

Canadian National Longitudinal Survey of Children and Youth (NLSCY)

Fathom workshop activity

For more information about the survey, see: <http://www.statcan.ca/Daily/English/990706/d990706a.htm>

Your instructor will tell you where to find the document.

This activity uses a collection containing responses by Ontario youths (age 10 to 13) to a survey as part of a long-term study conducted from 1994 through 1999. As is true of many surveys, the attributes are primarily categorical, having discrete answers.

The goal of this activity is to develop ways to characterize such categorical data.

The Data

1. Open the document, Ontario Youths.ftm.
2. Double-click the collection to open its inspector.

The document window should look similar to the one shown at right. The Comments pane in the inspector is a place to look for documentation about the data, and a place to put documentation when you create your own collections.

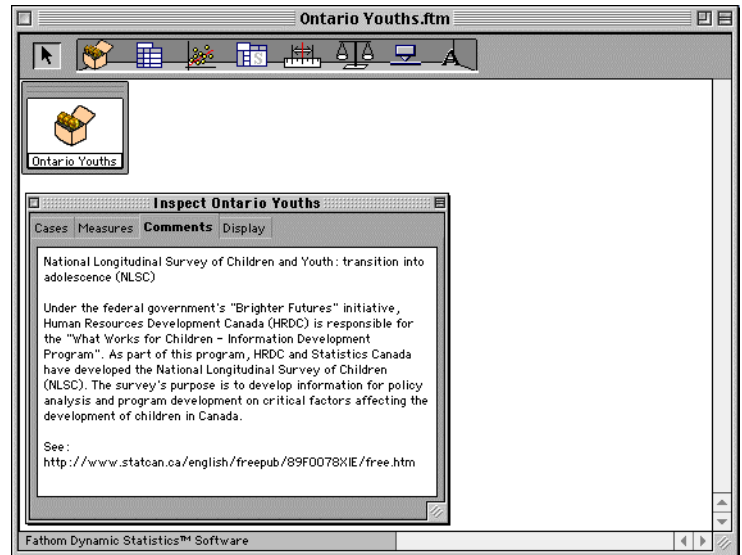
3. Choose the Cases tab in the inspector.

The Cases pane displays information about cases in the collection one at a time. In this collection both names and values of attributes are very long, so we would like to make their columns wider so they become more readable.

4. Click and drag the boundary between Attribute and Value so that its column becomes wider.
5. Repeat this for the Value column.

The inspector should look similar to the one shown below-right.

Notice that the mouse pointer changes when you are positioned over the column boundary.



Attribute	Value	Formula
Sex	F	
How_do_yo...	3. I like so...	
How_well...	Average	
How_impor...	Somewhat...	
I_feel_safe...	Some of th...	
I_have_a_l...	5. true	
I_get_alon...	5. true	
Other_kids...	3. Someti...	
Most_other...	4. Mostly...	

Attribute	Value
Sex	F
How_do_you_feel_about_school	3. I like school a bit
How_well_are_you_doing_in_...	Average
How_important_to_do_well...	Somewhat important
I_feel_safe_at_school	Some of the time
I_have_a_lot_of_friends	5. true
I_get_along_with_kids_easily	5. true
Other_kids_want_me_to_be_t...	3. Sometimes false/Sometim

6. Scroll through the list of attributes.
7. Click on the right arrow button at the bottom of the inspector to look at different cases.

Notice that the notation to the right of the arrow buttons lets you know that there are 1046 cases in this collection.

Bar Charts and Ribbon Charts

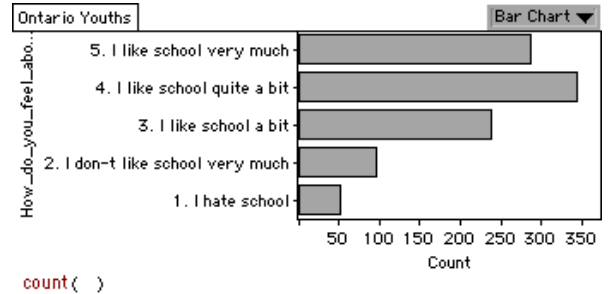
Attribute names in Fathom may not have spaces because spaces mess up formulas. Using underscore characters is one way to make attribute names legible.

Let's look at the second attribute, `How_do_you_feel_about_school`. Make a conjecture: What distribution of answers do you think you're going to find?

8. Make a new, empty graph in the document.

9. Click on the attribute `How_do_you_feel_about_school` and drag it to the *y*-axis of the empty graph.

You should get a bar chart. We placed the attribute on the *y*-axis rather than the *x*-axis because the bar labels are more readable that way.



10. Resize the graph until it looks similar to the one shown at right.

How does your conjecture hold up against the actual data?

11. Move the mouse on top of one of the bars and look in status bar in the lower left corner of the window. You should see an indication of the number and percentage of cases in that bar, as shown at right.

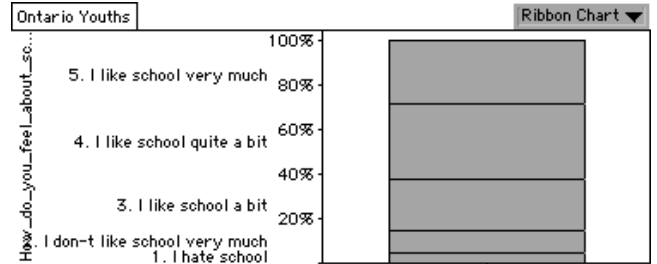
We're going to compare a bar chart of the responses to this survey question with a ribbon chart representation.

Fathom orders categories alphabetically. Here the category numbers insure that the alphabetic order makes logical sense.

12. Make a second bar chart just like the first.

13. Choose **Ribbon Chart** from the popup menu in the top left corner of the graph.

Your second graph should look similar to the one shown at right.



14. Move the mouse on top of the bar and look in the status bar.

As it does with a bar chart, the status bar displays the number and percentage of cases in the category under the mouse.

Use the following questions as an aid to thinking about the strengths and weaknesses of each graph.

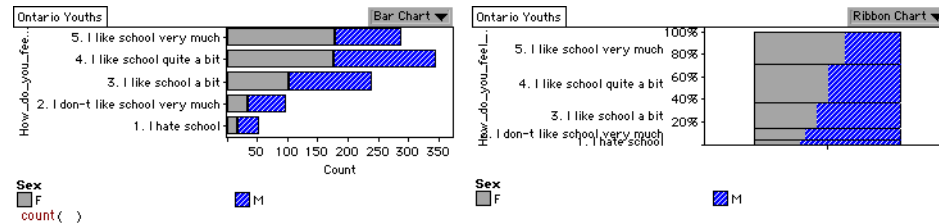
- With which graph can you more easily identify the category with the most responses? Can it be done with the other graph?
- With which can you visually estimate the percentage of responses in a given category? Can it be done with the other graph?
- Which graph makes it easy to visually estimate the percentage of youths that like school either very much or quite a bit? Can it be done with the other graph?

The Sex Difference—Splitting the Bars

What will be the difference between males and females in this age group with regard to liking school? Make a conjecture.

15. Click and drag the attribute `Sex` into the *middle* of each of the graphs.

The two graphs should look similar to those shown below. Note that the legend identifies which portion of each bar is male and which female.



Each graph has its strengths and weaknesses.

- With which graph can you most easily determine the category containing the highest *number* of females?
- With which graph can you most easily determine the category containing the highest *proportion* of females?
- Which graph makes it easy to decide if there is a trend? (What is that trend?)

To what extent is your conjecture about males versus females supported by the data?

Tables

Let s display the same information in tabular form.

*You can drag a summary table from the tool shelf or use the **Summary Table** command in the **Insert** menu.*

16. Make an empty summary table in the document.
17. Drag the `How_do_you_feel_about_school` attribute into the row area of the table and the `Sex` attribute into the column area of the table.

		Sex		Row Summary
		F	M	
How_do_you_feel_about_school	1. I hate school	17	37	54
	2. I don't like school very much	35	64	99
	3. I like school a bit	103	138	241
	4. I like school quite a bit	177	169	346
	5. I like school very much	180	109	289
Column Summary		512	517	1029

18. Resize the table until it looks similar to the one shown at right.

The table reports a total of 1029 cases. But originally we determined there were 1046 cases. What s going on?

By default Fathom gives you the count for each cell. What kinds of questions can you answer with that information?

19. Choose **Add Formula** from the **Insert** menu.
20. In the resulting formula editor open the **Special** category.
21. Double-click on `rowProportion` so that it gets entered into the formula.
22. Click the **OK** button.
23. Resize the table so that you can see all its values at once.

Special
caseIndex
columnProportion
columnTotal
expected
false
grandTotal
pi
rowProportion

Verify the trend you saw with the ribbon chart using the numbers reported in the table.

These three representations bar chart, ribbon chart, and summary table each have their strengths and weaknesses. Which did you find most useful for this investigation?

Extra

- For what other questions is there considerable difference between the sexes? For what attributes is there little difference?
- When you drop a categorical attribute on the *x*-axis and another categorical attribute on the *y*-axis, you get a breakdown plot. Investigate and characterize such plots.