Figure A.1: Household income and racial background demographics for the comScore versus census data from the same period (CPS 2017 [35]).

A APPENDIX

A.1 comScore Data Demographics Compared to Census Estimates

Before preprocessing we compare the demographics reported in the comScore data to census population estimates from the same time period (2017) from the U.S. Census Bureau American Community Survey (ACS) [34] and Current Population Survey (CPS) [35]. We compare estimates at the household level. These comparisons are shown in Figures A.2 and A.1. When comparing the distribution of the comScore population by U.S. state there is a Pearson correlation of 0.988 (p=0.000). When comparing racial background and household income groups there is a Pearson correlation of 0.979 (p = 0.021) and 0.971 (p=0.029), respectively. Note that even though the data are highly correlated with U.S. demographics with respect to racial background and household income, we create data panels that are even more representative of the U.S. population for use in our t-closeness analyses that evaluate the relationships between demographics and cohort groupings. The data panels are created using stratified random sampling (without replacement), where strata are defined by the joint distribution of the racial background and household income demographic groups.

A.2 Distribution Of The Unique Set Of Domains Per Machine, Per Week

Figure A.3 shows a box plot of unique domains per machine for each week in our dataset (shown up to 50). Overall the median number of domains per week is less than the 7-domain cutoff, shown by a dashed horizontal line

A.3 Domain Visit Frequencies By Race And Income

In order to satisfy the comScore data terms of use, we avoid showing web traffic data connected to named domains.

Table A.1 shows the relative visit frequencies for the top 100 domains, for the aggregated overall population and by subpopulations partitioned by racial background, and then by household income group. These data represent those used in the Chi-Squared tests described in Section 5.1 that are used to test browsing differences between demographic groups. Domains are ordered by their frequency in the overall population, in descending order. Domain names are not shown. Table A.2 shows the names of the top D=100 domains in alphabetical order.
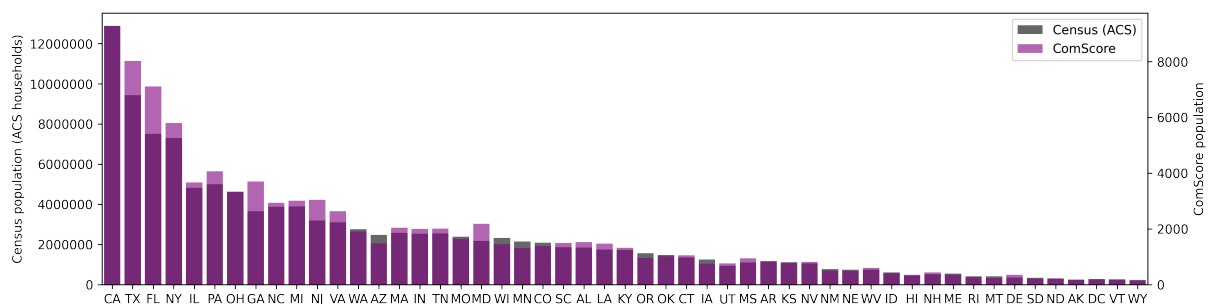
Figure A.2: State population for the comScore versus census data from the same period (CPS 2017 [35]).
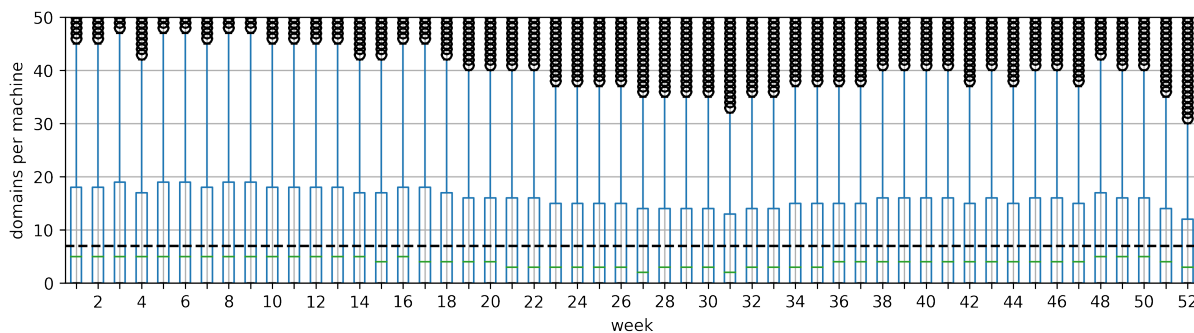

Figure A.3: Box plot of unique domains per machine for each week in our dataset (shown up to 50). Overall the median number of domains per week is less than the 7-domain cutoff, shown by a dashed horizontal line.

Table A.1: Relative visit frequencies for the top 100 domains, for the overall population and by subpopulations partitioned by demographic group. Values for the randomized comparison group used in the Chi-squared tests are also shown. Domains are ordered by their frequency in the overall population, in descending order.

domain	overall	random control	white	black	asian	other	less than $25,000	$25,000 - $75,000	$75,000 - $150,000	$150,000 or more
1	2.32%	2.31%	2.20%	2.44%	2.52%	2.75%	2.36%	2.20%	2.42%	2.54%
2	1.54%	1.54%	1.49%	1.76%	1.45%	1.63%	1.64%	1.52%	1.55%	1.43%
3	1.54%	1.52%	1.50%	1.87%	1.07%	1.65%	1.60%	1.53%	1.55%	1.41%
4	1.51%	1.50%	1.49%	1.72%	1.13%	1.60%	1.49%	1.50%	1.56%	1.42%
5	1.38%	1.37%	1.40%	1.35%	1.23%	1.42%	1.53%	1.37%	1.36%	1.19%
6	1.26%	1.25%	1.09%	1.48%	1.62%	1.65%	1.45%	1.18%	1.18%	1.35%
7	1.11%	1.10%	1.05%	1.31%	0.95%	1.30%	1.17%	1.09%	1.10%	1.11%
8	0.85%	0.84%	0.89%	0.67%	0.90%	0.82%	0.74%	0.81%	0.98%	0.96%
9	0.37%	0.37%	0.40%	0.27%	0.34%	0.37%	0.35%	0.37%	0.41%	0.34%
10	0.34%	0.33%	0.32%	0.38%	0.30%	0.38%	0.35%	0.33%	0.34%	0.33%
11	0.31%	0.31%	0.31%	0.33%	0.27%	0.33%	0.33%	0.31%	0.30%	0.30%
12	0.30%	0.30%	0.34%	0.24%	0.18%	0.26%	0.23%	0.30%	0.36%	0.33%
13	0.29%	0.29%	0.29%	0.22%	0.45%	0.30%	0.27%	0.27%	0.33%	0.39%
14	0.28%	0.28%	0.27%	0.30%	0.26%	0.33%	0.33%	0.26%	0.27%	0.29%
15	0.26%	0.26%	0.29%	0.19%	0.19%	0.27%	0.28%	0.27%	0.27%	0.19%
16	0.25%	0.25%	0.27%	0.23%	0.20%	0.23%	0.27%	0.26%	0.26%	0.18%
17	0.25%	0.25%	0.27%	0.22%	0.21%	0.24%	0.22%	0.25%	0.29%	0.26%
18	0.25%	0.25%	0.26%	0.22%	0.22%	0.24%	0.21%	0.24%	0.29%	0.24%
19	0.24%	0.24%	0.26%	0.18%	0.20%	0.24%	0.22%	0.24%	0.27%	0.24%
20	0.22%	0.22%	0.20%	0.31%	0.21%	0.25%	0.26%	0.22%	0.21%	0.19%
21	0.21%	0.21%	0.15%	0.37%	0.28%	0.34%	0.27%	0.19%	0.18%	0.28%
22	0.20%	0.20%	0.20%	0.23%	0.21%	0.20%	0.19%	0.19%	0.22%	0.23%
23	0.20%	0.20%	0.19%	0.19%	0.24%	0.21%	0.18%	0.18%	0.21%	0.27%
24	0.20%	0.20%	0.19%	0.24%	0.21%	0.22%	0.23%	0.21%	0.19%	0.17%
25	0.18%	0.18%	0.19%	0.13%	0.24%	0.17%	0.13%	0.17%	0.23%	0.21%
26	0.17%	0.17%	0.16%	0.10%	0.21%	0.25%	0.13%	0.14%	0.16%	0.38%
27	0.17%	0.17%	0.16%	0.21%	0.11%	0.17%	0.20%	0.17%	0.15%	0.12%
28	0.16%	0.16%	0.16%	0.13%	0.17%	0.17%	0.13%	0.15%	0.18%	0.15%
29	0.15%	0.15%	0.13%	0.25%	0.13%	0.15%	0.18%	0.15%	0.14%	0.11%
30	0.15%	0.16%	0.14%	0.12%	0.28%	0.20%	0.15%	0.13%	0.15%	0.26%
31	0.15%	0.15%	0.15%	0.17%	0.16%	0.17%	0.16%	0.15%	0.16%	0.16%
32	0.15%	0.15%	0.15%	0.12%	0.25%	0.15%	0.11%	0.14%	0.19%	0.20%

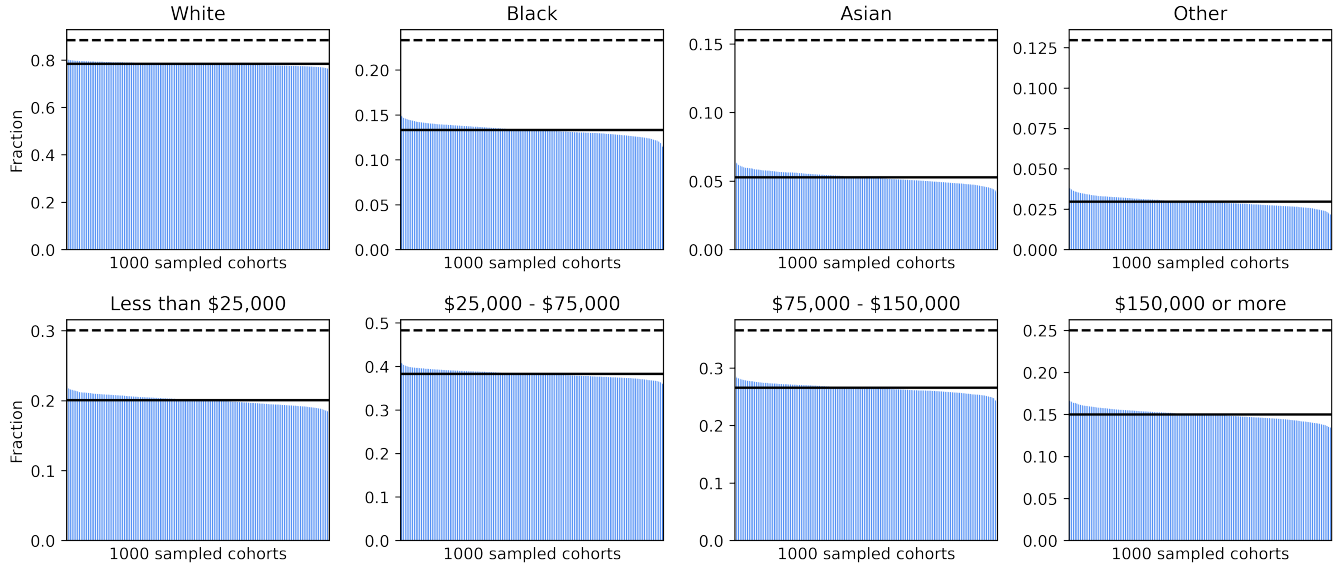|										
33	0.14%	0.14%	0.15%	0.14%	0.08%	0.12%	0.11%	0.14%	0.16%	0.15%
34	0.14%	0.14%	0.13%	0.16%	0.15%	0.14%	0.13%	0.13%	0.16%	0.17%
35	0.14%	0.13%	0.12%	0.16%	0.16%	0.15%	0.14%	0.13%	0.14%	0.15%
36	0.14%	0.14%	0.14%	0.12%	0.13%	0.15%	0.14%	0.14%	0.13%	0.13%
37	0.14%	0.14%	0.14%	0.12%	0.12%	0.14%	0.14%	0.14%	0.14%	0.14%
38	0.13%	0.13%	0.14%	0.12%	0.10%	0.12%	0.13%	0.14%	0.13%	0.10%
39	0.13%	0.13%	0.13%	0.14%	0.12%	0.12%	0.14%	0.12%	0.13%	0.13%
40	0.13%	0.13%	0.13%	0.07%	0.17%	0.20%	0.09%	0.11%	0.13%	0.32%
41	0.13%	0.13%	0.14%	0.13%	0.08%	0.12%	0.11%	0.13%	0.15%	0.15%
42	0.12%	0.12%	0.13%	0.11%	0.12%	0.12%	0.10%	0.12%	0.14%	0.13%
43	0.12%	0.12%	0.13%	0.11%	0.08%	0.11%	0.10%	0.12%	0.14%	0.10%
44	0.12%	0.12%	0.12%	0.08%	0.18%	0.13%	0.10%	0.10%	0.13%	0.20%
45	0.12%	0.12%	0.13%	0.09%	0.12%	0.11%	0.10%	0.12%	0.15%	0.15%
46	0.11%	0.11%	0.10%	0.11%	0.13%	0.12%	0.11%	0.10%	0.11%	0.14%
47	0.11%	0.10%	0.11%	0.07%	0.13%	0.10%	0.07%	0.09%	0.14%	0.16%
48	0.11%	0.11%	0.09%	0.16%	0.09%	0.16%	0.14%	0.11%	0.10%	0.11%
49	0.11%	0.11%	0.11%	0.13%	0.14%	0.13%	0.12%	0.11%	0.11%	0.14%
50	0.11%	0.11%	0.11%	0.08%	0.16%	0.10%	0.08%	0.09%	0.14%	0.17%
51	0.11%	0.10%	0.10%	0.12%	0.10%	0.11%	0.13%	0.11%	0.09%	0.08%
52	0.11%	0.11%	0.12%	0.09%	0.09%	0.09%	0.08%	0.10%	0.13%	0.16%
53	0.11%	0.11%	0.11%	0.12%	0.09%	0.12%	0.10%	0.11%	0.12%	0.11%
54	0.11%	0.11%	0.13%	0.07%	0.09%	0.10%	0.08%	0.11%	0.14%	0.11%
55	0.10%	0.10%	0.11%	0.08%	0.06%	0.08%	0.08%	0.10%	0.12%	0.11%
56	0.10%	0.10%	0.10%	0.06%	0.14%	0.10%	0.08%	0.09%	0.12%	0.13%
57	0.10%	0.10%	0.10%	0.10%	0.09%	0.10%	0.10%	0.09%	0.10%	0.11%
58	0.10%	0.09%	0.08%	0.13%	0.15%	0.11%	0.11%	0.08%	0.10%	0.12%
59	0.10%	0.10%	0.10%	0.10%	0.09%	0.10%	0.10%	0.10%	0.09%	0.09%
60	0.09%	0.09%	0.09%	0.06%	0.09%	0.08%	0.06%	0.08%	0.11%	0.12%
61	0.09%	0.09%	0.09%	0.06%	0.07%	0.08%	0.07%	0.08%	0.10%	0.11%
62	0.09%	0.09%	0.08%	0.07%	0.15%	0.10%	0.08%	0.08%	0.09%	0.15%
63	0.09%	0.09%	0.09%	0.09%	0.09%	0.09%	0.09%	0.09%	0.09%	0.09%
64	0.09%	0.09%	0.09%	0.10%	0.09%	0.10%	0.10%	0.09%	0.09%	0.10%
65	0.09%	0.09%	0.09%	0.10%	0.09%	0.10%	0.09%	0.09%	0.09%	0.09%
66	0.09%	0.09%	0.09%	0.12%	0.06%	0.09%	0.09%	0.10%	0.10%	0.08%
67	0.09%	0.09%	0.10%	0.04%	0.07%	0.06%	0.06%	0.08%	0.11%	0.10%
68	0.09%	0.09%	0.08%	0.11%	0.13%	0.10%	0.10%	0.07%	0.09%	0.11%
69	0.08%	0.08%	0.08%	0.07%	0.11%	0.08%	0.07%	0.08%	0.09%	0.09%
70	0.08%	0.08%	0.07%	0.09%	0.08%	0.08%	0.08%	0.08%	0.08%	0.08%
71	0.08%	0.08%	0.08%	0.08%	0.07%	0.08%	0.09%	0.09%	0.08%	0.07%
72	0.08%	0.08%	0.08%	0.07%	0.07%	0.07%	0.09%	0.08%	0.07%	0.06%
73	0.08%	0.08%	0.08%	0.07%	0.11%	0.09%	0.08%	0.08%	0.09%	0.09%
74	0.08%	0.08%	0.09%	0.08%	0.07%	0.07%	0.07%	0.09%	0.09%	0.07%
75	0.08%	0.08%	0.08%	0.05%	0.08%	0.07%	0.05%	0.06%	0.09%	0.12%
76	0.08%	0.08%	0.09%	0.03%	0.04%	0.06%	0.05%	0.07%	0.10%	0.10%
77	0.07%	0.07%	0.06%	0.07%	0.09%	0.08%	0.07%	0.06%	0.07%	0.08%
78	0.07%	0.07%	0.06%	0.10%	0.06%	0.06%	0.08%	0.07%	0.07%	0.06%
79	0.07%	0.07%	0.08%	0.05%	0.06%	0.06%	0.06%	0.07%	0.08%	0.07%
80	0.07%	0.07%	0.07%	0.04%	0.06%	0.06%	0.05%	0.06%	0.08%	0.10%
81	0.07%	0.07%	0.08%	0.07%	0.05%	0.06%	0.07%	0.08%	0.07%	0.05%
82	0.07%	0.07%	0.08%	0.05%	0.05%	0.06%	0.06%	0.07%	0.07%	0.06%
83	0.07%	0.07%	0.07%	0.10%	0.06%	0.08%	0.10%	0.07%	0.06%	0.06%
84	0.07%	0.07%	0.07%	0.06%	0.04%	0.08%	0.07%	0.07%	0.07%	0.07%
85	0.07%	0.07%	0.05%	0.12%	0.06%	0.07%	0.08%	0.06%	0.06%	0.05%
86	0.07%	0.07%	0.08%	0.04%	0.06%	0.06%	0.05%	0.07%	0.09%	0.07%
87	0.07%	0.07%	0.09%	0.04%	0.04%	0.06%	0.05%	0.07%	0.09%	0.07%
88	0.07%	0.07%	0.06%	0.07%	0.11%	0.07%	0.05%	0.06%	0.09%	0.09%
89	0.07%	0.07%	0.08%	0.04%	0.04%	0.05%	0.05%	0.07%	0.08%	0.07%
90	0.07%	0.07%	0.07%	0.06%	0.09%	0.07%	0.05%	0.06%	0.09%	0.08%
91	0.07%	0.07%	0.09%	0.04%	0.05%	0.06%	0.06%	0.07%	0.08%	0.08%
92	0.07%	0.07%	0.07%	0.08%	0.05%	0.07%	0.07%	0.07%	0.07%	0.06%
93	0.07%	0.07%	0.07%	0.04%	0.13%	0.06%	0.04%	0.06%	0.09%	0.09%
94	0.07%	0.07%	0.07%	0.06%	0.05%	0.06%	0.06%	0.07%	0.08%	0.08%
95	0.07%	0.07%	0.05%	0.11%	0.12%	0.10%	0.10%	0.06%	0.06%	0.06%
96	0.07%	0.07%	0.07%	0.08%	0.07%	0.07%	0.08%	0.07%	0.08%	0.07%
97	0.07%	0.07%	0.07%	0.07%	0.06%	0.07%	0.07%	0.07%	0.06%	0.05%
98	0.06%	0.06%	0.06%	0.06%	0.07%	0.07%	0.05%	0.06%	0.07%	0.09%
99	0.06%	0.06%	0.05%	0.11%	0.05%	0.08%	0.08%	0.06%	0.06%	0.05%
100	0.06%	0.06%	0.07%	0.04%	0.04%	0.05%	0.05%	0.06%	0.08%	0.07%

A.4 Baseline t-closeness Analysis With U.S. Population

For baseline t-closeness analysis we create a random panel matching the estimated size of the FLoC OT, where user demographic groups and cohort IDs are randomly assigned. We assign the household income groups and racial background groups so that their joint distribution matches the U.S. population estimates (CPS 2017 [35]). To estimate the size of the FLoC OT, we use our data to compute cohorts at varying levels of k and find a consistent relationship where (mean cohort size) / $k \sim 1.5$. The FLoC OT had $k = 2000$ and 33,872

cohorts. From this data we estimate the FLoC OT size with mean cohort sizes of 3000 (k=2000 x 1.5) and 101,616,000 (3000 x 33,872) user devices. We assign the 33,872 cohort IDs to the 101,616,000 users by randomly assigning the first 2000 for each cohort and then using a uniform distribution over cohort IDs for the remaining assignments. We then use this panel to do the following t-closeness check for the household income groups and racial background groups separately. For each cohort, we check for each demographic group whether the fraction of users in that demographic group exceeds the fraction in the general population by more than $t = 0.1$. No cohorts exceed this threshold. This is illustrated in Figure A.4.

Table A.2: Top 100 domain names in alphabetical order.

	domain		domain		domain		domain
1	247-inc.net	.	.	.	.	.	.
2	adobe.com	.	.	.	.	.	.
3	adp.com					.	.
4	amazon.com	26	espn.com	51	microsoftonline.com	.	.
5	aol.com	27	etsy.com	52	mozilla.org	.	.
6	apple.com	28	expedia.com	53	msn.com	76	taleo.net
7	ask.com	29	facebook.com	54	myway.com	77	target.com
8	att.com	30	foxnews.com	55	netflix.com	78	tripadvisor.com
9	bangcreatives.com	31	go.com	56	norton.com	79	tumblr.com
10	bankofamerica.com	32	google.com	57	nytimes.com	80	twitter.com
11	bestbuy.com	33	googlesyndication.com	58	office.com	81	verizonwireless.com
12	bing.com	34	homedepot.com	59	pandora.com	82	walmart.com
13	bongacams.com	35	hulu.com	60	paypal.com	83	washingtonpost.com
14	btrll.com	36	ibtimes.com	61	pinterest.com	84	weather.com
15	camdolls.com	37	imdb.com	62	pornhub.com	85	webmd.com
16	capitalone.com	38	indeed.com	63	pornhublive.com	86	wellsfargo.com
17	chase.com	39	instagram.com	64	quizlet.com	87	wikia.com
18	chaturbate.com	40	instructure.com	65	realtor.com	88	wikihow.com
19	citi.com	41	intuit.com	66	reddit.com	89	wikipedia.org
20	cnn.com	42	kohls.com	67	redtube.com	90	wordpress.com
21	comcast.net	43	linkedin.com	68	reference.com	91	xfinity.com
22	craigslist.org	44	live.com	69	roblox.com	92	xhamster.com
23	dropbox.com	45	livejasmin.com	70	s3xified.com	93	xhamsterlive.com
24	ebay.com	46	lowes.com	71	safesear.ch	94	xnxx.com
25	elbowviewpoint.com	47	macys.com	72	skype.com	95	xvideos.com
.	.	48	mapquest.com	73	smartadserver.com	96	yahoo.com
.	.	49	mcafee.com	74	steamcommunity.com	97	yelp.com
.	.	50	microsoft.com	75	steampowered.com	98	youporn.com
.	.	.	.	.	.	99	youtube.com
.	.	.	.	.	.	100	zillow.com

Figure A.4: t-closeness analysis for OT sized panel and U.S. population. For each cohort, we check for each demographic group whether the fraction of users in that demographic group exceeds the population mean by more than $t = 0.1$. For illustration, we show this analysis for 1000 randomly sampled cohorts. The solid black line indicates the fraction of each demographic group in the population. The dashed line indicates the threshold where t-closeness for $t = 0.1$ would be violated.