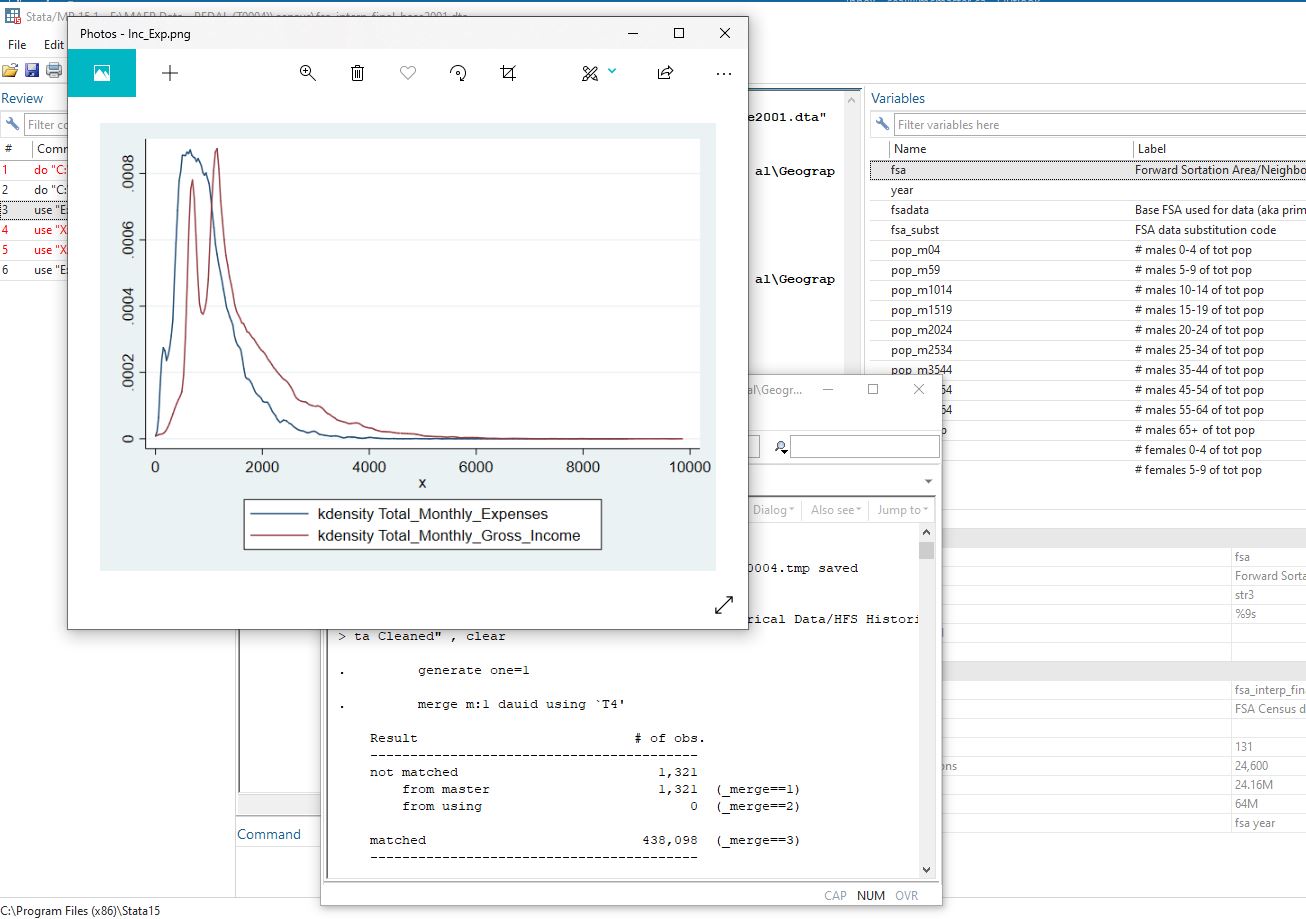
****

**Data Available in the SEAL Lab**

**McMaster University**

SEAL (Secure Empirical Analysis Lab)

**Spark: a center for Research Innovation**

Email: SEAL@mcmaster.ca

Web: seal.mcmaster.ca

Faculty of Social Sciences

McMaster University

1280 Main Street West LRW 5022

Hamilton, ON L8S 4K1

Release 1.0 March 2021

**Contents**

[**An Introduction to Data in the SEAL Lab** 3](#_Toc65678637)

[**Education Data Sets** 4](#_Toc65678638)

[OUAC (Ontario University Applications Center) Data 4](#_Toc65678639)

[Ontario Schools Data 5](#_Toc65678640)

[EQAO (Education Quality and Accountability Office) Data 7](#_Toc65678641)

[**Canadian Charity Data** 7](#_Toc65678642)

[CRA T3010 Canadian Charity Returns Data 7](#_Toc65678643)

[**Social Welfare Data** 8](#_Toc65678644)

[Food Bank Client and Usage Data 8](#_Toc65678645)

[**Supplementary Data** 9](#_Toc65678646)

[Public Use Restricted Files 9](#_Toc65678647)

[Public Use Files 9](#_Toc65678648)

# **An Introduction to Data in the SEAL Lab**

The SEAL Lab has data sharing agreements with a number of educational and social service organizations. This document describes some of the data that is available to researchers through the SEAL Lab.

All our research ready data have been cleaned and are linkable to other data. The cleaning process typically involves:

* Values and missing data corrections
* Substantive review
* Consistency checks

In some cases, substantial effort has been invested in hand checking and correcting data. We aim to be transparent in our data cleaning so that researchers are aware of how and what has been cleaned in a data set.

Most of the SEAL data is de-identified, however, postal codes are retained for linking between data sets. If data sets are to be merged, the postal codes are recoded and then the researcher can perform the file merges. THE FSA is always retained.

The education data sets can all be merged. Generally, all other data sets can be merged with any data set listed under **Supplementary Data**.

Terms of use licenses exist for some data sets listed under **Supplementary Data**. Contact the SEAL Lab for restrictions, especially if you are an international or non-academic researcher.

Researchers are welcome to bring their own data into the SEAL lab. Contact the SEAL Lab for assistance with secure FTP transfer into the lab.

# **Education Data Sets**

## OUAC (Ontario University Applications Center) Data

Applicants to Ontario Universities do not file their application directly to the university; rather they file it through OUAC, which acts as an intermediary to streamline the application process. Students submit their applications to OUAC, which forwards them to the universities. Once offers are made by the university, that information is communicated to the student through OUAC. OUAC differentiates between two types of applicants: 101 and 105. 101 students are those who apply directly from Ontario high schools, while 105 students are effectively all other applicants.

Important: At this time, SEAL has only cleaned and processed the 101 data. The cleaning process and final data described below contains only to the 101 students. The SEAL Lab is working on making available the 105 data.

**Holdings:** 101 data for the 1994/95 to 2016/17 schools years. Planned update for 101 applicants to the 2019/20 school year late in 2021. Plan to add 105 applicants from 1994/95 to 2019/20 school years in early 2022.

**Sample Student Information:**

Year of application

Student’s country of residence.

Country where the student attended a previous English Language school

Aboriginal status.

Aboriginal type for students who identify as Aboriginal

Codes for university-specific programs that student applied to

Previous postsecondary attendance

Best six average grade

Birth month and year

Course code for courses used in best-six average

Country of residence

Date and reason for application

Entry date into Canada

Applicant gender

Immigration status

Language of instruction at school

Marital status

High school information

**Sample Application Information:**

Rank of program choice

Count of students in year used to create cut offs

gcut85 – Top 85% (15th percentile) cutoff for variable uoff\_g

gcut95 – Top 95% (5th percentile) cutoff for variable uoff\_g

Student offer and response

Full or part time study

**A detailed data dictionary for the OUAC data is available from the SEAL lab.**

## Ontario Schools Data

The Ontario schools data is collected from a variety of sources and is aggregated into several data sets.

**Holdings:** 1990/91 to 2014/15 school years. Planned update to 2019/20 school year in early 2022. A data dictionary upgrade is planned for early 2022.

**Sample Board information:**

|  |
| --- |
| Academic year |
| Board identification number |
| Board name |
| Board language |
| Board type |
| Total enrolment |
| Address and postal code |
|  |
|  |

**Sample School Information:**

|  |
| --- |
| School level |
| Board identification number |
| School identification number |
| School name |
| School type |
| Language |
| Total Enrollment  Class size by grade and gender |
| **School Catchment Information (defined by postal code):**  School identification number  School name  Grade  Postal code  Male enrollment  Female enrollment  Total enrollment |

**Sample Student Information:**

|  |
| --- |
| Grade  Gender |
| Male enrolment |
| Female enrolment |
| ESL status |
| Special needs status |

**Teacher Information:**

|  |
| --- |
| Depersonalized ID |
| School |
| Gender of teacher |
| Age |
| Position |

**School and Board Expulsion and Suspension Information:**

|  |
| --- |
| Student grade |
| Exceptionality of student |
| Number of times incident occurred |
| Number of male students |
| Number of female students |

**A codebook for the Ontario Schools data is available from the SEAL Lab.**

## EQAO (Education Quality and Accountability Office) Data

The current holdings of EQAO data cover the 2004/05 to 2013/14 school years. The data contains student level EQAO performance and the student level survey results. Cohort analysis across Grades 3, 6 and 9 EQAO results is feasible.

Sample information is not provided in this document given the confidential nature of this information and existing data sharing agreements. Contact the SEAL Lab to discuss data queries.

# **Canadian Charity Data**

## CRA T3010 Canadian Charity Returns Data

The T3010 is the tax return that must be filed by all registered Canadian charities under the Income Tax act. Canadian charities are tax-exempt, but the CRA collects T3010 data and renders it publically available so that governments, private donors, and the general public know what charities are doing and how charities are using their resources.

The main penalty for not filing a tax return is the revoking of a charity’s charitable status. Having a charitable status allows the charity to be tax-exempt and to be able to issue tax receipts that allow donors to reduce the amount of income tax they pay.

As of June 1, 2019, returns could be filed on paper or electronically, but prior to then, all returns were filed on paper. Charities all have different fiscal period end dates, and the T3010 return must be filed within 6 months of the charity’s fiscal period end date.

The T3010 itself collects information on a charity’s programs, fundraising, financial accounts, employees and volunteers, activities outside Canada, and political activities. Charities are also mailed Form TF725, the Registered Charity Basic Information Sheet. The TF725 is mailed to charities already filled out with information from the CRA’s records. Charities are asked to review the TF725 and report any changes or inaccuracies each year, and attach it to their T3010. Information from the TF725 is in the IDENT tables in the raw data. Charities also need to include Form T1235, the Directors, Trustees and Like Officials Worksheet, which collects a list of the charity’s directors and trustees. If applicable, charities also submit a list of the people and organizations to whom they made donations in Form T1236, Qualified donees worksheet/ Amounts provided to other organizations.

The data cleaned and hosted here is the T3010, TF725, T1235, and T1236. The RC232 Ontario Corporations Information Act Annual Return and T2081 Excess Corporate Holdings Worksheet for Private Foundation’s forms are also part of a charity’s tax return in some circumstances, but data from those forms is not cleaned or hosted here.

The SEAL Lab T3010 data is currently available for download from the SEAL Dataverse site. Check the SEAL website at seal.mcmaster.ca to get access to this data.

**A data dictionary is available for download from the SEAL Lab Dataverse site.**

# **Social Welfare Data**

## Food Bank Client and Usage Data

This is the administrative data from a food bank hub in a metropolitan area in Ontario, that supports a network of the ten major food banks. It covers April 1 2014 through December 2020, and contains an observation for each food bank visit during that period with the following exceptions. Some food banks started tracking user information after April 1 2014. Other food banks have opened, closed, relocated, or changed their target audience (e.g. only youth or only seniors) during this period. Some organizations have “food pantries” which provide a small amount of food to act as emergency supplies for people in urgent need. The major food banks operate on a much larger scale, typically serving over 200 households with a bigger variety of food designed to meet nutritional needs. This data contains all information about food bank users over this period.

The data covers the following broad areas:

Food bank visited and data of visit

Access to special food bank programs e.g. Christmas gift program

Family characteristics such as number and type of household head(s), children, extended families, economic households composed of unrelated individuals

Income sources and amount by various types e.g. ODSP, WI, wage income etc.

Household spending e.g. rent, utilities, food etc.

Socio-economic characteristics of all individuals within the household group e.g. immigration status, marital status, gender, age, housing type, disability etc.

The data may be matched to DA level census profile data.

Sample information is not provided in this document given the confidential nature of this information and existing data sharing agreements. Contact the SEAL Lab to discuss data queries. We have a confidential and public-use data dictionary.

# **Supplementary Data**

## Public Use Restricted Files

The SEAL Lab has a variety of public use restricted files that may be used to match population demographic and socioeconomic characteristics to various data sets based on postal code (typically DA level information). The SEAL Lab will match the population data at the record level data in the data set. Typically, postal codes are converted to FSAs after merging to ensure the de-identification of individuals.

SEAL clients are welcome to import their own matching data, or the lab can obtain the data on behalf of the researcher.

The following Statistics Canada data is typically used for matching purposes:

Various CANSIM tables

Census/NHS Profile Tables based on Federal Electoral Districts, FSA, CS, CD, CT or DA

Census/NHS PUMF Files

PCCF post-1995

Access requires that the researcher be eligible for access through the Statistics Canada DLI project at their university or licensed through CHASS.

## Public Use Files

The following data sets are unrestricted:

US Census PUMF data 1990, 2000 and 2010 plus the ACS post-2010

CTaCS

GIS Census Data - Canada