Ad Goal Results

Writeup of current results. Intentionally not formatted to ACM standard to show off larger figures more easily.

Data

Currently, all 377,721 ads from 2022 Facebook are included. This includes candidates, outside groups, and then also a few sponsors such as akoreanshop.com that probably shouldn't be in there.

Open questions: Should we also include 2020? For the analysis of purchase ads, it would be interesting to see how they compare to 2020 when they weren't banned yet. Also, our training data is from 2020, so reviewers will know that we have the data. Also, what about Google (note that we don't have training data, of course)?

Model

The models are binary random forest classifiers for each goal. For publication in a CS conference, we should probably do this at least with a BERT model instead. LLMs are sort of debatable here – on the one hand, prediction is not their strong suit and BERT and friends are still used in that domain, but LLMs are good at few-shot inference, and since we only have a training set of about 3k ads or so, this probably falls under that category. But that might be more trouble than it's worth.

Performance scores are here, I'll import them into Latex when we have our final models.

There is an important distinction here to how we've treated ad goal so far. We always had individual binary classifiers for each goal, but the variable we mostly used was an aggregated one, where an ad could only ever have one goal, determined by which goal had the highest probability. But in doing so we're throwing away information - for example it is interesting to see which ads are GOTV but not persuade because they tend to be from non-partisan groups just asking citizens to vote, without advocating for either side (see below for that). Also, it's not quite fair – persuade is by far the most common category, so an ad that has a 0.95 probability of being donate is MUCH more distinctly a donate than an ad with 0.95 probability of being persuade is a persuade ad, since the persuade model is much more prone to giving high scores. So we end up overpredicting persuade. And finally, this is not political science, so it's not like it matters that the different categories might be a little iffy for a regression model. Note that what I'm doing here effectively amounts to a multilabel model (which we use for issues), which we also might have done instead. But given the extremely high correlation between Learnmore and Persuade (see below), I'm wondering whether that might cause it to overpredict those ("this looks like a Learnmore ad - so it must be a Persuade ad too!").

Basic Descriptive Statistics

See Table 1. 71% of ads are persuade, and 63% are learnmore. That's the vast bulk of the ads. Then donate makes up 14%. I also created the category Nogoals for when all goal variables for an ad are 0. I noticed that a lot of these are about Ukraine (so it makes sense that they have no goal aimed at the American elections), though Warnock also has some. Contact is super rare. Despite the fact that Facebook banned them, some purchase ads do still exist (see below).

The third and fourth column of the table are spend. Note that when an ad had multiple goals, I evenly distributed the ad's spend between them. This is part of why 71% of ads are Persuade, but only 41% of ad spending is on Persuade - because it gets split between Learnmore and Persuade, whereas for the counts, I count both separately.

	Ad Count	Proportion Ads	Spend	Proportion Spend
DONATE	53493.00	0.14	20366746.00	0.07
CONTACT	58.00	0.00	43788.33	0.00
PURCHASE	4830.00	0.01	1958006.42	0.01
GOTV	17867.00	0.05	6101780.58	0.02
EVENT	551.00	0.00	89482.38	0.00
POLL	2328.00	0.01	1213016.92	0.00
GATHERINFO	10045.00	0.03	12538416.12	0.05
LEARNMORE	238335.00	0.63	94925214.17	0.35
PERSUADE	268859.00	0.71	113940942.08	0.41
NOGOALS	33807.00	0.09	23636496.50	0.09

Table 1: Ad count and spend, by goal.

Purchase Ads

Page	Spend	Proportion Spend
Akoreanshop.com	474999.50	0.20
Republican Dogs	272975.00	0.11
Screaming Freedom Apparel	178484.50	0.07
Defend The Second	120848.00	0.05
8 Billion Trees	118840.50	0.05
Proud Patriots	104731.50	0.04
Texas Precision Optics, Inc.	84999.50	0.04
Behavioral Health Center Abuse	84997.00	0.04
HealthInsurance.net	70678.50	0.03
Thien Moc Huong - Agarwood Jewelry	64999.50	0.03
I Love My Freedom	61531.00	0.03
Newsmax	54988.00	0.02
ObamacarePlans.com	44606.00	0.02
Ben Shapiro	42499.50	0.02
Flag and Cross	41840.50	0.02

Table 2: Purchase ad sponsors, sorted by their spend.

See Table 2. Even though purchase ads are now banned, there are still some advertisers running them. Some of them clearly aren't political advertisers, but others are. Most of them seem conservative/MAGA, but some Democratic ones like Beto O'Rourke and AOC are also on the list, a little further down than the top 15 shown here. Obviously, we could get a lot more interesting results through a comparison with 2020.

Sponsors

See Figure 1 for the goal distribution of top sponsors (according to ad spend). Almost everyone has lots of Persuade/Learnmore. Noteworthy that the ones who have substantial Donate ads are almost all candidates.

Figures 2-4 show top sponsors of donate, GOTV and Persuade ads (I excluded the rest because of the size of the figures). Notable result: the third highest sponsors of GOTV ads is 'Meta'?

I looked at sponsors who run ads that are GOTV ads but not persuade (Table 3). The background being that in some other democracies, it's not unheard of for elected officials to call on people to vote, no matter who they're going to vote for, just in the name of electoral participation. But in the U.S., my impression is that this is much more rare, and this table seems to bear that out – most of the sponsors that do so are nonpartisan groups. There are a few Democrats in there though (and I don't think it would be too much of a stretch to interpret that as them being more committed to democracy), but no Republicans as far as I can tell.

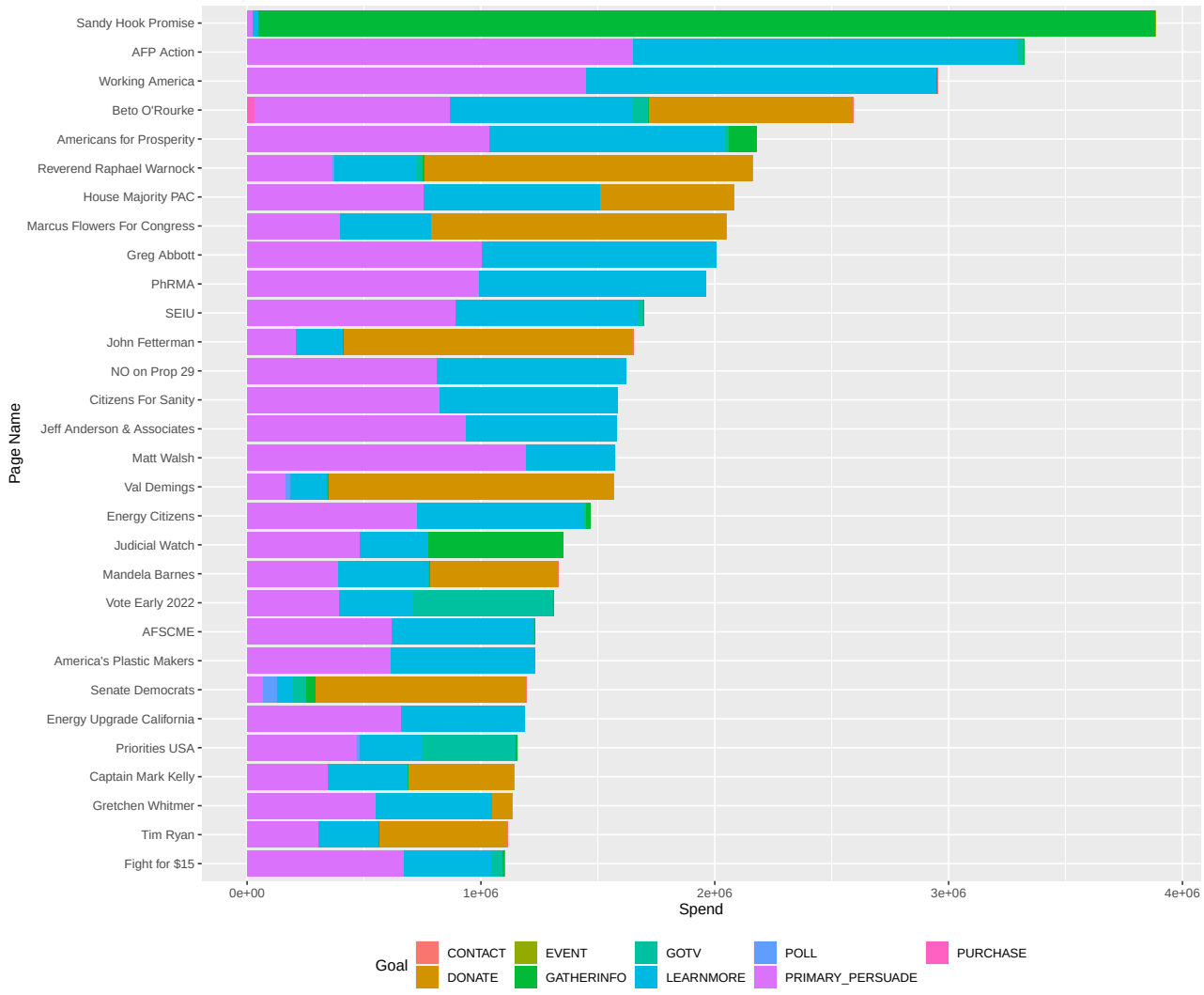

Figure 1: Goal by sponsor.

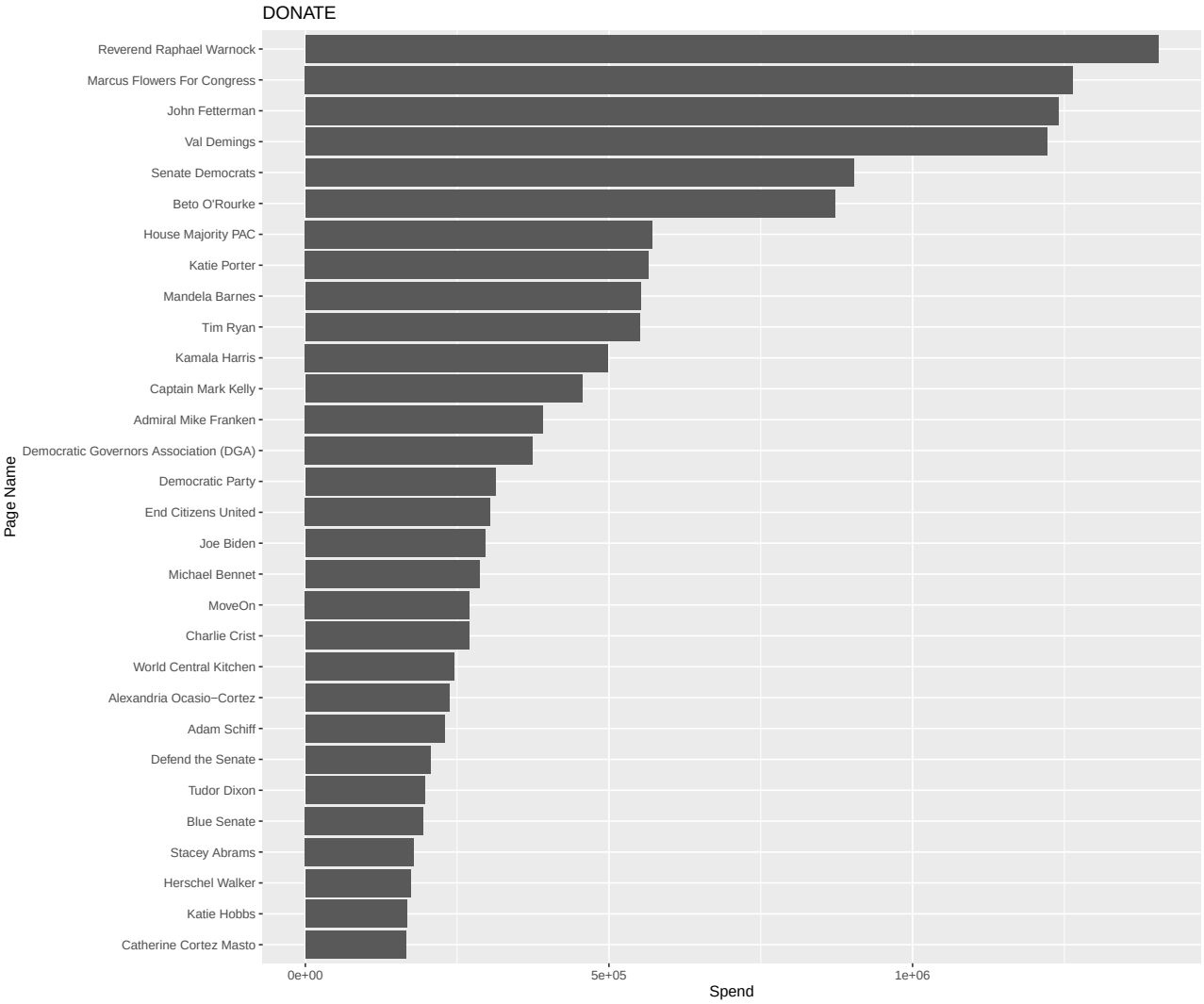

Figure 2: Top sponsors of donate ads.

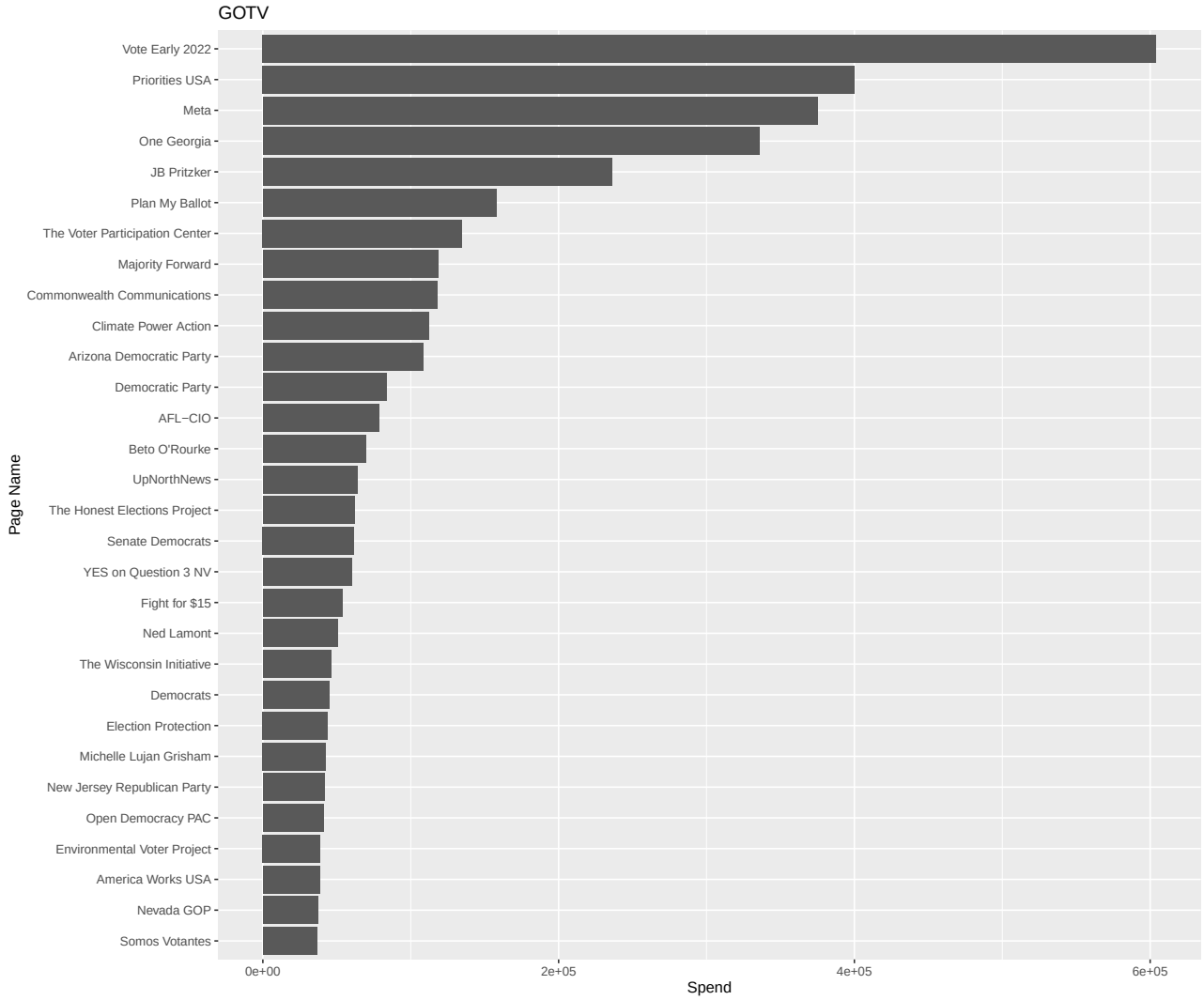

Figure 3: Top sponsors of GOTV ads.

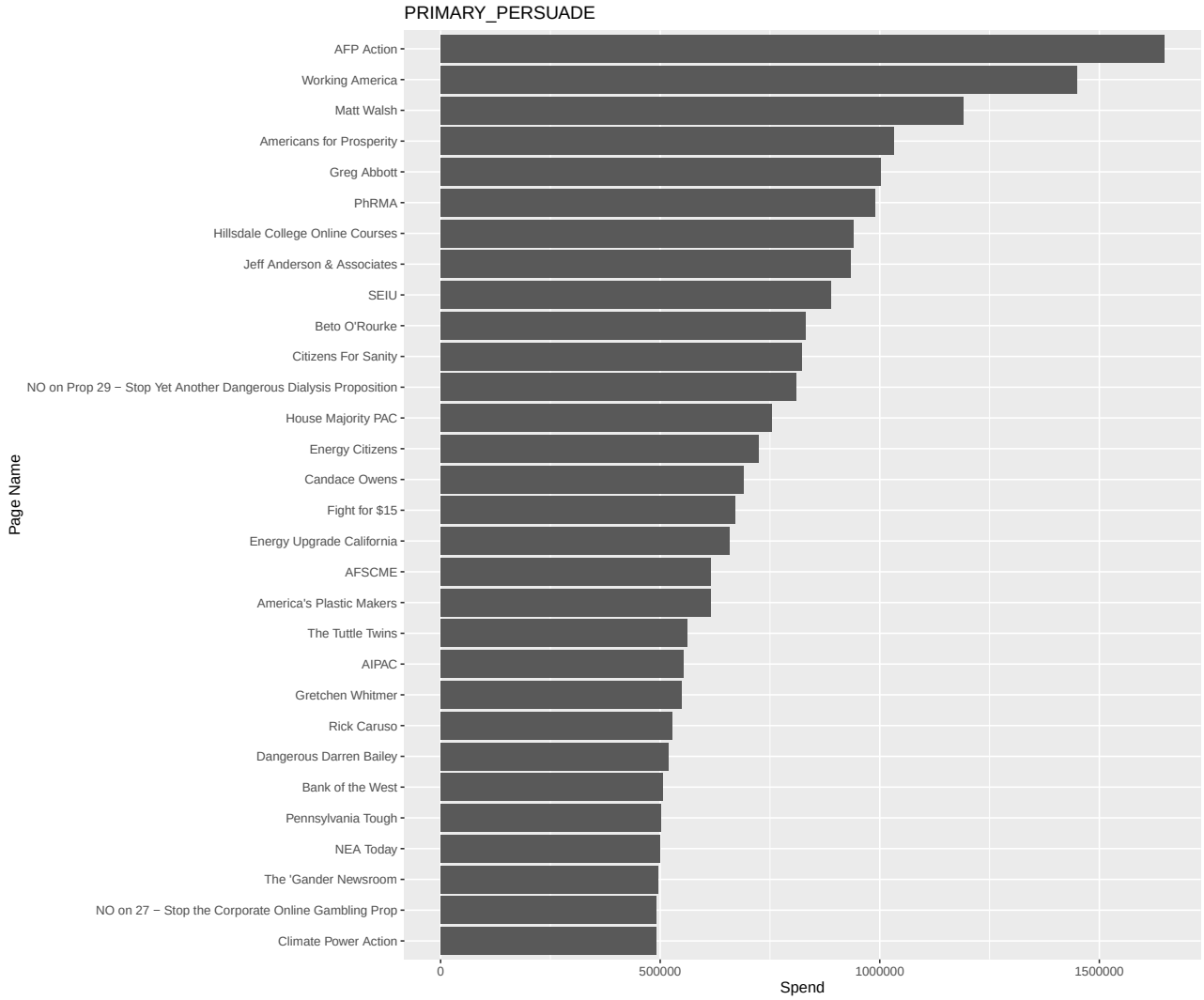

Figure 4: Top sponsors of persuade ads.

Sponsor	Ads
Vote Early 2022	424
One Georgia	216
Priorities USA	120
Power the Polls	119
The Honest Elections Project	65
The Voter Participation Center	53
Plan My Ballot	42
VOTEPROCHOICE	42
Democratic Party	36
CLEAR For PA	32
ProgressVA	32
redefinED atlanta	32
Speak Up For Education and Kids	27
JB Pritzker	25
Progress Georgia	25

Table 3: Sponsors who run ads that are GOTV without also being Persuade.

Age

See Figure 5. We've never really looked at age before (as far as I'm aware), but there are quite big differences here. Basically, it's textbook resource model of political participation – donation ads are targeted at older people, while GOTV ads have the youngest average. Also note the bump at the younger end of the Poll distribution (maybe Gen Z likes taking online polls or something?).

Note that I created the age data by using the demographic distribution, summing over genders, and then using the mean of every age category. So 29.5 for 25-34, 39.5 for 35-44, and so on. For 65+ I used 69.5. Also note that there is a 13-17 age group, which is interesting because they can't even vote (and I imagine there might even be some legal restrictions for what can be targeted at them?). But they do seem to be getting a few Gather Info ads.

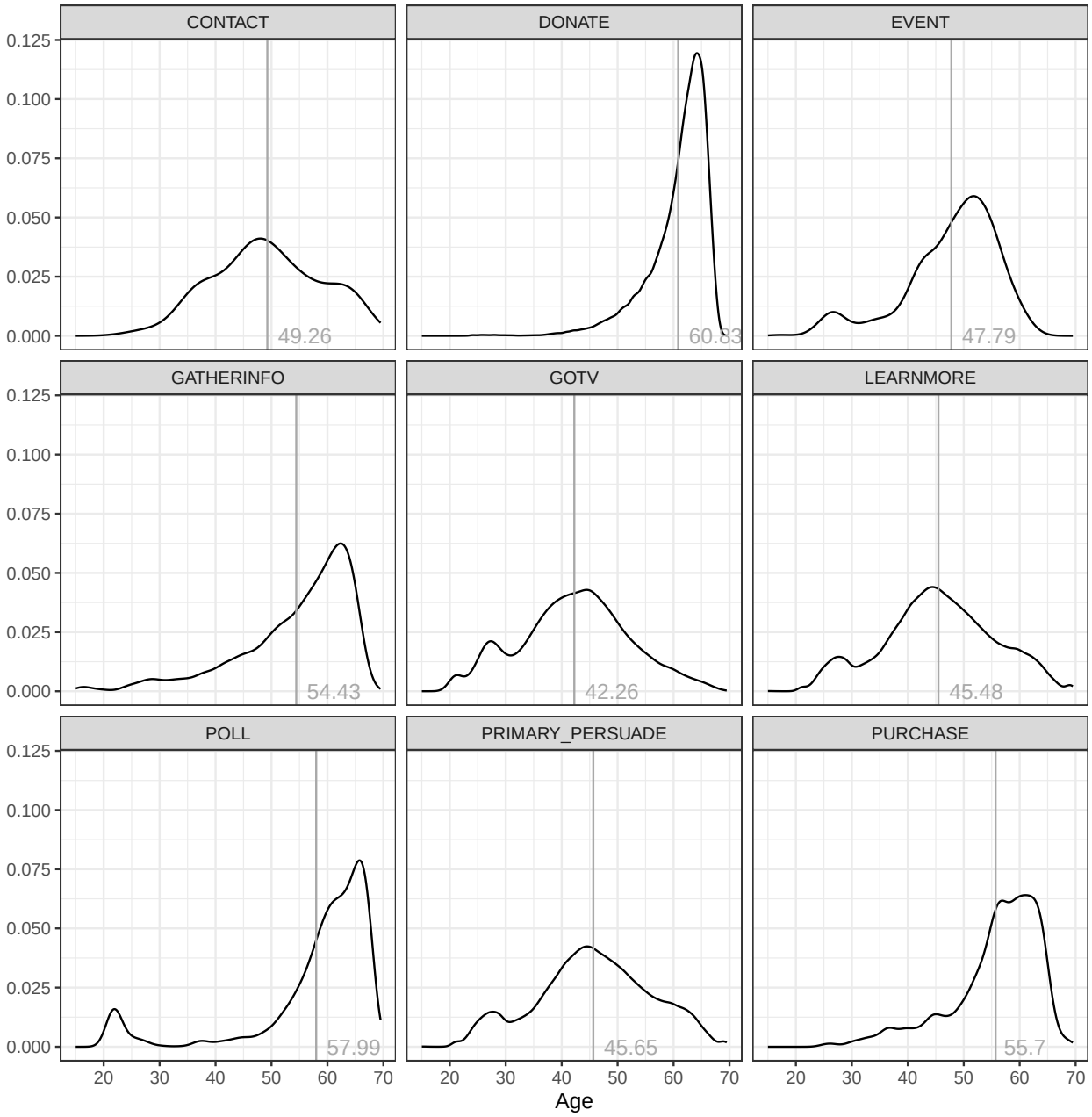

Figure 5: Density of ad target age, by goal. The vertical grey line and the number alongside it is the mean.

Ad Text

Table 4 shows the words most associated with each goal (using the method described in Monroe et al. (2008), see also Jurafsky and Manning for an introduction to this method aimed at NLP audiences). ((Sometimes these words are emojis, like the American flag in purchase. These are unicode and not displayed correctly by Latex (currently just showing whitespace), we'll have to figure that out later)).

For donate, the top words are unsurprising as they relate directly to donating. secure.actblue is the primary donation platform of the Democratic party, so that also makes sense. A little further down are the words race and deadline, indicative of a common strategy in donation ads – asking for money quickly in order to meet a fundraising deadline.

Contact ads are rare, so the words here are very directly related to the few campaigns that do run them, especially Maggie Hassan.

Purchase ads seem to be primarily focus on selling MAGA-themed swag.

GOTV ads, unsurprisingly focus on voting. Noteworthy here is how high 'early' voting is on the list. This is likely related to the Democratic strategy of focusing on early and mail-in (mail is the 8th word) voting in the context of Donald Trump's voter fraud claims. November 8th and Tuesday are on the list too though, indicating a focus on election-day voting as well.

The top words for events are largely unsurprising. For poll, I am rather curious why Exxon appears three times. Gatherinfo is a bit all over the place. Learnmore and persuade have effectively the same top words, which is unsurprising given that we have a lot of evidence that they're basically the same thing (see below).

Finally, nogoals focuses a lot on the Ukraine war.

donate	contact	purchase	gotv	event	poll	gatherinfo	learnmore	persuade	nogoals
donate	hassan	trump	vote	saturday	survey	sign	vote	vote	unhcr
help	tell	limited	early	rally	poll	add	paid	state	now
chip	www.onenationamerica.org	edition	voting	tickets	response	name	state	paid	poll
$	nation	proud	ballot	greet	take	petition	de	county	donate
now	3495		election	rsvp	exxonmobil	demand	district	district	trump
can	sen	patriots	november	meet	responses	marriage	representative	city	$
>	202-224-3324	2024	georgia	join	official	censorship	8th	8th	urgent
secure.actblue.com	inflation	claim	mail	pm	approve	>	candidate	8	ukraine
campaign	resort	flag	plan	ticket	missing	alliance	city	school	races
beto	reckless	dogs	8th	september	recycling	penn	8	representative	emergency
win	operas	hat	day	attention	approval	survivors	authorized	de	refugees
donation	onenationamerica.org		make	trump	exxon	signatures	house	council	difference
race	prevention		tuesday	13570	exxchange	action	county	early	state-level
deadline	4589	free	find	secure.winred.com	mobil	now	la	police	america
need	dc	printed	person	arlington	corporation	greene	council	nov	across
defeat	french	ships	union	updates	respond	tell	y	candidate	update
today	spending	shirt	staceyabrams.com	october	commonally	act	police	plan	breaking
grassroots	s	confirm	october	joint	biden	biden	el	voted	country
democratic	maggie	extremely	return	>	president	moveon.org	union	local	help
goal	140	christmas	one	event	>	protect	voted	judge	close

Table 4: Top 20 words most associated with a given goal.

Correlation between goals

The result that stands out here (see Figure 6) is that Persuade and Learnmore ads (which, mind you, make up about 70% of all ads) have a .94 Pearson's R correlation. That's definitely a bit strange. The picture also doesn't change much if we look at co-occurrences of predictions rather than predicted probabilities (see Table 5). Out of 377k ads. 234k are both Learnmore and Persuade. This is also (mostly) not the fault of a model overpredicting the most common category, we see a similar pattern in the training data, even if not quite as pronounced (Table 6). GOTV also has a bit of a correlation with Persuade (and Learnmore), which is not that surprising given that most GOTV ads aim to get people to get out and vote for a specific candidate, which is part of the Persuade definition. Also, every single one of the extremely rare contact ads is also Learnmore and Persuade (in the inference set). We should consider whether any of this is a problem. Also, remember that Learnmore is a little weird anyway since the coders see information the model can't – the Learn More button. Doesn't seem to hinder the model much, though. But for that reason we had excluded it in the fall.

	DONATE	CONTACT	PURCHASE	GOTV	EVENT	POLL	GATHERINFO	LEARNMORE	PERSUADE
DONATE	403	0	34	12	147	84	74	10	43
CONTACT	0	52	0	0	0	0	0	52	52
PURCHASE	34	0	711	0	1	0	0	611	389
GOTV	12	0	0	15877	4	0	5	11779	15717
EVENT	147	0	1	4	422	0	0	185	264
POLL	84	0	0	0	0	436	6	198	311
GATHERINFO	74	0	0	5	0	6	1162	826	580
LEARNMORE	10	52	611	11779	185	198	826	234736	233749
PERSUADE	43	52	389	15717	264	311	580	233749	238428

Table 5: Goal co-occurrences.

	DONATE	CONTACT	PURCHASE	GOTV	EVENT	POLL	GATHERINFO	LEARNMORE	PERSUADE
DONATE	572	0	2	11	0	2	1	24	58
CONTACT	0	43	0	0	0	1	7	22	13
PURCHASE	2	0	174	2	0	1	2	17	2
GOTV	11	0	2	308	2	1	36	165	215
EVENT	0	0	0	2	26	0	4	2	2
POLL	2	1	1	1	0	64	0	3	6
GATHERINFO	1	7	2	36	4	0	213	20	44
LEARNMORE	24	22	17	165	2	3	20	1494	1242
PERSUADE	58	13	2	215	2	6	44	1242	1684

Table 6: Goal co-occurrences, training set.

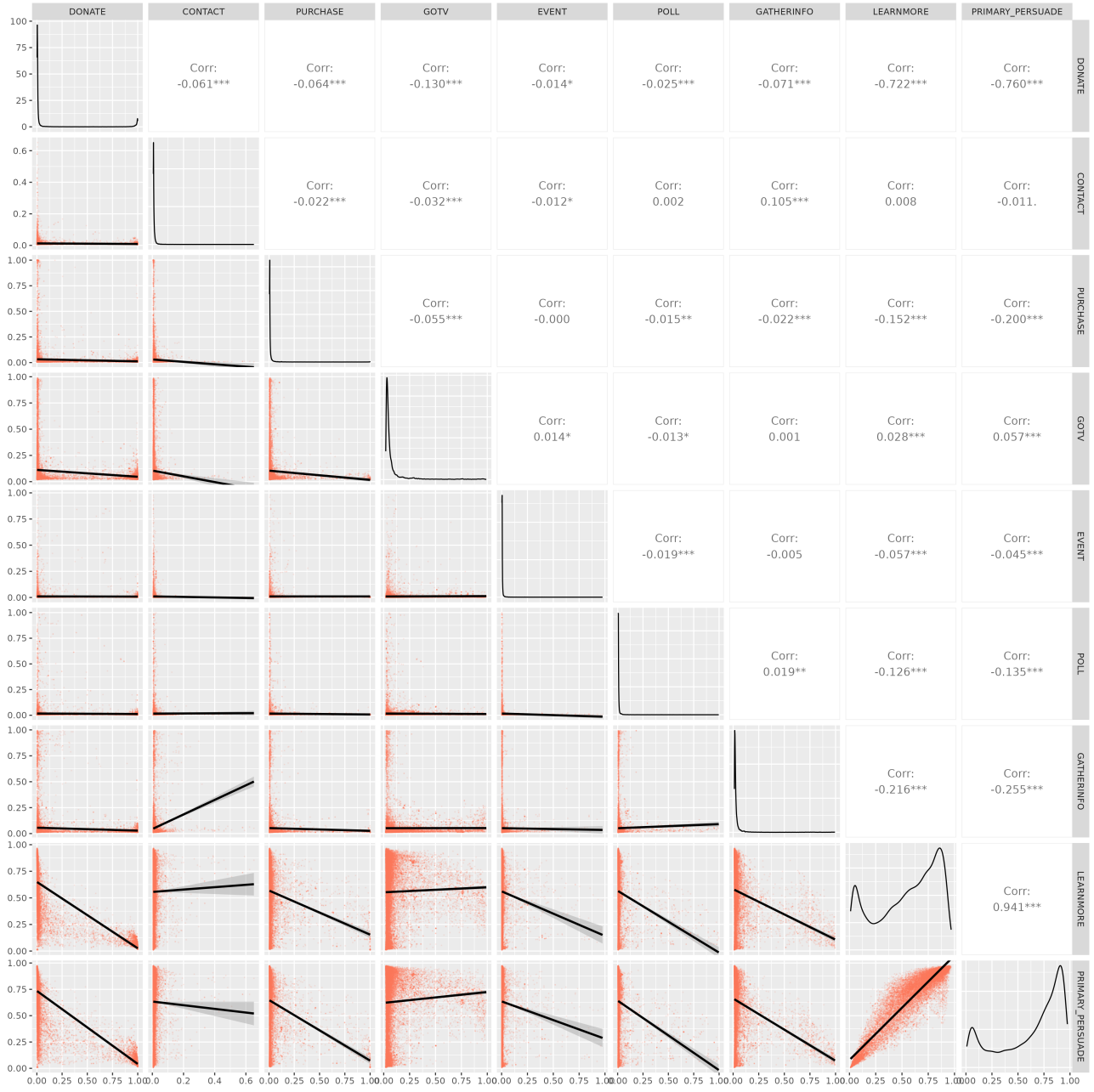

Figure 6: Correlations between goals. Densities by goal are on the diagonal.