Gaspard Pairault *2024*

Spatial relationships between the urban environment and mental health:

A focus on Greater Sydney using *Movement and Places*, a new dataset released by the NSW Government

PHRAME

Supervisors:
Luke Knibbs
Joe Van Buskirk

Sydney Local Health District
The University of Sydney

Background

With growing cities, there is an increasing interest in the influence of the urban environment on both physical and mental health (Krefis et al., 2018). Conceptually, researchers agree that environmental characteristics like access to healthcare, education, employment, green spaces or the safety of the neighbourhood should affect mental health both at the individual and community level (Reuben et al., 2022). However, systematic reviews show that changes in the urban environment with transport infrastructure, green space or urban regeneration plans have limited direct effects on mental health (Moore et al., 2018). The fast-growing urban populations are increasing the demand for housing and putting pressure on different components of the urban environment including transport networks and infrastructure (UN, 2015). Australian cities are no exception with 2023 being the largest year in urban population growth across all states and territory capitals (ABS, 2024). The rapid growth of the urban population across Sydney is associated with different rises in poverty across suburbs of the east and west, ultimately widening the socio-economic gap (Flavel et al., 2022). Statistical projections of population growth, urbanisation rate and climate change have driven the design and construction of liveable cities, which consider the health of the citizens and environment as their most important asset (WHO, 2016). As the size and population of cities grow, transport networks including roads and streets need to be upgraded to maximise the connectivity and mobility of people, ultimately reducing various inequalities.

The NSW Government implemented a new framework called Movement and Place (MaP) which aims to address some of the challenges faced by the urban environment in Sydney (MAP, 2021). The Decision Board of MaP consists of multiple governmental organisations including Transport for NSW, the Department of Planning and Environment and NSW Health. Two main objectives are clearly articulated in the framework. First, to contribute to the transport network of public spaces where people can live healthy, productive and social lives. Secondly, to enhance transport and provide appropriate space allocation for people and goods to move safely and efficiently, ultimately connecting different places and people. The philosophy of the project relies on acknowledging that transport networks (roads and streets) have a dual function by moving people and goods while being important public spaces where daily life unfolds. Previous guidelines similar to the MaP framework were published under the name of “Road Planning Framework” in 2017 by the NSW Government but the availability of a dataset with indicators to describe the urban environment is a new feature. The placement project is built around this dataset as no analysis has previously been done with it and it could be very useful in a public health context. The dataset consists of 36 environment indicators which are grouped first into 10 built environment performance outcomes and then into 5 distinct themes to cover all the concepts around movement and places. Each of the 36 environmental indicators has one or multiple metrics that can be used as variables for any data analysis. While the environmental indicators are not designed to be health-oriented, they have the potential to be associated with health. For example, a higher proportion of tree canopy in the neighbourhood was associated with lower psychological distress at the individual level after adjusting for possible confounders in Greater Sydney (Astell-Burt and Feng, 2019).

The Urban Liveability Index (ULI) is a different measure that is commonly used to describe the urban environment. The ULI is a statistical model developed by the Australian Urban Observatory that combines over 50 different environmental indicators to obtain a single index (AUO, 2022). The index has been designed to capture the spatial variation of the urban environment within cities and assess inequities in Australia. (Higgs et al., 20222). Researchers have chosen the indicators within the ULI to reflect the social determinants of health, which are important for physical and mental health. The index is designed using a compression algorithm so that scoring high in some indicators does not compensate too much for having poor performance in other environmental indicators (Higgs et al., 2019). An index like the ULI can summarise the complexity of the urban environment but sacrifices precision with potentially oversimplified results (EUJRC, 2008), The ULI contains a lot of indicators and does not provide any concrete description of the urban environment. For this reason, I will investigate the possible use of the MaP data instead of the ULI for urban and environmental epidemiology studies in Greater Sydney.

Research looking at the association with the urban environment and mental health is relatively new and results tend to vary depending on the study area. Better findings are necessary in Greater Sydney to quantify the association between different characteristics of the urban environment and mental health. At the community level, the proportion of people with high psychological is publicly available for each suburb of Greater Sydney on the Torrens University website (PHIDU, 2021). This data is obtained by model estimates from individual health surveys using the 10-item Kessler Psychological Distress scale (K10). The K10 is used to measure the risk of psychological distress at the individual level (Kessler et al., 2003). Overall, high-quality data are available to measure the spatial relationship between the urban environment and mental health in Greater Sydney.

Aim

During my placement, I used spatial statistics modelling to determine if the indicators from the Movement and Place dataset can be used to describe the urban environment and capture any association with mental health in Greater Sydney. There were three objectives within the aim of this placement:

1. To assess the association between MaP and ULI without direct relation to health
2. To assess the relationship between two datasets that describe the urban environment (ULI and MaP) and a modelled health outcome in Greater Sydney, in this case, high psychological distress as measured by the K10
3. To assess the robustness of these purely urban environmental models with the inclusion of an established confounder, e.g. Socio-economic factors

Methods

1. Software for spatial analysis: QGIS and R

During my placement at PHRAME, I performed different spatial statistics methods using QGIS and R. Spatial data stores the same information as non-spatial data but also has a location component that allows the data to be visualised on a map. Spatial data of different datasets were visualised using QGIS which is a software that allows to visualise spatial data. QGIS was useful for exploring the spatial distribution of different indicators in Greater Sydney but also to visualise possible associations with multiple indicators. QGIS was also used to visualise the different analyses performed with R. R is a programming language for statistical analysis commonly used in epidemiology and biostatistics. All the analysis during this placement was performed with R which included summarising the different datasets, fitting different regression models and using graphical visualisation to analyse the results (see Appendix to access R code)

2. Preliminary data exploration of MaP

2.1 Defining spatial level and study area

Spatial data can be stored as polygon data where each polygon stores a particular value and all polygons within a given area can be analysed. Statistical Areas Level 2 (SA2s) will be the polygons of interest during this project. SA2s are generally purpose areas built by the Australian Bureau of Statistics (ABS) to represent a community that interacts together socially and economically (ABS, 2021). SA2s have on average a population of 10,000 people and can be compared to what is commonly described as suburbs in Greater Sydney. Consequently, all the data will be summarised for each SA2 of Greater Sydney and the analysis will be performed between SA2s. The study area also had to be defined early on to avoid any unnecessary work. Indeed, the MaP data is available for most of NSW, but a smaller area was going to be used for the analysis. This had to be decided early on to avoid a situation where data would be summarised for some SA2s and not included in the analysis. Sydney Local Health District was too small as the analysis needed to capture all the inequities within Sydney. Greater Sydney which starts south from Newcastle and ends north from Wollongong while extending from the most urbanised environment in NSW to the Blue Mountains and protected wilderness areas. The division of Greater Sydney into SA2s was visualised using QGIS (Fig 1.a). The analysis focused on the association between urban characteristics and community mental health so including areas with low-density populations like NSW National Parks would not be relevant. Each SA2 of Greater Sydney was ordered by population density in R and the bottom 10% of the data was excluded from the analysis. The study area now contains all SA2 within Greater Sydney with a population density greater than 215 pers/km^2 and was visualized using QGIS (Fig 1.b). Overall, 335 SA2s were selected so for each variable used in this analysis, there will be 335 observations.

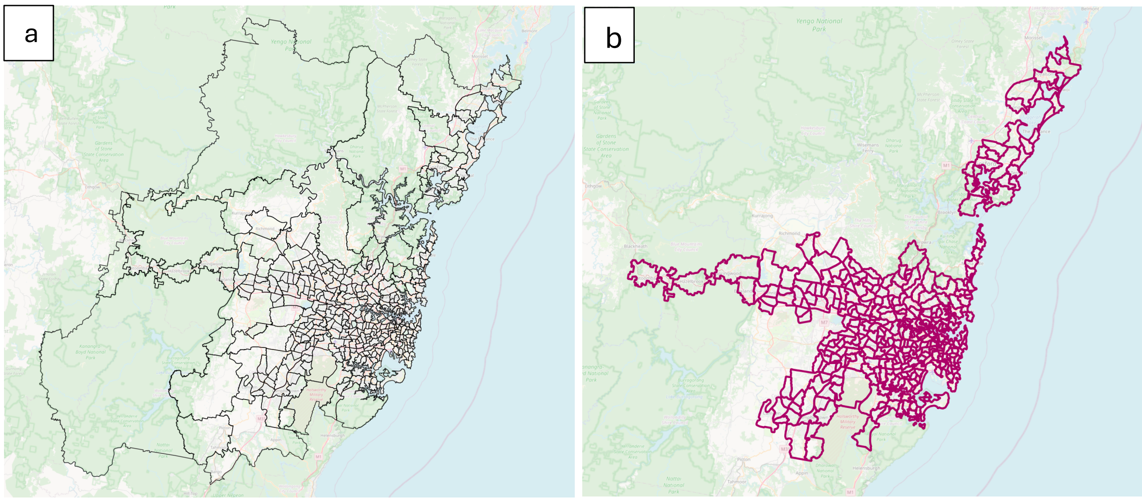


Figure 1 – Division of Greater Sydney in SA2s. a) All SA2s b) Selected SA2s

2.2 MaP and ULI early visualisation with QGIS

The environmental indicators from MaP are publicly available and can be downloaded from their website. A metadata spreadsheet was first built with Excel to better understand each metric's aim and meaning (see Appendix). There were 42 individual metrics and for each one, I recorded different information including a short description, the provenance of the metric, the unit of the metric as well as the type of geospatial data which was either a polygon or line type (Table 1). Understanding the provenance of the data is always a good measure, but it was especially important at this stage of the placement to be familiar with this new MaP dataset. As mentioned earlier the ULI also contains many environmental indicators. Data provenance was used to ensure that the indicators in MaP were different to those that were used for the calculation of the ULI. If this was the case and both datasets were originally the same metrics that had been summarised in different ways, then the project would be meaningless. Essentially, I would have been fitting statistical models to the same data and overfitting would have been very likely. It was not clear how much work I would have the time to do so the first block of the placement relied on determining if the MaP dataset can provide any information on the urban environment that the ULI cannot. To obtain promising results, the MaP dataset had to be different from the ULI. After comparing all indicators from MaP and those that were used to build the ULI, 3 indicators were excluded from the analysis. However, the overall trend was that the MaP data provided much more depth about different aspects of the urban environment than the ULI.

Metric Name	Type of geospatial data	Metric Unit	Data Source
Crime density	Polygon	Count per 100,000 people	BOSCAR crime Statistics
FSI and casualty crash rates	Line	Count per km	Centre for Road Safety database
Difference in temperature relative to non-urban reference	Polygon	Temperature	CSIRO 2018-19 UHI Estimates NSW Urban Heat Island to Modified Mesh Block 2016
Proportion of public space	Polygon	Percentage	DPIE, Public Space Location and Access + NSW Spatial Services
Jobs density	Polygon	Count per km	Employment count (EMP 2016) from TPZ19 employment projection table
Percentage of heritage area	Polygon	Percentage	EPI Heritage (DPIE) State Heritage Register Curtilage (DPIE) Commonwealth Heritage List (DEE) National Heritage List (DEE) Australia, World Heritage Areas (DEE) TfNSW S170 Heritage and Conservation Register.
Average building height	Polygon	Metres	Geoscape Buildings
Density of tree cover	Polygon	Percentage	Geoscape® Tree tiff and surface cover.
Pedestrian access to inland waterbodies	Line	Metres	NSW Hydroarea
Number of local living needs within walking distance	Line	Count within 800m radiu	NSW Spatial Services Point of interest
Public transport accesibility level	Polygon	Index	TfNSW (2021), GTFS from AbS Mesh Block 2016
Street lighting	Line	Count per km	TfNSW Road Track Path Network TfNSW Road Asset Management.
Distance to cycling network	Line	Metres	TfNSW Road Track Path Network + TfNSW, NSW Cycleways (August 2020)
Walkable distance to primary school	Line	Metres	TfNSW, School Gates

Table 1 – Metadata of a subset of the environmental indicators present in the MaP dataset

Polygons and line spatial objects from MaP could then be visualised in their original format using QGIS (Fig 2). The polygon metrics contained three subtypes (Mesh Blocks, Suburbs and Travel Zone) which are different ways of classifying the urban environment in Sydney and looking at the spatial variation of different characteristics at a small and localised scale. Each polygon has a different value which summarises the environment within the polygon. The line metrics on the other hand all corresponded to street or road segments. Each line has a value summarising the appropriate urban environment within the segment. Early visualisation of the raw data from MaP in QGIS allowed me to get an idea of the trend in the distribution of each metric indicator in Greater Sydney. It was also an insightful way to observe missing, incorrect or inaccurate missing data. This was a crucial validation step to identify anything missing or seemingly invalid in the data. Many line metrics did not align perfectly with the map of Sydney streets, so they had to be downloaded from the MaP website at a higher resolution.

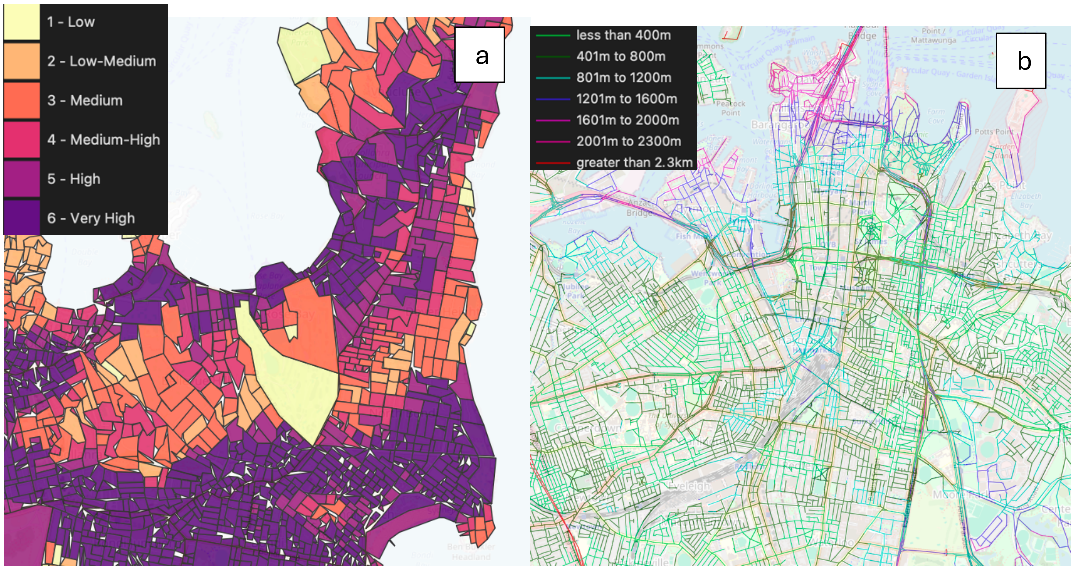


Figure 2 – Visualisation of spatial data from MaP indicators using QGIS.
a)Distribution of public transport accessibility, in the Bondi area, a polygon object stored at the Mesh Block level. b) Distribution of distance from primary school in the city of Sydney, an example of line spatial data

3. Data preparation and cleaning

The MaP dataset had to be summarised at the SA2 scale using R. Metrics that were stored as polygons were summarised first. Intersections between the incoming data (MaP) and the target polygon (selected SA2s) were performed and then summarised at the SA2 level using means. These steps were repeated for each selected SA2 and each selected environmental indicator. This process was very time-consuming as it could not be automated and had to be done step-by-step for each of the 15 polygon metrics. It is also a computationally expensive operation with each intersection taking several minutes. Line metrics were alo summarised at the SA2-scale but using a different statistical method approach. Line metrics in the MaP dataset can be divided into continuous and categorical.

Continuous metrics like crash rate were summarised using a weighted mean. Street segments of the same SA2 were grouped and the values were averaged by using the length of each street segment as a weight. Longer street segments had a larger impact on the SA2 value than smaller street segments. Overall, 9 metrics were summarised that way. The two categorical indicators were waterways and primary schools and described the distance between the street segment and the closest waterway or primary school. The distance was not continuous and was given as an interval such as between 400 and 800m. For each SA2, we computed the proportion of streets with a waterway or primary school within 800m. I picked the threshold of 800m as it corresponds to a 10-minute walk and was used in the ULI model (Higgs et al., 2019). Overall, 26 environmental indicators were selected from the MaP dataset and each was summarised for the 335 selected SA2 within Greater Sydney. The summary for the SA2 of Surry Hills is provided as an example (Table 2).

The two other datasets, ULI and K10, also had to be summarised at the SA2 scale. The ULI is available at the SA1 level for Greater Sydney on the AUO website. SA1 that form the same SA2 were combined and the average ULI was obtained with R. The 10-item Kessler Psychological Distress scale (K10) can be used to measure the risk of psychological distress at the individual level (Kessler et al., 2003). The K10 contains 10 questions each worth 5 points about anxiety and depressive symptoms felt by an individual. An individual with a score greater than 22 out of 50 is classified as experiencing high psychological distress. The community K10 score is available on the PHIDU website at the Population Health Area (PHA) scale. This indicator corresponds to the proportion of people with high psychological distress (K10 >22) in each PHA. PHA and SA2s are relatively similar with one-to-one one-to-two or one-to-three links between PHA and SA2s. Consequently, the proportion of highly distressed people can be summarised using K10 at the SA2 level. One limitation is that when the relationship is one-to-two or one-to-three, each of the

SA2_CODE21	117031336
SA2_NAME21	Surry Hills
UHI_mean	3.01
BuildingDensity_mean	2.67
BuildingHeight_mean	13.1
TreeCanopy_mean	0.207
PTAL_mean	98.8
PublicSpace_mean	8.64
PopDensity_mean	12117
LocalJobs_mean	41277
PopGrowth_mean	52538
EmploymentGrowth_mean	384
CultureHeritage_mean	391
HousingDiversity_mean	4
ModeShare_mean	0.799
ImperviousSurface_mean	92.8
CommunitySafety_mean	27837
LocalLiving_mean	3.83
LandDivision_mean	49.2
StreetSpacePedestrian_mean	35.2
RoadSafety_mean	2.27
PrimarySchool_mean	0.124
Waterways_mean	0.245
CrashRate_mean	11.1
BusStopDensity	0.289
StreetAspectRatio_mean	0.805
BusDelay_mean	134

Table 2 – Summarised environmental indicators for one SA2

two or three SA2s will have a similar K10 score but different environmental indicators variable from the MaP data. The ULI and community K10 distributions in Greater Sydney were visualised at the SA2 scale in QGIS.

4. Statistical Analysis with linear regression models

4.1 Overview of modelling strategy

After obtaining a clean dataset with values for each selected SA2 of Greater Sydney, I was able to fit different linear regression models. The three relationships I was interested in quantifying are MaP and ULI, ULI and community K10, and MaP and community K10. For the models involving MaP, I was interested in selecting the best combination of environmental indicators to measure the association with ULI and community K10 which were the outcome variables for the first and final model respectively. Multiple linear regression was used for those models while simple linear regression was fitted to measure the association between ULI and community K10.

4.2 Multiple Linear regression

4.2.1 Variable Transformation

Multiple linear regression is built on fundamental assumptions that ensure the validity of the results. The core assumption is that the relationship between each predictor and the outcome variable should be linear. We can first visualise the relationship using a scatter plot to test this assumption. The lack of a linear relationship can be treated by using some mathematical transformation on the dependent variable. A residual versus fitted plot can be obtained after fitting a simple linear regression model using each environmental indicator or its transformation to one of the outcomes (ULI or community K10). The plot can be used to assess linearity, homoskedasticity (variance of the residual is constant) and the absence of outliers for each model. Different logarithmic-based transformations were performed on each environmental indicator to maximise linearity while keeping the interpretability of the model in mind. When variables could not be transformed using a mathematical function, they were transformed into binary or more complex categorical variables using quantile or inflection points.

4.2.2 Variable Selection and model fitting

A correlation matrix was computed using the original data from MaP to assess the correlation coefficient between each predictor and ULI and ensure no multicollinearity between the predictors. The correlation matrix can be visualised using a correlogram and was used in combination with VIF to exclude indicators that can lead to multicollinearity. A step-by-step approach can be used to construct our final model to predict ULI or community K10 using the best combination of environmental indicators from MaP at the SA2 scale (Fig 3). First simple linear regression models were fitted individually using each remaining indicator to predict one of the outcome variables (ULI or K10). All the indicators with a p-value smaller than 0.25 at the univariate level, were included in a first multiple linear regression model. At every step, I used a function that removed one of the selected variables from the multiple regression model and computed the AIC. The AIC is an estimator of prediction error and can be used to compare the performance of similar models (Wagenmakers et al., 2004). The variable that led to the biggest decrease in AIC was removed to obtain a new model. Each model was then compared to check that there was not a change in either regression coefficient, adjusted R^2 or the number of observations that were too important and could justify the difference in AIC. These steps were repeated until dropping another variable did not lead to a reduction in AIC. Lastly, I checked whether adding any of the excluded variables due to having a p-value of 0.25 at the univariate levels would affect the performance of the model. Finally, the summary from the model can be interpreted to describe the relationship between the selected environmental indicators and the outcome variable.

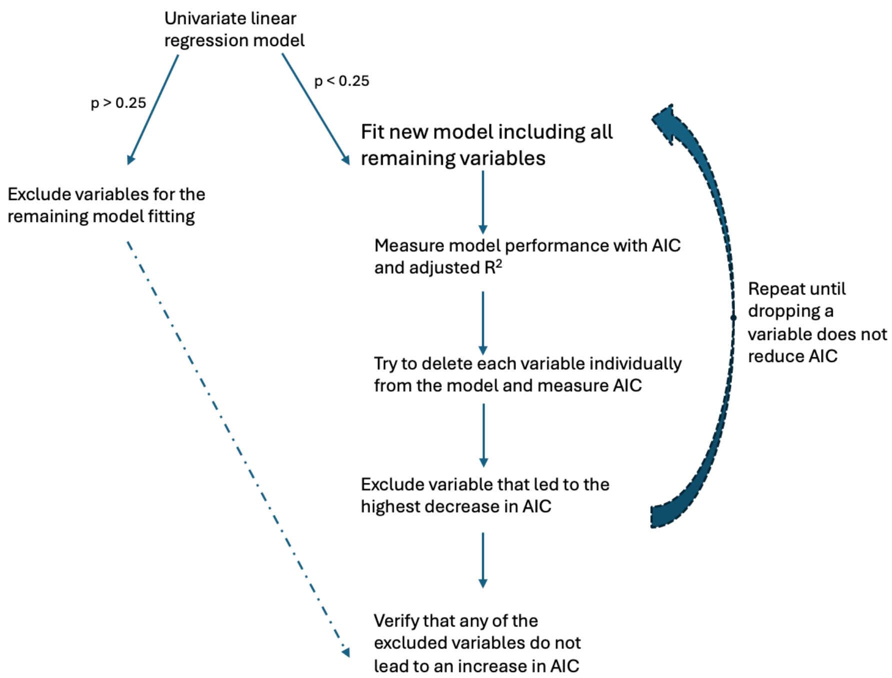


Figure 3 – Overview of indicators selection strategy for multiple linear regression models

4.2.3 Model interpretation, performance and sensitivity analysis

The different regression models were interpreted using the coefficients when the p-value was smaller than 0.05. When relevant, the data was scaled so that a multiple linear regression's different coefficients could be compared. Scaling is a data transformation technique to compare data that is not measured the same way and is appropriate here as the environmental indicators from MaP all have different units. The final model can be fitted again using the scaled data to compare the different coefficients. Model performance was assessed using adjusted R^2 where a score of 1 indicates a perfect fit of the model on the data. Sensitivity analysis was performed in the final model that looks at the association between indicators of MaP and community mental health. Sensitivity analysis is used to determine how the results of a model can be allocated to other sources such as unmeasured confounders. The Index of Economic Resources (IER), a measure of socio-economic status at the community level was introduced as a new variable in the multiple linear regression model (ABS, 2016). IER is an index released by the ABS that is available at the SA2 level and measures access to economic resources including income and housing. Each SA2 of Greater Sydney is assigned a number between 1 and 4 where a higher number indicates greater access to economic resources. Change in the sign of coefficients of the indicators when IER is included or excluded reveals how much the association is due to socio-economic factors. If the sign of the coefficient does not change and the indicators remain significant in the sensitivity mode, it provides robust evidence that there is an association between the urban environment and mental health at the SA2 level.

5. Ethical issues statement

Ethics approval is not required as the analysis is based on urban indicators from public datasets and area levels such as SA2 that cannot be used to identify specific individuals. A potential ethical issue would be if the results from the statistical model specifically name and identify a population with a high environmental factor that could be perceived negatively. A statement such as 'Crime density and crash rate are high in suburb X which lead to high psychological distress in the local population' could potentially be offensive. Instead, the results will be formatted more generally with statements such as 'SA2s with a high crime density and high crash rate tend to also have a high psychological distress at the community level'. The idea is to minimise the stigmatisation of neighbourhoods and suburbs by focusing more on the value of the spatial data than its precise location in Greater Sydney.

Results

1. Trends in summarised data at the SA2 level

1.1 Urban Liveability Index

The ULI ranges from 87.8 to 108.7 among the selected SA2s of Greater Sydney. The observed mean is 99.7 which is close to 100, the threshold defined for the AUO to separate high and low liveability scores. SA2s that are near a beach or bay tend to have a high liveability score. These SA2s are located in the eastern suburbs, Mosman and Northern beaches but also near Canda Bay, Inner West or the city of Sydney (Fig 4). SA2s of the West of Greater Sydney tend to have lower scores especially near the Blue Mountains or in the southwest near Camden. SA2s of the West also show a higher variability than SA2s of the East as seen in Paramatta where the ULI is similar to Coogee-Clovelly while the neighbour SA2s has one of the lowest ULI across Greater Sydney (Fig 5).

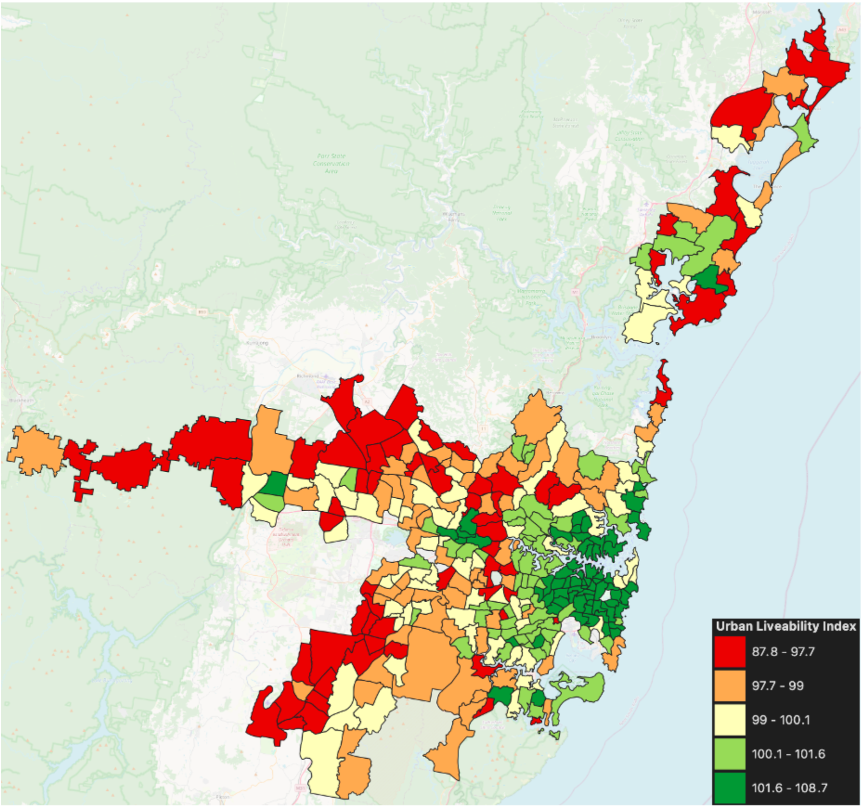


Figure 4 – Visualisation of ULI at the SA2 level in Greater Sydney using QGIS

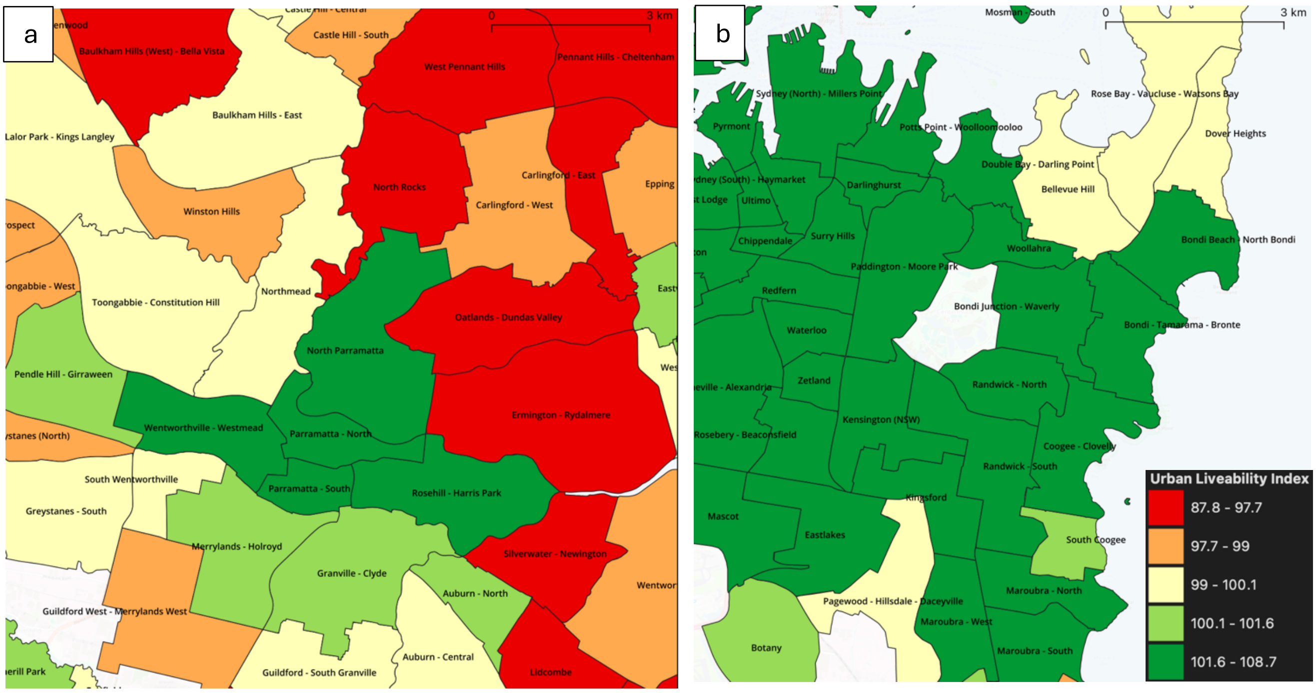


Figure 5 – Variability of ULI between SA2s of the East and West in Greater Sydney.
a) ULI in SA2s surrounding Paramatta in Western Sydney. b) ULI in SA2s of Sydney East

1.2 Community K10

The proportion of people with high psychological distress varies from 8.4% to 34.3% within the selected SA2s of Greater Sydney. SA2s of the East and North that have a high ULI tend to have a low proportion of highly distressed people (Fig 6). The highest proportion of distressed people tend to be in SA2s of the West near Liverpool, Fairfield, Parramatta or Blacktown. Spatial autocorrelation seems higher in the community K10 dataset than in the ULI. That is neighbour SA2s tend to have similar community K10 scores in all areas of Greater Sydney.

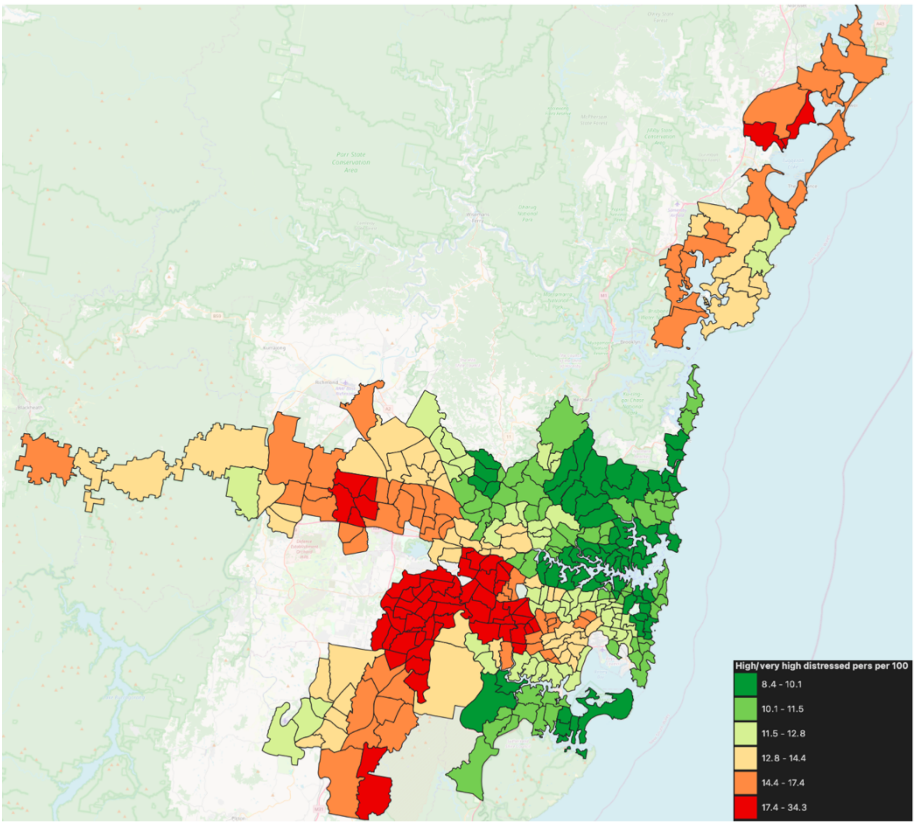


Figure 6 – Visualisation of community mental health at the SA2 level in Greater Sydney

2. Regression of MaP and ULI

Scatter plots to visualise the association between each indicator and the ULI were performed with R (Fig 7). If the scatter plot showed a linear pattern and the fitted vs residual plot showed a random distribution of the residuals near the y = 0 line, then the assumptions of linearity and homoskedasticity are respected. In such instances, the environmental indicators were not transformed. In the event of assumption violation, a log transformation was applied to adjust the linear pattern (Fig 8). If the indicator had a minimum value of 0, log(x+1) was used as log(0) is not defined mathematically (Changyong et al., 2014). Lastly, if log transformation could not lead to a linear pattern, the environmental indicator was categorised using several methods. The inflection point was determined visually for the urban heat and tree canopy indicators and the values were split into two categories, above or below the inflection point (Fig 9). Population growth was categorised into quartiles and the cycling lane indicator had a majority of zero so it was transformed to binary.

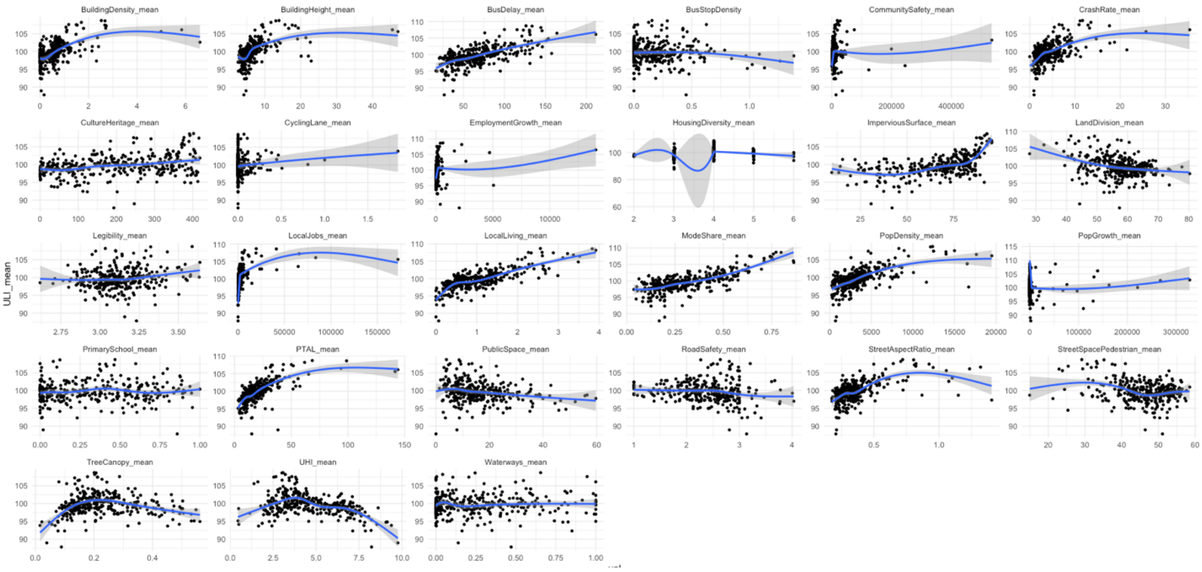


Figure 7 – Univariate association with environmental indicators and ULI

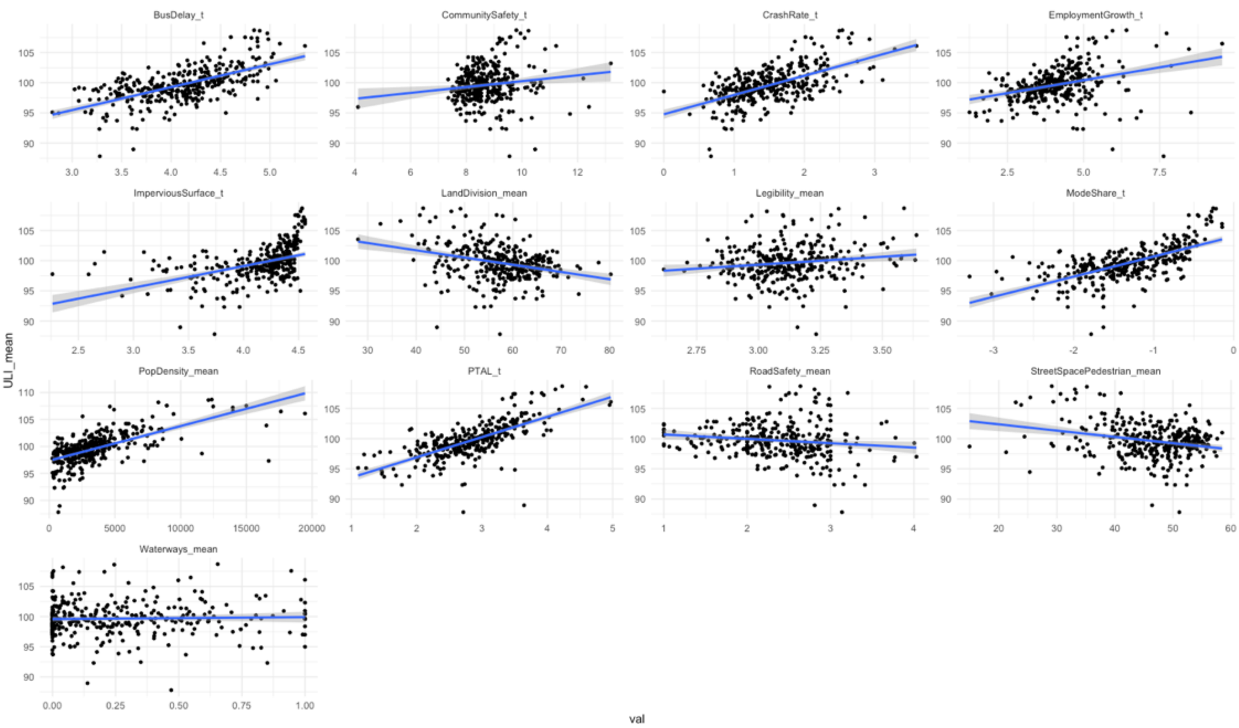


Figure 8 – Linear regression of log-transformed and non-transformed indicators with ULI

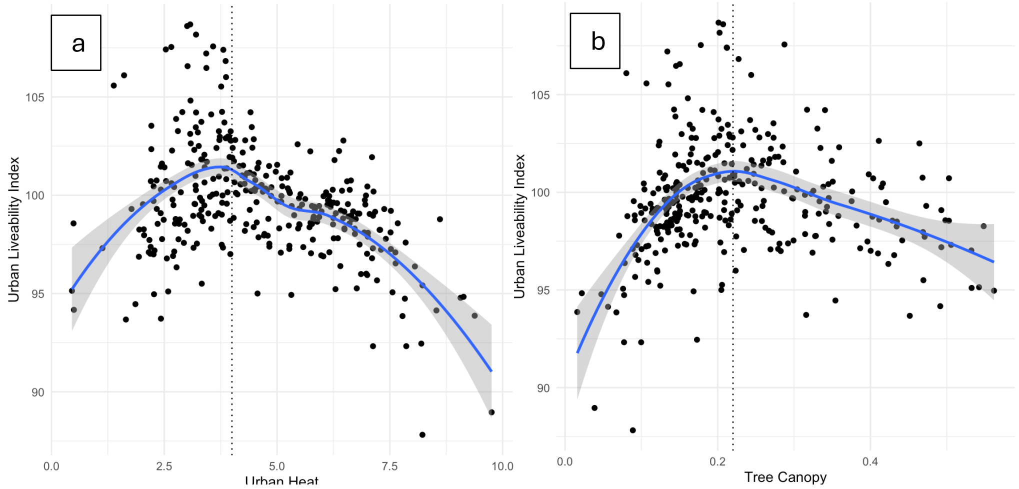


Figure 9 – Visualisation of inflection point to categorise urban heat and tree canopy

No transformation	x -> log(x)	x -> log(x+1)	Categorised
Pedestrian Space Legibility Culture Heritage Land Division Population Density Public Space Road Safety Waterways Primary School	Building Height Community Safety Employment Growth Impervious Surface Local Jobs Street Aspect Ratio Bus Delay Mode Share PTAL	Building density Crash Rate Local Living Bus Stop Density	UHI: inflection point (binary) Tree Canopy: inflection point (binary) Cycling Lane: 0 or >0 (binary) Population Growth: Categorised using quartiles

Table 3 – Summary of transformation applied to environmental indicators to respect linearity assumption

Multicollinearity, which occurs if two environmental indicators from MaP have a high correlation with one another, had to be avoided. The correlogram reveals that variable including Building Height and Building density tends to be highly positively correlated with other indicators (Fig 10). Land Division tend to also be negatively correlated with many variables. Ultimately, multicollinearity was assessed using VIF and any indicator with a VIF greater than five was omitted from the analysis.

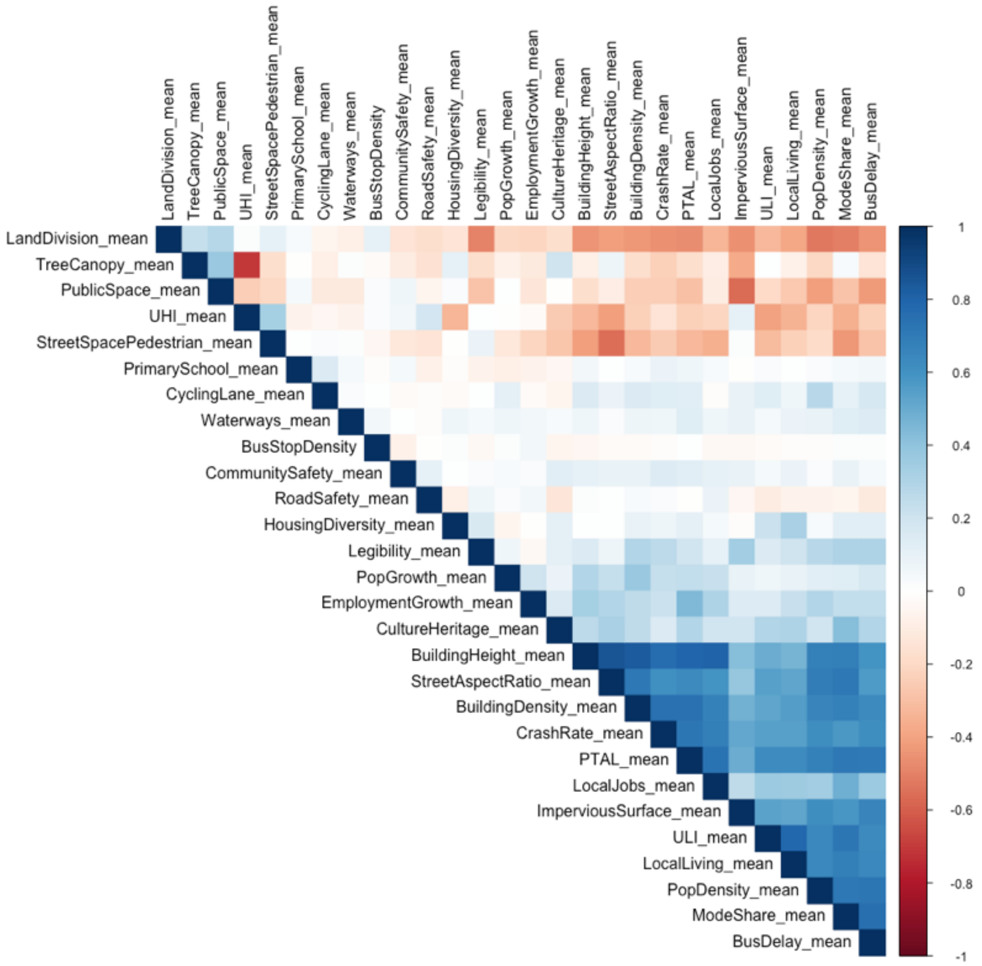


Figure 10 – Correlogram to visualise multicollinearity

Now that all variables that do not respect the model assumptions were limited, I could fit the linear regression model to predict the ULI using MaP. Out of all the indicators, 9 were selected in the modelling process to predict the ULI at the SA2 scale (Table 4). The adjusted R^2 is 0.70 which indicates a good model fit and performance. Indicators with a positive coefficient like crash rate, mode share, PTAL or Impervious surface are associated with an increase in ULI when the value of these indicators increases at the SA2 level.

Indicator (reference)	*Indicator Unit (range or categories)*	*Estimated coefficient*	*P-value*
Intercept	-	98.19	<0.01
Street Space Pedestrian	% (14.9 – 58.5)	-0.04	0.02
Legibility	Legibility level (2.6 – 3.6)	-1.62	0.01
Urban Heat Island (Low)	°C		
- High	(> 4)	-1.18	<0.01
Population Growth (Low)	%		
- Normal	(11-38)	0.08	0.77
- High	(38-167	-0.14	0.65
- Very High	(>167)	-1.01	<0.01
log(Crash Rate +1)	Crashes/km/year (0-35)	1.45	<0.01
log(Employment Growth)	% (3.5-14034)	-0.33	0.01
log(Mode Share)	% (0.04-0.86)	1.15	<0.01
log(PTAL)	Index (3-145)	2.01	<0.01
log(Impervious Surface)	% (9.6-95.4)	1.01	0.01
Overall	Index (70-130)		< 0.01

Table 4 – Summary of multiple linear regression model between indicators of MaP and ULI

For this model, I was more interested in the strength of the association than the direction. That is I wanted to determine which indicators from MaP played the most important role in model performance. R^2 and AIC score when each of the 9 variables is removed individually can be compared to the full model's performance to determine the weight of each indicator (Figure 11). The indicator with the highest impact on both model performance score

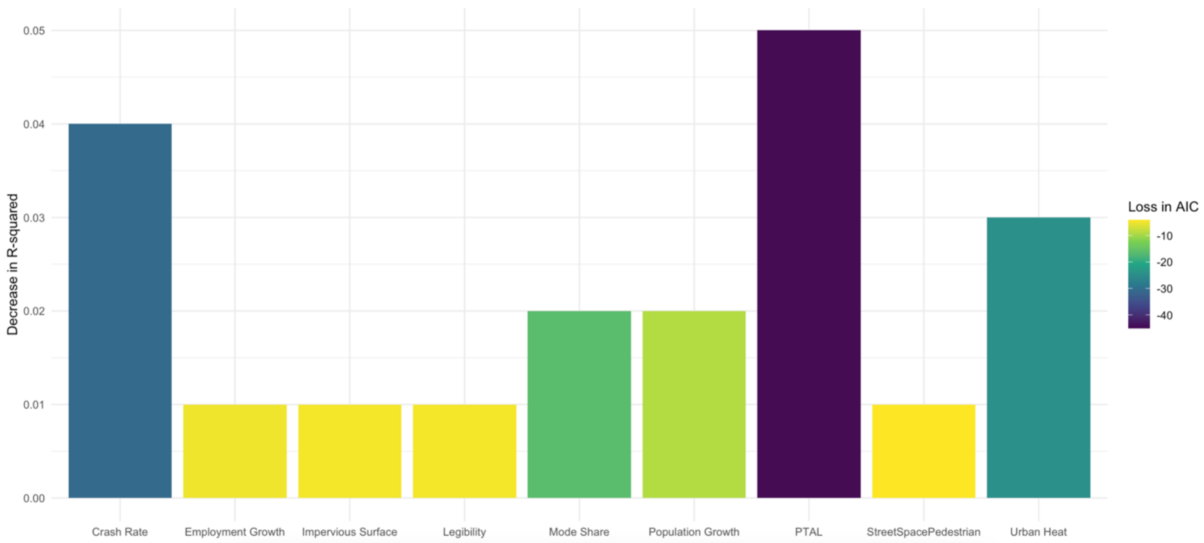


Figure 11 – Influence of each indicator on the model performance based on chance in R^2 and AIC

is PTAL (public transport accessibility level) followed by crash rate and urban heat.

3. Regression of ULI and community K10

I fitted two simple linear regression models which yielded similar results between urban liveability score and mental health in males and females (Fig 12). As the ULI of SA2 increases, the proportion of people with high psychological distress tends to decrease. That is as the liveability and quality of the urban environment improve in Greater Sydney, mental health at the community level tends to improve. The p-value for the coefficient and intercept in both models are close to zero and thus the results are significant. The two models are:

$$Male\ \widehat{community}\ K10\ = 32.93 - 0.23 \times\ ULI_{SA2}$$

$$Female\ \widehat{community}\ K10\ = 47.25 - 0.34 \times\ ULI_{SA2}$$

The intercept can be interpreted as the proportion of people with high psychological when the ULI of the SA2 is zero. The minimum value of the ULI in this dataset is 84 so the intercept cannot perfectly be interpreted but it reveals that overall, the proportion of high psychological distress is higher in females than males. The absolute value of the coefficient is 0.23 in males and 0.34 in females. Consequently, the same increase in ULI leads to a larger improvement in psychological distress in females than in males. Overall, SA2 with higher liveability scores had better community mental health. However, the liveability index contains over 40 different environmental indicators and this model does not inform us which indicator has the highest association with mental health.

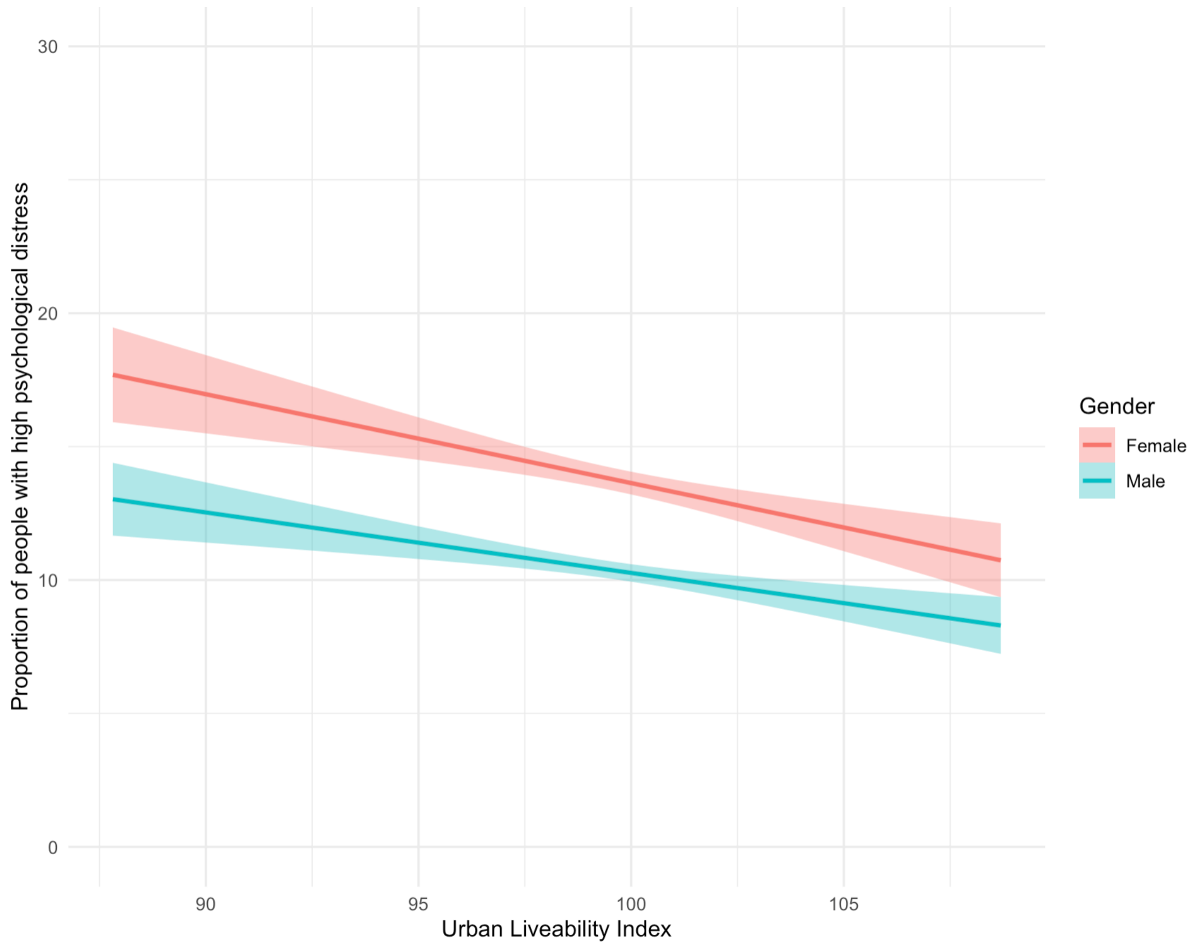


Figure 12 – Linear regression to measure the association with ULI and community mental health across genders

4. Regression of MaP and community K10

4.1 Final model without socio-ecoconmic status

Multiple linear regression and variable selection identified 6 indicators that have a significant association with mental health at the SA2 level (Table 5). The adjusted R^2 of the model was 0.5. This indicates that 50% of the variability in mental health can be explained by using only those 6 indicators from the urban environment at the SA2 level. This is the main finding of the placement and indicates that there is a strong and significant association between the urban environment and community mental health in Greater Sydney. Among the 6 variables, 4 are related to transport infrastructure while the 2 others are more related to the urban landscape.

Indicator	*Indicator Unit (range or categories)*	*Estimated coefficient*	*CI (95%)*	*P-value*
Overall				<0.01
Intercept		18.82	(15.07, 22.57)	<0.01
Mode Share	% (0.04-0.86)	-4.43	(-8.14, -0.73)	0.02
Tree Canopy	% (2.6 – 3.6)	-8.07	(-12.95, -3.19)	<0.01
Urban Heat Island	°C (0.4 – 9.8)	0.32	(0.03, 0.61)	0.03
log(Crash Rate +1)	Crashes/km/year (0-35)	4.65	(3.77, 5.52)	<0.01
log(BuildingHeight)	% (3.5-14034)	-4.04	(-5.44, -2.64)	<0.01
log(PTAL)	Index (3-145)	-1.18	(-2.13, -0.24)	0.01

Table 5 – Summary of multiple linear regression model between indicators of MaP and K10

All the environmental indicators have different metric units, some are log-transformed while some are not, so it is not evident which indicators have the greatest association with mental health here. Once the data was scaled, the standardised coefficient and their confidence interval could be compared with each other (Fig 13). Road crash rate is the indicator with the highest association with mental health in the model. SA2s with high crash rates tend to be SA2s where the proportion of high psychological distress based on K10 score is high. SA2s with high urban heat, that is SA2s that experience much warmer temperatures than rural areas of NSW are also associated with high psychological distress. The coefficient is still significant for urban heat using the 0.05 threshold for the p-value, but the effect on mental health is smaller than the road crash rate. SA2s where public transport was highly accessible and where the population used sustainable transport often correspond to SA2s with a low proportion of highly distressed people. A high proportion of tree canopy was one of the variables with the strongest association with low psychological distress at the community level. Surprisingly, high average building height is strongly associated with low psychological distress at the SA2 level.

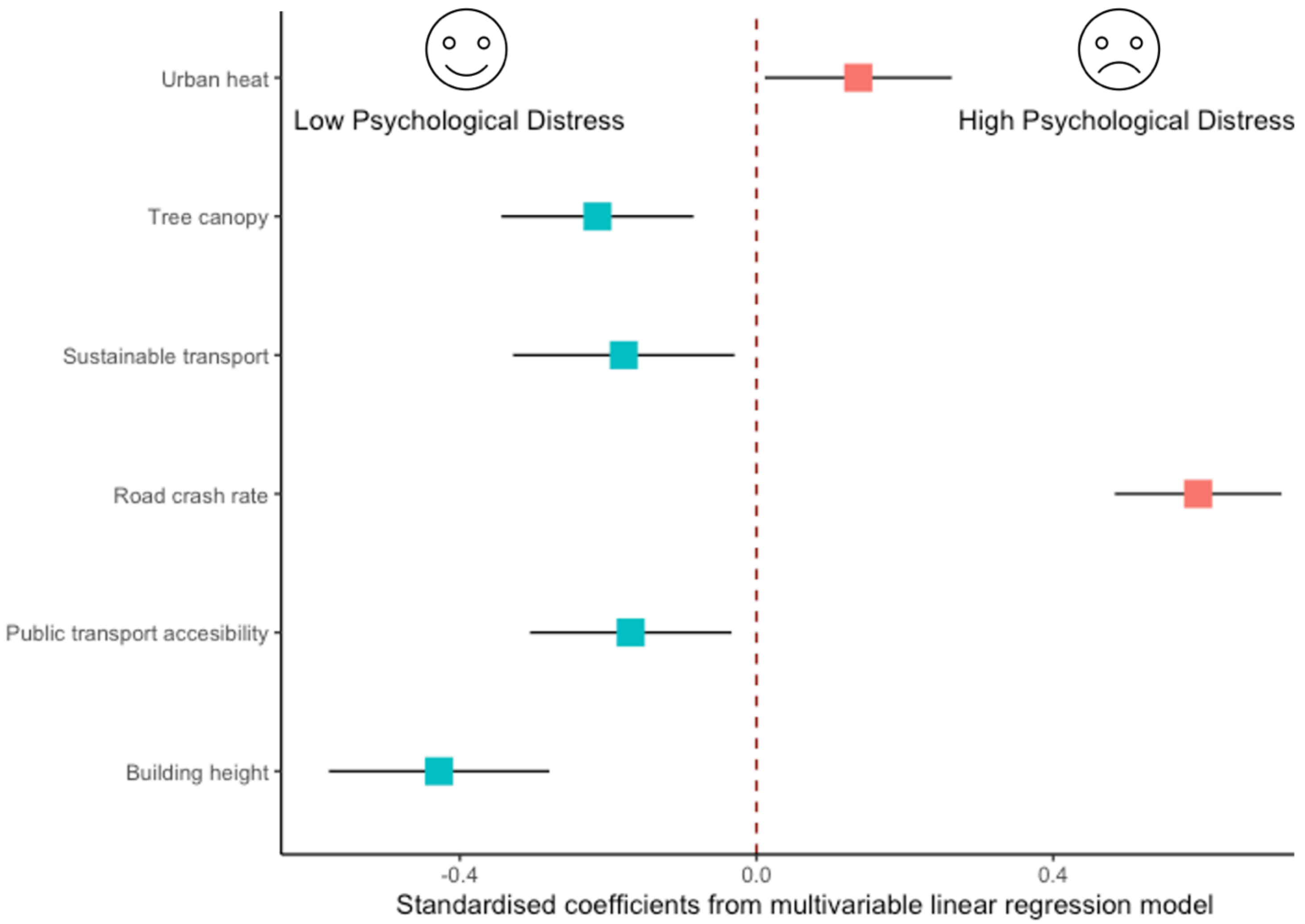


Figure 13 – Visualisation of coefficients and 95% confidence interval for the association between environmental indicators and community mental health using scaled data

4.2 Sensitivity analysis: adjusting for socio-economic factors

The aim of the placement was mainly to centre the analysis around the association between MaP and mental health while acknowledging that a lot of confounders are not included in the analysis. Socioeconomic status is known to be directly associated with mental health at the community level. The IER was obtained at the SA2 level and implemented into the final model as a sensitivity analysis (Table 6). As expected, the IER was significant in the model and greater access to economic resources was associated with lower psychological distress at the community level. Out of the 6 indicators, tree canopy was the only variable that did not remain significant. Mode share (sustainable transport), urban heat, public transport accessibility, building height and crash rate all had a significant association with community K10 score even when adjusting for socio-economic factors. The R^2 of the model was boosted from 0.50 to 0.68 indicating that 68% of the variability in community mental health can be explained by those variables and IER accounted for 18% of the variance. In the model, IER is the variable most associated with mental health. The overall effect of adding SES in the model is that the coefficients were scaled down but the direction of the effect remained the same. High urban heat and a high crash rate at the SA2 are still associated with a higher proportion of psychological distress. However, the variables play a smaller role in the association as some of the weight was transferred to the IER variable. Mode share, which measures the use of sustainable transport such as walking or cycling, had a higher effect on mental health in the sensitivity model.

Indicator (reference)	*Estimated coefficient*	*CI (95%)*	*P-value*
Overall			<0.01
Intercept	23.25	(20.10, 26.41)	<0.01
Mode Share	-7.61	(-10.66, -4.55)	<0.01
Tree Canopy	-2.72	(-6.80, 1.36)	0.19
Urban Heat	0.43	(0.19, 0.67)	<0.01
log(PTAL)	-1.01	(-1.76, -0.25)	0.01
log(Building Height)	-2.99	(-4.16, -1.83)	<0.01
log(Crash rate +1)	2.09	(1.30, 2.89)	<0.01
IER Quartile (1)			
- 2	-3.02	(-3.82, -2.23)	<0.01
- 3	-4.78	(-5.60, -3.97)	<0.01
- 4	-6.03	(-6.93, -5.13)	<0.01

Table 6 – Summary of sensitivity analysis between indicators of MaP and K10 by adjusting for socioeconomic factors

Strengths and limitations

This placement was centred around the MaP dataset to explore the relationships with other datasets, yet the results from the models are significant and the analysis appeared robust. The spatial area for the models was rigorously selected to obtain meaningful results in an urban environment. Including low-density SA2s may have exaggerated the effect of some urban indicators on community mental health. The main outcome variable, community mental health, was quantified using K10. The K10 is a meaningful and trusted indicator compared to other census-based surveys which have been criticised in Australia (Klein et al., 2020). The MaP dataset has also been meticulously summarised at the SA2 scale by using different spatial statistics transformations. For socioeconomic factors, IER was used which is a trusted index developed by the ABS to summarise information at the community level. Poor quality data can often be a limited factor in implementing the results of statistical models (Haug et al., 2011). However, obtaining data of high quality at the SA2 level was a priority in this placement, especially for MaP where preparing the data for the analysis was quite time-consuming. The strength of the findings relies on the quality of the data but also on the rigorous statistical methods that were applied. Each variable was individually analysed to determine if it needed to be transformed and was excluded from the analysis when including it would violate any of the modelling assumptions. Regarding the model fitting, each combination of variables to measure the association with one of the outcomes was tested individually. Methods like stepwise regression that automatically find the best combination in one step were avoided. These methods have been criticised for yielding results that can be incorrect (Gary, 2018).

Using statistical methods that produce clear and understandable results, applicable to public health, also strengthens the findings. Methods like linear regression allow the results to be easily interpretable. Other machine learning algorithms like k-means clustering may have yielded better model performance to predict the association between environmental indicators and mental health. Although some performance might be lost by using linear regression due to the assumption made on the data distribution, the results are easily interpretable. I also tried to interpret the results from the different regressions in a very general way. Instead of having a predictive approach by quantifying every effect, the results are presented broadly with statements such as ‘SA2s with high Crash rate also tend to have a high proportion of distressed people’. Overall, the strength of the project relies on resisting the urge to use more advanced and intriguing techniques and applying simple techniques correctly and rigorously to yield interpretable and simple results.

There are some limitations to this project and they should be fully acknowledged. The main limitations in this project can come from the data or the methods that were selected to analyse the data. The project is built around the MaP dataset where no previous analysis has been published to check whether the data can be useful in a public health context. A lot of confounders for mental health are not included in the model and we know it by design. The absence of socio-economic factors in the model was the biggest limitation but they were implemented in the model during the final week of the placement. However, a lot of other confounders are not included. A lot of individual socioeconomic factors such as level of education, relationship status, or salary are highly associated with psychological distress (Mair et al., 2008). Summarising these individual factors for SA2s with dense populations may be challenging. However, implementing these confounders in some ways in the model could potentially scale down the effect of numerous indicators and make them not significant as it was observed with tree canopy when socioeconomic factors were included. Another possible limitation in the use of the data is whether summarising the information at the SA2 scale which contains an average of 10,000 people could be a problem for some indicators. In general, it should not be a limitation, but it can cause problems when we are focusing on the results of the model in a small and diverse area. This can be the case in many gentrified SA2s where diverse people in terms of socioeconomic factors cohabit in the same area. Future research could potentially check if fitting the same model at a smaller scale like SA1 which contains about 400 people would yield the same results.

Regarding the statistical methods, implementing multiple linear regression on spatial data might have some limitations. All the statistical assumptions that could be tested were respected such as linearity, homoscedasticity or normality of residuals. Another fundamental assumption of linear regression is that each observation, that is each SA2 should be independent. However, due to the division of Sydney into Councils and LGA, we can expect nearby SA2s to have similar environmental indicators. In such instances, the observations would not be independent. This was partly tested for using Moran's Index which tests for spatial autocorrelation (see Appendix). This could bias the results toward or away from the null hypothesis depending on each indicator. Other models that are designed for spatial data may have been more statistically appropriate, however they are harder to implement than linear regression.

Outcomes and significance

Overall, two different outcomes of the project are significant for future work conducted by PHRAME for the Sydney Local Health District or The University of Sydney. The first outcome is the data cleaning and summary at the SA2 scale performed on three different datasets. The MaP dataset has not previously been used in any published analysis. Furthermore, it is only available as a spatial object in a very dense format and at a small geospatial scale, making it hard to use it straight away for analysis. However, I was able to summarise all the indicators neatly and completely at the SA2 level. The data could be very helpful for PHRAME if any of these environmental indicators need to be included as a variable in a different model. The method used for cleaning the data is also very valuable. Indeed, the R code can easily be modified to either summarise the same indicator at a different scale like SA1 or SA3 or to summarise different spatial objects at the SA2 level. The code that summarises each indicator of MaP could also be applied to other urban areas of NSW that have been excluded from the analysis like Wollongong or Newcastle. Having some statistics available at different levels is always useful in epidemiology, so the summary of both the ULI and community K10 at the SA2 scale could also easily be used for other research by PHRAME. Consequently, even when omitting the different regression models, the data cleaning process can be quite significant to help any future research by avoiding these time-consuming tasks. All of the indicators that have been summarised at the SA2 scale can be used for future analysis.

The second outcome of this project which is significant for PHRAME is the results from the three different regression models. A promising association was found between environmental indicators and mental health at the community level in Greater Sydney which is relevant for the Sydney Local Health District. Moreover, a list of five environmental indicators can be provided which are significant even when adjusting for socioeconomic factors. While there is no implied causation between these indicators and health outcomes, they could be considered individually to guide future urban planning projects in Greater Sydney that try to take mental health into account when presenting new projects. These 5 indicators could conceivably be used as a proxy for mental health which can be hard to collect at a large scale. Collecting mental health data is often time-consuming, and expensive and the data can largely be a misrepresentation of the true community mental health as the level of psychological distress in Australia is often underestimated (Klein et al., 2020). Furthermore, further investigation may establish similar relationships with other health outcomes, suggesting that environmental indicators from MaP could potentially be used as a proxy for health outcomes that are hard to collect.

Critical reflection on the placement

During my placement with PHRAME, I applied different spatial analysis methods to study the relationship between the urban environment and mental health at the community level. I learned to use QGIS, which was new and improved my coding skills with R. I worked with different spatial data with which I had little experience. I was planning on learning more about spatial data for a while as they are often used in epidemiology, so the placement was a great opportunity. I enjoyed the early stages of data cleaning to summarise the indicators of MaP at the SA2 level. It required a lot of critical thinking to ensure that the use of different summary methods such as weighted mean was appropriate. It also was a great opportunity to familiarise myself with QGIS and visualising spatial data which was quite challenging in the early stages. I had to try the different settings of the software to perform the spatial operations that I was looking for and do a lot of background research in spatial. Working in person at the Public Health Unit with Joe was quite helpful in overcoming different technical issues. As I was working in the same office as him, he was able to guide me through the different steps to follow with QGIS. It was also very valuable to do some days of the placement in person, as I could easily present my code with the different findings of the analysis. We were able to have back-and-forth discussions to decide what statistical methods were the most appropriate, whether there were any limitations and what possible improvements could be made to the analysis. No previous analysis or data cleaning had been done with the MaP data and I was not familiar with analysing spatial data with QGIS and R. Consequently, it was hard for my supervisors and myself to determine how much of the work I was going to be able to do within the semester. This led us to have a very simple approach in the early stages of the placement and then try to account as much as possible for the limitations as we realised that I would have enough time to perform all the objectives. I had a meeting every Friday with Luke and Joe to talk about the different findings and present the work that I did during the week. This was very valuable as I learnt a lot about how to present statistical models in a public health context. After all the analysis skills that I developed during this placement, this is the most significant thing that I learnt. From a mathematical and statistical perspective, I have been taught about maximising the performance and accuracy of models. From a public health perspective, the approach was very different. Of course, the models needed to be accurate and useful. However, the emphasis was more on simplifying the findings and ensuring that they could be explained in a very succinct and general way. Overall, this placement was really enjoyable, I am grateful for this opportunity, and I am looking forward to working on other statistical analysis projects in a public health context.

Appendix

Code Availability

MaP data cleaning
ULI data cleaning
MaP vs ULI analysis
Community K10 analysis

Supplementary files

Metadata MaP
Notebook during placement
Early Project description
Summary of statistical analysis

Additional methods and findings

- Table One

The ULI variable was categorised into normal, high and low liveability index to obtain descriptive statistics of the different environmental indicators from MaP and see if any association with ULI can be observed using a table.

	Low (N=100)	Normal (N=129)	High (N=106)
CrashRate_mean			
Mean (SD)	2.68 (1.59)	4.01 (2.88)	6.41 (4.98)
Median [Min, Max]	2.12 [0.446, 7.25]	3.33 [0, 21.5]	5.12 [1.10, 35.3]
BusStopDensity			
Mean (SD)	0.222 (0.261)	0.204 (0.240)	0.165 (0.180)
Median [Min, Max]	0.159 [0, 1.26]	0.131 [0, 1.38]	0.116 [0, 0.665]
Missing	4 (4.0%)	1 (0.8%)	2 (1.9%)
StreetAspectRatio_mean			
Mean (SD)	0.293 (0.130)	0.308 (0.120)	0.474 (0.224)
Median [Min, Max]	0.271 [0.176, 1.41]	0.282 [0.183, 1.15]	0.407 [0.204, 1.32]
BusDelay_mean			
Mean (SD)	50.1 (20.7)	60.2 (25.1)	92.0 (33.2)
Median [Min, Max]	45.9 [16.4, 131]	56.2 [20.6, 158]	88.6 [32.0, 212]
Legibility_mean			
Mean (SD)	3.09 (0.141)	3.11 (0.172)	3.16 (0.177)
Median [Min, Max]	3.08 [2.70, 3.50]	3.10 [2.62, 3.63]	3.14 [2.79, 3.64]
CyclingLane_mean			
Mean (SD)	0.0170 (0.0395)	0.0400 (0.107)	0.0656 (0.219)
Median [Min, Max]	0 [0, 0.193]	0 [0, 0.807]	0 [0, 1.85]

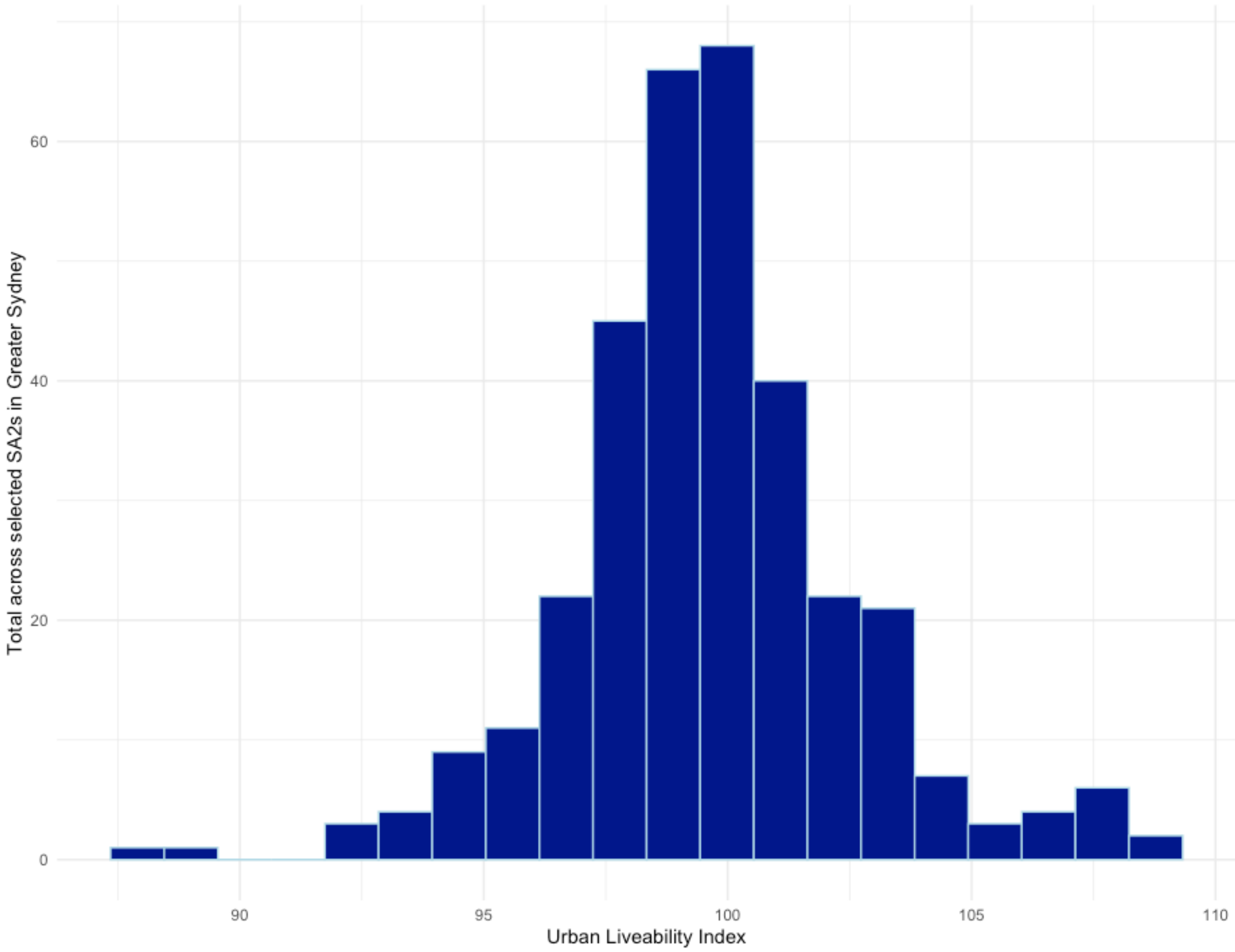
	Low (N=100)	Normal (N=129)	High (N=106)
ULI_mean			
Mean (SD)	96.6 (1.91)	99.5 (0.603)	103 (2.06)
Median [Min, Max]	97.2 [87.8, 98.5]	99.5 [98.5, 100]	102 [101, 109]
UHI_mean			
Mean (SD)	5.26 (2.30)	4.88 (1.61)	3.85 (1.05)
Median [Min, Max]	5.54 [0.454, 9.76]	5.09 [0.489, 8.61]	3.83 [1.38, 7.10]
BuildingDensity_mean			
Mean (SD)	0.294 (0.356)	0.450 (0.395)	1.06 (1.02)
Median [Min, Max]	0.160 [0, 2.32]	0.470 [0, 2.14]	0.779 [0, 6.59]
BuildingHeight_mean			
Mean (SD)	4.84 (2.18)	5.42 (2.55)	9.38 (6.36)
Median [Min, Max]	4.46 [2.56, 22.9]	4.97 [2.79, 22.1]	7.43 [3.16, 47.0]
TreeCanopy_mean			
Mean (SD)	0.228 (0.147)	0.212 (0.101)	0.224 (0.0784)
Median [Min, Max]	0.169 [0.0160, 0.562]	0.174 [0.0817, 0.503]	0.212 [0.0803, 0.503]
PTAL_mean			
Mean (SD)	12.1 (5.89)	16.0 (5.25)	33.1 (21.7)
Median [Min, Max]	11.1 [3.01, 38.3]	15.4 [6.13, 32.2]	28.5 [6.65, 144]
PublicSpace_mean			
Mean (SD)	17.8 (11.9)	17.8 (9.90)	12.3 (7.83)
Median [Min, Max]	15.2 [0.0162, 59.6]	15.6 [2.83, 56.4]	10.7 [1.31, 42.9]
PopDensity_mean			
Mean (SD)	2090 (1970)	2910 (1760)	5720 (3500)
Median [Min, Max]	1760 [243, 16700]	2710 [323, 11600]	4900 [433, 19500]
LocalJobs_mean			
Mean (SD)	796 (903)	1370 (1730)	8440 (20800)
Median [Min, Max]	530 [37.8, 5190]	887 [201, 14900]	2760 [280, 172000]
PopGrowth_mean			
Mean (SD)	3850 (17800)	3960 (22400)	11600 (47600)
Median [Min, Max]	27.9 [-6.00, 130000]	26.9 [-9.80, 221000]	66.3 [-67.5, 328000]
EmploymentGrowth_mean			
Mean (SD)	152 (551)	69.4 (92.2)	357 (1460)
Median [Min, Max]	37.8 [3.50, 5050]	44.3 [5.33, 644]	103 [6.67, 14000]

	Low (N=100)	Normal (N=129)	High (N=106)
CultureHeritage_mean			
Mean (SD)	171 (135)	162 (119)	256 (112)
Median [Min, Max]	129 [0, 413]	145 [0.0487, 397]	267 [1.79, 419]
HousingDiversity_mean			
Mean (SD)	3.55 (0.781)	3.78 (0.704)	3.98 (0.433)
Median [Min, Max]	3.01 [2.00, 6.00]	4.00 [2.01, 6.00]	4.00 [3.00, 5.00]
ModeShare_mean			
Mean (SD)	0.207 (0.0995)	0.264 (0.107)	0.471 (0.169)
Median [Min, Max]	0.194 [0.0369, 0.533]	0.246 [0.0779, 0.576]	0.477 [0.0632, 0.864]
ImperviousSurface_mean			
Mean (SD)	57.4 (19.6)	65.6 (16.3)	77.9 (14.4)
Median [Min, Max]	57.1 [9.62, 91.5]	70.5 [14.1, 86.7]	81.3 [15.4, 95.4]
CommunitySafety_mean			
Mean (SD)	10400 (27100)	8560 (9350)	16400 (55000)
Median [Min, Max]	4770 [61.3, 245000]	5170 [1750, 46900]	6320 [1550, 534000]
LocalLiving_mean			
Mean (SD)	0.539 (0.376)	0.924 (0.426)	1.73 (0.772)
Median [Min, Max]	0.476 [0, 1.47]	0.874 [0.154, 2.52]	1.55 [0.440, 3.92]
LandDivision_mean			
Mean (SD)	58.7 (6.75)	59.6 (6.51)	53.3 (7.61)
Median [Min, Max]	59.1 [41.9, 80.2]	60.1 [39.4, 77.7]	53.8 [28.0, 69.2]
StreetSpacePedestrian_mean			
Mean (SD)	46.4 (6.17)	47.9 (7.06)	44.0 (7.88)
Median [Min, Max]	46.8 [21.3, 57.5]	50.1 [15.0, 57.6]	44.6 [22.9, 58.5]
RoadSafety_mean			
Mean (SD)	2.56 (0.516)	2.29 (0.583)	2.30 (0.564)
Median [Min, Max]	2.60 [1.19, 4.03]	2.26 [1.00, 4.01]	2.32 [1.00, 3.67]
Missing	7 (7.0%)	10 (7.8%)	6 (5.7%)
PrimarySchool_mean			
Mean (SD)	0.309 (0.280)	0.330 (0.289)	0.351 (0.277)
Median [Min, Max]	0.233 [0, 0.953]	0.275 [0, 1.00]	0.362 [0, 0.937]
Missing	8 (8.0%)	12 (9.3%)	3 (2.8%)
Waterways_mean			
Mean (SD)	0.264 (0.294)	0.263 (0.270)	0.287 (0.284)
Median [Min, Max]	0.188 [0, 1.00]	0.189 [0, 1.00]	0.214 [0, 1.00]
Missing	7 (7.0%)	2 (1.6%)	3 (2.8%)

<u>Supplementary Table 1 – Statistical description of each MaP indicator for low, normal and high ULI at the SA2 level in Greater Sydney</u>

- ULI histogram

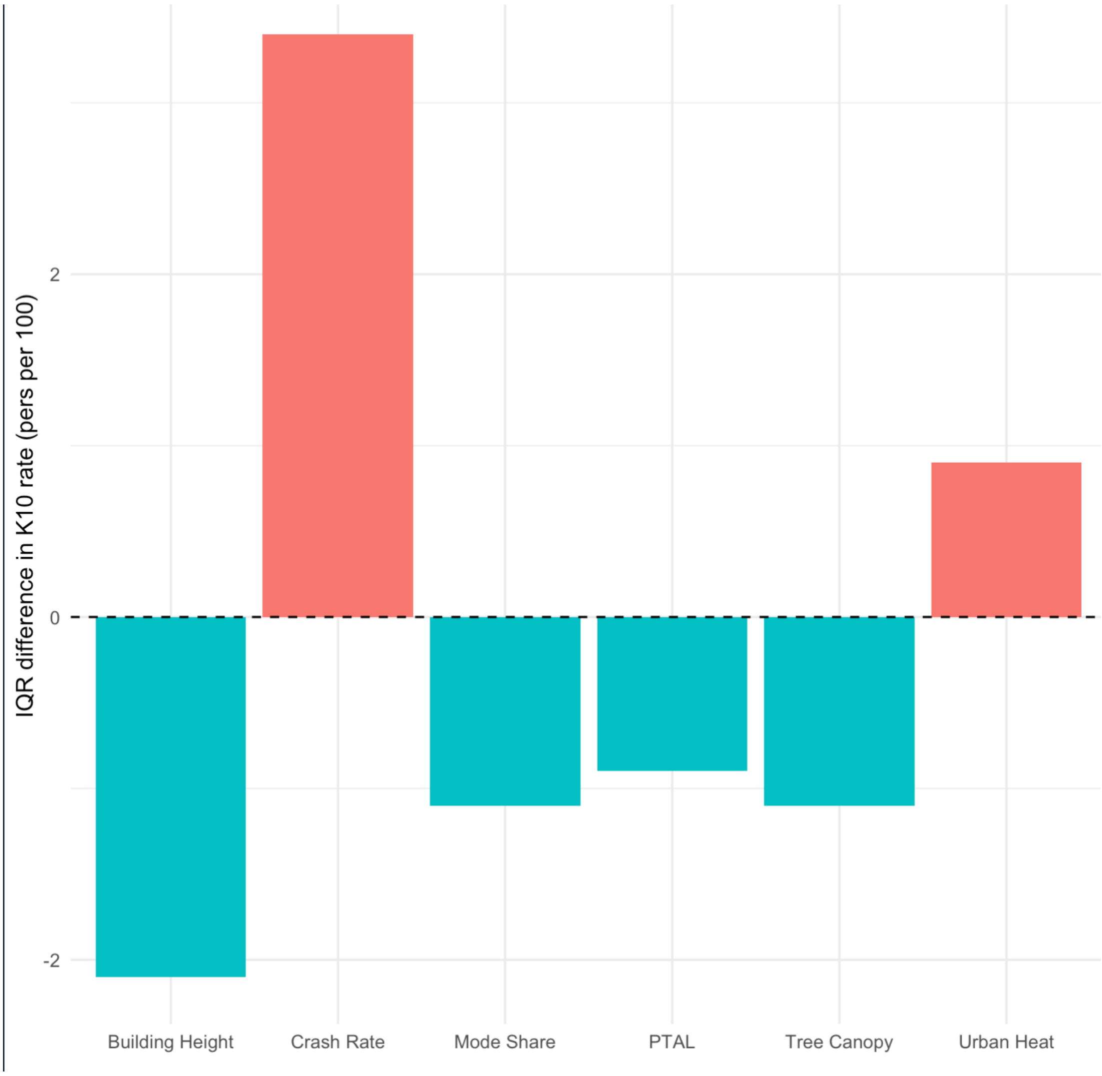
The normality of the dependent variable is not an assumption in linear regression, still, a histogram was fitted with R to visualise the distribution of the ULI. As expected from Higgs et al (2019), the variable is normally distributed with a mean near 100.



Supplementary Figure 1 – Distribution of ULI at the SA2 level in Greater Sydney

- Visualisation of model findings

Instead of using standardised coefficients to investigate the effect of each indicator on mental health, I plotted the change in average K10 score across the interquartile range (IQR). The interpretation is the same with crash rate and urban heat associated with a higher proportion of distressed persons.



Supplementary Figure 2 – Effect of each indicator on community mental health

- Moran's Index

Moran's Index was used to check if nearby SA2s were strongly spatially autocorrelated which could affect the generalisation of the results. The p-value is smaller than 0.05 and we can reject the null hypothesis. Spatial autocorrelation is not an issue in the model.

Moran I statistic standard deviate	Observed Moran I	Expected Moran I	p-value
4.86	0.164	-0.015	< 0.001

Supplementary Table 2 – Summary of Moran test between MaP and ULI

References

ABS, 2016. Socio-economic indexes for areas (SEIFA), Australia. Available at: https://www.abs.gov.au/statistics/people/people-and-communities/socio-economic-indexes-areas-seifa-australia/latest-release#index-of-economic-resources-ier

ABS,2021. Statistical Area Level 2. *Australian Statistical Geography Standard (ASGS) Edition 3.*Available at: https://www.abs.gov.au/statistics/standards/australian-statistical-geography-standard-asgs-edition-3/jul2021-jun2026/main-structure-and-greater-capital-city-statistical-areas/statistical-area-level-2

ABS, 2024. Capital city growth the highest on record. Media Release. Available at: https://www.abs.gov.au/media-centre/media-releases/capital-city-growth-highest-record#:~:text=Melbourne%20(up%20167%2C500)%20and%20Sydney,of%20the%20series%20in%201971.

Astell-Burt, T. and Feng, X., 2019. Association of urban green space with mental health and general health among adults in Australia. JAMA network open, 2(7), pp.e198209-e198209.

AUO, 2022. Urban Liveability Index. *Australian Urban Observatory.* Available at: https://auo.org.au/portal/metadata/urban-liveability-index/

Changyong, F.E.N.G., Hongyue, W.A.N.G., Naiji, L.U., Tian, C.H.E.N., Hua, H.E. and Ying, L.U., 2014. Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, *26*(2), p.105.

UN, 2015. World urbanization prospects: the 2014 revision by United Nations. *Department of Economic and Social Affairs in New York*

European Union and Joint Research Centre, 2008. Handbook on constructing composite indicators: methodology and user guide. OECD publishing.

Flavel, J., McKee, M., Tesfay, F.H., Musolino, C., Freeman, T., Van Eyk, H. and Baum, F., 2022. Explaining health inequalities in Australia: the contribution of income, wealth and employment. Australian Journal of Primary Health, 28(6), pp.474-481.

Gary, S., 2018. Step away from stepwise. *Journal of Big Data*, *5*(1), pp.1-12.

Higgs, C., Badland, H., Simons, K., Knibbs, L.D. and Giles-Corti, B., 2019. The Urban Liveability Index: developing a policy-relevant urban liveability composite measure and evaluating associations with transport mode choice. International journal of health geographics, 18, pp.1-25.

Higgs, C., Alderton, A., Rozek, J., Adlakha, D., Badland, H., Boeing, G., Both, A., Cerin, E., Chandrabose, M., De Gruyter, C. and De Livera, A., 2022. Policy-relevant spatial indicators of urban liveability and sustainability: Scaling from local to global. Urban Policy and Research, 40(4), pp.321-334.

Kessler, R.C., Barker, P.R., Colpe, L.J., Epstein, J.F., Gfroerer, J.C., Hiripi, E., Howes, M.J., Normand, S.L.T., Manderscheid, R.W., Walters, E.E. and Zaslavsky, A.M., 2003.Screening forserious mental illness in the general population. Archives of general psychiatry, 60(2), pp.184-189.

Klein, J.W., Tyler-Parker, G. and Bastian, B., 2020. Measuring psychological distress among Australians using an online survey. *Australian journal of psychology*, *72*(3), pp.276-282.

Krefis, A.C., Augustin, M., Schlünzen, K.H., Oßenbrügge, J. and Augustin, J., 2018. How does the urban environment affect health and well-being? A systematic review. Urban science, 2(1), p.21.

Mair, C., Roux, A.D. and Galea, S., 2008. Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. *Journal of Epidemiology & Community Health*, *62*(11), pp.940-946.

MaP, 2021. Movement and Place. *NSW Government*. Available at : https://www.movementandplace.nsw.gov.au/

Moore, T.H., Kesten, J.M., López-López, J.A., Ijaz, S., McAleenan, A., Richards, A., Gray, S., Savović, J. and Audrey, S., 2018. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health & place, 53, pp.237-257.

PHIDU, 2021. Social Health Atlas of Australia Population Health Area. *Data Workboos from Torrens University Australia*. Available at: https://phidu.torrens.edu.au/social-health-atlases/data#social-health-atlas-of-australia-population-health-areas

Reuben, A., Manczak, E.M., Cabrera, L.Y., Alegria, M., Bucher, M.L., Freeman, E.C., Miller, G.W., Solomon, G.M. and Perry, M.J., 2022. The interplay of environmental exposures and mental health: setting an agenda. Environmental health perspectives, 130(2), p.025001.

Wagenmakers, E.J. and Farrell, S., 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*, pp.192-196.

World Health Organization, 2016. Health as the pulse of the new urban agenda: United Nations conference on housing and sustainable urban development, Quito, October 2016.