AUSTRALIAN URBAN HEALTH INDICATORS PROJECT

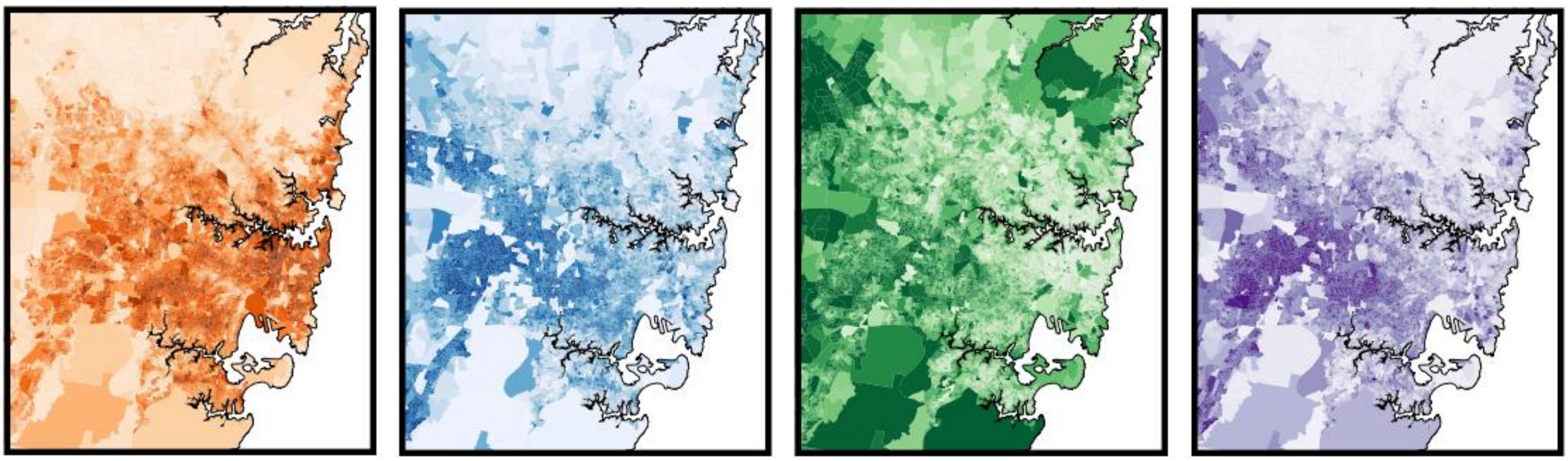
(AusUrb-HI)

Cross-NCRIS National Data Assets Program

Methodology and Outcomes Report

Authored by

Dr Hao Chen, Dr Aiden Price, Ms Flavia Barar, Dr Derrick Lopez, and Dr Moein Mehrolhassani



AURIN
aurin.org.au

CONTENTS

1. CASE STUDY INFORMATION

1.1. BACKGROUND

The Australian Urban Health Indicators (AusUrb-HI) project is an initiative aiming to enhance our understanding of urban health dynamics by developing new data assets. This venture encompasses three pivotal case studies: Cancer Determinants, Heat Health Vulnerability, and Urban Liveability & Health. These studies aim to amalgamate data from diverse domains like health, socio-economic status, the environment, climate, and urban planning to offer a comprehensive, spatially-explicit insight into urban population health. The resultant indicators are designed to assist health professionals, urban planners, and policymakers in crafting precise, targeted strategies. This report delves into the Heat Health Vulnerability case study, detailing the datasets used, the methodologies applied, the outcomes of the study, its impacts, lessons learned, and future directions.

Table 1.1: Project participants and contact information.

Lead Organisation	Australian Urban Research Infrastructure Network (AURIN) *University of Melbourne*
Partner Organisations	Queensland University of Technology (QUT) Royal Melbourne Institute of Technology (RMIT) University of Western Australia (UWA)
Case Study Contact Persons	Moein Mehrolhassani *moein.mehrolhassani@unimelb.edu.au*

Extreme heat is increasingly recognized as a significant public health hazard. In Australia, heatwaves are the leading cause of death among all natural hazards. The anticipated rise in extreme heat events due to climate change, coupled with a projected population surge to around 49.2 million by 2066, poses a significant challenge to the nation's health infrastructure. This challenge is compounded by rapid urbanization, densification, and the escalating demand for new housing in Australian urban areas. These factors underscore the critical need for climate-sensitive urban planning and design strategies to mitigate the health impacts of extreme heat.

Current heat vulnerability indicators lack the integration of detailed health data and often overlook urban morphology aspects crucial in assessing population vulnerability to extreme heat. This study introduces an advanced Heat Health Vulnerability Index (HHVI), leveraging linked health data, demographic factors, environmental conditions, and urban morphology to derive comprehensive spatial layers at a fine-grained level (SA1). These layers encompass exposure, sensitivity, and adaptive capacity, pinpointing regions particularly susceptible to heat and examining the interplay between human health and the built environment. A case study in New South Wales, Australia, demonstrates the indicator's effectiveness in guiding future urban planning to enhance health outcomes and promote climate-resilient urban spaces.

1.2. ACHIEVEMENT OF CASE STUDY AIMS

This study successfully achieves its primary objective: developing an HHVI that offers a nuanced perspective on population vulnerability to extreme heat. This HHVI is structured into spatial layers of sub-indices of exposure, sensitivity, and adaptive capacity. The generated HHVI was then verified by individual SA1 level linked health data via a time-stratified case-cross study design by comparing the lowest and highest quartiles of vulnerability to hospital admissions during heatwave and non-heatwave days. The study maps areas at high risk of heat stress, correlating these regions with human health and environmental factors. This enables the HHVI to reveal critical elements that influence broader spatial vulnerability patterns, playing a pivotal role in identifying and mitigating risks in the most vulnerable areas, thus fostering healthier, more resilient urban environments.

The study produces HHVI for each SA1 region within the case study area using Principal Component Analysis (PCA). Multiple indices are produced to assess differences in downscaling techniques and are calculated over the two most-recent census years (2016 and 2021) to assess changes in vulnerability over time. The index methodology utilizes multiple PCA factors scaled by explained variability to produce more complex and data-driven indices from up to 44 indicators of heat exposure, population sensitivity, and adaptive capacity. Notable initial findings indicate that (a) all downscaling methods produce highly correlated indices (>0.95); (b) both 2016 and 2021 indices correlate well with the Australian Bureau of Statistics' Index of Relative Socio-economic Disadvantage (IRSD); (c) remoteness behavior is unexpected, with inner regional SA1s experiencing lower vulnerability, on average, than major city SA1 regions; and (d) the difference noted in (c) is caused almost entirely by a consistent increase in heat exposure in major city SA1 regions. As a result of the PCA methodology, these vulnerability indices are a product of their inputs and their relevant metadata (collection year, case study geographical extent, and resolution). However, the consistency over time and across downscaling techniques suggests that the result is a robust index, capable of capturing important aspects of heat impacts and at-risk communities.

It is worth noting that accessing linked patient health data, however, presents its own set of challenges, often involving lengthy and complex processes that vary across Australian jurisdictions. For this project, data from the New South Wales Ministry of Health was utilized, primarily because it was the sole state able to provide data at the SA1 level. Following approval from data custodians and the Ethics committee, the Centre for Health Record Linkage (CHeReL) facilitates the linking of multiple databases, making the data available in the the Secured Unified Research Environment (SURE). Researchers then proceed to analyze this data within SURE, though it is important to note that the linked SA1 data remains confined to this secure environment. Only the resulting spatial indicator is extracted for dissemination and is set to be published on the websites of the project's participating organizations.

2. METHODOLOGY

In this chapter, we outline the foundational strategies and processes guiding the case study in developing targeted indicators. Key sections detail the methodological spectrum of our project, from the development to the validation of vulnerability indices and health outcome indicators. This includes procedures for data acquisition, data cleansing and alignment, indicator generation, and outcome validation. Additionally, we discuss various considerations in the design of methodology and decision-making processes throughout the case study. We also address ethical considerations in handling sensitive health data and highlight the various data sources forming the foundation of our analysis.

2.1. PROJECT OVERSIGHT

Understanding the requirements of stakeholders, including collaborators and beneficiaries, is crucial for delivering a set of high-value urban health indicators that are fit for purpose. Working with the population health and social science research communities is also essential to shape and scope the data integration methods. This collaboration ensures significant downstream impact and uptake. A Project Steering Committee, comprising the NCRIS facilities and Collaborators, was established during the project's initiation stage. This committee is responsible for the successful delivery of all work packages (WPs). Figure 2.1 shows the technical architecture design and work package of the project. Additionally, a Project Advisory Committee was formed, including leading researchers, policymakers, and practitioners in population health, environmental and socio-economic determinants of health, and urban health indicators. This committee convened post-workshop to review the project's broad progress in its initial months and provide specific feedback and guidance on the outcomes of the scoping workshop.

2.2. CASE STUDY AREA

The determination of the study area was discussed in the project initiation stage. Initially, our preference was for Queensland, but we expected to face limitations in the granularity of the geographical data available; the smallest unit we were expecting to access was

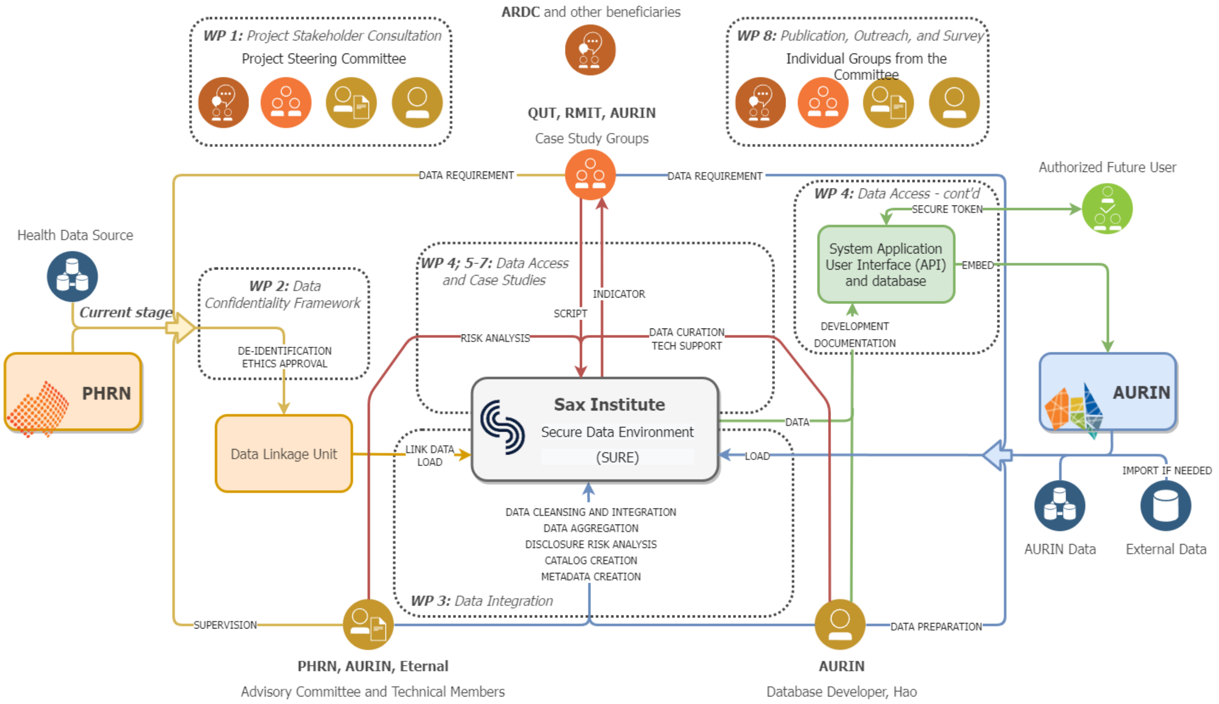


Figure 2.1: AusUrb-HI technical architecture design and work package diagram.

at the level of Statistical Area Level 2 (SA2), rather than Statistical Area Level 1 (SA1). This constraint resulted in a re-evaluation of suitable case study locations.

The initial analysis led us to exclude certain regions: Western Australia and South Australia were deemed unsuitable due to their low population densities outside of capital cities, which would pose challenges for our study in regional areas. Similarly, Victoria was set aside to avoid reinforcing a perceived "Melbourne-centric" focus in our research. Tasmania's cooler climate and the Northern Territory's lack of comprehensive data further narrowed our options.

In comparison, New South Wales emerged as the most viable choice. It was also identified that there was a potential opportunity to request and obtain SA1 level data in NSW, provided our case study was compelling. Achieving this would be a significant milestone. Furthermore, NSW presented an ideal setting for a pilot project due to its numerous populated regional areas and several large cities beyond Sydney.

Aligning with the Australian Urban Observatory (AUO)'s Urban Liveability Index (ULI) study areas, which encompasses the top five urban areas, was also a key consideration. Notably, the Newcastle-Maitland region is regarded as a contiguous urban corridor, with Newcastle's outward expansion and subsequent peripheral growth forming a connection with Maitland. This alignment with established urban study areas further justified the selection of NSW as our primary focus for the project. Accordingly, the final selected case study regions are shown in Figure 2.2 with their respective populations shown below:

- Greater Sydney - 4.8 million
- Wollongong - 0.3 million
- Newcastle - 0.17 million
- Maitland - 0.5 million
- Tweed heads - 0.65 million
- Albury - 0.05 million

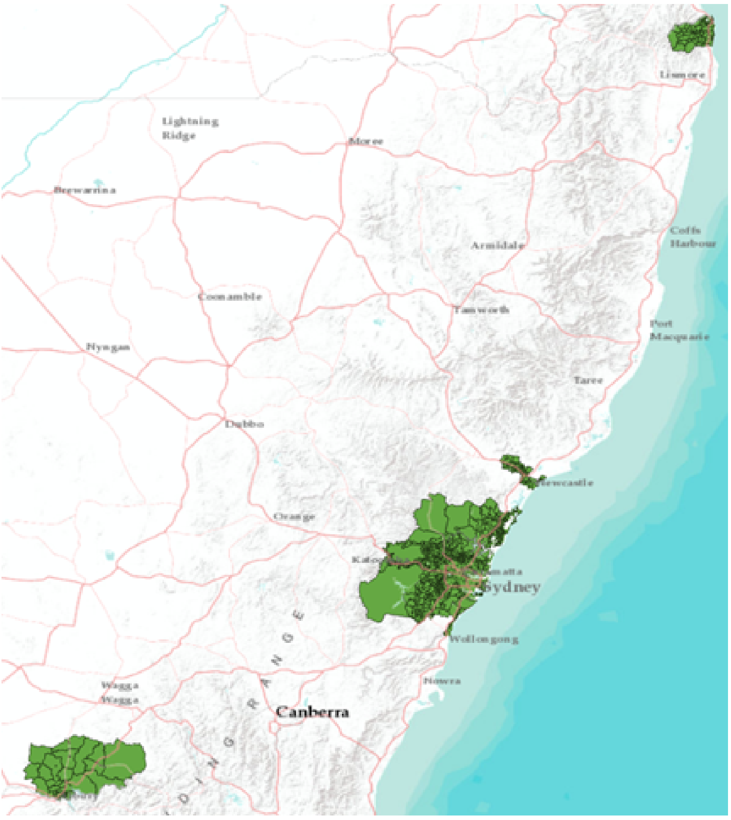


Figure 2.2: The study area defined for the HHVI and ULI case studies.

2.3. DATA ACQUISITION

Figure 2.3 illustrates the major data transaction action stages for the case studies involving datasets from three distinct sources: AURIN data collections, individual-level linked health data, and external datasets not already included in AURIN collections. The specifics of these sources will be expounded upon later in this section. The diagram delineates the following four stages:

- **Data Collection, Cataloguing, and Metadata Generation:** This preliminary stage encompasses gathering vector and raster data from AURIN databases and external sources, as well as health data from CHeReL, the NSW data linkage unit. This data is catalogued and accompanied by metadata to enhance FAIRness.
- **Data Cleansing and Pre-Processing:** The acquired data undergoes cleansing and pre-processing through the AURIN Developer Machine, utilizing databases such as PostgreSQL as well as GIS software like ArcGIS Pro. Python scripts and Jupyter Notebooks were employed to facilitate the process.
- **Data Integration, Risk Analysis, and Processing:** Subsequently, the processed data is uploaded from AURIN to SURE virtual machine (VM), alongside directly uploaded linked health data submitted by CHeReL in parallel. The upload process is monitored by the SURE curators and data custodians, as shown in Figure 2.4. The process results in an integrated dataset ready for subsequent analysis and processing.
- **Case Studies for Health Indices Generation:** In the final stage, the integrated data is used in case studies to develop health indices output.

2.3.1. General Data Confidentiality and Ethics Framework

This subsection provides a timeline with explanations as an overview of the major stages and activities involved in applying for and accessing the NSW linked health data. It illustrates the complexity and iterative nature of data confidentiality and ethics in applying for the data.

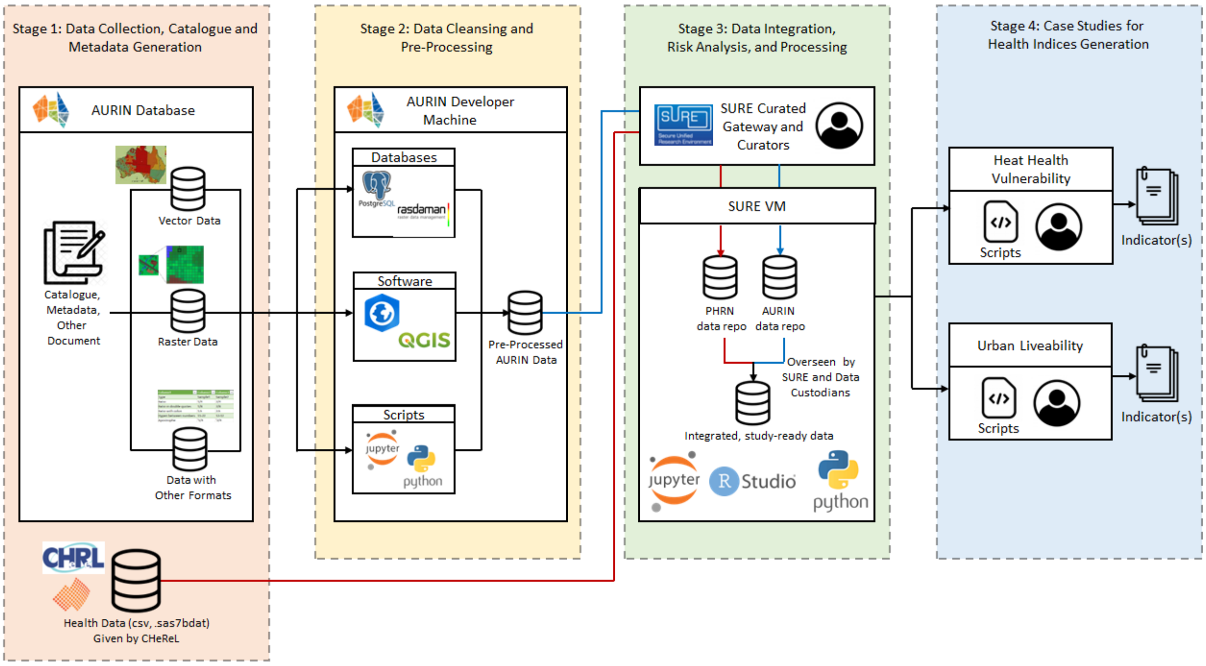


Figure 2.3: The data workflow diagram for HHVI and ULI.

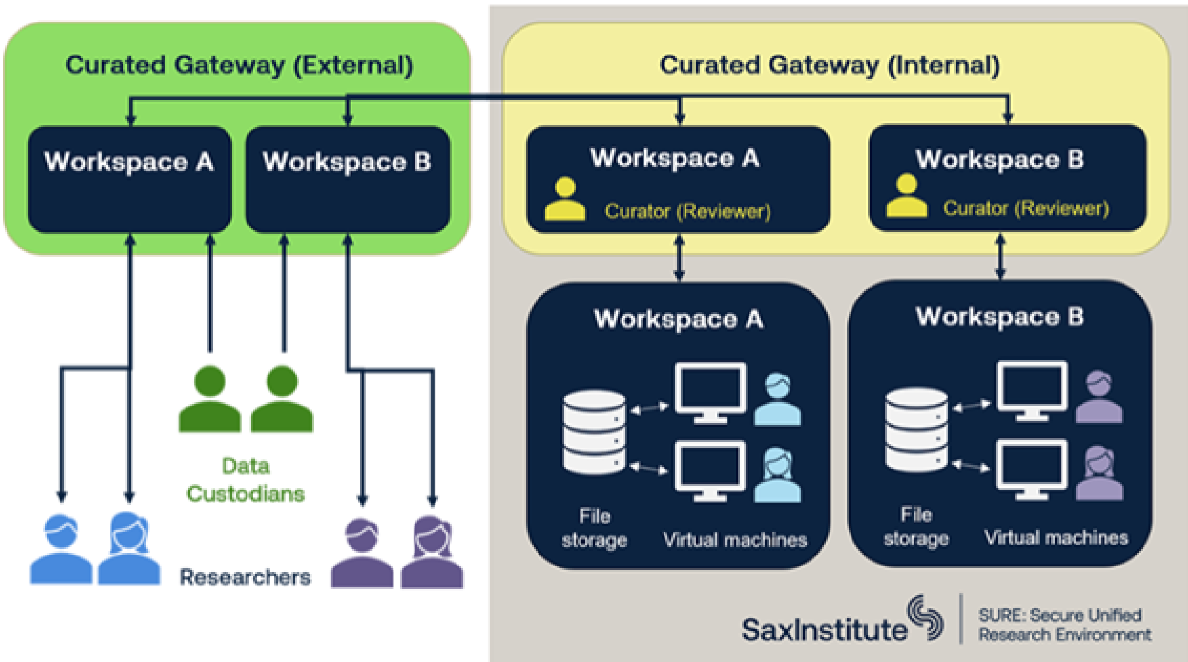


Figure 2.4: The Secure Unified Research Environment.

- **Preparation Phase:** The initial stage of the project, beginning on January 27, 2022, was dedicated to preparing for the NSW CHeReL data linkage application. Before that, there was a scoping stage that resulted in the change of study area from QLD to NSW. This phase was for laying the groundwork for the formal applications later on, involving the gathering and preparation of necessary documents and identifying the requirements set by CHeReL.

- **Application Submission and Revision:** Starting on March 30, 2022, this phase marked the submission of the first version of the combined protocol application to CHeReL along with the necessary supporting documents. CHeReL is also responsible for forwarding the documents to the relevant data custodians and acting as the coordinating entity. It was followed by a period of iterative revisions based on the feedback received from CHeReL. This process was essential for refining the project's aims and ensuring compliance with data linkage protocols.

- **Ethics and Regulatory Compliance:** Commencing on May 26, 2022, the project team focused on the ethics application as instructed by CHeReL. This involved the completion and submission of an application to the NSW Population and Health Services Research Ethics Committee (PHSREC), ensuring adherence to ethical standards and regulatory requirements.

- **Data Custodian Approval and Technical Feasibility:** Beginning on July 25, 2022, the project team submitted the final version of the application to CHeReL to obtain the technical feasibility letter as well as to present the application to Data Custodians for their review and approval. This step was fundamental in confirming the project's technical feasibility and securing the necessary data custodian endorsements.

- **Waiting and Follow-up:** The next stage starting on August 2, 2022, was characterized by waiting and obtaining various approvals and feedback. During this time, the team engaged in several follow-up activities, such as organizing meetings and revising the application based on the ongoing input and queries from CHeReL and other stakeholders. 22 Feb 2023 - Obtained approval for RBDM-deaths and COD-URF datasets. 30 March 2023 - Obtained approval for APDC and EDDC datasets. 18 May 2023 - Ethics approval obtained.

- **Data Acquisition:** Starting on May 25, 2023, this phase started waiting for the linked data extract file to be linked and sent to SURE. During the waiting period, the team focused on creating data dictionaries for data cleaning and preprocessing the AURIN data in parallel, preparing the datasets for detailed analysis.
- **Data Analysis:** Finally, beginning on July 14, 2023, the project advanced to the data analysis stage as the linked health data has been uploaded to SURE and is awaiting approval by the curators. on July 28, the data was approved and loaded into SURE. This final phase marked the transition from preparatory activities to the core analytical work.

The complete weekly-updated data acquisition status spreadsheet is accessible via AusUrbHI GitHub Page as a file named *AusUrb-Hi - Data Acquisition Status - By Case.xlsx* under the folder *documents/linked_health_data_documents*.

2.3.2. Data Catalogue Compilation

This section describes details of data acquisition steps for each source.

2.3.3. Linked Health Data Collections

This subsection describes the linked health datasets requested, the application process (including different applications submitted to different approvers, lessons learnt, reasons for delays, and limitations), as well as all the relevant documents to the applications.

Relevant documents regarding the linked health datasets and their applications as described below can be found under the same linked health data documents folder, with each sub-folder and the included documents shown below.

linked_health_data_documents
|- Linked Health Datasets (linked health datasets and detailed variables applied)
|- CHeReL (NSW data linkage unit application documents)
|- Ethics (PHSREC ethics application documents)
|- SURE (SURE application documents)
|- Approvals (any approval documents obtained from above items)
|- Amendment (documents regarding change of research investigator or datasets)

Four linked health datasets were requested in the case study, including Admitted Patient Data Coll (APPC), Emergency Department Data Coll (EDDC), Cause of Death Unit Record File (CODURF), and Registry of Births and Deaths Marriages Registration (RBDM). CHeReL provided four spreadsheets with individual variables in the preparation phase, so that individual variables can be checked as requested, and the spreadsheet were returned together with the other document to be submitted back to CHeReL. The spreadsheet names corresponding to each requested linked health dataset and the number of entries we finally obtained are (with requested variables documented in the corresponding GitHub folder):

20220711 nsw-admitted-patient-data-coll_DL.xlsx (**APDC**): 298,120,88 records for 4,540,451 patients
20220711 nsw-emergency-department-data-coll_DL.xlsx (**EDDC**): 23,330,967 records for 4,273,272 patients
20220711 cause-of-death-unit-record-file-6-29-11-2016.xlsx (**CODURF**): 221,787 records for 221,751 patients
20220711 nsw-registry-of-births-deaths-and-marriages-death-registrations-06-03-20.xlsx (**RBDM**): 264, 594 records for 264,530 patients

2.3.4. AURIN and External Raw Data Collections

This subsection describes AURIN datasets used in the case study, grouped by different categories. It also includes datasets that were collected from external sources (e.g., land surface temperature raster collections). The complete metadata and catelogue at attribute level for each datasets can be found again on the AusUrbHI GitHub Page, under the folder *documents/data_catalogue*, or Appendix A.

AURIN Socio-Demographic Datasets This category includes data obtained from the AURIN Data Provider and the ABS website (via the GeoPackages downloader). ABS datasets cover a wide range of topics relevant to the Australian economy, population, environment, and society. The NATSEM (National Centre for Social and Economic Modelling) datasets are a collection of data resources used for social and economic modeling, including information about demographics, income, employment, education, health, housing, and various other aspects of society.

- ABS census 2016 and 2021 datasets
- ABS non-census datasets, e.g., SEIFA, Income, Family & Community, etc.
- NATSEM social and economic indicators

AURIN Health-Related Datasets This category includes aggregated data (rather than individual-level linked data) that are relevant to population health, as well as health-related service location and accessibility data. The Public Health Information Development Unit (PHIDU) datasets are comprehensive collections of data focusing on a wide range of public health-related topics with information on various aspects of health, social, and economic well-being. The National Health Services Directory (NHSD) in Australia is a comprehensive directory of health services and providers, designed to help people find, access, and understand healthcare services.

- PHIDU Prevalence of Selected Chronic Diseases, Conditions, and Health Risk Factors
- ABS Health & Disability statistics
- NHSD Health Service Locations by service type

AURIN Built and Natual Environment Data This category includes built and environmental datasets provided by Geoscape Australia accessed through AURIN. Geoscape Australia is an initiative that provides comprehensive spatial data covering the entire geography of Australia. This data is detailed, in high quality, and is designed to support applications from urban planning to logistics, telecommunications, and emergency management.

- Buildings - roof height, material, cooling, building area, levels, swimming pools
- Surface Cover - % of bare earth, road, vegetation, built-up, water, building ,etc.

- Greenspace, Trees, Hydrology (in point, polyline, and polygon features), and Railway Stations

Land Surface Temperature Data The Longpaddock SILO dataset is a comprehensive collection of climate data managed by the Queensland Government in Australia. It's primarily designed for agricultural and environmental applications.

- Maximum Daily Temperature raster
- Minimum Daily Temperature raster

Urban Livability Data The RMIT AUO livability data is a collection of information that aims to measure and analyze the livability of urban areas. This data includes a range of metrics and indicators that are essential for understanding the quality of life in cities. It is created considering liveability domains including transport, social infrastructure, employment, walkability, housing, green infrastructure, and ambient environment.

2.4. DATA CLEANSING, PREPROCESSING, AND ALIGNMENT

All scripts for data cleansing and preprocessing can be found on the AusUrbHI GitHub Page, under the folder *scripts*.

2.4.1. Vector Data Harmonization through Concordance and Disaggregation

The AURIN vector data, provided in shapefile and geojson formats as outlined in Section 2.3.4, undergoes a meticulous refinement process to make it relevant for the specified study area. This refinement includes filtering based on SA1 codes to ensure the data aligns with the geographic focus of the study. An essential step in this process is the spatialization of data lacking inherent geospatial attributes, such as coordinates or geometries. This is achieved by integrating the non-spatial data with the geometry information from the ABS ASGS standard shapefiles corresponding to specific geolevels, like SA1 or SA2. This integration ensures that datasets initially devoid of spatial attributes are effectively rendered usable for spatial analysis.

The alignment of data from different years is a crucial aspect of this process, especially considering the variations in ASGS standards over time. Concordance files provided by the ABS play a vital role here. As depicted in Figure 2.5, they help illustrate how the same SA1 code might be represented by different polygons in different years, such as in 2016 and 2021, showcasing the dynamic nature of geographical boundaries over time. Moreover, some datasets are originally based on broader geolevels than SA1, such as SA2 or Population Health Area (PHA) (for PHIDU data). For these datasets, a disaggregation process is employed to bring the data down to the SA1 level. This disaggregation utilizes three primary methods: dividing the data by the number of smaller geolevel areas within the larger one, using population figures as a weighting factor, and in some cases, not dividing the data values during the disaggregation.

Figure 2.5: SA1 data alignment through concordance mapping

2.4.2. Temperature Raster Data Cube Analysis

This procedure is aimed at deriving the Excess Heat Factor (EHF), a well-established method commonly used to identify days with significant heat compared to both short- and long-term temperature averages. In selecting data sources for our study, we evaluated several options. Satellite imagery, such as that from MODIS, provides data with a spatial resolution ranging from 250 meters to 1 kilometer, offering daily, 8-day, 16-day, and monthly temporal resolutions. Landsat-8, with its 15 to 100 meters resolution and 16-day capture cycle, offers high spatial detail but less frequent updates. The Bureau of Meteorology (BOM) in Australia supplies weather station-based data with varying spatial resolutions, ranging from daily to annual temporal resolutions. Longpaddock SILO offers processed weather station data daily, effectively balancing spatial and temporal needs.

Considering our requirement for daily land surface temperature data to derive EHF values, we selected Longpaddock SILO data. Its daily temporal resolution is essential for calculating EHF, identifying excess heat days, and detecting heatwaves. Additionally, the fact that this data is processed and ready-to-use facilitates more efficient analysis. While it might lack the raw detail of satellite data, it provides a comprehensive solution for our study's needs. Importantly, unlike satellite data, Longpaddock SILO data is not affected by cloud cover, which can otherwise compromise the accuracy of the process.

The key processing steps for our analysis began with downloading maximum and minimum daily temperature NetCDF files from 2010 to 2022. We then merged these files into a comprehensive data cube, as illustrated in Figure 2.6. We refined the data to focus on our study area, applying a buffer of 0.5 and filling in any missing values. Subsequent steps involved calculating the Excess Heat Index (EHI), as well as EHF, and the 90th percentile temperatures (tx90, tn90) over the data cube, as shown in Figure 2.7. We then conducted an analysis at the SA1 level, focusing on variables such as the number of excess heat days and average temperatures, and finally converted our findings into a shapefile for geographical visualization and further analysis.

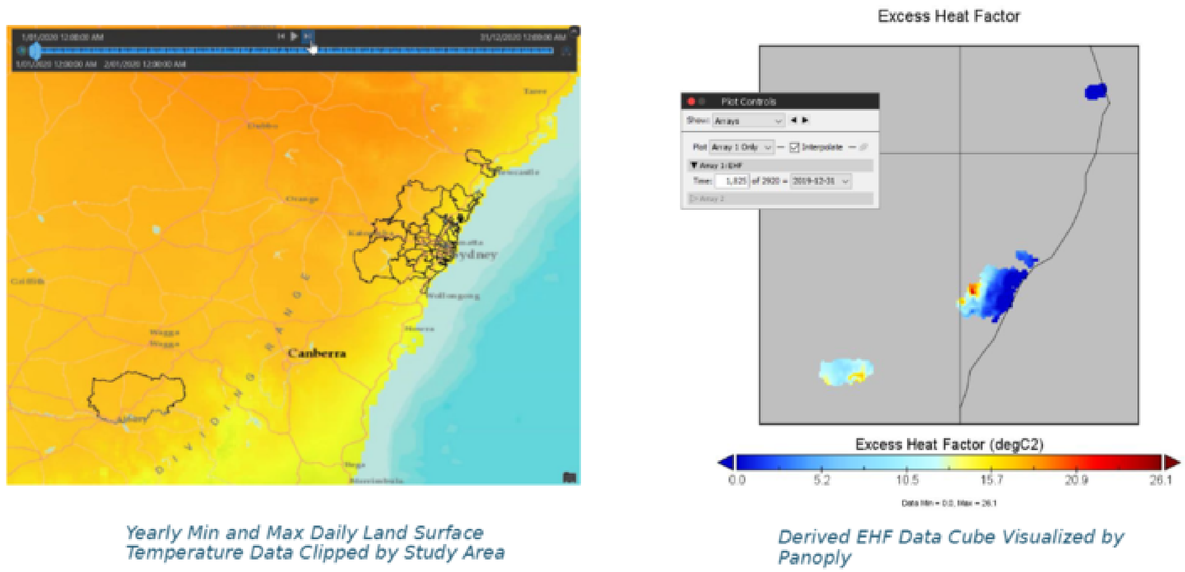


Figure 2.6: Creating minimum and maximum land surface temperature data cube for the study period.

2.4.3. Building Polygon Transformation

Geoscape's building data offers a comprehensive overview of various building attributes, catering to a broad range of applications in urban planning, real estate, and environmental studies. This dataset includes unique identifiers and timestamps, indicating the creation,

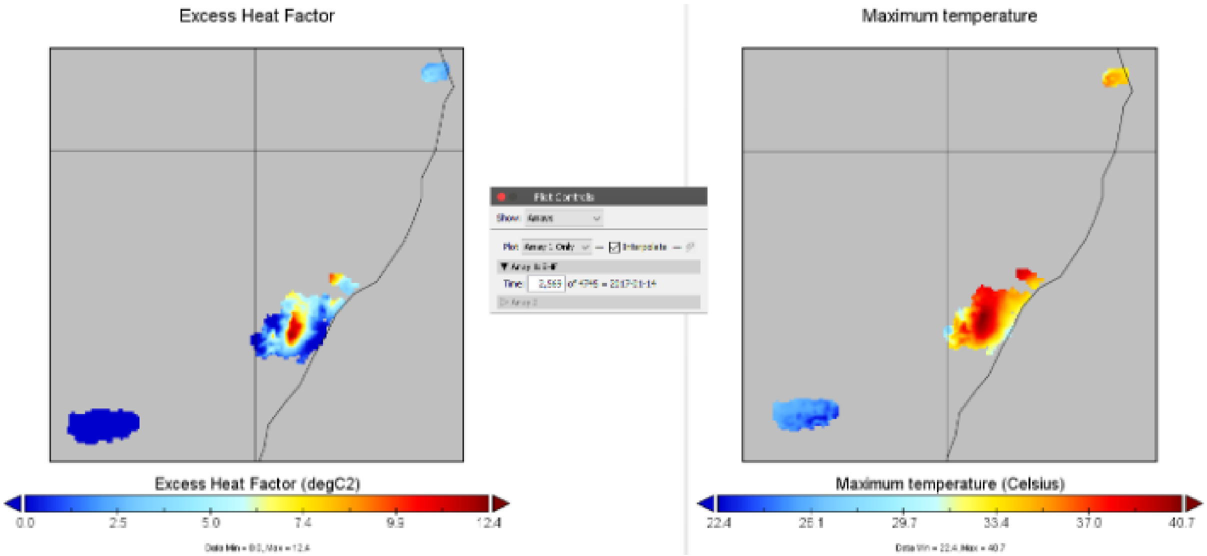


Figure 2.7: Deriving Excess Heat Factor using the land surface temperature data cube.

modification, and last capture dates of each building's data. It also provides reliable information on data source and quality. Essential geographical details such as the building's location, state, and precise coordinates are included. Physical characteristics like height, elevation, roof materials, type, color, and shape are meticulously detailed, offering insights into structure and design. The dataset encompasses information on environmental and planning features, including the presence of swimming pools or solar panels and local planning zone designations. It also covers technical data like size, shape, and capacity, including area and volume. The dataset is regularly updated, with review and verification dates for various features, ensuring data accuracy and timeliness. Furthermore, it includes details on the building's interaction with its natural environment, such as tree overhangs, enhancing its utility for environmental impact assessments.

To transform Geoscape's raw building data (which represents each building as a polygon, including its attribute values) into meaningful statistics at the SA1 level, spatial analyses were conducted using Python and ArcGIS. This process aggregated building data into SA1 level shapefiles, as illustrated in Figure 2.8. Relevant attributes were selected from an urban morphology perspective. For instance, building density, a critical statistic, was determined by calculating the number of buildings per unit area in each SA1 area. This measure is particularly useful in urban planning and development studies. Similarly, the average number of floors per building in an SA1 area was calculated by averaging the 'estimated levels' attribute of buildings, providing an indication of whether the urban landscape is dominated by high-rise or low-rise structures. This approach was applied similarly for other attributes. To calculate the percentage of buildings with access to swimming pools, the number of buildings with swimming pools in each SA1 area was counted and divided by the total number of buildings in that area. The analysis of roof materials involved categorizing buildings based on their roof material and determining the percentage of each type within an SA1 area, which could be useful for studies on urban heat islands or environmental planning.

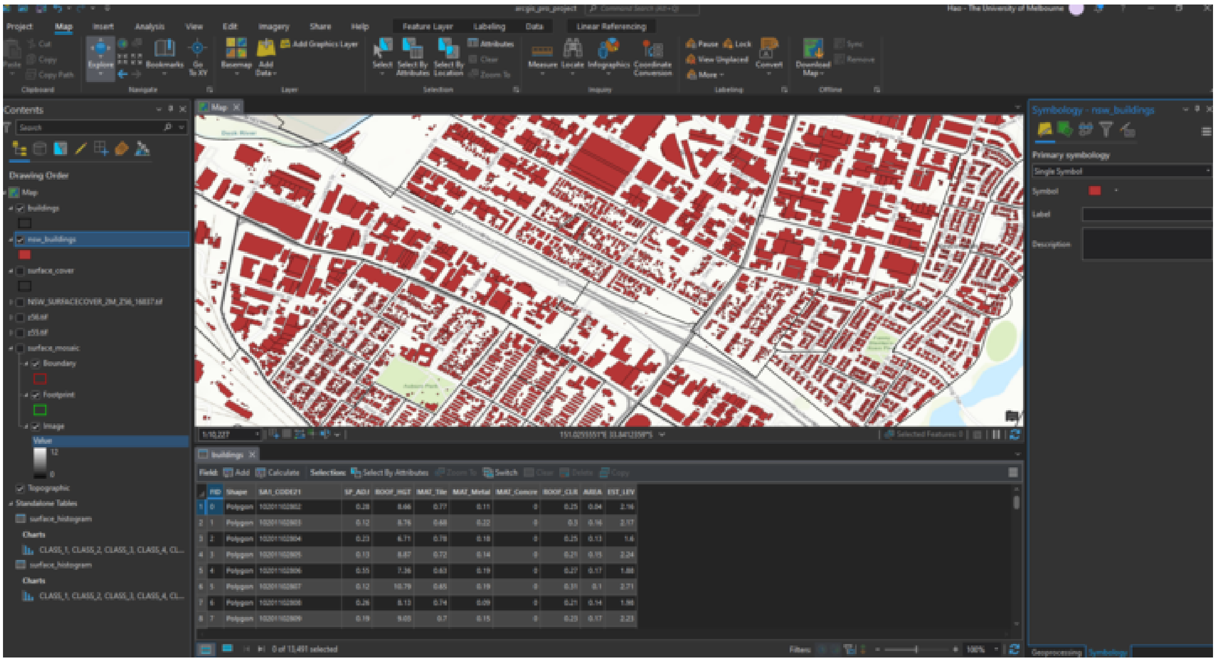


Figure 2.8: Transforming raw Geoscape building polygon data to statistical information at SA1 level.

2.4.4. Landcover Raster Statistics

The Geoscape land use dataset provides raster images that represent the utilization of various land areas, crucial for understanding the dynamics of both urban and rural environments. This dataset features classifications of land types such as vegetation, built-up areas, water bodies, road networks, etc. The processing of Geoscape's land use data revolves around converting these raster images into a shapefile. This shapefile represents different land use percentages, characterizing various land use types within each SA1 region. Using Python and ArcGIS software, the raw data was transformed to highlight patterns and trends in land use across different regions.

The process began by selecting relevant 2m surface rasters in New South Wales and two 30m surface rasters for the study area, as illustrated in Figure 2.9. The 30m rasters were preprocessed, clipped to fit the study area, and resampled to a 2m resolution. A mosaic dataset was then created in GDA94, incorporating all rasters with a specific mosaic rule to prevent duplication in overlapping regions. Subsequently, a land use shapefile was derived, involving symbology reclassification and zonal histogram analysis. The data was then converted into shapefile format and underwent manual validation and field name adjustments to ensure final accuracy.

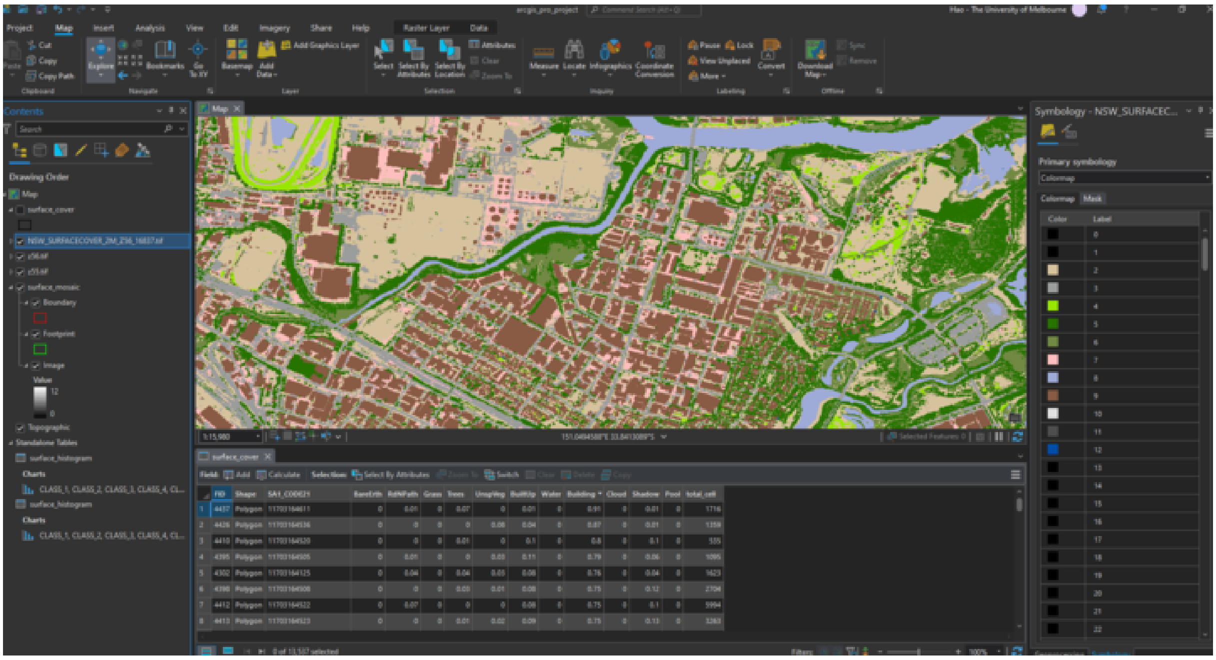


Figure 2.9: Transforming raw Geoscape building polygon data to statistical information at SA1 level.

2.4.5. Road Network Accessibility Analysis

Finally, to analyze datasets on points of interest, such as the accessibility of NHSD service locations and train stations from individual SA1s, we used network distance accessibility spatial analysis. This approach was chosen over a Euclidean distance-based method (straight-line distance on a map) for practical reasons. For instance, if a point of interest (like a station) is on the opposite side of a river from an SA1 and there's no bridge, it may not be as accessible compared to situations where no natural barriers exist. Network analysis considers road network spatial data, measuring accessibility via total travel distance through the road network.

The methodology involved handling an OpenStreetMap line shapefile and an NHSD service location point shapefile. Initially, we standardized the coordinate reference system across all files and refined the shapefiles to the study area. In ArcGIS Pro, we created a feature dataset within a geodatabase, added the line data, and then constructed a network dataset, as shown in Figure 2.10. A service area layer was then established with service locations loaded, and a service area analysis was conducted. The resulting polygons were saved as separate shapefiles.

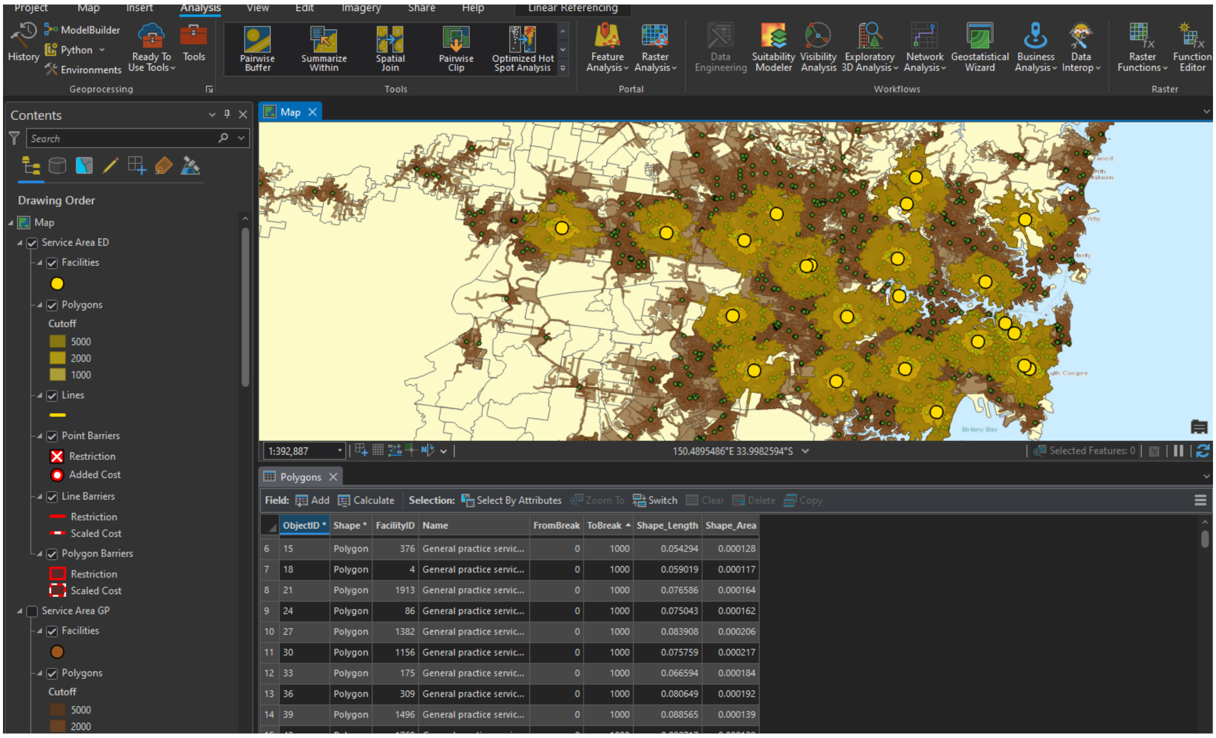


Figure 2.10: Service accessibility measurement through network distance analysis.

2.5. VULNERABILITY INDICES

Over the last few decades, it has become increasingly popular to characterise geographical inequality among populations by defining and comparing population vulnerability metrics. As government agencies seldom have the resources with which to carry out an effective assessment individually, the most popular measures of vulnerability are those which condense a wide array of data into a single value per region. These measures are referred to as vulnerability indices, or composite indicators of vulnerability. Examples of prominent vulnerability indices include the early social vulnerability index for environmental hazards [1], the social vulnerability index first used in the United States for emergency management and response [2], and the socio-economic indices for areas used in Australia [3].

In an era where the frequency and severity of heatwaves continues to increase, it is of great importance to understand how climate, landscapes, and populations contribute to population health risk. Due to the complex relationships between the environment and humans, no single investigation into vulnerability indices will provide a holistic and comprehensive answer to general environmental health vulnerability [4]. Despite this, local government agencies need information to support vulnerability assessments to effectively develop policy, apply interventions, provide emergency disaster management support, and aid their respective community's ability to adapt to changes in the environment [5]. Hence, relevant data items have been utilised in the creation of AusUrb-HI vulnerability indices, henceforth referred to as the HHVI.

While population vulnerability can be difficult to define due to differences in population characteristics and geographical composition, a framework provided by the Intergovernmental Panel on Climate Change (IPCC) in 2007 [6] suggests the consideration of three primary aspects of vulnerability: exposure, sensitivity, and adaptive capacity. Exposure is a direct measure of household, community, or population exposure to a certain event (extreme heat, cold, air pollution, etc.). Sensitivity is defined as the susceptibility of a household, community, or population to the exposure. Finally, adaptive capacity captures the capabilities of a household, community, or population to cope with or recover from the impact of the exposure. This framework is widely used in the creation of environmental vulnerability indices internationally as a way to determine relevant characteristics and their role in or impact on population vulnerability [7, 8, 9, 10, 11]. Hence, this framework is utilised in the development of AusUrb-HI HHVI methodology as part of the AusUrb-HI heat vulnerability case study, focused on the five largest cities in the state of New South Wales: Sydney, Wollongong, Newcastle - Maitland, Tweed Heads and Albury (see Section 2.2). The SA1 vulnerability indices produced in this case study combine a range of existing Australian data assets as well as a small number of data assets and indicators derived by AURIN to better represent and communicate aspects of population vulnerability (see Section 2.4). Note that all data items acquired and/or created for the vulnerability index aspect of this case study adhere to (or have been modified to adhere to) the Findable, Accessible, Interoperable, Reusable (FAIR) principles [12].

2.5.1. Index Data Considerations

The role of a vulnerability index is to draw value from many underlying variables to indicate regions with multi-faceted risk profiles at a glance, without required scrutiny of said underlying data. As such, a priori selection and exclusion of variables plays a key role in the value of a vulnerability study. For AusUrb-HI, this variable selection process reflected on key international literature, existing Australian data infrastructure and consultation with Australian data custodians to explore new or emerging data assets relevant for use in the creation of SA1 heat health vulnerability indices. Existing climate vulnerability indices use socio-economic indicators common to many general-purpose vulnerability indices [2, 3] and also incorporate climate factors with available data relevant to the study area [13]. More recently, indicators have also been considered against the above aspects of exposure, sensitivity, and adaptive capacity [6] as a way to compare similar contributions to heat health vulnerability [7, 10].

Data Used in Literature

There are a range of studies in this field, with data selections varying due to data availability, geographical resolution, and scope. For brevity, we refer to commonly included indicators from a select number of studies, representative of the field [5, 7, 8, 9, 10, 11, 13, 14, 15, 16]. Indicators are all given as proportions, unless otherwise specified. Socio-economic indicators include elderly population (aged 65 and over), living alone, high school education, under poverty level or low income, mean or median family income, mean or median household income, infant population (ages 4 and under), single parents, provision of unpaid childcare, persons with disabilities or requiring significant assistance for daily activities, non-English speakers, minority status or indigenous, living in mobile homes, renting or with a mortgage, median home loan repayments or rent payments, existing health conditions, existing risk conditions (e.g., smoking, obesity, alcohol consumption), internet access, gender, poor housing conditions, overcrowding, relocated recently (within the last year or five years) and access to a vehicle. Climate indicators include land surface temperature, mean maximum or minimum temperature, days exceeding various temperature thresholds, population density, building age, high-rise dwellings, outdoor workers, impervious surfaces, canopy cover, vegetation cover, distance to water, access to medical services (e.g., psychiatry, hospitals, general practices, emergency departments), proximity to or volume of water bodies, building materials and road density.

Theme	Indicator
Urban Exposure	Average Roof Height (m) Building Density Cool Roofing (%) Road and Path Density
Natural Cooling	Grass (%) Canopy Cover (%) Other Vegetation (%) Minimum Distance to Water Body (m) Minimum Distance to River m)
Heat Exposure	Temperature Deviation Percentile (mean) Temperature Deviation Percentile (max) Excess Heat Factor (mean) Excess Heat Factor (max) Excess Heat Days (no.)

Table 2.1: AusUrb-HI exposure indicators and their associated themes.

Data Used in AusUrb-HI

The items used in the AusUrb-HI vulnerability index closely follow those used in the literature, with a number of differences resulting from data availability, either altogether or at the appropriate spatial or temporal resolutions. These items provided by sub-index in Tables 2.1, 2.2 and 2.3 and are notably divided further into themes, which are commonly used to merge highly-related data items and minimise the potential for bias in the final vulnerability index [7, 13].

Future Data Considerations

This study considered a priori a range of data items which were excluded during the data exploration phase of the study. While these were excluded from this iteration of AusUrb-HI, there may be ways of representing these items in future which better highlight their relationship with population vulnerability. Items that may offer value in future include method of travel to work (walking, cycling, etc.), indigenous breakdowns (Torres Strait Islander, Aboriginal or both), proportion born in Australia, persons in non-standard housing, dependent students, financial indicators from the National Centre for Social and Economic Modelling (NATSEM), access to emergency departments or

Theme	Indicator
Household Composition	Infants, 0-4 (%)
	Elderly, 65+ (%)
	Indigenous (%)
	English Not Primary Language (%)
	Persons Requiring Assistance (%)
	Persons Providing Unpaid Assistance (%)
	Living Alone (%)
	Single Parent (%)
Other Sensitivity Indicators	Population Density
	Renters (%)
	Machinery Operators (%)
	Labourers (%)
	Houses Requiring Extra Rooms (%)
Socio-economic Status	Median Family Weekly Income ($)
	Low Income (%)
	Unemployed (%)
	Not in Labour Force (%)
Existing Conditions	Heart, Stroke, Vascular Disease (ASR)
	Diabetes (ASR)
	Psychological Distress (ASR per 100)
	High Blood Pressure (ASR per 100)
Risk Factors	Overweight (ASR per 100)
	Obese (ASR per 100)
	Smokers (ASR per 100)
	High Alcohol Consumption (ASR per 100)

Table 2.2: AusUrb-HI sensitivity indicators and their associated themes. Note the acronym age-standardised rate (ASR).

Theme	Indicator
Other Adaptive	Completed High School (%)
	Household Access to One or More Vehicles (%)
	Internet Access (%)
	Swimming Pools (%)
Liveability	Liveability Index

Table 2.3: AusUrb-HI adaptive capacity indicators and their associated themes.

general practices, and access to transport. Additionally, further exploration of exposure indicators and measures of adaptive capacity such as cool places (grocery stores, shopping centres) or building characteristics (e.g., building age) will improve those aspects of the vulnerability index which currently contain fewer data items.

2.5.2. Handling of Missing Data

Not all data are available at the desired SA1 spatial resolution and for the desired annual time points (2016, 2021) for the vulnerability index. However, some of those data are available at other spatial resolutions or times. To create the most informative vulnerability index possible given the available data, the following considerations were addressed.

- Temporal imputation: Data is available, but not at either or both of 2016 and 2021.
- Spatial disaggregation: Data is available, but not at the SA1 spatial resolution.
- Spatial imputation: Data is available at the SA1 resolution, but some regions have missing data entries.

Temporal Imputation

Where data is not available at either 2016 or 2021 but it available for other years, the indicator values have been averaged and assigned to both the 2016 and 2021 indicators. Averaging has occurred for chronic disease and risk factor data from the Public Health Information Development Unit (PHIDU), as well as for crowded dwelling data from the Australian Bureau of Statistics (ABS). Single year assignments have occurred for building, surface cover, and hydrology data provided by Geoscape, as well as the Liveability index provided by the Royal Melbourne Institute of Technology (RMIT). Finally, where data is available for one year but not the other, the same data has been used for both years. This is the case for internet access, which was collected in the 2016 Australian Census, but not collected in 2021.

Spatial Disaggregation

In some cases, data are not available at SA1 but are available at lower resolutions. To make use of the variation in these data items, the following simple imputation methods were considered.

- Population-weighted disaggregation (PD): Region counts distributed according to underlying SA1 region population distributions.
- Even-ratio disaggregation (ED): Region counts evenly divided by the number of underlying SA1 regions.
- Disaggregation with no division (ND): Region counts untouched and applied constantly to the underlying SA1 regions,

Due to clear flaws with any of the above methods, the vulnerability index methodology is replicated for each of the above methods, as well as a fourth index (OM) which simply omits data items which are not available at the SA1 geographical resolution. These imputation methods were considered for chronic disease and risk factor data from PHIDU, as well as for crowded dwelling data from ABS.

Spatial Imputation

Where an indicator appropriately exists or has been created through one of the above methods, some entries of the data may still be missing for a variety of reasons related to the indicator or data custodian, including the data collection extent, survey participation, suppression due to privacy concerns, or uncertainty in estimations. The literature relating to data that is missing completely at random (MCAR) is extensive [17]. However, as only a small number of data entries were missing for the vast majority of AusUrb-HI indicators (approximately 1%), we use mean imputation (MI) in this study.

2.5.3. Vulnerability Index Methodology

AusUrb-HI vulnerability index methodology largely follows a typical application of PCA for the purpose of creating an index [13, 17], with the objective being to explain the variance of observed data through a few linear combinations of the original data. For brevity, the full PCA process is omitted in this report. Many of the references in this section utilise PCA, but we recommend pursuing the composite indicator handbook for more information on the specifics of PCA with regard to vulnerability indices [17]. The AusUrb-HI vulnerability index creation process (aside from the specifics of PCA) is shown in Figure 2.11 and is described as follows.

- Exploratory analysis is carried out, with relevant adjustments made.

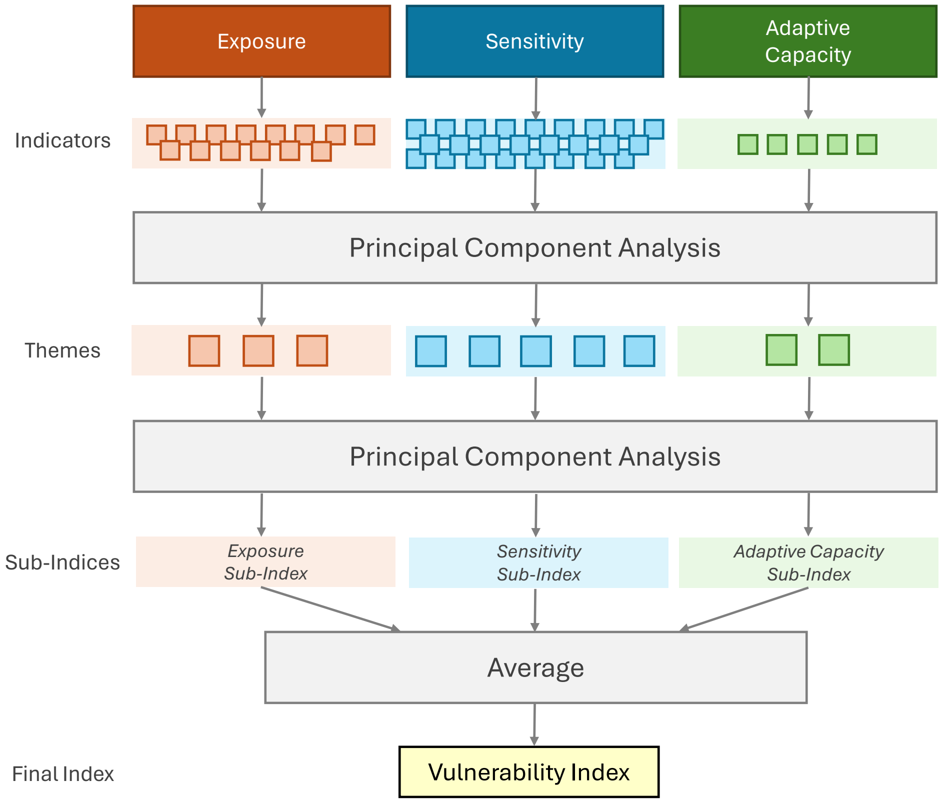


Figure 2.11: Flowchart displaying AusUrb-HI vulnerability index methodology.

- Data items are transformed, where necessary, such that each indicator has a positive relationship with vulnerability. Data items inverted include cool roofing percentage, grass percentage, other vegetation percentage, median income, high school completion percentage, vehicle access percentage, internet access percentage, swimming pool access percentage and the liveability index.
- Data items are also transformed such that each variable is approximately normally distributed. Log transformations were carried out on average roof height, unemployment rate, population density, percentage of dependent students, percentage of people providing unpaid care, percentage of people requiring assistance for core activities, elderly percentage, infant percentage, percentage of people living alone, percentage of single parents, indigenous population percentage, internet access percentage, high school completion percentage, vehicle access percentage and swimming pool access percentage.
- Data items are arranged into themes exhibiting either consistently high or

low correlation to minimise potential bias introduced by highly correlated indicators.

- PCA is applied to data items within each theme, which produces a set of principal components (new dimensions), each with factor loadings (proportions of existing dimensions) and eigenvalues (a numbers which represents the variation explained its associated principal component). Once these outputs are acquired, index calculations proceed as follows.
  - Principal components are accepted until at least 85% of the variance in the original data is explained by the new dimensions.
  - An additional principal component is selected if it explains at least 10% of the remaining variation in the original data and if it has an eigenvalue greater than 1.
  - If PCA is being applied to two variables, only the first principal component is selected.
  - Factor loadings are then scaled by their explained variance percentage, relative to the full variation explained by the remaining principal components.
  - For each principal component, factor loadings are added together to assess whether the increases in each new dimension will result in increases to vulnerability. If necessary, should the sum of the factor loadings be negative, we invert individual dimensions (i.e., all factor loadings for the principal component in questions) such that each dimensions contributes as much as possible towards the overall vulnerability measure.
  - The resulting factor loadings are then summed across principal components to give a single weight value for each underlying variable.
  - Finally, the relevant weights are multiplied through each variable and the resulting data are added together, resulting in a single index, with increasing values inherently leading to increases heat health vulnerability.
- This process is then applied to all themes within each sub-index (exposure, sensitivity, adaptive capacity), in the same manner as above.

- To avoid PCA declaring relative weights for each of the key exposure, sensitivity and adaptive capacity aspects of vulnerability, the sub-indices are not input into the above PCA process. Instead, no weighting is applied and the final vulnerability index is instead calculated by averaging the three sub-index scores.
- In addition to raw index scores, the final vulnerability index is additionally converted to percentiles across the full case study area.

2.5.4. Demonstration of Index Calculation Process: 2016 Urban Exposure Theme Index (ED)

We begin by creating a range of visualisations to determine data patterns and visible correlation structures to assess theme groupings and to assess whether inversions or transformations are necessary. Figure 2.12 shows the result of this assessment: histograms showing a range of roughly normally distributed variables.

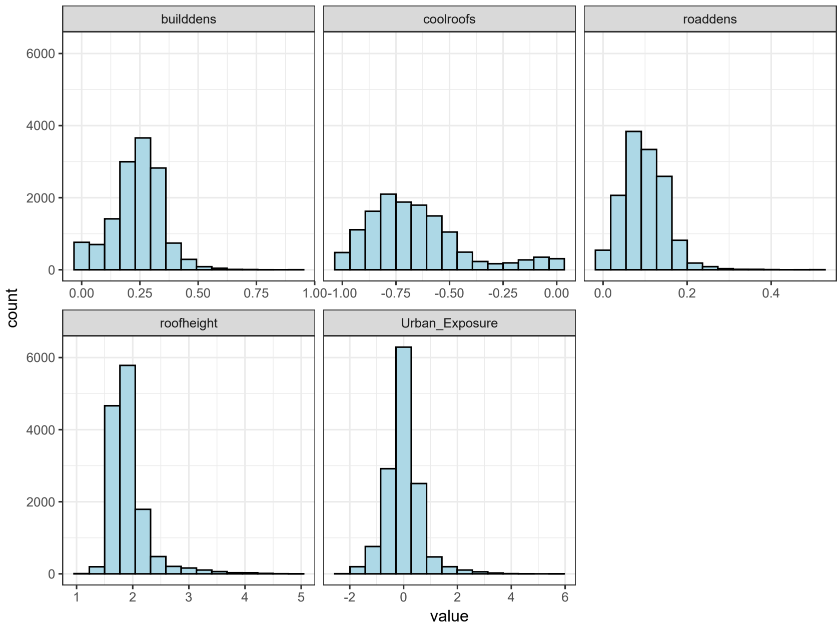


Figure 2.12: Histograms of post-processed data and resulting theme index.

Note the resulting urban exposure theme index histogram is also presented in Figure 2.12 to demonstrate the result of the methodology.

The presence of strong correlation among subsets of PCA inputs can significantly impact the value of the resulting principal components. As such, care must be taken so that there is either weak or strong correlation between all input variables used in the

PCA calculation. The correlation plot shown in Figure 2.13 shows limited correlation among the input variables in the creation of the urban exposure theme index. Note, again, that the resulting urban exposure theme index is also shown in Figure 2.13 to demonstrate the result of the methodology.

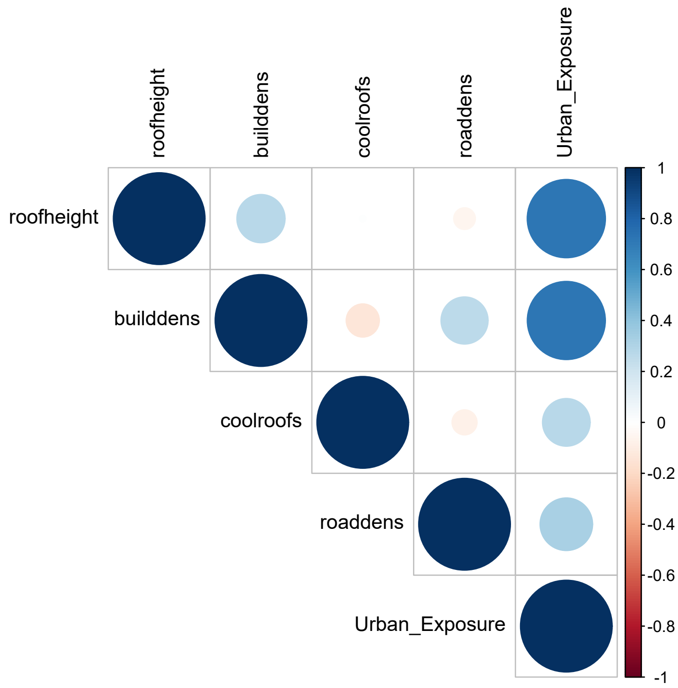


Figure 2.13: Correlation plot of theme index and underlying variables.

The post-processed data is then used as input into PCA, with the calculation process shown in Figure 2.14. With four input variables (roof height, road density, cool roofing percentage and building density), there are four initial principal components. The first three are required to surpass the 85% threshold and while the fourth principal component meets the threshold of 10% variance explained, it does not meet the eigenvalue requirement (omitted for brevity, see final row to observe removal of PC4).

The factor loadings of the first 3 principal components are then scaled by their remaining explained variation. For example, factor loadings of PC1 in Figure 2.14 are scaled by $0.3494/0.8505 = 0.4108$. The factor loadings are then summed together which, in the case of PC1, reveals that the entire dimension should be inverted. All scaled and potentially inverted factor loadings are summed over each input variable to give the bottom-right "IND" variable weights which will then be multiplied through each underlying variable and summed to produce the urban exposure theme index.

Visuals for the remaining themes, sub-indices, and overall vulnerability index can

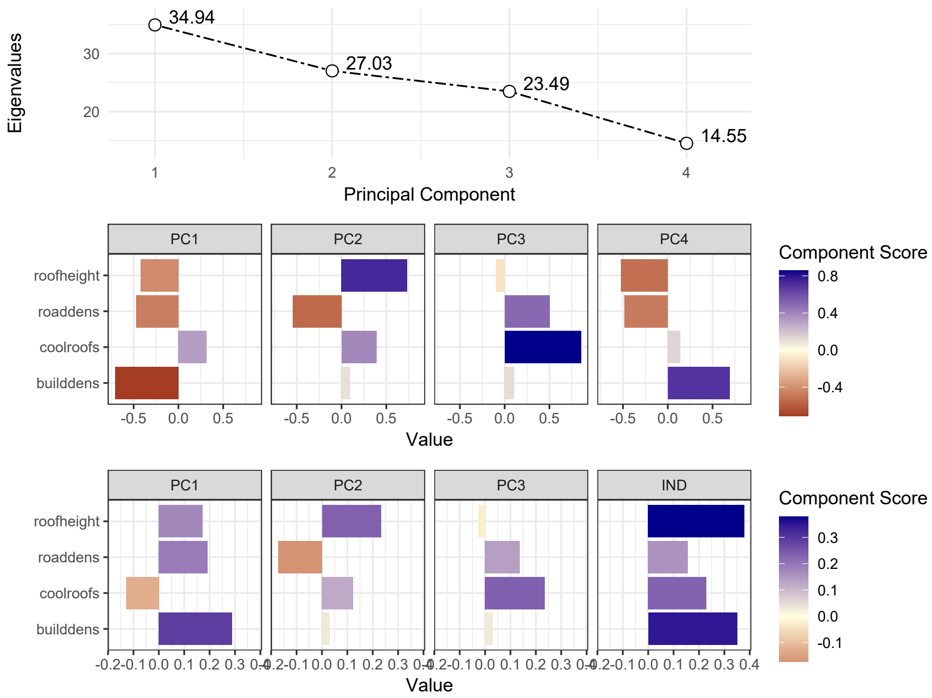


Figure 2.14: Line plot (top) shows the variation explained by each principal components, and bar plots (middle) show principal component factor loadings. Bottom set of bar plots demonstrate the index calculation process. First, factor loadings are scaled by relative variation explained and flipped if their sum is negative. Second, irrelevant principal components are removed. Finally, the weights are summed across principal components to give the resulting index.

be found in Appendix C. For further discussion on the resulting 2016 and 2021 indices across imputations methods, see Section 3.2.

2.6. LINKED HEALTH OUTCOME ANALYSIS

We received the hospital, emergency department, and death records linked by CHeReL in September 2023. To ensure compliance with the request we submitted to CHeReL, we performed simple descriptive statistics to confirm the conformity of the data provided. Due to the sensitive nature of these records, they are stored in SURE. All data transfers, both inbound and outbound, undergo careful curation before processing.

The size of the datasets was beyond the capacity of the standard virtual machine available in SURE. Consequently, we requested an upgrade to a Power VM within SURE, which incurred additional costs. This upgrade markedly enhanced the data processing speed.

Utilizing the land surface temperature data in SURE, we analyzed the number of individual hospital admissions during heatwave periods compared to non-heatwave days. Through these analyses, we explored various epidemiological study designs, ultimately selecting the time-stratified case-crossover as the most fitting for our study (refer to Figure 2.15). Employing this design, in alignment with existing literature, we observed an increase in the number of individuals (distinct from the number of admissions, as some cases involved transfers leading to multiple records) with emergency hospital admissions. The results will be discussed in the next chapter.

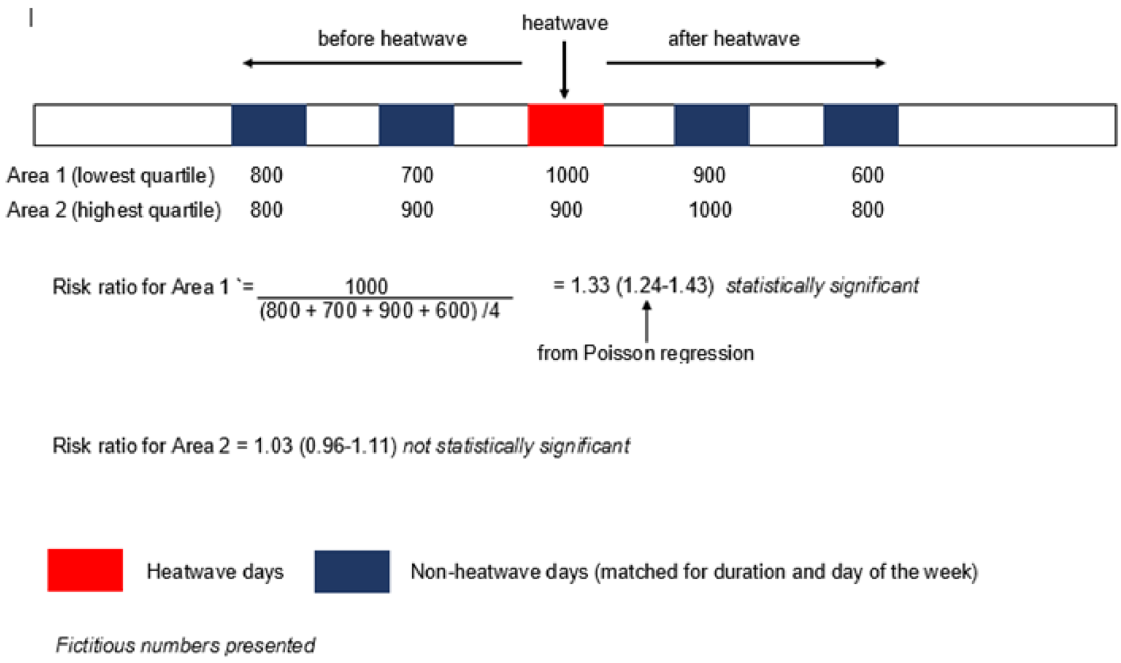


Figure 2.15: Time-stratified case-crossover study design and calculation of relative risk of adverse events during a heatwave using linked health data from NSW Health

Subsequently, we validated several heat vulnerability indicators developed using environmental data and Australian Bureau of Statistics data. Currently, our findings indicate that most of these indicators do not correlate with adverse health outcomes during heatwaves. Ongoing investigations are being conducted to resolve the inconsistencies between these indicators and the health outcomes. As all results require approval from a delegate of the data custodian and further clearance from NSW Health, we are not in a position to disclose specific numbers at this juncture.

3. DESCRIPTION OF CASE STUDY OUTPUTS

This chapter outlines the output data catalogue, followed by visualizations and analyses of vulnerability indicators, subindicators, and their correlation with health outcomes. Subsequently, the benefits of these outcomes, including the identification of stakeholders and beneficiaries are presented.

3.1. DATA CATALOGUE

This subsection provides a detailed overview of the data repository, metadata, and the means to access the data. The datasets, available through the *AURIN Data Provider (ADP)*, encompass a comprehensive collection of metrics and indicators pertinent to heat health vulnerability. The data spans various dimensions, including demographic factors, environmental conditions, and urban morphology, structured to facilitate in-depth analyses of exposure, sensitivity, and adaptive capacity in relation to extreme heat. Due to the length of the tables to be included, they have been moved to Appendix B.

Table B.1 details the range of individual data files from the case study output collection, indicating the specific content and focus of each file. These files are available in multiple formats, including spatialized versions like Geojson and Shapefile, to cater to diverse analytical needs and are downloadable from the ADP.

Table B.2 offers a granular view of the dataset by providing metadata for individual attributes within these data files. This includes the name and a succinct description of each attribute, allowing users to understand the specific data points and their relevance to the broader study.

Finally, Table B.3 encompasses part of the metadata for the entire data collection, with entries that are only for technical purposes omitted. It serves as a comprehensive data record that is stored within the AURIN database. This metadata includes key details such as authorship, dataset title, abstract, data lineage, research associations, and licensing information.

3.2. VULNERABILITY INDEX DISCUSSION

The aim of a vulnerability index is to condense a wide range of data into a single comparative representation of vulnerability across space and/or time. However, as with any other variable or indicator, it is important to be critical of the result, to compare against other measures used for similar purposes and to summarise trends and outliers.

Four versions of the AusUrb-HI HHVI have been constructed, each considering different options for how to include low resolution data alongside other SA1 variables in the index calculation process. These approaches aim to transform Local Government Area (LGA) level existing condition data (e.g., high blood pressure) and risk factor data (e.g., smoker) to the SA1 geographical resolution so that the variation across these areas can be considered in the index. Correlation between 2016 and 2021 vulnerability indices using each method (equal divide (ED), no divide (ND), population divide (PD) and omission (OM)), are compared in Figure 3.1.

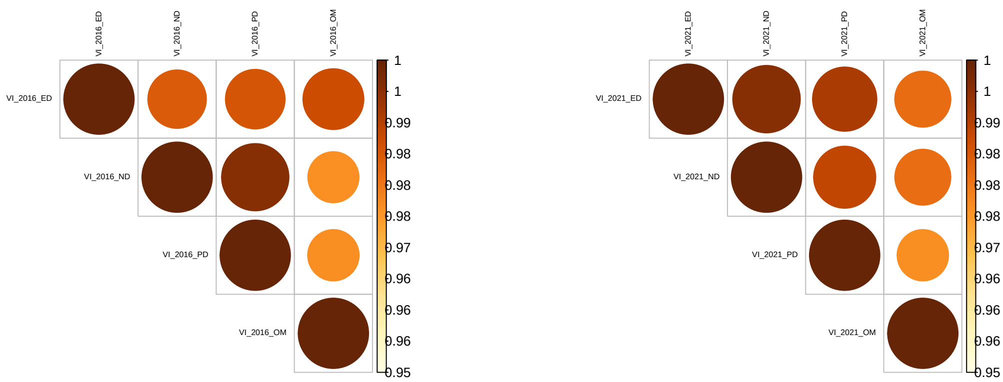


Figure 3.1: Comparison of correlation between the 2016 (left) and 2021 (right) vulnerability indices by imputation methodology.

Notably, the legend indicates that all vulnerability index correlations across both years exceed 0.95, indicating significant overlap between imputation approach. While imputation has been considered with the aim of including more indicators of population vulnerability, it is clear from Figure 3.1 that the resulting HHVI are equivalent. As such, to avoid potential bias introduced to the index via imputation method, we will proceed to discuss only the omission (OM) version of the HHVI and recommend this version be used or referred to in future. It is important also to note that only simplistic methods were considered for this index and that more complex methods spatial imputation methods will yield different but likely similar results.

Rather than reproducing all calculation visuals (from Section 2.5.4 and Appendix C) to discuss trends between the 2016 and 2021 omission approach, we instead present a comparison of sub-index and overall vulnerability index correlations in Appendix D. Figure D.1 shows only slight differences in correlation between sub-indices and the vulnerability index from 2016 to 2021.

Exposure sub-index correlation, shown in Figure D.2, is not so consistent between years, with correlation between exposure and roof height, cool roofing percentage, grass percentage and temperature deviation all decreasing in 2021. There are other interesting observations to be drawn from Figure D.2, such as the negative correlation between heatwave metrics (maximum and average excess heat factor) and both of cool roofing and canopy cover.

Sensitivity sub-index correlation, shown in Figure D.3, appears largely consistent across years, with exception of the unpaid assistance variable, which becomes marginally negatively correlated with sensitivity in 2021. Similarly, adaptive capacity correlations, shown in Figure D.4, are consistent between years. One notable observation in Figure D.4 is that vehicle access is negative correlated with liveability index, indicating that households are less likely to own a vehicle if they live in a location with small walking distance to services. Additionally, when comparing Figure D.4 to overall vulnerability index correlation in Figure D.5, we note that vehicle access is negatively correlated with the adaptive capacity sub-index but is moderately positively correlated with HHVI.

Overall correlation is consistent across years in Figure D.5 though, notably, the heat exposure theme and underlying variables exhibit low correlation with the overall index, compared to the stronger correlation observed with the exposure sub-index in Figure D.2. This may be a preferred result for applications in intervention planning or emergency management, where baseline population characteristics are highly correlated (i.e., the correlation between sensitivity and HHVI) and exposure and adaptive capacity contribute to better understand where these aspects of vulnerability are compounded in certain communities.

To better understand spatial patterns in the index and sub-indices across both 2016 and 2021, we present a series of choropleth plots (Figures 3.2, 3.3, 3.4, 3.5, 3.6 and 3.7) covering each of the population centres and displaying the omission HHVI as well as each sub-index (exposure, sensitivity and adaptive capacity).

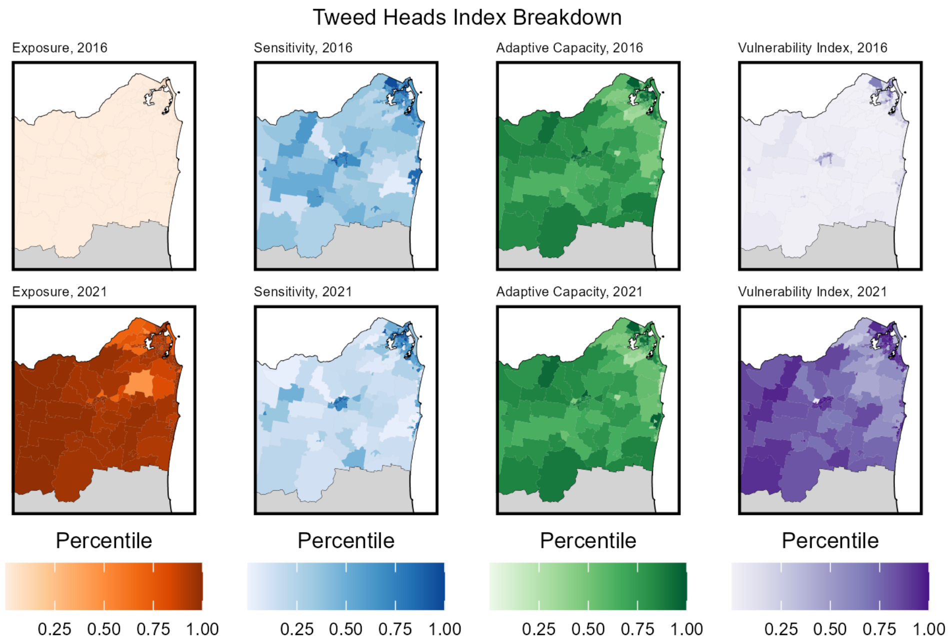


Figure 3.2: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on Tweed Heads.

The index breakdown for Tweed Heads, presented in Figure 3.2, shows that exposure has increased significantly across the region between 2016 and 2021. This increases the corresponding vulnerability index across the whole of Tweed Heads. However, the most vulnerable regions are located mostly in urban regions in both 2016 and 2021. The Wollongong visuals in Figure 3.3 are largely consistent across sub-indices and years. One observed difference is in Belambi (middle right of Wollongong case study area), which has a low exposure sub-index but high values in each other visual.

Greater Sydney, captured in Figures 3.4 and 3.5 show changes in northern Sydney's exposure, and highlight the increased adaptive capacity risk in north Sydney. Adaptive capacity trends in inner Sydney also differ from exposure and sensitivity, a likely result of liveability differences in inner city regions.

Newcastle and Maitland visuals in Figure 3.6 are very consistent between 2016 and 2021 and also exhibit similar patterns across sub-indices. The most vulnerable populations are observed in inner Maitland (top left), with inner Newcastle (bottom right) possessing a consistent average vulnerability index. Finally, Albury visuals in Figure 3.7 show the opposite trend in exposure compared to Tweed Heads, with a significant decrease throughout the Albury case study area in 2021 compared to 2016. High vulnerability is otherwise consistent in the Albury population centre (south), with a small

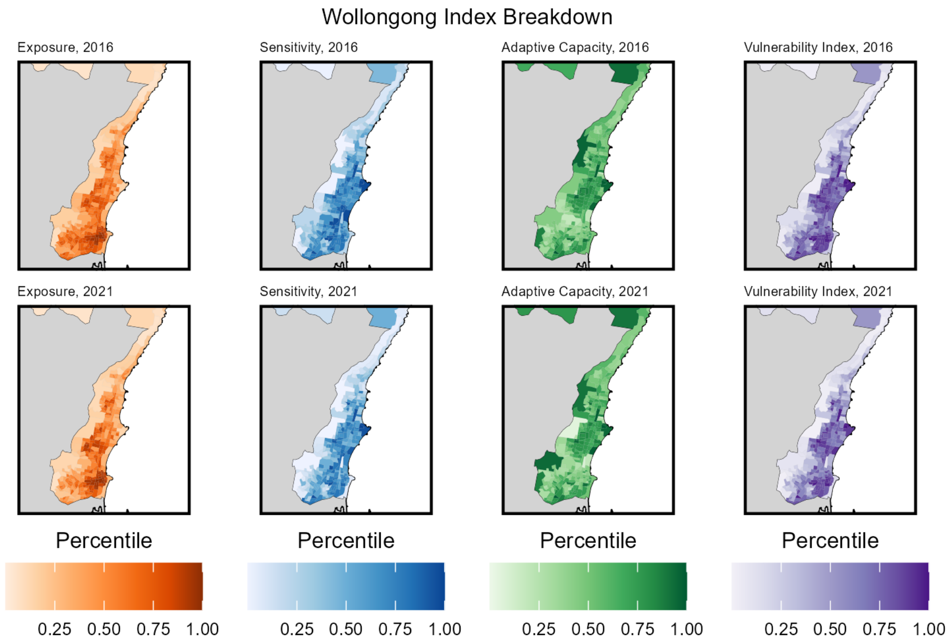


Figure 3.3: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on Wollongong.

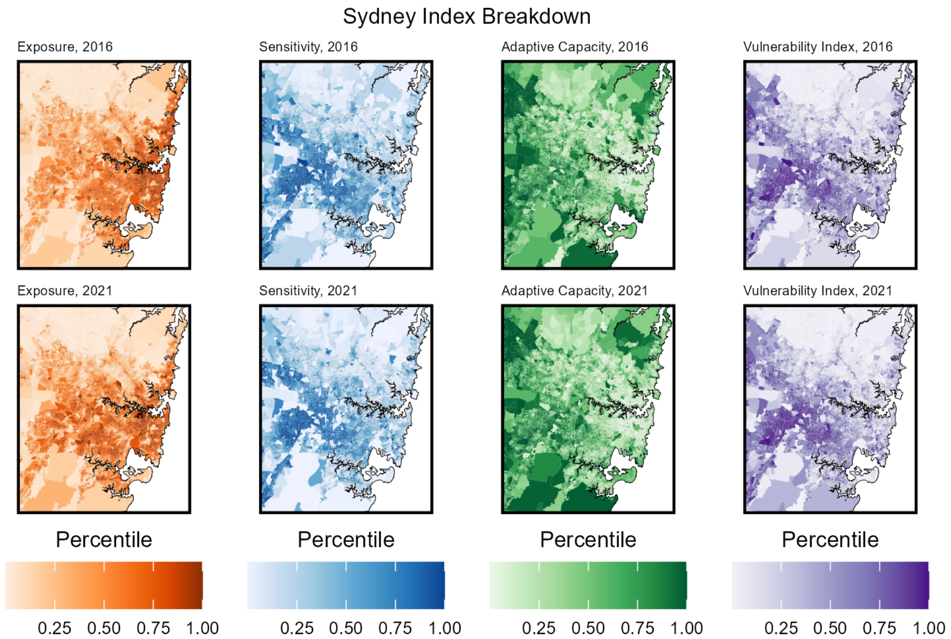


Figure 3.4: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on inner Sydney.

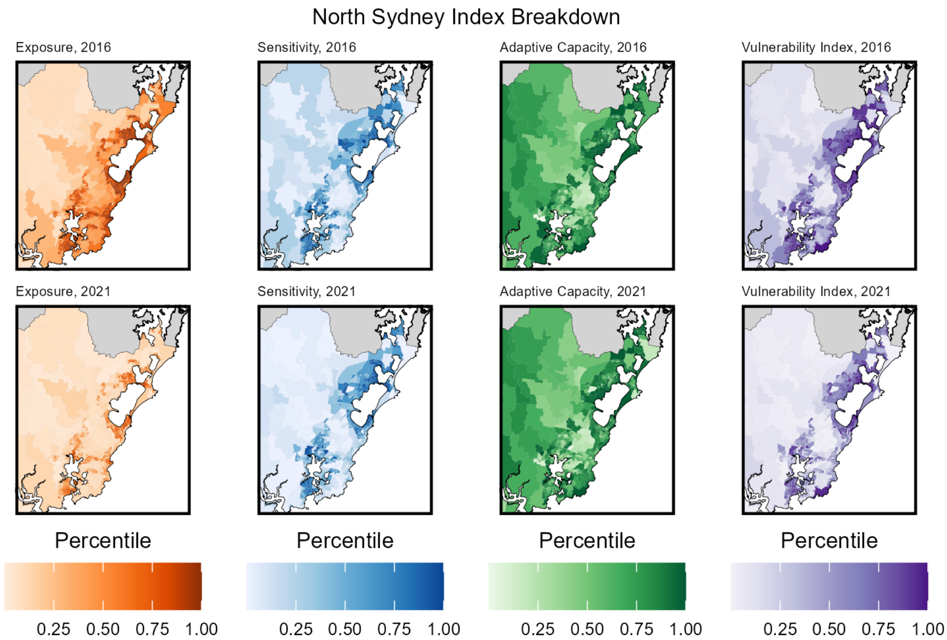


Figure 3.5: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on North Sydney.

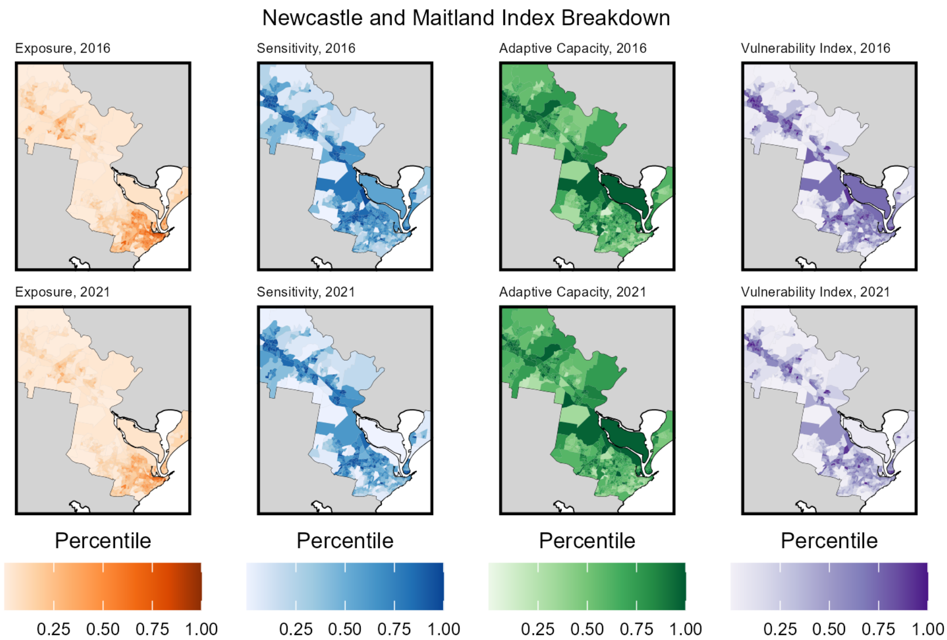


Figure 3.6: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on Newcastle and Maitland.

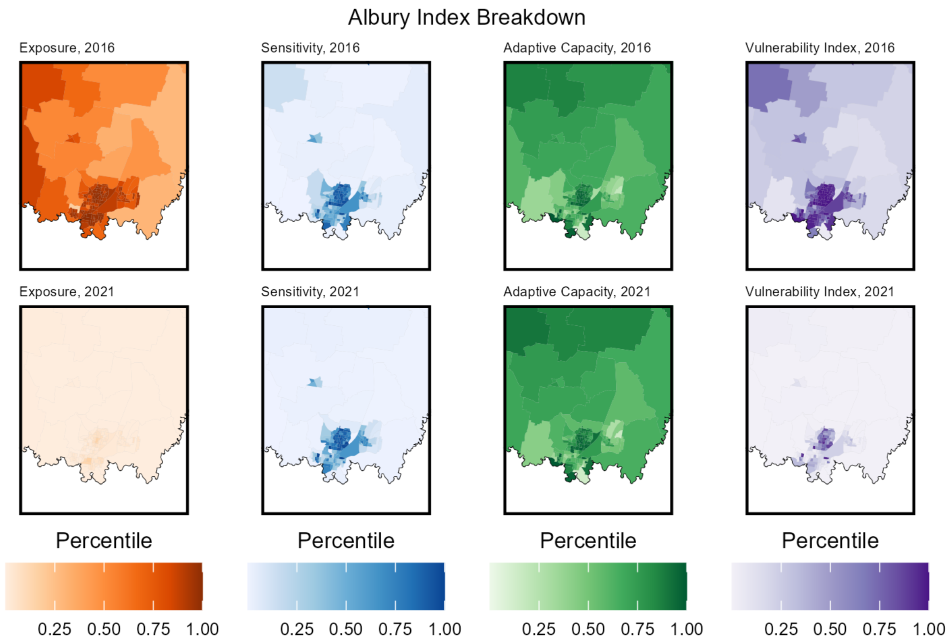


Figure 3.7: Choropleth plots of each sub-index and the overall vulnerability index in both 2016 and 2021, focused on Albury.

pocket of increased vulnerability to the north in Jindera.

As noted in Section 2.5.3, monitoring population vulnerability in Australia is not a novel practice. The socio-economic indices for areas, or SEIFA, are indices commonly used to monitor inequality across Australia [3]. In Figure 3.8, we produce correlations between all versions of the 2021 HHVI and the 2021 index of relative socio-economic disadvantage (IRSD) from the SEIFA collection. This index is originally inversely proportional to disadvantage, so we have inverted the index such that high values are equivalent to high disadvantage to match the direction of the HHVI.

The correlations between all HHVI and IRSD in Figure 3.8 range between $0.75 - 0.8$. The significant overlap in indices is due to common data contained within the AusUrb-HI sensitivity sub-index, with differences largely due to exposure and adaptive capacity variables as well as the weighting determined by the AusUrb-HI methodology. These differences represent scenarios in which socio-economic inequality aligns with one or both of exposure and adaptive capacity inequality across regions, making the AusUrb-HI HHVI of potential value to extreme heat intervention activities and emergency management planning with a focus on heat events.

The differences between the 2021 omission HHVI and the 2021 IRSD are visualised over the case study area in Figure 3.9. Additionally, these differences are summarised by case study area using boxplots in Figure 3.10.

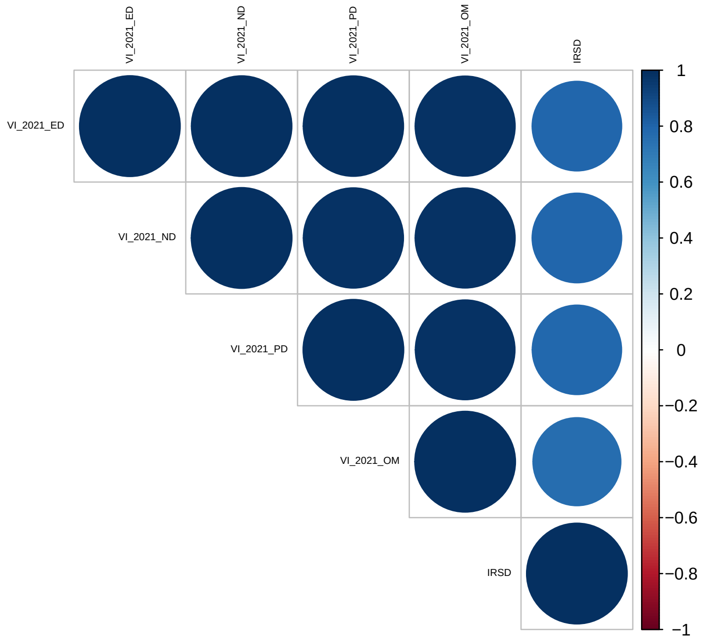


Figure 3.8: Comparison of correlation between the 2021 HHVI and the 2021 index of relative socio-economic disadvantage from the ABS SEIFA collection.

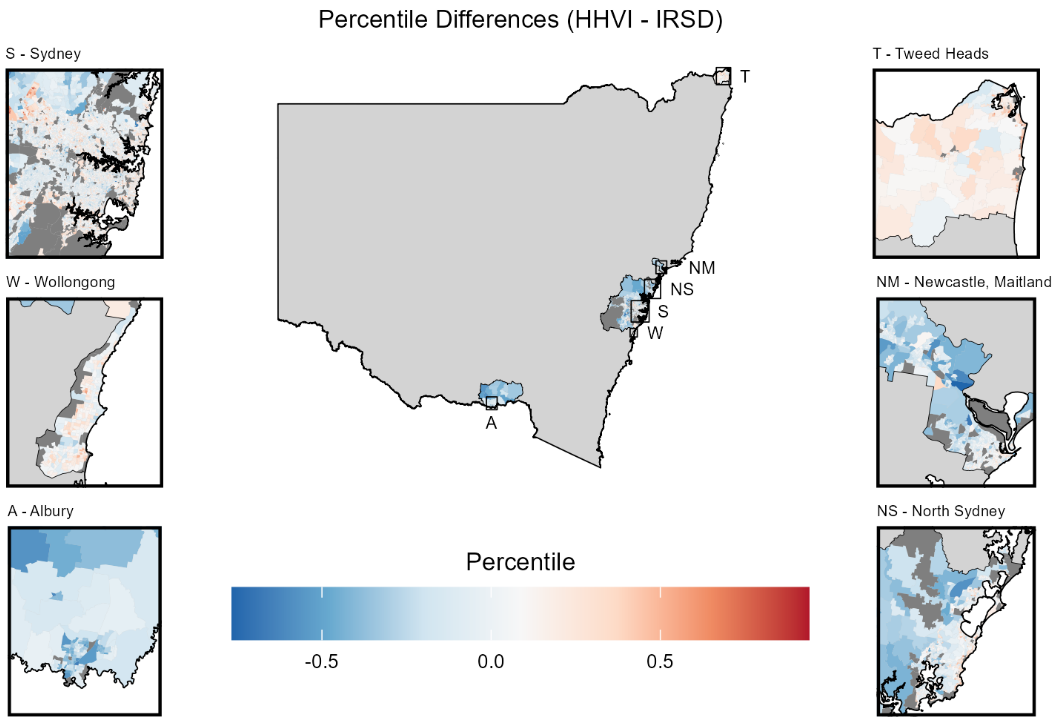


Figure 3.9: Choropleth map representing difference between 2021 HHVI percentiles and 2021 IRSD percentiles. Positive values indicate that a region's AusUrb-HI HHVI value is higher than the IRSD value for that region (i.e., we classify the region as more vulnerable), and vice versa.

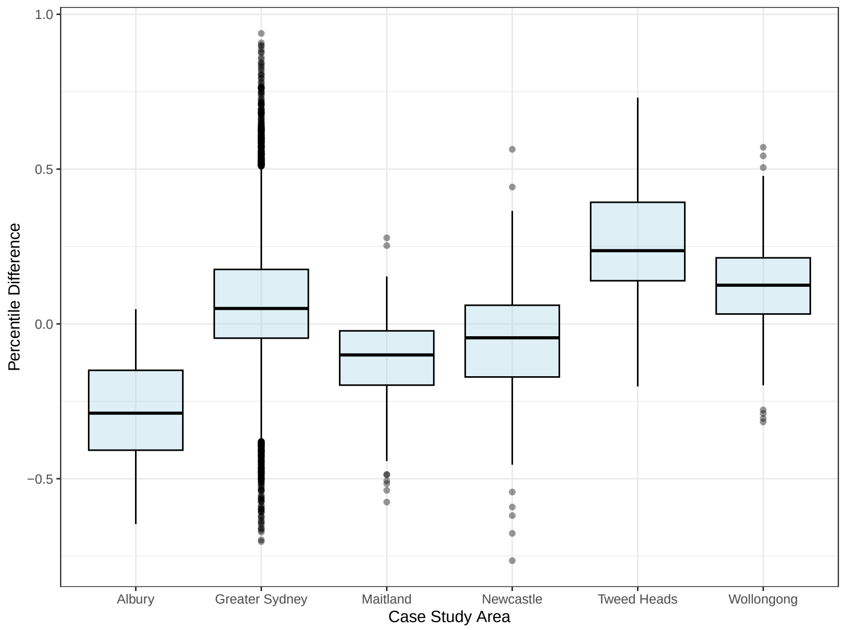


Figure 3.10: Boxplots representing difference between 2021 HHVI percentiles and 2021 IRSD percentiles. Positive values indicate that a region's AusUrb-HI HHVI value is higher than the IRSD value for that region (i.e., we classify the region as more vulnerable), and vice versa.

Figures 3.9 and 3.10 show that most HHVI values are consistent with IRSD values, especially in urban Maitland and Newcastle, and Tweed Heads. The HHVI places Tweed Heads and Wollongong SA1 regions as typically more vulnerable than the IRSD suggests. The HHVI also suggests that Albury is typically less vulnerable when compared to IRSD. Greater Sydney exhibits a large variation of differences, with HHVI indicating a slightly higher vulnerability overall. The boxplot for Greater Sydney in Figure 3.10 contains a number of outliers which could be the focus of future assessment (to determine in greater detail why the disagreement between HHVI and IRSD is so large in those outlier areas).

In addition to comparisons of raw IRSD scores, users of SEIFA indices often consider inequality deciles, derived from SEIFA index percentiles across Australia alongside the accessibility/remoteness index of Australia (ARIA+), an indication of accessibility of regions to service centres. We visually summarise how the AusUrb-HI 2021 omission sub-indices and HHVI vary with both ARIA+ and IRSD in Figures 3.11, 3.12, 3.13 and 3.14. Given our case study areas are focused around population centres, the vast majority of case study area data are categorised in ARIA+ as either "major city" or "inner regional". As such, we have excluded the small number of "outer regional" entries for visualisations.

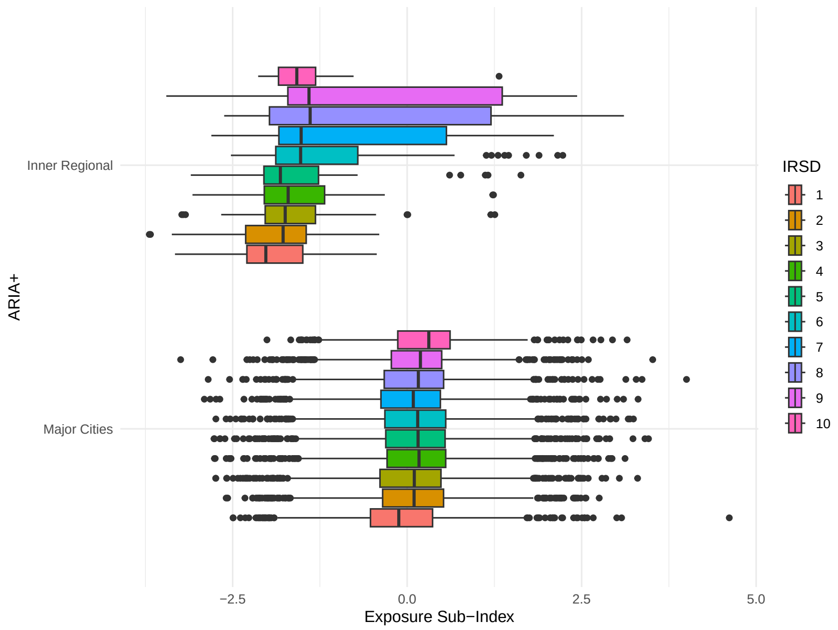


Figure 3.11: Boxplots of exposure sub-index grouped by ARIA+ and IRSD.

Figure 3.11 shows that major city exposure indices have a large variation in major city regions, but are typically higher (i.e., indicate higher vulnerability) than the surrounding inner regional areas. Equally notable is that the exposure sub-index appears largely unaffected by IRSD, with high disadvantage regions displaying a similar spread to low disadvantage regions in the same remoteness category.

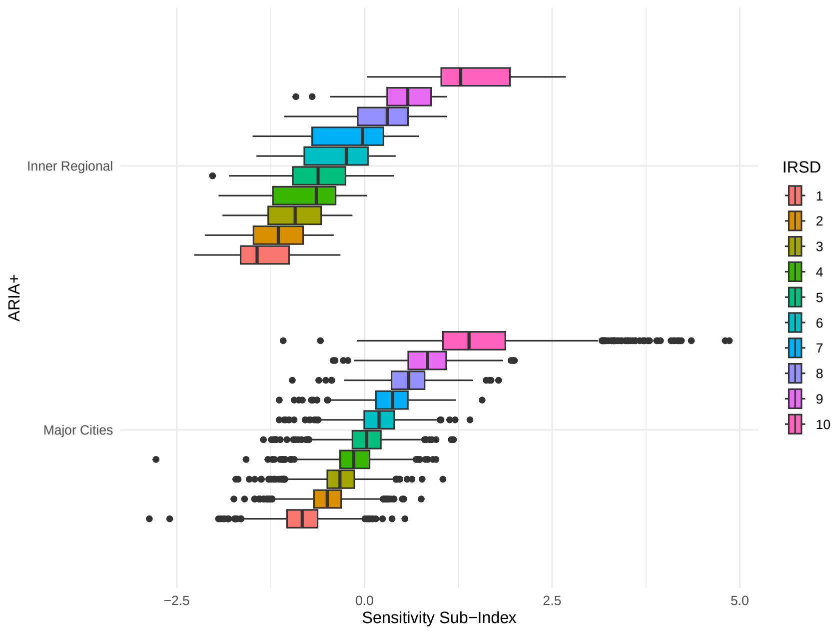


Figure 3.12: Boxplots of sensitivity sub-index grouped by ARIA+ and IRSD.

Figure 3.12 follows a similar trend with respect to remoteness; a surprising result that inner regional areas appear less vulnerable than major city regions. The IRSD trend is expected for the sensitivity sub-index given a number of underlying sensitivity variables capture population socio-economics. It is clear from this figure that the AusUrb-HI sensitivity sub-index agrees with IRSD with regards to move disadvantaged populations.

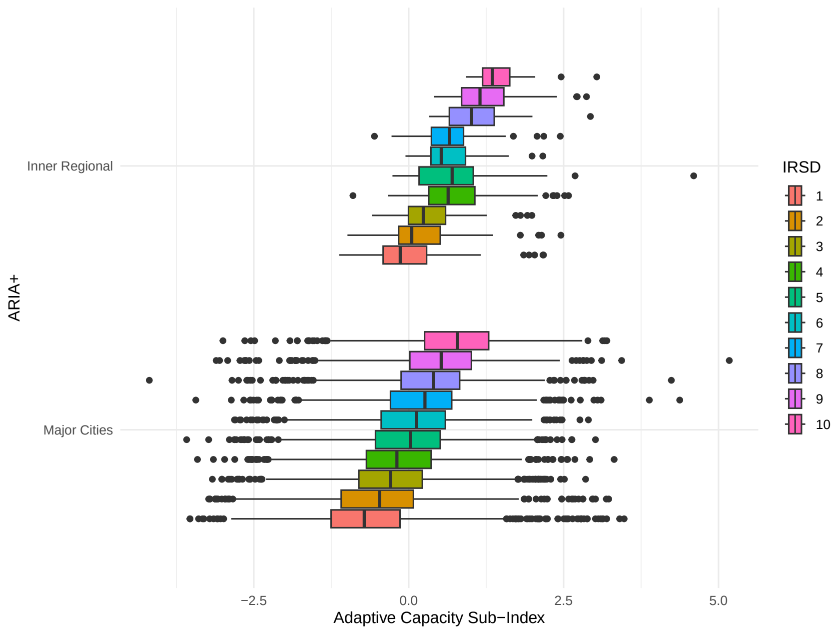


Figure 3.13: Boxplots of adaptive capacity sub-index grouped by ARIA+ and IRSD.

Figure 3.13 shows that adaptive capacity is generally lower in major city regions compared to the surround inner regional areas as expected. The trend of large variation in major city sub-index values continues, with a general increase in vulnerability as IRSD increases for both major cities and inner regional areas.

Boxplots of HHVI in Figure 3.14 show that inner regional areas typically exhibit substantially less vulnerability than major city areas though, again, there is significant variation within the major city areas. Since this index is a combination of many factors, the trends are complex to explain. The high variance may be due to a wide variety of socio-demographic inhabitants in population centres, or due to differences in climate and urban development differences across the NSW case study areas. The difference between inner regional and major city areas may be impacts by those same considerations, or even complex correlations that relate to measured variables but are not captured as part of the index calculation. Despite no deliberate attempt to calculate an urban heat island effect, combined information from exposure variables may be contributing to this difference across remoteness levels for exposure, though an increased study area with

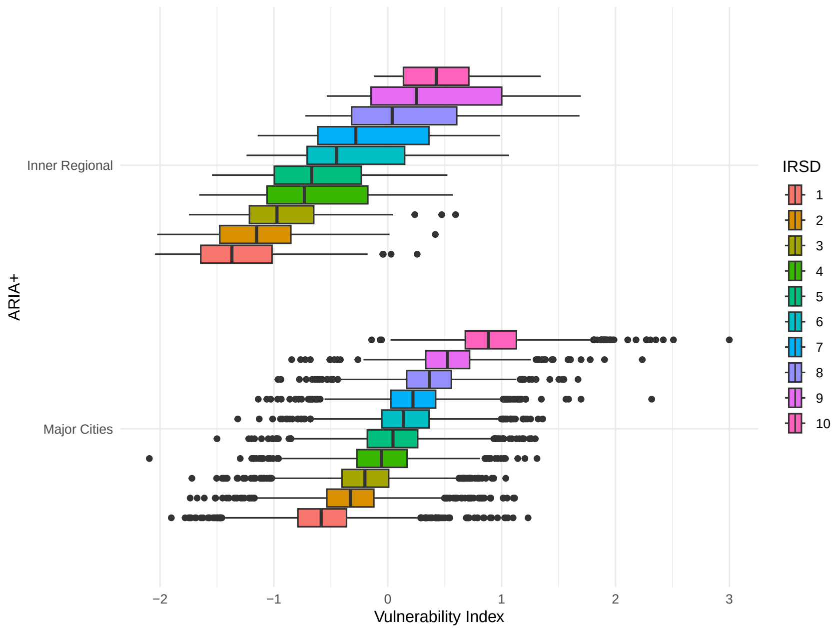


Figure 3.14: Boxplots of HHVI grouped by ARIA+ and IRSD.

more remote regions in future would help to further suggest this urban heat island effect.

Overall, the vulnerability indices produced through AusUrb-HI show the evolution of heat vulnerability over time and can be explored further in the future and used by relevant decision makers to better understand and act on inequalities in population vulnerability. The breakdown of indices at this fine scale enables longer-term planning with respective to sensitivity variables and development of intervention strategies with respect to cool roofing and provision of different ways for communities to overcome intense heat events (e.g., provision of increased cool places, green places, increasing vegetation, etc.

3.3. LINKED HEALTH DATA VALIDATION

Health indicators included emergency admissions during selected periods for all causes, cardiovascular disease (ICD codes I00-I99), myocardial infarction (ICD-10-AM codes: I21.x), acute coronary syndrome (ICD codes I20.0, I 21.x), coronary heart disease (ICD codes I20-I25), cerebrovascular disease (ICD codes (I60-I69), transient ischaemic attack (ICD code G45), heat-related conditions (ICD codes T67, X30), heart failure (ICD code I50). As the number of adverse events during the narrow window of the heatwave period is relatively small in epidemiological terms, we have to pool data across several years to obtain more precision in our measures of health outcomes. Most studies have pooled counts from 5-10 years of data. If future studies are planned, we will request longer-term data (we only asked for 5-years of data from CHeReL).

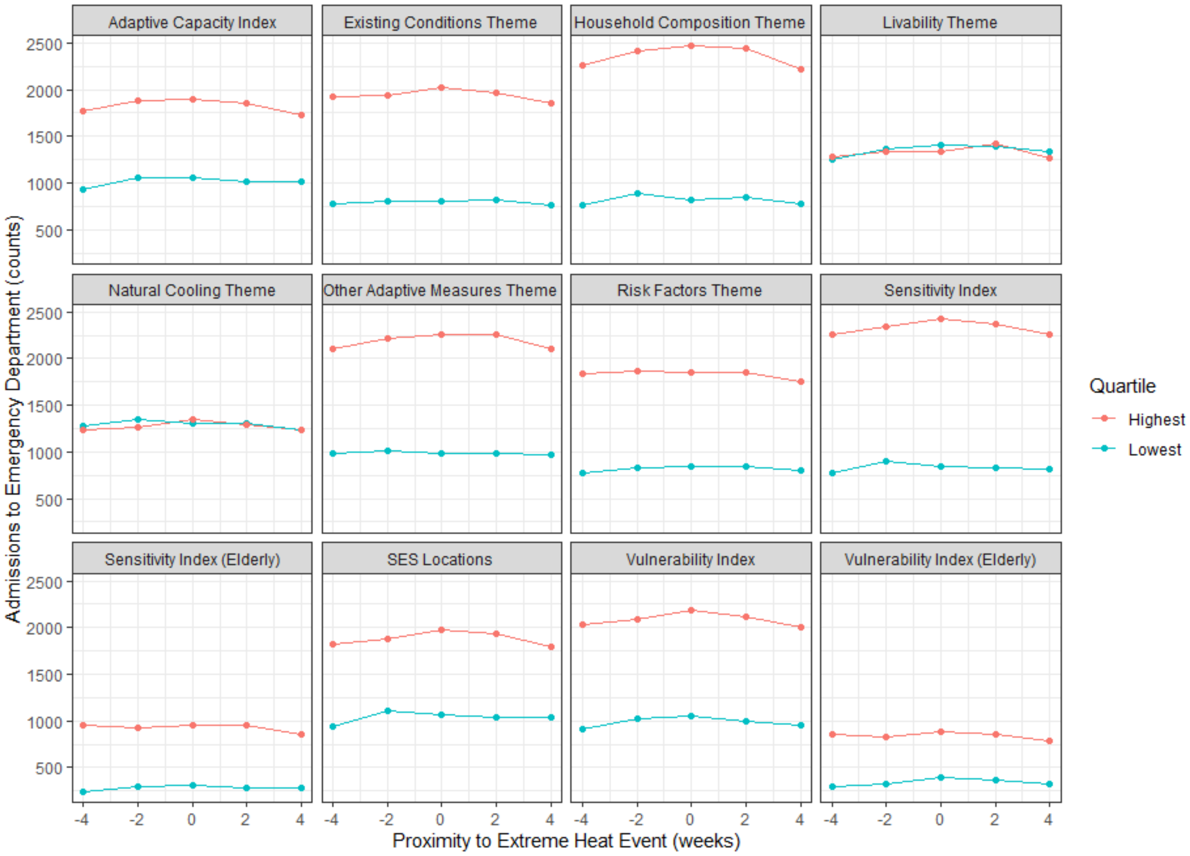


Figure 3.15: Admissions to Emergency Department count.

3.4. BENEFITS OF THIS PROJECT/CASE STUDY

The case study produces Australian Urban Heat Health Vulnerability indices combining data on vulnerable populations, climate, environmental factors, existing health conditions, and heat-related deaths. This indicator contributes to a better understanding of urban health vulnerabilities to heatwaves and extreme temperatures, aligning with the Australian National Environmental Science Programme 2020 Climate Systems and Sustainable Communities hubs. The insights gained will be invaluable in assessing the broader environmental impacts of climate change on health in Australia, informing policy-related assessments like the Australia State of the Environment Urban Environment report and individual State reports.

The broader impact of the case study is substantial, with beneficiaries including universities, government agencies, NGOs/NFPs, and local planning authorities across

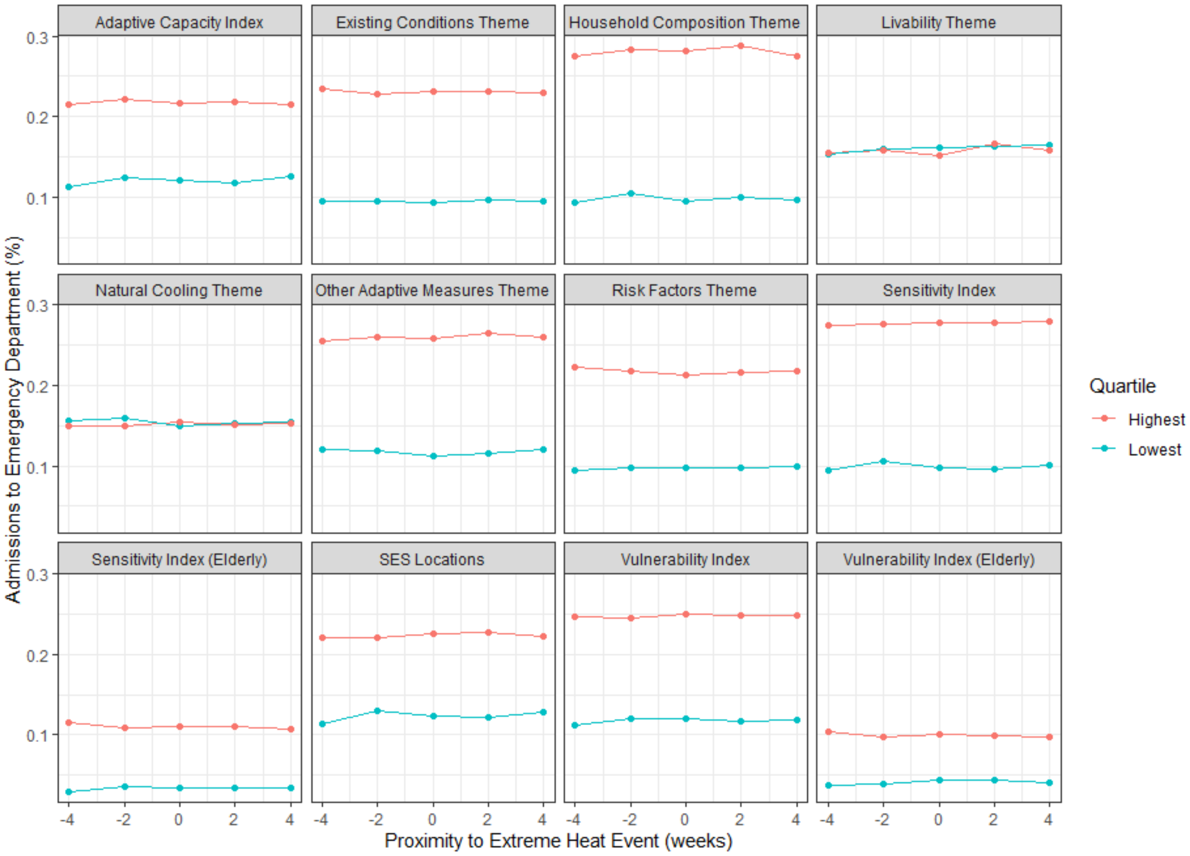


Figure 3.16: Admissions to Emergency Department percentage.

Australia. These new indicators and metrics will provide critical data for evidence-based policymaking and planning in health and social infrastructure. The outcomes are expected to lead to more liveable neighbourhoods, enhanced health services, targeted interventions, and a reduction in preventable diseases and their associated economic costs. The case study's outputs will also inform key reports like the State of the Environment 2021 Urban Environment report and the Social Infrastructure chapter of the 2021 Infrastructure Australia Plan. Moreover, the accessibility to these new indicators and methods will stimulate further research and attract additional investment from various governmental and research organizations, broadening the project's scope and impact.

4. LESSONS LEARNED

This chapter discusses the insights and experiences gathered from the case study. It reflects on the strategies that went well, identifies areas for improvement, and lists key learnings that can benefit future projects.

4.0.1. What Went Well?

Team Collaboration and Expertise The project benefited immensely from strong teamwork, characterized by members with specialized domain knowledge in various aspects such as GIS, environmental science, and epidemiology.

Effective Communication and Problem-Solving Regular catch-up meetings every week and monthly steering committee meetings were crucial in addressing immediate challenges and planning for upcoming tasks. These gatherings enabled the team to stay aligned, tackle roadblocks promptly, and maintain momentum throughout the project's lifecycle.

Engagement with the Broader Research Community Participation in conferences and workshops provided valuable opportunities for exchanging ideas, receiving feedback, and staying abreast of the latest research developments. These interactions enriched the project with fresh perspectives and potential avenues for innovation.

Supportive External Partnerships The cooperation with CHeReL stood out as exceptionally beneficial. Their responsiveness and support significantly eased the process of applying and accessing the linked health datasets.

4.0.2. What Could be Improved?

Streamlining Data Access Processes The project encountered long data application procedures that delayed data acquisition and analysis. While effective mitigation strategies were in place, the project still faced delays due to identified risks.

Streamlining these processes, possibly through policy changes, could substantially reduce the time and effort required to initiate and execute similar studies in the future.

Adaptability to Unforeseen Changes The necessity to alter the study area, methodology, and linked data handling procedures due to reasons such as data unavailability and low counts at SA1 level underscores the need for flexibility in research design. Future initiatives could incorporate modular frameworks that allow for adjustments without significant disruptions to the project timeline or objectives.

4.0.3. What was Learned?

Deducing Difficulty in Accessing Health Data The project underscored the difficulties in accessing linked health data, a frustration compounded by the fragmented landscape of data custodianship in Australia. Efforts to streamline these processes must continue, leveraging the work of organizations like RADiANT and PHRN.

Data Quality and Analysis Limitations The low counts in SA1 level linked health data presented substantial challenges, limiting the ability to derive significant insights. This experience highlights the importance of robust data quality and the need for innovative analytical methods that can accommodate such limitations.

Infrastructure and Resource Requirements The high computing demands for data processing and analysis in the SURE environment brought attention to the importance of planning for adequate computational resources. Budget allocations should account for these needs to avoid potential bottlenecks.

4.1. FEEDBACK TO AURIN/ARDC (IN CONFIDENCE)

The project team extends its gratitude to AURIN and ARDC for their support and collaboration.

5. FUTURE WORK

This chapter discusses the future work from three perspectives: technical developments, collaborations, and dissemination.

5.1. TECHNICAL DEVELOPMENTS

Technical developments will focus on expanding the capabilities and reach of the current study. Anticipated outcomes include the integration of new datasets, the application of novel analytical methods, and the exploration of methodologies to work directly with linked health data. This will potentially benefit researchers, policymakers, and urban planners by providing more detailed insights into health vulnerabilities associated with urban heatwaves.

Next Steps in Analysis The pilot project focused on emergency department and hospital records, requiring harmonization of diagnosis codes (SNOMED in ED data and ICD in hospital records). This effort is in the early stages, with discussions ongoing with potential partners, including health institutions and research bodies. Funding opportunities from health research foundations and government grants are being explored.

Next Steps in Data Acquisition Future projects aim to include ambulance data for location-specific health event analysis, contrasting with hospital data that only indicates the patient's residential location. Additional datasets such as the Perinatal Death Review Database and Mental Health Ambulatory Data Collection will be considered. The use of National Integrated Health Services Information (NIHSI) will also be explored, providing a comprehensive view of health services data across Australia. These developments are still in the planning phase, with potential support from health data agencies and research organizations.

5.2. COLLABORATIONS AND DISSEMINATION

Collaborations will be pivotal in advancing the research and application of urban heat-health vulnerability indices. Engagements with academic institutions, health organizations, and international research teams will be pursued.

Chicago Team and Other Collaborations Ongoing discussions with Northwestern University and the National Centre for Healthy Ageing (NCHA) aim at establishing a collaborative framework for shared research initiatives. Grant applications to the NHMRC, Heart Foundation, and US funding bodies are in preparation, focusing on data harmonization and joint studies on urban health vulnerabilities. Support from future partners, including statements of collaboration, will strengthen these applications.

Access to Health Data in Australia Improving access to health data in Australia is a critical aspect of future work. Efforts will be directed towards advocating for streamlined data access processes, engaging with data custodians, and participating in policy discussions. The aim is to reduce barriers to accessing linked health data, thereby facilitating more efficient and impactful research.

Dissemination/Documentation/Maintenance Future work will also involve continuous dissemination of findings, maintenance of the data asset, and documentation of methodologies and results. AURIN Data Provider will serve as the key platform for sharing resources, ensuring the project's long-term impact and accessibility. Engagement strategies will include workshops, conferences, and online platforms to reach a broad audience and foster community involvement in urban health research.

BIBLIOGRAPHY

[1] S. L. Cutter, B. J. Boruff, W. L. Shirley, Social vulnerability to environmental hazards, in: Hazards vulnerability and environmental justice, Routledge, 2012, pp. 143–160.

[2] B. E. Flanagan, E. W. Gregory, E. J. Hallisey, J. L. Heitgerd, B. Lewis, A social vulnerability index for disaster management, Journal of homeland security and emergency management 8 (1) (2011).

[3] Australian Bureau of Statistics, Technical paper: Census of population and housing: socio-economic indexes for areas SEIFA, Australian Bureau of Statistics, Canada (2001).

[4] A. Dwyer, C. Zoppou, O. Nielsen, S. Day, S. Roberts, Quantifying social vulnerability: a methodology for identifying those at risk to natural hazards, Geoscience Australia Canberra (2004).

[5] A. Price, K. Mengersen, F. Barar, R. Gupta, C. Davis, AusEnHealth methods report, Tech. rep., FrontierSI (2022).

[6] M. L. Parry, Climate change 2007-impacts, adaptation and vulnerability: Working group II contribution to the fourth assessment report of the IPCC, Vol. 4, Cambridge University Press, 2007.

[7] M. Beaty, B. Varghese, Reducing illness and lives lost from heatwaves, Australian Bureau of Statistics, Canada (2022).

[8] J. Yu, K. Castellani, K. Forysinski, P. Gustafson, J. Lu, E. Peterson, M. Tran, A. Yao, J. Zhao, M. Brauer, Geospatial indicators of exposure, sensitivity, and adaptive capacity to assess neighbourhood variation in vulnerability to climate change-related health hazards, Environmental Health 20 (2021) 1–20.

[9] A. El-Zein, F. N. Tonmoy, Assessment of vulnerability to climate change using a multi-criteria outranking approach with application to heat stress in Sydney, Ecological Indicators 48 (2015) 207–217.

[10] University of New South Wales (UNSW), Heat vulnerability index for Sydney, https://cityfutures.ada.unsw.edu.au/cityviz/heat-vulnerability-index-sydney/, [Online; accessed August-2021] (2021).

[11] L. Inostroza, M. Palme, F. De La Barrera, A heat vulnerability index: spatial patterns of exposure, sensitivity and adaptive capacity for Santiago de Chile, PLOS one 11 (9) (2016) e0162464.

[12] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (1) (2016) 1–9.

[13] C. E. Reid, M. S. O'neill, C. J. Gronlund, S. J. Brines, D. G. Brown, A. V. Diez-Roux, J. Schwartz, Mapping community determinants of heat vulnerability, Environmental health perspectives 117 (11) (2009) 1730–1736.

[14] C. E. Reid, J. K. Mann, R. Alfasso, P. B. English, G. C. King, R. A. Lincoln, H. G. Margolis, D. J. Rubado, J. E. Sabato, N. L. West, et al., Evaluation of a heat vulnerability index on abnormally hot days: an environmental public health tracking study, Environmental health perspectives 120 (5) (2012) 715–720.

[15] D. P. Johnson, A. Stanforth, V. Lulla, G. Luber, Developing an applied extreme heat vulnerability index utilizing socioeconomic and environmental data, Applied Geography 35 (1-2) (2012) 23–31.

[16] K. C. Conlon, E. Mallen, C. J. Gronlund, V. J. Berrocal, L. Larsen, M. S. O'Neill, Mapping human vulnerability to extreme heat: A critical assessment of heat vulnerability indices created using principal components analysis, Environmental health perspectives 128 (9) (2020) 097001.

[17] J. R. C.-E. Commission, et al., Handbook on constructing composite indicators: methodology and user guide, OECD publishing, 2008.

A. INPUT DATA CATALOGUE

Continued on the next page.

Table A.1: Input Data Sets

Dataset	Provider	Link	Spatial and Temporal Details
Prevalence of selected chronic diseases and conditions (estimates)	PHIDU	https://data.aurin.org.au/dataset/tua-phidu-phidu-estimates-chronic-disease-pha-2017-18-pha2016	2017-2018, NSW, PHA
Prevalence of selected health risk factors (estimates)	PHIDU	https://data.aurin.org.au/dataset/tua-phidu-phidu-estimates-risk-factors-adults-pha-2017-18-pha2016	2017-2018, NSW, PHA
Home and Community Care Program: Clients living alone, 2014/15 - by PHA, LGA, PHN	PHIDU	https://data.aurin.org.au/dataset/tua-phidu-phidu-home-community-care-program-pha-2014-15-pha2016	2014-2015, NSW, PHA
SEIFA	ABS	https://data.aurin.org.au/dataset/au-govt-abs-seifa-irsad-aust-sa1-2011-sa1	2011, 2016, 2021, NSW, SA1
G01 Selected Person Characteristics by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g01-selected-person-characteristics-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
G02 Selected Medians and Averages-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g02-selected-medians-and-averages-census-2016-sa1-2016	2016, 2021, NSW, SA1
G18 Core Activity Need for Assistance by Age by Sex-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g18-core-activity-need-assist-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G07 Indigenous Status by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g07-indigenous-status-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G30 Number of Motor Vehicle by Dwelling Structure-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g30-number-motor-vehicles-by-dwelling-census-2016-sa1-2016	2016, 2021, NSW, SA1
G39 Dwelling Structure by Household Composition and Family Structure	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g39-dwelling-struct-by-hsehold-and-family-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
G21 Unpaid Assistance to a Person with a Disability by Age by Sex-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g21-unpaid-asst-disability-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G22a Unpaid Child Care by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g22a-unpaid-child-care-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G22b Unpaid Child Care by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g22b-unpaid-child-care-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G23a Relationship in Household by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g23a-relationship-in-household-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G23b Relationship in Household by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g23b-relationship-in-household-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
G37 Dwelling Internet Connection By Dwelling Structure	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g37-dwelling-internet-by-dwelling-struct-census-2016-sa1-2016	2016, 2021, NSW, SA1
G43a Labour Force Status by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g43a-labour-force-status-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G43b Labour Force Status by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g43b-labour-force-status-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G45a Labour Force Status by Sex of Parents by Age of Dependent Children for One Parent Families	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g45a-lbr-frc-sts-by-prnt-sx-by-age-of-dep-chdrn-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
G45b Labour Force Status by Sex of Parents by Age of Dependent Children for One Parent Families	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g45b-lbr-frc-sts-by-prnt-sx-by-age-of-dep-chdrn-census-2016-sa1-2016	2016, 2021, NSW, SA1
G46a Non-School Qualification-Level of Education by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g46a-non-scl-qual-lvl-of-edu-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G46b Non-School Qualification-Level of Education by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g46b-non-scl-qual-lvl-of-edu-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G16a Highest Year of School Completed by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g16a-highest-yr-sch-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G16b Highest Year of School Completed by Age by Sex	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g16b-highest-yr-sch-by-age-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
G33 Tenure and Landlord Type by Dwelling Structure	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g33-tenure-landlord-type-by-dwelling-struct-census-2016-sa1-2016	2016, 2021, NSW, SA1
G28 Total Family Income (Weekly) by Family Composition	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g28-total-family-income-by-composition-census-2016-sa1-2016	2016, 2021, NSW, SA1
G58a Occupation by Hours Worked by Sex-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g58a-occup-by-hours-worked-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G58b Occupation by Hours Worked by Sex-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g58b-occup-by-hours-worked-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1
G59 Method of Travel to Work by Sex-Census	ABS	https://data.aurin.org.au/dataset/au-govt-abs-census-sa1-g59-method-of-travel-to-work-by-sex-census-2016-sa1-2016	2016, 2021, NSW, SA1

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	**Provider**	**Link**	**Spatial and Temporal Details**
Income (Including Government Allowances) (SA2) 2011-2019	ABS	https://data.aurin.org.au/dataset/au-govt-abs-abs-data-by-region-income-asgs-sa2-2011-2019-sa2-2016	2011-2019, NSW, SA2
Family & Community (SA2) 2011-2018	ABS	https://data.aurin.org.au/dataset/au-govt-abs-abs-data-by-region-family-and-community-asgs-sa2-2011-2018-sa2-2016	2011-2018, NSW, SA2
Health & Disability (SA2) 2011-2018	ABS	https://data.aurin.org.au/dataset/au-govt-abs-abs-data-by-region-health-and-disability-asgs-sa2-2011-2018-sa2-2016	2011-2018, NSW, SA2
Persons Born Overseas (SA2) 2011-2016	ABS	https://data.aurin.org.au/dataset/au-govt-abs-abs-data-by-region-persons-born-overseas-asgs-sa2-2011-2016-sa2-2016	2011-2016, 2021, NSW, SA2
Population Estimates by Age and Sex (SA2) 2017-2019	ABS	https://data.aurin.org.au/dataset/au-govt-abs-abs-regional-population-age-sex-sa2-2017-sa2-2016	2017-2020, NSW, SA2

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	Provider	Link	Spatial and Temporal Details
Social and Economic Indicators - Synthetic Estimates SA2 2016	NATSEM	https://data.aurin.org.au/dataset/uc-natsem-natsem-social-indicators-estimates-sa2-2016-sa2-2016	2016, NSW, SA2
Financial Indicators - Synthetic Estimates SA2 2016	NATSEM	https://data.aurin.org.au/dataset/uc-natsem-natsem-financial-indicators-synthetic-sa2-2016-sa2-2016	2016, NSW, SA2
Dependency rate SA2 2016	NATSEM	https://data.aurin.org.au/dataset/uc-natsem-natsem-social-indicators-dependency-rate-sa2-2016-sa2-2016	2016, NSW, SA2
Housing and transport 2016	PHIDU	https://data.aurin.org.au/dataset/tua-phidu-phidu-housing-transport-pha-2016-pha2016	2016-2020, NSW, PHA
Income support recipients	PHIDU	https://data.aurin.org.au/dataset/tua-phidu-phidu-income-support-pha-2017-20-pha2016	2017-2020, NSW, PHA
SILO - Australian climate data from 1889 to yesterday	SILO	https://www.longpaddock.qld.gov.au/silo/	2016-2021, NSW, raster
Geoscape buildings	Geoscape	https://geoscape.com.au/data-on-demand/	2020, NSW, polygon

Continued on next page

Table A.1: Input Data Sets (Continued)

Dataset	**Provider**	**Link**	**Spatial and Temporal Details**
Geoscape surface features	Geoscape	https://geoscape.com.au/data-on-demand/	2020, NSW, raster
Health services locations	HealthDirect	https://data.aurin.org.au/dataset/healthdirect-nhsd-point-november-2020-na	2020, NSW, point

B. OUTPUT METADATA

Continued on the next page.

Table B.1: Data File List and Description

File Name	Description
percVI_2016_ED.csv	percentile of vulnerability indices in 2016 by equal division disaggregation
percVI_2016_OM.csv	percentile of vulnerability indices in 2016 with disaggregation method omitted
percVI_2016_PD.csv	percentile of vulnerability indices in 2016 with population based disaggregation
percVI_2016_ND.csv	percentile of vulnerability indices in 2016 with no division disaggregation
percVI_2021_ED.csv	percentile of vulnerability indices in 2021 by equal division disaggregation
percVI_2021_OM.csv	percentile of vulnerability indices in 2021 with disaggregation method omitted
percVI_2021_PD.csv	percentile of vulnerability indices in 2021 with population based disaggregation
percVI_2021_ND.csv	percentile of vulnerability indices in 2021 with no division disaggregation
rawVI_2016_ED.csv	raw value of vulnerability indices in 2016 by equal division disaggregation
rawVI_2016_OM.csv	raw value of vulnerability indices in 2016 with disaggregation method omitted
rawVI_2016_PD.csv	raw value of vulnerability indices in 2016 with population based disaggregation
rawVI_2016_ND.csv	raw value of vulnerability indices in 2016 with no division disaggregation
rawVI_2021_ED.csv	raw value of vulnerability indices in 2021 by equal division disaggregation
rawVI_2021_OM.csv	raw value of vulnerability indices in 2021 with disaggregation method omitted

Continued on next page

Table B.1: Data File List and Description (Continued)

File Name	Description
rawVI_2021_PD.csv	raw value of vulnerability indices in 2021 with population based disaggregation
rawVI_2021_ND.csv	raw value of vulnerability indices in 2021 with no division disaggregation

Table B.2: Attribute Description

Attribute Name	Description
SA1_CODE21	9 digital SA1 code using 2021 ASGS
roofheight	Average Roof Height (m): The average height of roofs in meters.
builddens	Building Density: A measure of the density of buildings in a given area.
coolroofs	Cool Roofing (%): The percentage of roofs with cool roofing material.
roaddens	Road and Path (%): The percentage of the area covered by roads and paths.
grass	Grass (%): The percentage of the area covered by grass.
canopy	Trees (%): The percentage of the area covered by trees.
otherveg	Unspecified Vegetation (%): The percentage of the area with unspecified vegetation.
tempdevavg	Average Summer Percentile, Deviation from Reference Period (2010-2015): The average deviation from the reference period (2010-2015) for summer weather conditions.
tempdevmax	Maximum Summer Percentile, Deviation from Reference Period (2010-2015): The maximum deviation from the reference period (2010-2015) for summer weather conditions.
ehfavg	Excess Heat Factor (EHF) Average: The average excess heat factor, which indicates the level of excess heat in an area.

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	**Description**
ehfmax	Excess Heat Factor (EHF) Maximum: The maximum excess heat factor observed in an area.
nheatdays	Number of Excess Heat Days: The count of days with excess heat in a given area.
closewater	Distance to Water Body: The distance from a location to the nearest water body.
closeriver	Distance to River: The distance from a location to the nearest river.
popdens	Total Persons: The total number of people in a given area or dataset.
infants	Persons Age Groups 0-4 Years: The number of individuals in the age group of 0 to 4 years old.
elderly	Persons Age Groups 65-74 Years: The number of individuals in the age group of 65 to 74 years old.
indigenous	Aboriginal and/or Torres Strait Islander Persons (Total Persons): The total number of individuals who identify as Aboriginal and/or Torres Strait Islander.
engsl	Language Spoken at Home (Other Language - Persons): The number of individuals who speak a language other than the predominant language at home.

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	**Description**
medincome	Highest Year of School Completed (Year 12 or Equivalent - Persons): The number of individuals who completed Year 12 or its equivalent as their highest level of education.
assistreq	Median Total Family Income Weekly: The median (middle) value of the total weekly income earned by families in the dataset.
assistprov	Persons Total Has Need for Assistance: The total number of individuals who require assistance or support.
livalone	Persons Total Provided Unpaid Assistance: The total number of individuals who provide unpaid assistance or support to others.
lowincome	Persons Lone Person Total: The total number of individuals who live alone.
paralone	400-499 Total: The total count within the specified range (400 to 499).
rent	Total One-Parent Family: The total number of one-parent families.
unemp	Number of Motor Vehicles per Dwelling (No Motor Vehicles - Dwellings): The count of dwellings with no motor vehicles.
notinlab	Rented Total Total: The total number of rented dwellings.
machops	Internet Accessed from Dwelling Total: The total number of dwellings with internet access.

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	Description
labourers	Persons Total Unemployed Total: The total number of individuals who are unemployed.
crowding	Persons Not in the Labour Force Total: The total number of individuals who are not participating in the labor force.
heartsvd	Persons Machinery Operators and Drivers Total: The total number of individuals working as machinery operators and drivers.
diabetes	Persons Labourers Total: The total number of individuals working as laborers.
psych	Housing Suitability - Occupied Private Dwellings - Census Dwellings with Extra Bedrooms Needed: The number of occupied private dwellings that require additional bedrooms.
bloodpress	Estimated Number of People with Heart, Stroke, and Vascular Disease (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals with heart, stroke, and vascular disease per 100 people, based on modeling for the year 2017-18.
overweight	Estimated Number of People with Diabetes Mellitus (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals with diabetes mellitus per 100 people, based on modeling for the year 2017-18.

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	Description
obese	Estimated Number of People Aged 18 Years and Over with High or Very High Psychological Distress (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over experiencing high or very high psychological distress per 100 people, based on the Kessler 10 Scale, for the year 2017-18.
smokers	Estimated Number of People Aged 18 Years and Over Who Had High Blood Pressure (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over with high blood pressure per 100 people, based on modeling for the year 2017-18.
alcohol	Estimated Number of People Aged 18 Years and Over Who Consumed More Than Two Standard Alcoholic Drinks Per Day on Average (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over who consumed more than two standard alcoholic drinks per day on average per 100 people, based on modeling for the year 2017-18.
highschool	Estimated Number of People Aged 18 Years and Over Who Were Obese (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over who were obese per 100 people, based on modeling for the year 2017-18.

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	Description
vehicle	Estimated Number of People Aged 18 Years and Over Who Were Current Smokers (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over who were current smokers per 100 people, based on modeling for the year 2017-18.
internet	Estimated Number of People Aged 18 Years and Over Who Consumed More Than Two Standard Alcoholic Drinks Per Day on Average (Modelled Estimates) 2017-18 ASR per 100: The estimated count of individuals aged 18 and over who consumed more than two standard alcoholic drinks per day on average per 100 people, based on modeling for the year 2017-18.
swpools	Swimming Pool Adjacent (%): The percentage of dwellings with a swimming pool adjacent to them.
liveindex	Liveability Index: An index that measures the overall liveability of an area, taking into account various factors.
Urban_Exposure	Integrated index theme group: Urban_Exposure
Natural_Cooling	Integrated index theme group: Natural_Cooling
Heat_Exposure	Integrated index theme group: Heat_Exposure

Continued on next page

Table B.2: Attribute Description (Continued)

Attribute Name	Description
Other_Sensitive	Integrated index theme group: Other_Sensitive
Household_Composition	Integrated index theme group: Household_Composition
Socio-economic_Status	Integrated index theme group: Socio-economic_Status
Existing_Conditions	Integrated index theme group: Existing_Conditions
Risk_Factors	Integrated index theme group: Risk_Factors
Other_Adaptive	Integrated index theme group: Other_Adaptive
Liveability	Integrated index theme group: Liveability
Exposure	Exposure Heat Health Vulnerability subindex
Sensitivity	Sensitivity Heat Health Vulnerability subindex
Adaptive Capacity	Adaptive Capacity Heat Health Vulnerability subindex
Vulnerability Index	Heat Health Vulnerability Index

Table B.3: Dataset Metadata

Metadata Field	**Details**
Author(s)	AURIN
Owner(s)	University of Melbourne [ABN 84 002 705 224]
Title of Dataset	AusUrbHI Heat Health Vulnerability Indices for NSW Cities and Regions
Abstract	This dataset offers a nuanced perspective on population vulnerability to extreme heat at SA1 level, created from a blend of demographic factors, environmental conditions, and urban morphology datasets. The indices are structured into layers of exposure, sensitivity, and adaptive capacity.
Data Purpose	The dataset reveals elements that influence broader spatial vulnerability patterns, playing a pivotal role in identifying and mitigating risks in the most vulnerable areas, thus fostering healthier, more resilient urban environments.

Continued on next page

Table B.3: Dataset Metadata (Continued)

Metadata Field	**Details**
Lineage	The datasets was generated using AURIN datasets and additional external datasets (e.g., land surface temperature raster). Input datasets were harmonized through various processing techniques accordingly, such as disaggregation, raster to vector transformation, and concordance. The study produces Heat Health Vulnerability Indices (HHVI) for each SA1 region within the case study area using Principle Component Analysis (PCA). Multiple indices are produced to assess differences in downscaling techniques and are calculated over the two most-recent census years (2016 and 2021) to assess changes in vulnerability over time. The index methodology utilises multiple PCA factors scaled by explained variability to produce more complex and data-driven indices from up to 44 indicators of heat exposure, population sensitivity, and adaptive capacity. The generated HHVI was then verified by individual SA1 level linked health data via a time-stratified case-cross study design by comparing the lowest and highest quartiles of vulnerability to hospital admissions during heatwave and non-heatwave days. The study maps areas at high risk of heat stress, correlating these regions with human health and environmental factors.
Field of Research (FoR) Code	"490508 Statistical data science; 401302 Geospatial information systems and geospatial data modelling; 420602 Health equity"

Continued on next page

Table B.3: Dataset Metadata (Continued)

Metadata Field	**Details**
Geographic Extent, Geometry, Aggregation	level of aggregation: SA1 2021 ASGS
Research Associations	Funded by the Australian Research Data Commons (ARDC)
License type	CC BY-NC 4.0 Deed | Attribution-NonCommercial 4.0 International | Creative Commons
Conversion Page	https://aurin.org.au/resources/data/data-release-form/

C. EQUAL DIVIDE VISUALS

C.1. EXPOSURE: NATURAL COOLING

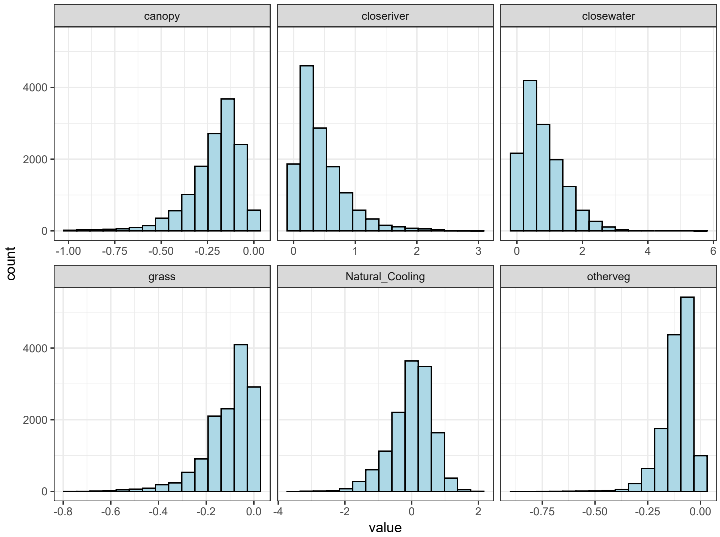


Figure C.1: Histograms of post-processed data and resulting theme index.

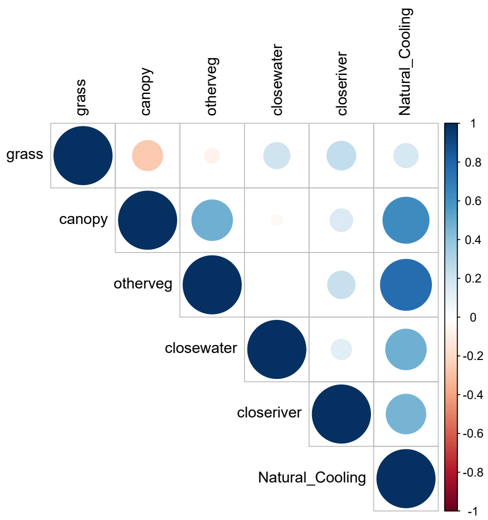


Figure C.2: Correlation plot of theme index and underlying variables.

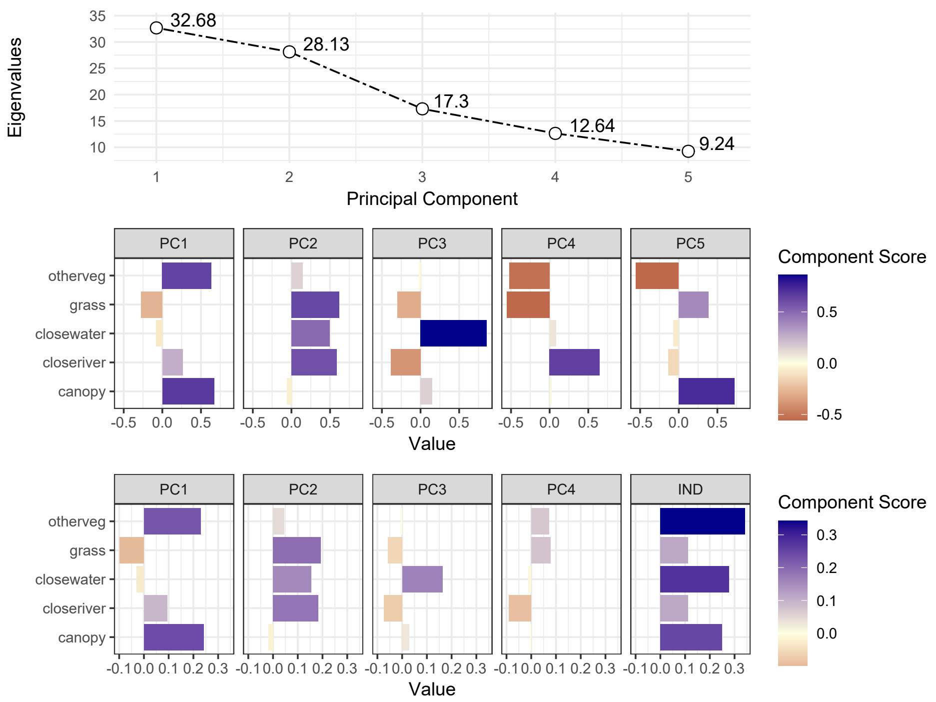


Figure C.3: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.2. EXPOSURE: HEAT EXPOSURE

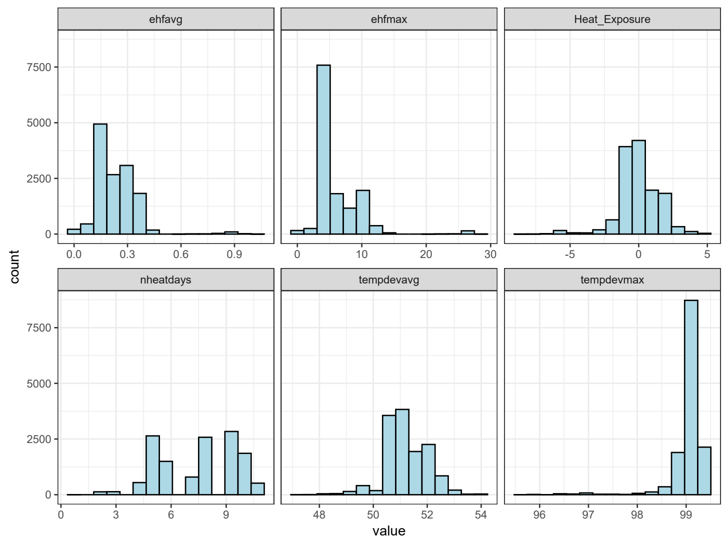


Figure C.4: Histograms of post-processed data and resulting theme index.

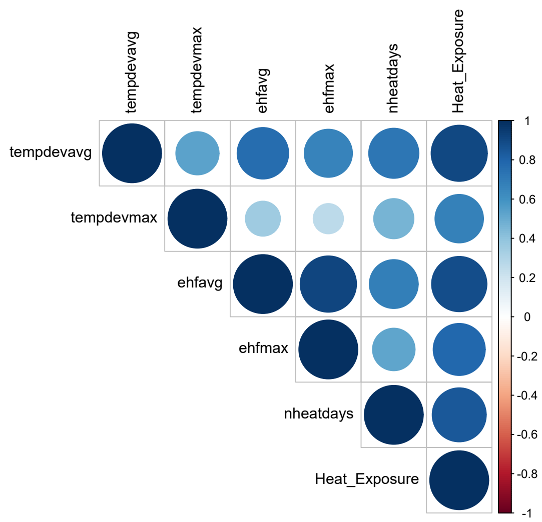


Figure C.5: Correlation plot of theme index and underlying variables.

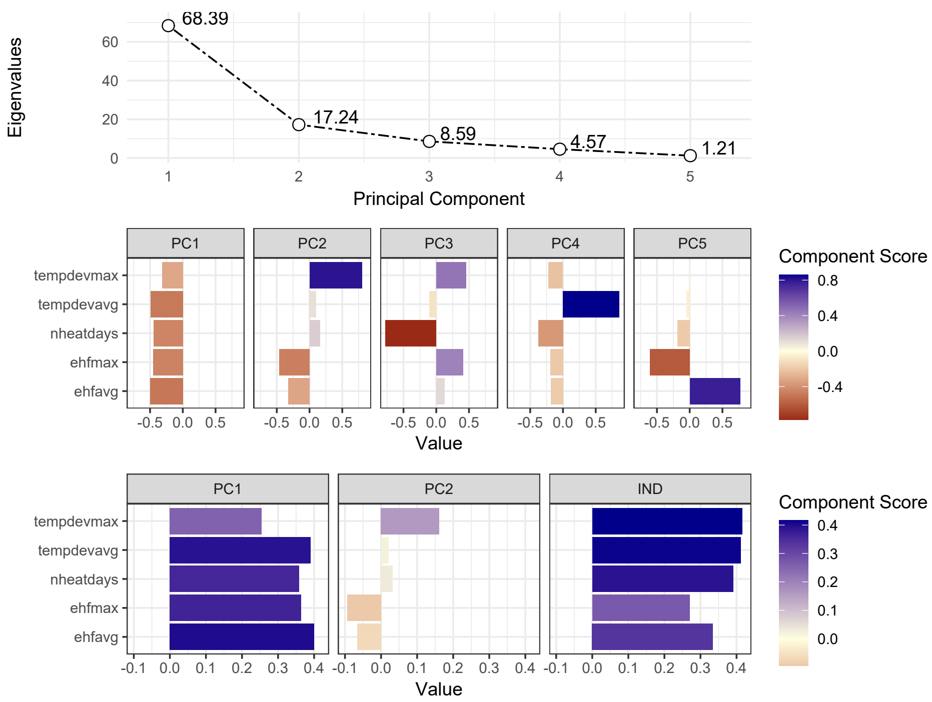


Figure C.6: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.3. EXPOSURE SUB-INDEX

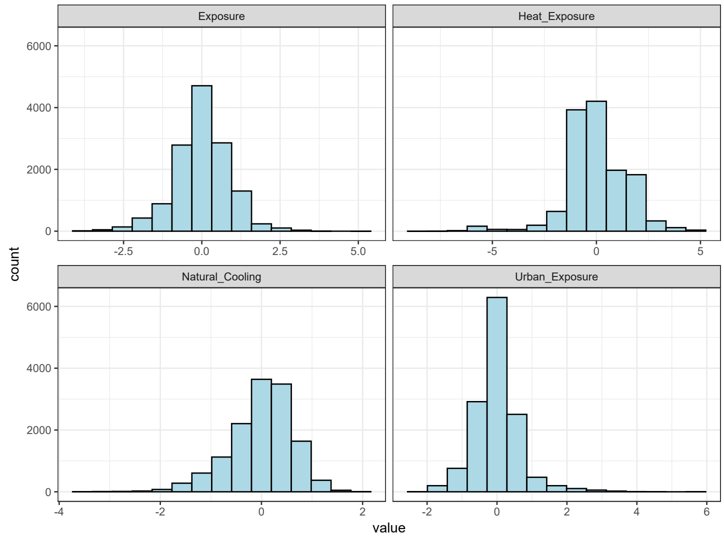


Figure C.7: Histograms of theme indices and resulting sub-index.

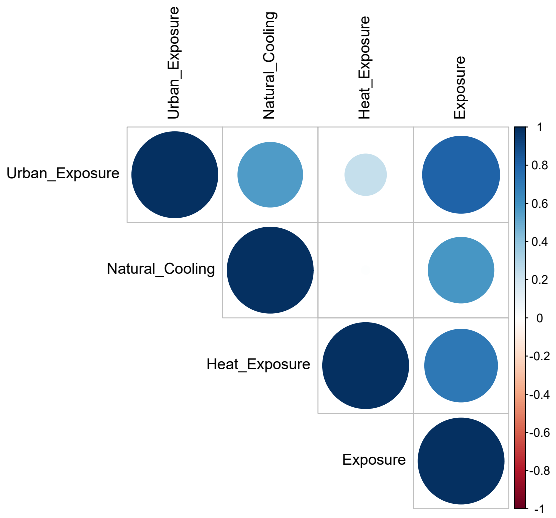


Figure C.8: Correlation plot of theme indices and resulting sub-index.

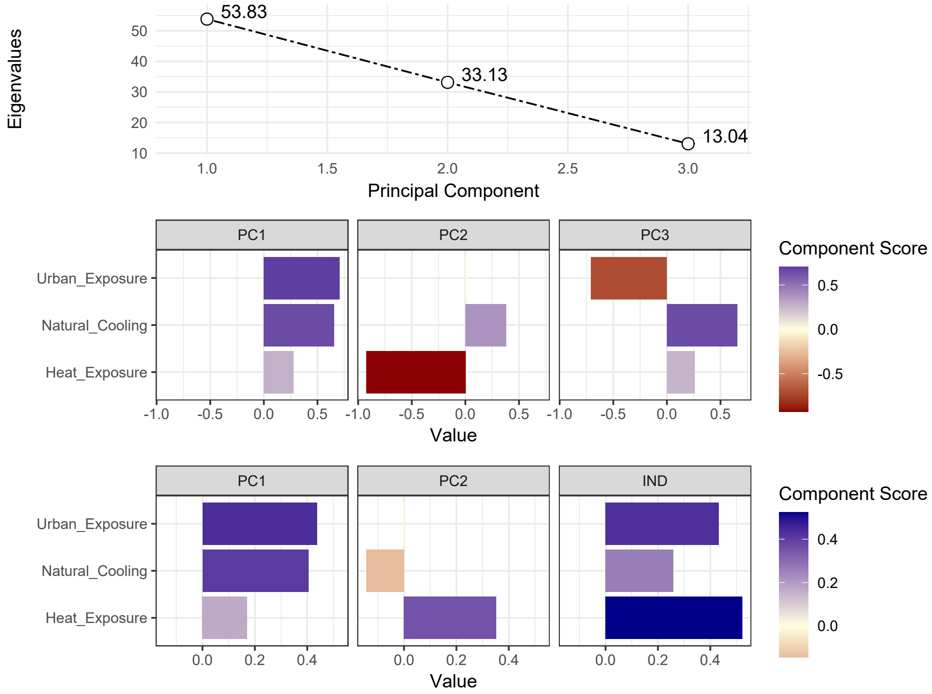


Figure C.9: Sub-index calculation breakdown. See Section 2.5.4 for more details.

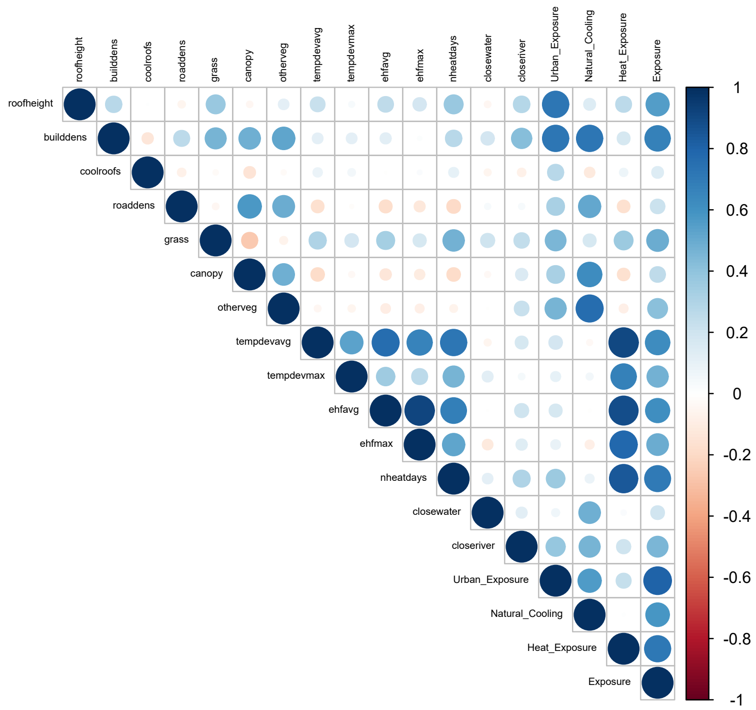


Figure C.10: Correlation plot of the exposure sub-index and all underlying variables.

C.4. SENSITIVITY: HOUSEHOLD COMPOSITION

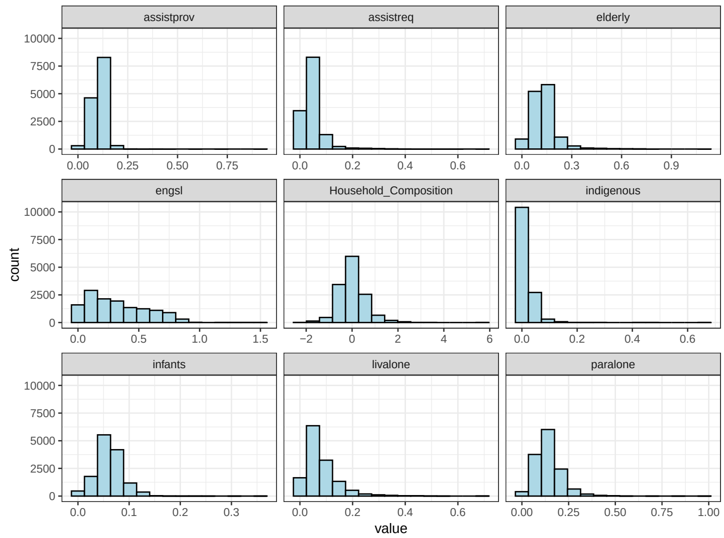


Figure C.11: Histograms of post-processed data and resulting theme index.

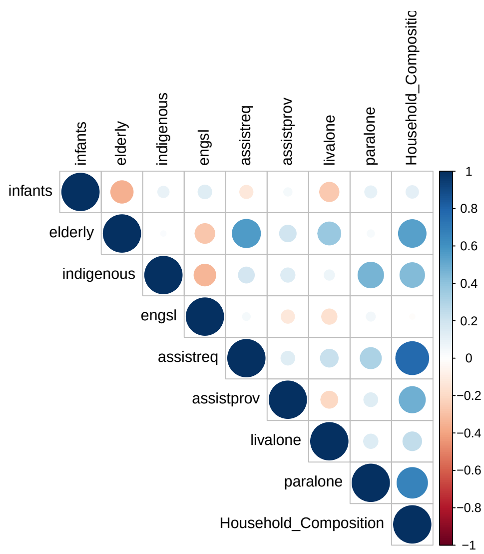


Figure C.12: Correlation plot of theme index and underlying variables.

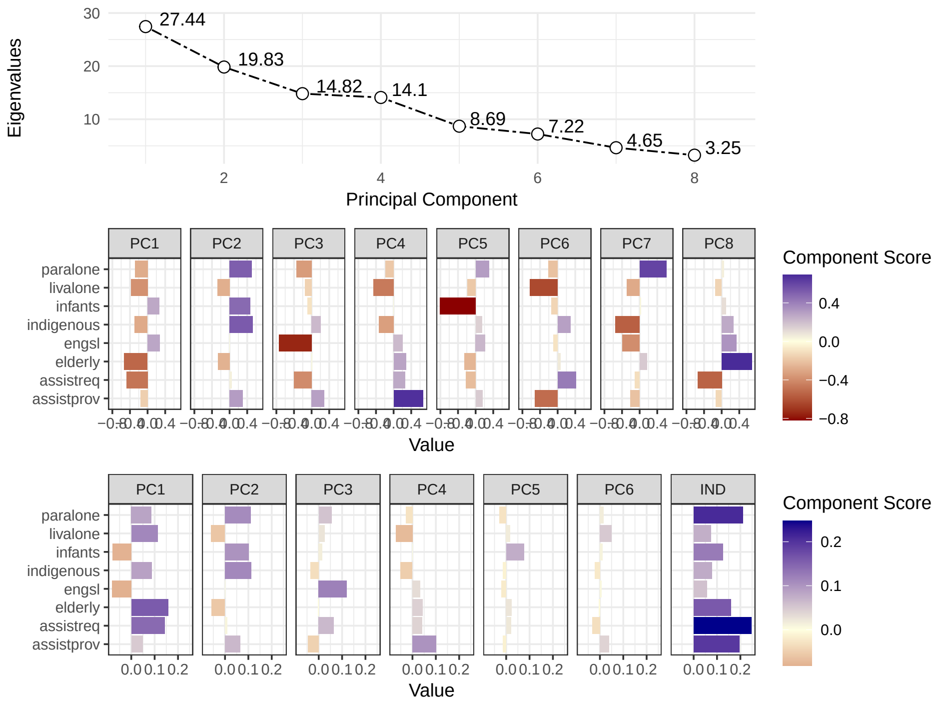


Figure C.13: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.5. SENSITIVITY: OTHER

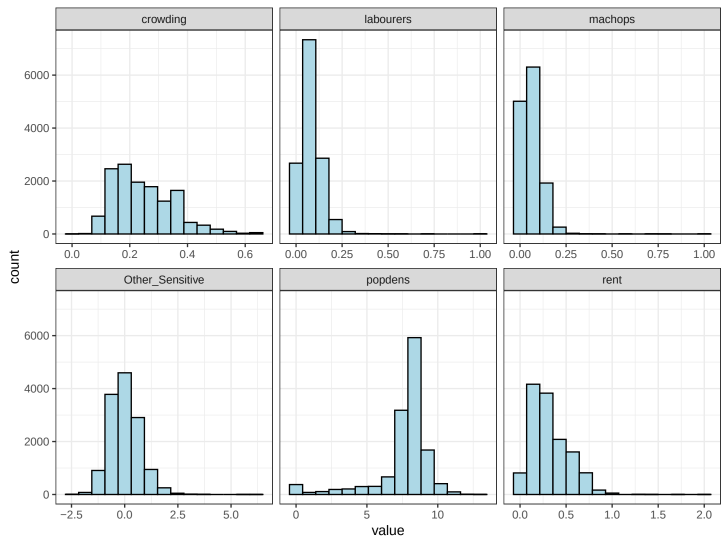


Figure C.14: Histograms of post-processed data and resulting theme index.

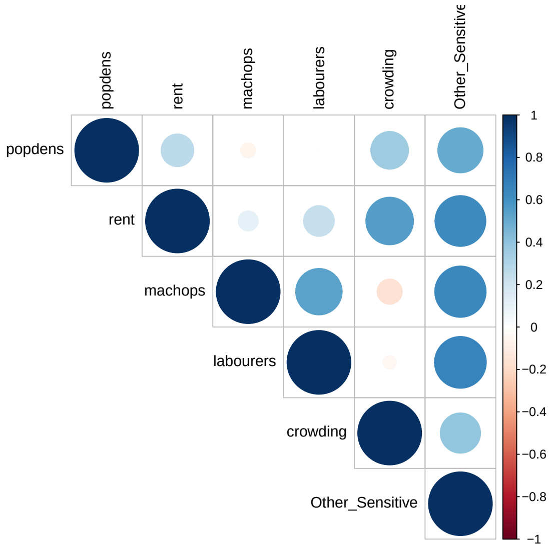


Figure C.15: Correlation plot of theme index and underlying variables.

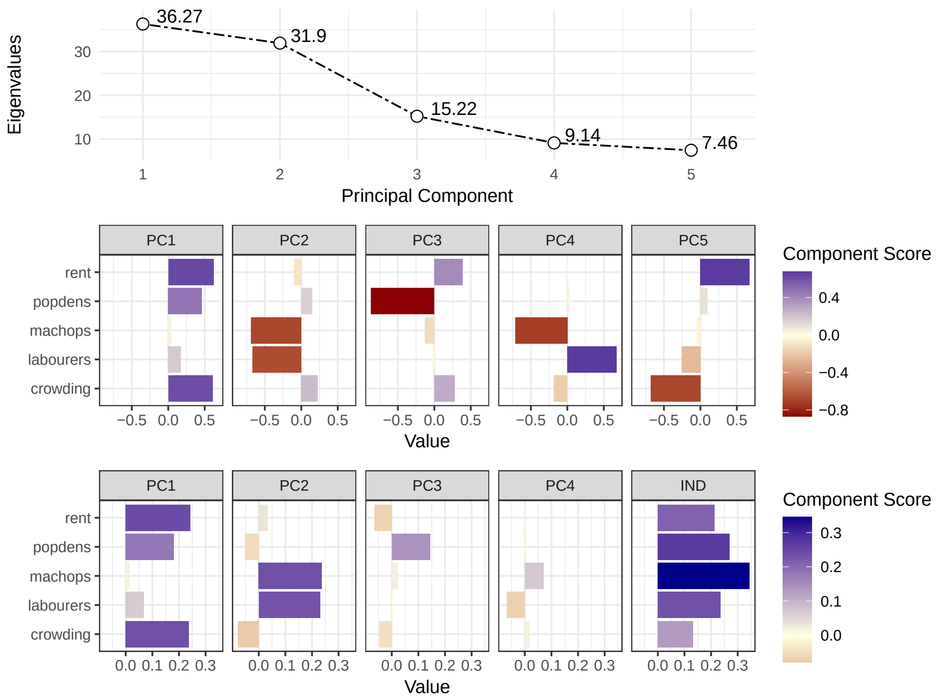


Figure C.16: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.6. SENSITIVITY: SOCIO-ECONOMIC STATUS

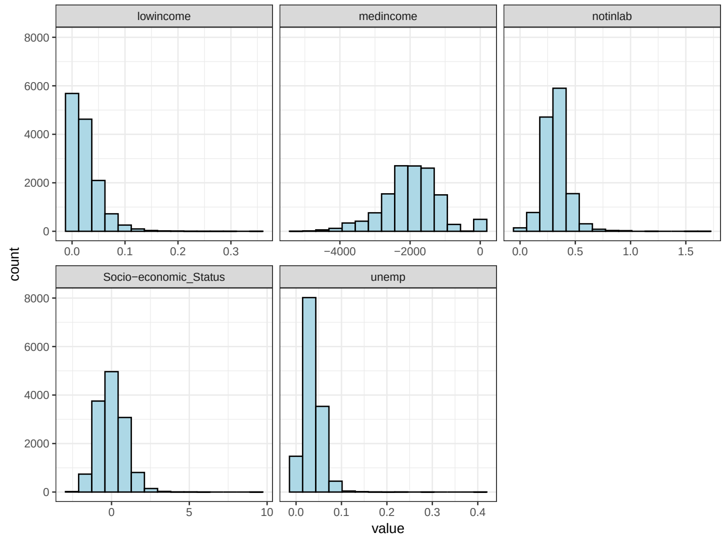


Figure C.17: Histograms of post-processed data and resulting theme index.

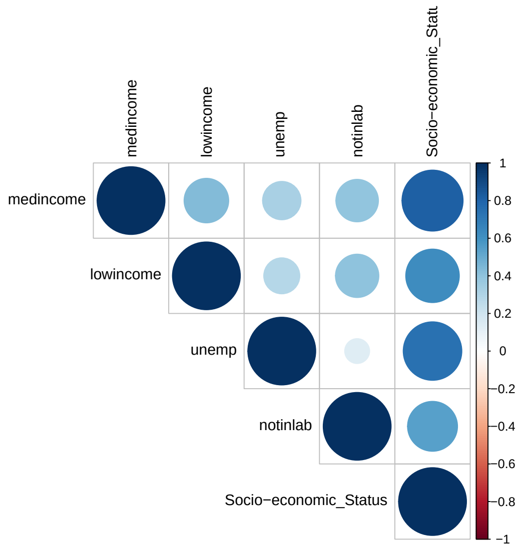


Figure C.18: Correlation plot of theme index and underlying variables.

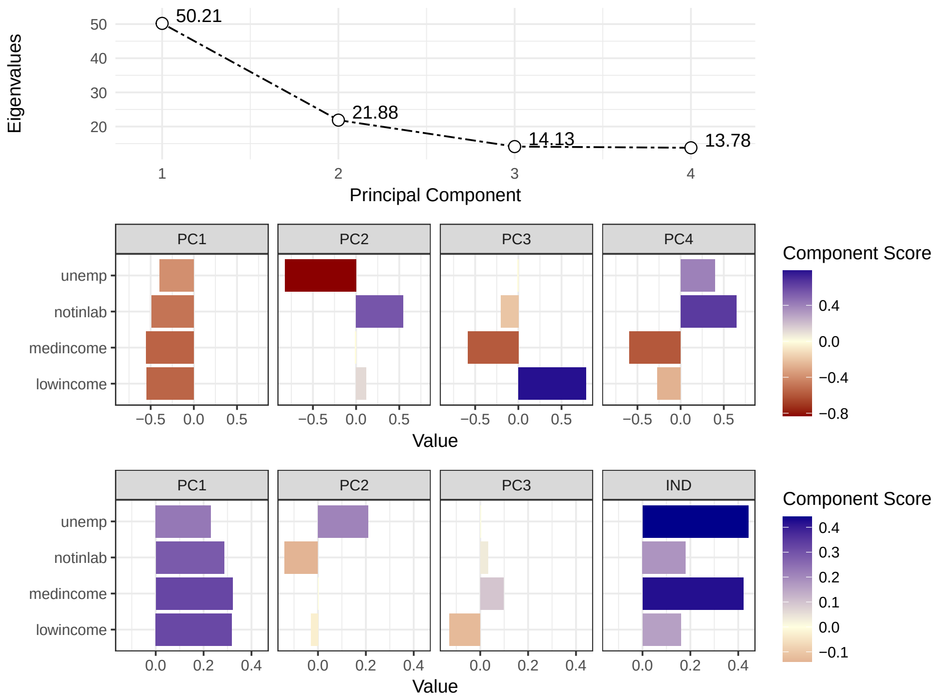


Figure C.19: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.7. SENSITIVITY: EXISTING CONDITIONS

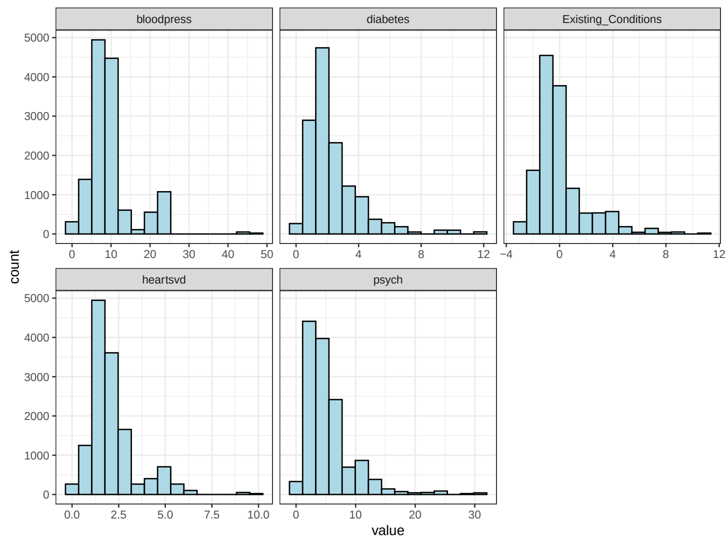


Figure C.20: Histograms of post-processed data and resulting theme index.

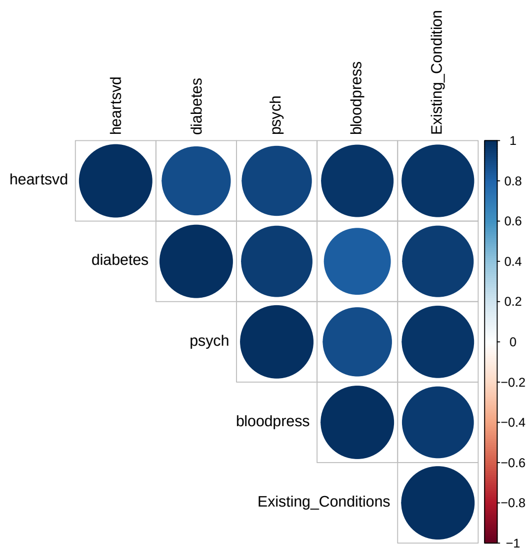


Figure C.21: Correlation plot of theme index and underlying variables.

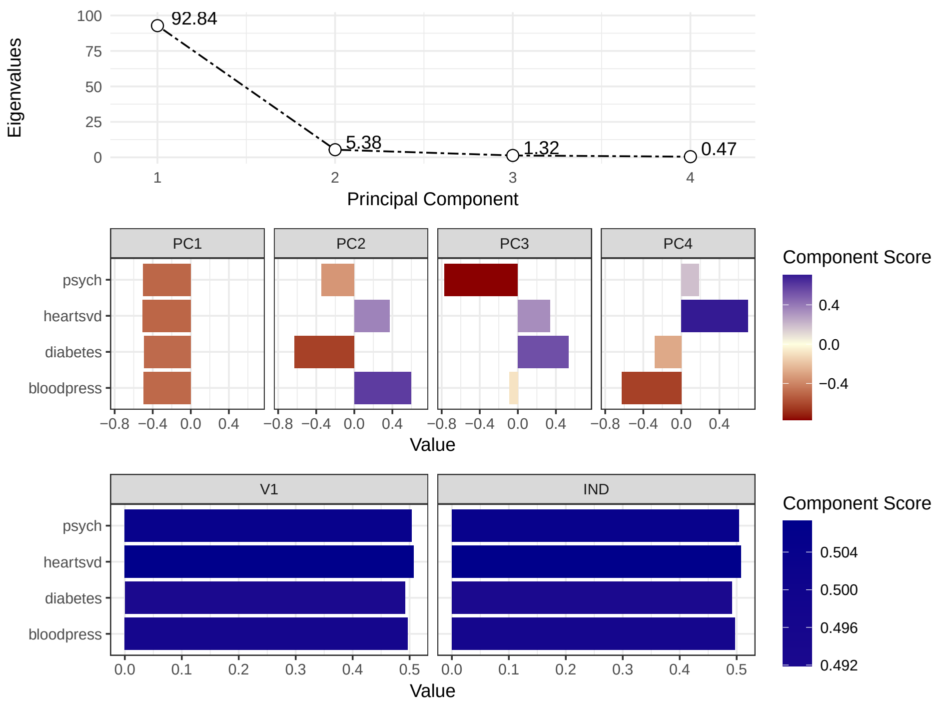


Figure C.22: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.8. SENSITIVITY: RISK FACTORS

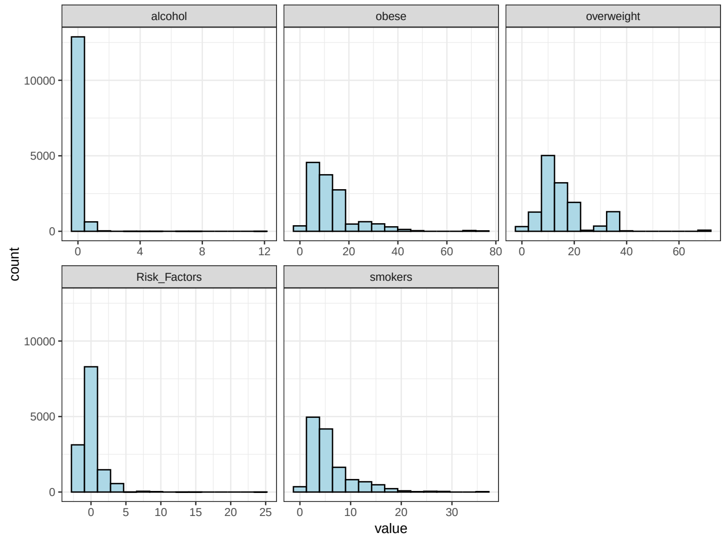


Figure C.23: Histograms of post-processed data and resulting theme index.

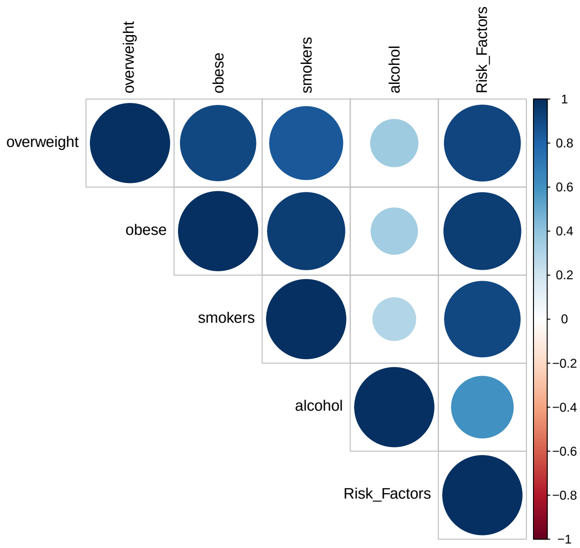


Figure C.24: Correlation plot of theme index and underlying variables.

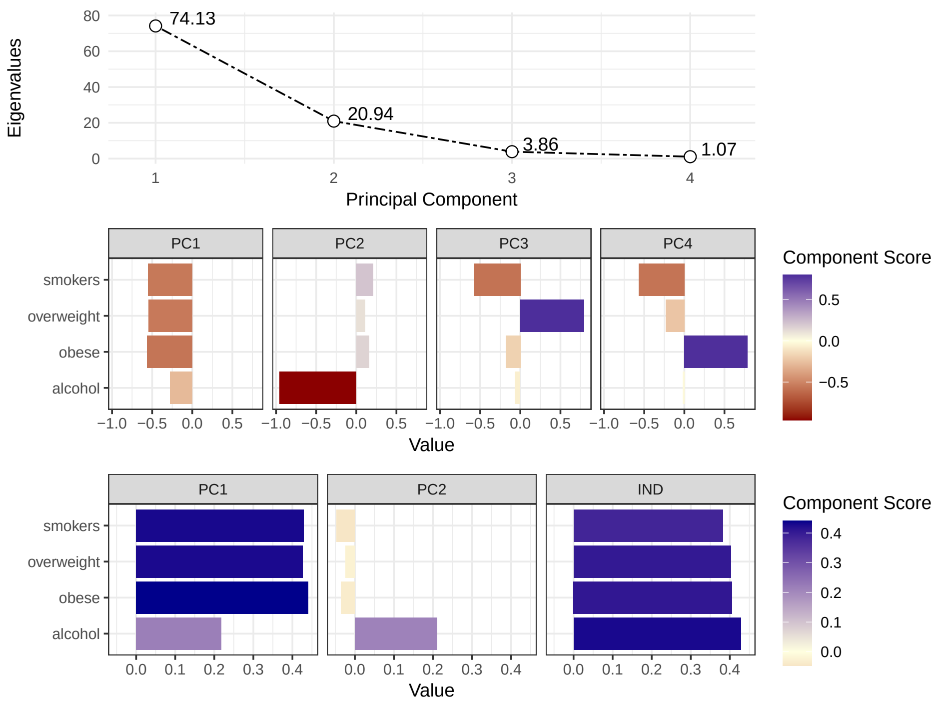


Figure C.25: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.9. SENSITIVITY SUB-INDEX

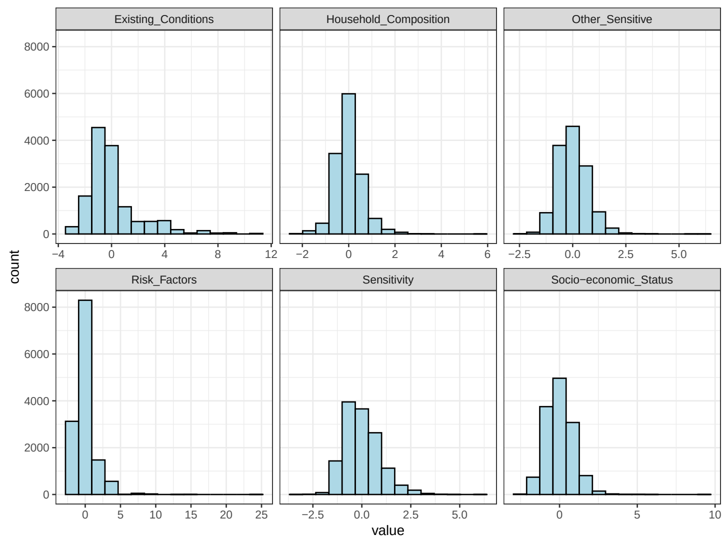


Figure C.26: Histograms of theme indices and resulting sub-index.

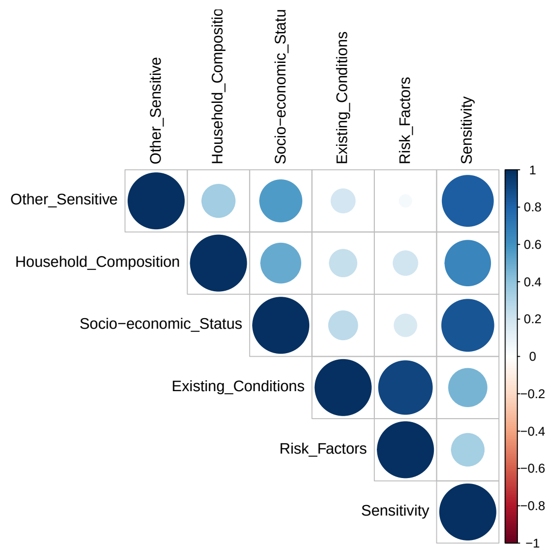


Figure C.27: Correlation plot of theme indices and resulting sub-index.

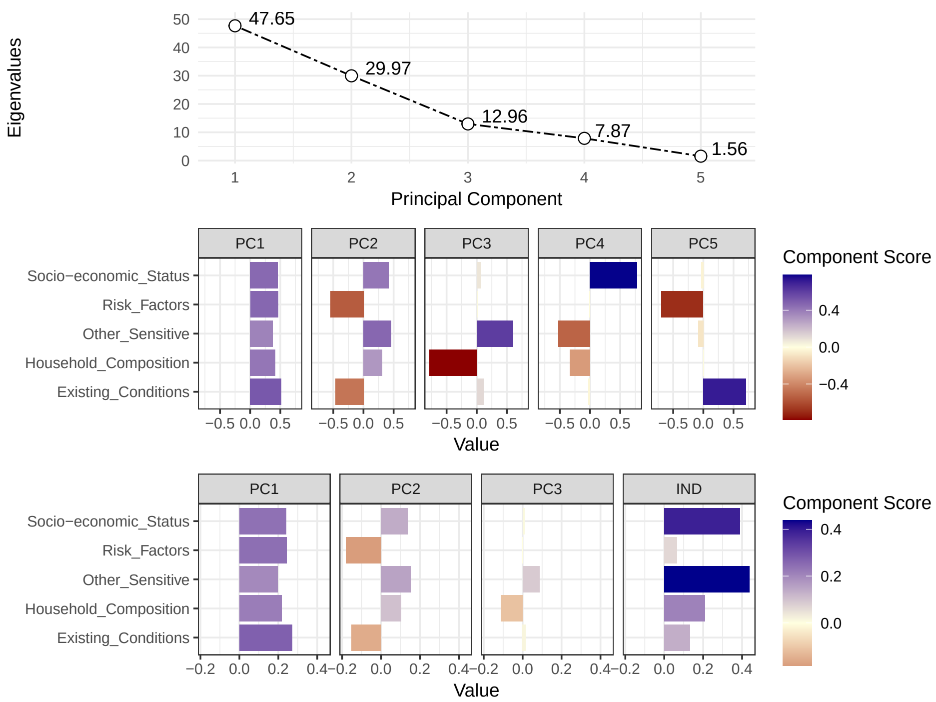


Figure C.28: Sub-index calculation breakdown. See Section 2.5.4 for more details.

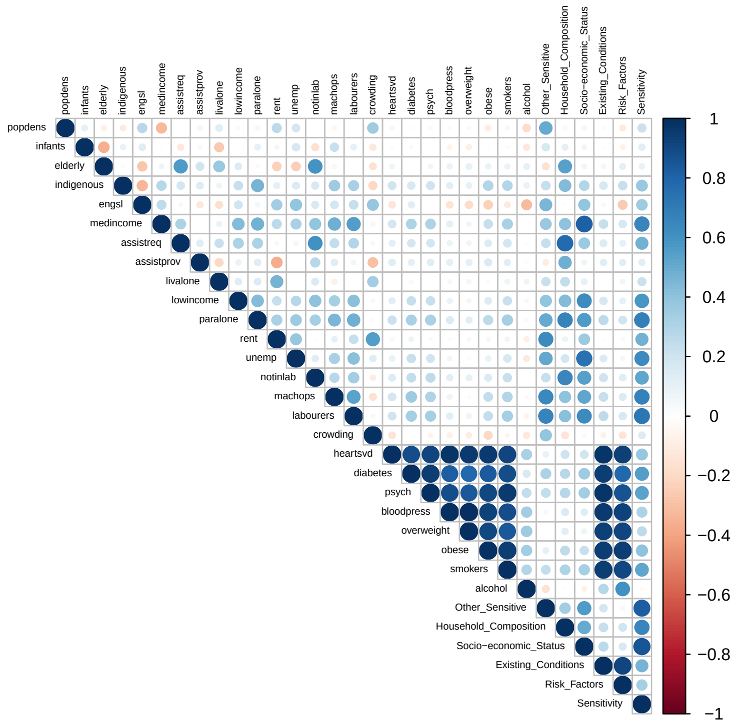


Figure C.29: Correlation plot of the sensitivity sub-index and all underlying variables.

C.10. ADAPTIVE CAPACITY: OTHER

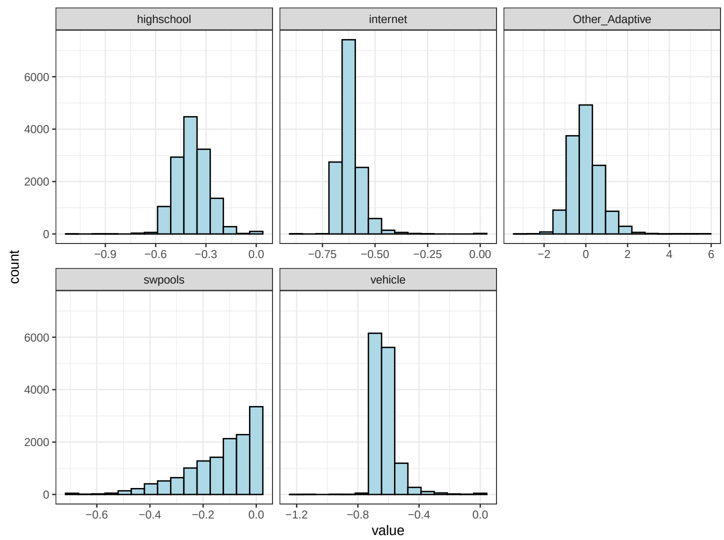


Figure C.30: Histograms of post-processed data and resulting theme index.

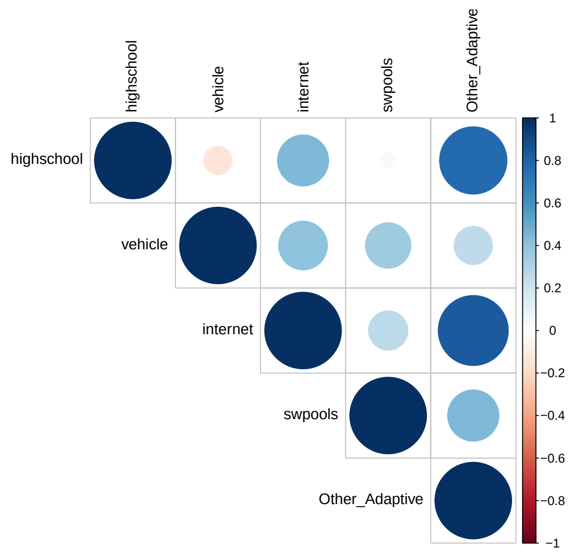


Figure C.31: Correlation plot of theme index and underlying variables.

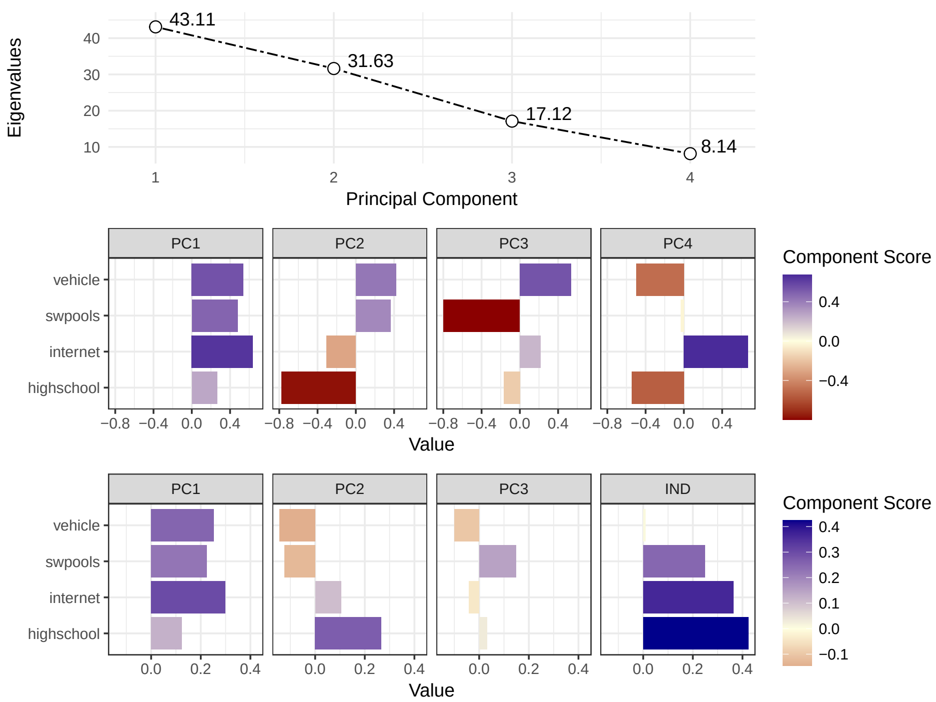


Figure C.32: Theme index calculation breakdown. See Section 2.5.4 for more details.

C.11. ADAPTIVE CAPACITY: LIVEABILITY INDEX

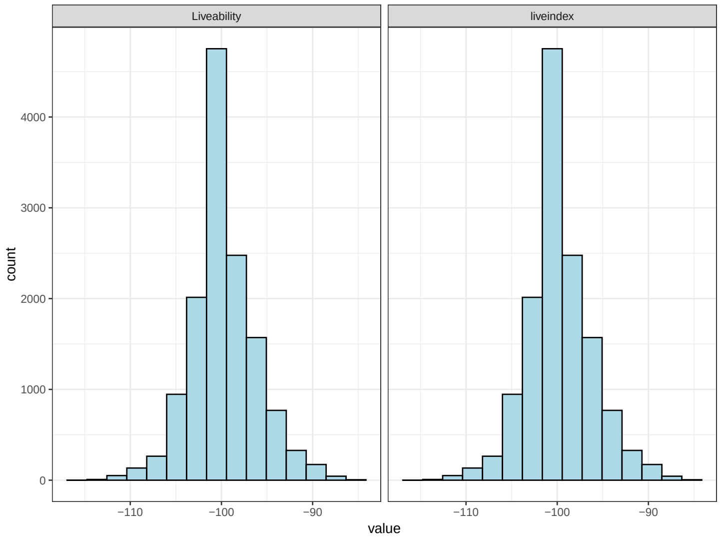


Figure C.33: Histograms of post-processed data and resulting theme index.

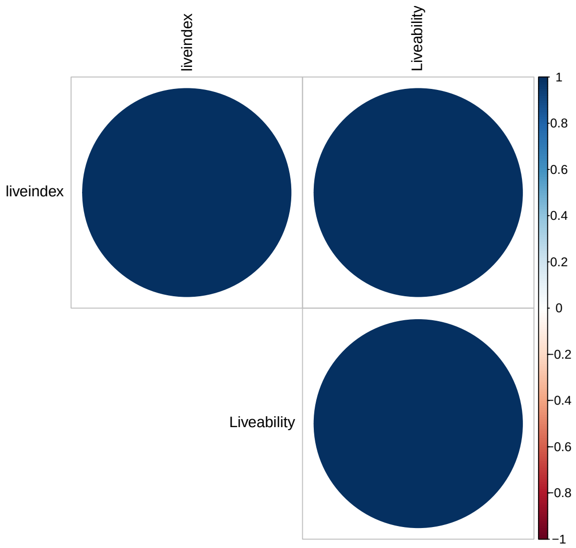


Figure C.34: Correlation plot of theme index and underlying variables.

PCA not performed.

PCA not performed.

PCA not performed.

Figure C.35: No PCA - only one variable in the theme index.

C.12. ADAPTIVE CAPACITY SUB-INDEX

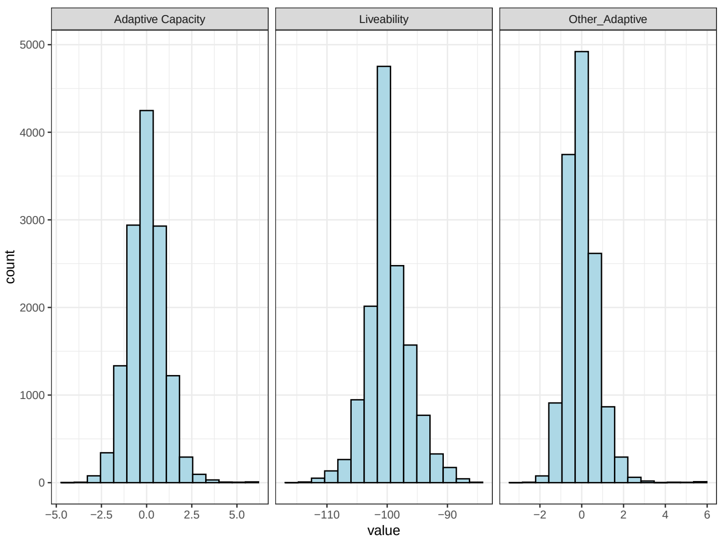


Figure C.36: Histograms of theme indices and resulting sub-index.

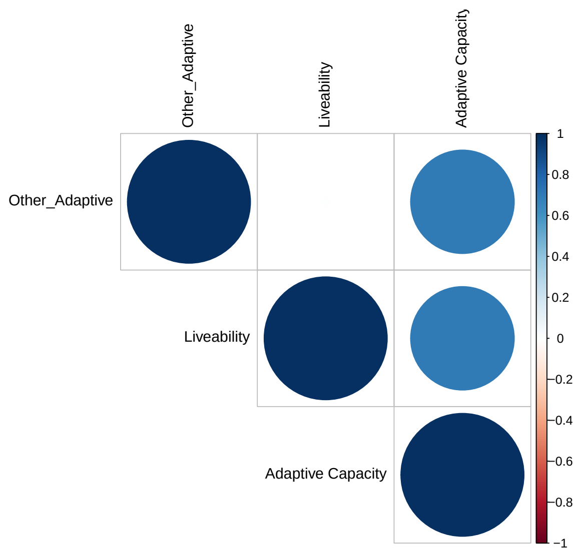


Figure C.37: Correlation plot of theme indices and resulting sub-index.

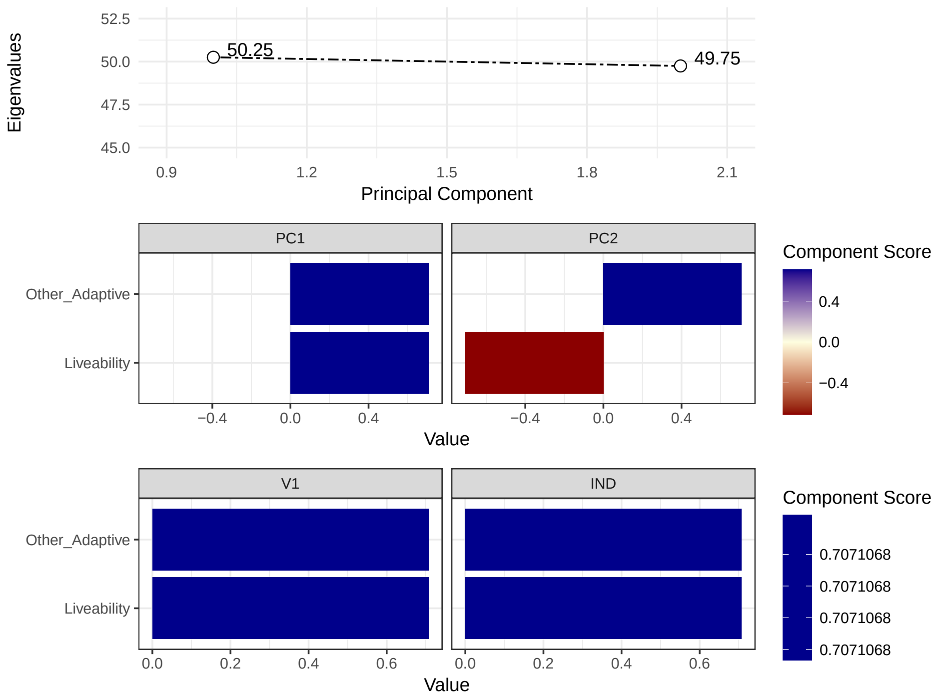


Figure C.38: Sub-index calculation breakdown. See Section 2.5.4 for more details.

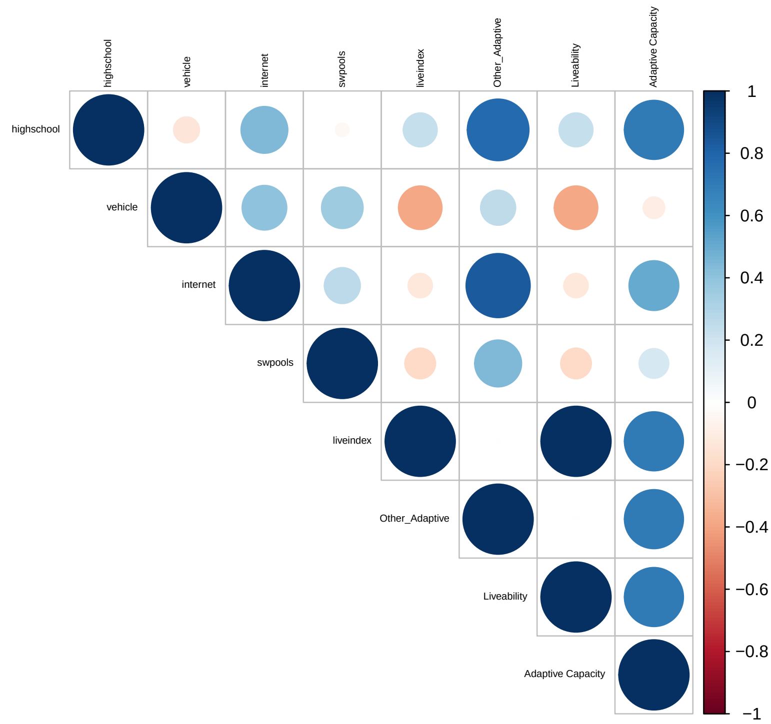


Figure C.39: Correlation plot of the adaptive capacity sub-index and all underlying variables.

C.13. VULNERABILITY INDEX

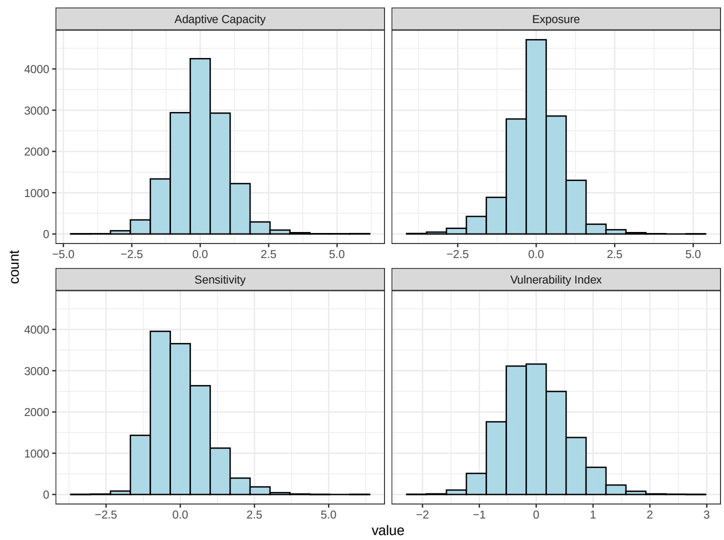


Figure C.40: Histograms of the sub-indices and resulting vulnerability index.

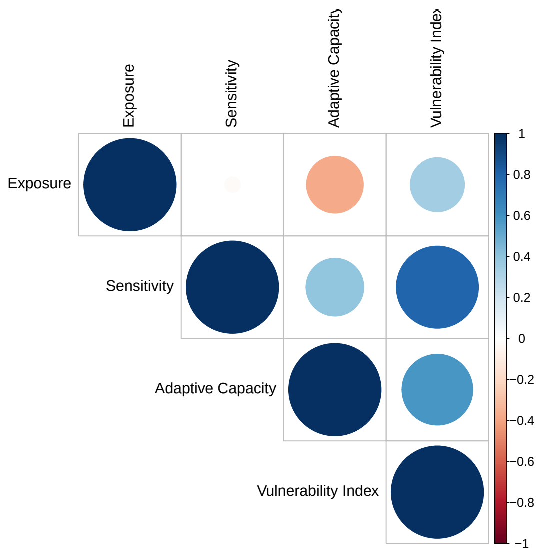


Figure C.41: Correlation plot of sub-indices and the vulnerability index.

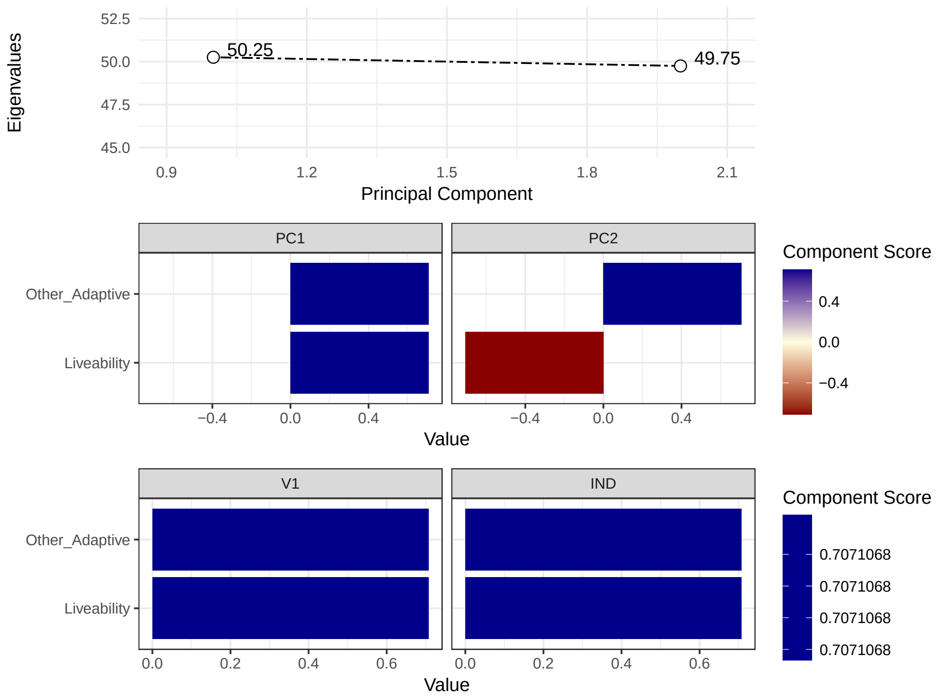


Figure C.42: No PCA - vulnerability index calculation carried out using equal weighting.

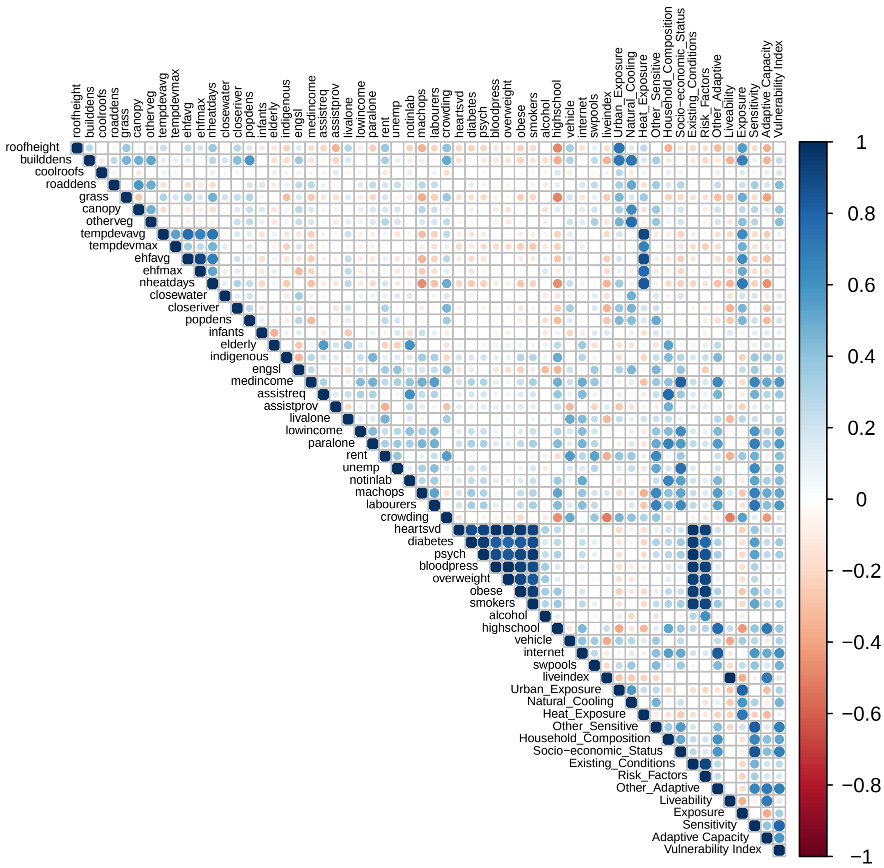


Figure C.43: Correlation plot of the vulnerability index and all underlying variables, their relevant themes and sub-indices.

D. OMISSION VISUALS

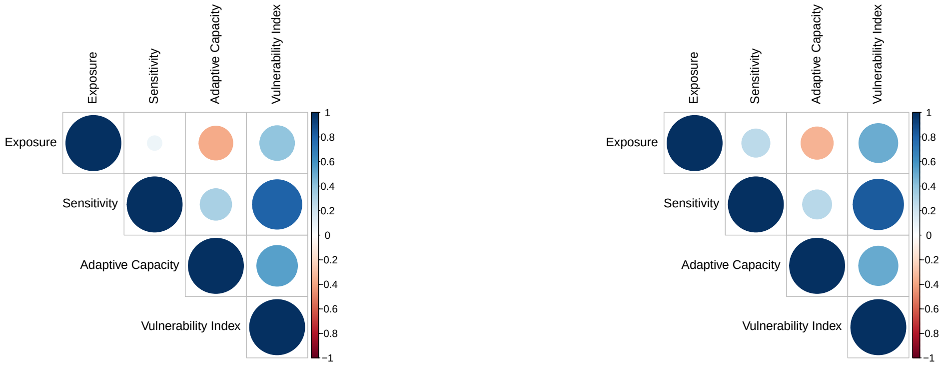
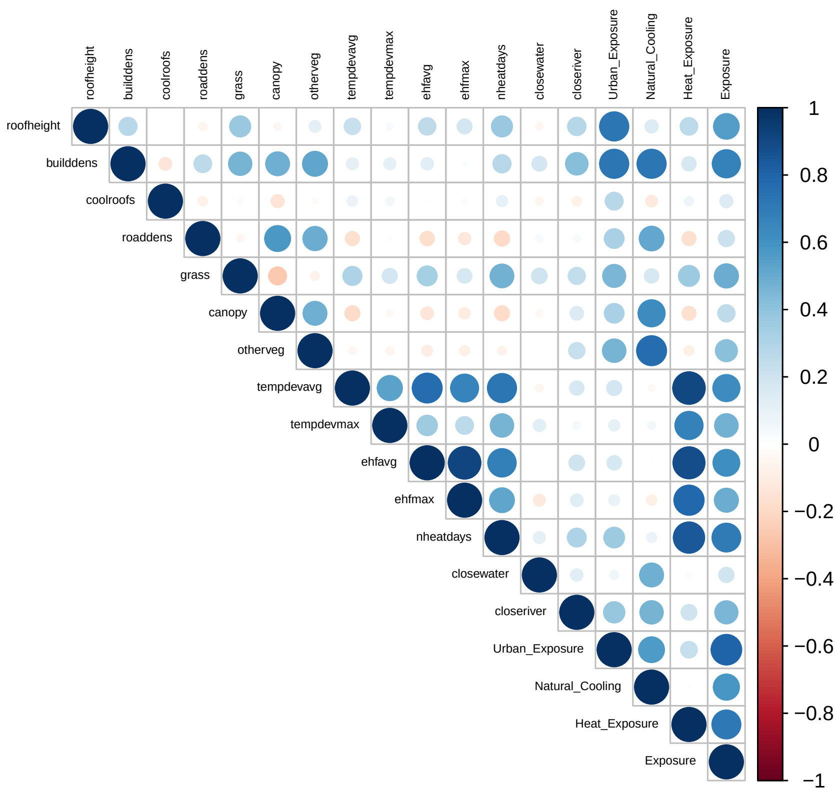


Figure D.1: Comparison of correlation between the 2016 (top) and 2021 (bottom) vulnerability indices and their sub-indices using the "omission" variation of the vulnerability index methodology.



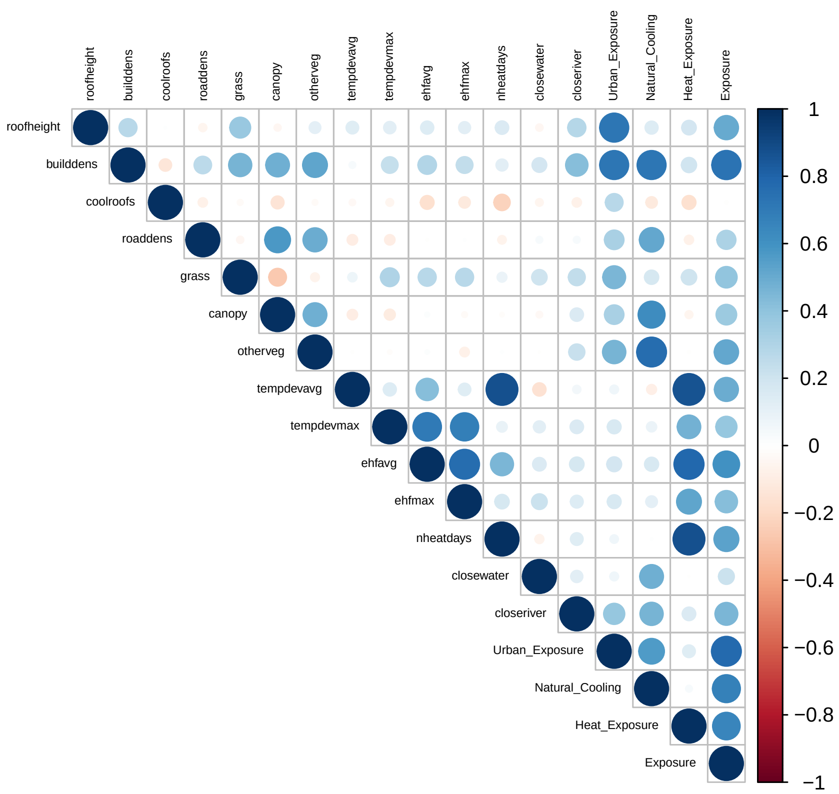
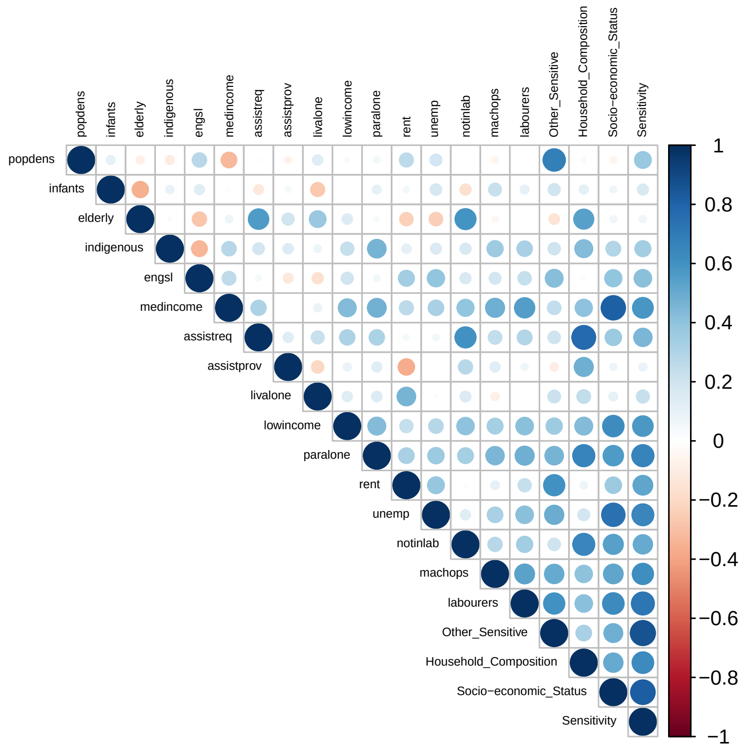


Figure D.2: Comparison of correlation between the 2016 (top) and 2021 (bottom) exposure sub-index, related sub-index themes, and underlying variables using the "omission" variation of the vulnerability index methodology.



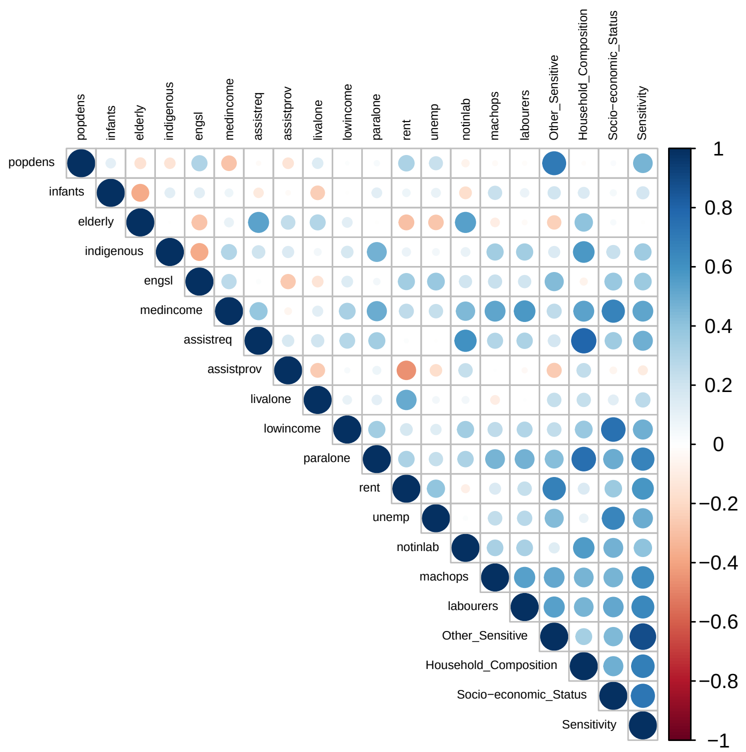


Figure D.3: Comparison of correlation between the 2016 (top) and 2021 (bottom) sensitivity sub-index, related sub-index themes, and underlying variables using the "omission" variation of the vulnerability index methodology.

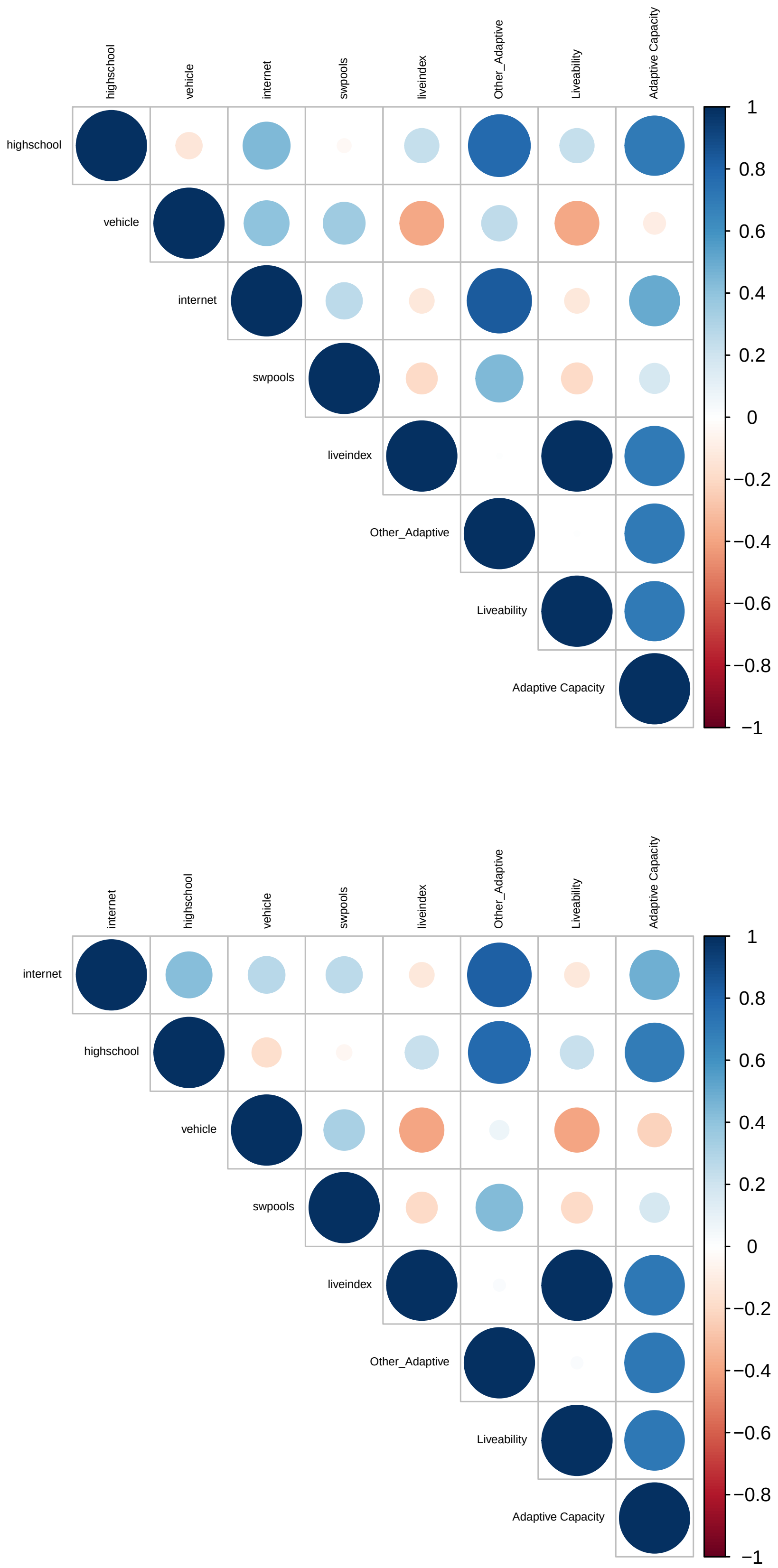


Figure D.4: Comparison of correlation between the 2016 (top) and 2021 (bottom) adaptive capacity sub-index, related sub-index themes, and underlying variables using the "omission" variation of the vulnerability index methodology.

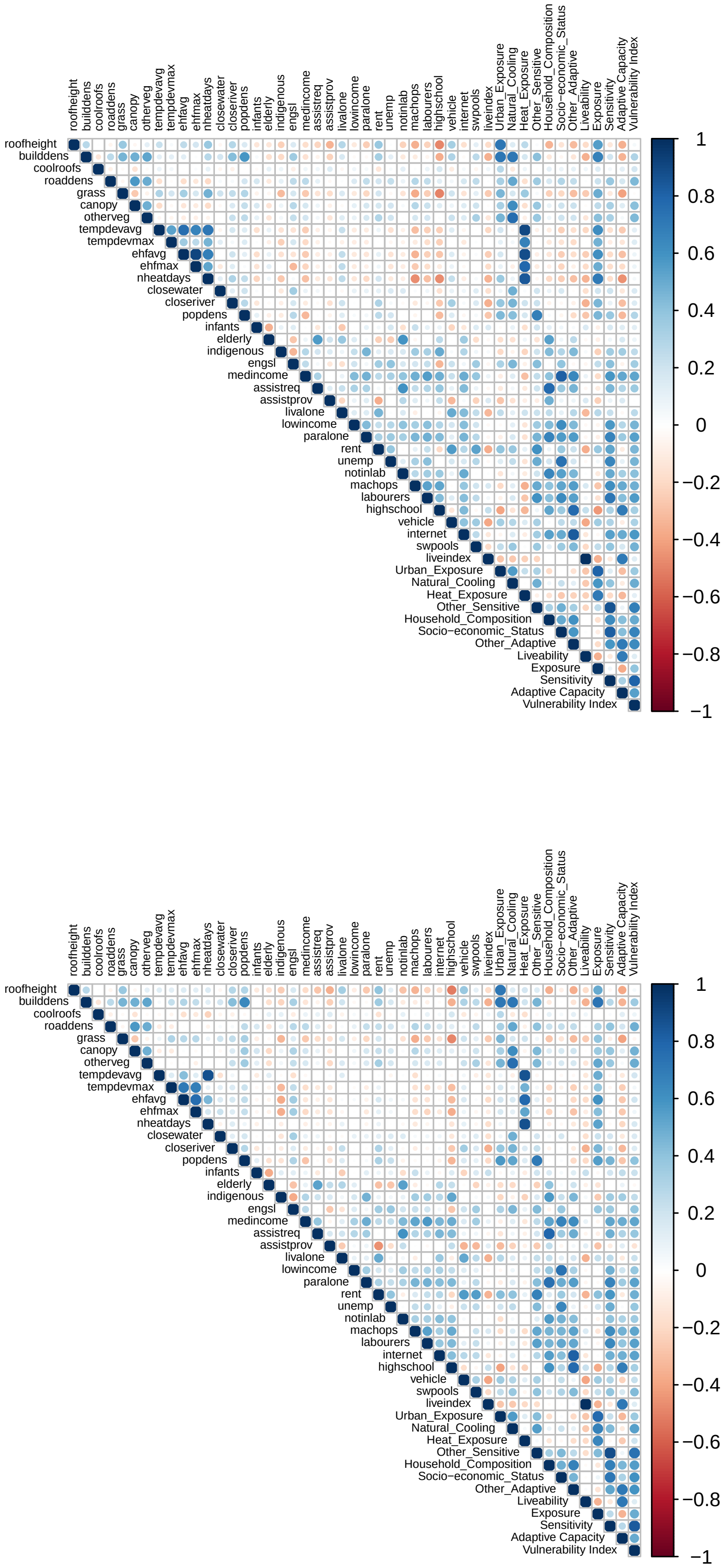


Figure D.5: Comparison of correlation between the 2016 (top) and 2021 (bottom) vulnerability indices and all related sub-indices, related sub-index themes and underlying variables using the "omission" variation of the vulnerability index methodology.