	Commuting Modes	Commuting from school				
		Car	Public	Wheels	Walk	Total
Commuting to school	Car	**58 (8.1%)**	**54 (7.6%)**	**1 (0.1%)**	**57 (8.0%)**	**170 (23.8%)**
		58 (8.1%)	54 (7.6%)	0 (0.0%)	57 (8.0%)	169 (23.7%)
		50 (7.0%)	72 (10.1%)	0 (0.0%)	60 (8.4%)	182 (25.5%)
		20 (2.8%)	48 (6.7%)	15 (2.1%)	94 (13.2%)	177 (24.8%)
	Public	**10 (1.4%)**	**190 (26.6%)**	**0 (0.0%)**	**30 (4.2%)**	**230 (32.3%)**
		8 (1.1%)	194 (27.2%)	0 (0.0%)	30 (4.2%)	232 (32.5%)
		6 (0.8%)	180 (25.2%)	0 (0.0%)	26 (3.6%)	212 (29.7%)
		37 (5.2%)	69 (9.7%)	34 (4.8%)	81 (11.4%)	221 (31.0%)
	Wheels	**0 (0.0%)**	**0 (0.0%)**	**27 (3.8%)**	**7 (1.0%)**	**34 (4.8%)**
		0 (0.0%)	0 (0.0%)	29 (4.1%)	6 (0.8%)	35 (4.9%)
		0 (0.0%)	0 (0.0%)	14 (2.0%)	17 (2.4%)	31 (4.3%)
		20 (2.8%)	23 (3.2%)	13 (1.8%)	40 (5.6%)	96 (13.5%)
	Walk	**3 (0.4%)**	**1 (0.1%)**	**0 (0.0%)**	**275 (38.6%)**	**279 (39.1%)**
		0 (0.0%)	0 (0.0%)	0 (0.0%)	279 (39.1%)	279 (39.1%)
		2 (0.3%)	0 (0.0%)	0 (0.0%)	276 (38.7%)	278 (39.0%)
		18 (2.5%)	70 (9.8%)	35 (4.9%)	96 (13.5%)	219 (30.7%)
	Total	**71 (10.0%)**	**245 (34.4%)**	**28 (3.9%)**	**369 (51.8%)**	**713 (100.0%)**
		66 (9.3%)	248 (34.8%)	29 (4.1%)	372 (52.2%)	715 (100.3%)
		58 (8.1%)	252 (35.3%)	14 (2.0%)	379 (53.2%)	703 (98.6%)
		95 (13.3%)	210 (29.5%)	97 (13.6%)	311 (43.6%)	713 (100.0%)

Table 1: Table 1 from the paper showing the counts and percentages for the original data and the three anonymization methods. Each group of four presents the data in order of Original (bold), SynDiffix, ARX, and SDV.

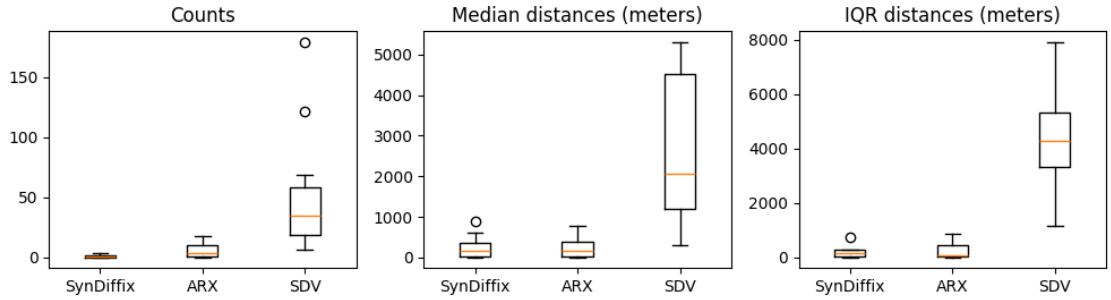
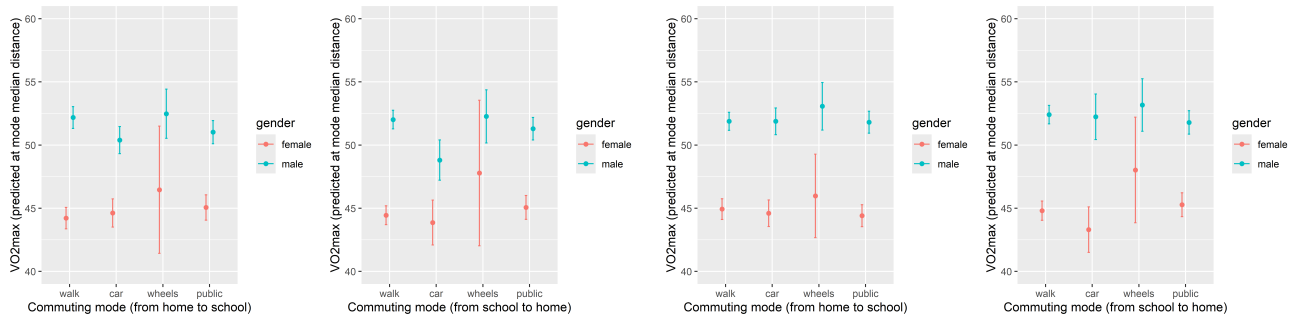


Figure 1: Absolute error of the three anonymization methods for the counts and distances in Tables 1 and 2. What we see here is that, for counts, SynDiffix is extremely accurate, but ARX is very accurate as well. SynDiffix and ARX are of equal quality for median and IQR distances. SDV is quite bad.

Commuting group	**From home to school**		**From school to home**	
	N (%)	**Distance (IQR)**	**N (%)**	**Distance (IQR)**
Car	**170 (24%)**	**3133 (3945)**	**71 (10%)**	**3615 (3896)**
	169 (24%)	3745 (4218)	70 (10%)	2725 (4184)
	182 (26%)	3758 (3915)	58 (8%)	3910 (3800)
	177 (25%)	7602 (8467)	95 (13%)	3934 (7362)
Public	**230 (32%)**	**4782 (4296)**	**245 (34%)**	**4996 (4033)**
	232 (33%)	4637 (4087)	245 (34%)	5029 (4055)
	212 (30%)	4973 (4193)	252 (35%)	5140 (3686)
	221 (31%)	5690 (8320)	210 (29%)	2249 (5174)
Wheels	**34 (5%)**	**1366 (2211)**	**28 (4%)**	**1444 (2369)**
	36 (5%)	1097 (2355)	30 (4%)	1243 (1626)
	31 (4%)	1356 (1378)	14 (2%)	2235 (3245)
	96 (13%)	6671 (8472)	97 (14%)	2741 (5282)
Walk	**279 (39%)**	**799 (789)**	**369 (52%)**	**973 (1043)**
	279 (39%)	784 (731)	368 (52%)	961 (1037)
	278 (39%)	805 (795)	379 (53%)	954 (1062)
	219 (31%)	5498 (8697)	311 (44%)	2374 (6068)
Total	**713 (100%)**		**713 (100%)**	
	716 (100%)		713 (100%)	
	703 (99%)		703 (99%)	
	713 (100%)		713 (100%)	

Table 2: Table 2 from the paper showing the counts and distances in meters (median and IQR) for the original data and the three anonymization methods. Each group of four presents the data in order of Original (bold), SynDiffix, ARX, and SDV. Note that the original distances median and IQR don't perfectly match those of the original Table 2 because of differences in the way median and IQR were calculated (Python versus R).



(a) Original Plot

(b) SynDiffix Plot

(c) ARX Plot

(d) SDV Plot

Figure 2: Comparison of the VO2max data. Here we see that ARX matches very closely with the original data. SynDiffix is quite close for female, but for reasons I don't understand yet, does somewhat bad for the car commute for males. Otherwise, though SynDiffix is pretty good. SDV is again quite bad. What will be important is whether the correct conclusions can be drown from the data in spite of the error.

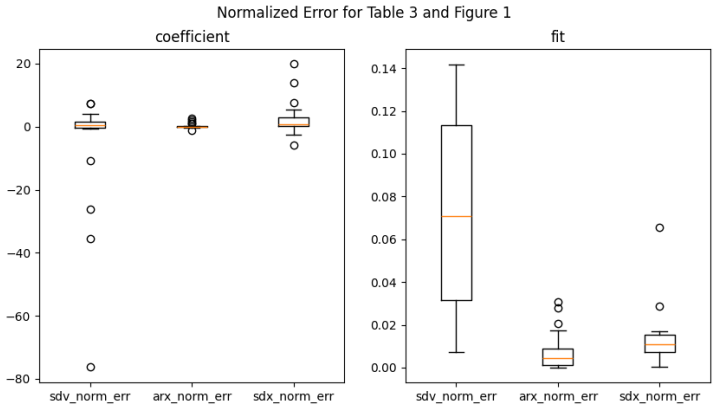


Figure 3: Normalized error for coefficients and fit for Figure 2. (Note that this plot isn't prettified yet.) This reflects the quality we see in Figure 2. SynDiffix clearly has more error than ARX.

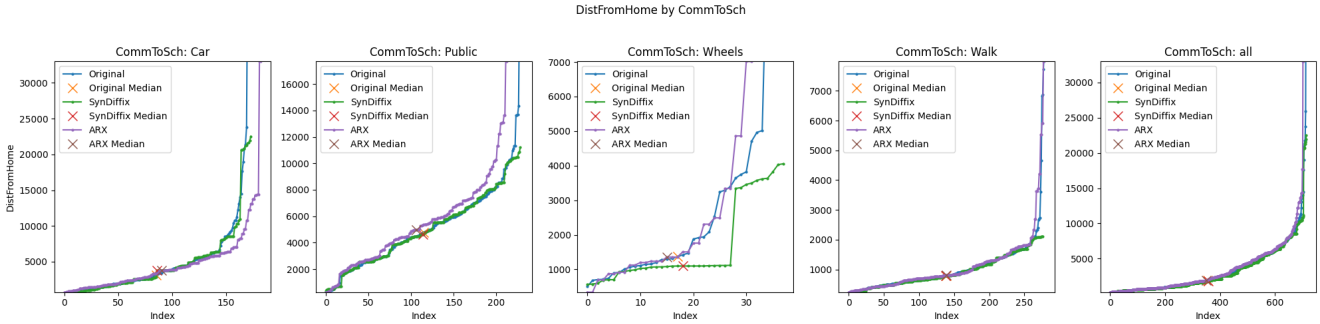


Figure 4: Distance from home distributions, by commuting type. Median distances marked with an X. I made this plot just to better understand where median distance errors were coming from for SynDiffix. There are two problems for SynDiffix. First, we adjust the "outlier" data points because they strictly speaking might break anonymity. Second, there are very few Wheels datapoints, and SynDiffix struggles with that.