GREEN Grid Data Processing:

Household Attributes File

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

Last run at: 2018-07-27 15:23:26

About

Report circulation:

- Public - this report is intended to accompany the data release.

License

Notices:

Citation

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

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

History

Specific history of this code:

- https://github.com/dataknut/nzGREENGridDataR/commits/master/dataProcessing/surveys/createHouseholdAttributes.Rmd

Requirements:

This report uses the original nz GREEn grid household survey and meta data.

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 (2019-202).

Introduction

The prupose of this report is to:

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

Load data

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

Location	Freq
Hawkes Bay	20
New Plymouth	24
NA	0

In total we have 44 households in two sample areas.

Data description

```
## hhAttributesDT
##
##  9  Variables      44  Observations
## ---------------------------------------------------------------------------
## sample
##        n  missing distinct
##       44        0        2
##
## Value      Powerco  Unison
## Frequency       24      20
## Proportion   0.545   0.455
## ---------------------------------------------------------------------------
## hhID
##        n  missing distinct
##       44        0       42
##
## 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
##       44        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
##       44        0        2
##
## Value        Hawkes Bay New Plymouth
## Frequency            20           24
## Proportion        0.455        0.545
## ------------------------------------------------------------------------
## nAdults
##        n  missing distinct     Info     Mean      Gmd
##       43        1        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        2        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        2        3     0.42   0.2143   0.3775
##
## Value          0     1     2
## Frequency     35     5     2
## Proportion 0.833 0.119 0.048
## ------------------------------------------------------------------------
## notes
##        n  missing distinct
##        4       40        4
##
## Disconnected 15/01/2015. Re-used (1, 0.25), Re-user (1, 0.25), Re-user.
## Then disconnected 02/04/2016 (1, 0.25), Unusual & specialist energy tech
## configuration. Disconnected 28/03/2016. Re-used. (1, 0.25)
## ------------------------------------------------------------------------
## r_stopDate
##        n  missing distinct
##        3       41        3
##
## Value      2015-01-15 2016-03-28 2016-04-02
## Frequency           1          1          1
## Proportion      0.333      0.333      0.333
## ------------------------------------------------------------------------
```

Describe data

NA usually means not known.

Number of adults

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

Number of teenagers

	Hawkes Bay	New Plymouth	NA
0	16	19	0
1	2	3	0
2	2	0	0
NA	0	2	0

Number of children

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

Notes

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
Disconnected 15/01/2015. Re-used	0	1	0
Re-user	0	1	0
Re-user. Then disconnected 02/04/2016	0	1	0
Unusual & specialist energy tech configuration. Disconnected 28/03/2016. Re-used.	0	1	0
NA	20	20	0

Summary

The cleaned data for the 44 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 key columns of the household attributes file. The data can be linked to the gridSpy data using hhID.

The purpose of the newID is to enable the flagging of re-used grid spy units. As an example unit rf_15 was re-used in a different household. Data users should therefore make sure that the correct household data (rf_15a or rf_15b) is linked to the grid spy data (coded rf_15) at the correct date.

Note also that data exists in the grid spy power demand data for whom no household data exists (e.g. rf_01 & rf_02).

hhID	Location	nAdults	newID	r_stopDate	notes
rf_06	New Plymouth	2	rf_06	NA	NA
rf_07	New Plymouth	2	rf_07	NA	NA
rf_08	New Plymouth	2	rf_08	NA	NA
rf_09	New Plymouth	2	rf_09	NA	NA
rf_10	New Plymouth	2	rf_10	NA	NA
rf_11	New Plymouth	2	rf_11	NA	NA
rf_12	New Plymouth	1	rf_12	NA	NA
rf_13	New Plymouth	2	rf_13	NA	NA
rf_14	New Plymouth	1	rf_14	NA	NA
rf_15	New Plymouth	1	rf_15a	2015-01-15	Disconnected 15/01/2015. Re-used
rf_15	New Plymouth	2	rf_15b	2016-04-02	Re-user. Then disconnected 02/04/2016
rf_16	New Plymouth	2	rf_16	NA	NA
rf_17	New Plymouth	2	rf_17a	2016-03-28	Unusual & specialist energy tech configuration. Disconnected 28/03/2016. Re-used.
rf_17	New Plymouth	NA	rf_17b	NA	Re-user
rf_18	New Plymouth	2	rf_18	NA	NA
rf_19	New Plymouth	1	rf_19	NA	NA
rf_20	New Plymouth	2	rf_20	NA	NA
rf_21	New Plymouth	2	rf_21	NA	NA
rf_22	New Plymouth	2	rf_22	NA	NA

hhID	Location	nAdults	newID	r_stopDate	notes
rf_23	New Plymouth	1	rf_23	NA	NA
rf_24	New Plymouth	2	rf_24	NA	NA
rf_25	New Plymouth	1	rf_25	NA	NA
rf_26	New Plymouth	2	rf_26	NA	NA
rf_27	New Plymouth	2	rf_27	NA	NA
rf_28	Hawkes Bay	2	rf_28	NA	NA
rf_29	Hawkes Bay	2	rf_29	NA	NA
rf_30	Hawkes Bay	2	rf_30	NA	NA
rf_31	Hawkes Bay	3	rf_31	NA	NA
rf_32	Hawkes Bay	2	rf_32	NA	NA
rf_33	Hawkes Bay	2	rf_33	NA	NA
rf_34	Hawkes Bay	3	rf_34	NA	NA
rf_35	Hawkes Bay	2	rf_35	NA	NA
rf_36	Hawkes Bay	1	rf_36	NA	NA
rf_37	Hawkes Bay	2	rf_37	NA	NA
rf_38	Hawkes Bay	2	rf_38	NA	NA
rf_39	Hawkes Bay	2	rf_39	NA	NA
rf_40	Hawkes Bay	2	rf_40	NA	NA
rf_41	Hawkes Bay	2	rf_41	NA	NA
rf_42	Hawkes Bay	2	rf_42	NA	NA
rf_43	Hawkes Bay	2	rf_43	NA	NA
rf_44	Hawkes Bay	2	rf_44	NA	NA
rf_45	Hawkes Bay	2	rf_45	NA	NA
rf_46	Hawkes Bay	2	rf_46	NA	NA
rf_47	Hawkes Bay	3	rf_47	NA	NA

Runtime

Analysis completed in 0.64 seconds (0.01 minutes) using knitr in RStudio with R version 3.5.0 (2018-04-23) running on x86_64-apple-darwin15.6.0.

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)
- readxl - reading .xlsx (Wickham and Bryan 2017)
- nzGREENGridDataR - 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 10.13.6
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions
## 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] Hmisc_4.1-1            Formula_1.2-3          survival_2.42-3
##  [4] lattice_0.20-35        dplyr_0.7.5            nzGREENGrid_0.1.0
##  [7] hms_0.4.2              lubridate_1.7.4        data.table_1.11.4
## [10] kableExtra_0.9.0       stringr_1.3.1          knitr_1.20
## [13] readr_1.1.1            ggplot2_2.2.1          rmarkdown_1.10
## [16] nzGREENGridDataR_0.1.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.17        prettyunits_1.0.2   png_0.1-7
##  [4] assertthat_0.2.0    rprojroot_1.3-2     digest_0.6.15
##  [7] R6_2.2.2            cellranger_1.1.0    plyr_1.8.4
## [10] backports_1.1.2     acepack_1.4.1       evaluate_0.10.1
## [13] highr_0.7           httr_1.3.1          pillar_1.2.3
## [16] RgoogleMaps_1.4.2   rlang_0.2.1         progress_1.2.0
## [19] lazyeval_0.2.1      readxl_1.1.0        rstudioapi_0.7
## [22] geosphere_1.5-7     rpart_4.1-13        Matrix_1.2-14
## [25] checkmate_1.8.5     proto_1.0.0         splines_3.5.0
## [28] foreign_0.8-70      htmlwidgets_1.2     munsell_0.5.0
## [31] compiler_3.5.0      pkgconfig_2.0.1     base64enc_0.1-3
## [34] htmltools_0.3.6     nnet_7.3-12         openssl_1.0.1
```

```
## [37] tidyselect_0.2.4    htmlTable_1.12      gridExtra_2.3
## [40] tibble_1.4.2        viridisLite_0.3.0   crayon_1.3.4
## [43] grid_3.5.0          gtable_0.2.0        magrittr_1.5
## [46] scales_0.5.0        stringi_1.2.3       mapproj_1.2.6
## [49] reshape2_1.4.3      bindrcpp_0.2.2      sp_1.3-1
## [52] latticeExtra_0.6-28 xml2_1.2.0          rjson_0.2.20
## [55] RColorBrewer_1.1-2  tools_3.5.0         ggmap_2.6.1
## [58] glue_1.2.0          purrr_0.2.5         maps_3.3.0
## [61] jpeg_0.1-8          yaml_2.1.19         colorspace_1.3-2
## [64] cluster_2.0.7-1     rvest_0.3.2         bindr_0.1.1
```

References

Dowle, M, A Srinivasan, T Short, S Lianoglou with contributions from R Saporta, 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, and Jennifer Bryan. 2017. *Readxl: Read Excel Files.* https://CRAN.R-project.org/package=readxl.

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.