Security and Privacy Final Project – Phase 2

Group D – Adversaries 😈

Jakub Mráz, jamr@itu.dk

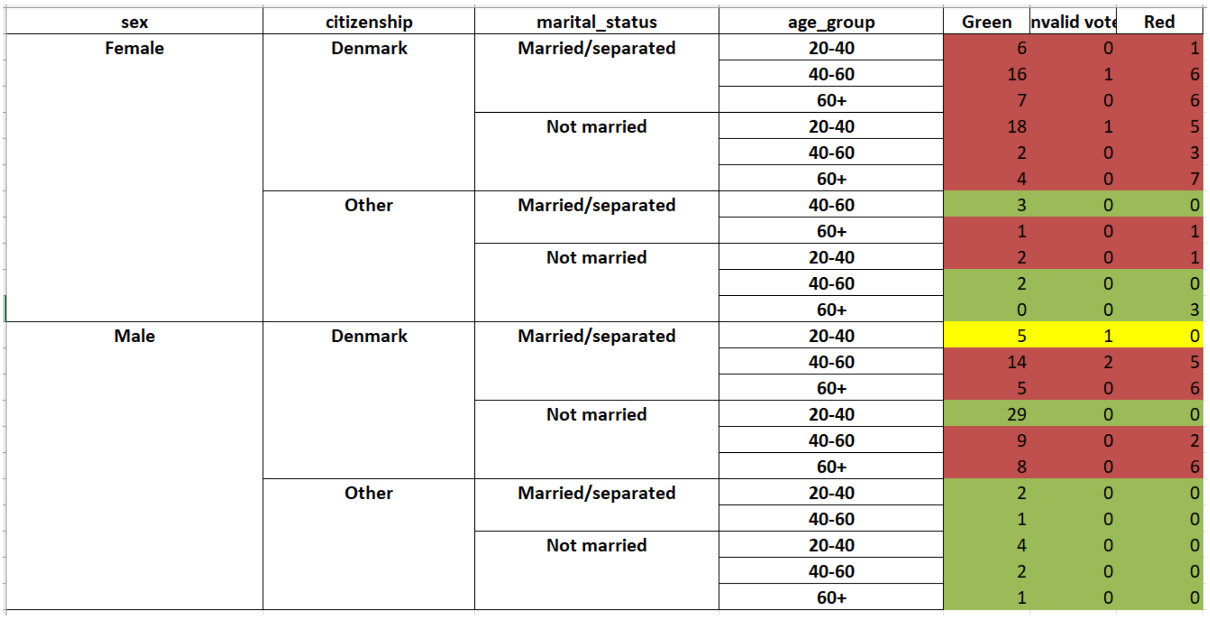
Mark Assejev, maass@itu.dk

Deanonymization technique

To deanonymize the dataset using the population registry, we first filtered the latter using the leaked set of names of people that participated in the study, leaving us with only names and information that were also present in the anonymised dataset.

Second, we applied the same groupings and transformations to the registry as were applied to the anonymised dataset. We could not perform a simple join, as the opposing group achieved 2-anonnymity on their dataset at the cost of completely forgoing the possibility for numerous data analyses.

Our strategy was to identify every possible combination of quasi-identifiers to find those that voted exclusively for one party.



sex	citizenship	marital_status	age_group	Green	nvalid vote	Red
Female	Denmark	Married/separated	20-40	6	0	1
			40-60	16	1	6
			60+	7	0	6
		Not married	20-40	18	1	5
			40-60	2	0	3
			60+	4	0	7
	Other	Married/separated	40-60	3	0	0
			60+	1	0	1
		Not married	20-40	2	0	1
			40-60	2	0	0
			60+	0	0	3
Male	Denmark	Married/separated	20-40	5	1	0
			40-60	14	2	5
			60+	5	0	6
		Not married	20-40	29	0	0
			40-60	9	0	2
			60+	8	0	6
	Other	Married/separated	20-40	2	0	0
			40-60	1	0	0
		Not married	20-40	4	0	0
			40-60	2	0	0
			60+	1	0	0

With this information, we could see that, for example, everyone in the group Male, Denmark, Married/separated, 20-40 voted Green, allowing us to deanonymize all 29 of them. We did this for all combinations coloured with green. The yellow row contains one invalid vote, allowing us to say for certain that none of them voted Red, though not for certain that they voted green. Those were deanonymized as having voted “not red”.

In total, we were able to deanonymize the votes of **50 people** out of 200 entries, making it a critical **25% vulnerability**.