Proportional Venn Diagrams

Two Properties

Consider a population that has attached to it two yes or no properties. We can, for example, take our population of drivers from the insurance simulation in unit 1. Each driver either had a history of impaired driving (yes or no) and passed a driver education course (yes or no). This means that in general, each person can fall into one of four categories as described by the table below:

<table>
<thead>
<tr>
<th>Impaired &amp; Took Course</th>
<th>Not Impaired &amp; Took Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impaired &amp; No Course</td>
<td>Not Impaired &amp; No Course</td>
</tr>
</tbody>
</table>

Suppose that our population had divided itself as follows:

- 100 Drivers Drive Impaired and **HAVE NOT** taken a driver education course.
- 25 Drivers Drive Impaired but have taken a driver education course.
- 25 Drivers **DO NOT** Drive Impaired and **HAVE NOT** taken a driver education course.
- 100 Drivers **Do Not** Drive Impaired but have taken a driver education course.

There are many ways to depict these 4 categories proportionally (and thus give visual meaning to the data). One way is to view each category as square of area at most 1. Then the second and fourth categories are a quarter the size of the first and third categories. A $1 \times 1$ square has area 1 and so a $\frac{1}{2} \times \frac{1}{2}$ square has area $\frac{1}{4}$. This gives the following

Suppose we wish to construct a proportional Venn Diagram of the following hypothetical results for Lollipops Versus Gummy Bears:

(A) 125 Little Girls like Lollipops but not Gummy bears.
(B) 60 Little Girls like Gummy Bears but not Lollipops.

(C) 20 Little Girls do not like Lollipops or Gummy Bears.

(D) 200 Little Girls love both Lollipops and Gummy bears.

How should we make each of the four categories in proportion?

**SOLUTION** Let us assume that the biggest square which represents 200 children is a unit square (i.e. $1 \times 1$). Then the 115 Girls in category $A$ should be in a square which has area $\frac{115}{200} = \frac{5}{8}$. Such a square would need to have sides of length $x$ where

$$x^2 = \frac{5}{8} \Rightarrow x = \frac{\sqrt{5}}{2\sqrt{2}} \approx 0.79$$

Similarly, girls in Category $(B)$ would produce a square with area $\frac{60}{200} = \frac{3}{10}$ implying sides of length $y$ with

$$y^2 = \frac{3}{10} \Rightarrow y = \frac{\sqrt{3}}{\sqrt{10}} \approx 0.55$$

Finally, besides offering them chocolate, the girls in category $(C)$ need a square with area $\frac{20}{200} = \frac{1}{10}$. This will have sides of length $z$ with

$$z^2 = \frac{1}{10} \Rightarrow z = \frac{1}{\sqrt{10}} \approx 0.32$$

And so we get the following:
Three Properties

Now suppose that we have 3 yes or no properties that can be applied to our population. For example, suppose we want to know for each student at the Whamo Zamo Kid Academy whether they

- Are good at sports **YES** or **NO**
- Are reading at or above their grade level **YES** or **NO**
- Are doing math at or above their grade level **YES** or **NO**

With three categories and two possibilities for each we have \(2 \times 2 \times 2 = 8\) total classification possibilities which we will encode as follows. The table below also provides a hypothetical distribution for the kids at the Whamo Zamo Kid Academy:

<table>
<thead>
<tr>
<th>CODE</th>
<th>Description</th>
<th>Number at Whamo Zamo</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZERO</td>
<td>Good at Nothing</td>
<td>70</td>
</tr>
<tr>
<td>MATH</td>
<td>Good at Math ONLY</td>
<td>30</td>
</tr>
<tr>
<td>READ</td>
<td>Good at Reading ONLY</td>
<td>90</td>
</tr>
<tr>
<td>SPORTS</td>
<td>Good at Sports ONLY</td>
<td>25</td>
</tr>
<tr>
<td>MR</td>
<td>Good at Math &amp; Reading but NOT Sports</td>
<td>60</td>
</tr>
<tr>
<td>MS</td>
<td>Good at Math &amp; Sports but NOT Reading</td>
<td>100</td>
</tr>
<tr>
<td>SR</td>
<td>Good at Sports &amp; Reading but NOT Math</td>
<td>10</td>
</tr>
<tr>
<td>ALL</td>
<td>Advanced at Everything</td>
<td>100</td>
</tr>
</tbody>
</table>

Below we indicate a Venn Type Structure (not yet a proportional Venn) with regions where each of the 8 possibilities can exist.

Notice how each attribute of Sports, Math, and Reading each control one aspect of the figure:

- Reading is the top Row (so Bottom row is Not Reading)
- Math is the left column (so Right Column is Not Math)
- Sports is the inner square (so outside the inner square is Not Sports)

Now let us compute the changes in size needed to make this a proportional Venn Diagram. We will start with the inner square. Then in the same spirit as we followed for the 2 category discussion we note that **All, SR, SPORT, MS** are squares with respective sides 10, \(\sqrt{10} \approx 3.16\), 5, and 10. Now as the other regions are not squares it takes a little more work to force them to have the correct area. Consider the following diagram in which the quantity \(x\) is always known and in our case is given by the values 10, \(\sqrt{10}, 5, \text{ and } 10\) just mentioned.
Notice that this L shaped region consists of 2 rectangles of dimensions $x$ by $y$ and 1 square of dimension $y$ by $y$. Its area is therefore $y^2 + 2xy$.

In each case we must solve the relevant quadratic. Notice that we are always discarding the negative solutions.

**MR**

$$y^2 + 20y - 60 = 0 \Rightarrow y = 4\sqrt{10} - 10 \approx 2.65$$
READ
\[ y^2 + 2\sqrt{10}y - 90 = 0 \Rightarrow y = 10 - \sqrt{10} \approx 6.84 \]

ZERO
\[ y^2 + 10y - 70 = 0 \Rightarrow y = \sqrt{95} - 5 \approx 4.75 \]

MATH
\[ y^2 + 20y - 30 = 0 \Rightarrow y = \sqrt{130} - 10 \approx 1.40 \]

And so our Final diagram is as shown below: