

GREEN Grid Data Processing

Create Household Attributes File

Ben Anderson (b.anderson@soton.ac.uk, @dataknut)

Last run at: 2018-06-21 16:55:37

Contents

1	Citation	2
2	About	3
2.1	Circulation	3
2.2	History	3
2.3	Support	3
3	Introduction	4
4	Requirements:	4
5	Load data	4
5.1	Data description	4
6	Describe data	6
6.1	Number of adults	6
6.2	Number of teenagers	6
6.3	Number of children	6
6.4	Outlier flag	6
7	Summary	7
8	Runtime	7
9	R environment	7
	References	8

1 Citation

If you wish to use any of the material from this report please cite as:

- Anderson, B. (2018) GREEN Grid Data Processing: Create Household Attributes File, Centre for Sustainability, University of Otago: Dunedin.

This work is (c) 2018 the University of Southampton.

2 About

2.1 Circulation

Report circulation:

- Restricted to: NZ GREEN Grid project partners and contractors.

2.2 History

Generally tracked via our git.soton repo:

- history
- issues

Specific history of this code:

- <https://git.soton.ac.uk/ba1e12/nzGREENGrid/tree/master/analysis/ev>

2.3 Support

This work was supported by:

- The University of Otago;
- The University of Southampton;
- The New Zealand Ministry of Business, Innovation and Employment (MBIE) through the NZ GREEN Grid project;
- SPATIALEC - a Marie Skłodowska-Curie Global Fellowship based at the University of Otago's Centre for Sustainability (2017-2019) & the University of Southampton's Sustainable Energy Research Group (2019-2022).

We do not 'support' the code but if you have a problem check the issues on our repo and if it doesn't already exist, open one. We might be able to fix it :-)

3 Introduction

The prupose of this report is to:

- create a household attribute file that can be linked to the project power monitoring data

The resulting cleaned data has *no* identifying information such as names, addresses, email addresses, telephone numbers and is therefore safe to share across all partners.

The data contains a unique household id which can be used to link it to the NZ GREEN Grid time use diaries and dwelling/appliance surveys. With some additional non-disclosure checks it should also be safe to archive all of these linkable datasets for re-use via the UK reshare service.

4 Requirements:

- GREEN Grid household suveys and metadata files

5 Load data

In this section we load metadata from /Users/ben/Syncplicity Folders/Green Grid Project Management Folder/Gridspy/Master list of Gridspy units.xlsx.

```
## Location
##   Hawkes Bay New Plymouth      <NA>
##           21           24           0
```

In total we have 45 households in two sample areas.

5.1 Data description

```
## metaDT
##
##   9 Variables      45 Observations
## -----
## sample
##      n missing distinct
##      45      0        2
##
## Value      Powerco  Unison
## Frequency      24      21
## Proportion  0.533  0.467
## -----
## hhID
##      n missing distinct
##      45      0        44
##
## lowest : rf_06 rf_07 rf_08 rf_09 rf_10, highest: rf_43 rf_44 rf_45 rf_46 rf_47
## -----
## newID
##      n missing distinct
##      45      0        44
##
```

```

## lowest : rf_06 rf_07 rf_08 rf_09 rf_10, highest: rf_43 rf_44 rf_45 rf_46 rf_47
## -----
## Location
##      n missing distinct
##      45      0      2
##
## Value      Hawkes Bay New Plymouth
## Frequency      21      24
## Proportion      0.467      0.533
## -----
## nAdults
##      n missing distinct      Info      Mean      Gmd
##      43      2      3      0.544      1.907      0.412
##
## Value      1      2      3
## Frequency      7      33      3
## Proportion 0.163 0.767 0.070
## -----
## nChildren0_12
##      n missing distinct      Info      Mean      Gmd
##      42      3      4      0.858      0.881      1.107
##
## Value      0      1      2      3
## Frequency      21      9      8      4
## Proportion 0.500 0.214 0.190 0.095
## -----
## nTeenagers13_18
##      n missing distinct      Info      Mean      Gmd
##      42      3      3      0.42      0.2143      0.3775
##
## Value      0      1      2
## Frequency      35      5      2
## Proportion 0.833 0.119 0.048
## -----
## outlierFlag
##                                     n
##                                     1
##                                     missing
##                                     44
##                                     distinct
##                                     1
##                                     value
## Unusual & specialist energy tech configuration
##
## Value      Unusual & specialist energy tech configuration
## Frequency      1
## Proportion      1
## -----
## removed
##      n missing distinct
##      20      25      13
##
## Value      3/6/1015      42019      42089      42166      42171      42296      42322
## Frequency      1      1      1      1      1      1      1

```

```
## Proportion    0.05    0.05    0.05    0.05    0.05    0.05    0.05
##
## Value         42330    42457    42462    42532    42821    live
## Frequency      1        1        1        1        1        8
## Proportion     0.05    0.05    0.05    0.05    0.05    0.40
## -----
```

6 Describe data

NA usually means not known.

6.1 Number of adults

	Hawkes Bay	New Plymouth	NA
1	1	6	0
2	16	17	0
3	3	0	0
NA	1	1	0

6.2 Number of teenagers

	Hawkes Bay	New Plymouth	NA
0	17	18	0
1	2	3	0
2	2	0	0
NA	0	3	0

6.3 Number of children

	Hawkes Bay	New Plymouth	NA
0	11	10	0
1	2	7	0
2	5	3	0
3	3	1	0
NA	0	3	0

6.4 Outlier flag

These may have been set for any number of reasons and mean the monitoring data should be used with caution.

	Hawkes Bay	New Plymouth	NA
Unusual & specialist energy tech configuration	0	1	0
NA	21	23	0

7 Summary

The cleaned data for the 45 households has been saved as a .csv file to:

- /Volumes/hum-csafe/Research Projects/GREEN Grid/Clean_data/safe/survey/

The following table shows the first few rows of the household attributes file:

sample	hhID	newID	Location	nAdults	nChildren0_12	nTeenagers13_18	outlierFlag	removed
Unison	rf_28	rf_28	Hawkes Bay	2	3	0	NA	NA
Unison	rf_29	rf_29	Hawkes Bay	2	1	0	NA	live
Unison	rf_30	rf_30	Hawkes Bay	2	0	0	NA	NA
Unison	rf_31	rf_31	Hawkes Bay	3	0	0	NA	live
Unison	rf_32	rf_32	Hawkes Bay	2	2	0	NA	NA
Unison	rf_33	rf_33	Hawkes Bay	2	2	1	NA	live

The data can be linked to the gridSpy data using hhID and/or newID.

8 Runtime

Analysis completed in 2.11 seconds (0.04 minutes) using knitr in RStudio with R version 3.5.0 (2018-04-23) running on x86_64-apple-darwin15.6.0.

9 R environment

R packages used:

- base R - for the basics (R Core Team 2016)
- data.table - for fast (big) data handling (Dowle et al. 2015)
- readr - for csv reading/writing (Wickham, Hester, and Francois 2016)
- Hmisc - for describe (Harrell Jr, Charles Dupont, and others. 2016)
- knitr - to create this document & neat tables (Xie 2016)
- nzGREENGrid - for local NZ GREEN Grid project utilities

Session info:

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.5
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
```

```
## [1] en_GB.UTF-8/en_GB.UTF-8/en_GB.UTF-8/C/en_GB.UTF-8/en_GB.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] knitr_1.20      Hmisc_4.1-1      Formula_1.2-3
## [4] survival_2.42-3 lattice_0.20-35   readr_1.1.1
## [7] lubridate_1.7.4 ggplot2_2.2.1     dplyr_0.7.5
## [10] data.table_1.11.4 nzGREENGrid_0.1.0
##
## loaded via a namespace (and not attached):
## [1] progress_1.2.0      tidyselect_0.2.4    xfun_0.1
## [4] reshape2_1.4.3      purrr_0.2.5         splines_3.5.0
## [7] colorspace_1.3-2    htmltools_0.3.6     yaml_2.1.19
## [10] base64enc_0.1-3     rlang_0.2.1         pillar_1.2.3
## [13] foreign_0.8-70      glue_1.2.0          RColorBrewer_1.1-2
## [16] readxl_1.1.0        bindrcpp_0.2.2      bindr_0.1.1
## [19] plyr_1.8.4          stringr_1.3.1       cellranger_1.1.0
## [22] munsell_0.5.0       gtable_0.2.0        htmlwidgets_1.2
## [25] evaluate_0.10.1     latticeExtra_0.6-28 highr_0.7
## [28] htmlTable_1.12      Rcpp_0.12.17        acepack_1.4.1
## [31] checkmate_1.8.5     backports_1.1.2     scales_0.5.0
## [34] gridExtra_2.3       hms_0.4.2           digest_0.6.15
## [37] stringi_1.2.3       bookdown_0.7        grid_3.5.0
## [40] rprojroot_1.3-2     tools_3.5.0         magrittr_1.5
## [43] lazyeval_0.2.1      tibble_1.4.2        cluster_2.0.7-1
## [46] crayon_1.3.4        pkgconfig_2.0.1     Matrix_1.2-14
## [49] prettyunits_1.0.2   assertthat_0.2.0    rmarkdown_1.10
## [52] rstudioapi_0.7      R6_2.2.2            rpart_4.1-13
## [55] nnet_7.3-12         compiler_3.5.0
```

References

- Dowle, M, A Srinivasan, T Short, S Lianoglou with contributions from R Saporita, and E Antonyan. 2015. *Data.table: Extension of Data.frame*. <https://CRAN.R-project.org/package=data.table>.
- Harrell Jr, Frank E, with contributions from Charles Dupont, and many others. 2016. *Hmisc: Harrell Miscellaneous*. <https://CRAN.R-project.org/package=Hmisc>.
- R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Wickham, Hadley, Jim Hester, and Romain Francois. 2016. *Readr: Read Tabular Data*. <https://CRAN.R-project.org/package=readr>.
- Xie, Yihui. 2016. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://CRAN.R-project.org/package=knitr>.