Australasian Transport Research Forum 2013 Proceedings
2 - 4 October 2013, Brisbane, Australia
Publication website: http://www.patrec.org/atrf.aspx

Workers, adventurers, explorers: uncovering activity patterns in Melbourne

Nicole Ronald

Faculty of Information and Communication Technologies, Swinburne University of Technology, Hawthorn, Australia

Email for correspondence: nronald@swin.edu.au

Abstract

The bulk of transport modelling is based on peak hour travel and focuses on the daily commute. However, this only considers when and where people are, and also ignores the growing travel demand outside peak hours. Understanding the nature of activities people are participating in is also important in order to understand daily patterns and changing behaviour.

Using daily activity-travel data provided by 16,000 households and 41,000 individuals in Melbourne and regional Victoria in 2009-10, this paper uses the k-means clustering technique to uncover similar groups of individuals. Both weekday and weekend activities are explored. This is also combined with socio-demographic data to identify different types of behaviours.

Although there is still support for a three-category classification on weekdays (work, education, other) and a two-category classification on weekends (home, other), extending the number of clusters to seven and six respectively allows more detail on the larger “other” clusters. In particular, we can look at the different socio-demographic details dominant in each cluster to determine patterns.

More differences between regions were expected, however this was not apparent from the clustering. Some differences, in that those who live in inner Melbourne were more likely to be working or socialising in the evening, were seen, but these were rare.

The results contribute to increasing our understanding of how people move around Melbourne, by providing more detail with regard to where and when people are travelling and what people are doing over the course of a day. Future work involves replicating the process for Brisbane and Sydney and comparing the outcomes, leading to better understanding of activity and travel behaviour for Australian cities.

1. Introduction

The bulk of transport modelling and analysis is based on peak hour travel and focuses on the daily commute. However, this only considers when and where people are, and also ignores the growing travel demand outside peak hours. Understanding the nature of activities people are participating in is also important in order to understand daily patterns and changing behaviour.

Looking at people's activities throughout the day, it is generally assumed that people are workers/commuters (so travelling in the AM and PM peak), students (so travelling in or slightly outside these times), or neither of the above. Using activity-travel data collected in Melbourne and regional Victoria during 2009/10, this paper sets out to explore whether this generalisation is still reasonable. How are Victorians spending their days?

In order to achieve this aim, this paper uses the k-means clustering technique to uncover similar groups of individuals. Both weekday and weekend activities are explored. This is also combined with socio-demographic data to identify different types of behaviours.

The paper is set out as follows. Firstly, we look briefly at the history of activity-based modelling and analysis and the use of data for these types of analyses. The methodology is then presented, comprising of the dataset used and transformations made, as well as the clustering algorithm. Results are then presented for both weekdays and weekends, before a discussion.

2. Background

Historically, the main elements of transport analysis has been *where* are people travelling and *when* are they travelling. While these are perfectly valid questions, *why* is also a pertinent question. Which activities do people need or want to engage in? This is one of the main reasons behind moving away from a trip-based model, which cannot completely capture the reasoning behind why people are travelling, towards activity-based models.

Activity-based models "aim at predicting which activities are conducted where, when, for how long, with whom, the transport modes involved and ideally also the implied route decisions" (Arentze and Timmermans, 2000). This is different to the traditional trip-based models that focus on single trips, rather than trip chains. It also emphasises why the travel is being undertaken. Certain activity-based models also model the choice process, as opposed to choice models where only the outcome is modelled.

Hagerstrand (1970) is often cited in articles on activity-based approaches as one of the first to recognise that the focus should be more on people rather than locations. This strand of research, known as time geography, is still developing and influencing the activity-based approach (for examples, see Neutens et al. 2011). In the late 1980s, Kitamura (1988) discussed the current state of activity-based analysis and looks at the contribution to the science of travel behaviour and as a planning tool.

While activity-based models are gaining interest in the US (as discussed previously at ATRF (Davidson et al., 2011), interest in Australia has been less enthusiastic, possibly due to the large amount of data required for validation and confidence in the model. Even without moving to an activity-based approach in modelling, analysing activities will offer insight into why people are travelling. Activity-travel surveys being undertaken around the country (see Inkabaran and Kroen (2011) for a list and a review of international survey methodology) provide a wealth of data that can be used to explore changes in travel and activity behaviour.

In general, large amounts of data are currently being collected on travel behaviour (Miller, 2008), ranging from vehicle counts and public transport passenger boardings to volunteered data via websites and survey participation. These data sets need to be processed in some way that it is useful to planners and decision-makers. On way of doing this is by clustering: using algorithms to look for patterns in the dataset. Jiang et al. (2012) undertook clustering on an activity-travel dataset collected in Chicago to discover the underlying patterns. Moiseeva (2013) used clustering to explore patterns in location knowledge acquisition by newcomers to a city, using data collected via GPS.

Dodson et al. (2010) applied clustering to the South-East Queensland household travel survey to identify transport-disadvantaged subgroups (low-income sole parents, working poor, students, single retired females with and without licences, and partner retired elderly without licences). The trip rate, mode share, the number of kilometres traveled by mode, and trip purpose for each cluster was then explored. This research used a similar methodology to Ryley (2006), who explored the usage of non-motorised modes in Edinburgh, UK: in both cases clustering was undertaken on socio-demographic variables.

3. Methodology

The methodology behind this work comprises several steps. Firstly, we describe the data used. The purposes were altered slightly to combine similar purposes, and these aggregate purposes are investigated further. The data was then transformed using the revised purposes, ready for clustering.

The aim of the paper is to explore Melbourne's travel patterns. As such, there is no methodological advance in this work. The key methodology is described in Jiang et al. (2012), who used k-means clustering to analyse Chicago's travel patterns.

3.1 Data

For this paper, data collected by the Department of Transport, Planning and Local Infrastructure, Victoria was used. The Victorian Integrated Survey of Travel and Activity (VISTA) is an ongoing data collection, with previous collection rounds in 2007/08 and 2009/10, and a current collection due for release next year. For this paper the 2009/10 round is used. This dataset comprises 41,626 individuals in 16,269 households located in Melbourne, Geelong and regional centres (Ballarat, Bendigo, Shepparton, and Latrobe). Of these, 30,283 individuals were surveyed on a weekday and 11,343 on a weekend.

Households are approached to be part of the survey, and subsequently complete a questionnaire regarding all of their travel movements for one day. They are asked to provide the start and end time of each trip, the end-purpose of the trip, how long they spent at their destination, the mode they took, and whether they travelled with anyone else. Socio-demographic data is also collected at the household and person level, comprising the LGA location of the household, car ownership, type of household (e.g., single person, couple, family) and age/gender of each household member.

A non-identifiable subset of the data for 2007/08 and 2009/10 is publicly available from http://www.transport.vic.gov.au/research/statistics/victorian-integrated-survey-of-travel-and-activity. This paper makes use of publicly available variables only. For example, this dataset notes whether travel takes place on a weekday or weekend only – there is no further differentiation between days of the week.

3.2 Purposes

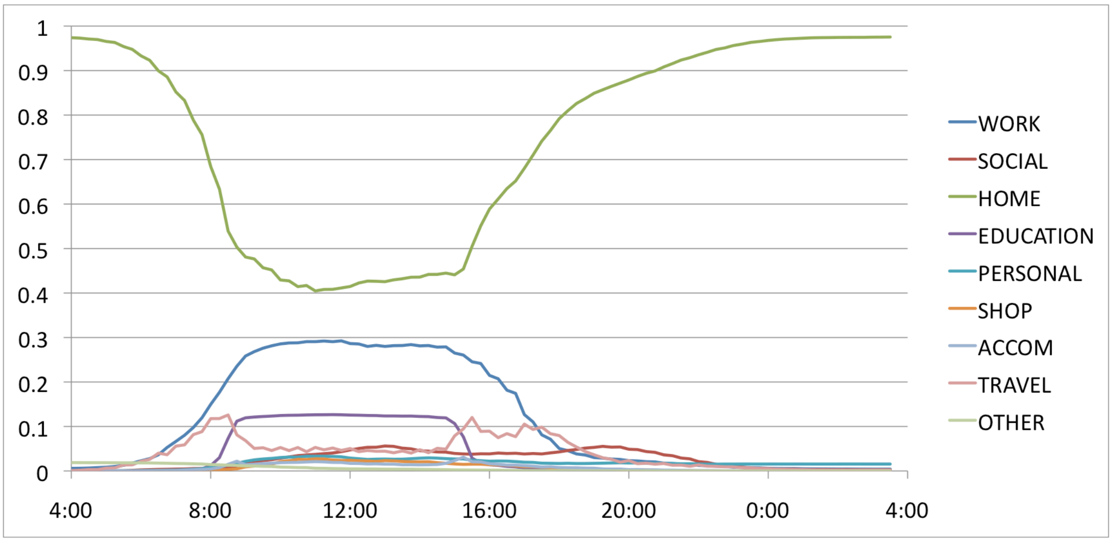
The VISTA dataset uses thirteen purposes for activities. As some of these are similar and to reduce the chance of over-clustering, these were combined into nine purposes for our clustering algorithm (as shown in Table 1).

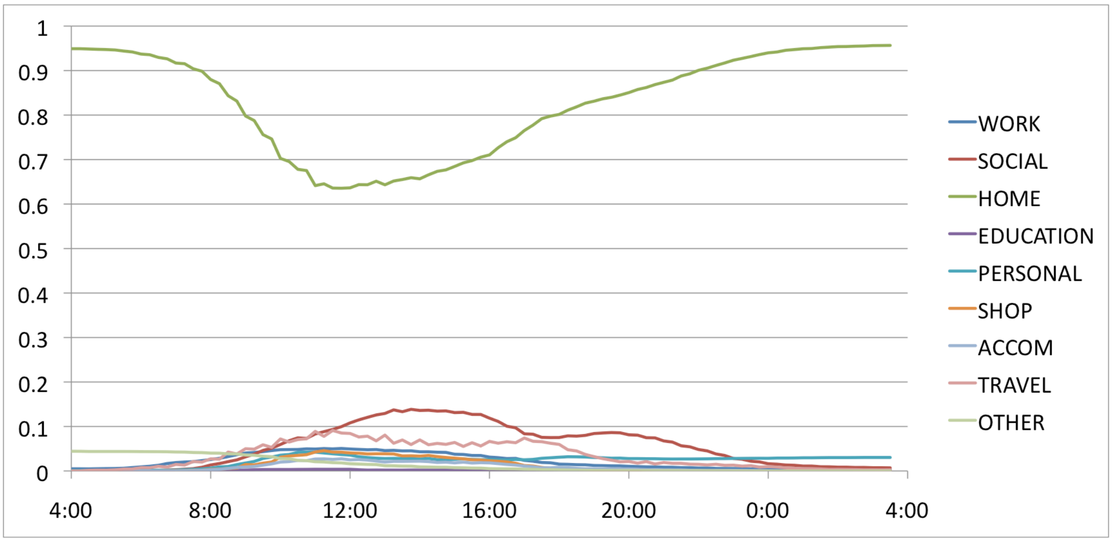
Table 1 Correspondence between VISTA purposes and clustering purposes

	VISTA purposes	Clustering purposes
1	Go home	Home
2	Work Related	Work
3	Education	Education
4	Social, Recreational	Social
5	Personal Business, Pick-up or Deliver Something	Personal business
6	Buy Something	Shopping
7	Pick-up or Drop-off Someone, Accompany Someone	Accompaniment
8	Change Mode	Travel
9	Other, Unknown	Other

Using these purposes, we can look at the activities being undertaken across Melbourne and regional centres for a weekday and weekend (Figure 1). Note a similar chart was produced in the VISTA summary of 2007/08 data looking at weekday activities (VISTA 2008) – these charts are for 2009/10 and are not weighted.

Figure 1 Weekday (above) and weekend (below) activities in Melbourne and Victoria





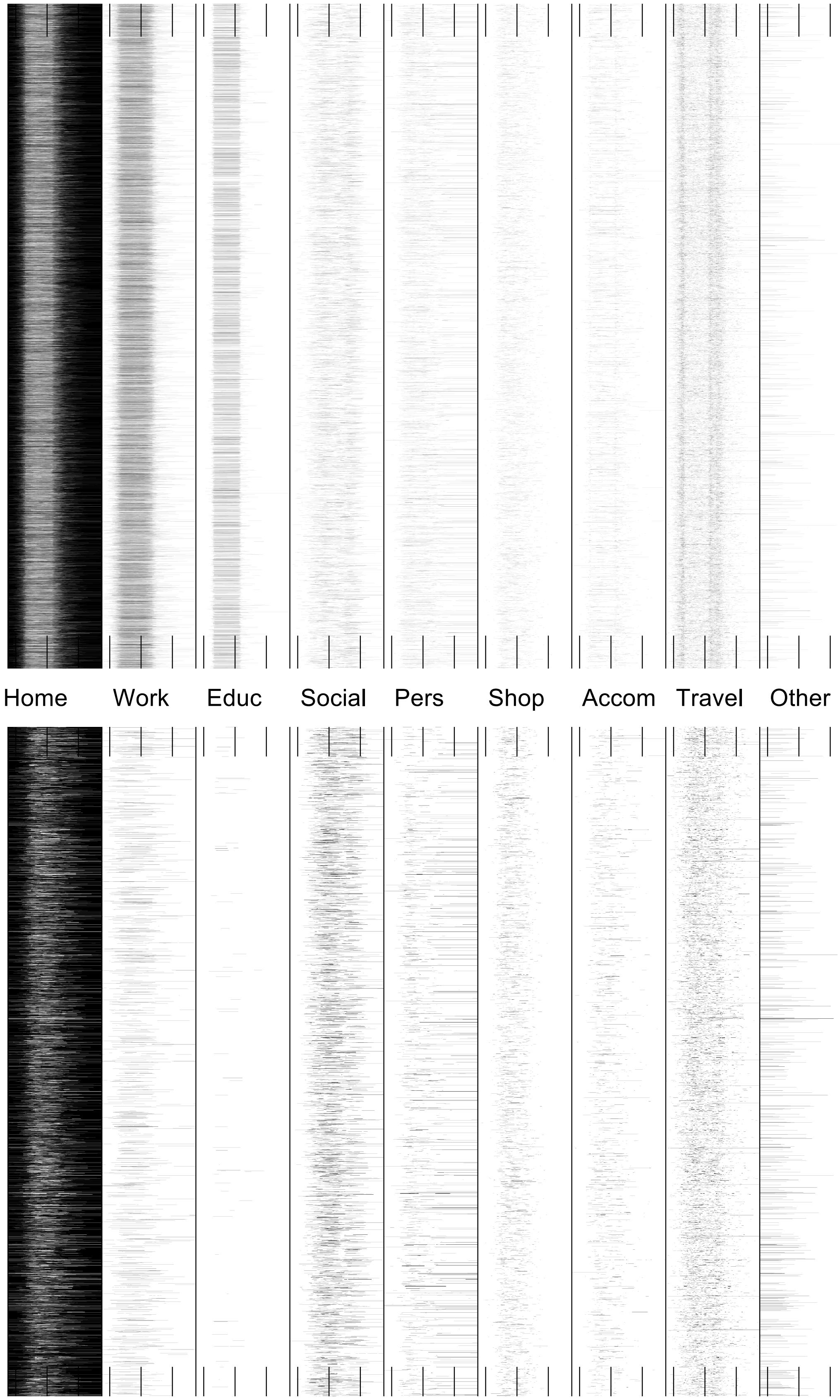
3.3 Methodology

K-means clustering is a way to cluster data to look for patterns. For this work, the open source package ELKI (Achtert et al., 2012) was used to perform the clustering.

In order to cluster, we transform the dataset so that each person has a string of zeros and ones representing their travel and activities for the day (from 4am – 4am). The activities were divided into nine categories across 15 minute intervals. This means that for each person surveyed, the transformed data set contained a string of 864 values (9 categories × 24 hours × 4 15-minute intervals), e.g., 96 integers representing whether the person was home (1) or not (0) during the day, 96 integers representing whether the person was at work or not, and so on. Note that no socio-demographic data is included in this transformed data set, nor the actual locations of where the activity took place (neither spatial (e.g., Melbourne CBD) nor categorical (e.g., workplace)) – only the activities and times during the day.

The transformed data for both weekday and weekend are shown in Figure 2. This shows the daily activities recorded for each person in the survey: each row represents a person's activity string in 15 minutes intervals from 4am to 4am. On the weekdays there is a clear work/education time period, as well as defined AM and PM travel peaks. On the weekend social activities are more prominent.

Figure 2 The transformed data for weekday (above) and weekend (below). Note ticks are placed at 6:00, 14:00, and 22:00 for each purpose.



The clustering process takes an input value for the number of clusters, and processes each input from the dataset to determine which cluster it fits best. This is done by calculating the "distance" between the mean of the cluster as it currently stands and the input. The input will be assigned to the cluster that it is "closest" to.

Note that we have to provide the number of expected clusters to the program – this means that the program cannot tell us what the optimal number of clusters is for our dataset. We can resolve this by clustering across a range of values, and then evaluating the clusters to see if it is an appropriate output. Two indices – the Dunn index (Dunn, 1973) and the silhouette index (Rousseeuw, 1987) – help us achieve this by measuring how close the data points within each cluster are, and how far apart the clusters are from each other.

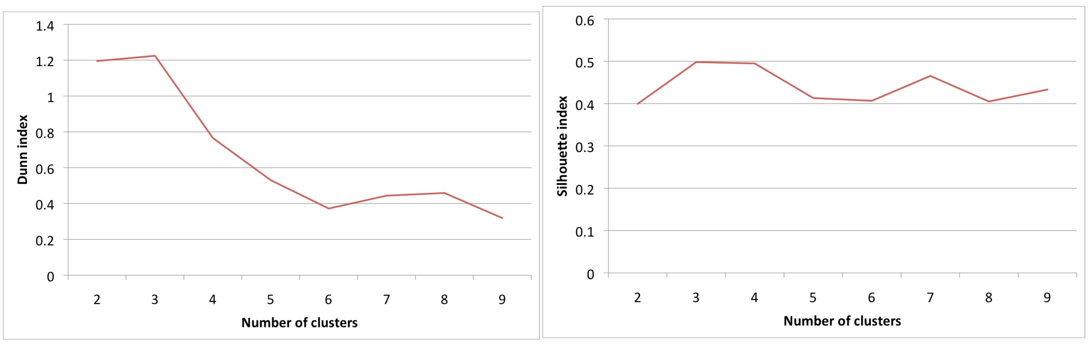
4. Results

In this section, the results for both weekday and weekend will be presented. Firstly, we determine the best number of clusters to analyse further. We then describe our new clusters, demonstrate how these clusters perform in comparison to our three base clusters of workers, students, and other, before identifying some pertinent socio-demographic differences between clusters.

4.1 Weekday

For a weekday, the Dunn and silhouette indices (see Figure 3) indicate that three clusters is optimal, however the silhouette index indicates that this is a very small improvement over four clusters. As we want to explore beyond the three clusters that we already know about, the next best clustering is seven.

Figure 3 Dunn and silhouette indices for weekdays.



In order to explore the characteristics of each cluster, we can firstly explore the crossover between three and seven clusters.

Table 2 The correspondence between three and seven clusters

Cluster	Count	Overall	Other	Workers	Students
Overall			52.21%	30.07%	17.72%
1	1364	4.50%	74.56%	17.08%	8.36%
2	3685	12.17%	0.00%	0.00%	100.00%
3	8468	27.96%	0.00%	99.99%	0.01%

4	2106	6.95%	72.22%	4.27%	23.50%
5	10911	36.03%	100.00%	0.00%	0.00%
6	522	1.72%	12.64%	21.84%	65.52%
7	3227	10.66%	71.12%	6.29%	22.59%

From Table 2 we can see that the three clusters lead us to describe 30% of the sample as workers, 17.72% as students, and 52.21% as other. This also corresponds to Jiang et al. (2012) who determined 53.90% of their sample to not be workers or students. This is a large amount to simply dismiss when considering activity patterns. It is clear that in our expanded set of clusters, cluster #3 is daytime workers, and cluster #2 is students.

The role of the other clusters can be clarified by looking at how different each cluster is to the sample overall. Looking at gender, age, and main work/study activity, we can start to put together a picture of the clusters. Table 3 shows the breakdown of each cluster.

Table 3 Properties of weekday clusters vs. socio-demographic variables. The asterisks denote if there is significant different with the sample (= 1%, *-5% level, proportion test)**

	Sample / Cluster	1	2	3
Female	51.55%	51.17%	50.18%	43.32%**
Young (18-35)	20.56%	32.26%**	9.69%**	30.15%**
Middle-aged (35-55)	37.17%	42.30%**	1.52%**	61.94%**
Older (55+)	19.55%	16.64%**	0.11%**	7.57%**
Student	21.11%	17.01%**	97.45%**	1.59%**
Keeping house	5.44%	4.91%	0.03%**	0.43%**
Retired	14.29%	9.31%**	0.03%**	0.07%**
Full-time work	34.47%	37.54%*	0.84%**	81.24%**

	4	5	6	7
Female	59.21%**	54.68%**	51.34%	59.34%**
Young (18-35)	18.42%*	17.99%**	25.10%*	12.24%**
Middle-aged (35-55)	32.38%**	31.56%**	34.10%	33.25%**
Older (55+)	31.10%**	31.46%**	14.56%**	27.36%**
Student	18.47%**	14.54%**	27.39%**	9.79%**
Keeping house	8.40%**	9.59%**	3.64%	9.33%**
Retired	24.31%**	26.75%**	9.77%**	22.06%**
Full-time work	17.76%**	18.53%**	32.95%	13.94%**

#1 Afternoon/evening workers/evening explorers: 4.5% of the sample, these people have a peak mean at work around 6pm. They are more likely younger, more likely to live in inner Melbourne, and more likely to be in a household with 2+ cars.

#2 Students: this cluster is most likely to be children who have a regular school day from roughly 9am-3pm. They are less likely to live in inner Melbourne, and are clearly most likely to be from a family household. They comprise 12.17% of the sample.

#3 Regular workers: this cluster follow more regular working hours, and are more likely to be travelling in the traditional AM and PM peaks. They are more likely younger or middle-aged, and more likely to be in a household with 2+ cars. They are just under 28% of the sample, the second-largest cluster.

#4 Afternoon explorers: almost half of this cluster are engaged in a social activity at 1pm. Some of them work in the afternoon, and some undertake personal activities with a peak around 3:30pm. These people are more likely female, are keeping house or retired, and are from a single/couple household. This cluster consists of just under 7% of the sample.

#5 Stay-at-home: this cluster spends most of the day at home and is the largest cluster with 36.03% of the sample. Most are keeping house/retired, however some may also be work-from-home. This cluster were more likely to live in a no- or one-car household.

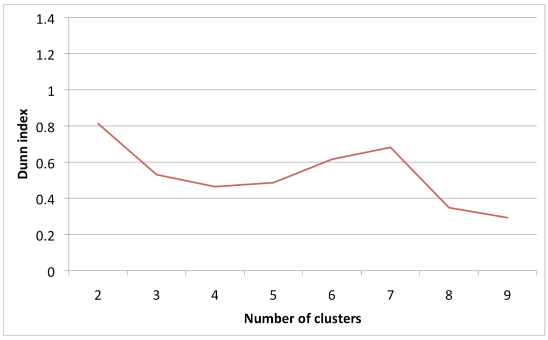
#6 Overnight adventurers: this cluster consists of those who did not return home at the end of the day (which is 4am in our data). They are more likely to be young students, and also live outside Melbourne. It is the smallest cluster at 1.72% of the sample.

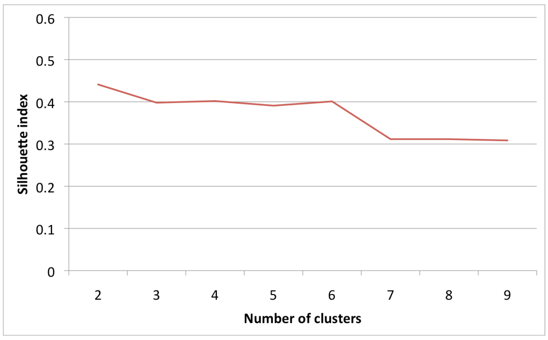
#7 Morning workers/explorers: 10.66% of the sample were identified as morning workers/adventurers. Most have left work or finished their social activities by 1pm. They are more likely to be female, older, and keeping house or retired.

4.2 Weekend

For a weekend, the Dunn and silhouette indices (Figure 4) indicate that two clusters is optimal, as opposed to three. As this will not give us a particularly good analysis, the next best option is six clusters.

Figure 4 Dunn and silhouette indices for the weekend





It appears that two clusters is better for the weekend than three – the influence of workers is not as dominant.

Again, we can see how the expanded set of clusters compares with two clusters (Table 4).

Table 4 The correspondence between two and six clusters

Cluster	Count	Overall	At-home/morning out	Out-and-about
Overall			69.58%	30.42%
1	1526	13.45%	13.37%	86.63%
2	1817	16.02%	53.00%	47.00%
3	383	3.38%	9.66%	90.34%
4	319	2.81%	0.63%	99.37%
5	6314	55.66%	99.97%	0.03%
6	984	8.67%	38.11%	61.89%

Staying at home again seems to be the dominant cluster, which corresponds to cluster #5 in the expanded set. The other clusters can be described as follows:

#1 Daytime explorers: the majority of this cluster (13.45% of the sample) is engaged in a social activity at 2:30pm, signifying a long lunch or shopping expedition. They are more likely to be from family households.

#2 Morning workers/explorers: this cluster are mostly involved in work in the morning, however could also be engaged in other activities. They are generally on their way home by the early afternoon. This cluster is most likely to be students, possibly engaged in school sports, or middle-aged.

#3 Overnight adventurers (returning home) and #4 Overnight adventurers (leaving home): due to the way that the data is transformed, not being home at the start of the day and not being home at the end of the day are classified differently. Cluster #3 also includes those who did not give their activity at the start of the day. As these two clusters together comprise just over 6% of the sample, it is not a huge amount, however may require more investigation of the underlying data.

#5 Stay-at-home: around 56% of the sample identify as staying at home, or undertaking some shorter activities. These people are more likely to be elderly.

#6 Evening explorers/workers: almost 9% of the sample are socialising at 7:30pm, although some are also working at this time. As the dataset is separated by weekend and weekday, it can be assumed that these evenings cover Saturday and Sunday evening, however most social activities would take place on Friday and Saturday evenings. This cluster is more likely to the younger, live in inner Melbourne and live close to train stations.

5. Discussion

More differences between regions were expected, however this was not apparent from the clustering. Some differences, in that those who live in inner Melbourne were more likely to be working or socialising in the evening, were seen, but these were rare. A more detailed investigation of, for example, work travelling times, may reveal differences.

There are limitations with the data set. The categorisation of days could be improved so that Friday and Saturday evenings could be explored in more detail. Some variables collected in the survey are missing from the publicly available data set, for example, income (at the household level) and occupation. How to deal with those who have not provided a start and end activity for their day also needs to be investigated further.

6. Conclusions and further work

This paper has used VISTA data to uncover activity-travel patterns in Victoria. Going beyond the usual classification of worker, student, and other, we have identified different types of workers and also adventurers. While these classifications fit neatly into the original classifications, it provides an indication of different types of travellers, as well as providing more detail for the 50+% of travellers who would otherwise be classified as other.

Combining socio-demographic variables with our clusters show that there is some difference in terms of home region: not only obvious differences such as students are more likely to come from family households, but that those with car access are more likely to go out during the evening.

In terms of processing power, if better resources come to hand, then the interval could be decreased to 5 minutes. This could improve our clustering outcomes, as shorter trips will not be excluded. Repeating the process with 2007/08 VISTA data could also show changes in behaviour over time. Most interesting though would be to replicate the process with data available in New South Wales and South East Queensland, to see if there are any differences between cities in Australia. Data for SEQ is readily available, and is already in a format amenable to processing and analysis; NSW data could be obtained with permission, however would require some effort to process.

While this is preliminary work, the scope of further exploration is broad. Understanding how people spend their time is of importance to planners and decision-makers, for example, in developing solutions to overcome social exclusion, influencing mode share, and managing congestion. This analysis could be used to check travel activity forecasts generated by models, as well as predict where people are likely to travel.

Acknowledgements

The author is grateful to Prof Leon Sterling for his support in undertaking this research, the Department of Transport, Planning and Local Infrastructure (DTPLI), Victoria for the provision of the data and their assistance, and the anonymous reviewers for their helpful comments.

References

Achtert, E, Goldhofer, S, Kriegel, H-P, Schubert, E, and Zimek, A (2012) Evaluation of Clusterings – Metrics and Visual Support *Proceedings of the 28th International Conference on Data Engineering (ICDE)*

Arentze, T and Timmermans, H (2000) *ALBATROSS : a learning based transportation oriented simulation system* Eindhoven, Netherlands: EIRASS

Davidson, B, Vovsha, P, and Freedman, J (2011) New advancements in activity-based models *Australasian Transport Research Forum 2011 Proceedings*

Dodson J, Burke M, Evans R, Gleeson B, and Sipe N (2010) Travel behavior patterns of different socially disadvantaged groups: Analysis of household travel survey data for a dispersed metropolitan area *Transportation Research Record: Journal of the Transportation Research Board* 2163, 24–31

Dunn, J C (1973) A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters *Journal of Cybernetics 3 (3)*, 32–57

Hagerstrand, T (1970) What about people in regional science? *Papers in Regional Science 24 (1)*, 6 – 21

Inbakaran C and Kroen A (2011) Travel surveys – review of international survey methods *Australasian Transport Research Forum 2011 Proceedings*

Jiang, S Ferreira, J and Gonzalez, M C (2012) Clustering daily patterns of human activities in the city *Data Mining and Knowledge Discovery 25*, 478–510

Kitamura, R (1988) An evaluation of activity-based travel analysis *Transportation 15 (1)*, 9 – 34

Miller H (2008) Transport 2.0: Meeting grand transportation challenges with geographic information science. *ArcNews, Winter 2008/2009*: http://www.esri.com/news/arcnews/winter0809articles/transport-20.html

Moiseeva A (2013) *Experience the City: Analysis of Space-Time Behaviour and Spatial Learning* PhD thesis, Eindhoven University of Technology

Neutens, T, Schwanen, T and Witlox, F (2011) The prism of everyday life: Towards a new research agenda for time geography *Transport Reviews 31 (1)*, 25–47

Rousseeuw, P J (1987) Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis *Computational and Applied Mathematics 20*, 53–65

Ryley T (2006) Use of non-motorised modes and life stage in Edinburgh *Journal of Transport Geography 14(5)* 367 – 375

VISTA (2008) 2007-08 Summary of survey results Available from http://www.transport.vic.gov.au/__data/assets/pdf_file/0014/31280/VISTA-07-Summary-Brochure.pdf, last accessed 12 June 2013