WHICH AUSTRALIAN HOUSEHOLDS SPEND THE MOST ON ELECTRICITY

COMPUTER SCIENCE RESEARCH PROJECT

COMP7801

DONE BY

MR SAISRIKAR PARUCHURI

44751575

UNDER THE GUIDANCE OF

DR. MARK PHILIP GRIFFIN

THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Introduction

- This project is to find the Australian households that spend the most on Electricity usage.
- This project uses some variables from the HILDA dataset to explore the variation of the electricity usage between households.
- To validate the chosen variables this project uses regression analysis on a univariate basis on how likely the variable is related to the electricity usage.
- This project applies statistical analysis like Factor analysis, Linkage Algorithm and K-means clustering on variables related to the electricity usage.

Literature Review

Electricity Usage:

- The research indicates that the adolescents in some households led to higher electricity usage because adolescents frequently use IT appliances that were positively correlated with high household electricity usage [1].
- The researchers suggest that the use of behavioral information provides more detailed information of the condition of electricity usage [1].
- They have taken sample households monitored over a year at an interval of 5 minutes during the common peak time of electricity usage[2].

1. E. Matthies, 2016
2. I. Dent, 2014

Household/Family Structure

- Small households of 2 or 3 people will typically use less energy than a larger household of 4 or 5 people[3]
- In order to calculate the Sydney households electricity consumption Lenzen et al. [4] has considered household size, age, income, degree of urbanity, and lifestyle.

Number of residents	Average Annual Electricity Bill
1	$1,426.08
2	$1,722.96
3	$1,993.84
4	$2,028.72
5	$2,576.28
6+	$2,634.36

[3] Blue 2018

3. C. BLUE, 2018
4. M. Lenzen et al. 2004

Employment

For instance, a study revealed that employment conditions, age and gender influenced the amount of time they spent at home (Brasche et al. 2005).

5. Sabine Brasche et al. 2005

Location

- Queensland electricity bills are higher compared to New South Wales electricity bills[3]
- Annual costs for cheapest electricity plans for 2 person household on selected postcode in each state is given below[3]

State	Cheapest Electricity Bill
NSW	$1,571.91
Victoria	$1,317.38
Queensland	$1369.27
SA	$2203.09

[3] C. Blue

3. C. BLUE, 2018

HILDA Dataset

- HILDA->Household Income Labour and Dynamics in Australia
- Publicly available de-identified data.
- HILDA[6] survey has followed approximately 17,000 people since 2001 and it is nationally representative.
- This project uses HILDA Stata data released in 2015.
- Some variables are chosen based on electricity usage from HILDA to estimate the households that spend the most on electricity usage and explore the variation of electricity between households.

ID	Pseudo Name	Gender	Year of Death	Test Result
1	DE41B4C0-EBC2	Male	1950-1959	+ve
2	F83246A6-EBC2	Male	1960-1969	-ve
4	0B56335A-EBC3	Male	1950-1959	-ve
6	0272584A-EBC3	Female	1970-1979	-ve
7	13525340-EBC3	Female	1960-1969	+ve
9	2BC03654-EBC3	Male	1970-1979	-ve
10	3812895C-EBC3	Male	1970-1979	-ve
11	4ABCDEFE-EBC3	Female	1960-1969	-ve
12	52C8B96A-EBC3	Male	1950-1959	+ve
13	59191DD2-EBC3	Female	1970-1979	-ve
14	61A29D52-EBC3	Male	1960-1969	-ve

6. N. Watson, 2012

RStudio

- This project uses RStudio for statistical computing (Scatterplots, Boxplots, Linear Regression etc) using the HILDA variables.
- RStudio [7] is an integrated development environment for the R programming language.

7. RStudio Team, 2016

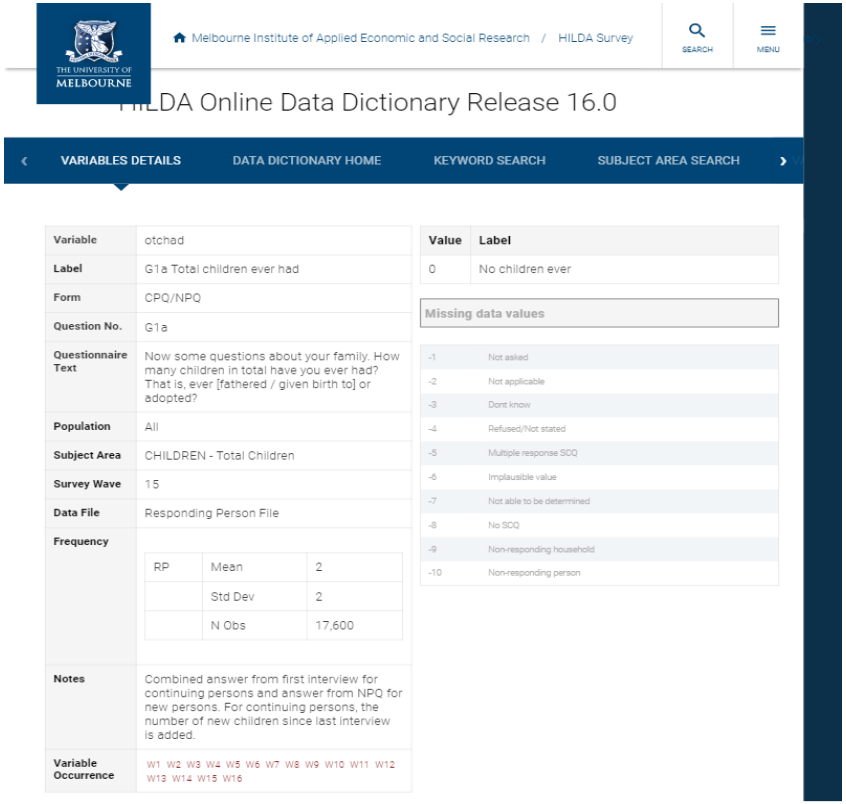
Variables

This project uses all the below variables to explore variation of electricity between households.

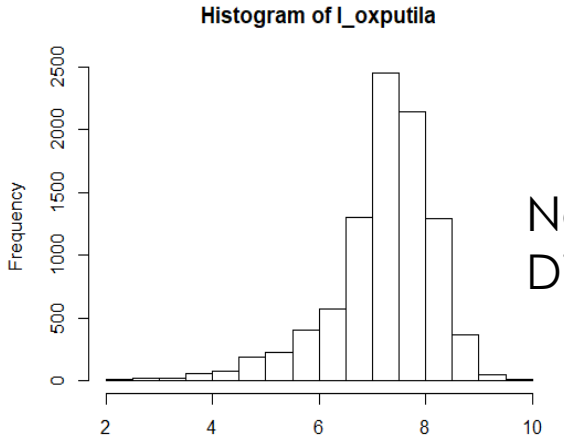
- Financial wages and salaries
- Home Repairs
- Physical Functioning
- Physical Activity
- Outdoor Tasks
- Unpaid work
- Annual household expenditure - electricity bills, gas bills and other heating fuel

HILDA Online Data Dictionary

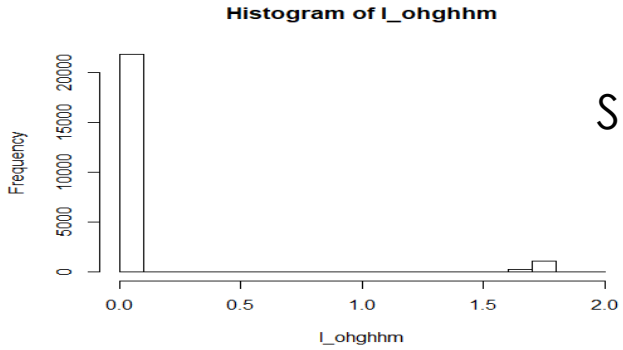
- Online data dictionary is helpful to find the variable names by their respective question number and subject area, and shows where the variable is stored in the Stata data.



Histogram of Variables



Normal Distribution



Skewed

- Generally, some variables have missing values in the HILDA data, and this might bias the regression analysis.
- Histograms provide a validation of the data distribution.
- Before regression analysis check the variable histogram

Evaluation of Variables

- To validate the chosen variables this project uses Linear regression analysis on a univariate basis on how likely the variable is relevant to the electricity usage.
- All the variables are compared with oxputila (Annual household expenditure) for validation

Example: Assume l_olshrcom ~l_oxputila

```
> summary(lm(l_oxputila~l_olshrcom))

Call:
lm(formula = l_oxputila ~ l_olshrcom)

Residuals:
    Min      1Q  Median      3Q     Max
-5.0103 -0.3736  0.1253  0.5534  2.3863

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.25250    0.02483 292.062   <2e-16 ***
l_olshrcom   0.02622    0.01609   1.629    0.103
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9049 on 4811 degrees of freedom
  (12793 observations deleted due to missingness)
Multiple R-squared:  0.0005514, Adjusted R-squared:  0.0003437
F-statistic: 2.654 on 1 and 4811 DF,  p-value: 0.1033

>
```

Analysis from RStudio using HILDA Data

Linear Regression

Variable	Estimate	P value	Continuous /Categorical
Financial year wages and salaries	0.118	<0.001	Continuous
Home Repairs	-0.301	<0.001	Categorical
Physical Functioning	-0.742	0.003	Continuous
Physical Activity	-0.081	0.026	Continuous
Outdoor Tasks	0.046	<0.001	Continuous
Unpaid work	0.025	<0.001	Continuous

Analysis from HILDA Data

Advanced Statistical Methods

- This project applies advanced statistical analysis that is factor analysis, k-means clustering and linkage algorithm to the selected variables from the multi-variate linear regression analysis.

Factor Analysis

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called **factors**.

```
Loadings:
                    MR1    MR2    MR3    MR4    MR5    MR6
Income               0.457 -0.419 -0.164
Paying for housework        0.238 -0.147  0.386
Bathing or Dressing -0.249 -0.307  0.219  0.153  0.164
Physical Activity    0.248  0.366 -0.145 -0.143  0.170
Home Maintenance     0.601         0.136  0.192
Unpaid work          0.189  0.203  0.446         -0.114

                 MR1   MR2   MR3   MR4   MR5   MR6
SS loadings    0.730 0.503 0.335 0.240 0.092 0.000
Proportion Var 0.122 0.084 0.056 0.040 0.015 0.000
Cumulative Var 0.122 0.206 0.261 0.301 0.317 0.317
```

Result of Factor Analysis

K-Means Clustering

k-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Linkage Algorithm: The linkage criterion determines the distance between sets of observations as a function of the pairwise distances between observations. Some commonly used linkages in this project are:

Single Linkage:

Single linkage defines the distance between 2 clusters as the minimum distance found between one case from the first cluster and one case from the second cluster.

Complete Linkage:

Complete linkage will search for the maximum distance between the cases from one cluster to another. This linkage will solve the chaining problem, but it creates another problem. If all the cases are near in 2 clusters and only one case is far from the other cases then it will no longer merge 2 clusters.

Average Linkage

To overcome the problems of single and complete linkage. The average linkage will consider the average of the distance between the cases and then it will decide whether to merge the clusters or not. It will provide more accurate results.

Discussion

Variable	Estimate
Financial year wages and salaries	0.118
Home Repairs	-0.301
Physical Functioning	-0.742
Physical Activity	-0.081
Outdoor Tasks	0.046
Unpaid work	0.025

1. E. Matthies et al. 2016
8. P. K. Narayan et al. 2005
9. Daftlogic.com, 2018

- Income or employment reducing electricity consumption could lead to a fall in income and/or employment [8].
- So, increase in financial year wages and salaries increases electricity consumption. Generally, unpaid work like volunteering, social activities etc are more likely done by teenagers because their need some experience to find a job. Several investigations have reported that when the number of adolescents in a household increases, residential electricity consumption increases as well [1].
- Outdoor workers in a household get higher electricity bills. For instance, because of lawn moving work they must regularly charge the battery of lawnmower depends on the number of houses they do lawn moving, which consumes more electricity usage. Lawnmower consumes minimum 1000W (which is equal to home air conditioner) and maximum 1400W [9].

Conclusion

- The linear regression analysis show that the list of selected variables show the variation of the electricity usage. The above analysis suggest that household members who earn high income, spending time in outdoor tasks and unpaid work will get higher electricity bills.
- Where as, people spending their time in home repairs, physical functioning and physical activity will get less electricity bills.

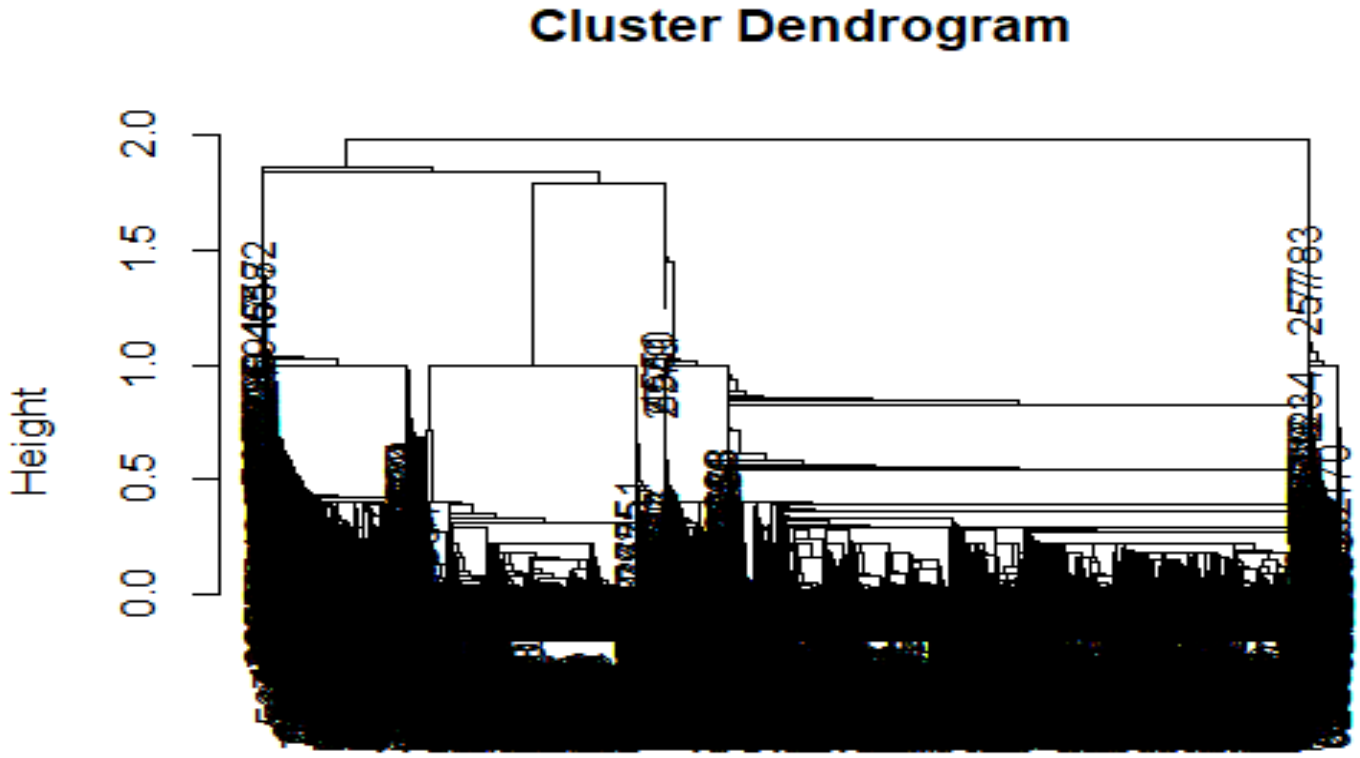
References

1. E. Matthies, M. Nachreiner, and H. Wallis, "Adolescents and electricity consumption; Investigating sociodemographic, economic, and behavioral influences on electricity consumption in households. (Special Section on the European Union: Markets and Regulators)," *Energy Policy*, vol. 94, pp. 224-234, 2016.
2. I. Dent, T. Craig, U. Aickelin, and T. Rodden, "An Approach for Assessing Clustering of Households by Electricity Usage," 2014.
3. C. BLUE. (2018, January 15). What is the average electricity bill. Available: https://www.canstarblue.com.au/energy/electricity/average-electricity-bills/
4. M. Lenzen, C. Dey, and B. Foran, "Energy requirements of Sydney households," Ecological Economics, vol. 49, no. 3, pp. 375-399, 2004/07/01/ 2004.
5. S. Brasche and W. Bischof, "Daily time spent indoors in German homes – Baseline data for the assessment of indoor exposure of German occupants," *International Journal of Hygiene and Environmental Health*, vol. 208, no. 4, pp. 247-253, 2005/07/20/ 2005.
6. N. Watson and M. P. Wooden, "The HILDA survey: a case study in the design and development of a successful household panel survey," Longitudinal and Life Course Studies, vol. 3, no. 3, pp. 369-381, 2012.
7. RStudio Team (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.
8. P. K. Narayan and R. Smyth, "Electricity consumption, employment and real income in Australia evidence from multivariate Granger causality tests," *Energy policy*, vol. 33, no. 9, pp. 1109-1116, 2005.
9. Daftlogic.com. (2018). *Power Consumption of Typical Household Appliances*. [online] Available at: https://www.daftlogic.com/information-appliance-power-consumption.htm [Accessed 11 Oct. 2018].

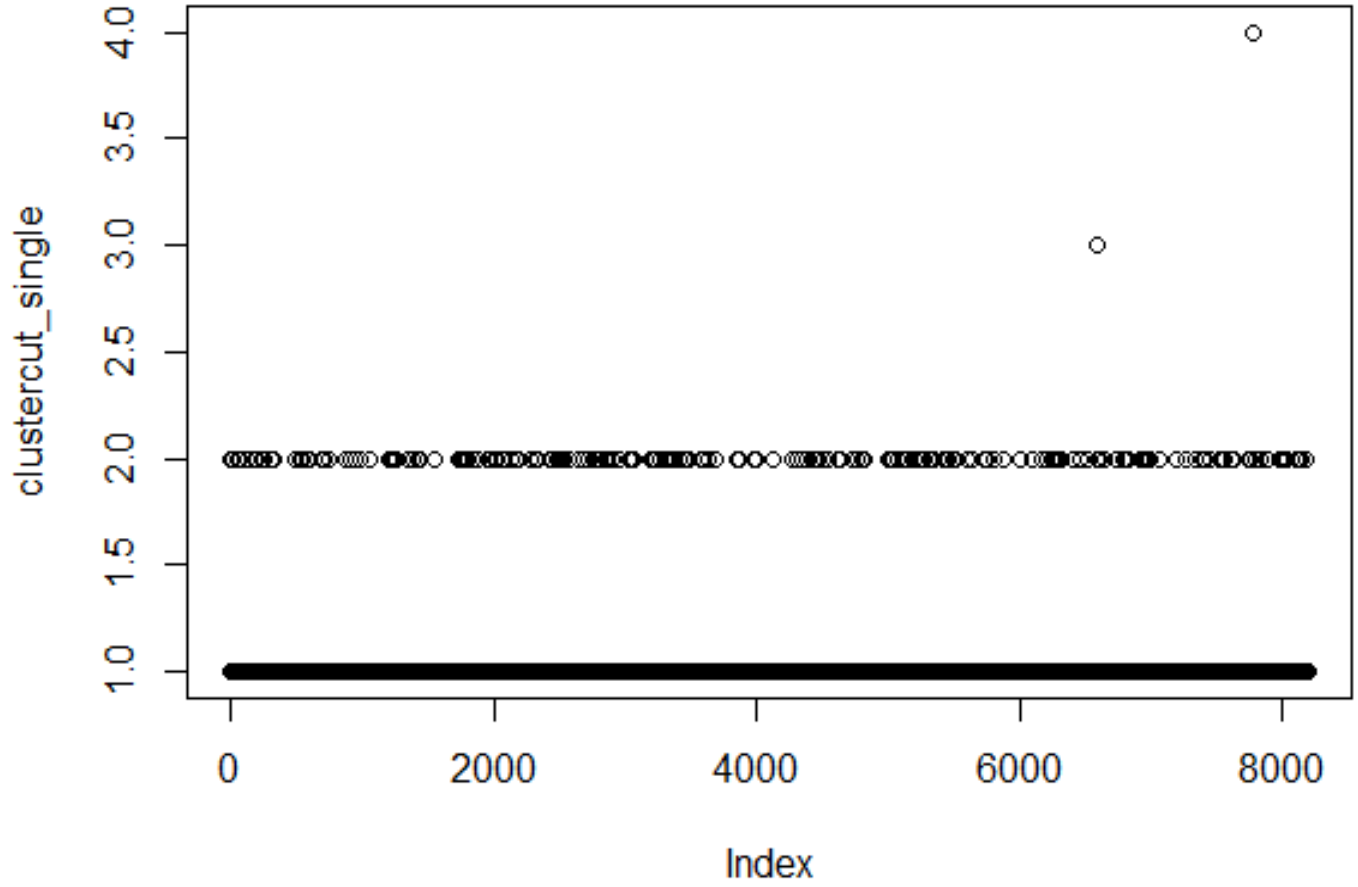
Questions

Appendix

Dendrogram of Single Linkage



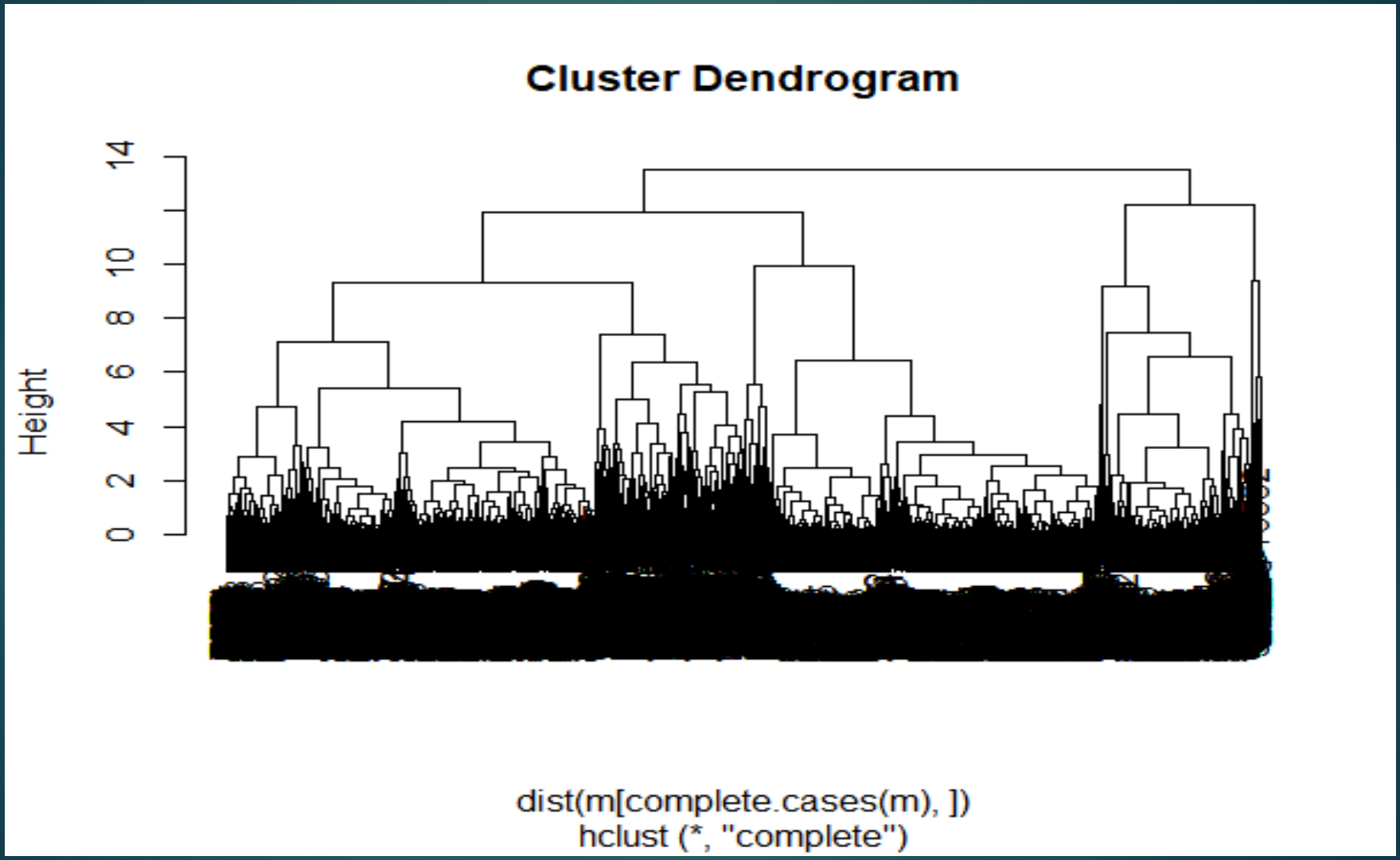
Single Linkage Cluster Cut



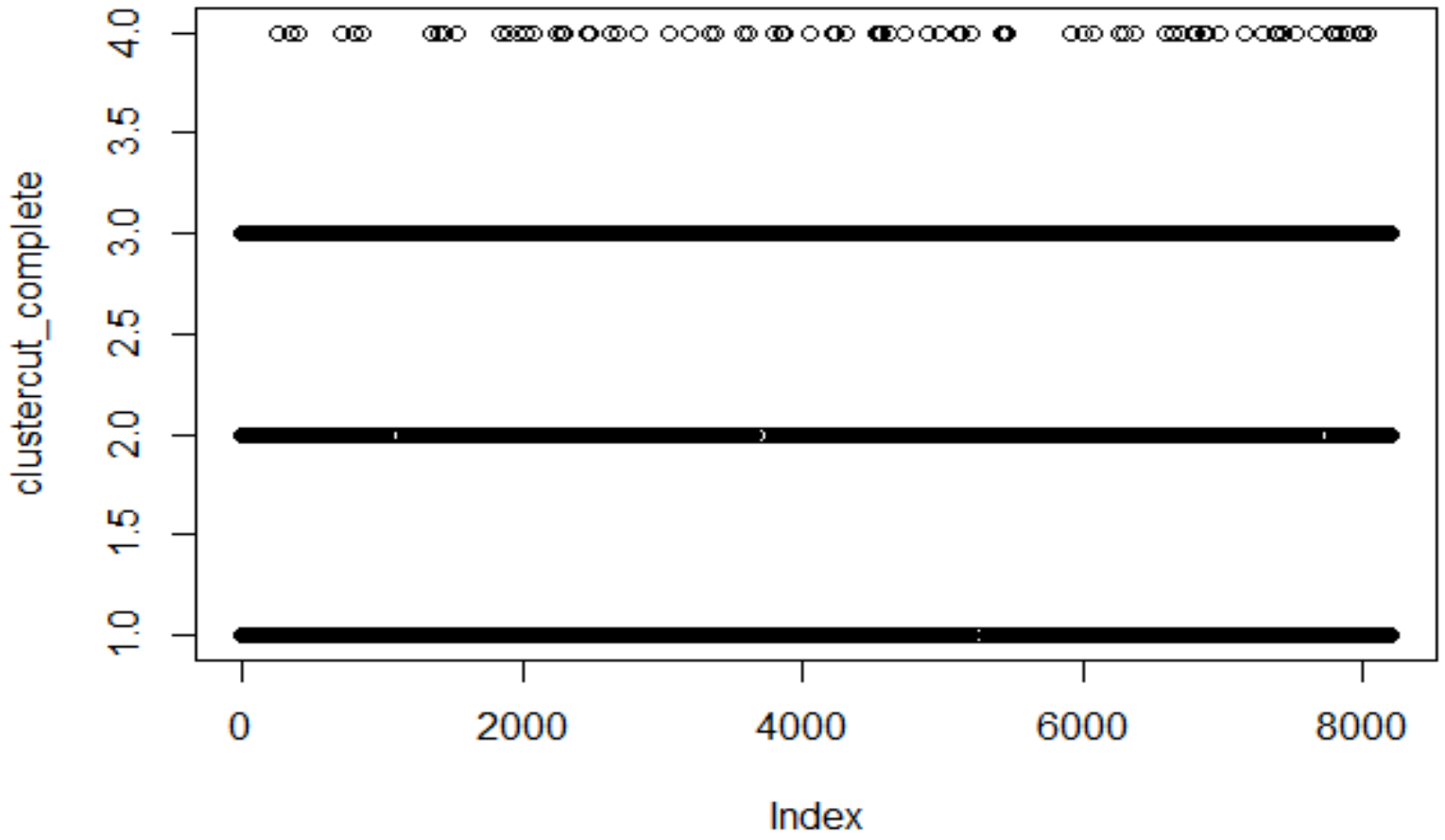
Mean of Single Linkage Cluster Cut

	1	2	3	4
Income	10.55264	10.54723	10.67631	10.89674
Bathing or dressing yourself	1.409268	1.410237	1.386294	1.386294
Physical activity	1.442345	1.43239	1.386294	1.522261
Outdoor tasks	3.865793	3.832714	5.484797	2.853555
Unpaid work	1.065767	1.067311	0	5.341647

Dendrogram of Complete Linkage



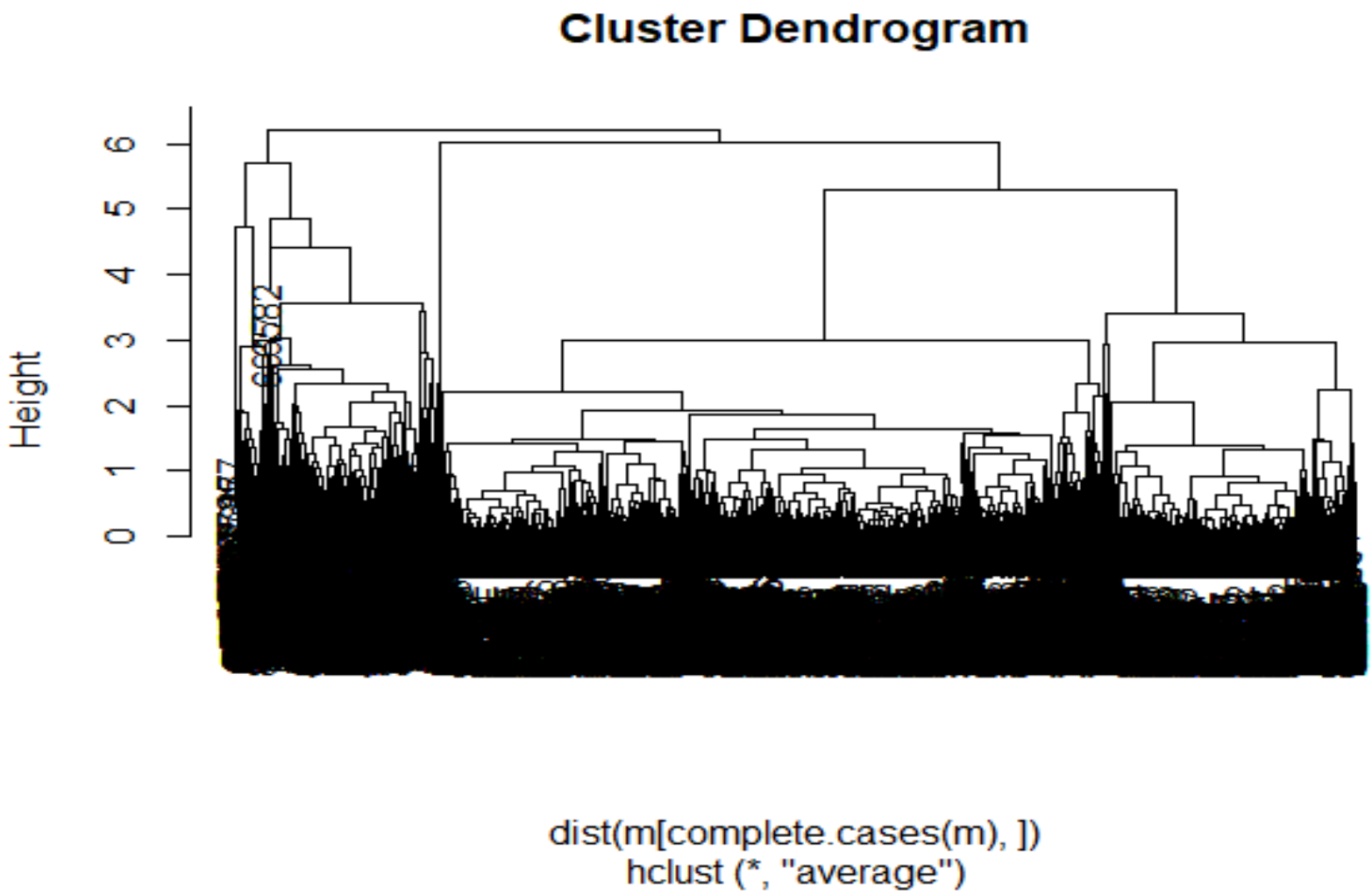
Complete Linkage Cluster Cut



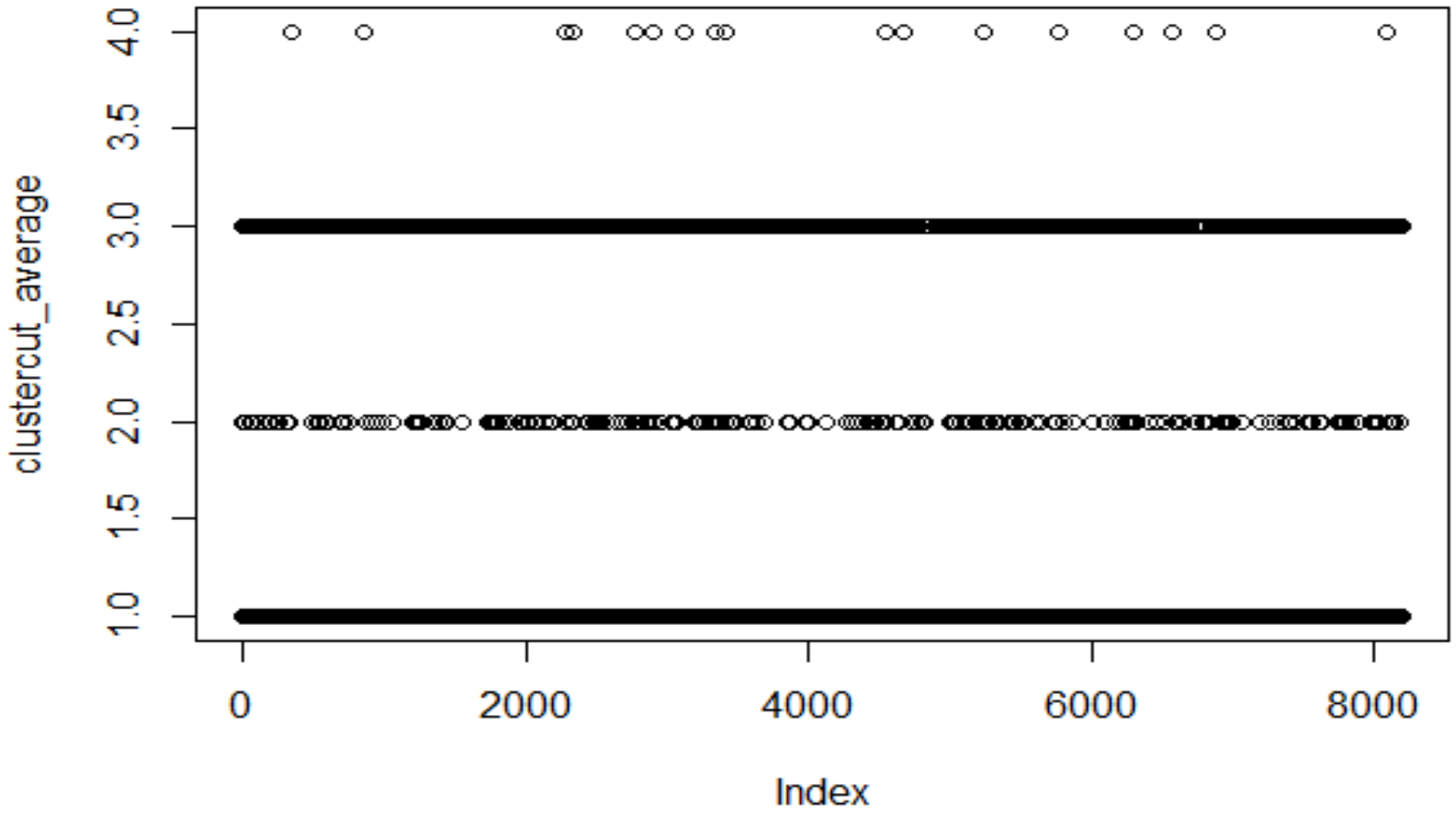
Mean of Complete Linkage Cluster Cut

	1	2	3	4
Income	10.58141	10.5027	10.54669	10.50935
Bathing or dressing yourself	1.409499	1.40847	1.409344	1.411716
Physical activity	1.446056	1.43331	1.442919	1.389232
Outdoor tasks	3.851015	3.849842	3.876243	3.986299
Unpaid work	1.03539	1.015707	1.094764	1.44589

Dendrogram of Average Linkage



Average Linkage Cluster Cut



Mean of Average Linkage Cluster Cut

	1	2	3	4
Income	10.54737	10.54835	10.58383	10.49556
Bathing or dressing yourself	1.409139	1.410134	1.409982	1.407889
Physical activity	1.44298	1.432063	1.439123	1.436675
Outdoor tasks	3.846073	3.832045	3.970951	3.974043
Unpaid work	1.048939	1.084276	1.156093	1.084059

Thank You