

Codebook for 2020 Scottish Index of Multiple Deprivation Dataset

Brian Fogarty

This is the codebook for the 2020 Scottish Index of Multiple Deprivation dataset (`simd2020.csv`) used in *Quantitative Social Science Data with R: An Introduction* (2nd Edition). The only variable not from original 2020 SIMD dataset is `urban`. Note: missing values are represented by `*` in the csv file. Information about the variable values is provided when necessary.¹

- `Data_Zone`: datazone number.
- `Intermediate_Zone`: intermediate zone area.
- `Council_area`: council area of datazone.
- `Total_population`: population.
- `Working_age_population`: working age population.
- `Income_rate`: percentage of people who are income deprived; need to multiply by 100 to get the percentage out of 100%. This variable measures the percentage of people in a datazone receiving income support, employment and support allowance, job seekers allowance, guaranteed pension credits, and child and working tax credits; essentially financial support from the government for low income and unemployed individuals.
- `Income_count`: raw count of people who are income deprived. This variable measures the number of people in a datazone receiving income support, employment and support allowance, job seekers allowance, guaranteed pension credits, and child and working tax credits; essentially financial support from the government for low income and unemployed individuals.
- `Employment_rate`: percentage of people who are employment deprived; need to multiply by 100 to get the percentage out of 100%. This variable measures the percentage of people in a datazone receiving incapacity benefits, employment and support allowance, or severe disablement allowance; essentially financial support from the government due to being unable to work.
- `Employment_count`: raw count of people who are employment deprived. This variable measures the percentage of people in a datazone receiving incapacity benefits, employment and support allowance, or severe disablement allowance; essentially financial support from the government due to being unable to work.
- `CIF`: standardised ratio of comparative illness factor. This variable measures the number of people in a datazone receiving one or more of disabled living allowance (DLA), attendance allowance, incapacity benefit (not receiving DLA), employment support allowance, and severe disablement allowance. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.
- `ALCOHOL`: standardised ratio of hospital stays related to alcohol misuse. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.
- `DRUG`: standardised ratio of hospital stays related to drugs misuse. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.
- `SMR`: standardised ratio of mortality. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.

¹Scottish Government, UK Open Government Licence. <https://www.gov.scot/Topics/Statistics/SIMD>.

- **DEPRESS**: percentage of people prescribed psychiatric medication for anxiety, depression, or psychosis; need to multiply by 100 to get the percentage out of 100%.
- **LBWT**: percentage of infants of low birth weight ($< 2.5\text{kg}$); need to multiply by 100 to get the percentage out of 100%.
- **EMERG**: standardised ratio of emergency hospital stays. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.
- **Attendance**: school pupil attendance percentage; need to multiply by 100 to get the percentage out of 100%.
- **Attainment**: attainment of school leavers (as a score). The score is based on school leavers' highest level of qualification for pupils who attend publicly funded secondary schools.
- **no_qualifications**: standardised ratio of people with no education qualifications. A value of 100 indicates the Scottish average for datazones of the same age and gender profile.
- **not_participating**: percentage of 16-19 year olds not participating in education, employment, or training; need to multiply by 100 to get the percentage out of 100%.
- **University**: percentage of 17-21 year olds entering university; need to multiply by 100 to get the percentage out of 100%.
- **drive_petrol**: average drive time to a petrol station in minutes.
- **drive_GP**: average drive time to a GP surgery in minutes.
- **drive_post**: average drive time to a post office in minutes.
- **drive_primary**: average drive time to a primary school in minutes.
- **drive_retail**: average drive time to a retail centre in minutes.
- **drive_secondary**: average drive time to a secondary school in minutes.
- **PT_GP**: public transport travel time to a GP surgery in minutes.
- **PT_post**: public transport travel time to a post office in minutes.
- **PT_retail**: public transport travel time to a retail centre in minutes.
- **Broadband**: percentage of premises without access to superfast broadband (at least 30Mb/s download speed); need to multiply by 100 to get the percentage out of 100%.
- **crime_count**: number of recorded crimes of violence, sexual offences, domestic housebreaking, vandalism, drugs offences, and common assault.
- **crime_rate**: recorded crimes of violence, sexual offences, domestic housebreaking, vandalism, drugs offences, and common assault per 10,000 people.
- **overcrowded_count**: number of people living in households that are overcrowded.
- **nocentralheat_count**: number of people living in households with no central heating.
- **overcrowded_rate**: percentage of people living in households that are overcrowded; need to multiply by 100 to get the percentage out of 100%.
- **nocentralheat_rate**: percentage of people living in households with no central heating; need to multiply by 100 to get the percentage out of 100%.
- **urban**: whether datazone is urban or rural (dummy variable).
 0. Rural
 1. Urban