

## 2011 CART SURF → NZ income survey

[http://www.stats.govt.nz/tools\\_and\\_services/microdata-access/nzis-2011-cart-surf.aspx](http://www.stats.govt.nz/tools_and_services/microdata-access/nzis-2011-cart-surf.aspx)

### Data summary

- 29447 observations (synthesized people)
- Eight variables – five categorical, 1 grouped numeric (age bands) and two numeric variables

*NOTE: This is a large data set suitable for exploratory use with Achievement Standards 91035 (1.10), 91036 (1.11), 91264 (2.9), 91581 (3.9) and 91582 (3.10)*

### BACKGROUND ON DATA SET

Classification and regression trees (CART) synthetic unit-record files (SURF)

This data set ...

- Is based on data from the June quarter 2011 NZ Income Survey (NZIS)
- can be treated as a simple random sample from the general New Zealand population aged 15 and over
- contains 29471 records across eight variables (replicating the size of the actual NZIS)
- NOTE – a sub-SURF containing 500 observations is available from Statistics New Zealand at the web address above

### VARIABLES

For each record (synthesised person) the 2011 CART SURF contains the following variables:

Variable name	Variable type	Code	Description
<b>ethnicity</b>	categorical	Maori Pacific Asian MELAA Other NZ European	Ethnicity of individual
<b>region</b>	categorical	Northland Auckland Waikato Bay of Plenty Gisborne Bay Taranaki Manawatu Wellington Nelson Canterbury Otago	Region that the individual currently lives
<b>gender</b>	Categorical	male female	Gender of individual

Variable name	Variable type	Code	Description
<b>agegp</b>	grouped numeric		Age of individual, grouped into <u>five year age-bands</u> from 15 years to 65 plus years
<b>qualification</b>	categorical	none school vocational degree other	Highest qualification of individual
<b>occupation</b>	categorical	manager professional technician community clerical sales machinery labourer other none	Category of main income source of individual (based on ANZSCO Level 1)
<b>hours</b>	numeric		weekly hours worked from all wages and salary jobs <u>excluding self-employment</u>
<b>income</b>	numeric		gross weekly income from <u>all</u> sources

## Interpreting the variables

Ethnicity information has been collapsed to one **Ethnicity** variable using the Ministry of Education priority reporting for student ethnicity system ([http://www.educationcounts.govt.nz/data-services/collecting-information/code\\_sets/ethnic\\_group\\_codes](http://www.educationcounts.govt.nz/data-services/collecting-information/code_sets/ethnic_group_codes)). The MoE priority order of ethnicities is

1. Maori
2. Pacific
3. Asian
4. MELAA (Middle Eastern, Latin American, African)
5. Other
6. NZ European

This method of reducing ethnicity is unlikely to result in a variable that represents the ethnic diversity and structure of the New Zealand population. For further details see:

[http://www.stats.govt.nz/browse\\_for\\_stats/population/census\\_counts/review-measurement-of-ethnicity/~media/Statistics/browse-categories/population/census-counts/review-measurement-ethnicity/understanding-working-ethnicity-data.pdf](http://www.stats.govt.nz/browse_for_stats/population/census_counts/review-measurement-of-ethnicity/~media/Statistics/browse-categories/population/census-counts/review-measurement-ethnicity/understanding-working-ethnicity-data.pdf)

The **income** variable is gross weekly income from all sources, not just wages and salary. This means that if you divide income by hours you do **not** get hourly rates for jobs. Losses are recorded as negative values.

The **hours** variable excludes self-employment. This means you cannot assume that everyone who is recorded as working zero hours is either unemployed or not in the labour force.