Addition Rule—Scottish Children

You will need
- **Scottish Children.ftm**

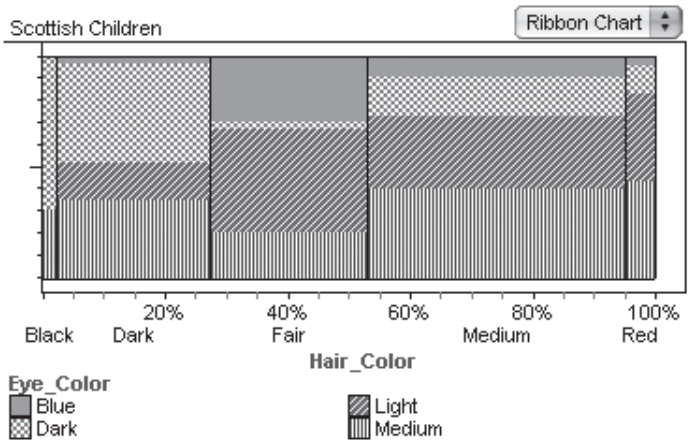
In this activity you'll use sample data from a survey of Scottish children to explore when you can find $P(A \text{ or } B)$ if you know $P(A)$ and $P(B)$.

EXAMINE DATA

1. Open the Fathom document **ScottishChildren.ftm.** You will see a collection of 500 children from Scotland who were randomly selected to participate in a survey.

The last attribute, *Subject,* is the number this child was in the study. This is a random sample of the original sample of 5387 children.

2. Double-click the collection to show the inspector. Make a ribbon chart for *Hair_Color,* then drag *Eye_Color* into the *interior* of the graph. (*Eye_Color* is now the legend.) The entire bar represents all the children. Each slice represents a different hair color. The four sections of each slice represent the four different eye colors.



Cases: 500 Selected: 111

3. Click on one section of the ribbon chart. Hold down the Shift key and click on additional sections. Now move your cursor over the *collection* (not the graph) and notice that in the lower-left corner of the Fathom window you can read the number of selected cases.

Use this method to answer these questions.

Q1 How many children have *either* medium *or* red as their hair color?

Q2 How many children have *either* blue eyes *or* medium hair color?

INVESTIGATE

Addition Rule

4. Define a new attribute, *IsMedium,* with the formula Hair_Color="Medium". This attribute will be true if the child's hair color is medium. Similarly, define

an attribute *IsRed* that will be true if the child's hair color is red, and define an attribute *IsBlueEyed* that will be true if the child has blue eyes.

Formulas such as count(IsBlueEyed) will be useful.

5. Go to the **Measures** panel of the inspector and define the measures *Medium* (the number of children whose hair color is medium), *Red* (the number of children whose hair color is red), and *BlueEyed* (the number of children who have blue eyes).

Q3 From the values of these three measures alone, can you determine the number of children whose hair color is medium or red? If so, explain how to do it. If not, what additional information do you need?

Q4 From the values of these three measures alone, can you determine the number of children who have blue eyes or who have medium as their hair color? If so, explain how to do it. If not, what additional information do you need?

In the formula editor, use the **or** button on the keypad for the logical expression "or."

6. Define the measures *MediumOrRed* (the number of children whose hair color is medium or red) and *BlueEyedOrMedium* (the number of children who have blue eyes or who have medium as their hair color). If possible, use the measures that already exist as part of your formulas.

Q5 Write your formulas for the measures you defined in step 6 and record their values.

Q6 True or false: $P(Medium \text{ or } Red) = P(Medium) + P(Red)$. Explain.

Q7 True or false: $P(BlueEyed \text{ or } Medium) = P(BlueEyed) + P(Medium)$. Explain.

Q8 What conclusions can you draw so far from doing this activity?

Extending the Rule

You should have found that with the measures *BlueEyed* and *Medium* alone, you can't determine the number of children who are blue-eyed or who have medium as their hair color.

In the formula editor, use the **and** button on the keypad for the logical expression "and."

7. Define a new measure, *BlueEyedAndMedium,* which counts the number of children who are blue-eyed and have medium as their hair color.

Q9 Write your formula for the measure you defined in step 7 and record its value.

Q10 From the values of *BlueEyed, Medium,* and *BlueEyedAndMedium* alone, can you determine the number of children who are blue-eyed or who have medium as their hair color? If so, explain how to do it. If not, what additional information do you need?

8. Make a summary table with *Eye_Color* for the rows and *Hair_Color* for the columns. The summary table is equivalent to a two-way table.

Scottish Children

		Hair_Color					Row Summary
		Black	Dark	Fair	Medium	Red	
Eye_Color	Blue	0	5	38	21	1	65
	Dark	9	55	5	36	3	108
	Light	0	21	59	69	9	158
	Medium	4	44	26	85	10	169
Column Summary		13	125	128	211	23	500

S1 = count( )

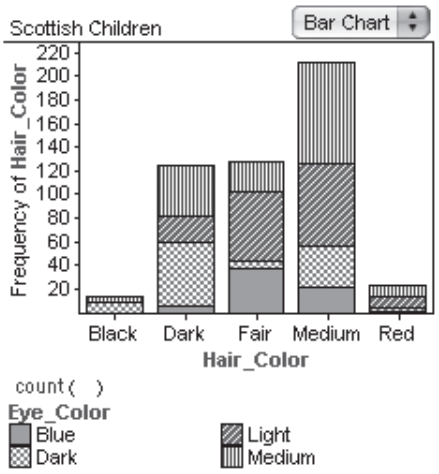
Q11 If you select a child at random from your collection, what is the probability that the child has blue eyes or has medium as a hair color?

Q12 What is the probability that a randomly selected child has black hair?

EXPLORE MORE

1. Make a second ribbon chart, with *Eye_Color* on the horizontal axis and *Hair_Color* as the legend attribute. Select sections of this graph and predict which sections will highlight in the first ribbon chart.

2. Make a bar chart of *Hair_Color* and drag *Eye_Color* into the interior of the graph, or make a bar chart of *Eye_Color* and drag *Hair_Color* into the interior of the graph. Explore how each section of the bar chart corresponds to each section of the ribbon chart. Compare and contrast the two types of graphs.



You can change the order of the categories by dragging a label.

3. Make a breakdown plot with *Hair_Color* on the horizontal axis and *Eye_Color* on the vertical axis. Determine which sections of the breakdown plot correspond to which cells of the summary table.

Addition Rule—Scottish Children

Activity Notes

Objectives

- Recognizing that $P(A \text{ or } B) = P(A) + P(B)$ only for disjoint categories
- Extending the addition rule to $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ when A and B are not disjoint

Activity Time: 30–50 minutes

Setting: Paired/Individual Activity or Whole-Class Presentation (use **ScottishChildren.ftm** for either)

Statistics Prerequisites

- Definition of probability
- Familiarity with the concepts of "or" and "and" (not the formulas)

Statistics Skills

- Addition rule for disjoint events
- Working with the definition of probability
- General addition rule
- Calculating probabilities with graphs and two-way tables
- *Optional:* Stacked bar charts (Explore More 2)

AP Course Topic Outline: Part III A

Fathom Prerequisites: Students should be able to make bar and ribbon charts and to define attributes and measures.

Fathom Skills: Students define formulas using the logical expressions "or" and "and," use a summary table as a two-way table, work with ribbon charts to calculate probabilities, and work with legends in ribbon charts. *Optional:* Students make stacked bar charts (Explore More 2) and breakdown plots (Explore More 3).

General Notes: This important activity helps students see why you can't always add $P(A)$ and $P(B)$ to get $P(A \text{ or } B)$. It also shows practical ways to extend the addition rule to compute $P(A \text{ or } B)$ when A and B are not disjoint. Fathom eases the computational burden of this activity, allowing students to focus on the concepts. The collection of 500 children is actually a random sample of a much larger sample of 5387 children. (Source: D. J. Hand et al., *A Handbook of Small Data Sets* (London: Chapman and Hall, 1994), 146. Their source: L. A. Goodman, "Association models and canonical correlation in the analysis of cross-classifications having ordered categories," *Journal of the American Statistical Association* 76 (1981): 320–334.))

Students write logical expressions to count the number of children with various values for attributes. This activity also reviews the ribbon chart, an important tool for reasoning about sets.

EXAMINE DATA

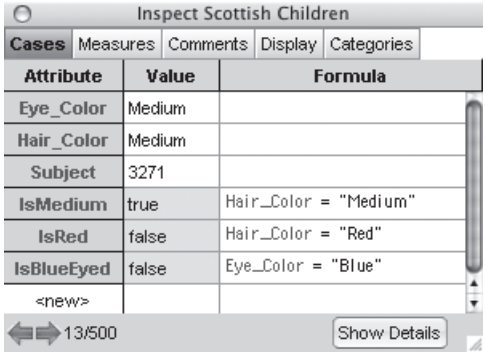
3. A common mistake students make here is dropping *Eye_Color* on the vertical axis instead of in the graph's interior. If they make this mistake, they can choose either **Edit | Undo** or **Graph | Remove Y Attribute.**
4. Because answering questions based on a ribbon chart can appear simple but become confusing, encourage students to spend time gaining understanding of ribbon charts.

Q1 234 children

Q2 255 children

INVESTIGATE

5. Here are the formulas for the complete set of attributes for the Scottish Children collection.



Q3 The categories *Medium* and *Red* are disjoint, so the answer can be found with the sum count(IsMedium)+count(IsRed). This value is equivalent to count(IsMedium or IsRed).

Q4 The categories overlap. You need to know how many have blue eyes and select *Medium* as their hair color. Hence, count(IsBlueEyed)+count(IsMedium) is not equivalent to count(IsBlueEyed or IsMedium).

6.–7. Here is the complete set of measures. The last, *BlueEyedNotMedium,* is not explicitly requested in the activity but is useful in understanding that if you know this quantity, then you can add it to *Medium* to get *BlueEyedOrMedium.* You may want to challenge students to use the alternative "not" statement (*MediumNotBlueEyed*) and explain how to use it to find *BlueEyedOrMedium.*

Inspect Scottish Children

Cases | **Measures** | Comments | Display | Categories

Measure	Value	Formula
Medium	211	count(IsMedium)
Red	23	count(IsRed)
BlueEyed	65	count(IsBlueEyed)
MediumOrRed	234	count(IsMedium or IsRed)
BlueEyedOrMedium	255	count(IsBlueEyed or IsMedium)
BlueEyedAndMedium	21	count(IsBlueEyed and IsMedium)
BlueEyedNotMedium	44	count(IsBlueEyed and ¬IsMedium)

Q5–Q7 The formulas are given in the inspector above, as are the values.

Q6 True; they are disjoint (see explanation in Q3).

Q7 False; they overlap (see Q4).

Q8 The conclusion to be drawn so far is that only for disjoint categories does *count*(*A* or *B*) = *count*(*A*) + *count*(*B*).

Q9 See the inspector above for the formula. There are 21 children who are blue-eyed and have medium as their hair color.

Q10 Students should conclude that *BlueEyedOrMedium,* previously defined by the formula count(IsBlueEyed or IsMedium), is equivalent to *BlueEyed* + *Medium* − *BlueEyedAndMedium.*

8. The summary table that students make here is exactly analogous to the ribbon chart they made in step 2.

Because Q11 and Q12 ask for probabilities and not numbers, you might want to show students how to change the summary table's default formula from counts to proportions. Double-click count() and enter count()/GrandTotal. This will give the proportion of cases in each cell.

Q11 The probability can be calculated two ways. Using the formula:

$$P(\textit{BlueEyed} \text{ or } \textit{Medium}) = P(\textit{BlueEyed}) + P(\textit{Medium}) - P(\textit{BlueEyed} \text{ and } \textit{Medium}) = \frac{65}{500} + \frac{211}{500} - \frac{21}{500} = \frac{255}{500} = \frac{51}{100}$$

Using the summary table, count just the shaded cells, counting 21 once. So

$$P(\textit{BlueEyed} \text{ or } \textit{Medium}) = \frac{5 + 38 + 21 + 1 + 36 + 69 + 85}{500} = \frac{255}{500} = \frac{51}{100}$$

You could also use the total of the column "Medium" and add to that the shaded cells in the "Blue" row that are not in the "Medium" column:

$$P(\textit{BlueEyed} \text{ or } \textit{Medium}) = \frac{211 + 5 + 38 + 1}{500} = \frac{51}{100}$$

Scottish Children

		Hair_Color					Row Summary
		Black	Dark	Fair	Medium	Red	
Eye_Color	Blue	0	5	38	21	1	65
	Dark	9	55	5	36	3	108
	Light	0	21	59	69	9	158
	Medium	4	44	26	85	10	169
Column Summary		13	125	128	211	23	500

S1 = count( )

Q12 13/500

DISCUSSION QUESTIONS

If possible, ask the following questions while projecting the Fathom document in front of the class.

- How could you use this ribbon chart to find the number of children with blue eyes?
- Which sections of the ribbon chart do you need to select to find the number of children whose hair color is *Red* or *Medium*?
- Which sections of the ribbon chart do you need to select to find the number of children who have light eyes or who have red as their hair color?
- What formulas did you use for measures *MediumOrRed* and *BlueEyedOrMedium*? How many different yet equivalent formulas can you write?

- What did you conclude in Q10?
- If there are two not necessarily disjoint events *A* and *B*, how can you compute *P*(*A* or *B*)?

EXPLORE MORE

2. The fundamental difference is that the ribbon chart shows percentages whereas the bar chart shows actual quantities.

3. This breakdown plot is in the same order as the summary table.