# 2011 and 2006 Canadian Census Tables

The 2011 Canadian census is distributed by Statistics Canada [in CSV format for various geographic region types](http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/details/download-telecharger/comprehensive/comp-csv-tab-dwnld-tlchrgr.cfm?Lang=E). The 2006 Canadian census is archived [in Beyond 20/20 format](http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/prof/92-591/index.cfm?Lang=E). Dissemination area granularity is delivered on request.

Geographic boundary files are also distributed by Statistics Canada [in GML format](http://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-eng.cfm). Currently the geographic boundaries are imported and tied to the 2011 census information. Tying 2006 geographic boundaries to 2006 census data will be performed in the near future.

The 2006 census contained more questions on the form and is more intriguing. The 2011 census is more recent and applicable to current datasets. The intention of mapping and importing these datasets into similar MySQL schemas is that both may be more accessible for our use. It should be noted that a degree of inaccuracy is artificially injected into this data by Statistics Canada to protect people’s privacy. The rounding process is [described by Statistics Canada](http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/help-aide/N2.cfm?Lang=E). These tables are not an exhaustive import of the census, only the tables which seemed immediately relevant for energy systems. More tables can be imported upon request.

# Glossary of Terms

***Forward Sortation Area (FSA)*** – The first three characters of a Canadian postal code are known as the forward sortation area. These are fairly large regions typically with populations in the scale of thousands. [Details of the FSA structure are available](http://en.wikipedia.org/wiki/Forward_sortation_area#Forward_sortation_areas).

***Dissemination Area (DA)*** – The census organizes the distribution of its form according to dissemination area and is smallest type of geographic region available. [A more complete definition and the structure of the dissemination area ID is provided by Statistics Canada](http://www.statcan.gc.ca/pub/92-195-x/2011001/geo/da-ad/def-eng.htm).

**Geographic Markup Language (GML)** – It is a common geographic information systems (GIS) file type which describes the layout of regions and other descriptors. There are many nuances to the file type which this table structure simplifies into coordinates of longitude/latitude.

# Schema Description

**census2006da** – 2006 census results grouped by dissemination area

**census2011da** – 2011 census results grouped by dissemination area

**census2011fsa** – 2011 census results grouped by forward sortation area

# Table Descriptions

Notable characteristics of tables will be described below. This is not intended to be an exhaustive description of each table and its columns. It simply outlines the details that may not be immediately obvious to a technical user.

**dissemination\_area** – Describes a dissemination area. Notable columns outlined below.

* da\_id – An eight-digit number which uniquely describes a DA. Statistics Canada defines the structure of this number, linked in Glossary of Terms. It use used as the primary key (PK) of dissemination\_area.
* data\_quality\_flags – The 2006 census contained several data quality flags for each DA’s results. Complete descriptions are in the Appendix. In future modifications to the schema, these may be more explicitly defined and parsed into new tables.

**polygon\_patch** – An artifact of the geographic boundaries being disseminated via GML file, the polygon nomenclature was maintained in the structuring of this database. An FSA or DA may be comprised of many separate polygons, illustrated in Figure 1. Together these “patches” may be grouped into a set of polygons which comprise the geographic region.

* polygon\_patch\_id – Primary key of the table arbitrarily created to uniquely identify rows.
* da\_id – Foreign key tied to dissemination\_area.da\_id which supports the one-to-many relationship between dissemination\_area and polygon\_patch.

**coordinate** – A polygon\_patch is made up of many coordinates which comprise the border of a polygon (ie. geographic region part). A polygon can be comprised of 10-1,000+ coordinates.

* coordinate\_id – Primary key of the table created to uniquely identify rows. Ordering is important because lines are drawn between coordinates to draw the border of a polygon. Each border will start and end at the same coordinate.
* path\_type – A polygon\_patch may have an exterior border and an interior border. The interior border cuts a region out of the exterior border (ie. a “doughnut hole”). This relationship is illustrated in Figure 1.

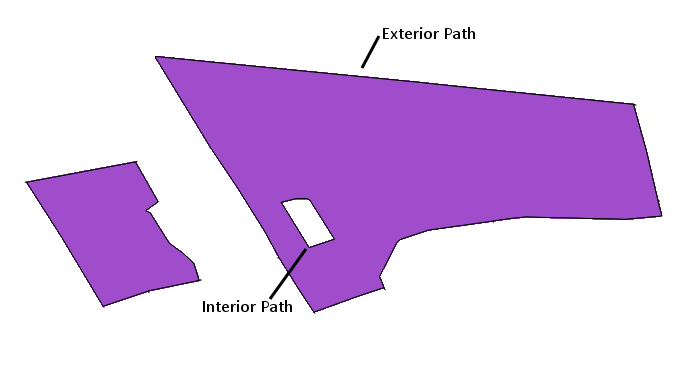


Figure 1. Two polygon\_patches make up a geographic region. One polygon\_patch has both an exterior path\_type of coordiantes and an interior path\_type of coordiantes.

**age** – Census question, “Total population by age groups.” 2011 ages were given as men and women combined. 2006 ages were only given as separate men and woman totals which were summed during import into the common MySQL schema.

**children\_at\_home** – Census question, “Total number of children at home.”

**dwelling** – Census question, “Total number of occupied private dwellings.”

**household\_income** – Census question, “Household income in 2005 of private households.” It should be noted that many other categorizations of income existed in the 2006 census including splits by gender, individual income, after-tax income, and more. This data may be parsed into the tables at a future date (ie. if requested).

**household\_size** – Census question, “Total number of private households by household size.” In the 2006 census the possible answers were 1, 2, 3, 4-5, 6+. In the 2011 census the possible answers were 1, 2, 3, 4, 5, 6+.

* persons\_4 – For the 2006 census this column is blank.
* persons\_5 – For the 2006 census this column is blank.
* persons\_4\_to\_5 – For the 2006 census the response to the question was stored. For the 2011 census the sum of persons\_4 and persons\_5 have been stored in persons\_4\_to\_5 for convenience even though it is redundant.

**household\_type** – Census question, “Total number of private households by household type.”

**labour\_force\_activity** – Census question, “Total population 15 years and over by labour force activity.”

**marital\_status** – Census question, “Total population 15 years and over by legal marital status.”

**size\_of\_families** – Census question, “Size of census family.”

**structural\_type** – Census question, “Total number of occupied private dwellings by structural type of dwelling.”

**work\_commute\_transportation** – Census question, “Total employed labour force 15 years and over with usual place of work or no fixed workplace address by mode of transportation.”

# Appendix

## Data Quality Flags (Pulled from Beyond 20/20 File)

2006 Census – Data quality flags and descriptions  
  
Data quality flags for place of residence  
Digit Description  
1st (0XXXX) Incomplete enumeration flag  
2nd (X0XXX) 100% data quality flag  
3rd (XX0XX) Population and dwelling counts error flag  
4th (XXX0X) 20% sample data quality flag  
5th (XXXX0) 2001 adjusted population flag  
  
Flag description  
Incomplete enumeration flag  
0 = Default.  
1 = Incompletely enumerated Indian reserve or Indian settlement (suppressed).  
2 = Excludes census data for one or more incompletely enumerated Indian reserves or Indian settlements.  
  
100% data quality flag  
0 = Default.  
1 = Data quality index showing, for the short census questionnaire (100% data), a global non-response rate higher than or equal to 5% but lower than 10%.  
2 = Data quality index showing, for the short census questionnaire (100% data), a global non-response rate higher than or equal to 10% but lower than 25%.  
3 = Data quality index showing, for the short census questionnaire (100% data), a global non-response rate higher than or equal to 25% (suppressed).  
  
Population and dwelling counts error flag  
0 = Default.  
1 = An error exists in the 2006 population and dwelling counts for this area. For further details, please refer to the population and dwelling counts data section of the 'Notes' file.  
2 = In 2001, the population and/or dwelling counts for this census subdivision were found to be incorrect. Since it is not possible to make changes to the 2001 Census data presented in these tables, the 2001 data should be used with caution. For further details, please refer to the population and dwelling counts data section of the 'Notes' file.  
3 = Both the 2006 and 2001 population and/or dwelling counts for this area were found to be incorrect. Since it is not possible to make changes to the census data presented in these tables, these counts should be used with caution. For further details, please refer to the population and dwelling counts data section of the 'Notes' file.  
  
20% sample data quality flag  
0 = Default.  
1 = Data quality index showing, for the long census questionnaire (20% sample data), a global non-response rate higher than or equal to 5% but lower than 10%.  
2 = Data quality index showing, for the long census questionnaire (20% sample data), a global non-response rate higher than or equal to 10% but lower than 25%.  
3 = Data quality index showing, for the long census questionnaire (20% sample data), a global non-response rate higher than or equal to 25% (suppressed).  
  
2001 adjusted population flag  
0 = Default.  
1 = 2001 adjusted count; most of these are the result of boundary changes.  
Note: The 100% and 20% sample data quality flags do not apply to the population and dwelling counts.  
  
Data quality flag for place of work  
Digit Description  
4th (XXX0X) 20% sample data quality flag  
  
Flag description  
20% sample data quality flag  
0 = Default.  
3 = Data quality index showing, for the long census questionnaire (20% sample data); a global non-response rate higher than or equal to 25% (suppressed); and most of the employed labour force working in the area, also reside in the same area.