

# How to use Census data on E-STAT with Fathom for a data analysis project

## Is it worth completing university?

What the Census tells us about the correlation between education and income

by Joel Yan, Statistics Canada, [joel.yan@statcan.ca](mailto:joel.yan@statcan.ca), 1-800-465-1222

### Assignment

Is it worth completing university? Investigate the correlation between the proportion of people with a completed university degree and average income using census data for census tracts in a large urban centre in Canada. We do this on E-STAT by creating a scatter plot of the relationship between average incomes versus the percent of the population with a university degree for census tract areas. Then we import the data into Fathom in order to compute the regression equation and correlation statistics.

E-STAT has over a thousand census characteristics for each census tract (and municipality) in Canada that can be manipulated within E-STAT and imported into 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*)

---

**Note:** The E-STAT portion of this activity is also available on the Statistics Canada web site in English at <http://www.statcan.ca/english/Estat/guide/income.htm> and in French at [http://www.statcan.ca/francais/Estat/guide/income\\_f.htm](http://www.statcan.ca/francais/Estat/guide/income_f.htm)

## 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. Select “2001 Census, 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 – Oshawa ( 68 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”
  - “With bachelor’s degree of 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”

### *Expressing Data as Percentages*

10. E-STAT has basic computation and mapping capability. 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 at the bottom of the screen. We then click the radio button under the table “Data as % of 1st characteristic” and click the “Redisplay As” tab. This produces a table showing the schooling level as a percentage of the total population. Note that when we compute the percentages of the total population by level of schooling, the income values in the table remain unaffected. This is because E-STAT recognizes which variables are absolute counts (e.g. the schooling level) and which are relative variables (e.g. income) and

computes percentages based on the data type. For details, click on the Help button on the left side bar, under “Search Census”.

### Sorting the Data

11. We can also sort the table in ascending order by any of the characteristics in the table. For example, to sort by average employment income, we click the radio button “Sort on data for the 2nd characteristic” and then click the “Redisplay As” tab.

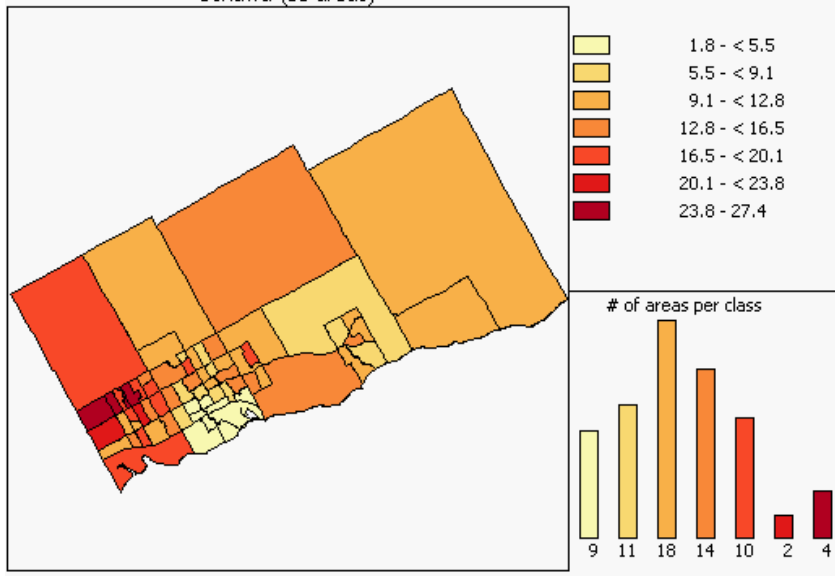
Questions: What are the minimum and maximum values for average employment income by census tract within the selected centre? \_\_\_\_\_

How might education level affect the average employment income? \_\_\_\_\_

### Mapping the Data

12. Maps help us visualize and better understand the areas for which we are looking at data. To produce a statistical map, we click the Map icon under the graph. By default this produces a colour shaded map of the first characteristic using 7 classes, as shown below. This type of map can be useful in identifying census tract areas with the selected characteristics. Furthermore, the frequency histogram to the lower right of the map can be used to assess the distribution of the data for the selected areas (e.g. normal, skewed, bimodal) E-STAT offers three different types of maps (proportional circle, dot, and shaded) and several options for selecting class intervals and colours. For details, check the Help on the left side bar under “Search Census”.

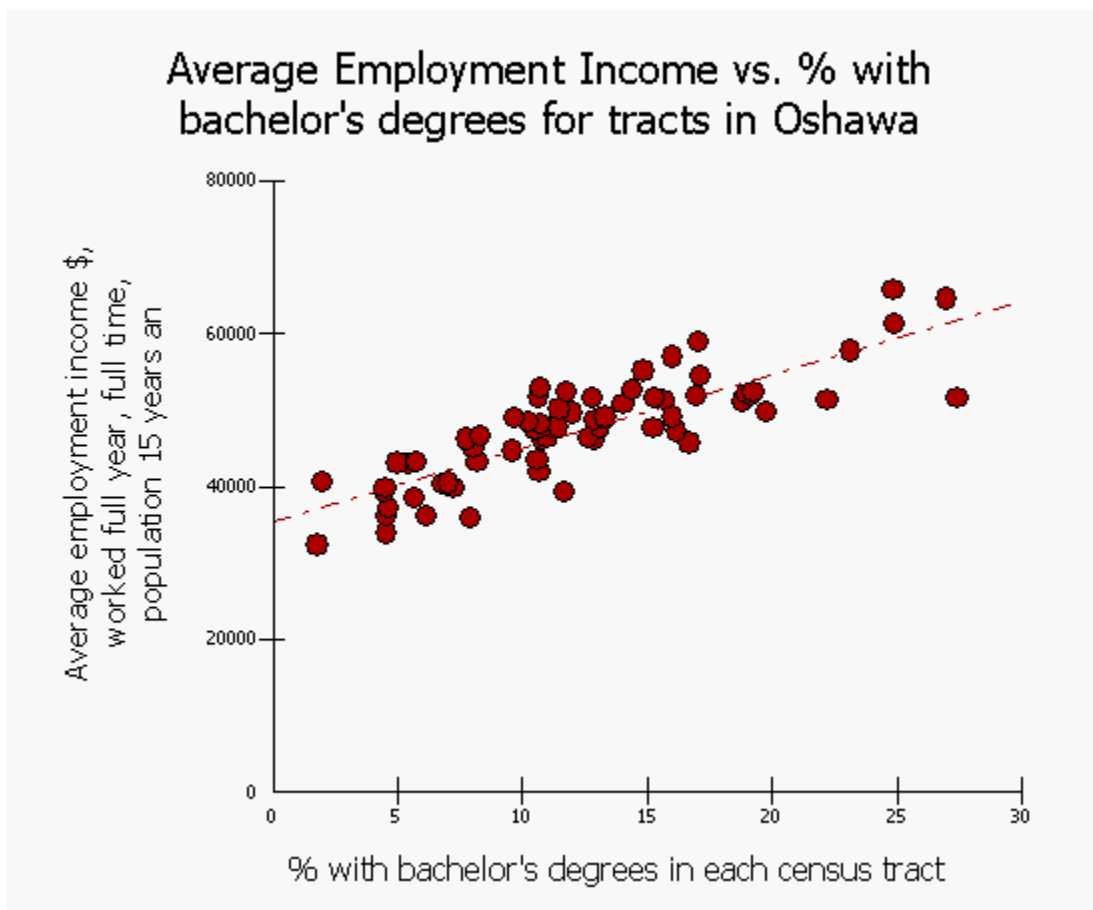
With bachelor's degree or higher, university, population 20 years and over by highest level  
Oshawa (68 areas)



**Analysis question:** After looking at the frequency histogram to the right of the map, describe the distribution of census tract values for schooling level within the selected urban centre. \_\_\_\_\_

### Producing a scatter graph

13. E-STAT can also produce a scatter graph, showing the relationship between any two variables. In this exercise, we want to see if there is a relationship between schooling levels and average employment income for census tracts. To produce a scatter graph with the line of best fit shown, click the “Scatter Graph” icon at the bottom of the screen.
14. If desired, we can modify the graphs labels, by clicking ‘Specify titles’ under the graph. For example, enter “Average Employment Income vs. % with bachelor’s degrees for tracts in Oshawa” in the Title box. Then enter “% with bachelor's degrees in each census tract” in the Subtitle box. Click on ‘Redraw using these titles’. The resulting scatter graph is shown below.



15. Questions involving analysis of this graph:  
Is the line of best fit a reasonable model for predicting the behaviour of average income as a function of % of university graduates within an area? \_\_\_\_\_  
What does the slope of this line of best fit represent? \_\_\_\_\_  
What is the approximate slope of the line for your selected centre? \_\_\_\_\_

**Hint:** Additional details on [Scatterplots](#) are available in [Statistics: Power from Data!](#).

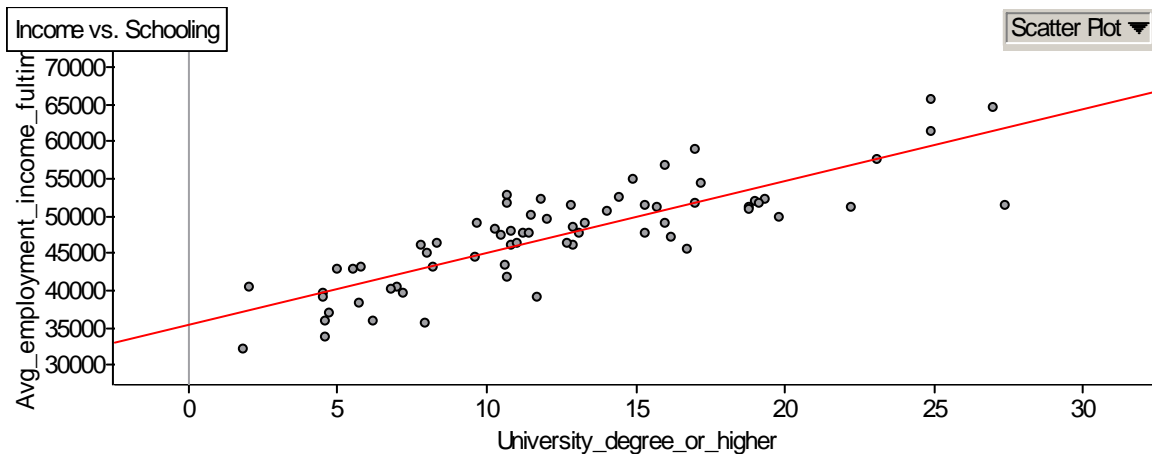
16. E-STAT will let us easily produce scatter plots with lines of best fit for assessing potential relationships between any two variables. However, E-STAT can not currently output the equation of the line of best fit or the  $r^2$  value. To obtain these statistics we can export the data from E-STAT to another package, such as a spreadsheet, graphing calculator, or analytical software such as Fathom.
17. Scroll down and select from **Choose an output format**: “Spreadsheet.WK1  
Geo=Rows”
18. Within the **Downloading** box select “Open” and press OK
19. When the data opens within your spreadsheet program, highlight and copy the columns of interest. Then highlight all the rows with data by clicking “Ctl A” and then “Ctl C”.

### ***Load the Data into Fathom***

20. Switch to Fathom. (If Fathom isn't already running, you will need to launch it.)
21. In a new document, make a new empty collection.
22. With the collection selected, chose Paste Cases from the Edit menu.
23. Make a case table for the collection (for example by choosing Case table from the Insert menu)
24. You may want to edit the second attribute name to census\_tract. Since the attribute names are quite long, it is helpful to shorten the other attribute names as well. In this case, we shortened the names to University degree or higher, and Avg\_Employment\_income\_fulltime.
25. Change the name of the collection to ‘Income vs. Schooling’
26. Save your Fathom document by choosing Save from the File menu.

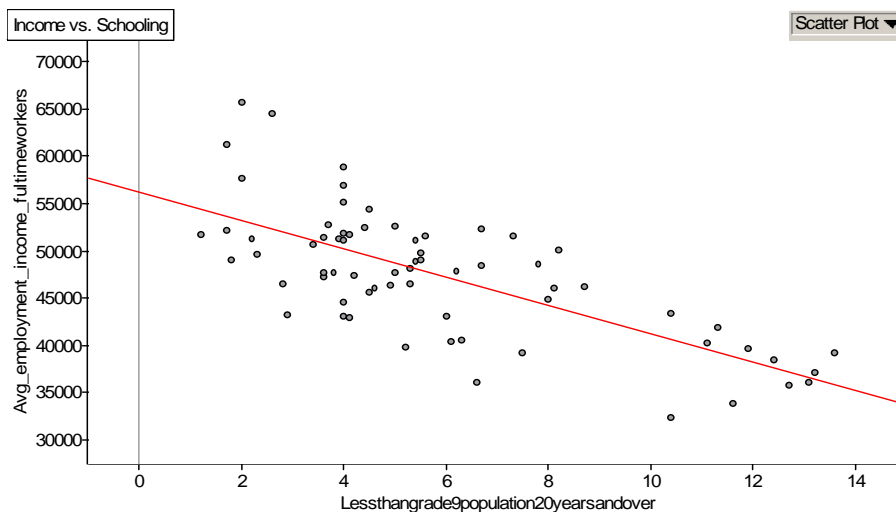
### ***Graphing the Data***

27. Make a graph showing the variation in average income versus the schooling level.
28. Using the **Graph** pull-down menu overlay a Least-Squares Line (as shown below). This gives you the equation for the line of best fit, as well as the coefficient of variation ( $r^2$ ).



$$\text{Avg\_employment\_income\_fulltimeworkers} = 959\text{University\_degree\_or\_higher} + 35500; r^2 = 0.70$$

29. Write a description of the pattern you see on this graph. Write a possible explanation for the shape observed in this distribution.
30. You can also add a residual plot to help you assess the quality of the fit of the least squares line.
31. Repeat the entire process and analysis using the percent of the population with less than grade 9 education. This will involve going back to E-STAT to retrieve new data. The resulting graph is shown below.



$$\text{Avg\_employment\_income\_fulltimeworkers} = -1500\text{Lessthangrade9population20yearsandover} + 56200; r^2 = 0.50$$

Notes: This analysis could be repeated at any geographic level. Using E-STAT for example, you could extract the same variables for the 49 counties, districts, and regional municipalities in Ontario and repeat the analysis. For further analysis of the relationship, look at the 2001 Census results on the relationship between education and income. You will find this at

<http://www12.statcan.ca/english/census01/products/highlight/Earnings/Index.cfm?Lang=E>. Select Census on the top blue bar, the Data on the left bar, then Highlight tables.

Also on the OCDSB website at [http://www.ocdsb.edu.on.ca/teacher\\_res/secondary/tecint/mdm4u.htm](http://www.ocdsb.edu.on.ca/teacher_res/secondary/tecint/mdm4u.htm)  
Also on the Fathom website at [http://www.keypress.com/fathom/sample\\_activities](http://www.keypress.com/fathom/sample_activities)