Commuting to school	Commuting Modes	Commuting from school				
		Car	Public	Wheels	Walk	Total
	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.
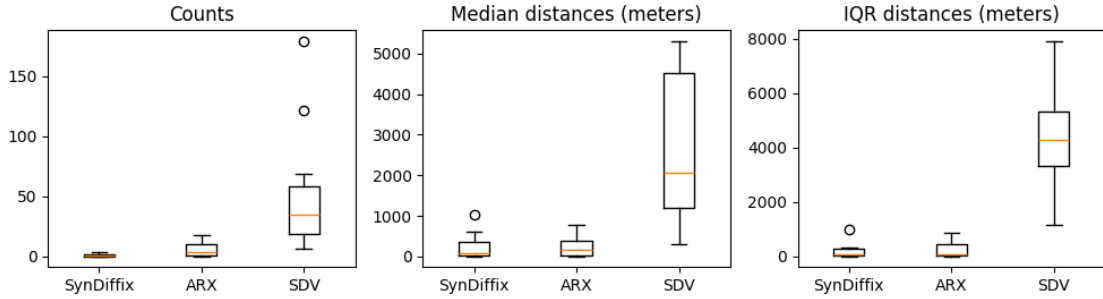

Figure 1: Absolute error of the three anonymization methods for the counts and distances in Tables 1 and 2. What we see here is that, for counts, SynDiffix is extremely accurate, but ARX is very accurate as well. SynDiffix and ARX are of equal quality for median and IQR distances. SDV is quite bad.

1

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%)	3754 (4215)	70 (10%)	2584 (3560)
	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%)	4712 (4228)	245 (34%)	5011 (4097)
	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%)	1094 (2251)	30 (4%)	1337 (1376)
	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%)	787 (737)	368 (52%)	960 (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 original 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).

Variables	Adjusted model			
	From home to school		**From school to home**	
	Coefficient	**95% CI**	**Coefficient**	**95% CI**
Constant	**36.42*****	**(28.17, 44.67)**	**36.63*****	**(29.11, 44.15)**
	26.54***	(16.07, 37.01)	32.89***	(26.25, 39.54)
	33.19***	(25.82, 40.56)	35.36***	(28.65, 42.07)
	56.08***	(45.07, 67.09)	49.18***	(40.44, 57.91)
Commuting group				
Car	**-6.49**	**(-15.92, 2.94)**	**-15.13***	**(-26.88, -3.39)**
	0.34	(-10.42, 11.09)	-12.67*	(-22.82, -2.52)
	-7.28	(-15.49, 0.92)	-17.72**	(-29.65, -5.8)
	-9.17	(-21.3, 2.95)	2.4	(-6.44, 11.24)
Public	**-0.08**	**(-9.06, 8.9)**	**-3.19**	**(-11.27, 4.88)**
	-0.39	(-11.14, 10.36)	8.71**	(2.72, 14.71)
	3.21	(-4.59, 11.01)	-4.08	(-10.99, 2.84)
	-6.17	(-16.67, 4.32)	-2.57	(-9.06, 3.92)
Wheels	**3.0**	**(-16.24, 22.25)**	**15.66**	**(-4.09, 35.41)**
	-2.93	(-40.4, 34.54)	10.7	(-3.6, 24.99)
	3.88	(-11.83, 19.58)	17.16	(-4.9, 39.22)
	-8.69	(-23.52, 6.14)	1.48	(-7.47, 10.44)
Walk (ref)				
Interaction Commuting group x Distance				
Car x Distance	**0.58**	**(-0.04, 1.2)**	**1.25****	**(0.34, 2.17)**
	0.21	(-0.32, 0.75)	2.06***	(1.19, 2.93)
	0.79**	(0.24, 1.34)	1.38**	(0.44, 2.33)
	0.35	(-0.35, 1.06)	-0.28	(-0.94, 0.37)
Public x Distance	**0.06**	**(-0.49, 0.61)**	**0.33**	**(-0.21, 0.88)**
	0.3	(-0.21, 0.82)	0.03	(-0.39, 0.45)
	-0.04	(-0.52, 0.45)	0.37	(-0.1, 0.84)
	0.04	(-0.48, 0.56)	0.38	(-0.06, 0.82)
Wheels x Distance	**-0.09**	**(-1.79, 1.62)**	**-1.15**	**(-2.89, 0.6)**
	0.64	(-2.73, 4.0)	-0.13	(-1.36, 1.11)
	0.08	(-1.32, 1.48)	-1.41	(-3.35, 0.53)
	0.09	(-0.88, 1.07)	-0.04	(-0.75, 0.66)
Walk x Distance	**-0.02**	**(-0.62, 0.58)**	**0.03**	**(-0.42, 0.48)**
	0.36	(-0.55, 1.26)	0.66***	(0.3, 1.02)
	0.17	(-0.33, 0.68)	-0.04	(-0.42, 0.34)
	-0.63	(-1.28, 0.02)	-0.08	(-0.44, 0.28)

* p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001

Table 3: Part 1 (of 2) of the original paper's Table 3 showing the parameters (regression coefficients) of the linear model for prediction of VO2max by group and distance. Each group of four presents the data in order of Original (bold), SynDiffix, ARX, and SDV.

Variables	Adjusted model			
	From home to school		From school to home	
	Coefficient	95% CI	Coefficient	95% CI
Gender				
	7.97***	**(6.75, 9.19)**	**7.58*****	**(6.52, 8.63)**
Males	7.13***	(5.95, 8.3)	8.26***	(7.31, 9.2)
	8.19***	(7.2, 9.18)	7.45***	(6.6, 8.29)
	0.29	(-1.66, 2.25)	-0.36	(-1.99, 1.27)
Females (ref)				
Interaction Commuting group x Gender				
	-2.2*	**(-4.16, -0.24)**	**-2.63***	**(-5.23, -0.03)**
Car x Gender	0.18	(-1.65, 2.01)	-4.21**	(-6.97, -1.44)
	-2.32**	(-3.86, -0.77)	-2.3*	(-4.59, -0.01)
	-2.75	(-5.66, 0.16)	0.41	(-2.97, 3.78)
	-2.0*	**(-3.81, -0.2)**	**-1.35**	**(-2.99, 0.3)**
Public x Males	-0.65	(-2.31, 1.02)	-1.88*	(-3.35, -0.41)
	-2.56***	(-4.05, -1.07)	-1.53*	(-2.83, -0.22)
	-1.69	(-4.46, 1.07)	-2.18	(-4.77, 0.4)
	-1.95	**(-7.49, 3.6)**	**-3.09**	**(-9.31, 3.12)**
Wheels x Males	2.03	(-3.32, 7.38)	nan	(nan, nan)
	-4.06	(-10.16, 2.05)	nan	(nan, nan)
	-0.17	(-3.71, 3.38)	0.16	(-3.25, 3.56)
Walk x Males (ref)				
Covariates				
	0.08***	**(0.03, 0.12)**	**0.07*****	**(0.03, 0.11)**
MVPA	0.17***	(0.12, 0.22)	0.14***	(0.09, 0.18)
	0.08***	(0.04, 0.12)	0.07***	(0.03, 0.12)
	-0.06**	(-0.11, -0.02)	-0.06**	(-0.11, -0.01)
	0.43*	**(0.0, 0.85)**	**0.4**	**(-0.02, 0.82)**
Age	0.71***	(0.31, 1.11)	-0.05	(-0.43, 0.33)
	0.53**	(0.14, 0.92)	0.55**	(0.17, 0.93)
	-0.11	(-0.64, 0.42)	-0.08	(-0.61, 0.45)

* $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

Table 4: Part 2 (of 2) of the original paper's Table 3 showing the parameters (regression coefficients) of the linear model for prediction of VO2max by group and distance. Each group of four presents the data in order of Original (bold), SynDiffix, ARX, and SDV.

(a) Original Plot

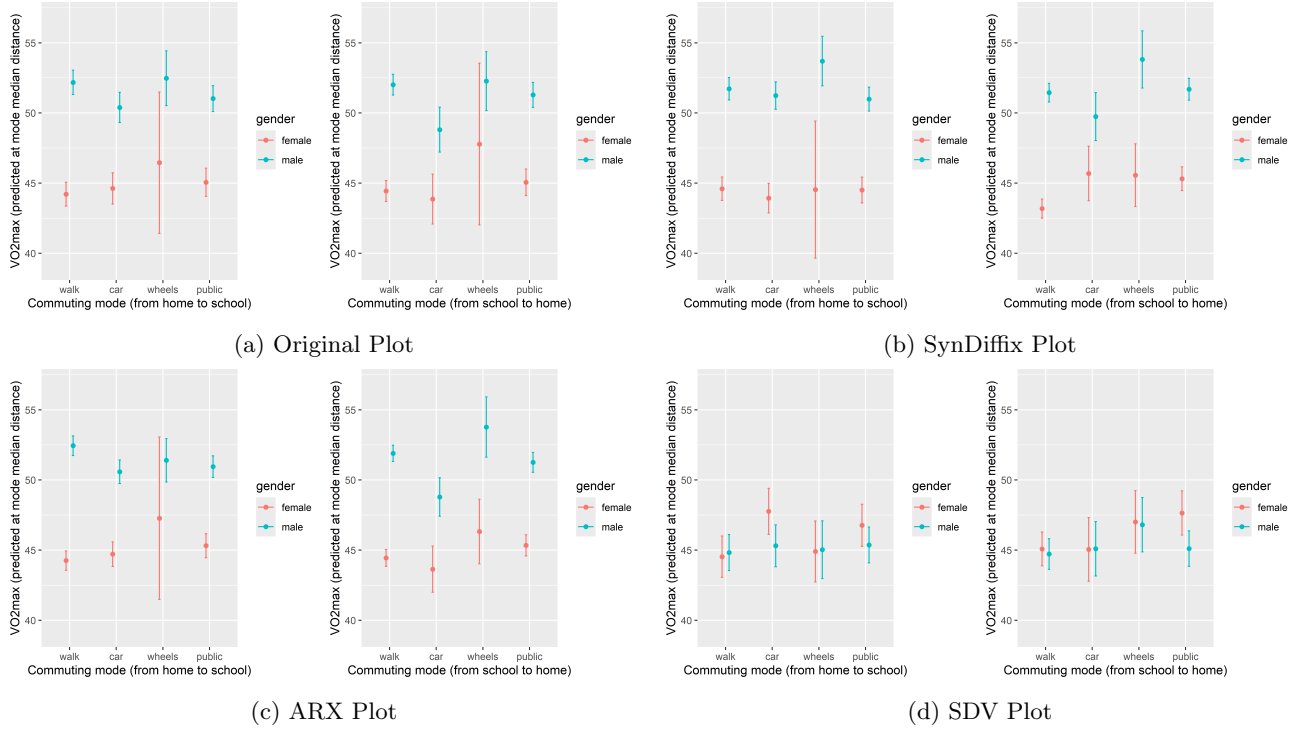(b) SynDiffix Plot

(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.
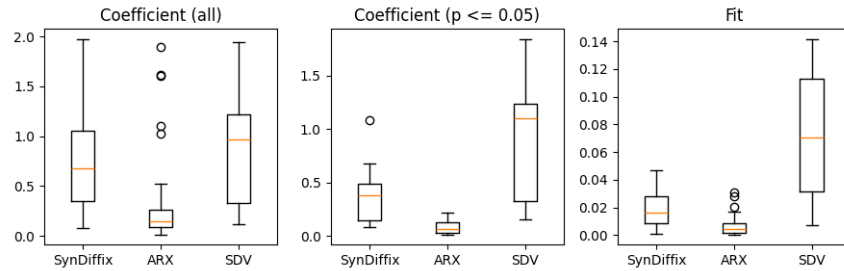

Figure 3: Normalized error for coefficients and fit for Figure 2. This reflects the quality we see in Figure 2. SynDiffix clearly has more error than ARX.

	SDX	ARX	SDV
Of the original 12 significant p-values, method is also significat	10 (83%)	12 (100%)	4 (33%)
Of the original 16 insignificant p-values, method is also insignificat	13 (81%)	13 (81%)	16 (100%)
Of the original 12 significant p-values, method matches	7 (58%)	8 (67%)	2 (17%)
Of the original 12 significant p-values, method off by 1	2 (17%)	3 (25%)	2 (17%)
Of the original 12 significant p-values, method off by 2	1 (8%)	1 (8%)	0 (0%)

Table 5: Error between each method's p-values and the original p-values. P-values are significant when $p \leq 0.05$. P-values are binned as $p \leq 0.001$, $0.001 < p \leq 0.01$, and $0.01 < p \leq 0.05$. Off by 1 means that the method's bin is one off from the original data's bin (both being significant). Off by 2 means that the method's bin is two off from the original data's bin.
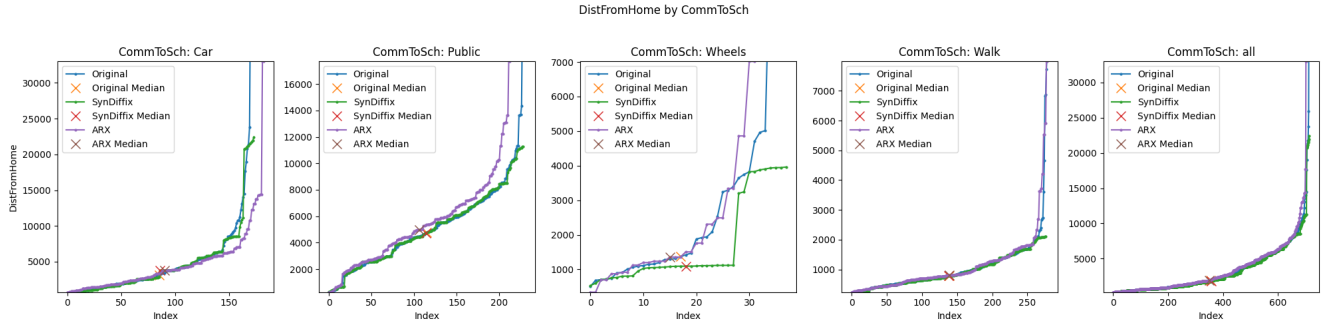

Figure 4: Distance from home distributions, by commuting type. Median distances marked with an X. I made this plot just to better understand where median distance errors were coming from for SynDiffix. There are two problems for SynDiffix. First, we adjust the "outlier" data points because they strictly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