

Urban Planning, Housing and Implications on Cultural Diversity

Domain

Community, Urban Planning & Real Estate

Question

This project seeks to provide an optimistic solution to the following questions:

- ◆ *Does housing policies and prices affect language and cultural diversity in Victoria?*
- ◆ *Could proper urban planning and altering real estate policies be useful to improve spatial segregation if present in Victoria?*

The Local Cities, State and Federal governments would likely be interested in the results of these questions regarding future planning for the socio-cultural environment of the country. The government departments of social services, immigration and infrastructure development as well as finance would most likely want answers to these long-term questions. The information could be used for urban planning and setting housing policies to promote increased cultural and linguistic diversity in Victoria for the sake of national identity and racial harmony.

Data

The following datasets were used:

- Victorian Property Sales Report – Median House by Suburb 2005-2015
Victorian State Government Open Data. Contains Data of the median property prices of each suburb from 2005-2015 with some calculated rate of change and growth rate.
URI: <https://www.data.vic.gov.au/data/dataset/victorian-property-sales-report-median-house-by-suburb-2004-2014>
- ABS Census Data Pack – T01 Selected Person Characteristics
Australian Bureau of Statistics Data. Contains Data of the population estimate for each area categorised by age, citizenship status, sex etc.
URI: <http://www.abs.gov.au/websitedbs/censushome.nsf/home/datapacks>
- ABS Census Data Pack – T10 Language Spoken at Home by Sex
Australian Bureau of Statistics Data. Contains Data of all known spoken languages in Australia grouped by area categorised by each census year, sex and language itself.
URI: <http://www.abs.gov.au/websitedbs/censushome.nsf/home/datapacks>

Pre-Processing

Languages Dataset:

Only the Data for Total Persons speaking the language were looked at instead of each sex. 6 languages were picked, one for English-only speakers, the others being the top five spoken languages other than English. They are Arabic, Chinese, Greek, Italian and Vietnamese. Work was done to combine language groups like Arabic type languages and Chinese type languages into individual ones for a more presentable result and map. Only data from the census year of 2011 was taken into consideration as there are no data for the Property Prices of 2001 suburbs.

A new csv file, *cleanlang*, was created from data of all persons of the 6 languages for 2011 after checking that there were no cells without data. The csv file *cleanlang* had its *region_id* which was previously just 9-digit SA2 (Statistical Area 2) 2011 Codes converted into region names which would have more meaning using the data from *region_id.csv* which contains the respective region names for each SA2 code in Victoria. The suburb names were then piped through some python code to standardise them into lower case to ensure uniformity amongst all the data sets.

Population Dataset:

The csv file came with population estimates grouped by categories such as age, sex, citizenship status, aboriginal status, birthplace etc. Only the columns 2011 for total persons count for each area were extracted along with the region IDs and placed into a new csv file. Some cleaning was done on region IDs with zero population as it would be better ignoring them rather than dividing by zero in later calculations, not impacting the mapping and visualisation of the population density of Victoria.

Similar process of standardisation was applied onto the *region_id* column of the population data set's new csv file, *cleanpop*, with lower case region names.

Property Sales Dataset:

After checking that there were no data for each suburb that was completely null, the excel file was first converted into a csv file for processing. A python function using pandas and numpy libraries was written to replace null values for some regions for certain years. The null values were replaced by the respective mean values for each row rather than a mean of all the housing prices for better accuracy.

After the csv file was successfully run through the function and each null value is successfully replaced, the column 2011 was extracted with the suburb names into a new clean csv file, *cleanhouse*, for later use. The region names were all standardised to lower case and with dashes and brackets removed using python functions.

After running through the 9-digit SA2 code in the *region_id* in both the *cleanlang* and *cleanpop* with a list of matching region names so that they match, the three datasets all had their respective region name columns normalised into lower case for easier manipulation.

Integration

Concentration and Confidence:

A function was written using the pandas and numpy library to calculate the density of each spoken language given the estimated population of the area. The function skips areas that do not appear on the data frame of the population estimate csv. It then matches each value of each language in that region with the total population estimate of the region using a loop and inner function. The concentration coefficient is then calculated with the simple formula:

$$C_{region} = \frac{Speakers_L}{Estimated\ Population_{region}}$$

The cleaned language data set (csv file) was then mapped out into maps for 2011's density for each language using ArcMap. This helps projects a rough visualisation of the density of a spoken language in every area but not yet the population density. The following map is an example of just one in many.

The cleaned language data set on the x-axis was plotted against the population data set on the y-axis on several scatter plots to visual any trends with increasing population.

Approximate similarity string matching was used with 1-gram matching the language and population

data sets with the housing prices data set as they do not share consistent region names and formats. The expected similarity of strings between area names was the reason behind 1-gram matching, a 90% match threshold was used to ensure maximum accuracy and no duplicates.

The concentrations for the matching region names were then put into a data frame along with the median new housing price of 2011 and then plotted on scatter plots.

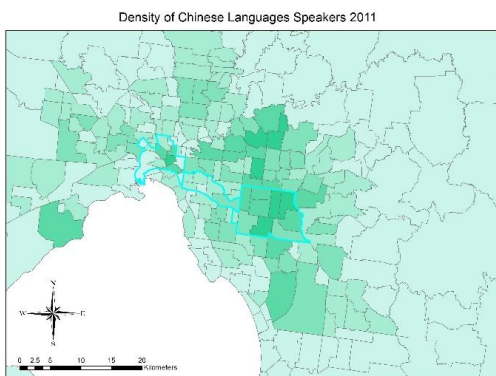


Fig 1: Density of Chinese Speakers in regions in Victoria 2011

Results

The scatter plots of the clean language and population data sets help visualise trends between each language and population. As seen in both figures 2 and 3, the English only speakers display

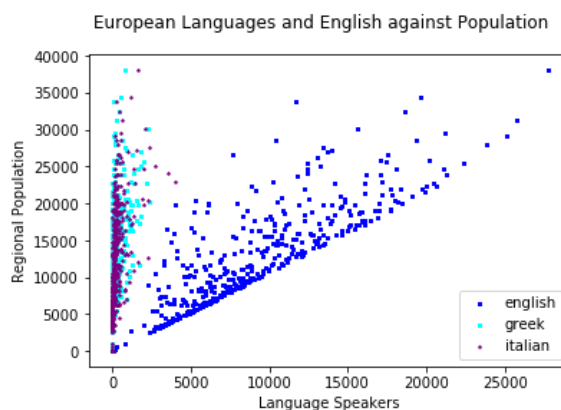
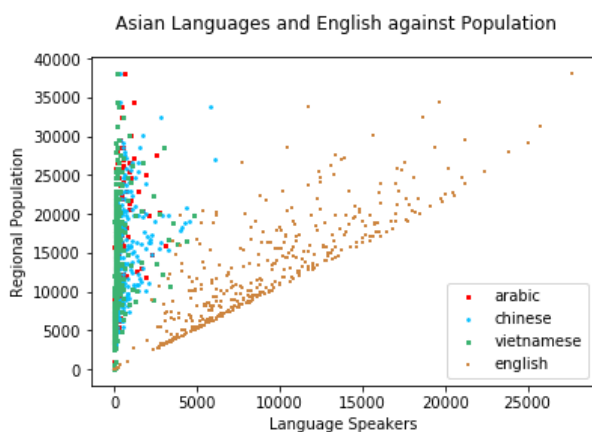


Fig 2: Asian languages against population in Victoria 2011 Fig 3: European languages against population in Victoria 2011

an almost linear correlation with population. This is no surprise being the lingua franca of Australia. The other European languages in figure 3, however, display a much lesser correlation than anticipated. This could be attributed to a dying Greek and Italian speaking population, perhaps turning to English only speakers (Table 1). The scatter plot used to map this out however, does not make it as easy to compare and visualise specific areas alongside the heatmaps.

The concentration correlation was calculated for each region and each language and mapped against housing prices. Fig 4 and 5 shows the different languages grouped as Asian Languages vs English and European Languages vs English. Fig 6 contrasts Italian and Chinese, the more spoken Asian and Non-English European languages and Fig 7 represents the spread of concentration of the top 3 spoken Asian Languages.

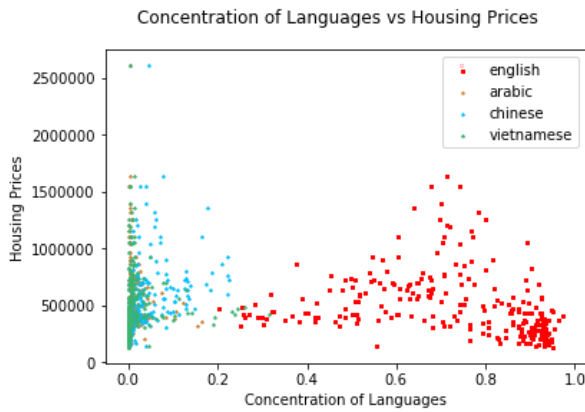


Fig 4: Concentration of Asian Languages vs Housing 2011

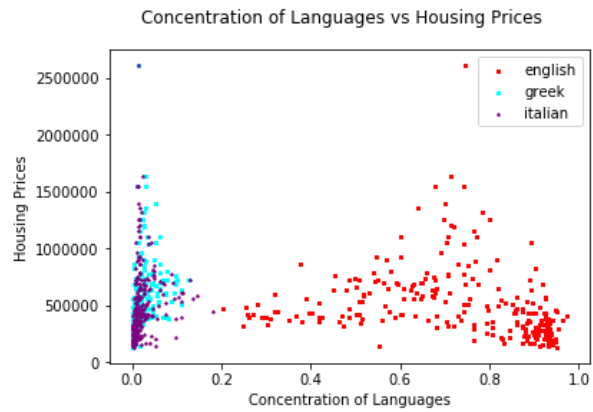


Fig 5: Concentration of European Languages vs Housing 2011

The English only speakers are clustered with high density speakers and cheaper housing prices compared to any other language speakers. This highlights some form of disparity between the housing prices. Although Greek and Italian failed expectations to exhibit a cluster similar to the English language, they are more spread out and do not seem to have a 'minimum' house pricing line as compared to the Asian languages (Figure 7). The Asian language speakers seem to have their density values clustered above an invisible line of around AUD\$400000 (Figure 7). This difference is evident when comparing Chinese against Greek in figure 6, with Chinese having an arbitrary line in the cluster.

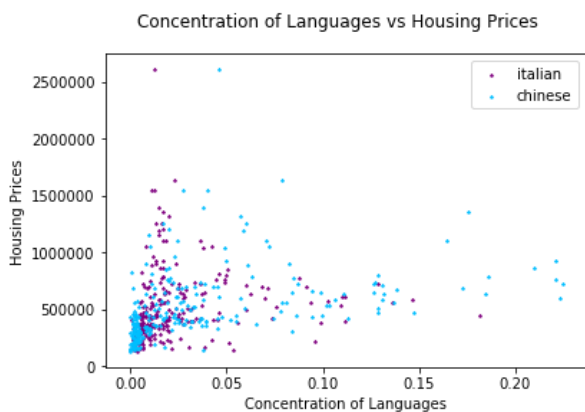


Fig 6: Chinese and Greek Languages vs Housing 2011

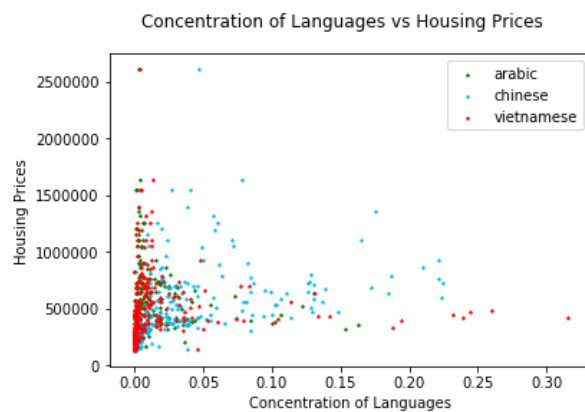


Fig 7: Concentration of Asian Languages Only vs Housing 2011

The Chinese language also showed further spread in its cluster in terms of language concentration in figure 7 compared to the other Asian languages and thus would be examined in detail more.

Case Analysis:

Looking at the heatmaps of Figure 1 and 8, certain areas were found to be concentrated in English-Only speakers and Chinese Languages speakers. Amongst these areas, two were selected for their relative similar population size.

Brighton East and Box Hill North having similar population.

Brighton East is slightly closer to Melbourne Central than Box Hill North.

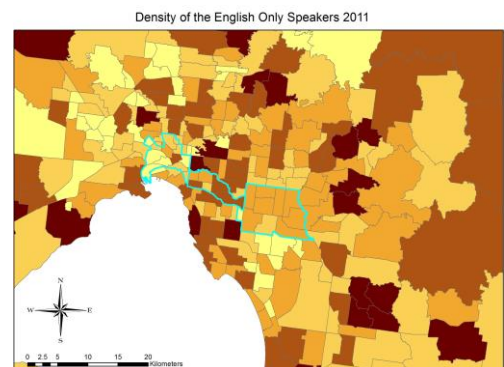


Fig 8: Density of English Only Speakers in regions in Victoria 2011

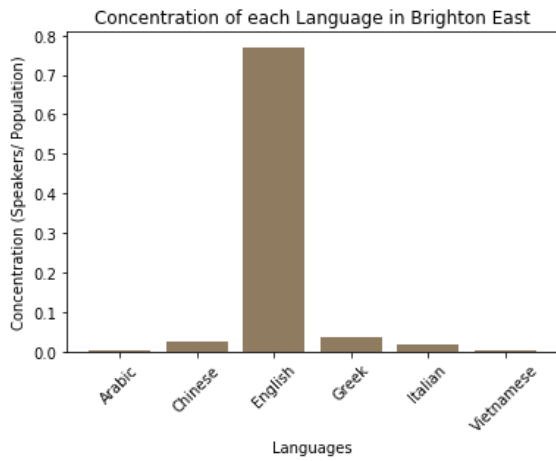


Fig 9: Concentration Languages Brighton East 2011

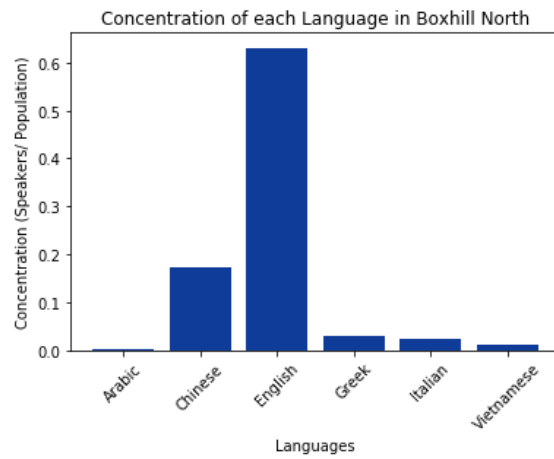


Fig 10: Concentration Languages Box Hill North 2011

Figure 9 and 10 contrasts the language densities of the two regions, showing significant European languages in Brighton East and Chinese speakers in Box Hill North. Table 1 shows the statistics of both regions. Despite the similar population size, Brighton East being a historically well-established area compared to developing Box Hill North, the prices of housing are almost twice the amount. The high concentration of Chinese speakers reflects the housing prices and region history.

Table 1: Statistics of Brighton East and Box Hill North 2011

Brighton East	Population: 15221
	Median Housing Price at 2011: AUD 1100000
	English Only Speakers Concentration: 0.770185927337
	Chinese Languages Speakers Concentration: 0.0240457262992

BoxHill North	Population: 16335
	Median Housing Price at 2011: AUD 683000
	English Only Speakers Concentration: 0.630976430976
	Chinese Languages Speakers Concentration: 0.17257422712

Areas with further away from Melbourne City and newer housing development areas are seeing large concentrations of culture pockets such as Chinese and Vietnamese language speakers. As potential new immigrants and temporary residents are not allowed to purchase pre-existing dwellings (ATO, 2017), they are forced to purchase newer ones or vacant land, often further away from the city central and in specific new developmental areas.

High prices of new dwellings in areas which are nearer to the city or full of older immigrants, English, Greek, Italian speakers, seem to drive newer immigrants away from those regions and into cheaper, developmental regions further away from Melbourne city as seen with our Brighton East and Box Hill North example.

Data Processing Benefits

Null values were removed and replaced by means would give a better estimate of the actual values compared to 0 or a fixed value. Calculating the density of languages using the language data and creating boxplots highlights trends density in certain languages in different suburbs.

As the issue addressed is about urban planning, mapping out the data of spoken languages and population density would be the best way to visualise these data sets. By comparing the densely-populated areas map and the individual maps for each language, we can locate the areas of concentrated language groups that would need to have some policies implemented depending on the density of the speakers. The ratio/percentage of the language spoken and of the population estimate for that area would be found for any area with a high concentration of one language. This would allow

easier visualisation rather than look at all the languages of Australia and every region. When shown on scatterplots, it gives a good visualisation of the contrasts between the trends of different languages.

Discussions and reflections

The Greek and Italian languages were initially expected to show trends more similar to English-Only speakers in their concentration compared to Asian languages. A small difference of the 'line' in the cluster was noticed but could be better visualised in other graphs or plots.

Another improvement that could be made to the code was the string matching 1-gram algorithm. It could be written in such a way where it takes into consideration with brackets or reverse wording order (e.g. West Melbourne, Melbourne West) rather than use a 90% match implementation.

The trends discovered are mostly limited to areas with high concentrations or low concentrations which were selected manually out of the heatmaps. A much better visualisation method could be obtained if the data was projected onto a 3-dimensional model which was too complex for this report.

Resolution

The analysis of the differences between the two regions highlights the disparity of housing prices and how it affects cultural and lingual diversity. The data represent trends in developmental and further areas where new housing is cheaper to have cultural pockets of 'newer' immigrants. It outlines the fundamental flaw in the housing policy by the Foreign Investment Review Board (FIRB, 2017) being only seemingly economically beneficial, however not socio-culturally beneficial to the lingual diversity of Victoria.

The State or Federal governments would be interested in the results in order to plan for long-term cultural and lingual growth of the country. It would also be a more pressing issue as more Asian immigrants arrive in Victoria. The governments may need to adjust their housing policies according to these data to avoid possibilities of cultural, lingual or even racial segregations.

Implemented Code

Apart from using ArcMap as a tool for mapping out the heatmaps of language speakers, the data wrangling and processing were done using Python. The libraries *pandas*, *numpy* and *matplotlib* were used for their data frame, numerical manipulation and plotting usages. The functions and codes were written in Jupyter Notebook files from scratch with no online code functions used or copied.

Notebooks were separated into two files, *test_wrangle* and *test_implement* with one processing clean and usable csv data from the raw data files for Housing, Population and Languages.

The other Python file was used for implementation. It read the clean csv files produced by *test_wrangle* such as *cleanhouse*, *cleanlang*, *cleanpop* and performed calculations on them to find the concentration and also data linkage to produce useful plots and graphs shown in the results.

Bibliography

Australian Government – Foreign Investment Review Board. 2017. *Residential real estate – Foreign Investment Review Board*. [ONLINE] Available at: <http://firb.gov.au/real-estate/>. [Accessed 1 May 2017].

Australian Taxation Office. 2017. *Foreign Investment Review Board approval*. [ONLINE] Available at: <https://www.ato.gov.au/Individuals/International-tax-for-individuals/Investing-in-Australia/Owning-real-property-in-Australia/>. [Accessed 2 May 2017].