EPC data:
Technical documentation

Creation date	2021-03-08
Edition	02
Author	Ellen Webborn
Project	Smart Energy Research Lab (SERL)
Organisation	University College London (UCL)

Introduction

This document describes the England and Wales Energy Performance Certificate (EPC) data collected for SERL participants, stored in the file *serl_epc_data_edition02.csv*. The data contains 80 columns and 2423 rows (one row per participant with available EPC data). This document lists the EPC variables available along with basic information about the values for each variable such as number of unique values and statistics for numerical variables. A guide to the variables is available here.

Data were collected with the Domestic Energy Performance Certificates API using the house number and postcode (details here). Where more than one EPC is registered for an address the most recent is provided.

The data have not been modified from the original source except for the removal of address data (replaced with our PUPRN (a unique identifier) used in the other datasets). Note that data quality analysis has not been performed for this dataset.

Data summary

Table 1 lists all variables currently available in the SERL EPC dataset. The number of unique values is given, alongside the R data class and an example value from the dataset.

For variables with fewer than 10 unique values in the EPC dataset, Table 2 shows the number of records with each value and the percent with this value (or non-value in the case of N/A or 'NO DATA!' etc.). We also include PUPRN to show the number of records.

Table 1: All EPC variables, the number of unique values found for each variable, the variable (R) class, and an example from the dataset. In some cases examples are invented (for statistical disclosure control) as the count for unique reads is very low; in these instances examples take the same format as the real data.

variable	n unique values	class	example
PUPRN	2,423	character	1ABC2DE3
currentEnergyRating	7	character	C
potentialEnergyRating	7	character	B
currentEnergyEfficiency	85	integer	71
potentialEnergyEfficiency	86	integer	84
propertyType	5	character	Bungalow
builtForm	7	character	Detached
inspectionDate	1,694	character	13/05/2008
localAuthority	304	character	E09000032
constituency	506	character	E14001013
lodgementDate	1,738	character	01/04/2009
transactionType	15	character	marketed sale
environmentImpactCurrent	92	integer	47
environmentImpactPotential	84	integer	90
energyConsumptionCurrent	476	integer	234
energyConsumptionPotential	400	integer	52
co2EmissionsCurrent	117	numeric	1.2
co2EmissCurrPerFloorArea	211	numeric	30
co2EmissionsPotential	115	numeric	5
lightingCostCurrent	187	integer	48
lightingCostPotential	127	integer	61
heatingCostCurrent	1,108	integer	270
heatingCostPotential	911	integer	356
hotWaterCostCurrent	304	integer	75
hotWaterCostPotential	200	integer	57
totalFloorArea	792	numeric	80.00
energyTariff	7	character	Single
mainsGasFlag	3	character	Y
floorLevel	18	character	Ground
flatTopStorey	3	character	Y
flatStoreyCount	8	integer	2
mainHeatingControls	38	integer	2106

variable	n unique values	class	example
multiGlazeProportion	62	integer	100
glazedType	9	character	double glazing, unknown install date
glazedArea	6	character	Much More Than Typical
extensionCount	6	integer	2
numberHabitableRooms	14	integer	3
numberHeatedRooms	14	integer	5
lowEnergyLighting	96	integer	67
numberOpenFireplaces	8	integer	0
hotwaterDescription	20	character	From main system
hotWaterEnergyEff	6	character	Good
hotWaterEnvEff	6	character	Good
floorDescription	69	character	(another dwelling below)
floorEnergyEff	6	character	Very Good
floorEnvEff	5	character	Good
windowsDescription	19	character	Mostly double glazing
windowsEnergyEff	6	character	Average
windowsEnvEff	6	character	Average
wallsDescription	91	character	Cavity wall, filled cavity
wallsEnergyEff	6	character	Poor
wallsEnvEff	6	character	Good
secondheatDescription	15	character	Room heaters, wood logs
sheatingEnergyEff	1	character	N/A
sheatingEnvEff	1	character	N/A
roofDescription	90	character	Pitched, 300+ mm loft insulation
roofEnergyEff	6	character	Good
roofEnvEff	6	character	Average
mainheatDescription	37	character	Boiler and radiators, mains gas
mainheatEnergyEff	6	character	Good
mainheatEnvEff	6	character	Good
mainheatcontDescription	32	character	Programmer and room thermostat
mainheatcEnergyEff	6	character	Average
mainheatcEnvEff	6	character	Average

variable	n unique values	class	example
lightingDescription	100	character	Low energy lighting in 18% of fixed outlets
lightingEnergyEff	7	character	Very Good
lightingEnvEff	6	character	Good
mainFuel	23	character	oil (not community)
windTurbineCount	4	integer	0
heatLossCorridoor	4	character	unheated corridor
unheatedCorridorLength	154	numeric	5.189
floorHeight	105	numeric	2.40
photoSupply	9	integer	0
solarWaterHeatingFlag	3	character	N
mechanicalVentilation	4	character	natural
constructionAgeBand	14	character	England and Wales: 1930-1949
lodgementDatetime	2,421	character	02/10/2010 14:51
tenure	6	character	owner-occupied
fixedLightingOutletsCount	48	integer	5
lowEnergyFixedLightCount	35	integer	4

Table 2: The number and percent of each value found in the dataset for each variable with fewer than 10 unique values found. Note that number is rounded down to the nearest 10 for statistical disclosure control, and percent is the rounded number as a percent of total (so may not sum to 100%).

variable	value	number (rounded)	percent
currentEnergyRating	A	0	0.0
	B	170	7.0
	C	690	28.5
	D	1,050	43.3
	E	390	16.1
	F	60	2.5
	G	20	0.8

variable	value	number (rounded)	percent
potentialEnergyRating	A	60	2.5
	B	1,130	46.6
	C	930	38.4
	D	200	8.3
	E	60	2.5
	F	10	0.4
	G	0	0.0
propertyType	Bungalow	330	13.6
	Flat	400	16.5
	House	1,630	67.3
	Maisonette	40	1.7
	Park home	0	0.0
builtForm	Detached	700	28.9
	Enclosed End-Terrace	30	1.2
	Enclosed Mid-Terrace	10	0.4
	End-Terrace	300	12.4
	Mid-Terrace	570	23.5
	NO DATA!	30	1.2
	Semi-Detached	750	31.0
energyTariff	NO DATA!	0	0.0
	Single	1,820	75.1
	Unknown	210	8.7
	dual	210	8.7
	dual (24 hour)	0	0.0
	off-peak 7 hour	0	0.0
	standard tariff	150	6.2
mainsGasFlag		150	6.2
	N	260	10.7
	Y	2,000	82.5
flatTopStorey		2,020	83.4
	N	230	9.5
	Y	160	6.6

variable	value	number (rounded)	percent
flatStoreyCount		2,330	96.2
	2	30	1.2
	3	30	1.2
	4	10	0.4
	5	0	0.0
	6	0	0.0
	7	0	0.0
	17	0	0.0
glazedType	INVALID!	0	0.0
	NO DATA!	150	6.2
	double glazing installed before 2002	760	31.4
	double glazing installed during or after 2002	740	30.5
	double glazing, unknown install date	600	24.8
	not defined	90	3.7
	secondary glazing	20	0.8
	single glazing	10	0.4
	triple glazing	0	0.0
glazedArea	Less Than Typical	0	0.0
	More Than Typical	50	2.1
	Much Less Than Typical	0	0.0
	Much More Than Typical	10	0.4
	NO DATA!	150	6.2
	Normal	2,180	90.0
extensionCount		150	6.2
	0	1,350	55.7
	1	660	27.2
	2	200	8.3
	3	30	1.2
	4	10	0.4

variable	value	number (rounded)	percent
numberOpenFireplaces		60	2.5
	0	2,000	82.5
	1	270	11.1
	2	60	2.5
	3	0	0.0
	4	0	0.0
	5	0	0.0
	7	0	0.0
hotWaterEnergyEff	Average	340	14.0
	Good	1,630	67.3
	N/A	10	0.4
	Poor	150	6.2
	Very Good	160	6.6
	Very Poor	110	4.5
hotWaterEnvEff	Average	290	12.0
	Good	1,680	69.3
	N/A	10	0.4
	Poor	190	7.8
	Very Good	160	6.6
	Very Poor	60	2.5
floorEnergyEff	Average	0	0.0
	Good	20	0.8
	N/A	1,290	53.2
	NO DATA!	1,010	41.7
	Poor	0	0.0
	Very Good	80	3.3
floorEnvEff	Average	0	0.0
	Good	20	0.8
	N/A	2,300	94.9
	Poor	0	0.0
	Very Good	80	3.3

variable	value	number (rounded)	percent
windowsEnergyEff	Average	1,290	53.2
	Good	750	31.0
	N/A	0	0.0
	Poor	130	5.4
	Very Good	110	4.5
	Very Poor	120	5.0
windowsEnvEff	Average	1,290	53.2
	Good	750	31.0
	N/A	0	0.0
	Poor	130	5.4
	Very Good	110	4.5
	Very Poor	120	5.0
wallsEnergyEff	Average	180	7.4
	Good	1,150	47.5
	N/A	0	0.0
	Poor	370	15.3
	Very Good	120	5.0
	Very Poor	580	23.9
wallsEnvEff	Average	180	7.4
	Good	1,150	47.5
	N/A	0	0.0
	Poor	370	15.3
	Very Good	120	5.0
	Very Poor	580	23.9
sheatingEnergyEff	N/A	2,420	99.9
sheatingEnvEff	N/A	2,420	99.9
roofEnergyEff	Average	430	17.7
	Good	1,000	41.3
	N/A	270	11.1
	Poor	110	4.5
	Very Good	270	11.1
	Very Poor	320	13.2

variable	value	number (rounded)	percent
roofEnvEff	Average	430	17.7
	Good	1,000	41.3
	N/A	270	11.1
	Poor	110	4.5
	Very Good	270	11.1
	Very Poor	320	13.2
mainheatEnergyEff	Average	200	8.3
	Good	1,940	80.1
	N/A	10	0.4
	Poor	60	2.5
	Very Good	120	5.0
	Very Poor	70	2.9
mainheatEnvEff	Average	110	4.5
	Good	1,980	81.7
	N/A	10	0.4
	Poor	50	2.1
	Very Good	150	6.2
	Very Poor	100	4.1
mainheatcEnergyEff	Average	770	31.8
	Good	1,280	52.8
	N/A	10	0.4
	Poor	130	5.4
	Very Good	70	2.9
	Very Poor	130	5.4
mainheatcEnvEff	Average	770	31.8
	Good	1,280	52.8
	N/A	10	0.4
	Poor	130	5.4
	Very Good	70	2.9
	Very Poor	130	5.4

variable	value	number (rounded)	percent
lightingEnergyEff		0	0.0
	Average	430	17.7
	Good	500	20.6
	N/A	0	0.0
	Poor	250	10.3
	Very Good	850	35.1
	Very Poor	370	15.3
lightingEnvEff	Average	430	17.7
	Good	500	20.6
	N/A	0	0.0
	Poor	250	10.3
	Very Good	850	35.1
	Very Poor	370	15.3
windTurbineCount		80	3.3
	-1	0	0.0
	0	2,320	95.7
	1	0	0.0
heatLossCorridoor	NO DATA!	2,020	83.4
	heated corridor	60	2.5
	no corridor	140	5.8
	unheated corridor	180	7.4
photoSupply		1,210	49.9
	0	1,190	49.1
	9	0	0.0
	20	0	0.0
	30	0	0.0
	35	0	0.0
	40	0	0.0
	45	0	0.0
	50	0	0.0

variable	value	number (rounded)	percent
solarWaterHeatingFlag		990	40.9
	N	1,410	58.2
	Y	10	0.4
mechanicalVentilation	NO DATA!	150	6.2
	mechanical, extract only	0	0.0
	mechanical, supply and extract	0	0.0
	natural	2,250	92.9
tenure		70	2.9
	NO DATA!	10	0.4
	owner-occupied	1,710	70.6
	rental (private)	260	10.7
	rental (social)	240	9.9
	unknown	100	4.1