they appeared.

Table A2. Top Five Most Mentioned Famous Females, by Collection

Collection	Rank	Name	Race	Mentions	Books
Mainstream	1	Eleanor Roosevelt	White	30	7
Mainstream	2	Martha Washington	White	9	6
Mainstream	3	Emily Dickinson	White	7	6
Mainstream	4	Shirley Temple	White	12	5
Mainstream	5	Rosa Parks	Black	43	4
Diversity	1	Rosa Parks	Black	157	27
Diversity	2	Harriet Tubman	Black	35	19
Diversity	3	Eleanor Roosevelt	White	42	18
Diversity	4	Coretta Scott King	Black	23	15
Diversity	5	Lena Horne	White	20	14
People of Color	1	Rosa Parks	Black	152	25
People of Color	2	Harriet Tubman	Black	35	19
People of Color	3	Eleanor Roosevelt	White	41	17
People of Color	4	Coretta Scott King	Black	22	14
People of Color	5	Lena Horne	White	20	14
African American	1	Rosa Parks	Black	44	11
African American	2	Coretta Scott King	Black	12	10
African American	3	Zora Neale Hurston	Black	21	9
African American	4	Lena Horne	White	14	9
African American	5	Harriet Tubman	Black	13	9
Ability	1	Judy Blume	White	2	2
Ability	2	Shirley Temple	White	12	1
Ability	3	Anna Lee	White	4	1
Ability	4	Avril Lavigne	White	4	1
Ability	5	Marilyn Vos Savant	White	4	1
Female	1	Betty Friedan	White	10	3
Female	2	Mary Pickford	White	5	3
Female	3	Billie Jean King	White	24	2
Female	4	Katharine Graham	White	14	2
Female	5	Gloria Steinem	White	13	2
LGBTQIA+	1	Alicia Keys	Multiracial	3	3
LGBTQIA+	2	Britney Spears	White	3	3
LGBTQIA+	3	Marilyn Monroe	White	3	3
LGBTQIA+	4	Julia Roberts	White	5	2
LGBTQIA+	5	Patsy Cline	White	3	2

Note: In this table, we show the five most frequently mentioned famous females in each collection, along with their race, the number of times they were mentioned, and the number of books in which they appeared.

Table A3. Top Five Most Mentioned Famous Males, by Collection

Collection	Rank	Name	Race	Mentions	Books
Mainstream	1	George Washington	White	152	32
Mainstream	2	Abraham Lincoln	White	270	25
Mainstream	3	Thomas Jefferson	White	71	15
Mainstream	4	John Adams	White	60	14
Mainstream	5	Benjamin Franklin	White	23	12
Diversity	1	Martin Luther King Junior	Black	282	51
Diversity	2	Abraham Lincoln	White	72	41
Diversity	3	George Washington	White	62	40
Diversity	4	Frederick Douglass	Black	131	30
Diversity	5	Langston Hughes	Black	109	30
People of Color	1	Martin Luther King Junior	Black	263	48
People of Color	2	Abraham Lincoln	White	70	39
People of Color	3	George Washington	White	58	37
People of Color	4	Frederick Douglass	Black	131	30
People of Color	5	Langston Hughes	Black	108	29
African American	1	Langston Hughes	Black	53	17
African American	2	Martin Luther King Junior	Black	130	16
African American	3	Malcolm X	Black	69	12
African American	4	Frederick Douglass	Black	43	12
African American	5	Duke Ellington	Black	25	12
Ability	1	Harold Pinter	White	78	2
Ability	2	Andy Warhol	White	4	2
Ability	3	Marco Polo	White	3	2
Ability	4	Duke Ellington	Black	2	2
Ability	5	Mark Twain	White	2	2
Female	1	John F. Kennedy	White	8	4
Female	2	Martin Luther King Junior	Black	19	3
Female	3	Jimmy Carter	White	15	3
Female	4	Richard Nixon	White	9	3
Female	5	Barack Obama	Black	5	3
LGBTQIA+	1	Alexander Hamilton	White	4	2
LGBTQIA+	2	Adam Lambert	White	3	2
LGBTQIA+	3	Alice Cooper	White	3	2
LGBTQIA+	4	James Dean	White	3	2
LGBTQIA+	5	Michael Jackson	Black	3	2

Note: In this table, we show the five most frequently mentioned famous males in each collection, along with their race, the number of times they were mentioned, and the number of books in which they appeared.

Table A4. Top Mentioned Famous Person, by Collection and Decade

Decade	Mainstream	Diversity	People of Color	African American	Ability	Female	LGBTQ
1920	James Fenimore Cooper *White Male* Charles Darwin *White Male* Mark Twain *White Male*						
1930	Abraham Lincoln *White Male*						
1940	Benjamin Franklin *White Male*						
1950	George Washington *White Male*						
1960	George Washington *White Male*						
1970	Claude Lorrain *White Male* Leonardo da Vinci *White Male*	Frederick Douglass *Black Male*	Frederick Douglass *Black Male*	Frederick Douglass *Black Male*			
1980	George Washington *White Male*	Franklin D. Roosevelt *White Male*	Franklin D. Roosevelt *White Male*	Paul Robeson *Black Male*			
1990	William Shakespeare *White Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*			
2000	Martin Luther King Jr. *Black Male*	George Washington *White Male*	George Washington *White Male*	Langston Hughes *Black Male*	Judy Blume *White Female*		
2010	George Washington *White Male*	Martin Luther King Jr. *Black Male*	Martin Luther King Jr. *Black Male*	Malcolm X *Black Male*	Andy Warhol *White Male*	John F. Kennedy *White Male*	Alicia Keys *Multiracial Female* Marilyn Monroe *White Female* Britney Spears *White Female*

Note: In this table, we show the top most uniquely mentioned famous figure in each collection by decade. When multiple names are listed for a collection within the same decade, it indicates that each of those people were tied for the most mentioned famous person in that collection-by-decade.

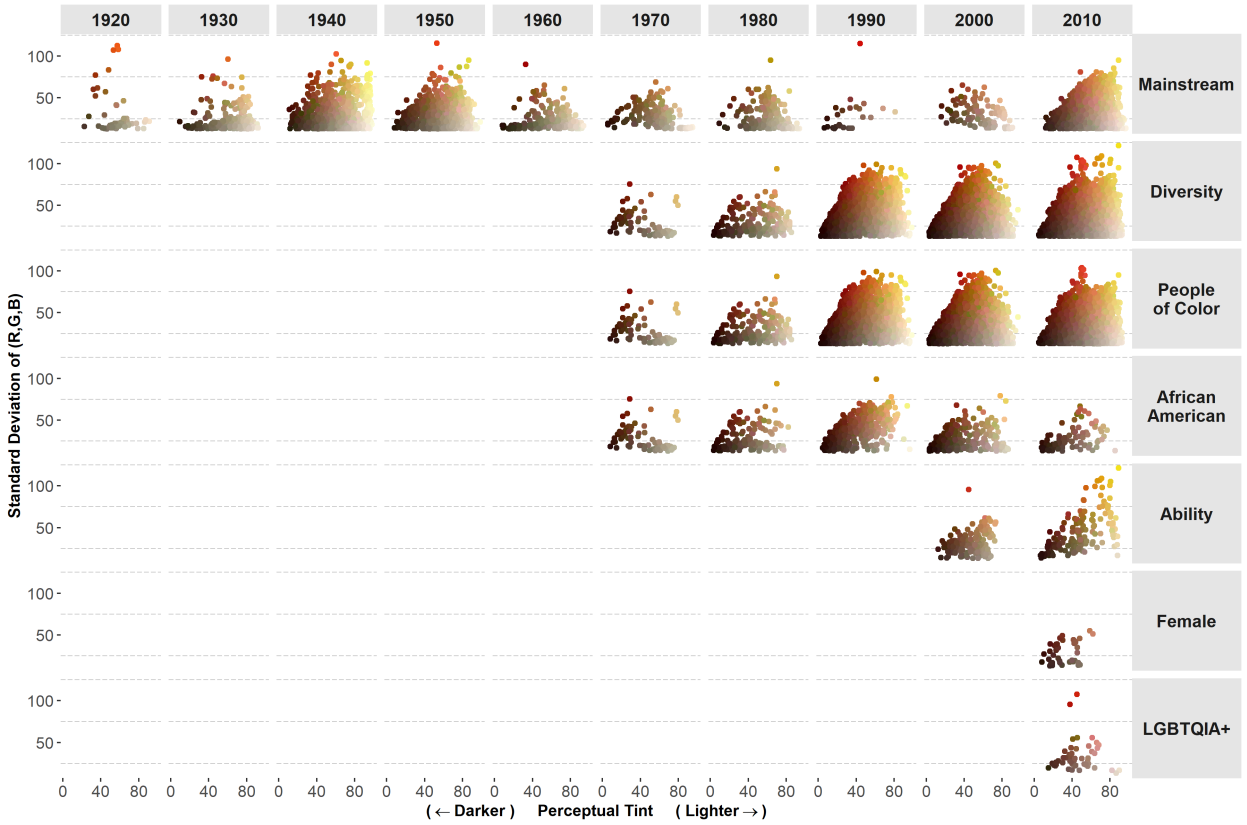
Table A5. Summary Statistics for Children's Book Purchases in Numerator Data

	All Children's Books		*Award-Winning Children's Books*	
Purchaser Demographics	N	Mean	N	Mean
Children				
Has Children	1,547,044	0.73	62,283	0.70
Race/Ethnicity				
Asian	1,506,152	0.06	60,633	0.06
Black/African American	1,506,152	0.04	60,633	0.07
Hispanic/Latino	1,506,152	0.06	60,633	0.08
White/Caucasian	1,506,152	0.81	60,633	0.75
Other Ethnicity	1,506,152	0.03	60,633	0.03
Gender				
Female	1,534,051	0.89	61,714	0.88
Male	1,534,051	0.10	61,714	0.11
Other	1,534,051	0.01	61,714	0.01
Sexuality				
Gay/Lesbian	1,111,247	0.01	41,943	0.02
Straight	1,111,247	0.82	41,943	0.81
Bisexual	1,111,247	0.03	41,943	0.03
Other Sexuality	1,111,247	0.01	41,943	0.01
Prefer Not to Answer	1,111,247	0.13	41,943	0.14
Income				
High Income	1,539,767	0.49	62,031	0.51
Mid Income	1,539,767	0.31	62,031	0.30
Low Income	1,539,767	0.20	62,031	0.19
Education				
Advanced Education	1,548,085	0.25	62,345	0.31
College Education	1,548,085	0.62	62,345	0.58
High School Education	1,548,085	0.12	62,345	0.09
Less than High School	1,548,085	0.02	62,345	0.02

Note: This table describes purchaser demographics for shopping trips in Numerator OmniPanel data between 2017 and 2020. The first two columns show the sample size and mean of purchaser demographic variables over all shopping trips that include the purchase of a children's book. The last two columns show the sample size and mean of purchaser demographic variables over all shopping trips that include children's books which won one of the awards in our sample.

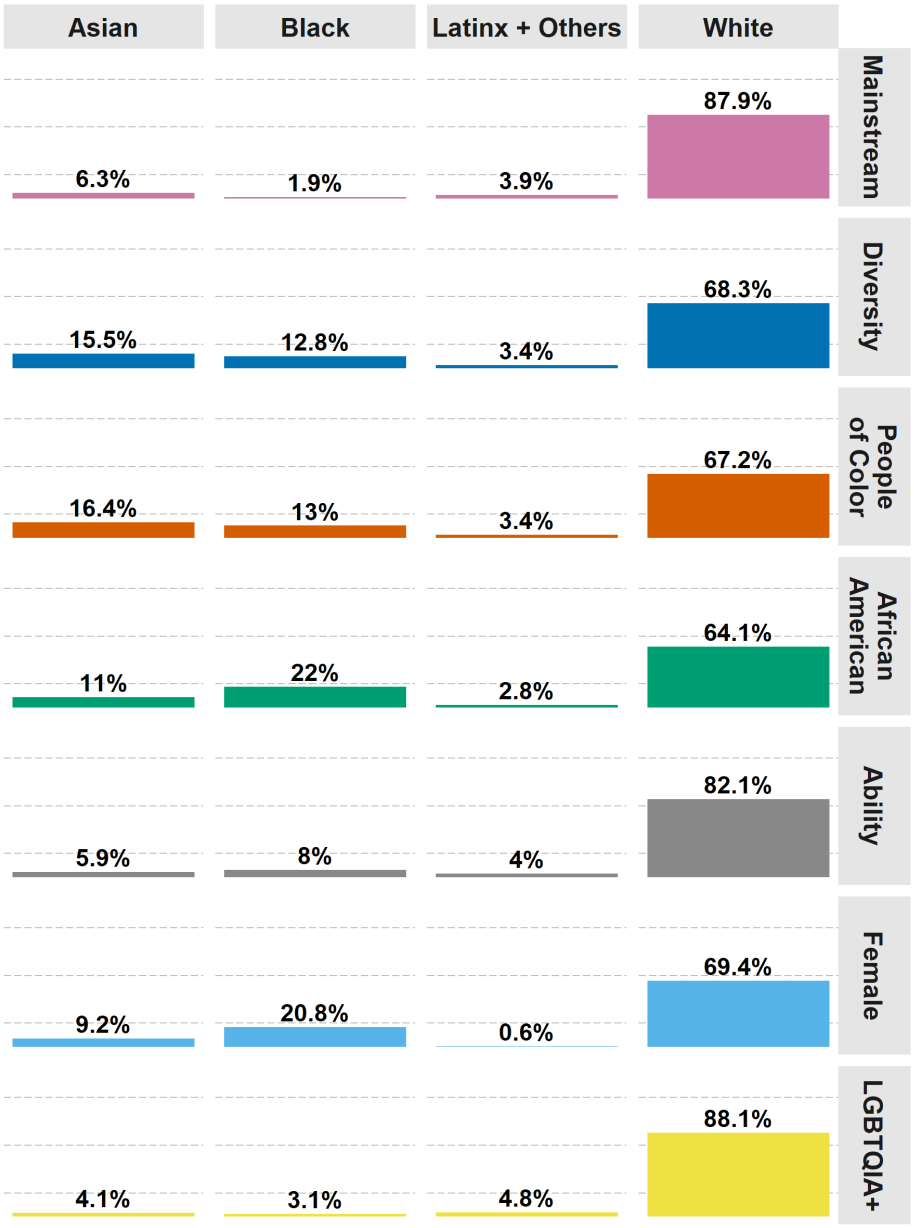
B Appendix Figures

Figure B1. Skin Color Data Over Time, Human Skin Colors



Note: In this figure, we show the representative skin colors for all detected faces with human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B) in each collection-by-decade. As described in Section III, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section III.B.

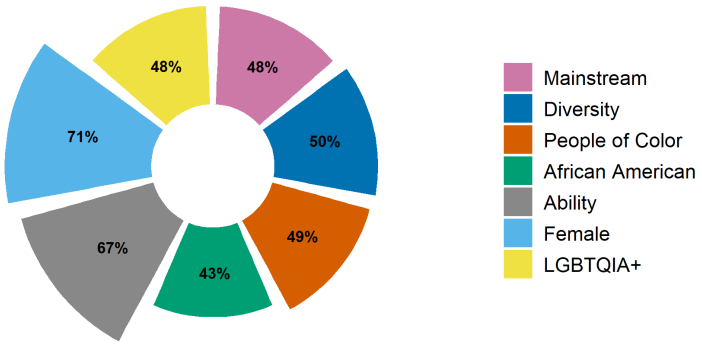
Figure B2. Most Pictured Characters Are Classified as White



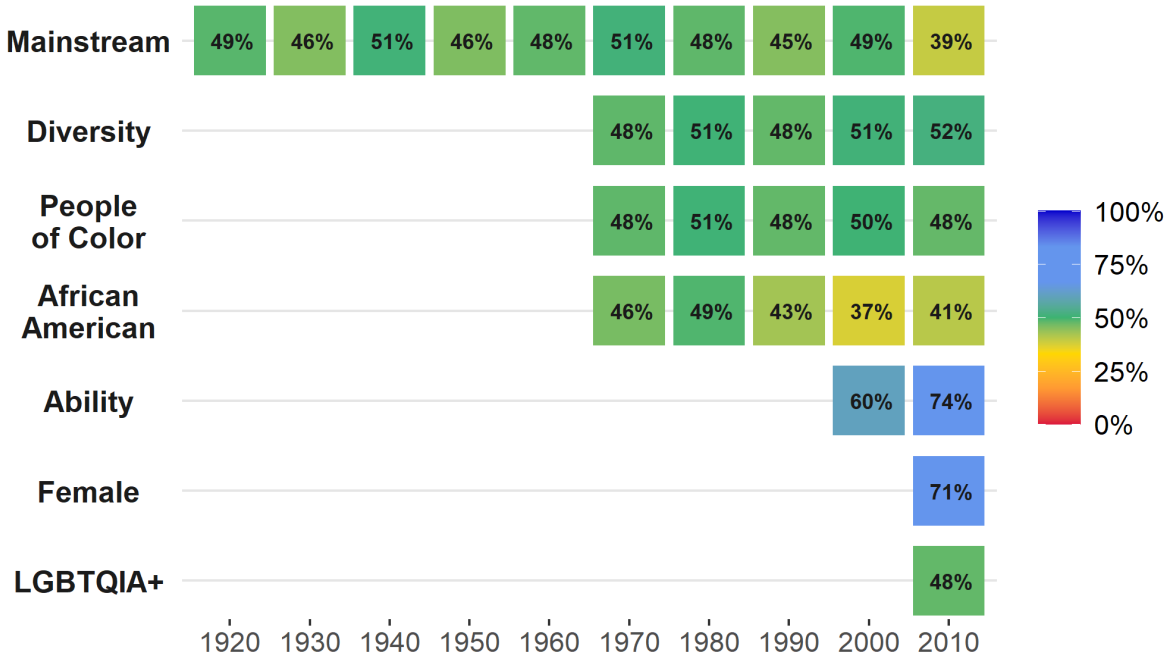
Note: In this figure, we show the proportion of faces in a book which our model labels as a given race averaged over all books in a collection. We detect faces using our face detection model (FDAI) described in Section III.A. Within these faces, we classify age and gender using an AutoML algorithm we trained using the UTKFace public data set.

Figure B3. Proportion of Detected Faces Which Are Female-Presenting

(a) Percent of Female-Presenting Faces Detected, Overall



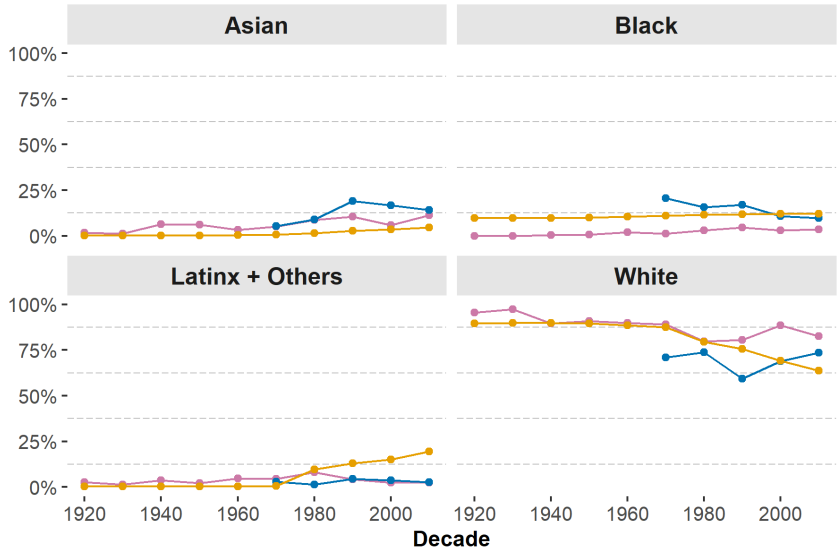
(b) Percent of Female-Presenting Faces Detected, Over Time



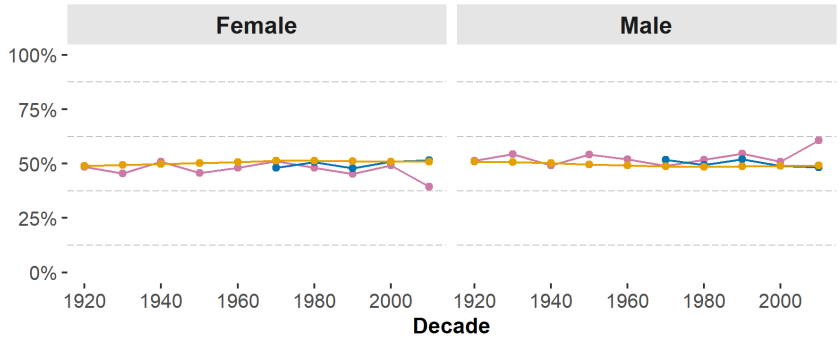
Note: In this figure, we find the proportion of faces in a book which our model labels as female. In Panel A, we show collection level averages of the proportion of female faces in a given book by averaging over all books in a collection. In Panel B, we show these values over time by averaging the proportion of female faces in a given book by each collection and decade.

Figure B4. Share of U.S. Population and Pictured Characters, by Identity

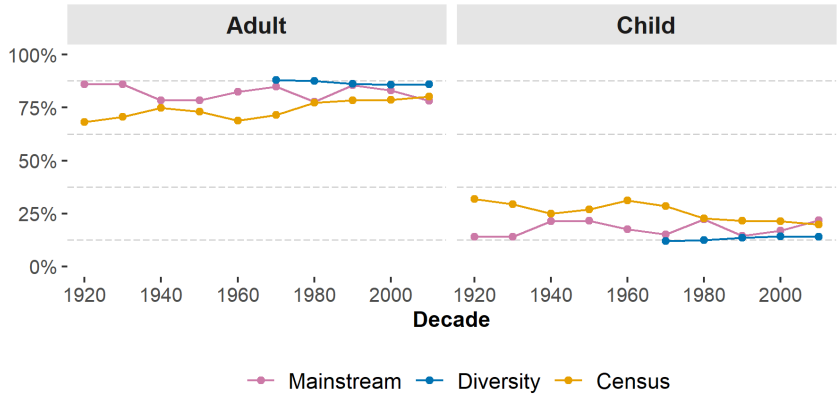
(a) Race/Ethnicity



(b) Gender



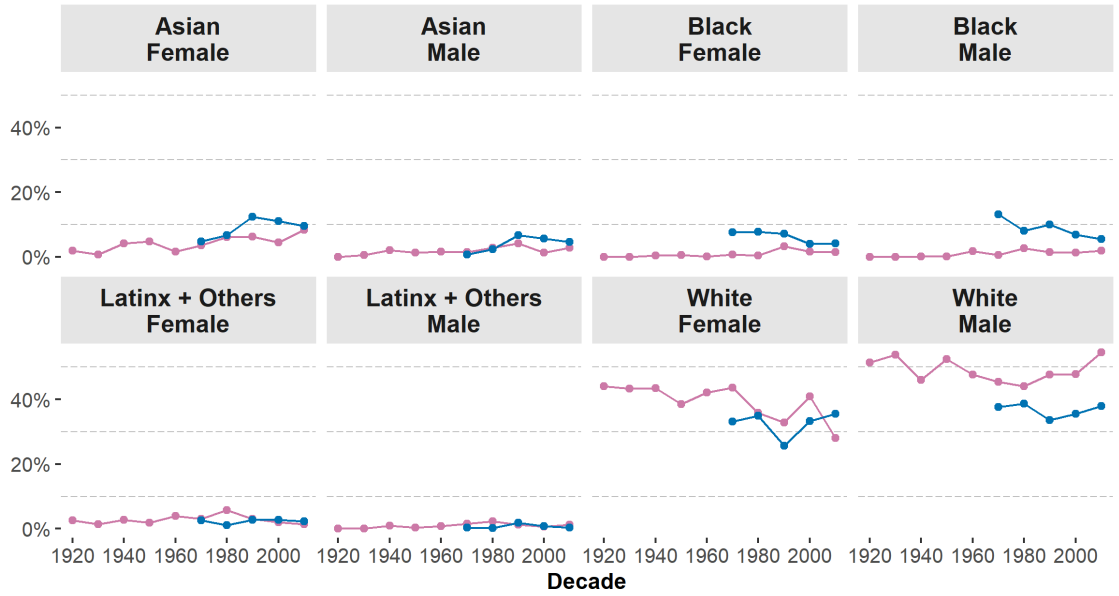
(c) Age



Note: We show the share of the U.S. population of specific identities mapped on the share of the pictured characters classified as a given identity in a given book averaged over all books in collection and decade. In Panel A, we show this by race/ethnicity. Each race/ethnicity category is constructed to be mutually exclusive as defined in Section V. In Panel B, we show this by gender. In Panel C, we show this by age group.

Figure B5. Proportion of Characters in Images and Text, by Race and Gender

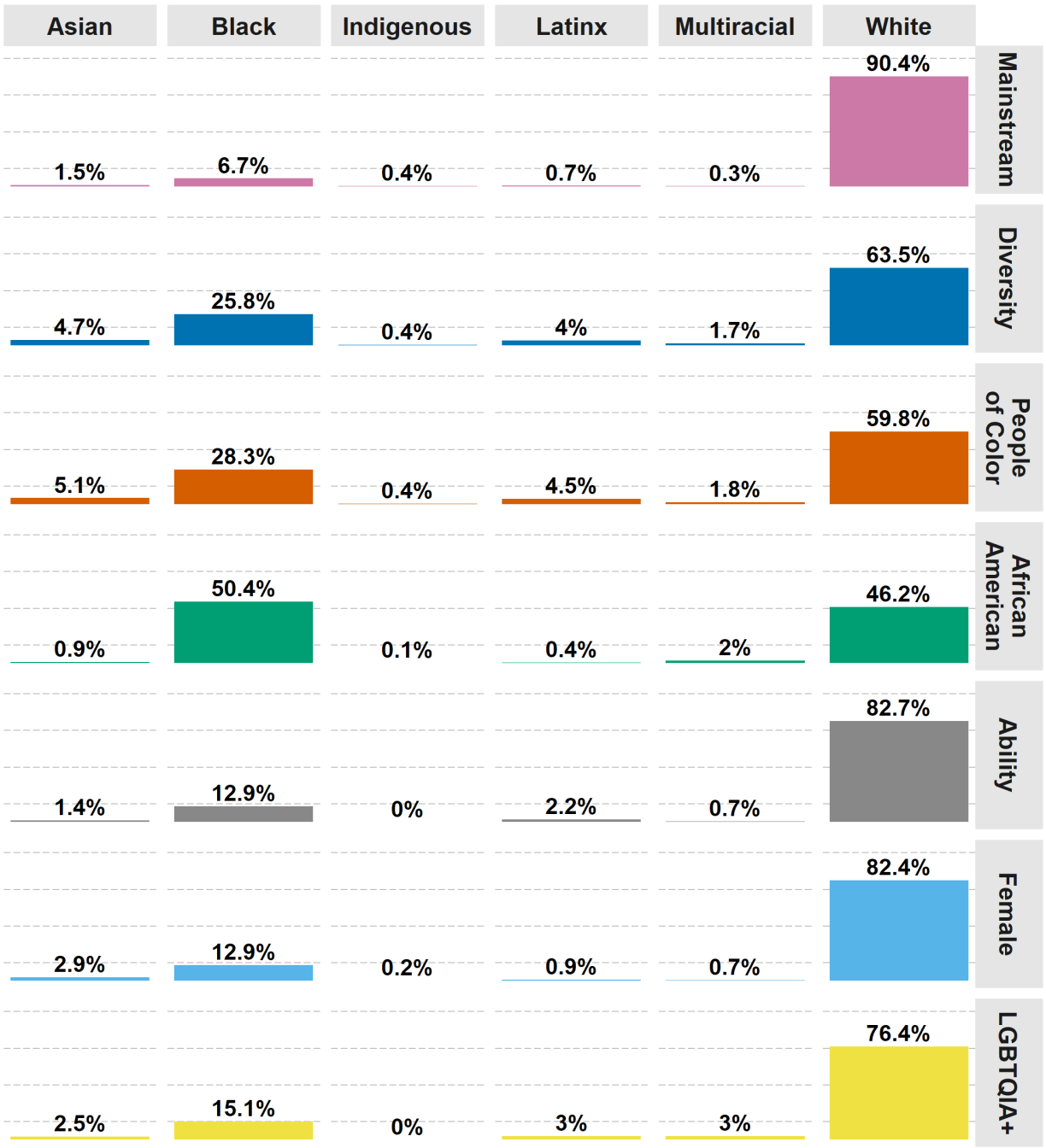
(a) Detected Faces (Images)



(b) Famous Figures (Text)

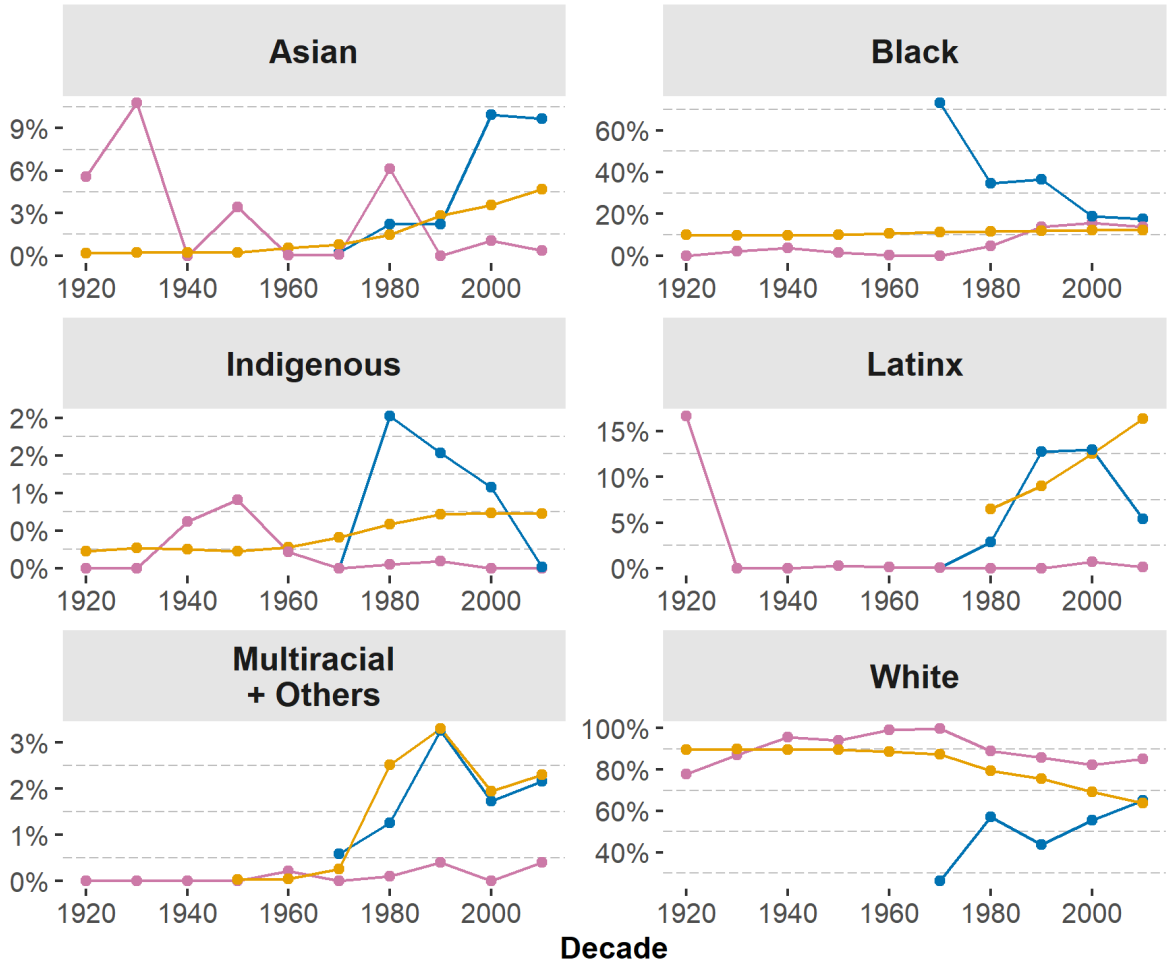
Note: In this figure, we show the share of the characters by race and gender in a given book averaged over all books in a collection and decade. In Panel A, we show this for detected faces in images. In Panel B, we show this for famous figures mentioned in the text.

Figure B6. Race Classifications of Famous Figures in the Text



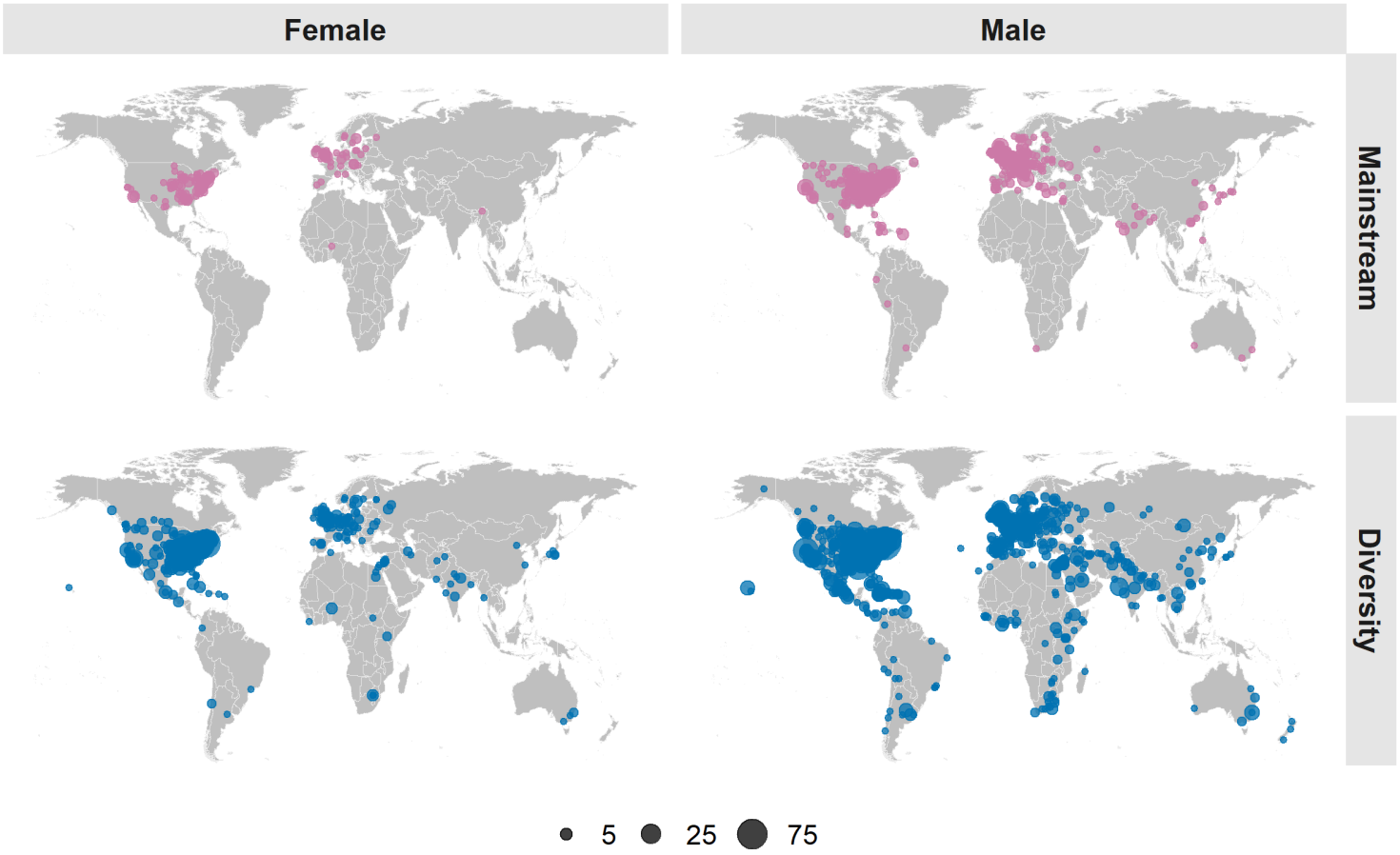
Note: In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be Black. We identify famous individuals using methods described in Section D.C.3. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial.

Figure B7. Share of U.S. Population and Famous People in the Text, by Race/Ethnicity



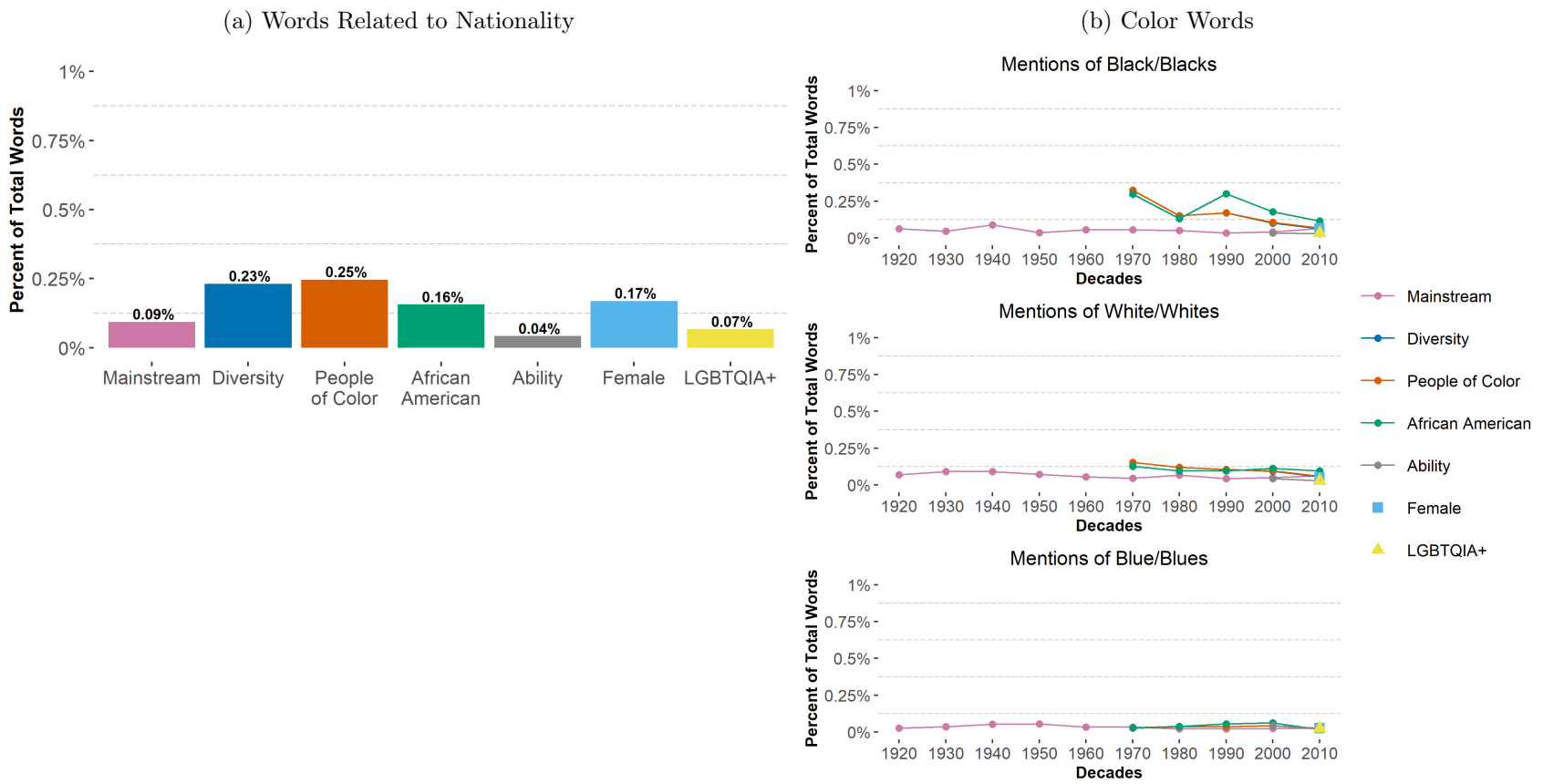
Note: In this figure, we show the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin was mentioned 3 times in a book and Jimmy Carter is mentioned 2 times, then 60 percent of the mentions of famous people in that book would be Black. We then show the average percentage breakdown over all books by collection and decade for the Mainstream and Diversity collections. We also show the share of the U.S. population by race/ethnicity for each decade as a comparison. We classify famous people using methods described in Section D.C.3. We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, we classify them as multiracial. Note that this is an analog to Figure 9, only with the y-axis collapsed to the maximum level for each race/ethnicity, respectively, to present easier to parse patterns for groups with lower levels of representation.

Figure B8. Birthplace of Famous Figures, by Gender



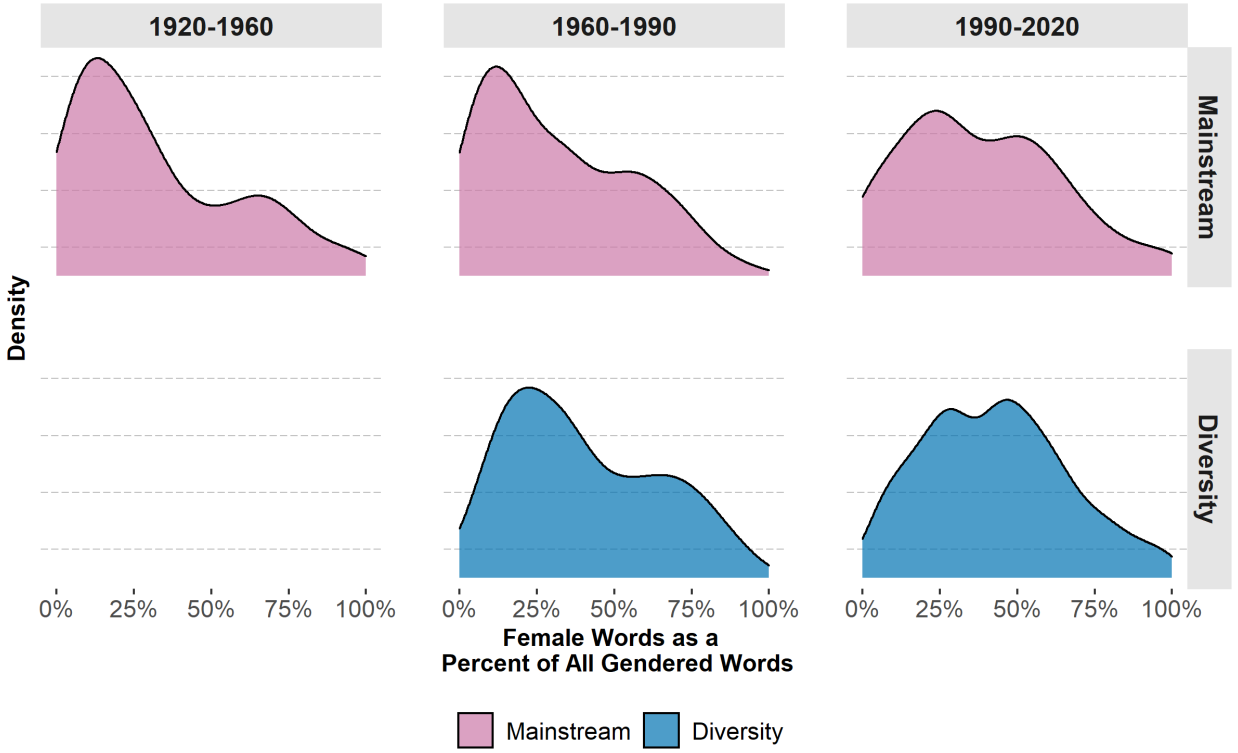
Note: In this figure, we show collection-specific measures of the birthplace of famous figures, separately for females and males. We identify famous individuals as well as their gender and birthplace using methods described in Section D.C.3. If the city/town they were born in was unavailable, we use birth country. Size of dots correspond to the number of famous characters born in a given location that are mentioned at least once in a given book and then aggregated across all books in a collection. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be Black. Note that this is an analog to Figure 10, only here with the maps shown separately for female and male famous figures.

Figure B9. Token-Based Proxies for Race: Nationality and Color



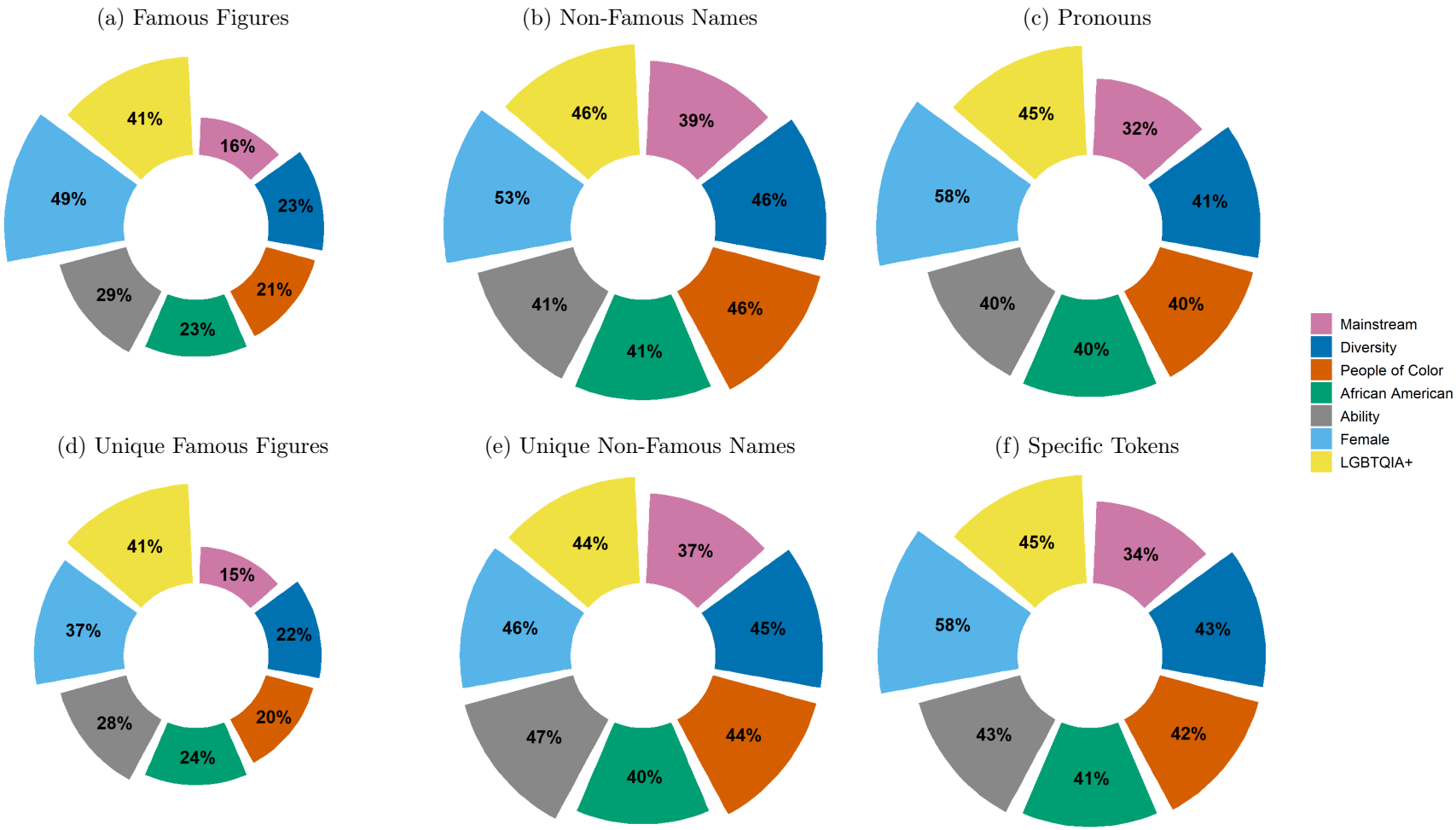
Note: In this figure, we show two measures of the representation of race in text: words related to nationalities and words related to color. In Panel A, we show collection-specific averages of the proportion of words in a book that relate to nationalities. In Panel B, we show collection-by-time averages of mentions of three color words: black, white, and blue – as a proportion of all words in our data. We generated the estimates using a pre-specified list of words (also known as "tokens," as described in Section D.C.1). We provide this list in the Data Appendix.

Figure B10. Distribution of Female Words as a Percent of All Gendered Words, Over Time



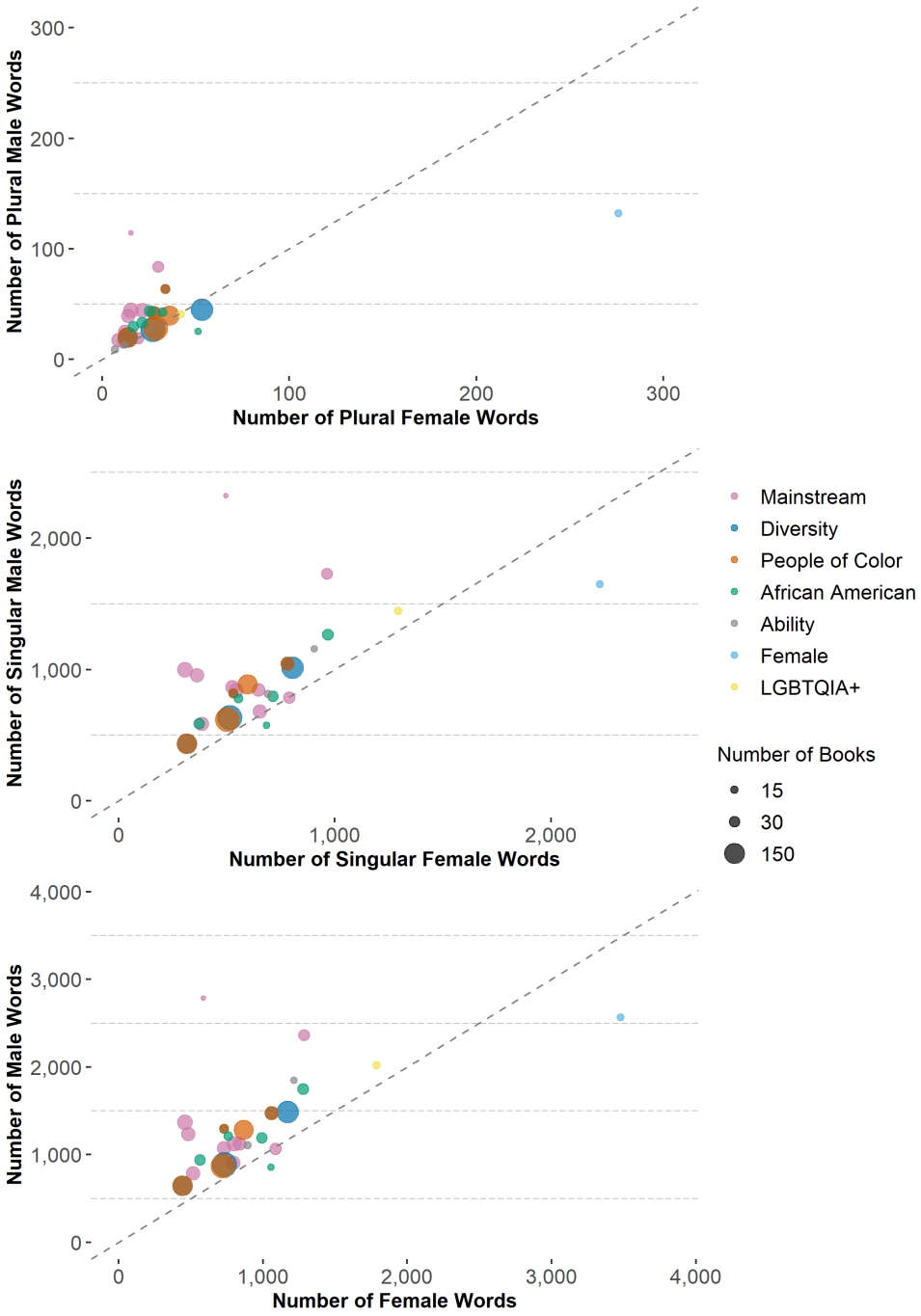
Note: In this figure, we show the distribution of female words as a percentage of all gendered words over time in the Mainstream and Diversity collections. In this case, gendered words encompass the total number of gendered first names, gender predictions of famous characters, gendered pronouns, and a pre-specified list of other gendered tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix.

Figure B11. Female Representation in Text, by Type of Word



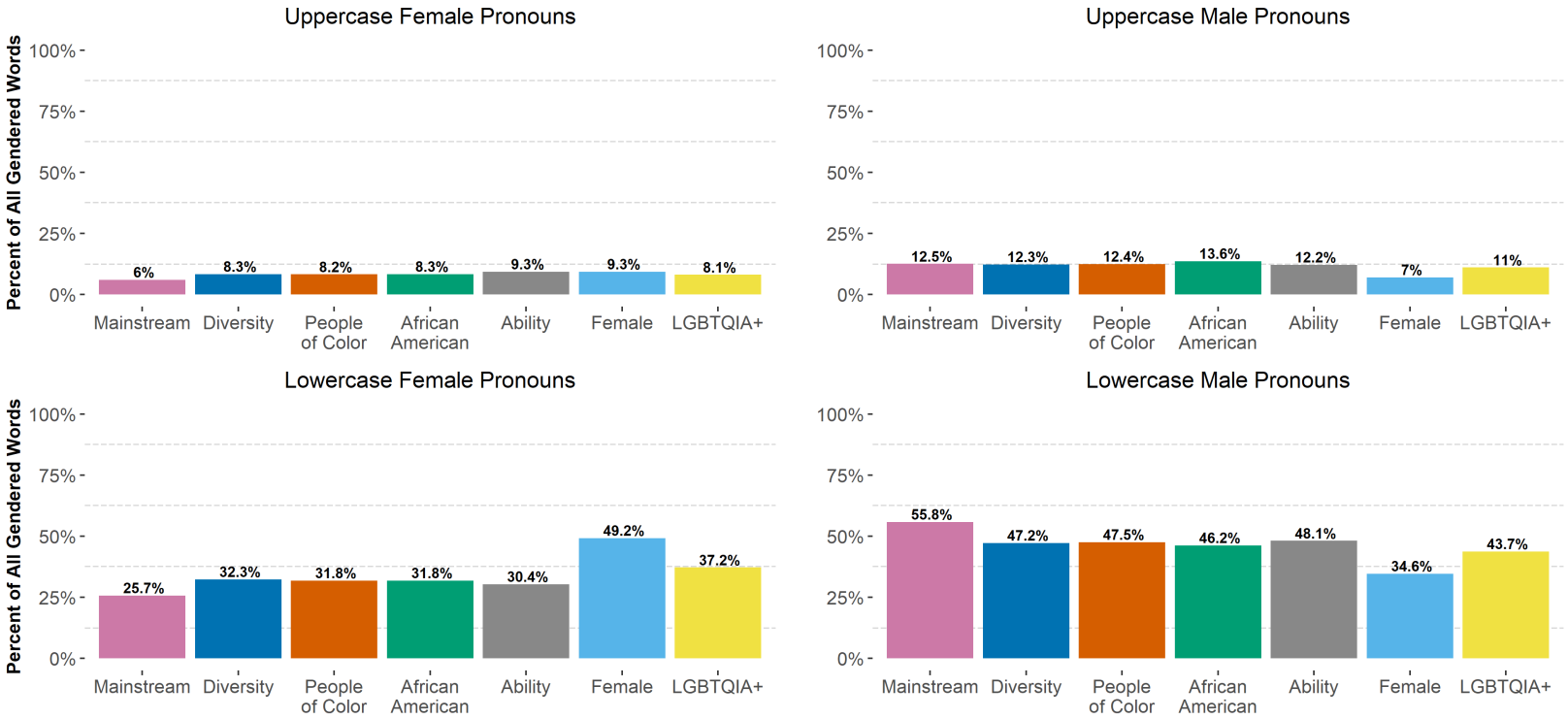
Note: In this figure, we show the proportion of female representation in the text by collection and type of word. In Panel A we find the percent breakdown of female famous people mentioned in a given book, averaged over all books in a collection. For example, if Aretha Franklin was mentioned 4 times in a book and Jimmy Carter is mentioned 2 times, then 60 percent of the famous people mentioned in that book would be female. In Panel B, we show the same thing as Panel A, but for mentions of non-famous names. Panel C shows the percentage of gendered pronouns which are female in a given book, averaged over all books in a collection. In Panel D we show the percentage breakdown of unique female famous people in a collection. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be female. In Panel E we show the same thing as Panel D but for unique non-famous names. Panel F, shows the percentage of female words from a pre-specified list of gendered words (tokens) such as queen or nephew (full list provided in Data Appendix) in a given book, averaged over all books in a collection.

Figure B12. Gender Representation, by Quantity of Individuals



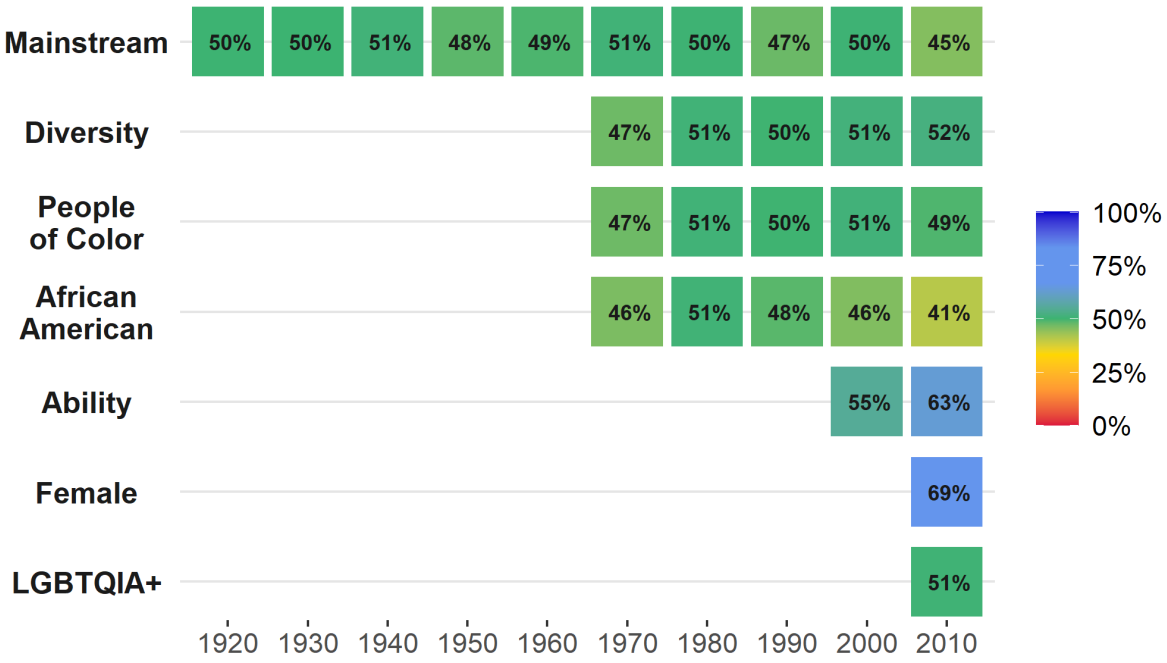
Note: In this figure, we show how gender representation in text varies by whether it is referring to an individual or a group of people; in other words, whether the representation of gender varies by presence of singular (individuals) or plural (groups of people) gendered words. We show collection-specific averages by decade. In the top plot, we show the number of plural male words vs. the number of plural female words; in the bottom plot, we show the number of singular male words vs. the number of singular female words. These male and female words were drawn from a pre-specified list of other gendered tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix. We also show the total number of male words and the total number of female words for reference.

Figure B13. Proportion of Females and Males Serving as Subjects and Objects of Sentences



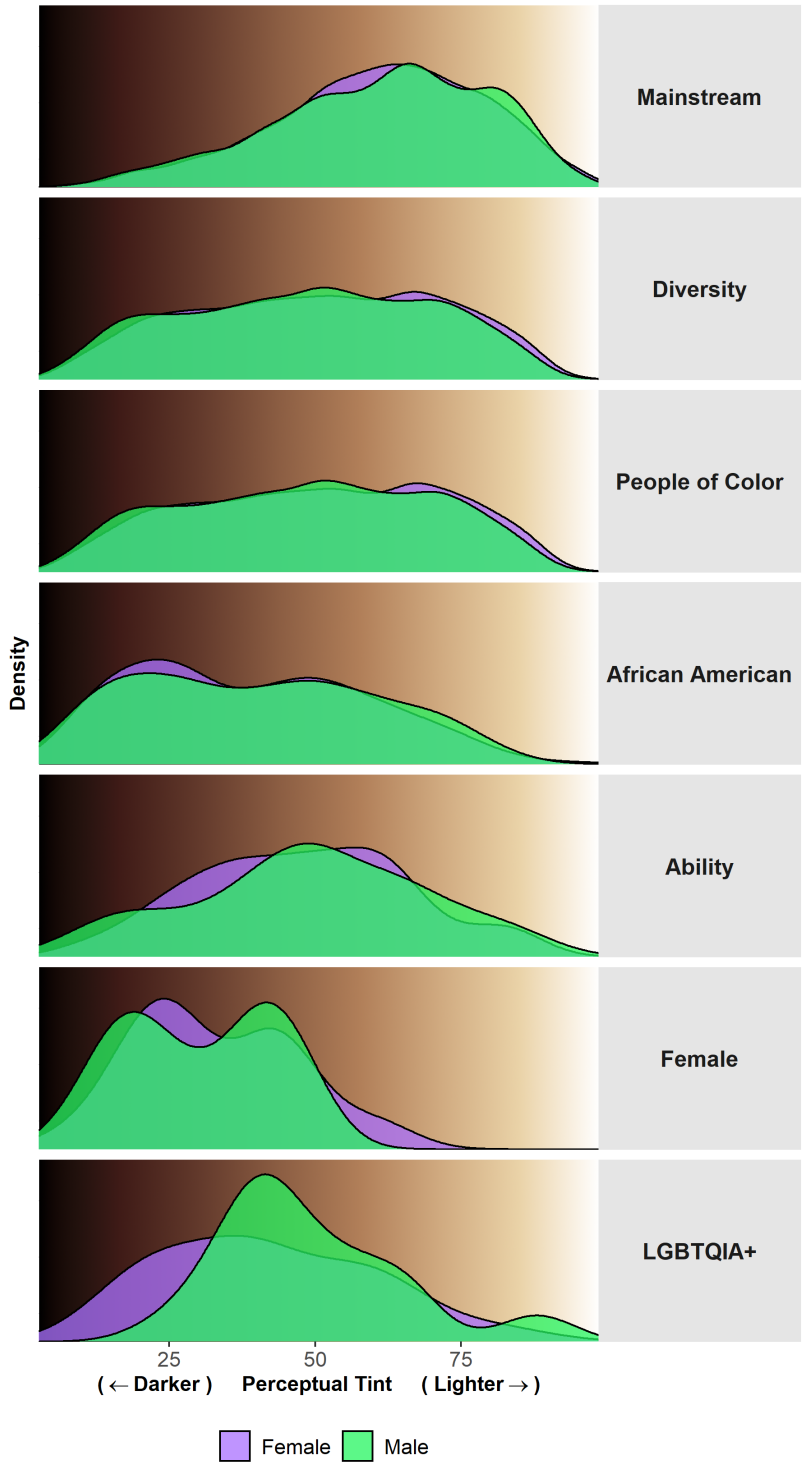
Note: In this figure, we plot the representation of gender by its location in sentences. The top two plots show the average proportion of all gendered pronouns in a book that are uppercase, and the bottom two plots show those that are lowercase. The left plots show the female-related pronouns, and the right plots show the male-related pronouns. We present these separately because an uppercase pronoun is more likely than a lowercase pronoun to be the subject, as opposed to the object, of the sentence in which it appears.

Figure B14. Average Probability a Face is Female, by Decade and Collection



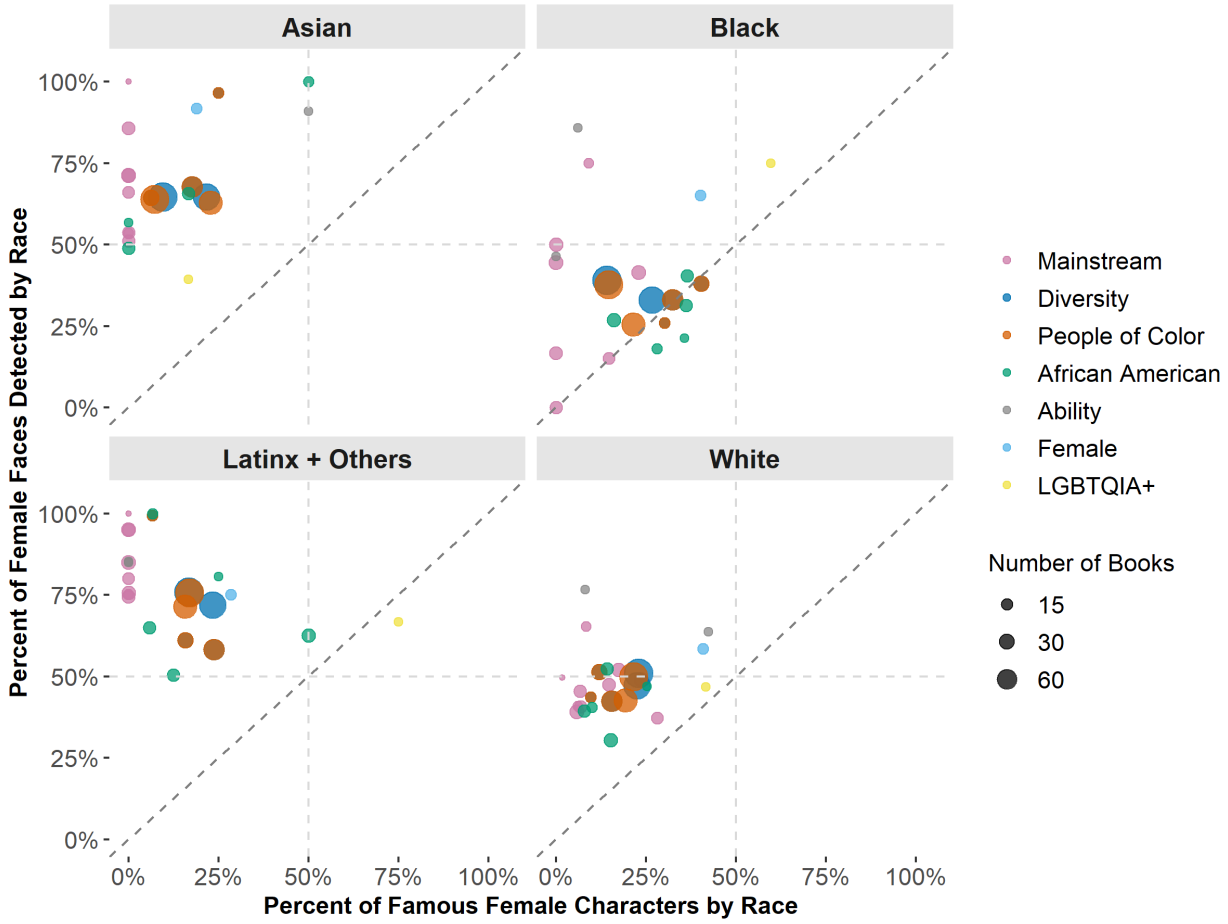
Note: In this figure, we present the average probability that a face was classified as being female in a given collection by decade. We classify gender using an AutoML algorithm trained on the UTKFace public data set.

Figure B15. Distribution of Skin Color, by Gender and Collection



Note: In this figure, we show the distribution of skin tint by gender. We detect faces using our face detection model (FDAI) described in Section III.A. Within these faces, we classify gender using an AutoML algorithm we trained using the UTKFace public data set. Skin tint is determined by the L* value of a face's representative skin color in L*a*b* space. We extract a face's representative skin color using methods described in Section III.B. These figures show the results for images that have identified human skin colors (polychromatic skin colors where R $\geq$ G $\geq$ B).

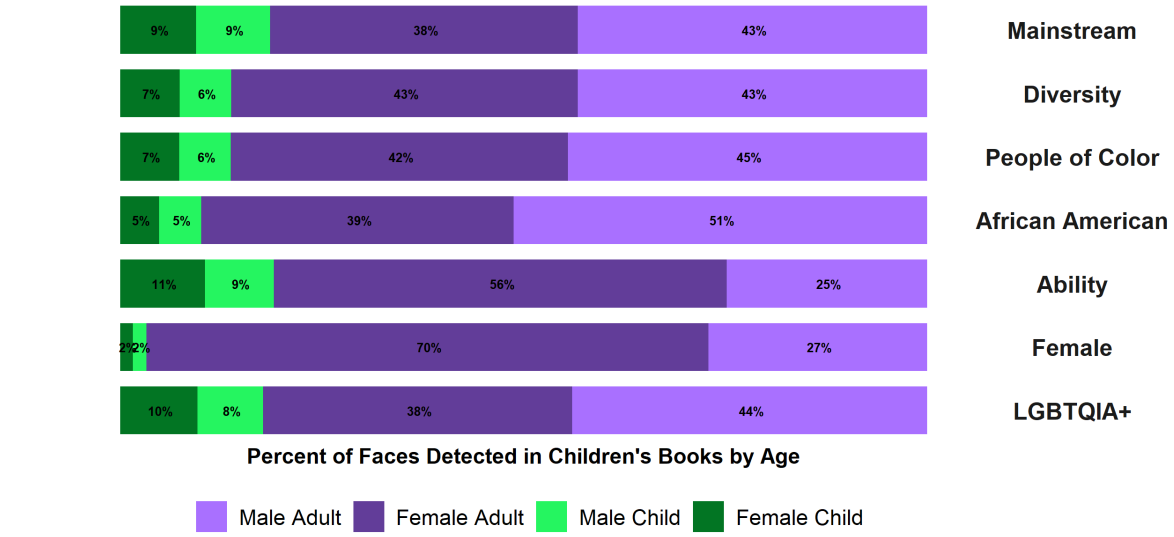
Figure B16. Race and Gender Representation in Images and Text



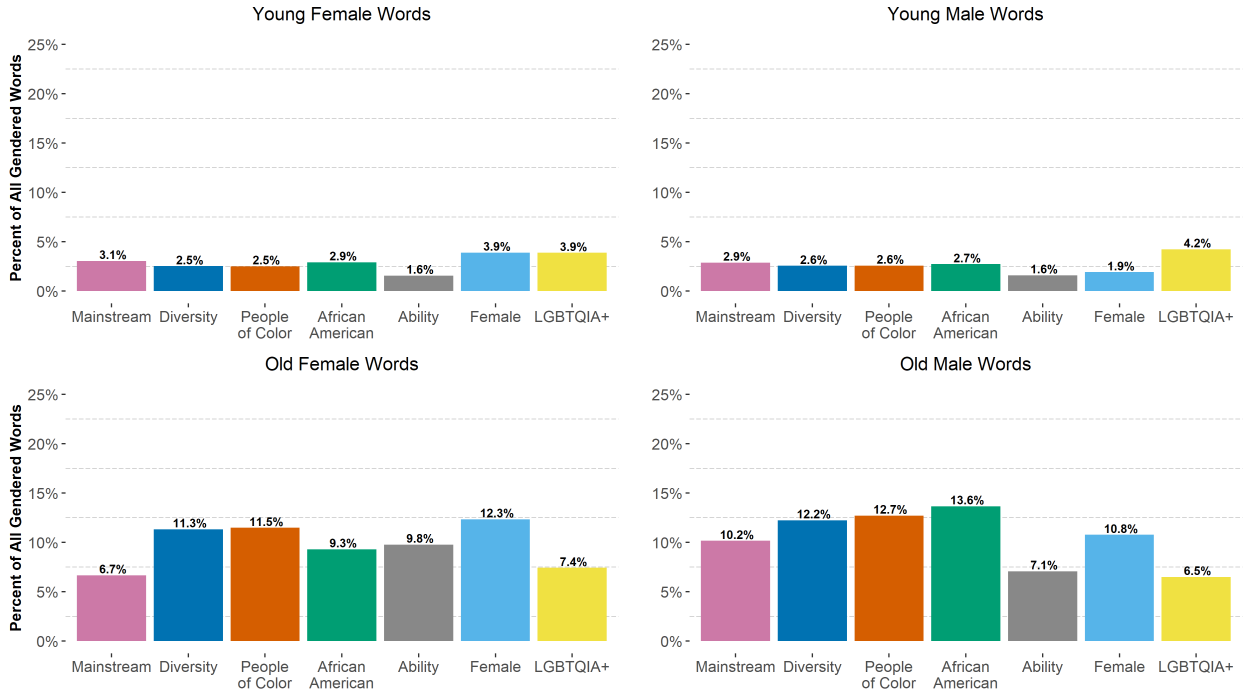
Note: In this figure, we plot female faces by race as a proportion of all faces with a given race classification on the y-axis and famous female characters by race as a proportion of all famous characters with a given race classification on the x-axis.

Figure B17. More Adults than Children in Images and Text

(a) Percent of Faces by Predicted Age Group and Gender

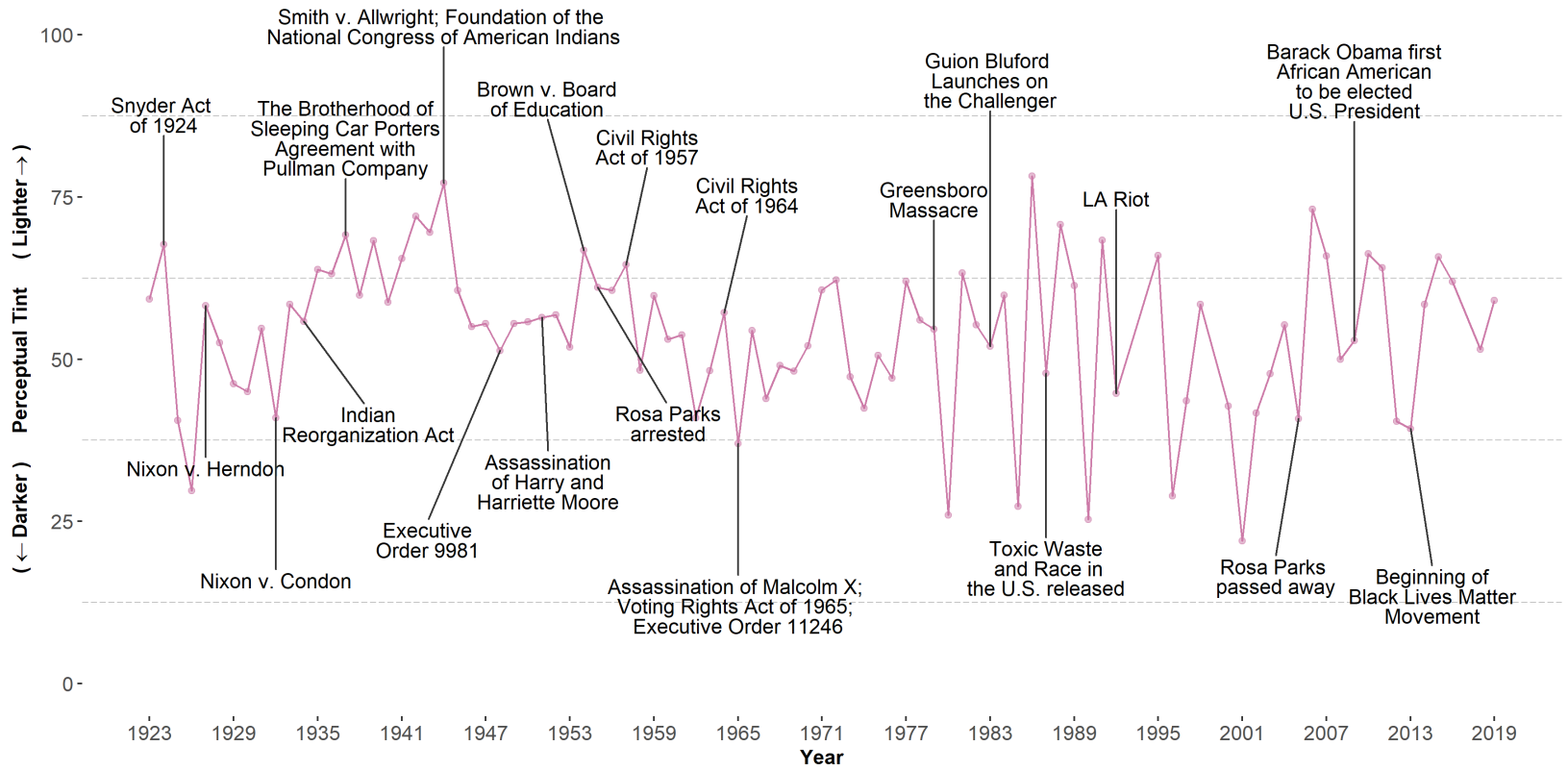


(b) Percent of Gendered Words by Age Group



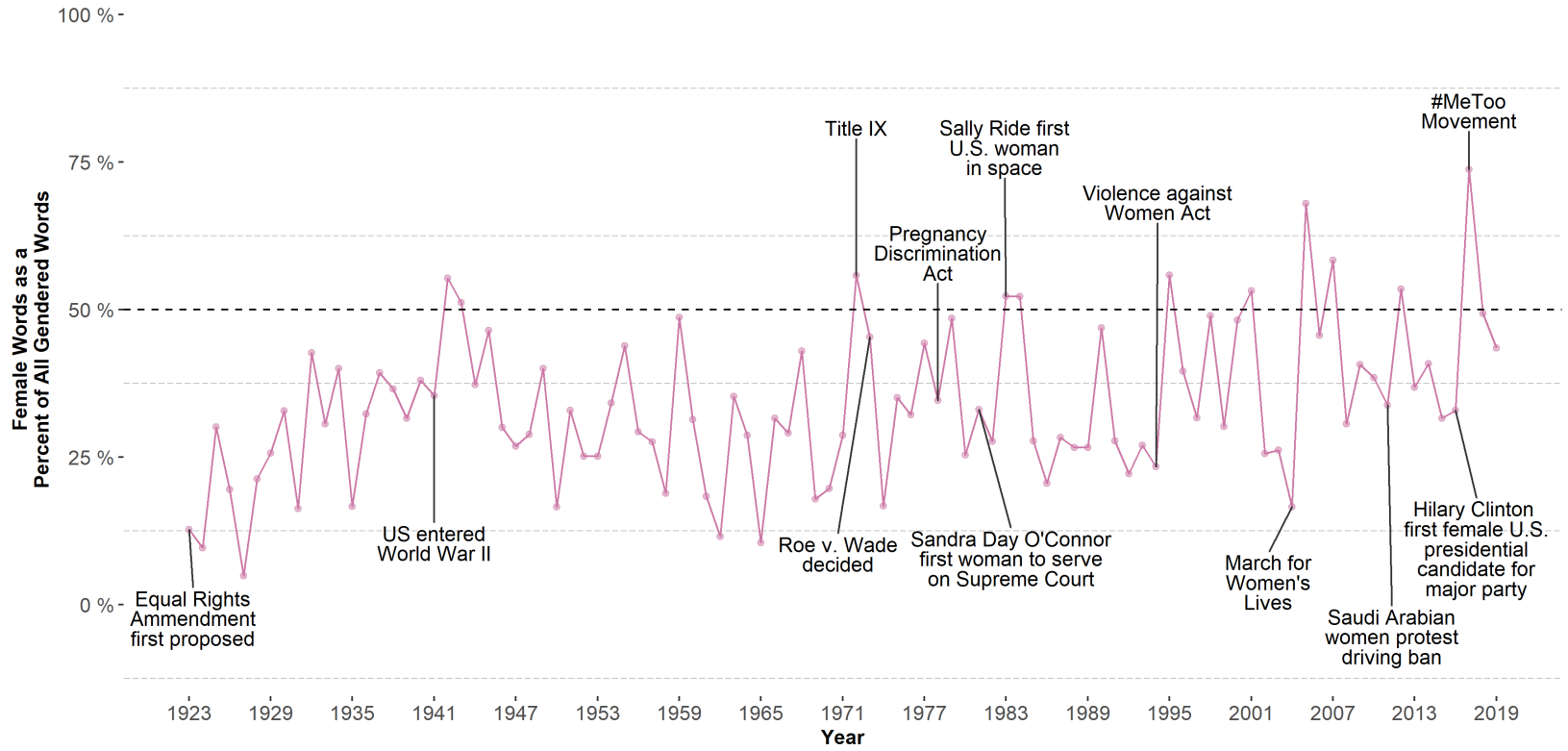
Note: In this figure, we show analysis of the representation of age and gender. In Panel A, we show analysis of predicted age and gender in the faces in images. Specifically, we plot the proportion of identified faces classified in each age (adult vs. child) and gender (female vs. male) category. In Panel B, we show analysis of age and gender in text. Specifically, we plot the proportion of words that refer to specific gender-age combinations (e.g., female adults or male children) as a percent of all gendered words in the book. Gendered words encompass the total number of gendered first names, gender predictions of famous characters, gendered pronouns, and a pre-specified list of other gendered tokens (e.g., queen, nephew). We list the pre-specified gendered tokens in the Data Appendix.

Figure B18. Mainstream Representation of Skin Color Throughout Historical Events



Note: In this figure we juxtapose measures of representation of skin color of pictured character faces from the Mainstream collection with the timing of salient historical events.

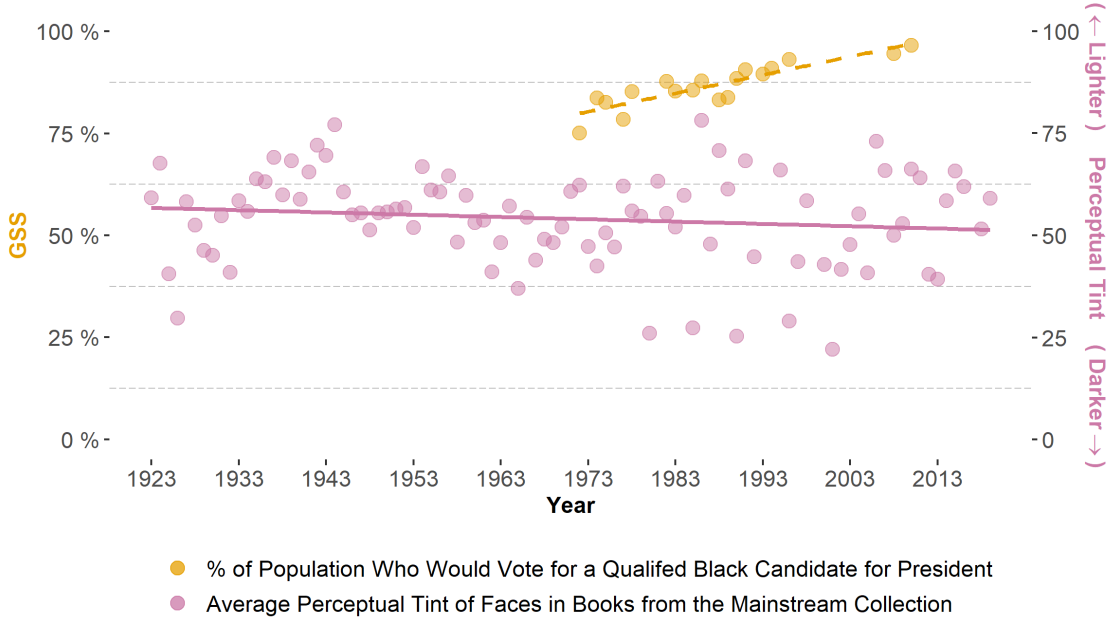
Figure B19. Mainstream Representation of Gender Throughout Historical Events



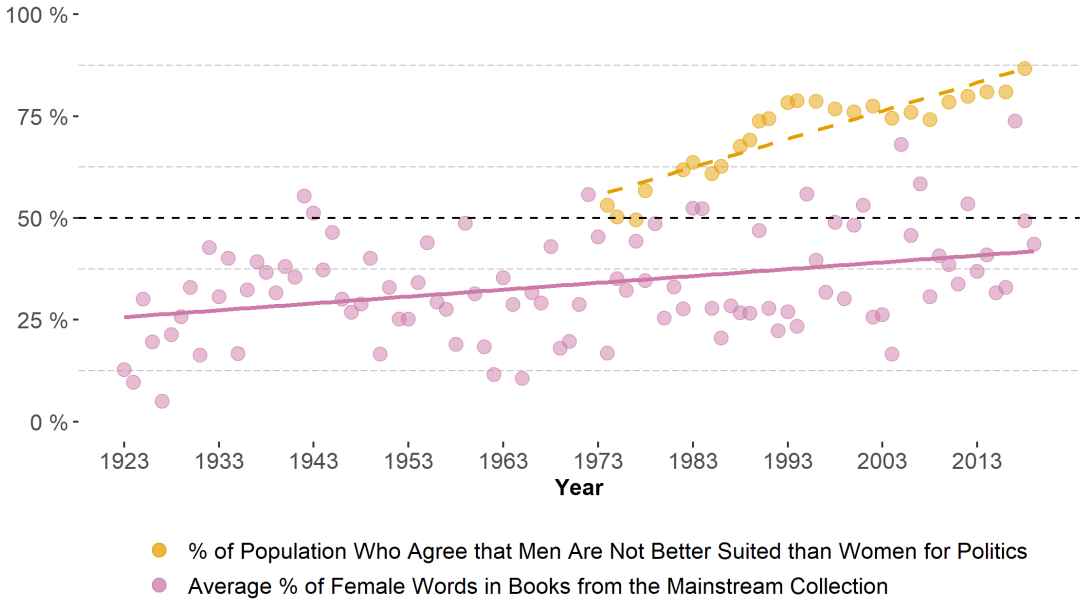
Note: In this figure we juxtapose textual measures of gender representation from the Mainstream collection with the timing of salient historical events.

Figure B20. Mainstream Representation and Social Attitudes Over Time

(a) Race



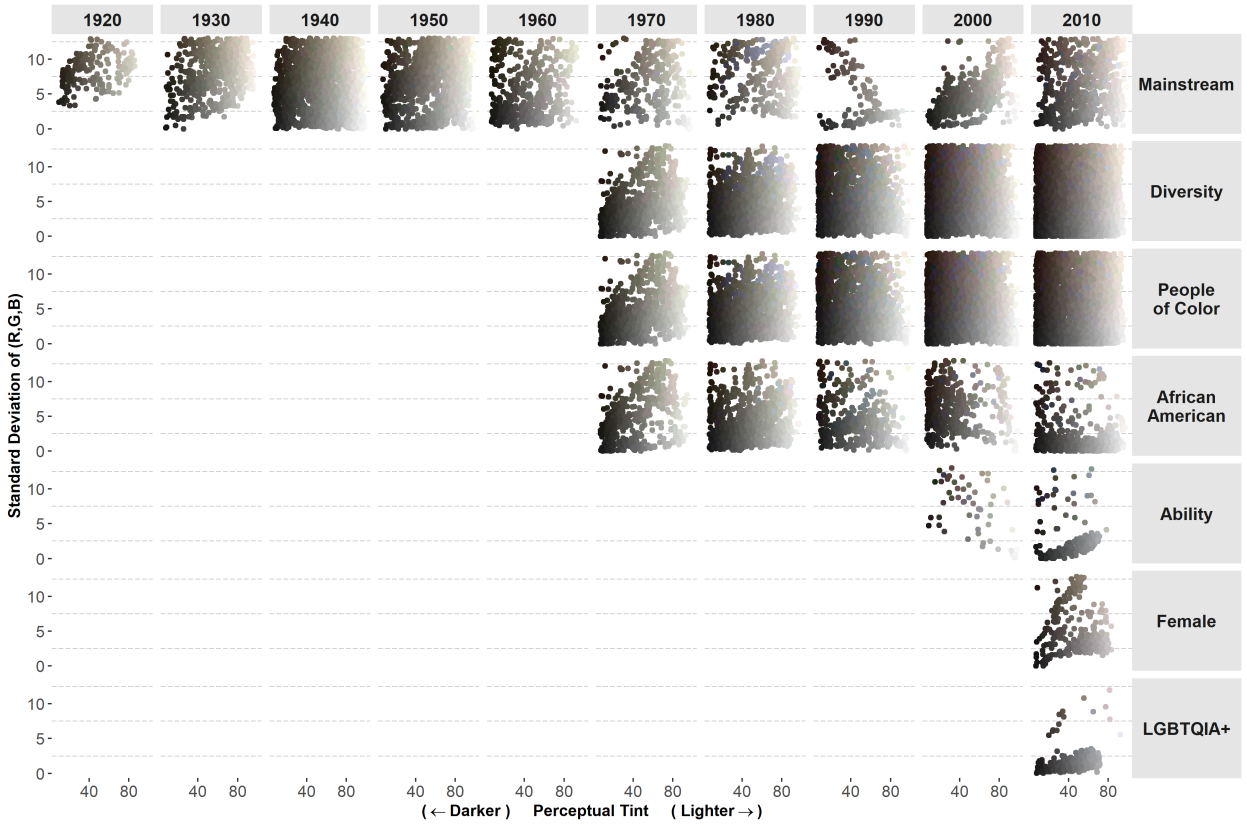
(b) Gender



Note: In this figure we compare trends in social attitudes with yearly representation in the Mainstream collection over time. In Panel A we show the proportion of respondents who would vote for a qualified Black candidate for president along with the average skin tint of faces found in books within the Mainstream collection by year. In Panel B we show the proportion of respondents who agree that men are not better suited than women for politics along with the average percent of female words in books within the Mainstream collection by year. Our data on social attitudes comes from the General Social Survey (GSS).

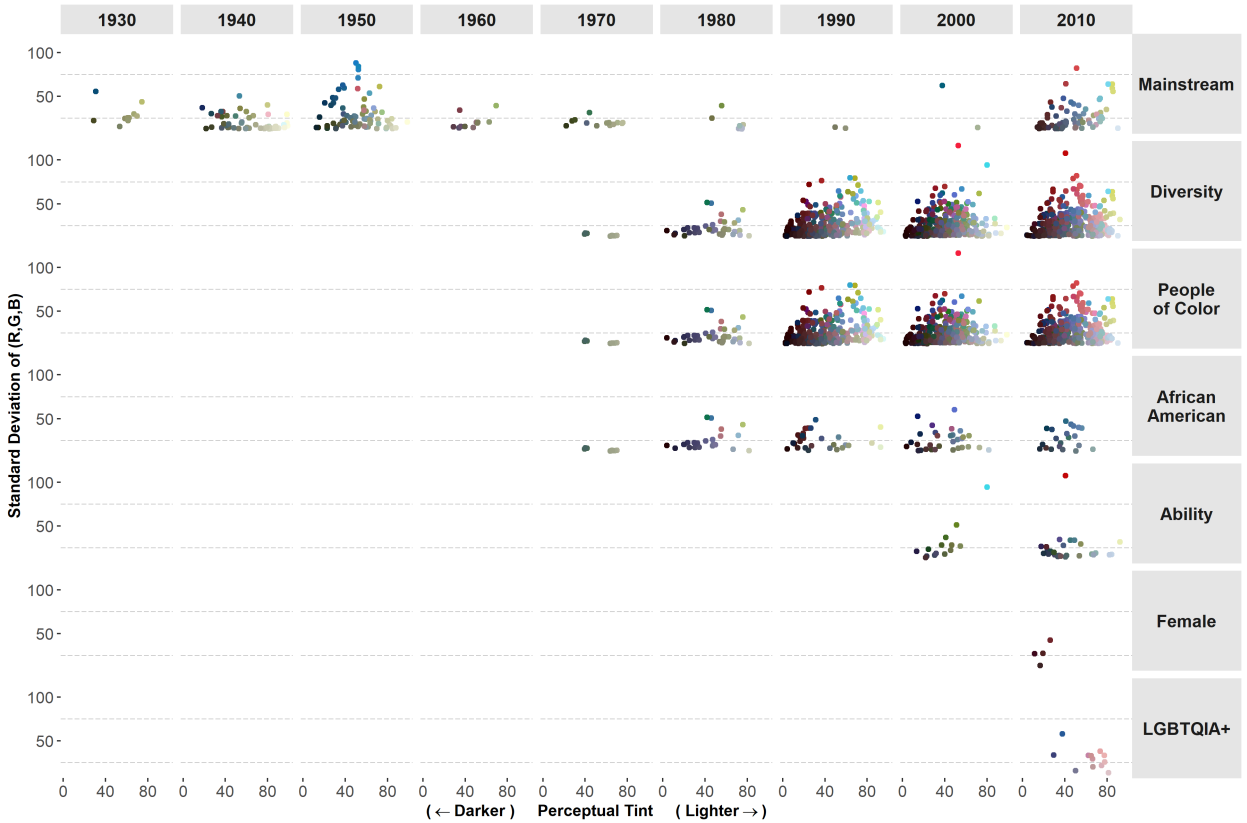
C Non-Typical Skin Color Appendix

Figure C1. Skin Color Data Over Time, Monochromatic Skin Colors



Note: In this figure we show an analog to Figure B1, here focusing on the representative skin colors for all detected faces with monochromatic skin colors (e.g., black and white) in each collection-by-decade. As described in Section III, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section III.B.

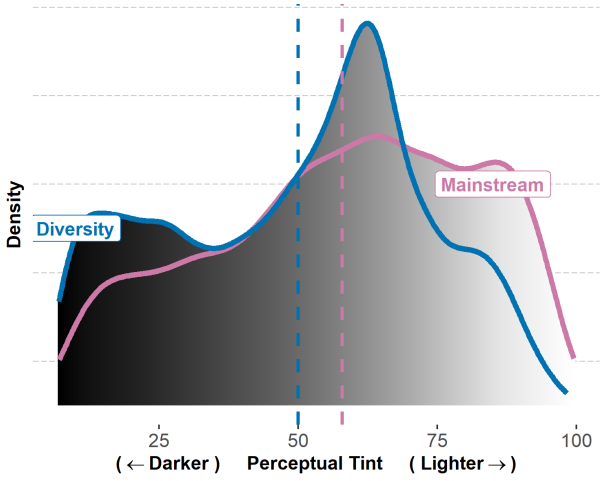
Figure C2. Skin Color Data Over Time, Polychromatic Non-Typical Skin Colors



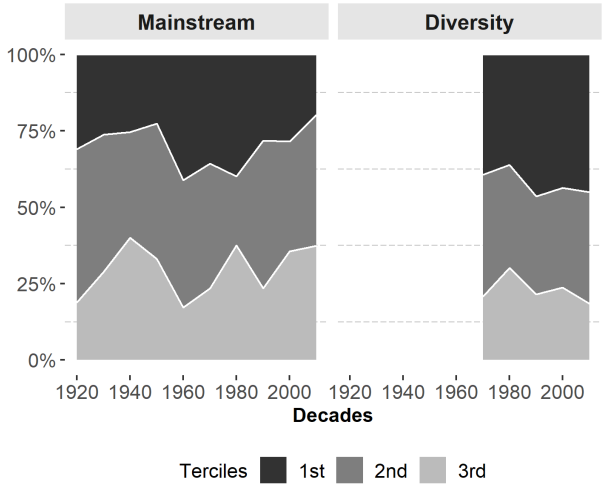
Note: In this figure, we show an analog to Figure B1, here focusing on the representative skin colors for all detected faces with non-typical skin colors (e.g., blue or green) in each collection-by-decade. As described in Section III, we use our face detection model (FDAI) trained on illustrations to classify faces in images. We determine a face's representative skin color using methods described in Section III.B. The data shown in this figure begin in the 1930s, as opposed to in the 1920s as in Figures B1 and C1 found in the Non-Typical Skin Color Appendix, because we detect no faces with polychromatic non-typical skin colors in books from the 1920s.

Figure C3. Skin Colors in Faces, by Collection: Monochromatic Skin Colors

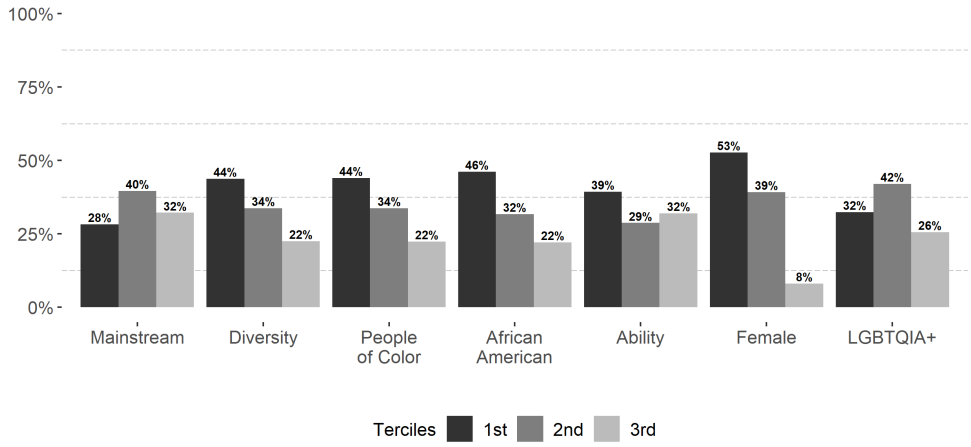
(a) Distribution of Skin Colors



(b) Mean Proportion in Each Tercile, Over Time



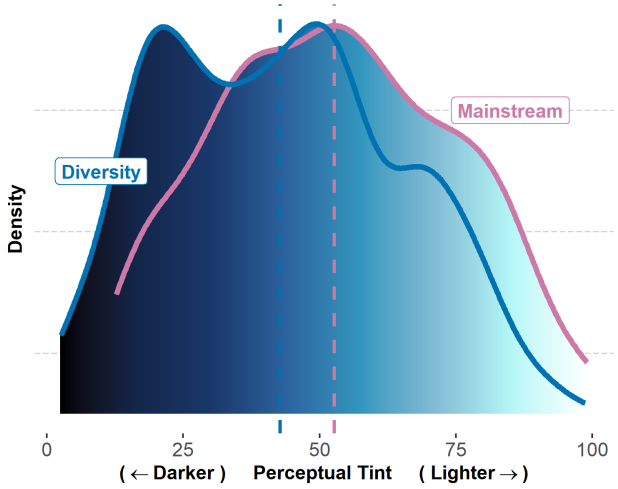
(c) Mean Proportion in Each Tercile, All Collections



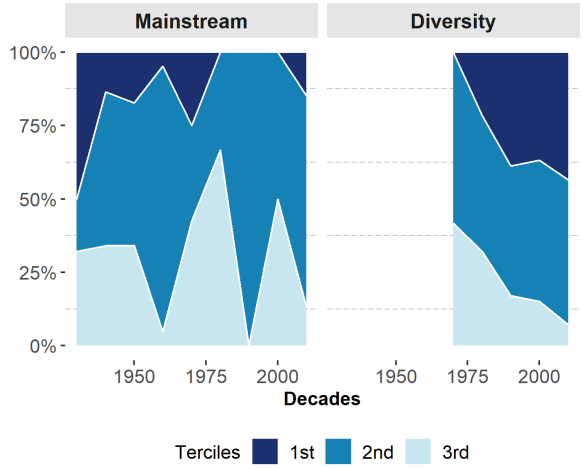
Note: This figure shows our analysis of the representative skin colors of the individual faces we detect in the images found in the books we analyze. This is an analog to Figure 5, only here we focus on monochromatic faces. Panel A shows the distribution of skin color tint for faces detected in books from the Mainstream and Diversity collections. The mean for each distribution is denoted with a dashed line. In Panel B, we show the average proportion of faces in each tercile, over time, for faces in the Mainstream and Diversity collections. Panel C shows the overall collection-specific average proportion of faces in each skin color tercile for each of the seven collections. Skin classification methods are described in Section III.

Figure C4. Skin Colors in Faces, by Collection: Polychromatic Non-Typical Skin Colors

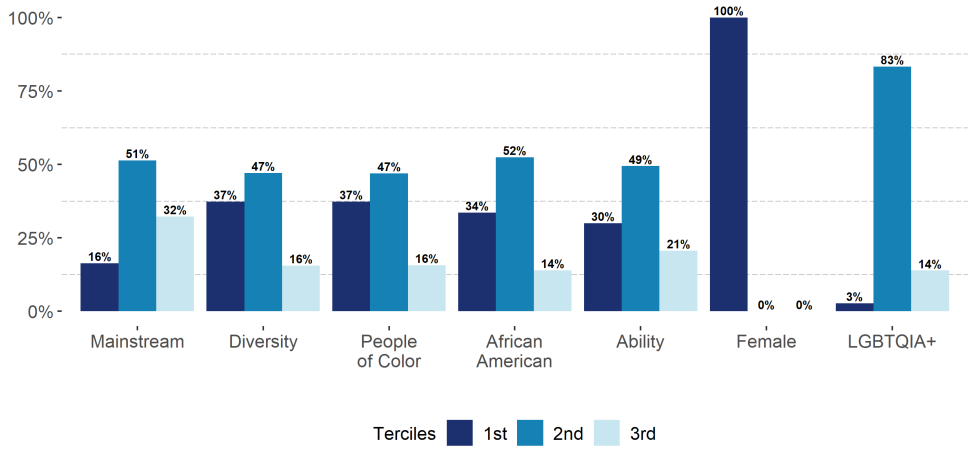
(a) Distribution of Skin Colors



(b) Mean Proportion in Each Tercile, Over Time

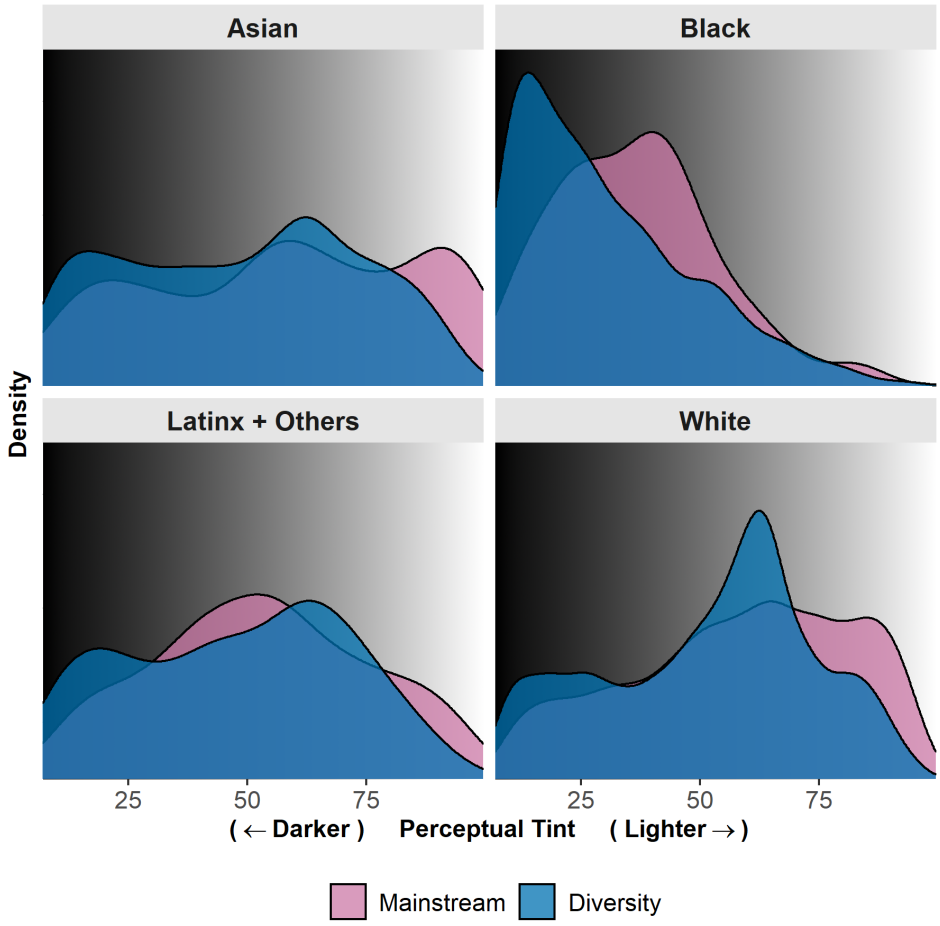


(c) Mean Proportion in Each Tercile, All Collections



Note: This figure shows our analysis of the representative skin colors of the faces detected in the books we analyze. This is an analog to Figure 5, only here we focus on faces that have non-typical skin colors. Panel A shows the distribution of skin color tint for faces detected in books from the Mainstream and Diversity collections. The mean for each distribution is denoted with a dashed line. In Panels B and C, we show the average proportion of faces in each tercile of the perceptual tint distribution across all books in a collection. In Panel B, we show the average proportion of faces in each tercile, over time, for faces in the Mainstream and Diversity collections. Panel C shows the overall collection-specific average proportion of faces in each skin color tercile for each of the seven collections. Skin classification methods are described in Section III.

Figure C5. Skin Color by Predicted Race of Pictured Characters: Monochromatic Faces



Note: This figure shows the distribution of skin color tint by predicted race of the detected faces in the Mainstream and Diversity collections. This is an analog to Figure 6, only here focusing on faces depicted in a monochromatic color scheme (e.g., black and white). Skin tint is determined by the L* value of a face's representative skin color in L*a*b* space. We extract a face's representative skin color using methods described in Section III.B. Race was classified by our trained AutoML model as described in Section III.C.

D Methods Appendix

D.A Comparing AI with Manual Content Analysis

In this section, we briefly describe the relationship our analysis has with manual content analysis. We first describe the key differences in scope and reach between our suite of computerized content analysis tools and the tools used by the field of manual content analysis. We then describe how we used manual content analysis to validate our measures. Finally, we conduct a cost-effectiveness analysis which highlights a key advantage of our approach – far greater reach in terms of the ability to measure representation in an entire book, to respond nimbly to changes in analysis plans, and significantly lower cost.

Our use of automated content analysis provides a series of key advantages over manual content analysis. The first important advantage is speed. Our suite of computerized analysis tools can process a very large amount of content in a short period of time. While previous, well-resourced efforts to conduct manual content analysis study the content of between fifty and three hundred books (Weitzman et al., 1972; Davis, 1984; Crisp and Hiller, 2011; Koss, Johnson and Martinez, 2018), our analysis included over one thousand books. Furthermore, in this analysis the most binding constraint was acquiring digitized versions of book content. Were our tools to be used by publishers, libraries, or other entities with access to large bodies or universes of relevant digitized books and other curricular materials, these analyses could be performed even more quickly and inexpensively on collections of books that were previously entirely infeasible to analyze because of speed and cost.

The second important advantage is scope. Our suite of tools is able to analyze all characters contained in the image and text. This is in important contrast to the vast majority of manual content analysis we are aware of, particularly those with larger sample sizes (e.g., in the low hundreds of books). These analyses are able to study a greater number of books by focusing on a smaller number of prominent features, such as the book's title, the images on its cover, and the identities of only the main characters, and such studies often explicitly mention making such restrictions in order to keep the costs of content analysis manageable (Kortenhaus and Demarest, 1993; McCabe et al., 2011; Koss, 2015; Koss, Johnson and Martinez, 2018).

In manual content analysis, the marginal cost of coder time increases with the addition of new dimensions of study. Furthermore, if re-analysis or new analysis is required after the initial analysis appears again, the fixed costs of identifying, hiring, and training coders are again incurred. For automated content analysis, the only cost is the computational power and the cost of adjusting the code, allowing for far greater flexibility and scope in addition to substantially lower costs.

A third advantage of computerized content analysis is reliability. In manual content analysis, inter-rater reliability is a core concern which increases with scale (Neuendorf, 2016; Krippendorff, 2018). In computer-driven analysis, however, these concerns do not vary with scale, as the traits of the coder are held constant.

The key historical advantage of manual content analysis has been its superiority in measuring more complex and nuanced understanding than those we capture here. Our focus in this study is primarily on measuring the presence of different identities, a domain for which computer-driven analysis is particularly suitable. Furthermore, recent advances in natural

language processing and computer vision have begun to make progress in incorporating the analysis of more complex features into the toolkit of computerized content analysis (Nenkova and McKeown, 2012; Ouyang and McKeown, 2014; Caliskan, Bryson and Narayanan, 2017; Garg et al., 2018).

Next, we describe our work to validate our tools using manual content analysis. Drawing from validation theory, we conducted traditional manual content analysis to validate our measures (Kane, 2013; Neuendorf, 2016). To do so, we hand-coded representations in 30 short stories and poems for children written and illustrated by a variety of authors and illustrators from a third grade reading textbook published in 1987. This helped us to evaluate the plausibility of our measures and also identify messages our tools failed to detect, clarifying limitations of computer-led content analysis.

Finally, we estimate a rough measure of the cost-effectiveness of our tools. It took approximately 40 hours to code the entire book (400 pages at an average of 6 minutes per page).[55] While the length of time needed to code "by hand" varies with the grade level of the books in our sample, we estimate that it would have taken us over 16,000 hours to hand-code the 162,872 pages in our sample of children's books. At an hourly wage of between $15 and $20, we estimate this work would have cost between $244,000 to $326,000.

D.B Images as Data

D.B.1 Image Feature Classification: Face Detection Methods

To train our face detection model, we split our manually labeled data set into training (80 percent of the data), validation (10 percent of the data, used for hyper-parameter tuning), and testing (10 percent of the data, used for evaluating the model).[56]

The manually labeled test data are kept separate from the training and hyper-parameter tuning algorithms.[57] The models compare results from the algorithms to the manual labels in the test data to evaluate the accuracy of the algorithms.

[55] Hand-coding of pages entails documenting a wide variety of features in image and, separately, text, which is a time- and detail-intensive process. Our estimate of six minutes per page represents a lower bound on the time needed to perform the type of analysis we conducted. In this case, for example, the manual coders did not count every token that could be related to gender, nationality, and color.

[56] The validation data are used for hyper-parameter tuning to optimize the model architecture. Hyper-parameter tuning involves "searching" for the optimal values of the hyper-parameters. Examples of hyper-parameters include learning rate, number of epochs (number of times the model goes through the whole data set), and different activation functions of the model that can be tuned to improve the accuracy of the model. FDAI is using Google Cloud infrastructure and functions to test different hyperparameter configurations and chooses the set of hyperparameters that maximize the model's accuracy.

[57] The manually labeled data for the face detection model came from data labeled by our research team. The manually labeled data for the feature classification model came from the UTKFace data set.

We use two specific parameters that are commonly used to evaluate the performance of this class of model: "precision" and "recall."[58] Precision is the proportion of items which are correctly assigned a label out of all items that *are assigned* that label. For example, precision for detected faces is the number of actual faces out of all regions in an image that our model classifies as a face (that might not always be a face). Recall, on the other hand, tells us the percentage of items that are correctly assigned a label out of all items that *should be assigned* that label. In the case of recall for faces, recall is the proportion of the number of correctly detected faces out of the actual number of faces in the book.[59] Formally:

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives}$$

The higher the precision, the fewer false positives the model produces. In other words, precision tells us from all the test examples that were predicted with a certain label, which ones are truly of that label? On the other hand, the higher the recall, the fewer false negatives the model produces. In other words, recall tells us, from all the test examples that should have had the label assigned, how many were actually assigned the label (Sokolova and Lapalme, 2009).

Our face detection model has 93.4 percent precision (6.6 percent of identified faces may not be true faces) and 76.8 percent recall (1 in 4 true faces may not be identified).

D.B.2 Image Feature Classification: Skin Segmentation Methods

Traditional skin segmentation methods assign a skin or non-skin label for every pixel of the cropped face image in which skin features are extracted. These labels are assigned using traditional image processing methods such as thresholding, level tracing, or watershed. These methods, however, face a number of challenges such as the need to take into account skin color (in)consistency across variations in illumination, acquisition types, ethnicity, geometric transformations, and partial occlusions (Lumini and Nanni, 2020). To deal with these issues, we isolate skin from non-skin parts of the detected face using a deep learning approach called Fully-Connected Convolutional Neural Network Continuous Conditional Random Field (FC-CNN CRF). An equivalent term for this is Fully-Convolutional Contin-

[58] AutoML has its own functions to calculate the precision and recall of the model. For our purposes, we use the precision and recall that were calculated on the test data. In other words, the model is run on the test data, and then the results generated by the trained model are compared to the results from the manually labeled test data.

[59] Sometimes "recall" is also referred to as "sensitivity."

uous Conditional Random Field. Our FC-CNN CRF method – by combining three different types of networks (an unary network, a pairwise network, and a continuous CRF network) – takes into account the local and global dependencies between the pixels, and considers the location of the pixels when assigning the skin label, preserving the region integrity. The CRF model parses the face image into semantic regions (e.g, eyes, eyebrows, and mouth) for further processing. This is integrated with an unary network for generating the feature map. The pairwise network is then used to learn the pixel-wise similarity based on neighbor pixels. Thus segmentation accuracy is greatly improved compared to traditional pixel-wise methods which do not take into account semantic regions, boundaries, and the correlations between neighbor pixels.

D.B.3 Image Feature Classification: Classifiying Skin Color Types

We classify the representative skin color for each detected face into one of three categories of skin color type: (1) monochromatic skin colors (e.g., greyscale, sepia), (2) polychromatic human skin colors (e.g., brown, beige), and (3) polychromatic non-typical skin colors (e.g., blue, green).

Monochromatic Classification. In the RGB color space, the closer the R, G, and B values are to each other, the less vibrant the color. For this reason, we classify a face as monochromatic if the standard deviation between the R, G, and B values associated with the weighted average of the face's top k skin colors is less than a threshold T. Thus, a given face i is classified as monochromatic using the following equation:

$$Monochromatic_i = \mathbb{1}\left[\sqrt{\frac{(R_i - \mu_i)^2 + (G_i - \mu_i)^2 + (B_i - \mu_i)^2}{3}} <= T\right] \tag{D1}$$

Where μ_i is equal to the average of the R, G, B values of face i.

Our process of choosing a threshold T proceeded as follows. First, we manually labeled a random sample of 2,836 detected faces (stratified by collection) as either monochromatic or polychromatic. We then calculated the mean squared error between the manual label and our predicted labels using the equation above for every integer value of T between zero and 100. We calculated the average of these mean squared errors using 1,000 bootstrapped samples. The threshold that minimized the mean squared error on average is given by $T = 13$; this provides a classification of images as being monochromatic or not that is 82.9 percent accurate, on average.

Polychromatic Classification. Once we have identified the monochromatic faces, we then separate the remaining faces into two polychromatic color types using the R, G, and

B values associated with the weighted average of a face's top k skin colors: (1) human skin colors and (2) polychromatic non-typical skin colors. This allows us to distinguish between humans and non-human characters who may have colorful skin tones (e.g., aliens, monsters, or characters found in Dr. Seuss books). Specifically, we classify the skin color of the face as a typical human skin color if R $\geq$ G $\geq$ B.[60] Otherwise, it is classified as a polychromatic non-typical skin color.

$$Human_i = [1 - Monochromatic_i] \times \mathbb{1}[R \geq G \geq B] \tag{D2}$$

$$NonTypical_i = [1 - Monochromatic_i] \times [1 - Human_i] \tag{D3}$$

We find this method of classifying the skin color of a face as human or non-typical to be 82.1 percent accurate using our set of 2,836 manually labeled faces.

To classify the darkness or lightness of pictured skin colors, we use the perceptual tint, or L* value, associated with the average of the k colors in L*a*b* space. This value ranges from zero to 100 where a value of zero represents the color black and a value of 100 represents the color white, and there is a range of colors in between.

D.B.4 Image Feature Classification: Race, Gender, and Age

Race Classification (Images). The model assigns the probability that a detected face is of a given race category: Asian, Black, Latinx + Others, or White.[61] Each identified face is assigned the race category which the model gives the highest predicted probability to.[62,63]

Gender Classification (Images). For each face detected, we predict the probability

[60]The boundaries of skin color regions in RGB space from an established pixel-based method of skin classification are defined as $R > 95$ and $G > 40$ and $B > 20$ and $\max\{R,G,B\} - \min\{R,G,B\} > 15$ and $|R - G| > 15$ and $R > G$ and $R > B$ (Vezhnevets, Sazonov and Andreeva, 2003). However, these rules for defining skin color regions are only focused on classifying skin color from photographs. We expand this region in RGB space to account for illustrated skin colors (such as pure white and yellow).

[61]The race labels in the original model were defined in the UTKFace data set and include: Asian, Black, Indian, Others (where "Others" includes Latinx and Middle Eastern) and White. We combine Asian and Indian predictions into a broader Asian category.

[62]Previously, many existing artificial intelligence models that classified putative race had a high error rate, both misclassifying the putative race of identified people and, in "one-shot" models that identify existence of people and their putative race simultaneously, misclassifying people as non-human (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). Much work has been done to acknowledge and address these disparities (Buolamwini and Gebru, 2018; Mitchell et al., 2019)

[63]Classifying race is an imperfect exercise that will yield imperfect algorithms with imperfect categories. Our analysis by race looks across collections within race, so any error within a race would be consistent across collections (i.e., Both the Mainstream and Diversity collections would classify people of the same race similarly.)

that the face is female- (or male-) presenting. We label a face as female if the predicted probability that the face is female-presenting is greater than 50 percent; otherwise, we label the face as male.

We recognize that these classifications are imperfect and focus only on the performative aspect of gender presentation, as they are trained based on how humans classify images. Future work should incorporate the classification of fluid and nonbinary gender identities.

Age Classification (Images). The model assigns the probability that a detected face is of a given age category (infant, child, teenager, adult, senior). We aggregate these categories into two bins: child and adult. We collapse the probabilities for infant and child into a single "child" bin and those for teenager, adult, and senior into a single "adult" bin. A face is classified as that of a child if the probability assigned to the age categories comprising the aggregated child bin is greater than 50 percent, and as that of an adult otherwise.

D.C Text as Data

In this section, we describe the tools we use to measure representation in the text of books. Social scientists have manually analyzed (i.e., by hand) the messages contained in text of printed material for centuries, a process which is highly resource intensive in terms of both labor and time (Neuendorf, 2016; Krippendorff, 2018). Recent work by economists and sociologists showcases how the computational speed and power of (super)computers can be harnessed to conduct automated text analysis, greatly accelerating the speed of work which would have traditionally been done manually (Gentzkow, Kelly and Taddy, 2019; Kozlowski, Taddy and Evans, 2019). We draw from this work and, in particular, a series of natural language processing tools that take bodies of text – e.g., from a book – and extract various features of interest. In Figure 3b, we show our process of extracting text from digitized books and then analyzing it; we refer to this as our "Text-to-Data Pipeline."

The first step in conducting this analysis is to use optical character recognition (OCR) to extract text from digital scans of books. We use the Google Vision Optical Character Recognition (GVOCR) tool for this task. We input the raw files into GVOCR, which identifies and separates the text in each file from the images (e.g., illustrations and photographs). It then applies its own OCR software to the text sections of the scans, converting the text into ASCII which then encodes each character to be recognized by the computer. This generates the text data we analyze.[64]

[64]There are other commonly used OCR interfaces. However, over the past five years, researchers have consistently identified Google Cloud Vision OCR as the best technology for converting images to text. In one study, Tafti et al. (2016) compare the accuracy of Google Docs (now Google Vision), Tesseract, ABBYY FineReader, and Transym OCR methods for over 1,000 images and 15 image categories, and found that

We clean these raw text data to remove erroneous characters and other noise generated by the OCR process, increasing the precision of our measurement of features in the text. The cleaning process removes numerical digits and line breaks but maintains capitalization, punctuation, and special characters. It also standardizes the various permutations of famous names (e.g., all variations of reference to Dr. Martin Luther King Jr. become "Martin Luther King Junior").

From these text data, we then derive several features. These features include: token (single word) counts[65], the presence of famous people, and the first names of characters. In the rest of this section, we describe how we use these features to construct measures of the representation of gender, race, and age in the text of each book.

D.C.1 Text Analysis: Token Counts

One branch of traditional content analysis consists of enumerating words that represent a particular attribute (Krippendorff, 2018). This process generates counts of different "tokens," which comprise a maximal sequence of non-delimiting consecutive characters.[66] In our context, a token is an individual word. We generated a set of tokens associated with identities related to gender, race, or age. The vocabulary used for each of these lists is shown in Section D.C.6 below. We aggregate counts of these words by their respective identity category (such as female or male) by book, generating our "token count" measures of the representation of each identity in each book (Neuendorf, 2016).

Gender (Token Counts). To calculate gender representation in token counts, we calculate the number of words with a gendered meaning. For example, female gendered tokens consist of titles and pronouns such as queen, aunt, girl, she, etc. Similar examples for male gendered tokens include husband, prince, son, himself, etc. [67]

Google Vision generally outperformed other methods. In particular, Google Vision's accuracy with digital images was 4 percent better than any other method. Additionally, the standard deviation of accuracy for Google Vision was quite low, suggesting that the quality of OCR does not drastically change from one image to the next. A test of OCR tools by programmers compared the performance of seven different OCR tools (Han and Hickman, 2019). This analysis also found Google Vision to be superior, specifically when extracting results from low resolution images. In another study that focused on comparing results from multiple image formats (including .jpg, .png, and .tif), Vijayarani and Sakila (2015) found that Google surpassed all other OCR tools. We also tested OCR using ABBYY FineReader and Google Tesseract. Our comparison of their performance relative to manual coding also showed GVOCR performed the best.

[65]This differs slightly from the notion of "lemma" counts, which are measures of the count of word stems. To understand the difference, take the words "father," "fatherly," and "fathered." If each word appeared once in a book, it would generate a token count of one for each word, but a lemma count of three for the lemma "father."

[66]We use the spaCy library to generate these counts, but we see similar patterns in our findings when we use NLTK instead.

[67]Traditional content analysis often restricts gendered words to pronoun counts. We show the sensitivity of our findings related to this construct by restricting the analysis to gendered pronouns only in Appendix

We show how gender representation varies on three additional dimensions: one, whether the gendered identity is represented by individuals (singular) or groups (plural); two, whether the character is placed as the subject or object of a sentence; and three, by the age of the gendered word. To analyze singular and plural representation separately, we separate gendered tokens into those referring to singular cases (e.g., daughter) and plural cases (e.g., daughters). To analyze whether the character is the subject or object of a sentence, we generate counts of the number of gendered pronouns that are capitalized versus lowercase, under the assumption that an individual who is the subject of a sentence is in a position of more active importance than the same character when used as the object and thus occupying a more passive role. To analyze representation of gender by age, we generate a list of "younger" gendered words (e.g., princess, boy) and "older" gendered words (e.g., queen, man).

Color (Token Counts). As another proxy for the analysis of race, we calculate the proportion of all words that refer to colors. For parsimony, in the paper we only show the words black, blacks, white, whites, blue, and blues.

Nationality (Token Counts). To calculate nationality representation in token counts, we calculate the proportion of all words that refer to nationalities. Note that we converted all multi-token nationalities to a single token analog.

D.C.2 Text Analysis: Named Entity Recognition

We also measure the representation of gender and race among named characters in these stories, be they fictional or historical. A series of studies show that exposure to salient examples of historical figures or celebrities from historically marginalized identities can lead to meaningful change in social attitudes towards people who hold that identity, and potentially are associated with changes in beliefs and academic performance among children who share that identity (Marx, Ko and Friedman, 2009; Plant et al., 2009; Alrababah et al., 2021). To identify characters in our text, we use a tool called Named Entity Recognition (NER).[68] NER identifies and segments "named entities," or proper nouns, starting with a pre-defined library of such entities and also identifying new entities through the application of neural nets. NER recognizes these entities in strings of text; applying NER to our data, we identify these entities and count how many times each specific named entity is mentioned

Figure B11. Our results are robust to this alternate specification.

[68]We run our NER analysis using the open-source software library spaCy, which employs convolutional neural networks for both text categorization and NER. Another commonly-used library for NER is NLTK, but it only recognizes single words for NER, whereas spaCy can recognize strings of words as a distinct entity. For example, "Martin Luther King" would be recognized as one entity in SpaCy but as three entities with NLTK ("Martin," "Luther," and "King").

in a given book. We then associate these frequency counts with identifiable traits of the people identified by NER, such as their race, gender, or place of birth. There are two types of named entities that we identify: (1) famous characters and (2) first names of characters.

D.C.3 Text Analysis: Famous Characters

To identify the instances of famous characters represented in books, such as Martin Luther King Junior or Amelia Earhart, we match all of the entities identified by NER that have at least two names (for example, a first and last name) with a pre-existing data set, Pantheon 2.0, that contains data from over 70,000 Wikipedia biographies which have a presence in more than 15 language editions of Wikipedia (Yu et al., 2016). This generates a data set of 2,697 famous people. To examine the race of the famous figures mentioned in the text, we count the number of famous people mentioned at least once in a book and sum over all books in a collection. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection and Jimmy Carter is uniquely mentioned in 2 books within the same collection, then 60 percent of the unique famous people mentioned in that collection would be Black. We count the number of unique books in which each famous person is mentioned as well as the number of times they are mentioned in each book.

Gender and Birthplace (Famous People). The Pantheon 2.0 data set contains information on the gender and birthplace of these famous people. We match these data to each famous figure identified from the NER in our data.[69]

Race (Famous People). We then manually code race for each identified person. This was conducted based on a manual internet search for each person, starting with Wikipedia.[70] We collapse the following identities: East Asian, Middle Eastern, and South Asian into the Asian category; North American Indigenous peoples and South American Indigenous peoples into the Indigenous category; and African American and Black African into the Black category. If an individual was coded as having more than one race, they were then classified as Multiracial.

D.C.4 Text Analysis: Character First Names

We also study the representation of gender among people who are named but not identified as "famous" using the methods described above. Using the named entities identified by the spaCy NER engine, we limit the sample to those entities categorized as a person and

[69]The Pantheon 2.0 curators run a classifier over the English text of the Wikipedia biographies to extract information such as place of birth and gender from each biography. Their classifier was trained on a data set called Pantheon 1.0 (Yu et al., 2016) which contains a subset of manually curated biographies.

[70]Note that coding of putative race is subject to the individual biases and perceptions of each human coder and may be classified with error.

remove the famous characters we found by applying the process described in Section D.C.3.[71] We then categorize the remaining named entities and construct a data set containing the name of each unique character and the number of times that character is mentioned in a given book.

Gender (Character First Names). To identify the gender of characters not identified as famous, we extract the first name of each remaining named entity and estimate the probability that the character is female using data on the frequency of names by gender in the U.S. population from the Social Security Administration. For example, if a character's first name is "Cameron," our estimated probability that the character is female is 9.16 percent because that is the proportion of people named "Cameron" in relevant Social Security data who are female. Our sample of "relevant" Social Security data include only data from years which overlap with the years in our sample of children's data.

If the predicted probability that a character is female is greater than 50 percent, we label that character as female. Otherwise, the character is labeled as male.[72] Using this method, we are able to make gender predictions for approximately 60,000 characters. To test how accurate these predictions are, we predicted the gender of each famous person in our data using their first names and compared these predictions to their gender identified using Wikipedia and found that our predictions were 96.35 percent accurate.[73]

We are not able to make a prediction for the remaining named entities. For example, characters such as "New Yorker" which the spaCy NER engine identified and labeled as a person will not receive a prediction because "New" does not appear as a first name in Social Security data.

D.C.5 Text Analysis: All Gendered Words

We aggregate all gendered mentions (gendered tokens (e.g., titles, pronouns, specific gender terms such as queen and husband), predicted gender of character first names, and gender of famous characters) to generate a composite measure of gender representation in text. We refer to this aggregate measure as "gendered words," or "words with a gender association."

[71] NER tags each entity with a different category: people, locations, currency, and more. This entity categorization (e.g., person, location) is not always correct, so there may be entities misclassified or missed overall. We do not use this categorization when identifying famous characters.

[72] We predict gender with the *gender* package available in R which uses Social Security Administration data (Mullen, 2020).

[73] We do not classify race using first names only. Other recent text analysis has shown that conventional methods for classifying race of names fail to accurately distinguish between Black people and White people (Garg et al., 2018).

The total number of female gendered words in book i is calculated as follows:

$$\begin{aligned}(\text{female words})_i = &(\text{total number of female-specific tokens})_i \\ &+ (\text{total number of mentions of famous female characters})_i \\ &+ (\text{total number of characters with female first names})_i\end{aligned}$$

D.C.6 Vocab Lists Used in Token Counts

The vocab lists containing all the words we use in our token counts are listed below. These lists may not be comprehensive.

Gendered Tokens. The gendered tokens we enumerate are as follows. Subset lists are used for the specific gendered token counts, gendered pronouns, singular/plural gendered token counts, younger/older gendered token counts and uppercase/lowercase pronouns.

Female. abuela, abuelita, actress, aunt, auntie, aunties, aunts, aunty, czarina, damsel, damsels, daughter, daughters, emperess, emperesses, empress, empresses, fairies, fairy, female, females, girl, girls, grandma, grandmas, grandmom, grandmother, grandmothers, her, hers, herself, housekeeper, housekeepers, ladies, lady, ma'am, madame, mademoiselle, mademoiselles, maid, maiden, maidens, maids, mama, mamas, mermaid, mermaids, miss, mlle, mme, mom, mommies, mommy, moms, mother, mothers, mrs, ms, nana, nanas, princess, princesses, queen, queens, she, sissie, sissy, sister, sisters, stepmother, stepmothers, titi, tsarevna, tsarina, tsaritsa, tzaritza, waitress, wife, witch, witches, wives, woman, women

Plural Female. aunties, aunts, damsels, daughters, emperesses, empresses, fairies, females, girls, grandmas, grandmothers, housekeepers, ladies, mademoiselles, maidens, maids, mamas, mermaids, mommies, moms, mothers, nanas, queens, sisters, stepmothers, witches, wives, women

Singular Female. abuela, abuelita, aunt, auntie, aunty, czarina, damsel, daughter, emperess, empress, fairy, female, girl, grandma, grandmom, grandmother, her, hers, herself, housekeeper, lady, maam, madame, mademoiselle, maid, maiden, mama, mermaid, miss, mlle, mme, mom, mommy, mother, mrs, ms, nana, princess, queen, she, sissie, sissy, sister, stepmother, titi, tsarevna, tsarina, tsaritsa, tzaritza, wife, witch, woman

Young Female. damsel, damsels, daughter, daughters, fairies, fairy, girl, girls, mademoiselle, mademoiselles, maiden, maidens, miss, princess, princesses, tsarevna

Old Female. abuela, abuelita, aunt, auntie, Auntie, aunts, aunty, czarina, emperess, emperesses, empress, empresses, grandma, grandmas, grandmom, grandmother, grandmoth-

ers, housekeeper, housekeepers, maam, madame, mama, mamas, mlle, mme, mom, mommies, mommy, moms, mother, mothers, mrs, nana, nanas, queen, queens, stepmother, stepmothers, titi, tsarina, tsaritsa, tzaritza, wife, witch, witches, wives, woman, women

Male. abuelito, abuelo, actor, boy, boys, bro, brother, brothers, butler, butlers, chap, chaps, czar, dad, daddies, daddy, dads, einstein, emperor, emperors, father, fathers, fellow, fellows, gentleman, gentlemen, granddad, granddads, grandfather, grandfathers, grandpa, grandpas, he, him, himself, his, hisself, husband, husbands, king, kings, knight, lad, lads, lord, lords, male, males, man, master, masters, men, merman, mermen, mr, paige, paiges, papa, papas, prince, princes, sir, sirs, son, sons, squire, squires, stepfather, stepfathers, tio, tsar, uncle, uncles, waiter, wizard, wizards

Plural Male. boys, brothers, butlers, chaps, daddies, dads, emperors, fathers, fellows, gentlemen, granddads, grandfathers, grandpas, husbands, kings, knights, lads, lords, males, masters, men, mermen, paiges, papas, princes, sirs, sons, squires, stepfathers, uncles, wizards

Singular Male. abuelito, abuelo, boy, bro, brother, butler, chap, czar, dad, daddy, emperor, father, fellow, gentleman, granddad, grandfather, grandpa, he, him, himself, his, hisself, husband, king, knight, lad, lord, male, man, master, merman, mr, paige, papa, prince, sir, son, stepfather, tio, tsar, uncle, wizard

Young Male. boy, boys, lad, lads, prince, princes, son, sons

Old Male. abuelito, abuelo, butler, butlers, czar, dad, daddies, daddy, dads, emperor, emperors, father, fathers, gentleman, gentlemen, granddad, granddads, grandfather, grandfathers, grandpa, grandpas, husband, husbands, king, kings, lord, lords, man, men, mr, papa, papas, sir, sirs, stepfather, stepfathers, tio, tsar, uncle, uncles, wizard, wizards

Racial Proxy Tokens. The tokens we use as proxies for race are as follows.

Colors. The color word tokens used as proxies for race and falsification words are the following: black, blue, brown, gold, golden, green, orange, pink, purple, red, silver, violet, white, yellow. For parsimony, in the paper we only show the words black, blacks, white, whites, blue, and blues.

Nationalities. afghan, african, albanian, algerian, american, andorran, angolan, antiguans, apache, argentinean, armenian, asian, australian, austrian, azerbaijani, bahamian, bahraini, bangladeshi, barbadian, barbudans, batswana, belarusian, belgian, belizean, beninese, bhutanese, bolivian, bosnian, brazilian, british, bruneian, bulgarian, burkinabe, burmese, burundian, cambodian, cameroonian, canadian, cape verdean, chadian, cherokee, chicana,

chicano, chicanx, chilean, chinese, choctaw, colombian, comoran, congolese, croatian, cuban, cypriot, czech, danish, djibouti, dominican, dutch, dutchman, dutchwoman, ecuadorean, egyptian, emirian, english, eritrean, estonian, ethiopian, fijian, filipino, finnish, french, gabonese, gambian, georgian, german, ghanaian, greek, grenadian, guatemalan, guinean, guinean, guyanese, haitian, herzegovinian, hispanic, honduran, hungarian, icelander, i-kiribati, indian, indonesian, iranian, iraqi, irish, irish, iroquois, israeli, italian, ivorian, jamaican, japanese, jordanian, kazakhstani, kenyan, kittian, korean, kuwaiti, kyrgyz, laotian, latina, latino, latinx, latvian, lebanese, leonean, liberian, libyan, liechtensteiner, lithuanian, lucian, luxembourger, macedonian, malagasy, malawian, malaysian, maldivan, malian, maltese, marinese, marshallese, mauritanian, mauritian, mexican, micronesian, moldovan, monacan, mongolian, mongols, moroccan, mosotho, motswana, mozambican, namibian, nauruan, navajo, nepalese, netherlander, nevisian, nicaraguan, nigerian, nigerien, ni-vanuatu, norwegian, ojibwe, omani, pakistani, palauan, panamanian, paraguayan, persian, peruvian, polish, portuguese, qatari, rican, romanian, russian, rwandan, salvadoran, samoan, saudi, scottish, senegalese, serbian, seychellois, singaporean, sioux, slovakian, slovenian, somali, spanish, sri-lankan, sudanese, surinamer, swazi, swedish, swiss, syrian, taiwanese, tajik, tanzanian, thai, timorese, tobagonian, togolese, tomean, tongan, trinidadian, tunisian, turkish, tuvaluan, ugandan, ukrainian, uruguayan, uzbekistani, venezuelan, vietnamese, welsh, yemenite, zambian, zealander, zimbabwean.

E Seattle Public Library Checkouts Data

To study the impact of being honored by the children's book awards we examine, we analyze data from the Seattle Public Library system on all public checkouts from the library between April 2005 and September 2017.[74] Awards are given near the end of January each year to books published in that year or the year before. We analyze checkout data for the award-winning books in our data, alongside all books belonging to the children's and junior book collections published in the year prior to the award, covering award years 2005 to 2017.

We collapse these to a data set of collection-by-day checkout likelihoods scaled by the number of books in the collection to generate a measure of the number of checkouts per book, per day, in each of the three collections. We limit checkout data for each book to the calendar year before the award was given and the two following calendar years.

To generate Figure 2, we re-center the checkout date according to its distance from the date in which the award is given for books published in that year. For example, books published in 2011 would be eligible for an award in 2012. Checkouts from before January

[74]These data are publicly available at https://data.seattle.gov/Community/Checkouts-by-Title/tmmm-ytt6; site accessed on April 15, 2021.

20th, 2012 (The first date of the ALA Midwinter Meeting in 2012) would be given negative values – for example, checkouts on January 10th, 2012, would be –10 days from January 20th, 2012. Checkouts after that date have positive values. Figure 2 shows the results of applying a 14-day moving average to each series of average collection-specific number of checkouts per day (divided by the number of books in that collection to account for the fact that the number of books per collection varies across the Mainstream, Diversity, and all other children's books) over the window of days to award spanning [–400 days, 730 days].

We quantify the post-award increase using a simple event study design. While not causal per se, this allows us to estimate more precisely how much more likely books in each collection are to be checked out after receipt of an award or honor, relative to the rest of the sample. To do so, we use the following equation:

$$checkouts_{cd} = \beta_1 Post + \beta_2 Post * Mainstream + \beta_3 Post * Diversity + \eta_c + \varepsilon_{cd}$$

The dependent variable is the number of checkouts, per book, in collection c on day d. We regress this on the following variables: whether the day is after January 20th ($Post$) (a noisy estimate of the date when the awards are announced each year); a set of fixed effects for each collection; and an interaction of the $Post$ variable with the $Mainstream$ and $Diversity$ collection variables. Our main coefficients of interest are β_2 and β_3.

We present our results in Table E1. This shows that after winning or being honored by an award, Mainstream books are approximately four times as likely as non-recognized children's books in the library to be checked out on any given day. We derive this from calculating the ratio of the post-award checkout rate for the Mainstream collection to that of the non-recognized books. For the Mainstream collection, this is the sum of the $Mainstream$ fixed effect, the constant (the $Diversity$ fixed effect) the coefficient on the "post-award" variable ($Post$), and the coefficient on the interaction term between $Post$ and the $Mainstream$ collection, which sums to approximately 0.483. The post-award checkout rate for non-recognized children's books in the library is the sum of the non-recognized children's books in the library fixed effect, the constant, and the coefficient on $Post$, which sums to approximately 0.120.

An alternate interpretation is that after winning the award, the Mainstream collection books are approximately 2.6 times more likely to be checked out than they were before. This is derived by dividing the sum of coefficients on $Post$, the interaction of $Mainstream$ and $Post$, the constant, and the $Mainstream$ fixed effect, by the $Mainstream$ fixed effect. We note that these should be interpreted as suggestive estimates; we define "pre-" and "post-" award using January 20th, an estimate of when news of the award announcements is likely

Table E1. Estimates of the Increase in Daily Checkouts After Receipt of Mainstream and Diversity Awards

Parameter	**Estimate**
Non-Recognized Children's Books in Library Fixed Effect	0.019***
	(0.004)
Mainstream Collection Fixed Effect	0.107***
	(0.006)
Diversity Collection Fixed Effect (constant)	0.075***
	(0.004)
Post	0.026***
	(0.005)
Post × Mainstream Collection	0.274***
	(0.007)
Post × Diversity Collection	0.017**
	(0.007)
Observations	3,375
Adjusted R^2	0.773

Notes: These parameters were generated using the equation given in this subsection of the Data Appendix estimated using data from the Seattle Public Library on daily checkouts. *p<0.1; **p<0.05; ***p<0.01

to reach readers, parents, and librarians. Its precise date varies from year to year.

For the Diversity awards, we see a slight change in checkout behavior after January 20th. This can be seen in our estimate of the interaction term between *Diversity* and *Post*, which is statistically significant, but small in magnitude - especially when compared to the coefficient on the interaction term between *Mainstream* and *Post*. Seen through the lens of the calculations above, after receiving an award, Diversity collection books are more than 1.7 percent *less* likely to be checked out than non-winners; this can be derived analogously, comparing the post-award checkout rate for the Diversity collection – the sum of the *Diversity* fixed effect, the coefficient on *Post*, and the coefficient on the interaction term between *Post* and the *Diversity* collection, which sums to approximately 0.118. The post-award checkout rate for non-winners is the sum of the *Non-winners* fixed effect and the coefficient on *Post*, which is approximately 0.120. Prior to receipt of the award, they were approximately 20 percent less likely to be checked out.

In Table E2, we present an alternative specification where we estimate a similar equation, only with separate parameters for award winners and honorees. This shows broadly similar results, with one exception: winning a mainstream award yields a premium that is 2.5 times as large as merely being an honoree. This is similar to the visual patterns we see in Figure 2 and, more specifically, the distinct post-award increases in checkouts we observe for winners and awardees, respectively.

Table E2. Estimates of the Increase in Daily Checkouts After Receipt of Mainstream and Diversity Awards and Honors

Parameter	**Estimate**
Non-Recognized Children's Books in Library Fixed Effect	0.026***
	(0.006)
Mainstream Winner Fixed Effect	0.117***
	(0.008)
Mainstream Honoree Fixed Effect	0.101***
	(0.008)
Diversity Winner Fixed Effect	0.001
	(0.008)
Diversity Honoree Fixed Effect (Constant)	0.064***
	(0.008)
Post	0.031***
	(0.007)
Post × Mainstream Winner	0.500***
	(0.010)
Post × Mainstream Honoree	0.205***
	(0.010)
Post × Diversity Winner	0.014**
	(0.010)
Post × Diversity Honoree	0.006**
	(0.010)
Observations	5,620
Adjusted R^2	0.775

Notes: These parameters were generated using the equation given in this subsection of the Data Appendix estimated using data from the Seattle Public Library on daily checkouts. This table is similar to Table E1, except that it separates award premia by whether books were named honorees for a given award, or recipients of the award itself. *p<0.1; **p<0.05; ***p<0.01

F Award Criteria

In this section we give the criteria for award selection for the Newbery and Caldecott awards and provides links to the criteria for the other awards.

F.A Caldecott Medal Criteria

Terms and criteria are listed below.[75] Note that the numbering and itemization follows the formatting as presented on the website and were not altered for consistency.

F.A.1 Terms

The Medal shall be awarded annually to the artist of the most distinguished American picture book for children published by an American publisher in the United States in English during the preceding year. There are no limitations as to the character of the picture book except that the illustrations be original work. Honor books may be named. These shall be books that are also truly distinguished.

The award is restricted to artists who are citizens or residents of the United States. Books published in a U.S. territory or U.S. commonwealth are eligible.

The committee in its deliberations is to consider only books eligible for the award, as specified in the terms.

F.A.2 Definitions

A "picture book for children" as distinguished from other books with illustrations, is one that essentially provides the child with a visual experience. A picture book has a collective unity of story-line, theme, or concept, developed through the series of pictures of which the book is comprised.

A "picture book for children" is one for which children are an intended potential audience. The book displays respect for children's understandings, abilities, and appreciations. Children are defined as persons of ages up to and including fourteen and picture books for this entire age range are to be considered.

"Distinguished" is defined as:

- Marked by eminence and distinction; noted for significant achievement.
- Marked by excellence in quality.
- Marked by conspicuous excellence or eminence.

[75] Terms and criteria downloaded exactly from https://www.ala.org/alsc/awardsgrants/bookmedia/caldecott on July 14, 2022.

- Individually distinct.
- The artist is the illustrator or co-illustrators. The artist may be awarded the medal posthumously.

The term "original work" may have several meanings. For purposes of these awards, it is defined as follows: "Original work" means that the illustrations were created by this artist and no one else. Further, "original work" means that the illustrations are presented here for the first time and have not been previously published elsewhere in this or any other form. Illustrations reprinted or compiled from other sources are not eligible.

"American picture book in the United States" means that books first published in previous years in other countries are not eligible. Books published simultaneously in the U.S. and another country may be eligible. Books published in a U.S. territory or U.S. commonwealth are eligible.

"In English" means that the committee considers only books written and published in English. This requirement DOES NOT limit the use of words or phrases in another language where appropriate in context.

"Published...in the preceding year" means that the book has a publication date in that year, was available for purchase in that year, and has a copyright date no later than that year. A book might have a copyright date prior to the year under consideration but, for various reasons, was not published until the year under consideration. If a book is published prior to its year of copyright as stated in the book, it shall be considered in its year of copyright as stated in the book. The intent of the definition is that every book be eligible for consideration, but that no book be considered in more than one year.

"Resident" specifies that author has established and maintains a residence in the United States, U.S. territory, or U.S. commonwealth as distinct from being a casual or occasional visitor.

The term, "only the books eligible for the award," specifies that the committee is not to consider the entire body of the work by an artist or whether the artist has previously won the award. The committee's decision is to be made following deliberation about books of the specified calendar year.

F.A.3 Criteria

In identifying a "distinguished American picture book for children," defined as illustration, committee members need to consider:

- Excellence of execution in the artistic technique employed;
- Excellence of pictorial interpretation of story, theme, or concept;
- Appropriateness of style of illustration to the story, theme or concept;
- Delineation of plot, theme, characters, setting, mood or information through the pictures;
- Excellence of presentation in recognition of a child audience.

The only limitation to graphic form is that the form must be one which may be used in a picture book. The book must be a self-contained entity, not dependent on other media (i.e., sound, film or computer program) for its enjoyment.

Each book is to be considered as a picture book. The committee is to make its decision primarily on the illustration, but other components of a book are to be considered especially when they make a book less effective as a children's picture book. Such other components might include the written text, the overall design of the book, etc.

Note: The committee should keep in mind that the award is for distinguished illustrations in a picture book and for excellence of pictorial presentation for children. The award is not for didactic intent or for popularity.

[Adopted by the ALSC board, January 1978. Revised, Midwinter 1987. Revised, Annual 2008.]

F.B Newbery Medal Criteria

Terms and criteria are listed below.[76] Note that the numbering and itemization follows the formatting as presented on the website and were not altered for consistency.

F.B.1 Terms

1. The Medal shall be awarded annually to the author of the most distinguished contribution to American literature for children published by an American publisher in the United States in English during the preceding year. There are no limitations as to the character of the book considered except that it be original work. Honor books may be named. These shall be books that are also truly distinguished.

2. The Award is restricted to authors who are citizens or residents of the United States.

[76] Terms and criteria downloaded exactly from https://www.ala.org/alsc/awardsgrants/bookmedia/newbery on July 14, 2022.

3. The committee in its deliberations is to consider only the books eligible for the award, as specified in the terms.

F.B.2 Definitions

1. "Contribution to American literature" indicates the text of a book. It also implies that the committee shall consider all forms of writing—fiction, non-fiction, and poetry. Reprints, compilations and abridgements are not eligible.

2. A "contribution to American literature for children" shall be a book for which children are an intended potential audience. The book displays respect for children's understandings, abilities, and appreciations. Children are defined as persons of ages up to and including fourteen, and books for this entire age range are to be considered.

3. "Distinguished" is defined as:

   - Marked by eminence and distinction; noted for significant achievement.
   - Marked by excellence in quality.
   - Marked by conspicuous excellence or eminence.
   - Individually distinct.

4. "Author" may include co-authors. The author(s) may be awarded the medal posthumously.

5. The term "original work" may have several meanings. For purposes of these awards, it is defined as follows:

   - "Original work" means that the text was created by this writer and no one else. It may include original retellings of traditional literature, provided the words are the author's own.
   - Further, "original work" means that the text is presented here for the first time and has not been previously published elsewhere in this or any other form. Text reprinted or compiled from other sources are not eligible. Abridgements are not eligible.

6. "In English" means that the committee considers only books written and published in English. This requirement DOES NOT limit the use of words or phrases in another language where appropriate in context.

7. "American literature published in the United States" means that books first published in previous years in other countries are not eligible. Books published simultaneously

in the U.S. and another country may be eligible. Books published in a U.S. territory, or U.S. commonwealth are eligible.

8. "Published. . . in the preceding year" means that the book has a publication date in that year, was available for purchase in that year, and has a copyright date no later than that year. A book might have a copyright date prior to the year under consideration but, for various reasons, was not published until the year under consideration. If a book is published prior to its year of copyright as stated in the book, it shall be considered in its year of copyright as stated in the book. The intent of the definition is that every book be eligible for consideration, but that no book be considered in more than one year.

9. "Resident" specifies that the author has established and maintains a residence in the United States, U.S. territory, or U.S. commonwealth as distinct from being a casual or occasional visitor.

10. The term, "only the books eligible for the award," specifies that the committee is not to consider the entire body of the work by an author or whether the author has previously won the award. The committee's decision is to be made following deliberation about the books of the specified calendar year.

F.B.3 Criteria

1. In identifying "distinguished contribution to American literature," defined as text, in a book for children,

   (a) Committee members need to consider the following:

   - Interpretation of the theme or concept
   - Presentation of information including accuracy, clarity, and organization
   - Development of a plot
   - Delineation of characters
   - Delineation of a setting
   - Appropriateness of style.

   Note: Because the literary qualities to be considered will vary depending on content, the committee need not expect to find excellence in each of the named elements. The book should, however, have distinguished qualities in all of the elements pertinent to it.

(b) Committee members must consider excellence of presentation for a child audience.

2. Each book is to be considered as a contribution to American literature. The committee is to make its decision primarily on the text. Other components of a book, such as illustrations, overall design of the book, etc., may be considered when they make the book less effective.

3. The book must be a self-contained entity, not dependent on other media (i.e., sound or film equipment) for its enjoyment.

   Note: The committee should keep in mind that the award is for literary quality and quality presentation for children. The award is not for didactic content or popularity.

Adopted by the ALSC Board, January 1978. Revised, Midwinter 1987. Revised, Annual 2008.

F.C Award Information for Diversity Collection

In this section, we provide the website describing each award and its selection criteria, accessed on July 15, 2022. Selection criteria vary by award. At a high level, they share two main goals. One is to recognize excellence in the content of the book. This goal, and the text of the various award criteria given in the links below, tracks closely with the main goals of the Caldecott and Newbery awards. The second goal is to recognize books who portray, recognize, or elevate a specific identity group, for example, people with disabilities or Hispanic Americans. These goals vary widely by award, as each award focuses on a specific identity.

- American Indian Youth Literature Award
  Site: ailanet.org/activities/american-indian-youth-literature-award
- Américas Award
  Site: claspprograms.org/pages/detail/65/About-the-Award
- Name: Arab American Book Award
  Site: arabamericanmuseum.org/book-awards/
- Asian/Pacific American Award for Literature
  Site: apalaweb.org/awards/literature-awards/literature-award-guidelines/
- Carter G. Woodson Book Awards
  Site: woodsonawards.weebly.com/
- Coretta Scott King Book Award

Site: ala.org/rt/emiert/cskbookawards/slction

- Dolly Gray Children's Literature Award
  Site: dollygrayaward.com/
- Ezra Jack Keats Award
  Site: degrummond.org/ezra-jack-keats-book-award-guidelin
- Middle East Book Award
  Site: meoc.us/book-awards.html
- Notable Books for a Global Society
  Site: clrsig.org/nbgs.html
- Pura Belpré Award
  Site: ala.org/alsc/awardsgrants/bookmedia/belpre
- Rise: A Feminist Book Project
  Site: risefeministbooks.wordpress.com/criteria/
- Schneider Family Book Award
  Site: ala.org/awardsgrants/awards/1/apply
- Skipping Stones Youth Honor Awards
  Site: skippingstones.org/wp/youth-honors-award/
- South Asia Book Award
  Site: southasiabookaward.wisc.edu/submission-guidelines/
- Stonewall Book Awards
  Site: ala.org/awardsgrants/awards/177/apply
- Tomás Rivera Mexican American Awards
  Site: education.txstate.edu/ci/riverabookaward/about.html