Spatial Synthetic Population Generation Using Simulated Annealing in Go

Hugh, Andi, Ric

Go (Golang)

A Compiled, Concurrent, Systems Programming Language

- **Type:** Statically-typed, compiled language with a C-like syntax but memory safety and garbage collection.
- **Compilation:** Compiles directly to machine code (not to a VM). **Produces a single, fast, dependency-free binary for Linux, Windows, macOS, and more.**
- **Concurrency Model:** Based on **CSP** (Communicating Sequential Processes). Lightweight **goroutines** and **channels** provide a simpler and more efficient alternative to traditional thread-based concurrency.
- **Systems Programming Focus:** **Designed for modern multi-core** and networked machines.
- **Much Much faster than Python and easter than JAVA**

"Do One Thing and Do It Well"

1. **Do One Thing and Do It Well:** Programs are small, focused tools that excel at a single, specific task.
2. **Write Programs to Work Together:** The output of one program should be able to serve as the input to another. This is enabled by...
3. **Everything is a Text Stream:** Use plain text as the universal interface. (JSON CSV …)

Creating synthetic area households for Scotland

Constraint data by area

geography_code	s1_hh_urban_rural %urban	s1_hh_urban_rural %rural	s2_hh_size%hhsize _1	s2_hh_size%hhsize _2
S00135307	0	62	22	22
S00135308	0	33	7	17
S00135309	0	71	26	32
S00135310	65	0	21	23

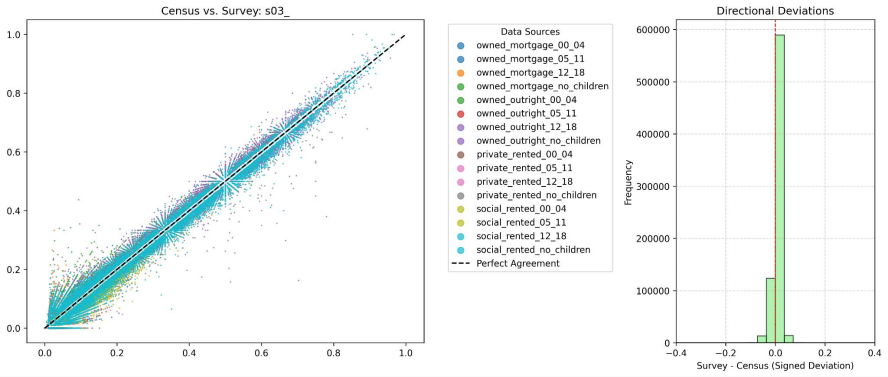
Household data (one hot)

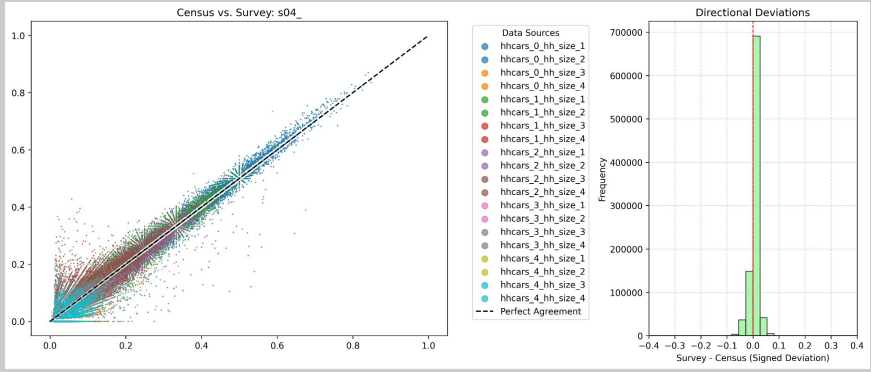
id	s1_hh_urban_rural%urban	s1_hh_urban_rural%rural	s2_hh_size%hhsize_1	s2_hh_size%hhsize_2
68006826	1	0	0	1
68013626	1	0	1	0
68020426	1	0	0	0
68027226	1	0	0	1
68047626	1	0	0	0

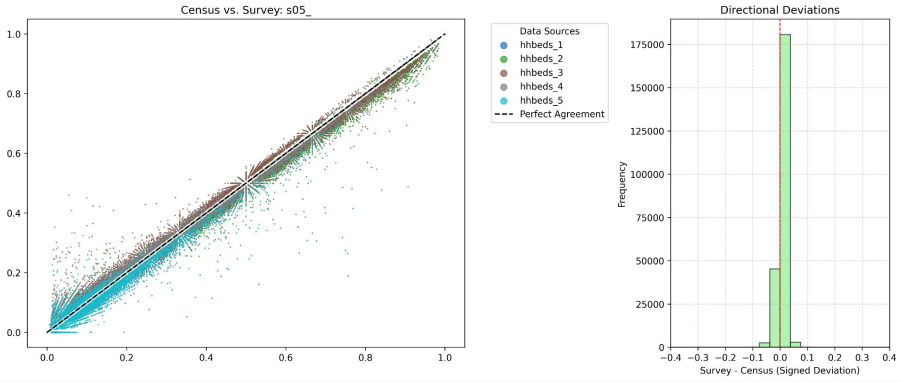
GoSynthPop population generation setup

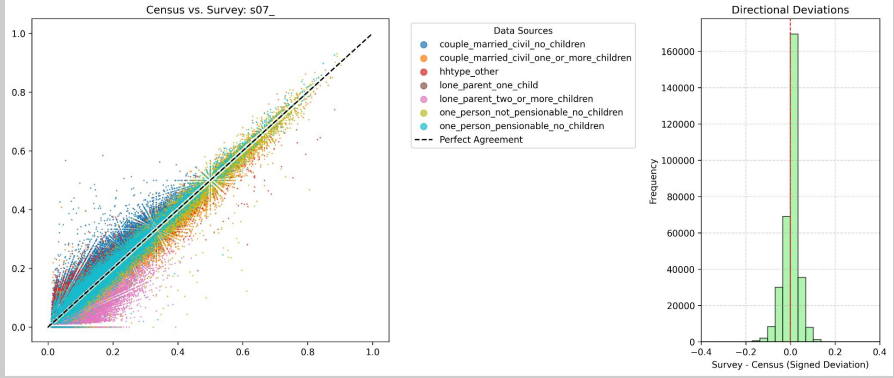
```
{
  "constraints": {
    "file": "data/census2022_all_go.csv"
  },
  "microdata": {
    "file": "data/us_hh_export_go.csv"
  },
  "output": {
    "file":
"results/synthetic_population_0608.csv"
  },
  "validate": {
    "file": "results/synthPopSurvay0608.csv"
  }
}
```

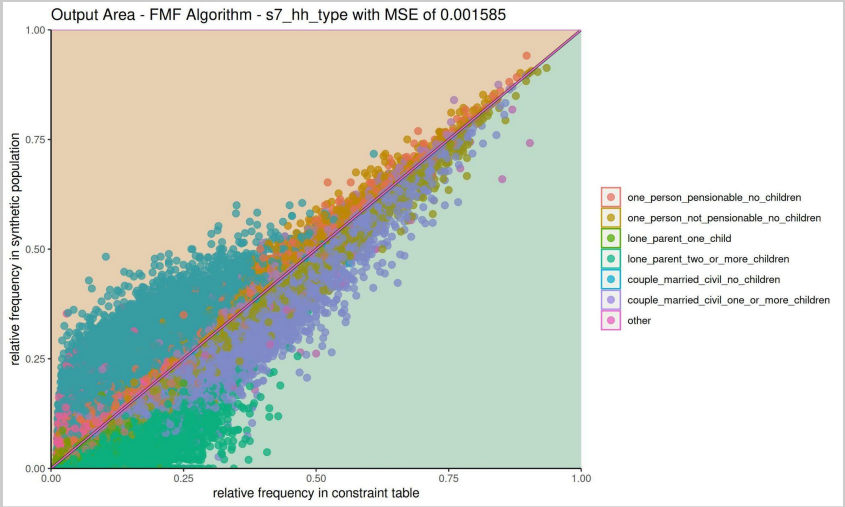
```
{
    "initialTemp": 5000.0,
    "minTemp": 0.00001,
    "coolingRate": 0.999,
    "reheatFactor": 0.8,
    "fitnessThreshold": 0.0001,
    "minImprovement": 0.0001,
    "maxIterations": 500000000,
    "windowSize": 1000,
    "change": 100000,
    "distance": "EUCLIDEAN",
    "useRandomSeed": "yes",
    "randomSeed": 42
}
```

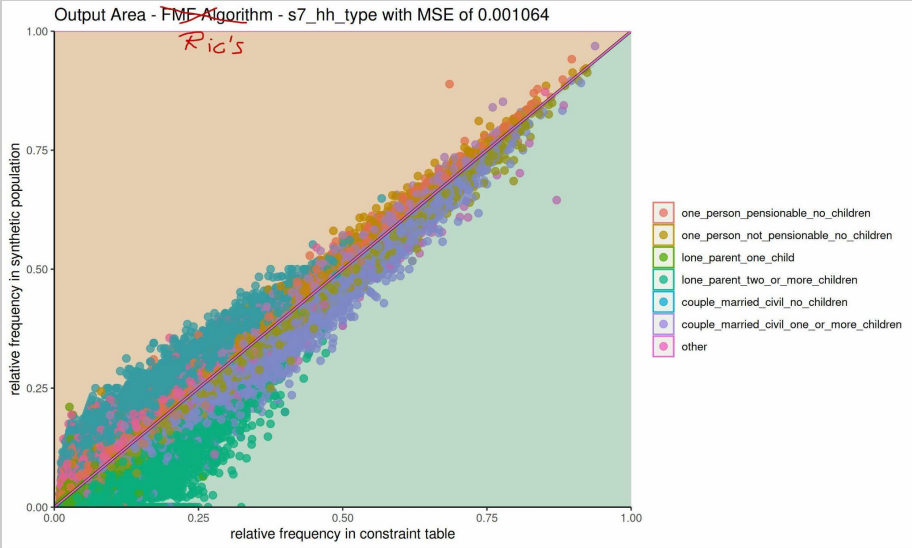

Census vs. Survey: s03_
Data Sources
owned_mortgage_00_04
owned_mortgage_05_11
owned_mortgage_12_18
owned_mortgage_no_children
owned_outright_00_04
owned_outright_05_11
owned_outright_12_18
owned_outright_no_children
private_rented_00_04
private_rented_05_11
private_rented_12_18
private_rented_no_children
social_rented_00_04
social_rented_05_11
social_rented_12_18
social_rented_no_children
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Census vs. Survey: s04_
Data Sources
hhcars_0_hh_size_1
hhcars_0_hh_size_2
hhcars_0_hh_size_3
hhcars_0_hh_size_4
hhcars_1_hh_size_1
hhcars_1_hh_size_2
hhcars_1_hh_size_3
hhcars_1_hh_size_4
hhcars_2_hh_size_1
hhcars_2_hh_size_2
hhcars_2_hh_size_3
hhcars_2_hh_size_4
hhcars_3_hh_size_1
hhcars_3_hh_size_2
hhcars_3_hh_size_3
hhcars_3_hh_size_4
hhcars_4_hh_size_1
hhcars_4_hh_size_2
hhcars_4_hh_size_3
hhcars_4_hh_size_4
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Census vs. Survey: s05_
Data Sources
hhbeds_1
hhbeds_2
hhbeds_3
hhbeds_4
hhbeds_5
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Census vs. Survey: s07_
Data Sources
couple_married_civil_no_children
couple_married_civil_one_or_more_children
hhtype_other
lone_parent_one_child
lone_parent_two_or_more_children
one_person_not_pensionable_no_children
one_person_pensionable_no_children
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Output Area - FMF Algorithm - s7_hh_type with MSE of 0.001585
relative frequency in synthetic population
relative frequency in constraint table
0.00
0.25
0.50
0.75
1.00
one_person_pensionable_no_children
one_person_not_pensionable_no_children
lone_parent_one_child
lone_parent_two_or_more_children
couple_married_civil_no_children
couple_married_civil_one_or_more_children
other


Output Area - FMF Algorithm - s7_hh_type with MSE of 0.001064
Ric's
relative frequency in synthetic population
relative frequency in constraint table
0.00
0.25
0.50
0.75
1.00
one_person_pensionable_no_children
one_person_not_pensionable_no_children
lone_parent_one_child
lone_parent_two_or_more_children
couple_married_civil_no_children
couple_married_civil_one_or_more_children
other

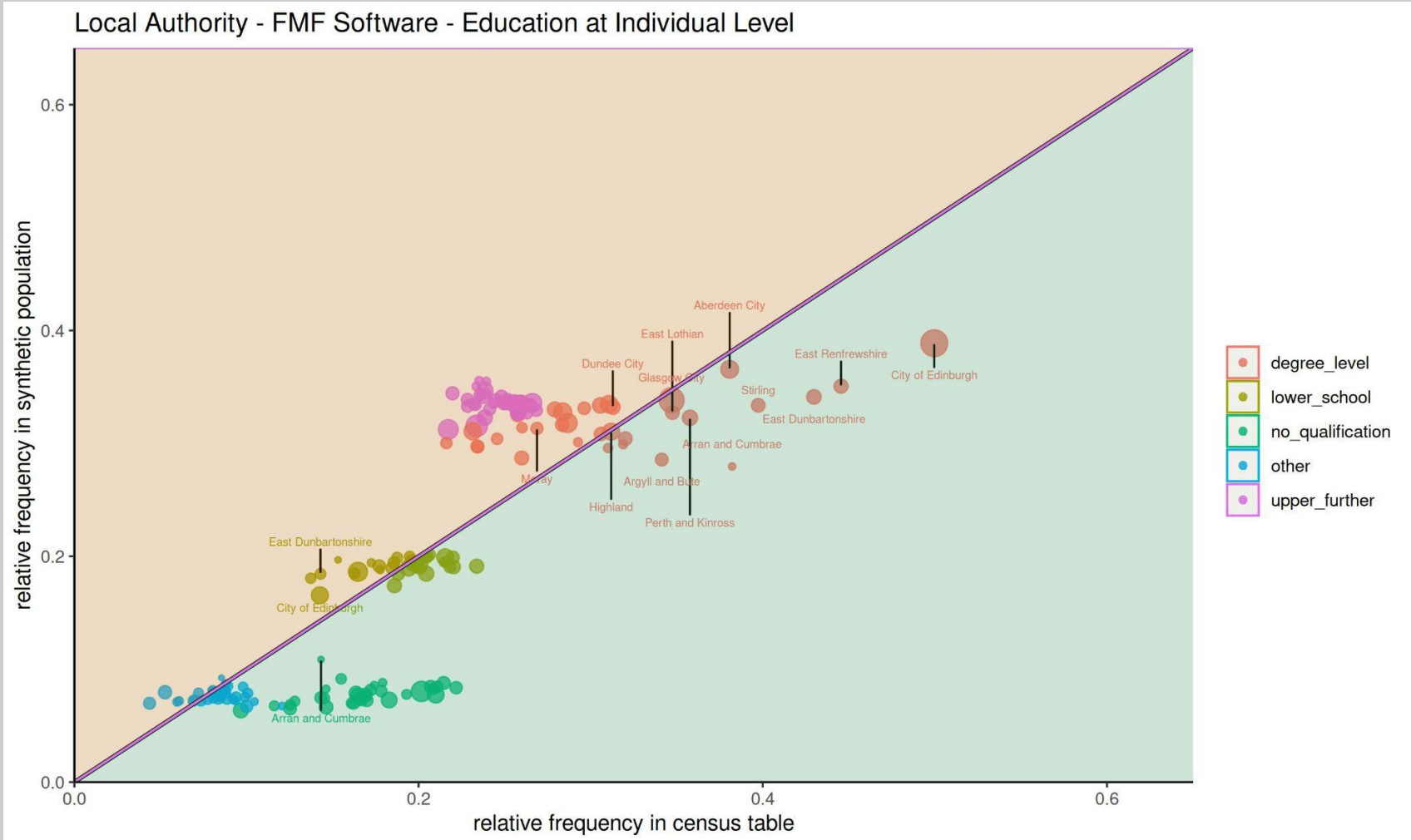
Creating synthetic households of individuals from household data

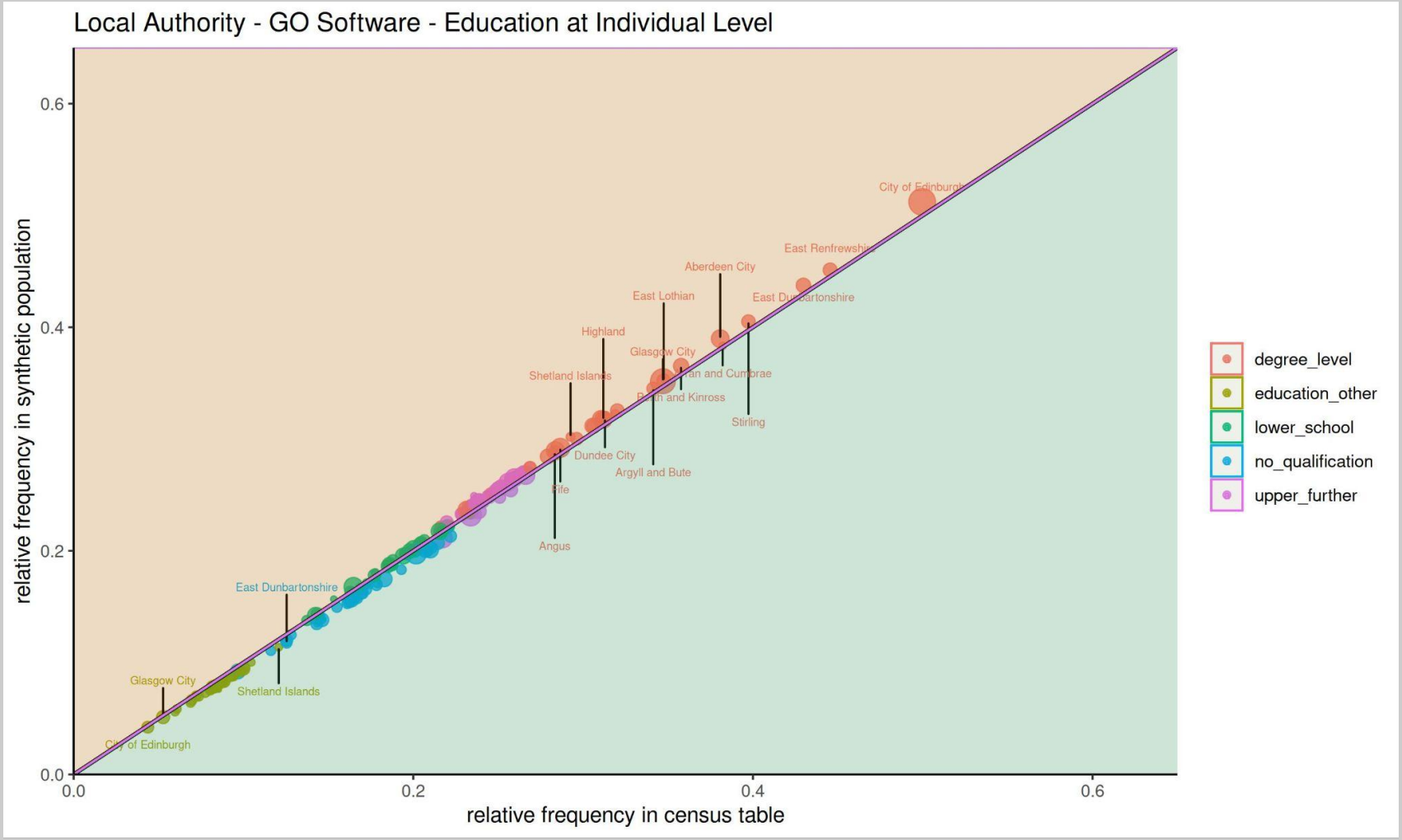
Constraint data by area

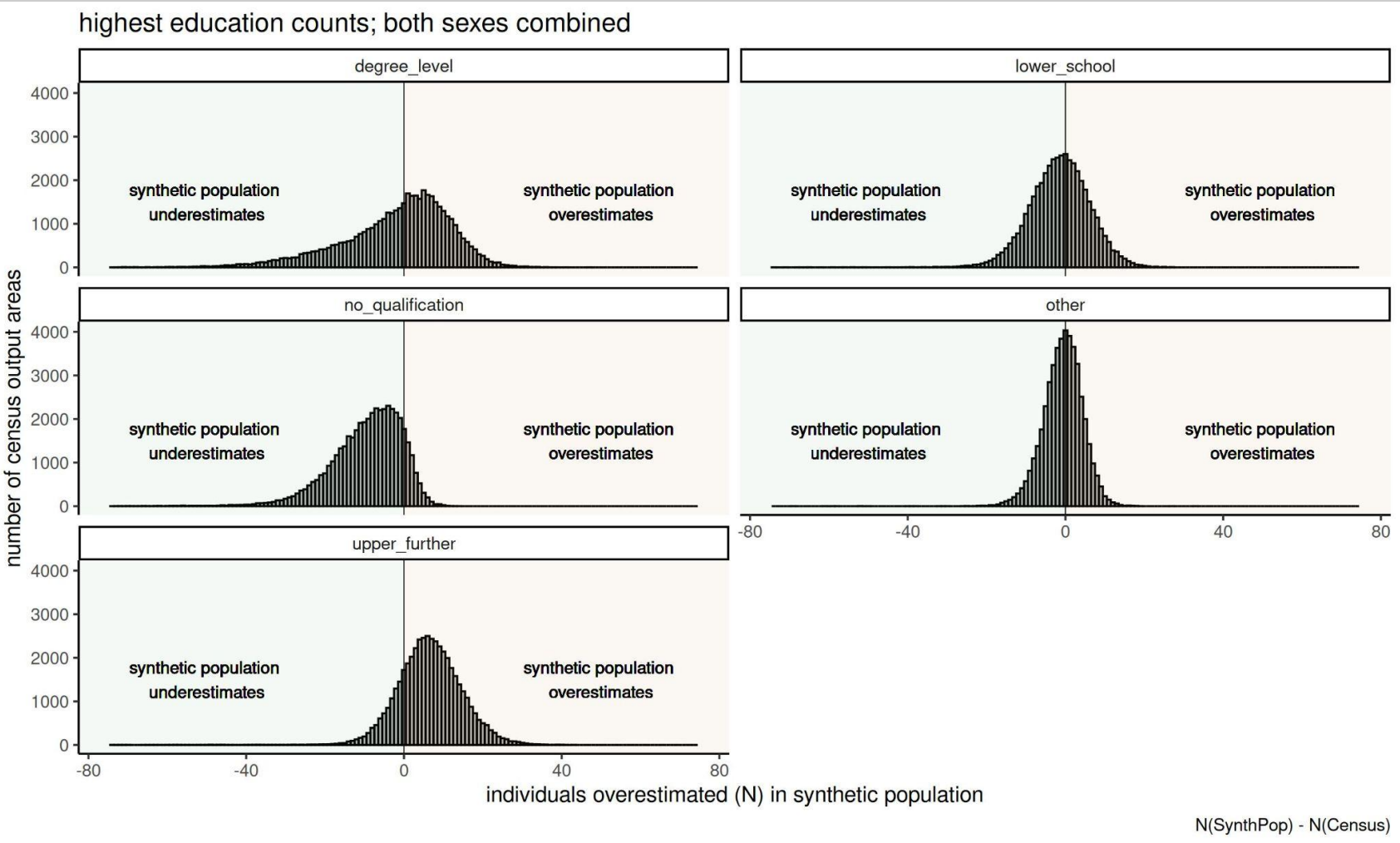
s13_unpaid_carer_hh_size%carer_2_hh_size_3	s14_edu_count_no_qualification	s14_edu_count_lower_school	s14_edu_count_upper_further	s14_edu_count_degree_level	s14_edu_count_other
3	13	24	27	44	10
1	7	8	7	24	8
3	27	21	18	64	15
2	11	39	40	31	15
1	12	17	27	52	9

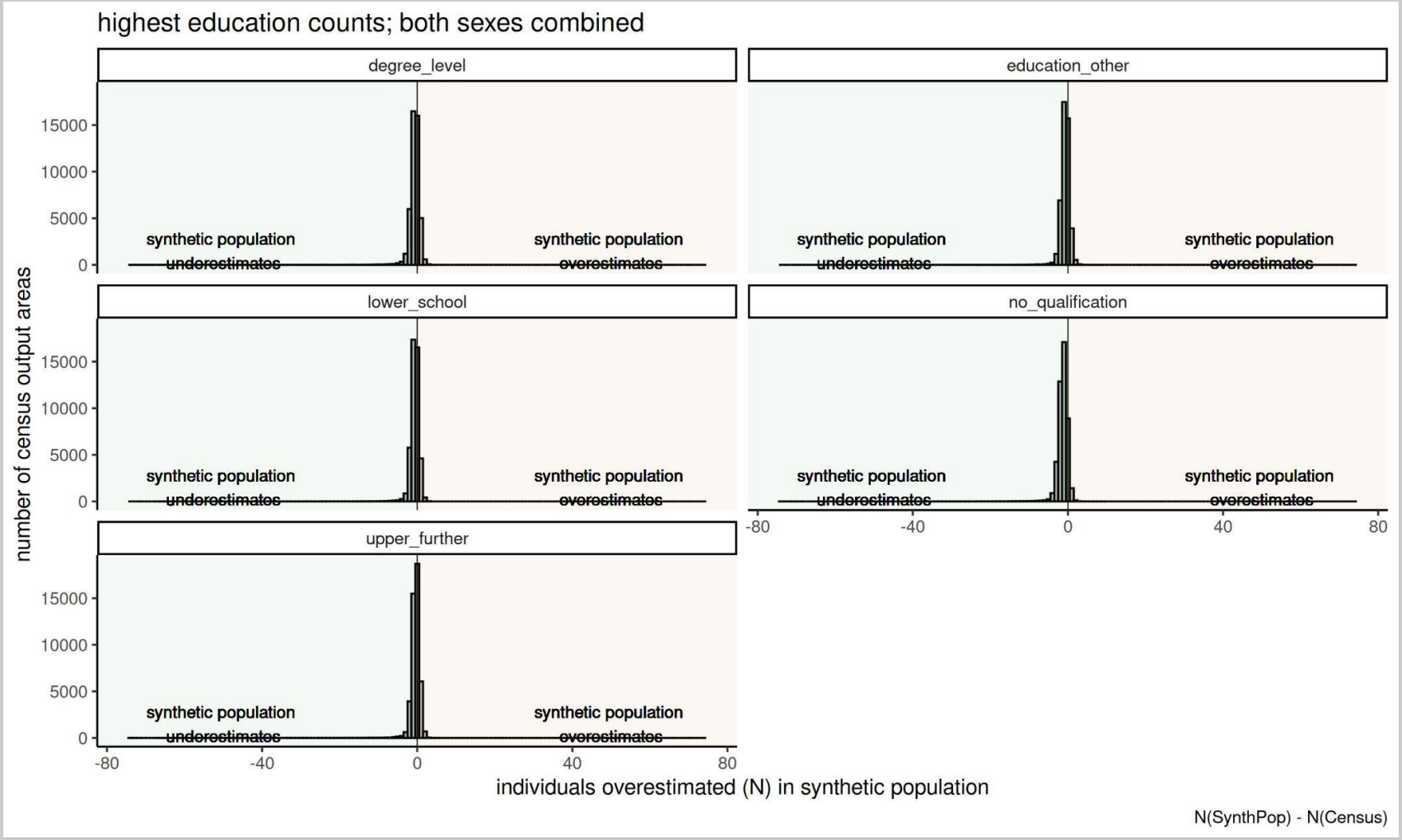
Household data

s13_unpaid_carer_hh_size%carer_2_hh_size_3	s14_edu_count_no_qualification	s14_edu_count_lower_school	s14_edu_count_upper_further	s14_edu_count_degree_level	s14_edu_count_other
0	0	2	0	0	0
0	0	0	0	0	1
0	0	0	1	1	0
0	0	0	1	1	0
1	0	0	0	2	0
0	0	1	2	1	1

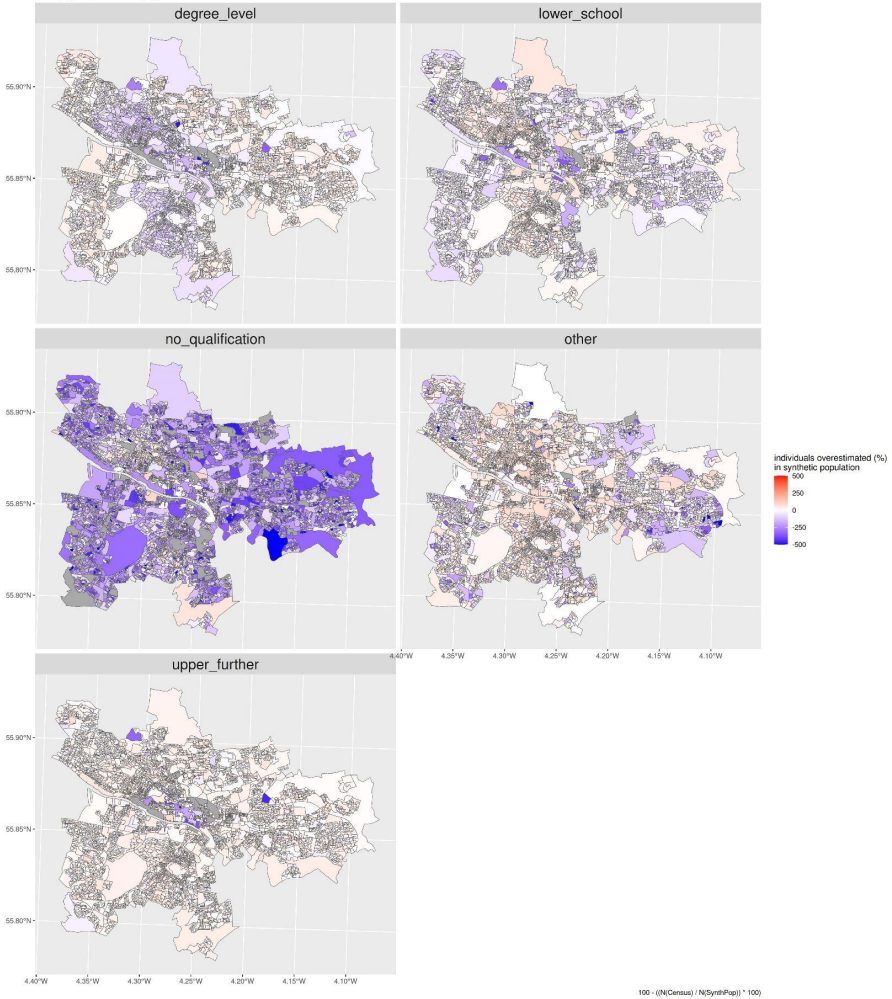

Local Authority - FMF Software - Education at Individual Level
relative frequency in synthetic population
relative frequency in census table
0.0
0.2
0.4
0.6
Aberdeen City
East Lothian
Dundee City
Glasgow City
East Renfrewshire
City of Edinburgh
Stirling
East Dunbartonshire
Arran and Cumbrae
Argyll and Bute
Highland
Perth and Kinross
East Dunbartonshire
City of Edinburgh
Arran and Cumbrae
degree_level
lower_school
no_qualification
other
upper_further


Local Authority - GO Software - Education at Individual Level
relative frequency in synthetic population
relative frequency in census table
0.0
0.2
0.4
0.6
City of Edinburgh
East Renfrewshire
Aberdeen City
East Lothian
East Dunbartonshire
Highland
Glasgow City
Shetland Islands
Stirling
Dundee City
Argyll and Bute
Fife
Angus
East Dunbartonshire
Glasgow City
Shetland Islands
City of Edinburgh
degree_level
education_other
lower_school
no_qualification
upper_further


highest education counts; both sexes combined
degree_level
lower_school
no_qualification
other
upper_further
synthetic population underestimates
synthetic population overestimates
number of census output areas
4000
3000
2000
1000
0
-80
-40
0
40
80
individuals overestimated (N) in synthetic population
N(SynthPop) - N(Census)

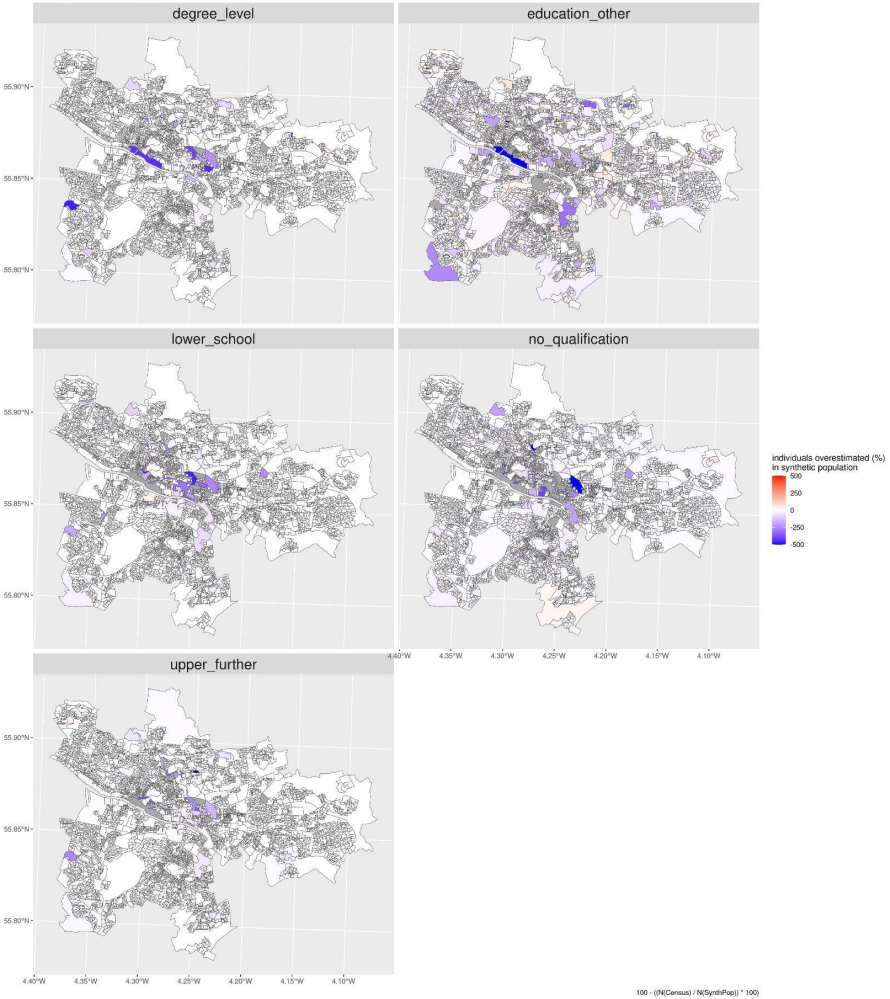

highest education counts; both sexes combined
degree_level
education_other
lower_school
no_qualification
upper_further
synthetic population underestimates
synthetic population overestimates
number of census output areas
0
5000
10000
15000
-80
-40
0
40
80
individuals overestimated (N) in synthetic population
N(SynthPop) - N(Census)

total population counts; by sex

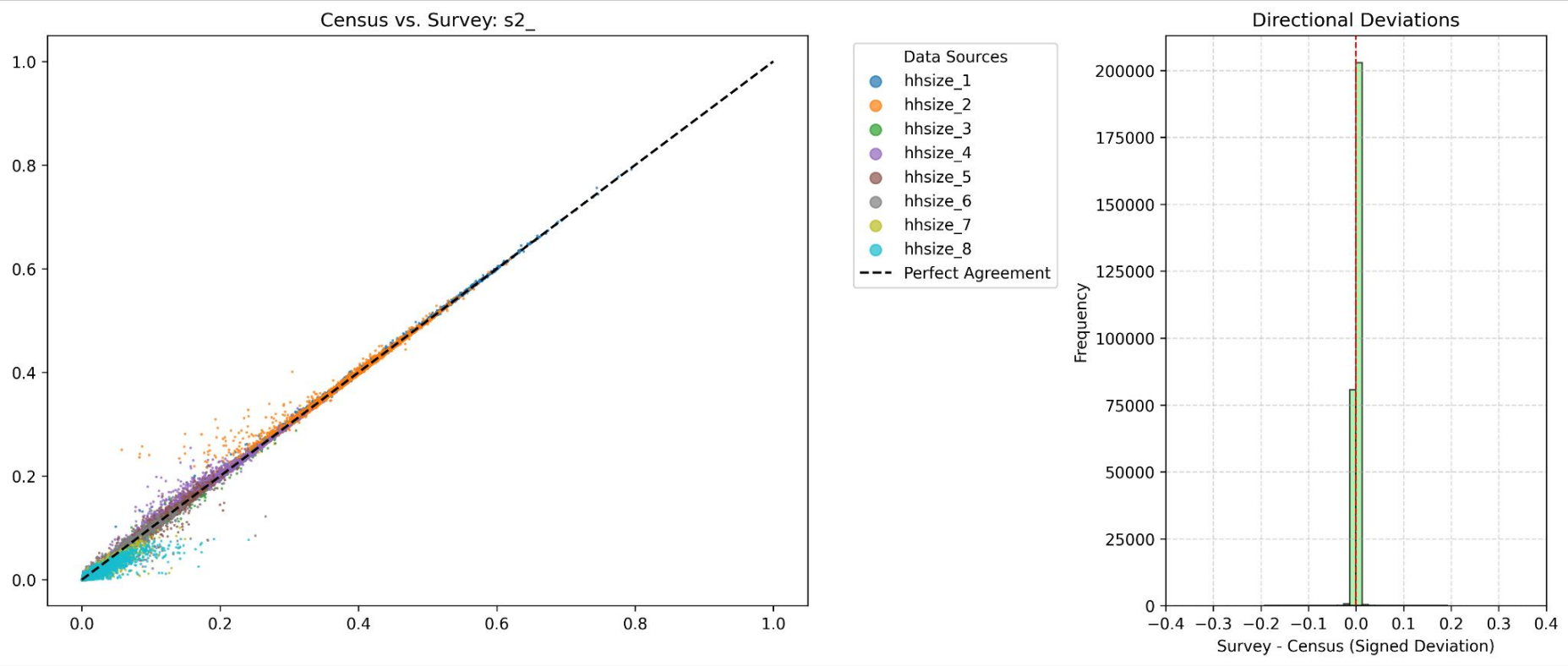


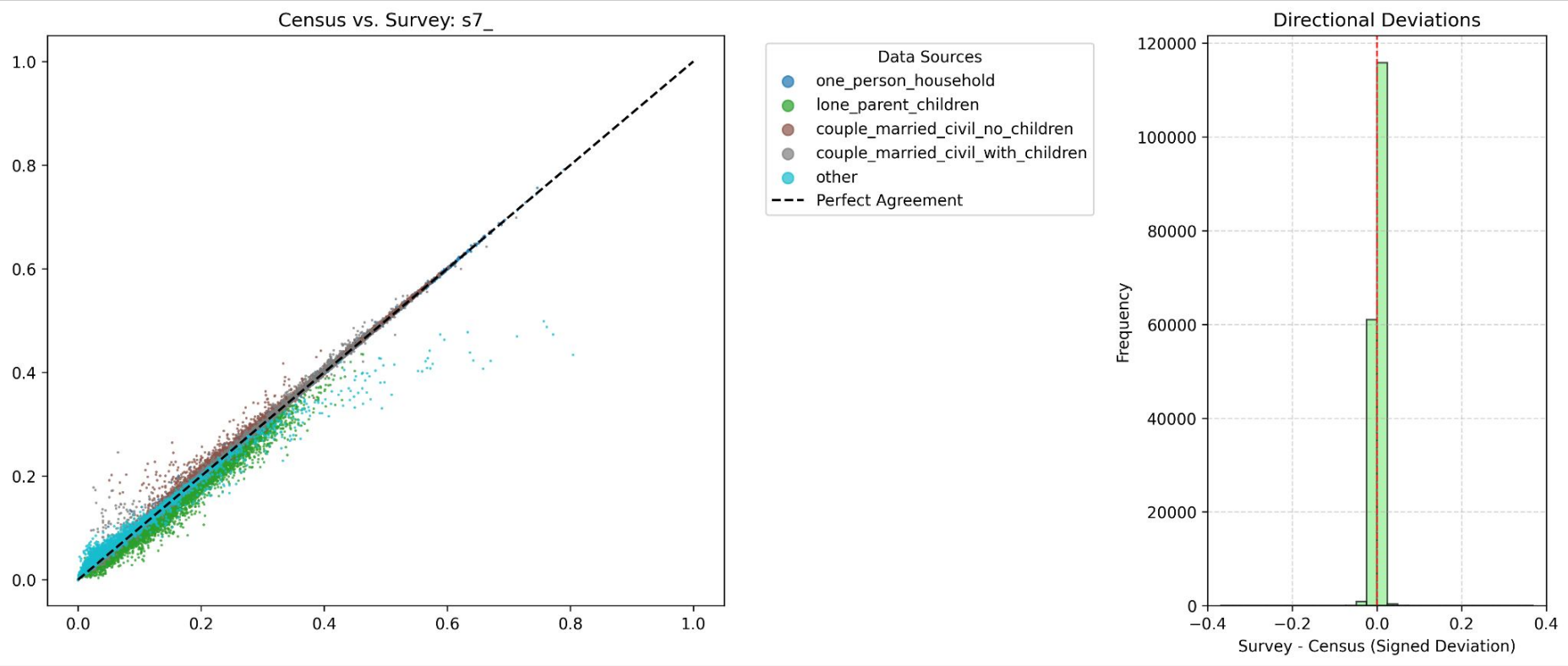
100 - ((N(Census) / N(SynthPop)) * 100)

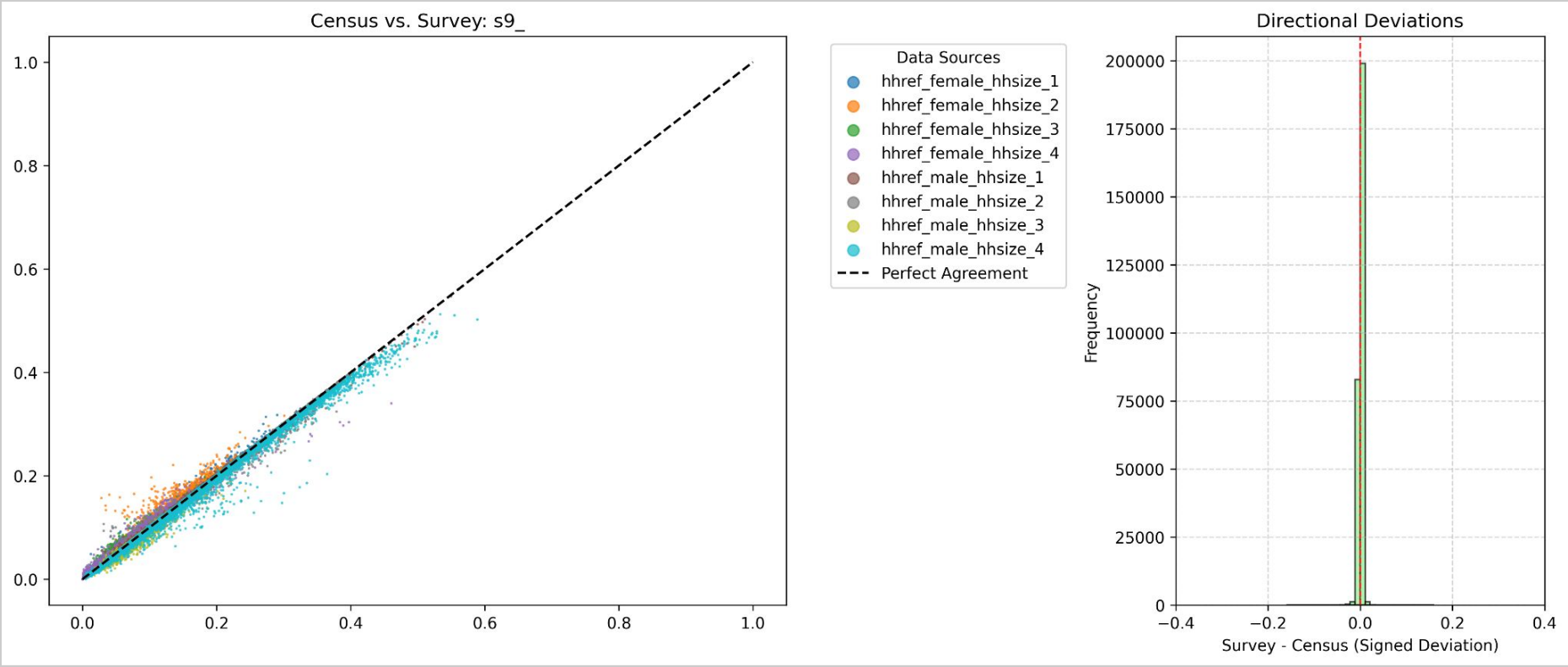
total population counts; by sex

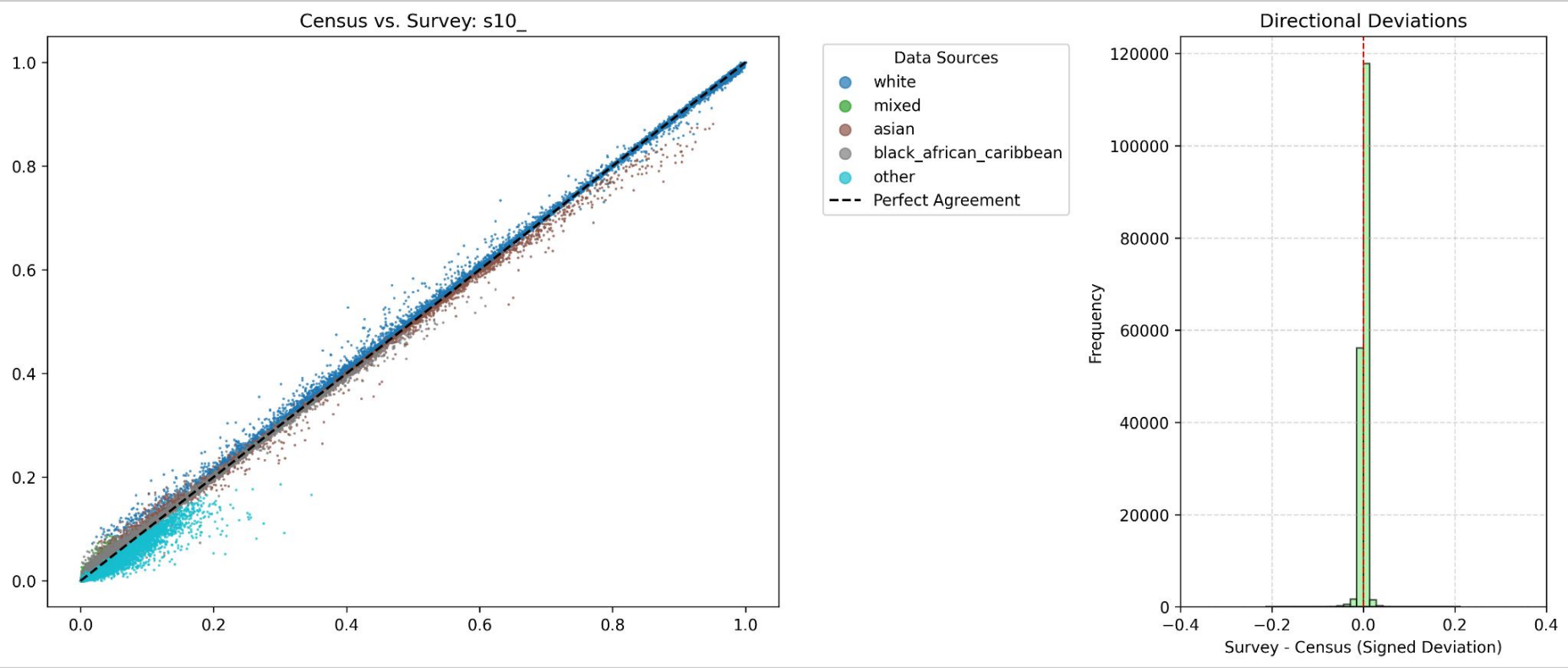


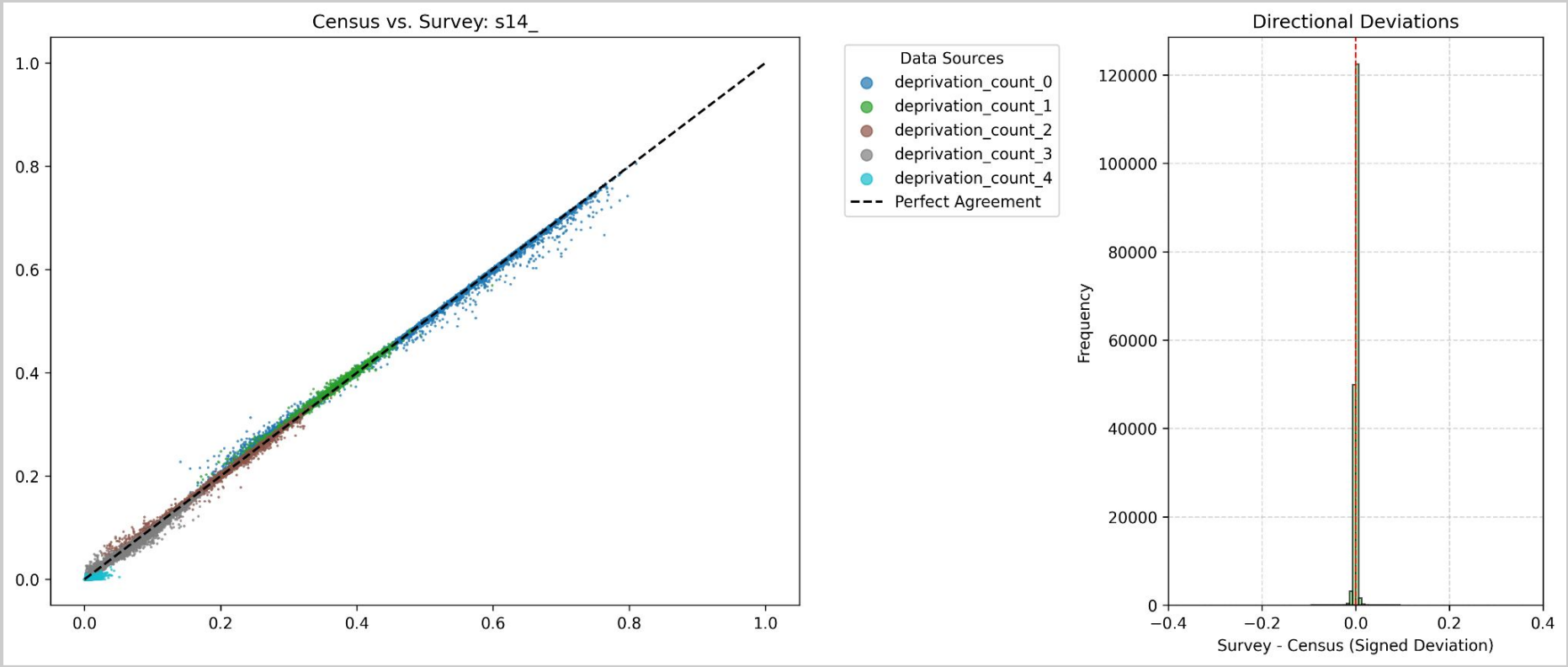
Creating synthetic area households for England and Wales


Census vs. Survey: s2_
Data Sources
hhsize_1
hhsize_2
hhsize_3
hhsize_4
hhsize_5
hhsize_6
hhsize_7
hhsize_8
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Census vs. Survey: s7_
Data Sources
one_person_household
lone_parent_children
couple_married_civil_no_children
couple_married_civil_with_children
other
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)


Census vs. Survey: s9_
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Data Sources
hhref_female_hhsize_1
hhref_female_hhsize_2
hhref_female_hhsize_3
hhref_female_hhsize_4
hhref_male_hhsize_1
hhref_male_hhsize_2
hhref_male_hhsize_3
hhref_male_hhsize_4
Perfect Agreement
Directional Deviations
Frequency
200000
175000
150000
125000
100000
75000
50000
25000
0
−0.4
−0.2
0.0
0.2
0.4
Survey - Census (Signed Deviation)


Census vs. Survey: s10_
Data Sources
white
mixed
asian
black_african_caribbean
other
Perfect Agreement
0.0
0.2
0.4
0.6
0.8
1.0
Directional Deviations
Frequency
120000
100000
80000
60000
40000
20000
0
−0.4
−0.2
0.0
0.2
0.4
Survey - Census (Signed Deviation)


Census vs. Survey: s14_
Data Sources
deprivation_count_0
deprivation_count_1
deprivation_count_2
deprivation_count_3
deprivation_count_4
Perfect Agreement
Directional Deviations
Frequency
Survey - Census (Signed Deviation)

#	Objectives	Key Deliverables	Start	End	Status
1.1	**To create synthetic individual population and household datasets**	1. An open-source synthetic population of GB individuals nested within households, which captures economic and health distributions in the population	1-Apr-24	31-Mar-26	
1.2	**To augment the population with relevant health and economic variables using matching techniques**	2. Open-source Python code for the (re)creation of the dataset and for the dynamic updating of the population	1-Apr-24	31-Mar-26	
1.3	**To use the synthetic population to analyse the distribution of economic and health outcomes for sub-groups and spatial zones**	3. Equity-focused analyses to support policy design	1-Apr-26	31-Mar-28	Not started

#	Objectives	Key Deliverables	Start	End	Status
1.1	**To create synthetic individual population and household datasets**	1. An open-source synthetic population of GB individuals nested within households, which captures economic and health distributions in the population	1-Apr-24	31-Mar-26	
1.2	**To augment the population with relevant health and economic variables using matching techniques**	2. Open-source GO code for the (re)creation of the dataset and for the dynamic updating of the population	1-Apr-24	31-Mar-26	
1.3	**To use the synthetic population to analyse the distribution of economic and health outcomes for sub-groups and spatial zones**	3. Equity-focused analyses to support policy design	1-Apr-26	31-Mar-28	Not started

https://ricci-colasanti.github.io/GoSynthPop/

Next ?

1. Individual education of England and Wales
2. Northern Ireland
3. Different household and individual constraints
4. Statistical analysis of output
5. I would like to develop GoSynthpop (UK-808) further
    a. Hugh's IPF (Anonymous populations)
    b. Other algorithms
        i. Genetic Algorithm
        ii. Particle swarm