DATA201 Group Project 2021

*Comparing Household Income and Crime Rates for New Zealand by Region*

**Group name: DHRL**

Ruben Castaing: 26996541

Harrison Allen: 97755656

Xiaodan Liang: 73190687

Junwei Liang: 91925811

Data Sources:

The data sourced for this project was retrieved from the Statistics New Zealand NZ.Stat table viewer, via the link http://nzdotstat.stats.govt.nz/wbos/index.aspx

The first dataset generated was **Annual Recorded Offences for the latest Calendar Years (ANZSOC)**.

This dataset contains a breakdown of reported offences across New Zealand by both type of offence (for example Homicide) and by policing district (for example Canterbury District). The dataset spans from 1994 to 2014.

Data in this dataset is derived from the New Zealand Police National Intelligence Application (NIA). The dataset defines the number of recorded offences, where a recorded offence is defined as a crime that has come to the attention of police.

This data set was selected as it gives a large and accurate nominal record of crime across New Zealand, and has a large enough time span to develop trends.

The second dataset was also generated using the NZ.Stat table viewer. This dataset **Household income by region, household type, and source of household income.** This dataset contains both the average and median incomes for households across New Zealand, broken down by regional council areas. The dataset spans 1996 to 2021.

The income data in this dataset is derived from question in the household labour force survey conducted by Stats NZ. This dataset is restricted to households where at least one member is in within the 18 to 64 year age range. Weekly household income is defined as the sum of weekly income of all people in the household from wages, salaries, self-employment and government transfers. The average is the total weekly household income, divided by the number of households.

This data set was selected as it provided a measure of wealth at a more individual level than other measures, for example .

Intended target of the data:

The intended target of these datasets was to investigate how levels of income may impact crime rates, and whether particular types of crime are more impacted than others.

Data Wrangling:

Initially, the income dataset required hand-labelling to allow for consistent naming of the regions within the dataset. Both CSV files were loaded into R, and the data frames were converted from a wide format to a long format. The function *reshape\_df* shown in the Jupyter notebook was developed to process the income data, by firstly converting the frame to long format, and renaming the regions to match the policing districts. This process required some districts to be joined to match the policing districts. For example, the Otago and Southland regional councils are covered by the Southern District policing region.

Once the income data frame was converted, the crime data frame was joined by matching year and region together. This allowed for a final data frame to be constructed that mapped average and median household income to numbers of offenses for each region by year.

Wrangling difficulties:

The data sets had different regions

Building the websites ended up being quite time consuming.

Achievements:

[Working GitLab.](https://eng-git.canterbury.ac.nz/jli328/201-group-project/-/blob/master)

Gitlab manage files between computers.

Found expected insights.

Website

Our website, saved on GitLab, communicates the findings we have discovered throughout the project.

A shiny app was developed showing interactive graphs between crime and income. But was difficult to integrate with CSS and hosting tools. The Shiny app was built in Jupiter Lab but converting the code into HTML requires set up on R studio.

Frustratingly, websites need a domain name and hosting which is difficult and costly to do. Initial