

How to Download Data and Import into SPSS using the SDA interface

This document describes the process of downloading a subset of microdata using the SDA interface version 3.+, and importing the subset into SPSS/Windows.

1. **Accessing files available in SDA.** From the listing of datafiles available in SDA, at <http://www.chass.utoronto.ca/datalib/major/sda.htm>, select an appropriate data file. For this example, we use the Census of Canada public use microdata files/'2001 public use microdata file of individuals'. Click on the 'X' under 'Data' to open SDA. In the SDA interface, from the Main Menu bar at the top of the screen, run you mouse over "Download" and select "Customized subset," as shown below:

The screenshot shows the SDA interface for the study "Census of Canada, 2001: public use microdata on individuals (Second rev ed 2006.04)". The "Download" menu is open, and "Customized Subset" is selected. The interface includes a "Variable Selection" section with a "Selected:" field and a "View" button. Below this is a "Copy to:" section with buttons for "Row", "Col", "Ctrl", and "Filter". The "Mode:" section has radio buttons for "Append" and "Replace". A tree view on the left lists various variables such as "Survey administration", "Geographic identifier", "Household type, size and relationships", "Dwelling characteristics", "Economic family status and size", "Census family status and size", "Respondent - demographic variables", "Household income", "Census family income", "Economic family LICO status", and "Sample weight". The "SDA Frequencies/Crosstabulation Program" section includes a "Help:" link and fields for "REQUIRED Variable names to specify" (Row, Column, Control), "OPTIONAL Variable names to specify" (Selection Filter(s)), and "Weight:" (weightp - Individual's weighting factor). The "TABLE OPTIONS" section includes checkboxes for "Percentaging:" (Column, Row, Total), "Statistics", "Question text", "Color coding", "Suppress table", and "Show Z-statistic". The "CHART OPTIONS" section includes a "Type of chart:" dropdown (Stacked Bar Chart), "Bar chart options:" (Orientation: Vertical, Horizontal; Visual Effects: 2-D, 3-D), "Show Percents:" checkbox, "Palette:" dropdown (Color, Grayscale), and "Size:" fields for width and height. At the bottom are "Run the Table" and "Clear Fields" buttons.

2. **Specifying the variables and cases to include in the subset.** In this example, we will select only those respondents who live in Ontario:

- Under "**Select Files to construct**", select "SPSS." (You may select more than one type simultaneously.)

SDA Customized Subset of Variables/Cases
 Help: [General](#)

Choose subset specifications below. Then press "Continue" at bottom of form.

Select FILE(S) to construct:

[Data file](#) (ASCII)
[Delimiter](#) between variables: None Blank Comma

[Codebook](#) for subset data (ASCII)

Data definitions for:
 SAS SPSS STATA DDI (XML) SDA (DDL)

- Under **"Select Cases to Include"**, enter "prov(35)" to filter just the cases for province of Ontario. These data files can be very large and take some time for the SDA website to create, so it's a good idea to use a filter to remove cases you are not interested in. You may filter cases by any variable in the datafile.

Select CASES to include:

[Selection Filter\(s\)](#): Example: *age(18-50)*

Under **"Select Variables to Include"**, there are two alternatives: to select 'All' or 'Some' of the variables in a major category, or to list individual variables by name in the dialogue box.

- Alternative 1: You may select 'All', 'Some', or 'None' of the variables in each major group; the default is 'None'. Click on the label of the major group to determine what variables it contains. When done selecting, click on the 'Continue' button to proceed to the next page:.

Select VARIABLES to include (individually and/or by group):
 (Note: CASEID is always included)

Enter names of [individual variables](#) to include:

You may select [All](#) or [Some](#) variables from the following groups:

All Some None [Survey administration](#)
 All Some None [Geographic identifier](#)
 All Some None [Household type, size and relationships](#)
 All Some None [Dwelling characteristics](#)
 All Some None [Economic family status and size](#)
 All Some None [Census family status and size](#)
 All Some None [Respondent - demographic variables](#)
 All Some None [Household income](#)
 All Some None [Census family income](#)
 All Some None [Economic family LICO status](#)
 All Some None [Sample weight](#)

- When selecting only 'Some' variables in a major category, using the list of major categories you will be prompted in the next screen to select variables in the major categories, for inclusion in the subset. Use <Ctrl> and click on more than one variable:

SDA Customized Subset of Variables/Cases
 Help: [General](#) / [Selecting multiple variables from a list](#)

Select the variables you want to include in your subset from the following groups. Then press "Continue" at bottom of form.

Select Variables from Groups:

Respondent - demographic variables

agep : Age
 sexp : Sex
 marstlp : Legal marital status
 marsthp : Marital status historical comparability indicator
 pobp : Place of birth (all Canada)
 pobpa : Place of birth (excluding Atlantic provinces and Territories)
 pobmthp : Place of birth of mother
 pobfthp : Place of birth of father
 genstpob : Generation status
 citothp : Citizenship other than Canadian (all Canada)

Household income

hhincp : Total household income groups (all Canada)
 hhincpa : Total household income groups (excl females, Atlantic provs & Territories)

- Alternative 2: In order to select variables specifically by variable name, you must know the correct name of each variable, as it has been defined in SDA. Enter the variable names of the variables in the dialogue box. In this example, we have selected the following variables: age (agep), sex (sexp), visible minority (visminp), ethnic group (ethnicr) and household income (hhincp). The resulting window should look like that shown below. Go to the bottom of this page and click "Continue."

Select VARIABLES to include (individually and/or by group):

(Note: CASEID is always included)

Enter names of [individual variables](#) to include:

agep sexp visminp ethnicr hhincp

You may select [All](#) or [Some](#) variables from the following groups:

All Some None [Survey administration](#)

All Some None [Geographic identifier](#)

All Some None [Household type, size and relationships](#)

All Some None [Dwelling characteristics](#)

- Go to the bottom of the page and click "Continue".

3. Checking the subset specification and creating the subset. The following window "Check the Subset Specifications" allows you to review the information you have entered and make sure it is correct:

Check the Subset Specifications

The specifications you have chosen are listed below. Please check that they are correct before continuing.

If the specifications are *NOT* correct

- Back up to a previous page;
- Correct your entries;
- Re-press the "Continue" button on that page.

If the specifications are correct, just press the "Create the Files" button below.
Please be patient if the original data file is large.

[Create the Files](#)

Files to create:

Data file(Delimiter = none)
SPSS syntax file

Filters for selecting cases:

provp(35)

Individual variables specified (including partial groups):

CASEID (always included)
agep
sexp
visminp
ethnicr
hhincp

Click the "Create the Files" button to continue.

4. Downloading the subset to your workstation. After a few moments, a screen titled "Download Files" will appear. Notice there are links to the files that have been created. You must now save the "Data file" and "SPSS syntax file" to your work station.

SDA Customized Subset of Variables/Cases
Help: [General](#)

Download Files

6 variables for 305013 cases in subset.
Data subsetting is now complete. The files are ready to view and/or download to your computer. The following procedures work with most browsers:

To view a file:
Click on its link.
To save it to disk while viewing it select the "File/Save As..." menu option.

To immediately save a file to disk (without viewing it first):
Right-click on the link, then click on "Save Target As..." (Internet Explorer) or "Save Link As..." (Firefox)
(On a Macintosh, click on the link with the option key held down.)

- [Data file](#)
- [SPSS syntax file](#)

Save the files to your hard drive by mousing over the each file link, clicking the right mouse button, and selecting "Save Target As" from the mouse drop down menu. Alternatively, click on each filename, and use 'File/Save as' in your web browser menu bar to save the file. The files will look like this:

Data file:

```
53634 37.036845 81516
53635 37.0368451111511
53636 36.9357257215 9
53637 37.016812512517
53638 37.036845 51515
53639 36.914764722513
53640 37.019517561523
53641 37.050256442516
53642 37.050256392523
53643 37.129732341515
53644 37.036845121515
53645 37.036845 41513
53646 36.9357257615 9
53647 37.1038031525 8
53648 37.016812492516
53649 37.016812472523
53650 36.935725721510
53651 37.0168125925 9
53652 37.019517541520
53653 37.050256422522
53654 37.050256352510
53655 37.129732401511
53656 37.019517471512
53657 37.103803172514
53658 36.914764652514
53659 36.9357256815 9
53660 37.019517521521
53661 37.036845121511
53662 36.992209 82511
53663 36.992209132514
53664 37.016812482513
53665 37.019517481513
53666 36.914764682516
```

SPSS syntax file:

```
/* SPSS Data Definition File
/* Created by ddlttox on May 07, 2007 (Mon 11:59 AM EDT)
/* DDL source file: "/web/htdocs/sdaweb/tmpdir/AAr5aG5m.txt".

TITLE "Census of Canada, 2001: individual pumf. Second revision.".
COMMENT DDL indicates that dataset record length (reclen) is 22 columns.

DATA LIST fixed records=1
  FILE="x" /* Replace 'x' with full path name of your input data file.
  /1 CASEID 1-6 weightp 7-16 (6) agep 17-18
    sexp 19 visminp 20 hhincp 21-22
.

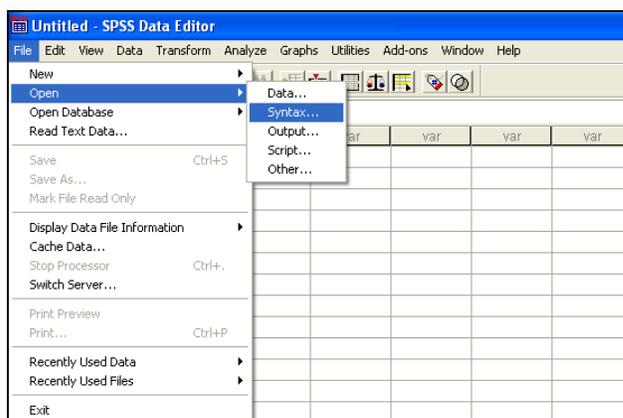
VARIABLE LABELS
  CASEID 'Case Identification Variable' /
  weightp 'Individuals weighting factor' /
  agep 'Age' /
  sexp 'Sex' /
  visminp 'Visible minority indicator' /
  hhincp 'Total household income groups (all Canada)' /
.

VALUE LABELS
  agep 85 '85 years of age and over' 98 'Not available' /
  sexp 1 'Female' 2 'Male' /
  visminp 1 'Chinese' 2 'South Asian' 3 'Black' 4 'Other visible minority'
  5 'Not a visible minority' 8 'Not available' /
  hhincp 1 'Loss' 2 'No income' 3 '$1 - $1,999' 4 '$2,000 - $4,999'
  5 '$5,000 - $7,999' 6 '$8,000 - $9,999' 7 '$10,000 - $14,999'
  8 '$15,000 - $19,999' 9 '$20,000 - $24,999'
  10 '$25,000 - $29,999' 11 '$30,000 - $34,999'
  12 '$35,000 - $39,999' 13 '$40,000 - $44,999'
```

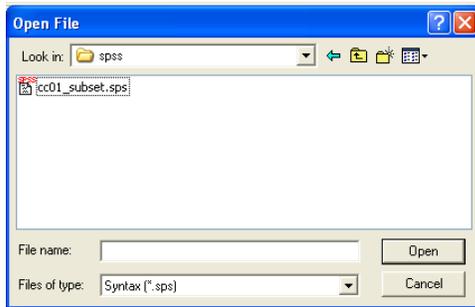
Select a subdirectory on your hard drive in which to save the files; we suggest **saving the data file with a .dat extension** (such as 'cc01_subset.dat') and the **SPSS syntax file with a .sps extension** (for example, 'cc01_subset.sps'). It be easier to remember at a later date which syntax file belongs with which data file if you save both files in the same subdirectory, and give both files the same root variable name – only the file extensions must be different.

5. Load the subset into SPSS.

- Open SPSS, select 'File/Open/Syntax' from the SPSS menu bar, and navigate to the subdirectory in which you saved the data and SPSS syntax file.



- Open the SPSS syntax file,



- On the 'Data list' command, enter the full path and filename of the data file that you saved (in 4. above). For example, here we have saved the data file in 'c:\temp\spss\cc01_subset.dat' and the SPSS syntax file in 'c:\temp\spss\cc01_subset.sps'. We need to edit the syntax file by deleting the 'X' in the 'Data list' statement, and replacing it with the path and filename of the data set:

```

/* SPSS Data Definition File
/* Created by ddlt0x on May 07, 2007 (Mon 11:59 AM EDT)
/* DDL source file: "web/htdocs/sdaweb/tmpdir/AAr5aG5m.txt".

TITLE "Census of Canada, 2001: individual pumf. Second revision."
COMMENT DDL indicates that dataset record length (reclen) is 22 columns.

DATA LIST fixed records=1
FILE="c:\temp\spss\cc01_subset.dat" /* Replace 'x' with full path name of your input data file.
/1 CASEID 1-6 weightp 7-16 (6) agep 17-18
sexp 19 visminp 20 hhincp 21-22
.

VARIABLE LABELS
CASEID 'Case Identification Variable' /
weightp 'Individuals weighting factor' /
agep 'Age' /
sexp 'Sex' /
visminp 'Visible minority indicator' /
hhincp 'Total household income groups (all Canada)' /
.

VALUE LABELS
agep 85 '85 years of age and over' 98 'Not available' /
sexp 1 'Female' 2 'Male' /
visminp 1 'Chinese' 2 'South Asian' 3 'Black' 4 'Other visible minority'
5 'Not a visible minority' 8 'Not available' /
hhincp 1 'Loss' 2 'No income' 3 '$1 - $1,999' 4 '$2,000 - $4,999'
5 '$5,000 - $7,999' 6 '$8,000 - $9,999' 7 '$10,000 - $14,999'
8 '$15,000 - $19,999' 9 '$20,000 - $24,999'
10 '$25,000 - $29,999' 11 '$30,000 - $34,999'
12 '$35,000 - $39,999' 13 '$40,000 - $44,999'
14 '$45,000 - $49,999' 15 '$50,000 - $54,999'
16 '$55,000 - $59,999' 17 '$60,000 - $64,999'
18 '$65,000 - $69,999' 19 '$70,000 - $74,999'
20 '$75,000 - $84,999' 21 '$85,000 - $99,999'
22 '$100,000 - $119,999' 23 '$120,000 or more'
99 'Not applicable' /

```

- Save the edited SPSS syntax file, and select 'Run/All' on the SPSS syntax editor menu bar.
- Once this syntax has run successfully, you can open the SPSS data editor window and check that the file has been read correctly. Then save the resulting SPSS system file (select 'File/Save as from the data editor menu bar), with a .sav extension (eg. 'c:\temp\spss\cc01_subset.sav').