

Beyond the *IMD – filling the data gap

Dr Mike Spencer
Dr Rosie Eccleston



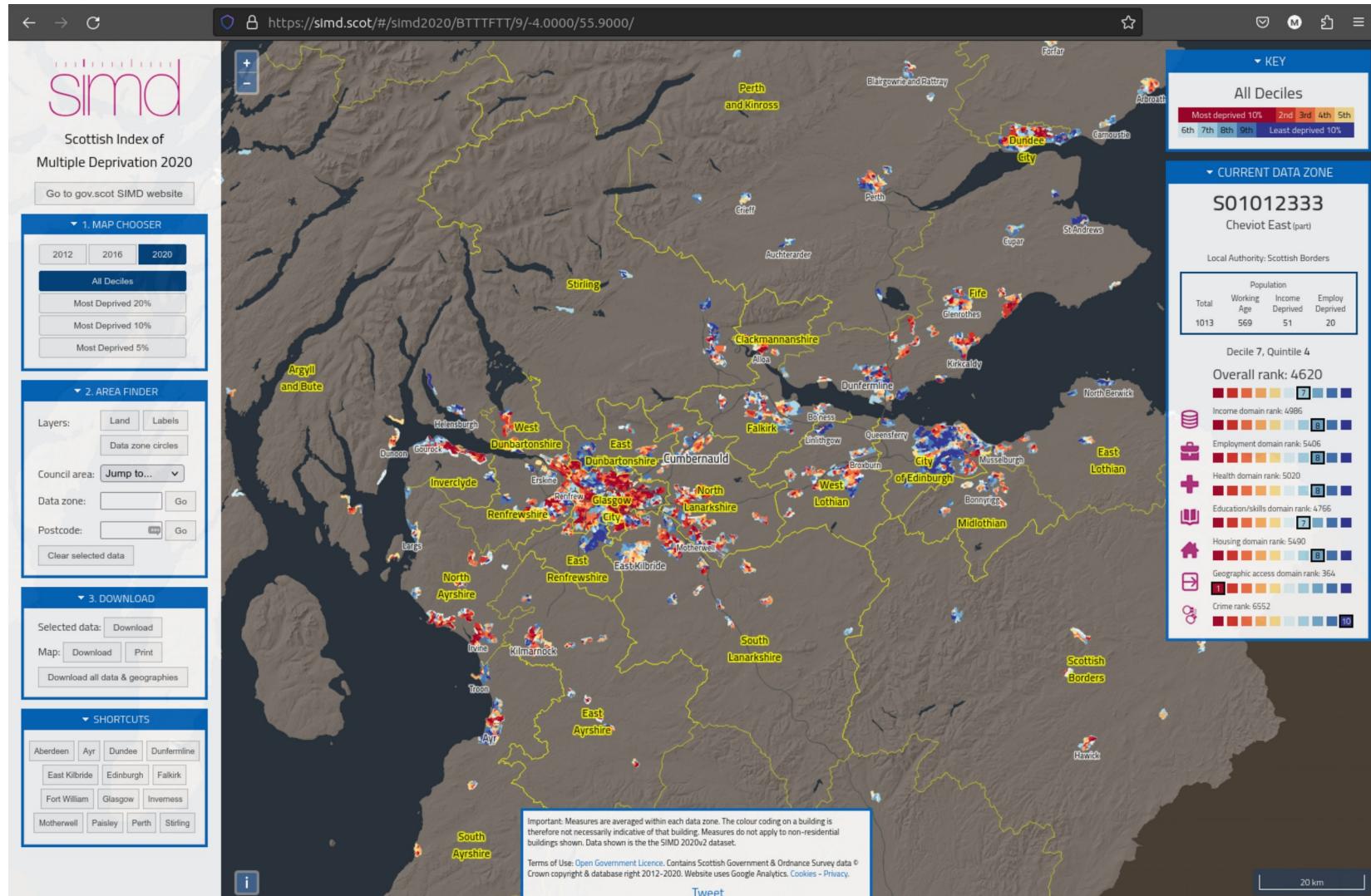
What is the *IMD?

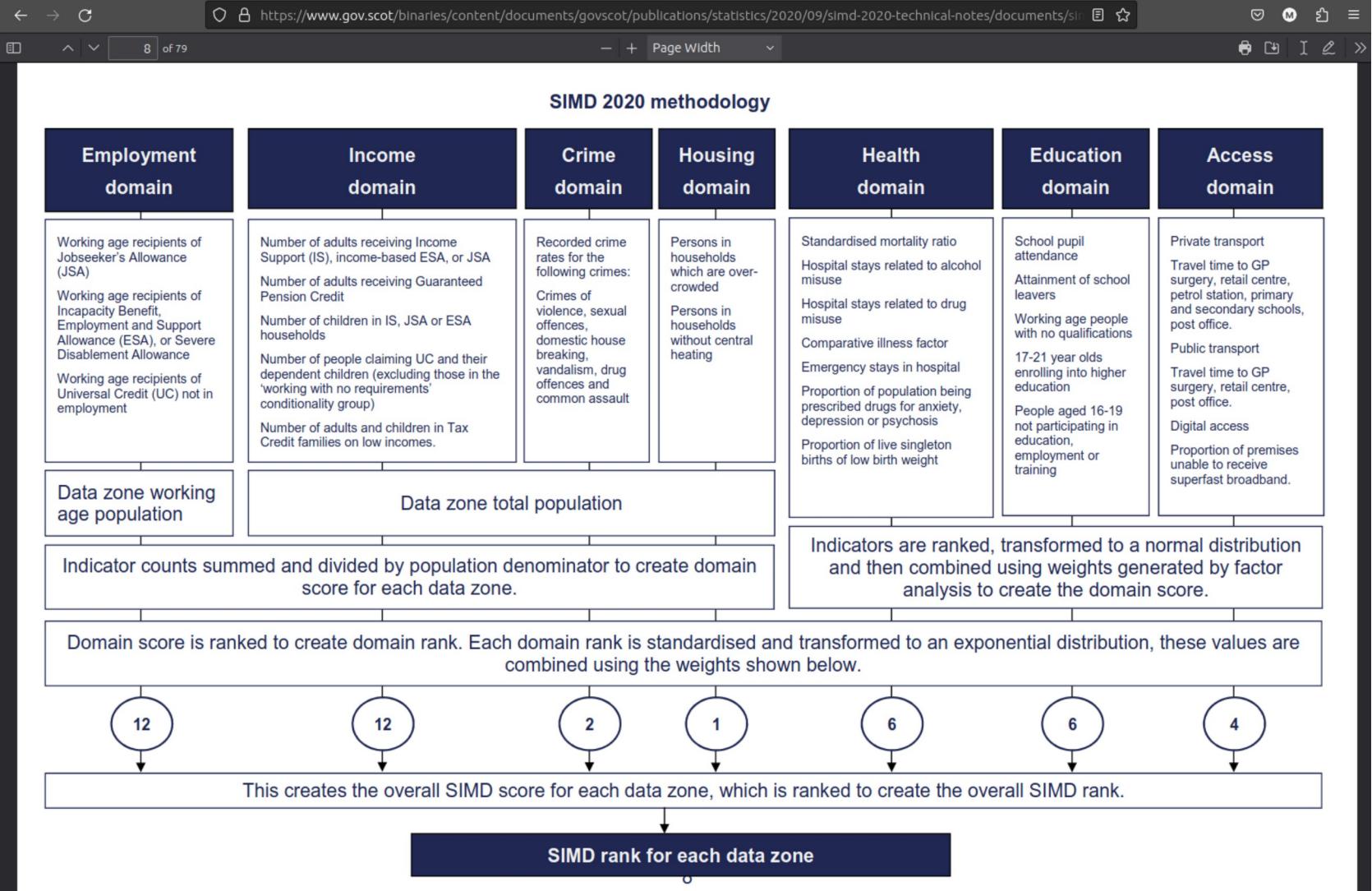
*IMD

- Different for England, Scotland & Wales
- Across health, income, access, education, etc.
- Used for policy and intervention
- National Statistics

Focus on SIMD







Limitations

- Infrequent updates
- A busy few years
- Labour intensive
- Benefits focused



An *IMD enhancement?

Desire

- Near real-time
- Administrative spatial resolution
- Discovers evolving situations

How can we help?

Partial solution

- Financial data
- Privacy first
- 5 safes
 - Projects, People,
 - Settings, Data,
 - & Outputs



What does this mean?

Toy data

```
> read_csv("data_in/finance_2019-01-06.csv") %>% glimpse()
```

Rows: 500000 Columns: 19

— Column specification —

Delimiter: ","

chr (2): cid, sex

dbl (16): income, income_salary, income_benefits, income_pension, income_investment, income_interest, income_other, expend...

date (1): end_of_this_period

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show_col_types = FALSE` to quiet this message.

Rows: 500,000

Columns: 19

\$ cid

<chr> "0372943541", "8718939101", "9576403582", "6434287673", "8384245418", "1877660945", "872680...

\$ sex

<chr> "M", "F", "F", "F", "M", "M", "F", "F", "F", "M", "M", "F", "M", "M", "F", "...

\$ end_of_this_period

<date> 2019-01-06, 2019-01-06, 2019-01-06, 2019-01-06, 2019-01-06, 2019-01-06, 2019-0...

\$ income

<dbl> 62384.298, 3984.488, 13459.366, 102496.911, 14291.991, 16236.704, 25905.508, 10583.545, 140...

\$ income_salary

<dbl> 12821.127, 59310.121, 17909.762, 13447.458, 16197.562, 26368.778, 18953.949, 78661.545, 299...

\$ income_benefits

<dbl> 1030.0490, 9868.1515, 1187.4295, 4605.2130, 743.6535, 834.0886, 2441.3281, 1011.7422, 2120...

\$ income_pension

<dbl> 12814.960, 38493.055, 8582.256, 24415.374, 6362.781, 1003.522, 5742.224, 7761.511, 3404.030...

\$ income_investment

<dbl> 69.43343, 90.77197, 263.00960, 78.18585, 26.86534, 59.99697, 198.48226, 52.14309, 161.82220...

\$ income_interest

<dbl> 31.69776, 385.90542, 23.60480, 188.74733, 367.46442, 103.29140, 78.88642, 367.41628, 98.814...

\$ income_other

<dbl> 1355.6848, 28890.9560, 2577.5952, 28988.3237, 5731.3177, 872.0811, 19111.8576, 1122.4333, 3...

\$ expenditure

<dbl> 59265.083, 3785.264, 12786.398, 97372.065, 13577.391, 15424.869, 24610.233, 10054.367, 1335...

\$ expenditure_committed

<dbl> 18715.289, 1195.346, 4037.810, 30749.073, 4287.597, 4871.011, 7771.652, 3175.063, 4218.116...

\$ expenditure_essential

<dbl> 31192.149, 1992.244, 6729.683, 51248.455, 7145.995, 8118.352, 12952.754, 5291.772, 7030.194...

\$ expenditure_qol

<dbl> 6238.4298, 398.4488, 1345.9366, 10249.6911, 1429.1991, 1623.6704, 2590.5508, 1058.3545, 140...

\$ expenditure_discretionary

<dbl> 3119.2149, 199.2244, 672.9683, 5124.8455, 714.5995, 811.8352, 1295.2754, 529.1772, 703.0194...

\$ expenditure_uncategorized

<dbl> 3119.2149, 199.2244, 672.9683, 5124.8455, 714.5995, 811.8352, 1295.2754, 529.1772, 703.0194...

\$ cash_balance_final

<dbl> 308.4810, 332.9998, 332.8288, 317.8296, 237.9787, 314.0453, 263.2046, 346.5375, 324.4015, 4...

\$ cash_min

<dbl> 99.71669, 142.70067, 139.40952, 93.90980, 155.48634, 99.35683, 133.75753, 118.08549, 79.321...

\$ cash_max

<dbl> 965.3130, 1056.2350, 1029.5984, 980.3929, 941.5497, 1011.3531, 948.6580, 1068.6138, 994.693...

Transforming individual data to statistics

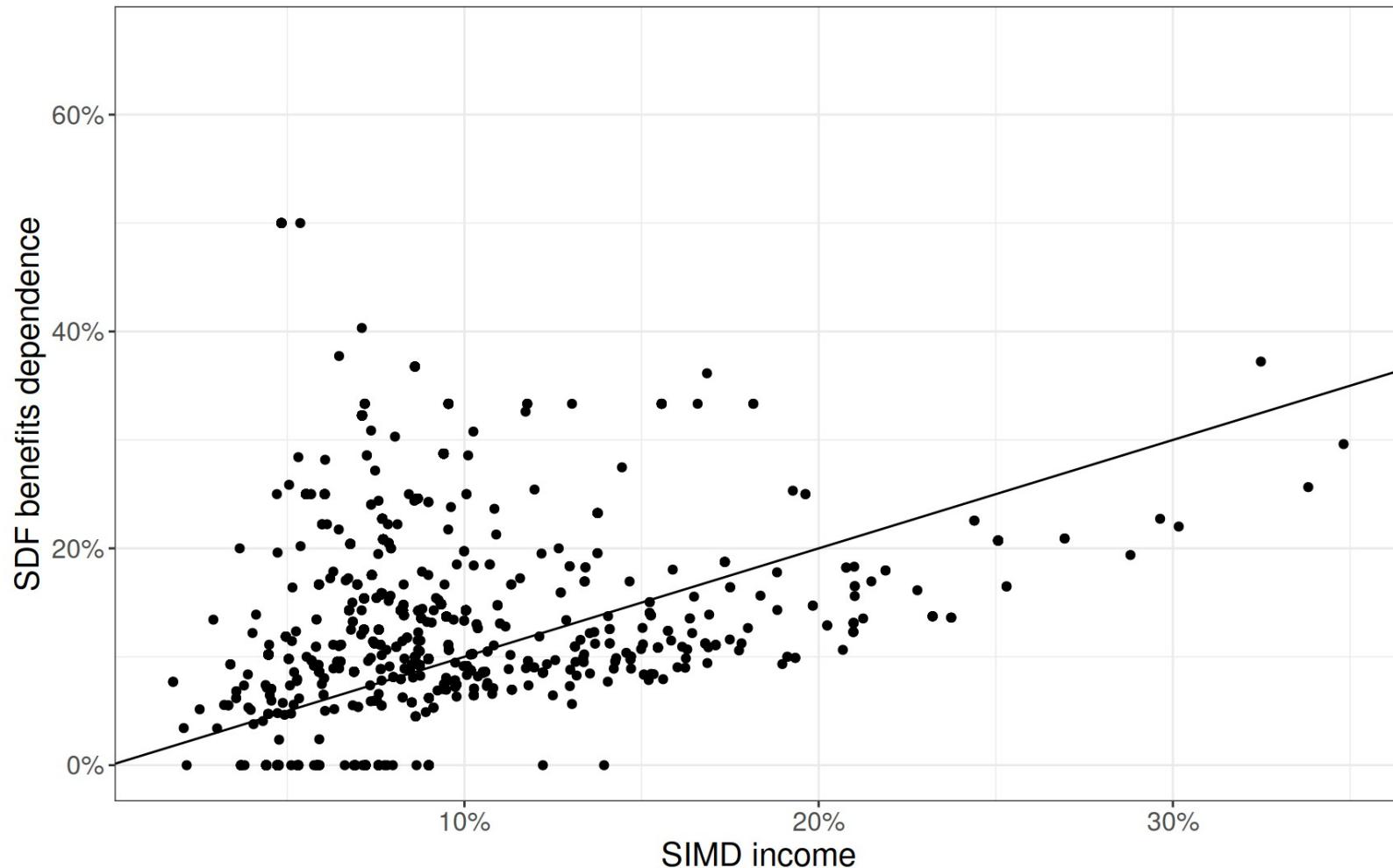
Sample metrics

- Benefits dependence
- Overdraft use
- Income

Caveats

- Work in progress!
- Small(ish) sample size
 - Please ask for more info
- Crude spatial joins

SIMD income compared to benefits dependence by postcode district

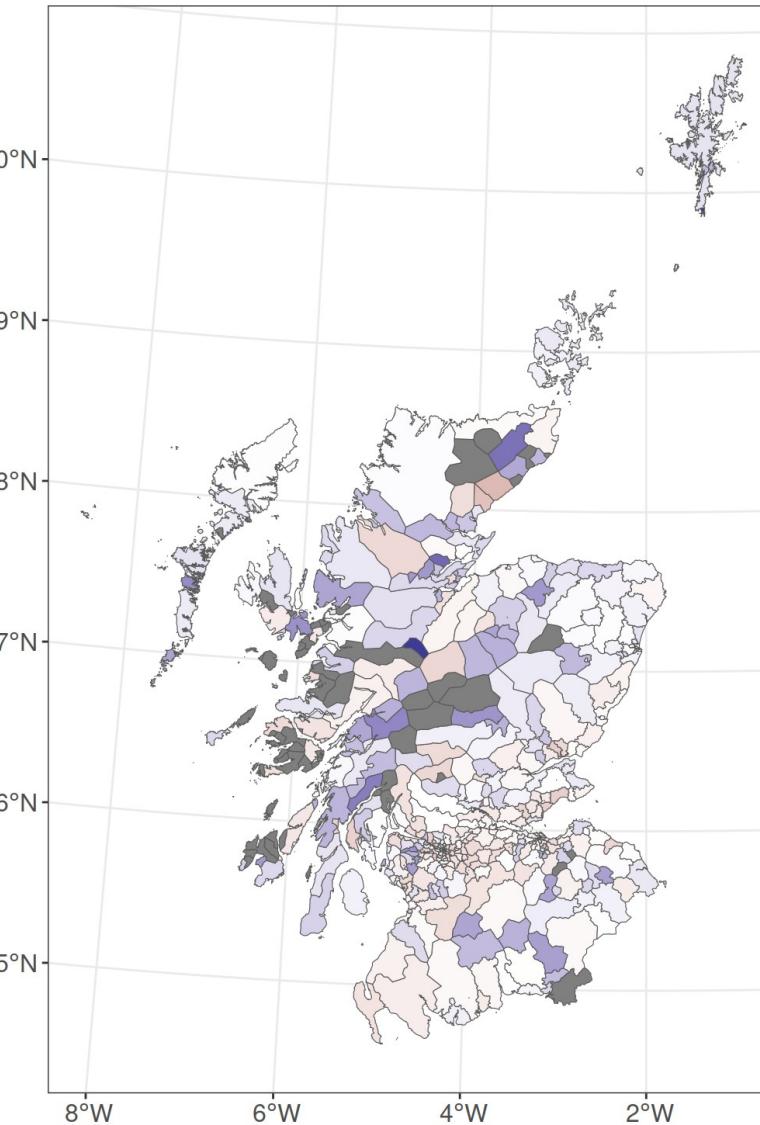


SIM

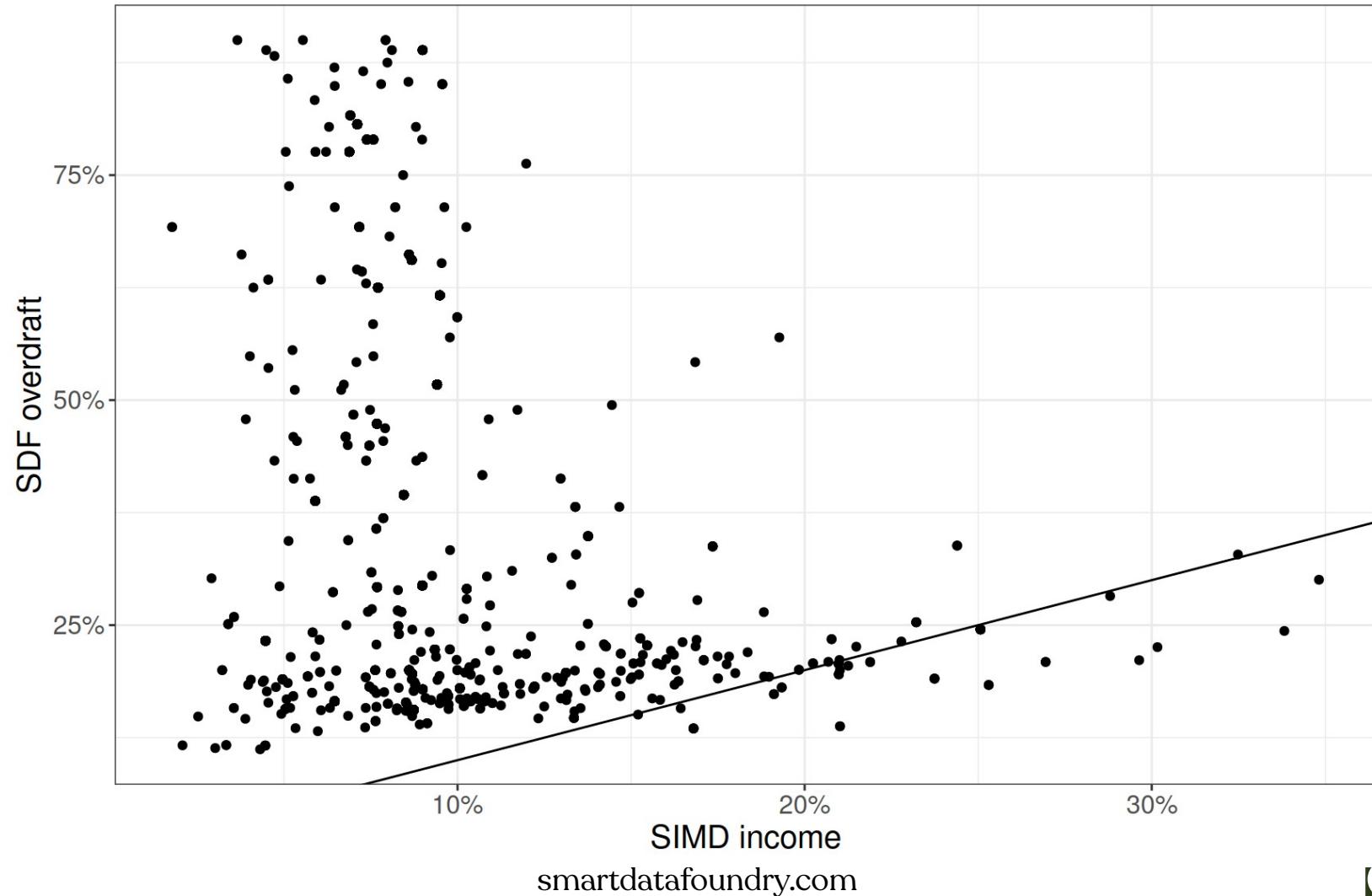
SDF benefits dependence

60%
40%
20%
0%

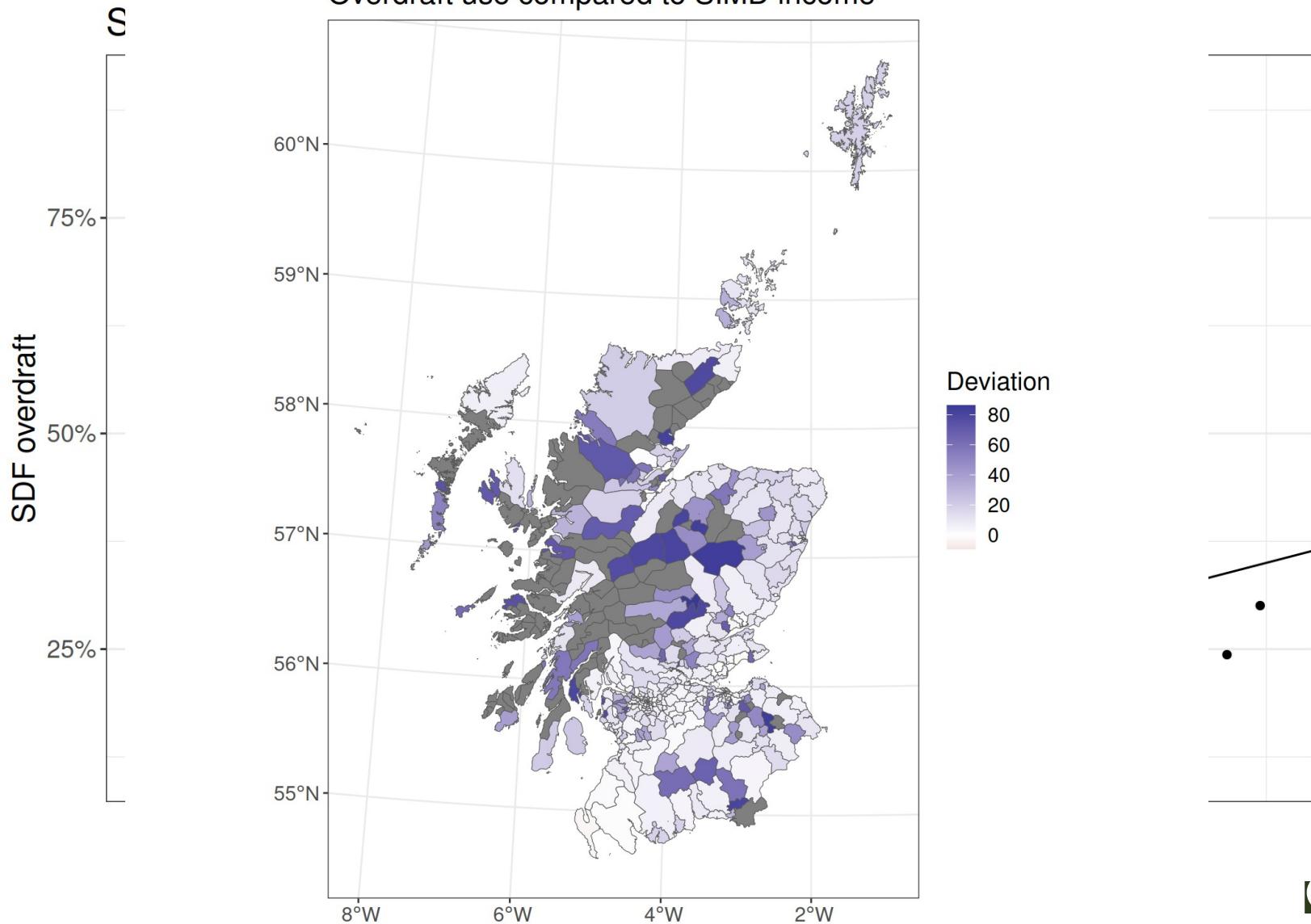
Benefits dependence compared to SIMD income



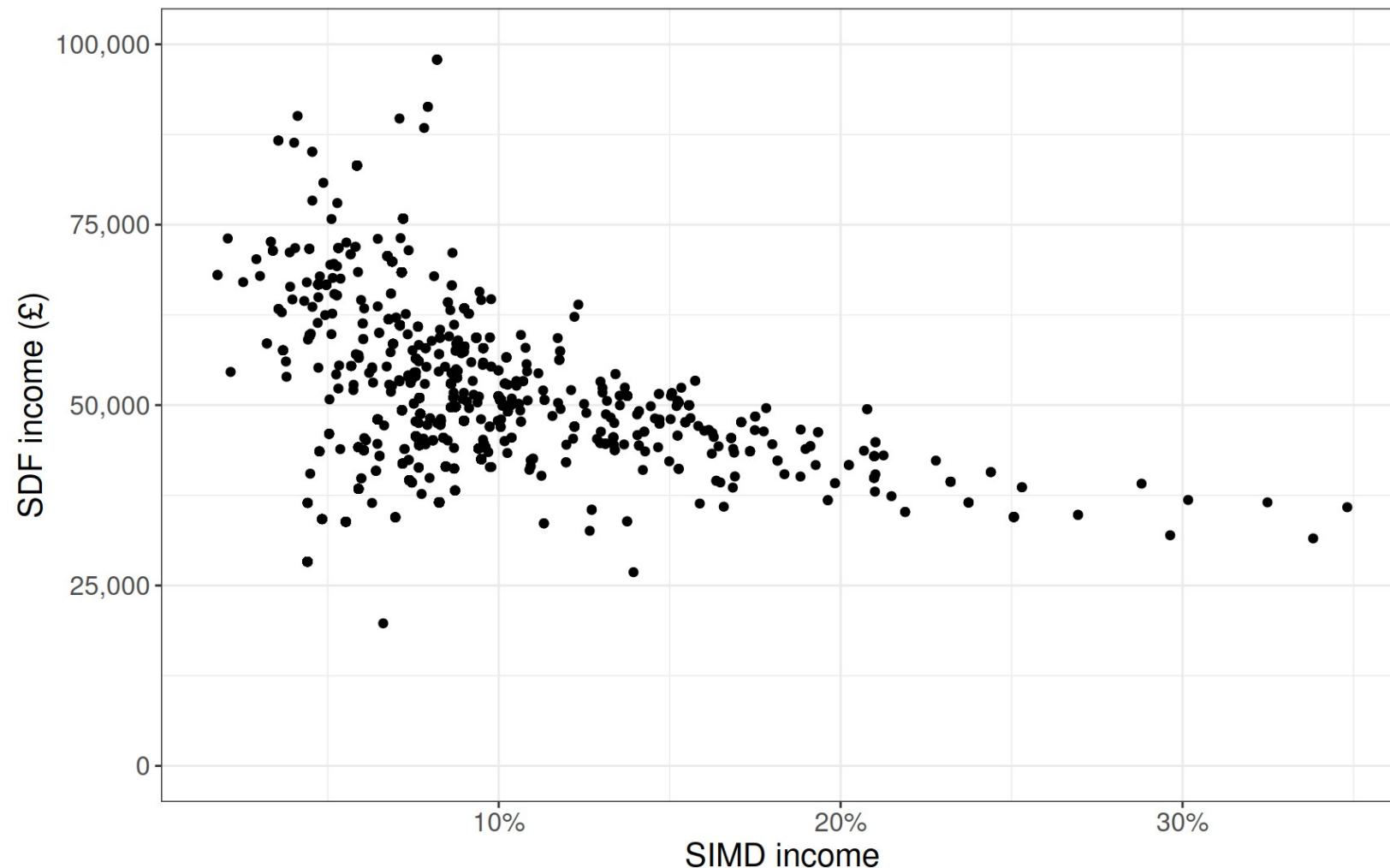
SIMD income compared to overdraft use by postcode district



Overdraft use compared to SIMD income

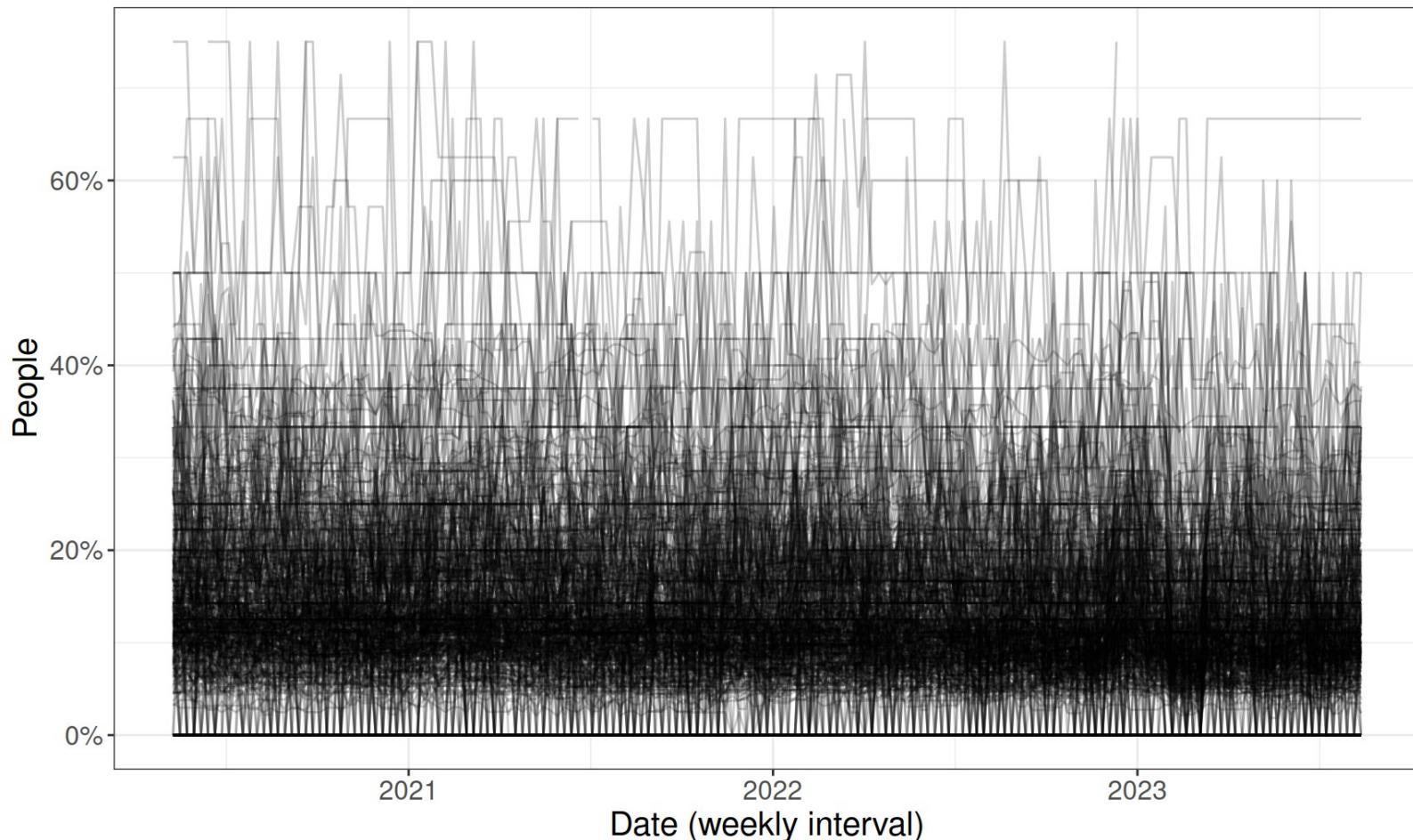


SIMD income compared to annual income by postcode district



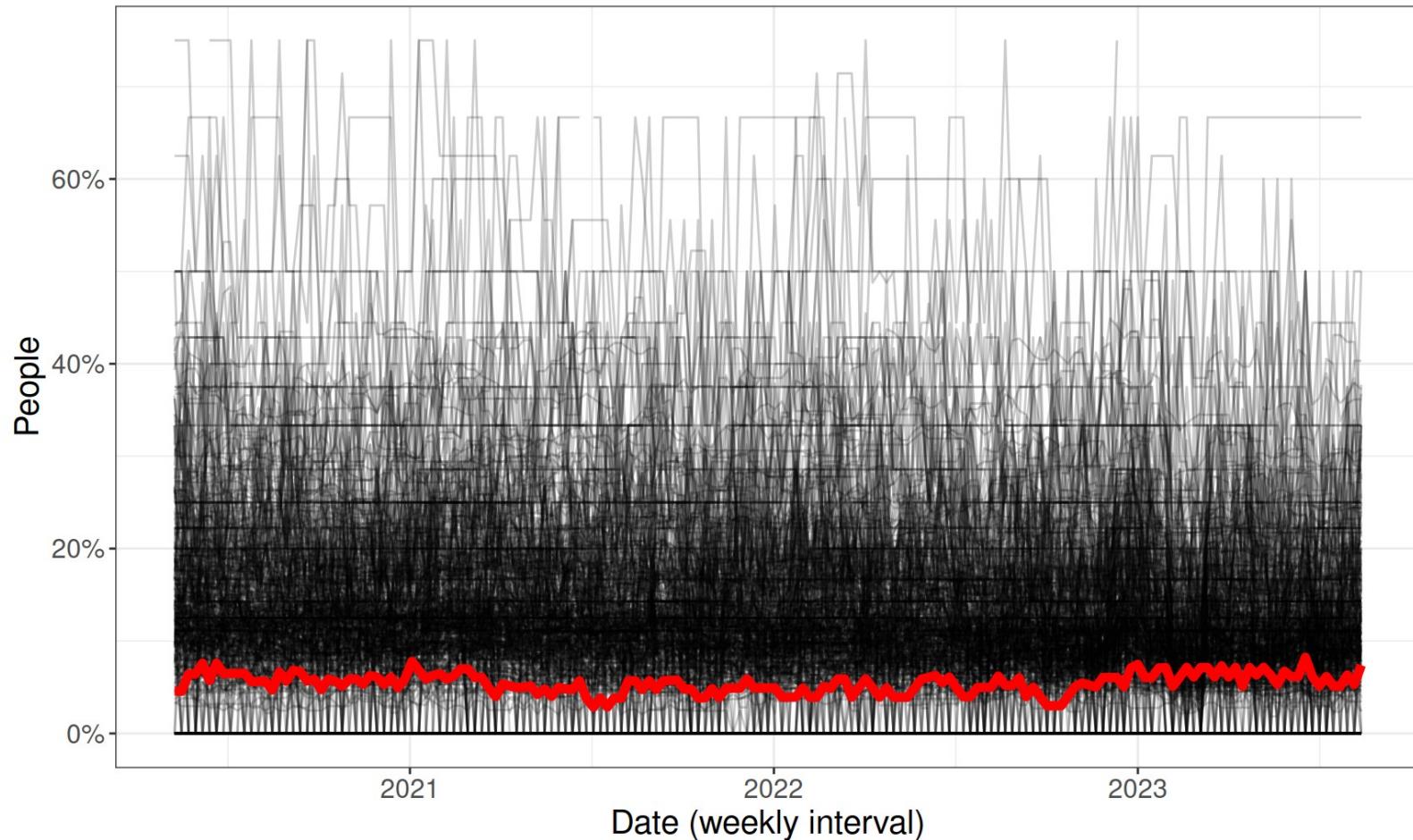
... as a time series

Benefit dependence in Scotland by postcode district

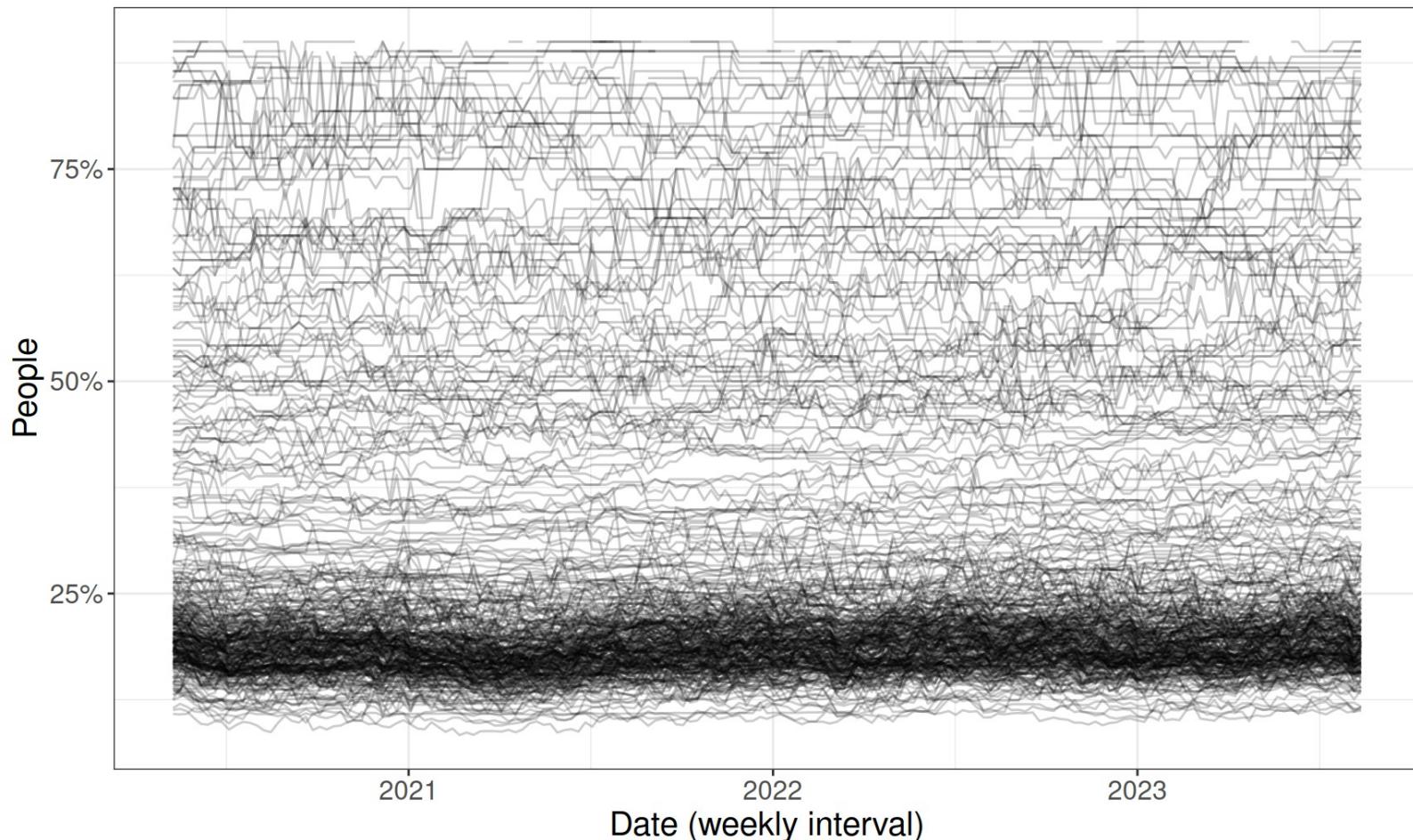


Benefit dependence in Scotland by postcode district

G12 postcode highlighted

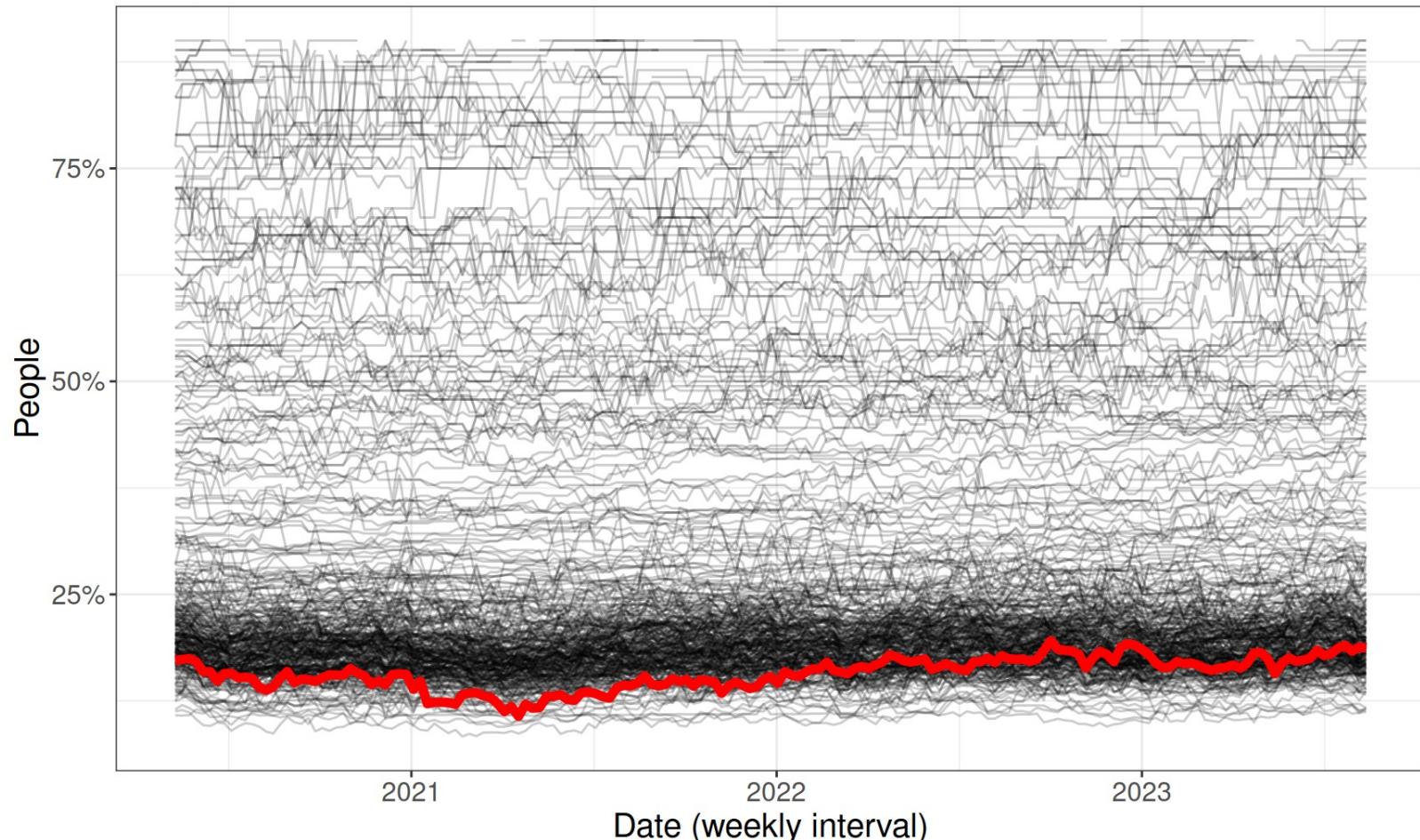


Overdraft use in Scotland by postcode district

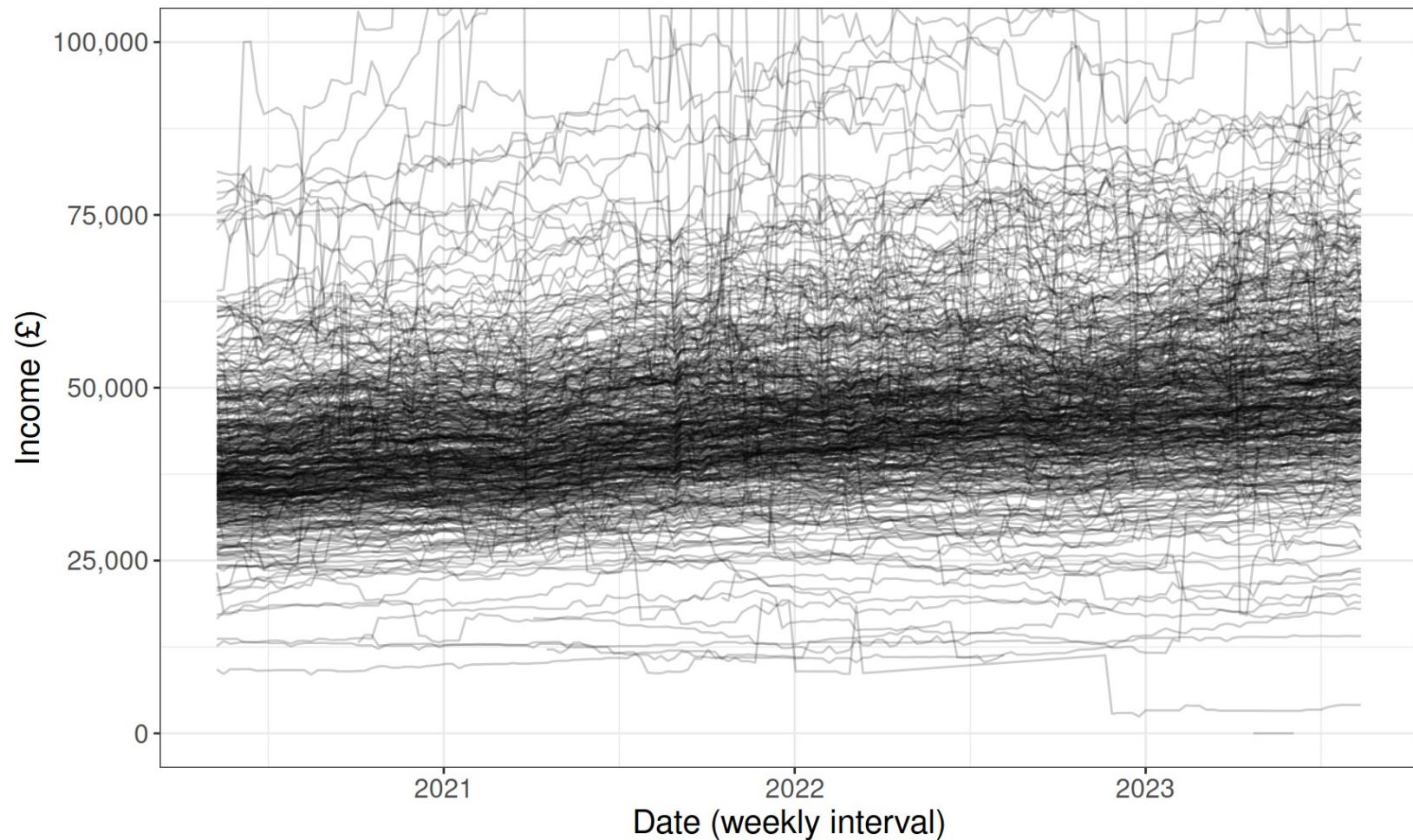


Overdraft use in Scotland by postcode district

G12 postcode highlighted

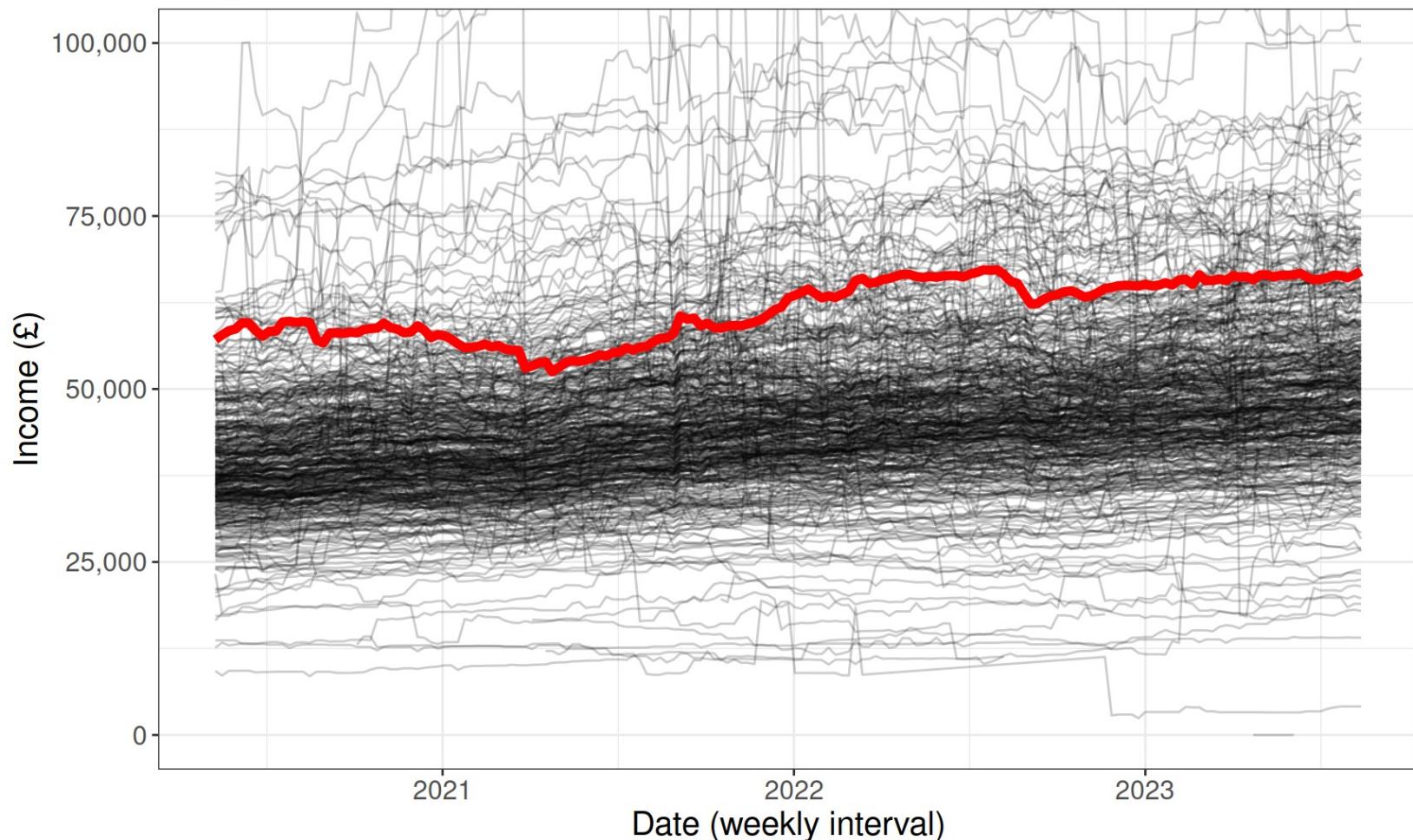


Income in Scotland by postcode district



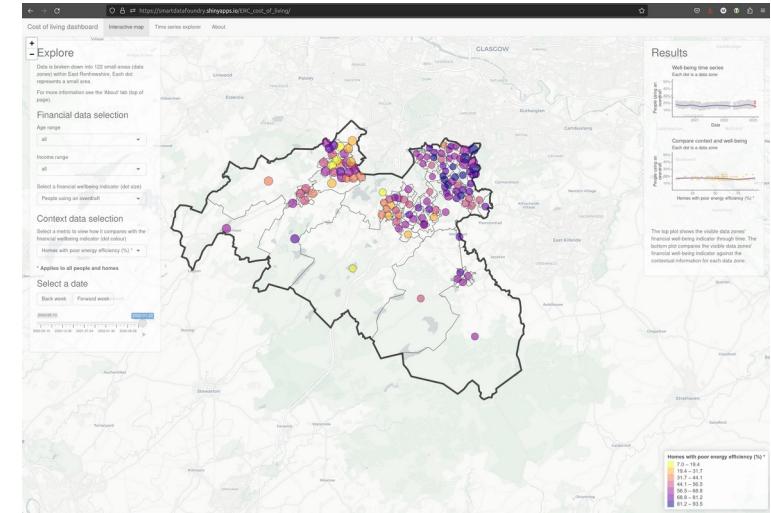
Income in Scotland by postcode district

G12 postcode highlighted



Use cases

- East Renfrewshire Council
- Interventions and baselines
- Canary in the coalmine
- Planning & policy



Future development

- England and Wales
- Other metrics
 - Net zero homes
 - Child poverty
 - Public health



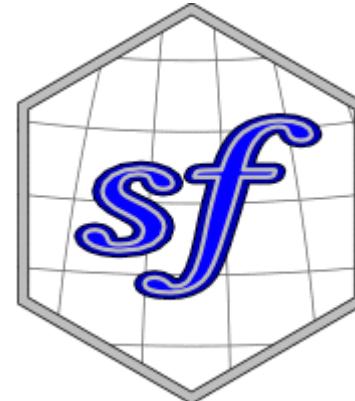
Maps with R

- Why?
- When?
- Is it easy?

Maps with R

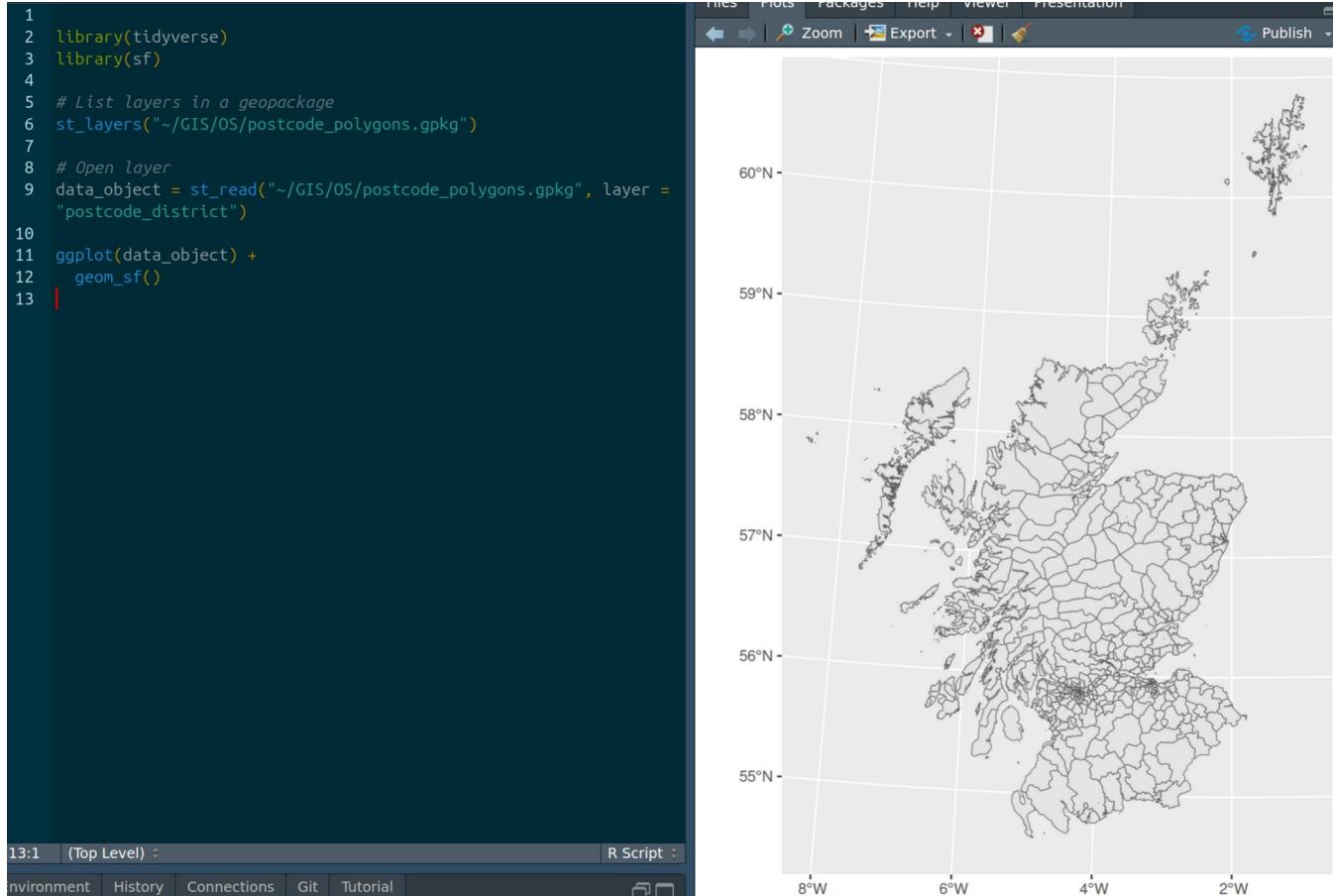
- Why?
- When?
- Is it easy?

<https://geocompx.org/>



Maps with R

```
1 library(tidyverse)
2 library(sf)
3
4
5 # List layers in a geopackage
6 st_layers("~/GIS/OS/postcode_polygons.gpkg")
7
8 # Open layer
9 data_object = st_read("~/GIS/OS/postcode_polygons.gpkg", layer =
  "postcode_district")
10
11 ggplot(data_object) +
12   geom_sf()
13 |
```



Maps with R

```
1
2 library(tidyverse)
3 library(sf)
4
5 # List layers in a geopackage
6 st_layers("~/GIS/OS/postcode_polygons.gpkg")
7
8 # Open layer
9 data_object = st_read("~/GIS/OS/postcode_polygons.gpkg",
10                       layer = "postcode_district")
11
12 ggplot(data_object) +
13   geom_sf()
14
```



Summary

- R is great for data visualisation maps!
- Finance data can supplement statistics publications
- Privacy and trust are key
- Collaborators and use cases welcome

Contact

- michael.spencer@smartdatafoundry.com
- [@mikerspencer@mastodon.scot](https://mastodon.scot/@mikerspencer)
- linkedin.com/in/mikerspencer/
- [@mikerspencer:matrix.org](https://matrix.org/@mikerspencer)
- smartdatafoundry.com