County-level Algorithmic Audit of Racial Bias in Twitter’s Home Timeline

Luca Belli[1], Kyra Yee[1], Uthaipon Tantipongpipat[1], Aaron Gonzales[1], Kristian Lum[1] and Moritz Hardt[2]
[1]Twitter, [2]Max Planck Institute for Intelligent Systems, Tübingen

Research question:
Is the racial composition of a US county associated with higher or lower visibility on Twitter’s Home Timeline?

Methodology

- Divide users into promoted and demoted, based on their normalized impressions, i.e.
  normalized_impressions = total_unique_impressions / ((1 + total_tweets_produced) × (1 + num_followers))
- Assign each user to a US county (or county equivalent)
- In each county, study the relationship between racial composition and share of promoted users

Important!

- Unit of analysis is County, not users
- More research is needed to understand if effects translate to users
- Different counties have different usage patterns
- The best way to analyze for bias based on a characteristic is to have that information but best way to ensure appropriate use is to never collect that data at all.

Limitations

- User assignment to a county
- Data loss
- Level of granularity
- Disparities between user base and Census population
- Definition of Race
- Amplification from other product surfaces

Analysis 1: Effect of racial composition on amplification

Fit a least squares linear regression model

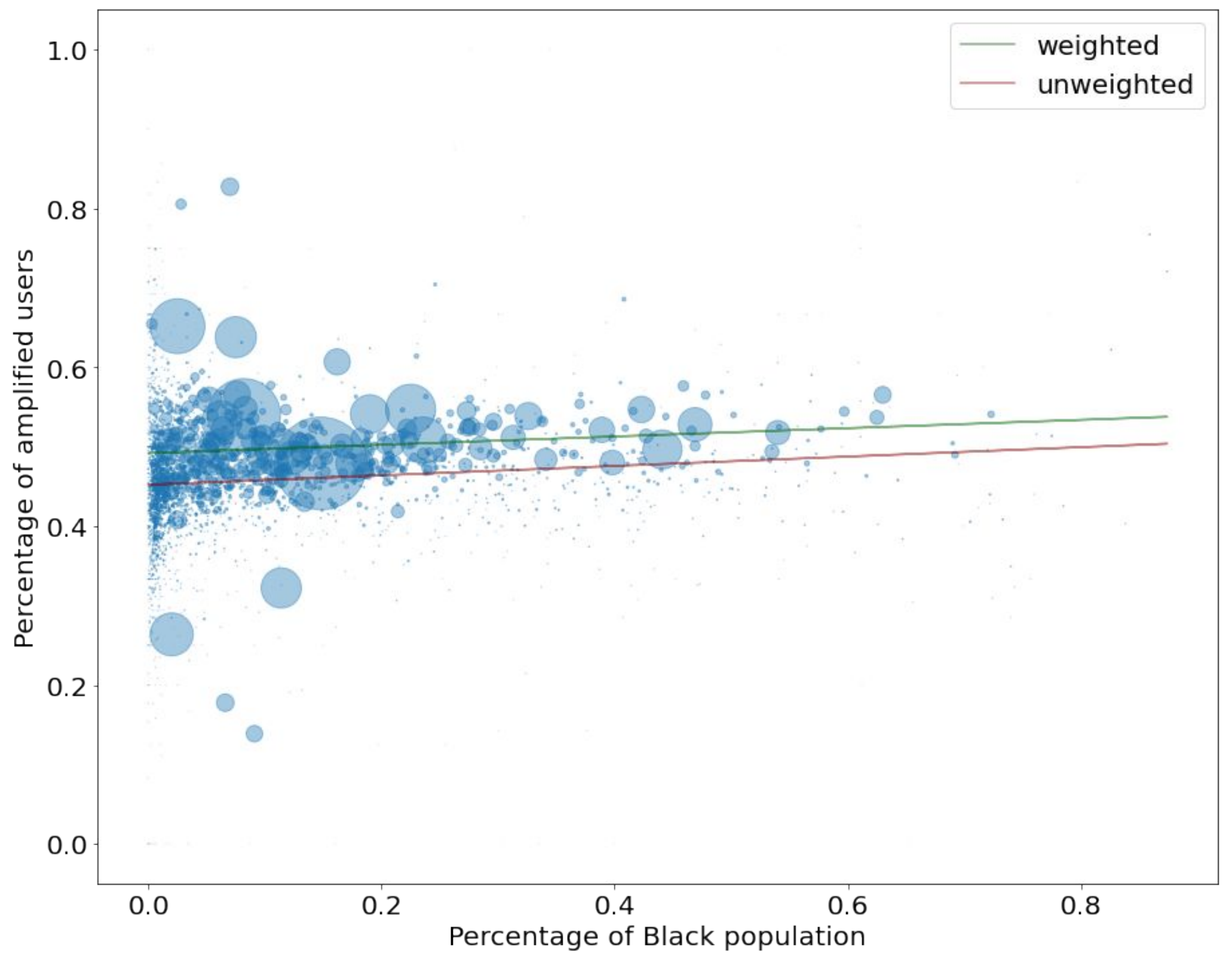
$$Y = \alpha + \beta X,$$

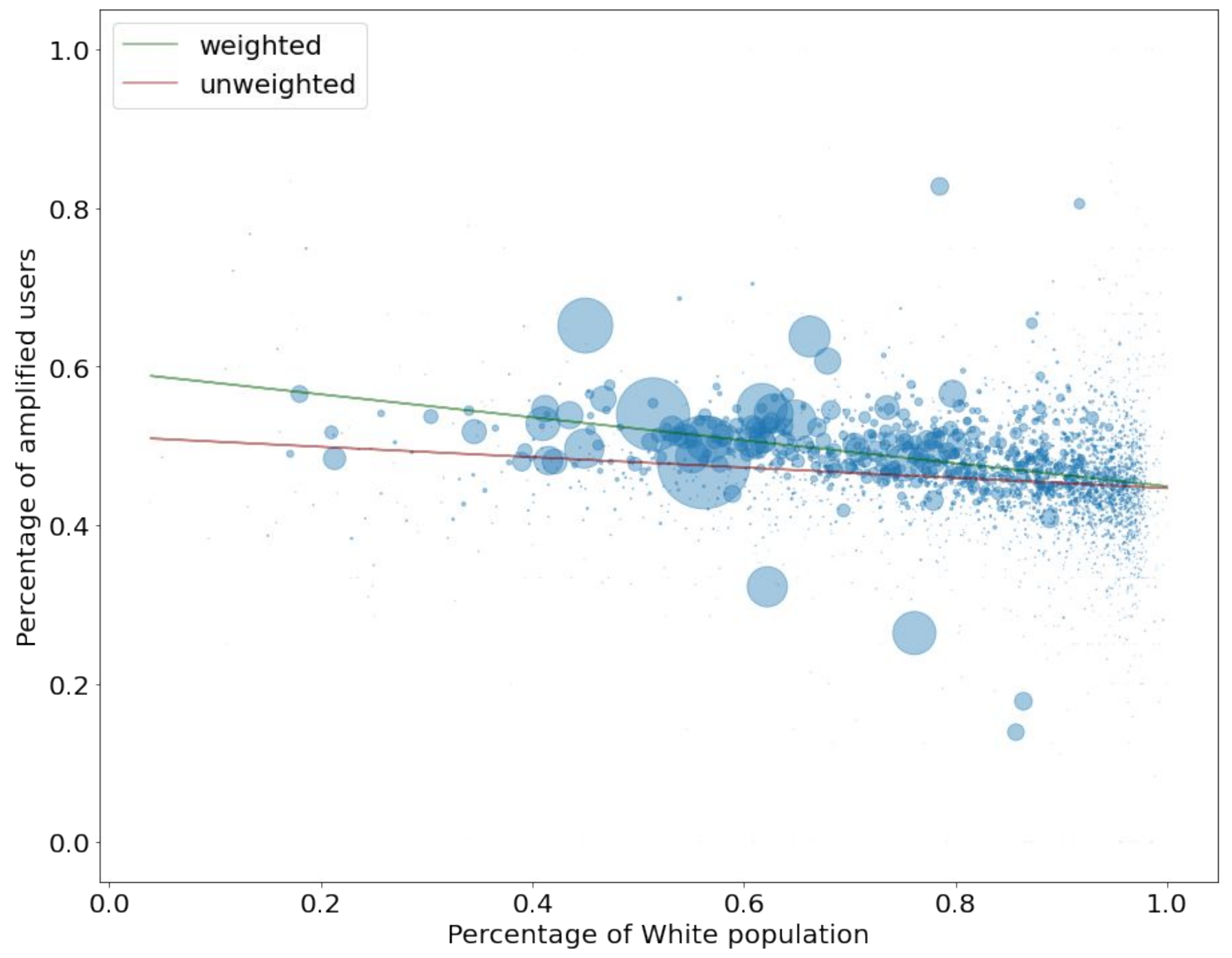
X = fraction of the county’s population in the given racial group
Y = share of amplified users in the county
Unit of analysis: county (i.e. each data point corresponds to one county).

β as an observational measure of bias. A positive coefficient indicates that fractional size of the group within a county is associated with a higher share of amplified users. A negative coefficient indicates the opposite.

Results

Independent variable	Statistic	Weighted	Unweighted
Percentage of Black population	Coefficient	0.0524	0.0593
	95% CI	[0.0322, 0.0726]	[0.0358, 0.0827]
	R^2	0.0083	0.0079
Percentage of White population	Coefficient	-0.1451	-0.0660
	95% CI	[-0.1609, -0.1294]	[-0.0862, -0.0458]
	R^2	0.0954	0.0131





Analysis 2: Distribution of amplified users by county

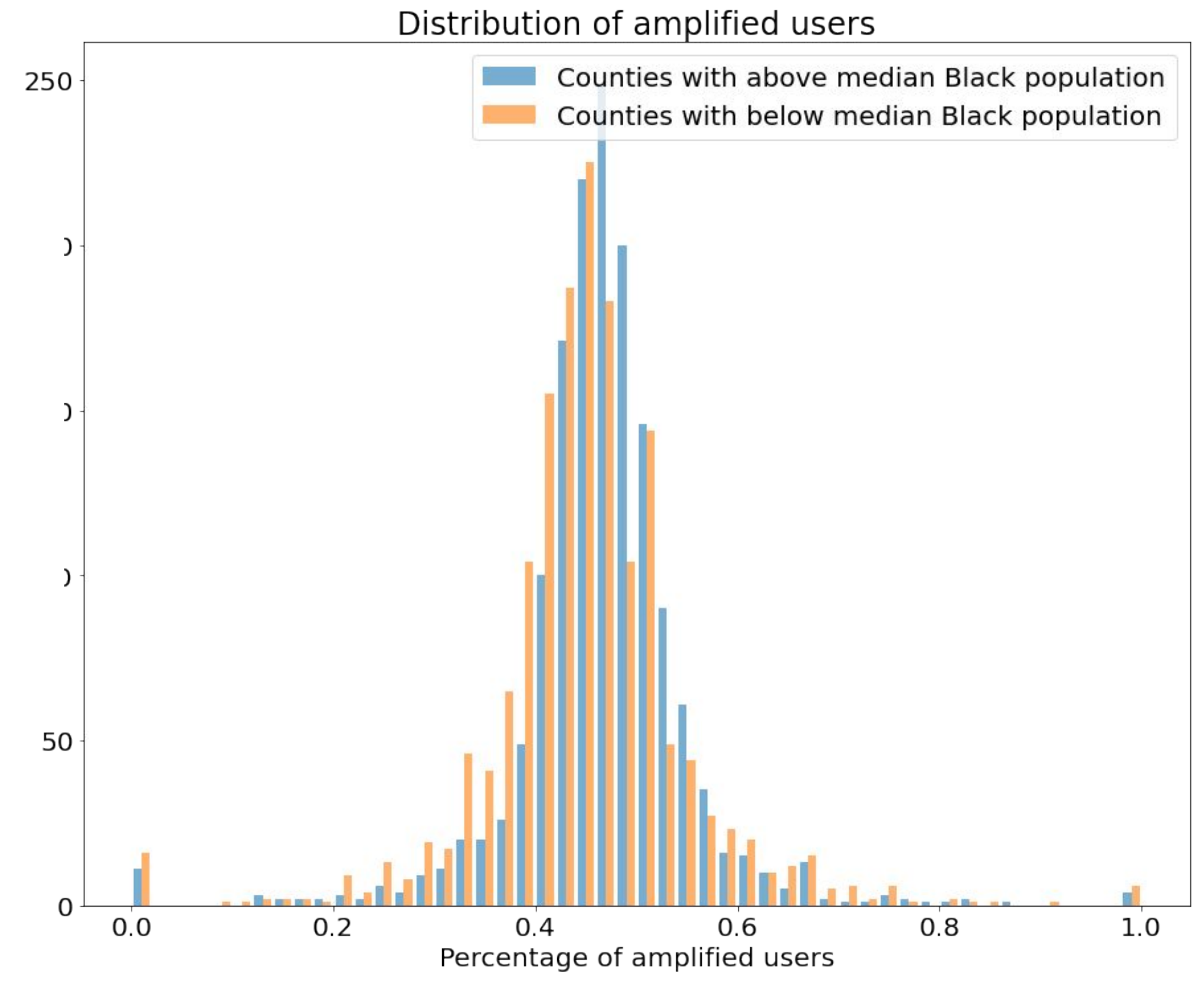
Divide the counties into two separate set: counties above and below the median of each racial group.
We then consider the histograms of the fraction of amplified users for each of the two cohorts.

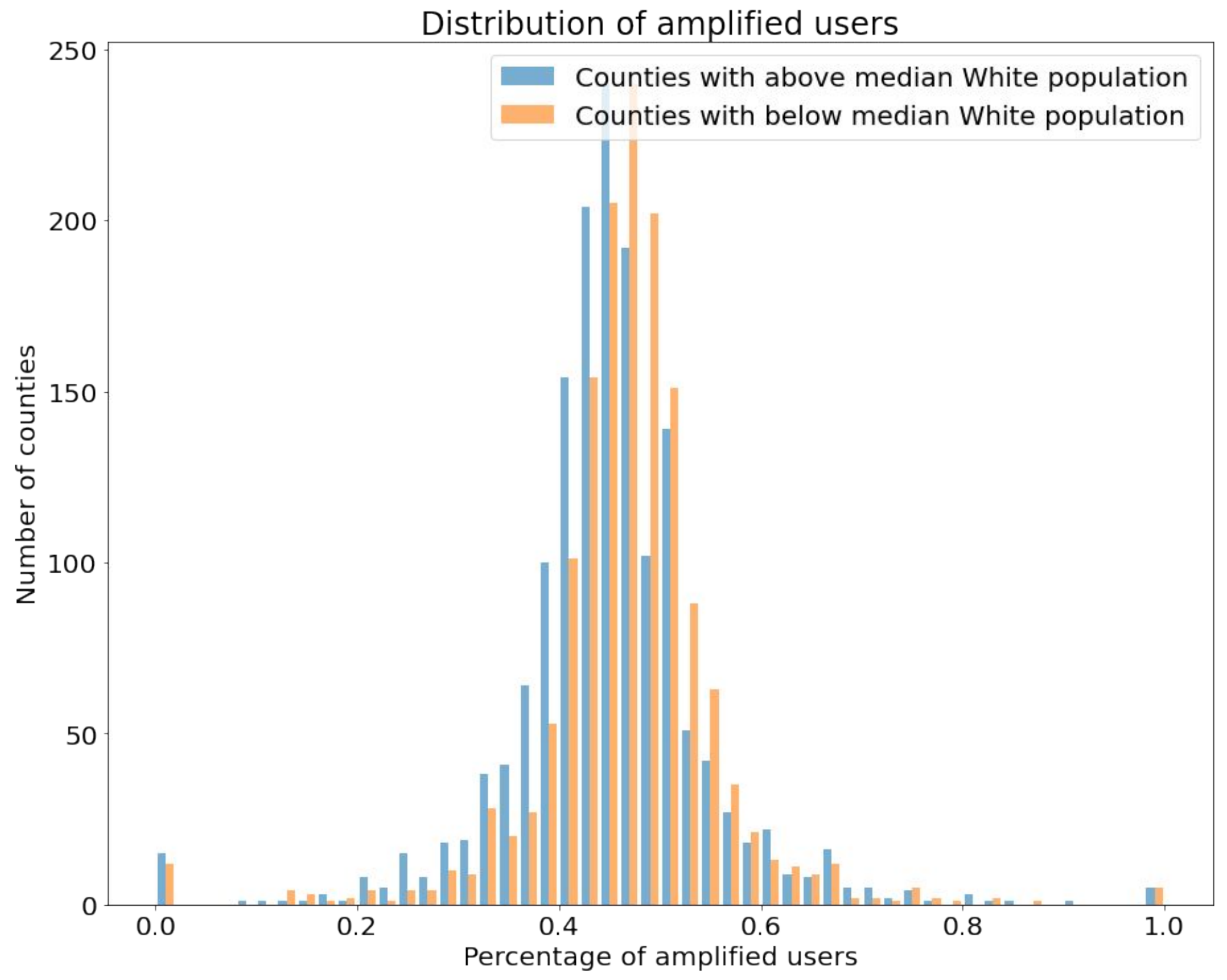
The difference between the two histograms is another indicator that racial composition of a county is associated with amplification.

Results

	Black		White	
	Above Median	Below Median	Above Median	Below Median
Mean	30.3800	31.6000	31.8200	30.1600
Variance	3659.1956	3028.6000	3276.5876	3415.0944
Std err	8.6416	7.8618	8.1773	8.3484

	Black	White
Total variation	0.172684	0.18839


Distribution of amplified users
Counties with above median Black population
Counties with below median Black population
Percentage of amplified users


Distribution of amplified users
Counties with above median White population
Counties with below median White population
Number of counties
Percentage of amplified users