



TE TAI ŌHANGA  
THE TREASURY

# Wellbeing in Aotearoa New Zealand: A Population Segmentation Analysis

Background paper to Te Tai Waiora:  
Wellbeing in Aotearoa New Zealand 2022

October 2022

**BACKGROUND PAPER TO  
TE TAI WAIORA: WELLBEING  
IN AOTEAROA NEW ZEALAND**

Wellbeing in Aotearoa New Zealand: A Population  
Segmentation Analysis

**MONTH/YEAR**

October 2022

**AUTHORS**

Sarah Crichton  
Principal Advisor  
Analytics and Insights  
The Treasury  
1 The Terrace, Wellington 6011  
Email: [sarah.crichton@treasury.govt.nz](mailto:sarah.crichton@treasury.govt.nz)  
Telephone: +64 4 890 7234

Hien Nguyen  
Modelling Analyst  
Analytics and Insights  
The Treasury  
1 The Terrace, Wellington 6011  
Email: [hien.nguyen@treasury.govt.nz](mailto:hien.nguyen@treasury.govt.nz)  
Telephone: +64 4 495 2843

**URL**

Treasury website at October 2022:  
<https://www.treasury.govt.nz/publications/background/wellbeing-aotearoa-new-zealand-population-segmentation-analysis>

**ACKNOWLEDGEMENTS**

This paper has benefited from input from many colleagues in the Treasury, including Tod Wright, Tim Hughes, Patrick Nolan and Diana Cook. Thanks to Keith McLeod, Simon Anastasiadis and Penny Mok for their comments on a draft of the paper. We adapted code written by our former colleagues, Keith McLeod, Steven Johnston and Simon Brown to analyse the GSS data. Any remaining errors are the sole responsibility of the authors.

**THE TREASURY**

The Treasury  
PO Box 3724  
Wellington 6008  
NEW ZEALAND  
Email [information@treasury.govt.nz](mailto:information@treasury.govt.nz)  
Telephone 64-4-472 2733  
Website [www.treasury.govt.nz](http://www.treasury.govt.nz)

## DISCLAIMER

The views, opinions, findings, and conclusions or recommendations expressed in this report are strictly those of the author. They do not necessarily reflect the views of the New Zealand Treasury, Stats NZ or the New Zealand Government. The New Zealand Treasury and the New Zealand Government take no responsibility for any errors or omissions in, or for the correctness of, the information contained in this Background Paper.

Access to the data used in this study was provided by Stats NZ in accordance with security and confidentiality provisions of the Data and Statistics Act 2022. The results presented in this study are the work of the author, not Stats NZ or individual data suppliers. These results are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ. For more information about the IDI please visit <https://www.stats.govt.nz/integrated-data/>

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes. Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.



# Executive Summary

This paper seeks to understand which factors are related to differences in subjective wellbeing within the New Zealand population aged 15 years and above. It uses data from the New Zealand General Social Survey (GSS) and regression tree analysis to identify the factors most strongly related to differences in subjective wellbeing (SWB). The tree analysis divides the population into groups or segments, whereby people in the same segment share a similar level of SWB and the factors most strongly related to SWB. The analysis shows how combinations of factors explain differences in SWB within the population and provides a person-centric view of wellbeing across multiple dimensions of wellbeing

We use self-reported life satisfaction (measured on a scale of 0-10) as a measure of subjective wellbeing and as the dependent variable in a regression tree analysis. We include the characteristics used to construct the wellbeing domains and sub-domains derived in earlier work on multi-dimensional wellbeing by McLeod (2018) based on the Treasury's Living Standards Framework. We also include additional characteristics available in the GSS and from linked administrative data. Having identified the population segments, we describe their socio-demographic and other characteristics. Owing to some significant changes to the health questions collected between 2016 and 2018, we include separate results for GSS 2018 and GSS 2014-2016.

Tree-based methods use a series of splitting rules to split the population into groups. The total population is first split into two groups (or *branches*) using the variable that most strongly differentiates subjective wellbeing in the population. Each branch is subsequently split into two, these are split again and so on. The tree-construction process continues until further splits do not explain significantly more variation in subjective wellbeing or when a minimum group (or *leaf*) size is reached. The final tree is a partition of the population into several groups or segments.

A key finding from this analysis is that mental health, having enough income to meet everyday needs, and trust in institutions or trust in other people, are the characteristics most strongly related to different levels of subjective wellbeing in the population aged 15 year or above. While our results are somewhat sensitive to the survey year, owing to changes to the health questions collected, we find that of all the characteristics we consider, mental health consistently appears at level 1 in the trees and is the characteristic that is most strongly related to different levels of subjective wellbeing in the population.

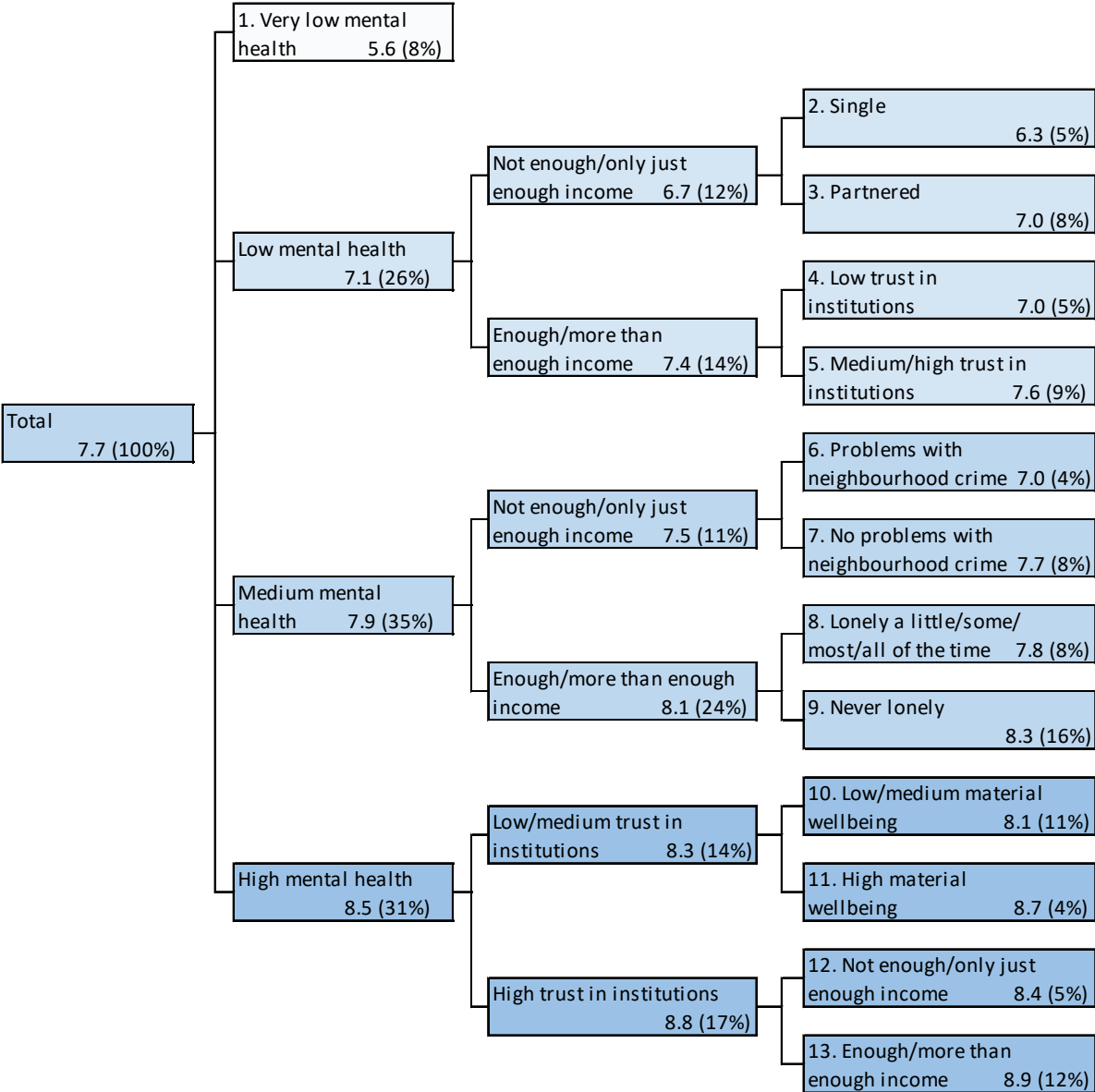
In our preferred tree for GSS 2018 (shown in the figure below), having enough income to meet every day needs and trust in institutions appear at level 2 in the tree, while partnership status, trust in institutions, problems with neighbourhood crime, loneliness, material wellbeing and having enough income appear at level 3. In our preferred tree for GSS 2014-2016, trust in people, trust in institutions and having enough income to meet every day needs appear at level 2, while loneliness, trust in institutions, having enough income and age appear at level 3<sup>1</sup>. We find that at levels 2 and 3 in the trees there is usually at least one other variable that could be substituted for the selected variable with very little or no loss of explanatory power.

---

1 Our findings align with previous studies that used regression analysis to examine the correlates of subjective wellbeing and found that subjective measures (including mental health and having enough income to meet everyday needs and necessities) are more highly correlated with life satisfaction than objective or demographic characteristics, including household income and education.

In our preferred tree for 2018, we segment the population aged 15 years and above into 13 segments, where average subjective wellbeing varies from 5.6 in segment 1 to 8.9 in segment 13, compared to an average subjective wellbeing of 7.7 in the total population overall.

**Regression Tree (GSS 2018)**



The population segment with the lowest average subjective wellbeing (5.6) is the 8% who have very low mental health (segment 1). This segment includes relatively more people who have a disability (22% compared to 8% in the population overall), females (60% compared to 51% in the population overall) and people who have received main benefits during the last year (17% compared to 9% in the population overall). The age, ethnicity and neighbourhood deprivation profile of the segment is similar to the population overall.

The segment with the next lowest average subjective wellbeing (6.3) is the 5% who have low mental health, who do not have enough income or who have only just enough income to meet their everyday needs and are single (segment 2). This segment includes relatively more Māori (28% compared to 13% in the population overall), females (64%), people aged 15 to 34 years (44% compared to 35%), people with a disability (15% compared to 8%), people who received main benefits during the last year (31% compared to 9%), people renting (58% compared to 34%) and people who live in the most deprived neighbourhoods (31% compared to 18%).

The segment with the highest average subjective wellbeing (8.9) is the 12% who have high mental health, high trust in institutions and have enough income or more than enough income to meet their everyday needs (segment 13). This segment includes relatively more males (56%), people aged 65 years or older (28% compared to 18% in the population overall) and Asian people (20% compared to 15% in the population overall).

While there are some differences in the distribution of socio-demographic characteristics across population segments, particularly for those segments with relatively low or high levels of subjective wellbeing, in general the differences are not very large. While disadvantaged groups are over-represented in the lower wellbeing segments, many are represented in the higher wellbeing segments as well. For example, while 34% of sole parents are in the two segments with lowest average subjective wellbeing (compared to 12% of the population overall), 13% are in the two segments with highest average subjective wellbeing (compared to 17% of the population overall).

By using a tree rather than a clustering method to segment the population, we have chosen to prioritise subjective wellbeing over the other wellbeing domains and create segments where individuals within the same segment share the characteristics that are most strongly related to different levels of subjective wellbeing in the population. A clustering method would have produced segments that are much less well differentiated on subjective wellbeing and more differentiated on other wellbeing domains.

Regression analysis (as opposed to regression tree analysis) has been used in several previous New Zealand studies to identify the factors that are related to (or correlated with) subjective wellbeing. We compared the two methods using GSS 2018 and found the results were consistent, although the relative importance of some variables in the models differed, reflecting the existence of interactions between some factors for particular population sub-groups. Our regression results for GSS 2018 largely confirm the results obtained in previous studies, with the degree of correlation between subjective wellbeing and other wellbeing domains varying considerably. The conclusions about the relative explanatory power of different factors and wellbeing domains varies across studies to some extent, reflecting differences in the choice of domains included and how they were constructed.





# Contents

- Executive Summary ..... i
- 1 Introduction .....2
  - 1.1 Summary of the literature .....2
- 2 Methodology and data .....7
  - 2.1 Methodology .....7
  - 2.2 Data .....9
- 3 Results .....12
  - 3.1 Regression tree analysis – GSS 2014-2016 .....12
  - 3.2 Regression tree analysis – GSS 2018 .....15
  - 3.3 Regression analysis – GSS 2018 .....21
- 4 Conclusion .....24
- References .....26
- Appendix A: Summary of relevant New Zealand literature .....27
- Appendix B: Variable descriptions .....28
- Appendix C: Alternative regression trees (GSS 2014-2016) .....35
- Appendix D: Alternative regression tree (GSS 2018) .....38
- Appendix E: Segment profiles (GSS 2018) .....39
- Appendix F: Regression results (GSS 2018) .....43

# 1 Introduction

This paper focuses on splitting the population into groups (or segments) whereby the population in each segment share the characteristics most strongly related to subjective wellbeing and a similar level of subjective well-being. Our approach is strongly influenced by work done by the Netherlands Institute for Social Research (Bijl et al., 2017) which used regression tree analysis to segment the population based on a life situation index. This study also builds on previous research in New Zealand that examined the correlates of subjective wellbeing (life satisfaction), including research by McLeod (2018), which used the Treasury's Living Standards Framework (LSF)<sup>2</sup> wellbeing domains and data from the GSS 2014-2016 to look at multi-dimensional wellbeing. We use the wellbeing domain and sub-domain variables from the McLeod study and add some additional variables from GSS and from administrative sources.

## 1.1 Summary of the literature

We are not aware of any previous New Zealand studies that have used regression tree analysis to understand the correlates of subjective wellbeing (SWB) and segment the population but there have been several studies which have used logit or linear regression to identify the correlates (sometimes referred to as determinants) of SWB that are highly relevant to our study. Appendix A provides an overview of the New Zealand literature we discuss in more detail below.

Most studies of SWB in Aotearoa New Zealand have used GSS data collected by Stats NZ. Although the question of which indicator/s best represent SWB is still debated, self-reported life satisfaction is a widely accepted measure in both the international and New Zealand literature. Dolan & White (2007) provides a comprehensive review of this indicator and its usage in the key fields, such as economics and psychology. In this paper we use the term subjective wellbeing refers to life-satisfaction.

Brown et al. (2012) used GSS 2008 and regression analysis to examine the determinants of life satisfaction. They found a significant positive relationship between life satisfaction and the 12-item Short Form Health Survey (SF-12) based measures of mental health and physical health, household equivalised gross income, home ownership, availability of help in a crisis, voluntary work, age (a U-shaped effect), female, being partnered and having a degree-level qualification, and a significant negative relationship with unemployment, feeling isolated, difficulty in expressing one's identity, feeling they do not belong to Aotearoa New Zealand and dissatisfaction with the quality of local facilities. They found no significant relationship between life satisfaction and ethnicity, having a child, having enough contact with family, having school or vocational qualifications (compared to having no qualifications), and having experienced a violent crime. They concluded that mental health, income, and social relationships had the strongest relationship with life satisfaction. R-squared (which reflects the degree of the variation explained by the covariates included in the regression) was not reported.

---

2 The Treasury's Living Standards Framework (LSF) is a key analytical and policy framework that focuses on the dimensions of life that matter most to New Zealanders and shape the quality of their lives. It was revised in 2021 to better reflect children's wellbeing and culture. The LSF is based on the Organisation for Economic Co-operation and Development (OECD)'s internationally recognised approach to measuring wellbeing. It has three levels: our individual and collective wellbeing, our institutions and governance and the wealth of Aotearoa New Zealand.

McLeod (2018) used GSS 2014-2016 to examine multi-dimensional wellbeing in Aotearoa New Zealand as expressed in the Treasury's LSF. He constructed eight wellbeing domains and sub-domains using measures available in the GSS and defined three levels of wellbeing (low, medium and high) for each domain. He considered the number of domains with low wellbeing versus the number of domains with high wellbeing and defined a cross-domain wellbeing score by subtracting the number of domains with low wellbeing from the number with high wellbeing (ie, to get a score from -8 to 8). A set of ordinary least squares (OLS) regression results showed that having low/high wellbeing in each domain was significantly associated with lower/higher life satisfaction. A regression analysis that included all eight domains, as well as sex, age group, ethnicity and family type showed that the explanatory power of each wellbeing domain varied considerably. Health<sup>3</sup> was the most significant domain, followed by income and consumption, civic engagement, cultural identity, social connections and housing. Knowledge and skills and safety were not significant. The r-squared for this regression analysis was 0.24.

Brown (2019) used GSS from 2008-2016 to examine mental health wellbeing. The paper included a regression analysis where low life satisfaction was the dependent variable and the explanatory variables comprised the domains and sub-domains defined by McLeod (2018), job wellbeing<sup>4</sup> and demographic controls. Low mental health had the strongest relationship with low life satisfaction, followed by low job wellbeing, low material wellbeing, low civic engagement, loneliness and low cultural identity. An analysis of high life satisfaction gave very similar results. The detailed regression results were not reported so it was not clear whether any of the other domains or sub-domains were significant.

Smith, Peach, & Cording (2019) investigated the relationship between multiple disadvantage and life satisfaction. Using GSS 2014-2016, they constructed measures of disadvantage for health, connectedness, material wellbeing, housing, safety and employment. The criteria used to define disadvantage in each wellbeing domain varied, with the requirement to meet at least one of, or at least two of, or all of the individual indicators specified in the domain. For example, health disadvantage was defined as having one or more of three indicators: poor physical health (SF-12 below 40), poor mental health (SF-12 below 40) or poor self-assessed general health. Material wellbeing disadvantage was defined as having both low household equivalised gross income (where low was defined as below 60% of the median) and lower levels of material wellbeing (0-10 on the MWI-9). The authors used OLS regression to show the relationship between an individual's life satisfaction and the six measures of disadvantage. Demographic controls for age, sex, ethnicity, family type and region were also included.

They found that health, connectedness, material wellbeing and housing were all significant, with health having the greatest explanatory power, followed by connectedness, material wellbeing, housing, and safety. The r-squared was relatively low at 0.189. The authors then included all two-way and three-way interactions between the domains. Only the interactions between health and housing, health and connectedness, and employment and connectedness, were significant and the r-squared increased slightly to 0.193. The combined effects were associated with reduced life satisfaction compared to the sum of their independent effects. The authors concluded that these effects were small, and the impacts of multiple disadvantages were additive.

---

3 The health domain comprised mental health and physical health (see Appendix A for the domain definitions).

4 Those classified as having low job wellbeing were either unemployed, employed and dissatisfied with their job, or not in the labour force and receiving a main benefit.

Carver and Grimes (2019) focused on whether household equivalised gross income (HEGI) or the Economic Living Standard Index (ELSI)<sup>5</sup> better explained life satisfaction using GSS 2012. The GSS 2012 used the short form version of ELSI and included three elements: essentials (the items people possess or consume), economising (the extent a household cut back its expenditure), and self-assessment (adequacy of family income, standard of living and satisfaction with standard of living). The first two elements (comprising 14 and eight indicators) were considered “objective ELSI” and the third (comprising three indicators) as “subjective ELSI”. Regression analysis showed that when both ELSI and HEGI were included in the same regression, HEGI was almost always insignificant. They also concluded that it was the subjective elements of ELSI that rendered income insignificant rather than the objective elements. From 2014 ELSI was replaced by the short form of the Material Wellbeing Index (MWI-9) which comprises nine of the original 24 indicators. Self-assessed income adequacy continued to be collected but self-assessed standard of living and satisfaction with standard of living were not.

In a recent paper, Haines and Grimes (2021) used GSS 2012 data to examine the determinants of life satisfaction of mothers with dependent children in material hardship. The short-form ELSI was used to measure material hardship. The 11 LSF wellbeing domains were included as explanatory variables. Several questions from the GSS were chosen as potential indicators to include for each domain. They used a) the first principal component of each domain and then b) one variable per domain (selecting the variable that had the strongest correlation with life satisfaction among mothers in hardship) as independent variables in the ordered logit regression of life satisfaction. Although the primary focus was on mothers with dependent children, the paper also included results for all males and all females separately, which are more relevant for our study.

In the principal component analysis, nine of the 11 domains were significant for both males and females. *Safely* was significantly associated with SWB for males, while *time use* was not significantly associated with life satisfaction for males or females. *Income and consumption and health* had the strongest relationship with life satisfaction for males, while for females, *income and consumption* and *social connections* had the strongest relationship. The pseudo r-squared was relatively low at 0.14 and 0.15 for males and females respectively. In the single variable analysis, only the *environment* domain (satisfaction with lakes, rivers, harbours, oceans and coastlines in your local area) was not significant. Feeling depressed and anxious, ELSI, satisfaction with where they are living, feeling safe walking home at night, trust in police, ease of expressing one’s own identity, satisfaction with knowledge and skills, labour force status, and degree of social isolation were all significantly associated with life satisfaction. The relative importance of these factors varied by gender to some extent. The reported r-squared was still relatively low at 0.164 and 0.185 for males and females respectively.

The New Zealand literature on SWB favours using OLS regression or ordered logit regression to identify the correlates of SWB. Mental health (when included) based on SF-12 or one or more components of it, was found to be highly correlated. Other factors, including subjective income adequacy, gross household income, material hardship, unemployment, cultural identity, feeling safe, housing and trust in police were identified as significant covariates in one or more studies. While the findings were fairly consistent across studies, the relative importance of individual covariates varied to some extent, reflecting differences

---

5 The ELSI was created by New Zealand’s Ministry of Social Development (MSD) to measure a household’s material living standards and has two versions. See Jensen et al. (2002, 2005) for detailed construction of the index.

in the variables included in the analyses and whether and how 'domain' variables were derived (usually several indicators were combined in some way to create a single domain variable). One notable feature in the previous studies was that r-squared were relatively low, in the range of 0.16-0.20.

Internationally, the key evidence on the characteristics associated with SWB has been accumulating since the early 1990s. We only provide a brief review here, focusing on the factors most frequently appearing in literature including income, health, education and other standard demographic characteristics, drawing heavily on Dolan et al. (2008) and Clarke (2018).

This literature suggests a U-shaped relationship between age and SWB, with SWB higher at younger and older ages and lowest in middle age. The relationship with sex is more ambiguous: some studies have found women were more satisfied with their life than men but others have found no difference once controls are included.

Research on the relationship between SWB and income find a positive relationship, with results consistent across different approaches, including cross-sectional survey data with demographic controls and panel data that controlled for unobserved personal traits (indicating a causal effect). Studies that have included relative income (defined in a range of different ways) have found that SWB is more strongly correlated with relative income than absolute income.

Studies consistently show a strong relationship between SWB and both physical health and psychological health, with the relationship stronger for psychological health than for physical health. Health is usually measured in surveys via subjective self-reports, and as such both health and SWB are contemporaneous subjective reports that are potentially contaminated by common factors that affects both SWB and health (such as mood or personality). Research using more objective measures of health (such as the number of specific health problems that the person experiences) also show a positive correlation between health and SWB.

The findings for education were mixed, with positive, negative and no relationship found in different studies. This may reflect that while education improves income and health, it may also raise expectations about the kind of life the individual aspires to. The concern with most cross-sectional studies is the correlation between education and unobserved personal traits or observed ones such as income. Most longitudinal surveys are not able to identify a significant effect because an individual's education level rarely increases during their time in the survey.

The findings for unemployment were consistent with strong and negative effects, including in fixed-effects models that controlled for individual heterogeneity. The effects of other activities on life satisfaction have also been considered in some studies, such as commuting, caring for others, community involvement and volunteering, exercise and religious activities.

The literature also identified other wellbeing dimensions associated with SWB, including personal attitudes to life (attitudes to circumstances, trust, political persuasion, and religion), relationships (marriage and intimate relationships, having children, seeing family and friends) and the wider economic, social and political environment (income inequality, unemployment rates, inflation, welfare system and public insurance, democracy, climate and the natural environment, safety, neighbourhood deprivation and urbanisation).

The Netherlands Institute for Social Research's *Social State of the Netherlands Report* (Bilj et al. 2017), used regression tree analysis to segment the population using Life Situation Index (LSI) as the dependent variable. The index combines eight social domains (health, housing, social and public participation, participation in sport, living standards, mobility, leisure activities and holiday behaviour).

They used a regression tree analysis to produce a segmentation of the population, to gain insight into the effects of the accumulation of disadvantage or advantage on life situation. The characteristics included in the tree were education, family type, health (chronic condition, disease or disability) and household disposable income decile. The Dutch adult population was categorised into 14 groups with differences in average LSI decreasing or increasing with the accumulation of disadvantage or advantage. For example, the average LSI was 84 for those with low education, lowest 20% of disposable household income and a chronic condition, disability or disease compared to 118 for those with high education, no chronic condition, disability or disease and highest 20% of income (compared to an average index value of 105 for the population overall).<sup>6</sup>

Regression tree analysis is mainly used to identify the key explanatory variables and any interactions between them. So, in addition to segmenting the population into groups, a regression tree analysis also identifies the characteristics that most strongly differentiate the dependent variable in the population. Regression trees are relatively intuitive and easy to understand. They provide a useful visualisation and communication tool, complementing the more common technique of regression analysis. We also include a regression analysis, which provides a more standard and complete analysis of the correlates of SWB. It has the advantage of identifying the estimated effect of a particular covariate having controlled for the effects of other covariates included in the model. We also compare the results with those of earlier New Zealand studies that have included regression analyses using earlier iterations of GSS.

Smith (2018) argued that the joint distribution of the wellbeing outcomes matters because we are interested in knowing whether having multiple disadvantages (ie, having a combination of circumstances such as poor health, low income and low social connections) is concentrated among relatively few people or not. Regression tree analysis helps answer this question.

Section 2 below describes the data and methods we use. Section 3 includes the main results and describes the distribution of wellbeing dimensions and socio-demographic characteristics across the population. Section 4 is the conclusion.

---

6 We treat subjective wellbeing (life satisfaction) as the dependent variable in our analysis, rather than a multi-dimensional composite measure like LSI, and include a much broader range of explanatory variables (both subjective and objective measures of wellbeing across multiple domains).

## 2 Methodology and data

This study uses regression tree analysis to identify the factors most strongly related to differences in SWB in the population. These factors divide the population into groups or segments, whereby people in the same segment share a similar level of SWB and the combination of factors most strongly related to differences in SWB. The analysis shows how combinations of factors explain variation in SWB across the population.

This paper builds on earlier Treasury research by McLeod (2018), which used GSS 2014-2016 data to look at the multi-dimensional aspects of wellbeing. We use the wellbeing domains and sub-domains he defined and update these to include GSS 2018. Owing to some significant changes to the health questions collected between 2016 and 2018, we include separate results for GSS 2018 and GSS 2014-2016.

Following previous research both in New Zealand and internationally, we use self-reported life satisfaction as a measure of SWB. We use the statistical technique of regression tree analysis to identify the factors that most strongly differentiate SWB in the population and to segment the population, whereby individuals in the same segment share a similar experience of wellbeing.

Tree-based methods use a series of splitting rules to segment the population into several distinct and non-overlapping groups. Regression tree methods identify the most important explanatory variables and any interactions between them. The population is first split into two groups (or branches) using the most important explanatory variable. From the first split, each branch is further split into two branches conditional on the previous split. Each branch is split further into two branches, then these are split into two and so on. The final tree is a partition of the population into a number of segments (or leaves).

The regression tree results are sensitive to the variables included in the analysis, as well as their specification. We include all the individual GSS questions that were used to derive the original sub-domains and domains, socio-demographic characteristics available in GSS and selected variables from linked administrative data. Having identified the population segments, we describe their demographic characteristics and wellbeing across the different wellbeing domains.

### 2.1 Methodology

A regression tree approach is one of several tree-based methods that use a series of splitting rules to segment the population into several non-overlapping sub-populations or regions ( $R_1, R_2, \dots, R_m$ ). A regression tree is used to predict a numerical response variable  $Y$ , where the explanatory variables  $X_1, X_2, \dots, X_n$  can be a mixture of binary, continuous or categorical variables.

The tree is first split into two *branches* using the variable split (or partition) that explains the most variation in  $Y$ . For a continuous variable  $X_k$ , the split is  $X_k < s$  versus  $X_k \geq s$ . The split of a categorical variable is of similar type. From the first split, each *branch* is further split into two branches conditional on the previous split. This is referred to as recursive binary splitting. The tree construction process continues until the pre-specified stopping criteria are reached. The endpoint of a tree is a partition of the population into a number of terminal nodes or *leaves*. The points along the tree where the splits occur are known as internal nodes.

At a given node (either internal or terminal), the predicted value and deviance can be calculated. The predicted value is simply the mean of response values of all observations that fall into the node. The deviance of node  $i$  is defined as:

$$D_i = \sum_{x_j \in R_i} (y_j - \mu_{R_i})^2$$

where  $y_j$  is the response value of the  $j^{\text{th}}$  observation and  $\mu_{R_i}$  is estimated mean of response values of all observations within that node. The sum of deviances  $D_i$  over all leaves gives the tree deviance.

Several pre-defined stopping rules can be applied. For example, a threshold on the minimum number of observations contained in a terminal node can be imposed, so that the growth of the tree will terminate when no split can be made without creating a new leaf containing fewer observations than this threshold. A cost complexity parameter  $c_p$  can also be introduced, and effectively replaces the deviance of the tree with a modified expression:

$$D(T) = \alpha |T| + \sum_{i=1}^{|T|} D_i$$

where  $\alpha = c_p D(T_0)$  and  $D(T_0)$  is the deviance of the tree with no splits, and  $|T|$  is the number of terminal nodes in the tree. In this case a new split is only made if it reduces the modified deviance  $D(T)$ , i.e., if the reduction in deviance due to this split outweighs the penalty due to the corresponding increase in the tree size  $|T|$ . By default, the `rpart` implementation of the algorithm calculates the threshold values of  $c_p$  at which an additional split is permitted and estimates the out-of-sample predictive performance of the tree of each corresponding size  $|T|$  by cross-validation.

We treat SWB (reported life satisfaction measured on a scale of 0 to 10) as the dependent or response variable and treat it as a continuous variable. The explanatory variables  $X_1, X_2, \dots, X_n$  we include are the wellbeing domains and sub-domains variables defined in McLeod (2018), responses to individual questions in GSS, and selected other variables collected in the survey or derived from linked administrative data.

To limit over fitting we specify  $c_p = 0.001$ . The cross-validated estimates of predictive performance reach a minimum at lower values of  $c_p$ , indicating that our trees are safe from over-fitting. For GSS 2018 we also require the terminal nodes to contain at least 300 out of 8,780 observations and for GSS 2014-2016 at least 500 out of 14,670. This is approximately 3.5% of the population. A segment smaller than this is likely to be too small to be of policy interest. These parameters generate trees with between 15 and 20 terminal nodes. Because the objective of this study is to segment the population into no more than 15 segments, we usually prune the trees to reduce the number of terminal nodes.

One advantage of the regression tree approach is the ease with which missing values are handled. If a splitting variable is unobserved for a given record the distribution at the node is used to assign it to one branch or the other (Venables & Ripley, 1999). This means that only records that have a missing value for the response variable or have missing values for *all* explanatory variables are excluded from the tree analysis.



## 2.2 Data

The results in this paper are derived from GSS 2014, 2016 and 2018. GSS 2020 was delayed to April 2021 owing to COVID-19<sup>7</sup> and the release of the unit record data in August 2022 was too late to be included in our analysis. The survey is carried out every two years and includes approximately 8,500 people aged 15 years and over who respond to a wide range of social and economic questions. Earlier rounds of GSS were conducted in 2008, 2010 and 2012. There was a significant re-design of the survey in 2014 with questionnaire changes leading to some inconsistencies in the measures we are focusing on. Notably, prior to 2014, life satisfaction was measured on a 5-point Likert scale (very dissatisfied to very satisfied), from 2014 onwards it was measured on a numeric scale of 0-10.

There was also a significant change in the health information collected between 2016 and 2018. Prior to 2018, the 12-item Short Form Health Survey (SF-12)<sup>8</sup> was collected, which provided derived measures of mental health and physical health. In GSS 2018, SF-12 was dropped, except for self-assessed general health status question which continued to be collected. From 2018 onwards, the World Health Organisation – Five Wellbeing Index (WHO-5)<sup>9</sup> which measures current mental wellbeing, was collected. In our analysis mental health based on either SF-12 or WHO-5 is consistently the characteristic most strongly related to (or correlated with) SWB in the population and hence we analyse GSS 2014-2016 and GSS 2018 separately to make sure mental health is included in our analysis. Our main results are based on GSS 2018 but we also include results for GSS 2014-2016 to demonstrate the sensitivity of our findings to changes in the information collected about mental health.

Many GSS questions are used in this analysis, including the variables used to derive the wellbeing domains and sub-domains and various socio-demographic characteristics. In addition, we administrative data in the Integrated Data Infrastructure (IDI) linked to GSS, to calculate other measures, including household equivalised disposable income (HEDI), benefit history and custody history. Appendix B lists the variables we include in the regression tree analysis.

---

7 Collection was suspended in August 2021 owing to the first community outbreak of the COVID-19 Delta variant. The achieved sample size was 3,500, compared to the original planned target of 8,500.

8 The SF-12 is a self-reported measure assessing the impact of health on an individual's everyday life. It is based on response to 12 items, which cover both physical and mental health. The questions pertaining to mental health are: During the last **four** weeks how much of the time have you: accomplished less than you would have like due to any emotional problems, such as feeling depressed or anxious; did you do work or other activities less carefully than usual due to any emotional problems, such as feeling depressed or anxious; felt calm and peaceful; felt downhearted and depressed; did you have a lot of energy; has your physical health or emotional problems interfered with your social activities, such as visiting friends and relatives?

9 WHO-5 is a short self-reported measure of current mental wellbeing based on responses to five statements. How often in the last **two** weeks have you felt: cheerful and in good spirits; calm and relaxed; active and vigorous; woken up feeling fresh and rested, felt your daily life has been filled with things that interest you?

## Subjective wellbeing measure

In this paper, we use life satisfaction, measured on a scale of 0 (least satisfied) to 10 (most satisfied) as a measure of SWB. This measure has been used by OECD (2011) and in many studies, including Brown et al. (2012), Brown (2019), Carver & Grimes (2019), Haines & Grimes (2021), and McLeod (2018), among others. The literature consistently shows correlations between SWB and other wellbeing dimensions. For example, OECD (2011) found correlations with income, jobs, housing, health, work and life-balance, education, social connections, civic engagement and governance, environment, and personal security. Although cognitive evaluation of the level of satisfaction with one's life is a widely accepted proxy of subjective wellbeing, the use of the measure has been criticised by some for being too narrow and that people are likely to change their assessment according to their experiences of life (Sen, 1999).

## Wellbeing domains and sub-domains

We include the wellbeing domains and sub-domains defined by McLeod (2018) with some minor modifications to reflect questionnaire changes between 2016 and 2018. Appendix Table B1 describes the domain and sub-domain definitions.

The Treasury's LSF (2018a, 2018b, 2021) includes 12 wellbeing domains that capture a comprehensive range of wellbeing dimensions for adult New Zealanders, one of which is SWB.<sup>10</sup> Of the other 11, eight domains were included by McLeod (2018), with job and earnings, environment and time use omitted. These were excluded because the relevant questions required to derive them were either not included in GSS or were not sufficient to construct the relevant domain.

Revisiting this we derive a 'job wellbeing' variable that combines labour force status and job satisfaction (for those employed). We include labour force status rather than job wellbeing in our preferred tree analyses, although we included job wellbeing in our regression analysis of GSS 2018. In addition, we removed *discrimination* from the 'social connections' domain because it falls outside the revised definition of the domain in the LSF 2021. We include self-assessed general health status because when SF-12 was replaced with WHO-5 in 2018 it was no longer possible to include the physical health sub-domain used in McLeod (2018).

A key feature of the domain and sub-domain variables derived by McLeod (2018) was that they were categorical variables with three levels (low, medium and high) defined somewhat arbitrarily with the proportion having low wellbeing assumed to be relatively low. In general, 'low' domain wellbeing reflects a person with 'low' wellbeing in at least one sub-domain, while 'high' domain wellbeing reflects a person with 'high' wellbeing in all sub-domains included in the domain. This approach is based on the assumption that a large proportion of the New Zealand adult population are doing quite well on each domain, with a large minority doing very well and a smaller minority reporting some level of difficulty. Appendix B Table B1 describes the domain and sub-domain definitions.

---

10 Twelve domains: cultural capability and belonging (cultural identity previously); engagement and voice (civic engagement and governance); environmental amenity (environment); family and friends (social connections); health; housing; income, consumption and wealth (income and consumption); knowledge and skills; leisure and play (time use); safety; work, care, and volunteering (jobs and earnings); subjective wellbeing.

Combining relevant questions to get sub-domain and domain measures is a largely subjective process that enables several dimensions to be captured in a single domain measure or derived variable. Many of the individual questions have multiple response categories, while others are continuous or ordered categorical variables. We found that the categorisation into three levels can materially reduce the amount of information captured, and in some cases, an individual question or indicator is more strongly correlated with SWB than the sub-domain or domains they are part of. For this reason, we include all the component questions or variables in our final analyses.

### **Demographic and other characteristics sourced from survey or linked administrative data**

We include the following socio-demographic characteristics collected in the survey: age, sex, disability status, ethnicity, family type, highest qualification, labour force status, tenure (home ownership), neighbourhood deprivation (NZDep) and region.

We supplement this with some characteristics sourced from linked administrative data: Household Equivalised Disposable Income (HEDI), poverty status (using 50% of median HEDI before housing costs (BHC50) threshold), benefit receipt and custodial sentences served during the year before interview date. Around 98% of primary respondents aged 15 years and over were linked to the IDI spine and 96% of primary respondents were in households where all members aged 18 years and over were linked to the spine. Because none of these administratively sourced variables were selected in the regression trees we decided not to derive and test any further measures. We only include a small number of administratively sourced characteristics, and many others could be derived and included. We include HEDI decile, BHC50 poverty and prior benefit receipt in the segment profiles. Where we include administrative data in our analysis, about 2%-3% of survey records are dropped owing to non-linkage. The results are weighted to represent New Zealand adult population and we do not adjust the survey weight to account for unlinked records.

# 3 Results

## 3.1 Regression tree analysis – GSS 2014-2016

In this section we describe how we arrive at our preferred regression tree result for 2014-2016. We start by including the eight wellbeing domain variables defined by McLeod (2018) as specified in Appendix B.3 and the demographic variables specified in Appendix B.1 (excluding disability status, which was not collected in GSS 2014). SF-12 mental health index and SF-12 physical health index are used to define the health sub-domains and combined to form the health domain. Wellbeing domain variables take three values: low; medium; high. Appendix C Figure C1 shows the resulting regression tree. The health domain appears at level 1 in the regression tree<sup>11</sup>, followed by age and the income and consumption domain at level 2, civic engagement, social connections, age, and income and consumption appear at level 3 and 4 in the tree<sup>12</sup>. The housing, knowledge and skills, and safety domains are not included in the tree. R-squared is 0.1979 for the tree with 12 segments.

This result is consistent with the regression analysis included in McLeod (2018), which found when all 8 wellbeing domains were included, along with demographic controls, health was the most significant domain, followed by income and consumption, then civic engagement and governance, cultural identity, social connections and housing. The knowledge and skills and safety domains were not significant.

Next, we include both domains and sub-domains (mental health, physical health, material wellbeing, income sufficiency, loneliness, contact with friends and family, discrimination, trust in people and trust in institutions), general health status and the same demographic characteristics as previously. Appendix C Figure C2 shows the resulting regression tree. Income sufficiency is labelled 'enough income' in this figure, reflecting the survey question and response categories the sub-domain it is based on. Loneliness is also described in terms of the response categories, rather than the three sub-domain levels: low; medium; high. Mental health appears at level 1 in the tree<sup>13</sup>, with income sufficiency (having enough income to meet every day needs) and material wellbeing at level 2. The other domains and sub-domains included in the tree were trust in people, trust in institutions, civic engagement, loneliness and general health status (at level 3 and 4 in the tree)<sup>14</sup>. Including the sub-domains significantly improves the r-squared to 0.256 for the tree with 14 segments. Including the mental health sub-domain explains nearly all the improvement in r-squared.

---

11 The first binary split is on the health domain, which separates the population into 2 groups (or branches); those with low health (average SWB=6.7, 20% of the population) and those with medium or high health (8.1, 80%). The next binary split separates out those with medium health (7.9, 58%) and those with high health (8.5, 22%). In Figure C1 we present this as a three-way split at level 1.

12 Segments 6, 7 and 8 are split further to give 15 segments, based on cultural identity, family type and civic engagement respectively (not shown in Figure C1).

13 The first binary split is on the mental health sub-domain, which separates the population into 2 groups (or branches); those with low mental health (average SWB=5.0, 11% of the population) and those with medium or high mental health (8.0, 89%). The next binary split separates out those with medium mental health from those with high mental health. In Figure C2 we present this as a three-way split at level 1.

14 Segment 13 is split further based on general health status to give 15 segments (not shown in Figure C2).

Next, we include the individual questions used to derive the domains and sub-domains with their original response categories. For example, for income sufficiency, we include each of the four categories separately (not enough, only just enough, enough, and more than enough income to meet every day needs). This allows the optimal variable splits to be identified by the algorithm rather than constraining them into the three categories (low, medium and high) in an arbitrary way. The institutional trust sub-domain combines five separate institutions together and we include each separately. Including SF-12 mental health and SF-12 physical health indices, material wellbeing index (MWI-9), and age as continuous variables rather than categorical variables may also improve the model fit. This was the case with r-squared increasing to 0.272 for the tree with 12 segments. The tree (not shown) has four splits on mental health at level 1. Trust in people and income adequacy are selected at level 2, with loneliness, trust in the health system, trust in police, and the civic domain selected at level 3.

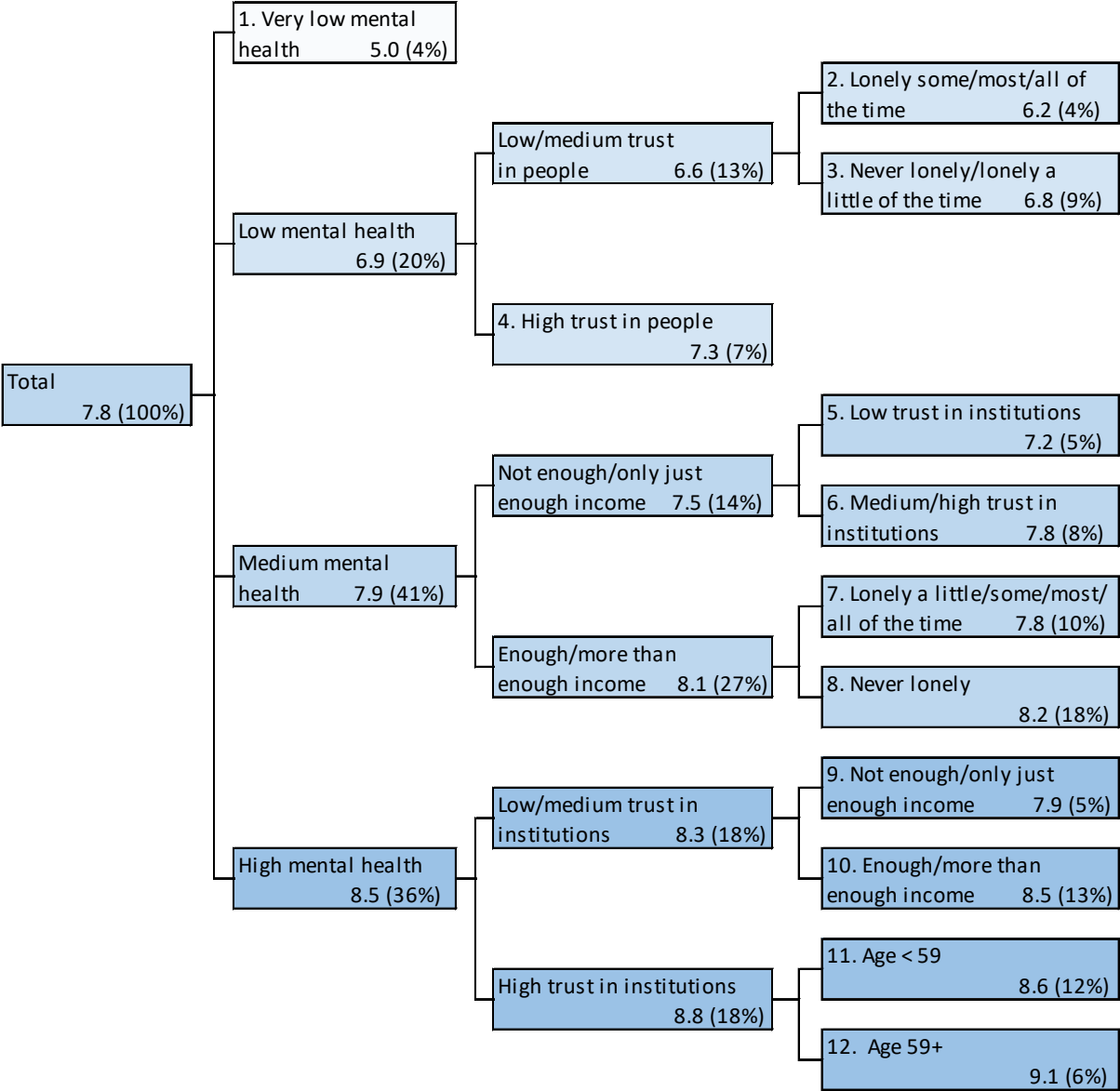
Given trust in the health system, trust in police, and the civic domain are included at level 3, we derived a new institutional trust measure by averaging trust in police, health and education. Appendix C Figure C3 shows the resulting tree with 16 segments at level 4. Replacing the five separate institution variables with the new derived variable, results in the derived variable being selected at level 2 in the tree (for those with high mental health), with income moving to level 3. It also appears at level 3 as expected, replacing trust in police. For those with medium mental health and enough or more than enough income, loneliness replaces trust in the health system at level 3. For those with high mental health, age now appears at level 3 (with people aged 59 years and over having higher SWB) where previously it appeared at level 4. Figure 1 shows the tree which is pruned to give 12 segments. R-squared is 0.272, the same as previously when individual institutions were included. When we exclude age, trust in people replaces age, and r-squared is unchanged at 0.272. This demonstrates that in some cases there is at least one other variable that can be substituted for the selected variable with little or no loss of explanatory power.

Many of the explanatory variables included in the analysis are correlated with each other to some extent. If an important explanatory variable is excluded from the tree analysis, a variable that is relatively highly correlated with it will likely replace it. For example, when mental health (SF-12 index) is excluded from the tree analysis, loneliness appears at level 1 in the tree, general health and income appear at level 2, trust in institutions and general health appear at level 3, trust in institutions, trust in people and general health appear at level 4. Loneliness and general health essentially replace mental health in the tree, with income, trust in people and age appearing as previously, although not necessary in the same place.<sup>15</sup> Excluding mental health (SF-12) reduces r-squared from 0.272 to 0.194, highlighting the importance of mental health in explaining variation in SWB.

---

<sup>15</sup> Those who are lonely a little, some, most or all the time and have poor or fair general health comprise segment 1 (7% of the population with average SWB of 6.2). Those who are lonely a little, some, most or all the time and have good to excellent general health and low trust in institutions comprise segment 2 (10%, 6.9). Those who are never lonely, have enough or more than enough income to meet their everyday needs, have high trust in institutions and high trust in people comprise the segment 15 (6%, 9.0).

**Figure 1 – Regression tree (GSS 2014-2016)**



Source: New Zealand GSS 2014-2016.

The first branch in Figure 1 comprises the 4% of the population with very low mental health, SF-12 < 28 (segment 1 with average SWB of 5.0).

On the second branch of the tree are those people with low mental health 28 ≤ SF-12 < 45. For this group, trust in people is the factor that most strongly differentiates SWB. Among the 13% with low mental health and low to medium trust in people (0-7), are the 4% who feel lonely at least some of the time (segment 2 with average SWB of 6.2) and the 9% who never feel lonely or feel lonely a little of the time (segment 3 with average SWB of 6.8). Those with low mental health and high trust in people (8+) comprise segment 4 with average SWB of 7.3.

On the third branch of the tree are those with medium mental health. For this group, having enough income to meet every day needs is the factor that most strongly differentiates SWB. Those with not enough or only just enough income to meet their everyday needs are split into two groups, the 5% who have low trust in institutions (average trust < 6.83<sup>16</sup>) comprise segment 5 (average SWB of 7.2) and the 8% who have medium to high trust in institutions (trust >=6.83) comprise segment 6 (average SWB of 7.8). For the group with enough or more than enough income, the 10% who feel lonely at least some of the time comprise segment 7 (average SWB of 7.8) and the 18% who never feel lonely comprise segment 8 (average SWB of 8.2).

On the fourth branch of the tree are those with high mental health. For this group, trust in institutions is the factor that most strongly differentiates SWB in this group. Those with low to medium trust in institutions (trust < 7.83<sup>17</sup>) are split into two groups, the 5% who don't have enough income or who have only just enough income comprise segment 9 (average SWB of 7.9) and the 13% who have enough income or more than enough income comprise segment 10 (average SWB of 8.5) Those with high trust in institutions (trust >= 7.83) are split into two groups based on age, the 12% aged under 59 years comprise segment 11 (average SWB of 8.6) and the 6% aged 59 years or older who comprise segment 12 (average SWB of 9.1).

## 3.2 Regression tree analysis – GSS 2018

Having selected our preferred tree for GSS 2014-2016, we now turn to GSS 2018 and include the individual questions that comprise the domains and sub-domains with their original response categories. Recall that the questions used to measure mental health changed in 2018 with WHO-5 replacing SF-12. The questions included are quite different, so they are likely measuring somewhat different aspects of mental health. Appendix D Figure 1 shows the resulting regression tree. The tree has five splits on mental health at level 1.<sup>18</sup> Having enough income and trust in police appear at level 2, with partnership status, family and friends (previously labelled social connections), loneliness and problem with neighbourhood crime appearing at level 3. We constrain the tree so that the 4% with the highest mental health are combined with the 27% with very good mental health (to give four splits on mental health at level 1). The variables selected at level 2 and 3 in this branch of the tree are unchanged. We choose four splits rather than five at level 1, because the reduction in r-squared is negligible (a reduction from 0.2666 to 0.2622) and we prefer a tree that is more similar to the GSS 2014-2016 tree.

---

16 Corresponds to a combined trust score (on a scale of 0 to 10) for the 3 institutions of 20.5, ie, a split into <=20 versus >=21.

17 Corresponds to a combined trust score for the 3 institutions of 23.5, ie, a split into <=23 versus >=24

18 This is the result of a series of binary splits, with the first split dividing the population into two groups (34% with low mental health and 66% with medium to high mental health), then a split of the 34%, into the 8% with very low mental health and the 26% with low mental health, and a split of the 66% into the 35% with medium mental health and the 31% with high mental health. A further split separates the 4% with very high mental health from the 27% with high mental health.

Next, we modify the family and friends domain by excluding the discrimination sub-domain, as this better reflects the domain in the LFS 2021. We exclude individual institutions and include the institutional trust variable (averaging trust in police, health and education). This replaces trust in police and for those with low mental health and enough income it also replaces the friends and family domain. The reduction in r-squared is negligible (a reduction from 0.2622 to 0.2620) and demonstrates that in some cases there is more than one variable that could be substituted for the selected variable with very little loss in explanatory power. Recall that we specify a minimum segment size of 300 (or 3.5% of the population). We prune the tree at level 3, to reduce the number of segments from 16 to 13, which does not materially reduce the explanatory power of the tree (r-squared reduces from 0.2671 to 0.2620), although statistically significantly more variation is explained by retaining these additional branches.<sup>19</sup> Figure 2 below shows the resulting tree. Some variables are described as taking the values: very low; low; medium; and high. These are specified in Appendix E Table E1.

The variables that do not appear in the tree (or are subsequently pruned) are contact with family, contact with friends, discrimination, the four housing sub-domains, the four safety sub-domains, highest qualification, region, age, sex, ethnicity, disability status, NZDep, tenure (home ownership) and the administratively sourced variables; household equivalised disposable income, poverty status, receipt of main benefits and custodial sentence history. Note this does not necessarily mean these variables are not significantly correlated with SWB, only that the variables included in the tree more strongly differentiate SWB, for the population sub-groups defined by the branches of the tree, than these variables do.

As we observed for 2014-16, at levels 2 and 3 in the trees there is usually at least one other characteristic that could be substituted for the selected characteristic with very little or no loss of explanatory power.

If mental health (WHO-5 index) is excluded from the tree analysis, loneliness appears at level 1 in the tree, general health and having enough income appear at level 2 and trust in institutions and general health appear at levels 3 and 4.<sup>20</sup> Excluding mental health (WHO-5) reduces r-squared from 0.262 to 0.194, highlighting the importance of mental health in explaining variation in SWB. The tree is very similar to the 2014-2016 tree where mental health (based on SF-12) was excluded and the explanatory variables included in the two analyses are the same.

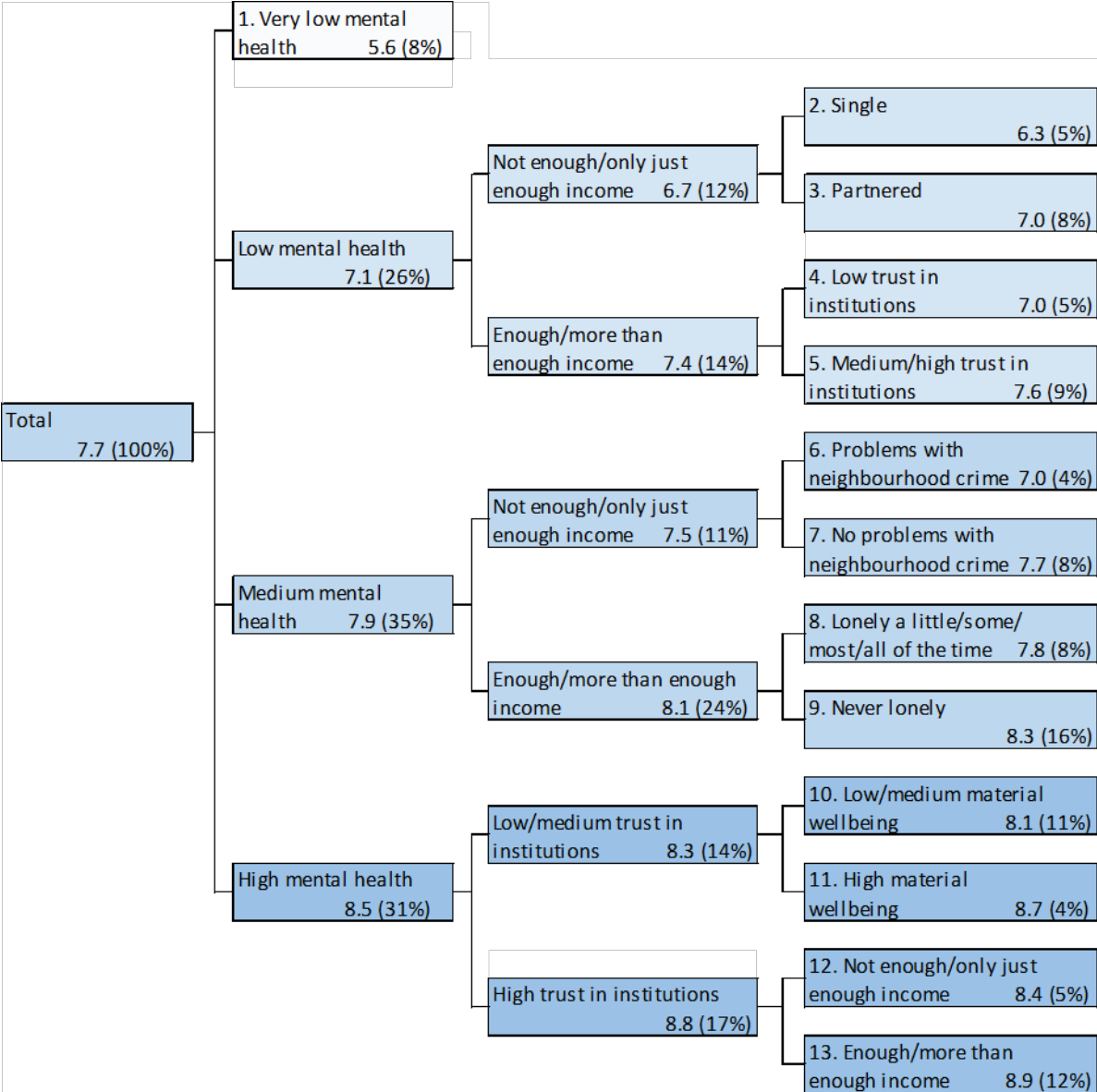
---

19 Segments 9, 10 and 13 were split in two based on general health status, house condition, and trust in institutions respectively.

20 Those who are lonely some, most or all the time and have poor or fair general health comprise segment 1 (4% of the population with average SWB of 5.6). Those who are lonely some, most or all the time and have good to excellent general health and low trust in institution comprise segment 2 (6%, 6.6). Those who are never or only a little lonely, have enough or more than enough income, have very good or excellent general health and have high trust in institutions comprise segment 11 (7%, 8.9).



**Figure 2 – Regression tree (GSS18)**



Source: New Zealand GSS 2018.

Comparing the preferred tree for 2018 (Figure 2) with that for 2014-2016 (Figure 1) we see many of the variables included in the trees are the same, and overall, the two trees are relatively consistent. Despite the differences in the mental health questions included, mental health still appears at level 1 of the GSS 2018 tree (with average SWB varying from 5.6 to 8.5), having enough income and trust in institutions appear at level 2, with income replacing trust in people at level 2 for those with low mental health. Those with very low mental health comprise segment 1 (although this is a larger segment than in 2014-2016).<sup>21</sup> The second main branch of the tree (those with low mental health) is the most different from GSS 2014-2016. People with low mental health, not enough or only just enough income and who were single, comprise segment 2. People with low mental health, not enough or only just enough income and who were partnered comprise segment 3. People with low mental health, enough or more than enough income and low to medium trust in institutions comprise segment 4. People with low mental health, enough or more than enough income and high

21 Suggesting that WHO-5 doesn't distinguish SWB as strongly at the bottom of the distribution as SF-12 does.

trust in institutions comprise segment 5. In the third and fourth main branches (medium and high mental health) income and trust in institutions are selected at level 2 in both trees, and some of the same characteristics are included at level 3 as previously, although not necessary in the same place.

## **Segment profiles GSS 2018**

In this section we describe the preferred tree for 2018 in more detail and the socio-demographic characteristics of the population segments. Similar analysis for GSS 2014-2016 is available on request.

The regression tree in Figure 2 shows that the population aged 15 years and above can be divided into four groups based on mental health (assessed using the WHO-5 index) with the average SWB varying from 5.6, 7.1, 7.9 and 8.5 respectively on a scale of 0-10. Three of the four branches are split into two and these are each split into two, to give a total of 13 segments.

The 8% of the population with very low mental health comprises segment 1. Those with low mental health are split into two groups based on income. Those reporting they did not have enough or only just enough income to meet their everyday needs (an average SWB of 6.7) are then split into two groups: segment 2 comprising single or sole parents reporting an average SWB of 6.3 (5% of the population) and segment 3 comprising couples with or without children reporting an average SWB of 7.0 (8% of the population). Those with low mental health, who had enough or more than enough income to meet their everyday needs (average SWB of 7.4) are split into 2 groups: segment 4, comprising those who had low to medium trust in institutions reporting an average SWB of 7.0 (5% of the population), and segment 5, comprising those who had high trust in institutions reporting an average SWB of 7.6 (9% of the population).

People with medium mental health are also split into two groups based on income. Those reporting they did not have enough or only just enough income to meet their everyday needs are then split into 2 groups: segment 6 comprising those reporting problems with neighbourhood crime, with an average SWB of 7.0 (4% of the population) and segment 7 comprising those reporting no problems with neighbourhood crime, with an average SWB of 7.7 (8% of the population). People with medium mental health, who had enough or more than enough income to meet their everyday needs were split into two groups: segment 8, those who were lonely at least a little of the time, reporting an average SWB of 7.8 (8% of the population) and segment 9, those who were lonely none of the time, reporting an average SWB of 8.3 (16% of the population).

People with high mental health are split into two groups based on level of trust in institutions. Those reporting low to medium trust in institutions are then split into two groups: segment 10 comprising those with low or medium material wellbeing (MWI-9 < 19) with an average SWB of 8.1 (11% of the population) and segment 11 comprising those reporting high material wellbeing (MWI-9 >= 19) with an average SWB of 8.7 (4% of the population). People with high mental health, who reported high trust in institutions are split into two groups: segment 12, those reporting they did not have enough or only just enough income to meet their everyday needs, with an average SWB of 8.4 (5% of the population) and segment 13, those reporting they had enough or more than enough income to meet their every-day needs, with an average SWB of 8.9 (12% of the population).

Table 1 below summarises the key characteristics of each population segment. The segments with the lower average SWB include a higher proportion of disadvantaged groups. Below we describe segments that have average SWB below the overall average of 7.7. All the segments are described in Appendix E Table E1, which also identifies where each segment materially differs from the demographic profile of population overall. Appendix E Table E2 describes the segment profiles across the various wellbeing domains and sub-domains defined in McLeod (2018).

This table shows that:

- Segment 1: (8% of the population with very low mental health and the lowest average SWB of 5.6) includes relatively more sole parents, disabled people, those receiving a main benefit, those with no qualifications, females and those aged 35-64 years.
- Segment 2: (5% of the population who are single, have low mental health and who do not have enough income or have only just enough income to meet every day needs, average SWB of 6.3) includes relatively more sole parents (45%), disabled people, females, Māori and Pacific peoples, those not in employment, those receiving a main benefit, those living in more deprived areas, those in poverty and renting their homes.
- Segment 3: (8% of the population who are partnered, have low mental health and not enough income or only just enough income, average SWB of 7.0) includes relatively more couples with children (65%), those aged 15-34, Pacific peoples, unemployed, those receiving main benefits, those with lower HEDI and those living in more deprived areas.
- Segment 4: (5% of the population who have low mental health, enough or more than enough income and have low to medium trust in institutions, average SWB of 7.0) includes relatively more people aged 35-64, those of European ethnicity, those employed and those with higher HEDI. It includes relatively fewer people who are unemployed and those with low HEDI.
- Segment 6: (4% of the population who have medium mental health, not enough income or only just enough income and live in a neighbourhood where crime is a problem, average SWB of 7.0) includes relatively more Māori and Pacific peoples, sole parents, those not in employment, those receiving main benefits, living in more deprived areas, renting their homes, those with lower HEDI and those in poverty. It includes relatively fewer people with high HEDI.

**Table 1 – Segment profiles – selected socio-demographic characteristics (column percentage)**

Segment		1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Average SWB		5.6	6.3	7.0	7.0	7.6	7.0	7.7	7.8	8.3	8.1	8.7	8.4	8.9	7.7
Population percentage		7.9	4.5	7.5	4.9	9.2	3.5	7.7	7.9	15.6	10.8	3.5	4.8	12.3	100
Life satisfaction (SWB)	0 to 4 (least)	29	13	5	5	2	6	2	1	s	2	s	1	1	5
	5	22	20	15	10	5	11	7	4	3	5	1	5	1	7
	6	10	17	13	13	10	14	9	5	3	5	1	4	1	7
	7	19	25	31	32	26	23	21	26	14	16	10	11	6	19
	8	11	18	24	27	36	31	33	40	41	31	29	33	25	30
	9	5	3	6	7	13	8	14	14	21	20	20	17	25	15
	10 (most)	4	4	7	5	8	7	13	10	18	22	37	29	41	17
Sex	Male	40	36	49	51	48	48	50	41	52	48	61	50	56	49
	Female	60	64	51	49	52	53	50	59	48	52	40	50	44	51
Age	15-34	29	44	43	31	36	43	41	44	27	34	16	45	33	35
	35-64	53	41	46	54	48	46	42	44	52	46	52	38	40	47
	65+	18	14	11	14	16	11	17	12	21	19	32	17	28	18
Disability		22	15	10	9	6	7	7	5	6	4	3	4	4	8
Family type	Couple with children	38	S	65	44	47	43	45	36	44	42	30	46	39	41
	Couple without child	24	S	35	30	31	18	19	28	37	26	46	21	35	29
	Sole parent	15	45	S	6	5	16	13	8	5	11	6	12	5	10
	Not in a family nucleus	23	55	S	19	17	23	22	28	15	21	19	22	21	20
Ethnicity	European	71	66	57	82	79	62	51	75	81	68	83	45	70	70
	Asian	12	9	18	7	13	13	22	15	10	14	7	27	20	15
	Māori	13	28	16	16	8	20	14	13	10	14	10	11	8	13
	Pacific peoples	8	13	16	3	4	17	15	2	3	11	1	21	5	8
	MELAA*	2	2	1	2	1	s	2	2	1	1	s	3	2	1
	Other	3	2	1	2	1	1	1	1	1	2	2	3	s	2
Labour force status	Employed	58	56	67	77	73	69	62	74	75	68	71	58	68	68
	Unemployed	4	10	6	1	2	7	7	3	2	4	1	6	2	4
	Not in the labour force	37	34	27	22	25	25	31	23	23	29	28	36	30	28
Highest qualification	No qualification	24	22	16	13	12	14	16	8	12	16	16	16	14	15
	School level	32	34	34	32	32	37	39	33	26	35	27	41	32	33
	Tertiary below degree	23	21	27	29	24	23	23	26	28	25	29	21	24	25
	Degree or above	18	20	21	25	31	23	19	33	32	23	22	21	27	25
	Other	3	2	2	1	1	2	3	1	2	1	6	2	3	2
Tenure of household	Dwelling not owned	41	58	42	30	26	43	50	37	21	33	18	42	30	34
	Dwelling owned	47	36	48	57	56	46	43	50	62	52	59	47	54	52
	Dwelling in family trust	12	6	10	14	18	10	7	13	17	14	23	11	16	14
NZDep	1 (Least deprived)	19	10	16	20	28	13	14	24	30	21	29	21	27	22
	2	18	13	15	23	21	15	19	23	20	19	30	14	22	19
	3	23	18	17	18	20	13	17	16	20	17	16	13	21	18
	4	21	28	25	20	18	25	25	23	20	25	13	23	18	22
	5 (Most deprived)	19	31	28	19	13	33	24	15	10	18	13	31	12	18
Household equivalised disposable income	1 (Lowest income)	13	20	10	8	6	17	15	9	6	11	9	16	10	10
	2	14	13	11	6	8	10	12	8	9	13	10	18	12	11
	3	11	13	16	8	7	10	13	8	6	10	9	11	6	9
	4	10	11	12	9	5	15	14	7	7	10	7	11	8	9
	5	10	11	10	8	9	13	13	9	8	10	4	8	7	9
	6	10	13	10	11	12	11	9	11	9	9	8	11	8	10
	7	10	10	9	11	11	9	11	10	10	8	10	8	11	10
	8	9	4	10	11	11	9	5	11	12	11	13	7	13	10
	9	8	4	9	15	14	3	6	12	14	11	10	7	12	11
	10 (Highest income)	7	1	3	14	15	2	3	15	18	8	20	3	14	11
Poverty (BHC50)		12	17	10	6	6	17	13	8	5	9	7	14	9	9
Any main benefit receipt in the last year		17	31	14	8	5	17	11	6	3	9	2	13	4	9

\* Middle Eastern/Latin American/African

While sole parents and people with a disability are somewhat over-represented in the lower wellbeing segments many are also in higher wellbeing segments. Appendix E Table E3 shows how population sub-groups are spread across segments. For example, 13% of sole parents are in segment 1 (compared to 8% in the population overall), 21% are in segment 2 (compared to 5% in the population overall) and 7% are in segment 13 (compared to 12% in the population overall). The demographic profile of some segments looks very similar to that of the population overall particularly those where average SWB is close to the overall population average.

The segments with higher average SWB include a lower proportion of disadvantaged groups. For example, segment 13 has the highest average SWB of 8.9 and has relatively fewer sole parents and those receiving main benefits, and relatively more males, older people and Asian people. The segment comprises the 12% of the population who had high mental health, enough income or more than enough income and high trust in institutions.

Segments with lower average SWB have low wellbeing in many other wellbeing domains, and segments with higher SWB have higher wellbeing in other wellbeing domains.

### 3.3 Regression analysis – GSS 2018

In this section we present the results we obtained from a regression analysis of the GSS 2018 data. While regression analysis and regression tree analysis are similar in some respects, they differ in some important ways, which are discussed below. We want to see how consistent the regression and regression tree results are and how our regression results compared with those included in previous research that used GSS 2008 to 2016 data.

We include the same GSS variables in the regression and regression tree analyses with one exception.<sup>22</sup> The detailed results we obtained from both an ordinary least squares (OLS) regression and an ordered logit regression analysis of GSS 2018 are in Appendix F. We use SAS proc surveyreg and proc surveylogistic with jack-knife replicate weights to estimate the models, so the standard errors and tests of significance account for the complex survey design. The OLS and logit regression results are very similar and we only describe the results from the OLS regression here. One advantage of OLS regression is that the parameter estimates can be directly compared to differences in average SWB in different branches of the regression tree.

The most significant variables in the regression analysis are mental health (WHO-5), sex, Māori ethnicity, family type, loneliness, job wellbeing and income adequacy, which are all highly significant, with F-values corresponding to  $p < 0.0001$ . General health status, material wellbeing index (MWI-9), tenure (home ownership), ease of expressing one's cultural identity, age group and trust in the education system are also highly significant with F-values corresponding to  $0.0001 < p \leq 0.001$ . Highest qualification, trust in courts, trust in people, trust in media, contact with friends and contact with family are significant with F-values corresponding  $0.001 < p \leq 0.05$ .

---

22 Labour Force Status (LFS) was included in the tree analysis, but job wellbeing (which combined LFS with job satisfaction) was included in the regression analysis. We didn't include the administratively sourced variables in the regression analysis because none of them appeared in the regression tree. When we subsequently tested whether these variables were significant in a regression analysis, we found they were not.

The other variables included in the model are not statistically significant with F-values corresponding to  $p \geq 0.05$ . These variables are feeling safe waiting for or using public transport at night, feeling safe walking alone in your neighbourhood after dark, feeling safe at home by yourself at night, problems with neighbourhood crime, trust in police, trust in Parliament, trust in the health system, being of European, Asian or Pacific ethnicity, housing condition, cold house, problems with mould, household crowding, and having experienced discrimination.

The results from the regression analysis are generally consistent with those from the regression tree, although the relative importance of some variables differs. While related, the two methods differ in some important ways. Regression analysis identifies the effect of each covariate controlling for the other covariates included in the regression model, while regression tree analysis uses a series of binary splits, with each subsequent split conditional on the previous split.<sup>23</sup> The first binary split identifies the covariate split that most strongly differentiates SWB in the population. Regression tree analysis also identifies interactions between covariates if they exist. A particular covariate may not be significantly correlated with the dependent variable across the total population, but it may be significant within a specific population sub-group (or branch of the tree). The relative importance of variable generally differs when interactions exist for population sub-groups defined by the branches in the tree. In this case, neighbourhood crime is not significant in the regression analysis but appears in the tree for the sub-population with medium mental health and not enough or only just enough income (segments 6 & 7). Including and testing the significance of interactions in a regression analysis identifies interactions that are, on average, significant across the whole population. When we include interactions between the most significant variables, no interactions are significant at  $p \leq 0.05$ .

In the regression analysis, trust in police, health, and Parliament are not significant and trust in education, courts, and media are significant at  $0.01 < p \leq 0.05$ . In the regression tree analysis, trust in police, trust in education and trust in health appear in various trees and we combined these together (by averaging them) rather than including the individual institutions separately in our preferred trees.

While highly significant in the regression analysis, neither sex nor Māori ethnicity appear in the 2018 tree. This is because there are other variables that better differentiate SWB for the population sub-groups defined by the branches in the tree. The estimated effects of sex (0.30 higher SWB, on average, for females) and ethnicity (0.33 higher SWB, on average, for Māori compared to non-Māori) are smaller than the differences in average SWB in adjacent branches of the regression tree at levels 2 and 3 which differ by between 0.5 and 0.7.

Job wellbeing is also highly significant in the regression analysis. While not included in our preferred tree, it does originally appear at level 2 for people with high mental health, splitting the population depending on whether people were very satisfied with their job or not.

---

23 The tree construction process is very similar to using forward selection in a regression analysis.

Apart from material wellbeing, all other covariates significant at  $0.0001 < p \leq 0.05$  do not appear in the regression tree because there are other variables that differentiate the population more strongly at levels 1 to 3. The estimated effects in the regression for highest qualification, ease of expressing one's cultural identity, home ownership, age,<sup>24</sup> and contact with family and friends, were mainly quite small, although some were not. Finding it hard or very hard to express one's cultural identity compared to very easy, reduced SWB by -0.69, being unemployed compared to being employed, reduced SWB by -0.23 ( $p=0.07$ ), being in poor general health compared to being in excellent health, reduced SWB by -0.60. Note that while these estimated effects are large, the proportion of the population with these characteristics is relatively small which means that very little variation in SWB is explained by the effects overall. Only 1.9% of the population found it hard or very hard to express their cultural identity, 3.0% said their health was poor and 3.7% were unemployed.

In summary, the results from the regression analysis are generally consistent with those from the regression tree analysis, although the relative importance of some variables differ owing to differences in the methods. By construction, only a small number of covariates can be included in a tree with 12 to 15 leaves, with the covariate selected at each step the one that most strongly differentiates the population sub-group at a given branch in the tree. Some covariates that are highly significant in a regression analysis do not appear in the final tree and some variables included in the tree are not significant in the regression analysis. This is because regression trees identify interactions between variables that exist for a particular sub-group of the population (defined by the branches in the tree) but, on average, are not present across the entire population. In particular, neighbourhood crime is not significant in the regression analysis but appears in a branch of the tree at level 3, and trust in police and trust in the health system are not significant in the regression analysis but are included in the derived variable *trust in institutions*, which appears at levels 2 and 3 of the regression trees.

---

24 Age didn't appear in the tree for 2018, however, it did appear at level 3 in one branch of the tree for 2014–2016, where those with high mental health and enough or more than enough income, were split by age  $< 59$  versus age  $\geq 59$ .

## 4 Conclusion

This study uses data from the New Zealand General Social Survey (GSS) and regression tree analysis to identify the factors most strongly related to differences in subjective wellbeing (SWB) in the population aged 15 years and above. The tree analysis divides the population into groups or segments, whereby people in the same segment share a similar level of SWB and the factors most strongly related to SWB. The analysis shows how combinations of factors explain differences in SWB in the population and provides a person-centric view of wellbeing across multiple dimensions of wellbeing.

Our approach is strongly influenced by research by the Netherlands Institute for Social Research (Bijl et al., 2017), which used regression tree analysis to segment the population and illustrate how the accumulation of disadvantage and advantage explained variation in Life Situation Index. We apply this approach to life satisfaction in New Zealand, building on previous research over the last 10 years that has used GSS data to explore the correlates of SWB, including research by McLeod (2018) on multi-dimensional wellbeing. We include the variables used to construct the wellbeing domains and sub-domains defined in McLeod (2018) and other selected characteristics in the survey and from linked administrative data.

While the regression tree results are somewhat sensitive to the survey year, owing to changes in the questions used to assess mental health, we find that of all the characteristics we consider, mental health is the one that most strongly differentiates SWB in the population and consistently appears at level 1 in the trees for both 2018 and 2014-2016.<sup>25</sup> Having enough income to meet every day needs, trust in institutions and trust in people, appear at level 2. Partnership status, loneliness, neighbourhood crime, material wellbeing, having enough income, trust in institutions and age appear in one or both trees at level 3. However, we find that at levels 2 and 3 in the trees there is usually at least one other variable that could substitute for the selected variable, with no or very little loss in explanatory power.<sup>26</sup>

We find some differences in the distribution of socio-demographic characteristics across population segments, particularly for those with the relatively low or high average SWB, but in general, the differences are not large. The segments with the low average SWB include a higher proportion of disadvantaged groups, including more sole parents, those receiving main benefits, and those with a disability. While sole parents and people with a disability are somewhat over-represented in the lower wellbeing segments, many are also in higher wellbeing segments.

---

25 The changes to the mental health questions collected between 2016 and 2018 did not substantially alter the tree results, despite the different questions asked measuring somewhat different aspects of mental wellbeing. The two instruments used (WHO-5 and SF-12) are detailed in Appendix B.

26 There may be other variables that could be included (or derived) that would have greater explanatory power and displace some of the variables selected at levels 2 or 3 in the trees.



Our findings largely confirm the results obtained in previous studies in New Zealand that used regression analysis to examine the correlates of SWB. The degree of correlation between SWB and other wellbeing domains vary considerably, with some having little or no correlation and others moderately correlated. The conclusions regarding the relative importance of these differ to some extent, reflecting the sensitivity of results to the variables included in the analysis and whether and how domain variables are constructed. The approach we have taken is consistent with previous New Zealand studies that have included both objective and subjective measures, including mental health, as explanatory variables.<sup>27</sup>

Previous studies that included mental health found it was the strongly correlated with SWB. Some studies did not include the subjective measure of income adequacy (having enough income to meet everyday needs<sup>28</sup>) and used household equivalised gross income derived from survey responses instead. Consistent with Carver & Grimes (2019) we find that the subjective measure of income adequacy is much more strongly correlated with SWB than household equivalised disposable income derived from linked administrative data.

Previous studies have also found that subjective measures are more generally highly correlated with SWB than demographic or objective characteristics. Even when all the subjective and objective measures we consider are included in a OLS regression analysis, only a modest proportion of the variation in SWB is explained by the model (r-squared is 0.35 for GSS 2018). Of the variables we considered, mental health explains the most variation in SWB and omitting it from the analysis materially reduces the amount of variation explained by the model.<sup>29</sup>

While our approach is strongly influenced by the tree analysis included in the Netherlands report (Bilj et al. 2017), our results cannot be compared to those. Firstly, because that report used Life Situation Index as the dependent variable and not life satisfaction. The Life Situation Index combined eight domains into a single composite index<sup>30</sup> so the dependent variable in the two analyses is very different. Secondly, the explanatory variables included in the two analyses are quite different, with most of the variables that feature in our trees not included in their analysis.

We use regression tree analysis rather than clustering to do the population segmentation. While clustering techniques are much more commonly used, we choose to use a tree-based method because we want to prioritise SWB over other wellbeing domains (by treating it as the dependent variable in the analysis) and to create segments where individuals within a segment share the three characteristics that most strongly differentiate their SWB from the rest of the population. In clustering analysis there is no dependent variable, and this approach would result in segments that are much less well differentiated on SWB.

---

27 For example, Brown (2012) included mental health SF-12, Smith, Peach & Cording (2019) combined mental health SF-12 with other health measures to define an indicator of health disadvantage, while Haines & Grimes (2021) included the single component of SF-12 that is most strongly correlated with SWB.

28 How well does your (you and your partner's combined) total income meet your everyday needs for such things as accommodation, food, clothing and other necessities?

29 Excluding mental health (WHO-5 index) from the OLS regression reduces r-squared from 0.35 to 0.29. An OLS regression that includes only mental health has an r-squared of 0.24. Many of the variables included in the analysis are correlated with each other and omitting mental health from the tree analysis results in loneliness and general health status replacing mental health in the tree and a material reduction in r-squared from 0.26 to 0.19.

30 Health, housing, social and public participation, participation in sport, living standards, mobility, leisure activities and holiday behaviour.

# References

- Bijl, R., Boelhouwer, J., & Wennekers, A. (2017). Social state of the Netherlands 2017.
- Brown, S. (2019). Wellbeing and mental health: An analysis based on the Treasury's Living Standards Framework (No. 19/01). Analytical Paper.
- Brown, D., Woolf, J., & Smith, C. (2012). An empirical investigation into the determinants of life satisfaction in New Zealand. *New Zealand Economic Papers*, 46(3), 239-251.
- Carver, T., & Grimes, A. (2019). Income or consumption: Which better predicts subjective well-being? *Review of Income and Wealth*, 65, S256-S280.
- Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the Gallup World Poll. *Journal of Economic perspectives*, 22(2), 53-72.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology*, 29(1), 94-122.
- Dolan, P., & White, M. (2007). How can measures of subjective well-being be used to inform public policy? *Perspectives in Psychological Science*, 2, 71-85.
- Haines, L., & Grimes, A. (2021). What Matters for the Wellbeing of Mothers and Children in Material Hardship? Application of a Modified Indicator Framework. *Social Indicators Research*, 1-26.
- Jensen, J., Spittal, M., Crichton, S., Sathiyandra, S., & Krishnan, V. (2002). Direct measurement of living standards: The New Zealand ELSI scale. Wellington: Ministry of Social Development.
- Jensen, J., M. Spittal and V. Krishnan, ELSI Short Form User Manual for a Direct, Centre for Social Research and Evaluation Te Pokapū Rangahau Arotaki Hapori, Wellington, 2005.
- McLeod, K. (2018). Our people-multidimensional wellbeing in New Zealand. The New Zealand Treasury. Analytical Paper No. 18/04.
- OECD (2011). *How's Life? Measuring Well-being*, OECD Publishing, Paris, <https://doi.org/10.1787/9789264121164-en>.
- Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
- Smith, C. (2018). Treasury living standards dashboard: Monitoring intergenerational wellbeing. New Zealand Treasury, Wellington.
- Smith, C., Peach, E. K., & Cording, J. (2019). The impact of multiple disadvantage on subjective wellbeing: New Zealand Families. Ministry of Social Development.
- Stiglitz, J. E., Sen, A., & Fitoussi, J. P. (2009). Report by the commission on the measurement of economic performance and social progress.
- The Treasury. (2018a). Living Standards Framework: Introducing the Dashboard. New Zealand Government.
- The Treasury. (2018b). Our people, our country, our future. The living standards framework: Background and future work. New Zealand Government.
- The Treasury. (2021). The Living Standards Framework 2021. New Zealand Government.
- Venables, W. N., & Ripley, B. D. (1999). Tree-based methods. In *Modern applied statistics with S-Plus* (pp. 303-327). Springer, New York.

# Appendix A

## Summary of relevant New Zealand literature

Study	GSS	Methodology	Key relevant findings
Brown et al., “ <i>An empirical investigation into the determinants of life satisfaction in New Zealand</i> ”, New Zealand Economic Papers, 2012	2008	Regression of life satisfaction, with selected characteristics grouped under 6 domains (demographic, education, health, economic, social and community relationships, safety and security).	Mental health, household equivalised gross income, social and community relationships (partnered, help in a crisis, felt they belonged to New Zealand, difficult to express identity) and unemployment were most strongly correlated with life satisfaction.
McLeod, “ <i>Our people- multidimensional wellbeing in New Zealand</i> ”, Treasury analytical paper, 2018	2014-2016	Regression of life satisfaction with 8 wellbeing domains (each including one or more sub-domains).	Health (combined physical and mental health sub-domains), income and consumption, followed by civic engagement, cultural identity, social connections, housing were correlated with life satisfaction.
Brown, “ <i>Wellbeing and mental health: an analysis based on the Treasury’s Living Standards Framework</i> ”, Treasury analytical paper, 2019	2008-2016	Regression of low (and high) life satisfaction, which included mental health and loneliness.	Mental health, followed by job wellbeing, material wellbeing, civic engagement, loneliness, low cultural identity.
Smith, Peach, & Cording, “ <i>The impact of multiple disadvantage on subjective wellbeing: New Zealand families</i> ”, Ministry of Social Development, 2019	2014-2016	Regression of life satisfaction included interactions between the domains. Constructed measures of disadvantage for 6 domains: health, connectedness, material wellbeing, housing, safety, and employment.	Health, followed by connectedness, material wellbeing, housing, and safety. Health: poor physical health or poor mental health or poor general health. The effects of multiple disadvantages on SWB were largely additive (the few interactions effects that were significant were small).
Carver and Grimes, “ <i>Income of consumption: which better predicts subjective wellbeing</i> ”, Review of Income and Wealth, 2019	2012	Regression analysis of life satisfaction.	When ELSI and Household Equivalised Gross Income (HEGI) are included, HEGI is not significant. ELSI’s subjective elements (adequacy of family income and standard of living) rendered income insignificant rather than the objective elements (the items people possess or consume and the extent of economising behaviours).
Haines and Grimes, “ <i>What matters for the wellbeing of mothers and children in material hardship? Application of a modified indicator framework</i> ”, Social Indicators Research, 2021	2012	Regression analysis of life satisfaction for males and females separately, with selected characteristics grouped under 11 wellbeing domains. Principal component or single most significant variable in each domain were included.	Principal component analysis: 9 out of 11 domains are significant. Income and consumption most important for both males and females, then social connections for females and health for males. Single variable: feeling anxious/depressed, material wellbeing (ELSI), satisfaction with where living, feeling safe walking home at night, trust in police, ease of expressing one’s identity.

# Appendix B

## Variable descriptions

### 1 The GSS sourced demographic variables

- Age: 15-34, 35-64, and 65 years old and above.
- Sex: Male and female.
- Disability status: Whether the respondent was screened as disabled. This question was not included in GSS 2014.
- Ethnicity: European, Asian, Māori, Pacific, MELAA (Middle Eastern, Latin American, and African ethnic) and Other. Multiple response, where a respondent can belong to more than one ethnic group.
- Family type: Couple with children, Couple without children, Not in a family nucleus, Sole parent.
- Education level: No qualification, Level 1, 2, 3 certificate or an overseas secondary school certificate, Tertiary below degree (Level 4 certificate or Level 5, 6 diploma), Degree level and above (Level 7, 8, 9, 10) and Other (not elsewhere included).
- Tenure of household: Dwelling not owned by usual residents, dwelling fully or partly owned by usual residents, dwelling held in a family trust by usual residents.
- Labour force status: Employed, Unemployed, Not in labour force.
- NZ Deprivation index: Ten deciles of deprivation, reported in five quintiles from 1 (least deprived) to 5 (most deprived).
- Region: Auckland, Wellington, Northland group (Northland, Bay of Plenty, Gisborne) Rest of North Island, Canterbury, and Rest of South Island.
- Disability status: Stats NZ derived variable based on the following questions:  
Do you have difficulty:
  - seeing, even if wearing glasses
  - hearing, even if using a hearing aid
  - walking or climbing steps
  - remembering or concentrating
  - washing all over or dressing
  - communicating, for example, understanding or being understood using your usual language?

### 2 Administratively sourced variables

- Household equivalised disposable income (HEDI) decile: Where we can link all individuals in the respondent's household aged 18 years and above to the IDI spine, we use administrative data on taxable income (wages and salary, benefits, taxable income from IR3 returns (self-employment, rents, dividends and other sources) and non-taxable income (including non-taxable benefits and transfers, e.g., accommodation supplement and working-for-families) to calculate household equivalised disposable income decile, from 1 (Lowest income) to 10 (Highest income).

- Poverty status: Using HEDI, we define people to be in poverty if their HEDI is lower than 50% of the median HEDI (referred to as the Before Housing Cost 50 (BHC50) poverty measure).
- Number of months receiving any main benefits during the year before the survey interview date (for the primary respondent): Main benefits include sole parent support, supported living payment, job seeker work ready and job seeker health condition: No months, 1-3 months, 4-6 months, 7-9 months, and 10-12 months.
- Number of months spent any time in custody during the year before the survey interview date (for the primary respondent). No months, 1-3 months, 4-6 months, 7-9 months, and 10-12 months.

### 3 The individual GSS questions associated with the domains and sub-domains developed by McLeod (2018)

#### 3.1. Income and consumption

- **Material wellbeing:** Material Wellbeing Index (MWI-9) is scored from 0 to 20. It is derived from the following nine questions:

In the last 12 months, to what extent have you done any of the following things to keep costs down (not at all, a little, a lot):

- gone without fresh fruit or vegetables
- postponed or put off visits to the doctor
- done without, or cut back on, trips to the shops or other local places
- spent less on hobbies or other special interests than you would like
- put up with feeling cold
- delayed replacing, or repairing, broken or damaged appliances?

When buying, or thinking about buying clothes or shoes for yourself, how much do you usually feel limited by the money available? (not at all limited, a little limited, quite limited, very limited)

Imagine that you have come across an item that you would really like to have. This item costs \$300. It is not an essential item, it's an extra. If this happened in the next month, how limited would you feel about buying it?

In the last 12 months have you/you or your partner not paid electricity, gas, rates or water bills on time because of a shortage of money?

- **Income adequacy:** How well does your/you and your partner's combined total income meet your everyday needs for such things as accommodation, food, clothing and other necessities? Not enough, only just enough, enough, more than enough.

#### 3.2. Health: SF-12 (collected in 2008-2016). Stats NZ derives a **mental health** index and a **physical health** index (which take values from 1 to 100) based on the following 12 questions:

- In general, would you say your health is excellent, very good, good, fair or poor?
- Please tell me if your health now limits you in the following activities: moderate activities, such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.

- Please tell me if your health now limits you in the following activities: climbing several flights of stairs.
- During the past four weeks, how much of the time have you accomplished less than you would like as a result of your physical health?
- During the past four weeks, how much of the time were you limited in the kind of work or other regular daily activities you do as a result of your physical health?
- During the past four weeks, how much of the time have you accomplished less than you would like as a result of any emotional problems, such as feeling depressed or anxious?
- During the past four weeks, how much of the time did you do work or other regular daily activities less carefully than usual as a result of any emotional problems, such as feeling depressed or anxious?
- During the past four weeks, how much did pain interfere with your normal work, including both work outside the home and housework?
- During the past four weeks, how much of the time have you felt calm and peaceful?
- During the past four weeks, how much of the time did you have a lot of energy?
- During the past four weeks, how much of the time have you felt downhearted and depressed?
- During the past four weeks, how much of the time has your physical health or emotional problems interfered with your social activities, such as visiting friends, relatives?

**Mental health:** WHO-5 World Health Organisation Mental Wellbeing Index (collected 2018 onwards) derived from five questions: In the last two weeks, how often have you:

- Felt cheerful and in good spirits
- Felt calm and relaxed
- Felt active and vigorous
- Woken up feeling fresh and rested
- Felt that your daily life has been filled with things that interest you?

Response categories: all of the time, most of the time, more than half of the time, less than half of the time, some of the time, at no time. These are scored 0 (all of the time) through 5 (at no time), then multiplied by 4 to give a score between 0 and 100.

**General health** (collected in all years): In general, would you say your health is excellent, very good, good, fair or poor?

### 3.3. Housing

- **Housing condition:** How would you describe the condition of your house or flat?
- **Mould problem:** 2014-2016: Does your house or flat have no problem, a minor problem or a major problem with dampness or mould? 2018: Does any part of your house have mould growing on it, for example, on the walls, ceiling, window frames, curtains, blinds? (yes, no).
- **Cold problem:** In winter, is your house or flat colder than you would like?
- **Crowding:** How many bedrooms needed?

3.4. Knowledge and skills: What is your highest completed **qualification**? No qualification, Level 1, 2, 3 certificate or an overseas secondary school certificate, Tertiary below degree (Level 4 certificate or Level 5, 6 diploma), Degree level and above (Level 7, 8, 9, 10) and Other (not elsewhere included).

### 3.5. Social connections

- **Loneliness:** People who have contact with family and friends can still feel lonely sometimes, while those who have little contact may not feel lonely at all. In the last four weeks, how much of the time have you felt lonely?
- **Contact with family:** Please think about all the contact you have with your family or relatives [who don't live with you]. How would you describe the amount of contact you have with them?
- **Contact with friends:** Please think about all the contact you have with your friends [who don't live with you]. How would you describe the amount of contact you have with them?
- **Discrimination:** In the last 12 months have you been discriminated against?

### 3.6. Safety

- **Feeling unsafe:** This is derived from the following questions:  
Thinking about crime, how safe or unsafe do you feel:
  - at home by yourself at night
  - walking alone in your neighbourhood after dark
  - waiting for or using public transport, such as buses and trains at night
  - using the internet for online transactions?
- **Victim of crime:** In the last 12 months, were any crimes committed against you?
- **Neighbourhood crime:** This is derived from the following questions:  
Thinking about the last 12 months, have any of these things been a problem in your neighbourhood?
  - vandalism/graffiti
  - burglary/break-ins
  - assaults
  - harassment
  - people using or dealing drugs?

### 3.7. Civic engagement and voice

- **Trust in people:** On a scale of zero to ten, in general how much do you trust most people in New Zealand?
- **Trust in institutions:** This is derived from the following questions:  
Where zero is not at all, and ten is completely, how much do you trust:
  - the courts
  - the education system
  - the health system
  - the parliament

- the police?

3.8. **Cultural identity:** People in New Zealand have different lifestyles, culture and beliefs that express who they are. How easy or hard is it for you to be yourself in New Zealand?

#### 4 Wellbeing domain and sub-domain variables sourced from GSS

The domain and sub-domains variable described in Table B1 below are those from McLeod (2018) which refer to domains defined in the 2018 version of the LFS. These were used in 2014-2016 regression trees shown in Figure C1 and C2. The low, medium, and high wellbeing levels for the variables included in the trees shown in Figure 1 and Figure 2 (material wellbeing index, mental health, trust in people and trust in institutions) differ from those in this table.

In almost all cases where a domain has more than one sub-domain a person is considered to have low wellbeing in that domain if they have low wellbeing for **any** of the sub-domains, and a person is considered to have high wellbeing in that domain if they have high wellbeing for **all** sub-domains. The exception is the Safety domain, where a person is considered to have low wellbeing if they have more than one sub-domain in low wellbeing.



**Table B1 – Definitions of wellbeing domains and sub-domains used in 2014-2016 regression trees shown in Figure C1 and C2**

Domain	Sub-domain	Low wellbeing	Medium wellbeing	High wellbeing
Income and consumption	Material wellbeing	Material wellbeing index (MWI) is 0-7 out of 20	MWI is 8-17 out of 20	MWI is 18-20 out of 20
	Income sufficiency	Not enough money to meet every day needs	Only just enough money to meet every day needs	Enough or more than enough money to meet every day needs
Health	Mental health	< 36 on SF-12 mental health index	37-53 on SF-12 mental health index	>54 on SF-12 mental health index
	Physical health	< 36 on SF-12 physical health index	37-53 on SF-12 physical health index	>54 on SF-12 physical health index
Housing	Condition	Immediate repairs or maintenance needed	Some repairs or maintenance needed	Only minor repairs or maintenance needed
	Cold problem	House always too cold in winter	House sometimes or often too cold in winter	House never too cold in winter
	Crowding	Bedrooms needed	N/A	No bedrooms needed
Knowledge and skills	Qualifications	No qualification	School or lower-level tertiary qualification	Bachelor's degree or higher
Social connections	Loneliness	Lonely most or all of the time	Lonely a little or some of the time	Never lonely
	Friend and family contact	Not enough/too much contact with friends or family	Right amount of contact with friends or family but not both	Right amount of contact with friends and family
	Discrimination	Discriminated against in past year	N/A	Not discriminated against in past year
Safety	Feeling unsafe	Feels unsafe at home alone at night, walking home after dark, using public transport or doing online transactions	Does not feel unsafe in any listed situation, nor safe in all situations	Feels safe in all listed situations
	Victim of crime	Victim of crime in past year	N/A	Not a victim of crime in past year
	Neighbourhood crime	Problem with vandalism, burglaries, assaults, harassment or drugs in neighbourhood	N/A	No problem with vandalism, burglaries, assaults, harassment or drugs in the neighbourhood

Domain	Sub-domain	Low wellbeing	Medium wellbeing	High wellbeing
Civic engagement	Trust in people	Trust in most people in NZ is 0-4 out of 10	Trust in most people in NZ is 5-6 out of 10	Trust in most people in NZ is 7-10 out of 10
	Trust in institutions (including the courts, education system, health system, parliament, and police)	Low trust (0-4 out of 10) in more than one out of five institutions	Low trust (0-4 out of 10) in fewer than two and high trust (7-10 out of 10) in fewer than four institutions	High trust (7-10 out of 10) in at least four out of five institutions
Cultural identity	Able to be yourself in NZ	Very hard, hard, sometimes easy and sometimes hard	Easy	Very easy

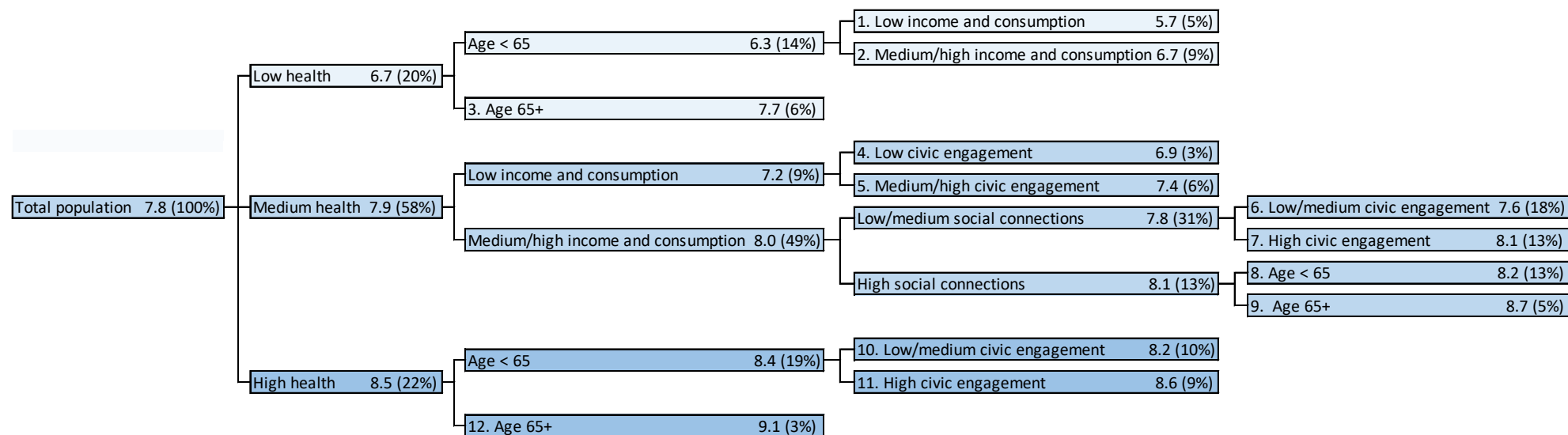
Note: In the case of Discrimination, Victim of crime, Neighbourhood of crime, there are only two levels: 'low' and 'high'. The housing domain includes three of the four sub-domains originally included in McLeod (2018).

Originally, we intended to do a pooled analysis of GSS 2014 to 2018, and because the response categories for 'problems with mould' changed between 2016 and 2018, we excluded the mould sub-domain and derived the housing domain based on the other three sub-domains (house condition, cold problem, and crowding).

# Appendix C

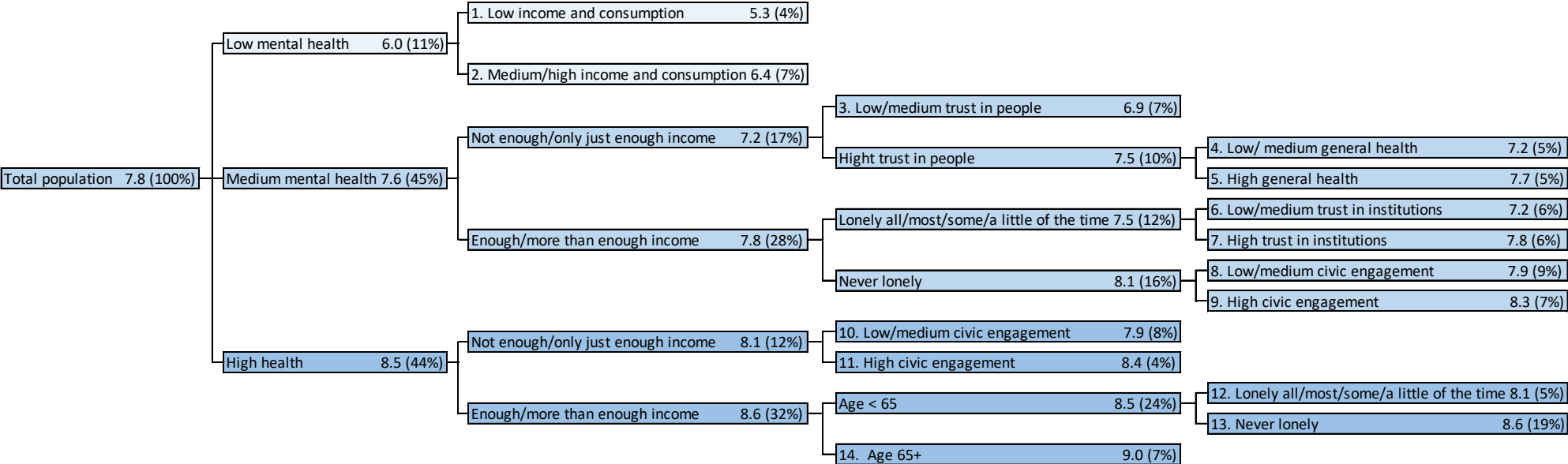
## Alternative regression trees (GSS 2014-2016)

Figure C1 – Alternative regression tree (GSS 2014-2016) based on the original wellbeing domains



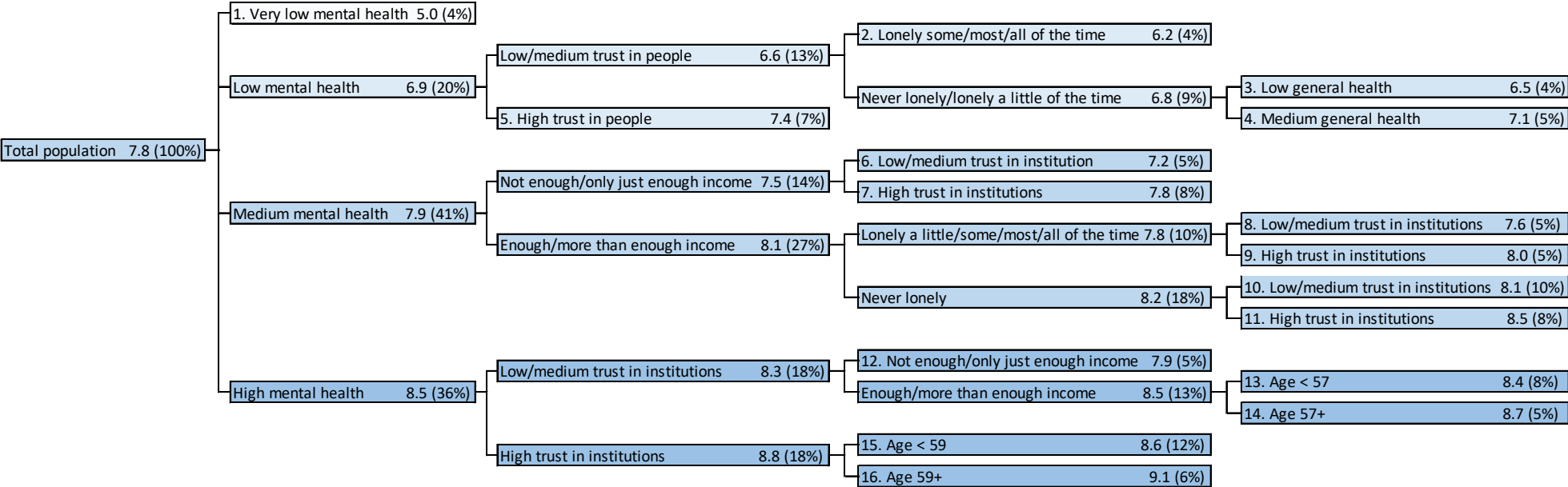
Note: The wellbeing domain variables are described in Table B1.

**Figure C2 – Alternative regression tree (GSS 2014-2016) based on the original domains and sub-domains**



Note: The wellbeing domain and sub-domain variables are described in Table B1.

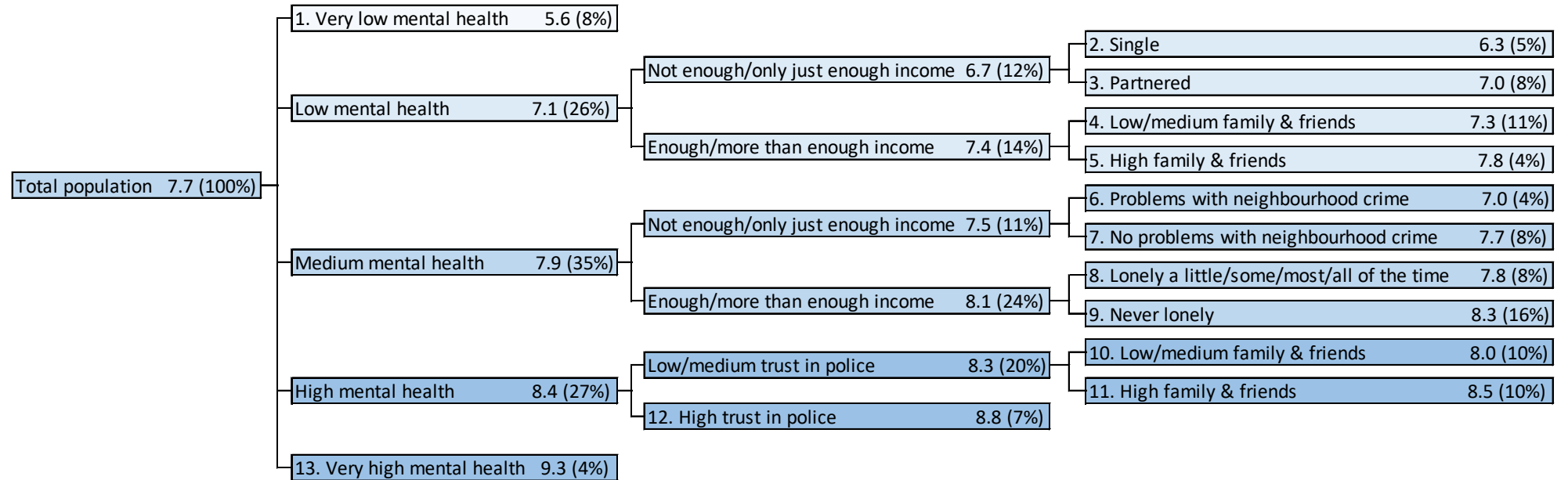
**Figure C3 – Regression tree (GSS 2014-2016) with 16 segments**



# Appendix D

## Alternative regression tree (GSS 2018)

Figure D1 – Alternative regression tree (GSS 2018)



# Appendix E

## Segment profiles (GSS 2018)

Table E1 – GSS 2018: Segment profiles

Segment	Key characteristics
1 Very low mental health ( <i>WHO-5</i> < 34)	<ul style="list-style-type: none"> <li>• Average SWB of 5.6/10, accounting for 8% of population.</li> <li>• More people aged 35-64 (53%), female (60%), sole parents (15%), with a disability (22%), having no qualification (24%), and fewer people having a degree or above (18%).</li> <li>• More people not in labour force (37%), receiving main benefits (17%) and receiving benefits for 10-12 months of the last year (11%).</li> </ul>
2 Low mental health (34 <= <i>WHO-5</i> < 58), not enough or only just enough income to meet every day needs, single or sole parents	<ul style="list-style-type: none"> <li>• Average SWB of 6.3/10, accounting for 5% of population.</li> <li>• More people aged 15-34 (44%), female (64%), Māori (28%), Pacific peoples (13%), sole parents (45%), with a disability (15%), and having no qualification (22%).</li> <li>• More people not in labour force (34%), unemployed (10%), living in the most deprived areas (31%) (only 10% were living in the least deprived areas), renting (58%), having low household equivalised disposable income (20% in the bottom income decile and only 1% in the top income decile), living in poverty (17%), receiving main benefits (31%) and receiving benefits for 10-12 months of the last year (22%).</li> </ul>
3 Low mental health (34 <= <i>WHO-5</i> < 58), not enough or only just enough income to meet every day needs, couples	<ul style="list-style-type: none"> <li>• Average SWB of 7.0/10, accounting for 8% of population.</li> <li>• More people aged 15-34 (43%), Pacific peoples (16%), couple with children (65%) and based in Auckland (42%).</li> <li>• More people unemployed (6%), living in the most deprived areas (28%), low household equivalised disposable income (16% in the third lowest income decile and only 3% in the top income decile), and receiving the main benefits (14%).</li> </ul>
4 Low mental health (34 <= <i>WHO-5</i> < 58), enough or more than enough income to meet every day needs, low trust in institutions (< 6.83)	<ul style="list-style-type: none"> <li>• Average SWB of 7.0/10, accounting for 5% of population.</li> <li>• More people aged 35-64 (54%) and European (82%). Other demographic characteristics largely resemble those of the population overall.</li> <li>• Fewer people unemployed (1%) and having low household equivalised disposable income (only 6% in the second lowest income decile).</li> </ul>
5 Low mental health (34 <= <i>WHO-5</i> < 58), enough or more than enough income to meet every day needs, medium to high trust in institutions (>=6.83)	<ul style="list-style-type: none"> <li>• Average SWB of 7.6/10, accounting for 9% of population.</li> <li>• Fewer sole parents (5%) but more people having a degree or above (31%).</li> <li>• More people living in the least deprived areas (28%) and having high household equivalised disposable income (only 6% in the bottom income decile).</li> </ul>
6 Medium mental health (58 <= <i>WHO-5</i> < 74), not enough or only just enough income to meet every day needs, problem with neighbourhood crime	<ul style="list-style-type: none"> <li>• Average SWB of 7.0/10, accounting for 4% of population.</li> <li>• More people aged 15-34 (43%), Māori (20%), Pacific peoples (17%), and sole parents (16%).</li> <li>• More people unemployed (7%), living in the most deprived areas (33%), with low HEDI (17% in the bottom income decile), living in poverty (17%), receiving main benefits (17%) and receiving benefits for 10-12 months of the last year (9%).</li> </ul>

Segment	Key characteristics
7 Medium mental health (58 <= WHO-5 < 74), not enough or only just enough income to meet every day needs, no problem with neighbourhood crime	<ul style="list-style-type: none"> <li>• Average SWB of 7.7/10, accounting for 8% of population.</li> <li>• More Pacific peoples (15%) and people based in Auckland (44%).</li> <li>• More unemployed (7%), renting a house (50%), and having low household equivalised disposable income (only 14% in the top 30% income group).</li> </ul>
8 Medium mental health (58 <= WHO-5 < 74), enough or more than enough income to meet every day needs, lonely	<ul style="list-style-type: none"> <li>• Average SWB of 7.8/10, accounting for 8% of population.</li> <li>• More people aged 15-34 (44%) and having a degree and above (33%).</li> </ul>
9 Medium mental health (58 <= WHO-5 < 74), enough or more than enough income to meet every day needs, never lonely	<ul style="list-style-type: none"> <li>• Average SWB of 8.3/10, accounting for 16% of population.</li> <li>• More people aged 35-64 (52%), European (81%), more couple without children (37%) but fewer sole parents (5%), and more people having a degree or above (32%).</li> <li>• Fewer people unemployed (2%) and having low income (6% in the bottom income decile), more people living in their own house (62%).</li> </ul>
10 High mental health (WHO-5 >= 74), low to medium trust in institutions (trust < 7.83), low to medium material wellbeing (MWI < 19)	<ul style="list-style-type: none"> <li>• Average SWB of 8.1/10, accounting for 11% of population.</li> <li>• Fewer people with a disability (4%).</li> </ul>
11 High mental health (WHO-5 >= 74), low to medium trust in institutions (trust < 7.83, high material wellbeing (MWI >= 19)	<ul style="list-style-type: none"> <li>• Average SWB of 8.7/10, accounting for 4% of population.</li> <li>• More people aged 35+ (52% aged 35-64 and 32% aged 65+), males (61%), European (83%), couple without children (46%), and fewer people with disability (3%).</li> <li>• Fewer people unemployed (1%); more people aged 65+ and still working (12%), living in the least 40% deprived areas (59%), and living in either their own house or one held in a family trust (82%).</li> </ul>
12 High mental health (WHO-5 >= 74), high trust in institutions (trust >= 7.83), not enough or only just enough income to meet every day needs	<ul style="list-style-type: none"> <li>• Average SWB of 8.4/10, accounting for 5% of population.</li> <li>• More people aged 15-34 (45%), Asian (27%, Pacific peoples (21%), and those based in Auckland (52%) and fewer people with a disability (4%).</li> <li>• More people unemployed (6%), not in labour force (36%), aged under 65 and not working (26%), living in the most deprived areas (31%), having low income (only 3% in the top income decile), receiving main benefits (13%) and receiving benefits for 10-12 months of the last year (8%).</li> </ul>
13 High mental health (WHO-5 >= 74), high trust in institutions (trust >= 7.83), enough or more than enough income to meet every day needs	<ul style="list-style-type: none"> <li>• Average SWB of 8.9/10, accounting for 12% of population.</li> <li>• More people aged 65+ (28%), males (56%), Asian (20%), couple without children (35%) but fewer sole parents (5%), fewer people with disability (4%).</li> <li>• Fewer people unemployed (2%), more people aged 65+ and still working (8%), more people living in the least deprived areas (27%).</li> </ul>

Notes: Trust in institutions is the average of trust in police, health, and education systems on a scale of 0 to 10. Other demographic characteristics largely resemble those of population overall if not specifically mentioned.



**Table E2 – GSS 2018 segment profiles: Wellbeing domains and sub-domains  
(column percentage)**

Segment		1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Average SWB		5.6	6.3	7.0	7.0	7.6	7.0	7.7	7.8	8.3	8.1	8.7	8.4	8.9	7.7
Population percentage		7.9	4.5	7.5	4.9	9.2	3.5	7.7	7.9	15.6	10.8	3.5	4.8	12.3	100
Income and consumption	Low	37	57	42	6	3	43	38	4	3	17	s	32	2	18
	Medium	46	41	55	62	64	57	61	68	54	68	6	65	45	55
	High	15	s	s	30	30	s	s	25	42	13	93	s	50	25
<i>Material wellbeing</i>	Low	29	44	32	6	3	31	24	4	3	10	s	16	2	13
	Medium	52	53	57	62	64	63	67	68	54	72	s	68	45	57
	High	17	2	8	30	30	5	7	25	42	17	100	13	51	28
<i>Income sufficiency</i>	Low	22	36	23	s	s	25	24	s	s	11	s	21	s	10
	Medium	34	64	76	s	s	75	76	s	s	29	6	78	s	27
	High	43	s	s	99	99	s	s	99	99	59	93	s	98	62
General health	Low	52	29	22	20	14	13	12	9	10	7	5	9	4	15
	Medium	30	41	40	44	37	40	33	31	29	21	21	24	16	30
	High	18	29	38	36	49	46	55	60	61	71	75	67	80	55
Housing	Low	24	33	25	13	10	29	23	11	9	17	4	28	9	16
	Medium	43	39	44	46	44	42	44	47	38	37	27	36	32	40
	High	28	20	27	36	41	26	28	35	50	40	69	32	55	39
Knowledge and skills	Low	24	22	16	13	12	14	16	8	12	16	16	16	14	15
	Medium	58	58	63	62	58	63	64	59	56	61	61	63	59	60
	High	18	20	21	25	31	23	19	33	32	23	22	21	27	25
Social connections	Low	45	37	35	31	30	26	26	35	18	22	12	19	15	26
	Medium	37	42	36	46	44	39	38	64	22	35	25	33	31	37
	High	17	21	29	23	26	34	35	s	59	41	63	46	53	36
<i>Loneliness</i>	Low	17	8	5	3	3	2	2	4	s	2	1	1	0	3
	Medium	47	55	45	48	44	43	36	96	s	29	15	26	22	35
	High	35	37	51	49	53	54	62	s	100	68	84	73	77	61
<i>Contact with friends and family</i>	Low	18	10	11	11	15	7	10	13	7	6	3	6	6	10
	Medium	36	29	31	36	33	34	32	37	27	26	19	26	23	30
	High	44	60	58	52	52	59	58	49	65	67	78	65	70	60
<i>Discrimination</i>		27	29	25	21	18	19	17	22	12	16	9	12	10	17
Safety	Low	28	31	23	23	19	54	3	15	13	17	10	10	8	17
	Medium	50	51	54	52	49	44	52	55	52	46	40	43	42	49
	High	21	16	23	23	31	s	43	28	35	36	49	45	48	33
Civic engagement	Low	45	38	29	47	8	31	19	15	16	28	29	8	3	21
	Medium	36	42	47	48	37	34	42	42	38	52	54	32	23	39
	High	18	20	24	5	54	35	38	43	46	20	18	60	74	39
<i>Trust in people</i>	Low	22	24	16	15	5	16	8	5	6	10	9	5	2	10
	Medium	29	34	31	37	19	25	28	24	22	29	20	19	12	24
	High	49	43	52	48	75	60	64	70	71	61	71	76	86	66
<i>Trust in institutions</i>	Low	37	28	20	40	4	25	15	11	12	24	25	3	1	17
	Medium	40	42	44	54	29	34	36	36	34	50	54	20	15	36
	High	23	30	35	5	67	41	48	52	53	26	20	77	83	47
Cultural identity	Low	31	24	25	20	14	20	21	16	11	13	8	15	7	16
	Medium	32	38	35	34	33	37	39	31	30	38	28	38	29	34
	High	36	38	39	46	53	43	39	53	59	49	64	48	64	50

Note: The wellbeing levels of low, medium, and high are defined in Appendix B Table B1.

**Table E3 – GSS 2018 segment profiles: Selected characteristics (row percentage)**

Segment		1	2	3	4	5	6	7	8	9	10	11	12	13
Average SWB		5.6	6.3	7.0	7.0	7.6	7.0	7.7	7.8	8.3	8.1	8.7	8.4	8.9
Population percentage		8	5	8	5	9	4	8	8	16	11	4	5	12
Life satisfaction	0-4 (least satisfied)	50	13	8	5	4	5	3	2	2	5	s	1	3
	5	24	12	15	7	7	5	8	4	6	7	1	3	1
	6	12	11	14	9	13	7	10	5	7	7	1	3	3
	7	8	6	12	8	13	4	9	11	11	9	2	3	4
	8	3	3	6	4	11	4	8	10	21	11	3	5	10
	9	3	1	3	2	8	2	7	7	22	14	5	5	21
	10 (most satisfied)	2	1	3	2	4	1	6	5	17	14	7	8	29
Sex	Male	6	3	7	5	9	3	8	7	17	11	4	5	14
	Female	9	6	7	5	9	4	8	9	15	11	3	5	10
Age	15-34	6	6	9	4	10	4	9	10	12	11	2	6	11
	35-64	9	4	7	6	9	3	7	7	17	11	4	4	10
	65+	8	4	4	4	8	2	7	5	17	11	6	5	18
Disability		23	9	10	6	8	3	7	5	13	5	1	3	6
Family type	Couple with children	7	s	12	5	10	4	8	7	16	11	2	5	11
	Couple without children	7	s	9	5	10	2	5	8	20	10	6	3	15
	Not in a family nucleus	9	12	s	5	8	4	8	11	11	11	3	5	12
	Sole parent	13	21	s	3	5	6	11	7	8	12	2	6	7
Ethnicity	European	8	4	6	6	10	3	6	9	18	10	4	3	12
	Asian	7	3	9	2	8	3	12	8	11	10	2	9	17
	Māori	8	10	10	6	6	5	9	8	12	12	3	4	8
	Pacific Peoples	7	7	15	2	5	7	14	2	7	14	1	12	8
	MELAA*	12	5	5	7	5	S	11	11	9	9	s	9	16
	Other	14	6	6	6	8	3	3	6	17	11	6	s	13
Labour force status	Employed	7	4	7	6	10	3	7	9	17	11	4	4	12
	Unemployed	8	12	11	1	5	6	15	6	7	10	1	8	7
	Not in the labour force	11	5	7	4	8	3	8	6	13	11	3	6	13
Highest qualification	No qualification	13	7	8	4	7	3	9	4	13	12	4	5	12
	School level	8	5	8	5	9	4	9	8	12	12	3	6	12
	Tertiary below degree	7	4	8	6	9	3	7	8	18	11	4	4	12
	Degree level and above	6	3	6	5	11	3	6	10	19	10	3	4	13
	Other	12	5	6	3	6	4	10	3	17	6	10	5	15
Tenure of household	Dwelling not owned	9	8	9	4	7	4	11	9	10	10	2	6	11
	Dwelling owned	7	3	7	5	10	3	6	8	19	11	4	4	13
	Dwelling family trust	7	2	6	5	12	3	4	7	20	11	6	4	15
NZDep	1 (Least deprived)	7	2	5	5	11	2	5	8	21	10	5	4	15
	2	7	3	6	6	10	3	8	9	16	10	5	3	14
	3	10	4	7	5	10	3	7	7	17	10	3	3	14
	4	8	6	9	5	8	4	9	8	15	13	2	5	10
	5 (Most deprived)	8	8	11	5	7	6	10	7	9	11	3	8	8
	Total	8	5	7	5	9	3	8	8	16	11	3	5	12
Household equivalised disposable income	1 (Lowest income)	10	8	7	4	6	6	11	7	9	11	3	7	12
	2	10	5	8	3	7	3	8	6	13	12	3	8	13
	3	10	6	13	5	7	4	10	7	10	12	3	6	8
	4	8	5	10	5	6	6	12	6	13	11	3	6	10
	5	8	5	8	4	9	5	10	8	14	12	2	4	10
	6	8	5	8	5	12	4	7	9	15	10	3	5	10
	7	8	4	7	6	10	3	8	8	17	9	4	4	13
	8	7	2	7	5	10	3	4	8	19	12	5	3	16
	9	6	2	6	7	13	1	4	9	22	11	3	3	14
	10 (Highest income)	5	0	2	7	13	1	2	11	27	8	7	1	16
Poverty (BHC50)		10	8	8	3	6	6	10	7	8	10	3	7	13
Benefit history last year		15	15	11	4	5	6	10	5	5	10	1	7	5

\* Middle Eastern/Latin American/African

# Appendix F

## Regression results (GSS 2018)

Table F1 – OLS Regression GSS 2018

Domain	Sub-domain (ref. category)	Category	Estimate	S.E	Pr >  t	F Value	DF	Pr > F
Demographic	<b>Ethnicity</b> (Non-Māori)	Māori	0.328 ***	0.067	<.0001	23.9	1	<.0001
	<b>Ethnicity</b> (Non-Pacific Peoples)	Pacific Peoples	0.080	0.107	0.456	0.6	1	0.4563
	<b>Ethnicity</b> (Non-European)	European	-0.114	0.084	0.176	1.9	1	0.1759
	<b>Ethnicity</b> (Non-Asian)	Asian	-0.070	0.105	0.505	0.5	1	0.5050
	<b>Sex</b> (Female)	Male	-0.296 ***	0.044	<.0001	44.8	1	<.0001
	<b>Age</b> (75+ years)	15 to 19 years	-0.158	0.124	0.2063	3.4	7	0.0025
		20 to 24 years	-0.163	0.145	0.2651			
		25 to 34 years	-0.260 **	0.116	0.0278			
		35 to 44 years	-0.366 **	0.117	0.0024			
		45 to 54 years	-0.386 **	0.119	0.0016			
		55 to 64 years	-0.436 ***	0.106	<.0001			
		65 to 74 years	-0.331 **	0.099	0.0012			
	<b>Family type</b> (Couple with children)	Couple without children	0.092	0.067	0.1710	14.8	3	<.0001
Sole parent		-0.329 ***	0.076	<.0001				
Not in a family nucleus		-0.264 ***	0.062	<.0001				
<b>Tenure of household</b> (Dwelling owned)	Dwelling not owned	-0.115 **	0.052	0.0287	6.9	2	0.0016	
	Dwelling held in a family trust	0.109 *	0.058	0.0636				
Knowledge and skills	<b>Highest qualification</b> (Degree or higher level qualification)	None	0.231 **	0.083	0.0062	3.2	5	0.0099
		Level 1	0.226 **	0.094	0.0177			
		Level 2, 3, overseas school	-0.009	0.057	0.8792			
		Others	0.058	0.140	0.6797			
		Levels 4-6 (tertiary below degree)	0.122 **	0.059	0.0428			
Health	<b>Mental health</b> (WHO-5 on a scale from 0-100)		0.030 ***	0.002	<.0001	359.4	1	<.0001
	<b>General health</b> (Excellent)	Very good	0.009	0.059	0.8790	6.4	4	0.0001
		Good	-0.055	0.071	0.4382			
		Fair	-0.263 **	0.099	0.0092			
Poor		-0.595 ***	0.174	0.0009				
Income	<b>Income adequacy</b> (More than enough money)	Not enough money	-0.598 ***	0.129	<.0001	9.2	3	<.0001
		Only just enough money	-0.287 ***	0.079	0.0004			
		Enough money	-0.062	0.050	0.2134			
<b>Material wellbeing</b> (MWI on a scale of 0 to 20)		0.023 ***	0.007	0.0006	12.4	1	0.0006	
Family and friends	<b>Loneliness</b> (None of the time)	A little of the time	-0.136 **	0.052	0.0098	13.2	4	<.0001
		Some of the time	-0.370 ***	0.073	<.0001			
		Most of the time	-0.924 ***	0.161	<.0001			
		All of the time	-0.296	0.213	0.1688			
	<b>Contact with friends</b> (About the right amount of contact)	Too much contact	0.075	0.156	0.6319	3.2	2	0.0454
		Not enough contact	-0.136 **	0.054	0.0136			
	<b>Contact with family</b> (About the right amount of contact)	Don't have a family	-0.193	0.261	0.4604	2.6	3	0.0558
		Too much contact	0.198	0.145	0.1738			
	Not enough contact	-0.117 **	0.057	0.0406				
Work	<b>Labour force status and job satisfaction</b> (Very satisfied with job)	Satisfied	-0.263 ***	0.056	<.0001	9.26	6	<.0001
		No feeling either way	-0.496 ***	0.082	<.0001			
		Dissatisfied	-0.568 ***	0.097	<.0001			
		Very dissatisfied	-0.510 *	0.266	0.0580			
		Not in the labour force	-0.343 ***	0.070	<.0001			
		Unemployed	-0.502 ***	0.142	0.0006			

**Table F1 (continued) – OLS Regression GSS 2018**

Domain	Sub-domain (ref. category)	Category	Estimate	S.E	Pr >  t	F Value	DF	Pr > F
Cultural capability and belonging	<b>Cultural identity</b> (Very easy)	Easy	-0.094 *	0.052	0.0726	5.5	3	0.0016
		Sometimes easy, sometimes hard	0.013	0.071	0.8540			
		Hard/Very hard	-0.687 ***	0.202	0.0010			
Engagement and voice	<b>Trust in people</b> (on a scale from 0-10)		0.036 **	0.017	0.0353	4.6	1	0.0353
	<b>Trust in media</b> (on a scale from 0-10)		-0.028 **	0.014	0.0408	4.3	1	0.0408
	<b>Trust in court</b> (on a scale from 0-10)		0.034 **	0.014	0.0153	6.1	1	0.0153
	<b>Trust in the education system</b> (on a scale from 0-10)		0.046 **	0.015	0.0034	9.0	1	0.0034
	<b>Trust in police</b> (on a scale from 0-10)		0.024	0.016	0.1286	2.4	1	0.1286
	<b>Trust in Parliament</b> (on a scale from 0-10)		0.008	0.014	0.5559	0.4	1	0.5559
	<b>Trust in the health system</b> (on a scale from 0-10)		0.019	0.014	0.1648	2.0	1	0.1648
Safety	<b>Crime committed against you</b> (No)	Yes	0.009	0.061	0.8819	0.0	1	0.8819
	<b>Problem with neighbourhood crime</b> (No)	Yes	-0.030	0.053	0.5690	0.3	1	0.5690
	<b>Feel safe at home by yourself at night</b> (Very safe)	Safe	-0.085 *	0.049	0.0872	1.1	5	0.3914
		Neither safe nor unsafe	-0.125	0.090	0.1674			
		Unsafe	-0.159	0.154	0.3040			
		Very unsafe	-0.531 *	0.304	0.0838			
	Not applicable	0.340	0.378	0.3698				
	<b>Feel safe waiting for public transport at night</b> (Very safe)	Safe	0.011	0.112	0.9228	1.9	5	0.1082
		Neither safe nor unsafe	0.033	0.107	0.7592			
		Unsafe	0.035	0.123	0.7786			
Very unsafe		0.387 *	0.198	0.0533				
Not applicable	0.116	0.095	0.2250					
<b>Feel safe walking alone in your neighbourhood at night</b> (Very safe)	Safe	-0.016	0.078	0.8363	1.3	5	0.2909	
	Neither safe nor unsafe	0.086	0.080	0.2819				
	Unsafe	-0.060	0.125	0.6300				
	Very unsafe	-0.142	0.166	0.3935				
Not applicable	-0.075	0.102	0.4629					
Housing	<b>House condition</b> (No repairs or maintenance needed right now)	Minor maintenance needed	0.005	0.057	0.9325	1.5	3	0.2270
		Some repairs and maintenance needed	-0.067	0.073	0.3665			
		Immediate repairs and maintenance needed	-0.236	0.144	0.1039			
	<b>Cold in winter</b> (No)	Always	0.171 *	0.090	0.0606	1.3	4	0.2908
		Often	-0.019	0.090	0.8384			
		Sometimes	0.044	0.051	0.3917			
		Have not spent a winter living in this house	-0.068	0.103	0.5094			
	<b>Mould problem</b> (No)	Yes	-0.041	0.051	0.4153	0.7	1	0.4153
<b>Crowding</b> (Two or more bedrooms spare)	Two or more bedrooms needed	0.066	0.220	0.7641	0.7	4	0.5697	
	One bedroom needed	0.087	0.130	0.5057				
	No bedrooms needed	0.042	0.063	0.5028				
	One bedroom spare	0.098	0.059	0.0983				
Other	<b>Discrimination</b> (No)	Experienced discrimination	0.050	0.056	0.3777	0.8	1	0.3777
Number of Observations			7917					
R-Square			0.353					