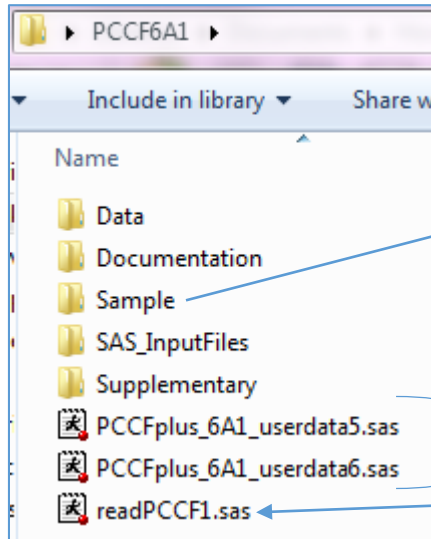


Overview: This guide illustrates how to convert an ASCII text file containing Postal Codes into a SAS dataset compatible with Statistics Canada's PCCF+ SAS program (version 6A1), and then how to run the PCCF+ program. SAS version 9+ and PCCF+ 6A1 must be loaded on your computer. If the PCCF+ files are 'zipped', you must 'unzip' them using **RIGHT-CLICK** → **EXTRACT ALL**. Once unzipped you should see something like this on your PC:

PCCF6A1 Directory:



<https://www.youtube.com/watch?v=ECoGzP26NbE>

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

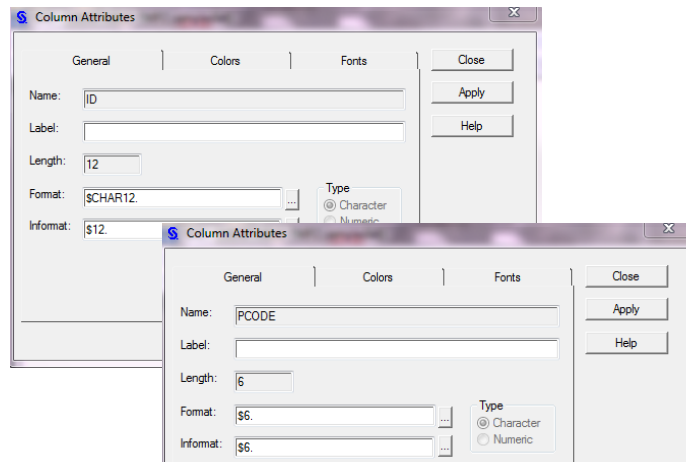
- **Data** – contains data used by PCCF+
- **Documentation** – user guides in English and French
- **Sample** – sample SAS datasets
- **SAS_InputFiles** – PCCF and Geocoding files used by PCCF+
- **Supplementary** – other files used by PCCF+

Sample SAS programs (originals provided by Statistics Canada).

SAS program to read ASCII data into SAS (*Appendix A, below*) – not provided by Statistics Canada.

Statistics Canada provides a **PCCF+ SAS program** that will work with a SAS dataset containing Postal Codes and ID numbers. This input file must have specific characteristics. For example, the *sample* SAS dataset provided by Statistics Canada takes the form shown below. We have to get *our* Postal Codes into this format.

	PCODE	ID
1		X00001
2		X00002
3		X00003
4		X00004
5	A0A1C0	1203123810
6	A0A1G0	1201122910
7	A0A1J0	1204046131
8	A0A1M0	1201013932
9	A0A1P0	1201228732
10	A0A1W0	1202007032
11	A0A1W0	1202113733
12	A0A1W0	1204014110



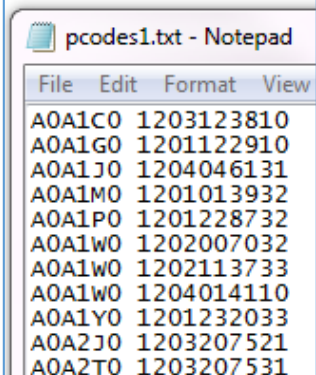
In this guide, we use a two-step process. **First**, we need to get our **Postal Codes** and **ID numbers** into a SAS dataset, similar to the one shown here, so it can be read by the PCCF+ SAS program. To do this, we'll use a simple SAS program (*provided in Appendix A*) to read in Postal Codes and ID numbers from a text file to create a new SAS dataset. **Second**, we'll modify the PCCF+ SAS program, provided by Statistics Canada, to read this new dataset.

Expected outputs will be: (i) a PDF file summary of the conversion (ii) a file containing your Postal Codes matched to Census geography, (iii) a 'problem' file.

A. Create a SAS dataset containing Postal Codes from a text file

We will use a SAS program to read in **Postal Codes** and **ID's** from an ASCII text file. We will use SAS to set the *length*, *format*, *name*, and *label* for each variable, to be compatible with the PCCF+ SAS program provided by Statistics Canada. In this example, the result of this process will be a SAS dataset named **'testdata1.sas7bat'** in the **'Sample'** directory.

1. Create a text file containing **6-character Postal Codes** (no space) and **up-to-15-character ID numbers**. This is your initial input data file. *Notepad is good for this*. Avoid word processors like MS-Word.
2. Write and run a **SAS program¹** to read these data into SAS and create a SAS dataset (named readPCCF1.sas in this example).



File	Edit	Format	View
AOA1C0	1203123810		
AOA1G0	1201122910		
AOA1J0	1204046131		
AOA1M0	1201013932		
AOA1P0	1201228732		
AOA1W0	1202007032		
AOA1W0	1202113733		
AOA1W0	1204014110		
AOA1Y0	1201232033		
AOA2J0	1203207521		
AOA2T0	1203207531		

SAS Program to read a text file into 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 an ID number, in each row
/* AOA1C0 1203123810
AOA1G0 1201122910
AOA1J0 1204046131
AOA1M0 1201013932
AOA1P0 1201228732
AOA1W0 1202007032
*/
/* The 'data' statement creates a dataset, 'testdata1', 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 statement ('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;
```

3. After running this program, you should find your Postal Codes in SAS dataset format in your **'Sample'** directory:

C:\Users\moonj\Desktop\PCCF6A1\Sample\testdata1.sas7bdat

This is where the PCCF+ SAS program will look for your Postal Code data in SAS format.

¹ See text version of SAS program in Appendix A. You can copy/paste this program into SAS, and then modify it to meet your needs.

B. Then use the PCCF+ SAS program to read this dataset, match Postal Codes with Census geography, and generate output.

1. Using SAS, open the PCCF+ version 6A1 SAS Program provided by Statistics Canada. In this program, there are several places you have make changes.

- Change the 'installation folder path' to the location where your SAS program and data are stored
- Change the 'libname statement' → path where your SAS Postal Code dataset is stored
→ "C:\Users\moonj\Desktop\PCCF6A1\Sample\"
- In the 'infile' statement, provide the name of the SAS dataset (no extension)
`%let inFile = testdata1;`
- Provide the name of an output file, which will be used to make the names of TWO files – one file showing Postal Codes matched to Census Geography, the other a '_problem' file
The filename used in this example is 'testdata1_out'
- Provide the path and filename for the summary PDF file, in this case:
"C:\Users\moonj\Desktop\PCCFplus\pccf6a1-fccp6a1\pccf6a1-fccp6a1\Sample\testdata1.pdf"

```
/* ***** */
/* 1. Set the installation folder path */
/* - Folder path must NOT have any spaces */
%let installDir = C:\Users\moonj\Desktop\PCCF6A1;

/* ***** */
/* 2. Specify the input data library and input file */
/* - Default for the input file is a .sas7bdat file (regular SAS file) */
/* - The file MUST contain the following fields: */
/* ID - Unique identifier, character field, 15 characters default */
/* PCODE - Postal Code, character field, 6 characters, no spaces */

* Data library (folder path where dataset is located);
libname inData "C:\Users\moonj\Desktop\PCCF6A1\Sample";
* Data file name (no file extension);
%let inFile = testdata1;

/* ***** */
/* 3. Output file name (the problem file is indicated with a "_problem" suffix) */
/* This output replaces the former HLTHOUT and GEOPROB files */
%let outName = testdata1_out;

/* ***** */
/* 4. Specify the output path and filename for the output summary (in PDF format) */
/* This output includes both the former HLTHOUT and GEOPROB files. */
%let pdfOutput = "C:\Users\moonj\Desktop\PCCF6A1\Sample\testdata1.pdf";

/* ***** */
/* 5. Version to run either Residential or Institutional coding (residential is default) */
/* - Indicate 0 if residential */
/* - Indicate 1 if institutional */
%let codeVersion = 0;

/* ***** */
/* 6. Code for old BC FSA moved by Canada Post. */
/* - Indicate 0 if NOT using old BC postal code locations (default) */
/* - Indicate 1 if using old BC postal code locations */
%let codeBC = 0;
```

Sample Program from Statistics

Change these last two steps as needed...

- Because the SAS dataset we created may not have the same 'format' as expected by the original Statistics Canada PCCF+ program, we may have to change another part of the program.

So, in this example, the *original* Statistics Canada program expects formats as follows:

PCODE \$6. ID \$15.

While *our* SAS dataset has the formats:

PCODE \$6. ID \$10.

NOTE: Check your data set for the actual format(s) for these variables. To do this, **open your SAS Postal Code dataset** → **right-click the ID column header** → **Choose 'Column Attributes'** → **take note of the 'Format'**. Use this in the PCCF+ SAS program.

So, we have to **change** the formats in the Statistics Canada sample program to match OUR dataset.

- Search the SAS program for the '**data input_data;**' statement.

Change the 'Format' statement to the formats associated with our SAS dataset. In this case, PCODE stays the same, but *ID must be changed from '\$15.'* to '*\$10.'*

```

/*****
data input_data;                               /* Read in data file with postal codes to be geocoded */
  * 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);
  FSA = upcase(substr(inPCode,1,3));
  LDU = upcase(substr(inPCode,4,3));
  if PCODE='' then delete;

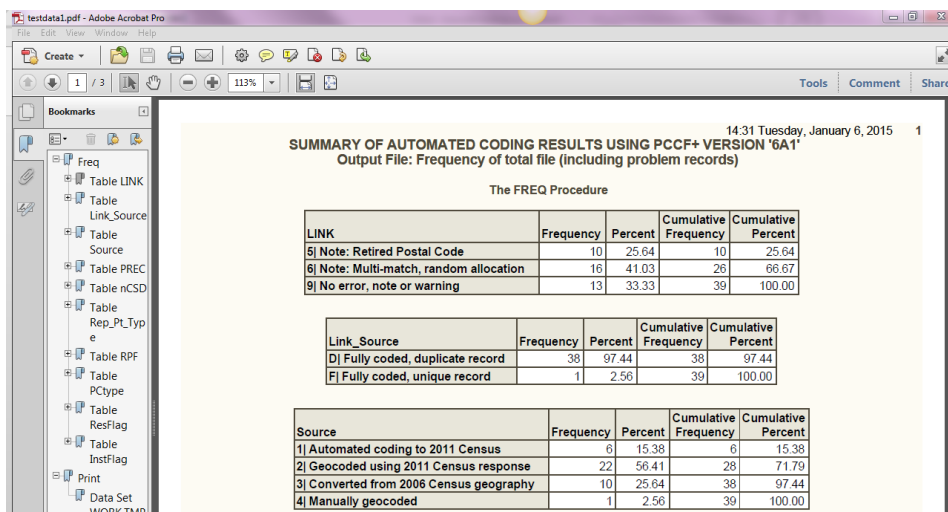
```

- Once you've made these changes, save the SAS program.

- Then, run the SAS program using the  icon.

- If the program runs successfully, you should get several files, including:

- In the 'Sample' directory, a summary in PDF format that looks something like this:



14:31 Tuesday, January 6, 2015 1

SUMMARY OF AUTOMATED CODING RESULTS USING PCCF+ VERSION '6A1'
Output File: Frequency of total file (including problem records)

The FREQ Procedure

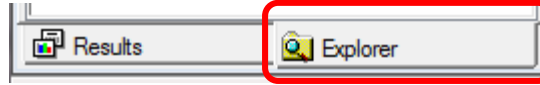
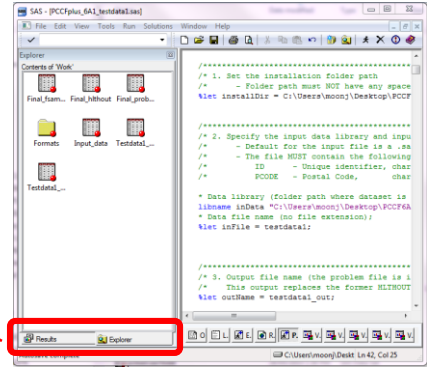
LINK	Frequency	Percent	Cumulative Frequency	Cumulative Percent
5] Note: Retired Postal Code	10	25.64	10	25.64
6] Note: Multi-match, random allocation	16	41.03	26	66.67
9] No error, note or warning	13	33.33	39	100.00

Link_Source	Frequency	Percent	Cumulative Frequency	Cumulative 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	6	15.38	6	15.38
2] Geocoded using 2011 Census response	22	56.41	28	71.79
3] Converted from 2006 Census geography	10	25.64	38	97.44
4] Manually geocoded	1	2.56	39	100.00

- b. **Two output files:** *first*, a Postal Code/Census Geography match file and *second*, a 'problem' file. These can be found using the 'Explorer' tab in SAS.

To see these two files, click on the 'Explorer' tab. Click into **Libraries** and then **Work** to find the files².



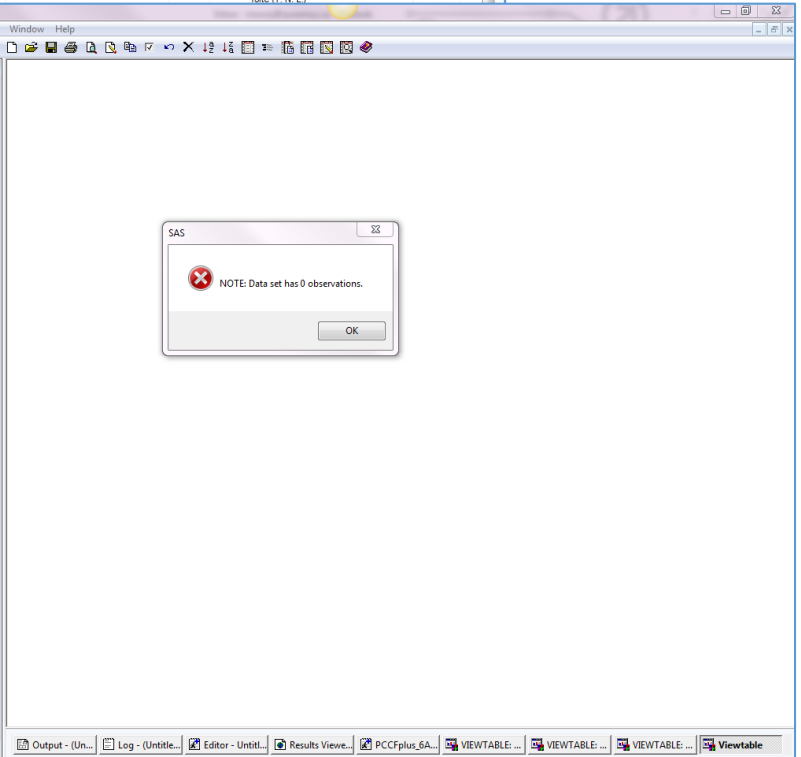
1

ID	PCODE	DAud	DB	DB_r2011	CSDud	CSDname	CMAJid	CMAtype	CMAName	CTName	Trac
1	1202084710	AOA3LO	10010192	02	F	1001519	10001	B	St. John's	0001.00	1
2	1204046131	AOA1JO	10010193	03	F	1001519	10001	B	St. John's	0001.00	1
3	1201013932	AOA1MO	10010500	09	F	1001339	10997	H	Division No. 1, Subd. G	9910.00	0
4	1201140231	AOA4GO	10010506	12	F	1001365	10997	H	Moderate metropolitan influenced zone (N.L.) / Zone d'influence métropolitaine modérée (T.-N.L.)	9910.00	0
5	1201122910	AOA1GO	10010535	08	F	1001409	10005	D	Bay Roberts	9910.00	0
6	1202007032	AOA1WO	10010555	11	F	1001415	10997	H	Moderate metropolitan influenced zone (N.L.) / Zone d'influence métropolitaine modérée (T.-N.L.)	9910.00	0
7	1201232033	AOA1YO	10010560	03	F	1001456	10996	G	Colliers	9910.00	0
8	1203207521	AOA2JO	10010581	05	F	1001485	10001	B	Conception Bay South	0301.02	1
9	1203207531	AOA2TO	10010583	02	F	1001485	10001	B	Conception Bay South	0301.02	1
10	1201082233	AOA3YO	10010603	02	F	1001485	10001	B	Conception Bay South	0300.01	1
11	1201018133	AOA3YO	10010607	01	F	1001517	10001	B	Paradise	0202.05	1
12	1202230631	AOA3YO	10010609	03	F	1001517	10001	B	Paradise	0202.05	1
13	1202185531	AOA3YO	10010667	01	F	1001485	10001	B	Conception Bay South	0300.01	1
14	1201042733	AOA3HO	10010580	01	F	1001551	10001	B	Pety Harbour-Maddox Cove	0100.03	1
15	1203123810	AOA1CO	10010684	12	F	1001557	10001	B	Bay Bulls	0110.10	1
16	1204011721	AOA4KO	10010686	07	F	1001559	10001	B	Willes Bay	0110.10	1
17	1202113733	AOA1WO	10010691	12	F	1001421	10996	G	Division No. 1, Subd. M	9910.00	0
18	1204014110	AOA1WO	10010692	41	F	1001421	10996	G	Division No. 1, Subd. M	9910.00	0
19	1203178910										
20	1201186931										
21	1202138031										
22	1201089033										
23	1204095210										
24	1203029210										
25	1203029231										
26	1201228732										

Right-click on the file icon and choose 'View in Excel' to export this file to Excel.

Work.testdata1_out
Contains matched Postal Codes and Census Geography

2
Work.testdata1_out_problem
Contains problem records that need attention.
NOTE: This may be an empty file if no problems are encountered



² Note: Filenames in these images may not match the example given, or in the YouTube video.

Appendix A. SAS Program to read ASCII data

In SAS, use **FILE** → **New Program**

Then **Copy/Paste** the following program into the blank window. Make any changes necessary to meet your local needs (i.e. *change libname and path, filename and path, etc.*).

```
/* 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 an ID number, in each row  */
/* A0A1C0 1203123810                                                                            */
/* A0A1G0 1201122910                                                                            */
/* A0A1J0 1204046131                                                                            */
/* A0A1M0 1201013932                                                                            */
/* A0A1P0 1201228732                                                                            */
/* A0A1W0 1202007032                                                                            */
/*                                                                                              */
/* The 'data' statement creates a dataset, 'testdata1', 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 statement ('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;
```