The Journal of Open Source Software

ausdex: A Python package for using Australian economic indexing data

Robert Turnbull[1] **and Jonathan Garber**[1]

1 Melbourne Data Analytics Platform, University of Melbourne

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Andrew Stewart

Reviewers:

- @konstantinstadler
- @potterzot

Submitted: 11 February 2022
Published: unpublished

Summary

The Australian Bureau of Statistics (ABS) publishes a variety of indexes for the Australian economic environment. These include the Consumer Price Index (CPI) used for calculating inflation and various indexes designed to measure socio-economic advantage. ausdex makes these data available in a convenient Python package with a simple programatic and command-line interfaces.

Statement of need

ausdex is a Python package for querying data produced by the ABS and returning them in a convenient format. Currently ABS data is typically housed in Microsoft Excel spreadsheets linked from the data catalogue. This package interfaces with a subset of the the data to provide an Application Programming Interface (API) to derived economic metrics. For example, we expose the Australian consumer price index data to create an inflation calculator similar to the cpi Python package for adjusting US dollars. In addition, we bring API access to ABS Socio-Economic Index Data for Areas (SEIFA) aggregated at the suburb level in Victoria. This allows for quick assessment of the socio-economic history of a suburb in Victoria from historical census data. These datasets are housed in different online repositories and are aggregated to different spatial extents since statistical geographic boundaries are redrawn from every census dataset. The `ausdex.seifa_vic` submodule eliminates the need for downloading and combining datasets from different data sources, and allows for time series comparisons tied to the current suburb geographic boundaries.

Socio-economic indexes aggregated from census data for Victoria

Since 1986 (see table 1), the ABS has generated "Socio-Economic Indexes For Areas" (SEIFA) following each census (Australian Bureau of Statistics, 2016a). These indexes are aggregations of socio-economic inputs from the census forms (i.e. household income, rental/mortgage price, educational level) at the "census district level" or "mesh level" (2006–current). Census districts, or mesh levels, are geographic areas statistically defined from the census data to be the largest scale (smallest) geographic building blocks of demographic and socio-economic data based on population distribution. These statistical geographies are redrawn after each census. The ABS aggregates these to other statistical "levels" of geographic area from the Australian Statistical Geography Standard (ASGS) (Statistical Areas Levels 1–4) suburbs (Australian Bureau of Statistics, 2016c) and local government areas in their "Data Cube" outputs.

However, there have been several new suburbs created during the SEIFA program, and datasets aggregated by suburb are not readily available for census data before 2006. To address this,

we used the current Victorian suburb areal polygons (Department of Industry, Science, Energy and Resources, 2019b) as the constant spatial areas over which we aggregate all previous census datasets. To overcome suburb names that are repeated, the suburb polygons were also overlaid with local government areas (Department of Industry, Science, Energy and Resources, 2019a) to distinguish duplicate suburbs.

For an example, fig. 1 shows the Index of Economic Resources (IER) scores in 2015 for all suburbs in Victoria.

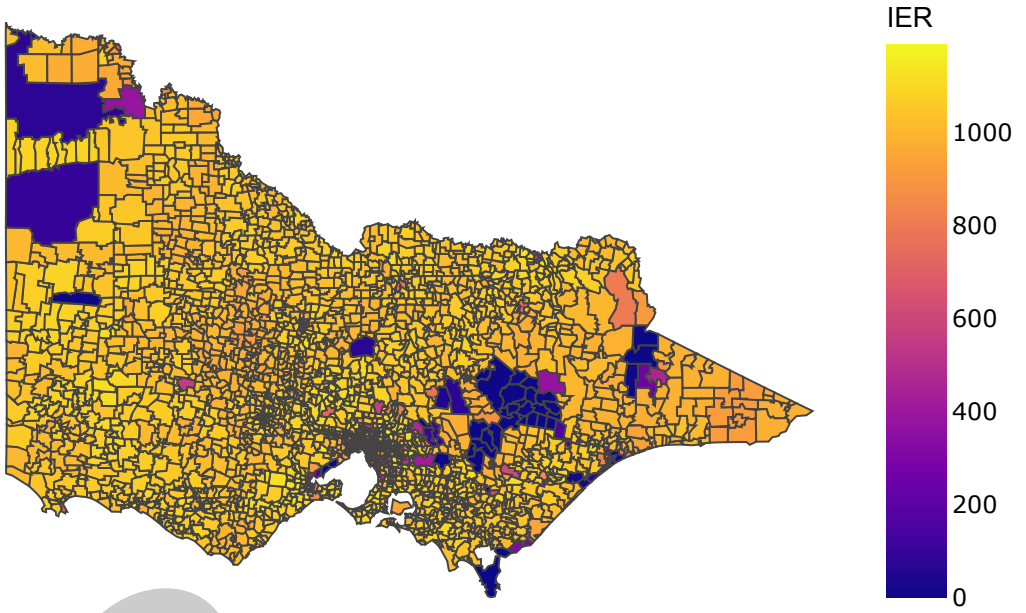

Figure 1: Figure 1

Figure 1: A choropleth map of Victorian suburbs representing the interpolated IER scores in 2015.

Index	Name	Years Published
IEO	Index of Education and Occupation	1986, 1991, 1996, 2001, 2006, 2011, 2016
IER	Index of Economic Resources	1986, 1991, 1996, 2001, 2006, 2011, 2016
IRSD	Index of Relative Socio-economic Disadvantage	1986, 1991, 1996, 2001, 2006, 2011, 2016
IRSAD	Index of Relative Socio-economic Advantage and Disadvantage	2001, 2006, 2011, 2016
UIRSA	Urban Index of Relative Socio-economic Advantage	1991, 1996
RIRSA	Rural Index of Relative Socio-economic Advantage	1991, 1996

Table 1: The SEIFA indexes and the years published.

Spatially aggregating the 1986–2006 datasets

For the SEIFA datasets from 1986 to 2006, we collected census district polygons from AURIN (Sinnott et al., 2015) and the ABS data repository (2006), together with associated aggregated SIEFA scores. A list of the data sources is shown in table 2. These census district level SEIFA scores were aggregated to the current suburb GIS datasets (Department of Industry, Science, Energy and Resources, 2019b) using the following steps:

1. Suburbs and census districts were both reprojected to EPSG:4326.
2. The polygons were unioned together, so the resulting polygon layer had an individual polygon for each overlapping census district and suburb (fig. 2).
3. The merged polygons were reprojected to a UTM projected coordinate system EPSG:32756. Note that this UTM coordinate system does not overlay the state of Victoria perfectly, but we are assuming that locally the measured areas are relatively consistent with each other.
4. The SEIFA scores for all of the census district parts within each suburb were aggregated using a weighted average, using the polygon area as the weight.

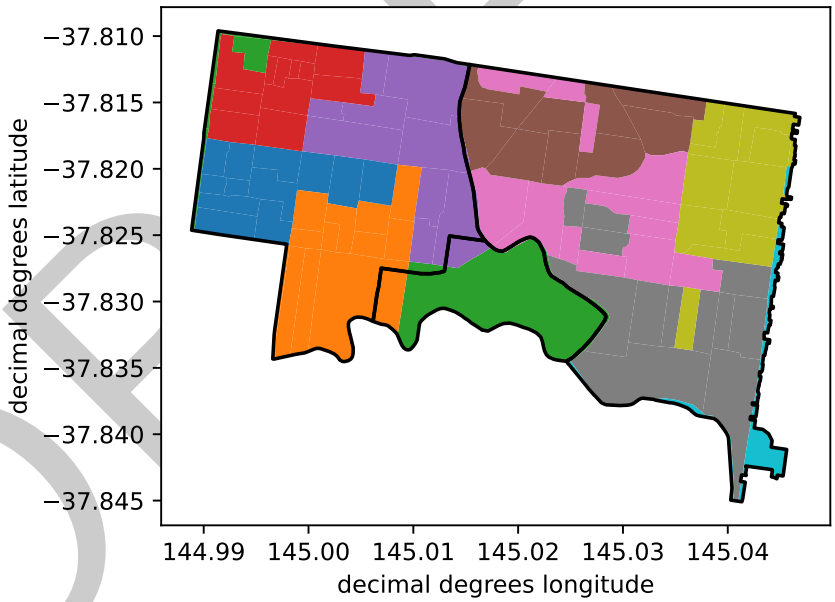

Figure 2: Figure 2

Figure 2: Map of three suburb outlines (black lines) for Richmond (left), Burnley (center), and Hawthorn (right) overlaying 1986 Census Districts (colored polygons with white boundaries). The census districts are colored according to the census district code. Note that these districts do not line up with suburb boundaries. The green district in the lower middle section spans parts of Richmond, and Cremorne. Likewise one of the orange and purple census districts spans two suburbs.

Spatially aggregating the 2011 and 2016 datasets

For the 2011 and 2016 datasets, we used the same procedure set out above, but started with a different statistical geographic dataset. We used Statistical Area Level 1 (SA1) aggregated

estimates of the SEIFA variables published as an ABS data cube, and GIS polygons of SA1 boundaries from the ASGS created for 2011 and 2016 to derive suburb aggregated datasets.

Data sources

Year	Dataset type	Dataset source
1986	Census district polygons and metrics	(Sinnott et al., 2015) wfs id: `AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1986`
1991	Census district polygons and metrics	(Sinnott et al., 2015) wfs id: `AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1991`
1996	Census district polygons and metrics	(Sinnott et al., 2015) wfs id: `AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1996`
2001	Census district polygons and metrics	(Sinnott et al., 2015) wfs id: `AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_2001`
2006	ABS census district shapefile	(Australian Bureau of Statistics, 2006a)
2006	ABS census district SEIFA metrics	(Australian Bureau of Statistics, 2006b)
2011	ABS SA1 Polygons	(Australian Bureau of Statistics, 2011b)
2011	ABS SA1 SEIFA metrics	(Australian Bureau of Statistics, 2011a)
2016	ABS SA1 Polygons	(Australian Bureau of Statistics, 2016c)
2016	ABS SA1 SEIFA metrics	(Australian Bureau of Statistics, 2016b)
All	VicMap suburb polygons	(Department of Industry, Science, Energy and Resources, 2019b)
All	VicMap Local Government Area Polygons	(Department of Industry, Science, Energy and Resources, 2019a)

Table 2. A list of data sources for `seifa_vic` submodule.

Inflation

The Consumer Price Index (CPI) is a weighted average price of a basket of goods and services for urban consumers (Australian Bureau of Statistics, 2018). The ABS issues the Australian CPI each quarter. The data are available from September 1948 onwards. The CPI values before decimalization of the Australian currency on February 14, 1966 are in understood according to the conversion rates specified in the 1965 Year Book of Australia such that £1 is equivalent to $2 (Commonwealth Bureau of Census and Statistics, 1965, p. 810).

To adjust prices for inflation, we assume that the ratio of prices for two dates is equal to ratio of the CPIs for those dates (Parkin & Bade, 2019, p. 811). This gives the formula:

$$\text{Price at time B} = \text{Price at time A} \times \frac{\text{CPI at time B}}{\text{CPI at time A}}$$

The ausdex package automates the process for downloading the latest version of the Australian CPI data from the ABS. The user enters a price, the original date (A) and the evaluation date (B) and it returns the adjusted price. The inputs can be scalar values or vectors as a `NumPy` array (Harris et al., 2020), a `pandas` series (McKinney, 2010) or a `Modin` `pandas` series (Petersohn et al., 2020). The size of the returned vector of prices is the same as the vector of original prices. If the original date or the evaluation dates are vectors instead of scalar values then these must be the same size as the vector of prices. Several scenarios for validation are in the automated tests and these have been compared with the Reserve Bank of Australia's inflation calculator.

Module Features

The components of the module work both from a simple command-line interface and through the API. The code style adheres to PEP 8 (Rossum et al., 2001) through the use of the Black Python code formatter. Automated tests run as part of the CI/CD pipeline and testing coverage is above 96%. The package is thoroughly documented at https://rbturnbull.github.io/ausdex/.

Acknowledgements

This project came about through a research collaboration with Vidal Paton-Cole and Robert Crawford (University of Melbourne). We acknowledge the support of our colleagues Aleksandra Michalewicz and Emily Fitzgerald.

This app uses the NCRIS-enabled Australian Urban Research Infrastructure Network (AURIN) Portal e-Infrastructure to access the following datasets:

- AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1986,
- AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1991,
- AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_1996,
- AU_Govt_ABS-UoM_AURIN_DB_3_seifa_cd_2001.

References

Australian Bureau of Statistics. (2006a). *Collection district digital boundaries, victoria (ASGC 2006) in ESRI shapefile format*. Statistical Geography - Australian Standard Geographical Classification (ASGC), Digital Boundaries.

Australian Bureau of Statistics. (2006b). *SEIFA, census collection districts, data cube only, 2006*. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia - Data only.

Australian Bureau of Statistics. (2011a). *Statistical area level 1, indexes, SEIFA 2011*. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia.

Australian Bureau of Statistics. (2011b). *Statistical area level 1 (SA1) ASGS ed 2011 digital boundaries in ESRI shapefile format*. Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas.

Australian Bureau of Statistics. (2016a). *2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia*. https://www.abs.gov.au/ausstats/abs@.nsf/mf/2033.0.55.001

Australian Bureau of Statistics. (2016b). *Statistical area level 1, indexes, SEIFA 2016*. Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia.

Australian Bureau of Statistics. (2016c). *Statistical area level 1 (SA1) ASGS ed 2016 digital boundaries in ESRI shapefile format*. Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas.

Australian Bureau of Statistics. (2018). *6461.0 - Consumer Price Index: Concepts, Sources and Methods*. https://www.abs.gov.au/ausstats/abs@.nsf/mf/6461.0

Commonwealth Bureau of Census and Statistics. (1965). *Official Year Book of the Commonwealth of Australia*. https://www.ausstats.abs.gov.au/ausstats/free.nsf/0/1485DE068ECD64A7CA257AF60014F8E1/$File/13010_1965_bk51.pdf

Department of Industry, Science, Energy and Resources. (2019a). *VIC local government areas - PSMA administrative boundaries GeoJSON*. VIC Local Government Areas - Geoscape Administrative Boundaries.

Department of Industry, Science, Energy and Resources. (2019b). *VIC suburb/locality boundaries - PSMA administrative boundaries GeoJSON*. VIC Suburb/Locality Boundaries - Geoscape Administrative Boundaries.

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Parkin, M., & Bade, R. (2019). *Macroeconomics: Australia in the Global Environment* (2nd Edition). Pearson Australia.

Petersohn, D., Macke, S., Xin, D., Ma, W., Lee, D., Mo, X., Gonzalez, J. E., Hellerstein, J. M., Joseph, A. D., & Parameswaran, A. (2020). Towards scalable dataframe systems. *Proc. VLDB Endow.*, *13*(12), 2033–2046. https://doi.org/10.14778/3407790.3407807

Rossum, G. van, Warsaw, B., & Coghlan, N. (2001). *Style guide for Python code* (PEP No. 8). https://www.python.org/dev/peps/pep-0008/

Sinnott, R. O., Bayliss, C., Bromage, A., Galang, G., Grazioli, G., Greenwood, P., Macaulay, A., Morandini, L., Nogoorani, G., Nino-Ruiz, M., Tomko, M., Pettit, C., Sarwar, M., Stimson, R., Voorsluys, W., & Widjaja, I. (2015). The Australia urban research gateway. *Concurrency and Computation: Practice and Experience*, *27*(2), 358–375. https://doi.org/10.1002/cpe.3282