

Exploring Relationships between Highest Level of Education and Income using Corel Quattro Pro

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

INTRODUCTION	1
CLASSROOM INSTRUCTIONS.....	1
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 COREL QUATTRO PRO	13
APPENDIX B - DATA SOURCE.....	13
APPENDIX C - MINISTRY EXPECTATIONS ADDRESSED BY THIS WORKSHOP.....	14

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? In this workshop we will look for a relationship between Highest Level of Education and Income. We will be working with real Statistics Canada data from the 1991 Census.

The workshop contains 3 parts:

In **Part A** we will use Corel Quattro Pro to create a scatter graph showing Income and Age for 3 groups with different levels of Education.

In **Part B** we will begin a process of trying to find the best fit trend line for the data. We will add a line of best fit to your graph and analyze its equation.

In **Part C** we will attempt to find a curve which fits the data better. We will begin by examining a polynomial of degree 2. Then you will choose and graph a third type of curve.

Classroom Instructions

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

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

CAN_Education.qpw

Parts A-B of the workshop 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.

Part A - Create a Graph of Age vs. Income

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

1. Open the file *CAN_Education.qpw* using Corel Quattro Pro.
2. Click the **high school** Sheet tab at the bottom of the window.

21	British Col	Not In CM	U4_Wife	2	Government U/	\$20,00	Person 1's
22	British Col	Not In CM	01_Husbar	2	Wages an	06 \$15,00	Person 1
23	British Col	Not In CM	01_Husbar	2	Other inco	12 \$45,00	Person 1

Navigation: high school / less than university / university / all / E / F / G

QuickCell: PUMF91159... NUM CAPS SCRL READY

The data we are going to graph is located in columns H (Age) and AI (Total Income).

1. Select the first 100 values in the Age (column H). For help on selecting data, see Appendix A. Do NOT select the entire column as you will not be able to generate the graph.

	A	B	C	D	E	F	G	H	I	J	K
1	Province	Census_M	Census_F	Number of Major_Sou	Total_Cen	Relationship	Age	Sex	Legal_Mar	Fertility	
2	British Col	Vancouver	13-Non fai	1 Non-family persons		Person 1	61	Female	Widowed	3	
3	British Col	Vancouver	13-Non fai	1 Non-family persons		Person 1	68	Female	Widowed	0	
4	British Col	Vancouver	13-Non fai	1 Non-family persons		Person 1	74	Female	Widowed	0	
5	British Col	Vancouver	13-Non fai	1 Non-family persons		Person 1	47	Male	Divorced		
6	British Col	Vancouver	04_Wife	2 Wages an	12 \$45,00	Person 1's	41	Female	Legally ma	0	
7	British Col	Vancouver	01_Husbar	2 Wages an	15 \$60,00	Person 1	38	Male	Legally married (not s		
8	British Col	Vancouver	12-Non fai	1 Non-family persons		Persons n	71	Male	Divorced		
9	British Col	Vancouver	01_Husbar	2 Governmei	06 \$15,00	Person 1's	82	Male	Legally married (not s		
10	British Col	Vancouver	10_Never-	3 Wages an	07 \$20,00	Person 1's	15	Male	Never married (single)		
11	British Col	Vancouver	04_Wife	4 Wages an	18 \$75,00	Person 1's	55	Female	Legally ma	5	
12	British Col	Vancouver	04_Wife	2 Wages an	05 \$10,000	Person 1's	51	Female	Legally ma	3	
13	British Col	Victoria	13-Non fai	1 Non-family persons		Person 1	78	Female	Widowed	2	
14	British Col	Victoria	01_Husbar	2 Wages an	14 \$55,00	Person 1	42	Male	Legally married (not s		
15	British Col	Victoria	04_Wife	6 Wages an	10 \$35,00	Person 1	29	Female	Legally ma	4	
16	British Col	Not In CM	13-Non fai	1 Non-family persons		Person 1	34	Male	Never married (single)		
17	British Col	Not In CM	04_Wife	2 Wages an	12 \$45,00	Person 1's	43	Female	Legally ma	3	
18	British Col	Not In CM	04_Wife	2 Wages an	08 \$25,00	Person 1's	50	Female	Legally ma	4	
19	British Col	Not In CM	04_Wife	2 Wages an	03 \$5,000	Person 1's	58	Female	Legally ma	2	
20	British Col	Not In CM	04_Wife	2 Investment	05 \$10,000	Person 1's	60	Female	Legally ma	1	
21	British Col	Not In CM	04_Wife	2 Governmei	07 \$20,00	Person 1's	75	Female	Legally ma	3	
22	British Col	Not In CM	01_Husbar	2 Wages an	06 \$15,00	Person 1	56	Male	Legally married (not s		
23	British Col	Not In CM	01_Husbar	2 Other inco	12 \$45,00	Person 1	69	Male	Legally married (not s		

Navigation: high school / less than university / university / all / E / F / G

QuickCell: PUMF91159... Sum=12099 Ave=40.73737374 Count=297 Max=85 Min=15

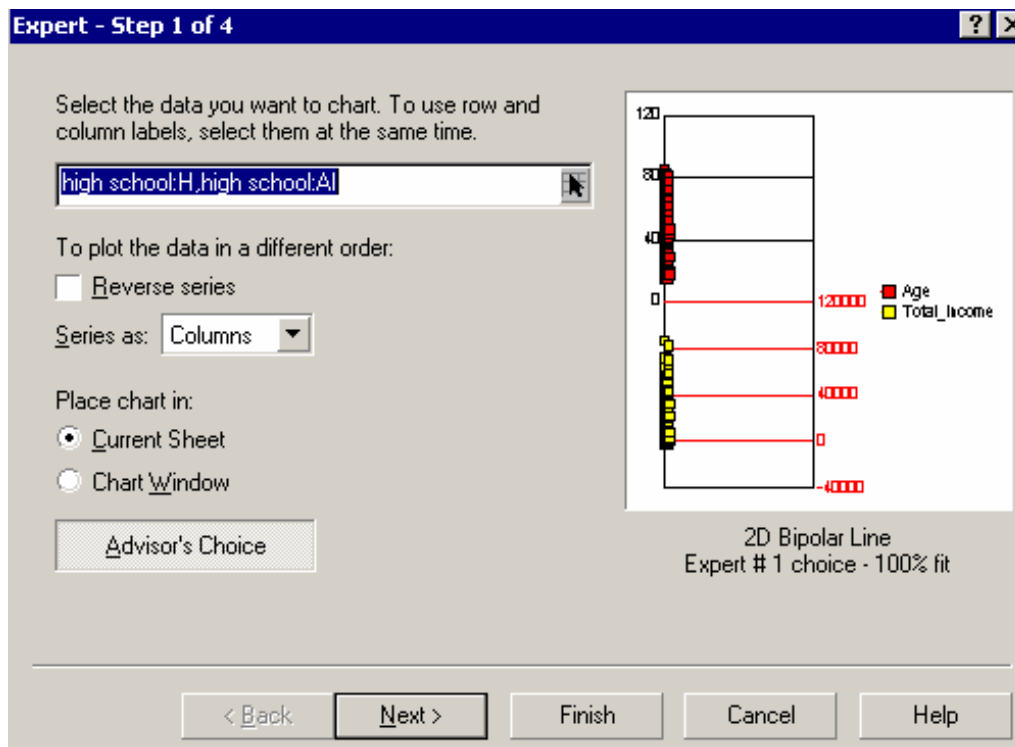
2. Holding the CTRL key, select the first 100 values for Total Income (column AI). By holding the Control key when selecting the Total Income data, the Age data remains selected.

NOTES:

The graph will **NOT** display properly if you select the columns in the reverse order AI, H.

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.

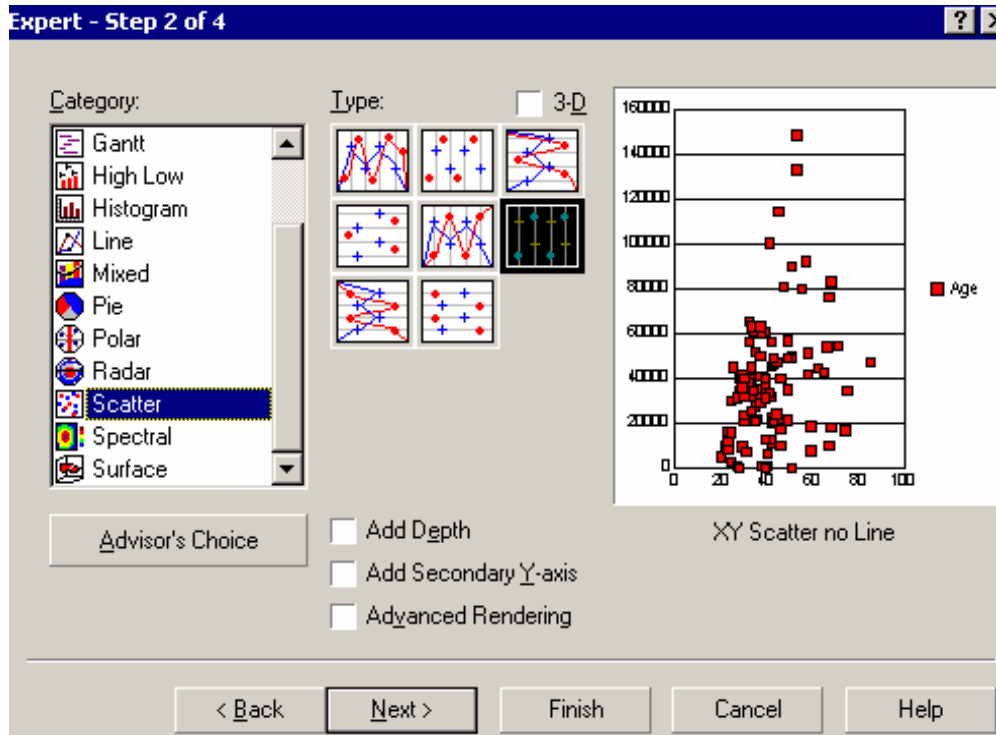
3. Click **Insert**→**Chart**. The Quattro Pro Chart Expert is launched.



Step 3.

4. Click **Next**.

5. Select: **Category:** *Scatter*
Type: *XY Scatter no Line*



6. Ensure that the **Add Secondary Y-axis** checkbox is **CLEAR**
7. Click **Next**

8. In the **Expert – Step 3 of 4** Window, enter the following titles:

Title: *Age vs. Income: High School*

X-axis: *Age (Years)*

Y-axis: *Income (CDN\$)*

Expert - Step 3 of 4

If you want a main title, a subtitle, a footnote, or an axis titles type them below:

Title
Income vs. Age: High School

Subtitle

X-Axis
Age (Years)

Y-Axis
Income (CDN\$)

Footnote

Income vs. Age: High Scho

Income (CDN\$)

Age

Label

< Back Next > Finish Cancel Help

9. Click **Next**.

10. Select an alternate colour scheme if you wish and click **Finish**.

Expert - Step 4 of 4

Choose a color scheme:

- No change
- Color Washes
- Deep Reds
- Default (Version 9)
- Gray Marble
- Icy Blues
- Tangerine

Click on Finish to create the chart

Income vs. Age: High Scho

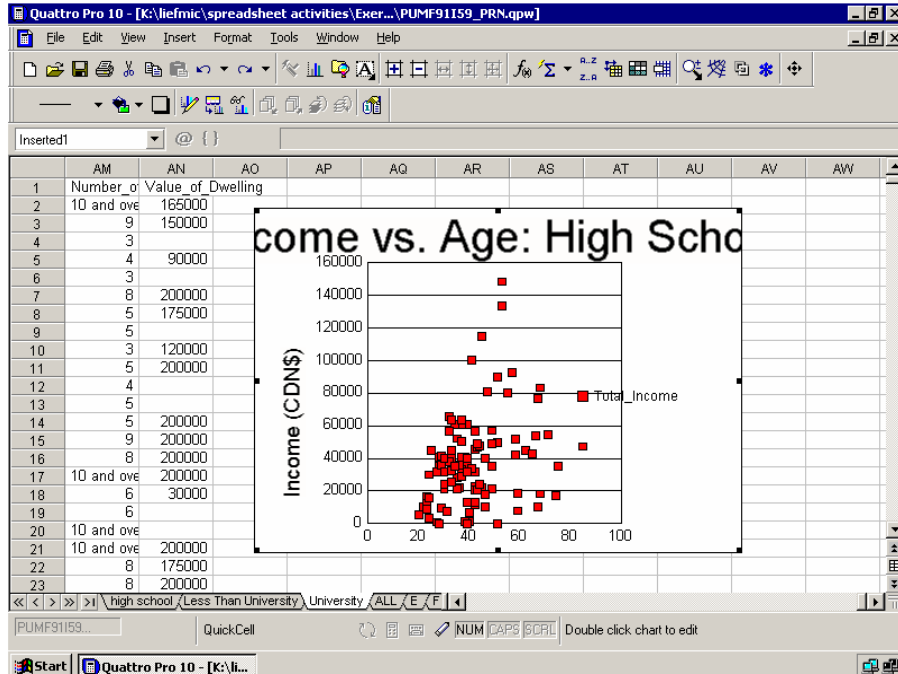
Income (CDN\$)

Age

Label

< Back Next > Finish Cancel Help

11. The cursor changes to a Chart icon. Scroll to the right of the data (column AO).
12. Drag the cursor **down** and to the **right** to draw a box. When you release the mouse button, your chart will appear in the box.



13. Save the file.

What is the range of ages? _____

What is the range of incomes? _____

Manually sketch a best fit trend line on the above graph.

What does the trend line tell us about one's income as he/she gets older?

14. Repeat steps 1-13 for the **university** and **less than university** (time permitting) tabs. Save the file after each graph is created.

TIPS:

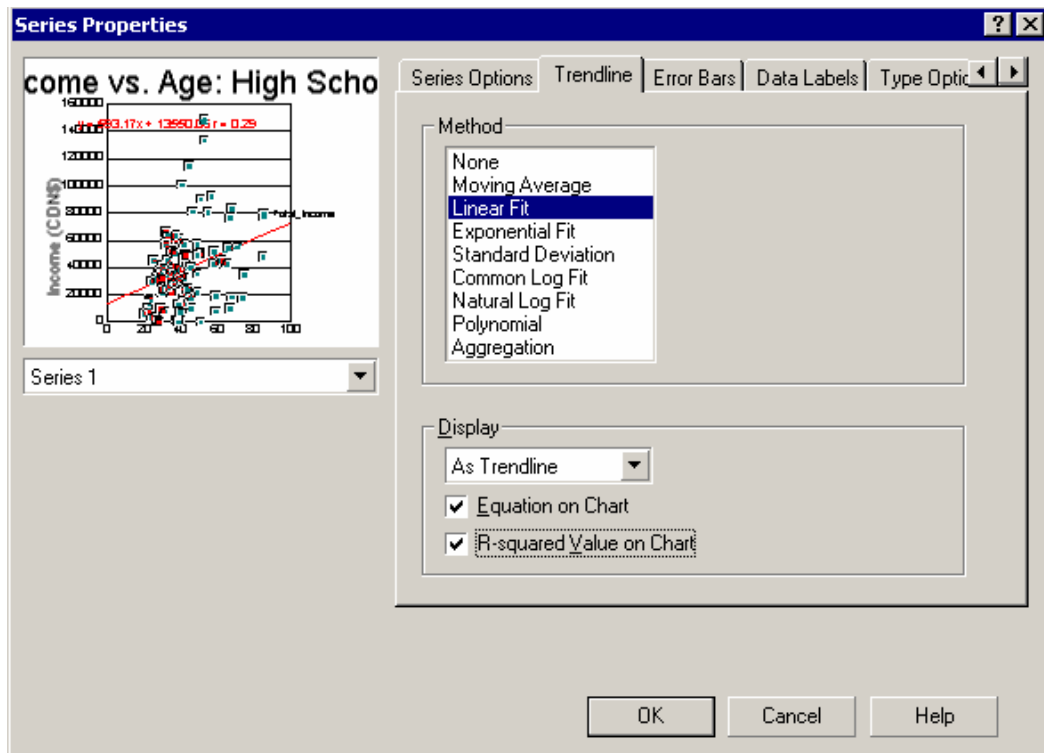
You can change the order of the Sheets by dragging-and-dropping the tabs.

You can rename a Sheet by double-clicking its tab and entering the new name.

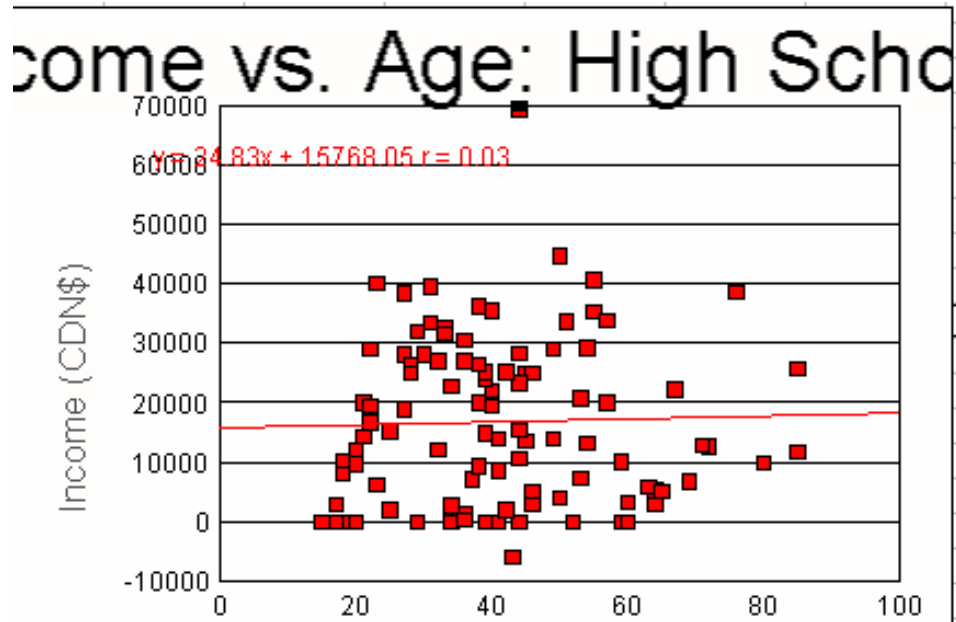
Part B – Graphing curves of best fit

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

1. Click on a point on the chart. The series turns light blue.
2. Right-click and select **Series Properties**.
3. Click the **Trendline** tab.
4. Select Method: **Linear Fit**.
5. Check the checkboxes for **Equation on Chart** and **R-squared Value on Chart**.



6. Click **OK**. The chart is updated.



7. Save the file.

How well does the line fit the data? _____

Does it look like the function you drew manually in Part A? _____

Read the values from the graph or use the equation of the line to predict the income for someone who is:

60 years old? _____ *80 years old?* _____

Is it a good estimate for income? _____

8. Repeat Steps 1-3 for the **less than university** (time permitting) and **university** charts.
9. 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?

A 3-year Bachelor's degree costs approximately \$36 000 in total. Based on the difference in salary for age 30 that you calculated above, how many years will it take to pay for a university degree?

Part C - Extension Exercises

Exercise 1 – Polynomial Trendlines

1. For ONE of the charts you created in Part B, repeat steps 1-6 of Part B to change the Trendline to Type **Polynomial** (order 2). You may wish to do this for the other charts as well. Save the file after each change.

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. Natural Log Fit, Polynomial of order >2 , others may be possible).

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>

The following attributes are numeric, permitting enhanced graphical and numerical analysis, such as scatter plots, regression and correlation:

- Attribute # 20) Knowledge of non-official languages
- Attribute # 31) Hours worked in the reference week
- Attribute # 32) Weeks worked in 1990
- Attribute # 35) Total income
- Attribute # 40) Value of dwelling

Examples of relationships involving these attributes that could be investigated include:

Working hard or hardly working: How many weeks did Canadians work in 1990?

Banker's Hours: Does the number of hours worked in 1990 vary by Industry? By Sex?

Equal pay: Does Total income vary by Sex?

For Sale: Does Value of dwelling vary across Ethnic origin? By Industry?

Categorical attributes

The remaining 35 variables are categorical, permitting cross-tabulations, frequency charts, ribbon charts, and graphical analysis.

Treating categorical attributes numerically

Several of the categorical attributes can be treated and graphed as if they were numeric attributes. However, be aware that there is loss of information when non-numeric values for the attribute are dropped. This function could be usefully applied to the following attributes:

- # 8) Age
- # 11) Fertility/number of children
- # 26) Years of university
- # 37) Owner's major payments (monthly)
- # 38) Monthly gross rent, and
- # 39) Number of rooms.

Appendix A – Selecting Data in Corel Quattro Pro

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 CTRL 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_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_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 Corel Quattro Pro 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: Income_vs_Age_by_Education_Quattro_Pro.doc
On Directory: K:\liefmic\spreadsheet activities\Exercise 2- Income vs. Age by Education
Last Updated: 12-Mar-04 by Michael Lieff