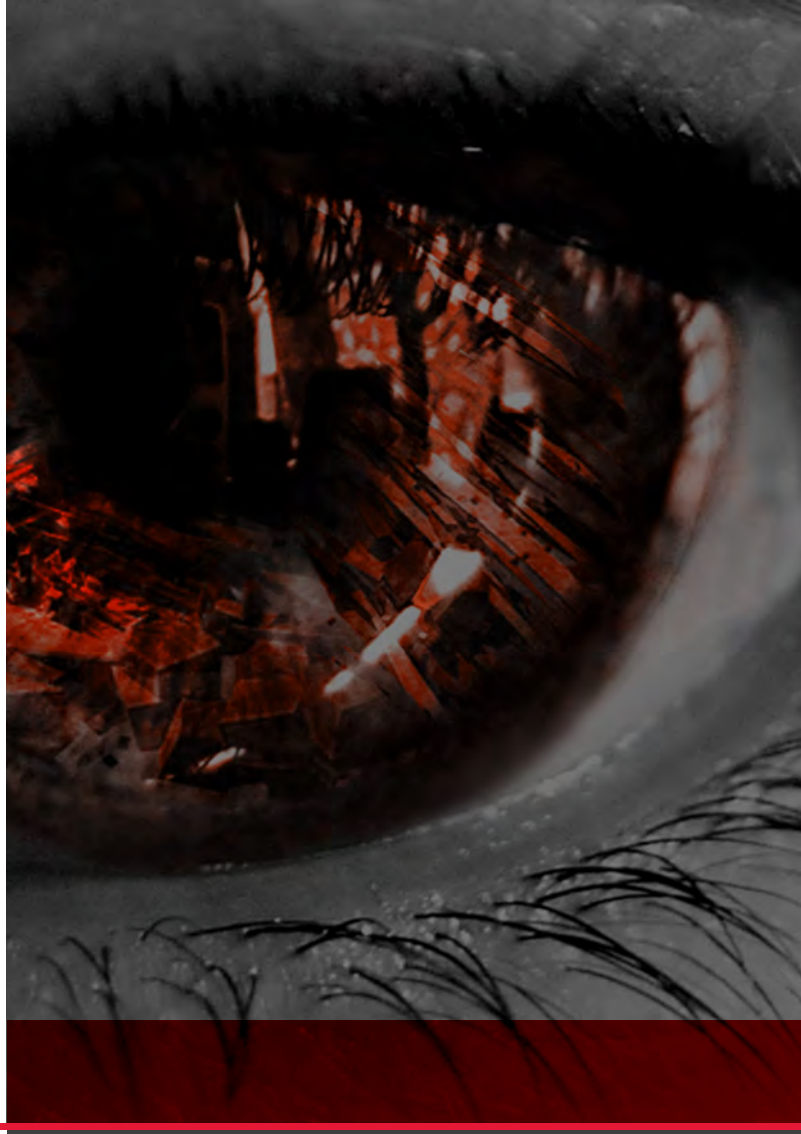

THE UNIVERSITY OF SYDNEY

The Henry Halloran Trust

Open Access Policy Analysis and Visualisation

Somwrita Sarkar,
Nicole Gurran,
Hanley Weng,
Stacey Miers,
Sachin Wasnik.

Urban Housing Lab
The University of Sydney

Incubator program
Final Report 2015

EXECUTIVE SUMMARY

Urban Lab@Sydney was awarded $30,876 by the Henry Halloran Trust in 2014, with the aim of establishing the beginnings of an open access data system that would enable analysis and visualization of headline indicators for the Australian housing market and policy space.

The Blue Sky UrbanLab@Sydney project has provided the opportunity to kick start and provide momentum for a long term machine learning and big data led information visualization research and dissemination programme around housing and larger urban planning issues for Sydney and New South Wales. It was envisioned as one of the first efforts of its kind in any university across Australia to develop an online, publicly relevant and accessible system for data analysis, information visualization and urban informatics, including a regular commentary series on housing policy and city science research in Australia.

This summary report documents the project's attained milestones. The various stages and achievements on data collection and analysis, research directions and progress, major research themes developed, and implementation stages are discussed. Finally, ongoing work and emerging plans for the future are presented. In summary, the visualization platform has been realized. A number of scholarly outputs were achieved. A number of new collaborations have been initiated at both academic and industry levels. The implementation of the Blue Sky project has been strengthened and provided continuity by the Henry Halloran Trust Research Incubator Fellowship awarded to the team leaders (Nicole Gurran and Somwrita Sarkar) in 2015, for 2015-17. Many of the results emerging from the Blue Sky UrbanLab@Sydney project have directly led to research and outcomes progress for the Incubator project. The outputs from the Incubator and all future projects will also be hosted on the online visualization platform that has been created as part of this Blue Sky project.

STAGE 1: DATA

One of the primary and continuously ongoing aspects of the Blue Sky project has been the collection of relevant open source data that informs housing policy research in Australia. The aim of the Blue Sky project was to (a) identify all potential sources of such data, along with identifying any potential road blocks to their use (such as particular data points not being available at all ABS and non-ABS spatial scales, or changes of geographies between two census periods, etc.), (b) developing a system to continuously track important data points, and (c) to track new "types" of data points in addition to traditional sources of data, such as newspaper content, twitter or social media based or physical sensors based data.

Three major gap areas were identified in this direction:

1. **Data availability and coordination:** A major gap area identified was that although open data spanning across several government and private sources is freely available, it is not available for combined analysis and visualization to an ordinary informed and interested member of the public. In this direction, the recently introduced Australian Urban Research Infrastructure Network (AURIN) strives to partially fulfill the gap, but usage of the AURIN platform is restricted to organizational users, and comparative visualizations are not easily performed. Some organizations and data points are listed as available, but cannot be loaded when the request is made. Further, AURIN serves purely as a data repository only: no analysis, or critical commentary is available.
2. **Data reliability:** Many data points are now being reported on a monthly basis, but the data quality and reliability is questionable. For example the housing approvals and completions data, as Monthly Monitors, from the NSW Department of Planning, was discovered to be

inaccurate and unreliable, with major gaps of communication and data flow between private bodies such as certifiers, local councils, and state level agencies. Many local councils, such as the City of Sydney, have made attempts to make approvals, completions, and construction activity data available, but such data is only sporadically available.

3. **Data led indicators and analysis:** While a lot of data is available, its use is limited. An analysis of scholarly and policy literature reveals limited use of this data, in ad-hoc ways. While the ABS has enormous amounts of data available, with some of it spatially visualized via the TableBuilder GIS mapping tool, there is limited work on understanding how existing data points may be combined to form indicators that can have policy relevance. A rich toolkit that incorporates both data points and indicators is available as the NSW Housing Toolkit (http://www.housing.nsw.gov.au/Centre+For+Affordable+Housing/NSW+Local+Government+Housing+Kit/), but using the toolkit data points and indicators as a framework to produce visual analytics has received limited focus.

Three primary areas of data collection were prioritized, focusing on prices, demand and supply of housing in Australia. Specifically, because of the apparently anomalous activity of the Sydney housing market in the recent months, specific focus was given to the collection of Sydney and New South Wales related data. Further, one of the initial aims of the project was also to extend spatially informed analysis of the housing market to very fine geographic scales, such as the LGA, suburbs, or latitude-longitude level precision data, since much of previous research focuses on aggregate levels of analysis.

With this view, the Table 1 outlines the data points that have been collected, and the sources from which they are being collected. The project has enabled a system, where the data points are now being continuously collected, and will continue to be collected even beyond the specific duration of this project.

Traditional as well as new non-traditional sources of data have been considered. For example, while the Australian Bureau of Statistics (ABS), NSW Department of Planning, and Housing NSW have been mined for known data points, new non-traditional data from social media, online blogs, and newsarticles have been collected for the development of new types of indicators and analysis of public sentiment and information dynamics in the housing market, and its effect on price dynamics.

Data Source	Data points and time intervals	Geographic units	Status
Australian Bureau of Statistics (ABS)	Approvals of new dwelling stock, monthly	State level	Collected, and continues each month
NSW Department of Planning	Approvals and completions of new dwelling stock, monthly	LGA level, Sydney metropolitan area	Collected and continues each month
NSW Housing	Rent and Sales reports, quarterly	LGA level, Sydney metropolitan area	Collected and continues each quarter
City of Sydney	Residential property monitors, monthly, for approvals, lodgements, and completions of new dwelling stock	City of Sydney, address level micro-data	Collected (only available upto 2013).
Twitter	Tweets related to the Sydney housing market, daily, hourly	Latitude-longitude level microdata	Collected and continues to be collected
Newsarticles, blogs, online text data	Textual data reporting on the Sydney housing market, daily, hourly	Various	Collected, mined, processed, and continuous

Table 1: Data collection sources, types of data and frequency of collection

STAGE 2: HEADLINE INDICATORS

With the data points that have now been collected, the team has planned a series of critical headline indicators that highlight the main points of the housing market in Sydney and New South Wales, and will be visualized geographically, hosted on the Urban Lab@ Sydney / Urban Housing Lab@ Sydney website. The website has been set up and activated at http://sydneyurbanlab.com/v0.5/pages/index.html and will continue to be updated regularly even after this project is completed. Results from the UrbanHousingLab Incubator (www.urbanhousinglab.com) project will also be visualized on this visualization platform. The visualizations are constantly updated and the team is working towards refining the clarity and interactive aspects of the data visualizations.

Some of the indicators derive from the pure data points, while others are derived from combinations of data points and statistical computations around them. The presentation of the indicators will be carried across two dimensions: understanding their geographic or spatial evolution and their temporal evolution.

The following "pure" data indicators had been finalized for the first stage and web based visualizations have either been developed are in the process of being developed:

1. Approvals of new dwelling stock at the state and LGA level, over time.
2. Completions of new dwelling stock at the state and LGA level, over time.
3. Sales data (median and quartiles) at the LGA and suburb levels over time
4. Rent data (median and quartiles) at the LGA and suburb levels over time
5. Demographics over time and space: population, immigration, family, income characteristics, etc.

In addition, many other pure and derived data points can be visualized using the same framework. For example, the NSW Housing Information Kit has a comprehensive list of detailed variables listed (http://www.nswlocalgovernmenthousingkit.com.au/hkit/dave_tables_vars_list.cfm).

In particular, the derived variables planned are around the measurements of:

1. Housing stress
2. Housing affordability
3. Inequality measures such as: a. Gini coefficient b. Theil index c. Atkinson index

Specifically, a primary emergent research motivation, given the still rising and unaffordable all time price highs in the Sydney market, is the theme of measuring inequality in the housing market and its effect on the future of affordability and stress: of access, opportunity of entering the market, housing stock distributions. These derived indicators have been finalized and developed, in the form of coordinated GIS shape files and related excel and csv based files, and one of the primary projects in the Urban Housing Lab incubator is to now host them up in final form on the visualization platform.

STAGE 3: DISSEMINATION AND FUTURE PROJECTS

For the final stages of the project, several dissemination strategies were planned, primarily focused around web based interactive visualizations, a series of short commentaries accompanying the visualizations, a quarterly or yearly white paper on the state of the Sydney/ NSW housing market, and journal papers.

Scholarly output from project

One journal paper has been submitted to Environment and Planning B, and is currently in review. The open access draft for this paper is available at: http://arxiv.org/abs/1509.00959. A second paper on the use of new types of data from wifi based sensors to track vehicular mobility in the city in real time resulted from one of the CIs collaborations with the School of IT. A preliminary paper was accepted and presented at the ACM International Conference on Knowledge Discovery and Data Mining (KDD 2015, Sydney), Urban Computing workshop, UrbComp15, available here: http://www2.cs.uic.edu/~urbcomp2013/urbcomp2015/papers/Infer-Traffic-Connectivity_Chawla.pdf This paper has now been invited to the journal *IEEE Transactions on Big Data* in extended form.

Two other papers are in an advanced stage and ready for submission by end of this year or beginning of next year. One of them is focused on charting the co-evolution of social media data from Twitter and newspaper content online and the Sydney housing market, proposing algorithms for using social media and newspaper data to predict housing price movement at an LGA or suburb level. Such research has been done for the stock market (for example see: http://arxiv.org/pdf/1406.7330.pdf), but has never been performed for the housing market, an extension that we are now finalizing. The second paper is focused on the rent and sales and approvals and completions data, and extends current regression based econometric models into machine learning models for continuous prediction of price, demand and supply activity in NSW and Sydney, at the LGA and suburb levels.

New collaborations and projects

At the beginning of the Blue Sky project, we were only focused on housing data, especially for Sydney and NSW. However, as unexpected collaborations emerged, and resulted in publication outputs, we are now extending our focus into relating housing aspects with transport and health related aspects, since internationally there is focus as well as data availability (data of new types such as sensor data on real time traffic mining) on this topic. For example, our UrbComp15 real time traffic mining paper (see above) has now resulted in a further collaboration with Woods Bagot Australia, a large urban design and architecture firm, who are keen to use the methods developed for real spatio-temporal trajectory mining to **understand real time pedestrian behavior** in the city. This understanding can be used to inform the design of more walkable and healthier neighborhoods.

We have been in conversation with the Department of Planning, NSW, to apply our machine learning and econometric based modeling to the problems of **housing demand and supply prediction** and inferring the **co-evolving relationships between prices, demand, supply, construction activity, internal migration and work-home travel patterns** for the Sydney metropolitan region.

We are planning both pilot and ARC Linkage applications through both of the new collaborations stated above.

UrbanLab@Sydney: Data visualization and infographics platform

The final deliverable for the Blue Sky project is the Urban Lab@Sydney data visualization and infographics platform. The platform is online at www.sydneyurbanlab.com and interactive visualizations of approvals, median sales prices across LGAs have been produced. The full list of pure and derived indicators have been finalized and developed, in the form of coordinated GIS shape files and related excel and csv based files, and one of the primary projects in the Urban Housing Lab incubator is to now host them up in final form on the visualization platform. Work is on-going in terms of refining the visualizations. An example screen shot of the output from the website is attached in Fig. 1. As was the aim, both the spatial as well as temporal information is captured in the interactive visualizations – users can zoom in and zoom out spatially, choose area definitions (in upcoming versions), and click through a time line to see temporal variations over space.

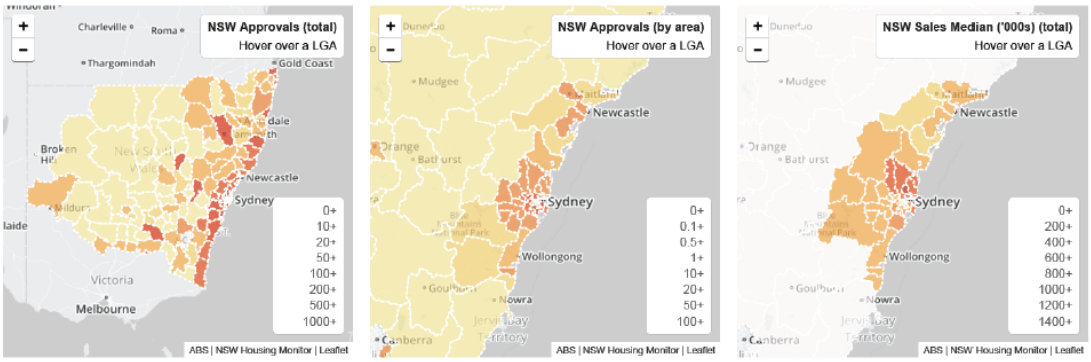

Figure 1: Interactive visualizations developed for approvals and sales in Sydney LGAs

Apart from the interactive visualizations, a separate output will be a regular (quarterly or 6 monthly) commentary series on housing prices, supply and demand in the Sydney metropolitan region, downloadable as pdfs. An example downloadable commentary pdf file is attached in Fig. 2.