

How to use Census data on E-STAT with *Excel* for a data analysis project

Is it worth completing university?

What the Census tells us about the correlation between Education and Income (using 2001 Census data)

by Joel Yan, Statistics Canada, mdm4u@statcan.ca, 1-800-465-1222

Assignment

Investigate the association between the proportion of people with low levels of education and average census tract income using the census data for London, Ontario. Using census tract data on E-STAT, create a scatter plot of the relationship between average income versus the percent of the population over 15 with less than grade 9 education for census tract areas. Then we import the data into Excel in order to compute the regression equation and correlation statistics.

E-STAT has over 1700 census characteristics for each census tract (and municipality) in Canada that can be manipulated within E-STAT and imported into Excel or Fathom to explore relationships. This activity is intended to provide an example of how this can be done.

Related Expectations for the Ontario Grade 12 Mathematics of Data Management Course:

- Solve problems involving complex relationships with the aid of diagrams (ODV.02, *Organization of Data for Analysis – Overall Expectations – page 49 in the Ontario curriculum*)
- Describe the relationship between 2 variables by interpreting the correlation coefficient (STV.04, *Statistics – Overall Expectations – page 52*)
- Organize and summarize data from secondary sources (e.g. the Internet) using technology (ST1.04, *Statistics – Specific Expectations on Collecting Data – page 52*)
- Calculate the correlation coefficient for a set of data using statistical software (ST4.02, *Statistics – Describing the relationship between 2 variables – page 52*)

Procedure

Finding the right table using Search Census

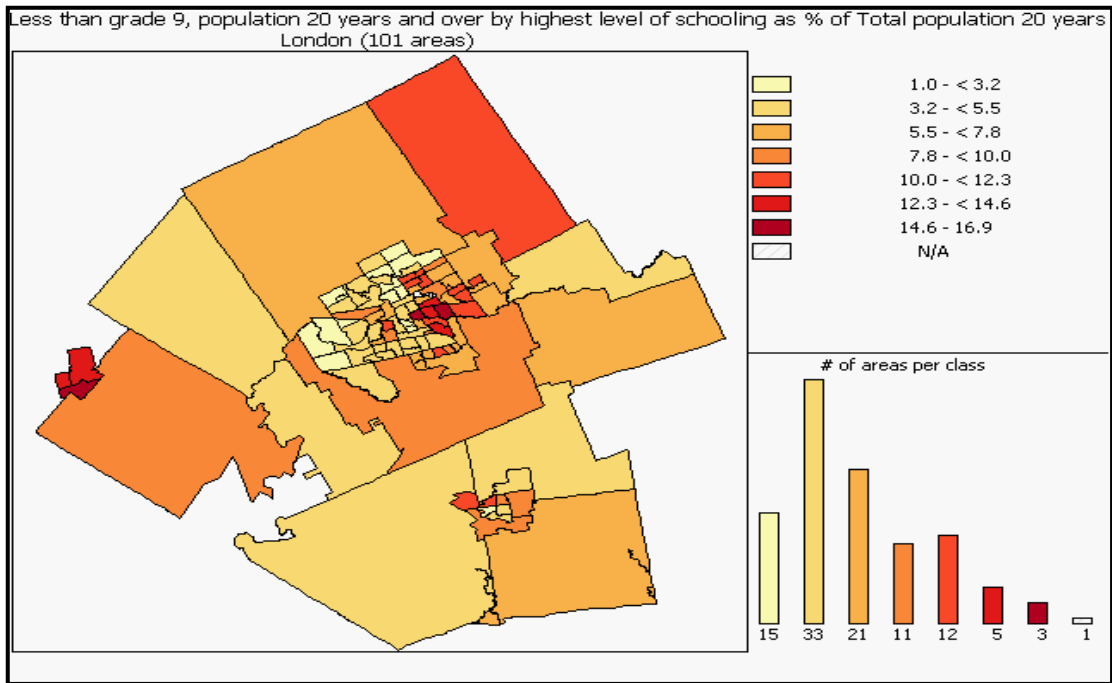
1. Start up E-STAT – at <http://estat.statcan.ca>
2. From the left tool bar click **Search Census**.
3. Under **Census**, select “2001 Census” and click “Go!”
4. Click the down arrow and select “2001 Census of Population (46 Large Urban Centres, [Census Tracts](#) (neighbourhoods))” and click “Go!”
5. Under **Profile selection**, select “2001 School Attendance, Education, Field of Study, Highest Level of Schooling and Earnings” and click “Go!”.

Geography and Characteristics selection

6. Under **Geography**, select “ Large Urban Centres in Ontario - 2001 – London (101 areas)” or another urban centre without too many census tracts
7. Under **Characteristics**, click “View Checklist”
8. Select the following variables from the checklist:
 - “Total population 20 years and over by highest level of schooling”
 - “Less than grade 9, population 20 years and over by highest level of schooling”
 - “With bachelor’s degree or higher, university, population 20 years and over by highest level of schooling”
 - “Average employment income \$, worked full year, full time, population 15 years and over with employment income”

Tip: Because the list of variables is long, use the “Find on this page” key from the Edit pull-down menu on your browser. For example, do a find on ‘schooling’, and then on ‘income’

9. Click the **Home** or **End** keys, and click on “Back to Main Selection Form”.
10. Scroll down and select from **Download to a file**: “Tab Delimited File”
11. Within the **File Download** box select “Save”.
12. Within the **Save As File name** box, select drive C, and enter the file name as “London Education vs. Income”. Click “Save”, then click “Close”.
13. E-STAT has basic computation and mapping capability. As an additional activity click on the **Go Back Arrow** to return to the “2001 School Attendance, Education, Field of Study, Highest Level of Schooling and Earnings.”
14. To see the geographic distribution of the first variable as a percent of the total population, we first click the “Table, Areas as rows” icon. Then click the radio button “Data as % of 1st characteristic” and click the “Redisplay as” tab. We now click the Map icon. By default this produces a colour shaded map with 7 classes. This can be useful in identifying census tract areas with the selected characteristics.



Load File in Excel

15. Open Excel
16. File / Open / for **File of type**: select “All Files“
17. Select “London Education vs. Income.TAB.”
18. The “Text Import Wizard” will appear; accept the default by pressing Next / Next / Finish
19. Your Excel worksheet should appear as follows:

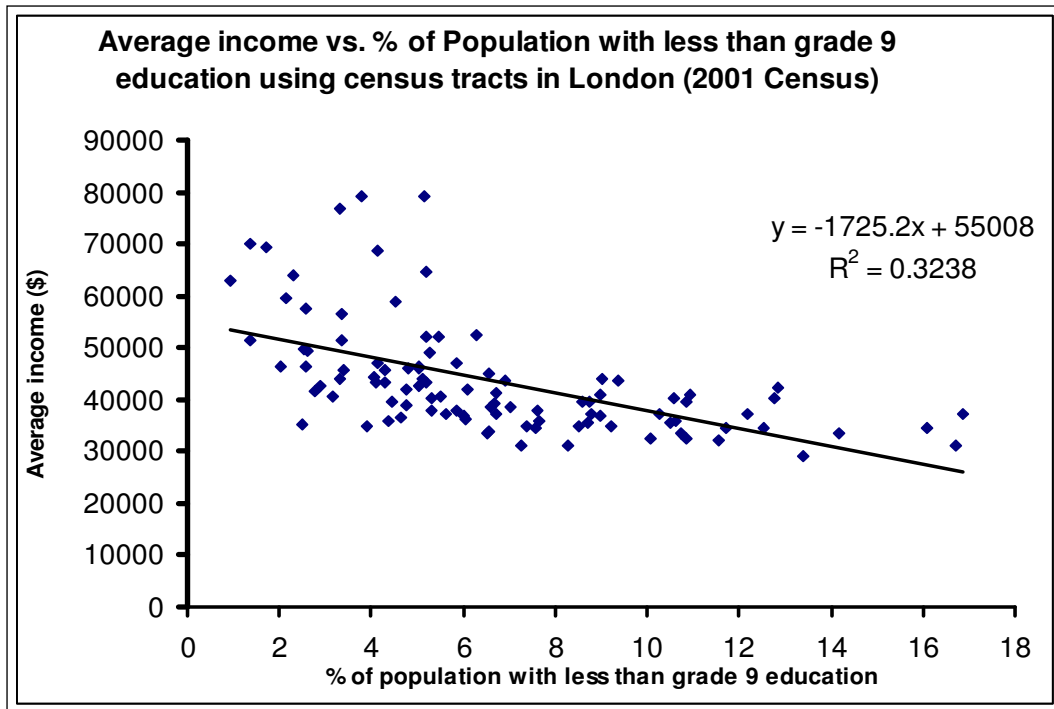
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Geo	#1	#2	#3	#4	#5								
2	5550001	1.01	5515	370	570	41239								
3	5550001	1.02	3160	240	240	37901								
4	5550001	1.03	2700	180	375	39239								
5	5550001	1.05	2885	175	350	36219								
6	5550001	1.06	3400	225	390	38605								
7	5550002	2.01	2620	230	240	37063								
8	5550002	2.02	2120	230	180	39549								
9	5550002	2.03	1475	70	220	42052								
10	5550002	2.04	1665	115	240	43765								
11	5550003	3	460	40	75	35684								
12	5550004	4.01	3820	155	700	44466								
13	5550004	4.03	2790	115	630	46963								
14	5550004	4.04	2640	140	335	40164								
15	5550005	5.01	4320	230	500	38024								
16	5550005	5.02	2920	140	665	45924								
17	5550005	5.03	3985	200	790	46471								
18	5550006	6.01	2800	95	635	45678								
19	5550006	6.02	3190	145	725	58928								
20	5550006	6.03	4730	45	1520	62966								
21	5550007	7.01	3935	85	970	59663								
22	5550007	7.02	2505	130	775	64647								
23	5550008	8	3635	50	1420	70077								

20. Scroll down to the bottom of the spreadsheet to examine the definitions of the columns.
21. Convert the values of column D as a percent of the value of Column C and place in column G. (*Hint Use: =D2/C2*100 to do this calculation. Then copy the result to all other entries in that column*)

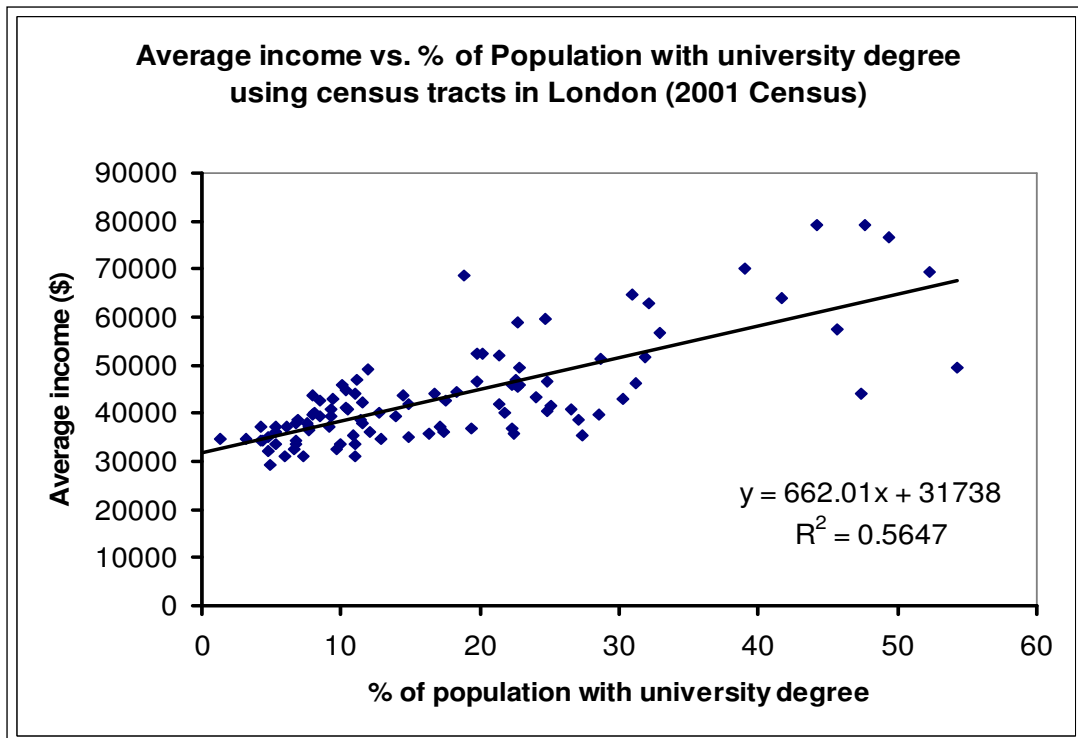
Create Scatter Plot and perform Regression within Excel

22. From the horizontal menu bar select Insert / Chart; this starts the “Chart Wizard”.
23. From “Standard Types” select “XY (Scatter)”.
24. From “Chart Sub-Type” select the first chart type.
25. Click on “Next” and click on “Series”.
26. Within the “Series” box there may be some indicated series, Remove any indicated Series; click on the “Add” tab.
27. Series 1 should be highlighted. To specify the Y-values, click on the coloured range square and with your cursor select the average income data within column F; note the “Chart source data” box at the top of the screen, click on the coloured range square.
28. To specify the X-values, click on the coloured range square and with your cursor select the % population data within column G; note the “Chart source data” box at the top of the screen, click on the coloured range square.
29. Click “Next”. Under **Chart Title**, type “Average Income vs. % of Population with less than grade 9 education using census tracts in London (2001 Census). Under **Value (X) axis**, type “% of population with less than grade 9 education”. Under **Value (Y) axis**, type “Average income (\$)”. Click “Finish”.
30. Click on the Chart and enlarge the graphic
31. Right click on any data point and select “Add trendline...”
32. From the **Add Trendline** box select “Linear”
33. Click on the **Options** tab and select “Display equation on chart” and “Display R-squared value on chart”

34. Continue with chart modifications and the following may be the possible result. Each point represents the values for one census tract in London.



35. If you follow the same procedure but graph % with a university degree vs. income the next graph may be the result.



File: mdm4u/Lesson 3c – Education & Income 2001 (Excel)
 Originator: John Kezys, Mohawk College, kezysj@mail.mohawkc.on.ca
 Updated: December 2004 by Joel to use 2001 Census data
 Latest revisions on August 26, 2005 by Mylène Abi-Zeid