Understanding time use via data mining: A clustering-based framework

Article *in* Intelligent Data Analysis · May 2018

DOI: 10.3233/IDA-173708

Jorge Rosales-Salas
Universidad Mayor

Sebastián Maldonado
University of Chile

Alex Seret
KU Leuven

Intelligent Data Analysis 22 (2018) 597–616
DOI 10.3233/IDA-173708
IOS Press

Understanding time use via data mining: A clustering-based framework

Jorge Rosales-Salas[a], Sebastián Maldonado[b,*] and Alex Seret[c]
[a]*Centro de Investigación en Sociedad Tecnológica y Futuro Humano, Facultad de Humanidades, Universidad Mayor, Av. Portugal N°351, Santiago, Chile*
[b]*Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes. Monseñor Álvaro del Portillo 12455, Las Condes, Santiago, Chile*
[c]*Generation Research. Lasarettsgatan 13, 89133 Örnsköldsvik, Sweden*

Abstract. In this work, a data mining framework is proposed to improve the understanding of how people allocate their time. Using a multivariate approach, we performed a clustering procedure, and subsequently a regression analysis to detect which variables influence individual time use for each cluster found. Results suggest that the impact of various sociodemographic variables on sleep and work depends significantly on the characteristics of the individuals analyzed. This suggests that inquiries into time allocation and individual behavior should no longer be limited to discussions focused only on single variables. Based on our results, we recommend that researchers advance their methodological analysis towards a multifactorial approach and include clustering as a fundamental step. Proper identification of the most significant variables involved in time allocation decisions would allow researchers to better analyze and interpret their data and results.

Keywords: Time use, sleep, work, clustering, data mining

1. Introduction

Multivariate data analysis based on predictive analytics has been very useful for decision-making in social sciences, such as sociology [6], education [54], and economy [16]. The success of these techniques relies on their ability to discover complex patterns in large datasets, which cannot be unveiled with the traditional statistical methods.

Within human behavior research, it can be argued that the study of time allocation is an essential aspect: given an overall temporal limitation – everybody has 24 hours per day – time use decisions reveal the relative value individuals assign to every activity performed.

There has been a common trend within the studies that focus primarily on statistical analyzes when examining the role of time in individual behavior: while the methodology presented, and the results obtained are multivariate, studies tend to conduct their discussions based on analyzing each independent variable by itself. These studies aim at answering questions such as: Is time allocation influenced by gender [47,48]? Is age a relevant factor when analyzing time use [19,49,65]? Is time allocation different

*Corresponding author: Sebastián Maldonado, Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes. Monseñor Álvaro del Portillo 12455, Las Condes, Santiago, Chile. Tel./Fax: +56 226181874; Email: smaldonado@uandes.cl.

for married couples [35,46,66]? Does activity duration vary depending on race/ethnicity [1,37]? Does the presence of children in the household impact parental time [7,20,31]?

The main purpose of this paper is to present a way for time use research to advance towards a multidimensional approach. This study selected one of the many techniques proposed by data mining that could reach this objective, namely, a clustering procedure aimed at selecting multifactorial sociodemographic groups and analyzing the variables that impact their time use behavior. To the best of our knowledge, no such study has been reported yet in the time use literature. This study will mainly focus on two activities, chosen by their length and daily relevance: *sleep* and *paid work*.

This study aims at providing an adequate procedure for expanding the multifactorial perspective from just the methodological framework to include analytical, descriptive and interpretative stages. In the next section, the literature on sleep and paid work is briefly described. The proposed clustering framework is covered in Section 3. Data, methods and results are presented and examined in Section 4. Finally, a summary of the principal findings and future developments is presented in Section 5.

2. Literature research

Descriptive and analytical contributions within time use literature can be found among different disciplines, analyzing various activities, and using diverse methods. The standard modelling strategy is the use of generalized linear methods for multivariate analysis (linear or logistic regression depending on the task) [6,55,67], and well-known statistical tests for descriptive analysis, including respondent segmentation based on one or two variables at time. For example, the computation of the mean time allocated to work or sleep for various groups defined by variables such as gender or age [18,43,58].

We believe that there are ways of moving forward, which should lead to a better understanding of individual time allocation. More specifically, we postulate that there are methodological and empirical gains if new elements are incorporated in the analysis, i.e. a clustering procedure. By clustering individuals by some of the most frequently used variables in time use research (e.g. age, gender, income, and number of children), the study of sleep and paid work time allocation could focus on other potentially relevant factors within a larger group of sociodemographic variables.

Let us briefly explore the relationships between independent variables and time allocated to sleep and paid work.

2.1. Sleep

For centuries, sleep has been considered to be a passive state, a simple suspension of activities that individuals must undertake on a daily basis to give them the necessary physical and mental conditions to keep performing other activities when awake. Even though sleep takes more time, on average, than any other single activity (individuals spend roughly a third of their lives asleep), it is generally viewed only as a mandatory activity from which to take time if needed, i.e. a trade-off activity. However, the true and active nature of sleep not only is fundamental to well-being, health, and productivity, but it also impacts significantly on how individuals allocate time to other important activities, such as work and leisure.

There are two main reasons, we believe, for the lack of sufficient appreciation of sleep as a driver of time assignment. First, there is the generalized public opinion that sleeping is “invariant” over time and between groups, due to the common assumption that, in general, people should sleep around 8 hours [52]. This contrasts with most research that states sleep required by individuals varies with age, lifestyle, and health factors [40]. Second, sleep duration analysis has been limited by poor measurements, such as inaccuracies in data collection methods, the differences between sleep time and time in bed, and the lack of acknowledgment of sleep interruptions during the night [8]. These factors, among others, might explain why the availability of time for activities with active participation (being awake) is usually considered to be a fixed quantity.

When studying the relationship between sociodemographic variables and sleep, gender and age differences have been a prolific area of research. In gender analysis, Burgard and Ailshire [58] assessed the explanations for gender differences in time allocated to sleep, analyzing its trade-off with paid and unpaid labor, and its relationship with the family life-course, analyzing whether women sleep more than men. Their results show that women slept more than men during most life course stages, a gap partially explained by work and family responsibilities. These results are complemented by evidence suggesting that women get less high-quality, uninterrupted sleep because they prioritize family needs, even during sleeping hours [24,54,56,57], and that women were less likely to be over sleepers than men [59]. Furthermore, gender differences have also been studied from different perspectives and within various areas of research, such as biomedical studies [5,15], actigraphy [10], time diary data [30,34,36,50], and social scientific studies of large, population-based samples [44].

Regarding age, studies have found this variable to be a significantly important one to analyze both the observed and the recommended sleep time. In [17], Tune reviewed habitual sleep patterns of more than 500 adults using time diaries, finding variations in mean sleep time of over 30 minutes across age groups. These results were confirmed in [32,33], whose results indicate that sleep time decreases with age. In 2015, the National Sleep Foundation presented recommendations regarding sleep time duration, stating that sufficient sleep duration requirements vary across the lifespan and from person to person, decreasing with age [38].

2.2. *Paid work*

Paid labor is an activity present in most time use research and has been taken as the basis of every time allocation decision as it is the only activity that can provide monetary income, which allows individuals to buy market goods, the only source of satisfaction recognized by economics at the beginning of the twentieth century. Within labor research, in addition to studying the relationship between individuals' characteristics and the quantity and quality of paid labor, working time has a special characteristic: its allocation depends on how the household manages its monetary budget and needs, so a two-worker household would allocate working time differently among its members if one member either suffers a job loss or a wage rate increase. In that case, the time assigned to work by one individual is now dependent on the income needed both by the individual and the household, and this household now faces the opportunity of having more than one worker, a complementary source of income. This analysis induces a new way of looking at the trade-off between work and non-work and a new look at the way individuals interact within the household: cooperation, bargain, unilateral decisions.

The predominant variable used to analyze, describe and compare paid labor behavior has been gender. In [21], Gershuny stated that women are spending more time engaged in paid work than ever before, specially mothers with young children, and that there are important differences between employed women and employed men. Additionally, in [22] Gornick concluded that across developed societies, women continue to face substantial gender differentials in labor participation rates, working hours, and income that grow even larger among parents. These inequalities are rooted in women's enormous responsibilities for childcare and domestic labor and have enduring effects that contribute to long-term inequalities in earnings and reinforce patterns of gender discrimination in the labor market [4,12,18,23,27,43,62,63].

When addressing household dynamics among couples, gender also plays an important role given that inequalities in paid working hours contribute to a power imbalance at home [2,60]. Dual-earner couples with fewer combined work hours will probably depend on small contributions from wives, and longer hours from husbands. Couples with a higher combined working time are likely to depend substantially on the contributions of wives, however the husbands of these couples are also likely to work longer hours and thus contribute less to household production [26]. In terms of intra-household dynamics, in [13,14] Becker analyzed a two-member household scenario, where each individual is characterized by his/her own preferences, introducing the idea of "caring" by assuming that the preferences of one individual are influenced by the other member's utility function; also, individual specific wage rates are introduced. However, there has been criticism against most collective labor supply models because they have been restricted to the case where both household spouses work [11]. This issue also shows the importance of gender division of labor: from the studies that consider the possibility of one working member, the focus is on the participation decision of women, but they all treat the decision of the other household member as exogenous. The incorporation of the husband's working time in a married woman's labor supply is often used to test for the presence of an added worker effect, according to which the spouse of an unemployed man would offer more labor than a woman with an employed spouse to compensate the income loss of the household [53].

The presence of children is another important variable that has showed gender division of labor. According to Gornick [22], parenting is far more likely to influence women participation in the work force than men across countries. Mothers have less working time than married women without children [25]. For husbands, the effects are much smaller and typically in the other direction: fathers, in general, work longer hours that other men. However, results show that there are significant differences between countries, suggesting that the national context (e.g. the structure of public policies and work institutions) can greatly influence women's ability to combine paid work and motherhood [26].

In terms of employment status, a large body of literature shows that rates of part-time work differ greatly across industrialized countries and remains the main source of women's employment status [22, 27,41,42]. Many employed wives, specially mothers with young children, seek part-time jobs to balance work and family. However, and according to [26], since the balance of part-time to full-time jobs is much higher in some countries, wives in countries with limited full-time opportunities see less chances of achieving equality in earnings and career mobility as well as working time.

Nonetheless, several developments in advanced societies may help, over time, to equalize the division of tasks towards a more egalitarian family: the social movement emphasizing the values of women's liberation; the increase of the general level of education driving more educated men to spend more time in domestic labor and more educated women to spend more time in paid work; the increase in women's participation in the active labor force; the reduction in the average number of children per family; and the use of technology, which enhances the productivity of household chores [45,51].

3. Proposed data mining framework for time use modeling

In this section, a clustering-based framework is presented for analyzing time use patterns across respondents. The reasoning behind this strategy is that a multidimensional scheme to group individuals leads to non-obvious insights for decision-making and allows the identification of various groups which are homogenous in terms of their characteristics and time allocation, but present important differences among clusters. Identifying these differences leads to relevant insights, since it allows decision-makers to understand which variables impact time use decisions for each group, going one step beyond the traditional questions that relate activities such as sleep or work with variables such as gender or age in a bivariate fashion. The clustering method proposed in this framework is K-means, the best-known approach for this task, which has been successfully used for intelligent pattern analysis (see e.g. [3,61]).

The proposed strategy has the following steps:

1. Select the most influential explanatory variables for the time-use study to conduct the clustering procedure. This selection should be done based on prior works on the topic.
2. Normalize numerical variables, e.g. by calculating standard scores according to Eq. (1). For each feature, compute its mean and standard deviation. Next, subtract the mean and divide the values by its standard deviation.
$$x_i^* = \frac{x_i - \text{mean}(X)}{\text{std}(X)} \tag{1}$$
3. For nominal variables, apply binarization via dummy coding to have all variables in the same range (between 0 and 1).
4. Apply K-means clustering to the selected variables to identify homogenous clusters of respondents. Run K-means for an increasing value of k, the number of predefined groups, and monitor the value of metrics designed to identify the optimal number of clusters, such as the overall within-cluster variance. Use the elbow rule to decide the final number of clusters.
5. For each cluster found in the previous step, apply linear regression with sleep as a dependent variable and the rest as independent variables. Stepwise variable elimination can be used to identify which ones are significant in explaining the time allocated to sleep.

Next, the K-means clustering algorithm and the elbow rule are detailed and formalized. The K-means approach defines k clusters *a priori*, with the purpose of finding the partition that minimizes the within-cluster distances (also referred to as within-point scatter). Formally, given an encoder C where $C(i) = k$ means that the i-th instance is assigned to the k-th cluster, $k = 1, \ldots, K$, *K-means* aims at minimizing:

$$\min_C W(C) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i:C(i)=k}\sum_{i':C(i')=k}||x_i - x_{i'}||^2 = \sum_{k=1}^{K} N_k \sum_{i:C(i)=k}||x_i - \bar{x}_k||^2 \tag{2}$$

where $\bar{x}_k$ corresponds to the mean vector related to the k-th cluster, while N_k is the number of examples in cluster k. Equation (2) is NP-Hard since the evaluation of all possible partitions of k clusters cannot be solved in polynomial time [27]. Therefore, the *K-means* method solves Eq. (2), heuristically in two steps, as follows:

Algorithm 1: K-means

Random initialization of cluster centroids

repeat

1. Given a set of centroids $\{m_1, \ldots, m_k\}$, the total within-point scatter is minimized by assigning each observation to the closest centroid. That is, the cluster to which a sample x_i is assigned is given by:
$$C(i) = \textit{argmin}_{1 \leqslant k \leqslant K}||x_i - m_k||^2 \tag{3}$$
2. For a given clustering partition, update the centroids by computing the means of the currently assigned clusters.

until Convergence is reached (assignments do not change)

Convergence is assured for the k-means algorithm, although it may find a suboptimal local minimum [64]. In case of mixed data types, such as numerical and dummy variables, the Euclidean distance used in Eq. (3) can be replaced by a more suitable measure, such as the Gower distance [29]. The main idea is to use a distance measure for each variable, such as simple matching for nominal attributes or the Manhattan distance for numeric ones, normalize these distances between 0 and 1, and finally aggregate all distances by averaging them.

The elbow rule selects a number of clusters so that the gain of including an additional cluster is very small. It works as follows: first, the within-cluster variance is plotted against the number of clusters for an increasing value of k. The first clusters usually capture a great percentage of the variance of the problem, but the marginal gain drops at some point, giving an angle in the plot (the "elbow"). The parameter k is then chosen at this point [9].

This study uses the linear regression analysis to find time use patterns. While there are more sophisticated estimation procedures, the focus of this paper is to shed a light on the potential benefits of a cluster-based analysis. Furthermore, linear regression has been used previously in time use research [6,28,55,67],[1] with studies adjusting their regressions by different variables (e.g. age, gender, ethnicity, income, education, number/presence of children), delivering what could be interpreted as the first step towards a clustering procedure [36]. However, their discussions are based in describing the relation between time allocation and one variable at a time.

The presented framework allows decision-makers to contrast differences among various groups of respondents, finding non-obvious patterns thanks to data mining techniques. In the next sections, the results of studying this framework using a well-known time use dataset, are described.

4. Data and results

4.1. Description of the dataset

This study used the Multinational Time Use Survey (MTUS). The MTUS is an ex-post harmonized cross-time, cross-national, comparative time-use database. It is constructed from national randomly sampled time-diary studies, with common series of background variables and total time spent in 69 activities. The MTUS provides information on individual time use, based on questionnaires in which individuals report their activities throughout the 24 hours of the day. This sample is limited to the 2003–2012 waves of the American Time Use Survey (ATUS). The original sample consisted of 130,610 individuals from the ATUS Survey between 2003 and 2012.

Given that not all of the harmonized background variables reported by MTUS were included in all of the surveys (more than 60 surveys in 23 countries), only the background variables reported for all 10 ATUS datasets were selected. Additionally, the respondents with incomplete data for those variables were removed, leading to a final sample of 117,504 individuals. The pre-processing step also included two variable transformations. On one hand, the ATUS information contained in the MTUS dataset included income as a nominal variable, with 16 categories ranging from 0 to over 150,000 US dollars. To transform this categorical variable to a numerical one, we used the average of the two values that define the range of each category to be the income reported by the individual. On the other hand, to identify the potential impact of the type of day (working day or weekend day), each day was assigned a value of 0 if it was a working day, or a value of 1 if it was a weekend day. This new variable was used as a control in the regression analyses.

[1] Other types of methodologies commonly used in the literature are logistic regression and odds ratios.

Table 1
Descriptive statistics for the entire sample by type of variable (ATUS 2003–2012)

Variable type	Variable description	Variable name	Mean (hours)	Std. deviation (hours)
Dependent	Time allocated to the activity	Sleep	8.74	2.23
		Work	2.77	4.06
			Category/value(s) most present	**Percentage**
Independent	Sociodemographic	Income	{**Low**, Medium, High, Highest}	44.0
		Number of children	{**0**, 1, 2, 3+}	54.5
		Age	{young, **middle aged**, old}	39.3
		Gender	{Men, **Women**}	56.6
		Household type	{one person, couple, **couple living with others**, other type}	40.8
		Household size	{0, **1**, **2**, 3...n} individuals living in the household	51.1
		Family status	{adult 18–39 with no co-resident child $<$ 18, adult 18+ living with 1 co-resident child $<$ 5, adult 18+ living with 1 co-resident child 5–17 none $<$ 5, **Aged 40+ with no coresident children under 18**}	43.3
		Unmarried child in parent home	{Yes, **No**}	94.1
		Single parent	{Yes, **No**}	91.0
		In a couple	{**Yes**, No}	56.5
		Citizen	{**Yes**, No}	92.4
		Employment status	{**Full time**, part-time, not in paid work}	52.4
		Student	{Yes, **No**}	93.3
		Retired	{Yes, **No**}	84.1
		Highest level of education	{Uncompleted secondary, completed secondary, **above secondary education**}	60.2
	Control (Temporal)	Type of day	{Week day, **Weekend**}	50.4
		Year	{**2003**, ..., 2012}	14.8
		Month	{**January**, ..., December}	10.1

The dependent variable of this study is sleep time (including sleep and naps, and imputed sleep) and working time (paid work – main job – not at home, paid work at home, and second or other job not at home) while the independent variables were classified in two categories: sociodemographic characteristics, and temporal variables. The latter group was formed by control variables since they are not of primary interest in this research.

In terms of sociodemographic variables, all of those that were fully reported by the individuals were selected. Table 1 shows the descriptive statistics with the mean and the standard deviation of the activities conducted, and the category with the most responses within the sociodemographic and temporal variables.

Household characteristics include household type, household size, family status, presence of an unmarried child in the parental home, and number of children at home. Individual characteristics included gender, age, whether the respondent was a single parent, if they were in a couple, if they were American citizens, and their highest level of education. Working arrangements are also included, such as their employment status, if they were students, or retired. Finally, to detect temporal effects, the regression included control variables: year, month and type of day (weekday or weekend), variables used by studies such as [36] to analyze differences in time allocation.

4.2. Clustering results

To use the clustering procedure presented in Section 3, we needed to select those variables that allow

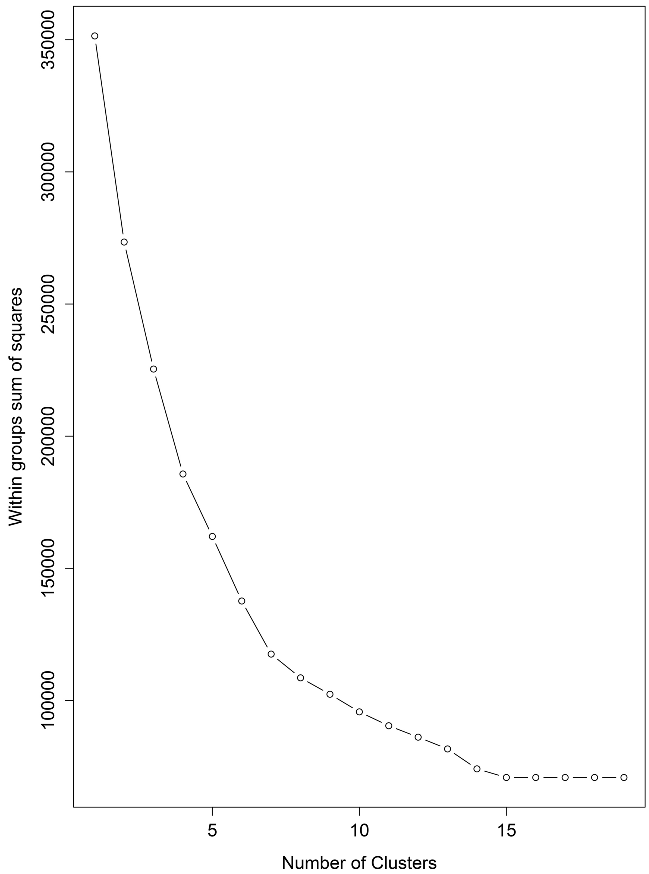

Fig. 1. Elbow rule for the K-means clustering procedure.

the best way to categorize the data. Within the time use literature, there are some issues related to specific variables that have consistently taken the lead role when analyzing the impact and/or relation they have on time allocated to both sleep and working time. These are: gender, age, income level, and the number of children [7,47,49,65]. Although other variables such as education level or employment could have been included in this step, we decided to use them in the regression step to strength the effect of the four aforementioned variables. It is important to notice that clustering algorithm usually assign the same importance to all variables, and the inclusion of many variables affects negatively the performance of the clustering model [64]. Given these variables, the K-means clustering procedure was conducted to identify homogenous clusters of respondents.

In Fig. 1, the total within-cluster distance is presented for an increasing number of clusters. It can be observed that this metric is high when few clusters are used, but it becomes steadily lower when using 15 clusters or more. The final number of clusters delivered by the elbow rule was 15, since the marginal decrease of this metric is very low when using 16 clusters instead of 15. In other words, there is very little gain in terms within-cluster variance if we use 16 clusters compared to 15.

Tables 2–5 and Annex A present the descriptive statistics for all clusters found using the K-means clustering algorithm.

Table 2 shows that the clustering procedure delivered homogeneous subsamples of individuals, in which every cluster was encompassed only by one gender and presented highly concentrated age, in-

Table 2
Descriptive statistics of clustering variables for all 15 clusters – percentages

Cluster	Sample	Gender		Age			Income				Number of children			
	size	Men	Women	Young (18–39)	Middle (40–59)	Old (60+)	Low (0–40)	Medium (40–80)	High (80–150)	Highest (150+)	0	1	2	3+
1	9569	100.0	0.0	97.5	2.5	0.0	46.6	39.5	13.9	0.0	66.7	33.3	0.0	0.0
2	8860	100.0	0.0	2.9	93.0	4.2	67.3	32.7	0.0	0.0	79.9	20.1	0.0	0.0
3	7597	0.0	100.0	59.5	39.9	0.7	0.0	17.6	82.4	0.0	0.0	23.3	62.6	14.0
4	12298	0.0	100.0	0.0	0.0	100.0	80.6	17.5	1.9	0.0	96.9	2.3	0.7	0.0
5	7636	100.0	0.0	6.0	72.2	21.7	0.0	23.5	76.5	0.0	74.9	24.9	0.1	0.0
6	5016	0.0	100.0	74.8	23.5	1.7	57.6	35.1	7.2	0.1	0.0	0.0	0.0	100.0
7	10110	0.0	100.0	59.4	39.7	0.8	66.0	34.0	0.0	0.0	0.0	39.4	60.6	0.0
8	6587	100.0	0.0	61.3	37.1	1.5	53.4	43.6	2.9	0.0	0.0	0.0	54.1	45.9
9	8390	0.0	100.0	99.3	0.6	0.0	51.3	36.5	12.2	0.0	63.0	37.0	0.0	0.0
10	11778	0.0	100.0	0.7	77.1	22.1	69.3	30.7	0.0	0.0	88.2	11.7	0.0	0.0
11	6268	100.0	0.0	51.8	47.2	1.1	0.0	18.3	81.2	0.4	0.0	15.7	66.9	17.4
12	3290	0.0	100.0	32.8	55.1	12.2	0.0	0.0	0.0	100.0	35.7	22.9	29.1	12.3
13	8757	100.0	0.0	0.0	0.0	100.0	66.6	29.7	3.6	0.0	96.0	3.2	0.8	0.0
14	3260	100.0	0.0	27.4	57.0	15.5	0.0	0.0	0.0	100.0	39.4	23.0	26.7	10.7
15	8088	0.0	100.0	5.1	71.6	23.2	0.0	29.3	42.9	27.8	79.1	20.6	0.3	0.0
Overall	**117504**	**43.4**	**56.6**	**36.1**	**39.3**	**24.5**	**44.1**	**28.0**	**22.5**	**5.6**	**54.5**	**18.6**	**17.6**	**9.2**

Table 3
Overall description of all 15 clusters

Cluster	Overall description
1	Young men with low income and no children living in the house
2	Middle aged men with low income and no children living in the house
3	Young women with high income and two children living in the house
4	Old women with low income and no children living in the house
5	Middle aged men with high income and no children living in the house
6	Young women with low income and three or more children living in the house
7	Young women with low income and two children living in the house
8	Young men with low income and two children living in the house
9	Young women with low income and no children living in the house
10	Middle aged women with low income and no children living in the house
11	Young men with high income and two children living in the house
12	Middle aged women with the highest income and no children living in the house
13	Old men with low income and no children living in the house
14	Middle aged men with the highest income and no children living in the house
15	Middle aged women with high income and no children living in the house

come, and number of children categories. Color coding shows the highest category/values in each variable in terms of the percentage of observations. Table 3 presents a descriptive summary of all the clusters regarding the selected variables, showing that every cluster presents a distinctive definition.

Continuing the overall description of clusters, Table 4 provides every cluster's average time allocation to sleep and work. Data shows a negative correlation of 0.6. Additionally, in terms of sleep, income was the segregating force, a result partially supported by [16], with low income men and women sleeping more than high income individuals. Regarding work, gender was the separating force, followed by age, with men always allocating more time to work than women, except for old men, which together with old women (clusters 4 and 13), were the clusters with the lowest time allocated to work and 2 of the highest sleep time allocations, which is supported by studies such as [36]. In a more specific note, when segmenting clusters by gender but pairing them with a matching cluster with the same income level and number of children – clusters 9 and 1, 4 and 13, 10 and 2, 3 and 11, 12 and 14, 15 and 5 and 6, 7 and

Table 4
Average daily hours of sleep and work time for all clusters

Cluster	Sample size	Sleep (hours)	Work (hours)
1	9569	8.90	3.89
2	8860	8.75	3.41
3	7597	8.47	2.68
4	12298	9.03	0.43
5	7636	8.27	4.09
6	5016	8.75	1.98
7	10110	8.96	2.46
8	6587	8.61	4.06
9	8390	9.23	3.07
10	11778	8.83	2.54
11	6268	8.22	4.48
12	3290	8.45	2.65
13	8757	8.99	0.92
14	3260	8.23	4.17
15	8088	8.41	3.08
Overall	**117504**	**8.74**	**2.77**
Correlation coefficient			**−0.6018**

Table 5
Measures of association and ANOVA test for sleep and work time between all clusters

	Measures of association				ANOVA	
	R	R Squared	Eta	Eta Squared	F	Sig.
Sleep*clusters	−0.028	0.001	0.133	0.018	152.096	0.000
Work*clusters	−0.009	0.000	0.294	0.087	796.840	0.000

8 – women always slept more and allocated less time to paid labor than their male counterparts, a result supported by [58].

We ran an ANOVA test and a Measures of Association test. Results show significant differences across clusters, as shown in Table 5, both for sleep and work.

In terms of the sociodemographic data, clustering results show the following: On one hand, the sample presented variables with the same predominant category regardless of clustering. On average, 91% of the respondents were not single parents; 94% of the individuals did not live with an unmarried child in the household; 93% of the individuals were citizens; 93% were not students; and 60% had an education above the secondary level. On the other hand, we found variables with a predominant category present in only some of the clusters. For example, in clusters 4 and 13, data shows that most of the individuals in those clusters were not in paid work (86% and 77% respectively), and were retired (76% and 65% respectively). Annex A complements this information by presenting the sociodemographic variables that showed a lower concentration of observations between categories, i.e. those with different predominant categories depending on the clusters: household type and size, family status, civic status, and employment status.

Some insights are worth noting: on average, whenever children were present in the households, they were always living with both parents; household size grows with the number of children living in the household; and there is a positive relation between the highest level of education of the individual and their income, regardless of age, gender and number of children.

4.3. Regression analysis results

To test and compare the proposed data mining framework, we first estimated a model with all the

Table 6
Overall model results for sleep

Variables	Standarized beta	t-stat	Sig.
Employment status – Full time	−0.164	−41.372	0.000
Number of children	−0.080	−10.881	0.000
Age	−0.071	−12.323	0.000
Employment status – Part time	−0.064	−18.825	0.000
Education – Less than secondary	0.055	17.046	0.000
Income	−0.052	−15.711	0.000
Family status – Aged 18+ living with 1+ coresident children 5–17, none < 5	0.040	7.245	0.000
Education – Above secondary	−0.035	−10.406	0.000
Household type – Married plus others	−0.033	−6.470	0.000
Household size	0.032	4.236	0.000
Student	−0.029	−9.372	0.000
Citizen	−0.028	−9.502	0.000
Unmarried children living in home	0.027	6.944	0.000
Family status – Aged 18 to 39 with no coresident children < 18	0.026	6.106	0.000
Family status – Aged 18+ living with 1+ coresident children aged < 5	0.026	4.440	0.000
Civic status	0.015	3.463	0.001
Gender	0.013	4.483	0.000
Retired	−0.013	−2.944	0.003
Adjusted R squared		**0.090**	
Durbin-Watson		**2.004**	

Table 7
Overall model results for work

Variables	Standarized beta	t-stat	Sig.
Employment status – Not in paid work	−0.520	−195.341	0.000
Employment status – Part time	−0.154	−64.370	0.000
Gender	−0.047	−20.163	0.000
Family status – Aged 40+ with no coresident children < 18	0.029	5.049	0.000
Age	−0.026	−6.509	0.000
Household size	0.026	4.493	0.000
Civic status	0.023	7.381	0.000
Citizen	−0.016	−7.222	0.000
Number of children	−0.014	−2.488	0.013
Unmarried children living in home	−0.011	−3.727	0.000
Family status – Aged 18 to 39 with no coresident children < 18	0.011	2.963	0.003
Student	−0.010	−4.169	0.000
Family status – Aged 18+ living with 1+ coresident children 5–17, none < 5	0.007	2.221	0.026
Adjusted R squared		**0.416**	
Durbin-Watson		**2.003**	

individuals (the "overall" model) to analyze which of the variables had a significant impact on time allocated to sleep and on time allocated to work. Therefore, we conducted a stepwise forward linear regression [17][2] for the whole sample (117,504 individuals) to estimate the time allocated to sleep and work with all the variables presented in Table 1 (sociodemographic and temporal factors). These results are presented in Tables 6 and 7.

In Tables 6 and 7 we observe the significant variables found in the multivariate linear regression for

[2]The stepwise forward regression involves starting with no variables in the model, then testing the incorporation of every variable, adding the variable – if any – that improves the model the most, and repeating this procedure until there is no variable that improves the model. The categorical variables were separated into dummies for each category.

the entire sample (control variables not reported)[3] ordered by higher standardized beta (absolute value). The variables with the highest influence over the time allocated to sleep are: employment status (working full time, part time, or not at all), number of children at home, age, the education level of the individual and income. Surprisingly, gender is not one of the most influential, even though most studies start their analysis by segmenting per gender. In the regression regarding work, the variables with the highest influence were employment status, gender, family status, age and household size. Surprisingly, income is not significant in this estimation. Results suggest that both age and employment status are the common driving force that leads to the allocation of time for these two important activities: sleep and work.

Results show the same relationship (in terms of its sign) between most sociodemographic variables and time allocated to sleep and work (e.g. a negative relationship between working part time and both sleeping and working time); however, gender, and the presence of an unmarried children in the house has a negative impact on working time while presenting a positive impact on sleep.

Next, for each cluster found, linear regressions were conducted with sleep and work as dependent variables and the rest as independent variables. The variables used for clustering – age, gender, income and number of children – were excluded from the analysis since they are assumed to be homogeneous within clusters, while the variables that were not significant in the overall models for sleep and work were included. Tables 8 and 9 show whether each variable was significant (and its relationship with sleep and work) in the "overall model" and in the cluster regressions, also including the respective adjusted R squared, and the final column shows on how many clusters that variable was significant.

There are many noticeable results shown in Tables 8 and 9. First, only 1 out of all the variables (employment status) was significant across all clusters for both activities. Moreover, education was significant for all clusters only in sleep, a result contrasted by the fact than when regression on work time, education was not significant in the overall model. Family status is only significant in clusters with an average of zero children. Civic status is only significant in clusters exclusively formed by women.

An interesting result is drawn from household size and household type. When analyzing sleep, household size had a positive impact on sleep in the overall model but a negative impact in the cluster regressions (young men and women with high income and an average of 2 children, and middle age women with the highest income of the sample and no children); while when analyzing work time, household size was only significant in the overall regression (with a positive relationship). Household type, in another hand, presented a different relationship with sleep depending on the cluster, with a one-person household having a positive impact on the sleep of individuals in cluster 14 (middle age men with the highest income and no children) and a negative impact on cluster 15 (middle age women with high income and no children), further showing the presence of gender divisions. When analyzing work time, however, household type was not significant in the overall model and only impacted 3 clusters, compared to the 8 clusters within the sleep analysis.

Another noteworthy result is drawn from the presence of an unmarried children in the house. This variable has a negative impact on working time while presenting a positive impact on sleep, a situation mirroring the relationship between gender and both sleep and work time. However, this variable is only significant in one cluster when analyzing work time (young men with low income and no children) while is significant for 9 clusters when analyzing sleep.

Results show that within the variables that were not significant in the overall models, whether the individual was a single parent is a significant variable in only 1 cluster for both sleep and work time (young women with low income and 2 or more children living in the house). Furthermore, education

[3]Type of day was one of the most influential independent variable for the overall model and for the 15 clusters.

Table 8
Significant variables for each cluster for sleep

	Total	Clusters															Number of
	sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	clusters
Education																	**15**
Above secondary	–	–		–		–	–	–	–	–	–				–	–	10
Less than secondary	+	+	+		+			+	+		+	+	+	+			9
Completed secondary			+		+					–				+			4
Employment status																	**15**
Full time	–	–	–		–	–	–	–	–	–	–	–	–	–	–		13
Not in paid work		+		+	+		+	+	+	+	+	+		+		+	11
Part time	–		–														1
Citizen	–	–	–	–		–	–		–			–	–	–		–	**10**
Unmarried child living in home	+	+		+			+	+	+	+		+	+		+		**9**
Student	–	–	–			–		–	–	–		–				–	**8**
Household type																	**8**
Other		+			+	+			+					+			5
One person household															+	–	2
Married with others	–									–							1
Married alone																	0
Family status																	**6**
Aged 18 to 39 with no coresident children under 18	+	+	+			+					+					+	5
Aged 18+ living with 1+ coresident children 5–17, none under 5	+										–				–		2
Aged 18+ living with 1+ coresident children aged under 5	+																0
Aged 40+ with no coresident children under 18																	0
Retired	–		–		–						–		+	–			**5**
Household size	+			–								–	–				**3**
Civic status	+							+			+						**2**
Single parent							+										**1**
Adjusted R square	**0.090**	**0.098**	**0.062**	**0.090**	**0.036**	**0.094**	**0.097**	**0.095**	**0.121**	**0.092**	**0.077**	**0.119**	**0.103**	**0.033**	**0.114**	**0.079**	

Table 9
Significant variables for each cluster for work

	Total	Clusters															Number of
	sample	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	clusters
Employment status																	**15**
Not in paid work	–	–	–	–	–	–	–	–	–	–	–	–	–	–		–	14
Full time		+	+	+	+	+		+	+		+		+	+	+		11
Part time	–						–			–		–			+	–	5
Family status																	**6**
Aged 18 to 39 with no coresident children under 18	+									+					+		2
Aged 18+ living with 1+ coresident children 5–17, none under 5	+					–										–	2
Aged 18+ living with 1+ coresident children aged under 5													–		–		2
Aged 40+ with no coresident children under 18	+										+						1
Citizen	–	–	–				–	–			–						**5**
Civic status	+						+			+	+		+			+	**5**
Education																	**3**
Above secondary										–				+			2
Completed secondary												+					1
Less than secondary																	0
Household type																	**3**
Other			–														1
Married with others											+						1
Married alone						–											1
One person household																	0
Student	–	–								–					–		**3**
Unmarried child living in home	–	–															**1**
Single parent								+									**1**
Retired																	**0**
Household size	+																**0**
Adjusted R squared	**0.416**	**0.351**	**0.396**	**0.393**	**0.405**	**0.428**	**0.360**	**0.372**	**0.361**	**0.358**	**0.405**	**0.435**	**0.456**	**0.445**	**0.490**	**0.431**	

is significant in 3 clusters when regressing on work time (against 15 clusters when regressing on sleep time). Retired is not significant variable in any cluster when regressing on work time, a fact contrasted by the 5 clusters where being retired was significant when analyzing sleep. Moreover, retired presented a positive relationship with sleep time in cluster 12 (middle age women with the highest income of the sample and no children), while presenting a negative relationship in the other 4 clusters.

In general, there are much less significant variables per cluster when analyzing work than sleep, showing that when constructing homogeneous groups of individuals by age, gender, income and number of children, different variables influence time allocated to different activities. This is contrasted by the fact that the explanatory power – represented by the adjusted R squared – of the overall model for work was much higher (0.416) than for sleep (0.090). Furthermore, when analyzing sleep, the explanatory power within the clusters presented a relevant variation ranging from 0.033 (cluster 13) to 0.121 (cluster 8); when analyzing work, this variation yields a distribution, ranging from 0.351 (cluster 1) to 0.490 (cluster 14).

This is the first study, to the best of our knowledge, to provide information about clustering within time use literature. Results shows that the impact of different variables on sleep and work time allocation is significantly related to the characteristics of the individuals being analyzed. This suggests that concerns about sleep and work behaviors should no longer be limited to individual analysis, e.g. by gender, by age group, or other single variables. These results provide the needed insights for future survey design to select and collect the necessary data while working with limited resources. The present findings may have significant implications. Based on these results, we recommend that time use researchers advance their methodological analysis towards a multifactorial approach and include clustering as a fundamental step. This could lead to beneficial changes in the reporting and interpretation of results.

However, these results should not be used to generalize to the larger American population. Residual confounding as well as additional unmeasured conditions may have influenced associations in this study. There is always the possibility of residual effects of control variables due to imprecision in construct measurement. We used education and income to measure SES. The use of additional indicators of SES, including more refined measures of education (more than 3 categories) and income (real values instead of imputation) may reveal other insights. It is possible that more detailed descriptions of home and family stressors might reveal important influences on sleep and work in future research.

5. Conclusions

This study proposes a clustering-based procedure aimed at evolving time use research towards multifactorial analysis. By clustering with some of the most frequently used variables in time use research – age, gender, income and number of children – the study of sleep and work time allocation could focus on other potentially relevant factors within the same group of sociodemographic variables.

After estimating results for the total sample, and for each of the 15 clusters obtained, evidence suggests that the impact of different variables on sleep and work allocation depends significantly on the characteristics of the individuals being analyzed. Even though this insight may appear to be common sense, most studies on time use and its relationship with sociodemographic characteristics focus only on multiple, but independent, bivariate discussions.

In addition to the most commonly studied variables within the time use literature – gender, age, income, and number of children – results show that the highest level of education, employment status and citizenship played the most important roles impacting sleep time across clusters, while household size, being in a couple or being a single parent were the least significant variables. Surprisingly, gender is not one of the most influential. Regarding work time, employment and family status had the most important roles, however both education and citizenship had far less impact than for sleep time. Results suggest that both age and employment status are the common driving force that leads to the allocation of time for these two important activities: sleep and work. Some additional interesting results appear: gender, and the presence of an unmarried children in the house, have a negative relationship on working time while a positive one on sleep.

As previously stated, these results provide the necessary insight for future survey design to select and collect only relevant data while working with limited resources. Proper identification of the most significant variables involved in the decision-making behind sleep time allocation would allow researchers to better analyze and interpret their data and results.

Further research could focus on expanding the discussion and interpretation of results using a clustering procedure. Moreover, the variables used in this study for clustering the total sample can be subject to modification; for example, the level of education and citizenship showed great impact across clusters while gender had one of the lower impacts in the overall regression on sleep and income had no impact in the overall regression on work. Another research line corresponds to the proposal of alternative approaches to our clustering-based framework. For example, decision trees such as CART can be useful for finding patterns in time use modelling thanks to their easy interpretation. Finally, it is worth acknowledging that this study could be further replicated with other activities other than sleep and work, such as leisure, household labor and travel.

References

[1] A. Rafferty, Ethnic penalties in graduate level over-education, unemployment and wages: evidence from Britain, *Work, Employment & Society* **26**(6) (2012), 987–1006.

[2] B. Hobson, No Exit, No Voice: Women's Economic Dependency and the Welfare State, *Acta Sociologica* **33**(3) (1990), 235–250.

[3] B. Peralta, P. Espinace and A. Soto, Enhancing K-Means using class labels. Intelligent, *Data Analysis* **17**(6) (2013), 1023–1039.

[4] C. Carrasco and A. Recio, Time, Work and Gender in Spain, *Time & Society* **10**(2–3) (2001), 277–301.

[5] C. Gale and C. Martyn, Larks and owls and health, wealth, and wisdom, *British Medical Journal* **317** (1998), 1675–1677.

[6] C. Van Klaveren, B. van Praag and H.M. van den Brink, A collective household model of time allocation: A comparison of native dutch and inmigrant households in the Netherlands, *CESifo Working Paper Series*, 1753, CESifo Group Munich, 2006.

[7] D. Sinai and L. Tikotzky, Infant sleep, parental sleep and parenting stress in families of mothers on maternity leave and in families of working mothers, *Infant Behavior and Development* **35**(2) (2012), 179–186.

[8] D. Thompson, *How Much Do We Sleep? Americans Don't Know*, The Atlantic, (2014), Retrieved from http://www.theatlantic.com/health/archive/2014/01/how-much-do-we-sleep-americans-dont-know/283148/ on October 10th, 2016.

[9] D.J. Ketchen, Jr. and C.L. Shook, The application of cluster analysis in Strategic Management Research: An analysis and critique, *Strategic Management Journal* **17**(6) (1996), 441–458.

[10] D.S. Lauderdale, K.L. Knutson, L.L. Yan, P.J. Rathouz, S.B. Hulley, S. Sidney and K. Liu, Objectively measured sleep characteristics among early-middle-aged adults: The CARDIA Study, *American Journal of Epidemiology* **164** (2006), 5–16.

[11] E. Duguet and V. Simonnet, *The participation of couples in the labor market: an econometric analysis*, Université de Paris I, 2003.

[12] F. Blau, M. Ferber and A. Winkler, *The Economics of Men, Women, and Work*, Upper Saddle River, N.J.: Prentice Hall, 1998.

[13] G. Becker, Altruism in the Family and Selfishness in the Market Place, *Economica* **48** (1981), 1–15.

[14] G. Becker, *A Treatise on the Family*, Cambridge, Mass: Harvard University Press, 1981.

[15] G. Burazeri, J. Gofine and J.D. Kark, Over 8 hours of sleep – marker of increased mortality in Mediterranean population: follow-up population study, *Croatian Medical Journal* **44** (2003), 193–198.

[16] G. Szalontai, The Demand for Sleep: A South African Study, *Economic Modelling* **23** (2006), 854–874.

[17] G.S. Tune, Sleep and wakefulness in normal adult humans, *British Medical Journal* **2** (1968), 269–271.
[18] I. Padavic and B. Reskin, *Women and Men at Work*, Thousand Oaks, California: Pine Forge Press, 2002.
[19] I. Roberts, Taking age out of the workplace: putting older workers back in? *Work, Employment & Society* **20**(1) (2006), 67–86.
[20] J. Boergers, C. Hart, J.A. Owens, R. Streisand and A. Spirito, Child sleep disorders: Associations with parental sleep duration and daytime sleepiness, *Journal of Family Psychology* **21**(1) (2007), 88–94.
[21] J. Gershuny, *Changing Times: Work and Leisure in Postindustrial Society*, Oxford: Oxford University Press, 2000.
[22] J. Gornick, Gender Equality in the Labor Market, in D. Sainsbury, ed., *Gender Policy Regimes and Welfare States*. Oxford: Oxford University Press, 1999.
[23] J. Gornick and J. Jacobs, Gender, the Welfare State, and Public Employment: A Comparative Study of Seven Industrialized Countries, *American Sociological Review* **63**(5) (1998), 688–710.
[24] J. Hislop and S. Arber, Sleepers wake! The gendered nature of sleep disruption among mid-life women, *Sociology* **37** (2003), 695–711.
[25] J. Jacobs and J. Gornick, Hours of Paid Work in Dual-Earner Couples: The U.S. in Cross-National Perspective, *Sociological Focus* **35**(2) (2002), 169–187.
[26] J. Jacobs and K. Gerson, *The Time Divide*, Harvard University Press, 2004.
[27] J. Rubery, M. Smith and C. Fagan, National Working Time Regimes and Equal Opportunities, *Feminist Economics* **4**(1) (1998), 71–101.
[28] J. van den Bulck, Television viewing, computer game playing, and internet use and self-reported time to bed and time out of bed in secondary-school children, *Sleep* **27**(1) (2004), 101–104.
[29] J.C. Gower, A general coefficient of similarity and some of its properties, *Biometrics* **27** (1971), 857–874.
[30] J.P. Robinson and G. Godbey, *Time for Life: The surprising ways Americans use their time*, University Park: The Pennsylvania State University Press, 1997.
[31] J.W. Yeung, Children's time with fathers in intact families, *Journal of Marriage and the Family* **63** (2001), 136–154.
[32] K. Morgan, *Sleep and Aging: A Research-Based Guide to Sleep in Later Life*, John Hopkins University Press, Baltimore, 1987.
[33] K.I. Hume, F. Van and A. Watson, A field study of age and gender differences in habitual adult sleep, *Journal of Sleep Research* **7** (1998), 85–94.
[34] L. Hale, Who has time for sleep? *Journal of Public Health* **27** (2005), 205–211.
[35] L. Ruppanner, Housework conflict and divorce: a multi-level analysis, *Work, Employment & Society* **26**(4) (2012), 638–656.
[36] M. Basner, K.M. Fomberstein, F.M. Razavi, S. Banks, J.H. William, R.R. Rosa and D.F. Dinges, American time use survey: sleep time and its relationship to waking activities, *Sleep* **30**(9) (2007), 1085–1095.
[37] M. Hall, J. Bromberger and K. Matthews, Socioeconomic Status as a Correlate of Sleep in African-American and Caucasian Women, *Annals of the New York Academy of Sciences* **896** (1999), 427–430.
[38] M. Hirshkowitz, K. Whiton, S.M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E.S. Katz, L. Kheirandish-Gozal, D.N. Neubauer, A.E. O'Donnell, M. Ohayon, J. Peever, R. Rawding, R.C. Sachdeva, B. Setters, M.V. Vitiello, J.C. Ware and P.J. Adams Hillard, National Sleep Foundation's sleep time duration recommendations: methodology and results summary, *Sleep Health: Journal of the National Sleep Foundation* **1**(1) (2015), 40–43.
[39] M.A. Efroymson, Multiple regression analysis, in A. Ralston and H.S. Wilf, eds, *Mathematical Methods for Digital Computers*, Wiley, 1960.
[40] National Sleep Foundation, *How much sleep do we really need*? (2016) Retrieved from https://sleepfoundation.org/how-sleep-works/how-much-sleep-do-we-really-need on October 10th, 2016.
[41] OECD (Organization for Economic Cooperation and Development), *Employment Outlook*, Paris: OECD, 1999.
[42] OECD (Organization for Economic Cooperation and Development), *Women and Structural Change: New Perspectives*, Paris: OECD, 1994.
[43] P. Odih, Gender, Work and Organization in the Time/Space Economy of 'Just-in-Time' Labour, *Time & Society* **12**(2–3) (2003), 293–314.
[44] P.M. Krueger and E.M. Friedman, Sleep duration in the United States: A cross-sectional population-based study, *American Journal of Epidemiology* **169** (2009), 1052–1063.
[45] R. Andorka, Time budgets and their uses, *Annual Review of Sociology* **13**(1) (1987), 149–164.
[46] R. Hertz, *More Equal than others: Women and Men in Dual-Career Marriages*, Berkeley: University of California Press, 1986.
[47] R. Leong, J. Lo and M. Chee, Self-reported sleep duration and levels of inflammatory markers: Similarities and differences across genders, *Sleep Medicine* **16** (2015), S61.
[48] R. Meadows, S. Arber, S. Venn and J. Hislop, Engaging with Sleep: Male definitions, understandings and attitudes, *Sociology of Health and Illness* **30** (2008), 696–710.
[49] R.K. Golley, C.A. Maher, L. Matricciani and T.S. Olds, Sleep duration or bedtime? Exploring the association between

sleep timing behaviour, diet and BMI in children and adolescents, *International Journal of Obesity* **37**(4) (2013), 546–551.

[50] S. Chatzitheochari and S. Arber, Lack of sleep, work and the long hours culture: evidence from the UK Time Use Survey, *Work, Employment & Society* **23** (2009), 30–48.

[51] S. Gronmo and S. Lingsom, *Sexual differences in household work: Patterns of time use change in Norway*, Presented at the 10^{th} World Congress of Sociology, Mexico City, 1982.

[52] S. Hegarty, *The myth of the eight-hour sleep*, BBC World Service, (2012), Retrieved from http://www.bbc.com/news/magazine-16964783 on October 10^{th}, 2016.

[53] S. Lundberg, The Added Worker Effect, *Journal of Labor Economics* **3** (1985), 11–37.

[54] S. Maldonado, G. Armelini and C.A. Guevara, An Econometric Model for Assessing University Enrollment and Marketing Efforts, *Intelligent Data Analysis* **21**(4) (2017), 945–962.

[55] S. Taheri, L. Lin, D. Austin, T. Young and E. Mignot, Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index, *PLoS Medicine* **1**(3) (2004), e62, 210-217.

[56] S. Venn, S. Arber, R. Meadows and J. Hislop, The fourth shift: exploring the gendered nature of sleep disruption among couples with children, *The British Journal of Sociology* **59**(1) (2008), 79–97.

[57] S.A. Burgard, The needs of others: gender and sleep interruptions for caregiving, *Social Forces* **89** (2011), 1189–1215.

[58] S.A. Burgard and J.A. Ailshire, Gender and Time for Sleep among US Adults, *American Sociological Review* **78**(1) (2013), 51–69.

[59] S.E. Cha and K.S. Eun, Gender Difference in Sleep Problems: Focused on Time Use in Daily Life of Korea, *Social Indicators Research* **119**(3) (2014), 1447–1465.

[60] S.M. Bianchi, L.M. Casper and P.K. Peltola, A Cross-National Look at Married Women's Economic Dependency, *Gender Issues* **17** (1999), 3–33.

[61] S.M. Savaresi and D.L. Boley, A comparative analysis on the bisecting K-means and the PDDP clustering algorithms, *Intelligent Data Analysis* **8**(4) (2004), 345–362.

[62] S-Y. Won, Gendered working-time arrangements and their policy implications: Korean experiences, *Time & Society* **21**(3) (2012), 285–307.

[63] T. Boje, Gender, Work Time and Flexible Employment, *Time & Society* **5**(3) (1996), 341–361.

[64] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, 2^{nd} ed. Springer-Verlag, 2009.

[65] W.B. Webb and C.M. Levy, Age, Sleep Deprivation, and Performance, *Psychophysiology* **19** (1982), 272–276.

[66] W.M. Troxel, D.J. Buysse, K.A. Matthews, H.M. Kravitz, J.T. Bromberger, M. Sowers and M.H. Hall, Marital/Cohabitation Status and History in Relation to Sleep in Midlife Women, *Sleep* **33**(7) (2010), 973–981.

[67] Y, Abe, Family labor supply, commuting time, and residential decisions: The case of the Tokyo Metropolitan Area, *Journal of Housing Economics* **20** (2011), 49–63.

Annex A. Descriptive statistics of demographic variables for all clusters

	Household type		Household size		Family status		Civic status		Employment status	
Cluster	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage
1	Couple living with others	35	3 individuals	31	Aged 18 to 39 with no coresident children under 18	65	Not in a couple	64	Full time	71
2	1 person household	45	1 individual	45	Aged 40+ with no coresident children under 18	77	Not in a couple	60	Full time	63
3	Couple living with others	93	4 individuals	55	Aged 18+ living with 1+ coresident children 5–17, none under 5	59	In a couple	89	Full time	53
4	1 person household	63	1 individual	63	Aged 40+ with no coresident children under 18	97	Not in a couple	74	Not in paid work	86
5	Couple living alone	40	2 individuals	44	Aged 40+ with no coresident children under 18	69	In a couple	74	Full time	79
6	Couple living with others	69	5 individuals	45	Aged 18+ living with 1+ coresident children aged under 5	60	In a couple	66	Not in paid work	45
7	Couple living with others	56	3 individuals	37	Aged 18+ living with 1+ coresident children 5–17, none under 5	67	In a couple	54	Full time	47
8	Couple living with others	90	4 individuals	43	Aged 18+ living with 1+ coresident children aged under 5	51	In a couple	87	Full time	79
9	Couple living with others	32	2 individuals	34	Aged 18 to 39 with no coresident children < 18	62	Not in a couple	62	Full time	57
10	1 person household	45	1 individual	45	Aged 40+ with no coresident children under 18	88	Not in a couple	64	Full time	48

	Household type		Household size		Family status		Civic status		Employment status	
Cluster	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage	Category/value most present	Percentage
11	couple living with others	96	4 individuals	59	Aged 18+ living with 1+ coresident children 5–17, none under 5	57	In a couple	91	Full time	90
12	Couple living with others	69	4 individuals	32	Aged 18+ living with 1+ coresident children 5–17, none under 5	41	In a couple	84	Full time	53
13	Couple living alone	46	2 individuals	50	Aged 40+ with no coresident children under 18	96	In a couple	54	Not in paid work	77
14	Couple living with others	68	4 individuals	31	Aged 18+ living with 1+ coresident children 5–17, none under 5	40	In a couple	82	Full time	83
15	Couple living alone	43	2 individuals	50	Aged 40+ with no coresident children under 18	74	In a couple	72	Full time	62
Overall	**Couple living with others**	**41**	**2 individuals**	**27**	**Aged 40+ with no coresident children under 18**	**43**	**In a couple**	**57**	**Full time**	**52**