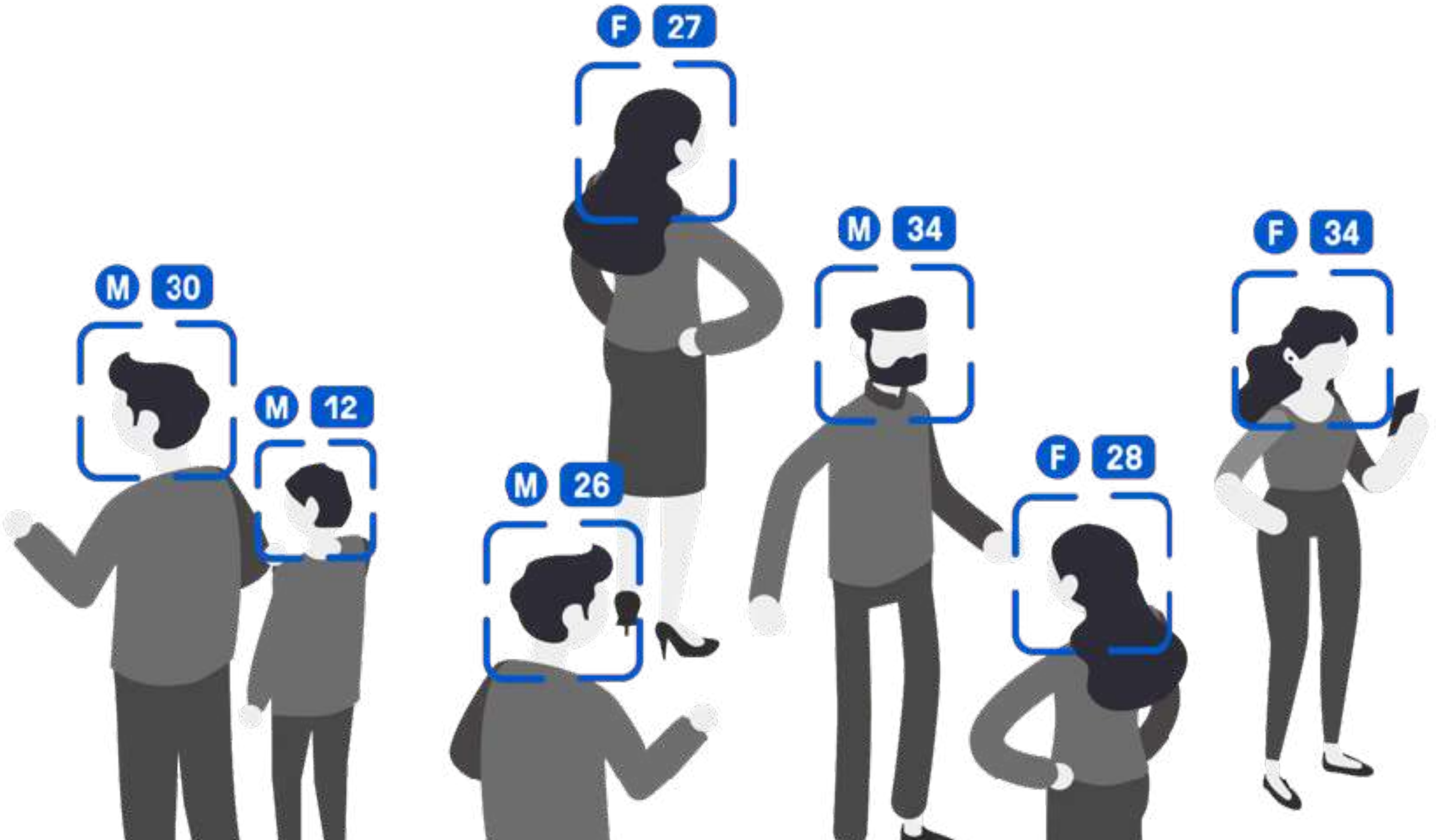
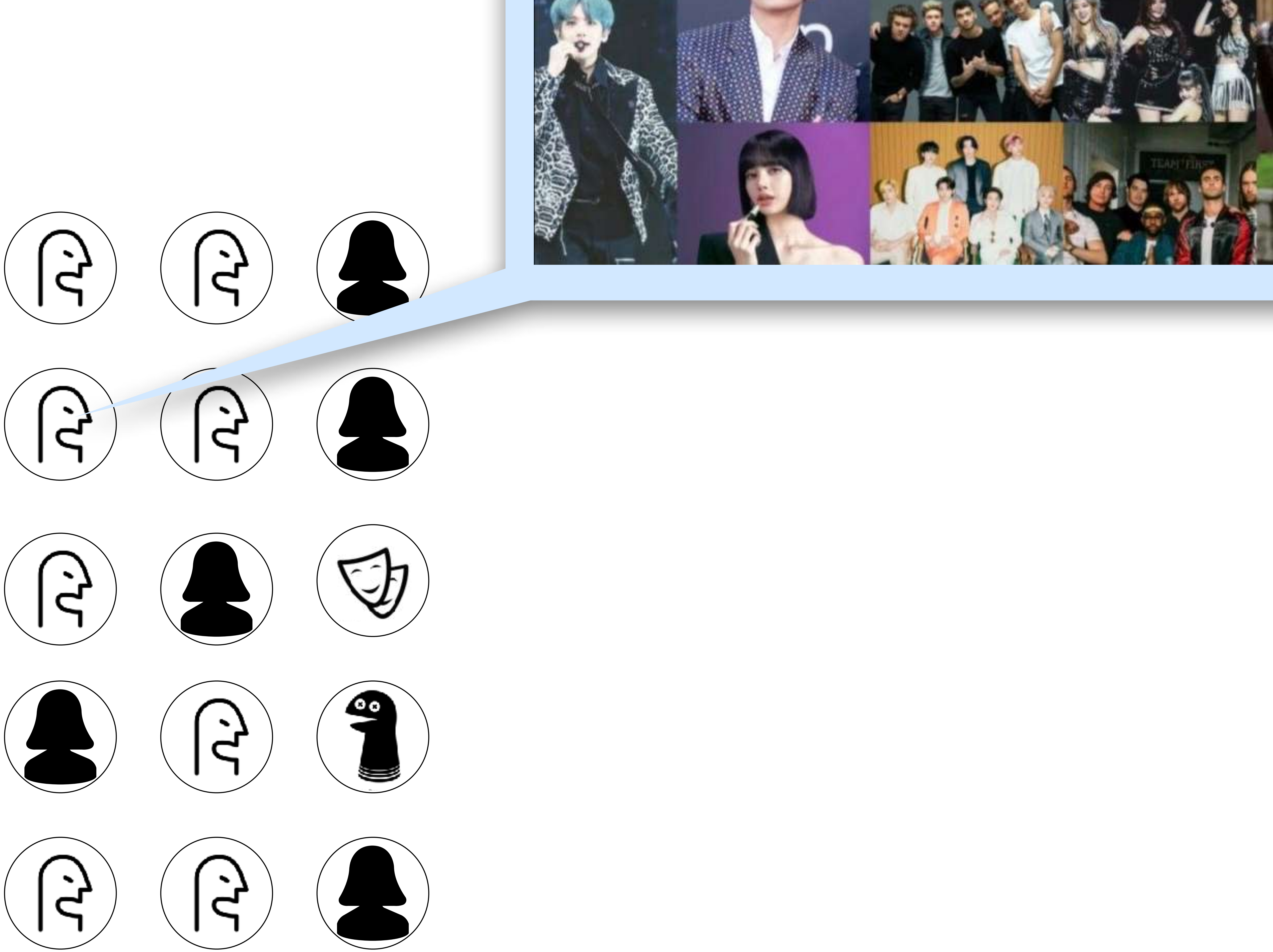
Characterizing Image Sharing Behaviors in US Politically Engaged, Random, and Demographic Audience Segments

Keng-Chi Chang · UC San Diego
Cody Buntain · University of Maryland





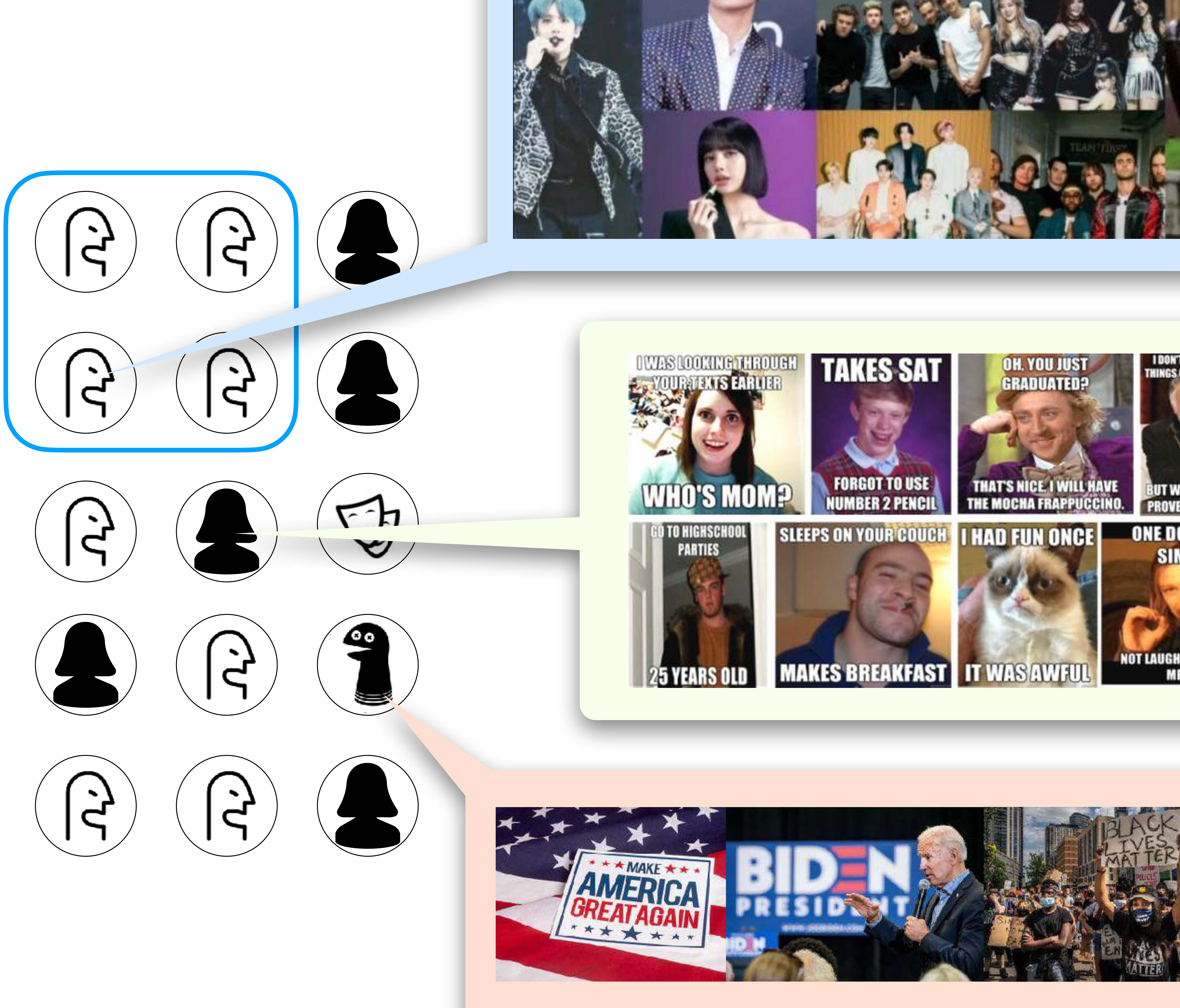

I WAS LOOKING THROUGH YOUR TEXTS EARLIER
WHO'S MOM?
TAKES SAT
FORGOT TO USE NUMBER 2 PENCIL
OH, YOU JUST GRADUATED?
THAT'S NICE. I WILL HAVE THE MOCHA FRAPPUCCINO.
GO TO HIGHSCHOOL PARTIES
25 YEARS OLD
SLEEPS ON YOUR COUCH
MAKES BREAKFAST
I HAD FUN ONCE
IT WAS AWFUL


I WAS LOOKING THROUGH YOUR TEXTS EARLIER
WHO'S MOM?
TAKES SAT
FORGOT TO USE NUMBER 2 PENCIL
OH, YOU JUST GRADUATED?
THAT'S NICE. I WILL HAVE THE MOCHA FRAPPUCCINO.
GO TO HIGHSCHOOL PARTIES
25 YEARS OLD
SLEEPS ON YOUR COUCH
MAKES BREAKFAST
I HAD FUN ONCE
IT WAS AWFUL
MAKE AMERICA GREAT AGAIN
BIDEN
BLACK LIVES MATTER

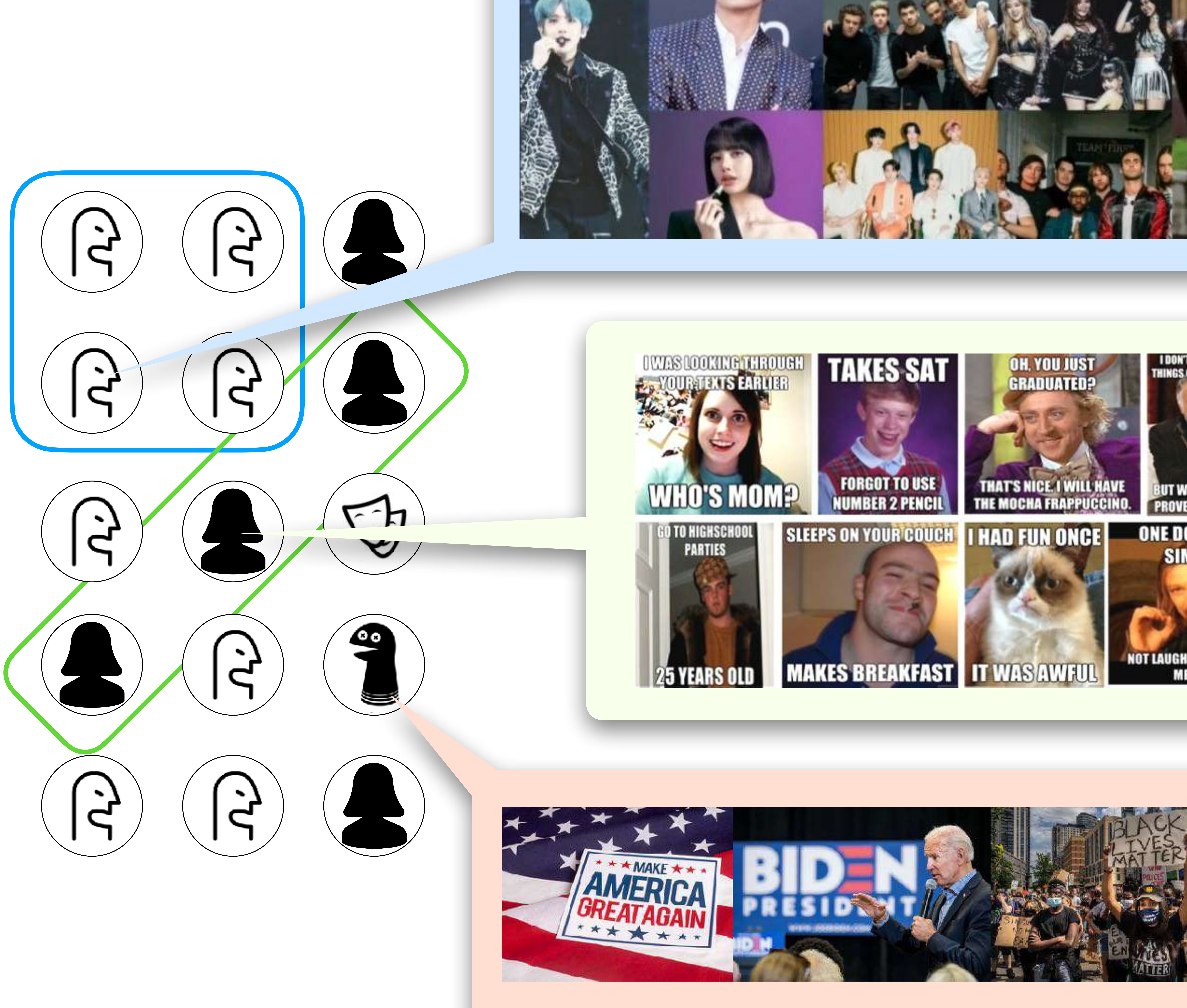
- People share **diverse imagery** on social media

- People share **diverse imagery** on social media
- **Demographics** are fundamental to understanding of public opinion

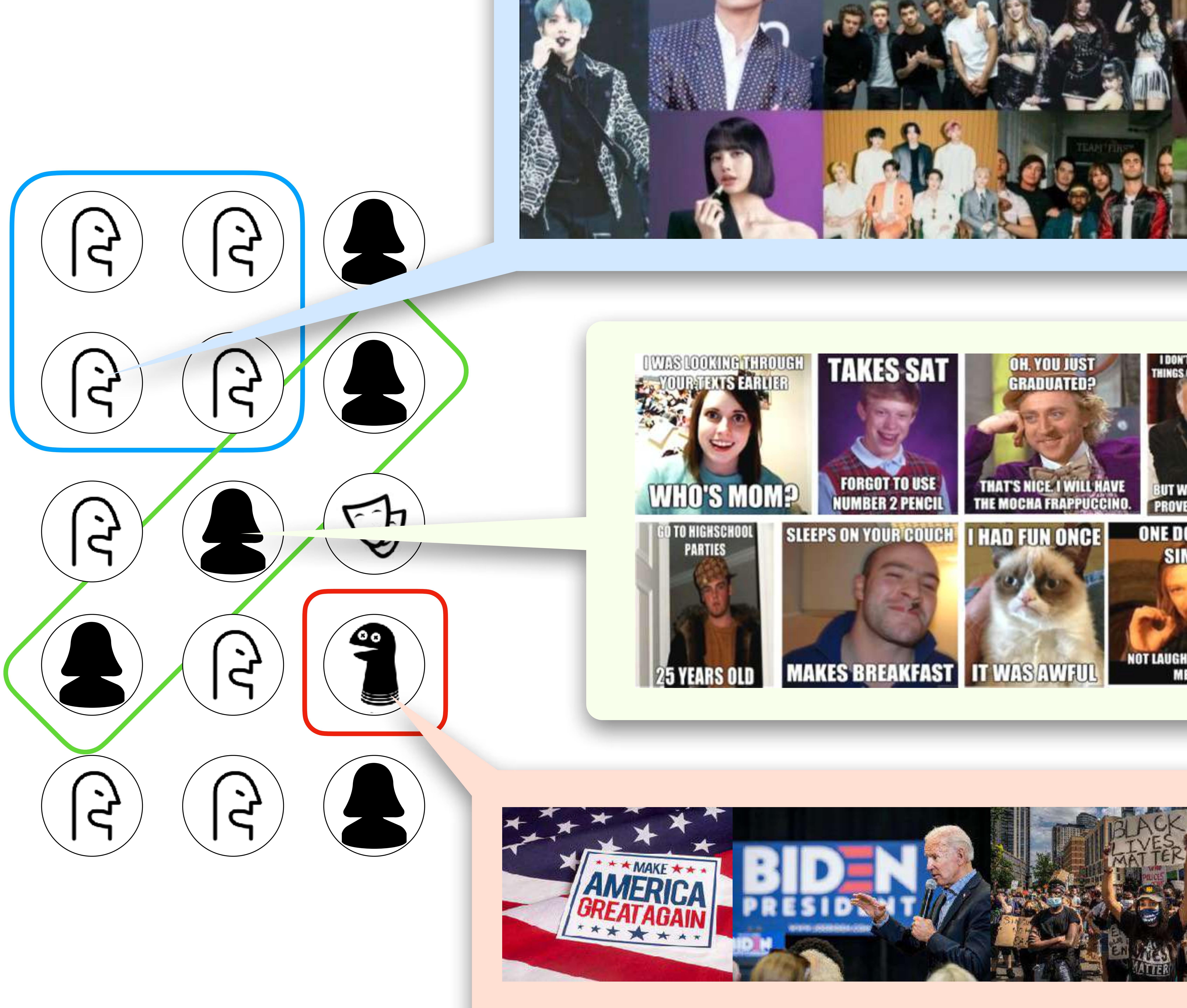
- People share **diverse imagery** on social media
- **Demographics** are fundamental to understanding of public opinion



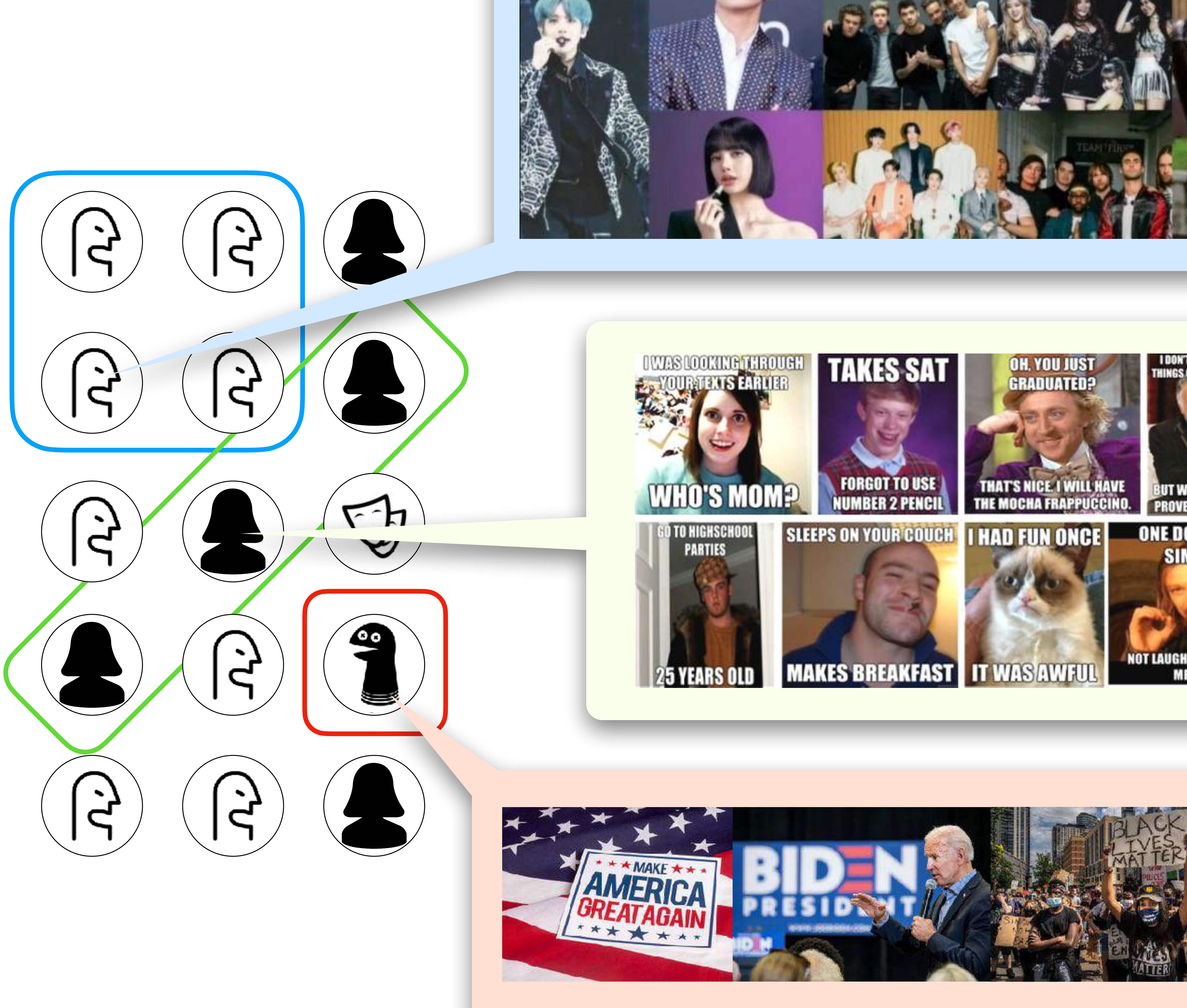
- People share **diverse imagery** on social media
- **Demographics** are fundamental to understanding of public opinion



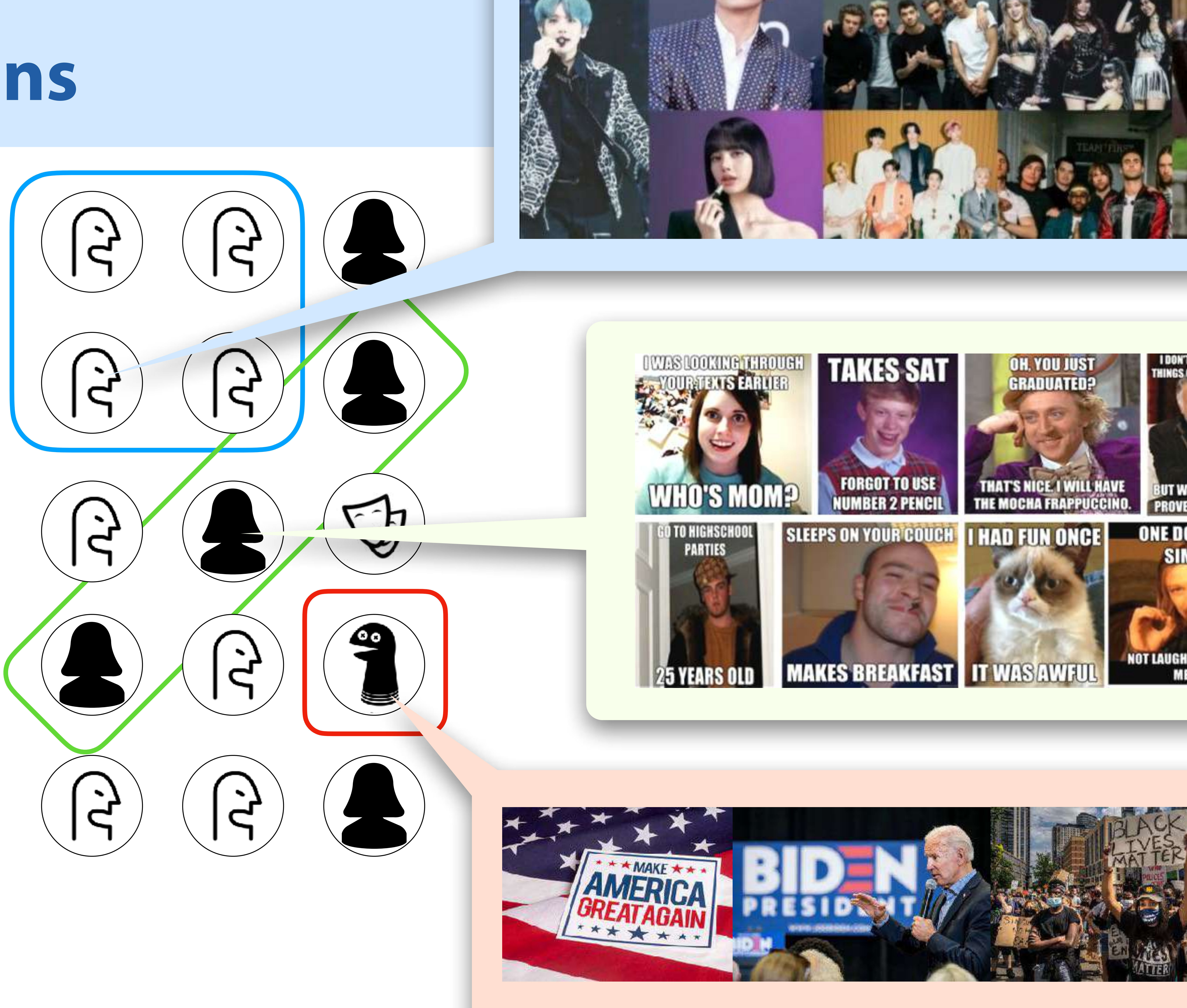
- People share **diverse imagery** on social media
- **Demographics** are fundamental to understanding of public opinion



- People share **diverse imagery** on social media
- **Demographics** are fundamental to understanding of public opinion
- **How to study their relationships?**

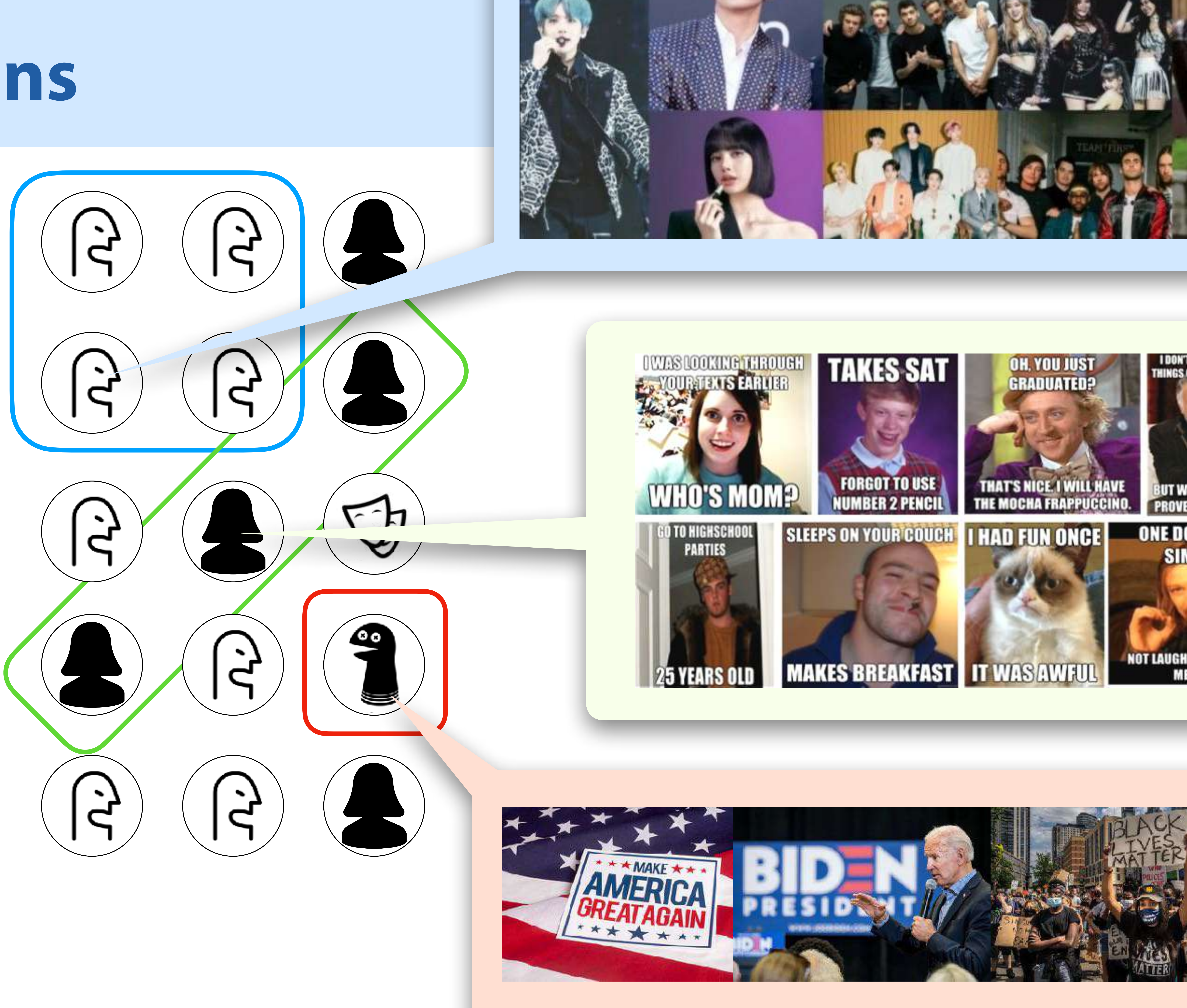


Research questions



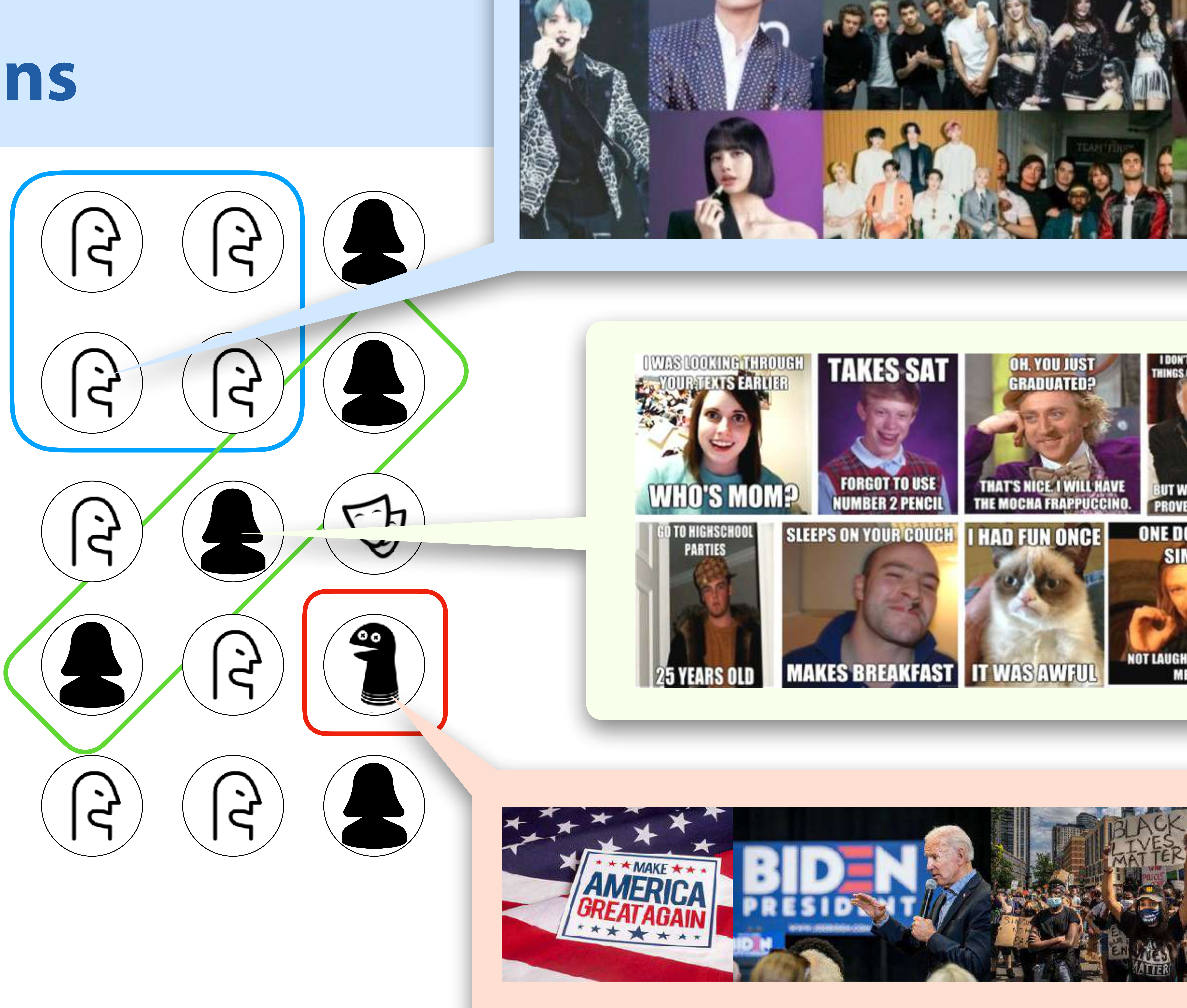
Research questions

- What types of image sharing behavior are predictive of the account's demographic backgrounds?

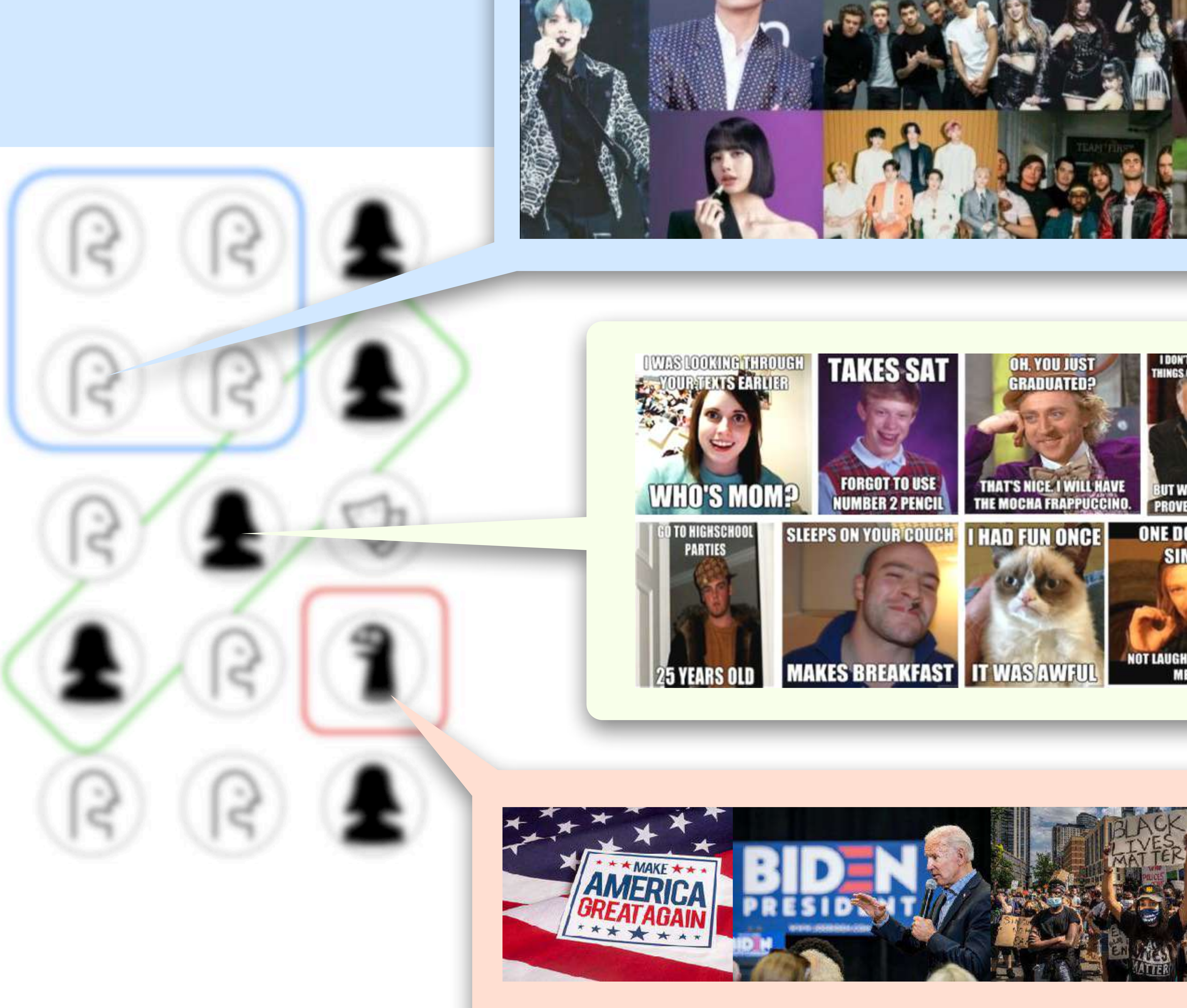


Research questions

- What types of image sharing behavior are predictive of the account's demographic backgrounds?
- Do politically engaged accounts share different types of imagery?



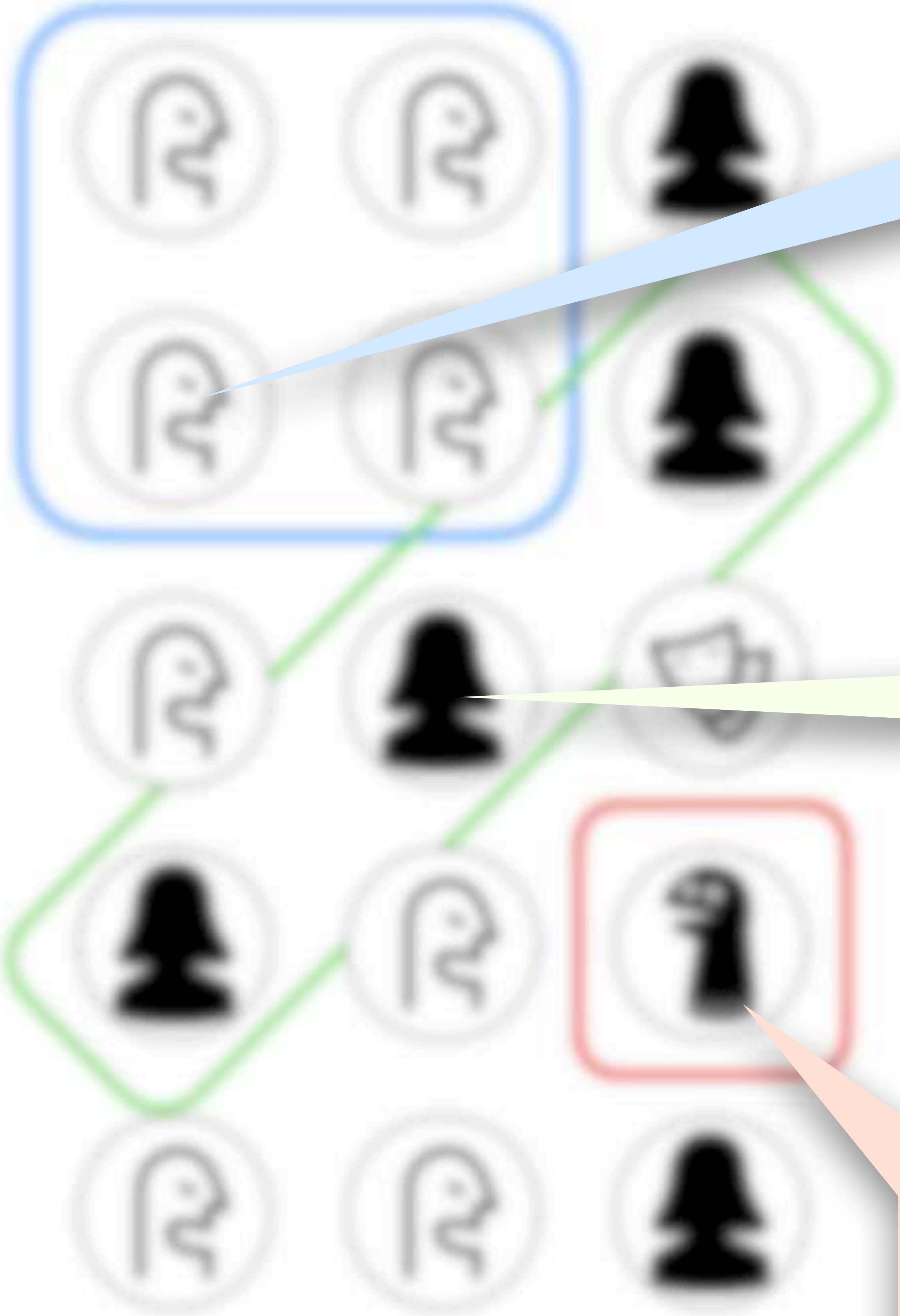
Challenge!!



I WAS LOOKING THROUGH YOUR TEXTS EARLIER
WHO'S MOM?
TAKES SAT
FORGOT TO USE NUMBER 2 PENCIL
OH, YOU JUST GRADUATED?
THAT'S NICE. I WILL HAVE THE MOCHA FRAPPUCCINO.
GO TO HIGHSCHOOL PARTIES
25 YEARS OLD
SLEEPS ON YOUR COUCH
MAKES BREAKFAST
I HAD FUN ONCE
IT WAS AWFUL
MAKE AMERICA GREAT AGAIN
BIDEN PRESIDENT
BLACK LIVES MATTER

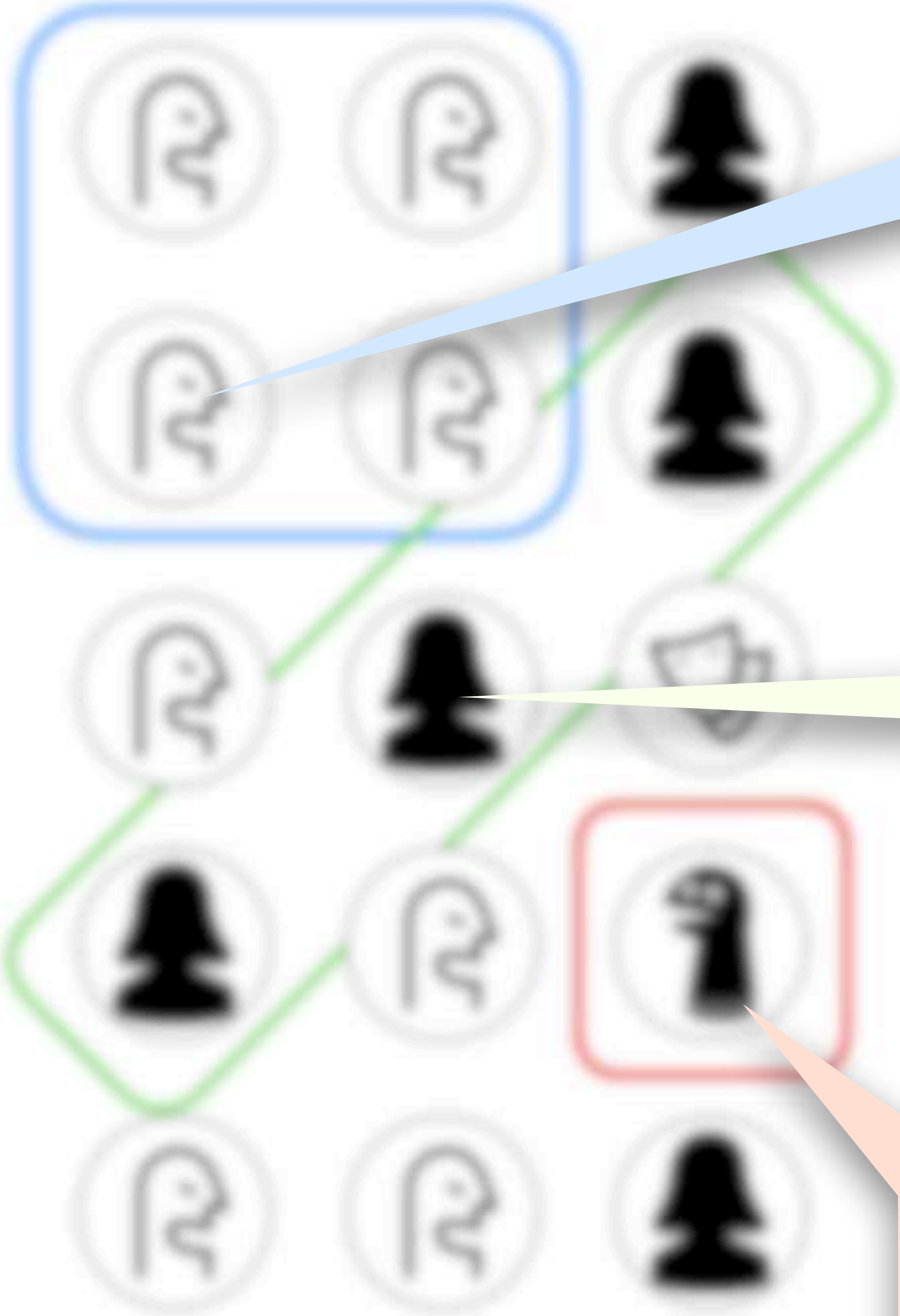
Challenge!!

- No ground truth for demographics!



Challenge!!

- No ground truth for demographics!
- Construct proxies for demographics from **profile pictures**



Demographic proxies from profile pictures

Demographic proxies from profile pictures

Demographic proxies from profile pictures

Demographic proxies from profile pictures

FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation

Kimmo Kärkkäinen
UCLA
kimmo@cs.ucla.edu

Jungseock Joo
UCLA
jjoo@comm.ucla.edu

Abstract

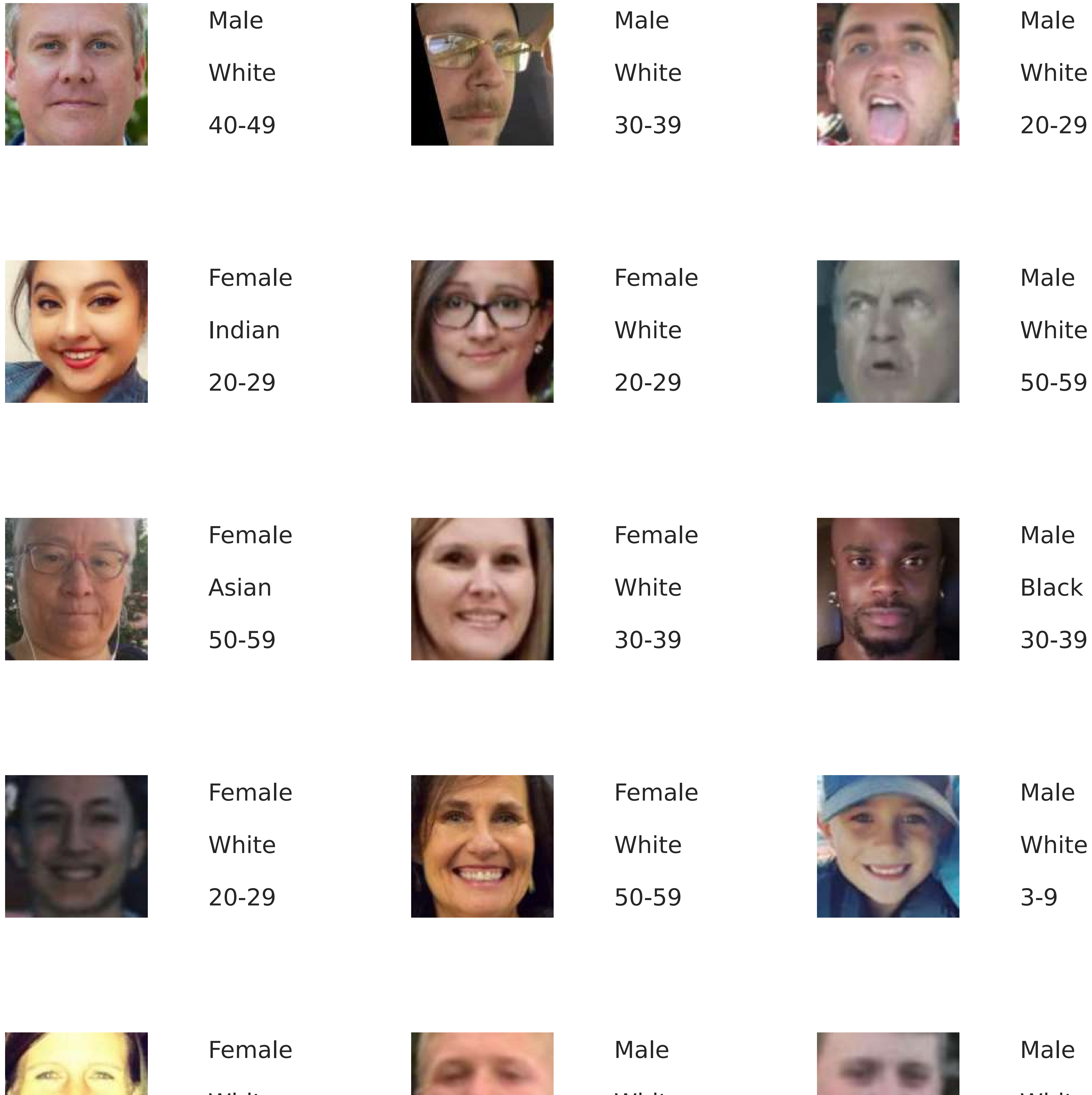
Existing public face image datasets are strongly biased toward Caucasian faces, and other races (e.g., Latino) are significantly underrepresented. The models trained from such datasets suffer from inconsistent classification accuracy, which limits the applicability of face analytic systems to non-White race groups. To mitigate the race bias problem in these datasets, we constructed a novel face image dataset containing 108,501 images which is balanced on race. We define 7 race groups: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino. Images were collected from the YFCC-100M Flickr dataset and labeled with race, gender, and age groups. Evaluations were performed on existing face attribute datasets as well as novel image datasets to measure the generalization performance. We find that the model trained from our dataset is substantially more accurate on novel datasets and the accuracy is consistent across race and gender groups. We also compare several commercial computer vision APIs and report their balanced accuracy across gender, race, and age groups. Our code, data, and models are available at

(around 80%), e.g. White, compared to "darker" faces, e.g. Black [40]. This means the model may not apply to some subpopulations and its results may not be compared across different groups without calibration. Biased data will produce biased models trained from it. This will raise ethical concerns about fairness of automated systems, which has emerged as a critical topic of study in the recent machine learning and AI literature [16, 11].

For example, several commercial computer vision systems (Microsoft, IBM, Face++) have been criticized due to their asymmetric accuracy across sub-demographics in recent studies [7, 44]. These studies found that the commercial face gender classification systems all perform better on male and on light faces. This can be caused by the biases in their training data. Various unwanted biases in image datasets can easily occur due to biased selection, capture, and negative sets [60]. Most public large scale face datasets have been collected from popular online media – newspapers, Wikipedia, or web search– and these platforms are more frequently used by or showing White people.

To mitigate the race bias in the existing face datasets, we propose a novel face dataset with an emphasis on balanced

Demographic proxies from profile pictures



**FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age
for Bias Measurement and Mitigation**

Kimmo Kärkkäinen
UCLA
kimmo@cs.ucla.edu

Jungseock Joo
UCLA
jjoo@comm.ucla.edu

Abstract

Existing public face image datasets are strongly biased toward Caucasian faces, and other races (e.g., Latino) are significantly underrepresented. The models trained from such datasets suffer from inconsistent classification accuracy, which limits the applicability of face analytic systems to non-White race groups. To mitigate the race bias problem in these datasets, we constructed a novel face image dataset containing 108,501 images which is balanced on race. We define 7 race groups: White, Black, Indian, East Asian, Southeast Asian, Middle Eastern, and Latino. Images were collected from the YFCC-100M Flickr dataset and labeled with race, gender, and age groups. Evaluations were performed on existing face attribute datasets as well as novel image datasets to measure the generalization performance. We find that the model trained from our dataset is substantially more accurate on novel datasets and the accuracy is consistent across race and gender groups. We also compare several commercial computer vision APIs and report their balanced accuracy across gender, race, and age groups. Our code, data, and models are available at

(around 80%), e.g. White, compared to "darker" fa
Black [40]. This means the model may not apply
subpopulations and its results may not be compare
different groups without calibration. Biased data w
duce biased models trained from it. This will rais
concerns about fairness of automated systems, w
emerged as a critical topic of study in the recent
learning and AI literature [16, 11].

For example, several commercial computer vis
tems (Microsoft, IBM, Face++) have been criticize
their asymmetric accuracy across sub-demographi
cent studies [7, 44]. These studies found that the c
cial face gender classification systems all perform b
male and on light faces. This can be caused by th
in their training data. Various unwanted biases i
datasets can easily occur due to biased selection,
and negative sets [60]. Most public large scale face
have been collected from popular online media –
pers, Wikipedia, or web search– and these platfo
more frequently used by or showing White people.

To mitigate the race bias in the existing face data
propose a novel face dataset with an emphasis on b

Data: Twitter timeline

Data: Twitter timeline

- **Random**: Random set of 5,000 accounts geolocated to the United States

Data: Twitter timeline

- **Random**: Random set of 5,000 accounts geolocated to the United States

- **Political**: Follow at least 5 US political Twitter accounts (e.g. Senate, House, Governors)

Data: Twitter timeline

- **Random**: Random set of 5,000 accounts geolocated to the United States
- **Political**: Follow at least 5 US political Twitter accounts (e.g. Senate, House, Governors)
- Collect shared images from their timeline

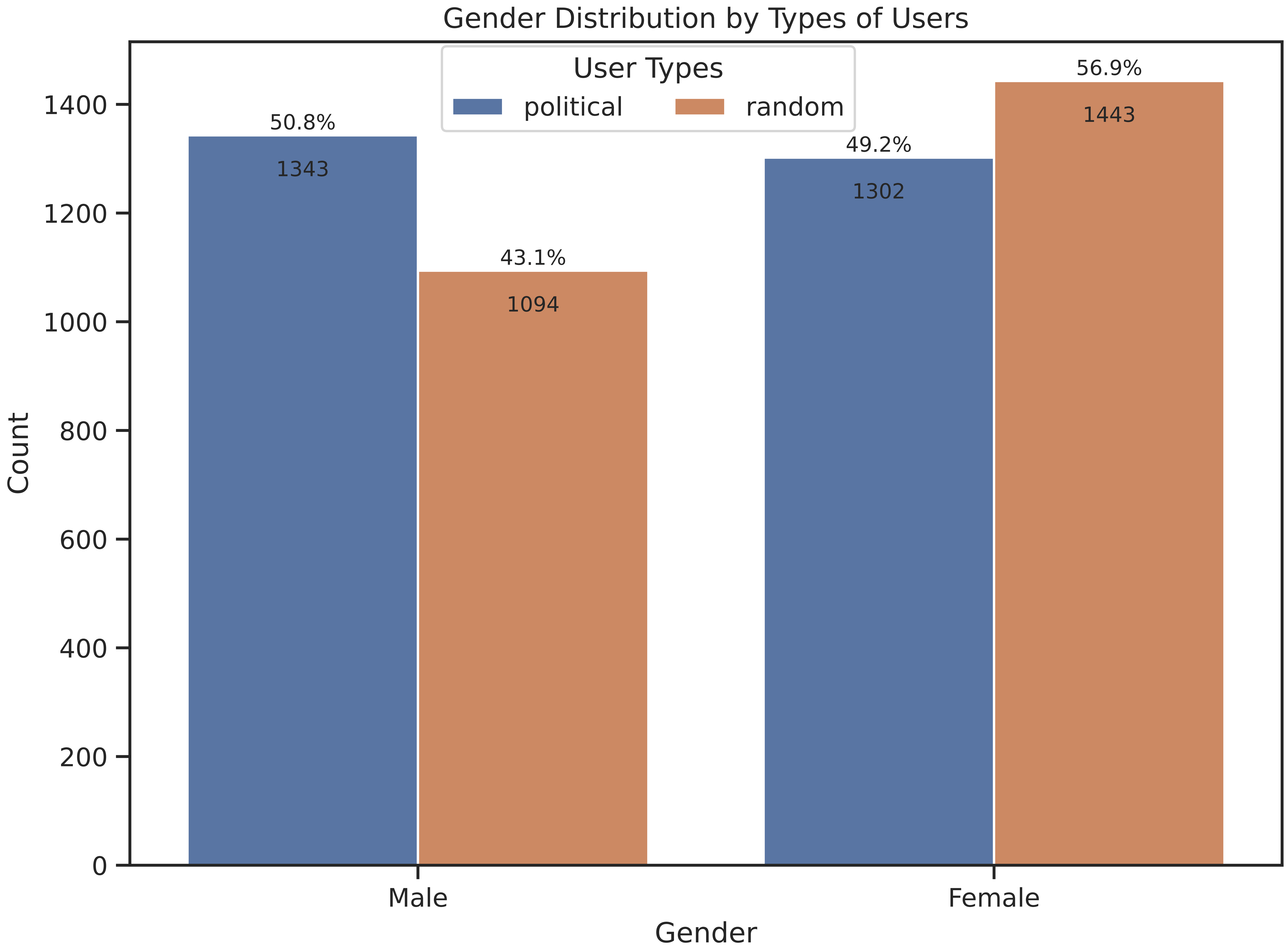
Data: Twitter timeline

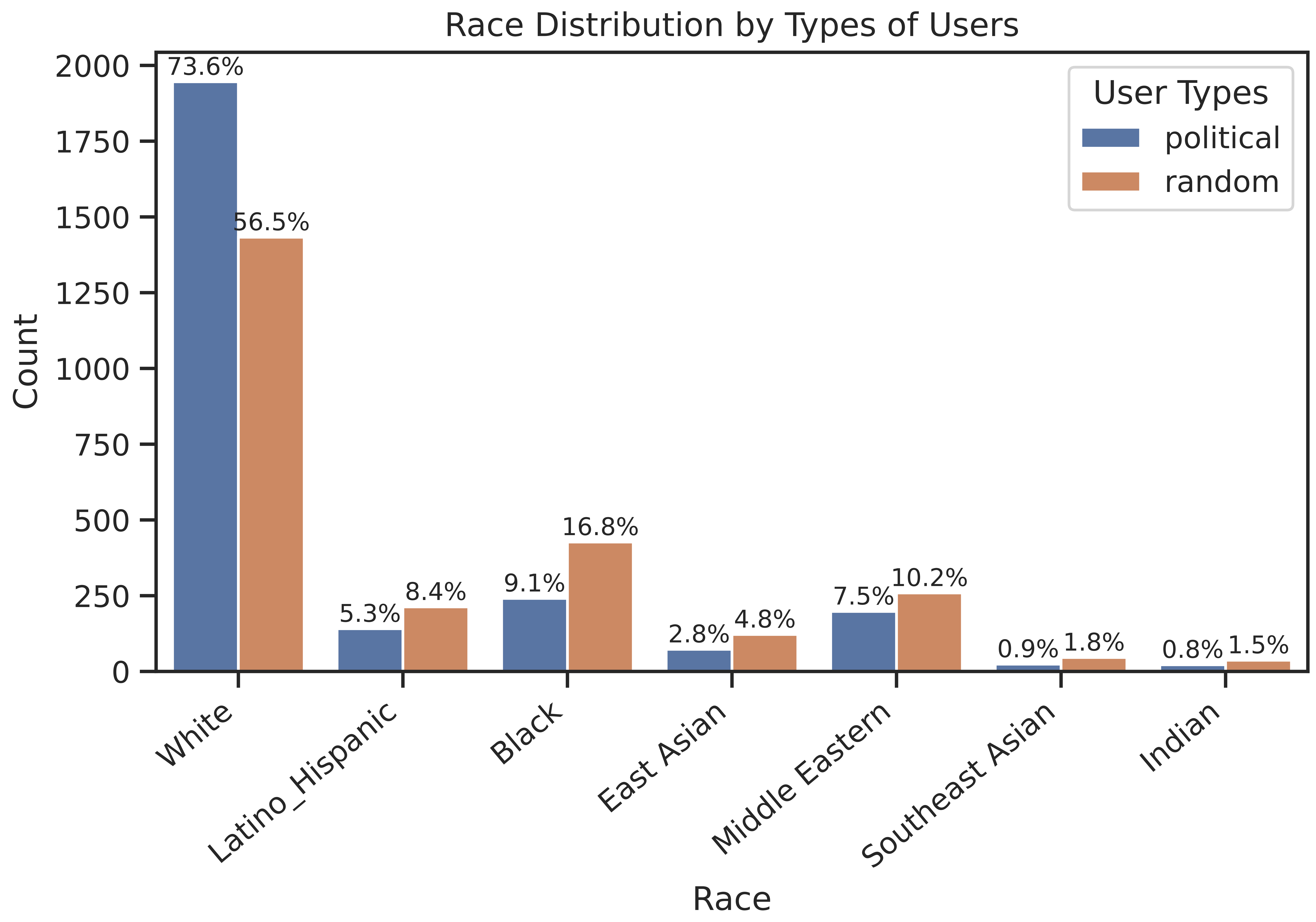
- **Random**: Random set of 5,000 accounts geolocated to the United States

- **Political**: Follow at least 5 US political Twitter accounts (e.g. Senate, House, Governors)

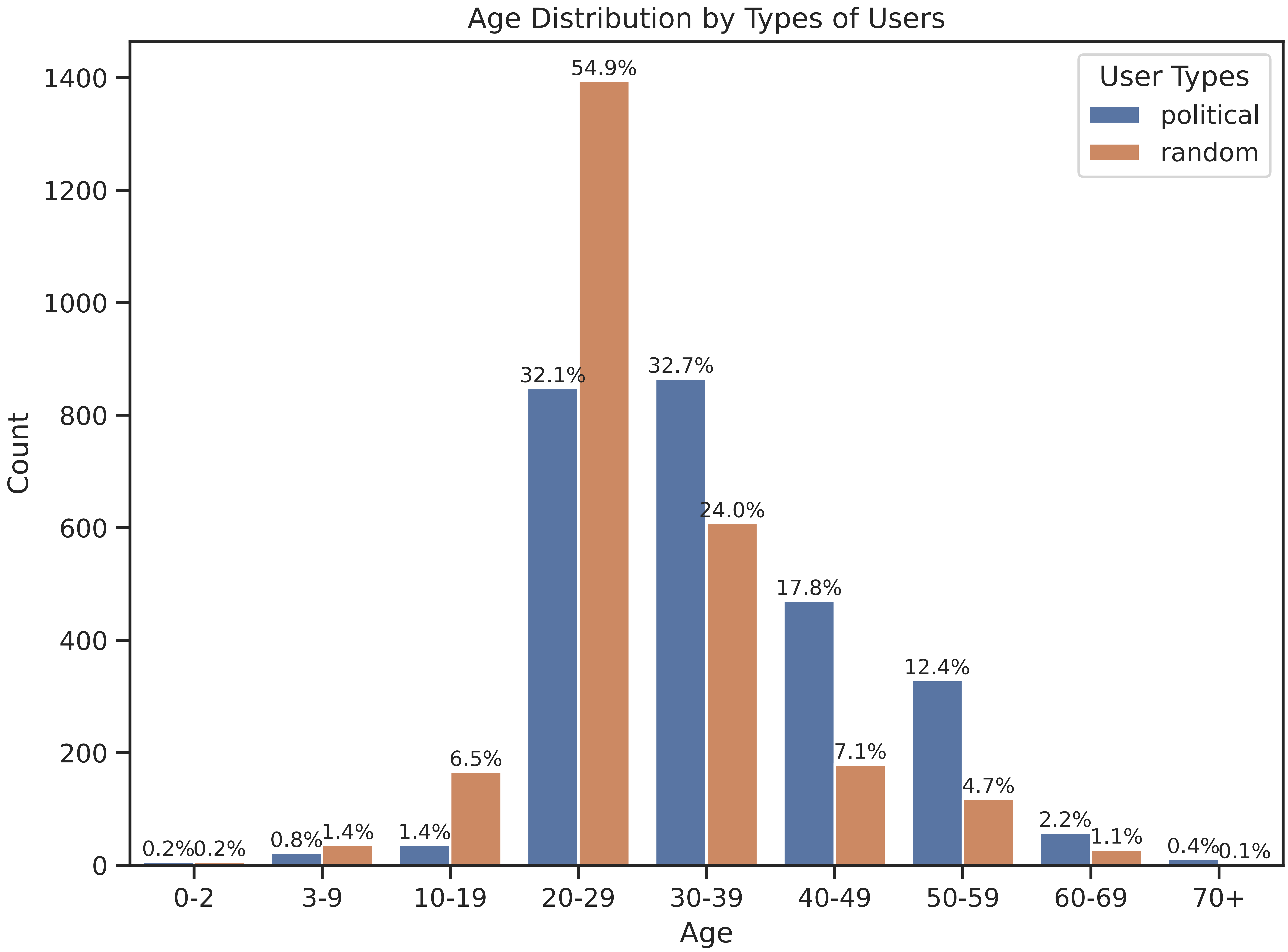
- Collect shared images from their timeline

Sample	Random	Political
# users	5,000	5,000
# tweets & retweets	31,038,705	35,932,231
% tweets	40%	37%
% retweets	60%	63%
% has image	18%	14%


Gender Distribution by Types of Users
User Types
political
random
50.8%
1343
43.1%
1094
49.2%
1302
56.9%
1443
Count
1400
1200
1000
800
600
400
200
0
Male
Female
Gender



Race Distribution by Types of Users
User Types
political
random
Count
Race
2000
1750
1500
1250
1000
750
500
250
0
73.6%
56.5%
5.3%
8.4%
9.1%
16.8%
2.8%
4.8%
7.5%
10.2%
0.9%
1.8%
0.8%
1.5%
White
Latino_Hispanic
Black
East Asian
Middle Eastern
Southeast Asian
Indian

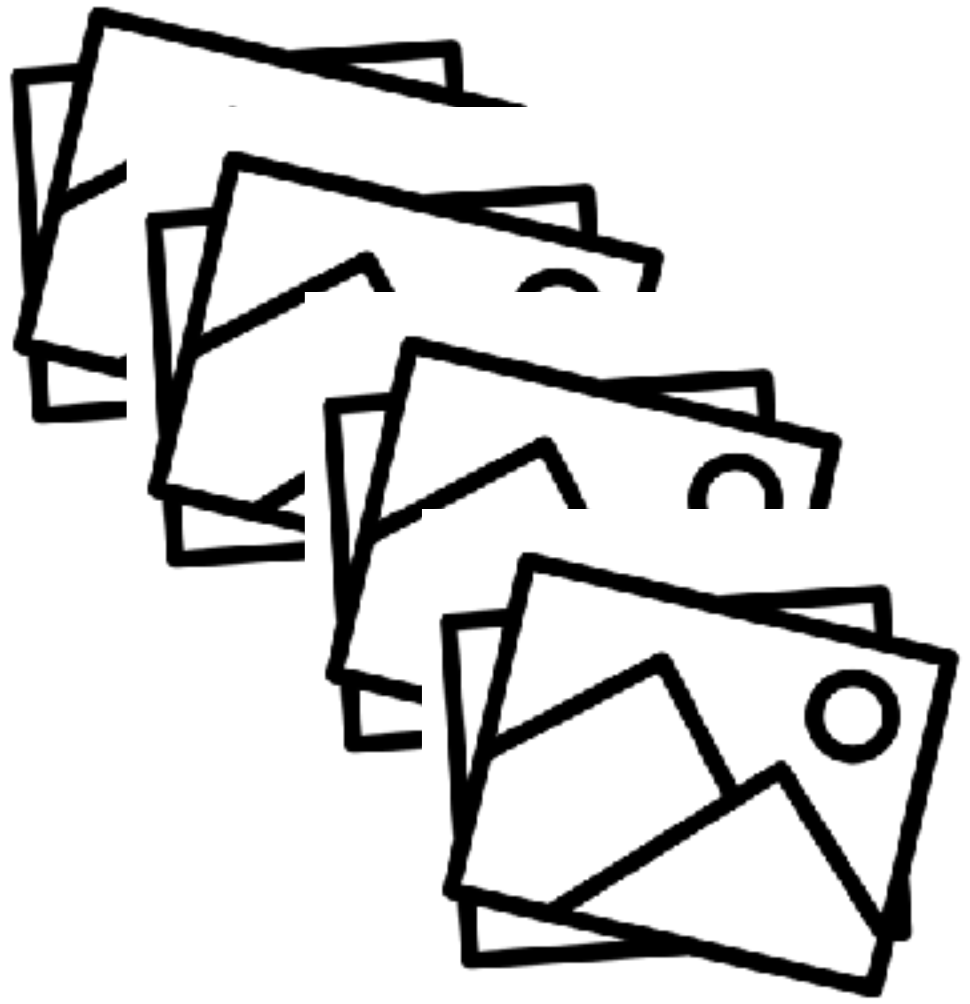


Age Distribution by Types of Users
User Types
political
random
Count
Age
0
200
400
600
800
1000
1200
1400
0-2
3-9
10-19
20-29
30-39
40-49
50-59
60-69
70+
0.2%
0.2%
0.8%
1.4%
1.4%
6.5%
32.1%
54.9%
32.7%
24.0%
17.8%
7.1%
12.4%
4.7%
2.2%
1.1%
0.4%
0.1%

Identify visually-similar imagery

Identify visually-similar imagery

10M images



Identify visually-similar imagery

10M images

Extract visual features

ResNet50

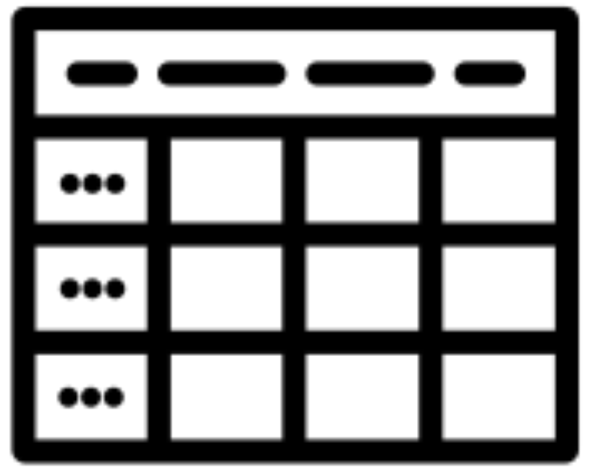
Identify visually-similar imagery

10M images

Extract visual features

ResNet50

10M x 2048-dim embedding



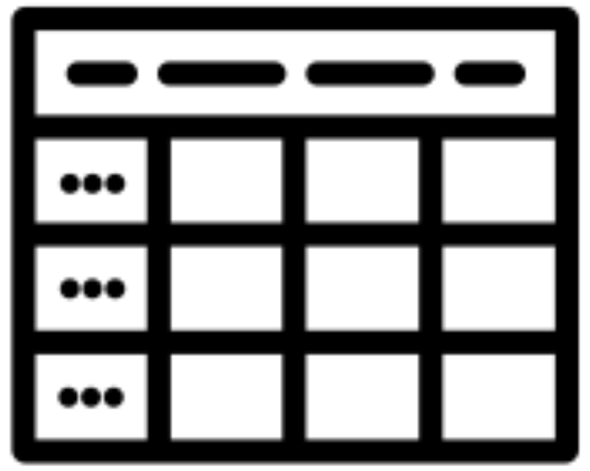
Identify visually-similar imagery

10M images

Extract visual features

ResNet50

10M x 2048-dim embedding



Cluster

K-means

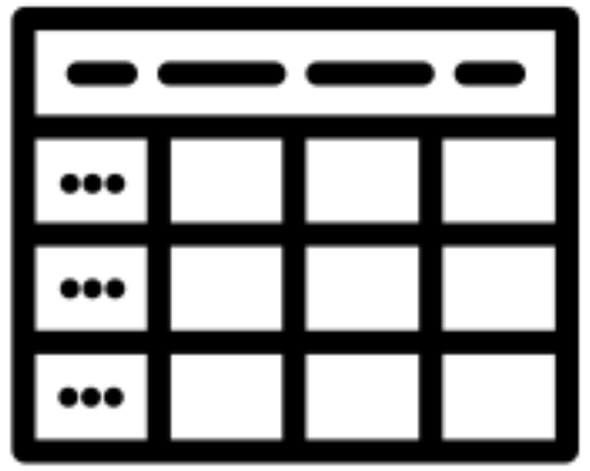
Identify visually-similar imagery

10M images

Extract visual features

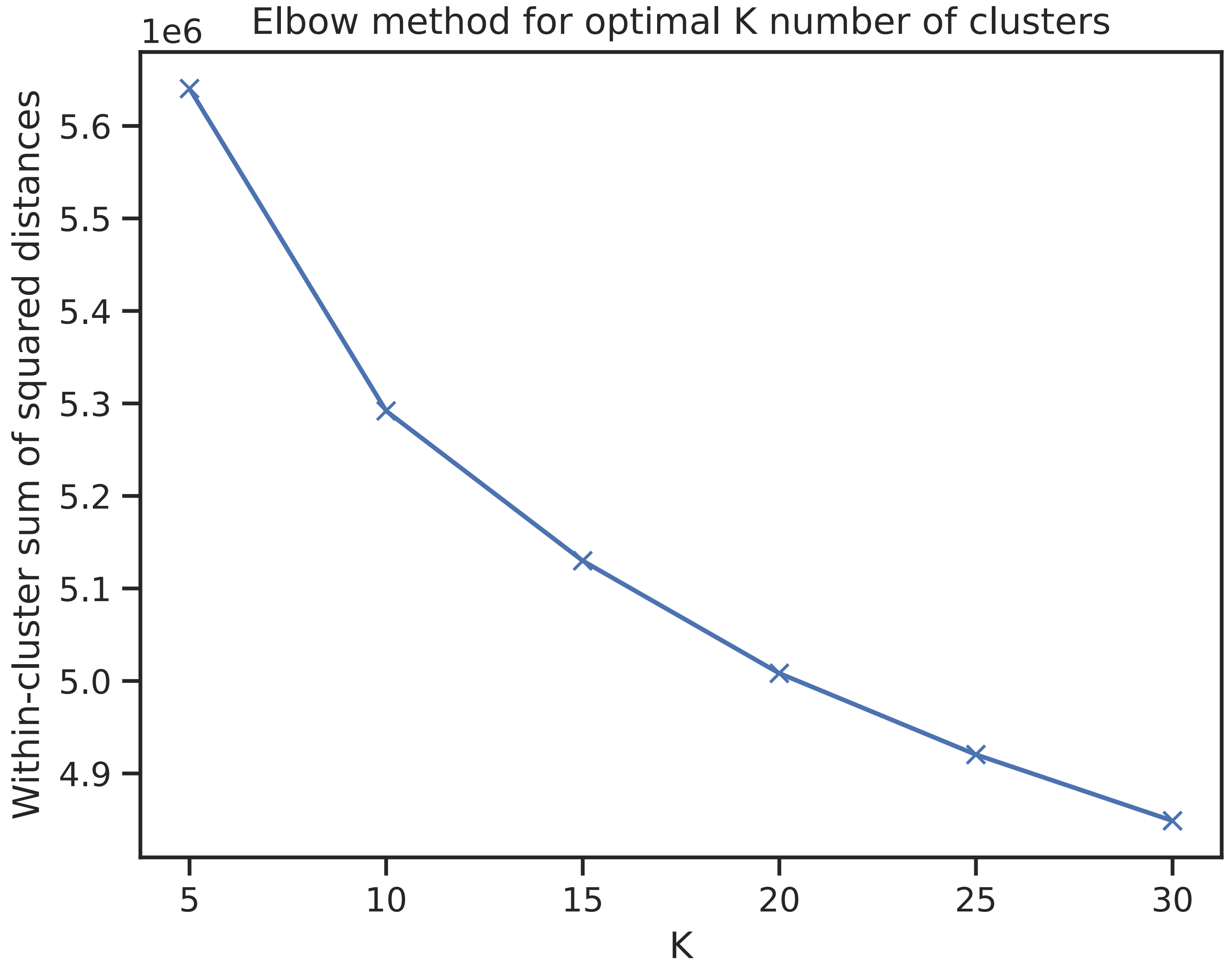
ResNet50

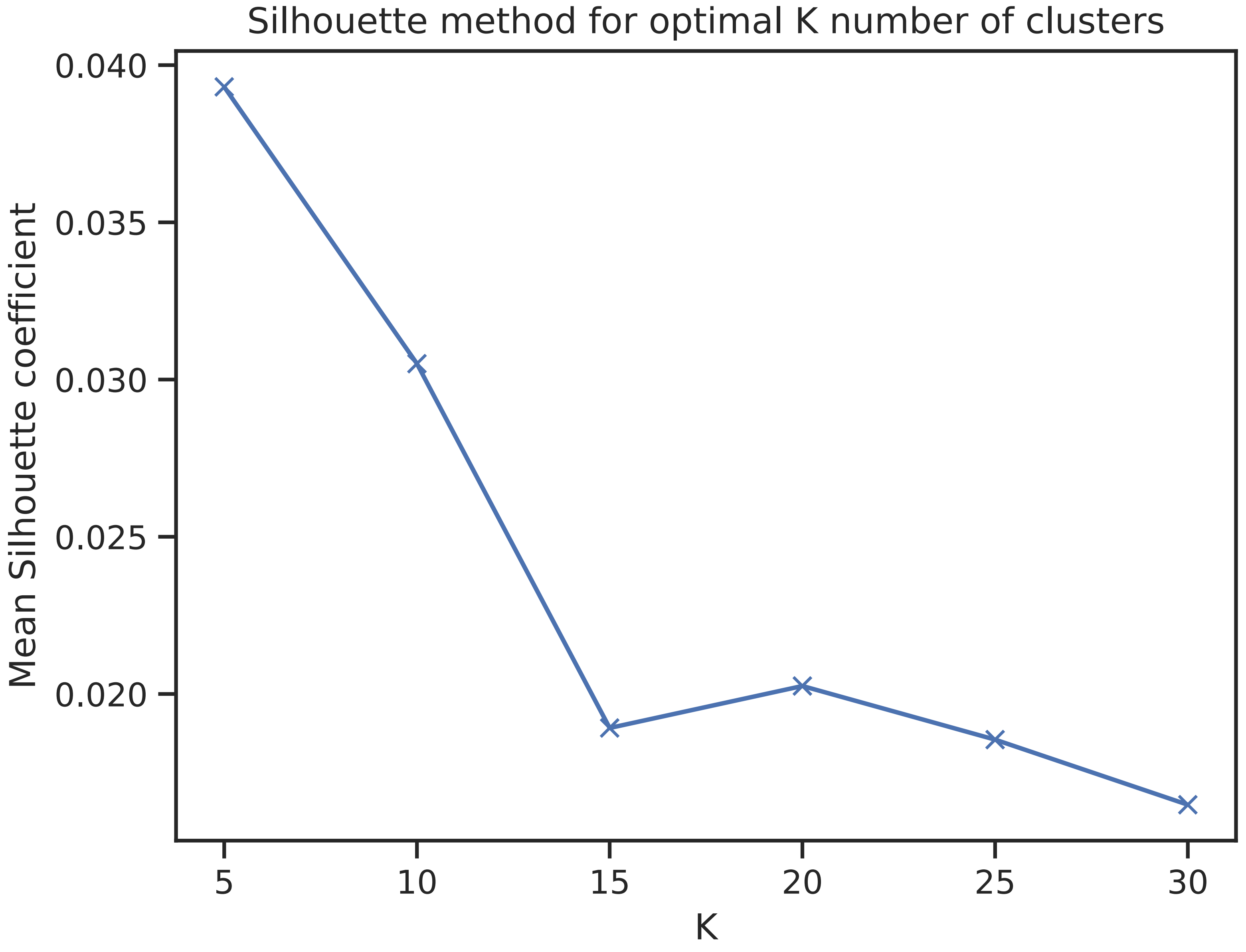
10M x 2048-dim embedding



Cluster

K-means





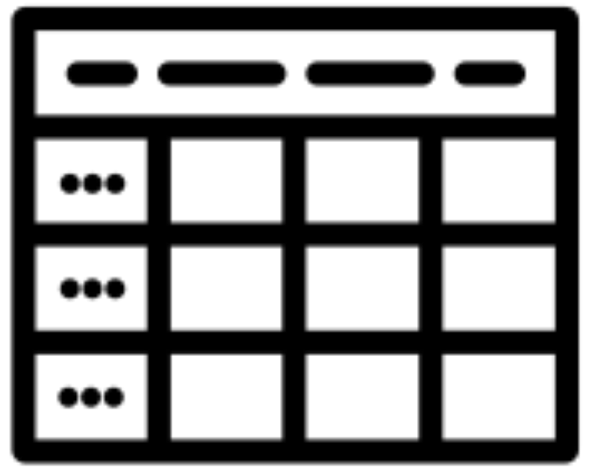
Identify visually-similar imagery

10M images

Extract visual features

ResNet50

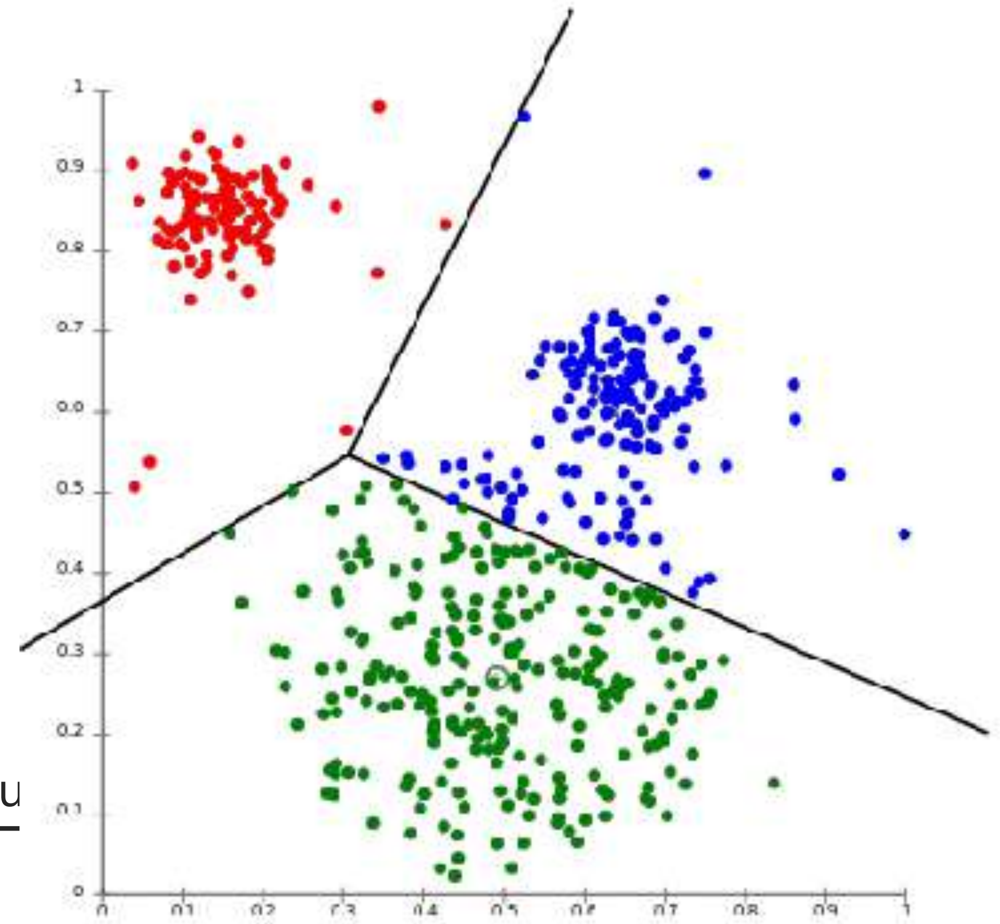
10M x 2048-dim embedding

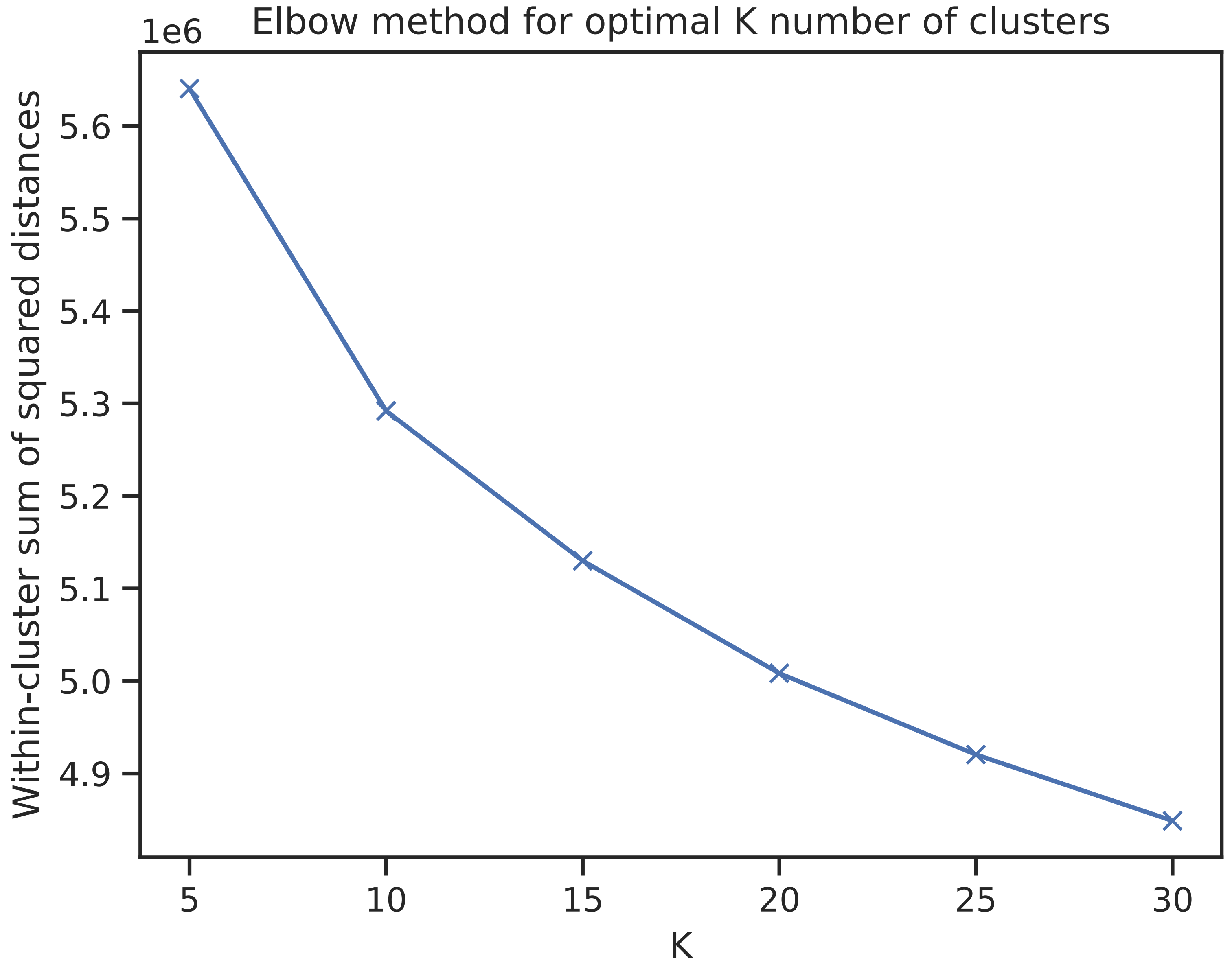


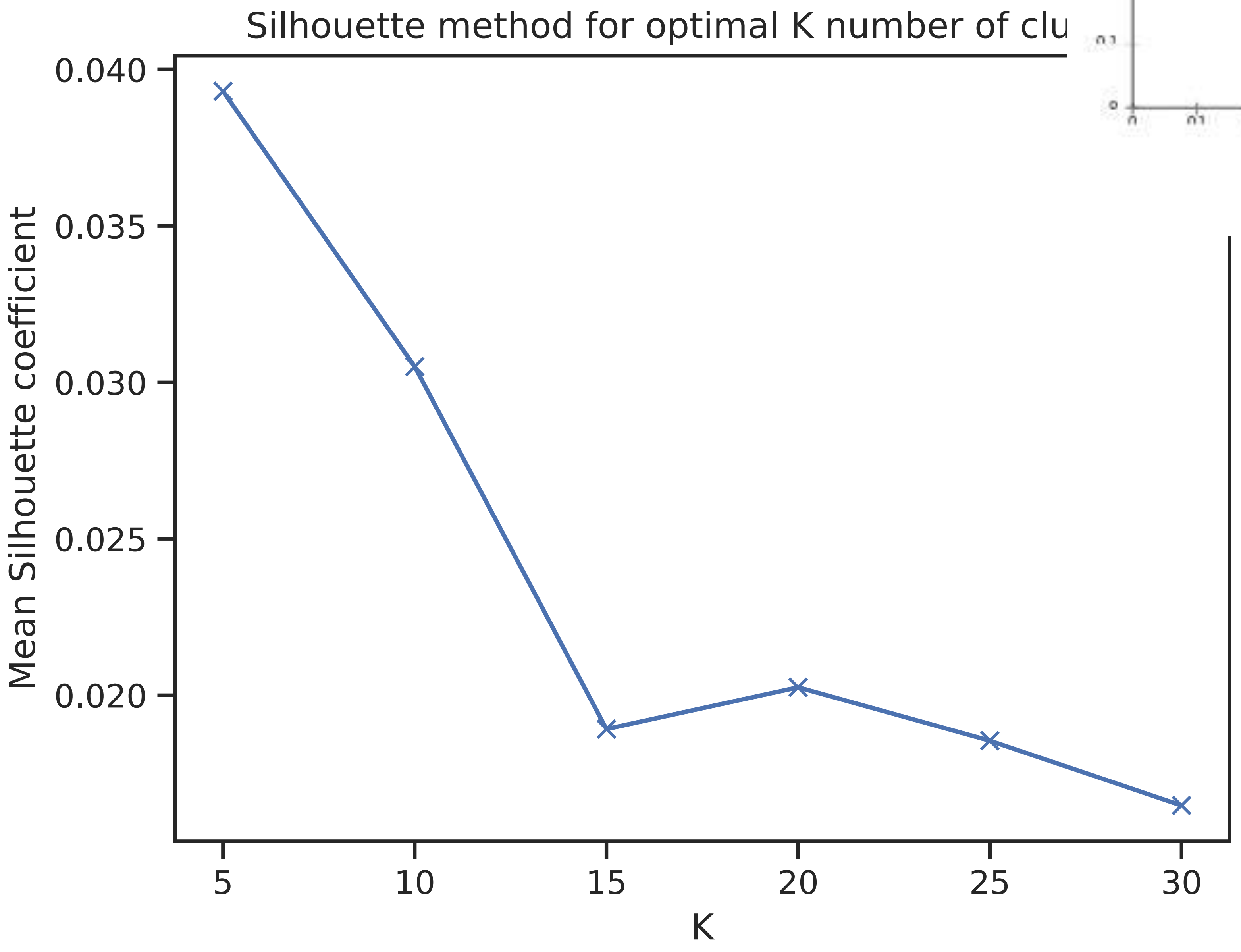
Cluster

K-means

K = 20 types of imagery







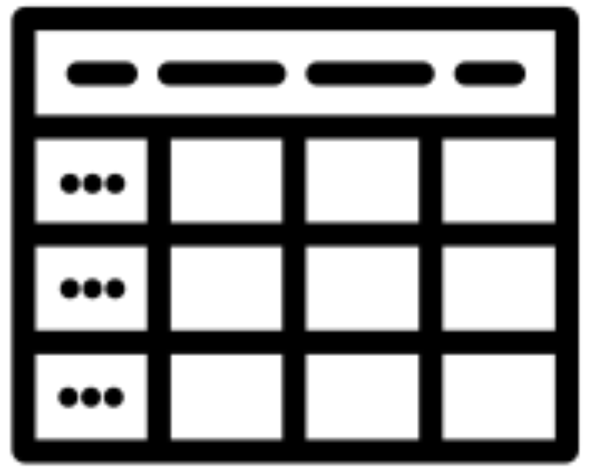
Identify visually-similar imagery

10M images

Extract visual features

ResNet50

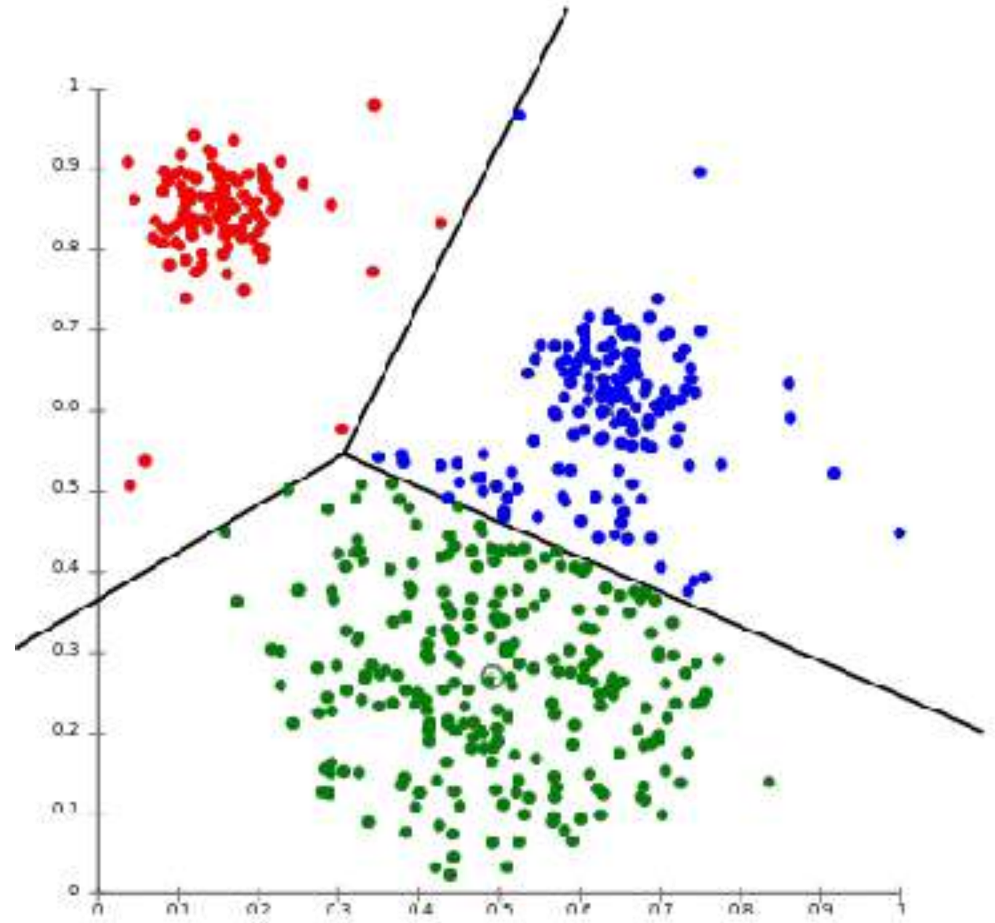
10M x 2048-dim embedding



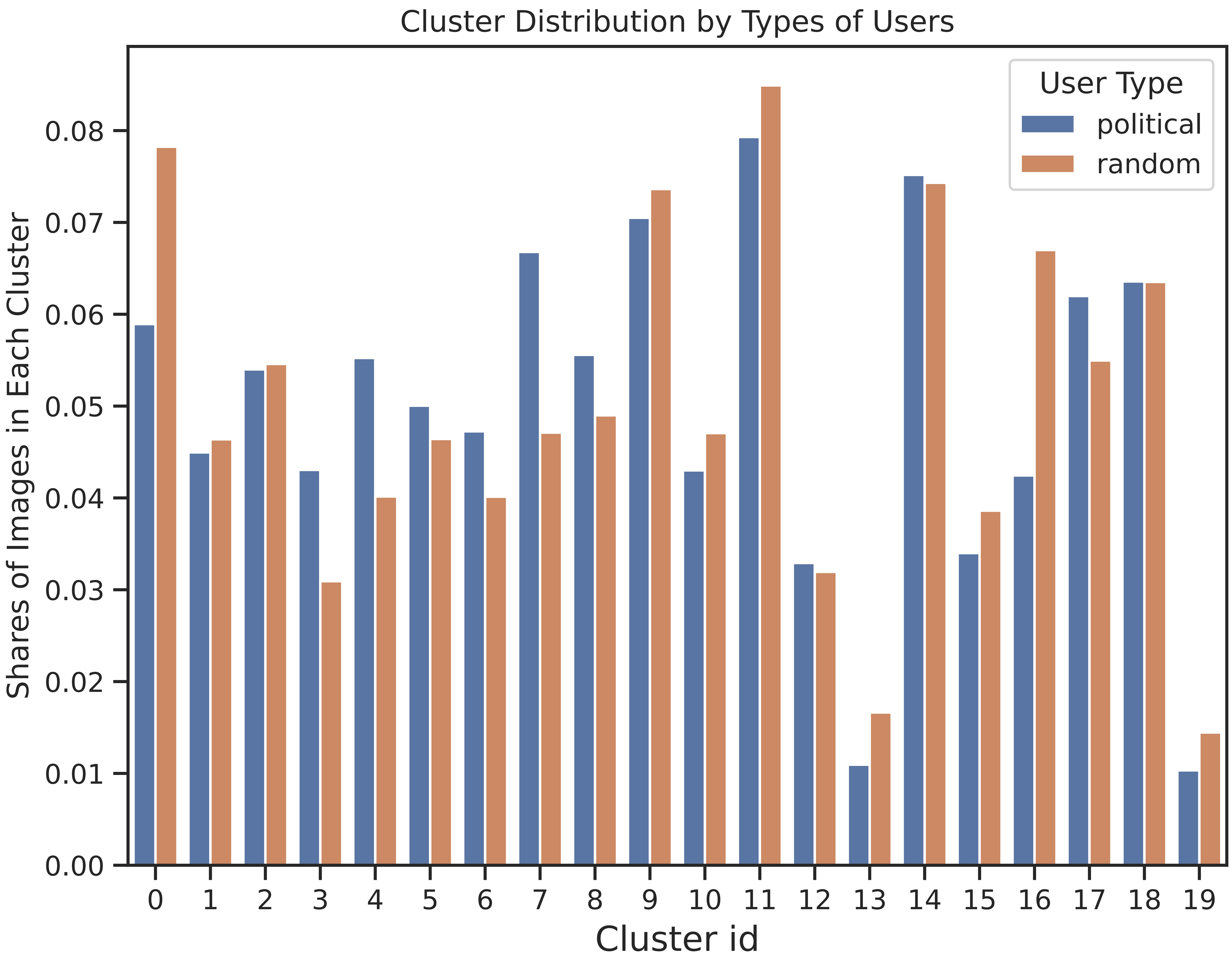
Cluster

K-means

K = 20 types of imagery



Distribution of clusters for each user


Cluster Distribution by Types of Users
User Type
political
random
Shares of Images in Each Cluster
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
Cluster id
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Which clusters are predictive of demography?

Which clusters are predictive of demography?

- Logistic regression at user level — For each user i:

Which clusters are predictive of demography?

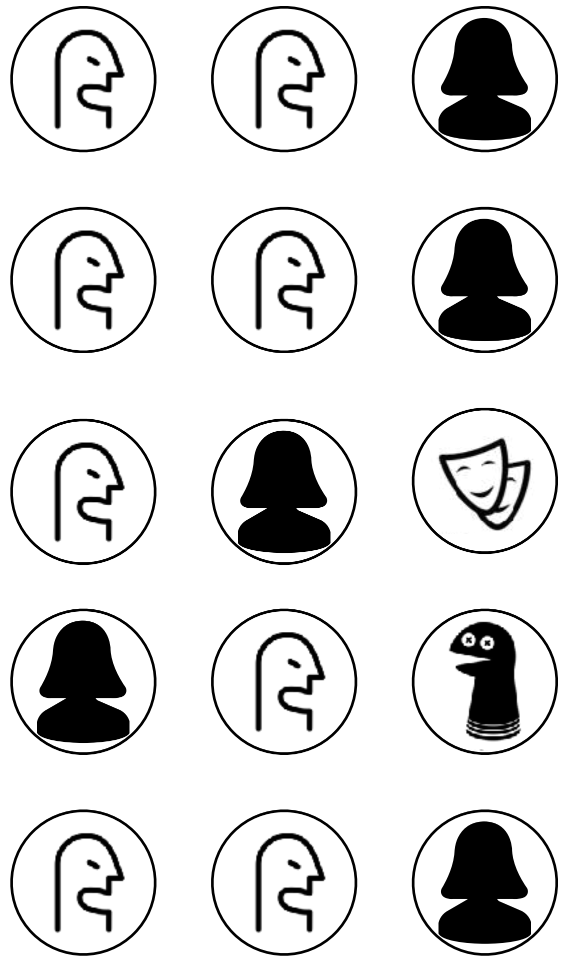
- Logistic regression at user level — For each user i:

$$\text{logit}\left(\text{demography}_i = d\right) = \sum_{k=1}^{20} \beta_k \frac{\text{\# images in cluster } k \text{ shared by user } i}{\text{\# images shared by user } i} + \varepsilon_i$$

Which clusters are predictive of demography?

- Logistic regression at user level — For each user i:

$$\operatorname{logit}\left(\text{demography}_i = d\right) = \sum_{k=1}^{20} \beta_k \frac{\#\text{ images in cluster } k \text{ shared by user } i}{\#\text{ images shared by user } i} + \varepsilon_i$$



Gender

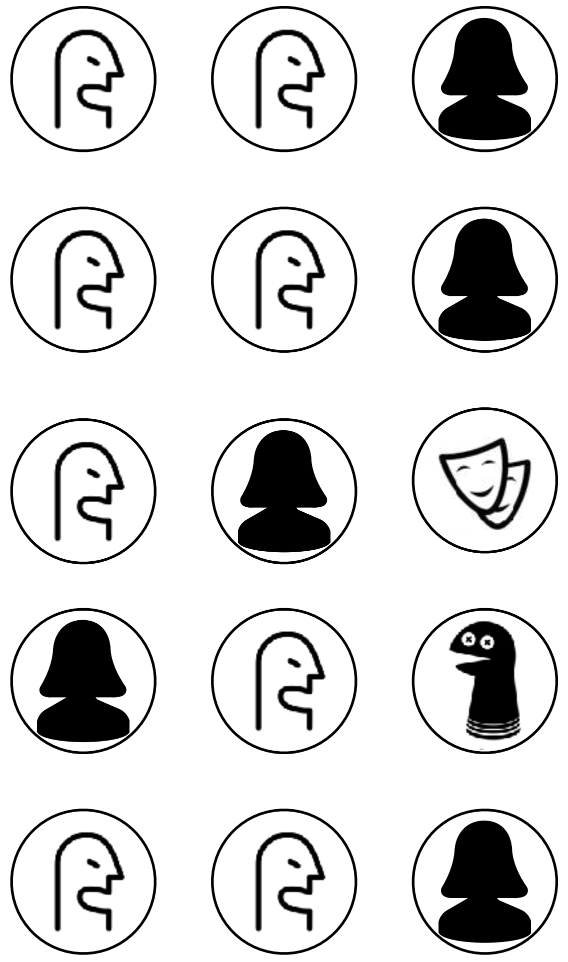
Race

Age

Which clusters are predictive of demography?

- Logistic regression at user level — For each user i:

$$\operatorname{logit}\left(\text{demography}_i = d\right) = \sum_{k=1}^{20} \beta_k \frac{\text{\# images in cluster } k \text{ shared by user } i}{\text{\# images shared by user } i} + \varepsilon_i$$

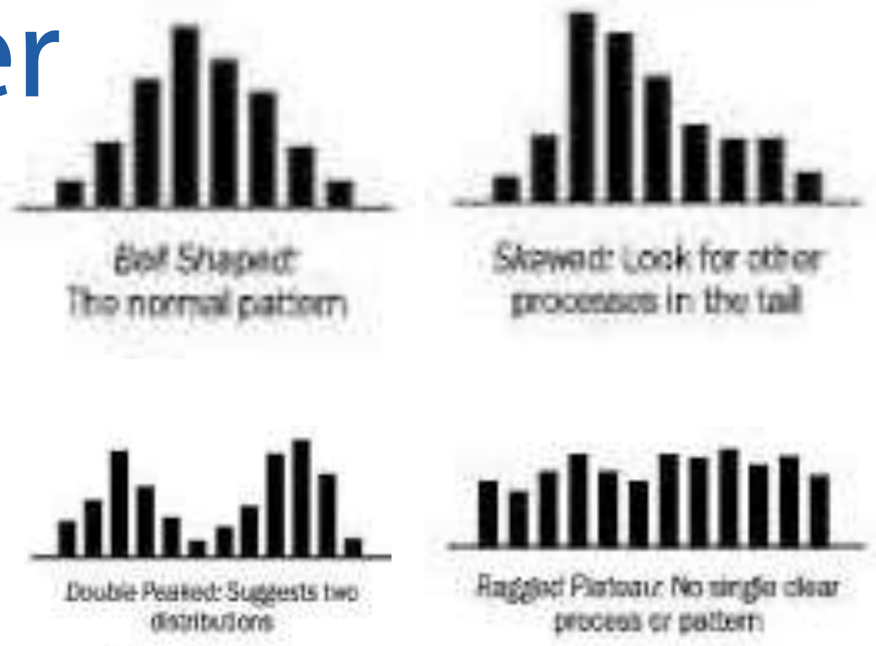


Gender

Race

Age

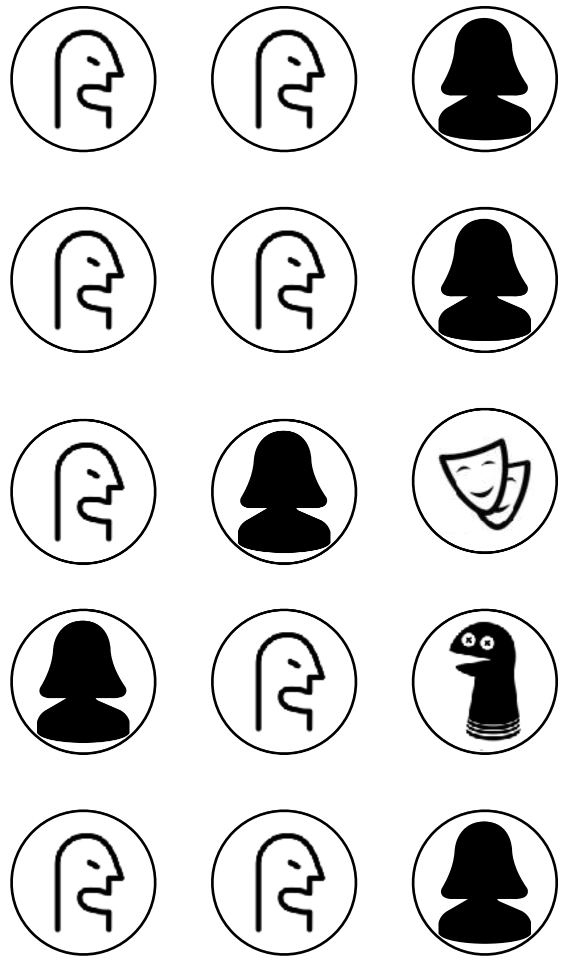
User-level cluster distribution



Which clusters are predictive of demography?

- Logistic regression at user level — For each user i:

$$\text{logit}\left(\text{demography}_i = d\right) = \sum_{k=1}^{20} \beta_k \frac{\#\text{ images in cluster } k \text{ shared by user } i}{\#\text{ images shared by user } i} + \varepsilon_i$$



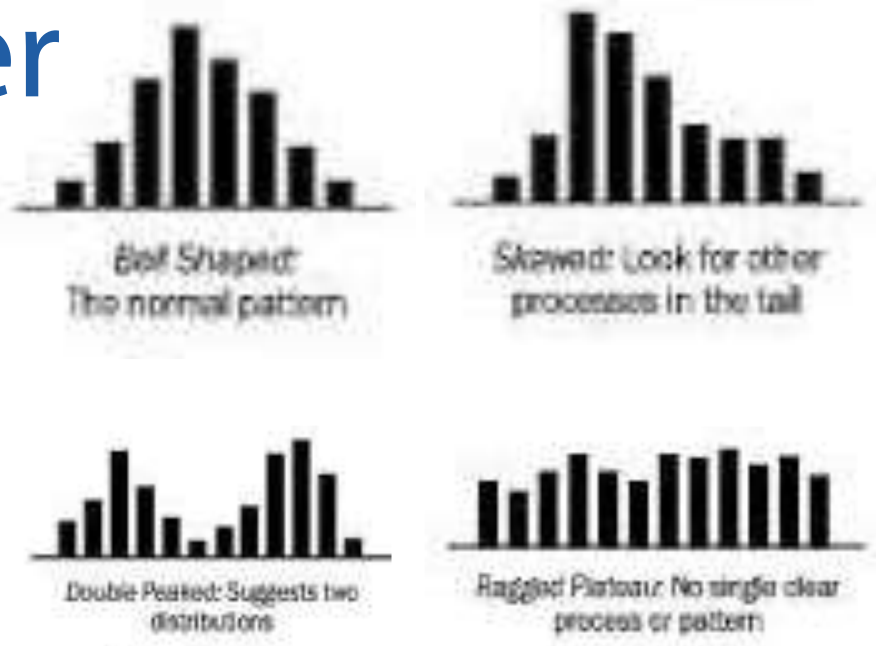
Gender

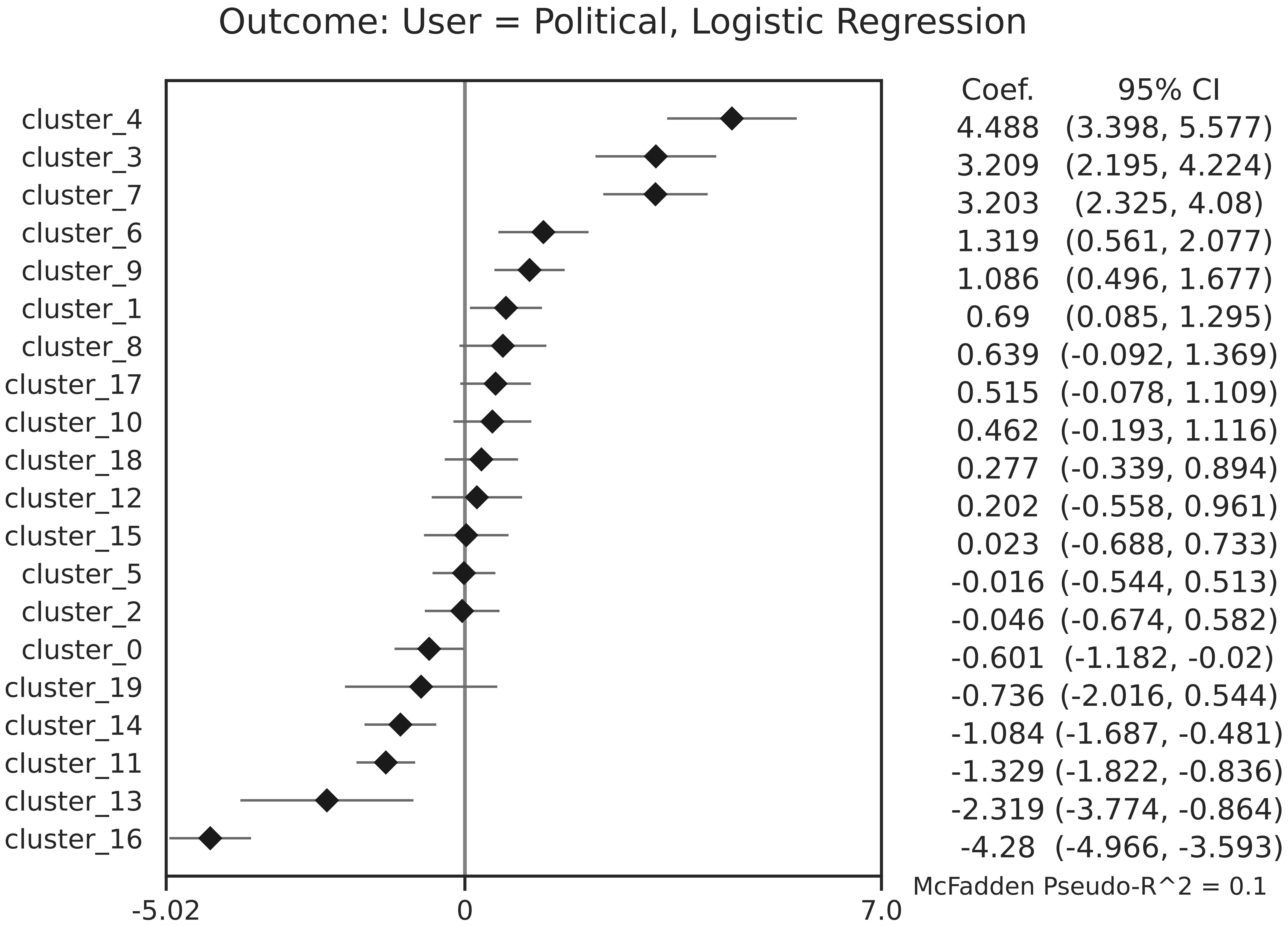
Race

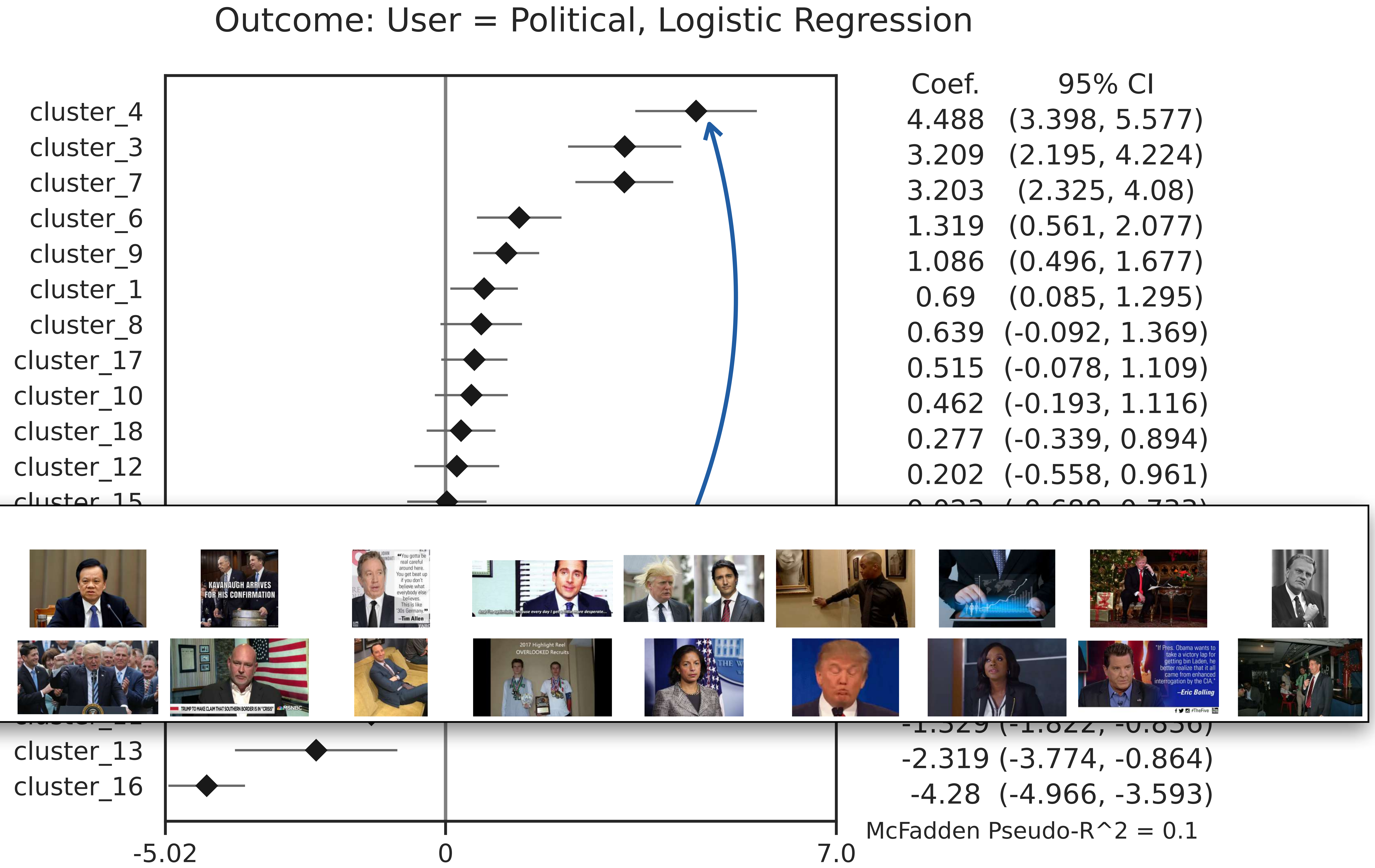
Age

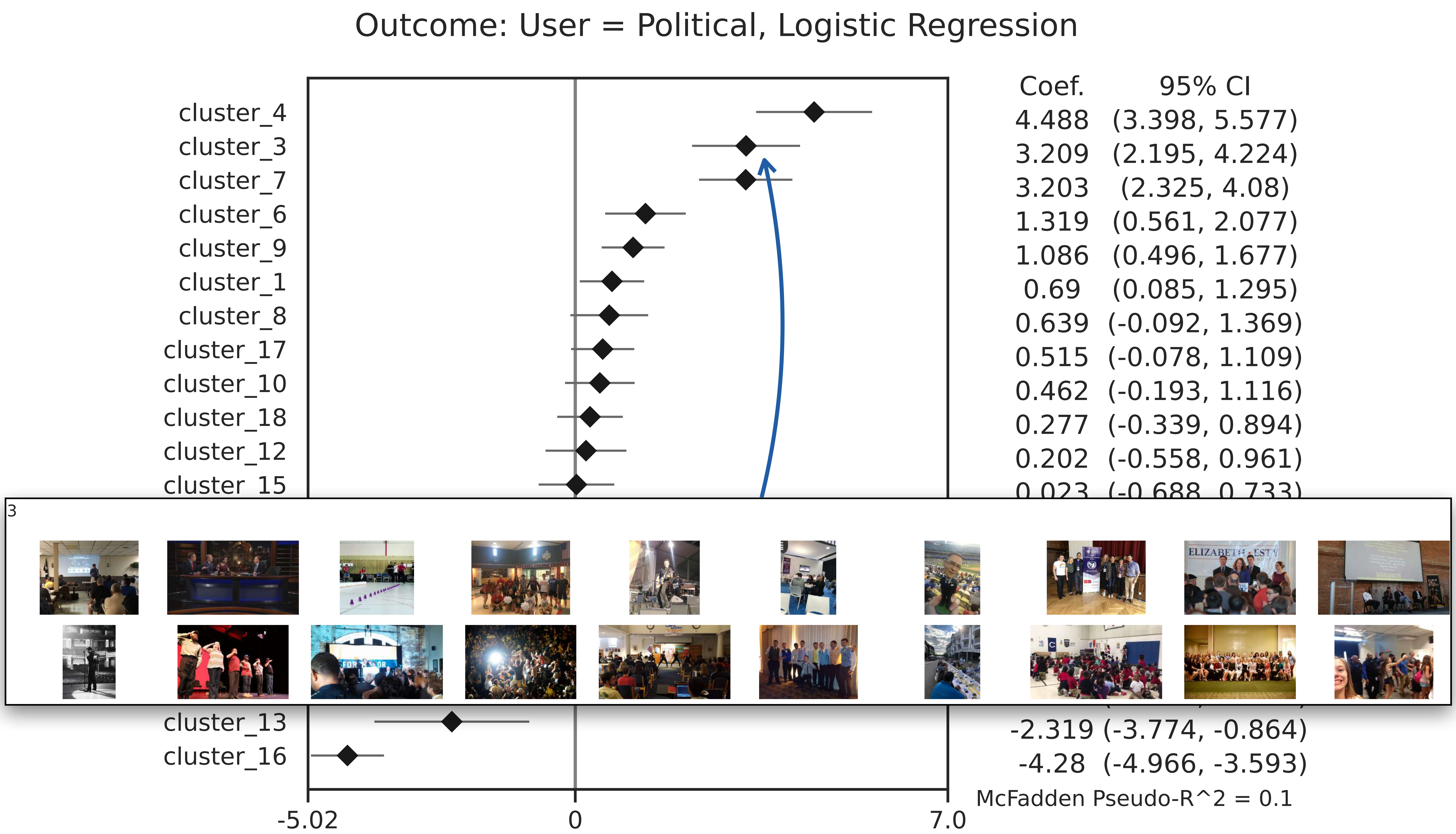
Predictive power of clusters k on demography d

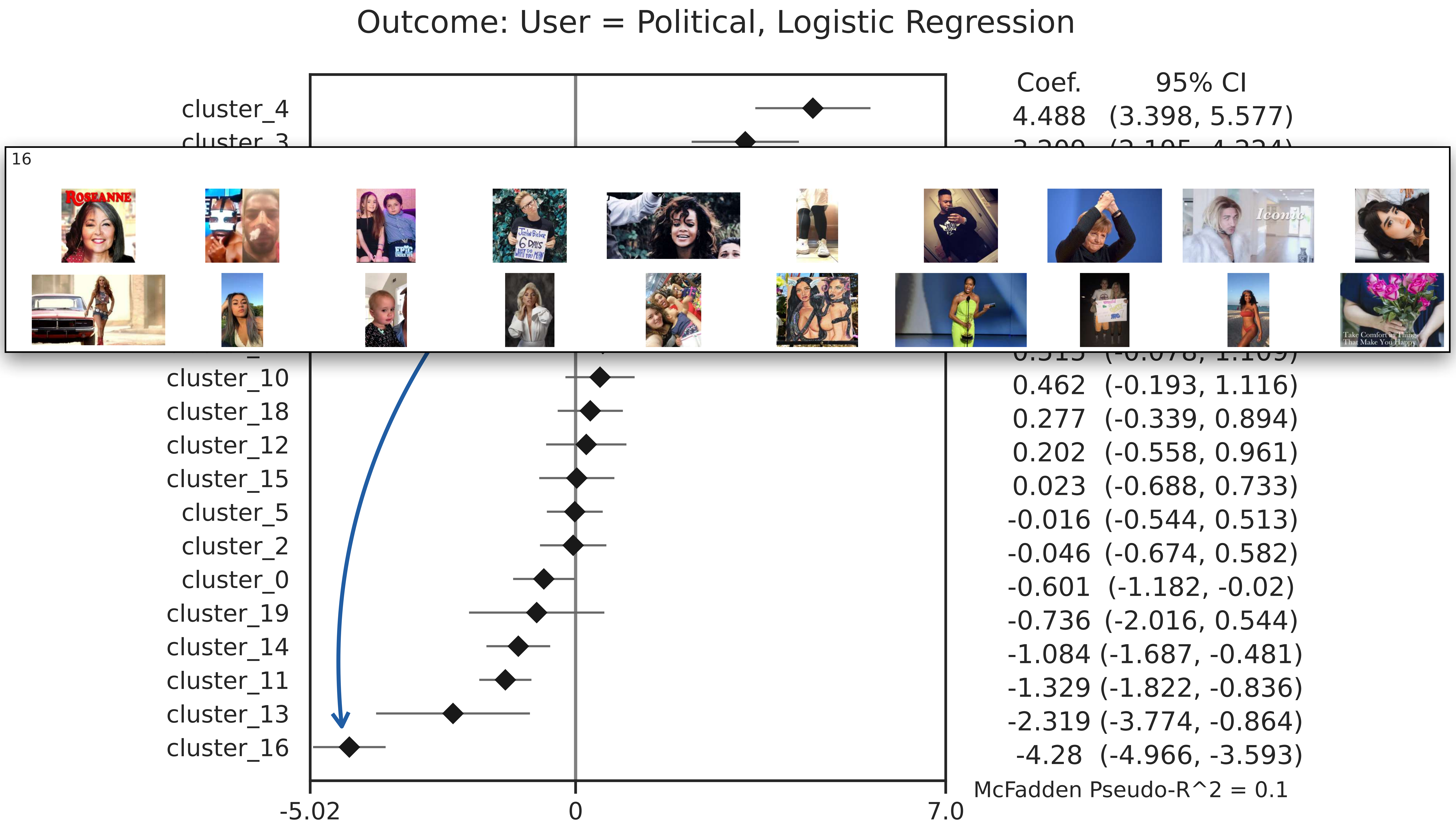
User-level cluster distribution

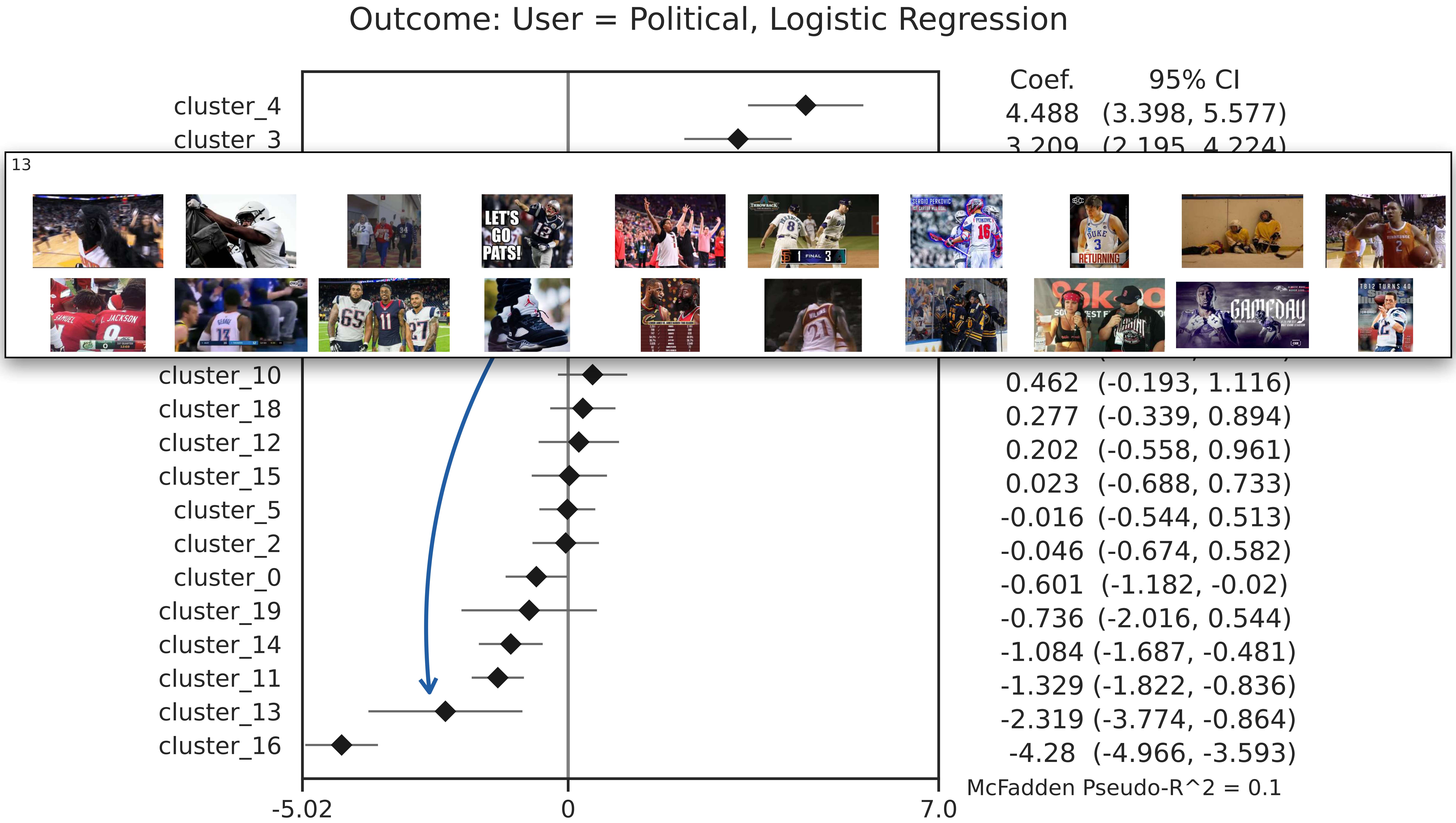



Outcome: User = Political, Logistic Regression
Coef. 95% CI
cluster_4 4.488 (3.398, 5.577)
cluster_3 3.209 (2.195, 4.224)
cluster_7 3.203 (2.325, 4.08)
cluster_6 1.319 (0.561, 2.077)
cluster_9 1.086 (0.496, 1.677)
cluster_1 0.69 (0.085, 1.295)
cluster_8 0.639 (-0.092, 1.369)
cluster_17 0.515 (-0.078, 1.109)
cluster_10 0.462 (-0.193, 1.116)
cluster_18 0.277 (-0.339, 0.894)
cluster_12 0.202 (-0.558, 0.961)
cluster_15 0.023 (-0.688, 0.733)
cluster_5 -0.016 (-0.544, 0.513)
cluster_2 -0.046 (-0.674, 0.582)
cluster_0 -0.601 (-1.182, -0.02)
cluster_19 -0.736 (-2.016, 0.544)
cluster_14 -1.084 (-1.687, -0.481)
cluster_11 -1.329 (-1.822, -0.836)
cluster_13 -2.319 (-3.774, -0.864)
cluster_16 -4.28 (-4.966, -3.593)
-5.02
0
7.0
McFadden Pseudo-R^2 = 0.1

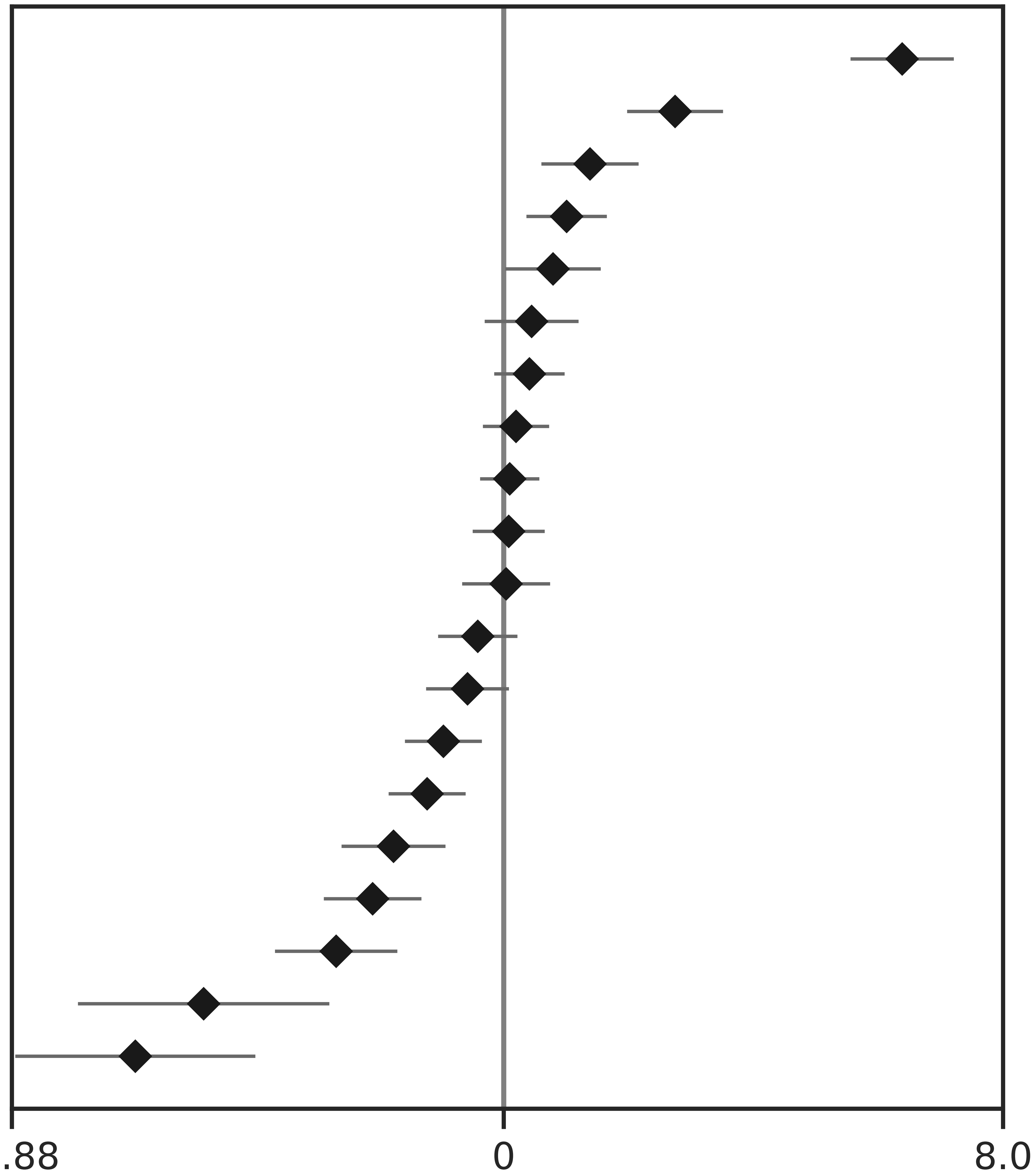

Outcome: User = Political, Logistic Regression
Coef.
95% CI
cluster_4
4.488
(3.398, 5.577)
cluster_3
3.209
(2.195, 4.224)
cluster_7
3.203
(2.325, 4.08)
cluster_6
1.319
(0.561, 2.077)
cluster_9
1.086
(0.496, 1.677)
cluster_1
0.69
(0.085, 1.295)
cluster_8
0.639
(-0.092, 1.369)
cluster_17
0.515
(-0.078, 1.109)
cluster_10
0.462
(-0.193, 1.116)
cluster_18
0.277
(-0.339, 0.894)
cluster_12
0.202
(-0.558, 0.961)
cluster_13
-2.319
(-3.774, -0.864)
cluster_16
-4.28
(-4.966, -3.593)
-5.02
0
7.0
McFadden Pseudo-R^2 = 0.1
4
KAVANAUGH ARRIVES FOR HIS CONFIRMATION
2017 Highlight Reel
OVERLOOKED Recruits
–Eric Bolling


Outcome: User = Political, Logistic Regression
Coef. 95% CI
cluster_4 4.488 (3.398, 5.577)
cluster_3 3.209 (2.195, 4.224)
cluster_7 3.203 (2.325, 4.08)
cluster_6 1.319 (0.561, 2.077)
cluster_9 1.086 (0.496, 1.677)
cluster_1 0.69 (0.085, 1.295)
cluster_8 0.639 (-0.092, 1.369)
cluster_17 0.515 (-0.078, 1.109)
cluster_10 0.462 (-0.193, 1.116)
cluster_18 0.277 (-0.339, 0.894)
cluster_12 0.202 (-0.558, 0.961)
cluster_15 0.023 (-0.688, 0.733)
3
ELIZABETH
cluster_13 -2.319 (-3.774, -0.864)
cluster_16 -4.28 (-4.966, -3.593)
McFadden Pseudo-R^2 = 0.1
-5.02
0
7.0


Outcome: User = Political, Logistic Regression
Coef.
95% CI
cluster_4
4.488 (3.398, 5.577)
16
Roseanne
Justin Bieber 6 Days
Iconic
Take Comfort in Things That Make You Happy
cluster_10
0.462 (-0.193, 1.116)
cluster_18
0.277 (-0.339, 0.894)
cluster_12
0.202 (-0.558, 0.961)
cluster_15
0.023 (-0.688, 0.733)
cluster_5
-0.016 (-0.544, 0.513)
cluster_2
-0.046 (-0.674, 0.582)
cluster_0
-0.601 (-1.182, -0.02)
cluster_19
-0.736 (-2.016, 0.544)
cluster_14
-1.084 (-1.687, -0.481)
cluster_11
-1.329 (-1.822, -0.836)
cluster_13
-2.319 (-3.774, -0.864)
cluster_16
-4.28 (-4.966, -3.593)
-5.02
0
7.0
McFadden Pseudo-R^2 = 0.1

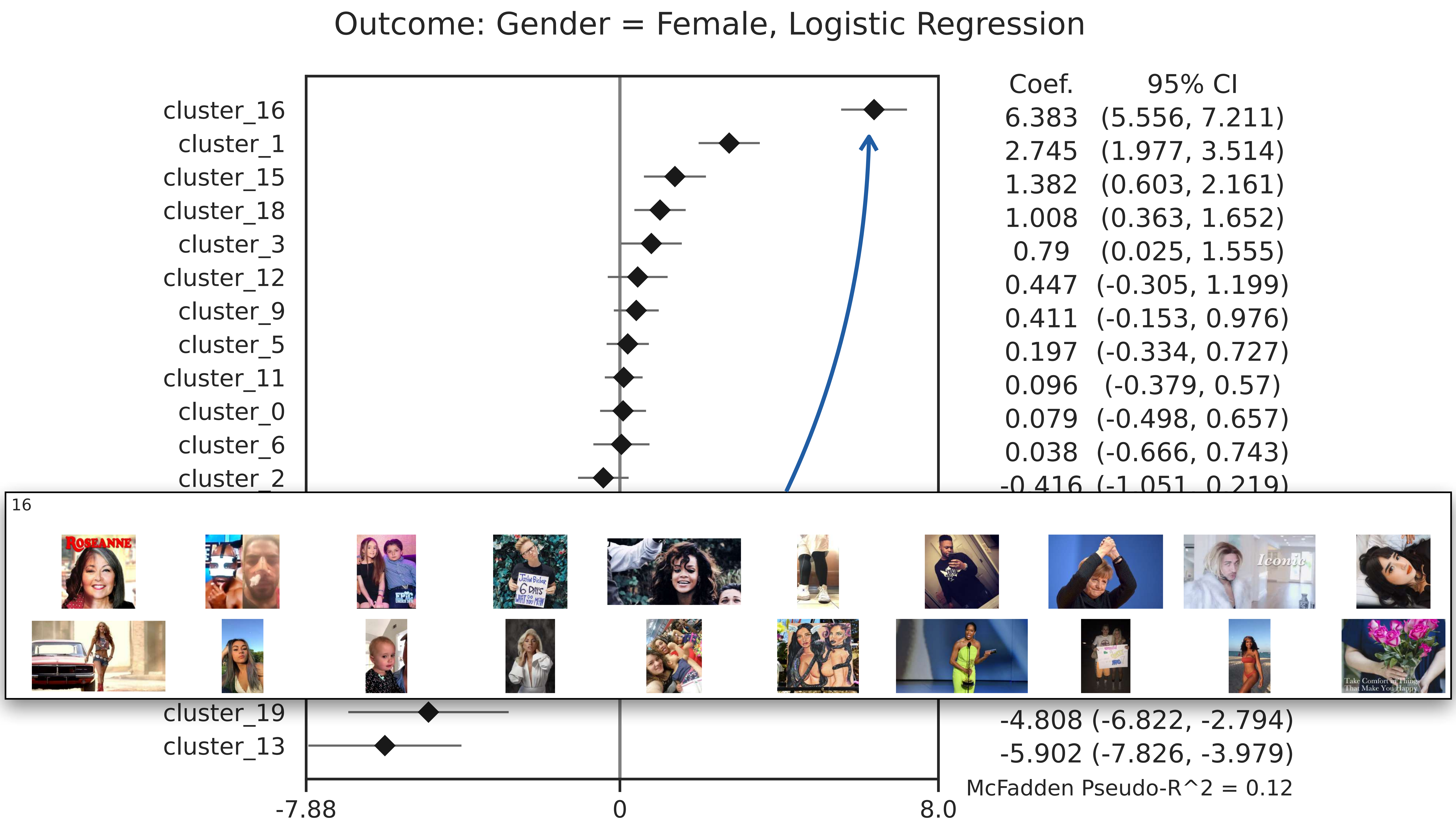

Outcome: User = Political, Logistic Regression
Coef. 95% CI
cluster_4 4.488 (3.398, 5.577)
cluster_3 3.209 (2.195, 4.224)
13
LET'S GO PATS!
DUKE
RETURNING
SERGIO PERKOVIC
THROWBACK
FINAL
L. JACKSON
SAMUEL
GAMEDAY
TB12 TURNS 40
cluster_10 0.462 (-0.193, 1.116)
cluster_18 0.277 (-0.339, 0.894)
cluster_12 0.202 (-0.558, 0.961)
cluster_15 0.023 (-0.688, 0.733)
cluster_5 -0.016 (-0.544, 0.513)
cluster_2 -0.046 (-0.674, 0.582)
cluster_0 -0.601 (-1.182, -0.02)
cluster_19 -0.736 (-2.016, 0.544)
cluster_14 -1.084 (-1.687, -0.481)
cluster_11 -1.329 (-1.822, -0.836)
cluster_13 -2.319 (-3.774, -0.864)
cluster_16 -4.28 (-4.966, -3.593)
McFadden Pseudo-R^2 = 0.1
-5.02
0
7.0

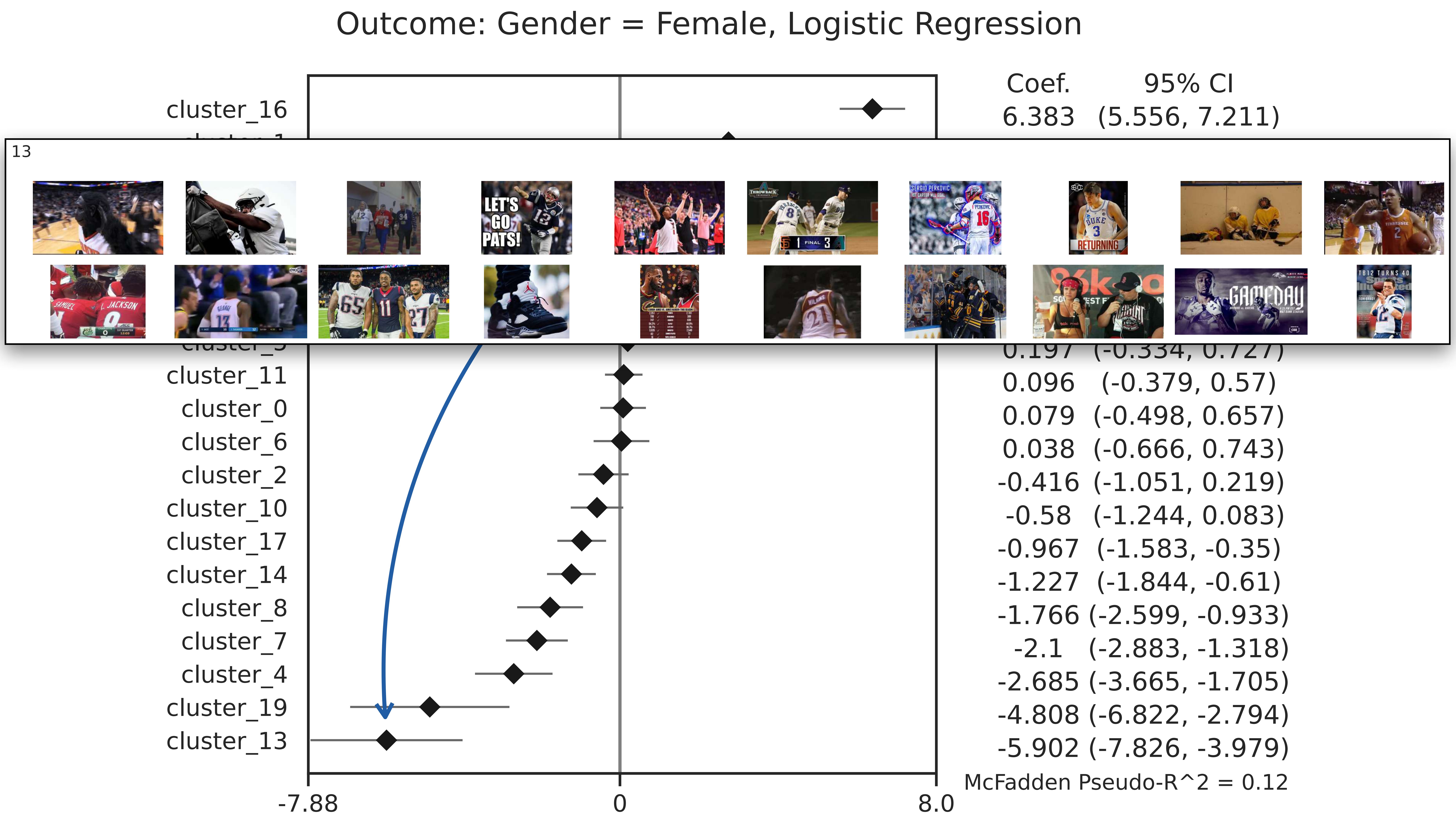
Outcome: Gender = Female, Logistic Regression

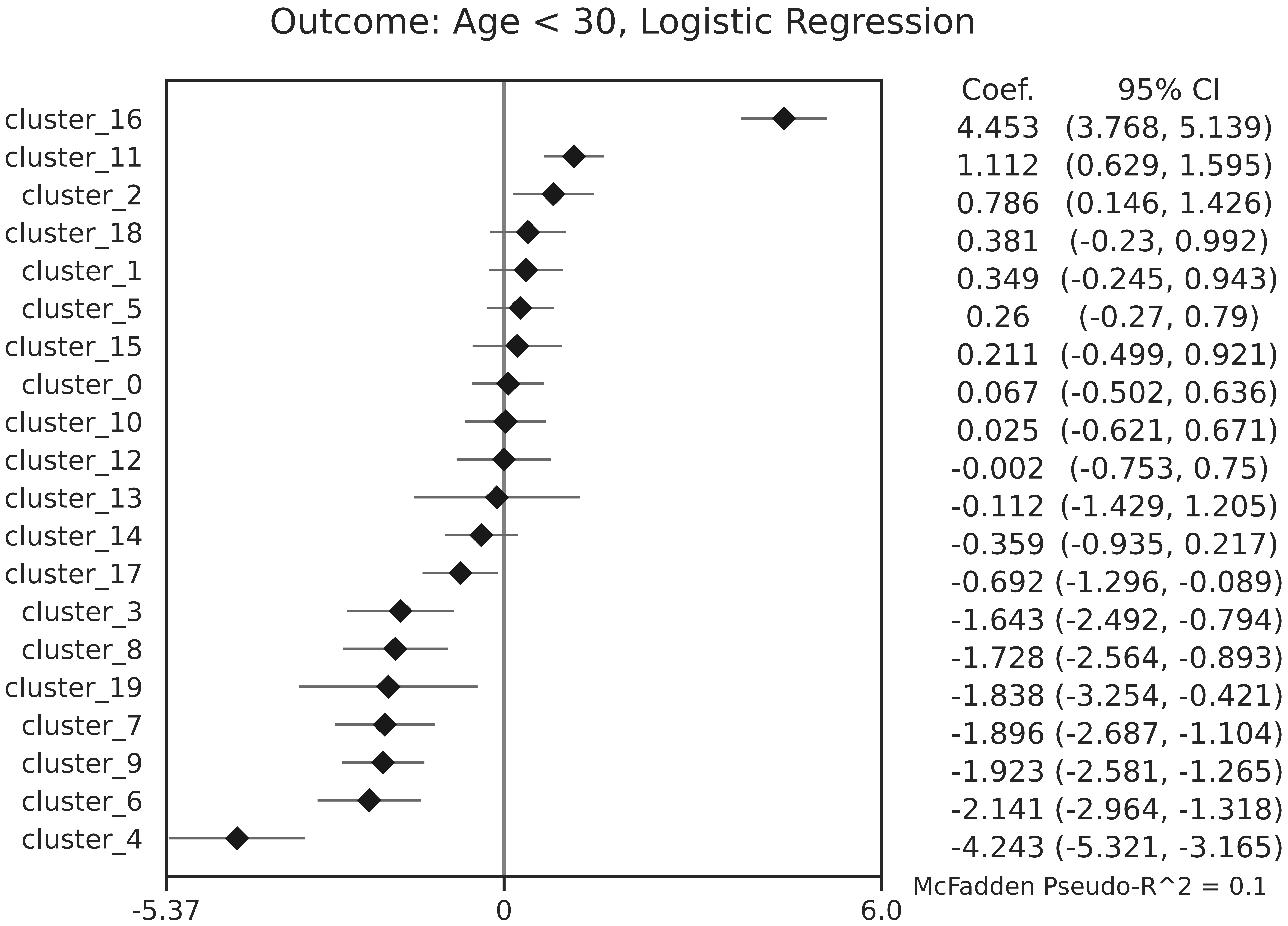


	Coef.	95% CI
cluster_16	6.383	(5.556, 7.211)
cluster_1	2.745	(1.977, 3.514)
cluster_15	1.382	(0.603, 2.161)
cluster_18	1.008	(0.363, 1.652)
cluster_3	0.79	(0.025, 1.555)
cluster_12	0.447	(-0.305, 1.199)
cluster_9	0.411	(-0.153, 0.976)
cluster_5	0.197	(-0.334, 0.727)
cluster_11	0.096	(-0.379, 0.57)
cluster_0	0.079	(-0.498, 0.657)
cluster_6	0.038	(-0.666, 0.743)
cluster_2	-0.416	(-1.051, 0.219)
cluster_10	-0.58	(-1.244, 0.083)
cluster_17	-0.967	(-1.583, -0.35)
cluster_14	-1.227	(-1.844, -0.61)
cluster_8	-1.766	(-2.599, -0.933)
cluster_7	-2.1	(-2.883, -1.318)
cluster_4	-2.685	(-3.665, -1.705)
cluster_19	-4.808	(-6.822, -2.794)
cluster_13	-5.902	(-7.826, -3.979)

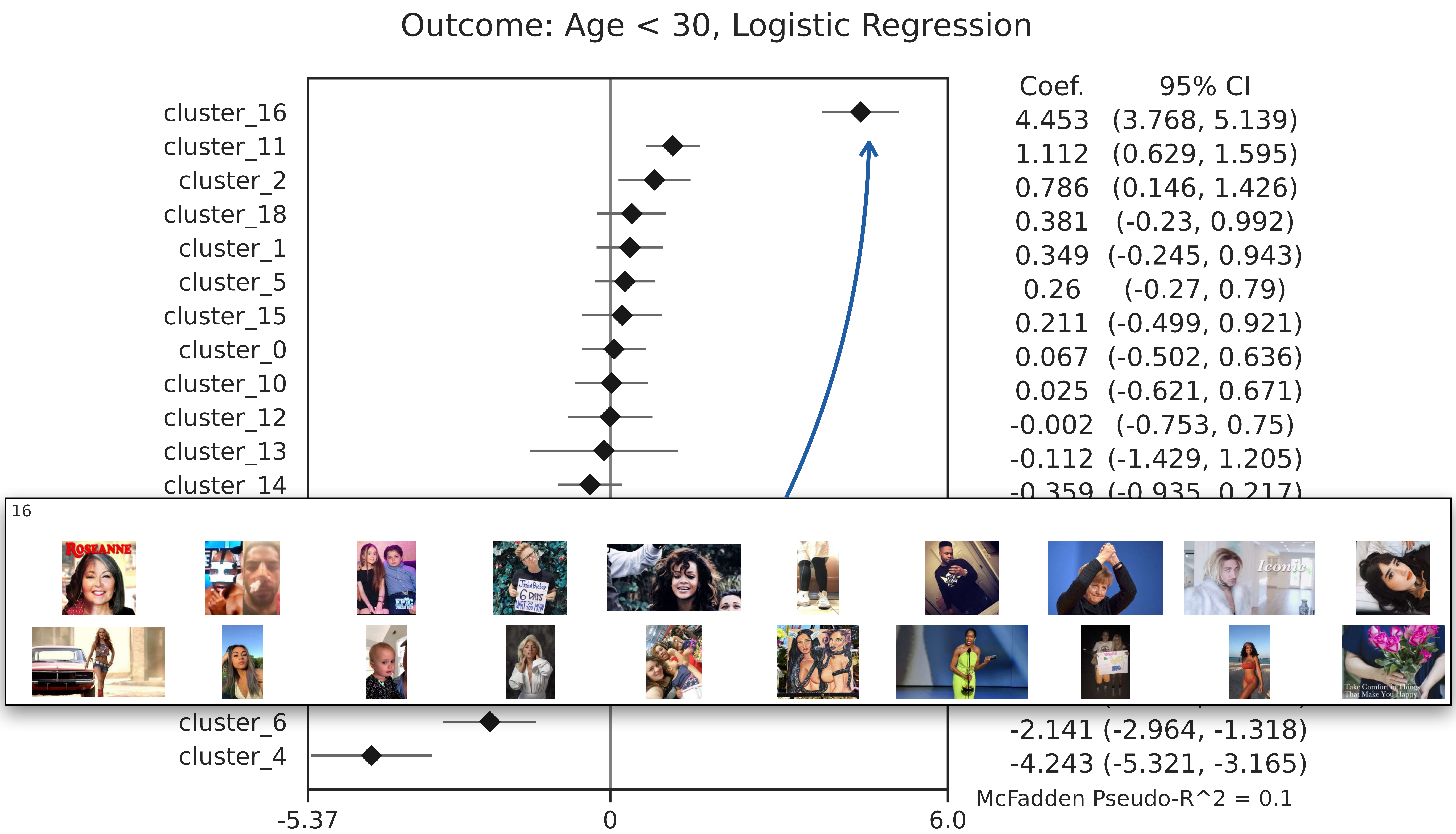
McFadden Pseudo-R^2 = 0.12


Outcome: Gender = Female, Logistic Regression
Coef. 95% CI
cluster_16 6.383 (5.556, 7.211)
cluster_1 2.745 (1.977, 3.514)
cluster_15 1.382 (0.603, 2.161)
cluster_18 1.008 (0.363, 1.652)
cluster_3 0.79 (0.025, 1.555)
cluster_12 0.447 (-0.305, 1.199)
cluster_9 0.411 (-0.153, 0.976)
cluster_5 0.197 (-0.334, 0.727)
cluster_11 0.096 (-0.379, 0.57)
cluster_0 0.079 (-0.498, 0.657)
cluster_6 0.038 (-0.666, 0.743)
cluster_2 -0.416 (-1.051, 0.219)
16
ROSEANNE
Iconic
Take Comfort in Things That Make You Happy
cluster_19 -4.808 (-6.822, -2.794)
cluster_13 -5.902 (-7.826, -3.979)
McFadden Pseudo-R^2 = 0.12
-7.88
0
8.0

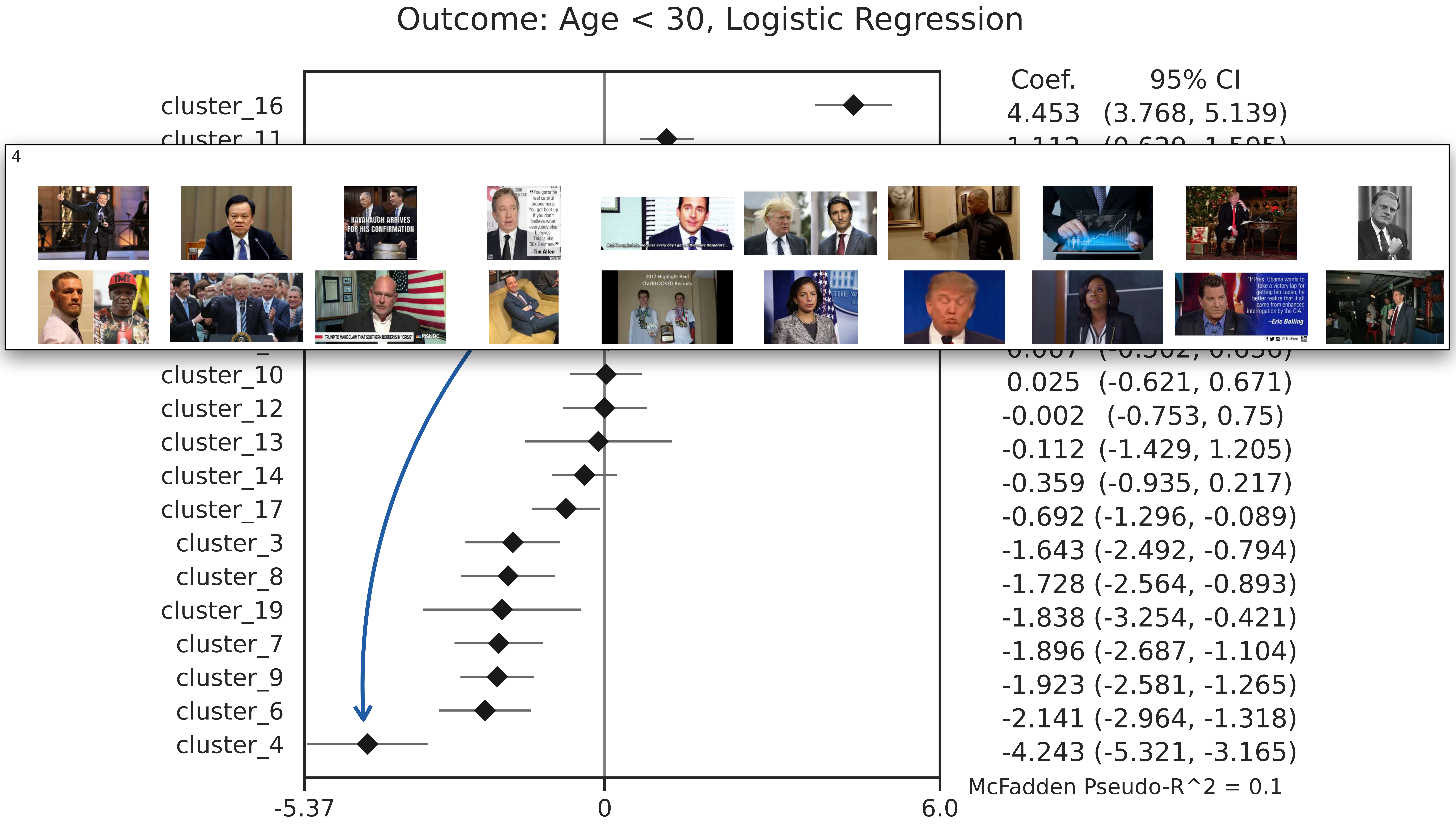

Outcome: Gender = Female, Logistic Regression
Coef.
95% CI
cluster_16
6.383
(5.556, 7.211)
13
LET'S GO PATS!
RETURNING
GAMEDAY
cluster_11
0.096
(-0.379, 0.57)
cluster_0
0.079
(-0.498, 0.657)
cluster_6
0.038
(-0.666, 0.743)
cluster_2
-0.416
(-1.051, 0.219)
cluster_10
-0.58
(-1.244, 0.083)
cluster_17
-0.967
(-1.583, -0.35)
cluster_14
-1.227
(-1.844, -0.61)
cluster_8
-1.766
(-2.599, -0.933)
cluster_7
-2.1
(-2.883, -1.318)
cluster_4
-2.685
(-3.665, -1.705)
cluster_19
-4.808
(-6.822, -2.794)
cluster_13
-5.902
(-7.826, -3.979)
McFadden Pseudo-R^2 = 0.12
-7.88
0
8.0



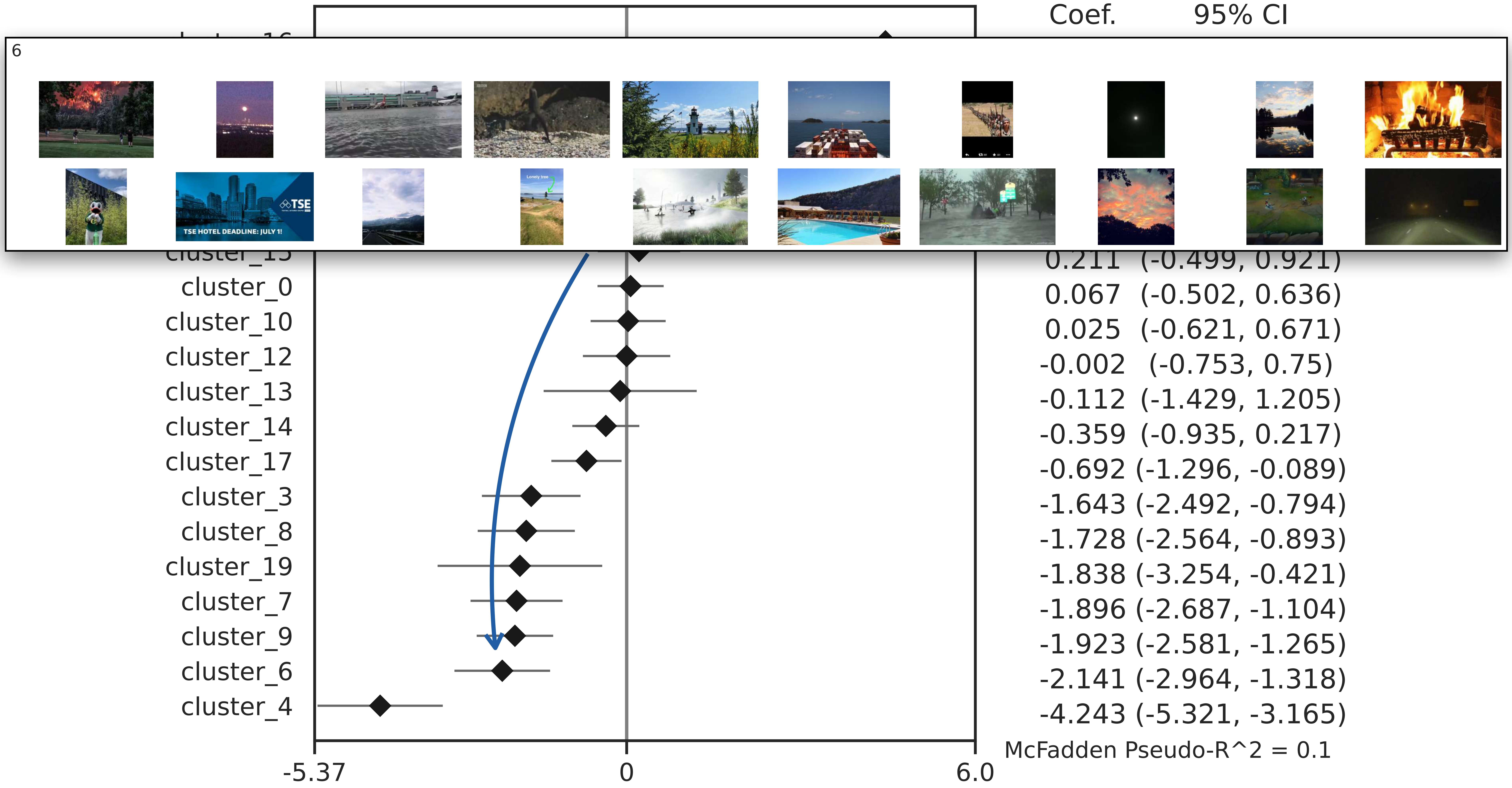
Outcome: Age < 30, Logistic Regression
Coef.
95% CI
cluster_16 4.453 (3.768, 5.139)
cluster_11 1.112 (0.629, 1.595)
cluster_2 0.786 (0.146, 1.426)
cluster_18 0.381 (-0.23, 0.992)
cluster_1 0.349 (-0.245, 0.943)
cluster_5 0.26 (-0.27, 0.79)
cluster_15 0.211 (-0.499, 0.921)
cluster_0 0.067 (-0.502, 0.636)
cluster_10 0.025 (-0.621, 0.671)
cluster_12 -0.002 (-0.753, 0.75)
cluster_13 -0.112 (-1.429, 1.205)
cluster_14 -0.359 (-0.935, 0.217)
cluster_17 -0.692 (-1.296, -0.089)
cluster_3 -1.643 (-2.492, -0.794)
cluster_8 -1.728 (-2.564, -0.893)
cluster_19 -1.838 (-3.254, -0.421)
cluster_7 -1.896 (-2.687, -1.104)
cluster_9 -1.923 (-2.581, -1.265)
cluster_6 -2.141 (-2.964, -1.318)
cluster_4 -4.243 (-5.321, -3.165)
-5.37
0
6.0
McFadden Pseudo-R^2 = 0.1

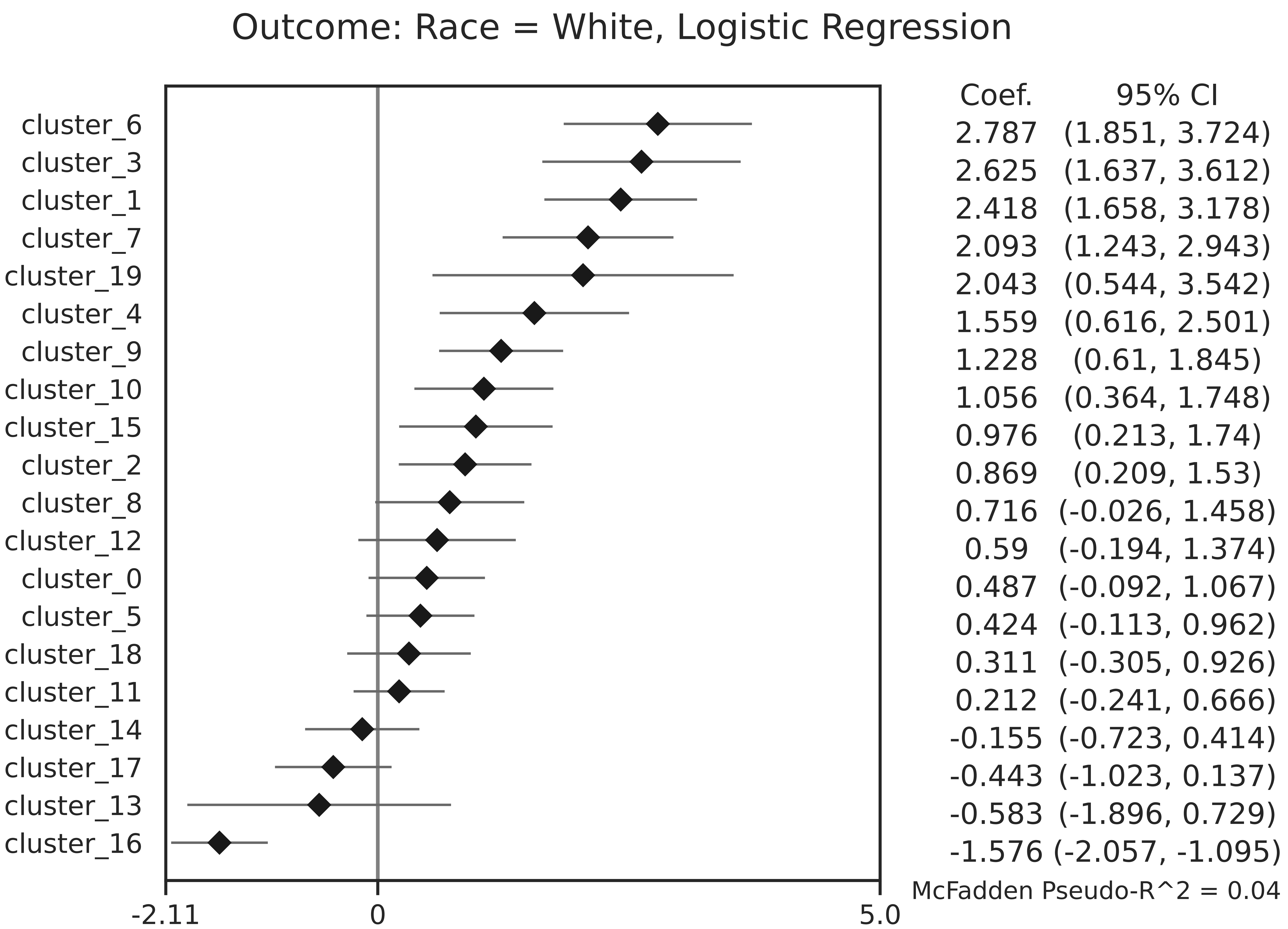


Outcome: Age < 30, Logistic Regression
cluster_16
cluster_11
cluster_2
cluster_18
cluster_1
cluster_5
cluster_15
cluster_0
cluster_10
cluster_12
cluster_13
cluster_14
cluster_6
cluster_4
-5.37
0
6.0
Coef. 95% CI
4.453 (3.768, 5.139)
1.112 (0.629, 1.595)
0.786 (0.146, 1.426)
0.381 (-0.23, 0.992)
0.349 (-0.245, 0.943)
0.26 (-0.27, 0.79)
0.211 (-0.499, 0.921)
0.067 (-0.502, 0.636)
0.025 (-0.621, 0.671)
-0.002 (-0.753, 0.75)
-0.112 (-1.429, 1.205)
-2.141 (-2.964, -1.318)
-4.243 (-5.321, -3.165)
McFadden Pseudo-R^2 = 0.1
16
ROSEANNE
Iconic

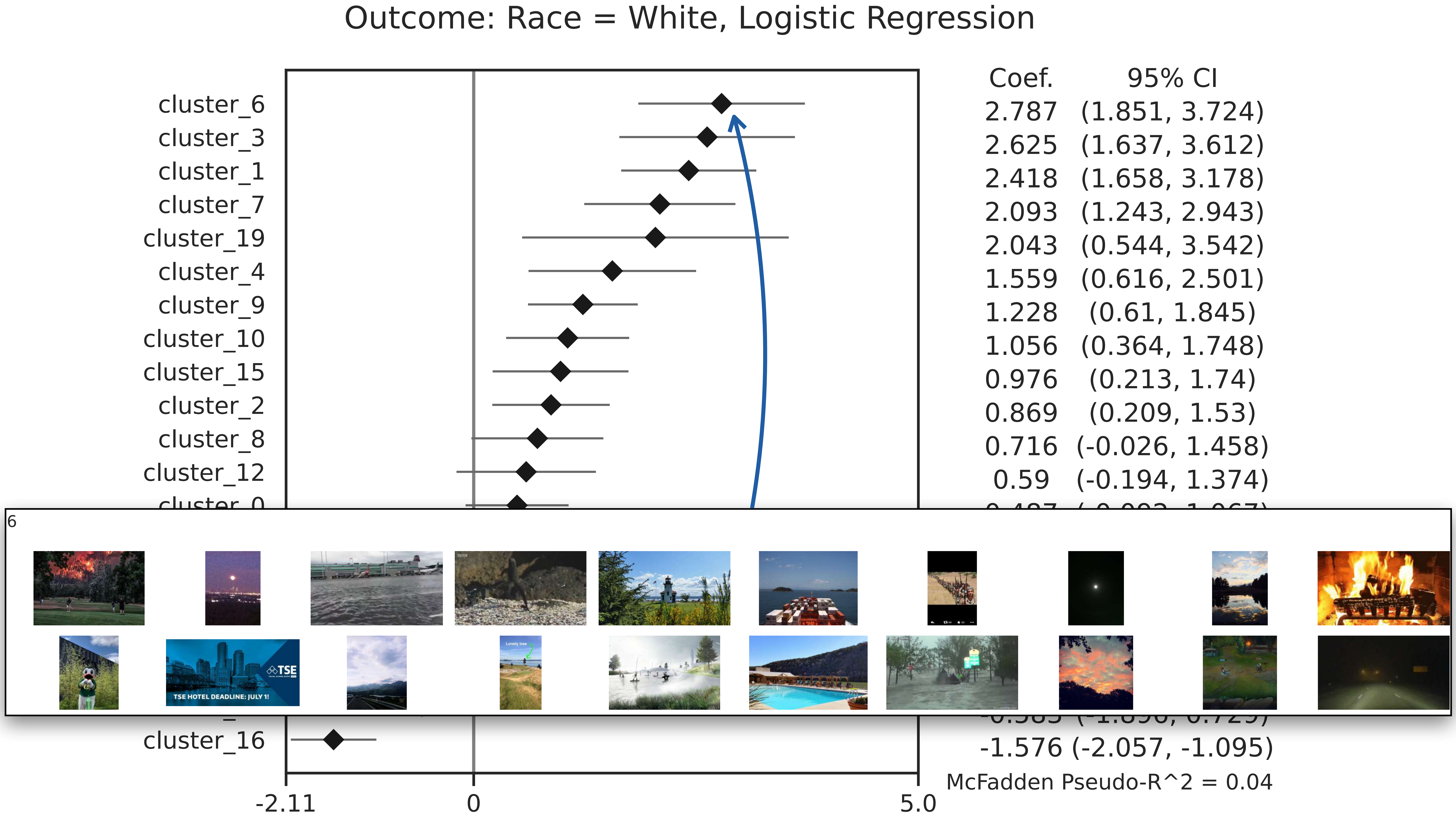

Outcome: Age < 30, Logistic Regression
Coef.
95% CI
cluster_16
4.453 (3.768, 5.139)
cluster_10
0.025 (-0.621, 0.671)
cluster_12
-0.002 (-0.753, 0.75)
cluster_13
-0.112 (-1.429, 1.205)
cluster_14
-0.359 (-0.935, 0.217)
cluster_17
-0.692 (-1.296, -0.089)
cluster_3
-1.643 (-2.492, -0.794)
cluster_8
-1.728 (-2.564, -0.893)
cluster_19
-1.838 (-3.254, -0.421)
cluster_7
-1.896 (-2.687, -1.104)
cluster_9
-1.923 (-2.581, -1.265)
cluster_6
-2.141 (-2.964, -1.318)
cluster_4
-4.243 (-5.321, -3.165)
-5.37
0
6.0
McFadden Pseudo-R^2 = 0.1
4
KAVANAUGH ARRIVES FOR HIS CONFIRMATION
2017 Highlight Reel
OVERLOOKED Recruits
"If Pres. Obama wants to take a victory lap for getting bin Laden, he better realize that it all came from enhanced interrogation by the CIA."
-Eric Bolling

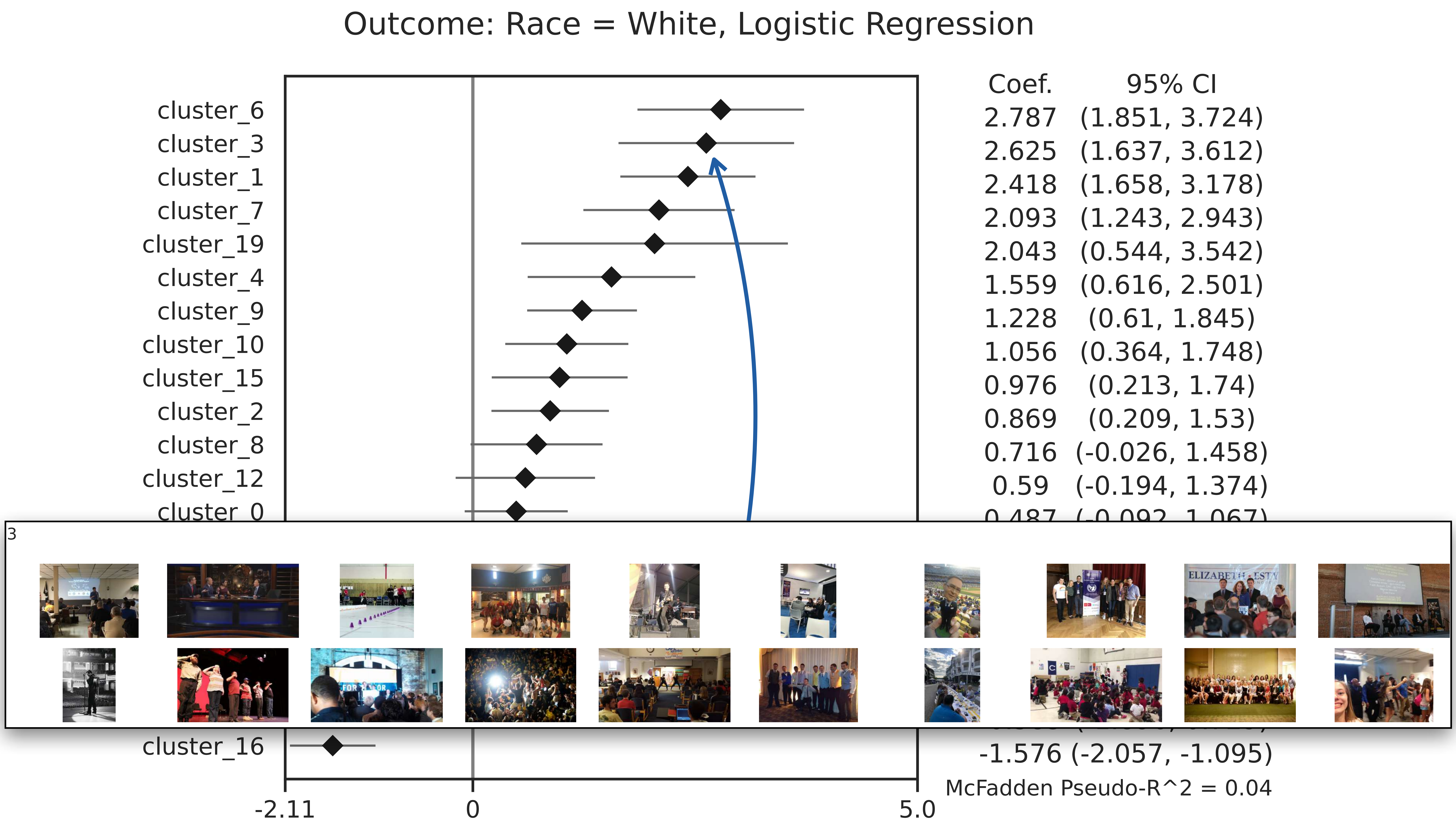
Outcome: Age < 30, Logistic Regression



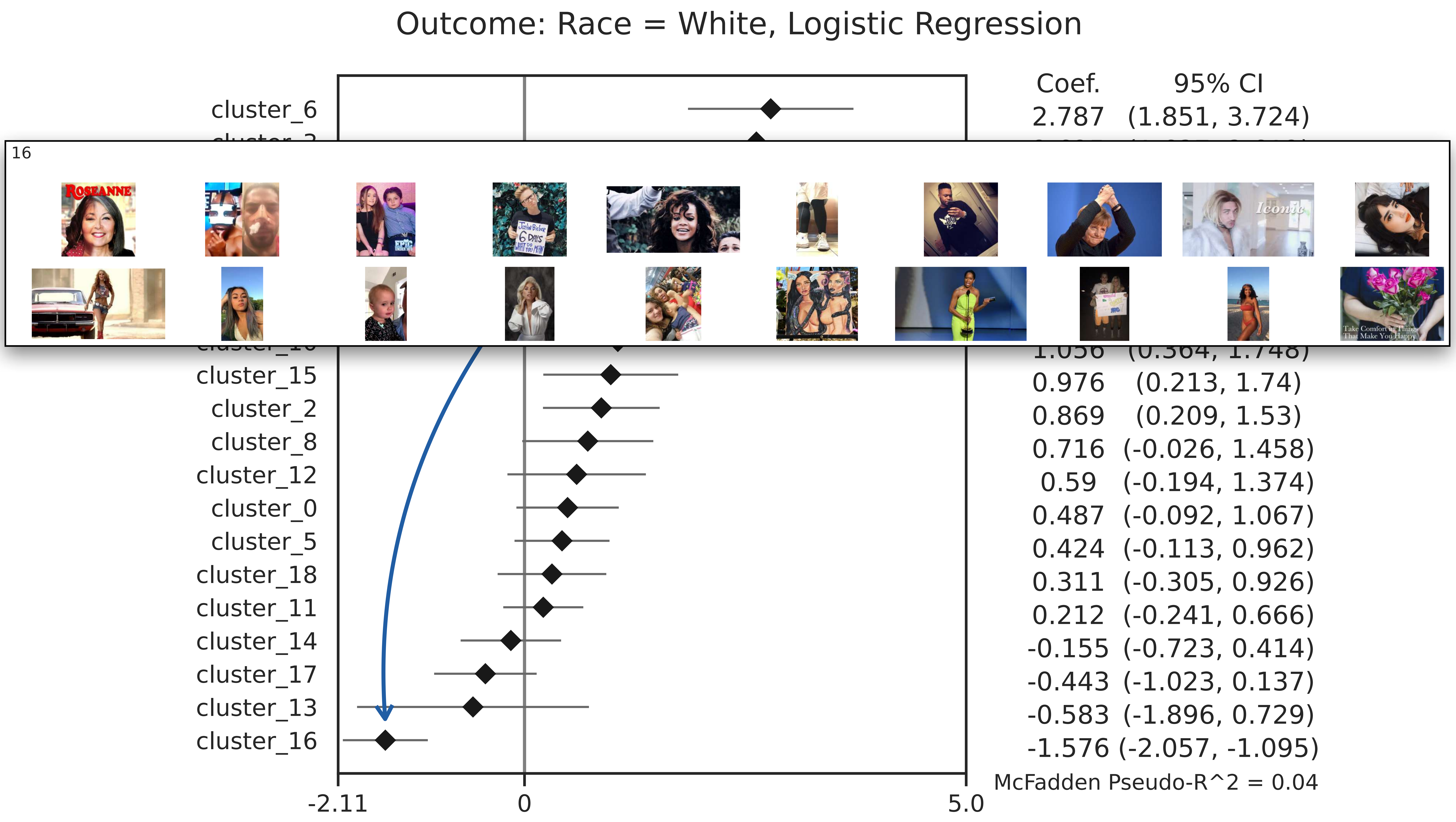


Outcome: Race = White, Logistic Regression
Coef. 95% CI
cluster_6 2.787 (1.851, 3.724)
cluster_3 2.625 (1.637, 3.612)
cluster_1 2.418 (1.658, 3.178)
cluster_7 2.093 (1.243, 2.943)
cluster_19 2.043 (0.544, 3.542)
cluster_4 1.559 (0.616, 2.501)
cluster_9 1.228 (0.61, 1.845)
cluster_10 1.056 (0.364, 1.748)
cluster_15 0.976 (0.213, 1.74)
cluster_2 0.869 (0.209, 1.53)
cluster_8 0.716 (-0.026, 1.458)
cluster_12 0.59 (-0.194, 1.374)
cluster_0 0.487 (-0.092, 1.067)
cluster_5 0.424 (-0.113, 0.962)
cluster_18 0.311 (-0.305, 0.926)
cluster_11 0.212 (-0.241, 0.666)
cluster_14 -0.155 (-0.723, 0.414)
cluster_17 -0.443 (-1.023, 0.137)
cluster_13 -0.583 (-1.896, 0.729)
cluster_16 -1.576 (-2.057, -1.095)
McFadden Pseudo-R^2 = 0.04
-2.11
0
5.0

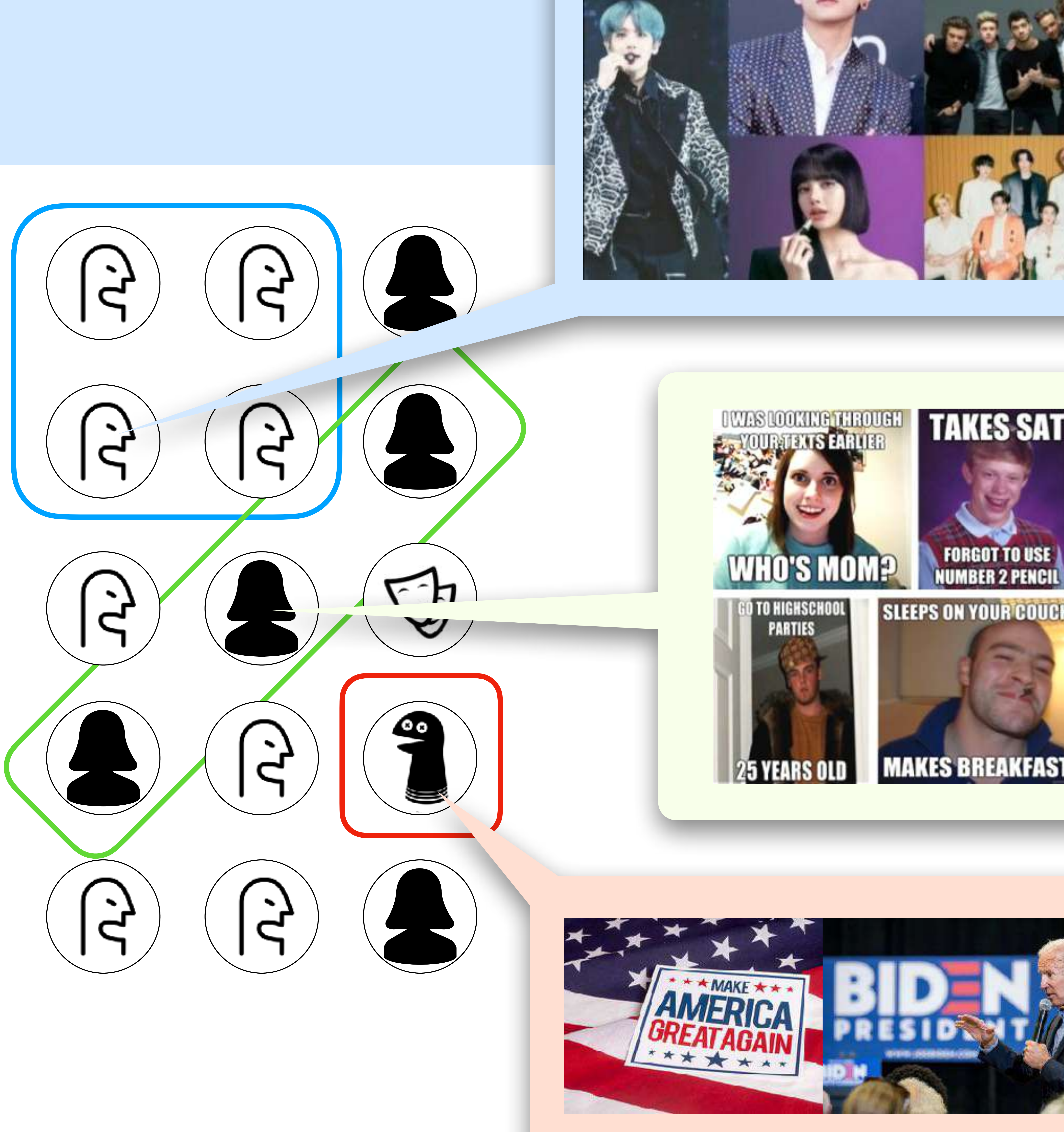

Outcome: Race = White, Logistic Regression
Coef.
95% CI
cluster_6
2.787 (1.851, 3.724)
cluster_3
2.625 (1.637, 3.612)
cluster_1
2.418 (1.658, 3.178)
cluster_7
2.093 (1.243, 2.943)
cluster_19
2.043 (0.544, 3.542)
cluster_4
1.559 (0.616, 2.501)
cluster_9
1.228 (0.61, 1.845)
cluster_10
1.056 (0.364, 1.748)
cluster_15
0.976 (0.213, 1.74)
cluster_2
0.869 (0.209, 1.53)
cluster_8
0.716 (-0.026, 1.458)
cluster_12
0.59 (-0.194, 1.374)
cluster_16
-1.576 (-2.057, -1.095)
McFadden Pseudo-R^2 = 0.04
-2.11
0
5.0
6
TSE
TSE HOTEL DEADLINE: JULY 1!



Outcome: Race = White, Logistic Regression
Coef.
95% CI
cluster_6
2.787 (1.851, 3.724)
cluster_3
2.625 (1.637, 3.612)
cluster_1
2.418 (1.658, 3.178)
cluster_7
2.093 (1.243, 2.943)
cluster_19
2.043 (0.544, 3.542)
cluster_4
1.559 (0.616, 2.501)
cluster_9
1.228 (0.61, 1.845)
cluster_10
1.056 (0.364, 1.748)
cluster_15
0.976 (0.213, 1.74)
cluster_2
0.869 (0.209, 1.53)
cluster_8
0.716 (-0.026, 1.458)
cluster_12
0.59 (-0.194, 1.374)
cluster_0
3
ELIZABETH ESTY
cluster_16
-1.576 (-2.057, -1.095)
McFadden Pseudo-R^2 = 0.04
-2.11
0
5.0


Outcome: Race = White, Logistic Regression
Coef. 95% CI
cluster_6 2.787 (1.851, 3.724)
16
ROSEANNE
Iconic
Take Comfort in Things That Make You Happy
1.056 (0.364, 1.748)
cluster_15 0.976 (0.213, 1.74)
cluster_2 0.869 (0.209, 1.53)
cluster_8 0.716 (-0.026, 1.458)
cluster_12 0.59 (-0.194, 1.374)
cluster_0 0.487 (-0.092, 1.067)
cluster_5 0.424 (-0.113, 0.962)
cluster_18 0.311 (-0.305, 0.926)
cluster_11 0.212 (-0.241, 0.666)
cluster_14 -0.155 (-0.723, 0.414)
cluster_17 -0.443 (-1.023, 0.137)
cluster_13 -0.583 (-1.896, 0.729)
cluster_16 -1.576 (-2.057, -1.095)
McFadden Pseudo-R^2 = 0.04
-2.11
0
5.0

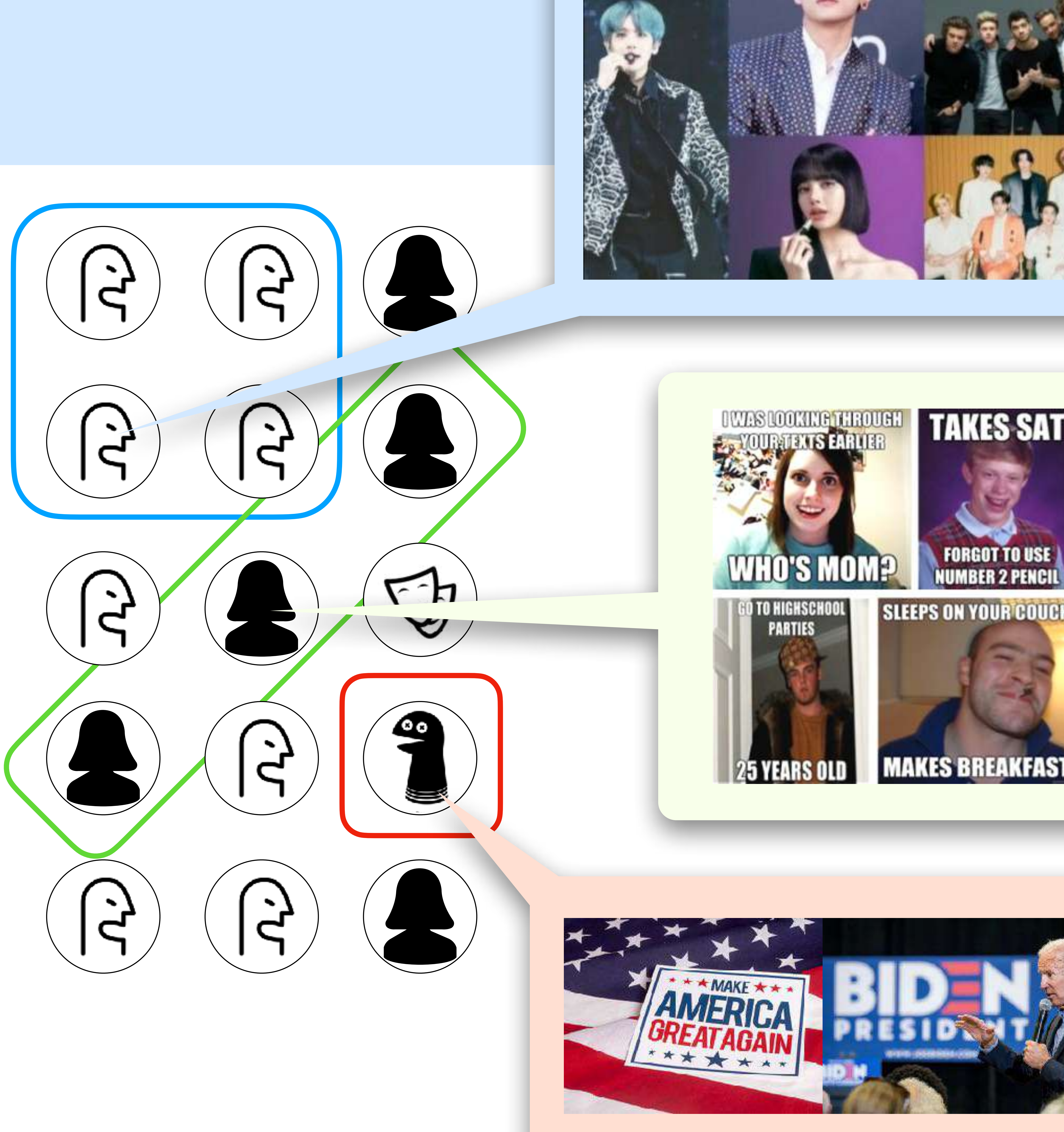
Findings



I WAS LOOKING THROUGH YOUR TEXTS EARLIER
WHO'S MOM?
FORGOT TO USE NUMBER 2 PENCIL
GO TO HIGHSCHOOL PARTIES
25 YEARS OLD
MAKE AMERICA GREAT AGAIN

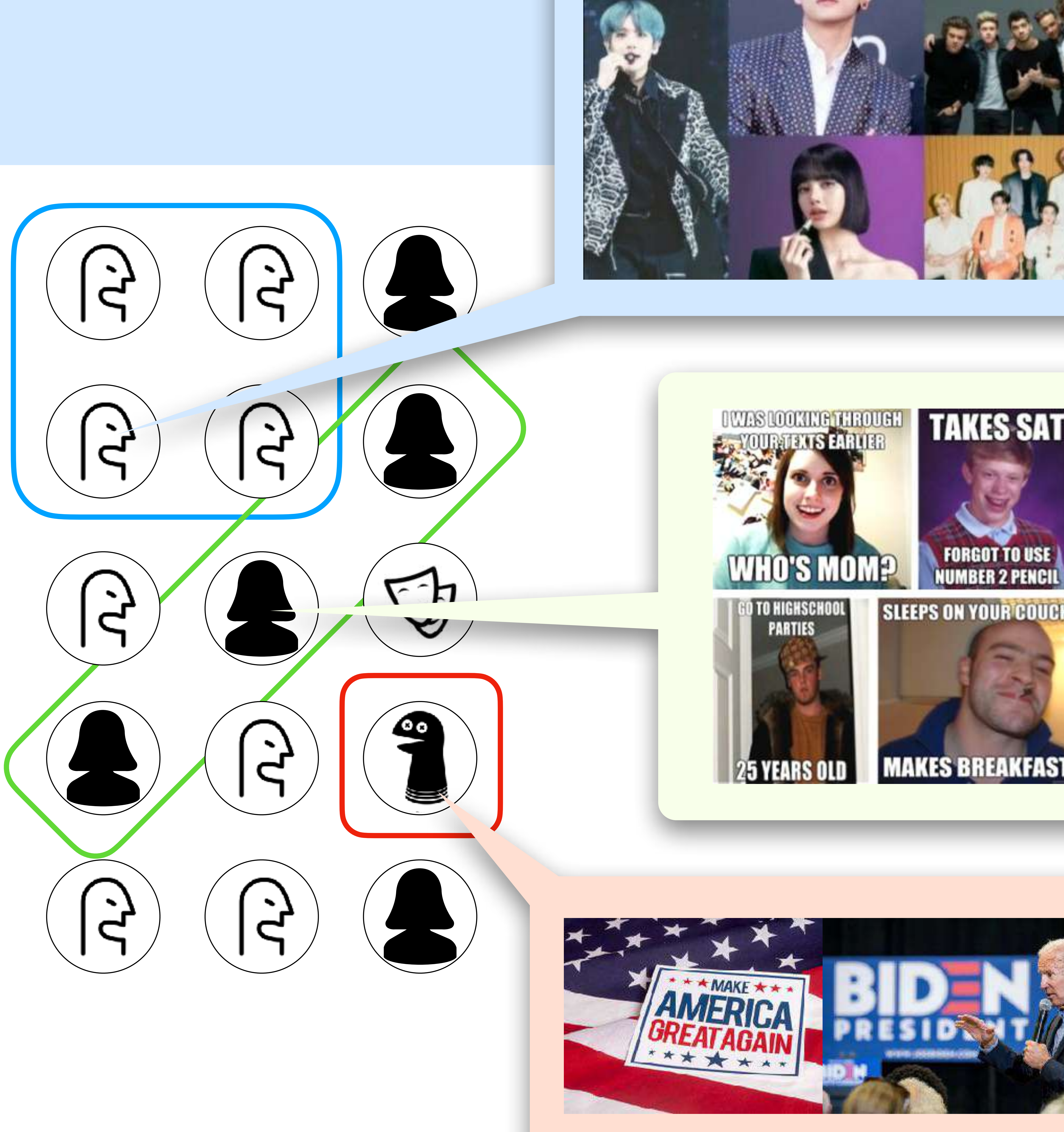
Findings

- Politically engaged and general audiences post largely similar distributions of political imagery



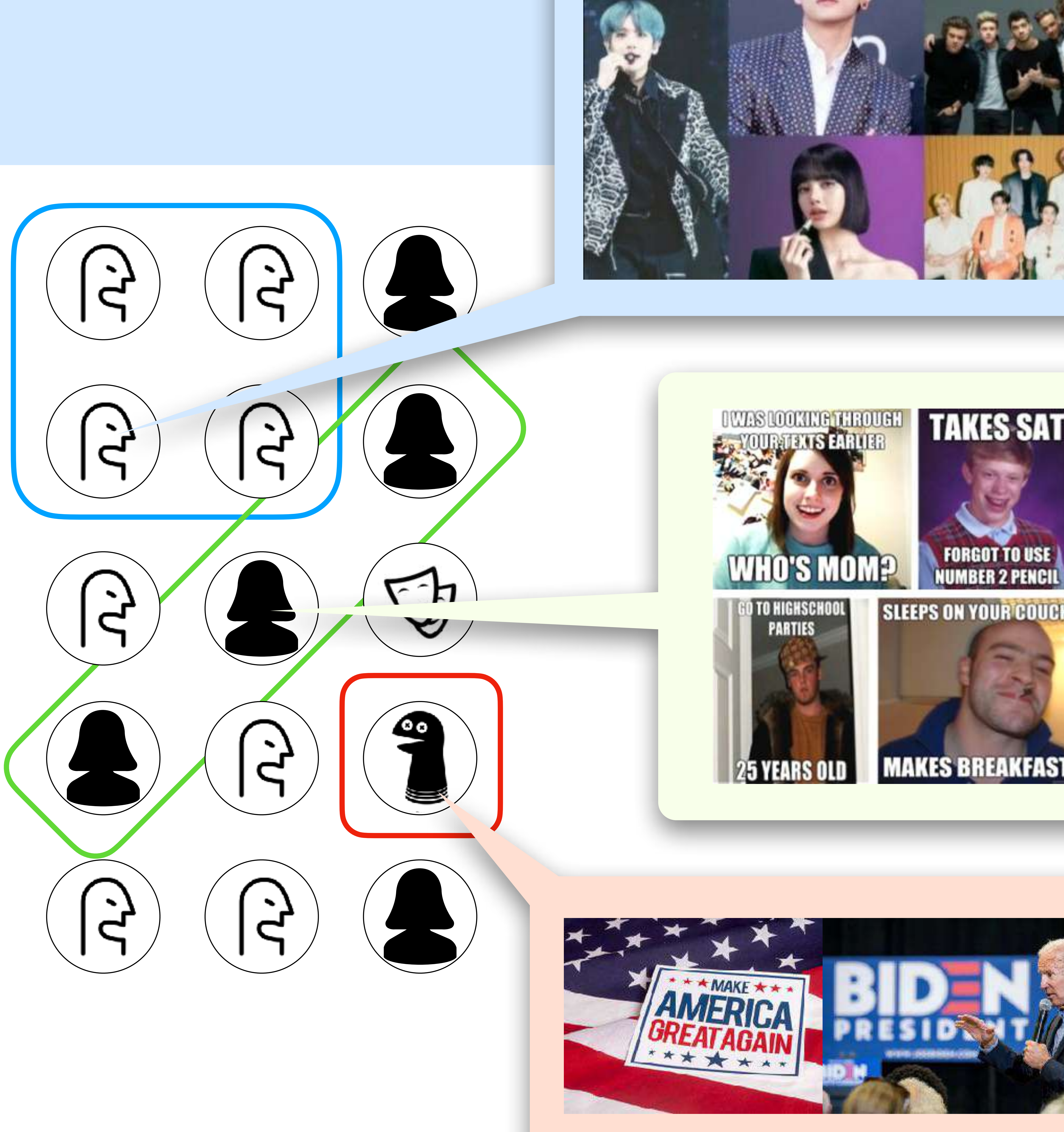
Findings

- Politically engaged and general audiences post largely similar distributions of political imagery
- Overall, around half of the clusters contain predictive information about the account's race, gender, age, and political engagement



Findings

- Politically engaged and general audiences post largely similar distributions of political imagery
- Overall, around half of the clusters contain predictive information about the account's race, gender, age, and political engagement
- Implications for "content-based" information targeting



Thank you!   Questions?

Keng-Chi Chang · @kengchichang · kechang@ucsd.edu
Cody Buntain · @codybuntain · cbuntain@umd.edu