

SOCI 424

# Canadian Statistics and Microdata for the Social Sciences

*Alex Guindon*  
*GIS and Data Services Librarian*  
*Concordia University*  
*[alex.guindon@concordia.ca](mailto:alex.guindon@concordia.ca)*

# What we will cover today

- ▶ Statistics vs. Microdata
- ▶ Sources of Microdata
  - ▶ Odesi (for Statistics Canada microdata)
  - ▶ Hands-on exercise on Odesi

# Types of data files

The background of the slide features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# What is microdata?

- ▶ Microdata consists of the data directly observed or collected from a specific unit of observation (individual cases, answers from one individual, from one respondent).

# Aggregate data vs. Microdata

## Aggregate data

- ▶ Every cell represents *added up* data (or frequencies), i.e. how many people living in Quebec identify themselves as South Asian, Chinese, Black, etc.
- ▶ These stats can be expressed in absolute numbers or percentages
- ▶ Beyond basic sums and percentages, no serious statistical analysis can be done with aggregated data

## Microdata

- ▶ Each row represents the answers of an *individual respondent*
- ▶ Often the numbers are codes, not actual values.
- ▶ Thus you need the documentation to understand the data
- ▶ Raw data is what should be used to conduct in-depth statistical analysis.

# Aggregate Data (2006 Census)

Ethnocultural Portrait of Canada - Data table - Windows Internet Explorer

http://www12.statcan.ca/census-recensement/2006/dp-pd/hlt/97-562/pages/page.cfm?Lang=E&Geo=CMA&Code=24&Table=

Links Dictionaries Delicious CLUES Concordia Directory Coutts - Login Databases Google Gmail Staff Page Wiki Microsoft Outlook Web Access

LibX CUL Keyword Search CLUES Catalogue - Concordia University Libraries Clear Scholar

Concordia Libraries Ethnocultural Portrait of ...

Statistics Canada  
www.statcan.gc.ca

Français Home Contact Us Help Search canada.gc.ca

Home > Census > Data products > Highlight tables > Ethnocultural Portrait of Canada

Visible minority groups, 2006 counts, for Canada, provinces and territories, and census metropolitan areas and census agglomerations - 20% sample data

Select a view

Quebec 2006 counts Submit

Page 1 of 2 Show all | Next >

Geographic name	Total - Population by visible minority groups	Total visible minority population <sup>1</sup>	South Asian <sup>2</sup>	Chinese	Black	Filipino	Latin American	Southeast-Asian <sup>3</sup>
Canada	31,241,030	5,068,090	1,262,865	1,216,570	783,795	410,695	304,245	239,935
Quebec	7,435,905	654,355	72,845	79,830	188,070	24,200	89,505	50,455
Alma, CA (Que.)	32,280	215	10	40	45	10	25	80
Amos, CA (Que.)	17,590	165	0	20	70	20	0	55
Baie-Comeau, CA (Que.)	29,455	150	0	35	65	0	10	10
Campbellton (Quebec part), CA	3,040	10	0	10	0	0	0	0
Cowansville, CA (Que.)	12,095	95	15	10	50	0	0	10
Dolbeau-Mistassini, CA (Que.)	14,390	70	0	30	10	10	0	10
Drummondville, CA (Que.)	76,400	1,660	10	70	240	0	1,095	105
Granby, CA (Que.)	66,800	1,285	0	130	305	20	460	100
Hawkesbury (Quebec part), CA	1,400	0	0	0	0	0	0	0
Joliette, CA (Que.)	42,515	705	30	105	285	0	185	55
La Tuque, CA (Que.)	15,070	105	0	20	50	0	0	15

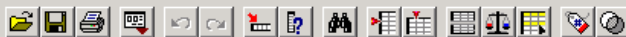
Done

Start In-box - Microsoft Outlook Ethnocultural Portrait... SOCI-498.docx - Micro... 9:58 AM

# Microdata (unaggregated)

SurveyData.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help




1 : ID1


	ID	Program	Affiliation	Age	LibSocial	LibIndiv	LibGroup	LibSilent	IndivFreq	IndivLocat	IndLibLocat	GrLocat	GrLibrL
1	1	JMSB MBA	2	2	1	4	4	1 4	2	0	5	0	
2	2	Political science	1	2	1	3	2	2 3	1	1	1	1	
3	3	Art history	2	2	1	5	2	5 2	0	0	0	1	
4	4	English	1	2	4	3	3	5 2	0	3	1	1	
5	5	Accounting	1	1	3	4	3	5 1	2	3	1	1	
6	6	ENCS	2	2	2	5	3	4 2	2	3	5	2	
7	7	JMSB	1	1	1	4	3	5 1	3	3	3	1	
8	8	Engineering	1	2	1	4	4	5 2	0	3	5	1	
9	9	JMSB, Finance	1	2	1	3	3	5 1	3	3	3	1	
10	10	n/a	5	4	2	5	4	5 1	5	3	5	2	
11	11	Sociology	1	1	1	5	3	5 1	2	1	0	1	
12	12	MSc. Administration	2	2	4	4	4	4 1	3	3	3	1	
13	13	ENCS	1	2	4	2	2	1 1	3	1	5	1	
14	14	JMSB	1	2	4	4	4	5 2	2	1	3	1	
15	15	Political science	1	1	5	1	1	1 3	6	3	1	2	
16	16	BComm	1	2	5	5	3	5 1	0	2	3	1	
17	17	Civil engineering	1	3	2	4	3	1 2	0	3	1	1	
18	18	JMSB	1	1	1	5	4	5 1	3	2	3	4	
19	19	Film studies	1	1	2	4	4	4 2	0	2	0	0	
20	20	Finance	1	1	2	5	5	4 1	3	0	3	1	
21	21	JMSB	1	3	1	5	4	5 2	3	2	3	4	
22	22	Accounting	0	1	4	4	4	4 1	0	3	3	4	
23	23	Leisure sciences	1	1	1	5	5	5 2	3	1	2	1	
24	24	CIISE	2	2	3	2	1	5 1	4	0	2	1	
25	25	JMSB	1	1	1	4	3	5 1	3	1	2	1	
26	26	Marketing	1	1	1	3	3	2 2	3	1	3	1	
27	27	Economics	1	1	3	4	4	5 2	0	3	0	1	
28	28	Marketing	1	2	2	4	4	3 2	0	3	5	1	
29	29	ECE	1	1	1	5	2	5 1	4	0	6	1	
30	30	n/a	2	4	4	2	2	2 2	0	1	1	1	
31	31	Child studies	1	1	1	5	4	5 2	3	2	1	1	


Data View Variable View


SPSS Processor is ready


Start






































10:13 AM

The background of the slide features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

# An Overview of statistics canada microdata



# Microdata from Statistics Canada

## 1. *Public Use Microdata Files (PUMF)*

- ▶ They are *samples* of the complete datasets
- ▶ Preserve the confidentiality of respondents
  - ▶ Nominal information is deleted
  - ▶ Some variables are suppressed
  - ▶ Some variable categories are collapsed
  - ▶ Information for smaller geographic areas (Census tracts, dissemination areas) may not be available
- ▶ PUMFs are NOT available for longitudinal surveys

## 2. *Master Files*

- ▶ Complete datasets

# Dataset documentation

Documentation includes:

- ▶ Survey methodology
  - ▶ Description
  - ▶ Target population
  - ▶ Sampling method (weighting)
  - ▶ Measures of data reliability
- ▶ Questionnaire
- ▶ List and description of variables
- ▶ You cannot properly analyse survey results without consulting the documentation

# Access to datasets

# How to access PUMFs

- ▶ PUMFs are made available to **students and faculty** through the **<odesi>** database
- ▶ Some files may not be in **<odesi>**
  - ▶ Ask the data librarian (that's me!) to download them for you

The background features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

How to explore statistics  
canada's collection?

# What dataset should I use?

There are two main ways to explore Statistics Canada surveys:

1. Statistics Canada's website

1. Includes ALL surveys
2. Browse by Key Resources - Definitions and Documentation - Surveys →
3. If no PUMF is available, you must use Master File at QICSS →

2. Browse <odesi> by topic

1. Only includes surveys that have PUMFs
2. This is the only option that will allow you to download the datasets

# <odesi>

- ▶ Statistics Canada's PUMFs are directly accessible through <odesi>.
- ▶ <odesi> can be accessed on campus or from home (login as a Concordia user)
- ▶ Users can download complete datasets or select specific variables and cases
- ▶ Survey documentation (user guides, codebooks) is also available

# <odesi>

Examples of surveys available on <odesi> :

- ▶ General Social Survey (most cycles)
- ▶ Census microdata (back to 1971)
- ▶ National Household Survey (2011)
- ▶ Ethnic Diversity Survey (2002)
- ▶ Labour Force Survey (1976 onward)
- ▶ Access and Support to Education and Training Survey (2008)



<odesi> demo

# Final words...

REMEMBER:

- ▶ Not all Statistics Canada's surveys have PUMFs  
(Master files are only available at the [QICSS](#))
- ▶ ... and some PUMFs may not be on <odesi>
  - ▶ Consult the complete list on the [DLI website](#)
  - ▶ To obtain these datasets, contact me: [alex.guindon@concordia.ca](mailto:alex.guindon@concordia.ca)