

# Exploring Categorical Data—The Technology Use Survey of Canada

## *Exploring Mathematics with Fathom Institute*

This activity is designed for workshops with teachers in Canada who have access to the data from the Statistics Canada General Social Survey, Cycle 14, on technology use among those age 15 and over living in a private household.

The purpose of this activity is to introduce teachers to Fathom as a tool for exploring categorical data. Though the activity could be adapted for classroom use, it was not designed for this purpose.

### Opening the Fathom Document\*

1. From within Fathom choose Open from the File menu.
2. Navigate to the TechUse.ftm document. It is most likely within the Statistics Canada Data folder within the Sample Documents folder.
3. Open this document.

It should look similar to the document shown at right.

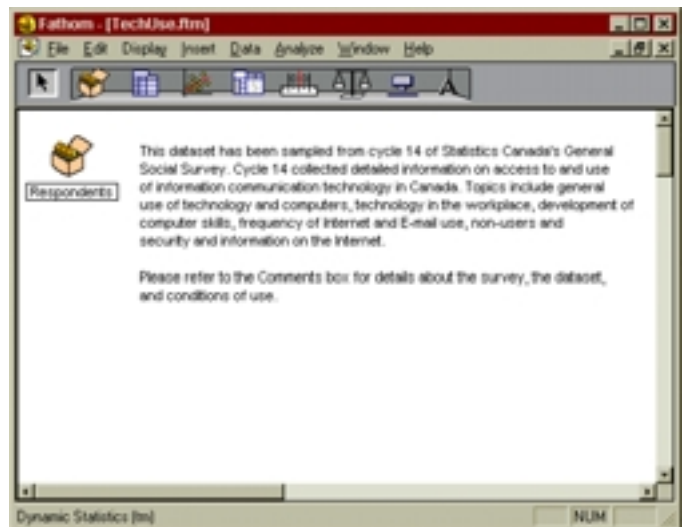
4. Read the text in the document in order to become familiar with the origin and nature of the data.

The text in the document takes up a lot of room. Let's hide it.

5. Click anywhere in the text.

The text object acquires a border to show that it is selected.

6. Choose Hide Text Object from the Display menu.



### Inspecting Individual Cases

The box with the gold balls labeled Respondents is a *collection*. It contains information about 1000 people or *cases*. These cases can be viewed one at a time or in a case table.

7. Click once on the collection (the box of gold balls).
8. As shown at left, move the mouse over the lower-right corner of the border.
9. Click and drag the mouse down and to the right to "open" the collection.

*The text object can be made visible again using the Show Hidden Objects command in the Display menu.*



\* The Fathom document used here has 1000 cases and 84 attributes. On a Macintosh, you should set the Fathom memory partition to at least 25 megabytes. If you experience working with the document to be overly slow and tedious, use the TechUseSmallA and TechUseSmallB documents instead because they have many fewer attributes.

Or choose Inspect Case from the Edit menu.

As the collection opens you should see some gold balls, as shown at right. Each gold ball represents one person who took the survey. The number below each ball is the identification number for that person.

10. Double-click the first of the gold balls.
11. Click the Cases tab near the top-right of the resulting *inspector* window.

The *inspector* for the collection displays the *attributes* and attribute *values* for the case.

Notice the arrows in the lower-left corner of the inspector. Next to the arrows it says, "1/1000," indicating that this is the first case out of 1000.

12. Click on the right-pointing arrow several times, to look at the values for other cases in the collection.
13. Make the inspector somewhat bigger by dragging the lower-right corner of the window.

There are 84 attributes for each case. Each attribute corresponds to a question on the original survey.

If you're looking at one of the small documents there are only 44 attributes.

14. Use the inspector's scroll bars to look through the list of attributes.

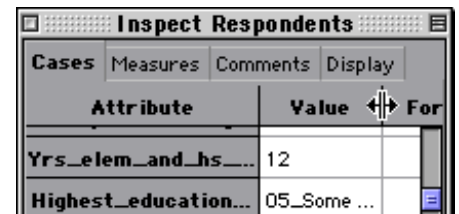
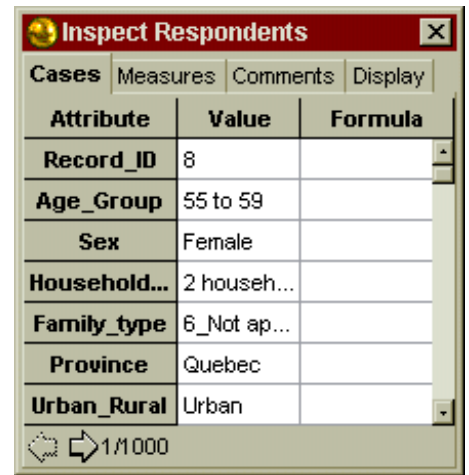
Many of the attribute names are longer than will display in the Attribute column, and many of the values do not fully display within the Value column.

15. Drag the boundaries between the columns (as shown at right) to make more room for attribute names and values.

The inspector "floats" above the document. It provides a convenient way to view data for individual cases without taking up space in the document.

16. Close the inspector (for now) by clicking in its close box.
17. Drag the lower-right corner of the collection towards its upper-right corner until it collapses.

You should now have plenty of space available in the document for the work that follows.



## Viewing Cases in a Case Table

Sometimes it is helpful to view data in a table such that each row is one case and each column holds the values of an attribute.

18. Click once on the collection so that it is selected.



19. Position the mouse over the case table icon in the tool bar. Click and drag a new case table into the document.

Your case table should look similar to the one shown at right.

20. Experiment with scrolling the case table both vertically and horizontally.
21. Experiment with resizing the case table by dragging its edges or corners.

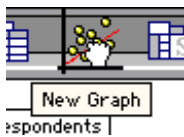
	Record_ID	Age_Gro...	Sex
1	8	55 to 59	Female
2	121	18 to 19	Female
3	125	60 to 64	Male
4	128	18 to 19	Female
5	136	80 years a...	Female

*The most common problems people experience in early stages of using Fathom comes from making the case table so big that there isn't room for other objects, especially graphs. Keep your case tables small!*

## Making Graphs

Usually graphs provide much more powerful tools for exploring data than do tables. Suppose we are interested in Internet connectivity and how it varies from one province to another.

22. Drag a graph from the tool bar into an empty place in the document.



The graph comes up empty because Fathom doesn't know what attribute to plot.

23. Click and drag the attribute `Connected_to_Internet` (the 8<sup>th</sup> one) to the x-axis of the graph.

Record_ID	Connected_to_Internet	Num
2	No	0
1	Yes	2

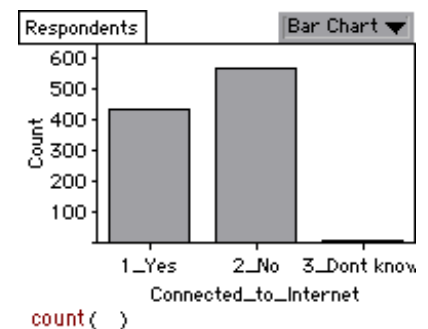


This attribute is *categorical*, so Fathom creates a bar chart with the height of each bar proportional to the number of cases in each category.

What does this graph tell you about internet connectivity?

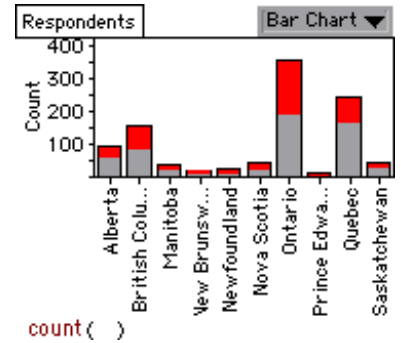
24. Make a second bar chart, this time with `Province` on the x-axis.
25. Click once on the bar labeled `1_Yes` in the first bar chart.

When you select cases in one view, Fathom shows them as selected in all views.



What does this graph tell you about internet connectivity across Canada's provinces?

Notice that because the counts vary greatly from one province to another, it is fairly difficult to compare the *proportion* of respondents from each province that have internet connectivity. A ribbon chart makes a much better tool for such comparisons.

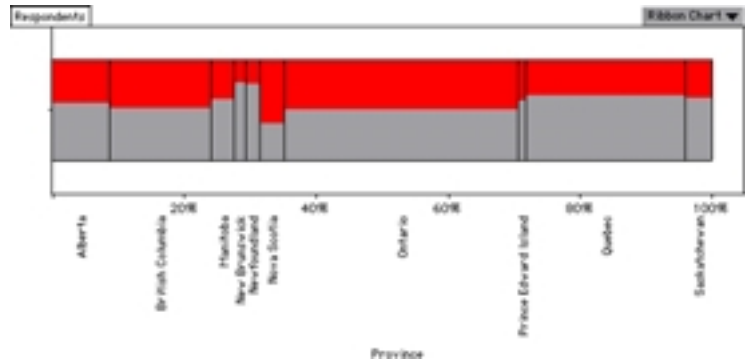


- Hold down the mouse on the Bar Chart popup menu in the upper-right corner of the bar chart of provinces and choose Ribbon Chart.

A ribbon chart allocates a strip along its length for each category. The width of the strip is proportional to the count. Since selection shows vertically in each strip, it is relatively easy to compare the proportion selected in different categories.

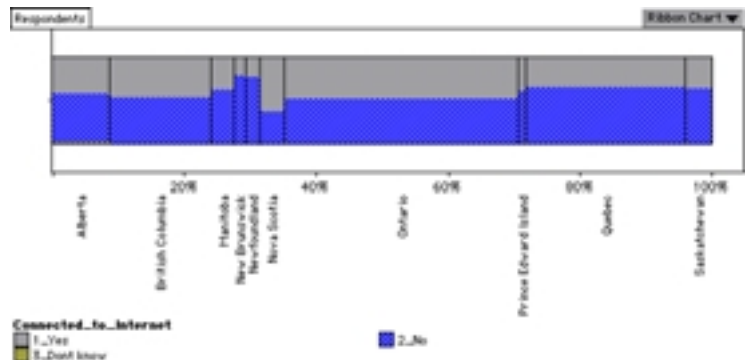
Which provinces have the highest and lowest proportions of respondents with internet connectivity?

- Get rid of the selection by clicking in white space in a graph.
- Drag the Connected\_to\_Internet attribute into the *middle* of the ribbon chart.



Dropping a categorical attribute into the middle of a ribbon chart splits the each category and shows a legend. This is more persistent and often more legible than using selection.

Note: Ribbon charts pay back many times over the time spent becoming familiar with them.



## Displaying Summary Statistics

Sometimes you want to have the actual numbers, not just a picture. Let's look at the information we generated using a ribbon chart, but this time in a summary table.



- Drag a summary table from the tool bar and place it in an empty area of the document.
- Drag the Connected\_to\_Internet attribute to the empty column header of the summary table (as shown at right).



31. Resize the summary table so that you can see all of its contents. It will look similar to the one shown at right.

Respondents	Summary Table			Row Summary
	Connected_to_Internet			
	1_Yes	2_No	3_Dont know	
	433	566	1	1000

*S1 = count( )*

We want to break these results down according to province.

32. Drag the Province attribute onto the summary table's empty row header.

33. Resize the resulting table so that you can see all the provinces (shown below).

Note that in the illustration below the table has been adjusted so that it is too narrow to show the third column which we're not very interested in anyway.

Respondents	Summary Table			Row Summary
	Connected_to			
	1_Yes	2_No		
Province	Alberta	37	50	88
British Columbia	74	91	155	
Manitoba	13	21	34	
New Brunswick	4	14	18	
Newfoundland	5	16	21	
Nova Scotia	24	14	38	
Ontario	173	102	355	
Prince Edward Island	4	6	10	
Quebec	65	150	243	
Saskatchewan	14	24	38	
Column Summary	433	566	1000	

*S1 = count( )*

What can you conclude from this table? Describe the coincidence for Saskatchewan and Nova Scotia.

This summary table is much like the bar chart we made in the previous section. It's fine for comparing counts, but not so good for comparing proportions. We can change that by changing the formula for the values computed in each cell.

34. Double-click on the formula displayed under the summary table.

This will open up Fathom's formula editor so that you can change what is computed.

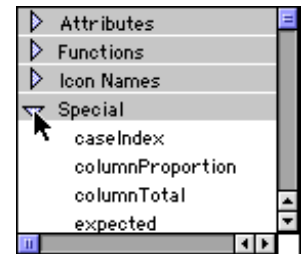
35. Click on the plus mark to the left of the special heading in the formula editor.

This should open up (as shown at right) a list of "special" quantities that Fathom can compute.

36. Scroll the list down until you can see the rowProportion entry.

37. Double-click the rowProportion entry causing it to be entered into the formula pane as shown at left.

38. Click the OK button at the bottom of the formula editor.



Respondents	Summary Table
	433

*S1 = count( )*

*S1 = count( )*

*On a Mac, the plus marks are triangles.*

*rowProportion*

The summary table recomputes. The top part of the result is shown at right.

Explain what the values in the cells represent. Rank the provinces from highest to lowest in internet connectivity.

Respondents	Summary Table		
	Connected_to_Int		
	1_Yes	2_No	
Province	Alberta	0.42045455	0.56818182
British Columbia	0.47741935	0.52258065	
Manitoba	0.38235294	0.61764706	
New Brunswick	0.22222222	0.77777778	

### Filter Cases

Sometimes we're only interested in a subset of the available data. Suppose, for example, we want to compare the internet connectivity of Ontario and Quebec without necessarily being distracted by data for the other provinces.

39. Click once on the ribbon chart of internet connectivity split by province.

40. Choose Add Filter from the Data menu.

This brings up the formula editor. The expression you create here will filter out those cases for which the value of the expression is false.

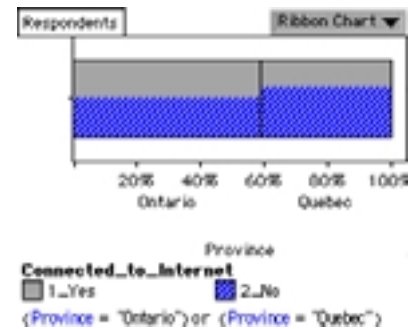
41. Create the expression shown at right.

`(Province = "Ontario") or (Province = "Quebec")`

Some hints:

- You can type the whole expression, or you can enter attribute names by double-clicking them in the list of attributes.
- When you type a left parenthesis, the right parenthesis appears automatically.
- When you type a double quote, its matching double quote appears automatically.
- Attribute names that are correctly entered appear in blue.
- There is no way to find the list of possible categories for an attribute.

If you got the correct expression for the filter, your graph should look similar to the one at right. With less going on, it's much easier to decide which province has greater internet access.



### Two things to know about filters:

- A filter on a table or graph affects only that table or graph. You might, for example, make two graphs, one of just the males and another of just the females by placing the appropriate filter on each graph.
- A filter on the collection affects *all* views of that collection. If all your work were going to be restricted to just Ontario and Quebec, it would make sense to put a filter on the collection rather than to have to put a filter on every display of the data.

## Challenges

1. Are people with higher incomes more or less likely to use the Internet for banking?
2. Characterize the differences between males and females with regard to computer use. Find at least two characteristics that do not conform to this general trend.
3. Use filters to restrict the number of people displayed in a case table to the single person that is most like yourself.