Paper 109-2011

Creating Proportional Venn Diagrams Using Google and SAS®

Hillary Kruger, RTI International, Durham, NC, USA

ABSTRACT

This paper presents a simple process to create proportional Venn diagrams within SAS® using the API for Google charts. Many features are available to create a Venn that suits any need.

INTRODUCTION

Data visualization has emerged as an important and engaging new topic in research. We often need to display as much information as possible in as little space as possible. One of the ways to achieve this is with proportional diagrams. While SAS® offers many complex and informative graph options, a proportional Venn diagram is not currently available. Assigned to create a proportional Venn, I investigated what was available and found Google charts. I created a proportional Venn using summary data and simple HTML (Hyper-Text Mark-up Language) code variation.

BACKGROUND

Venn diagrams show the intersection of data, which is quite useful to many scientific fields. Venn charts are relatively non-complicated, easy-to-understand charts. This simplicity can be misleading, however, if the data is not evenly distributed. The Venn diagram is an increasingly popular way to express a point, mathematical or not, as well as in research. This spike in interest may prove how Venn diagrams show quantitative data in a format that is easy for a non-scientific audience.

In some scenarios, the sample sizes aren't important in order to show a point, like in Figure 2. In others, and for most uses in scientific research, the size of the sample or population of comparison is important, if not crucial, to the scientific process. Figure 3 shows a traditional, squared Venn. Figure 3 may appear proportional at first glance, but upon closer inspection, the numbers are not represented by the size of the squares and rectangles. The information relayed by this chart may be numerically accurate, but it is still misleading.

Figure 2.

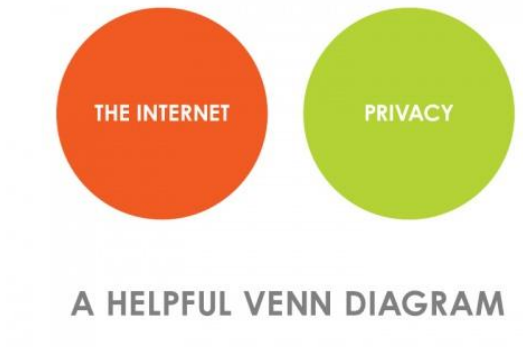
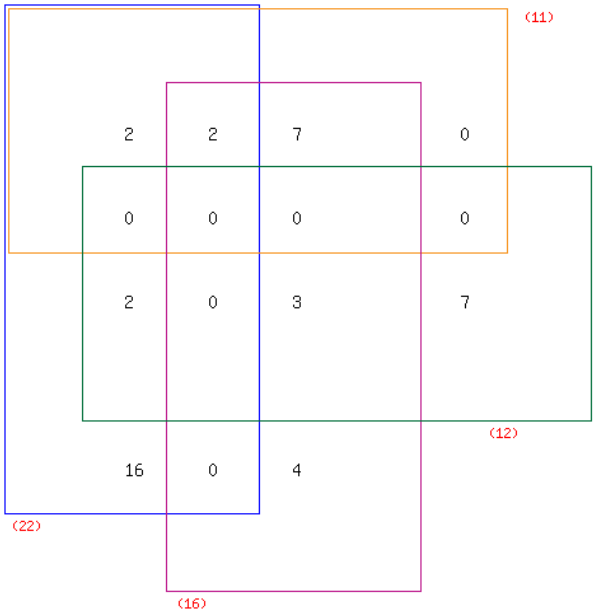


Figure 3.



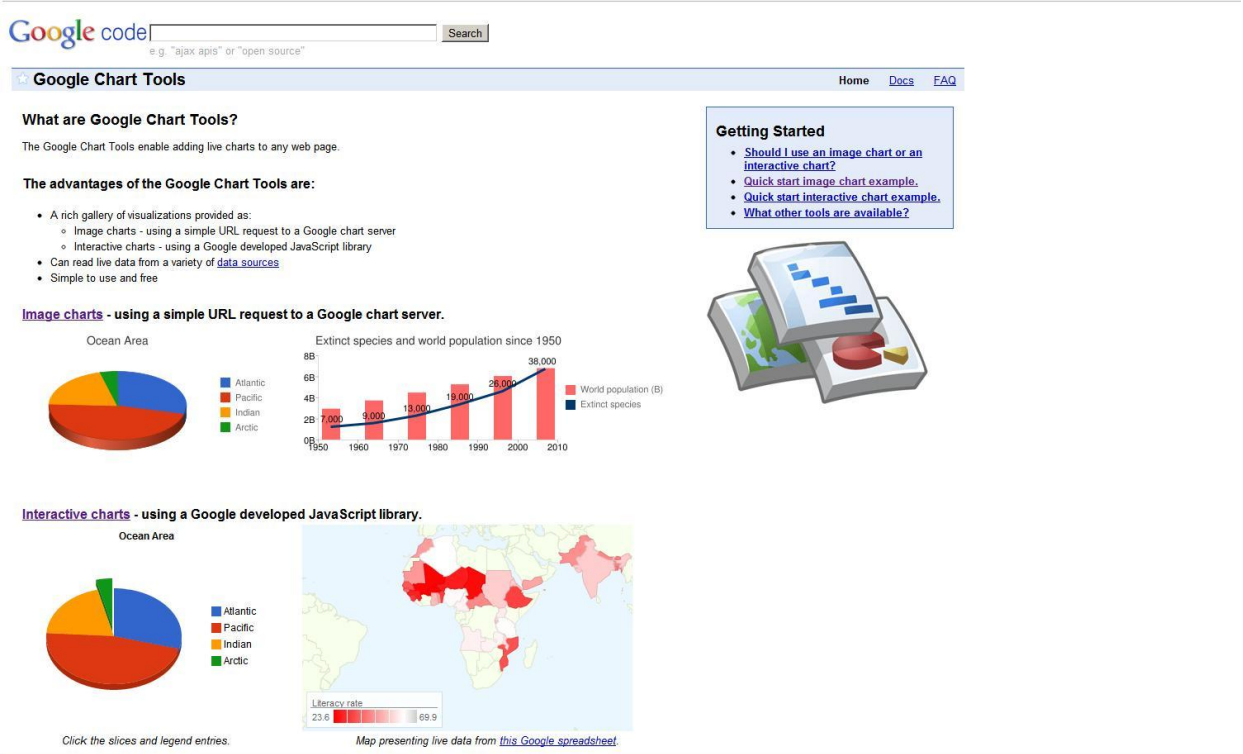
(Sources: Figure 2: http://datavis.tumblr.com/; Figure 3: http://foofish.blogspot.com/2007/10/short-survey-on-venn-diagrams-and.html)

GOOGLE CHARTS

Unlike Google maps, using these features do not require a key. There are no terms of service or use. There are two types of charts: image or interactive. Interactive charts are similar to SAS® charts with drill-down functionality. Here I will be discussing image charts – or charts that use a simple URL (Uniform Resource Locator).

A screen shot of the Google chart website (http://code.google.com/apis/charttools/) is shown in Figure 1.

Figure 1.



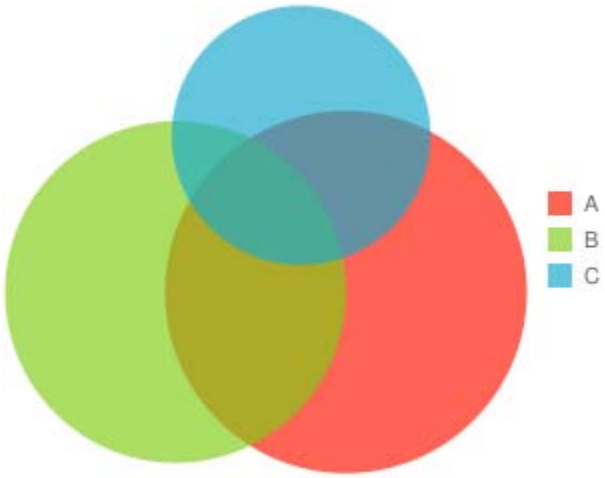
CODE

The following is a very basic implementation of Google chart code which creates the three-circle Venn displayed in Figure 4. (A more complex example is included in the appendix of this paper.)

```
%let A=35; %let AB=8; %let ABC=4;
%let B=30; %let AC=5;
%let C=15; %let BC=3;

%macro VennChart (size, type, col1, col2, col3, dat1, dat2, dat3, dat4, dat5, dat6,
     dat7, lab1, lab2, lab3);
data _null_;
file 'c:\VENN.html';
put '<img src="http://chart.apis.google.com/chart?
 chs='"&size."'
 &cht='"&type."'
 &chco='"&col1."','"&col2."','"&col3."'
 &chd=t:'"&dat1."','"&dat2."','"&dat3."','"&dat4."','"&dat5."','"&dat6."','"&dat7."'
 &chdl='"&lab1."'|'"&lab2."'|'"&lab3."'"
 width="300" height="300" alt="" />';
run;
%mend VennChart;
%VennChart (300x300,v,FF6342,ADDE63,63C6DE,&A.,&B.,&C.,&AB.,&AC.,&BC.,&ABC.,A,B,C);
```

Figure 4.



DETAILS

The program assumes you have the following information available, specific to three-circle Venns: intended size (in DPI or Dots Per Inch[2]), type (V for Venn), hexadecimal colors (one color for each large circle), detailed size overlap information (your data, in seven values, which should total 100), and label values. This file is saved as HTML, but can be saved in other formats once opened in a web browser. This capability means the diagram is easily saved and can be easily entered into documents or reports.

DISCUSSION

In order to create diagrams in Google charts, one merely has to find the information required. The size, colors, and labels are typically set by a standard implied by the audience (like a client or publication). They can easily be changed however, those details we will not address.

Most important in data visualization is the data itself. In many cases, the numbers needed to create the diagram can be found using a PROC FREQ. This is ultimately determined by the data, of course. It is left to the individual programmer to decide the best way to accomplish this.

There are seven important numbers included in the macro call. The first number is the percent of the total that A is exclusively A. The second is for B, the third for C. The fourth number is for the space that A and B share (or where that data overlaps). The fifth is for A and C and the sixth for B and C. The last number is the percent of total 'space' where all three A, B, and C intersect.

Others have created proportional Venns using a complex macro (see source section) but there are limitations using this process. Many casual SAS® users may be overwhelmed by the program. The Google API (Application Programming Interface) offers a quick, easy way to create accurate visualizations. This is something a SAS® user at any level can create and understand.

Once the data values are known and set (via %let statements), the user need only to apply formatting standards.

OTHER USES

Two-circle Venns are just as simple to create and require only a slight variation. The values for the third circle or "C" and its combinations are each set to zero, and the extra color is not included. Otherwise, the macro remains otherwise the same.

The hexadecimal color options allow the user to create full color charts. Google recommends varying the colors as much as possible for the best distinction between circles. However, I found grayscale charts to be just as informative.

It is possible to create the chart directly from your data source, though I feel the extra step of confirming numbers in a PROC FREQ to be useful. Confirming the overlapping percentages is always a good idea.

REFERENCES

Azimaee , Mahmoud. *KML Macro: Integrating SAS® and Google API and Its Application in Mapping Manitoba's Health Data on Google Earth and Google Map.* SAS Global Forum 2010.

Google Chart Tools / Interactive charts (aka Visualization API) Query Language Reference (Version 0.7). Available online: http://code.google.com/apis/visualization/documentation/querylanguage.html#plainText

Harris, Kriss. How *To Generate 2, 3 and 4 Way Venn Diagrams with Drill Down Functionality within 4 minutes!* SAS Global Forum 2008.

Li, Shiqun (Stan). *Using SAS® to Create Proportional Venn Diagrams.* SAS Global Forum 2009

Massengill, Darrell. *Google Maps and SAS/Graph®.* SAS Global Forum 2010.

Roehl, William G. *Bridging the Gap between the Google Analytics API and SAS®.* SAS Global Forum 2010.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name:	Hillary Kruger
Enterprise:	RTI International
Address:	3040 E. Cornwallis Road
City, State ZIP:	Durham, NC 27709
Phone:	(919)541-6243
E-mail:	hkruger@rti.org

APPENDIX

MACRO (VennMacro.sas)

```
%macro VennChart ( );
data _null_;
  set one;
  file &name. ;
  put '<img src="http://chart.apis.google.com/chart?chs='"&size."
  '&chdlp='"&lege."
  '&chtt='"&titl."
  '&cht='"&type."
  '&chmar='"&mar1."','"&mar2."','"&mar3."','"&mar4."
  '&chco='"&col1."','"&col2."','"&col3."
  '&chd=t:'"&dat1."','"&dat2."','"&dat3."','"&dat4."','"&dat5."','"&dat6."','"&dat7."
  '&chdl='"&lab1."'|' "&lab2."'|'"&lab3."
  'alt= '" "'/>'
;
run;
%mend vennChart;
```

MACRO CALL

```
%include "C:\VennMacro.sas";
filename VENN "C:\VENN.html";
data one; input var1; datalines; 1; run;
%let name = venn;            ** FILENAME;
%let type = v;               ** CHART TYPE - v=venn;
%let col1 = FF6342;          ** A COLOR;
%let col2 = ADDE63;          ** B COLOR;
%let col3 = 63C6DE;          ** C COLOR;
%let dat1 = 35;              ** SIZE OF A;
%let dat2 = 30;              ** SIZE OF B;
%let dat3 = 15;              ** SIZE OF C;
%let dat4 = 8;               ** SIZE OF A + B OVERLAP;
%let dat5 = 5;               ** SIZE OF A + C OVERLAP;
%let dat6 = 3;               ** SIZE OF B + C OVERLAP;
%let dat7 = 4;               ** SIZE OF A + B + C OVERLAP;
%let lab1 = Circle+A;        ** LABEL CIRCLE A;
%let lab2 = Circle+B;        ** LABEL CIRCLE B;
%let lab3 = Circle+C;        ** LABEL CIRCLE C;
%let widt = 300;             ** WIDTH IN DPI;
%let heig = 300;             ** HEIGHT IN DPI;
%let size = &widt.x&heig.;   ** WIDTH x HEIGHT;
%let lege = t;               ** LEGEND PLACEMENT - t=top;
%let titl = VENN+DIAGRAM;    ** TITLE;
%let mar1 = 5;               ** MARGIN - LEFT;
%let mar2 = 5;               ** MARGIN - RIGHT;
%let mar3 = 5;               ** MARGIN - TOP;
%let mar4 = 5;               ** MARGIN - BOTTOM;

%VennChart ();

run;
```

OUTPUT