Assignment 2 Report

Elements of Data Processing 2021 S1

Group 64
Jiles Denham, John Li, Mitch O'Grady & Ziheng Tang

Introduction

The amount of capital that households have access to can impact a wide range of key metrics for an advanced and mature society. Such metrics include livability, inclusiveness, health, and sustainability (LIHS metrics). By providing a Universal Basic Income (UBI) to all, we can ensure that everyone has access to a sustenance level of resources.

In addition to improving our society, we must also be proactive in ensuring that Victoria is properly prepared for the risk of mass unemployment and extreme inequality with automation looming on the horizon.

In this observational study, we investigate possible links between income and factors related to LIHS metrics. We also explore the Centrelink benefit scheme, and its adequacy as a social welfare system. As there are no open data studies on universal basic income in Victoria, we will be instead using income as a means to investigate the effects of providing a UBI to Victorians.

Datasets

Datasets containing metrics related to liveability, inclusiveness, health and sustainability in Victoria were chosen from **ABS** via **AURIN,** and from the **AEDC** (Australian Early Development Census). Income data was also sourced from ABS. This input data and all output data was aggregated by Statistical Area Level 2 (SA2).

Australian government sources were chosen for their accuracy and consistent data formats, which aided in effective processing. Critically, the above ABS data tracked Socio-Economic Indexes for Areas (**SEIFA**), a dedicated attempt to measure socio-economic advantage and disadvantage. This metric is broadly defined by the ABS as "... **people's access to material and social resources, and their ability to participate in society**."

Data for average weekly rent by SA2 was obtained from datasets constructed from census data, along with data on the Consumer Price Index (CPI) provided by the **ABS** website. Values for Youth Allowance, Newstart (JobSeeker) and rent assistance were retrieved from Centrelink payment guides provided by Government Services Australia.

Data Wrangling and Analysis

Data Cleanup and Preprocessing

2011 and 2016 datasets on a wide range of domains, including income, were retrieved from the AURIN portal. Datasets were exported as csv files, keeping only the columns that were of interest to us and cleaned of NaNs and trailing white space. The files from AURIN had columns in machine readable format, requiring conversion to readable labels using the JSON metadata. For some datasets, repeated entries were grouped together making it easier to analyze; e.g. one location could have multiple entries for each parameter. This was solved by using the location as the key to a dictionary, and mapping related data to that key as a list. This reduced a 40,000 lined csv to roughly 300 lines.

For the AEDC data that was downloaded, each metric was its own file and within each file, there were multiple columns for each year. This data is converted into a similar format as the other datasets that we have obtained in order to combine them together. To do this conversion, first the years are split out from the column names. Then the dataset gets converted into a stacked format where all the indexing information is in the columns similar to a database table. This makes it easy to concatenate the information from the multiple files together and to reshape via pivoting to be the same layout as the other datasets.

The datasets that we have all include a column for the SA2 numerical code. As SA2 codes come in two types (five digit and nine digit codes), in order to merge our data by SA2 codes, we needed to convert all nine-digit codes to their corresponding five-digit code.

Analysis of Income Correlations

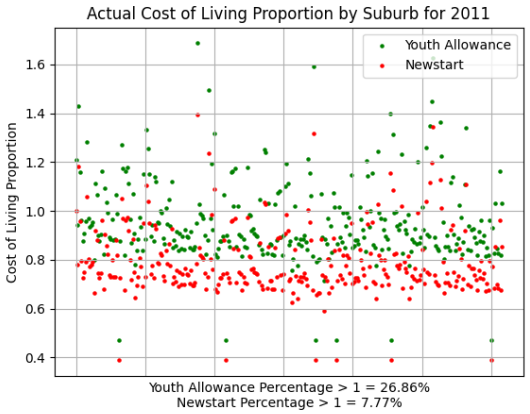
The preprocessed datasets are all linked together by SA2 Code. The final dataframes for each year that we run the analysis on has the SA2s as the index and metrics as the columns. The correlations between income and the individual metrics was computed using Pearson's Correlation method and highly correlated metrics were plotted as a scatter graph.

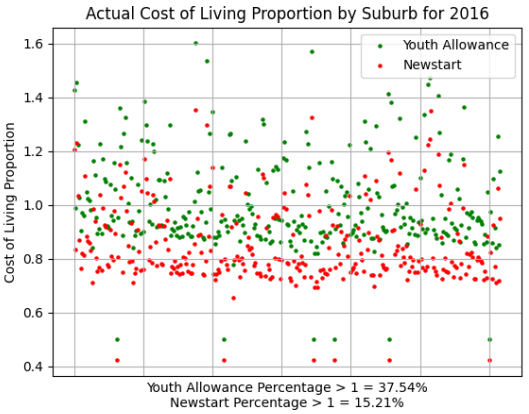
Analysis of Centrelink Benefits

For our analysis on Centrelink benefits, a cost of living (COL) proportion was calculated (expenses divided by welfare received). A proportion greater than 1 indicates that the payment for an individual in that year and SA2 is inadequate in covering their cost of living. The main challenge was obtaining a value for the average weekly rent per person, the largest expense for the vast majority of people receiving benefits. This was calculated from two datasets taken from census data provided by the **ABS**; number of bedrooms by dwelling type and weekly rent by dwelling type. By obtaining the weighted mean of bedrooms by dwelling type, dividing the weekly rent of each dwelling type by this mean, and then taking the weighted means of these weekly rent values for each SA2 and year, a figure for the average weekly rent per person was obtained.

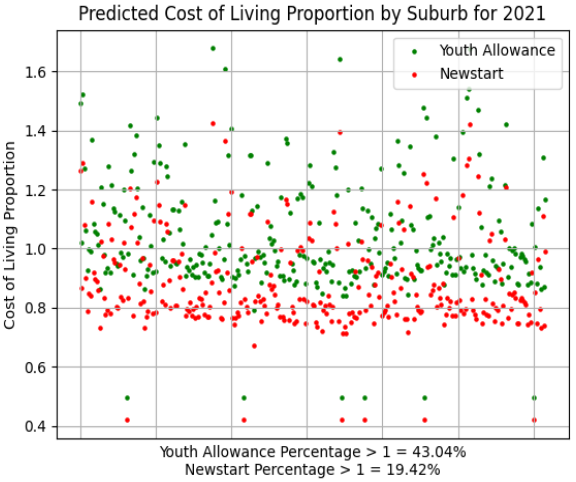
As there is no census data available for 2021, a predicted weekly rent value was calculated using quarterly consumer price index (CPI) data and the 2016 rent values. Regression was considered, but considered inappropriate as it would extrapolate beyond the scope of data observed.

No adequate data for living expenses on an individual basis was found. Therefore a frugal $75, $87.5, and $100 grocery bill and $35, $42.5 and $50 for miscellaneous expenses (transport, bills) was added to the weekly rent for 2011, 2016, and 2021 respectively, providing a value for weekly COL adjusted for inflation. Youth Allowance and Newstart (JobSeeker) payments were analysed, as these are the most common payments provided to renters that are not on the pension. Full rent assistance for each year was also added to the benefit amount. With this data, a COL proportion for both benefits was obtained for each SA2 and census year.





Year	Youth Allowance Amount (weekly)	Newstart Amount (weekly)
2011	$234.15	$283.1
2016	$260.135	$307.89
2021	$303.185	$357.335



From these plots it can be seen that despite increases in Youth Allowance, Newstart/Jobseeker and Austudy (same amount as Newstart) throughout each 5 year period, the percentage of SA2 areas that have a COL proportion greater than 1 increases. In particular, those who receive Youth Allowance (unemployed or full time studying under 22) have much more financial pressure put upon them despite having less skills for employment and full time study/apprenticeship commitments. Additionally it can be seen that the vast majority of SA2 areas leave very little expendable income from Youth Allowance (COL proportion > 0.9), with financial pressure alleviated only slightly for those on Newstart/Jobseeker.

Key Results

Strong, meaningful correlations with an increase in household income were found, including an:

- Increase in Relative Socio-economic Advantage Index (*r = 0.88*, ABS)
- Decrease in childhood developmental vulnerability (AEDC) for:
    - Language and cognitive function (*r = 0.49*)
    - Physical health and wellbeing (*r = 0.43*)
    - Communication skills and general knowledge (*r = 0.41)*
    - Emotional maturity (*r = 0.38*)

Demonstrated increase over time of geographical distribution where individuals who are solely reliant on Centrelink suffer from an income deficit

- Implies inadequacy in current social welfare to cover a basic standard of living
- Most financially pressured group consists of recipients under 22, arguably the most vulnerable group due to lack of employment experience and tertiary qualifications
- High COL proportion, and thus minimal expendable income removes ability to participate in society and higher susceptibility to our found correlations

Impact

Establishing such strong positive correlations between income and relative well-being indicates that the solution to many social problems may be to simply provide money to those who do not have enough of it to meet their needs.

Reducing developmental vulnerability in children has profound lifelong benefits. Intergenerational developmental vulnerability may also be reduced or even ended entirely if one or more generations are given enough resources to overcome their income-correlated limitations.

These results highlight the potentially immense value of a Basic Income and increased social welfare to Victorian social outcomes.

Limitations

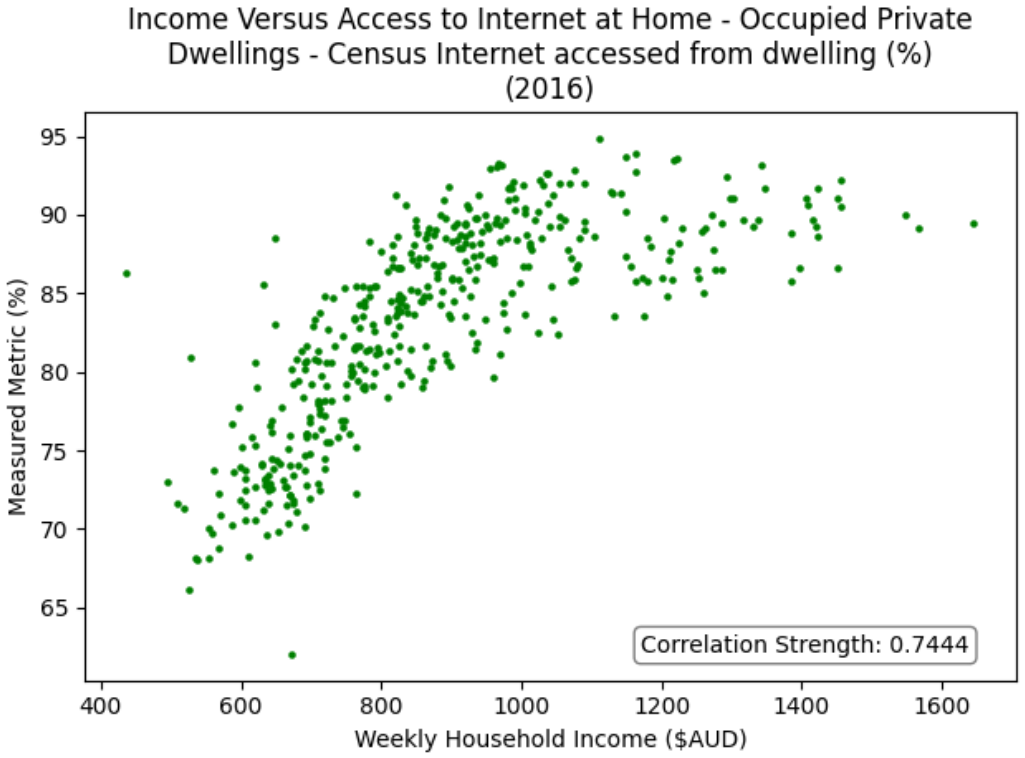
Our analysis is an observational study that uses ABS census data. This relies on individual citizens to supply accurate and honest responses to provided questions. Experimental studies could yield improvements. The real world also includes complexities not present in a proper scientific study where the participants will be chosen such that the only difference between groups is income. Because of these issues, we cannot identify a causal relationship between correlations we have uncovered.

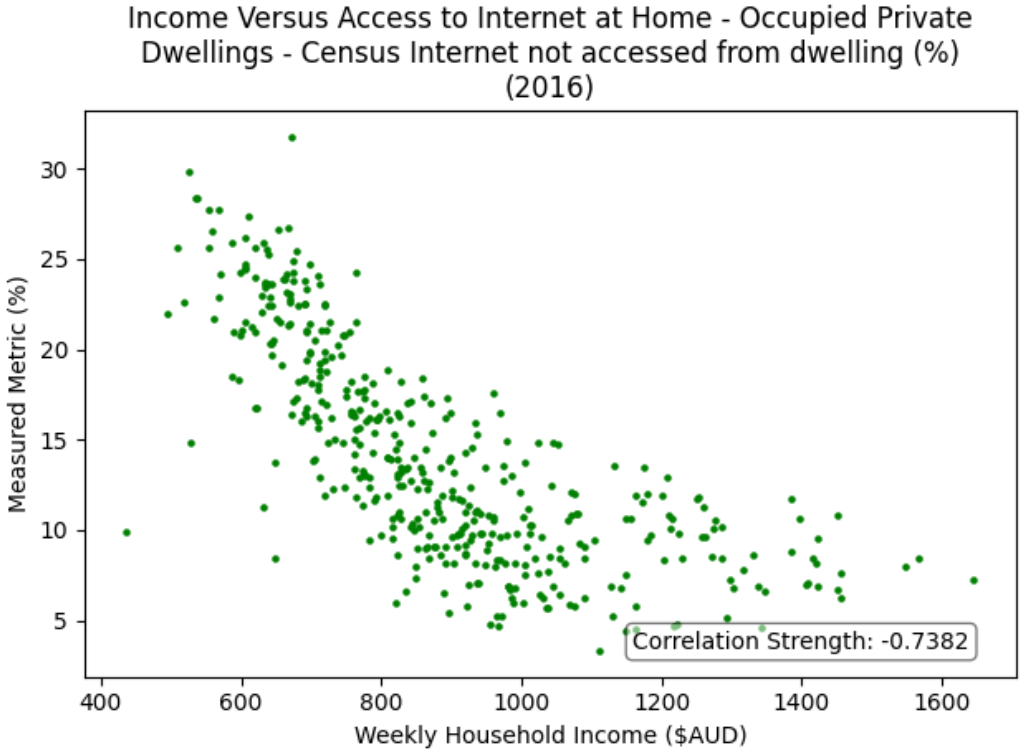
The data that we have is also aggregated by SA2 spatial regions. Although this may provide privacy for the census takers, it may also be an obstacle in establishing correlations, as correlations on an individual level may not be apparent when aggregated.

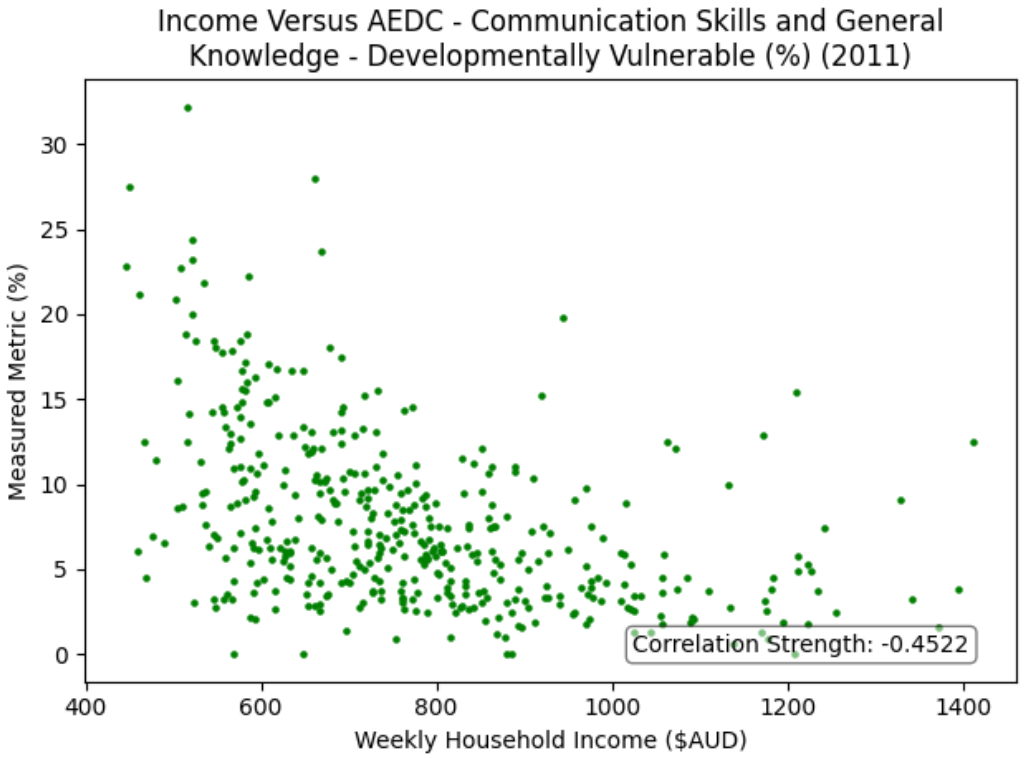
Regarding our Centrelink benefit analysis, our values for weekly average rent were inferred from the assumption that 1 bedroom meant 1 tenant. Therefore, these values don't account for couples or others sharing a room, which would decrease the weekly rent per tenant. Since there was no data for individual cost of living expenses, we had to extrapolate our values based on anecdotal data.

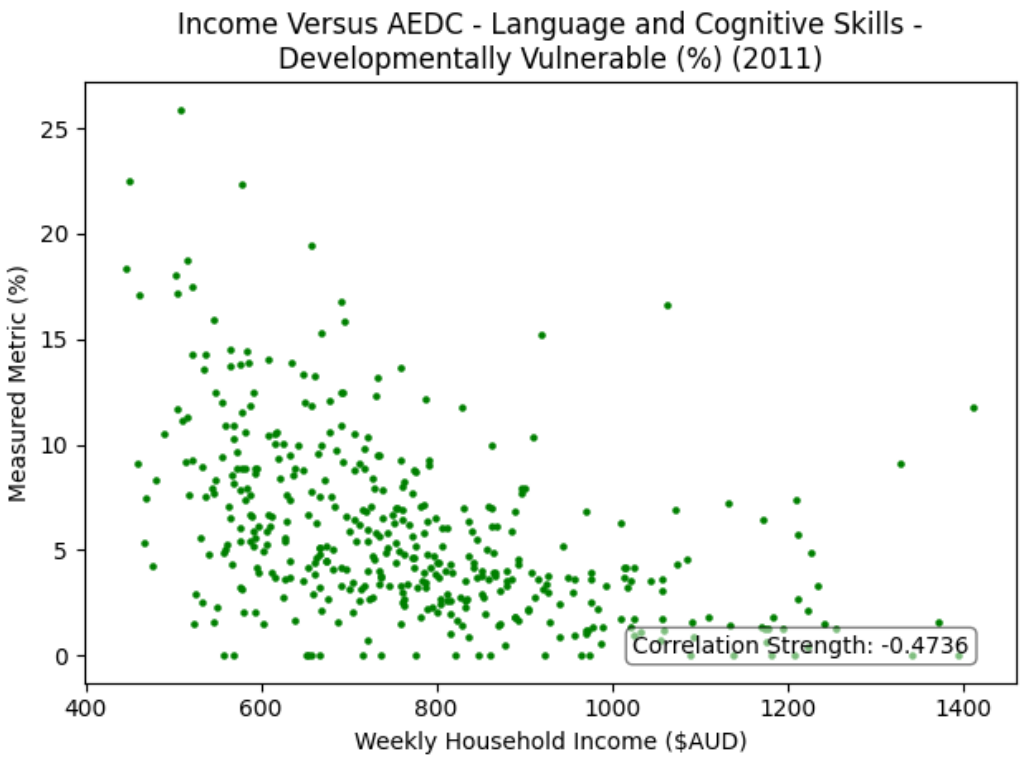
Appendix

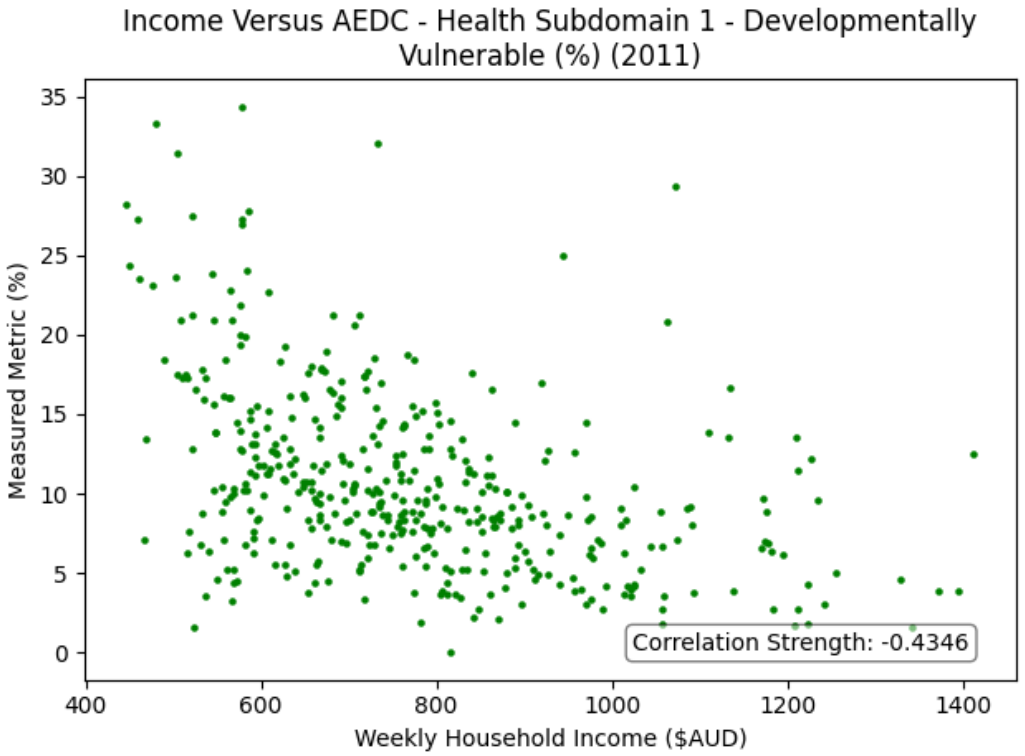
Correlation Graphs

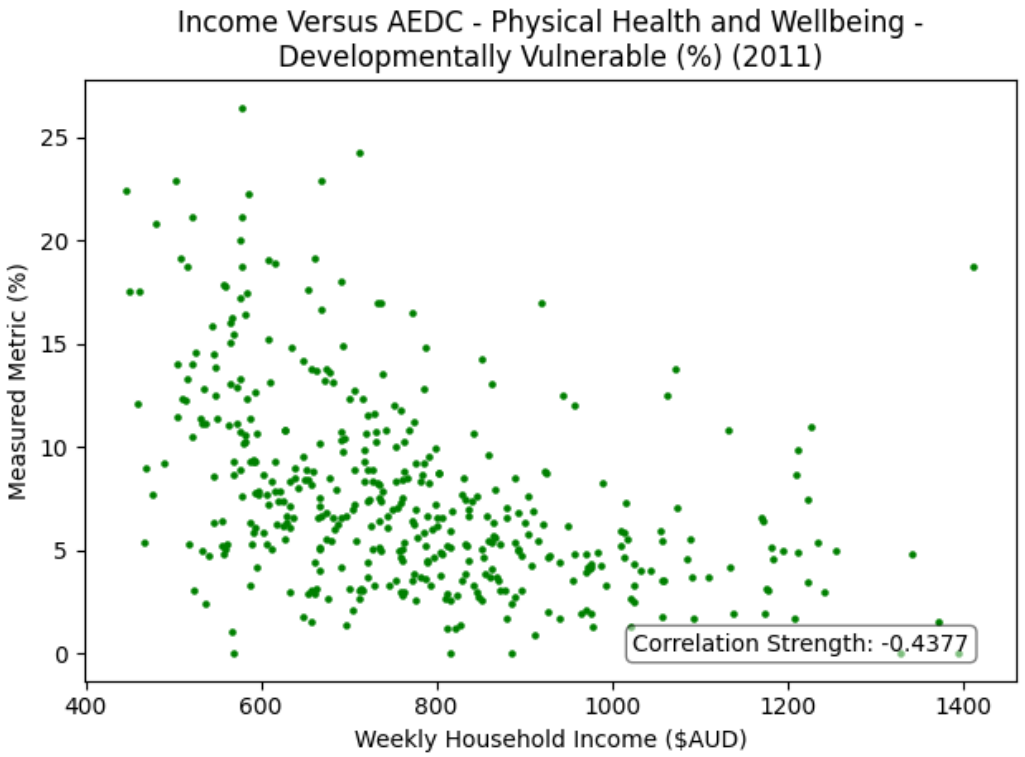


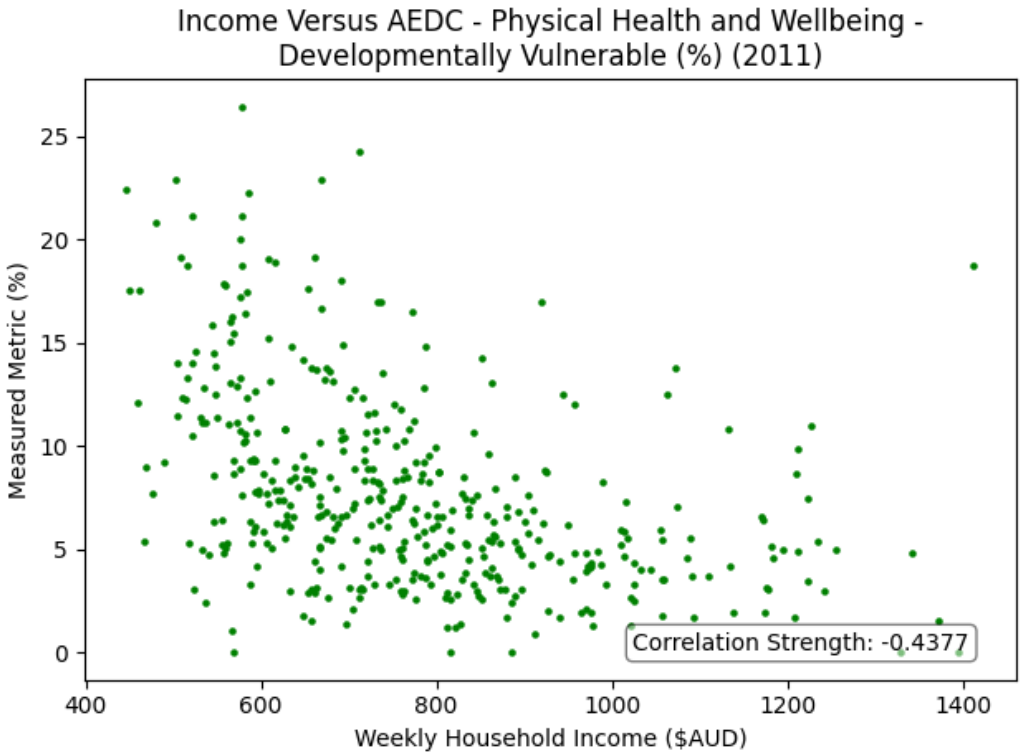


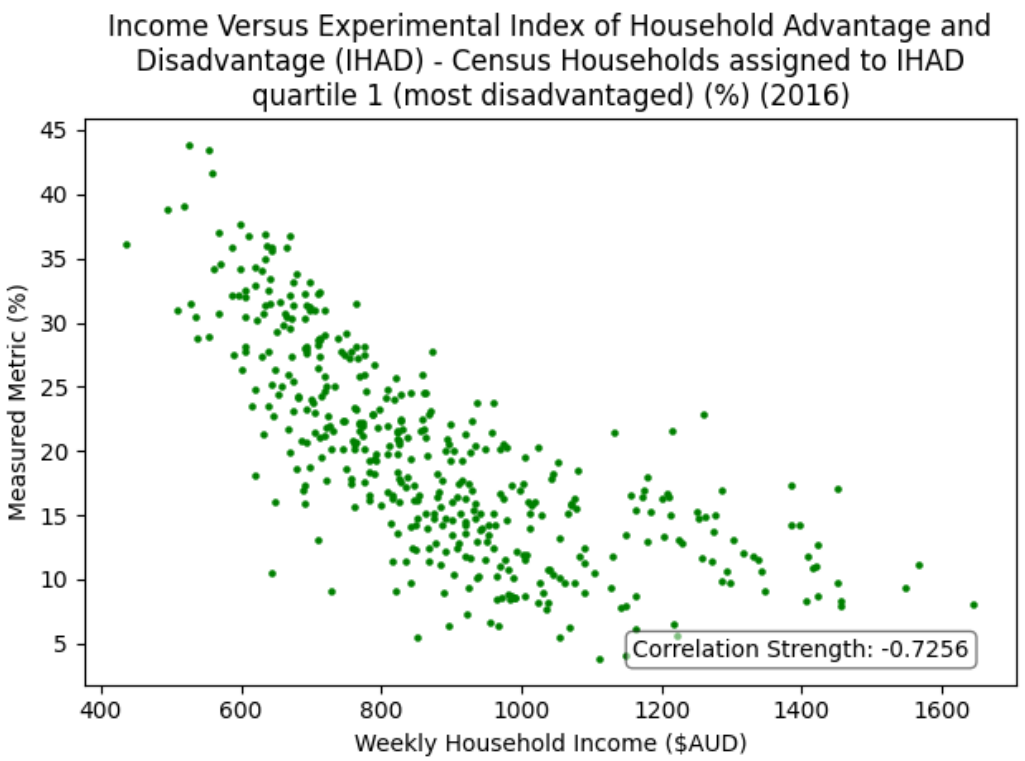

Income Versus AEDC - Communication Skills and General Knowledge - Developmentally Vulnerable (%) (2011)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.4522

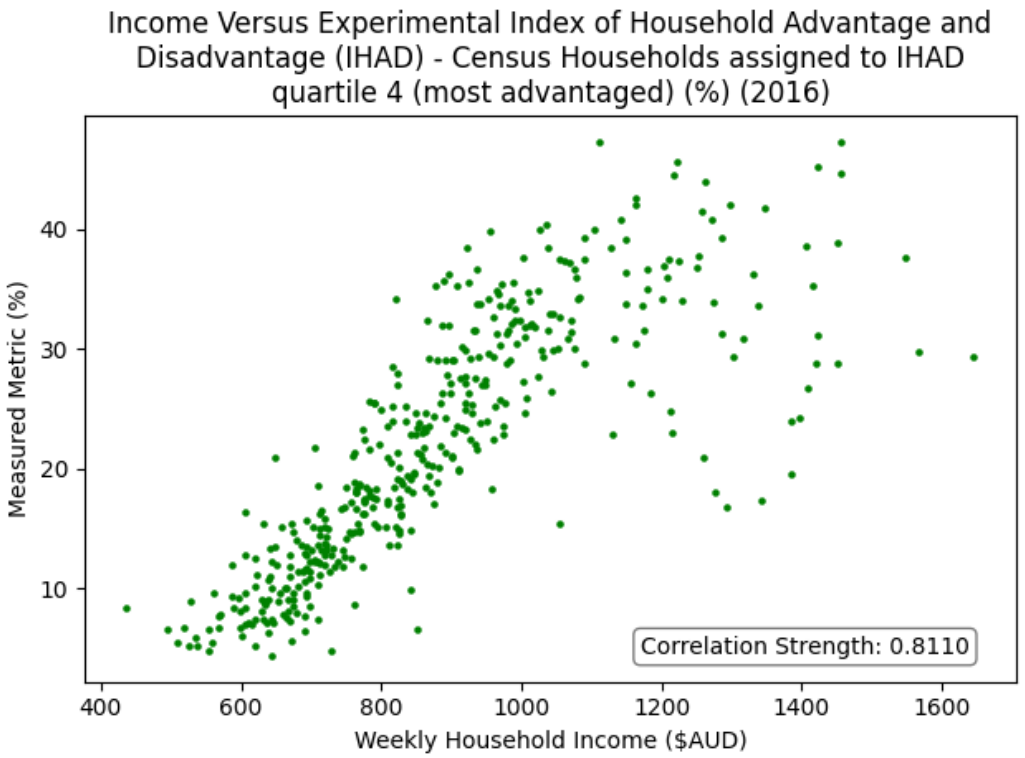

Income Versus AEDC - Language and Cognitive Skills - Developmentally Vulnerable (%) (2011)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.4736

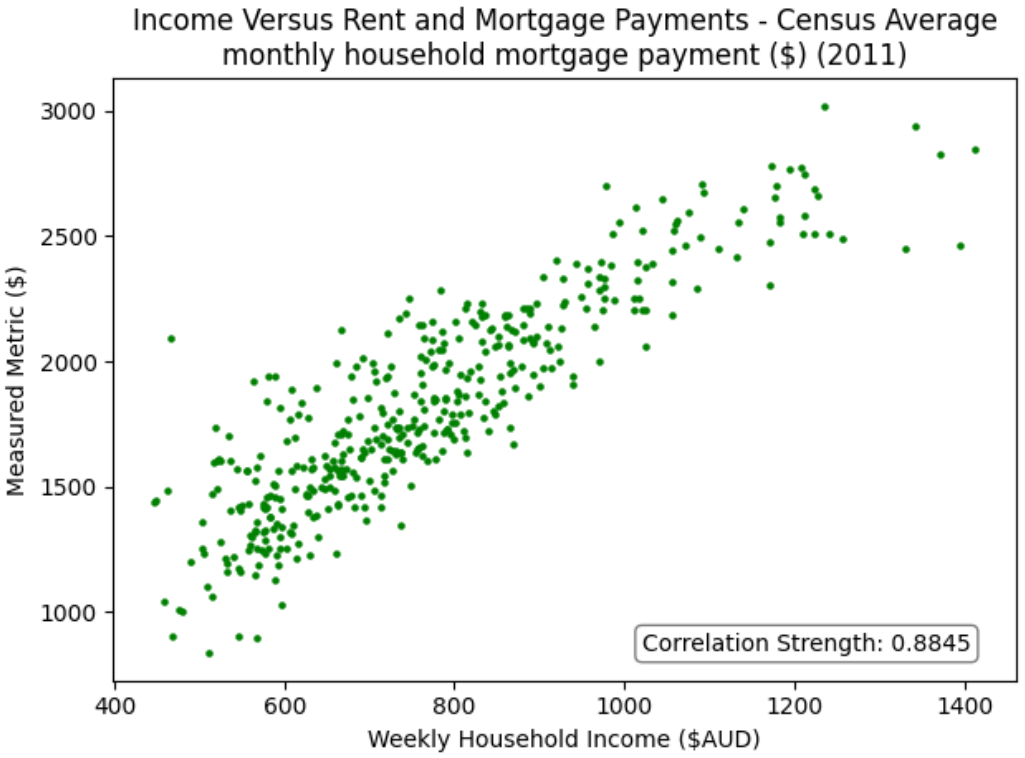

Income Versus AEDC - Health Subdomain 1 - Developmentally Vulnerable (%) (2011)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.4346
0
5
10
15
20
25
30
35
400
600
800
1000
1200
1400

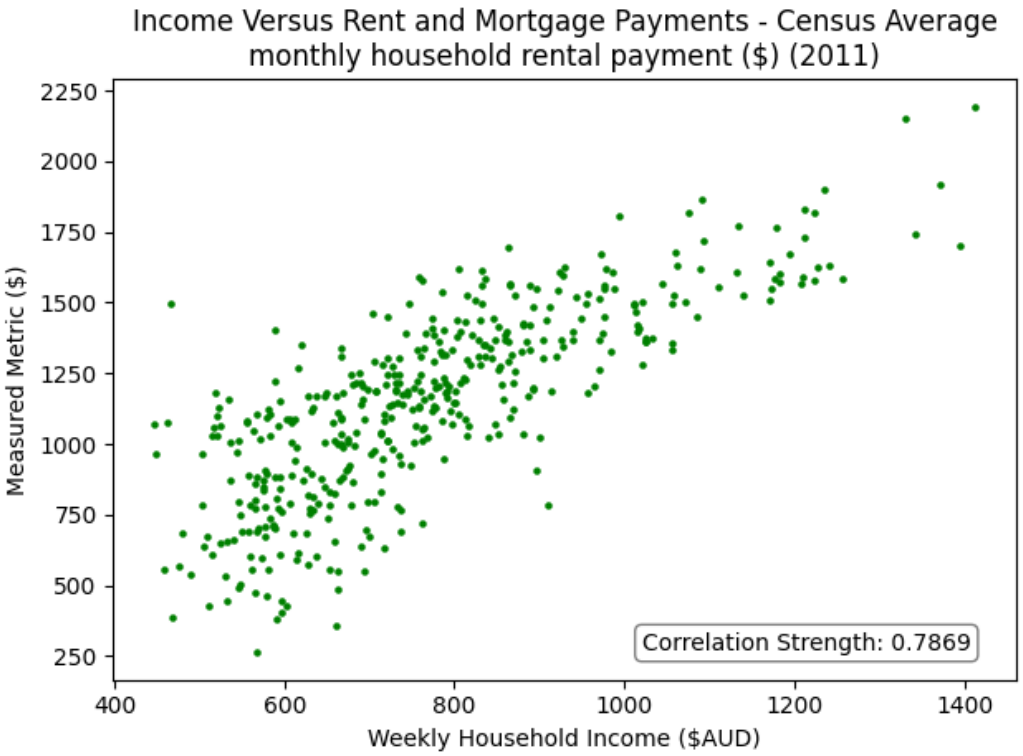

Income Versus AEDC - Physical Health and Wellbeing - Developmentally Vulnerable (%) (2011)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.4377
0
5
10
15
20
25
400
600
800
1000
1200
1400

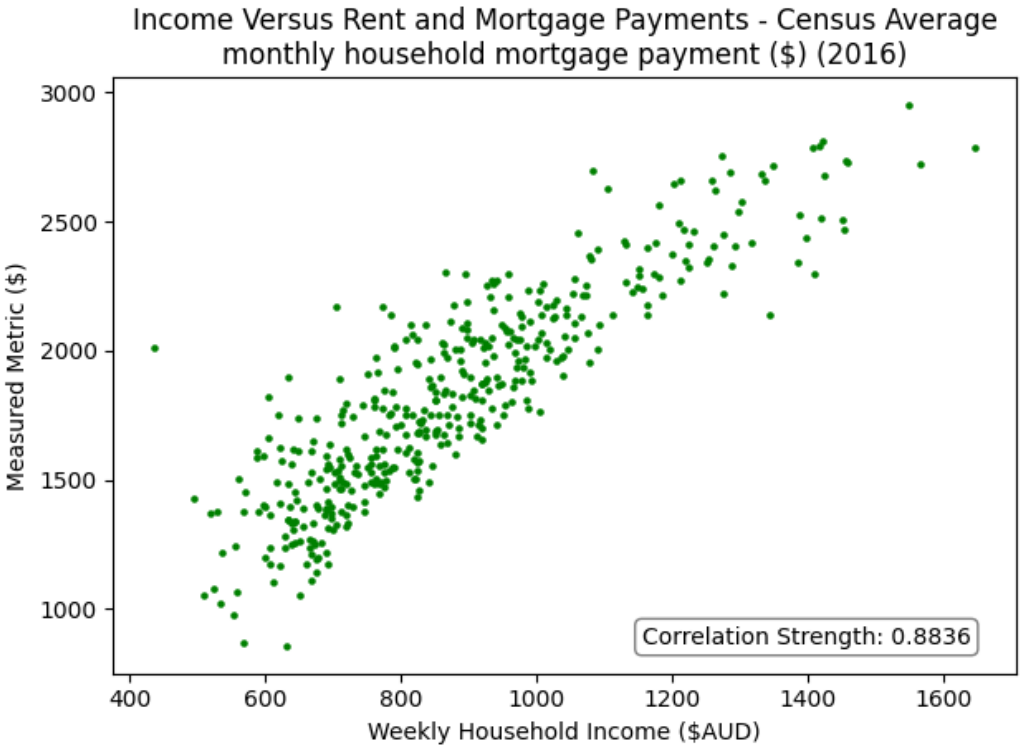

Income Versus AEDC - Physical Health and Wellbeing - Developmentally Vulnerable (%) (2011)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.4377
0
5
10
15
20
25
400
600
800
1000
1200
1400

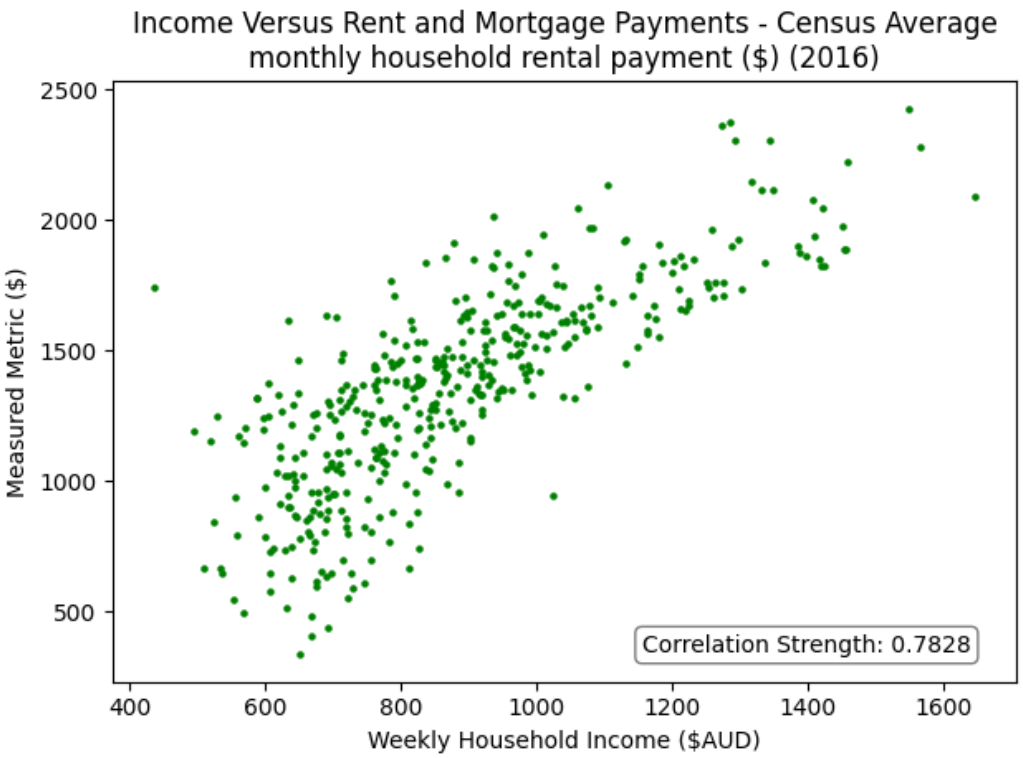

Income Versus Experimental Index of Household Advantage and Disadvantage (IHAD) - Census Households assigned to IHAD quartile 1 (most disadvantaged) (%) (2016)
Measured Metric (%)
Weekly Household Income ($AUD)
Correlation Strength: -0.7256
5
10
15
20
25
30
35
40
45
400
600
800
1000
1200
1400
1600


Income Versus Experimental Index of Household Advantage and Disadvantage (IHAD) - Census Households assigned to IHAD quartile 4 (most advantaged) (%) (2016)
Measured Metric (%)
10
20
30
40
400
600
800
1000
1200
1400
1600
Weekly Household Income ($AUD)
Correlation Strength: 0.8110


Income Versus Rent and Mortgage Payments - Census Average monthly household mortgage payment ($) (2011)
Measured Metric ($)
1000
1500
2000
2500
3000
400
600
800
1000
1200
1400
Weekly Household Income ($AUD)
Correlation Strength: 0.8845


Income Versus Rent and Mortgage Payments - Census Average monthly household rental payment ($) (2011)
Measured Metric ($)
Weekly Household Income ($AUD)
Correlation Strength: 0.7869
2250
2000
1750
1500
1250
1000
750
500
250
400
600
800
1000
1200
1400


Income Versus Rent and Mortgage Payments - Census Average monthly household mortgage payment ($) (2016)
Measured Metric ($)
Weekly Household Income ($AUD)
Correlation Strength: 0.8836
3000
2500
2000
1500
1000
400
600
800
1000
1200
1400
1600



For more plots, check out our github page.