

# Appendices

## A Appendix Tables

Table A1. Summary Statistics for Children's Book Purchasers in Numerator Data

<i>Purchaser Demographics</i>	<i>All Children's Books</i>		<i>Award-Winning Children's Books</i>	
	N	Mean	N	Mean
<b>Children</b>				
Has Children	1,547,044	0.73	62,283	0.70
Has Children Ages 0-5	1,188,039	0.23	47,782	0.14
Has Children Ages 6-12	1,188,039	0.46	47,782	0.38
Has Children Ages 13-17	1,188,039	0.23	47,782	0.35
<b>Race/Ethnicity</b>				
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
<b>Gender</b>				
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
<b>Sexuality</b>				
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
<b>Income</b>				
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
<b>Education</b>				
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 shows the sample size and mean of purchaser demographics for children's book purchases in Numerator OmniPanel data from 2017-2020. The first two columns include all children's book purchases. The last two columns include all purchases of a children's book which was recognized by one of the awards in our sample.

Table A2. 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 A3. 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 A4. 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 A5. Top Mentioned Famous Person, by Collection and Decade

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

*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.

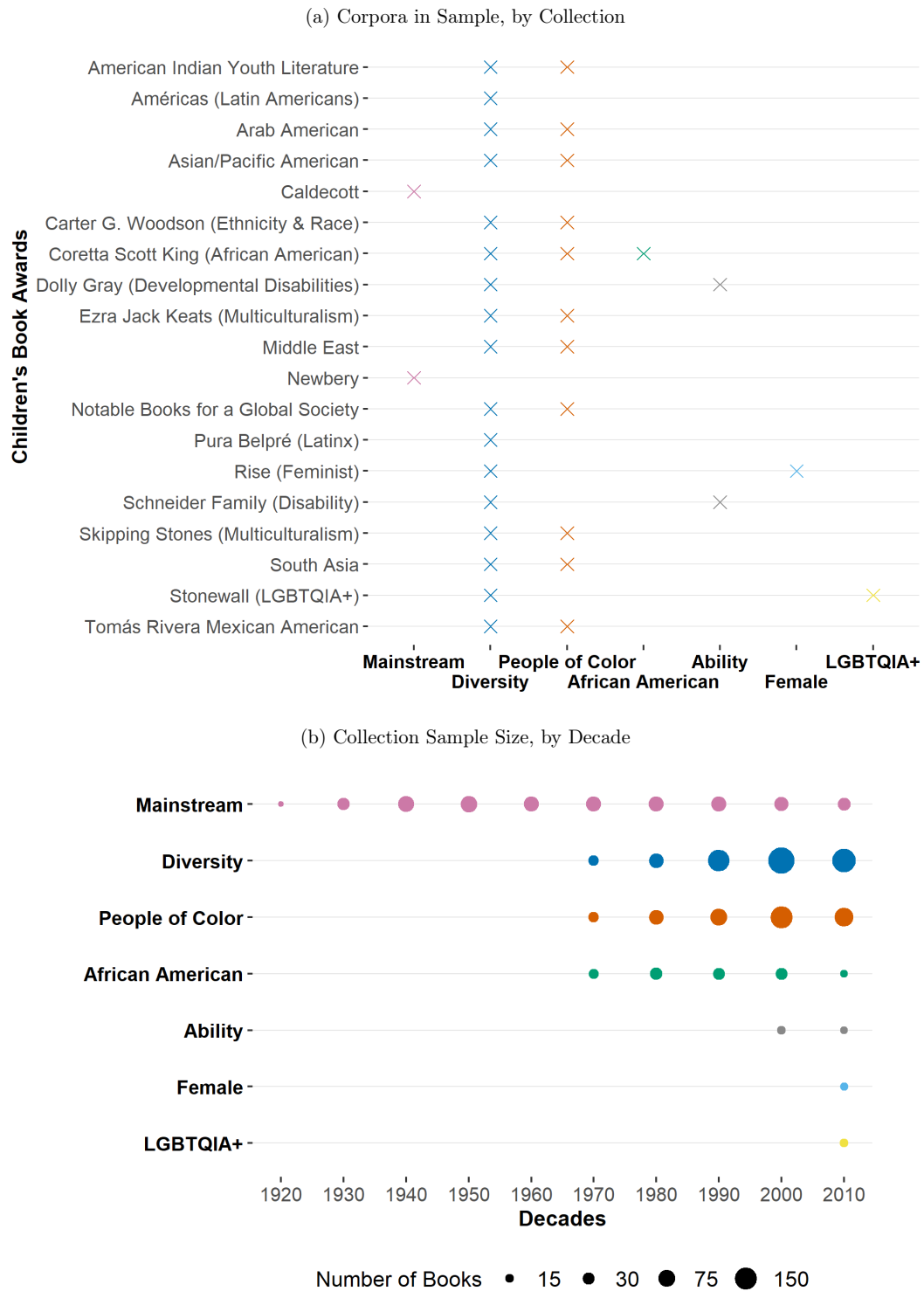
Table A6. Correlation between U.S. Demographics and Representation

	<i>Dependent variable: Percent of</i>			
	Faces	Famous People	Female Words	Images vs. Text
	(1)	(2)	(3)	(4)
<i>Panel A: Percent of Labor Force Participation</i>				
Females	-0.08 (0.15)	0.69** (0.29)	0.45*** (0.15)	-0.36 (0.22)
<i>Panel B: Percent of Population</i>				
Asian	1.49** (0.5)	-0.89 (0.73)		
Black	1.49*** (0.25)	5.15*** (1.36)		
Latinx	-0.53 (0.21)	0.03 (0.05)		
White	0.40** (0.16)	0.33 (0.24)		

*Note:* This table estimates the relationship between major demographic parameters (U.S. female labor force participation in Panel A and the racial composition of the U.S. population in Panel B) and representation in the images and text of children’s books from our Mainstream collection. We regress a measure of market share or market power – either population share of a given racial group or female labor force participation – for a given race or gender on a measure of their proportional representation in award winning children’s books over time. We show each coefficient from these bivariate regressions in this table, with standard errors in parentheses. For example, the first row and column shows the coefficient from a regression of the percentage of female labor force participation on the percentage of female faces in the Mainstream collection over time. Our data on female labor force participation is constructed by taking the yearly average over monthly unadjusted data between 1948-2019 from the U.S. Bureau of Labor Statistics and retrieved from FRED, Federal Reserve Bank of St. Louis. Our data on population breakdown by race is from 1920-2019 U.S. census data. Census information on the proportion of people who are Latinx comes from a response to a question regarding ethnicity and is not mutually exclusive to the other race categories. We construct each race/ethnicity category to be mutually exclusive; for example, we count an individual who identifies as Latinx and White in the Latinx category, not the White category. Census data on ethnicity are only available beginning in 1970.

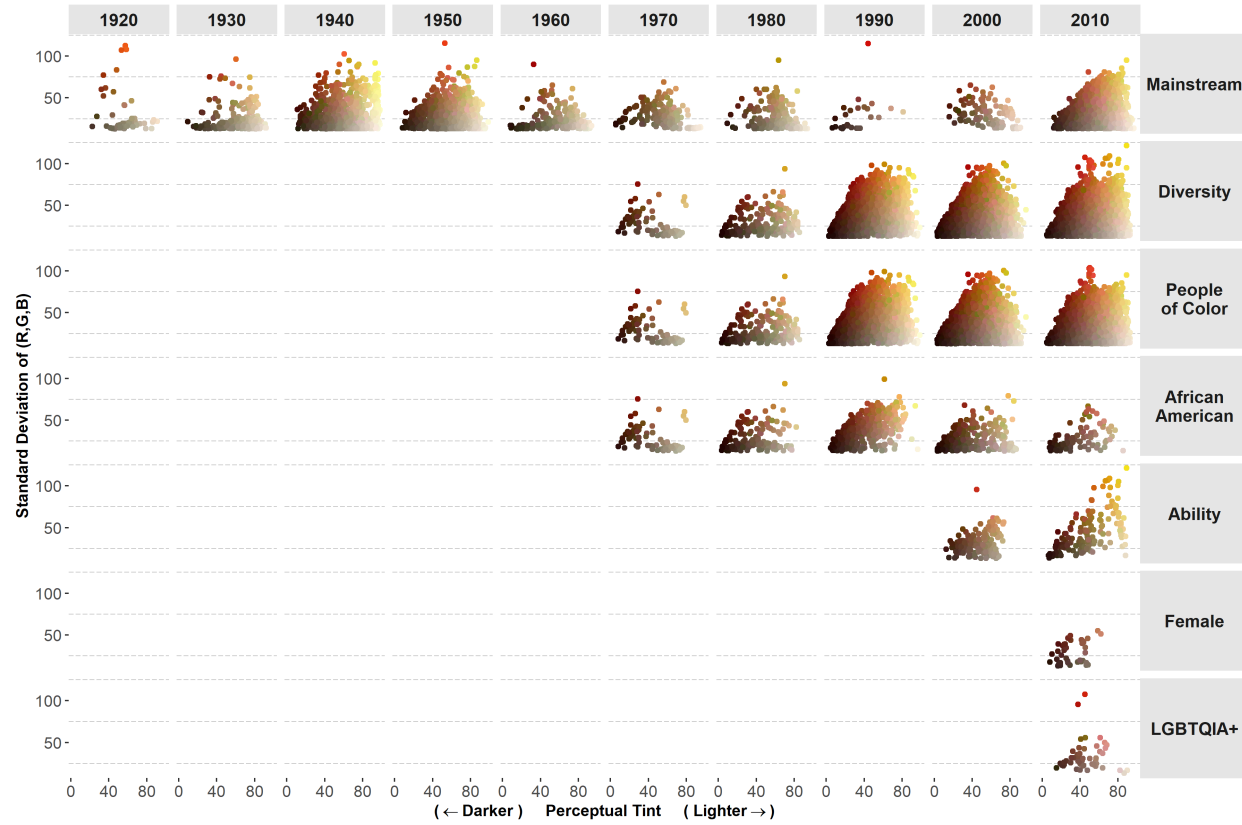
## B Appendix Figures

Figure B1. Books in the Sample



*Note:* This figure shows the main sources of data we use for our analysis. In Panel A, we list the book awards in our sample, along with the collections into which we group them in our analysis. In Panel B, we show our sample size in each collection, over time.

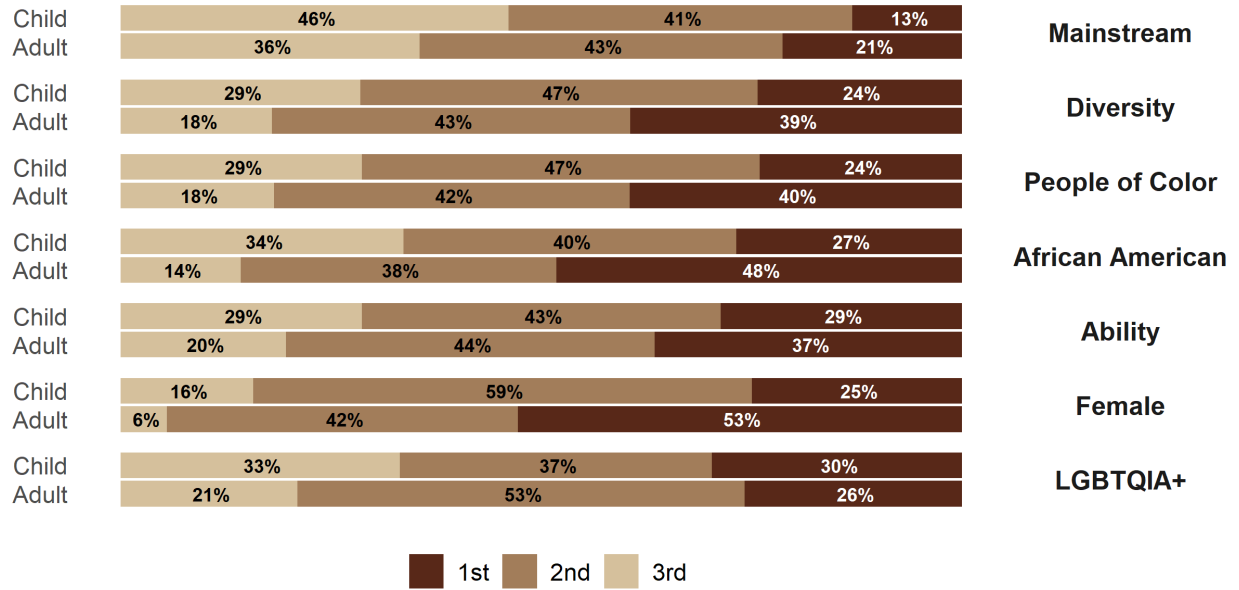
Figure B2. Skin Color Data Over Time, Human Skin Colors



*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 IV.A, 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 IV.A.2.

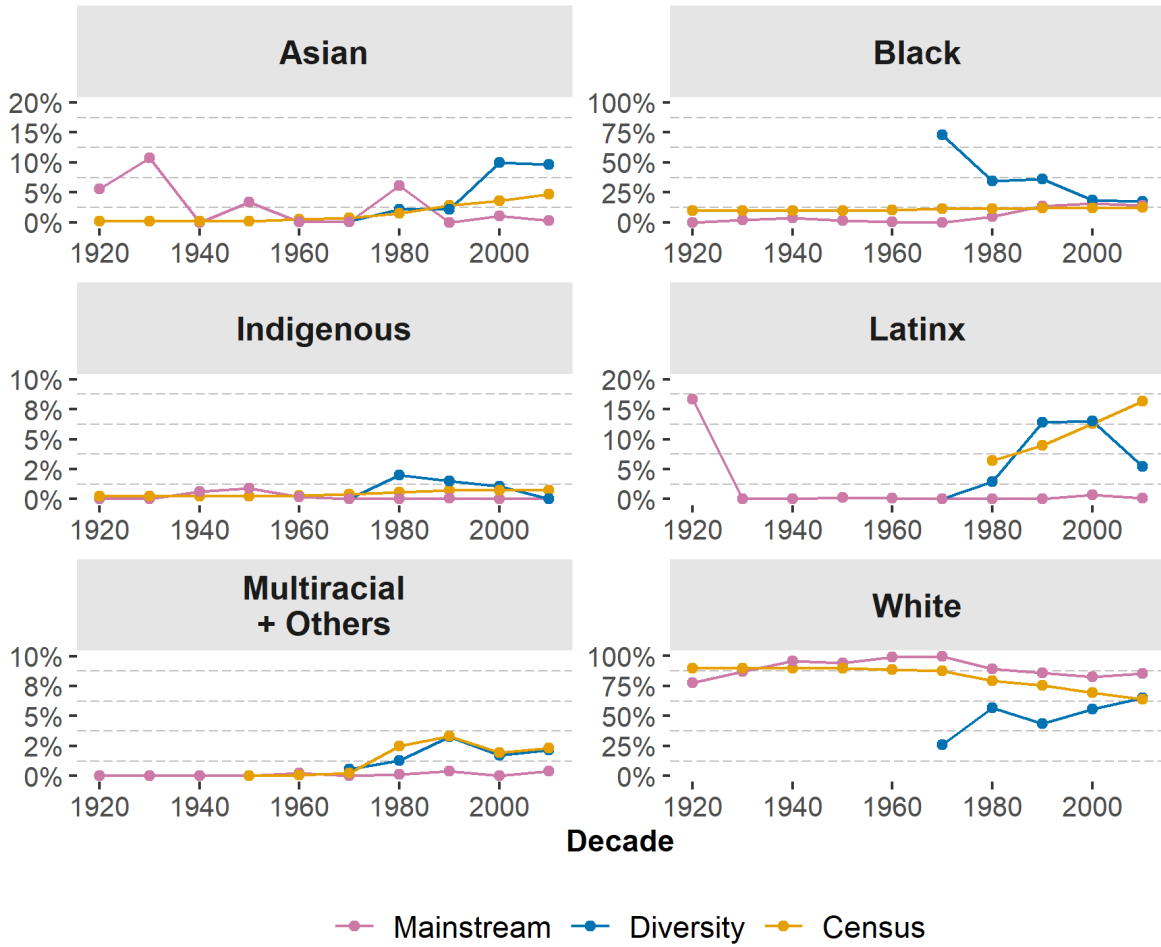


Figure B3. Skin Color Terciles by Age, by Collection



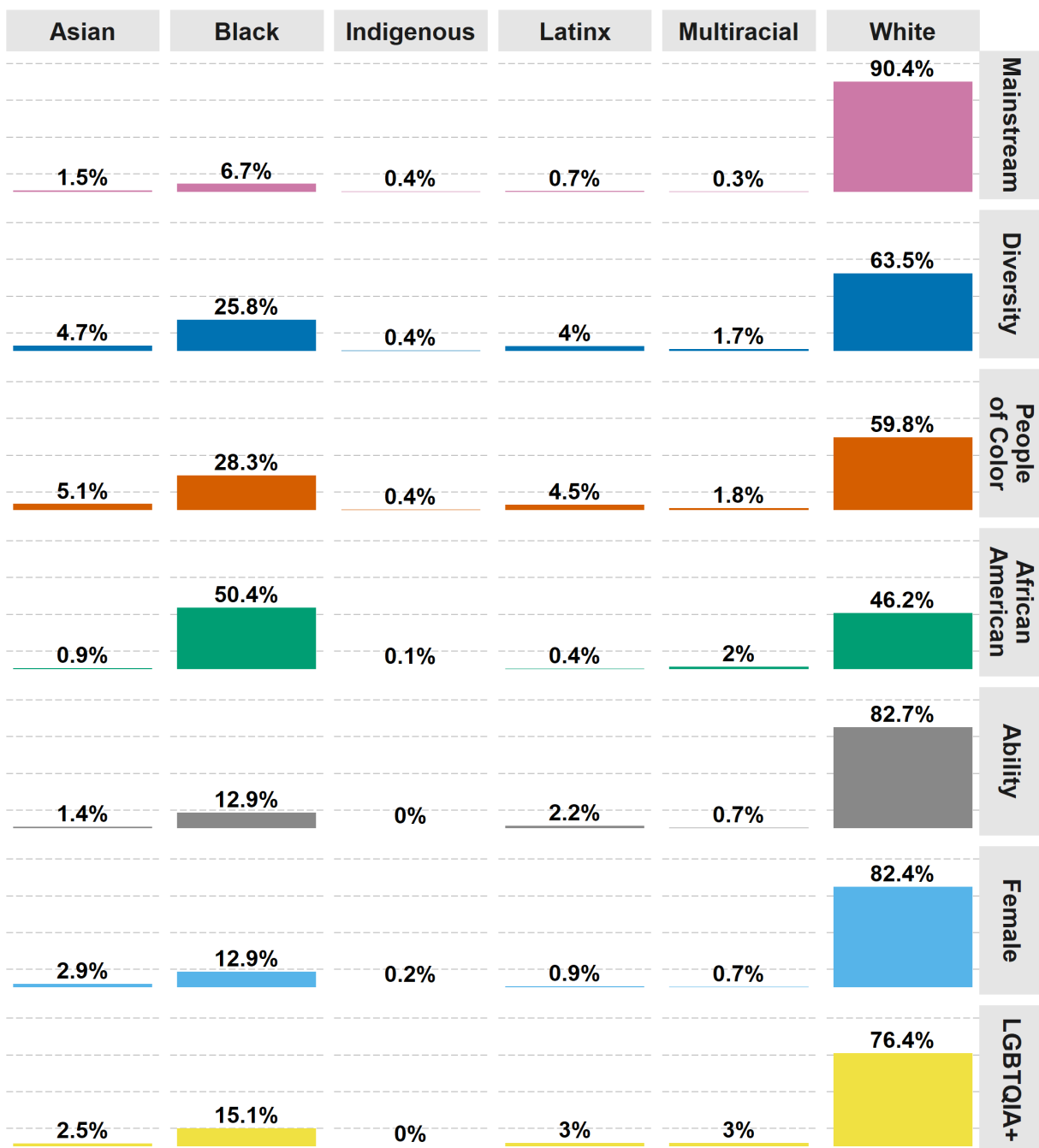
*Note:* In this figure, we show the proportion of faces in each tercile of the perceptual tint distribution by the classified age (adult vs. child) of the face. We detect faces using our face detection model (FDAI). Within these faces, we classify age using an AutoML algorithm we trained using the UTKFace public data set. Skin tint is determined by the  $L^*$  value of a face's representative skin color in  $L^*a^*b^*$  space. These figures show the results for images that have human skin colors (defined as polychromatic skin colors where  $R \geq G \geq B$ ).

Figure B4. Share of U.S. Population and Famous People in the Text, by Race/Ethnicity



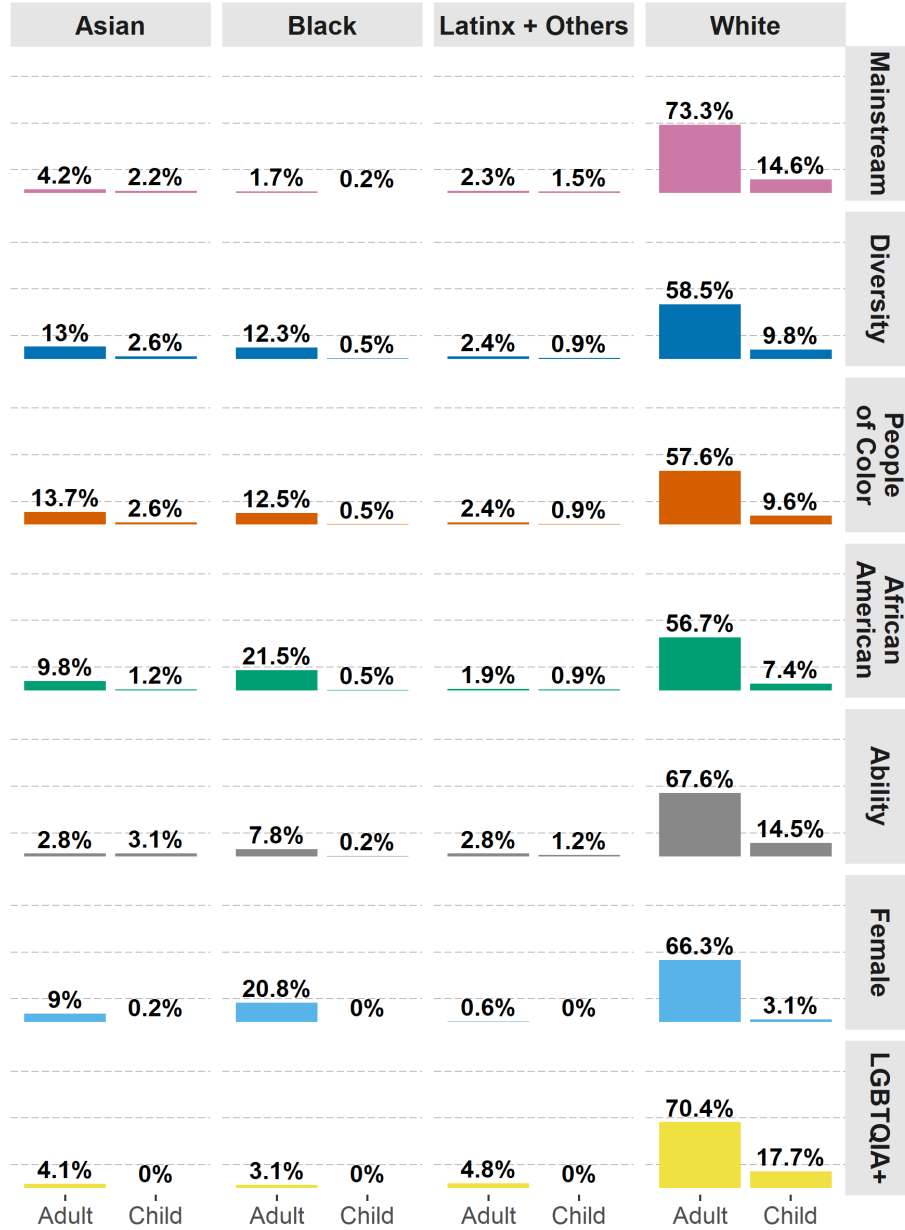
*Note:* In this figure, we show the percent breakdown of famous people mentioned in a given book by race/ethnicity. For example, if Aretha Franklin was mentioned 3 times in a book and Jimmy Carter is mentioned 2 times, then 60 percent of the mentions of famous people in that book would be Black. We then 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 IV.B. 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 6, 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 B5. Race Classifications of Famous Figures in the Text



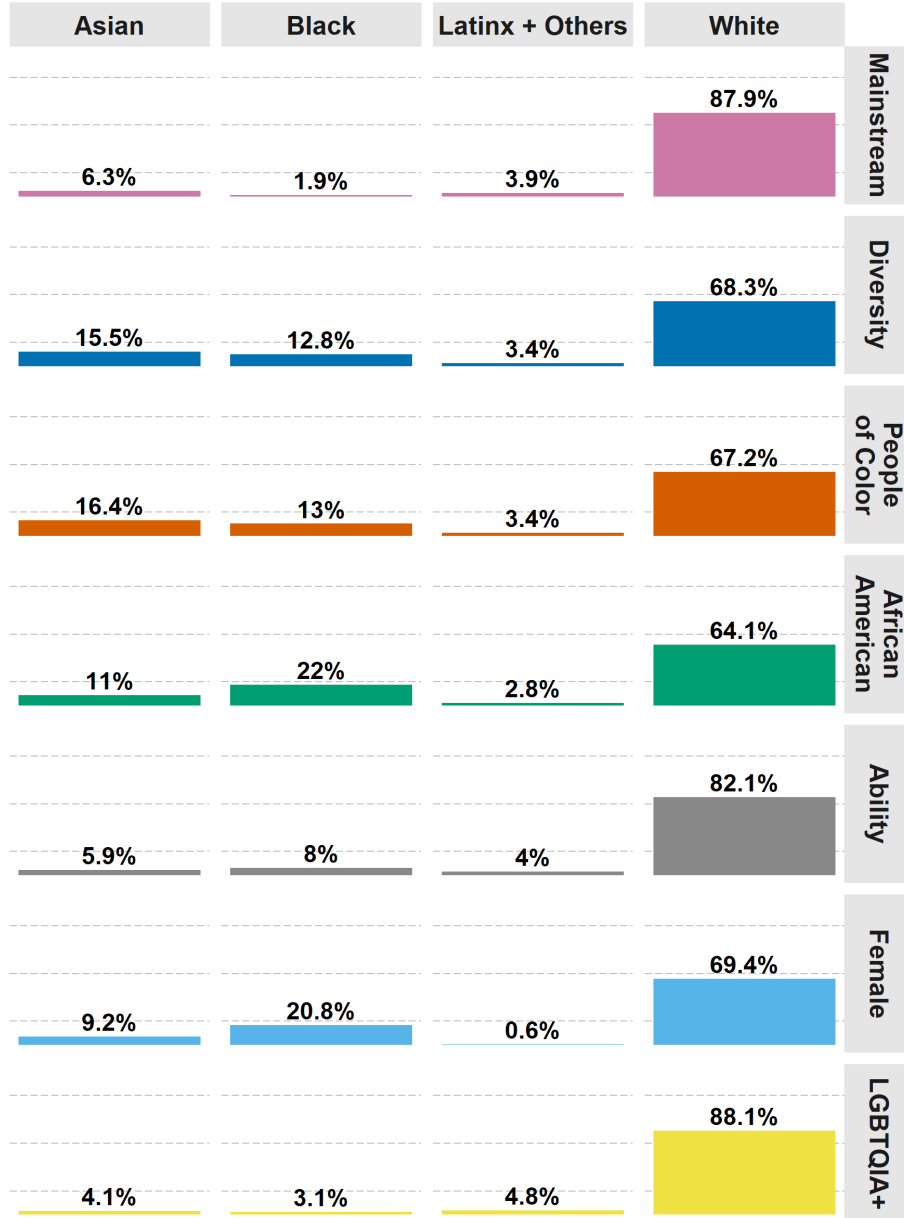
*Note:* In this figure, we count the number of famous people mentioned at least once in a given book and sum over all books in a collection. We then show the percentage breakdown of these famous people by race. For example, if Aretha Franklin was uniquely mentioned in 3 different books within a collection 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 IV.B. 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 B6. Race and Age Predictions of Pictured Characters



*Note:* In this figure, we show the proportion of detected faces in all collections by race and age predictions. Race and age were classified by our trained AutoML model as described in Section IV.A.3. See Appendix Figure B7 for the same figure broken down by race alone.

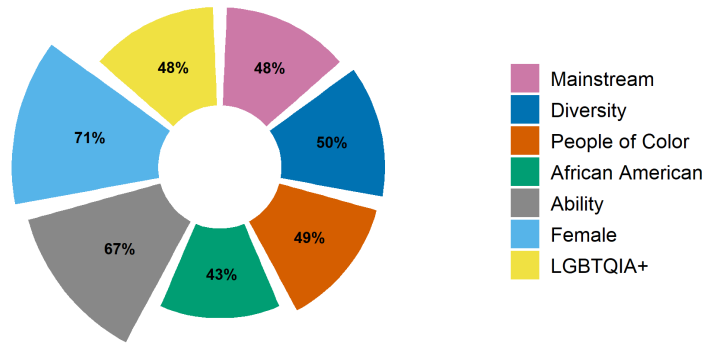
Figure B7. Most Pictured Characters Are Classified as White



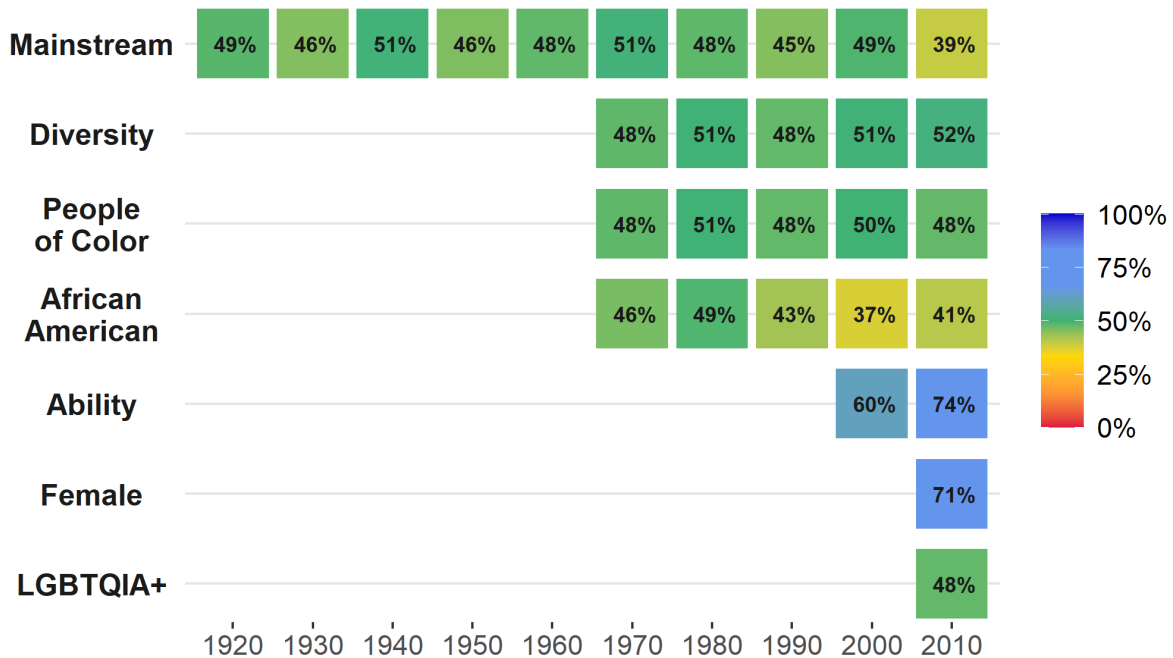
*Note:* In this figure, we show the proportion of faces in a book which our model labels as a given race averaged over all books in a collection. We detect faces using our face detection model (FDAI) described in Section IV.A.1. Within these faces, we classify race using an AutoML algorithm we trained using the UTKFace public data set.

Figure B8. 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 B9. Share of U.S. Population and Pictured Characters, by Identity



*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. In Panel B, we show this by gender. In Panel C, we show this by age group.

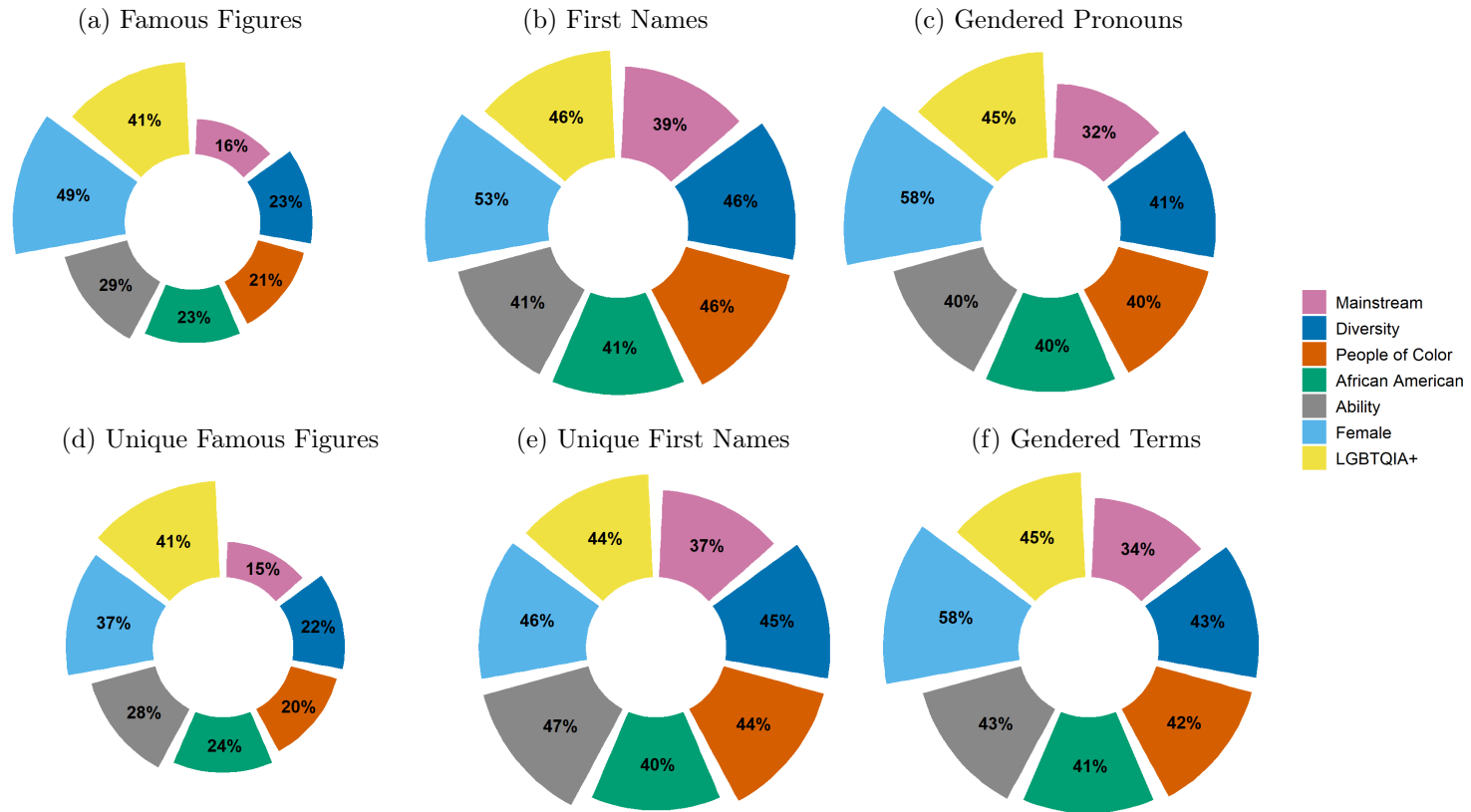
Figure B10. Proportion of Characters in Images and Text, by Race and Gender



*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.

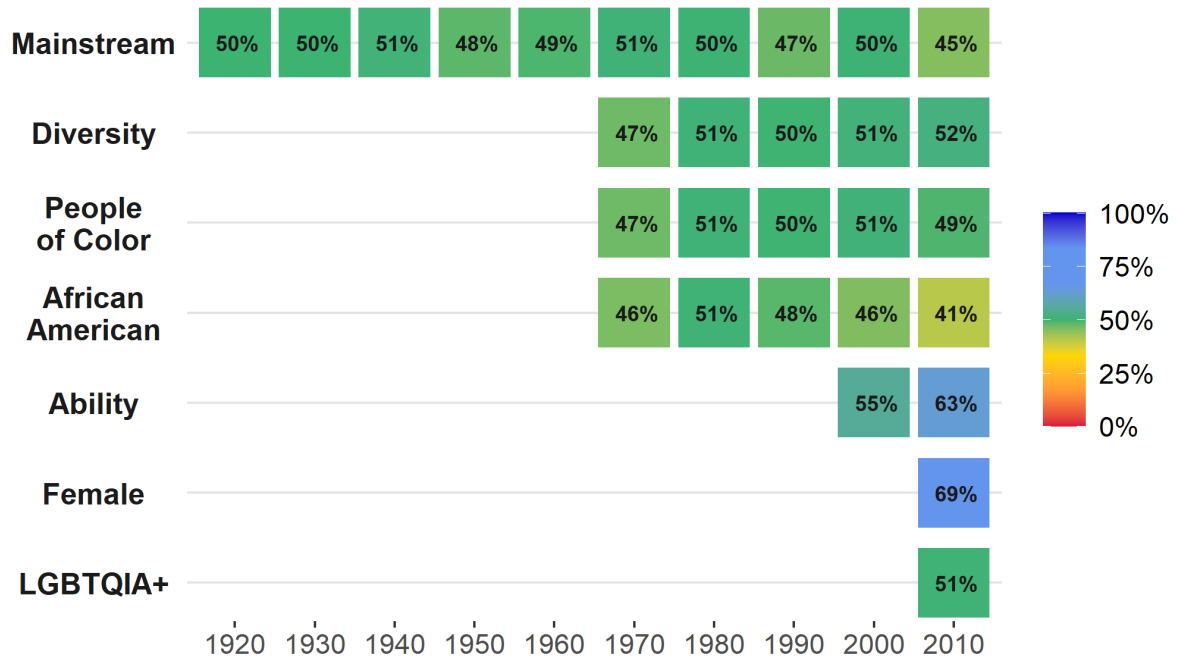


Figure B11. Female Representation in Text, by Type of Word



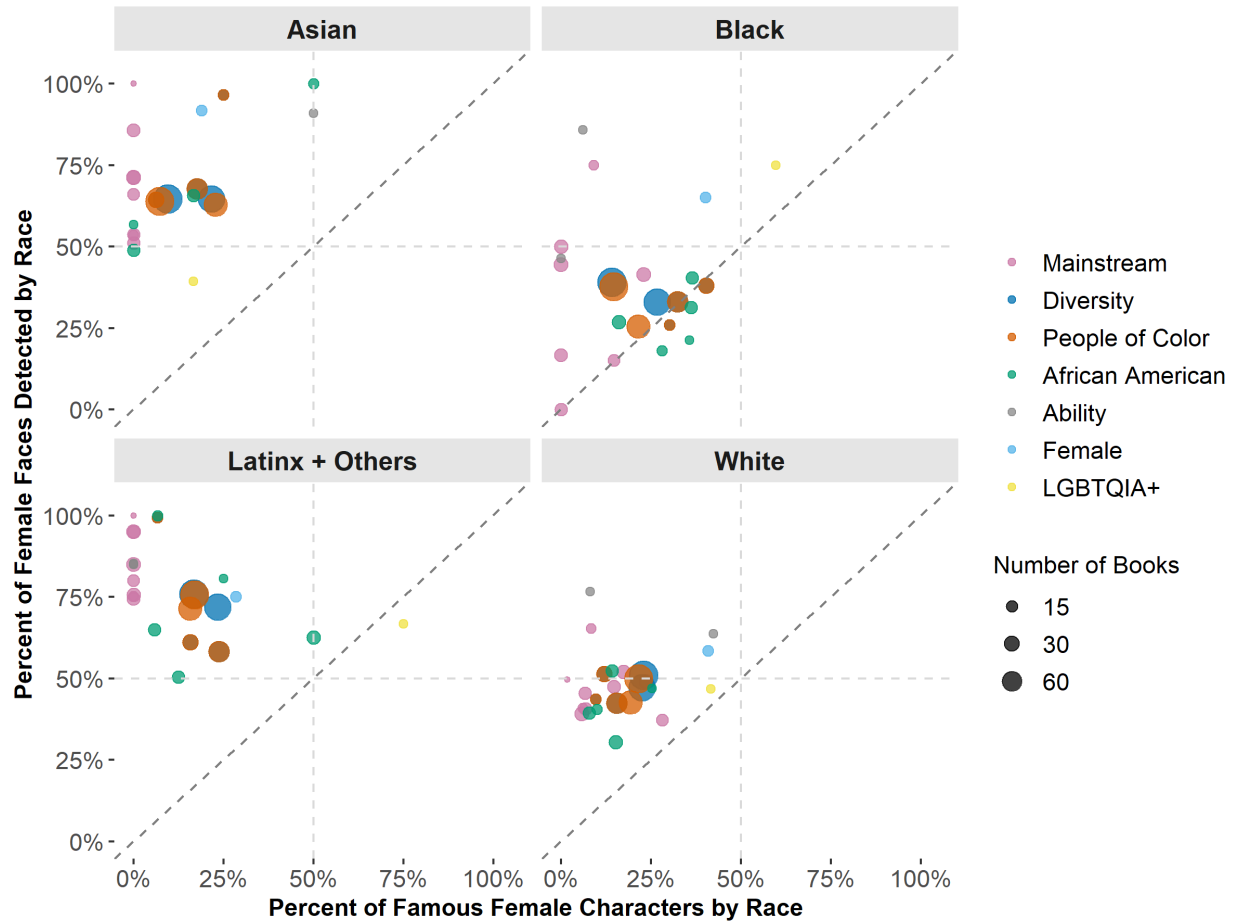
*Note:* We show the proportion of female representation in the text by collection and type of word. In Panel A, we find the percent breakdown of female famous people mentioned in a given book, averaged over all books in a collection. For example, if Aretha Franklin was mentioned 4 times in a book and 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 character first 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 character first names. Panel F shows the percentage of female terms such as queen or nephew (full list provided in Data Appendix) in a given book, averaged over all books in a collection.

Figure B12. 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.

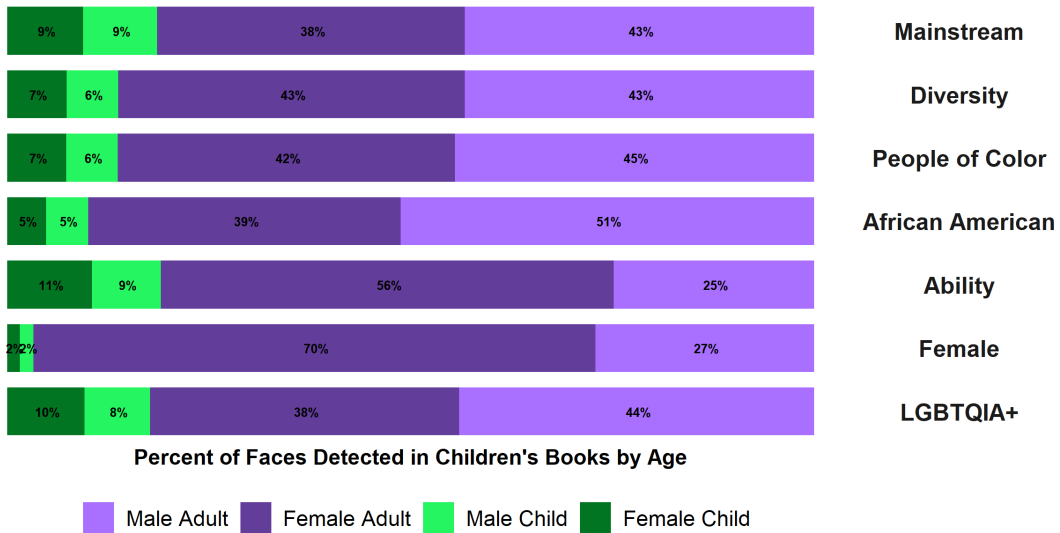
Figure B13. Race and Gender Representation in Images and Text



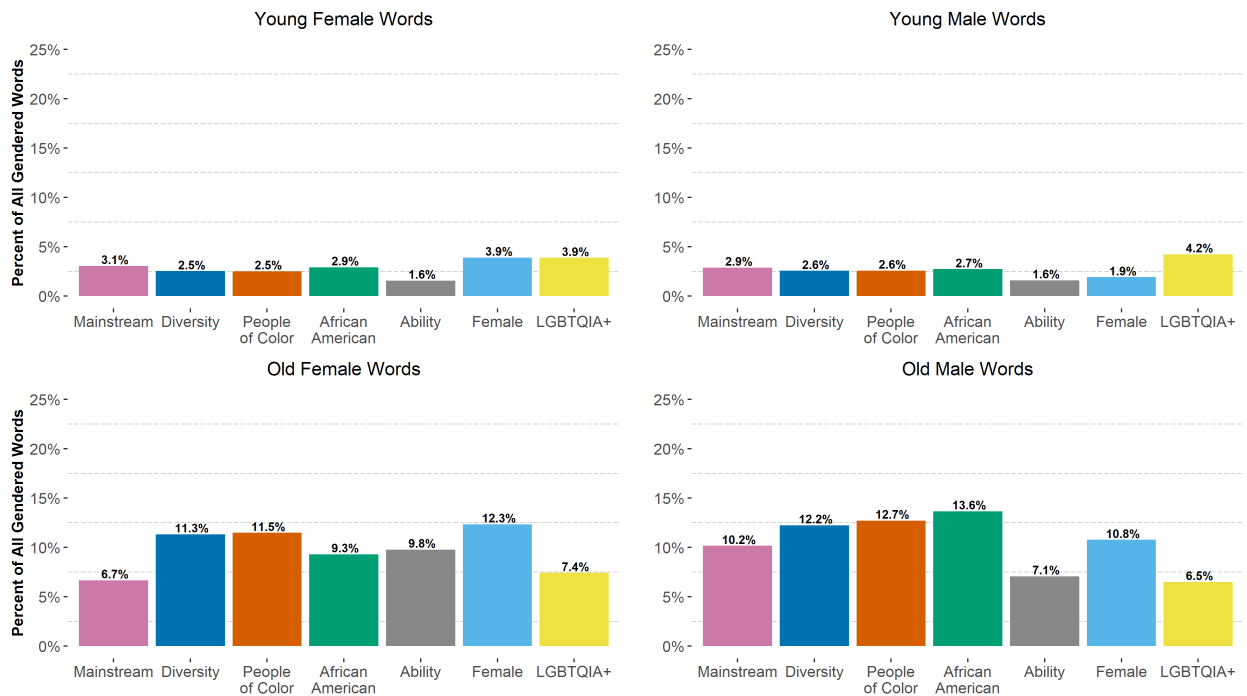
*Note:* In this figure, we plot female faces by race as a proportion of all faces with a given race classification on the y-axis and famous female characters by race as a proportion of all famous characters with a given race classification on the x-axis.

Figure B14. More Adults than Children in Images and Text

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

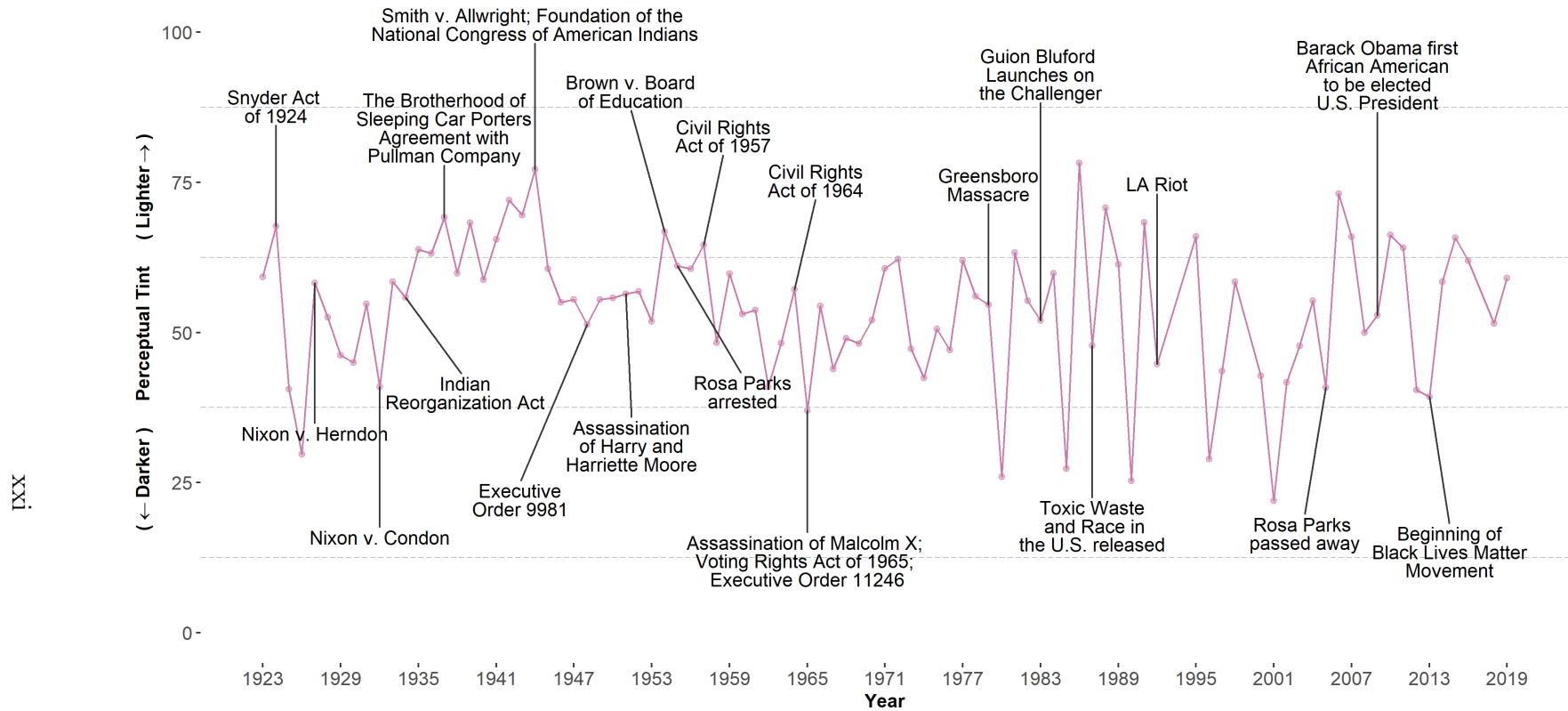


(b) Percent of Gendered Words by Age Group



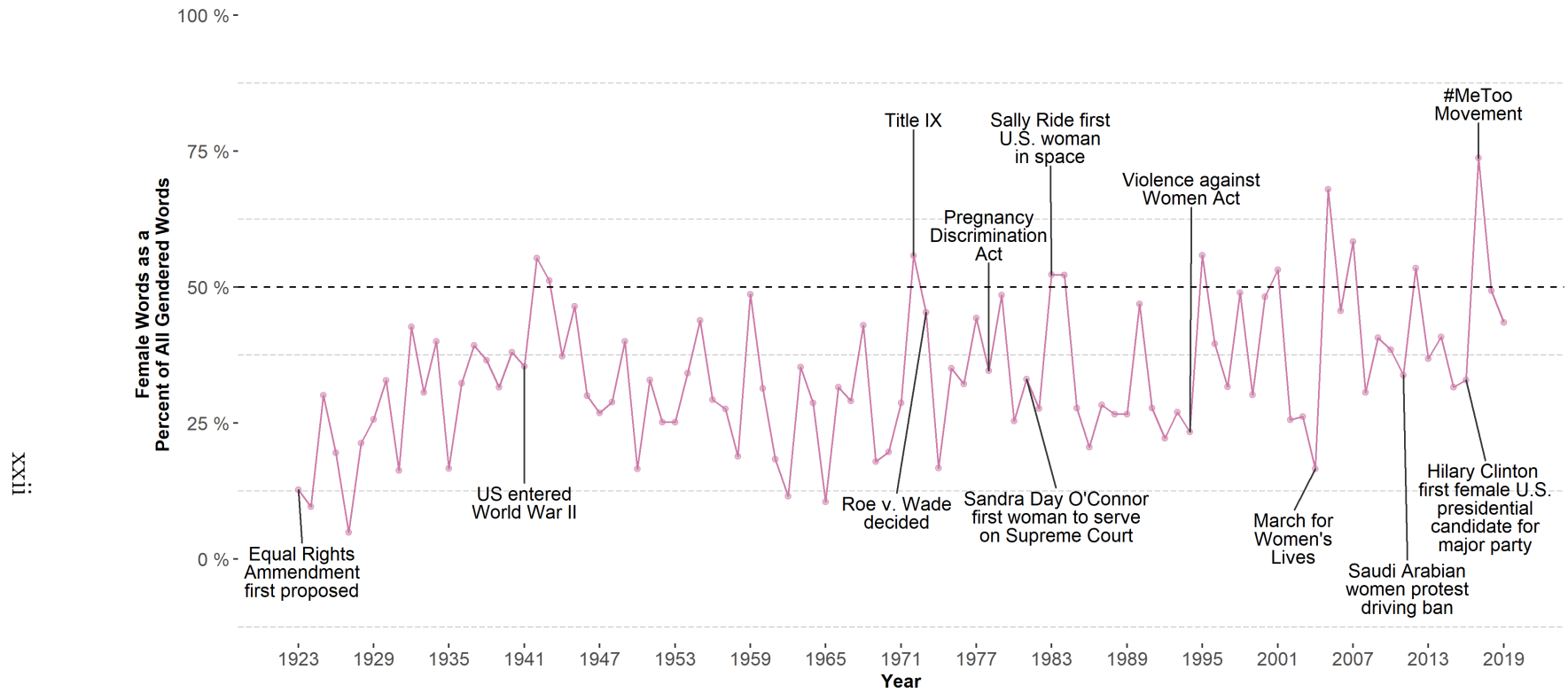
*Note:* In this figure, we show analysis of the representation of age and gender. In Panel A, we show analysis of predicted age and gender in the faces in images. Specifically, we plot the proportion of identified faces classified in each age (adult vs. child) and gender (female vs. male) category. In 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 terms (e.g., queen, nephew). We list the pre-specified gendered terms in the Data Appendix.

Figure B15. Mainstream Representation of Skin Color Throughout Historical Events



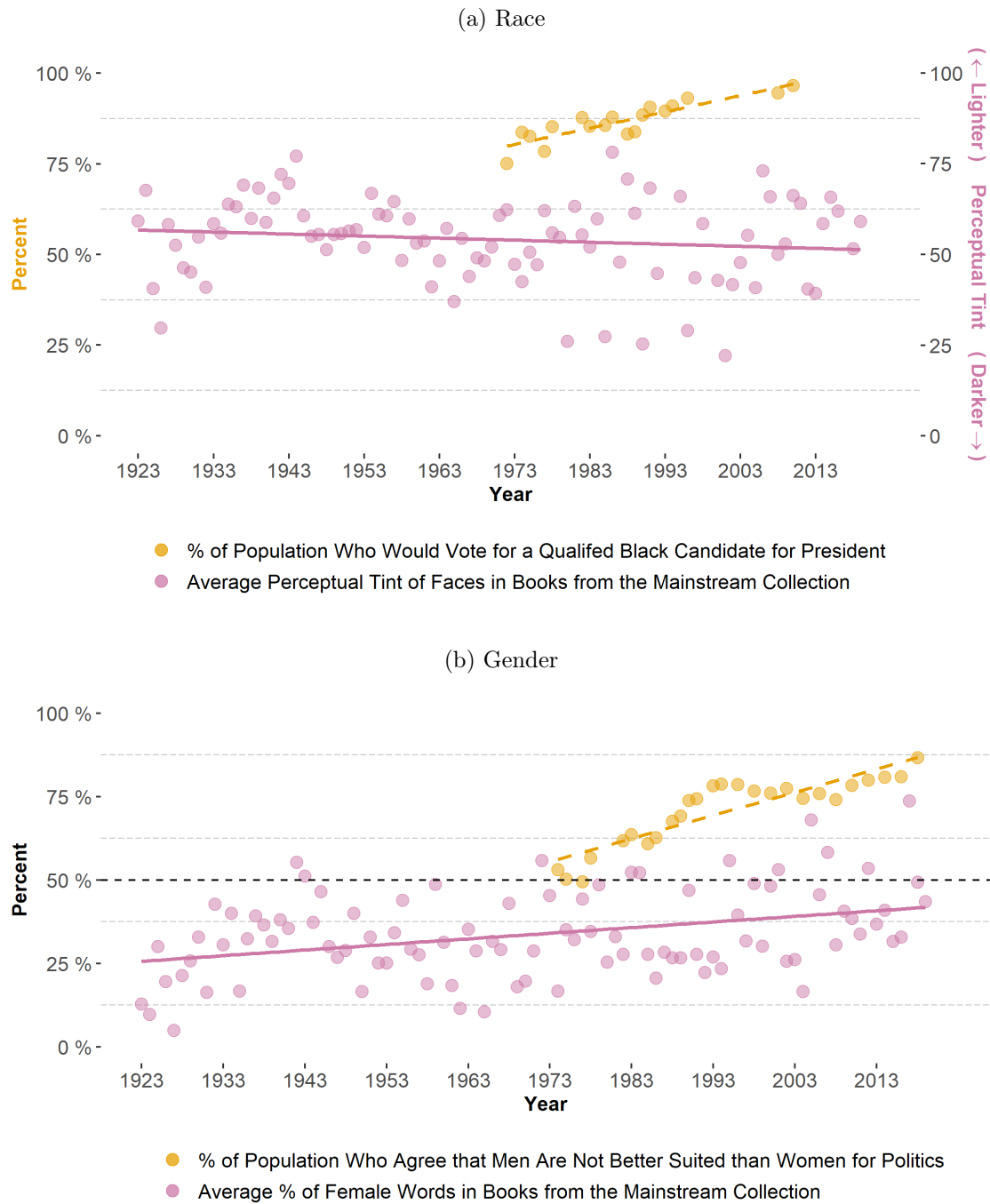
*Note:* In this figure we juxtapose measures of representation of skin color of pictured character faces from the Mainstream collection with the timing of salient historical events.

Figure B16. 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.

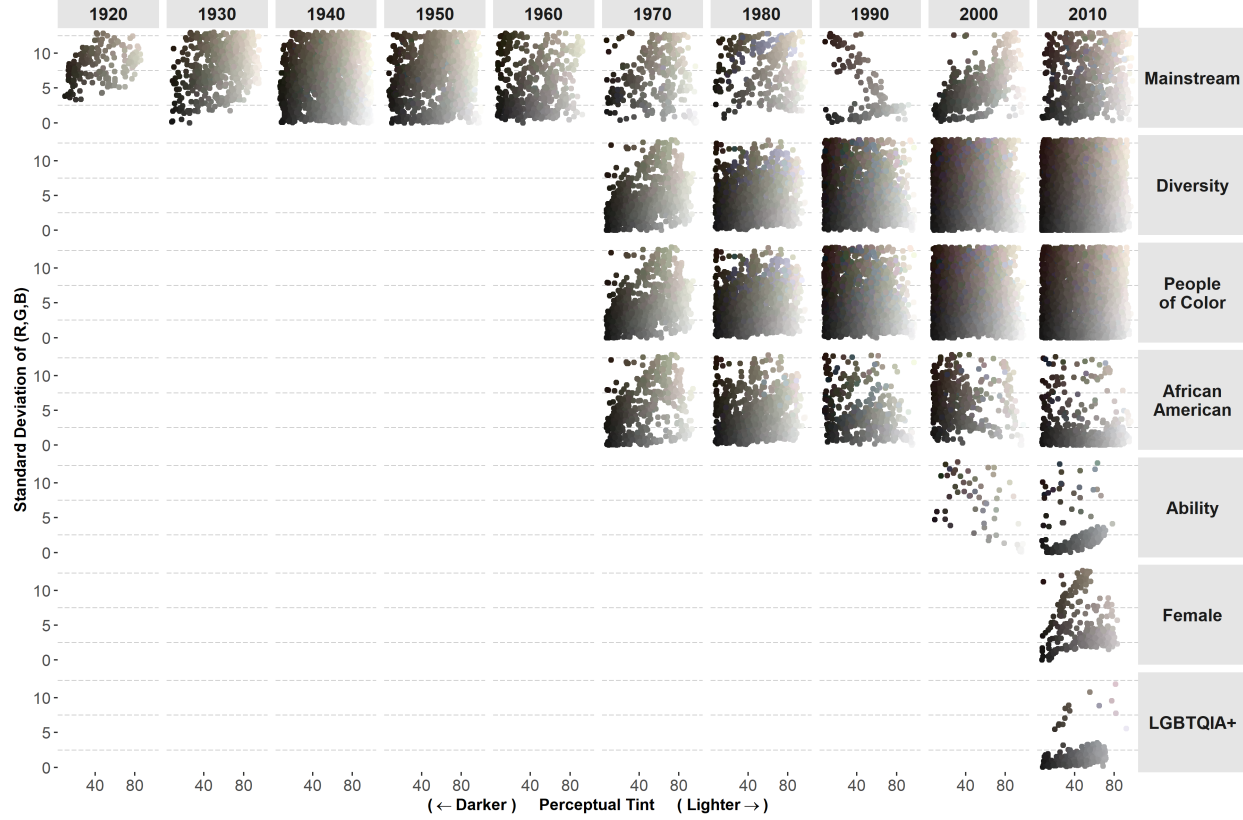
Figure B17. Mainstream Representation and Social Attitudes Over Time



*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

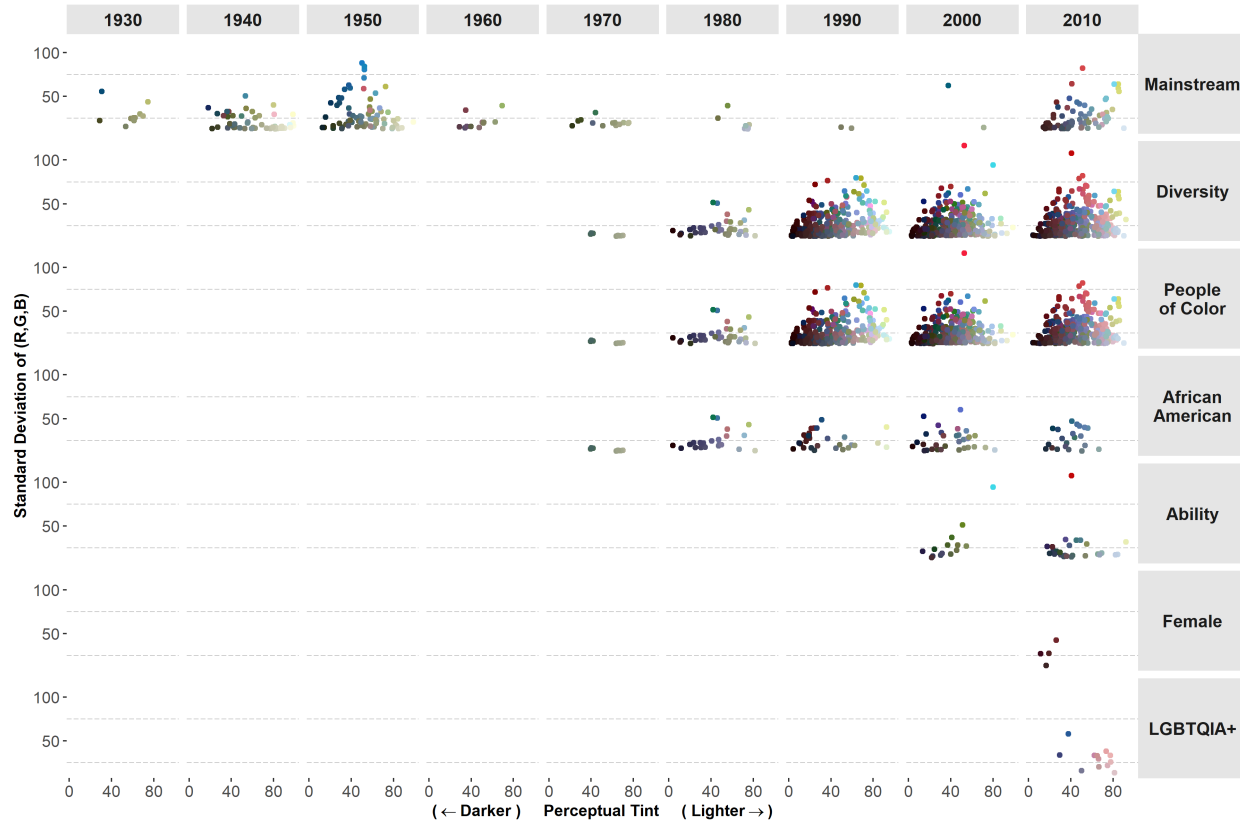
Figure C1. Skin Color Data Over Time, Monochromatic Skin Colors



*Note:* In this figure, we show an analog to Figure B2, 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 IV.A, 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 IV.A.2.

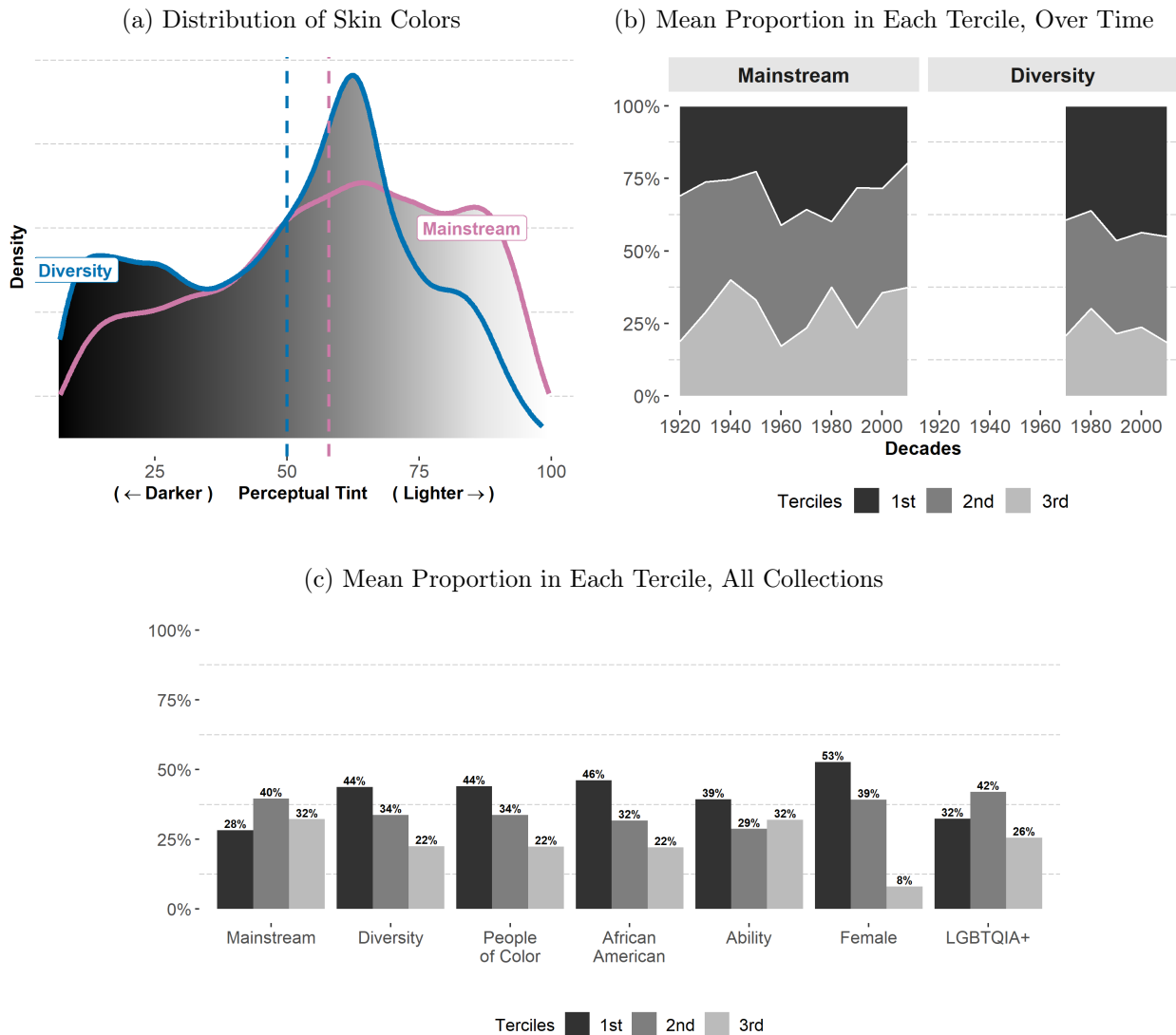


Figure C2. Skin Color Data Over Time, Polychromatic Non-Typical Skin Colors



*Note:* In this figure, we show an analog to Figure B2, 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 IV.A, 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 IV.A.2. The data shown in this figure begin in the 1930s, as opposed to in the 1920s as in Figures B2 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



*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 4, 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 IV.A.

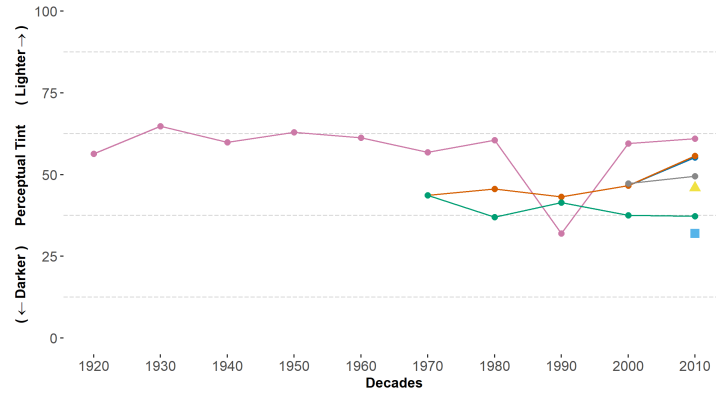
Figure C4. Skin Colors in Faces, by Collection: Polychromatic Non-Typical Skin Colors



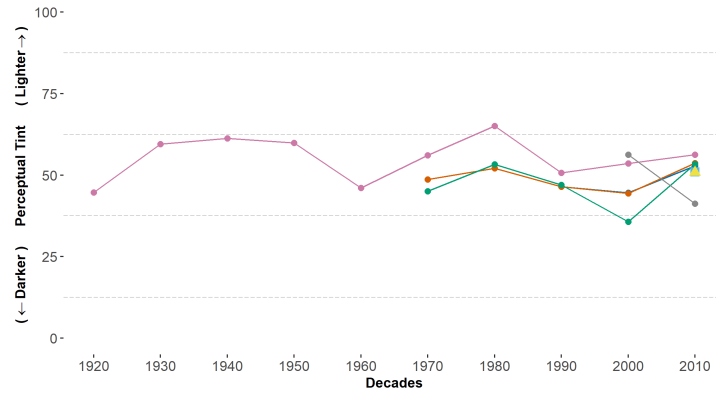
*Note:* This figure shows our analysis of the representative skin colors of the faces detected in the books we analyze. This is an analog to Figure 4, only here we focus on faces that have non-typical skin colors. Panel A shows the distribution of skin color tint for faces detected in books from the 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 IV.A.

Figure C5. Skin Colors over Time, by Collection

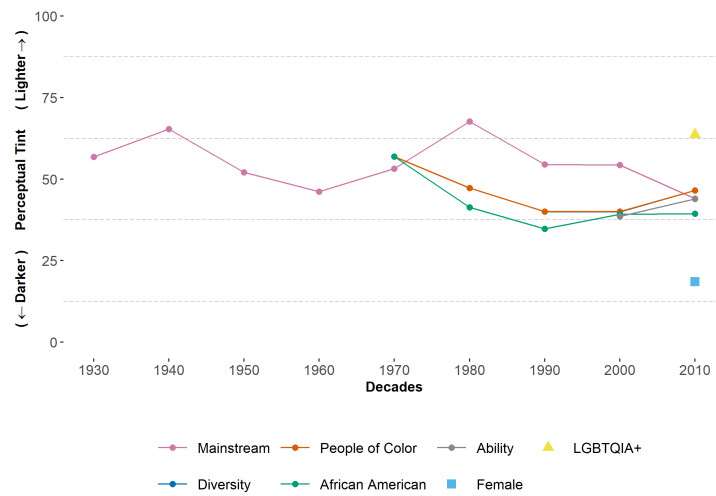
(a) Human Skin Colors



(b) Monochromatic Skin Colors



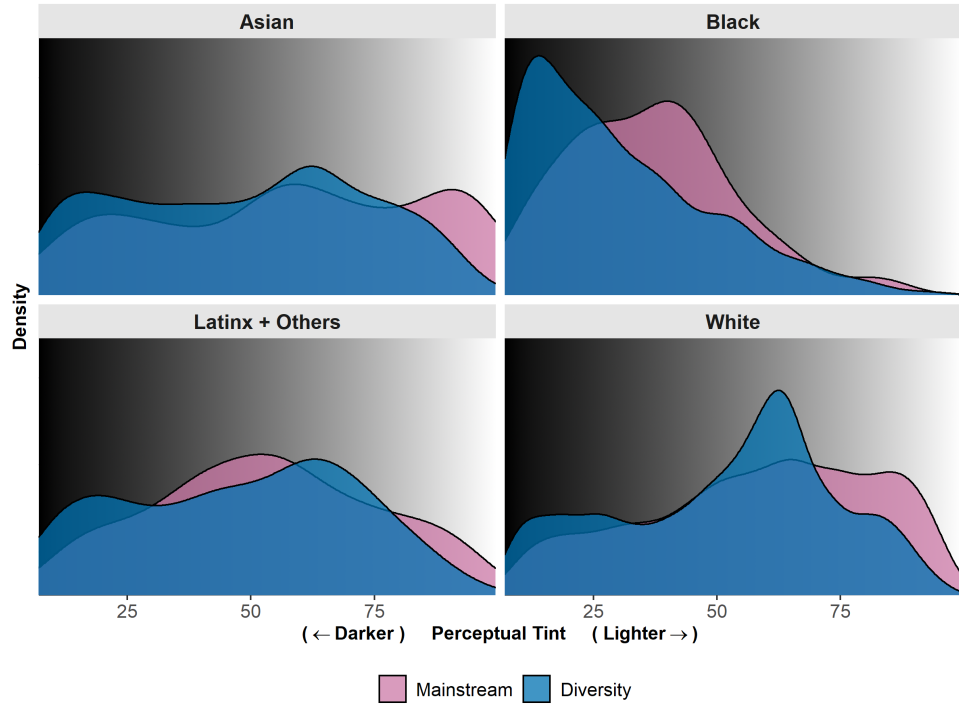
(c) Polychromatic Non-Typical Skin Colors



Note: This figure shows the average skin tint over time in our sample of award-winning children's books. We first take the average skin tint for all faces in a given book, then we average across all books in a given year. We separate the faces by skin color type, Panel A shows the average skin tint for all faces with human skin colors, Panels B and C show the same thing as Panel A but for Monochromatic and Polychromatic Non-typical Skin Colors, respectively.

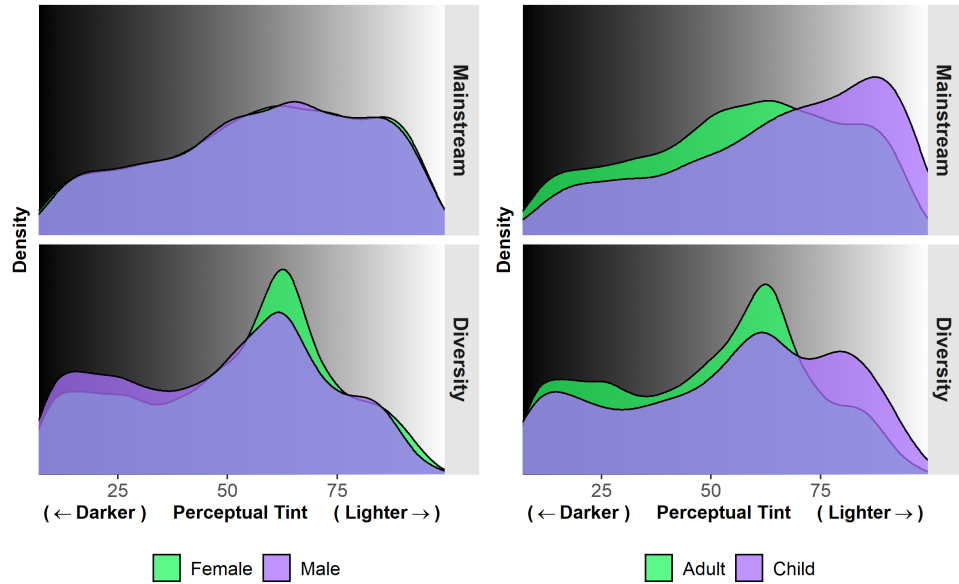
Figure C6. Skin Color by Predicted Race of Pictured Characters: Monochromatic Faces

(a) Skin Color Distribution by Race



(b) Skin Color Distribution by Gender

(c) Skin Color Distribution by Age



*Note:* This figure shows the distribution of skin color tint by predicted race, gender, and age of the detected faces in the Mainstream and Diversity collections. This is an analog to Figure 5, only here focusing on faces depicted in a monochromatic color scheme (e.g., black and white).

## D Methods Appendix

### D.A Comparing AI with Manual Content Analysis

In this section, we describe some of the biases inherent in manual and AI-run 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.

#### D.A.1 AI is Only Human

Historically, content analysis to measure representation has been done “by hand” using human coders (Bell, 2001; Neuendorf, 2016; Krippendorff, 2018). Such analysis provides deep understanding but can generally only be done on a small set of content and necessarily reflects human behavior and biases. While artificial intelligence tools also reflect bias in their training data and algorithms, they can be standardized, are more replicable, and can be applied to a much larger sample than manual content analysis permits. An additional advantage of this approach is that, by following a set of clearly-defined steps for implementation which rely primarily on computers, it minimizes variation in results stemming from researcher-specific biases or idiosyncrasies.

Our paper brings a set of artificial intelligence tools to bear on the field of content analysis. These tools are powerful, computer-driven methods. They are designed by humans and, in many cases, trained with initial human input. We use them because they offer a few key advantages. The first is scale: because algorithms are automated, they allow for analysis of a much larger set of content than would be possible using conventional, “by hand” methods. The second is adaptability: we can rapidly change one dimension of measurement and re-run the analysis at low cost. Were we to do this via hand-coding, the cost would increase linearly with each addition or adjustment (see Section D.A); with AI-based analysis, the marginal cost of such additions or adjustments is much lower.

Measuring representation in content via any means will generate some errors in measurement. In traditional content analysis, analysts may misclassify some images or text. If this occurs at random, this can be treated as standard measurement error, which would be captured via estimating inter-rater reliability (Neuendorf, 2016; Krippendorff, 2018). If, however, traits of the analyst systematically influence their coding, then error from misclassification may be non-classical, leading to a bias in expectation (Krippendorff, 1980). This can arise, for example, if an analyst’s identity (e.g., one’s race and/or gender) causes them to classify content differently than analysts of different identities (Boer, Hanke and He, 2018).

These same biases appear in AI models. Many AI models, including those we use, are trained using a set of data which are first labeled by humans. Furthermore, nearly all models are either fine-tuned, evaluated, or both, based on their performance relative to human classification. As a result, the bias in classical content analysis is “baked into the pie” for computer-driven content analysis (Das, Dantcheva and Bremond, 2018).

Most face detection models are trained using photographs of humans – particularly White humans – which could lead us to undercount people of color and illustrated characters if the model were less able to identify characters on which it was not trained. To address this, we trained our own face detection model using 5,403 illustrated faces from the Caldecott and Newbery corpora (discussed in Section IV.A.1). A similar problem with under-detection of certain types of faces could also appear in the skin segmentation process, as we relied upon a series of convolutional neural networks to identify skin, rather than on human-performed identification of the skin region of faces.

These issues persist when classifying features. In the case of gender, for example, all public data sets with labels for gender that we encountered have a binary structure, limiting classification to “female” or “male,” and neglecting to account for gender fluidity or nonbinary identities. Furthermore, intrinsic to these models is the general assumption that we can predict someone’s gender identity using an image of their faces (Leslie, 2020). Similar problems beset the task of classifying putative race (Fu, He and Hou, 2014; Nagpal et al., 2019; Krishnan, Almadan and Rattani, 2020). Resolving these problems is an active field of inquiry, and recent scholarship has suggested several promising paths forward for doing so (Buolamwini and Gebru, 2018; Mitchell et al., 2019).

While AI is a product of and therefore reflects human biases, human biases are also intrinsic to traditional “by-hand” content analysis. Manual coding necessarily can only reflect the biases of the individual coders. We observed that the identities of the manual labelers on our team led to non-classical error, particularly in the classification of race of the pictured characters in images. We therefore use multiple measures for each identity to try to understand the extent of this potential measurement error. For example, in addition to the manually coded putative race of famous figures, we examine also examine skin color of detected characters.

While we primarily use AI tools to study representation, we end this section by emphasizing that AI and manual coding provide complementary understanding of content. The tools we use are meant to rapidly estimate how a human might categorize these phenomena. They are motivated by human perception and, ultimately, their performance is also

evaluated based on how accurately they can determine how a human might perceive the representations in images and text. Our use of these tools depends on human input at each stage, from the conception of tools and the labelling of training data, to the evaluation of the tools’ accuracy and the way that we interpret their results. We see our efforts adding the strengths of recent advances in computational science to content analysis as a natural extension of the rich history of human-driven analysis in this field.

## **D.A.2 Validation**

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. Regardless of whether we use manual coding or computer vision, the broad patterns are the same. Over 50 percent of the characters/detected faces and gendered words are male and the skin colors depicted are skewed away from darker-skinned individuals.

## **D.A.3 Cost-Effectiveness**

It took approximately 40 hours to code the entire book (400 pages at an average of 6 minutes per page).<sup>64</sup> 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).<sup>65</sup>

---

<sup>64</sup>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.

<sup>65</sup>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.



The manually labeled test data are kept separate from the training and hyper-parameter tuning algorithms.<sup>66</sup> The models compare results from the algorithms to the manual labels in the test data to evaluate the accuracy of the algorithms.

We use two specific parameters that are commonly used to evaluate the performance of this class of model: “precision” and “recall.”<sup>67</sup> 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.<sup>68</sup> Formally:

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

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

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 and 76.8 percent recall.

## 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, ge-

---

<sup>66</sup>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.

<sup>67</sup>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.

<sup>68</sup>Sometimes “recall” is also referred to as “sensitivity.”

ometric transformations, and partial occlusions (Lumini and Nanni, 2020). 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: Classifying 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., grayscale, 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:

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

Where  $\mu_i$  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 tints (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  $R \geq G \geq B$ .<sup>69</sup> Otherwise, it is classified as a polychromatic non-typical skin color.

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

$$(D3) \quad NonTypical_i = [1 - Monochromatic_i] \times [1 - Human_i]$$

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

To train our feature classification model we use a publicly available labeled data set called UTKFace which is a large-scale face data set consisting of over 20,000 face images with age, gender, and ethnicity labels. The images cover large variation in pose, facial expression, illumination, occlusion, and resolution and cover a large age range of individuals (from 0 - 116 years old) (Zhang and Qi, 2017). We split this data set into three parts: training (80 percent of the data), validation (10 percent), and testing (10 percent). The resulting model has 90.6 percent precision and 88.98 percent recall in our testing data.

*Race Classification (Images).* The model assigns the probability that a detected face is of a given race category: Asian, Black, Latinx + Others, or White. 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. Each identified face is assigned the

---

<sup>69</sup>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).

race category which the model gives the highest predicted probability to.<sup>70,71</sup>

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

### D.C.1 Digitizing Text

To extract text from digital scans of books, we use the Google Vision Optical Character Recognition (GVOCR). 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.<sup>72</sup>

---

<sup>70</sup>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).

<sup>71</sup>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.)

<sup>72</sup>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 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

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”).

## D.C.2 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 Terms.** The gendered terms we enumerate are as follows.

***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, ma’am, 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

---

(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.

**Old Female.** abuela, abuelita, aunt, auntie, Auntie, aunts, aunty, czarina, emperess, emperesses, empress, empresses, grandma, grandmas, grandmom, grandmother, grandmothers, housekeeper, housekeepers, maam, madame, mama, mamas, mlle, mme, mom, mommies, mommy, moms, mother, mothers, mrs, nana, nanas, queen, queens, stepmother, stepmothers, titi, tsarina, tsaritsa, tzaritzza, 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

## E Seattle Public Library Checkouts Data

To study the impact of being recognized by the children’s book awards we examine, we analyze data from the Seattle Public Library system on all checkouts from the library between April 2005 and September 2017.<sup>73</sup> 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 measuring the number of collection-by-day checkouts, scaled by the number of books in the collection to generate a measure of the average number of checkouts per book, per day, in each of the three collections. We limit checkout data for each book to approximately one calendar year before the award was given and the two following calendar years.

To generate Figure 1, 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 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 1 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 average 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

---

<sup>73</sup>These data are publicly available at <https://data.seattle.gov/Community/Checkouts-by-Title/tmmm-ytt6>; site accessed on April 15, 2021.

*Diversity* collection variables.

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

Parameter	<i>Dependent variable:</i>
	Estimate
Non-Recognized Children’s Books in Library Fixed Effect	0.089*** (0.004)
Diversity Collection Fixed Effect	0.064*** (0.004)
Mainstream Collection Fixed Effect	0.165*** (0.004)
Post	0.032*** (0.005)
Post $\times$ Diversity	0.012* (0.008)
Post $\times$ Mainstream	0.277*** (0.008)
Observations	3,387
Adjusted R <sup>2</sup>	0.757

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

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 coefficient on the “post-award” variable (*Post*), and the coefficient on the interaction term between *Post* and the *Mainstream* collection, which sums to approximately 0.474. 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 and the coefficient on *Post*, which sums to approximately 0.121.

An alternate interpretation is that after winning the award, the Mainstream collection books are approximately 2.9 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*, 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 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 11 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 0.108. 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 0.121. Prior to receipt of the award, they were approximately 28 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 1 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, Disaggregated by Winners and Honorees

Parameter	<i>Dependent variable:</i>
	Estimate
Non-Recognized Children's Books in Library Fixed Effect	0.089*** (0.006)
Diversity Honoree Fixed Effect	0.063*** (0.006)
Diversity Winner Fixed Effect	0.065*** (0.006)
Mainstream Honoree Fixed Effect	0.165*** (0.006)
Mainstream Winner Fixed Effect	0.183*** (0.006)
Post	0.032*** (0.008)
Post $\times$ Diversity Honoree	0.006 (0.011)
Post $\times$ Diversity Winner	0.014 (0.011)
Post $\times$ Mainstream Honoree	0.205*** (0.011)
Post $\times$ Mainstream Winner	0.499*** (0.011)
Observations	5,645
Adjusted R <sup>2</sup>	0.743

Notes: 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.<sup>74</sup> 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.

---

<sup>74</sup>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 was previously recognized by 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.<sup>75</sup> 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.

---

<sup>75</sup>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 was previously recognized by 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](http://ailanet.org/activities/american-indian-youth-literature-award)
- Américas Award  
Site: [claspprograms.org/pages/detail/65/About-the-Award](http://claspprograms.org/pages/detail/65/About-the-Award)
- Name: Arab American Book Award  
Site: [arabamericanmuseum.org/book-awards/](http://arabamericanmuseum.org/book-awards/)
- Asian/Pacific American Award for Literature  
Site: [apalaweb.org/awards/literature-awards/literature-award-guidelines/](http://apalaweb.org/awards/literature-awards/literature-award-guidelines/)
- Carter G. Woodson Book Awards  
Site: [woodsonawards.weebly.com/](http://woodsonawards.weebly.com/)
- Coretta Scott King Book Award



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

## G Perspectives of Suppliers of Children’s Books

We complement our quantitative analysis of the supply and demand pressures on publishers’ choice of books with qualitative analysis of data from semi-structured, one-on-one interviews of professionals who currently work at or recently worked at libraries, publishing houses, and children’s bookstores, and/or who served on award selection committees. Our interviews began with a prompt that asked a series of questions, first about the processes they used to identify and select books, and then about their perception and understanding of the forces that shape the content of these books.

A few key themes arose from these conversations. The first theme is that many booksellers, publishers, and librarians wish to procure and promote books that highlight people from historically marginalized groups, particularly Black and Latina/o/x people. A common goal across librarians and booksellers was the desire to show children both potential versions of themselves, as well as potential versions of the world they grow up to inhabit. One professional who had served as both a librarian and a bookseller asserted that, when presenting books to children, librarians and booksellers alike wish “to provide each child with both a mirror and a window.” This paraphrases the description in Bishop (1990), which argues that the books we give to children should serve as mirrors, windows, and sliding glass doors - in other words, the books should show children visions of themselves, windows onto the reality they inhabit, and doors through which they can step to see imaginary futures they might inhabit, respectively.

The second theme is that, until recently, this desire to present children with both a mirror and a window was very difficult to meet. Several interviewees asserted that this difficulty arose from mainstream publishers not offering sufficient amounts of this content. This corresponds to the economic forces we study in Section VI, wherein books with greater representation of non-dominant societal groups will be under-supplied by the market. One interviewee – the owner of a decades-old children’s bookstore in a medium-sized midwestern city – lamented that before the mid-2010’s, requests to publishers for books representing people of color was met with the quip: “we don’t sell those books because those books don’t sell.” In response, motivated booksellers such as this professional sought out smaller publishers specializing in such content, such as Lee and Low, a publishing house founded in 1991 to address this shortcoming.<sup>76</sup>

To better understand the process through which books were selected for these awards,

---

<sup>76</sup>The #WeNeedDiverseBooks movement ([diversebooks.org](http://diversebooks.org)), started in roughly 2012, has also agitated and organized for more equitable representation in books. A relevant resource created to meet this need is the Diverse Book Finder, available at [diversebookfinder.org](http://diversebookfinder.org).

we also conducted semi-structured interviews with people involved in the selection committees. Committee members are selected by either election or appointment by the head of ALSC to serve for a one-year term. Committee members review books published in that year, vetting them based on a set of criteria specific to each award. At the end of the term, the committee convenes to discuss candidates and select honorees. Two key themes arose in these discussions: first, the criteria for selection are stable over time, despite the other secular changes in this period.<sup>77</sup> Second, the composition of the award committees generally comprise a circulating group of librarians, booksellers, and educators that refreshes every year.<sup>78</sup> As a result, these awards – particularly those in the Mainstream collection – are likely to reflect the equilibrium of supply from the publishing industry and demand from the annually rotating group of educators and booksellers selected to be on the committees, rather than the idiosyncratic tastes of a few individuals.

---

<sup>77</sup>We give these criteria for the Mainstream collection awards, and link to those in the Diversity collection, in Appendix F.

<sup>78</sup>According to ALSC bylaws for the Mainstream awards, individuals who served on a committee in one year were ineligible to serve on it in following several years.