

Exploring Relationships between Highest Level of Education and Income on the 1991 Census microdata using Microsoft Excel

Created by Michael Lieff (M@LIEFF.NET)
Faculty of Education, Queen's University

While on practicum at Statistics Canada
Feb. 23 – Mar. 12, 2004

Working under the direction of Joel Yan (joel.yan@statcan.ca)

Table of Contents

CLASSROOM INSTRUCTIONS.....	1
INTRODUCTION.....	2
PART A - CREATE A GRAPH OF AGE VS. INCOME.....	2
PART B – GRAPHING CURVES OF BEST FIT.....	7
PART C - EXTENSION EXERCISES.....	11
APPENDIX A – SELECTING DATA IN EXCEL.....	13
APPENDIX B - DATA SOURCE.....	13
APPENDIX C - MINISTRY EXPECTATIONS ADDRESSED BY THIS WORKSHOP.....	14

Note: A Quattro Pro version of this activity is also available. The file is called Lesson 9Age_vs_Income_by_Education_ Quattro_Pro_mlieff.doc

File: Lesson9 - Age_vs_Income_by_Education_ Excel.doc
Updated: 8-Mar-04 by Michael Lieff

Classroom Instructions

The lab is meant to be completed independently or in a small group. Groups of discussion questions could be taken up as a class.

To complete this exercise, participants will need a copy of the file

CAN_CMA_Education.xls

Parts A-B of the lab will take approximately an hour to complete – more if discussions around the questions occur. At the discretion of the leader, to save time, students can ignore the **less than university** group and use only the **high school** and **university** groups. Part C – Extension Exercises can be completed as homework or in the next class.

Introduction

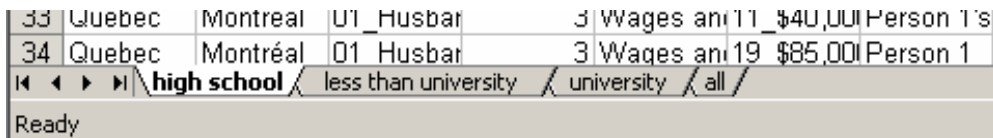
Most of us begin to have an income sometime during our teenage years. As we enter our twenties and progress through adulthood, our income increases. Education is a major factor influencing income.

How much does our education affect our income? Through this investigation we will examine the relationship between Highest Level of Education and Income using real data from the 1991 Canadian Census and Microsoft Excel.

Part A - Create a Graph of Age vs. Income

Our first task is to use Microsoft Excel to create a graph of Age vs. Income for the group whose highest level of education is high school.

1. Open the file *CAN_CMA_Education.xls* using Microsoft Excel.
2. Click the **high school** Worksheet tab at the bottom of the window.



33	Quebec	Montreal	U1_Husbar	3	Wages an	11	\$40,00	Person 1's
34	Quebec	Montréal	01_Husbar	3	Wages an	19	\$85,00	Person 1

Ready

High school is selected

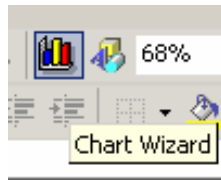
1. The data we are going to graph is located in columns H (Age) and AI (Total Income). First, select all of the values for Age (column H). For help on selecting data, see Appendix A.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Province	Census_M	Census_F	Number_of	Major_Sou	Total_Cens	Relationst	Age	Sex	Legal_Mar	Fertility	Religion	Place_of	Immigrant
2	Newfoundl	Not In CM	04	Wife	2	Govermmer	08	\$25,000	Person 1's	64	Female	Legally ma	5	Catholic(A
3	Newfoundl	Not In CM	02	Male c	2	Wages ani	12	\$45,000	Person 1	39	Male	Never married (single)	Anglican	29
4	Newfoundl	Not In CM	07	Never--	5	Wages ani	18	\$75,000	Person 1's	41	Male	Never married (single)	United Chu	29
5	Nova Scoti	Halifax	01	Husbar	2	Wages ani	11	\$40,000	Person 1	50	Male	Legally married (not s	Anglican	30
6	Nova Scoti	Halifax	01	Husbar	4	Self-emplo	10	\$35,000	Person 1	44	Male	Legally married (not s	Catholic(A	29
7	Nova Scoti	Halifax	11	-Non far 1	1	Non-family persons		Person 1's	80	Female	Widowed	2	United Chu	29
8	Nova Scoti	Not In CM	01	Husbar	2	Wages ani	09	\$30,000	Person 1	36	Male	Legally married (not s	Baptist	29
9	Nova Scoti	Not In CM	04	Wife	3	Self-emplo	10	\$35,000	Person 1's	39	Female	Legally ma	1	Catholic(A
10	New Bruns	Not In CM	13	-Non far 1	1	Non-family persons		Person 1	51	Female	Widowed	2	Baptist	30
11	New Bruns	Not In CM	13	-Non far 1	1	Non-family persons		Person 1	39	Male	Never married (single)	United Chu	29	
12	New Bruns	Not In CM	06	Female	2	Wages ani	09	\$30,000	Person 1	71	Female	Widowed	5	Catholic(A
13	New Bruns	Not In CM	01	Husbar	2	Wages ani	15	\$60,000	Person 1	30	Male	Legally married (not s	Catholic(A	29
14	New Bruns	Not In CM	01	Husbar	2	Investment	13	\$50,000	Person 1	76	Male	Legally married (not s	Catholic(A	29
15	New Bruns	Not In CM	01	Husbar	3	Wages ani	09	\$30,000	Person 1	46	Male	Legally married (not s	Presbyteri	29
16	New Bruns	Not In CM	01	Husbar	5	Govermmer	02	\$2,000	Person 1	46	Male	Legally married (not s	Catholic(A	29
17	Quebec	Québec	06	Female	2	Wages ani	08	\$25,000	Person 1	36	Female	Legally ma	1	Roman Ca
18	Quebec	Québec	04	Wife	3	Wages ani	11	\$40,000	Person 1's	34	Female	Legally ma	1	Roman Ca
19	Quebec	Québec	04	Wife	3	Wages ani	11	\$40,000	Person 1's	59	Female	Legally ma	2	Roman Ca
20	Quebec	Québec	05	Female	5	Wages ani	17	\$70,000	Person 1's	31	Female	Divorced	2	Roman Ca
21	Quebec	Québec	07	Never--	5	Wages ani	16	\$65,000	Person 1's	15	Male	Never married (single)	Roman Ca	04
22	Quebec	Québec	04	Wife	4	Wages ani	12	\$45,000	Person 1's	38	Female	Legally ma	2	Roman Ca
23	Quebec	Montréal	13	-Non far 1	1	Non-family persons		Person 1	54	Female	Divorced	4	Roman Ca	
24	Quebec	Montréal	13	-Non far 1	1	Non-family persons		Person 1	85	Female	Widowed	7	7 and ov	
25	Quebec	Montréal	10	Never--	2	Govermmer	06	\$15,000	Person 1's	64	Female	Never mari	0	Roman Ca
26	Quebec	Montréal	01	Husbar	2	Wages ani	10	\$35,000	Person 1	45	Male	Legally married (not s	Roman Ca	04
27	Quebec	Montréal	10	Never--	3	Wages ani	12	\$45,000	Person 1's	20	Female	Never mari	0	Roman Ca
28	Quebec	Montréal	04	Wife	3	Wages ani	14	\$55,000	Person 1's	36	Female	Legally ma	1	Roman Ca
29	Quebec	Montréal	04	Wife	3	Wages ani	14	\$55,000	Person 1's	41	Female	Legally ma	1	Roman Ca
30	Quebec	Montréal	04	Wife	3	Govermmer	09	\$30,000	Person 1's	69	Female	Legally ma	3	Roman Ca
31	Quebec	Montréal	10	Never--	3	Other inco	09	\$30,000	Person 1's	16	Male	Never married (single)	Roman Ca	04
32	Quebec	Montréal	07	Never--	3	Wages ani	20	\$100,000	Person 1's	21	Male	Never married (single)	Roman Ca	04
33	Quebec	Montréal	01	Husbar	3	Wages ani	11	\$40,000	Person 1's	33	Male	Legally married (not s	Roman Ca	04
34	Quebec	Montréal	01	Husbar	3	Wages ani	19	\$85,000	Person 1	44	Male	Legally married (not s	Roman Ca	04

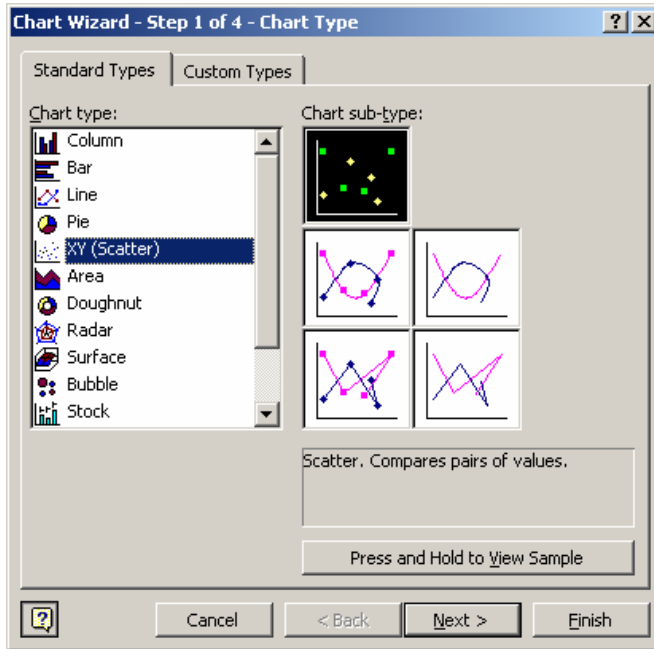
2. Holding the CTRL key, select all of the values for Total Income (column AI). By holding the Control key when selecting the Total Income data, the Age data remains selected.

NOTE: You may notice that some values for Total Income are blank. These will not show up on the graph, so it is not necessary to filter them out.

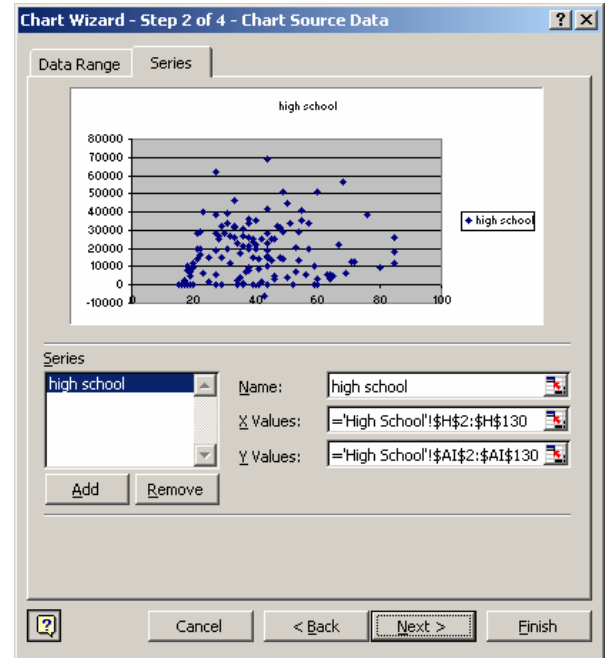
3. Click the Chart Wizard button in the Toolbar above.



- The Chart Wizard is launched.
Select **XY (Scatter)** from the **Standard Types** tab and click **Next**.



Step 6.



Step 7-9.

- On the **Chart Wizard – Step 2 of 4 – Chart Source Data** screen, click the **Series** Tab.
- In the **Name** text box enter “high school”
- Click **Next**.

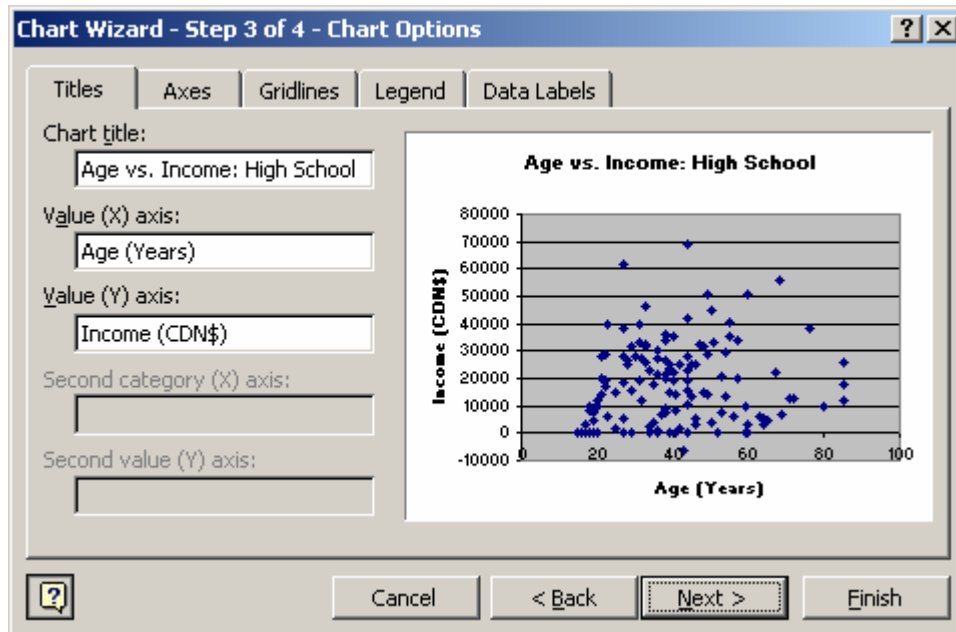
NOTE: There should only be one series listed in the **Series** list. If there is more than one Series, please click **Back** and double-check the **Chart type**.

8. In the **Titles** tab, enter the following titles:

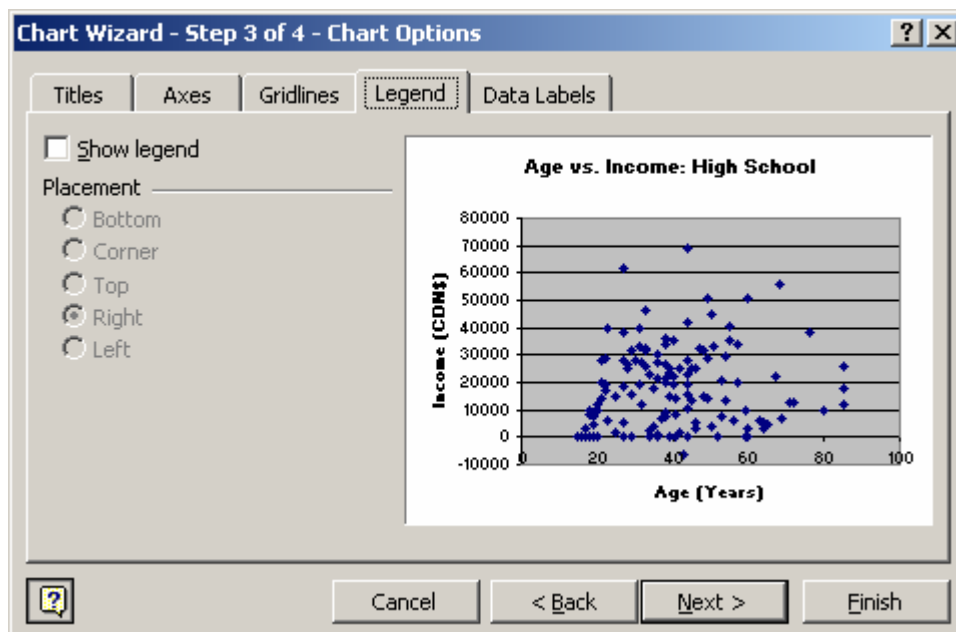
Chart Title: Age vs. Income: High School

Value (X) axis: Age (Years)

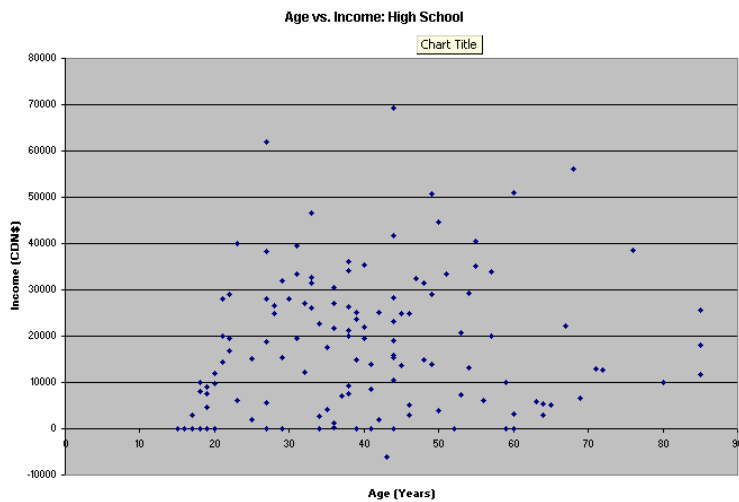
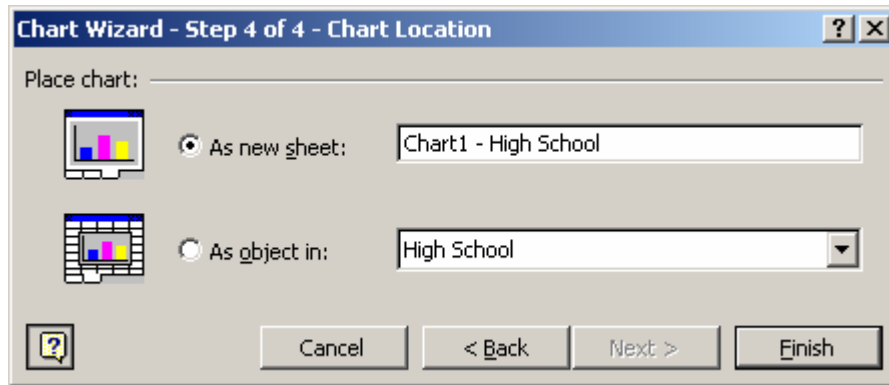
Value (Y) axis: Income (CDN\$)



9. Click the **Legend** tab.
Clear the **Show legend** checkbox.



10. Save the chart as a new sheet in your Excel Workbook. To do this, select the **As new sheet** radio button and enter the name **Chart1 - High School**
11. Click **Finish** to create the graph.



What is the range of ages? _____

What is the range of incomes? _____

Draw a line of best fit on the above graph Age vs. Income: High School. What does it tell us about one's income as he/she gets older?

12. Repeat steps 1-11 for the **less than university** (time permitting) and **university** tabs.

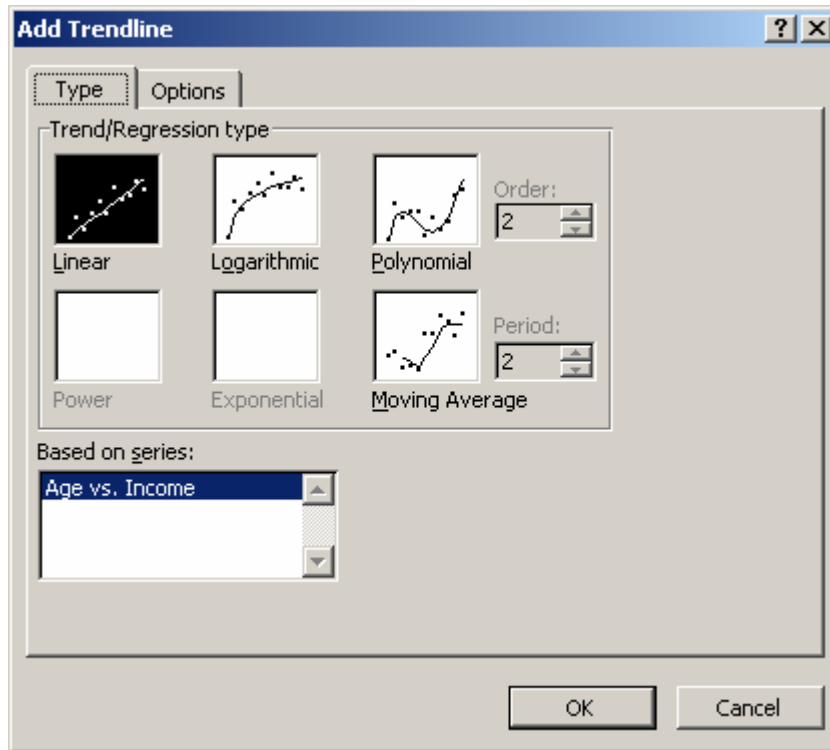
TIPS:

You can change the order of Worksheets by dragging-and-dropping the tabs.
 You can rename a Worksheet by double-clicking its tab and entering the new name.

Part B – Graphing curves of best fit

The second task we will complete is to draw and analyze curves of best fit.

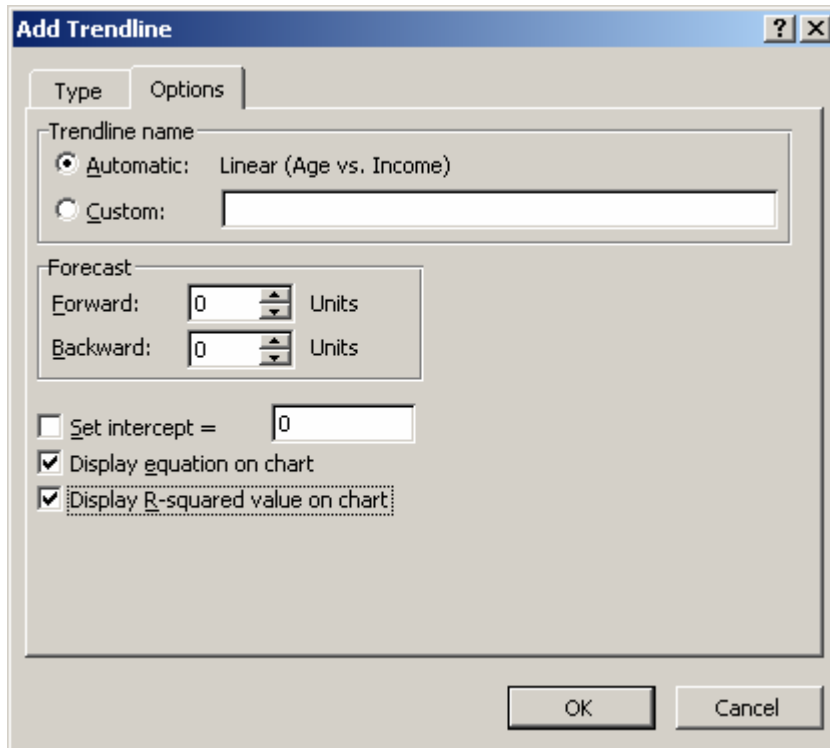
1. Select the **high school** tab and click anywhere on the high school chart.
2. On the Toolbar click **Chart**→**Add Trendline**. By default a Linear (straight line) type is selected.



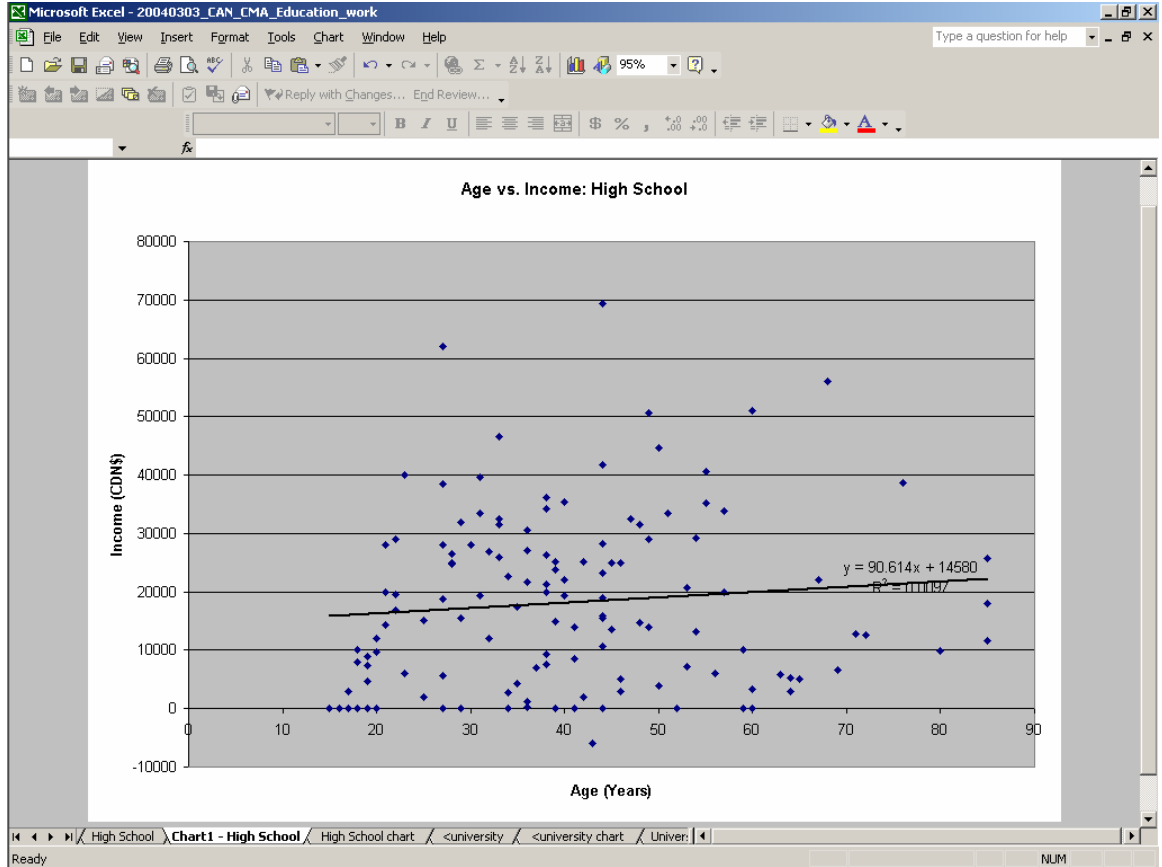
3. Click the **Options** tab and fill the checkboxes for:

- **Display equation on chart**
- **Display R-squared value on chart**

As in the picture below:



4. Click **OK** to create the Trendline.



TIP: Once created, you can drag-and-drop the equation box to another spot to make it more visible, and change the text colour to make it easier to read.

How well does the line fit the data? _____

Does it look like your diagram from Part A? _____

What income would this line predict for someone who is:

60 years old? _____ 80 years old? _____

Is it a good estimate for income? _____

5. Repeat Steps 1-3 for the **less than university** (time permitting) and **university** charts.
6. Use $I(x)$ to denote the function for Income and fill in the following table:

Highest Education	Equation of the line of best fit	R^2 value
High school	$I_H(x) =$	
Less than University	$I_L(x) =$	
University	$I_U(x) =$	

Compare the equations of the 3 lines. What trend do you see?

Calculate I_U , the income, earned by someone with a 3-year university degree, when they are 30, 45, and 60.

HINT: To find $I_U(30)$, substitute 30 into the university equation from your table above.

$$I_U(30) = \qquad I_U(45) = \qquad I_U(60) =$$

Compare this to the salary earned by someone on the “high school” category..

$$I_H(30) = \qquad I_H(45) = \qquad I_H(60) =$$

How much more money does the university degree holder earn when they are 45?

$$I_U(45) - I_H(45) =$$

Is this a significant difference?

In general, do the lines of best fit provide reasonable estimates of income? Can you think of a way to find a more accurate answer? What is a simple way to factor in the \$36 000 cost of a university degree?

Part C - Extension Exercises

Exercise 1 – Polynomial Trendlines

1. Repeat steps 1-6 of Part B for ONE of the charts you created, but create a Trendline of Type **Polynomial** (degree 2).

Is this curve more or less realistic than the Linear one created above? Why?

2. Create a third type of Trendline for your graph (e.g. logarithmic, polynomial of degree >2 , others may be possible).

TIP: You may wish to clear other Trendline(s) to make your graph easier to read. To erase a Trendline, select it and/or right-click and select **Clear**.

For which age ranges is your curve of best fit reasonable?

Justify why the curve fits by explaining the factors that may affect one's income in those ranges.

For which age ranges is your curve not reasonable?

Exercise 2 – Finding another relationship in the Census Microdata

1. Examine the 40 attributes provided in the 1991 Census microdata file. Find a relationship between two or more of these attributes. Use filters, or sort data and create separate tabs to examine the relationships between different groups. You can use scatter graphs, histograms, tables or other methods to communicate the relationship that you have discovered.

You can find descriptions of the attributes here:

<http://www.statcan.ca/english/kits/pumf91/1991public.htm#list>

Appendix A – Selecting Data in Excel

To select all of the data in a column, click on the column header.

To select a range of data you can use any of the three following methods:

1. Select the first cell, then:

Using the mouse: Holding the primary mouse-button (i.e. the left button for a right-handed mouse) drag the cursor to the last cell in the range.

Using the keyboard: Holding the SHIFT key, use the arrow keys to move the cursor to the last cell in the range.

2. Select the first cell; scroll to the last cell; while holding the SHIFT key, select the last cell in the range.
 3. While holding the CTRL key, select individual cells. This will ADD cells to the currently selected range, and can be used with methods 1) and 2) above.
-

Appendix B - Data Source

The data in the file *CAN_CMA_Education.xls* is based on the Statistics Canada 1991 Census Microdata. Links to the Census Microdata for Canada and by Province are located here: <http://www.statcan.ca/english/kits/pumf91/intropages.htm>

The *CAN_CMA_Education.xls* file was created through the following procedure:

1. Save the Canada data as an unformatted text (*.TXT) file.
2. Import the data into Microsoft Excel as a TAB-delimited text file.
3. AGE attribute: instances of the value “85 and over” were changed to “85” so that the field was numeric and could be used for graphing.

Descriptions of all 40 attributes can be found here:

<http://www.statcan.ca/english/kits/pumf91/1991public.htm#list>

Appendix C - Ministry Expectations addressed by this Workshop

Mathematics of Data Management (MDM4U):

Organization of Data for Analysis

Overall Expectations

- Organize data to facilitate manipulation and retrieval
- Solve problems involving complex relationship, with the aid of diagrams

Specific Expectations

Organizing data

- Create spreadsheet template to that facilitate the manipulation and retrieval of data from large bodies of information that have a variety of characteristics

Statistics

Overall Expectations

- standard techniques for collecting data
- analyze data involving one variable, using a variety of techniques

Specific Expectations

Collecting data

- Organize and summarize data from secondary sources (Internet)

File: Lesson 9 Age_vs_Income_by_Education_Excel.doc
On Directory: MDM4U Treasure Trove/ Lessons/
Updated: 8-Mar-04 by Michael Lieff, m@lieff.net