Analysis Report
on
The Impact of Including a Housing Stress Variable in SEIFA

A Report for NSW Treasury

Prepared by the Advanced Modelling and Optimisation Section

Australian Bureau of Statistics

September 2018

Disclaimer

Views expressed in this report are those of the Advanced Modelling and Optimisation Section and do not necessarily represent those of the Australian Bureau of Statistics. Where quoted or used, they should be attributed clearly to the Advanced Modelling and Optimisation Section.

Table of Contents

1. Executive Summary

This report presents the results of an investigation into the potential impact of including a measure of housing stress in the Index of Relative Socio-economic Disadvantage (IRSD).

In the analysis, a housing stress variable was created using data from the 2016 Census of Population and Housing. This variable was added to the list of candidate variables for IRSD, and then the standard process for creating the Socio-Economic Indexes for Areas (SEIFA) was applied.

The results of this analysis show that the impact of including a measure of housing stress on the overall distribution of index scores for NSW is small. However, there are a few areas with large changes in ranking. These areas tend to have a relatively high proportion of housing stress and relatively low proportions for the other indicators of disadvantage used to create the index.

The housing stress measure used in this analysis has some limitations, which should be taken into account when these results are considered. The limitations are described in section 3.2 of this report.

2. Introduction

The Socio-Economic Indexes for Areas (SEIFA) are created by the Australian Bureau of Statistics (ABS) every five years using data from the Census of Population and Housing. The most recent version is SEIFA 2016. SEIFA is initially calculated at the Statistical Area Level 1 (SA1) level, and indexes for other geographic levels are calculated by aggregating the SA1 level data. For more information about the concepts behind SEIFA and the methods used to create it, see the SEIFA 2016 Technical Paper, which is available to download at the following location:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument

SEIFA 2016 does not contain a direct measure of housing stress. The ABS considered incorporating a measure of housing stress into SEIFA 2006, but decided not to pursue this for various reasons. In particular, the fact that the Census did not collect information on Commonwealth Rent Assistance was considered to be a reason why it would be challenging to create an accurate measure of housing stress. For more information about the decision not to include a measure of housing stress in SEIFA, see Section 3.1.6 of the SEIFA 2006 Technical Paper, which is available to download here:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2039.0.55.0012006?OpenDocument

The NSW Treasury has asked the ABS to carry out an analysis of the potential impact of including a measure of housing stress on the Index of Relative Socio-economic Disadvantage (IRSD), which is one of the indexes in SEIFA. The aim of the analysis would be to assess how much the index would be changed by including the measure, with a particular focus on areas in NSW.

There are a number of different possible ways to measure housing stress. The measure used in this analysis is designed to calculate the number of dwellings where the housing costs make up 30% or more of the total household income, for lower income households only.

This report explains the details of the housing stress measure, describes the steps taken in the analysis process and presents the results. Some limitations and caveats are also discussed.

3. The Measure of Housing Stress

3.1 Details of the measure

The measure of housing stress used for this analysis was created from data collected in the 2016 Census of Population of Housing. The details of the measure were determined through discussion with NSW Treasury and the ABS's Living Conditions Statistics team. The measure is the proportion of households in the SA1 who belong to approximately the lowest 40% of households by equivalised household income and who spend 30% or more of their total household income on rent or mortgage costs.

Income is collected in ranges in the Census of Population and Housing, so the midpoint of the range indicated by the Total Household Income variable (HIND) was assigned to each dwelling for the purposes of calculating this measure.

The cut-off for determining the lowest 40% of households was calculated using the Equivalised Total Household Income variable (HIED). Cumulative counts of dwellings with stated, non-nil values for HIED were used to determine the cut-off for this measure. Again, HIED is expressed in ranges so the cut-off is an approximation. The cut-off used for the measure is households with HIED values from '02' to '07'. This cut-off captures about 44.6% of the households with stated, non-nil values of HIED in the Census.

Total household income, total equivalised household income and rent payments are all expressed in weekly values in the Census data, while mortgage payments are expressed in monthly values. The mortgage payments were converted to weekly values and the measure was calculated by comparing weekly rent or mortgage payments to the midpoint of the range of weekly total household income.

The measure was calculated at the dwelling level. This measure was then converted into a proportion at the SA1 level by dividing the number of dwellings with housing stress by the number of dwellings with stated responses to the Census variables used to calculate the housing stress measure.

For more information about the Census variables used to calculate this measure, see the 2016 Census Dictionary:

http://www.abs.gov.au/ausstats/abs@.nsf/mf/2901.0

3.2 Limitations of the measure

The questionnaire for the Census of Population and Housing is not focused on income. That is, the income question is only one among many other topics. For this reason, respondents to the Census may not provide answers that are as accurate as they would for surveys that have many questions about different types of income, such as the ABS's Survey of Income and Housing.

Also, income is collected only in ranges in the Census. Using the midpoint of the range as the income value for the dwelling will incorrectly classify some households as experiencing or not experiencing housing stress. However, there is no clear reason to suspect any systematic bias.

It should also be noted that there are various different conceptual ways to measure housing stress. The measure used in this analysis does not incorporate information about expenses other than housing, and also does not include information about any support that households may receive, such as Commonwealth Rent Assistance. In some cases, households with high mortgage costs may be choosing to pay more than they are required to, and the measure may not accurately reflect housing stress for those households. Also, there are some high-wealth households that have almost finished paying off a mortgage and report very low income, and the measure may not accurately reflect housing stress in those cases either. For these reasons, some measures of housing stress only include households who are renting.

3.3 Validation of the measure

After the measure was calculated at the SA1 level, the SA1s with the highest proportions of households experiencing housing stress according to this measure were checked. The results were consistent with Census data presented at the SA1 level in the online QuickStats tool:

http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20QuickStats

The housing stress measure was also mapped within Sydney to provide a visual representation of the distribution of SA1s with high proportions of housing stress according to this measure. Two of these maps are provided with this report and are described in Appendix 2.

4. Summary of Analysis

The following process was used for this analysis:

1. The housing stress indicator variable, as defined in section 3.1 of this document, was created at the dwelling level.

2. A denominator variable for the housing stress variable was created. This variable counts the in-scope dwellings in the SA1 with stated responses for the variables used to derive the housing stress variable, including weekly rent payments, monthly mortgage payments, household income and equivalised household income. If the denominator count for an SA1 is very small, that area will be excluded from receiving a SEIFA score.

3. The housing stress indicator variable and the corresponding denominator variable were used to calculate the proportion of dwellings in the SA1 where the household is in housing stress.

4. The housing stress proportion variable was added to the list of candidate variables for the Index of Relative Socio-economic Disadvantage (IRSD). The standard approach for calculating SEIFA using Principal Component Analysis (PCA) was then applied to the list of candidate variables. See the SEIFA 2016 Technical Paper for more details about the standard approach for calculating SEIFA.

5. The output from the PCA was then compared with the SEIFA 2016 IRSD to assess the impact of including the housing stress variable.

5. Results

5.1 <u>Excluded SA1s</u>

In the 2016 SEIFA, 55028 SA1s received IRSD scores, out of a total of 57523 SA1s in Australia. In this analysis, after the housing stress variable was included, 11 additional SA1s were excluded due to low denominator counts for the new housing stress variable. In other words, 55017 SA1s received scores in this analysis.

5.2 <u>Correlations between candidate variables</u>

In the standard process for creating SEIFA, the pairwise correlation between the candidate variables is checked before the PCA is run. If there is a pair of variables with a very high correlation, either positive or negative, then these two variables are likely to be capturing similar concepts and it will not be beneficial to include both variables in the analysis. For this reason, the absolute values of the correlations between the candidate variables is checked. If the absolute value of the correlation between two candidate variables is more than 0.8, then the two variables are considered and one of them is usually removed from the candidate variable list.

For this analysis, the correlation of the new housing stress variable with all the other variables was checked. There were no correlations whose absolute value was greater than 0.8, so none of the candidate variables were removed from the list.

5.3 <u>Impact on variable loadings</u>

The SEIFA creation process is iterative: the PCA is run, the variable with the lowest absolute loading is dropped from the candidate variable list and the PCA is run again. This process is repeated until there are no variables with an absolute loading of less than 0.3. When this process was run with the housing stress variable added to the IRSD candidate variable list, the CERTIFICATE, FEWBED and OCC_SALES_L variables were dropped due to low loadings. These are the same variables that were dropped from the 2016 IRSD.

The following table shows the variable loadings after the final PCA from this analysis, compared with the variable loadings from the 2016 IRSD.

Table 1: <u>Comparison of variable loadings</u>

Variable mnemonic	Variable loading from 2016 IRSD	Variable loading after including housing stress variable
INC_LOW	-0.91	-0.90
CHILDJOBLESS	-0.83	-0.83
NONET	-0.79	-0.77
NOYR12ORHIGHER	-0.77	-0.75
UNEMPLOYED	-0.75	-0.75
OCC_LABOUR	-0.74	-0.73
LOWRENT	-0.73	-0.72
HOUSING_STRESS	N/A	-0.69
ONEPARENT	-0.67	-0.69
DISABILITYU70	-0.67	-0.66
SEPDIVORCED	-0.55	-0.56
OCC_DRIVERS	-0.54	-0.53
OCC_SERVICE_L	-0.53	-0.54
NOCAR	-0.49	-0.50
OVERCROWD	-0.46	-0.47
NOEDU	-0.43	-0.44
ENGLISHPOOR	-0.30	-0.32

5.4 <u>Proportion of variance explained</u>

The proportion of variance explained by the 2016 IRSD index was 43.12%.

The proportion of variance explained in this analysis, with the addition of the housing stress variable, was 43.17%.

5.5 <u>Changes in ranking, percentiles and distribution of NSW SA1s by quintile</u>

This section presents some summaries of the impact on the SA1 index values.

The following two plots give an impression of the amount of change to the rankings of the SA1s due to the inclusion of the housing stress variable.

The first plot has a point for every SA1 that received an index score in this analysis. The ranking is calculated by sorting the SA1s by their index score. The SA1 with the lowest score is given a rank of 1, the SA1 with the second lowest score is given a rank of 2 and so on. The second plot is restricted to NSW SA1s, although the rankings are still calculated at the Australia level. These plots show that there is a general pattern to the maximum amount of change, but there are a few SA1s with larger decreases in their rankings after the housing stress variable is included.

Figure 1: <u>Plot of SA1 rankings: analysis with housing stress variable against 2016 IRSD</u>
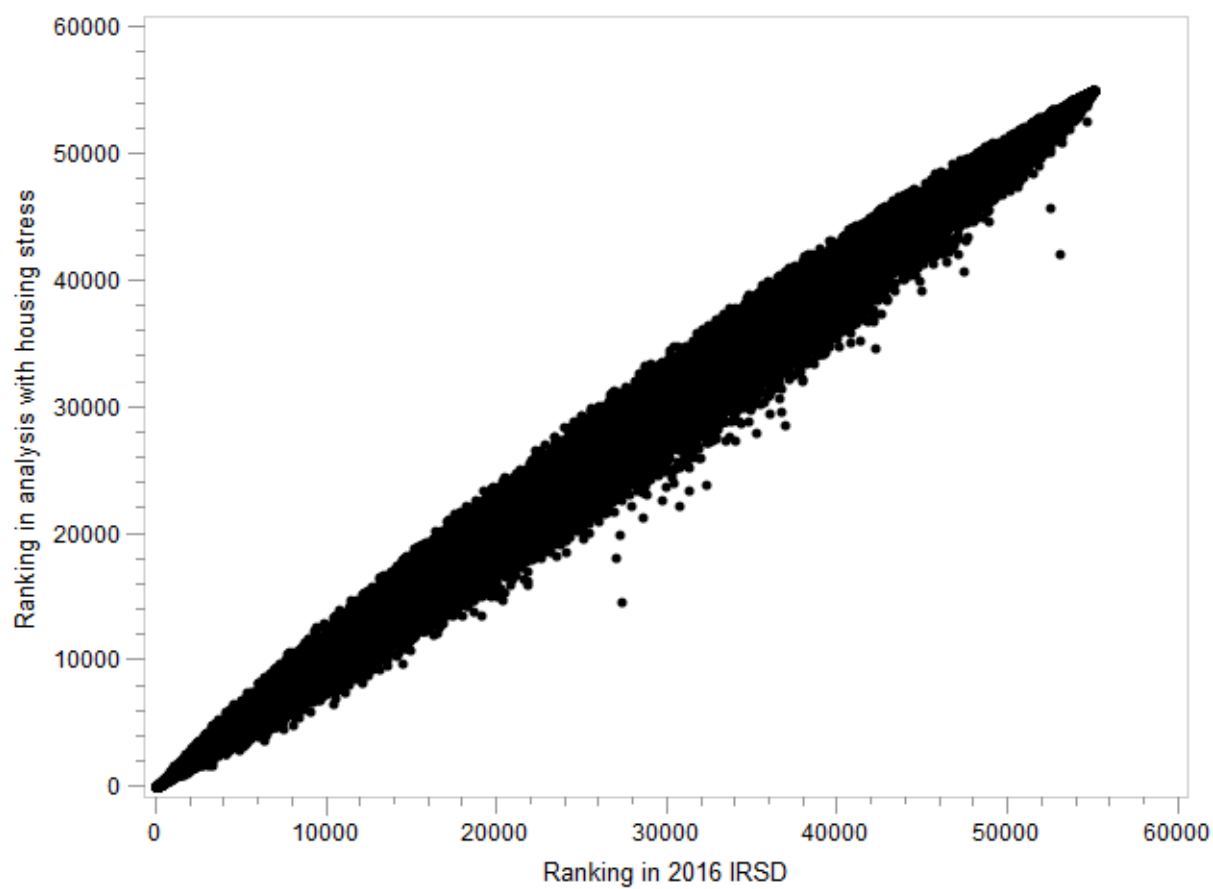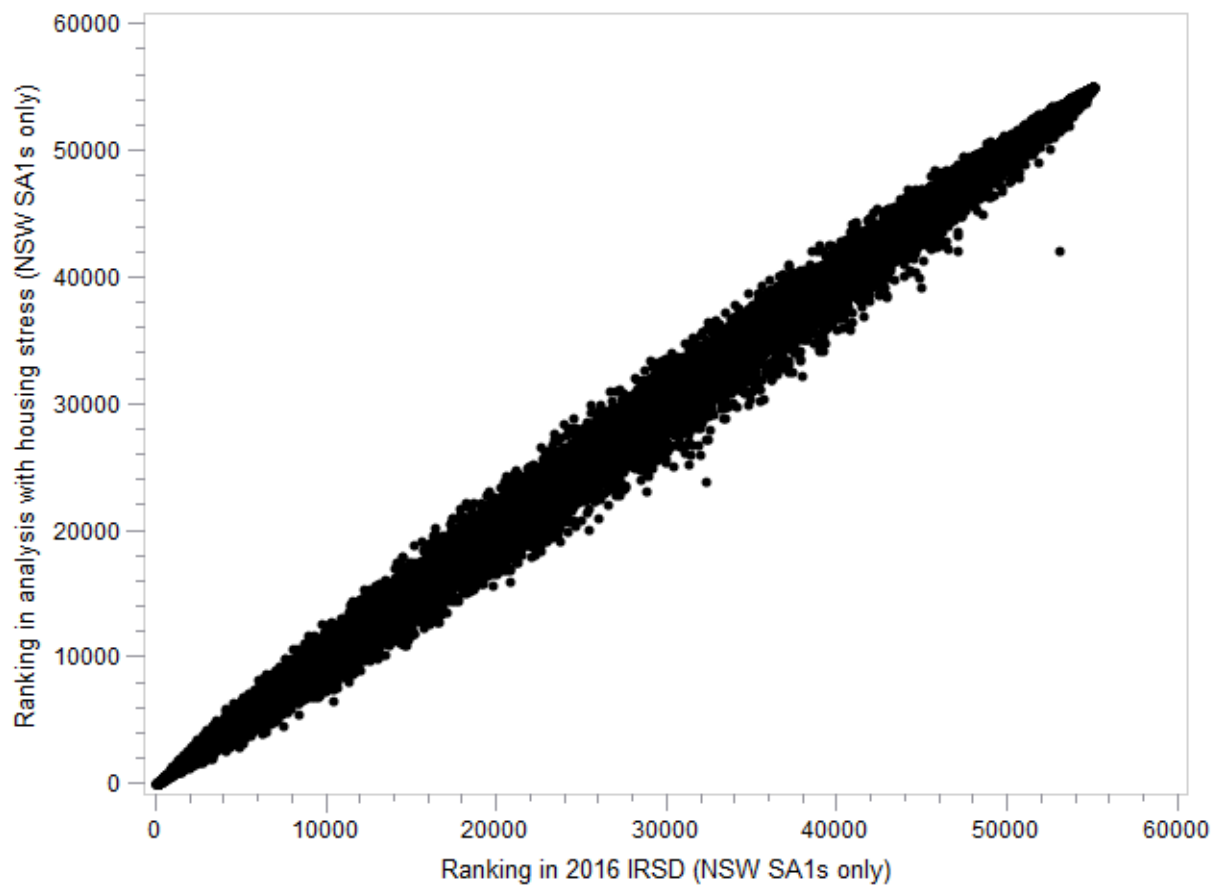
Figure 2: Plot of SA1 rankings: analysis with housing stress variable against 2016 IRSD, showing SA1s in NSW only

Another way to examine the impact is to tabulate the amount of change to the percentile of the SA1. The percentiles of the SA1s are calculated as follows: all areas are ordered from lowest to highest score, the lowest 1% of areas are given a percentile number of 1, the next lowest 1% of areas are given a percentile number of 2 and so on, up to the highest 1% of areas which are given a percentile number of 100. The following table contains counts of SA1s within groups determined by the amount of change to the SA1's percentile caused by including the housing stress variable.

Table 2: Distribution of changes to percentile after including housing stress variable

Size of change to percentile	Count of SA1s
Decrease of 10 or more	57
Decrease of 5 to 9	1975
Decrease of 1 to 4	17796
No change	13607
Increase of 1 to 4	20156
Increase of 5 to 9	1426
Increase of 10 or more	0

The following table shows the impact of including the housing stress variable on the distribution of NSW SA1s across the Australia-level quintiles. The percentages in the table are calculated as percentages of the total number of NSW SA1s receiving an index score.

Table 3: Comparison of the proportion of NSW SA1s falling into each quintile

Quintile	1	2	3	4	5
Percentage (count) of NSW SA1s: 2016 IRSD	21.63% (3810)	20.07% (3536)	18.09% (3186)	17.13% (3018)	23.07% (4064)
Percentage (count) of NSW SA1s: with housing stress variable included	21.86% (3850)	20.11% (3543)	17.74% (3124)	17.08% (3009)	23.17% (4082)

5.6 Analysis of areas with the largest changes

An analysis of the SA1s with large changes to their percentiles suggests that the biggest impact of including the housing stress variable will be on areas with relatively high housing stress but relatively low proportions for many of the IRSD variables.

Out of the 57 SA1s where the impact of including housing stress was a decrease in the percentile of 10 or more, only 12 SA1s originally had a percentile value of 50 or less. None of these SA1s originally had a percentile value of less than 35.

5.7 Analysis of impact on index at higher level geographies

The SEIFA scores are calculated at the SA1 level, but scores are then created for other geographic levels from population-weighted averages of the SA1-level scores. This aggregation process has been repeated to assess the impact of including a housing stress variable on the ranks at the Statistical Area Level 2 (SA2) and Local Government Area (LGA) levels. The following tables show the impact on the ranking of the SA2s and LGAs in NSW that had the lowest scores for the 2016 IRSD.

Table 4: Lowest scoring SA2s in NSW (according to 2016 IRSD)

SA2 Code	SA2 Name	Rank in 2016 IRSD	Rank after including housing stress variable	Percentile in 2016 IRSD	Percentile after including housing stress variable
11504	Ashcroft - Busby - Miller	32	29	2	2
11511	Cabramatta - Lansvale	35	34	2	2
11515	Fairfield	38	35	2	2
11313	Bidwill - Hebersham - Emerton	46	44	3	3
11513	Canley Vale - Canley Heights	49	48	3	3
11204	Tamworth - West	50	52	3	3
11601	Warwick Farm	54	50	3	3
11599	Lurnea - Cartwright	55	53	3	3
11480	Fairfield - East	59	54	3	3
11546	Port Kembla - Warrawong	62	64	3	3
11096	Walgett - Lightning Ridge	64	77	3	4
11583	Auburn - North	71	60	4	3
11155	Kempsey	75	79	4	4
11598	Liverpool	80	68	4	4
11512	Cabramatta West - Mount Pritchard	81	76	4	4
11316	Lethbridge Park - Tregear	85	84	4	4
11573	Lakemba	91	66	5	4
11516	Fairfield - West	94	80	5	4
11483	Guildford - South Granville	97	74	5	4
11158	Nambucca Heads	99	105	5	5

Table 5: Lowest scoring LGAs in NSW (according to 2016 IRSD)

LGA Code	LGA Name	Rank in 2016 IRSD	Rank after including housing stress variable	Percentile in 2016 IRSD	Percentile after including housing stress variable
11200	Brewarrina (A)	35	35	7	7
11700	Central Darling (A)	37	37	7	7
17900	Walgett (A)	39	38	8	7
12850	Fairfield (C)	42	40	8	8
12150	Coonamble (A)	48	46	9	9
14350	Kempsey (A)	56	52	11	10
11250	Broken Hill (C)	67	70	13	13
16610	Richmond Valley (A)	69	63	13	12
12950	Gilgandra (A)	75	83	14	16
15700	Nambucca (A)	76	69	14	13
17400	Tenterfield (A)	78	82	15	16
14550	Kyogle (A)	79	76	15	14
18020	Warrumbungle Shire (A)	83	93	16	18
14920	Liverpool Plains (A)	84	90	16	17
13010	Glen Innes Severn (A)	87	85	16	16
14200	Inverell (A)	89	87	17	16
11150	Bourke (A)	90	108	17	20
15300	Moree Plains (A)	94	92	18	17
14870	Lithgow (C)	105	104	20	20
12350	Cowra (A)	107	113	20	21

6. Appendices

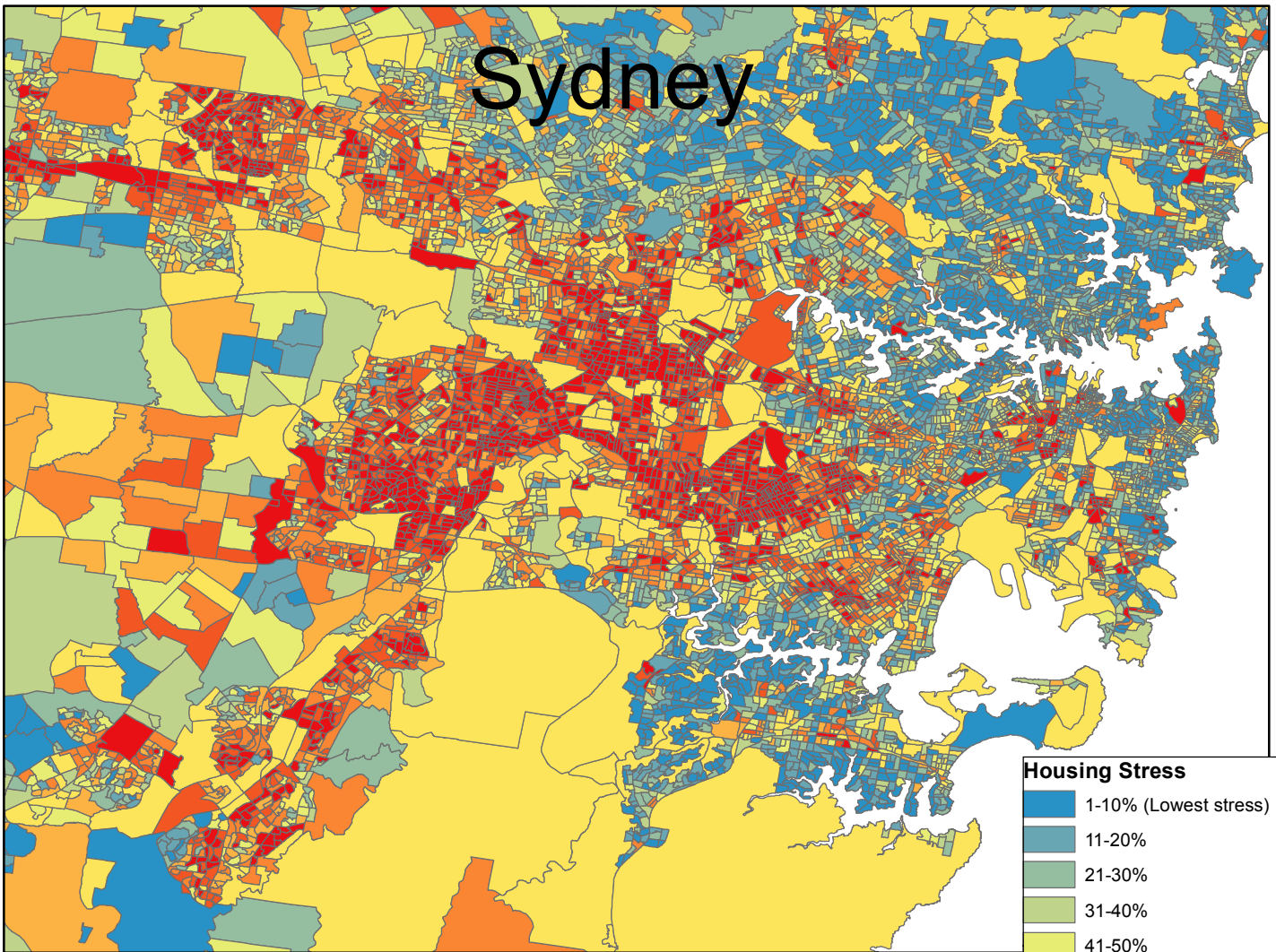6.1 <u>Attachment 1: Accompanying spreadsheet</u>

The spreadsheet that accompanies this report ("SA1 IRSD scores with housing stress.xls") presents results at the SA1 level across Australia. For comparison purposes, the spreadsheet contains the published IRSD 2016 scores (in the first 13 columns of the spreadsheet) and also the scores produced by the analysis described in this report.

A cell containing 'NA' indicates that the SA1 was excluded from receiving a score once the housing stress variable has been added.

6.2 <u>Attachment 2: Maps of distribution of housing stress variable in Sydney</u>

The accompanying PDF document "Sydney Housing Stress.pdf" shows the geographic distribution of the housing stress variable that was created for this project, across Sydney and Inner Sydney.

Sydney

Inner Sydney

Housing Stress

- 1-10% (Lowest stress)
- 11-20%
- 21-30%
- 31-40%
- 41-50%
- 51-60%
- 61-70%
- 71-80%
- 81-90%
- 91-100% (Highest Stress)