

**2011 CART SURF 🡪 NZ income survey**

<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)*

**OVERVIEW**

This data set is based on data from the June quarter 2011 New Zealand Income Survey (NZIS). More details on the background of the data set can be found at the [end](#background) of this document. Variable details are given on the following pages.

**VARIABLES**

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

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

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **Variable type** | **Code** | **Description** |
| **agegp\_cat** | categorical | |  | | --- | | 15 – 20 yrs | | 20 – 25 yrs | | 25 – 30 yrs | | 30 – 35 yrs | | 35 – 40 yrs | | 40 – 45 yrs | | 45 – 50 yrs | | 50 – 55 yrs | | 55 – 60 yrs | | 60 – 65 yrs | | 65+ yrs | | Age of individual, grouped into five year age-bands from 15 years to 65 plus years. |
| **agegp\_mid** | grouped numeric |  | Age of individual, grouped into five year age-bands from 15 years to 65 plus years. Midpoint of age-band given as shown below   |  |  | | --- | --- | | Age band | midpoint | | 15 – 20 yrs | 18 | | 20 – 25 yrs | 23 | | 25 – 30 yrs | 28 | | 30 – 35 yrs | 33 | | 35 – 40 yrs | 38 | | 40 – 45 yrs | 43 | | 45 – 50 yrs | 48 | | 50 – 55 yrs | 53 | | 55 – 60 yrs | 58 | | 60 – 65 yrs | 63 | | 65+ yrs | 68 | |
| **agegp\_cat\_10yrs** | categorical | |  | | --- | | 15 – 25 yrs | | 25 – 35 yrs | | 35 – 45 yrs | | 45 – 55 yrs | | 55 – 65 yrs | | 65+ yrs | | Age of individual, grouped into ten year age-bands from 15 years to 65 plus years. |
| **agegp\_mid\_10yrs** | grouped numeric |  | Age of individual, grouped into ten year age-bands from 15 years to 65 plus years. Midpoint of age-band given as shown below   |  |  | | --- | --- | | Age band | midpoint | | 15 – 25 yrs | 20 | | 25 – 35 yrs | 30 | | 35 – 45 yrs | 40 | | 45 – 55 yrs | 50 | | 55 – 65 yrs | 60 | | 65+ yrs | 70 | |
| **qualification** | categorical | none  school  vocational  degree  other | Highest qualification of individual |
| **occupation** | categorical | manager  professional  technician  community  clerical  sales  machinery  labourer  other  none | Category of main income source of individual (based on ANZSCO Level 1) |
| **hours** | numeric |  | weekly hours worked from all wages and salary jobs excluding self-employment |
| **income** | numeric |  | gross weekly income from all 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>). 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.

Please note that the prioritising of ethnicity is not preferred practice by Statistics New Zealand, but in this case helps simplify the double-counting problem for the data set to be suitable for teaching. 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>

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.

**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