



Statistics Canada Data Tables Information Sheet

[Caveat](#) - this information / data has been assembled by Data for Good volunteers, from Statistics Canada(StatCan) website plus other sources. We hope you will find this a useful reference!

[Change Log](#) - see Appendix B, for change log, we will keep this document evergreen

Statistics Canada Data Tables Information Sheet	1
Useful Links	Error! Bookmark not defined.
Statistics Canada Data Tables	4
What is Accessible as Data?	4
How to Search Data?	5
Downloading, visualizing	6
CSV/TAB	6
SDMX (XML) (SDMX-ML)	6
Web Data Service (for developers)	6
Program Packages (for developers)	6
Data Visualization	7
Geographic Levels	7
Focus on Homelessness	7
What we will work with from Micro-Datathon #2	7
Tips & Best practices	9
Citing References	9
How to ask StatCan a question	10

Background

Data for Good Calgary has had a few online, do at your own pace datathons and has called them “Micro Datathons”. There was one related Mini Datathon that took place in 2020 and is referenced many times here because it deals with Statistics Canada Census data which is a related and can be used to create Data Tables and there are data sets in that datathon that we will utilize and bring into our workshop or challenges for this collaboration.



To share all of our information and have it accessible in one spot we will reference and provide links to data.world and relevant projects/datasets data for good has shared on that platform. These documents are public and open however the platform may require you to sign-up to view detail and at no cost.

The Challenge instruction sheet will reiterate how to sign-up if you are unfamiliar with using data.world.

Useful Links

StatCan Home or start page 2016:

<https://www.statcan.gc.ca/eng/start>

The main page for data search and retrieval can be found here:

<https://www150.statcan.gc.ca/n1/en/type/data?MM=1>

The site has been re-designed recently and an explained video on how to navigate the pages can be viewed from the prompt on the “Data” page under “What’s new” and the link is given below (accessed on Jan 11 2021 - it may change in the future):

<https://www.statcan.gc.ca/eng/sc/video/new>

This video is a great, just over 5 minute video on using the new Data Table web pages to search and download information:

Wilfred Laurier University Library. (2018). *Finding statistics using StatCan's data portal* [Video file]. Retrieved from https://youtu.be/azTz_rVRsZQ.

Census Program Geography. These “shapes” are also used to make Data Tables at certain geographical detail.

[Census geography \(statcan.gc.ca\)](https://www.statcan.gc.ca/eng/sc/geography)

January 28 2021 Presentation by Ian (links and video), to Data for Good Calgary (Video):

slides:



General description of “Statistics” and good sections on Census and StatCan Data Tables (formerly CANSIM) from Kwantlen Polytechnic University

[Start here - Statistics - LibGuides at Kwantlen Polytechnic University \(kpu.ca\)](https://kpu.ca/libguides/statistics)

(Statistics LibGuide by Chris Burns (from KPU Library) is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.)

Information on Data Table (CANSIM) Vectors. This resource provides a general description but is mostly an R programming resource from Mountain Maith

[Retrieving individual Statistics Canada vectors • cansim \(mountainmath.github.io\)](https://mountainmath.github.io/cansim/)

Statistics Canada Page to search by Vector when that information is known:

[Search by vector \(statcan.gc.ca\)](https://statcan.gc.ca/search-by-vector)

Concordance between old CANSIM table numbers and new ProductID associated with Data Tables from Statistics Canada:

[Concordance between CANSIM table numbers and the new Product ID \(statcan.gc.ca\)](https://statcan.gc.ca/concordance-between-cansim-table-numbers-and-the-new-product-id)

Python StatsCan Package related

A python library for reading data from Statistics Canada

[A python library for reading data from Statistics Canada — stats can documentation \(stats-can.readthedocs.io\)](https://statscan.readthedocs.io/en/latest/)

The Python library allows you to check data table release information and this is detail behind the releases:

[Developers \(statcan.gc.ca\)](https://statcan.gc.ca/developers)

Statistics Canada Developers Guide including Data Table metadata description

[Full Table Download \(CSV\) User Guide \(statcan.gc.ca\)](https://statcan.gc.ca/full-table-download-csv-user-guide)

The Python libraries make data table retrieval easy. These are the Web Data Service REST API information:

[Web Data Service \(WDS\) User Guide \(statcan.gc.ca\)](https://statcan.gc.ca/web-data-service-wds-user-guide)

Data Specific web reference Pages

Real-time CANSIM tables from Statistics Canada site:

[Real-time CANSIM tables \(statcan.gc.ca\)](https://statcan.gc.ca/real-time-cansim-tables)



For coders:

Statistics Canada github site:

[Statistics Canada - Statistique Canada \(github.com\)](https://github.com/StatisticsCanada)

Statistics Canada Data Tables

We have talked in Microdatathon #2 about the Canadian Census and explored data from it. The Census occurs every five years. Data Tables takes information from ongoing data sources which include origins like:

- Surveys
- Administrative data gathered by Statistics Canada like
 - Data gathered from other government agencies
 - CRA (Canadian Revenue Agency) tax return information
 - Vital statistics from Provincial governments
 - From agreements with other non-governmental data providers

These tables often involve more than one dimension and are produced for a certain geographic granularity.

What is Accessible as Data?

There are categories of data:

1. Tables - these are probably what we are most interested in as data analysts/scientists. They have a ProductID (old CANSIM number) reference and are associated with a Geography.
2. Profiles of a community or region - these seem to include open databases for specific domains. Examples of some of these are:
 - a. Various products from Census can be found here. The Census Profile , 2016 is classified as one of these and has a Catalogue no. reference and in this case its own web developer data service.
 - b. "The Open Database of Healthcare Facilities" and is an example of data released under the Open Government License - Canada.
 - c. There are 280 items and they can contain information different from a and b. Perhaps we need to explore this category and learn more about it?

3. Thematic Maps - These are map based representations and include PDF and HTML versions and could link to computer files to enable mapping. The Census 2016 is also referenced here and we have used the map shapes for Micro-Datathon #2 for example.
4. Public use microdata - Non-aggregated data associated with certain topic areas. There may have been many recently released for public availability. "Crowdsourcing: Impacts of COVID-19 on Canadians' Experiences of Discrimination Public Use Microdata File" ([Crowdsourcing: Impacts of COVID-19 on Canadians' Experiences of Discrimination Public Use Microdata File \(statcan.gc.ca\)](#)) is a product delivered by "EFT" - Statistics Canada's electronic file transfer service. This is a group of data that data for good might explore at a later date?
5. Data Visualization - These include Visualizations created by various tools and embedded into web pages. A plotly example - [Weekly death counts: Interactive tool \(statcan.gc.ca\)](#) and a Power BI example - [Labour Force Survey in brief: Interactive app \(statcan.gc.ca\)](#). There is quite a variety of tools used to create these visuals and now with the new web organization and search it might be a good visualization exercise to explore and learn from the variety of displays that are accessible here.

How to Search Data?

Before using programming languages to access Statistics Canada tables, you need to know what you are looking for. The Python and R packages can help you with that but if you really are exploring then you might want to use the web pages to find a table that will work for you. Data can be searched using a keyword to filter results and narrow down your choices. Even without keywords there are selections you can make to target data items. You can select choices related to one or more of these categories:

1. Subject - Some key subject categories. In microdatathon #2 we looked at core housing and here we could select "Housing" to see if we see the same information or additional data we could use that might help to give us more information on homelessness.
2. Geography - this allows you and can inform you on what geographic level of detail is available with current filters applied.
3. Frequency - This allows you to assess how current the data since data tables can be created on a schedule, the most in demand are being created more frequently.
4. Survey or Statistical Program - This list is incredibly large and if you look at the where the data tables originate from you get a sense of how additional information from the census comes from and the variety of sources.
5. Content - this is used to exclude archived content.

Downloading, visualizing

Data can be viewed as HTML, downloaded in several formats, or viewed graphically:
An example of download formats is provided [here](#).

CSV/TAB

- Stores tabular data in plain text form. Fields for each record of data are separated by a comma (CSV). There are options to limit export to the data displayed or to download entire data and metadata. There are special options of database loading.

SDMX (XML) (SDMX-ML)

- SDMX-ML is a statistical and metadata exchange standard for the electronic exchange of statistical information. In order to make use of the data and metadata within the files, the files have to be processed and/or transformed. Examples include loading them into a database; converting them into application-specific processing format or processing them into a specific presentation format.

Web Data Service (for developers)

- <https://www12.statcan.gc.ca/wds-sdw/cpr2016-eng.cfm>

Program Packages (for developers)

- Python - (repeated from above) - [Web Data Service \(WDS\) User Guide \(statcan.gc.ca\)](#)
- R
 - statscanr package : A Pipeline Package for Statistics Canada Web Data Services
 - [statscanr-package: statscanr: A Data Pipeline Package for Statistics Canada Web... in andrewidyck/statscanr: A Data Pipeline Package for Statistics Canada Web Data Services \(rdrr.io\)](#) (just found this on 1/11/2020 - PFJ)
 - [statCanR \(dal.ca\)](#)
 - cansim package
 - [Working with Statistics Canada Data in R, Part 1: What is CANSIM? – Data Enthusiast's Blog](#)
 - [CRAN - Package cansim \(r-project.org\)](#)



- [Retrieve and work with public Statistics Canada data tables in R • cansim \(mountainmath.github.io\)](https://mountainmath.github.io)

Data Visualization

Since we are focusing on Python then the packages that relate to Python visualizations are needed.

If you are a Python coder then you will be using your preferred IDE.

If you are new to coding or using Python to add different visuals or read using the statscan Python package then you will need version 3.7 Python and these libraries loaded to use Python in Power BI (for example):

- pip install pandas
 - Gives you numpy (at a version that works properly), pytz and pandas
- pip install matplotlib
 - Gives you kiwisolver, pillow, cycler, matplotlib
- pip install stats-can

Geographic Levels

See Micro-Datathon #2 documentation

Focus on Homelessness

What we will work with from Micro-Datathon #2

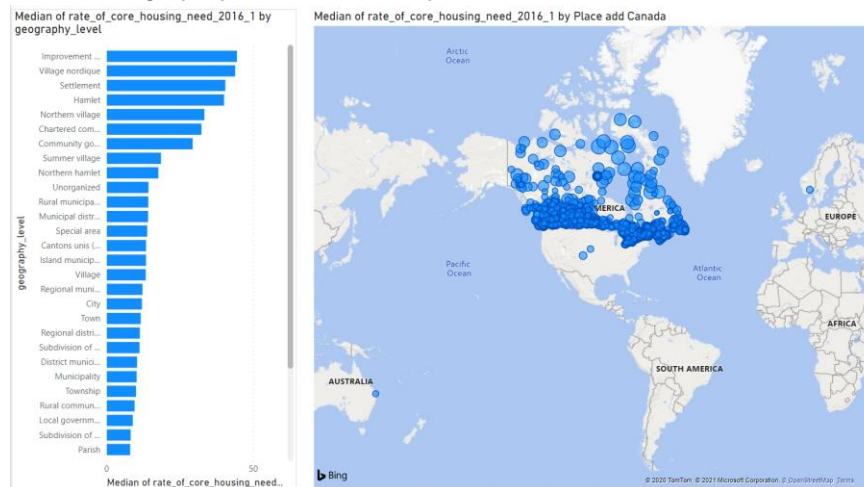
In Micro-Datathon #2 we looked at “Core Housing need” and these other datasets. I have only included some of that detail. Please look at the original document in data.world (see link at the top of document). These are visuals of the data and we will use the Python library to pull data that might help us learn more. The visuals are only meant to show what data might exist and how it may be used - they are not “final” outputs of analysis.

1. **Core housing need** <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/chn-biml/index-eng.cfm>

Statistics Canada, in collaboration with the Canada Mortgage and Housing Corporation (CMHC), presents information on 'core housing need.' [Housing suitability](#) identified whether the dwelling had enough bedrooms according to its size and composition.

Housing adequacy was assessed based on the Dwelling condition not being reported in need of major repairs. A shelter-cost-to-income ratio of less than 30% was required to deem the housing affordable.

Data is largely by Canada and by Province, some breakdown to CMA level.



2. Homeless Shelter Capacity

Table 14-10-0353-01 Homeless shelter capacity, bed and shelter counts for emergency shelters, transitional housing and violence against women shelters for Canada and provinces, Employment and Social Development Canada annual (number)

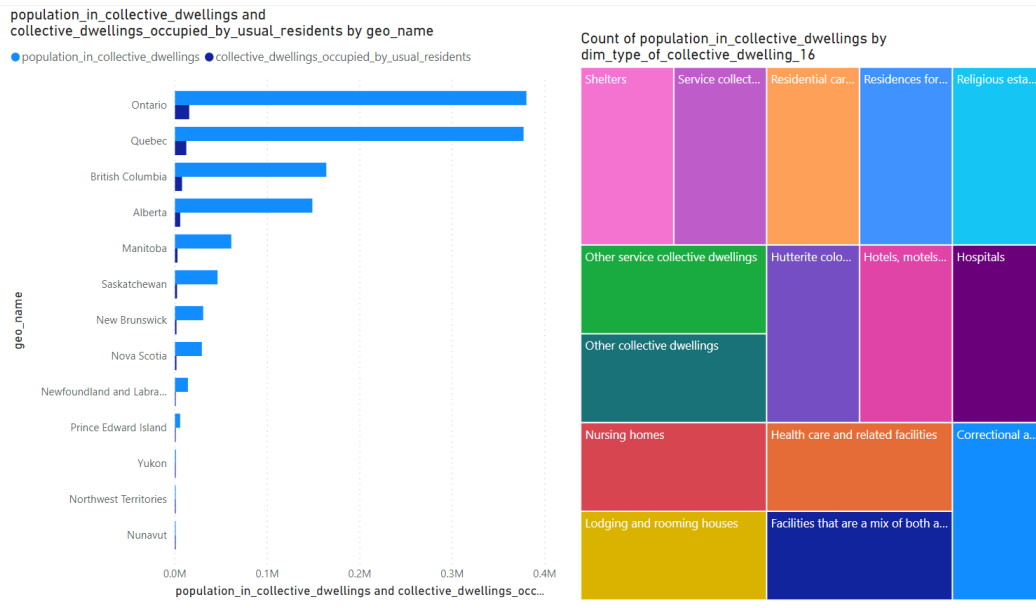
<https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410035301>



3. Individuals reporting a collective dwelling (including Shelters)

- Summary: [98-400-X2016019](https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=4&LANG=E&A=R&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=01&GL=-1&GID=1234492&GK=1&GRP=1&O=D&PID=109538&PRID=10&PTYPE=10944)
- Download of data, by Canada and by province
<https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dt-td/Rp-eng.cfm?TABID=4&LANG=E&A=R&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=01&GL=-1&GID=1234492&GK=1&GRP=1&O=D&PID=109538&PRID=10&PTYPE=10944>

[5&S=0&SHOWALL=0&SUB=0&Temporal=2016&THEME=116&VID=0&VNAMEE=&VNAMEF=&D1=0&D2=0&D3=0&D4=0&D5=0&D6=0](#)



Tips & Best practices

This section has been repeated from Micro-Datathon #2.

Citing References

Example Citation: Statistics Canada. 2017. Calgary [Economic region], Alberta and Alberta [Province] (table). Census Profile. 2016 Census. Statistics Canada Catalogue no. 98-316-X2016001. Ottawa. Released November 29, 2017.

Example on a StatCan page: <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1410035301>



How to ask StatCan a question

- Use the little speech bubbles, in the lower right corner of StatCan pages