Finding Canadian Data

A BRIEF INTRODUCTION

Martin Chandler
martin.chandler@mcgill.ca

About this presentation

Download it here: https://tinyurl.com/CanDataMcGill

What is “data”?

Terminology

Statistics vs Microdata vs Geospatial data

Finding Statistics

Finding Microdata

Finding Geospatial Data

Learning Outcomes

By the end of this presentation, you will be:

- better able to identify types of data
- better able to find data
- better able to ask questions about data

What is data?

Data is facts, in different forms, about a particular subject, used to answer a question or to make an argument.

This can include numbers (numerical data), information about locations (geospatial data), or words about something (descriptive or otherwise)

Eg: The number of couples introduced through the pineapple export business is a point of data (being 1, as far as I know)

The number of dogs living in a particular household is another data point

Some terminology

Data: facts, or factual information, used for reasoning or analysis

Data point: one fact or unit of study

Dataset: A collection of facts, often gathered together to be manipulated

Home ID	Postal Code	Dog	Breed	Temperament
001	M6R 1N6	0	N/A	N/A
002	H4C 3C5	1	Beagle	Friendly
003	B2R 1S2	2	Dachshund; Terrier	Cute; Derpy
004	L2S 2A7	0	N/A	N/A

Data point

Dataset

Some termin

Variable: A characteristic tha . May be numeric (“35”) or descriptive

Case: The unit of analysis (ie llected, the microdata file will have 1000

Variable

Case

Home ID	mperament
001	A
002	endly
003	te; Derpy
004	A

Statistics

Calculated figures produced using methods developed through modes of inquiry – counts, totals, averages, means, etc.

Eg: Approval rates for governing party

Number of households with dogs, in a given area

Microdata

A dataset with individual pieces of recorded information

Eg: Individual responses to a survey

Student test scores by schools

Exact breakdown of spending for an event

Which should I use, statistics or microdata?

“How much/many?” vs “Why?”

Eg: “I need data on the number of Gujarati speakers in Montreal.” You want statistics!

Or: “I’m exploring the relationship between radon and lung cancer in Montreal”. You want microdata!

Statistics vs Microdata

- Respondents had an average of 0.75 dogs per household
- 33% of dogs surveyed were derpy

Home ID	Postal Code	Dog	Breed	Temperment
001	M6R 1N6	0	N/A	N/A
002	L2S 2A7	1	Beagle	Friendly
003	B2R 1S2	2	Dachshund; Malamute	Cute; Derpy
004	H4H 1V1	0	N/A	N/A

Why does this matter?

Different platforms and tools for finding

Different restrictions and requirements for use (e.g. privacy)

Different methods and software for working with the information

(A little) more terminology

Numeric data – data based on numbers; usually in spreadsheets/CSVs

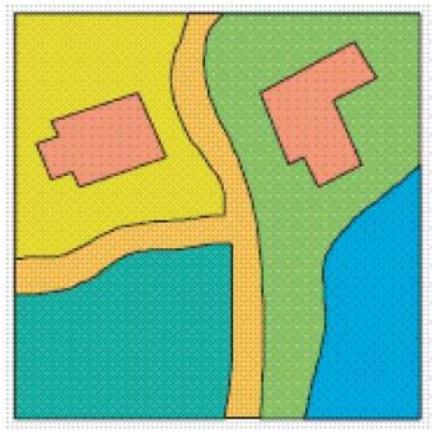
COL0	COL1	COL3	COL5	COL6	COL7	COL8	COL9	COL10	COL11	COL12
2050000	12	205	0	403390	347435	345	376245	21380	225940	111270
2050001	12	205	1	3868	3575	5	3740	80	1925	1170
2050002	12	205	2	6194	5270	5	5895	155	3185	1845
2050003	12	205	3	2955	2405	5	2135	515	1515	895
2050004	12	205	4.01	3466	2765	5	2560	955	2535	915
2050004	12	205	4.02	4771	3865	0	3440	1205	2845	1515
2050005	12	205	5	1808	1380	0	1730	135	885	680
2050006	12	205	6	3129	2360	0	2615	210	1705	880
2050007	12	205	7	1859	1340	0	1430	205	805	770
2050008	12	205	8	2778	2225	0	2055	660	1670	910
2050009	12	205	9	2357	1850	0	1555	615	1585	480
2050010	12	205	10	5036	4140	0	4320	680	3080	1415
2050011	12	205	11	5631	4365	0	4780	680	3770	1205
2050012	12	205	12	2482	1830	0	2245	140	1550	535
2050013	12	205	13	2561	2085	5	2515	80	1490	740
2050014	12	205	14	4248	3475	5	3795	140	2100	1380
2050015	12	205	15	4829	4255	5	4400	420	2615	1480
2050016	12	205	16	3766	3190	0	3400	325	2265	940
2050017	12	205	17	2914	2515	0	2795	160	1545	845

(A little) more terminology

Geospatial data – data that involves a location

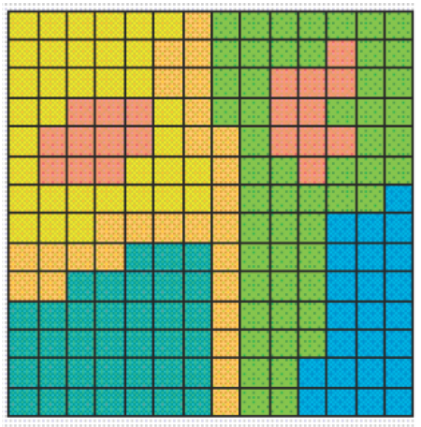
Vector data

points, lines, & polygons on XY coordinates



Raster Data

pixels/grid cells, usually square



FED_NUM	NID	FEDNUM	ENNAME	FRNAME	PROV	DECPOPCNT	USESDERPY	USESBOUFFON	WTF
24001	{AB0EAE36-C1EC-49F9	24001	Abitibi--Baie-James--Nur	Abitibi--Baie-James--Nunav	QC	85475	138	630	60
24002	{6C60176A-0BCC-4B37	24002	Abitibi--TÃ©miscamingu	Abitibi--TÃ©miscamingue	QC	102794	156	872	540
24003	{66747292-7B75-4BC1-	24003	Ahuntsic-Cartierville	Ahuntsic-Cartierville	QC	110473	881	73	77
24004	{94538187-0A87-4207-	24004	Alfred-Pellan	Alfred-Pellan	QC	98045	846	470	133
24005	{D97AF803-B05B-4ED4	24005	Argenteuil--La Petite-Nat	Argenteuil--La Petite-Natio	QC	94208	167	939	46
24006	{7E5624C1-5FAD-465E-	24006	Avignon--La Mitis--Matan	Avignon--La Mitis--Matane-	QC	74547	206	745	0
24007	{F4A433CE-8E4D-450A	24007	Beauce	Beauce	QC	106337	281	978	114
24008	{754BB4E4-8F8E-4E89-	24008	Beauport--Limoilou	Beauport--Limoilou	QC	92944	869	141	131
24009	{5E243D4C-2EE8-4E8C-	24009	BÃ©cancour--Nicolet--Sa	BÃ©cancour--Nicolet--Saur	QC	93779	150	623	807
24010	{D1F455BD-FFA2-4060	24010	Bellechasse--Les Etchemi	Bellechasse--Les Etchemins	QC	112385	415	598	143
24011	{5737F403-FDED-41DD	24011	Beloeil--Chambly	Beloeil--Chambly	QC	109955	649	516	100
24012	{9FEF06FB-4B7F-4B89-	24012	Berthier--MaskinongÃ©	Berthier--MaskinongÃ©	QC	98590	16	704	126
24013	{BF246EF7-4729-45BD-	24013	ThÃ©rÃ¨se-De Blainville	ThÃ©rÃ¨se-De Blainville	QC	98499	861	61	44
24014	{1711BCE7-E928-41B4-	24014	Pierre-Boucher--Les Patri	Pierre-Boucher--Les Patriot	QC	95326	473	355	73
24015	{644A53FD-09EC-43E3-	24015	Bourassa	Bourassa	QC	100286	649	347	104
24016	{19AECA7D-7860-4DE3	24016	Brome--Missisquoi	Brome--Missisquoi	QC	98616	37	823	507
24017	{0C426B17-CDCD-401A	24017	Brossard--Saint-Lambert	Brossard--Saint-Lambert	QC	100828	727	504	3
24018	{0A31B4B1-7972-4A54-	24018	Rimouski-Neigette--TÃ©	Rimouski-Neigette--TÃ©m	QC	84809	233	638	142
24019	{4EA9158D-8269-42E5-	24019	Charlesbourg--Haute-Sai	Charlesbourg--Haute-Saint-	QC	103331	646	80	68
24020	{57ABE396-D834-4AC5	24020	Beauport--CÃ´te-de-Beau	Beauport--CÃ´te-de-Beaupr	QC	92496	460	648	41
24021	{30BC4AFF-9804-4219-	24021	ChÃ¢teauguay--Lacolle	ChÃ¢teauguay--Lacolle	QC	92169	457	564	113
24022	{400EF08D-8993-4CBC-	24022	Chicoutimi--Le Fjord	Chicoutimi--Le Fjord	QC	81501	201	722	41
24023	{05F8541F-CD56-4523-	24023	Compton--Stanstead	Compton--Stanstead	QC	101946	30	872	111
24024	{091A4423-615E-4AFB-	24024	Dorval--Lachine--LaSalle	Dorval--Lachine--LaSalle	QC	106886	632	66	135
24025	{042D5699-FD0B-4524-	24025	Drummond	Drummond	QC	98681	307	634	933
24026	{CF02F254-5D4A-43BF-	24026	GaspÃ©sie--Les ÃŽles-de	GaspÃ©sie--Les ÃŽles-de-l	QC	78833	480	549	194

Where to find data*

Numeric Data Libguide

Geospatial Data Libguide

*Please note: LibGuide structure/layout are subject to change

Top 3 resources (according to Martin):

NUMERIC:

Statistics: CHASS Census Analyzer

Microdata: <odesi>

GEOSPATIAL:

Géoindex

Census of Population

Conducted every 5 years

Comprehensive of the population of Canada (mostly)

Legislated by the *Statistics Act*

Data is collected via mandatory “short form” (100% of population) and “long form”* (20% of population)

Available through the StatsCan website, or through the CHASS Census Analyser

*In 2011, the long form was voluntary, and conducted under the title “National Household Survey”.

Census of Population

Covers topics such as:

- Age
- Sex
- Marital Status
- Dwellings and households
- Language
- Citizenship
- Aboriginal peoples*
- Income
- Labour market activity
- Occupation
- Education
- Mode of transportation to work
- House and shelter costs
- Place of birth
- Religious affiliation

*Depending on the individual/nation/reserve

Other surveys and statistical programs

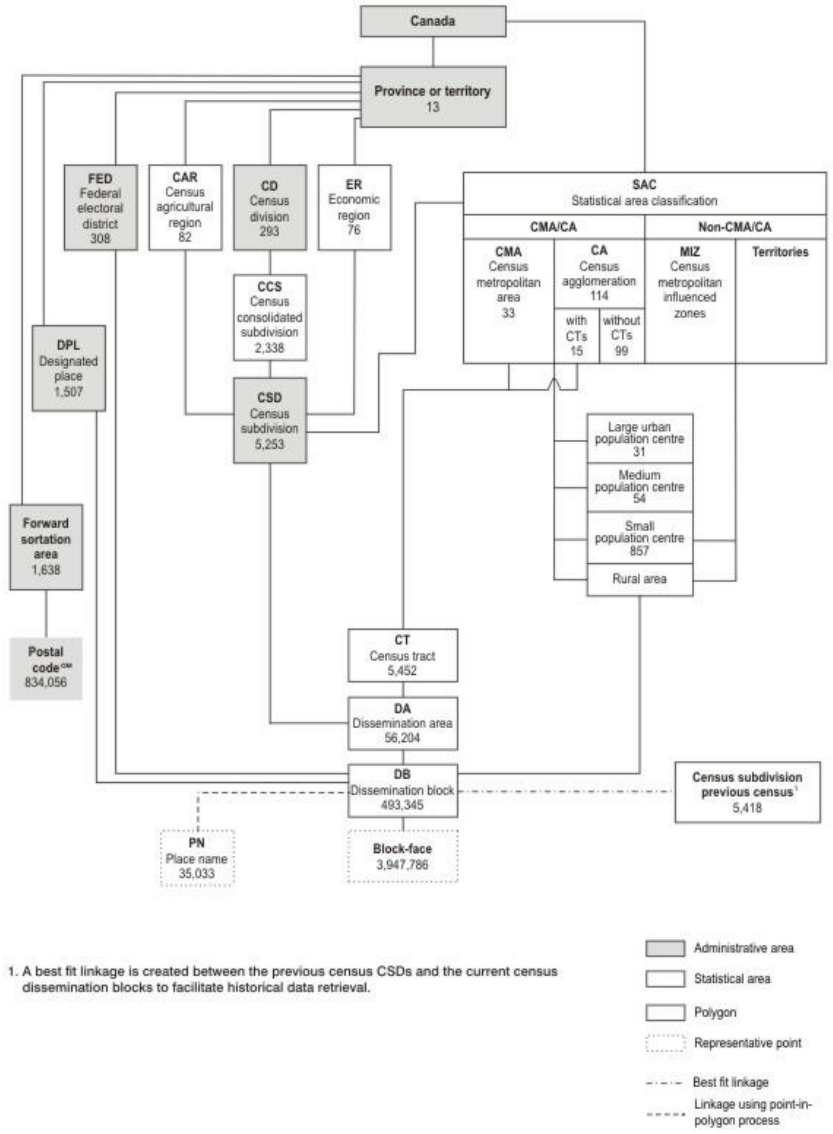
Statistics Canada also has 350+ active and inactive surveys and databases

Includes topics such as transportation, travel & tourism, crime & justice, children & youth, business, health, agriculture, and more

Available through the StatsCan Website

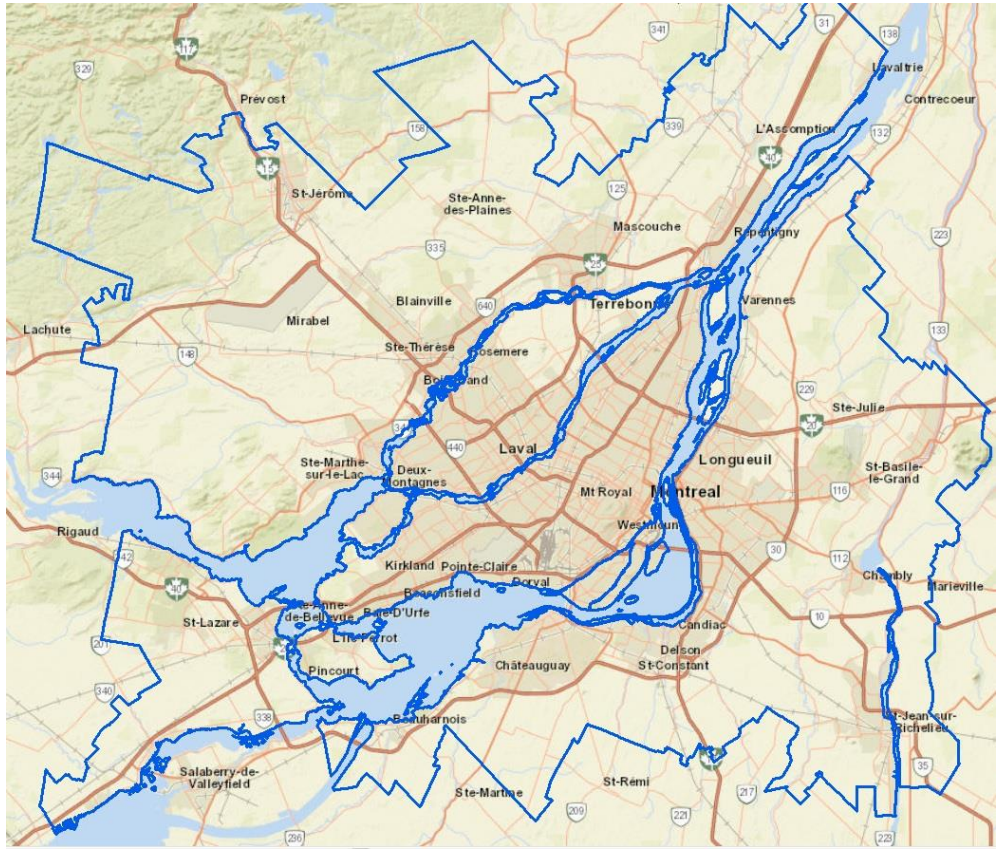
Census Geography

Statistics Canada's hierarchy of standard geographic units



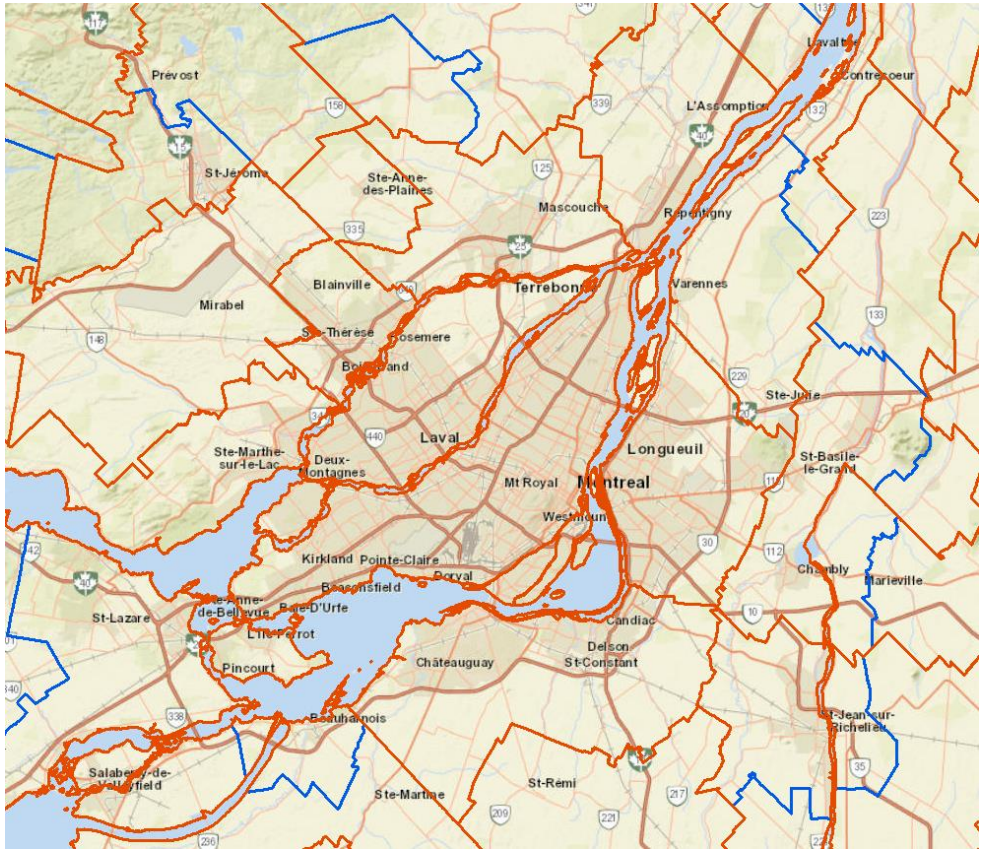
Census Geography

Census Metropolitan Area (CMA)



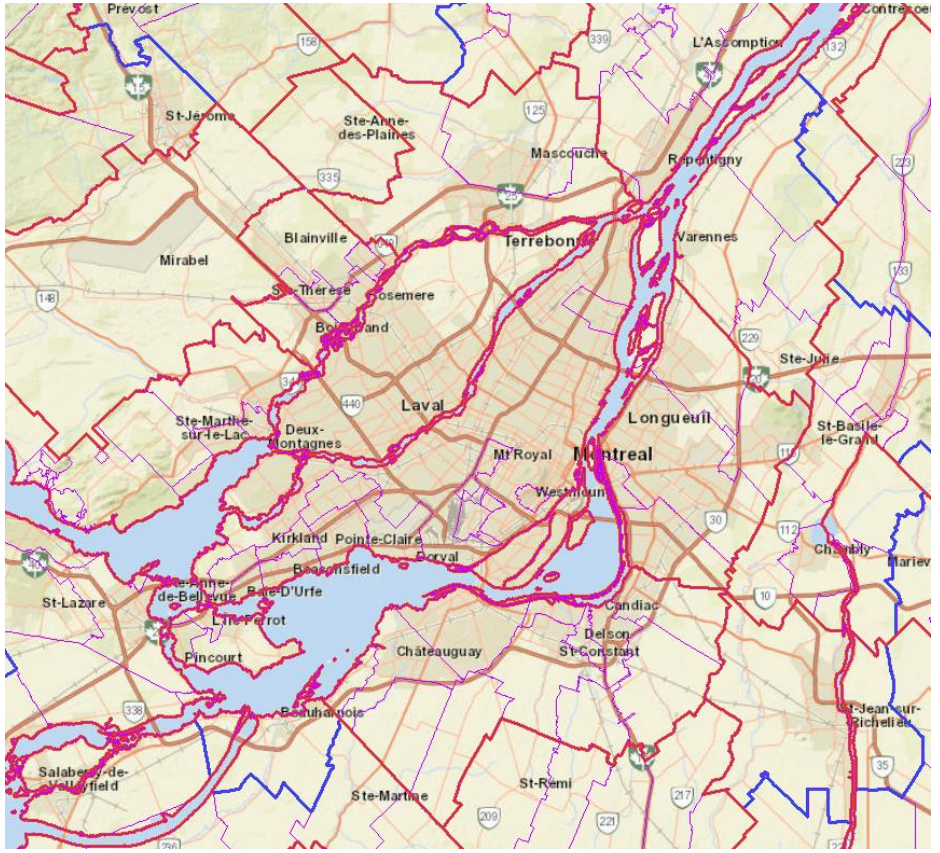
Census Geography

Census divisions



Census Geography

Census subdivisions

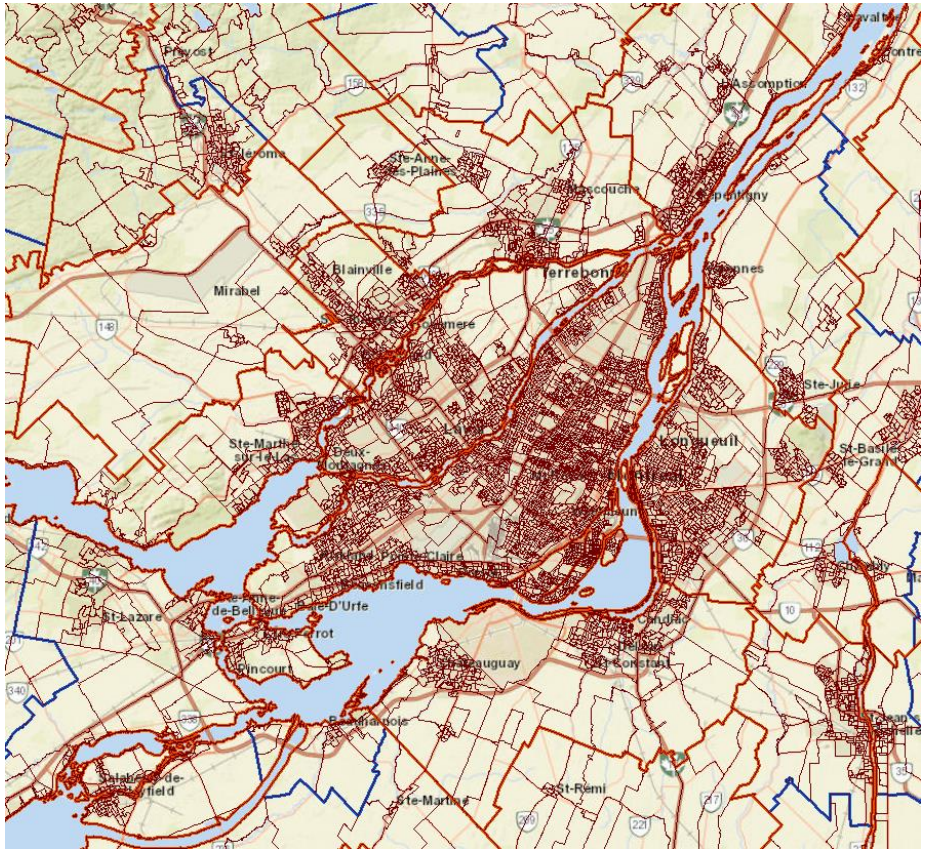


Census Geography

Census tracts

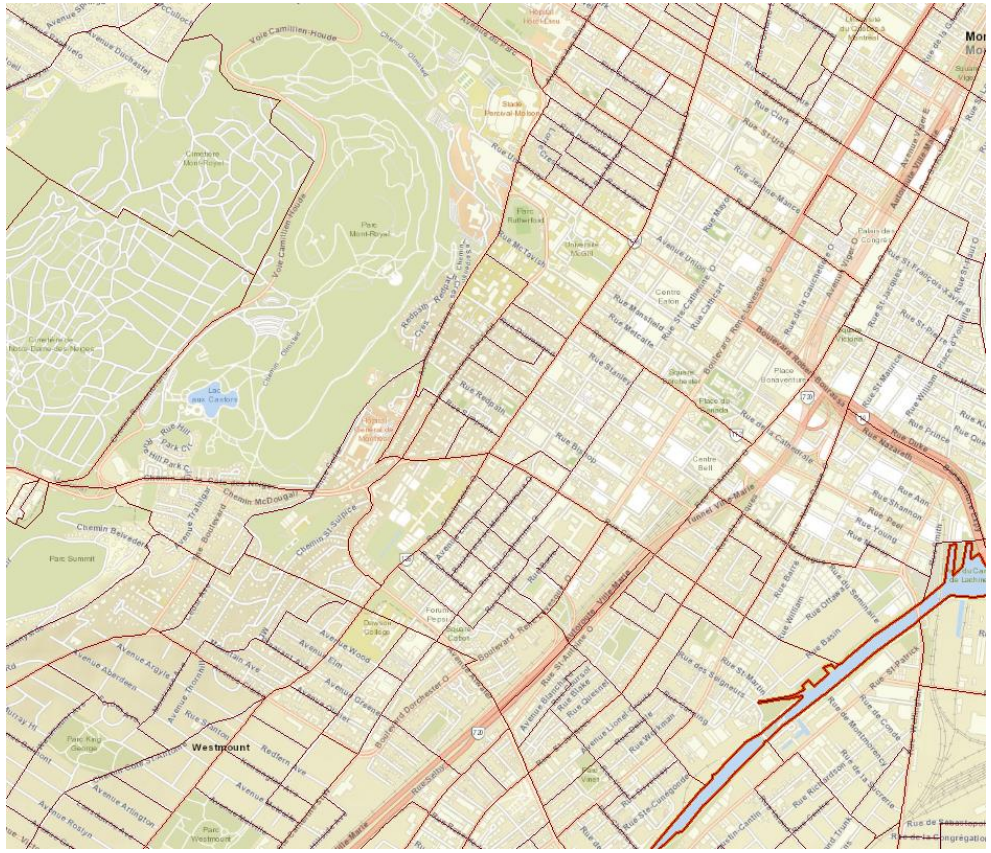
Census Geography

Dissemination areas



Census Geography

Dissemination areas (close up)



Demo: CHASS Census Analyser

CHASS Census Analyzer

I’m looking for some data...

- “I’m trying to figure out which province has the most Gujarati speakers in 2016.”
- Statistics, or microdata?
- Statistics!

Exercise

Using the CHASS Census Analyzer, download 2 tables with the following information:

- Census year: 2001 and 2016
- Geography: Census tracts within Montreal
- Variables:
  1. Total population
  2. Mother tongue: Gujarati
  3. Total Population by Home Language

Feel free to discuss and/or work together, and ask if you have any questions!

Microdata

All of the surveys Statistics Canada engages in produce microdata

However, not all of that microdata is available for use through the StatsCan website (some is)

Some is available through data portals, some through the Data Liberation Initiative (DLI), and some through Research Data Centres (eg QICSS)

Demo: <odesi>

<odesi>

I’m looking for some data…

- “I’m exploring the relationship between income level and diabetes.”
- Statistics, or microdata?
- Microdata!

Exercise

You've been asked to find some data to help explore the relationship between self-employed business owners and feelings of isolation at work.

Using <odesi>, search for possible sources of microdata on this topic. Browse through the variables

If you want to do some cross-tabulation of variables, great!

Feel free to discuss and/or work together, and ask if you have any questions!

Demo: GéoIndex

GéoIndex

I’m looking for some data…

- “I’m trying to find data of where the school boards in Quebec are.”
- What sort of data would this be?
- Geospatial!

Exercise

Using Geoindex, search for data showing different property lots on the island of Montreal from 2015, and from 1912.

Feel free to discuss and/or work together, and ask if you have any questions!

Survey

Please fill in this survey, to improve future sessions!

https://tinyurl.com/FindDataSurv

Questions?

martin.chandler@mcgill.ca

Sources

Esri, Inc. (March 14, 2019). “Raster is faster but vector is corrector”. Teach with GIS: An Implementation Guide for Teachers, Schools, and Districts. Redlands, CA: Esri, Inc.

Ham Sandwich. (August 24, 2019). “How much does Commander Data weigh?” Accessed Nov 29, 2019 from https://scifi.stackexchange.com/questions/217969/how-much-does-commander-data-weigh

Statistics Canada. (2011). Hierarchy of standard geographic units. Accessed Jan 6, 2020 from https://www150.statcan.gc.ca/n1/pub/92-195-x/2011001/other-autre/hierarch/h-eng.htm

Statistics Canada. (2016). 2016 Census - Boundary files. Accessed Jan 6, 2020 from https://www12.statcan.gc.ca/census-recensement/2011/geo/bound-limit/bound-limit-2016-eng.cfm

Trimble, L. (2017). Finding Canadian Statistics & Data. Presentation for graduated student library assistants, Map & Data Library, University of Toronto.

Wikipedia contributors. (November 2, 2019). Contagion (Star Trek: The Next Generation). In *Wikipedia, The Free Encyclopedia*. Accessed December 12, 2019, from https://en.wikipedia.org/w/index.php?title=Contagion_(Star_Trek:_The_Next_Generation)&oldid=924271549