#### **OUEEN'S UNIVERSITY LIBRARY**



# PCCF+

## Postal Code Conversion File Plus

ACCOLEDS, University of Calgary, December, 2015

Jeff Moon
Data Librarian &
Academic Director, Queen's Research Data Centre
Queen's University Library







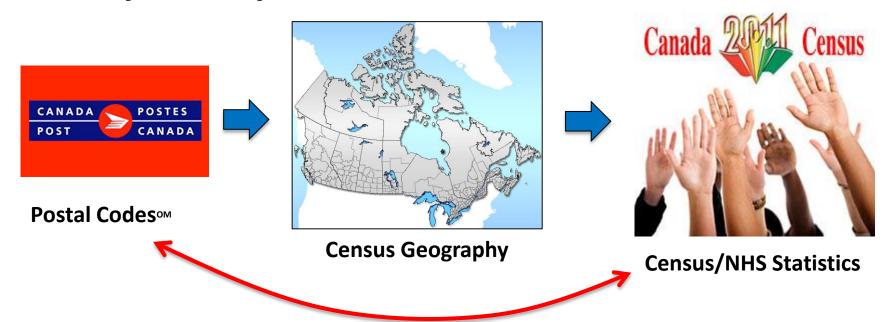
- A SAS<sup>©</sup> control program
- Associated datasets derived from the PCCF
- A Postal Code<sup>oM</sup> population weight file
- Geographic Attribute File
- Health Region boundary files
- Other supplementary data

#### What does it do?



# PCCF+ automatically assigns a range of Statistics Canada's standard geographic areas based on Postal Codes

#### And why would you want to do this?





#### **PCCF** versus **PCCF**+

The PCCF uses an SLI or 'Single Link Indicator' for matching. The single link indicator identifies the geographic area with the majority of dwellings assigned to a particular Postal Code.

The PCCF+ uses population-weighted random allocation for Postal Codes that link to more than one geographic area.

Saeeda Khan Health Analysis Division, Statistics Canada

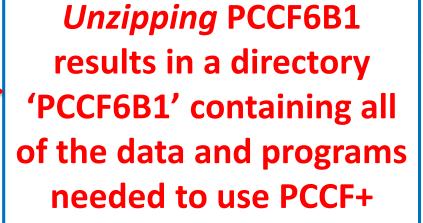


#### Consider using PCCF+ rather than PCCF-SLI if any of the following apply

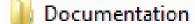
- You want to use variables present on the PCCF+ which are not present in regular PCCF
- Your file is less than perfect with respect to postal codes
- You want help to evaluate the quality of the postal code on your data file
- The "vintage" of the postal codes on your file spans more than one census
- You want to do better coding in rural areas
- If your file includes postal codes used by residents of "incompletely enumerated Indian Reserves"

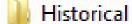
#### We'll be working with PCCF6B1

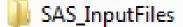
**Download from DLI and unzip** 





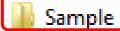




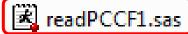


Supplementary

📗 Other SAS programs



PCCFplus\_6B1\_ACCOLEDS.sas



#### Sample data files

this is where we'll put our TEXT file containing Postal Codes

#### **Modified SAS program**

based on Statistics Canada program

SAS program to convert TEXT file to SAS

not provided by Statistics Canada

## Working with PCCF+



#### Two steps

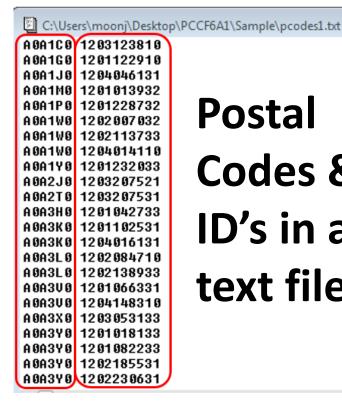
- 1. Convert a text file containing Postal Codes into SAS dataset format
- 2. Run the PCCF+ program against this dataset

# Step 1.1 Start with a text file containing Postal Codes

#### **Input File Specifications:**

#### **Postal Codes:**

- Can be sorted or unsorted
- Must each have a unique ID, up to 15 chars
- Must be 6 characters, no spaces/hyphens
- Must be in UPPER CASE
- May be duplicated, but need different ID's
- It is important to note that in constructing input data, duplicate PCODEs should be retained and provided separate IDs.



**Postal** Codes & ID's in a text file

#### **UPPER CASE** needed



#### Results with 2 Postal Codes containing lower case letters

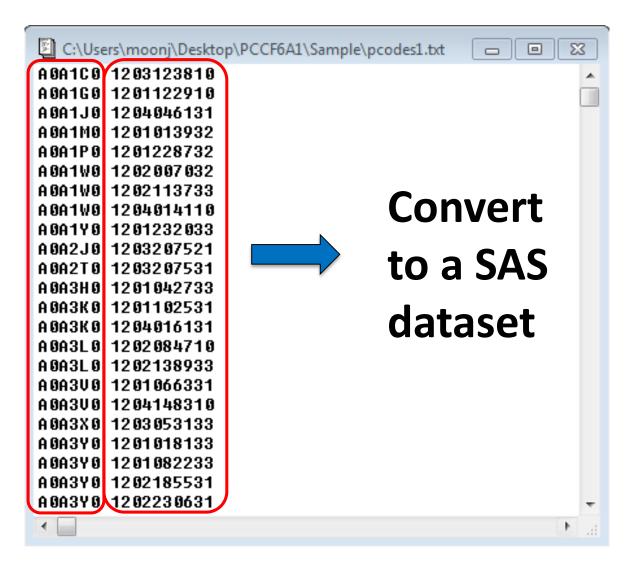
LINK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0  Error: No match to PCCF	2	5.13	2	5.13
5  Note: Retired Postal Code	10	25.64	12	30.77
6  Note: Multi-match, random allocation	16	41.03	28	71.79
9  No error, note or warning	11	28.21	39	100.00

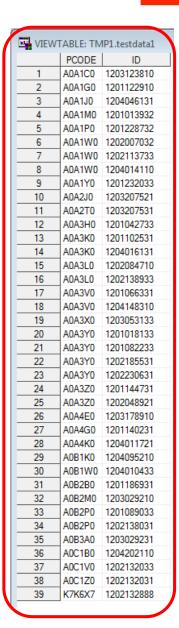
#### Same file with NO lower case letters... NO ERRORS

LINK	Frequency		Cumulative Frequency	Cumulative Percent
5  Note: Retired Postal Code	10	25.64	10	25.64
6  Note: Multi-match, random allocation	16	41.03	26	66.67
9  No error, note or warning	13	33.33	39	100.00

You can use EXCEL (or OPENREFINE) to convert to UPPER CASE

## **Step 1.2 Convert text file to SAS**





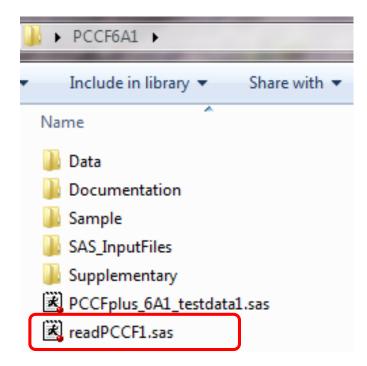
CANADA

POSTES



## Let's do this together

- 1. Open the SAS program 'readPCCF1.sas'
- 2. Modify the SAS program
- 3. Run the SAS program



C:\Users\TFDLLC221\Desktop\ACCOLEDS\_2015\PCCF6B1

#### Open, modify, and run 'readPCCF1.sas'



C:\Users\TFDLLC221\Desktop\ACCOLEDS\_2015\PCCF6B1

format PCODE \$6. ID \$10.;

label PCODE=PCODE ID=ID:

proc print;

```
readPCCF1.sas
   /* 'libname' sets 📆 a place (myownlib) on my hard drive where my da
   /* versus only reeping data for the duration of the SAS session
   libname myownlib 'C:\Users\moonj\Desktop\PCCF6A1\Sample';
     'filename' assigns a nickname (pcodes) to the text file
   filename pcodes "C:\Users\moonj\Desktop\PCCF6A1\Sample\pcodes1.txt";
   /* In this example, the data file is called 'pcodes1.txt'
      the file 'pcodes1.txt' must have a 6 digit Postal Code, a spe
      'data' statement & SAS
     commands to read Postal Codes
     into a SAS dataset
      Because of the two-part name, the dataset 'testdata1.sas7bda
   /* location derined by the libname statement ('C:\Users\moonj\D
   data myownlib.testdata1;
       length PCODE $ 6 ID $10;
                                  * defines variable lengths;
       infile pcodes;
                                  * tells SAS where to look for the data:
       input PCODE $ ID $;
                                  * inputs the data;
```

\* formats the variables:

ш

\* labels the variables:

'libname' statement to set up a place on the hard drive (myownlib) to store SAS datasets

'filename' statement to create a 'nickname' (pcodes) for the text file containing Postal Codes

#### Open, modify, and run 'readPCCF1.sas'



#### readPCCF1.sas

run;

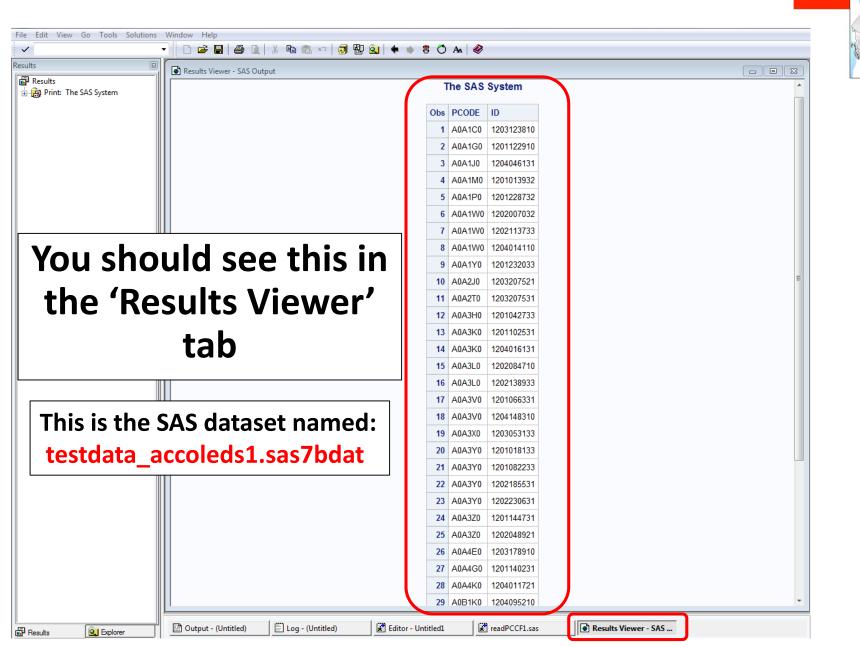




d Postal Codes and ID numbers into SAS for use with PCCF+ Version 6A1 libname myownlib 'C:\Users\TFDLLC221\Desktop\ACCOLEDS\_2015\PCCF6B1\Sample'; libname myownlib 'C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1\Sample'; containing data & tells SAS where to find it \*/ 'filename' assigns a nickname (pcodes) to the text file filename pcodes "C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\pcodes1.txt"; /\* In this example, the data file is called 'pcodes1.txt' /\* the file 'pcodes1.txt' must have a 6 digit Postal Code, a space, and an ID number, in each row filename pcodes "C:\Users\TFDLLCxxx\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\pcodes1.txt"; A0A1M0 1201013932 **DOUBLE QUOTES** A0A1P0 1201228732 A0A1W0 1202007032 The 'data' statement Because of the two-p data myownlib.testdata accoleds1; v hard drive in the /\* location defined by data myownlib.testdata accoleds1; \* defines variable lengths; length PCODE \$ 6 ID \$10; infile pcodes; \* tells SAS where to look for the data; input PCODE \$ ID \$; \* inputs the data; format PCODE \$6. ID \$7.; \* formats the variables: label PCODE=PCODE ID=ID; \* labels the variables; proc print;

#### **RESULT: a SAS dataset containing Postal Codes**

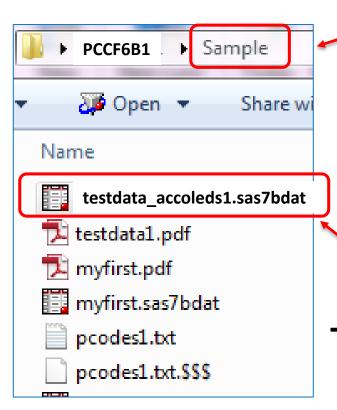
POSTES



#### Step 1 completed.

C:\Users\TFDLLC221\Desktop\ACCOLEDS\_2015\PCCF6B1\Sample





In your "Sample" directory
you should have a SAS dataset named

testdata\_accoleds1.sas7bdat

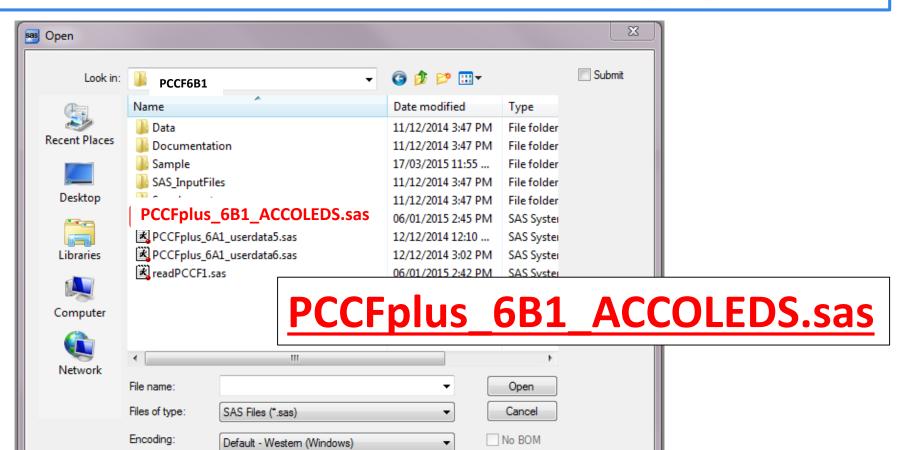
This contains *your* Postal Codes, and an ID variable. Next, we'll use the PCCF+ SAS program to match your Postal Codes to Census Geography

# Step 2. Next, use the PCCF+ SAS program to read this Postal Code dataset



From the main PCCF6B1 directory, open the PCCF+ SAS program:

C:\Users\TFDLLC221\Desktop\ACCOLEDS\_2015\PCCF6B1\PCCFplus\_6B1\_ACCOLEDS.sas



```
/* 1. Set the installation folder path
                                                                                                       POSTES
         - Folder path must NOT have any spaces
  %let installDir = C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1;
%let installDir = C:\Users\TFDLLC221\Desktop\ACCOLEDS_2015\PCCF6B1;
          - Default for the input file is a .sas7bdat file (regular SAS file)
libname inData "C:\Users\TFDLLC221\Desktop\ACCOLEDS_2015\PCCF6B1 \Sample\";
   * Data library (folder path where dataset is located);
   libname inData "C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\";
   * Data file name (no file extension permitted here);
   $let inFile = testdata accoleds1;
   *%let inFile = cchs pcodes;
                                  %let inFile = testdata_accoleds1;
   /* Code to read text file into a SAS dataset - remove stars at beginning of each line to active this code */
   *filename sampdat "&installDir.\Sample\sampledat.txt"; /* input data in text format */
   *data inData.&infile.;
       infile sampdat lrecl=256 stopover pad;
       input
                    PCODE
                                  $6.
                                              /* Postal Code
                                $15.;
   *run:
     %let outData = C:\Users\TFDLLC221\Desktop\ACCOLEDS_2015\PCCF6B1\Sample;
   * Output data library (folder path where processed data files will be output);
```

```
$let outData = C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\;
```

This output replaces the former HLTHOUT and GEOPROB files %let outName = accoleds out;

```
%let outName = accoleds_out;
```

```
%let pdfOutput = "C:\Users\TFDLLC221\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\summaryresults accoleds.pdf";
      INIS OUTPUT INCLUDES DOTH THE FORMER HAIHOUF AND GROPKUB FILES.
```

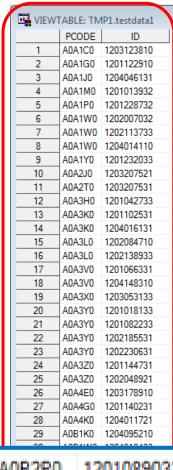
```
$let pdfOutput = "C:\Users\moonj\Desktop\ACCOLEDS 2015\PCCF6B1\Sample\summaryresults accoleds.pdf";
```

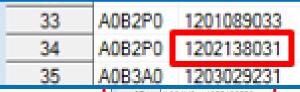
One more possible change needed...

# Your 'ID' variable may be a different length than the *default* in Statistics Canada's SAS program.

#### **Statistics Canada's SAS program:**

But our 'ID' variable is 10 characters long





37 A0C1V0 1202132033 38 A0C1Z0 1202132031 39 K7K6X7 1202132888

# Count the characters in your 'ID' (or just check how you read 'ID' into SAS)



#### readPCCF1.sas

```
readPCCF1.sas
                                                                                                            /* SAS Program to read Postal Codes and ID numbers into SAS for use with PCCF+ Version 6A1
   /*
   /* 'libname' sets up a place (myownlib) on my hard drive where my datasets will be stored
   /* versus only keeping data for the duration of the SAS session
   libname myownlib 'C:\Users\moonj\Desktop\PCCF6A1\Sample';
   /* 'filename' assigns a nickname (pcodes) to the text file containing data & tells SAS where to find it */
   filename pcodes "C:\Users\moonj\Desktop\PCCF6A1\Sample\pcodes1.txt";
   /* In this example, the data file is called 'pcodes1.txt'
   /* the file 'pcodes1.txt' must have a 6 digit Postal Code, a space, and In ID number, in each row
   /* A0A1C0 1203123810
      A0A1G0 1201122910
      A0A1J0 1204046131
      A0A1M0 1201013932
      A0A1P0 1201228732
      A0A1W0 1202007032
   /* The 'data' statement creates a dataset, 'te data1', in the library 'myownlib'
   /* Because of the two-part name, the dataset 'testdata1.sas7bdat' will be stored on my hard drive in the
   /* location defined by the libname strement ('C:\Users\moonj\Desktop\PCCF6A1\Sample')
   data myownlib.testdata1;
       length PCODE $ 6 ID $10;
                                     defines variable lengths:
       infile pcodes;
                                    * tells SAS where to look for the data:
       input PCODE $ ID $;
                                   * inputs the data;
       format PCODE $6. ID $10.;
                                   * formats the variables:
       label PCODE=PCODE ID=ID:
                                   * labels the variables:
   proc print;
   run:
```

#### We must be brave here...

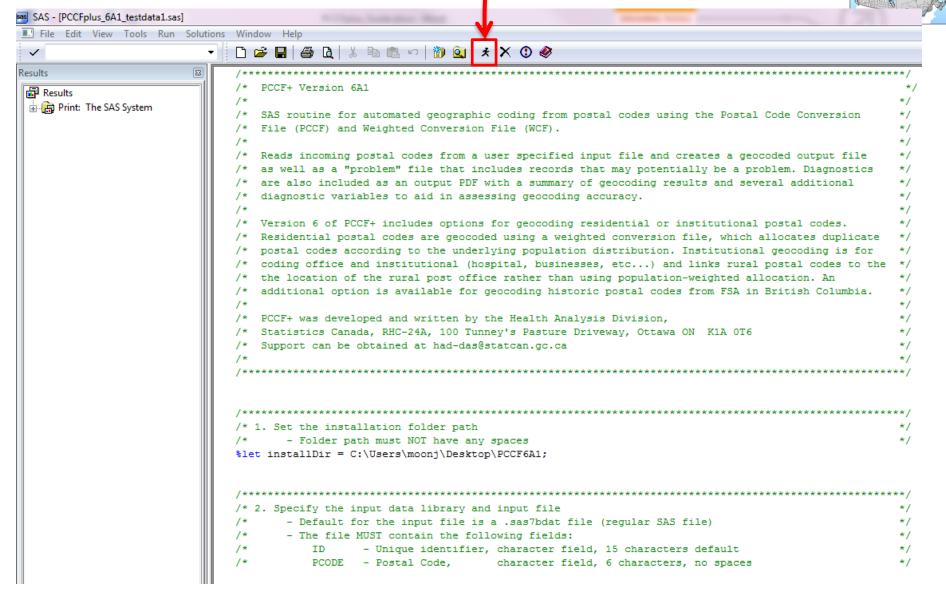


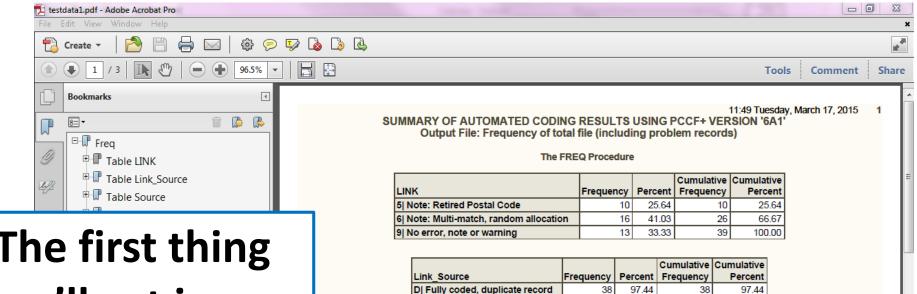
```
EDITS NOT NORMALLY REQUIRED BELOW THIS POINT
 /* Some options are specified below that can be edited
         - Random seed value
         - Length of unique identifier (ID) field
                                                  /* Read in data file with postal codes to be geocoded
data input data;
     * Unique identifier variable (ID) can be changed here;
     format ID $10. PCODE $6. FSA $3. LDU $3.; /* CHANGED INPUT FORMAT FOR ID TO '$10.' FROM '$15.'
     set inData.&inFile. (rename=(PCODE=inPCode));
     PCODE = compress(inPCode);
           = upcase(substr(inPCode, 1, 3));
     FSA
           = upcase(substr(inPCode, 4, 3));
     if PCODE='' then delete:
```

format ID \$10. PCODE \$6. FSA \$3. LDU \$3.;

We're now ready to run the PCCF+ program!







F| Fully coded, unique record

# The first thing we'll get is a PDF 'Summary of Automated Coding Results...'

Source	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1  Automated coding to 2011 Census	7	17.95	7	17.95
2  Geocoded using 2011 Census response	21	53.85	28	71.79
3  Converted from 2006 Census geography	10	25.64	38	97.44
4  Manually geocoded	1	2.56	39	100.00

2.56

100.00

PREC	Frequency		Cumulative Frequency	
5  1 or more DAs (DMT=H-Z)	38	97.44	38	97.44
7  1 DA (DMT=ABEG)	1	2.56	39	100.00

nCSD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1  1 CSD	19	48.72	19	48.72
2  2 CSDs	10	25.64	29	74.36
3  3 CSDs	4	10.26	33	84.62
4  4 CSDs	1	2.56	34	87.18
5  5 CSDs	1	2.56	35	89.74
6  6 CSDs	1	2.56	36	92.31
7  7 CSDs	3	7.69	39	100.00

Rep_Pt_Type	Frequency		Cumulative Frequency	
1  Block-face	3	7.69	3	7.69
2  Dissemination block	2	5.13	5	12.82
3  Dissemination area	34	87.18	39	100.00

## Links made:

#### Link source:

## More link details:

# Other information...

#### The FREQ Procedure

LINK	Frequency		Cumulative Frequency	Cumulative Percent
5  Note: Retired Postal Code	10	25.64	10	25.64
6  Note: Multi-match, random allocation	16	41.03	26	66.67
9  No error, note or warning	13	33.33	39	100.00

			Cumulative	Cumulative
Link_Source	Frequency	Percent	Frequency	Percent
D  Fully coded, duplicate record	38	97.44	38	97.44
F  Fully coded, unique record	1	2.56	39	100.00

Source	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1  Automated coding to 2011 Census	7	17.95	7	17.95
2  Geocoded using 2011 Census response	21	53.85	28	71.79
3  Converted from 2006 Census geography	10	25.64	38	97.44
4  Manually geocoded	1	2.56	39	100.00

			Cumulative	Cumulative
PREC	Frequency	Percent	Frequency	Percent
5  1 or more DAs (DMT=H-Z)	38	97.44	38	97.44
7  1 DA (DMT=ABEG)	1	2.56	39	100.00

nCSD	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1  1 CSD	19	48.72	19	48.72
2 2 CSDs	10	25.64	29	74.36
3  3 CSDs	4	10.26	33	84.62
4  4 CSDs	1	2.56	34	87.18
5  5 CSDs	1	2.56	35	89.74
6  6 CSDs	1	2.56	36	92.31
7  7 CSDs	3	7.69	39	100.00

Rep_Pt_Type	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1  Block-face	3	7.69	3	7.69
2  Dissemination block	2	5.13	5	12.82
3  Dissemination area	34	87.18	39	100.00

#### There are two more 'output' files

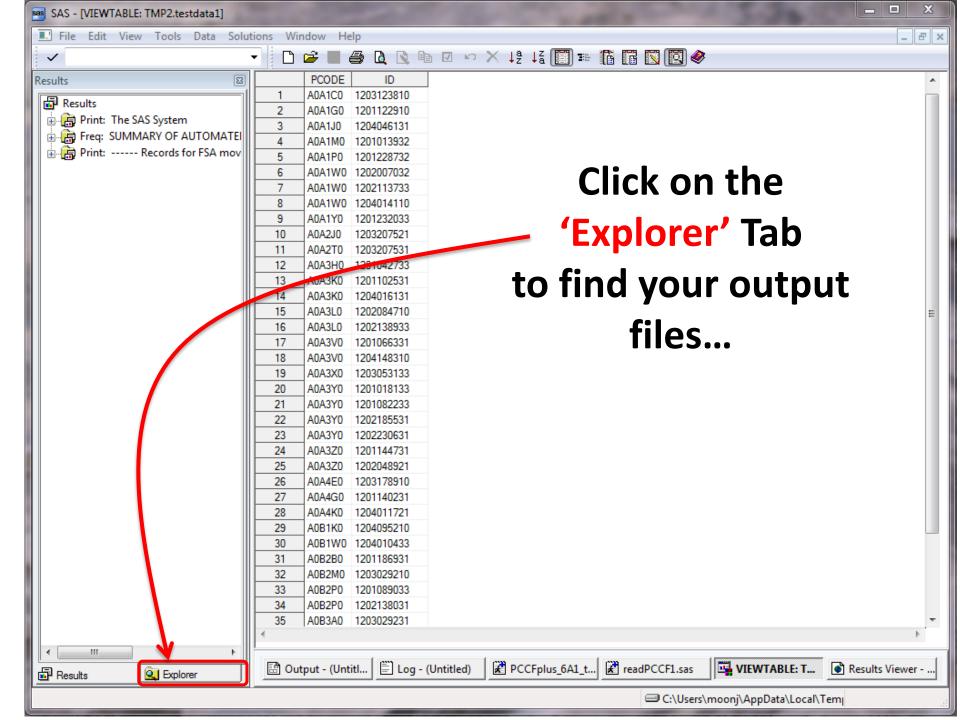


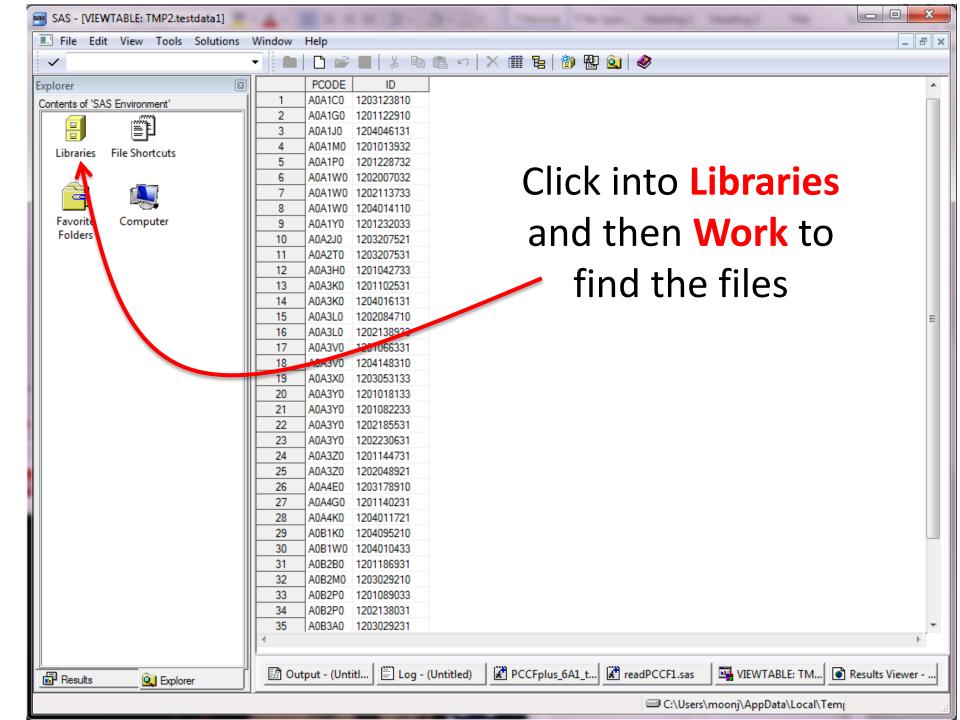
*First*, a

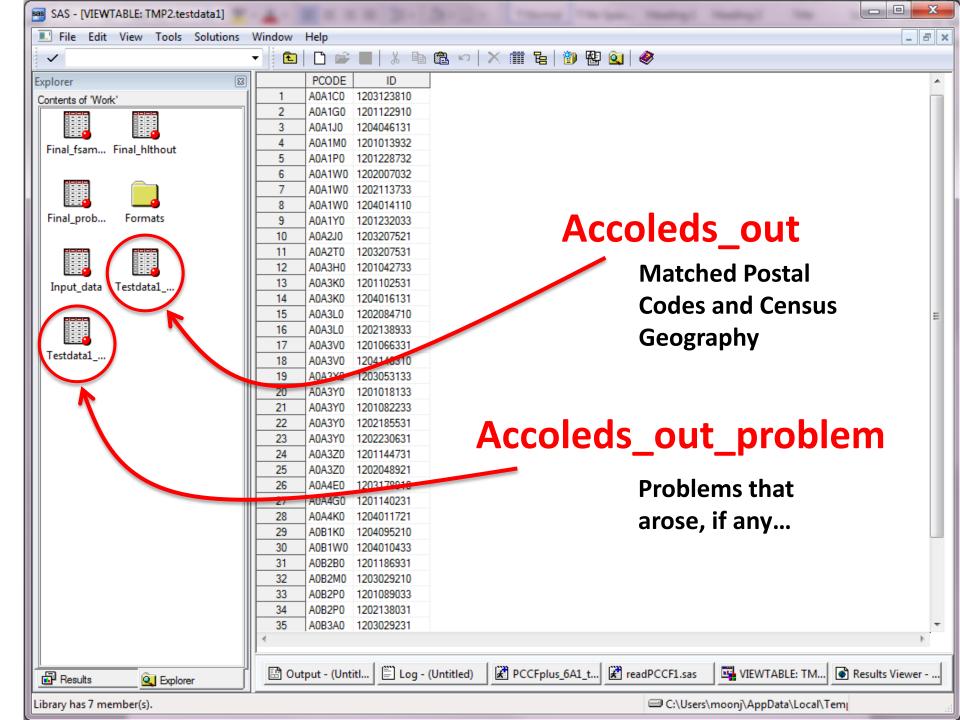
# Postal Code/Census Geography match file

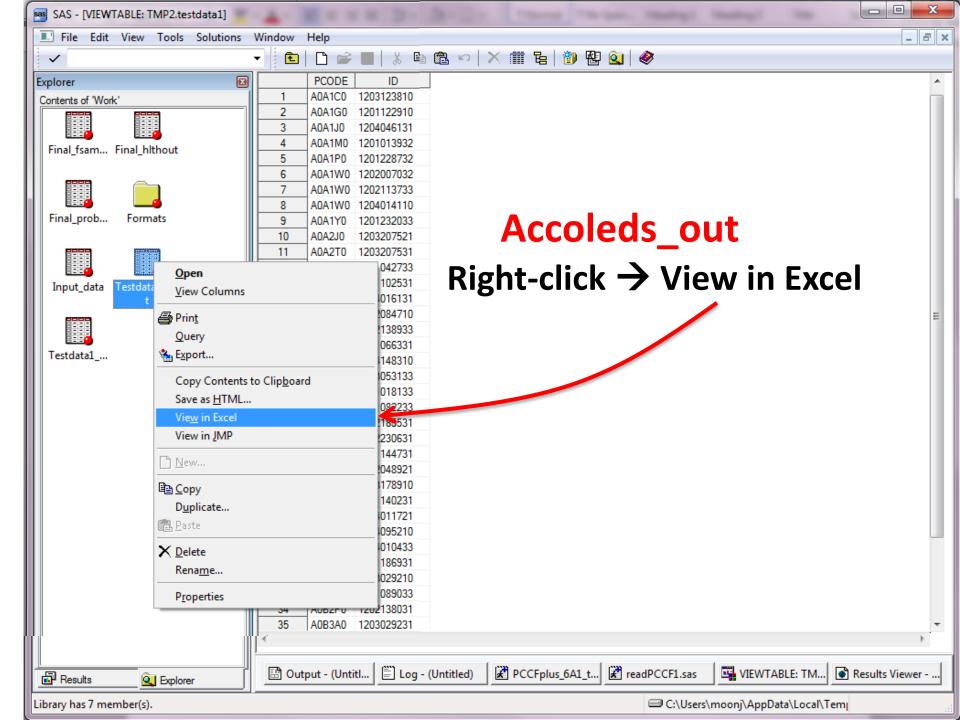
[OUR ULTIMATE GOAL]

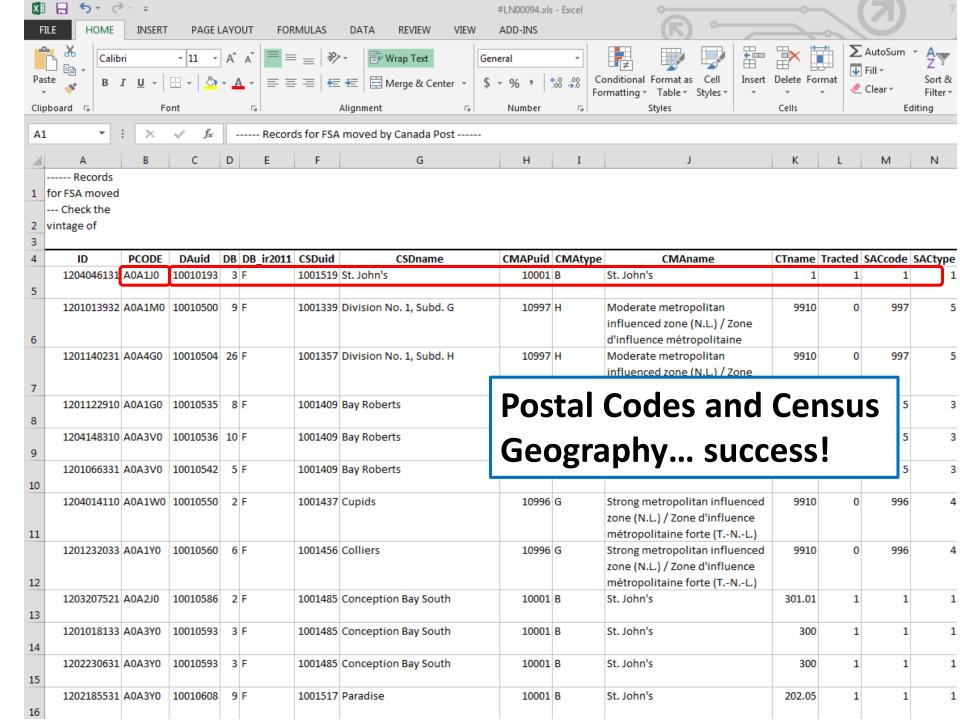
Second, a problem file













## And there is a *lot* of Census Geography

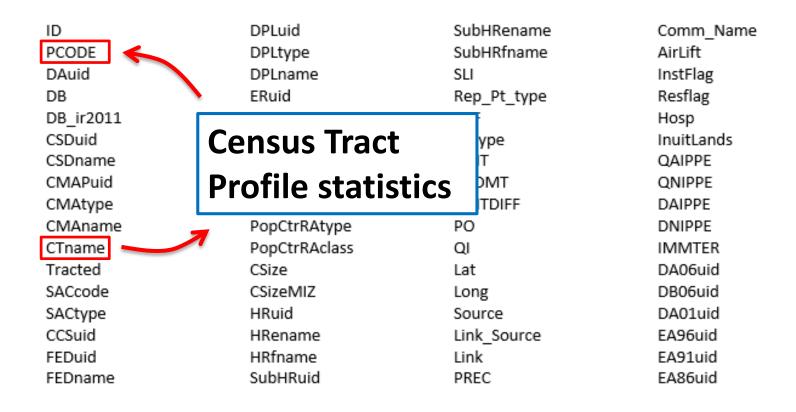
DPLuid	SubHRename	Comm_Name
DPLtype	SubHRfname	AirLift
DPLname	SLI	InstFlag
ERuid	Rep_Pt_type	Resflag
ERname	RPF	Hosp
CARuid	PCtype	InuitLands
CARname	DMT	QAIPPE
PopCtrRAPuid	H_DMT	QNIPPE
PopCtrRAname	DMTDIFF	DAIPPE
PopCtrRAtype	PO	DNIPPE
PopCtrRAclass	QI	IMMTER
CSize	Lat	DA06uid
CSizeMIZ	Long	DB06uid
HRuid	Source	DA01uid
HRename	Link_Source	EA96uid
HRfname	Link	EA91uid
SubHRuid	PREC	EA86uid
	DPLtype DPLname ERuid ERname CARuid CARname PopCtrRAPuid PopCtrRAtype PopCtrRAclass CSize CSizeMIZ HRuid HRename HRfname	DPLtype SubHRfname DPLname SLI ERuid Rep_Pt_type ERname RPF CARuid PCtype CARname DMT PopCtrRAPuid H_DMT PopCtrRAname DMTDIFF PopCtrRAtype PO PopCtrRAclass QI CSize Lat CSizeMIZ Long HRuid Source HRename Link_Source HRfname Link

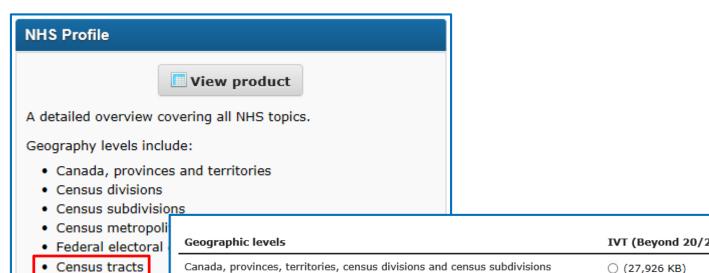


You can use Census geography to find Census demographics, and relate these to your Postal Codes

CANADA

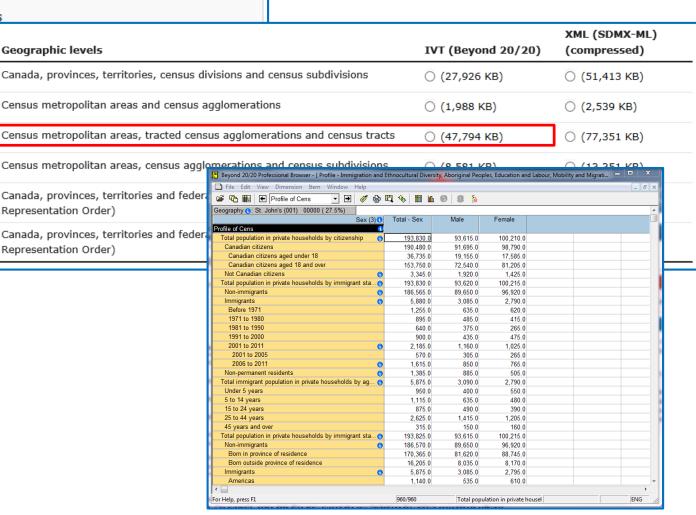
POSTES





Representation Order)

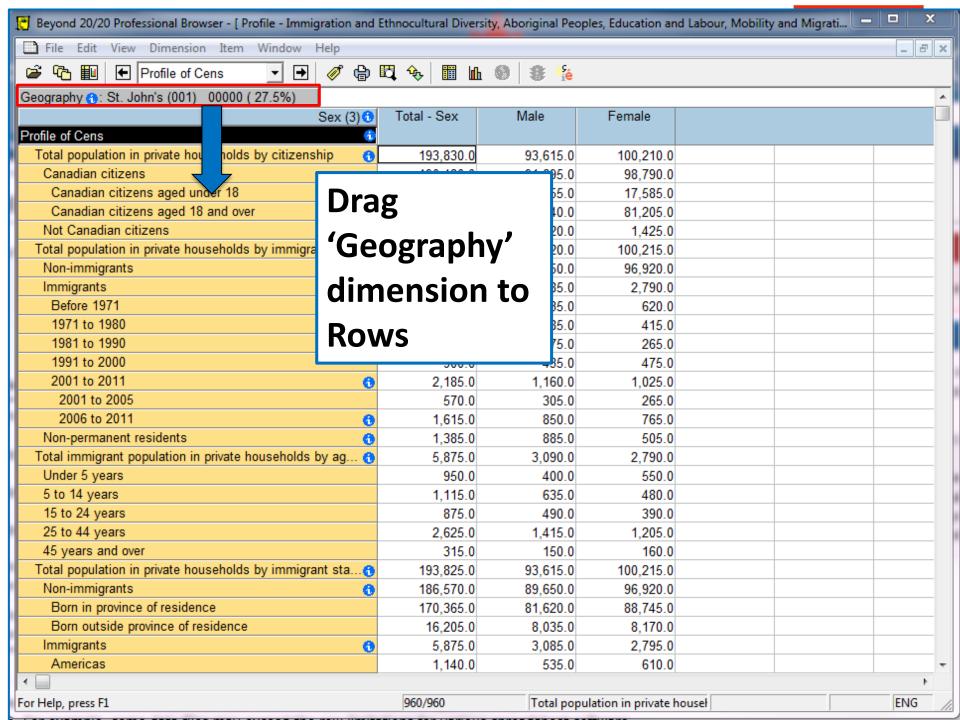
Representation Order)

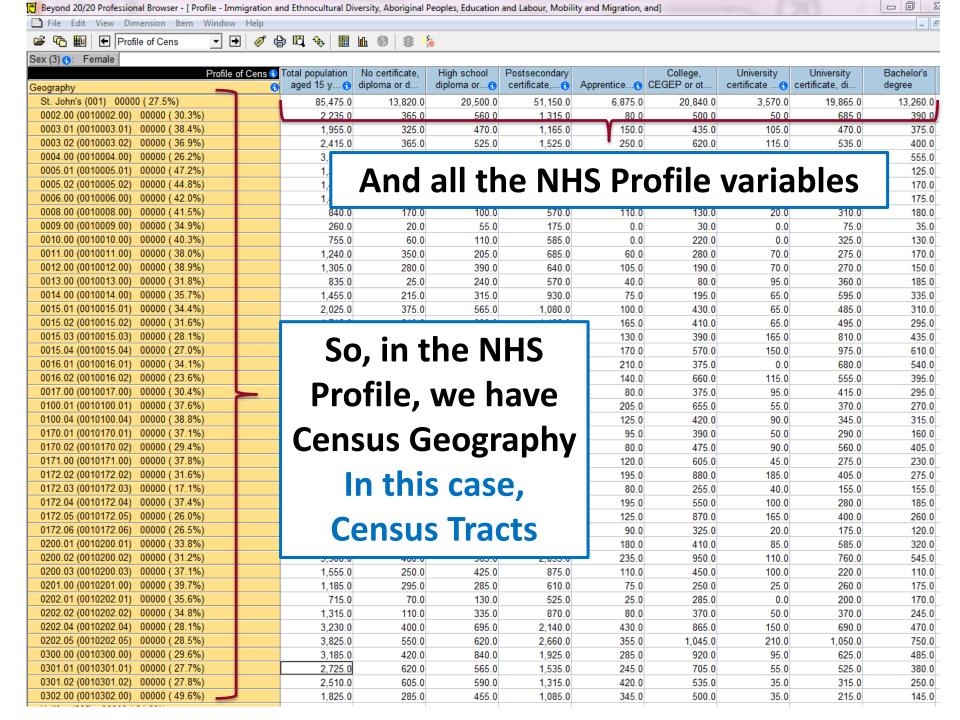


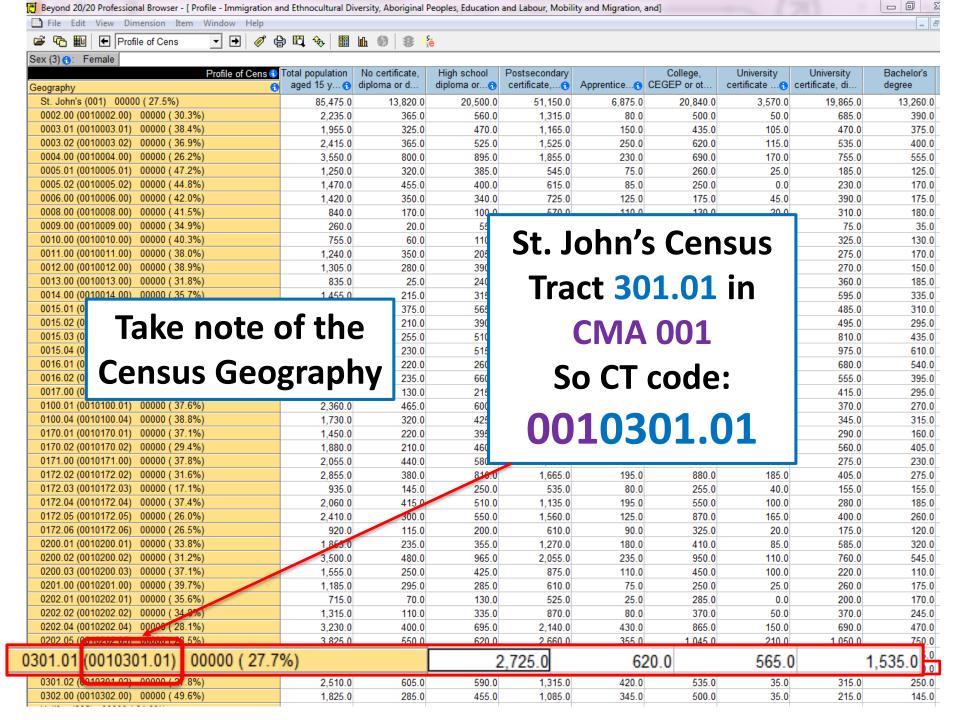
POSTES

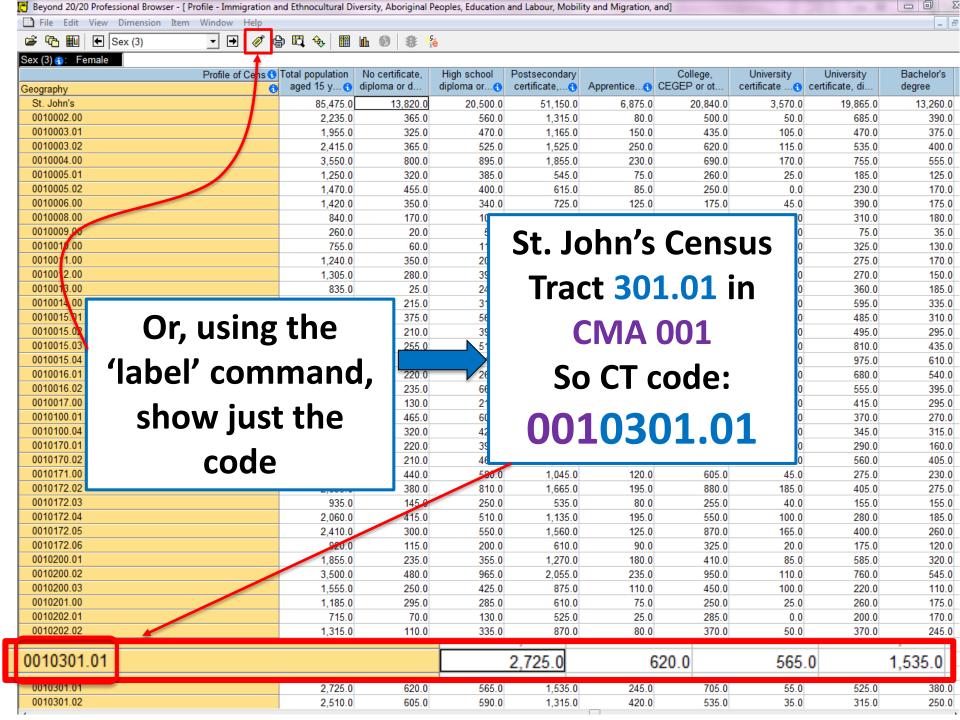
CANADA

CANADA





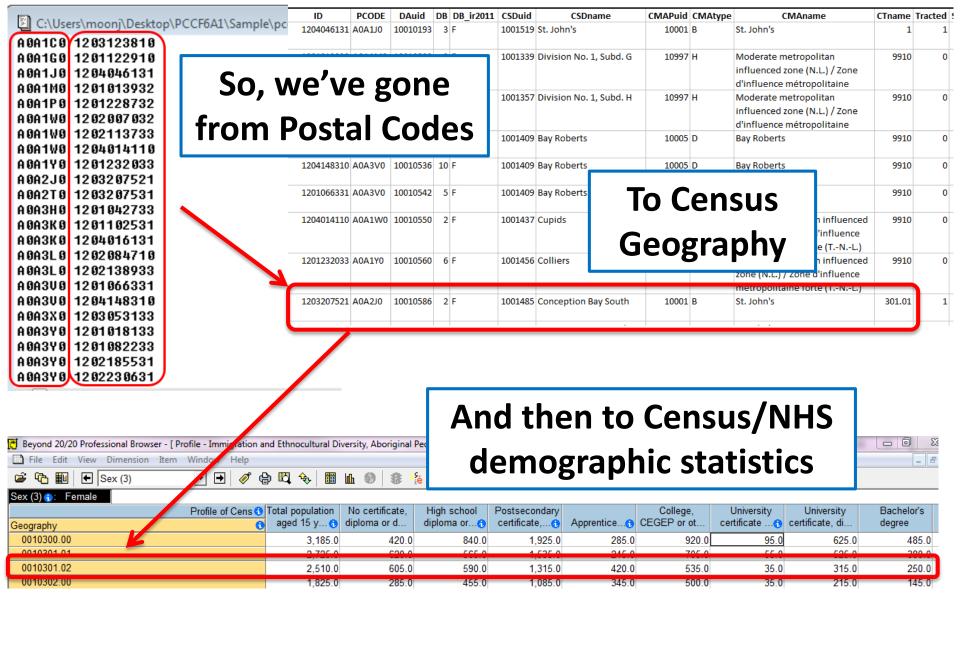




# Now, we take that Census Geography, and link it to our Postal Codes...



ID	PCODE	DAuid	DB	DB_ir2011	<b>CSDuid</b>	CSDname	CMAPuid	CMAtype	CMAname	CTname	Tracted
1204046131	A0A1J0	10010193	3	F	1001519	St. John's	10001	В	St. John's	1	1
1201013932	A0A1M0	10010500	9	F	S	t. John's	Cens	sus	Moderate metropolitan nfluenced zone (N.L.) / Zone d'influence métropolitaine	9910	0
1201140231	A0A4G0	10010504	26	F	1	Tract 30:		n	Moderate metropolitan nfluenced zone (N.L.) / Zone d'influence métropolitaine	9910	0
1201122910	A0A1G0	10010535	8	F	1	CMA	OOT		Bay Roberts	9910	0
1204148310	A0A3V0	10010536	10	F	1	So CT o	ode:		Bay Roberts	9910	0
1201066331	A0A3V0	10010542	5	F		001030	າ1	1	Bay Roberts	9910	0
1204014110	A0A1W0	10010550	2	F		00103	71.0		strong metropolitan influenced zone (N.L.) / Zone d'influence métropolitaine forte (TNL.)	9910	0
1201232033	A0A1/0	10010560	6	F	1001456	Colliers	10996	G	Strong metropolitan influenced zone (N.L.) / Zone d'influence métropolitaine forte (TNL.)	9910	0
1203207521	A0A2J0	10010586	2	F	1001485	Conception Bay South	10001	В	St. John's	301.01	1







http://library.queensu.ca/data/how\_to\_guides

Postal Code Conversion File Plus (PCCF+)

- <u>PCCF+ Guide</u> —: How to take Postal Codes in a text file, convert to a SAS dataset, and then run PCCF+ to match Postal Codes with Census geography
- readPCCF1.sas: Program to convert Postal Codes in a text file into a SAS dataset



Note: filenames & paths used in this guide may not match those used in YouTube.

**PCCF+ Reference Guide from Statistics Canada:** 

http://www.statcan.gc.ca/pub/92-154-g/92-154-g2013001-eng.htm



**Questions?**