Innovative spatial analysis for comparing Mashhad in Iran with Melbourne in Australia

Ray Wyatt[1], **Mohammad Rahim Rahnama**[2]

ray.wyatt@unimelb.edu.au
rahnama@ferdowsi.um.ac.ir

Abstract: We demonstrate an approach that has considerable potential for more effectively analysing urban structures. We focus upon the different neighbourhoods' different levels of liveability which, we argue, can be encapsulated using two types of statistic - "opportunity indicators" and "success indicators". For opportunity indicators we mapped different zones' "accessibility to shops" and "accessibility to jobs", and for success indicators we mapped "average income", "percentage of adults employed", "tree cover" and "lack of traffic"'. We then used what we claim to be an advance upon normal clustering methods – neural clustering, in order to cluster zones into groups that are similar. Finally, we mapped such clusters using an innovative, graphical-communication method known as face charts. By closely inspecting the maps of clustered face charts, rather than using more conventional methods, one gains more insight into the similarities and differences between our two cities.

Introduction

In geography, comparing places has a long tradition, and the best studies are those which address copious details into which the reader can delve and so appreciate how each locality is structured. The downside of such studies, however, is that readers risk becoming so submerged within this detail that they fail to gain any sort of insightful understanding. Hence in order that readers do not

[1] Department of Resource Management & Geography,University of Melbourne, Australia

[2] Department of Geography,Mashhad Ferdowsi University, Iran

lose their focus, it is often more productive to simply present the important data. But how does one determine which data is important?

The answer to this question will always depend upon why one wants to compare the places. For example, if one wants to compare them in terms of environmental sustainability, then data about environmental attributes will obviously be more important than data in the form of, say, commercial attributes. In other words, the type of focus that one adopts is what dictates which data is important.

In this paper we compare two very different cities – Mashhad in north-eastern Iran and Melbourne in south-eastern Australia, and we have unilaterally decided that our reason for doing this is to compare the daily lives on citizens living within the different neighbourhoods of each place. It then follows that one type of important data will be people's relative access to facilities – the higher the level of access the more one is, in theory, able to live comfortably. Here we label such data as "opportunity indicators". The other type of data that impacts upon residents' quality of life is that indicating their neighbourhood's level of success. If one lives where people have made sure that they are well provided for, then the chances of that neighbourhood having a high level of amenity are strong. We label this kind of data "success indicators".

Before proceeding, readers should note the massive difference in scale between Mashhad and Melbourne. Melbourne's population is about ۵۰٪ more than Mashhad's, but because it sprawls so much its total area is several times greater than Mashhad's. To emphasise this, the two cities' sizes, at the same map scale, are shown in Figure ۱. It follows that in this paper, the maps of Mashhad will be at a much more detailed scale than the maps of Melbourne.

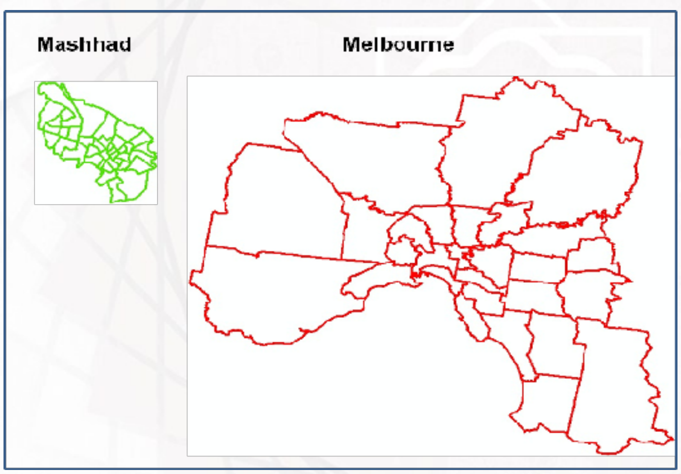


Figure ۱ – Spatial extent of Mashhad compared to that of Melbourne

Data used

As well as being driven by overall focus, data usage is also determined by availability. In this instance data about each zone's number of shops was available in both cities, in the form of a cartographic source in Melbourne (Marshall, ۲۰۱۱) and in the form of "area devoted to commercial purposes" in Mashhad (see Figure ۲). As such, our shopping-opportunities data was fairly approximate and not strictly comparable, but since our focus was on relative rather than absolute levels of accessibility this was not a major problem.

Another readily available indicator of opportunity, albeit available only in Melbourne, was the number of jobs located within each statistical zone, as recorded by the Australian "journey to work" census of ۲۰۰۱. Although this data is a little dated, jobs distributions are notoriously slow-changing and again, we were interested only in relative patterns and so this data sufficed for our purposes.

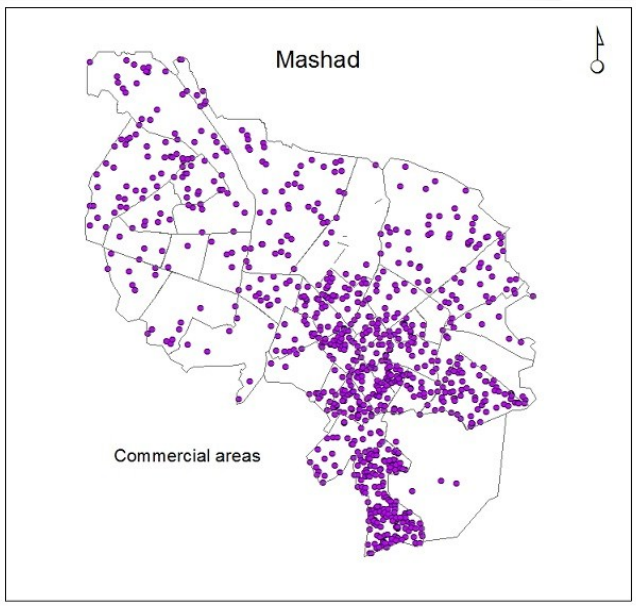


Figure ۲ – Approximate distribution of 'shops' in Mashhad

Note that it would be a huge mistake to approximate each neighbourhood's accessibility to shops and jobs by simply counting up how many of these are located there. Since people travel across the city to reach shops and jobs, total accessibility to such opportunities depends not only upon the local supply but also upon how many there are in other zones and how far the latter are away. That is, to accurately gauge each neighbourhood's true accessability to opportunities, area-specific data needs to be interpolated across the whole city to form some kind of continuous, accessibility surface.

Enter Geographic Information Systems (GIS). These allow current researchers to construct such a surface with comparative ease, and so most of them do just that. The problem is, however, that many do it with scant regard for the number of arbitrary assumptions that are implicit within the method that they use. It is as if modern technology's power tempts them to apply it in an unthinking way, and this brings with it the danger of false conclusions being drawn.

For instance, those who interpolate zone data from zones' centroids might choose to use kriging, inverse distance weighting, rubber sheeting or whatever, but when each of these approaches generates a different surface, it is very difficult to judge which one constitutes the true picture. Similarly, those who use GIS to calculate some sort of Hansen accessibility measure for each neighbourhood are left with the problem of how to determine a suitable exponent for the distance factor. Short of conducting a full behavioural survey to calibrate one, it is impossible to say which value will be the most valid.

In this paper, therefore, we revert to a method that was first proposed as far back as ١٩٢٦ (see Yuill, ١٩٧١), more than half a century before GIS appeared – the standard deviational ellipse. We use one to approximate the spread of the opportunities and then, in order to approximate its accessibility to the total distribution, we simply measure each zone's distance to this ellipse's centroid and its two foci. The sum of such distances is likely to be every bit as accurate as all manner of sophisticated, accessibility measures based upon dangerously arbitrary and often unrealised assumptions.

Our logic here is that the simplest and least risky way to measure any zone's accessibility would be to calculate its straight-line distance to the distribution's centre of gravity. But this brings with it the implicit assumption that opportunities are spread radially outwards from the centroid, which is seldom true in reality.

The standard deviational ellipse, therefore, better describes most distributions than a circle does. It is optimally oriented to encompass around ٦٧٪ of the distribution, provided that the latter is roughly Gaussian in nature, and the relevant ellipse for shops in Mashhad is shown on the left of Figure ٣. Our ellipses were actually calculated using a GIS, and although our particular package (*ESRI*) does not plot ellipses' foci, these are easy enough to approximate, as shown.

Each zone's distance to the centroid *and* to the ellipse's two foci is surely a more accurate approximation of its accessibility than is its distance to the centroid only, and the result - each zone's estimated 'access to shops', is shown on the right of Figure ٣.

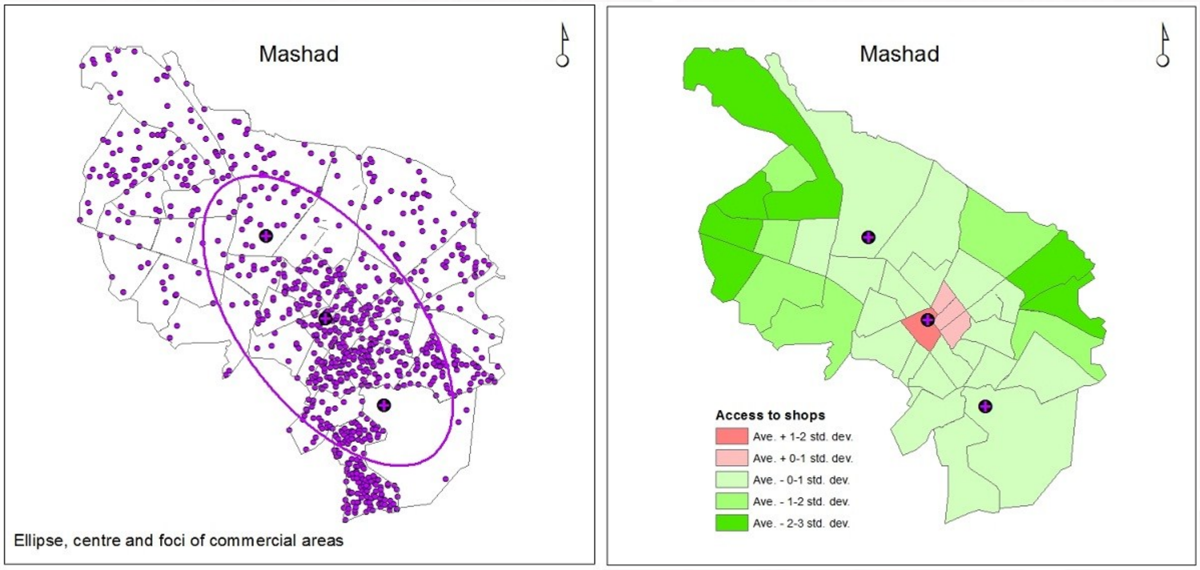
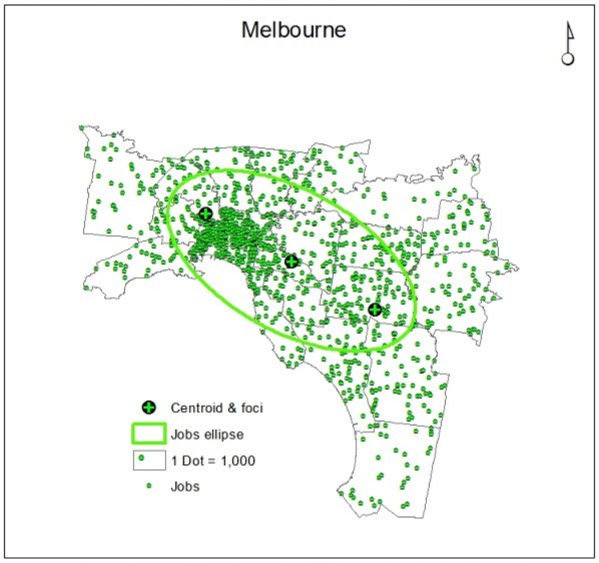


Figure ٣ – Standard deviational ellipse for shops in Mashhad (left), and zones' estimated access to shops (right)

Note that in Melbourne we might have used a more sophisticated method to measure each zone's accessibility to jobs. This is because as well as having data on the spatial distribution of jobs, we also had data on the spatial distribution of the residents holding those jobs – both are shown in Figure ٤; and one's chances of being employed depend not only upon one's distance from jobs but also upon the number of competing workers who live closer to those jobs.

Hence one might use a GIS to buffer around the centroid and the two elliptical foci of the jobs distribution, and for each band assign the number of competing workers who live closer to the jobs distribution than this band is. However, this is quite a challenging GIS operation to perform, and due to time limitations it was not carried out for this paper. It remains a topic for further work, and here we simply found each zone's distance to the centroid and to the two foci of the jobs distribution – in exactly the same way that we calculated accessibility to shops in Mashhad.



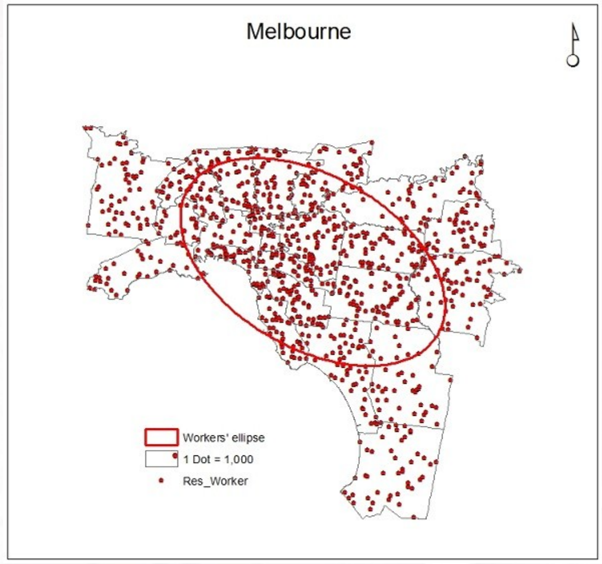


Figure ٤ – Distribution of jobs and workers' residences in MelbourneIn

terms of "success" data, in Mashhad we had available two measures – "average income" and "percentage of the working age population (١٥ to ٦٤ years) that is employed". In Melbourne these two data sets were also available, and in addition we used two other measures – "tree cover" and "traffic absence". The income and employment statistics were taken from the demographic Census, and the data on trees and absence of road traffic was, again, cartographically estimated from the same source that was used to estimate access to shops (Marshall, ٢٠١١).

Mapping the data

It proved possible, therefore, to investigate each city's urban structure, using three variables in Mashhad and six in Melbourne. The results for Mashhad are shown in Figure ٥ and those for Melbourne are shown in Figure ٦. All maps use the same colours for zones' scores in terms of the latter's number of standard deviations from the average score.

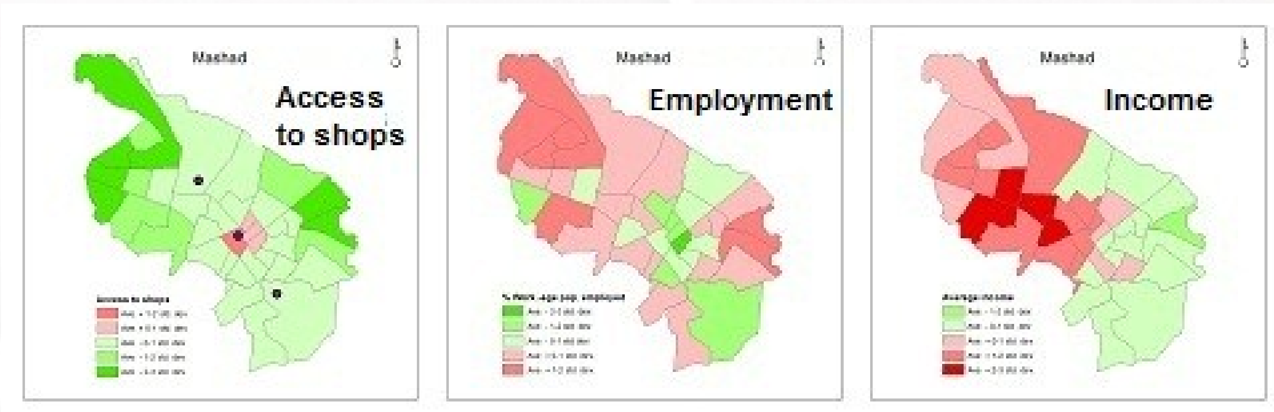


Figure ۵ – Patterns in Mashhad

For Mashhad, the picture is reasonably clear, chiefly because we only need to note three variables. Areas of high shops access are located around the central area, with access dropping off more rapidly towards the northwest and towards the southwest. Also, there is a belt of lower employment through the middle of the city just east of the central, affluent neighbourhoods, and the zones with the highest employment rates are at the edges of the city. Moreover, there is a definite pattern of lower incomes in the east and higher incomes in the west.

For Melbourne with its six indicators the picture is more complicated. Yet one can still see a number of things, such as high accessibility to shops being in central areas and running northwest to southeast, just like in Mashhad. Moreover, the areas with the highest employment rates are in the inner- and outer-eastern suburbs, and the highest income levels are immediately south and southeast of the central area, with low income zones located in the industrial outskirts both towards the northwest and the southeast.

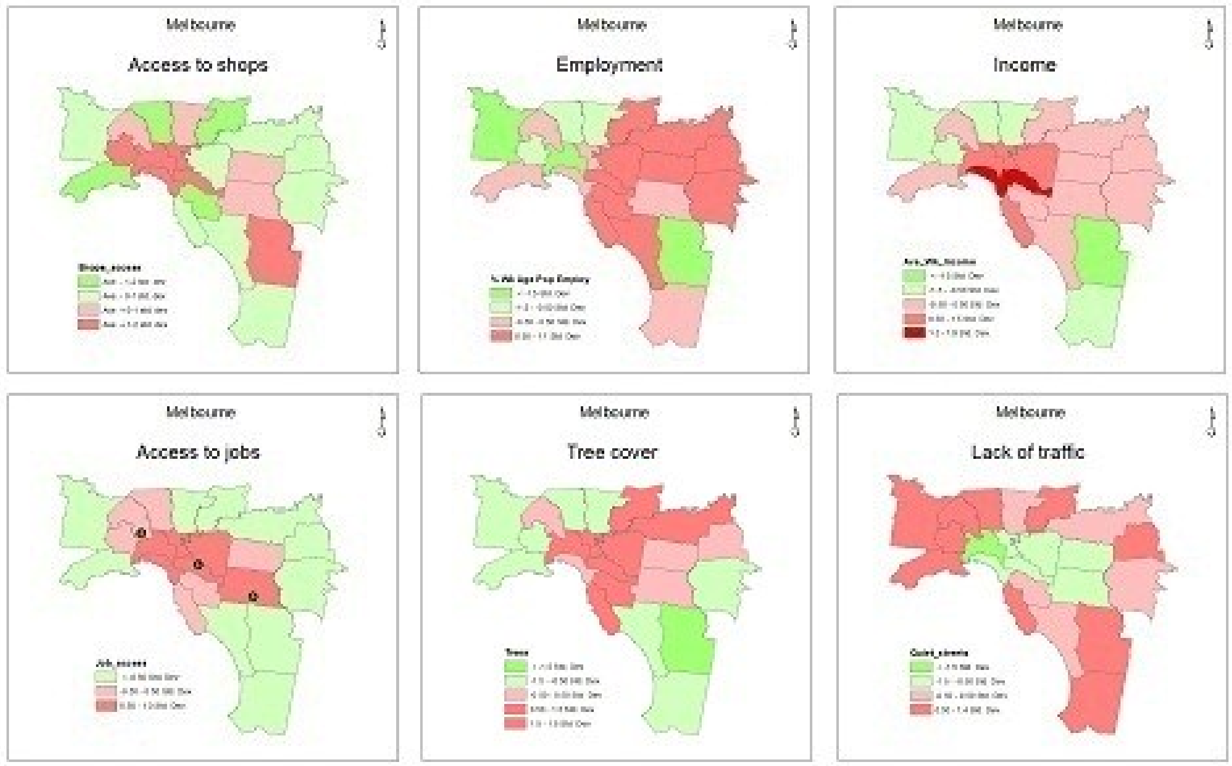


Figure ٦ – Patterns in Melbourne

Also, high accessibility to job opportunities runs through the central area with a northwest to southeast orientation, with levels dropping off more quickly towards the northeast and the southwest, which again, is similar to the pattern of access to shops in Mashhad. Finally and somewhat surprisingly, tree cover is greatest in the established areas of the central and inner eastern and southern suburbs, along with some suburbs on the north-eastern fringe. Lack of traffic is almost a mirror image of this, with outer areas scoring much more strongly than middle- and inner-ring suburbs.

Yet these coloured maps, although they are separately useful, are difficult to keep in mind all at once. This makes it hard, or even impossible, to appreciate the way in which they are blended within each zone. That is, holistic appreciation of the aggregated nature of each zone's amenity is elusive. So in order to really understand each city's morphology we need to employ some kind of summarizing, multivariate technique. Accordingly, we employed cluster analysis, as described in the next section.

Clustering the data

It needs to be noted that we did not use conventional, aggregative cluster analysis. The problem with the latter is that it begins by joining the most similar pair of entities, then the next most similar pair of entities/groups of entities, and so on until there are only a few clusters left. This means that clusters are formed no matter what – even when there is no strong clustering tendency within the data set. As such, the final clusters could well contain a diverse range of entities, which somewhat defeats the original purpose of cluster analysis - to formulate groups so that between-groups heterogeneity and within-groups homogeneity are both maximised.

Accordingly, one of the authors wrote an *Excel macro* program that employs a different clustering method – Kohonnen clustering based on the principles of the self-organising map (Haykin, ۱۹۹۷). This approach sets up an artificial, neural network that gradually assesses whether or not there is clustering apparent within the data, and it only generates clusters if there is. Expressed the other way around, if this method finds that there is no evidence of clustering, then it will simply not come up with any clusters.

The way it works is as follows. Every multi-dimensional entity, such as a zone with six indicator scores, is expressed as a string of numbers that have been standardised to values between zero and unity. Such vectors are then compared, one by one and several times, to say, ۱۰۰ destination cells that each contain, in this case, six random numbers between zero and unity. When the cell that most closely corresponds to each vector is found, that entity is then allocated to that destination cell. But this happens only after the numbers in the winning cell have been altered, very slightly, so that they more closely approximate the vector's numbers, and its neighbouring cells are likewise altered, albeit to a lesser degree. Such alterations make it more likely that this vector, and vectors that are similar to it, will be allocated to this cell, or to its neighbours, during the next pass of the data through the program.

Hence after the complete data set has been passed through the allocation process many times, the weights within any particular region of the destination matrix's cells are gradually massaged to approximate vectors that are similar to each other, and the latter are located there. That is, entities/vectors that are eventually assigned to a similar section of the destination matrix's cells are designated to be a cluster.

Obviously, if all entities are vastly different, the destination matrix will not organise itself into differentiated regions, and so there will be no clustering of entities. Conversely, whenever a group of vectors ends up in the same region of the destination matrix, one can be sure that they really do constitute a valid cluster of similar entities.

Yet the downside of this rather clever technique is that the user needs to experiment with various adjustments of parameters, such as those that determine just how much a destination cells' numbers should be altered whenever a vector lands in it, and those that indicate how many neighbouring cells should be altered and by how much. Moreover, since the method starts off with random numbers assigned to each destination cell, it often gives different results for the same set of data inputs. This is common with neural network-based methods, and it necessitates running the program many times and eventually adopting an “average looking” clustering result.

This was done with both the Mashhad and Melbourne zones/entities. The results are shown in Figure ٧. These two maps are certainly simpler than the multiple maps shown above in Figures ٥ and ٦. Yet they only have meaning if one remembers some of the details learned from Figures ٥ and ٦. If one cannot remember such details, then Figure ٧’s maps are nothing more than colours for different clusters of zones. One needs background knowledge to interpret the colours.

For example, the brown coloured cluster in Mashhad roughly corresponds to the broad swath of low-income zones located just east of the central area plus some poorer areas on the western fringe. Moreover, the brown coloured cluster in Melbourne comprises two municipalities with heavy tree cover, high employment rates and high incomes.

Yet remembering all of the details of Figures ٥ and ٦ can frequently be impossibly difficult. The result is that readers usually remember only some of the details. Moreover, different people remember different details, and so different people will interpret Figure ٧ in different ways.

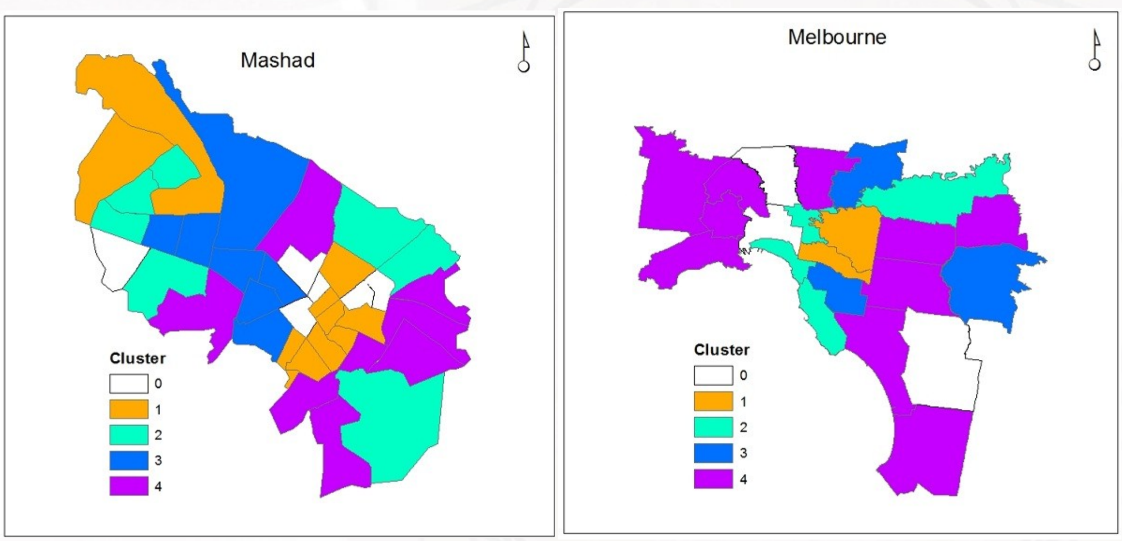


Figure ٧ **–** Simplified urban morphologies of Mashhad and Melbourne

In fact, this tendency for people to have selective memory for details is well known in psychology, and it is the chief reason why a simple and methodical method, like a goals-achievement matrix, always outperforms acknowledged world experts in tasks such as diagnosing diseases, predicting firms' business success and foreshadowing students' academic performances. The human experts always place undue importance upon those indicators which their experience has convinced them are the more important ones, and they pay less heed to other indicators (Dawes, ١٩٧٩; ١٩٨٦).

What we need, therefore, is some method that is in between the impossibly detailed, multiple maps of Figures ٥ and ٦ on the one hand, and the over-simplified, singular maps of Figure ٧ on the other. We need a method which shows the clusters of zones, but which also details zones' characteristics upon which such clustering has been based.

Accordingly, we here present an innovative way of showing, on the same map, both the clusters and the nature of each zone within each cluster. This method is known as “face charts”. It was developed by one of the authors as an effective means for holistically summing up the nature of multi-faceted entities (Wyatt, ٢٠٠٨), and it is described in the next section.

Face charting the data

The logic behind face charts is that the human brain has not evolved to enable it to quickly and effectively absorb the messages implicit within tables of numbers or coloured maps. But humans have evolved to take in the messages conveyed by the human face, almost instantly and usually very accurately. It would seem useful, therefore, to utilize images of faces as a receptacle for communicating complicated data. The latter's holistic messages would at least be assimilated faster than if the same data were conveyed using a bar chart, histogram or star plot (Wyatt, ٢٠١٢).

This idea has been around for some time – ever since Chernoff (١٩٧٣) used face diagrams to convey complicated, multi-dimensional data entities, and he has had many imitators since (eg Dorling, ٢٠١١). According to the authors, however, such efforts all run foul of the human tendency to interpret faces in terms of age, sex, race or emotional state.

That is, people cannot help seeing some face diagrams as male and others as female; or seeing some as Caucasian and others as Negroid, Asian or whatever; or as old rather than young; or as happy, sad, frightened, surprised or whatever. Such interpretations tend to cloud the method's ability to act as a coldly objective communicator of complicated data.

Accordingly, the face diagrams used here have been so standardised that they cannot be interpreted in terms of age, sex, race or emotion. This has been achieved by making every facial feature a standard, circular shape. To distinguish such images from “face diagrams” they are designated by the more objective-sounding term of “face charts”.

Figures ٨ and ٩ show screen dumps, one for Mashhad and one for Melbourne, generated by a face charts-drawing program written by one of the authors in the *Visual Basic* computer programming language. In this program a face chart is used to show each zone's totality of scores – three scores for each zone in Mashhad and six for each zone in Melbourne.

Specifically, the size of the eyes is proportional to the average income of people in that zone; the size of the nose corresponds to that zone's accessibility to shops and the size of the mouth indicates that zone's employment rate. Moreover, in the Melbourne map each zone's accessibility to jobs is indicated by the size of its face chart's hat; the zone's extent of tree cover is proportional to the size of the ears, and its absence of traffic is indicated by the size of its neck. Finally, to make things clearer, whenever a score is in the top third of all scores its facial feature is automatically coloured green and whenever it is in the bottom third its facial feature is coloured red.

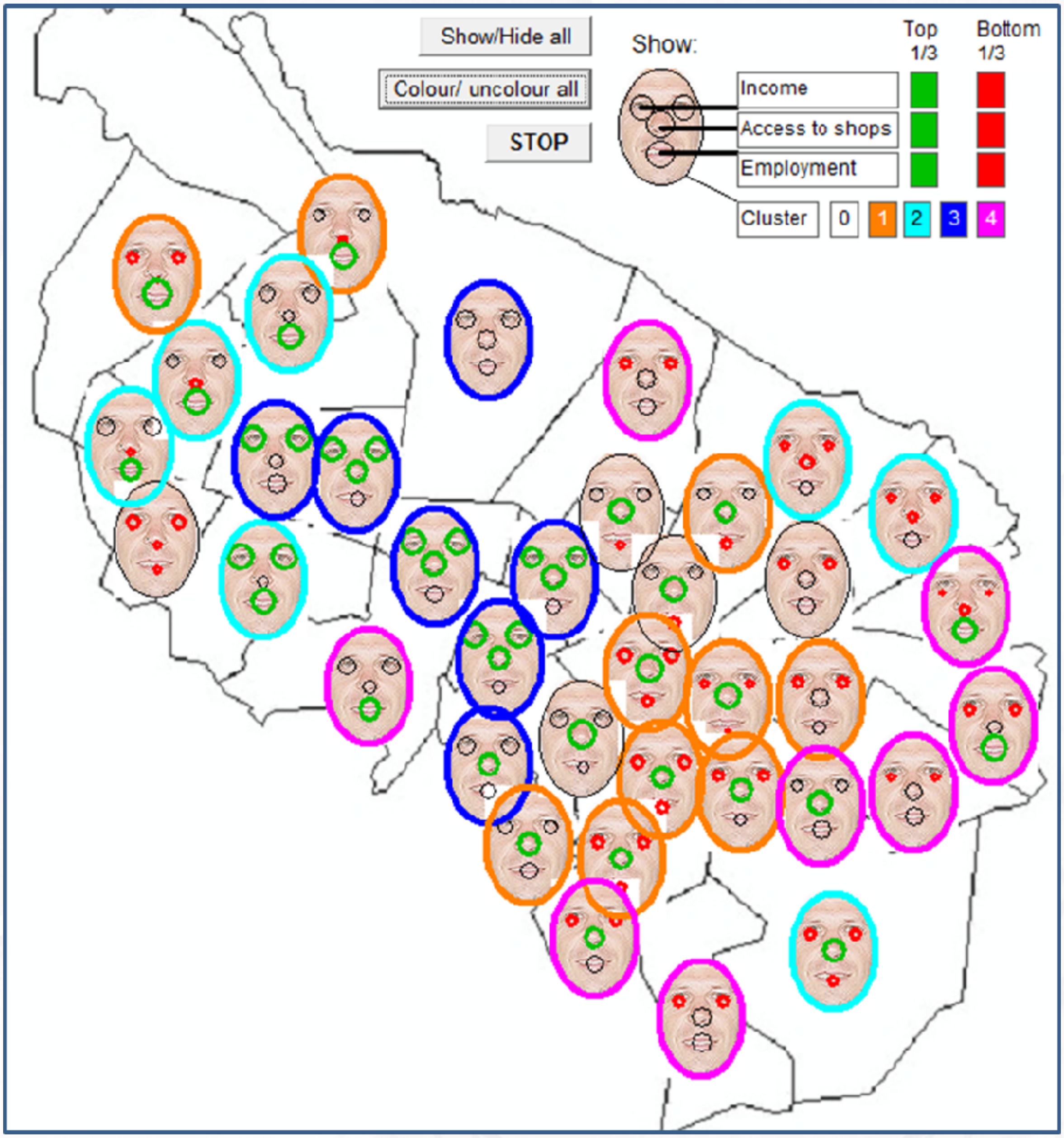


Figure ۸ – Morphology of Mashhad using face charts

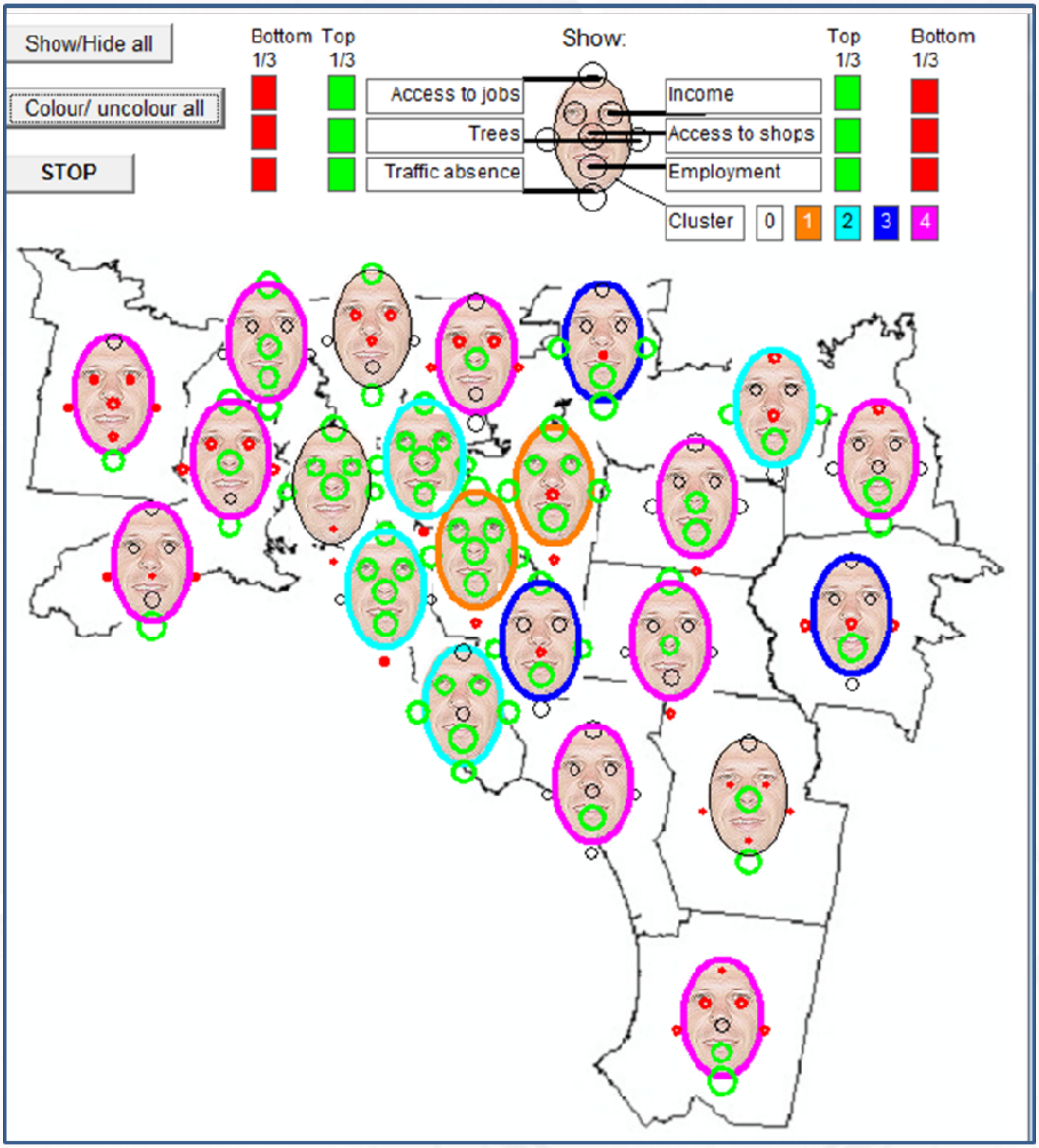


Figure ۹ **–** Morphology of Melbourne using face charts

Such face charts allow one to instantly see why the clusters came out in the way that they did. For example, in the Mashhad map the faces of the zones in the brown cluster all have small eyes, and so this cluster obviously comprises low-income zones. Conversely, the dark blue cluster comprises zones of high income and usually high accessibility to shops, as shown by their large eyes and usually large noses respectively. Moreover, the purple cluster comprises zones that are low-income and mostly remote from shops, as shown by all the face charts' small eyes and noses, and the light blue cluster is somewhat mixed, suggesting that it may have split into two clusters had the program been left to run a little longer, although its zones' generally small noses indicate a remoteness from shopping opportunities.

In short, zones in the different Mashhad clusters can be described as follows:

- dark blue cluster (۳) - *high* income, mostly *high* shop access, *low* employment;
- brown cluster (۱) – *low* income, *high* shop access, *low* employment;
- light blue cluster (۲) – *low* income, mostly *low* shop access, -;
- purple cluster (٤) – *low* income, -, -.

It should be noted that since there are only three variables on which to base clustering in Mashhad, its clusters are bound to be somewhat approximate and preliminary. If more variables were used, the clustering may well have become firmer and more reliable. Also, three zones did not really fit into a cluster at all – the white ones shown in Figure ٥ above, which, incidentally, uses the same colours for equivalent clusters in Figures ٨ and ٩.

Note that in the Melbourne map there are also some zones which did not fit into any cluster, including the central-city municipality and two others at the suburban fringe, and zones in the Melbourne clusters can be summarised as follows:

- light blue cluster (۲) – *high* income, mostly *high* shop access, *high* employment, mostly *high* job access, mostly *good* trees, mostly *poor* traffic;
- brown cluster (۱) - *high* income, -, *high* employment, *high* job access, *good* trees, *poor* traffic;
- dark blue cluster (۳) - *low* income, *low* shop access, *high* employment, mostly *low* job access, mostly *good* trees, mostly *poor* traffic;
- purple cluster (٤) – *low* income, -, -, mostly *low* job access, *poor* trees, -.

Note that the dark blue cluster ۳ in Mashhad has similarities with the light blue cluster ۲ in Melbourne. They are similarly located close to the commercial centre, and their zones have high average incomes and high access to shops. The only difference is that in Mashhad these zones' employment rates are low whereas in Melbourne they are high.

Note also that it is not so much the zones in the different clusters that are of the most interest. More instructive questions are prompted by examining those zones which do not fit into any cluster. In this sense, the method we have presented acts similarly to a regression analysis in which the trend of the vast majority of entities is of some interest but the outliers are more interesting still.

For example, why is the un-clustered zone near the western edge of the Mashhad map in Figure ٨ not a member of the light blue cluster like its neighbouring zones are? It has low income and low access to shops, but all of the zones in the light blue cluster which also have low incomes and low access to shops actually have fairly high employment rates whereas it does not. As such, this zones goes against what we would normally expect to find in this section of Mashhad; its employment rate is unexpectedly low.

Likewise, why is the northernmost zone in the Melbourne map not a member, like its neighbours, of the purple cluster? There are two members of the latter which, like it does, have both high access to jobs and good scores for lack of traffic, but they both have high access to shops whereas this un-clustered zone does not. Again, this zone is an outlier in terms of what one would expect to find in this part of Melbourne. Its access to shops is lower than expected.

Conclusions

Apart from our assertion about the superiority of neural clustering over conventional, aggregative cluster analysis, our conclusions are twofold.

Firstly, we suggest that Figures ٨ and ٩ lead to many more insights and hypotheses because they are far clearer than both the detailed maps of Figures ٥ and ٦ and the cryptic maps of Figure ٧. They strike a balance between the data overload of Figures ٥ and ٦ on the one hand, and the over simplification of Figure ٧ on the other. These maps actually show all of the information from all previous maps, yet in a way that is nowhere near as cognitively demanding.

In addition, clarity is enhanced by these maps being the output of an interactive program. If we want, for example, a map showing only access to shops, everything can be turned off (erased) except for each zone's nose. Moreover, it is important to remember that relative circle sizes actually convey more information than do simplified, coloured maps like those in Figures ٥ and ٦.

Secondly, despite our two cities being half a world apart and vastly different in size, environment, culture, function and history, they have some similarities. For example, both appear to have a strong commercial centre and some suburbs near the central area are characterised by high levels of income and shops access. These are towards the west in Mashhad (cluster ۳) and towards the east in Melbourne (cluster ۲).

Moreover, in both places the areas with the highest employment rates are towards the city's outskirts both in the east and in the west, and in both cities there are several low-employment zones scattered around the fringe. Finally, Mashhad has a cluster of low-income and low-employment zones along the eastern border of its central zone (cluster ۱) while Melbourne has a similar, albeit slightly less differentiated area immediately to the west and northwest of its central municipality (part of cluster ٤).

Such results are in no way definitive. This paper has only sought to illustrate the potential of the innovative methods used. It concedes that a full and rigorous comparison of our two cities requires a much more comprehensive collection of statistics from both places. Nevertheless, once such statistics have been assembled, the methods described above will have very strong potential for prompting hitherto undiscovered insights into our two cities' urban morphologies.

References

Dorling, D. (۲۰۱۱) - http://www.sasi.group.shef.ac.uk/thesis/chapter۸.html, accessed in ۲۰۱۱.

Chernoff, H. (۱۹۷۳) - Using faces to represent points in k-dimensional space graphically, *Journal of the American Statistical Association,* ٦۸.

Dawes, R.M. (۱۹۷۹) - The robust beauty of improper linear models, *American Psychologist*, ۳٤, ٥۷۱-٥۸۲.

Dawes, R.M. (۱۹۸٦) - Proper and improper linear models, *International Journal of Forecasting*, ۲, ٥-۱٤.

Haykin, S. (۱۹۹۹) - *Neural networks - A comprehensive foundation* (۲nd ed.), Prentice-Hall, New York.

Marshall, K. (۲۰۱۱) – Our liveable city, *The Melbourne Magazine* (No. ۸٦, Dec.), The Age Newspaper, Melbourne, ٤۰-۷٦.

Wyatt, R. (۲۰۰۸) – Face charts: a better method for visualizing complicated data, in Xiao, Y. & Thij, E. (eds.) - *Proceedings of the IADIS International Conference on Computer Graphics and Visualization*, Amsterdam, The Netherlands (on disk; pp. ٥۱-٥۹).

Yuill, R.S. (۱۹۷۱) – The standard deviational ellipse: an updated tool for spatial description, *Geografiska Annaler. Series B, Human Geography,* ٥۳ (۱), ۲۸-۳۹.