WHAT WE TEACH ABOUT RACE AND GENDER: REPRESENTATION IN IMAGES AND TEXT OF CHILDREN'S BOOKS

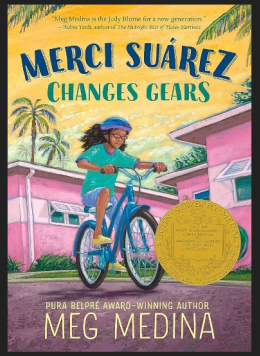
Paper by ANJALI ADUKIA
ALEX EBLE
EMILEIGH HARRISON
HAKIZUMWAMI BIRALI RUNESHA
TEODORA SZASZ

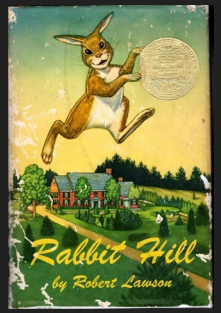
31/08/2023

Presentation by Magnabosco Sara Gondorf Torben Gauthier-Villars Arthème

DATASET

Two collections of books :

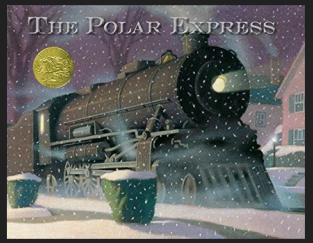




Mainstream Collection :

- Newbery award (since 1922)
- Caldecott award (since 1938)

495 books from 1923 to 2019

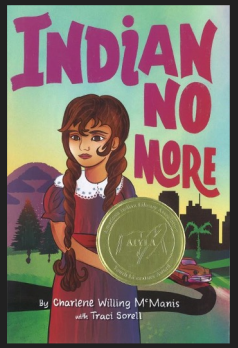


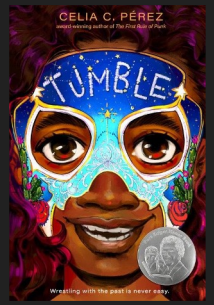
Diversity Collection :

From 16 awards :

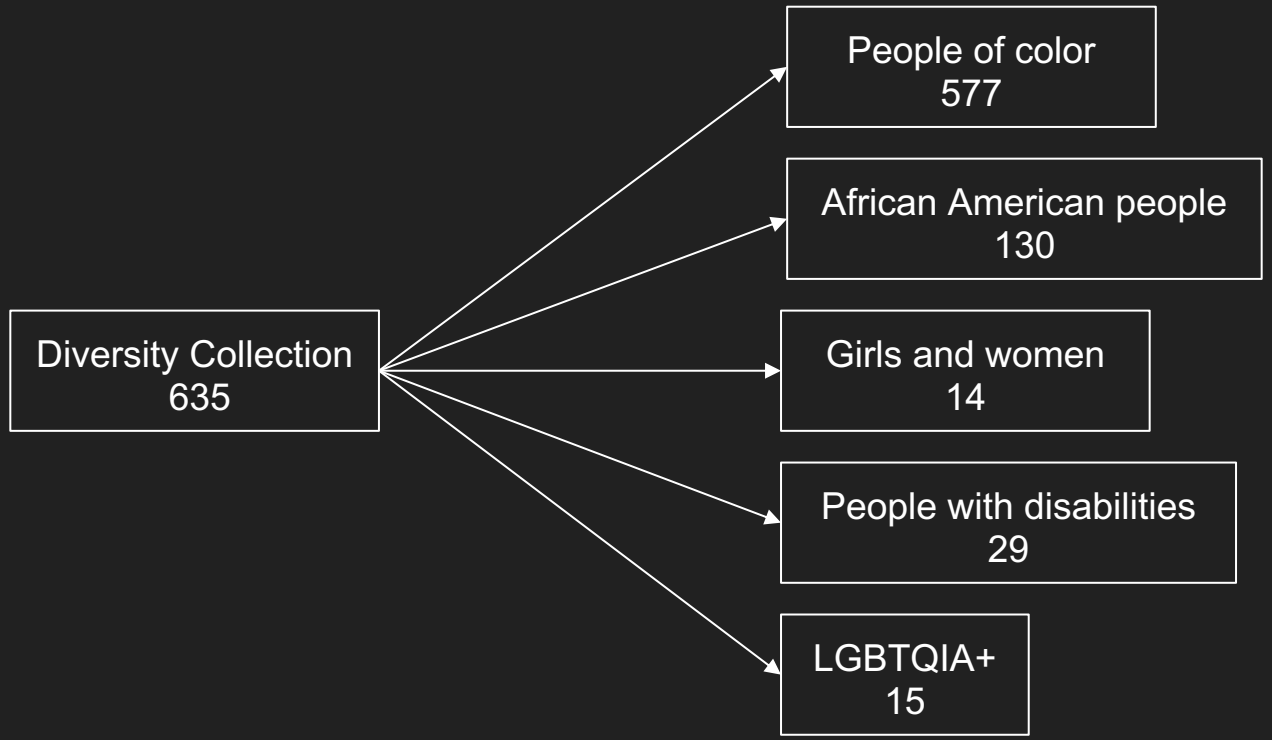
- American Indian Youth Literature
- Notable Books for a Global Society
- Asian/Pacific American Award for Literature
- …

635 books from 1971 to 2019





DATASET



(One book can be in multiple groups)

METHODS AND MODELS

IMAGE FEATURE CLASSIFICATION

Goal :

From illustrated characters or pictures predict :

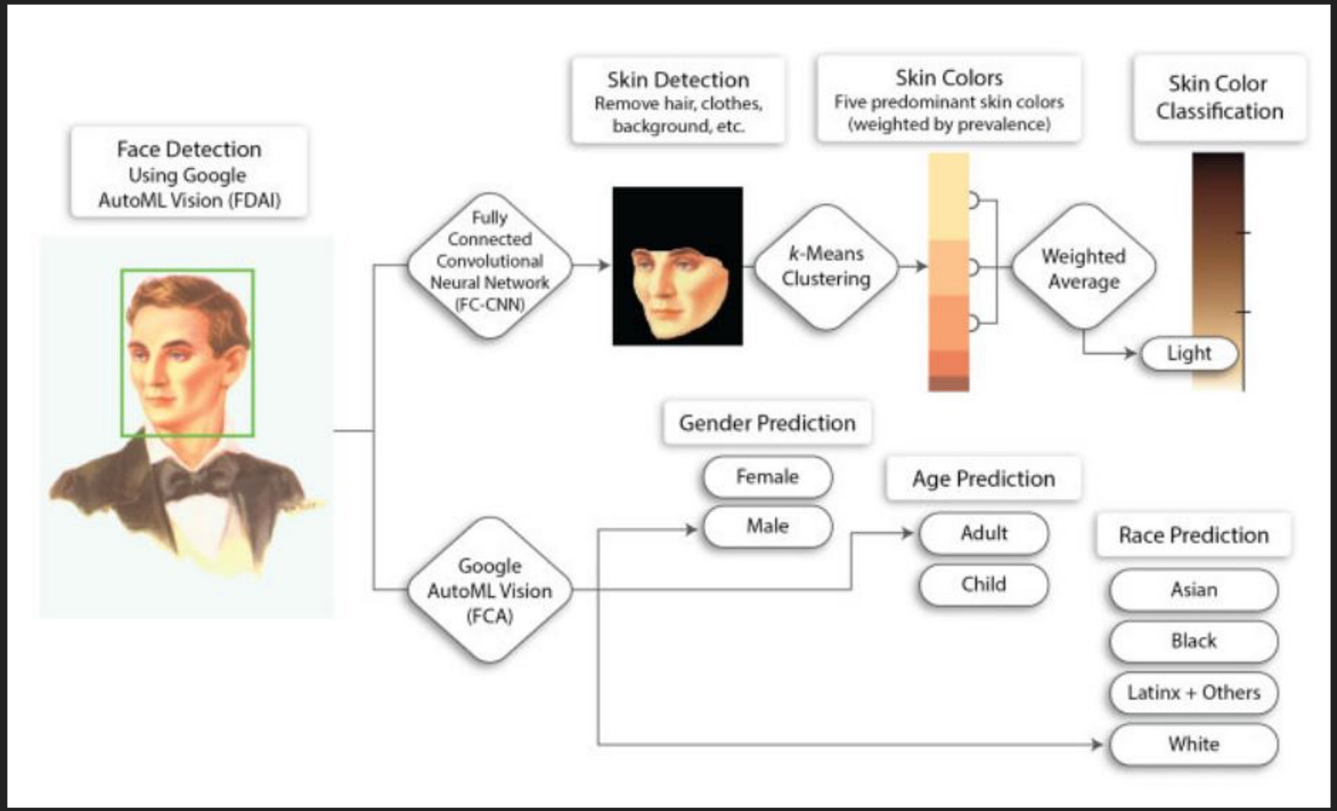
- Skin color
- Gender
- Race
- Age

Example :

Hero Boy (Chris) from the 1985 book *The Polar Express*

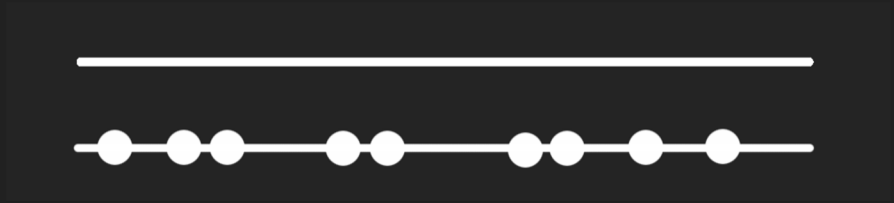
Skin color : Light
Gender : Male
Race : White
Age : Child

IMAGE TO DATA PIPELINE

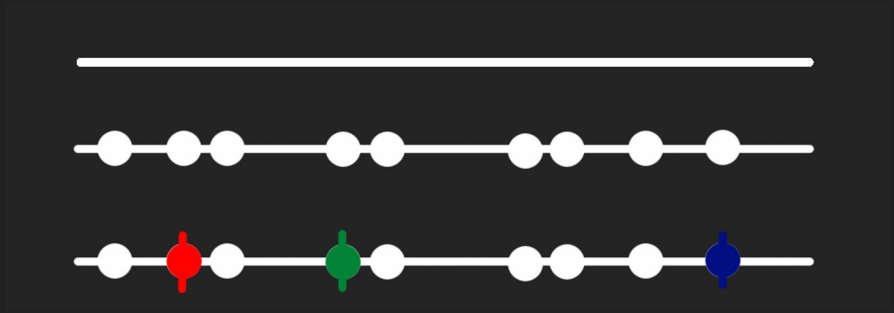


K mean Clustering - Example in 1D with k=3

K mean Clustering - Example in 1D with k=3

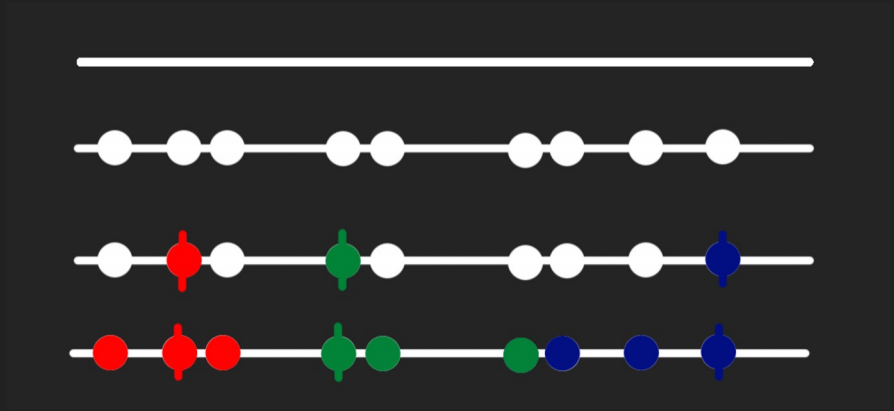


K mean Clustering - Example in 1D with k=3



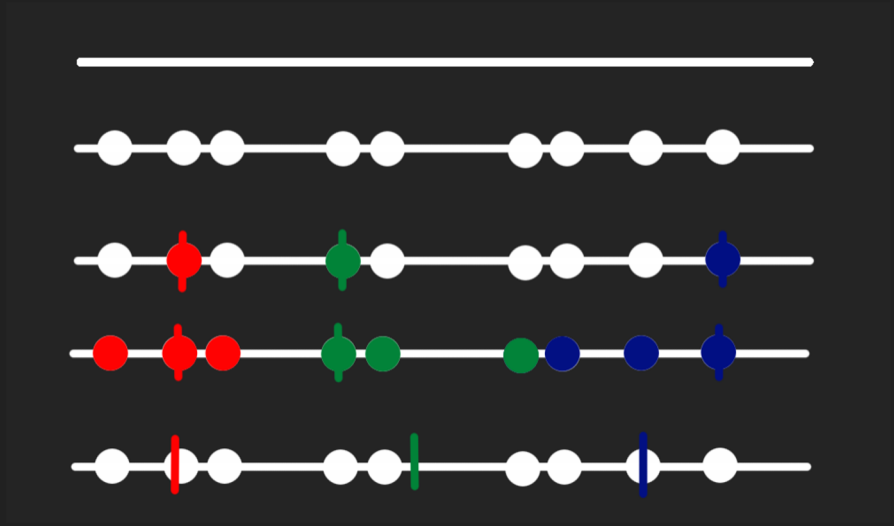
Pick random points as centroid of class.

K mean Clustering - Example in 1D with k=3



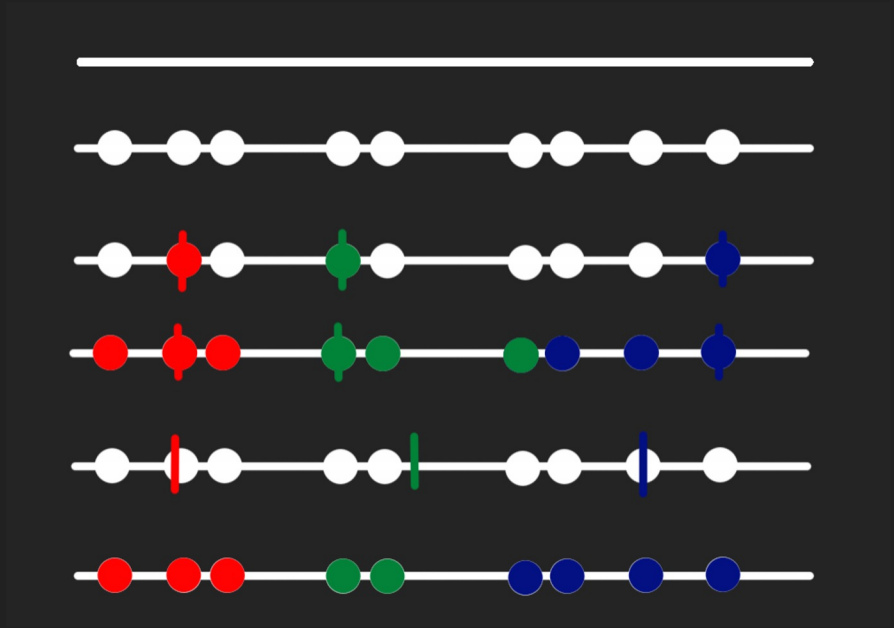
Classify all the others point with the class of the closest centroid.

K mean Clustering - Example in 1D with k=3



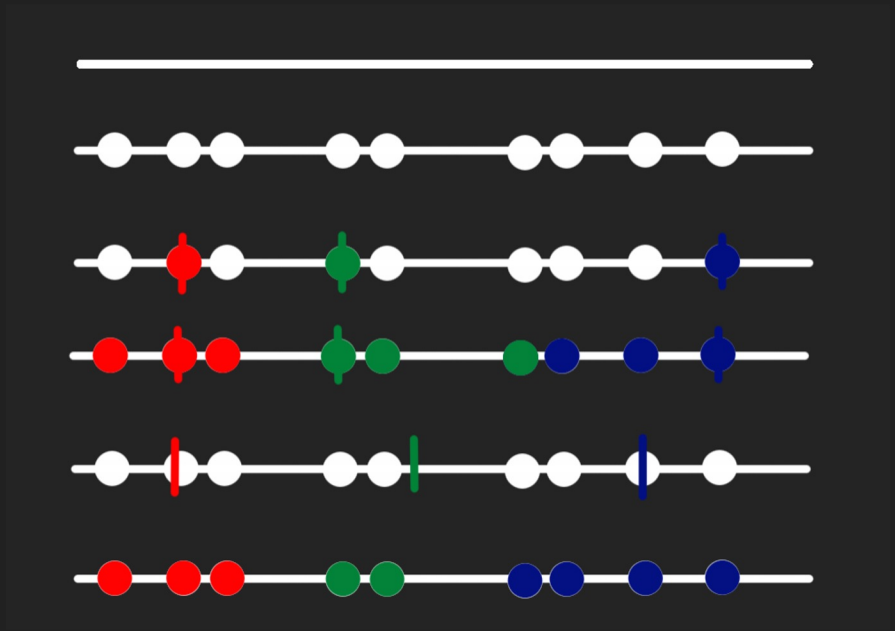
Compute new centroid by taking the mean in each class.

K mean Clustering - Example in 1D with k=3



Re-classify the points sing the nearest centroid as before.

K mean Clustering - 2D exemple



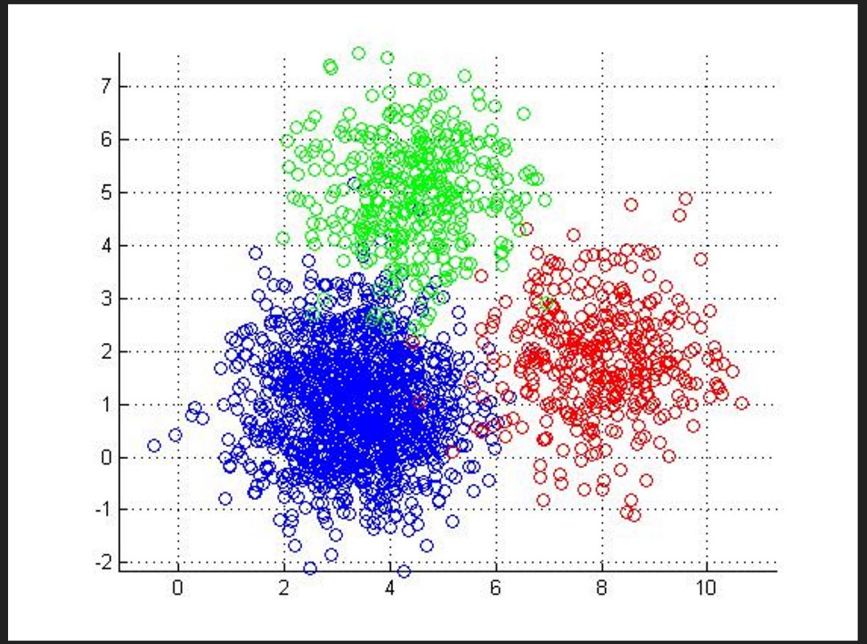
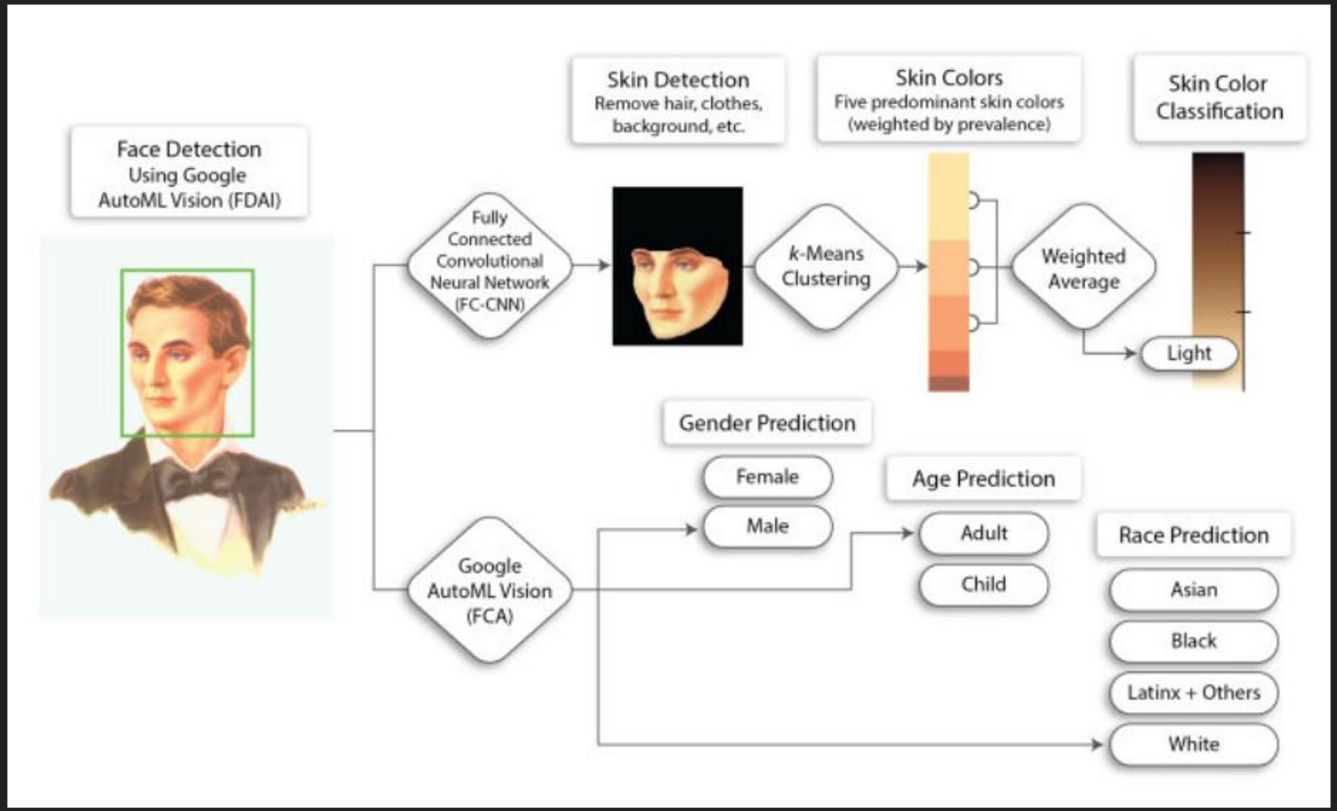


Image from Prateek Joshi on Perpetual Enigma

IMAGE TO DATA PIPELINE

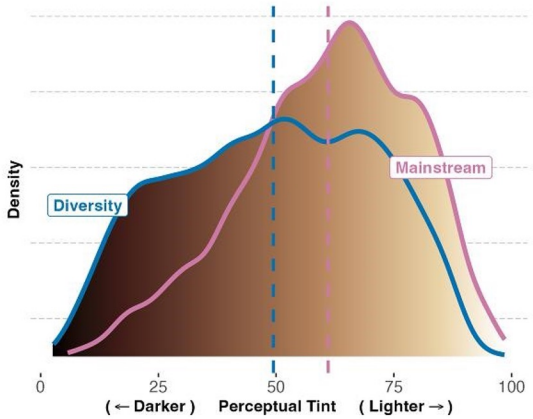


Text as Data

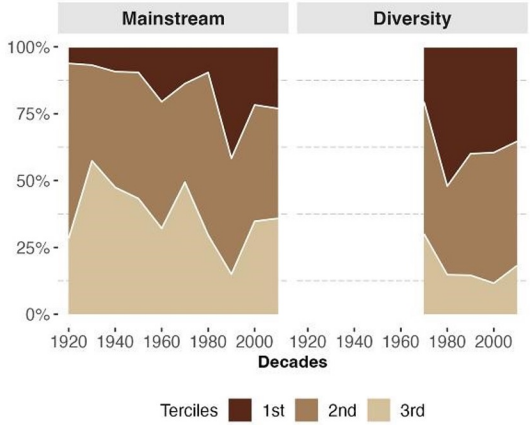
- Token Counts (Gender and Age)
- Named Entity Recognition (Race and Gender)
  - Famous individuals
  - Character first names

RESULTS

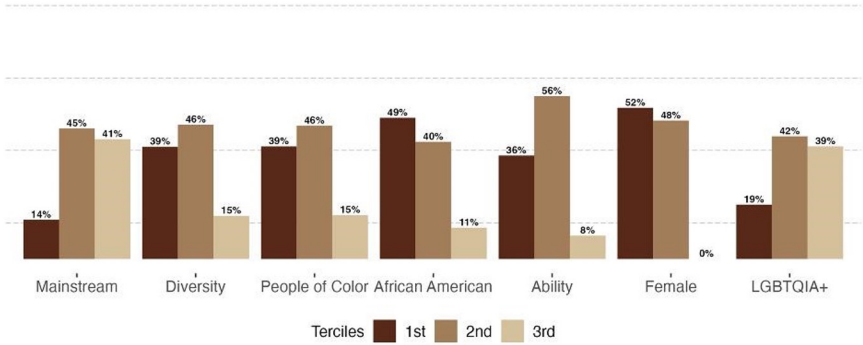
(A) Distribution of Skin Colors



(B) Mean Proportion in Each Tercile, Over Time

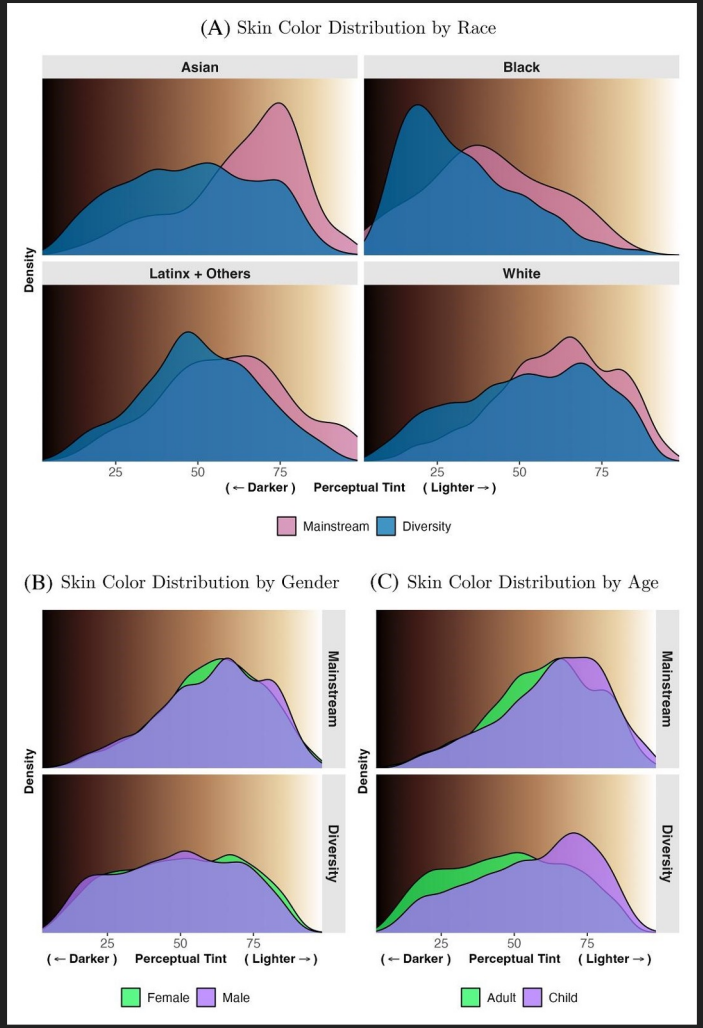


(C) Mean Proportion in Each Tercile, All Collections



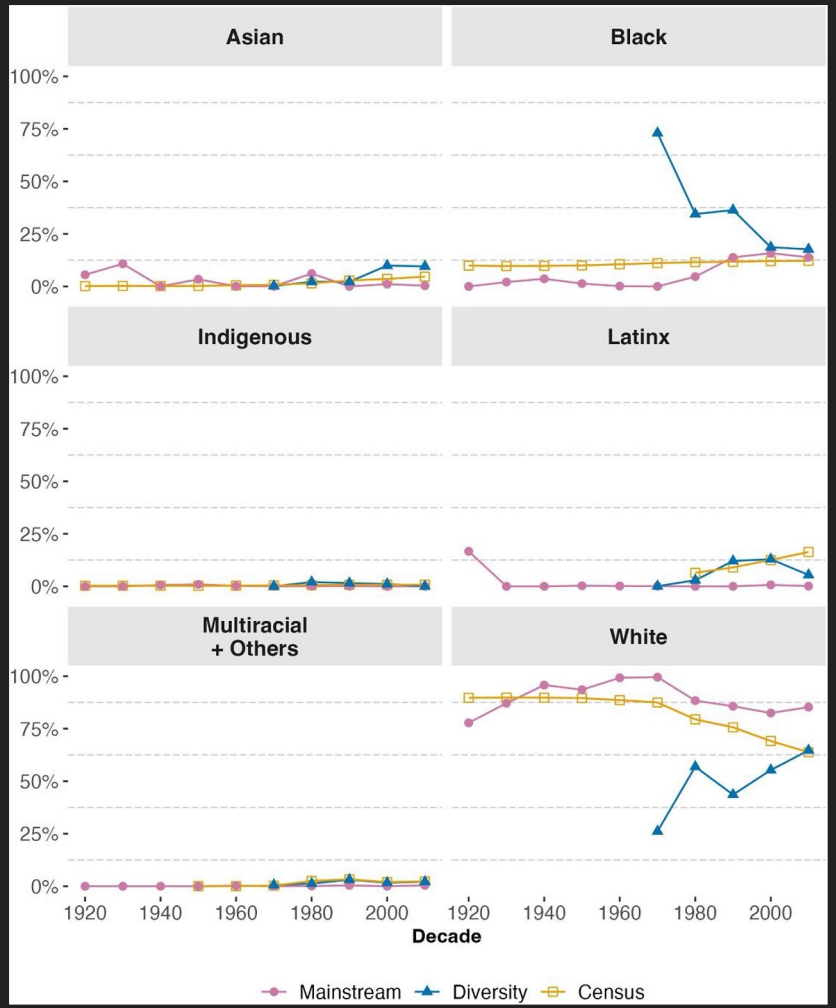
Skin Color

- Mainstream: lighter skin
- Children: lighter skin



Putative Race

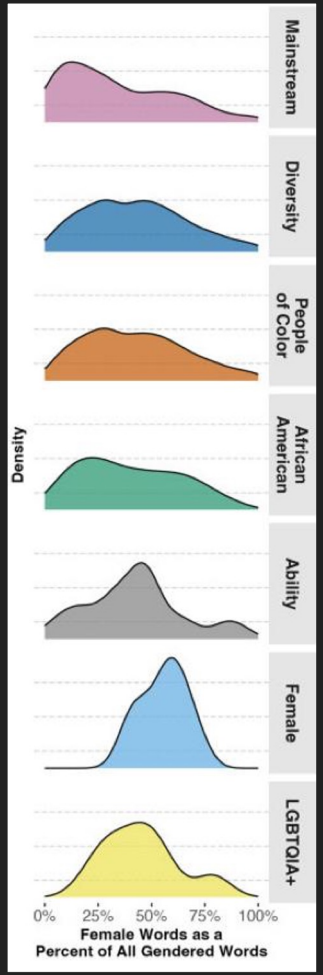
- Famous people
- African American 50% black
- Other collections 7 - 29% black
- Other groups highly underrepresented



GENDER

Analysis of the texts

- Between 34% and 45% overall
- Exception: 56 % in female category
- Improvement over time
- Famous figures: 85% men in Mainstream
- Still more male famous figures in in female category

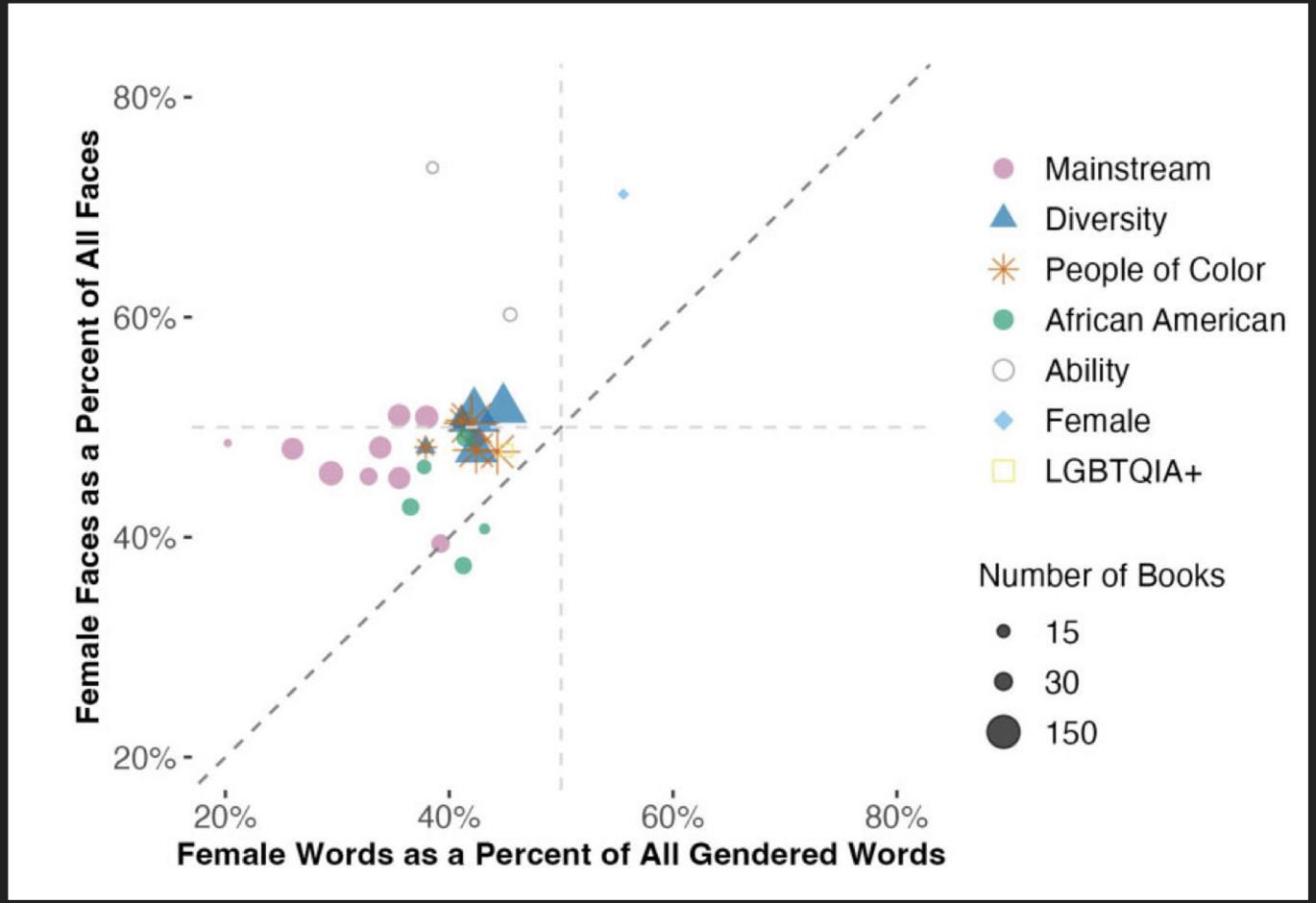


GENDER

Analysis of the images

- Overall less than 50%
- Exception: Female category 71%, Ability category 67%
- Consistent over time

Women are more likely to be seen than heard


80%
60%
40%
20%
Female Faces as a Percent of All Faces
20%
40%
60%
80%
Female Words as a Percent of All Gendered Words
Mainstream
Diversity
People of Color
African American
Ability
Female
LGBTQIA+
Number of Books
15
30
150

AGE

- 3%-19% images of children
- 17%-32% words referring to children
- Regardless of gender, adults are overrepresented (relative to U.S. population) in all categories

ECONOMIC AND SOCIAL FACTORS

Demand

- Utility from homophily
- Status quo bias

Supply

- Tyranny of the market
- Pricing-in of representation

ECONOMIC AND SOCIAL FACTORS

Utility from homophily

- Consumers with Daughters: 2% more images containing women, 3% more text containing women
- Consumers with Sons: symmetric purchases in terms of books with a lower proportion of female representation

- Black/Latinx people more likely to buy books containing characters of darker skin tone

- Adults buying children's could correlate to their overrepresentation

ECONOMIC AND SOCIAL FACTORS

Status quo bias

- School libraries reflect surrounding areas beliefs
- Purchases male focuses even though most purchasers are women
- People with daughters still purchase books containing less than 50% female words

Implication being that it is more important for sons to read male representing literature then it is for women to read female representing literature

ECONOMIC AND SOCIAL FACTORS

Tyranny of the market

- Overrepresentation of white famous figures (relative to white U.S. population)
- Library stock containing double the amount of mainstream books compared to the diverse books

Pricing-in of representation

- Diverse books priced 22% higher on average

Collection	Number of checkouts (1)	Mean checkouts per title (2)	Number of unique titles (3)	Mean library copies per title (4)
Panel A: Seattle Public Library Inventory and Checkouts				
Mainstream	388,357	991	392	14.0
Diversity	248,860	212	1,176	7.0
All other children's books	17,027,557	238	71,590	5.6
People of color	225,851	216	1,045	7.0
African American	37,367	217	172	8.3
Female	7,272	97	75	6.5
Ability	14,170	301	47	7.7
LGBTQIA+	8,295	251	33	9.3

Panel B: Average price and copies purchased in Numerator OmniPanel	Number of copies sold (1)	Mean book price (2)	Number of unique titles (3)	Mean copies sold per title (4)
Mainstream	40,854	$7.66	493	83
Diversity	35,553	$9.34	1,067	33
All other children's books	1,683,406	$7.42	97,866	17
People of color	26,899	$9.51	880	31
African American	9,081	$9.95	149	61
Female	4,892	$8.68	120	41
Ability	2,834	$8.70	55	52
LGBTQIA+	2,838	$9.07	34	83

HISTORICAL TRENDS AND REPRESENTATION

- Temporary change in representation after major historical events

- Positive significant relation between market share - of various racial groups and gender groups - and representation in children's books

CONCLUSION

Limitations

- Single component to larger societal process
- Not facturing in depiction of identities
- One of many child specific medias
- Disregard to nonbinary people and the multifacetedness of race

provides a complement to traditional “by-hand evaluation” and helps to improve future efforts to understand inequalities in representation and their consequences

DISCUSSION

Questions

- Why do you think children are more likely to be portrayed as white in comparison to adults?