

**Canada**

**Statistics Canada**

**2006 Census of Population [Canada] Public Use  
Microdata File (PUMF): Individuals File (version 2)**

**Study Documentation**

June 28, 2018

# Metadata Production

<b>Metadata Producer(s)</b>	Social Science Data Centre (SSDC) , Queen's University
<b>Production Date</b>	June 27, 2010
<b>Version</b>	SP XML Version 2.0
<b>Identification</b>	pumf-95M0028-E-2006-individuals

# Table of Contents

<a href="#">Overview.....</a>	<a href="#">4</a>
<a href="#">Scope &amp; Coverage.....</a>	<a href="#">4</a>
<a href="#">Producers &amp; Sponsors.....</a>	<a href="#">6</a>
<a href="#">Sampling.....</a>	<a href="#">6</a>
<a href="#">Data Collection.....</a>	<a href="#">6</a>
<a href="#">Data Processing &amp; Appraisal.....</a>	<a href="#">7</a>
<a href="#">Accessibility.....</a>	<a href="#">7</a>
<a href="#">Rights &amp; Disclaimer.....</a>	<a href="#">8</a>
<a href="#">Files Description.....</a>	<a href="#">9</a>
<a href="#">pumf_95M0028_E_2006_individuals_F1.....</a>	<a href="#">9</a>
<a href="#">Variables Group(s).....</a>	<a href="#">10</a>
<a href="#">Demography.....</a>	<a href="#">10</a>
<a href="#">Variables Description.....</a>	<a href="#">11</a>
<a href="#">pumf_95M0028_E_2006_individuals_F1.....</a>	<a href="#">12</a>

**2006 Census of Population [Canada] Public Use Microdata File (PUMF):  
Individuals File (version 2) (2006 Census PUMF: Individuals File)**  
*Fichier des particuliers, (fichiers de microdonnées à grande diffusion) du Recensement de 2006*

Overview	
Type	Census of Population
Identification	pumf-95M0028-E-2006-individuals
Version	<p>Production Date: 2010-05-27 Version May 27, 2010</p> <p><u>Notes</u></p> <p>Revised data file and SAS file were received from Statistics Canada to address problems with the following two variables:</p> <p>Variable MFS&lt;br /&gt; A discrepancy involving labelling in the PUMF dataset and codebook has been detected. The 'no postsecondary qualifications' category of the MFS – 'Major Field of study' (based on the MFS Classification – Historical) variable is reported as having the value 11 in the codebook, while the 'no postsecondary qualifications' category is represented by the value 12 in the dataset.</p> <p>Variable ATTSCH&lt;br /&gt; A discrepancy in the SAS program source code has been reported. The SAS program source code contains the mnemonic 'SCHATT' to describe 'attendance at school,' the proper mnemonic for 'attendance at school' is 'ATTSCH.'</p> <p>Both of these issues have been corrected in Version 2</p>
Series	<p>The Census Public Use Microdata Files (PUMF) contain samples of anonymous responses to the Census questionnaire. The files have been carefully scrutinized to ensure the complete confidentiality of the individual responses. PUMFs enable the development of statistical information about Canadians, the families and households to which they belong, and the dwellings in which they live. Microdata files are unique among census products in that they give users access to non-aggregated data. Each file contains anonymous individual responses on a large number of variables. The PUMF user can group and manipulate these variables to suit his/her own data and research requirements. Tabulations not included in other census products can be created, or relationships between variables can be analysed using different statistical tests. PUMFs provide quick access to a comprehensive social and economic database about Canada and its people. However, to ensure the anonymity of the respondents, geographic identifiers have been restricted to the provinces/territories and large metropolitan areas.</p>
<p><b>Abstract</b></p> <p>The 2006 Census public use microdata file (PUMF) on individuals contains 844,476 records, representing 2.7% of the Canadian population. These records were drawn from a sample of one-fifth of the Canadian population (sample data from questionnaire 2B). The 2006 PUMF includes 123 variables. Of these, 102 variables, or 83%, come from the individual universe and 21 variables, or 17%, are drawn from the family, household and dwelling universes. The file does not include people living in institutions.</p> <p>Since the 1971 Census, Statistics Canada has traditionally produced three public use microdata files: the Individual File, the Family File and the Household and Dwelling File. To meet users' needs and allow international comparison of PUMFs, Statistics Canada has decided to produce two files for the 2006 Census: The Individual File and the Hierarchical File. The latter file will contain combined data from the family, household and dwelling universes.</p>	
Kind of Data	Census data
Unit of Analysis	Individual

## Scope & Coverage

### Scope

Reduced level of detail

- this product contains only information on the largest census metropolitan areas and the provinces. Yukon Territory, the Northwest Territories and Nunavut are grouped under the term 'Northern Canada.'
- the data have been aggregated in such a way as to preserve confidentiality while, at the same time, providing as much detail as possible in order to maintain the analytical value of the file. For example, the data on occupation do not indicate 'physician', but rather the more general category 'occupations in medicine and health.' This category also includes other medical occupations, such as 'nurse.'

Data not available<br /> For a few records, the codes for certain variables were changed to indicate Not available, so as to guarantee data confidentiality. Users must make sure to exclude them from their calculations.

Lower and upper income limits<br /> The PUMF contains lower and upper income limits. Thus, the data on total income and sources of income are adjusted proportionally.

#### New Features

The content of the 2006 PUMF is largely the same as that of the 2001 PUMF. However, various changes should be noted, resulting from new questions in the 2006 Census and more generally from improvement of the content of the file. Note that the 2006 PUMF does not contain, as in 2001, variables with two levels of content: more detailed content for Quebec, Ontario and the West and less detailed content for the Atlantic provinces and the territories. Because the duplication of variables did not entail an increase in content, duplicate variables were eliminated from the 2006 PUMF and replaced by a single variable with content for all of Canada.

- Changes in the content of the 2006 Census questionnaire<br /> New variables were inserted to reflect the content of the 2006 Census questionnaire
- Income<br /> For some respondents, Statistics Canada has income data obtained directly from Canada Revenue Agency. As a result, the 2006 PUMF now includes the variables 'after-tax income' and 'income tax paid.'
- Schooling<br /> The 2B questionnaire for the 2006 Census contains two new questions on the location of study and the major field of study based on the Classification of Instructional Programs (CIP) Canada 2000. The 2006 PUMF includes these new variables, which indicate, firstly, the province or territory (in Canada) or the country (outside Canada) where the highest certificate, diploma or degree was obtained and secondly, the major field of study.
- Religion<br /> The question on religion was not asked in the 2006 Census.
- Improvement of the content of the file<br /> Change in the content of certain variables
- Geography<br /> In 2006, four new census metropolitan area groupings were added: 'Brantford - Guelph - Barrie,' 'Kingston - Peterborough,' 'Kelowna - Abbotsford' and 'Moncton - Saint John.'
- Ethnic origin<br /> The variable 'ethnic origins (EthDer),' derived from single responses and some multiple responses, includes six categories more than the 2001 variable ETHNI. The multiple response categories remain the same, but six single response categories were added for ethnic origins that showed a sizable increase in 2006.
- Non-official languages<br /> The 2006 variable 'knowledge of non-official languages - derived - single responses and multiple responses' replaces the 15 'Knowledge of...' variables used in 2001. The 2006 version identifies 25 different single response categories and one multiple response category for non-official languages.
- Other new variables in the 2006 PUMF<br />
- Labour market activities. This variable (WrkAct) indicates the number of weeks in which persons worked for pay or in self-employment, in 2005 (see details in the section labour market activities).
- Journey to work. This section includes a new variable, province of work (PWPR). It represents the geographic location of the place of work of persons aged 15 and over, excluding institutional residents, who worked since January 1, 2005 (see detailed definition in the section journey to work).
- Visible minority. The variable VisMinH (visible minority population, historical) is inserted in the 2006 PUMF.
- Estimation. In addition to the weighting variable (WEIGHT), the 2006 PUMF has eight weighting factors for replicates, for purposes of estimating sampling variability (WT1-WT8).

<b>Keywords</b>	Aboriginal, Age, Birthplace, Citizenship, Commuting, Dwelling, Education, Employment, Ethnic origin, Family, First Nation, Full time work, Government income, Household, Immigration, Income, Investment, Labour force, Language, Marital status, Mobility, Mother tongue, Part time work, Pension, Self-employment, Shelter, Unpaid work, Visible minority, Wages
<b>Topics</b>	Education, Labour, Ethnic diversity and immigration, Population and demography, Income, pensions, spending and wealth, Families, household and housing, Languages, Society and community

<b><u>Time Period(s)</u></b>	2006
<b><u>Countries</u></b>	Canada
<b><u>Geographic Coverage</u></b> Canada, Provinces, Territories, CMA	
<b><u>Universe</u></b> <p>Included: all Canadian citizens and landed immigrants who have a usual place of residence in Canada or who are abroad, either on a military base or on a diplomatic mission. The file also includes data on non-permanent residents of Canada, that is, persons who hold a student authorization or an employment authorization or a minister's permit, or who are refugee claimants, and members of their family living with them.</p> <p>Excluded: institutional residents, residents of incompletely enumerated Indian reserves or Indian settlements, and foreign residents (foreign diplomats, members of the armed forces of another country stationed in Canada and residents of another country temporarily visiting Canada).</p>	

<b>Producers &amp; Sponsors</b>	
<b>Primary Investigator(s)</b>	Statistics Canada
<b>Other Producer(s)</b>	Statistics Canada

<b>Sampling</b>
<b><u>Sampling Procedure</u></b> <p>The microdata sample for individuals is selected using a three-phase sampling plan. The first sampling phase consists of the sample of one-fifth of the population (20% sample data). This is a cluster sample. It consists of all households who completed the long questionnaire in the census. This sample was divided into two parts representative of Canada in order to create two sampling frames used to select the microdata samples. The first frame was used to select microdata from the individuals file. The second frame was used to select microdata from the hierarchical file. The third phase consisted in selecting records from the individuals file. The final sample represents 2.7% of the target universe.</p>
<b><u>Weighting</u></b> <p>WEIGHT = weighting of individuals</p> <p>The microdata file contains a record for each selected unit in the sample. Each record contains a certain number of characteristics or variables described in Chapter 2. Therefore, each of these units represents, on average, many other units that are not part of the sample. To represent all these other units in the estimation process, the file contains a variable called 'WEIGHT' (weighting factor for individuals), which corresponds to the number of units (including the selected unit) represented by each record in the file. WEIGHT still has the same value: 36.99457415.</p> <p>The weighting factor therefore indicates the number of times a record must be repeated to obtain population estimates. For example, to estimate the number of persons who speak Chinese at home in Canada in the target universe, it is necessary to total the weighting factors of all records belonging to this category in the file.</p>

<b>Data Collection</b>	
<b>Data Collection Dates</b>	start 2006-05-16 end 2006-05-16
<b>Time Period(s)</b>	start 2006-05-16 end 2006-05-16
<b>Data Collection Mode</b>	Paper questionnaire and online questionnaire (<a href="http://www.census2006.ca" target=new">www.census2006.ca</a>)

## Data Processing & Appraisal

### Other Processing

Users must refrain from publishing unweighted tables and from conducting analyses based on unweighted data from the microdata file. They must also make sure to exclude from their calculations all values that are unavailable or not applicable.

### Estimates of Sampling Error

The 'coefficient of variation' is a measure frequently used to determine the degree of sampling variability. This is simply the relationship of the standard error of an estimate to the value of that estimate or, in other words, the standard error expressed as a percentage of the targeted estimate.

The sampling plan must be taken into account in computing the sampling error. The Individuals File does not contain all the necessary information. In order to estimate this sampling error, we propose an approximate method called the 'random groups method.' This method, which is described in detail in Chapter 2 of the book *Introduction to Variance Estimation* (Wolter, K. M., *Introduction to Variance Estimation*, Springer Series in Statistics, Springer-Verlag, New York, 1985.), is easy to apply. One of its features is that it tends to overestimate the sampling error for small estimates. This results in a conservative procedure for testing significant differences.

The principle is as follows: the sample was divided into eight replicates, each representative of the sample. These replicates or portions are defined by their weighting factors, WT1, WT2, ... , WT8, for example, the fourth replicate is the set of records for which WT4 is greater than 0. The values for a given replicate weighting factor is 0 if the record is not part of the replicate for this factor or  $8 * \text{WEIGHT}$  (eight times the value of the weighting factor).

After calculating the desired estimate with all records as in Section B.2, the following calculations are required:

- Recalculate the same estimates but this time based only on the different replicates. This yields eight different estimates based on the same concept.
- Calculate the average of the eight estimates. When the estimate is based on a limited number of records, some replicates may be empty. In this case, assign the value 0.
- Calculate the sum of the squared deviations between the estimates for the replicates and the average obtained in (2).
- Divide the number obtained in (3) by 56 and extract the square root of the result. The number resulting from this operation is an estimate of the standard error of the targeted estimate.
- Divide the number obtained in (4) by the targeted estimate. The result is the coefficient of variation.
- One can calculate a confidence interval that would be accurate 19 times out of 20 by removing the standard error twice from the targeted estimation for the lower bound and by adding the standard error twice to the targeted estimation for the upper bound.

## Accessibility

<b>Access Authority</b>	Data Liberation Initiative (DLI) (Statistics Canada) , <a href="http://www.statcan.gc.ca/dli-ild/dli-idd-eng.htm">http://www.statcan.gc.ca/dli-ild/dli-idd-eng.htm</a>
<b>Contact(s)</b>	Data Liberation Initiative (DLI) (Statistics Canada) , <a href="http://www.statcan.gc.ca/dli-ild/dli-idd-eng.htm">http://www.statcan.gc.ca/dli-ild/dli-idd-eng.htm</a>
<b>Distributor(s)</b>	Data Liberation Initiative

### Access Conditions

DLI License

### Citation Requirements

The publishing of analysis and results from research using any of the data products is permitted in research communications such as scholarly papers, journals and the like. The authors of these communications are required to cite Statistics Canada as the source of the data, and to indicate that the results or views expressed are those of the author/authorized user and are not those of Statistics Canada. Permission to include extracts of these data in textbooks must be obtained from the Licencing Section of Statistics Canada's Marketing Division. (from Section 9, DLI License Agreement)

Rights & Disclaimer	
Copyright	Copyright (c) Statistics Canada, 2010



# Files Description

Dataset contains 1 file(s)

pumf_95M0028_E_2006_individuals_F1	
# Cases	844476
# Variable(s)	124
<p><b><u>Version</u></b>  Version May 27, 2010 Statistics Canada  Revised data file and SAS file were received from Statistics Canada to address problems with the following two variables:  Variable MFS&lt;br /&gt; A discrepancy involving labelling in the PUMF dataset and codebook has been detected. The 'no postsecondary qualifications' category of the MFS – 'Major Field of study' (based on the MFS Classification – Historical) variable is reported as having the value 11 in the codebook, while the 'no postsecondary qualifications' category is represented by the value 12 in the dataset.  Variable ATTSCH&lt;br /&gt; A discrepancy in the SAS program source code has been reported. The SAS program source code contains the mnemonic 'SCHATT' to describe 'attendance at school,' the proper mnemonic for 'attendance at school' is 'ATTSCH.'  Both of these issues have been corrected in Version 2</p>	
<p><b><u>Processing Checks</u></b>  Missing values were declared  Values were added for the following variables:&lt;br /&gt; <ul style="list-style-type: none"> <li>• CHDBN</li> <li>• CQPPB</li> <li>• EICBN</li> <li>• EMPIN</li> <li>• GOVTI</li> <li>• GROSRT</li> <li>• GTRFS</li> <li>• HRSWRK</li> <li>• INCTAX</li> <li>• INVST</li> <li>• MRKINC</li> <li>• OASGI</li> <li>• OMP</li> <li>• OTINC</li> <li>• RETIR</li> <li>• SEMPI</li> <li>• TOTINC</li> <li>• TOTINC_AT</li> <li>• VALUE</li> <li>• WAGES</li> <li>• WKSWRK</li> </ul> </p>	

# Variables Group(s)

Dataset contains 18 group(s) total - showing a subset of 1

Group Demography							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	AGEGRP	[AGEGRP] Age groups	discrete	numeric-2.0	843676	800	Date of Birth

# Variables Description

**Dataset contains 124 variable(s) total - showing a subset of 1**

# File : pumf\_95M0028\_E\_2006\_individuals\_F1

## # AGEGRP: [AGEGRP] Age groups

<b>Information</b>	[Type= discrete] [Format=numeric] [Range= 1-21] [Missing=*/88]
<b>Statistics [NW/ W]</b>	[Valid=843676 / 31211434.34 ] [Invalid=800 / 29595.659 ]
<b>Universe</b>	Total population, excluding institutional residents
<b>Literal question</b>	Date of Birth
<b>Interviewer's instructions</b>	If exact date is not known, enter best estimate
<b>Notes</b>	Refers to the age at last birthday (as of the census reference date, May 16, 2006). This variable is derived from date of birth. Derived variable: Question 3

Value	Label	Cases	Weighted	Percentage (Weighted)
1	0 to 4 years	45815	1694906.4	5.4%
2	5 to 6 years	18869	698050.6	2.2%
3	7 to 9 years	30056	1111908.9	3.6%
4	10 to 11 years	21910	810551.1	2.6%
5	12 to 14 years	34492	1276016.9	4.1%
6	15 to 17 years	35128	1299545.4	4.2%
7	18 to 19 years	22090	817210.1	2.6%
8	20 to 24 years	55920	2068736.6	6.6%
9	25 to 29 years	53356	1973882.5	6.3%
10	30 to 34 years	54559	2018387.0	6.5%
11	35 to 39 years	59427	2198476.6	7.0%
12	40 to 44 years	70380	2603678.1	8.3%
13	45 to 49 years	69950	2587770.5	8.3%
14	50 to 54 years	63340	2343236.3	7.5%
15	55 to 59 years	55902	2068070.7	6.6%
16	60 to 64 years	42618	1576634.8	5.1%
17	65 to 69 years	32902	1217195.5	3.9%
18	70 to 74 years	27797	1028338.2	3.3%
19	75 to 79 years	23052	852798.9	2.7%
20	80 to 84 years	15724	581702.7	1.9%
21	85 years and over	10389	384336.6	1.2%
88	Not available	800	29595.7	

*Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.*