

1
2
3
4
5
6
7
8
9
10
11
12
13
14

**A User's Guide to Diamond Graph:
2-D Graphical Display to Equally Present Effects of Two
Categorical Predictors on Continuous Outcomes**

Xiuhong Li and Alvaro Muñoz

Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health,
Baltimore, Maryland 21205

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Abstract

Diamond Graph (2003), a two-dimensional (2-D) graphical display, was proposed to equally present the effects of two categorical predictors on a continuous outcome. The key feature of Diamond Graph is to replace the parallelepiped volume in three-dimensional (3-D) bar graphs with a polygon, whose area and middle vertical length represent the outcome. Unlike 3-D bar graphs, Diamond Graph is invariant to rotations, and avoids the problem of outcomes being concealed by others in joint categories. It also circumvents the limitations of other 2-D alternatives when the primary objective is to equally present the effects of two categorical predictor variables on a continuous outcome. This paper provides three examples that apply Diamond Graphs to different data types, including proportions, incidence rates, and relative risks. The corresponding S-plus programs used to implement the Diamond Graphs in these examples are also included. The original paper of Diamond Graph published in The American Statistician (2003) and the S-Plus programs for plotting Diamond Graphs in the paper can be accessed at <http://www.statepi.jhsph.edu/~xili/diamondgraph.html>.

Key Words: Graphical methods; Diamond graphs; Three-dimensional bar graphs; Two-dimensional projections; S-plus programs; Statistics; Epidemiology.

1. Introduction

In biomedical and epidemiological studies, some factors are equally important to the outcome of interest. There is a need to graphically convey how two categorical variables (predictors or risk factors) jointly contribute to a continuous outcome. One way is to apply three-dimensional (3-D) bar graphs that are extensively used in biomedical journals (Klag, Whelton, Randall, et al. 1996; Huang, Hankinson, Colditz, et al. 1997; Farahmand, Michaëlsson, Baron, et al. 2000). However, it is well known that 3-D bar graphs are subject to false perception and interpretation, and confined to data showing non-overlapping trends, i.e., none of the outcomes in joint categories is concealed by others (Cleveland and McGill, 1984; Wilkinson, 1999; Harris, 1999).

The alternative way is to apply two-dimensional (2-D) graphs, such as mosaic (Hartigan and Kleiner, 1981, 1984; Friendly, 1994; Wilkinson, 1999), grouped bar graphs (Tufte, 1983), grouped dot plots and framed rectangle charts (Cleveland and McGill, 1984), and Trellis display (Becker, Cleveland, and Shyu, 1996). However, none of above-mentioned 2-D graphs is suitable for *equally* presenting the relationships between a continuous outcome and each of two categorical predictors in a single plot. The Diamond Graph presented here was developed to overcome these limitations of both 2-D and 3-D graphs (Li, Buechner, Tarwater, and Muñoz, 2003).

Diamond Graph projects 3-D bar graphs into 2-D whereby the third dimension is replaced with a polygon, whose area and middle vertical length represent the outcome. In Section 2, we describe the design and methods of Diamond Graph. Three examples of using Diamond Graphs to display different data types (i.e., proportions, incidence rates, and relative risks) with S-plus programs are provided in Section 3, followed by a discussion in Section 4.

The original paper of Diamond Graph published in *The American Statistician* (Li,

1 Buechner, Tarwater, and Muñoz, 2003) and the S-Plus programs for plotting Diamond Graphs in
2 the paper can be accessed at <http://www.statepi.jhsph.edu/~xili/diamondgraph.html>.

3 4 **2. Design and Methods**

5 In Diamond Graph, we replace the parallelepiped volume in 3-D bar graphs with a
6 polygon in each cell of the 2-D grid defined by categories of the two predictors. Because the
7 primary purpose of Diamond Graph is to provide *equal* representation of the effects of two
8 predictors on a continuous outcome in two dimensions, we first use square cells rotated 45°
9 clockwise to construct the grid resembling a diamond to achieve equal representation for two
10 predictors (Figure 1). Once the square cells are established, the volumes of the parallelepipeds are
11 compressed into the areas of a polygon within each cell (i.e., three dimensions are projected to
12 two dimensions). In addition, because it is easier to perceive lengths than areas (Cleveland and
13 McGill, 1984), we use a polygon whose middle vertical length also represent the outcome.

14 To facilitate the description of how to construct the Diamond Graph, we use the case
15 where the outcome is proportional data. We develop a function for drawing the square cells of
16 the grid and the shaded areas defining the polygons. In our function, p represents the proportion
17 describing the outcome in a given cell and ℓ is the length of the diagonal for each square cell. The
18 shaded polygon located at the center of each square cell in Figure 1 is constructed based on three
19 attributes: (i) the middle vertical length of the polygons ($V(p) = p\ell$); (ii) the middle horizontal
20 length of the polygons ($H(p) = (0.5 + 0.5p)\ell$); and (iii) the horizontal length of the top and
21 bottom of the polygons ($h(p) = (0.5 - 0.5p)\ell$). The areas of the shaded polygons are given by
22 $[H(p) + h(p)] V(p)/2 = p\ell^2/2$, which is exactly p of the full area of the square cell given by $\ell^2/2$. We
23 select this type of polygon to represent the outcome because all its three attributes have linear
24 relationship with the outcome p , and the shaded polygons have constant shape (Li, Buechner,

1 Tarwater, and Muñoz, 2003). We also indicate the numeric value of outcome as a label at the
2 center of shaded polygons to further facilitate the graphical display. Figure 1 shows
3 representations of proportional data ($p = 0.10, 0.25, 0.50, \text{ and } 0.75$) as shaded areas that indicate
4 the percentages (10, 25, 50, and 75) of the shaded area of each square cell.

5 An equivalent and simpler description for the construction of the polygon is as follows:
6 starting at the center of the square cell, identify the two points on the vertical diagonal away from
7 the center where the distance between each point and the center is $p\%$ of the distance between the
8 top or bottom vertex and the center. At the levels of the above-mentioned two points, the length of
9 the top and bottom horizontal lines of the polygon simply equals to half of the full length of the
10 horizontal lines at the same levels from side to side of the square cell. After locating the top and
11 bottom lines of the polygon, from the four end-points draw four lines parallel to the sides of
12 square cell to construct the polygon.

13 To plot outcomes ranging from 0 to any positive value (e.g., incidence rates, relative risks,
14 odds ratios), we first transform the outcomes into proportional values by dividing each value of
15 the outcome by the highest one. These derived proportions are used to plot polygons that
16 represent the relative magnitude of outcomes. Instead of derived proportions, the exact numeric
17 values of the outcome are used as the labels at the center of the polygons to conserve and convey
18 the information inherent in the original scale of the outcome.

19
20
21

3. Examples

3.1 Graph of Proportions

22 In this example, we create Figure 2 (a reprint of Figure 3 in Li, Buechner, Tarwater, and
23 Muñoz, 2003) to depict the likelihood of developing AIDS in six years according to the amount of
24 HIV in the Blood and degree of immune deficiency (Mellors, Muñoz, Giorgi, et al. 1997). The
25

1 data presented in Figure 2 is as following:

Likelihood of developing AIDS in 6 years			Degree of Immune Deficiency*				
			Minor	Mild	Moderate	Severe	Very severe
Amount of HIV in blood (copies per ml)	Very high	(HIV >30,000)	67%	67%	78%	89%	96%
	High	(10,001-30,000)	37%	55%	55%	73%	73%
	Medium	(3,001 - 10,000)	14%	37%	37%	37%	NA‡
	Low	(501 – 3,000)	17%	17%	17%	NA	NA
	Very low	(<500)	2%	10%	NA	NA	NA

2 * Measured by CD4 cell count; minor: CD4 cell count >750, mild: 501 – 750, moderate: 351 – 500, severe: 201 –
 3 350, very severe: <=200 cells/mm³. ‡NA: No subjects in this category because low viral load does not cause severe
 4 immune deficiency.

6 The S-Plus program for implementing Figure 2 is shown as below:

```

7 #-----
8 # 1. Obtain the number of categories of the two predictors;
9 # row = number of categories of the predictor on the southwest side of Diamond Graph;
10 # col = number of categories of the predictor on the southeast side of Diamond Graph;
11 #-----
12 row <- 5
13 col <- 5
14
15 #-----
16 # 2. Select a square region for plot;
17 #-----
18 par(pty="s")
19
20 #-----
21 # 3. Assign the coordinate of the lowest point of Diamond Graph to (0,0), and define
22 # the ranges for x and y axes based on the number of categories of two predictors.
23 #-----
24 plot(0,0, xlim=c(0-row, col), ylim=c(0,row+col),axes=F, type="n", xlab="", ylab="")
25
26 #-----
27 # 4. Obtain the coordinate of the center of each square cell (cx,cy) with diagonal= 2;
28 # i.e., the distance from the center to a vertex of a square cell equals to 1.
29 #-----
30 for (i in 1:row)
31 {
32     if (i == 1){
33         cx <- seq(0,(col-1) )
34         cy <- seq(1,col)
35     }
36     else{
37         cx <- c(cx, seq( (1-i),(col-i) ) )
38         cy <- c(cy, seq(i,(col+i-1)) )
39     }
40 }
41
42 #-----
43 # 5. Read in the outcomes, which start from the value in the cell 1 (at bottom)
44 # to cell 2, to cell 3, and finally to cell 4 (at top), as depicted below in
45 # the 2x2 graph.
46 # /4\ <- top
47 # / \/\
48 # \3/\2/
49 # \1/ <- bottom
50 #
51 # Empty cells (with no subjects) are assigned -1 and will not be displayed.
52 # Only positive outcomes will be plotted as polygons in square cells.

```

```

1 #-----
2 outcome <- c(.02,.1, -1, -1, -1,
3             .17,.17,.17, -1, -1,
4             .14,.37,.37,.37, -1,
5             .37,.55,.55,.73,.73,
6             .67,.67,.78,.89,.98)
7
8 #-----
9 # 6. If data are not proportional, convert the outcomes to proportions -
10 #   each value is divided by the highest one within that outcome.
11 #   If data are proportional, no need to convert, which is the case for Figure 2.
12 #-----
13 p <- outcome
14
15 #-----
16 # 7. Calculate the 'scale' based on the number of categories of two predictors;
17 #   We have found scaling all proportions (p) by 'scale' useful in the
18 #   graphical illustration when p is high.
19 #-----
20 scale <- 1.005 - 0.005*(row+col)
21 p[p<1] <- p[p<1] *scale
22 p[p==1] <- p[p==1]*scale+0.005
23
24 #-----
25 # 8. Draw the outline of the square cells.
26 # 1) draw 45 degree lines from southwest to northeast
27 # 2) draw -45 degree lines from southeast to northwest
28 #-----
29 for (i in 0:row) lines(c(-i,col-i),c(i,col+i),lwd=0.01)
30 for (i in 0:col) lines(c(i, -row+i), c(i, row+i),lwd=0.01)
31
32 #-----
33 # 9. Calculate the coordinates of the six vertices of inside polygons;
34 #   start clockwise from northwest vertex as (x1, y1).
35 #
36 #   x1,y1 _____ x2,y2
37 #   /               \
38 #  x6,y6 /           \ x3,y3
39 #   \               /
40 #   \               /
41 #   x5,y5 _____ x4,y4
42 #-----
43 x1<- cx - (1-p)/2
44 x2<- cx + (1-p)/2
45 x3<- cx + (1+p)/2
46 x4<- cx + (1-p)/2
47 x5<- cx - (1-p)/2
48 x6<- cx - (1+p)/2
49 y1<- cy + p
50 y2<- cy + p
51 y3<- cy
52 y4<- cy - p
53 y5<- cy - p
54 y6<- cy
55
56 #-----
57 # 10. Plot a polygon inside each square cell, and then print the value of the outcome
58 #     at the center in each square cell.
59 #
60 #     Here we use cex= 0.7*scale for the labels. Other applications may
61 #     need to use a different factor so that labels can be read easily.
62 #-----
63 n <- row*col
64 for (i in 1:n) {

```

```

1  if (p[i]>0)
2    polygon(c(x1[i],x2[i],x3[i],x4[i],x5[i],x6[i]),
3            c(y1[i],y2[i],y3[i],y4[i],y5[i],y6[i])), density=-.1, border=F, col=5)
4
5  if (outcome[i]>=0){
6    # In Windows system, S-plus needs 0.03 to be added to cy so that lable is properly
7    centered
8    text(cx[i], cy[i]+0.03, outcome[i]*100, cex=0.7*scale, lwd=2)
9  }
10 }
11
12 #-----
13 # 11. Below are the commands to print labels for two predictors.
14 #     It requires trial and error for the proper location;
15 #     but users can skip the following steps and add the labels in PowerPoint.
16 #-----
17
18 # print labels on the southwest side of Diamond Graph
19 text(c(-5.4,-5.4), c(1.8,1.8), adj=0, cex=.9,"Amount of HIV")
20 text(c(-5.0,-5.0), c(1.5,1.5), adj=0, cex=.9,"in Blood")
21 text(c(-0.7,-0.7), c(.35,.35), adj=1, cex=.9,"very low")
22 text(c(-1.7,-1.7), c(1.35,1.35), adj=1, cex=.9,"low")
23 text(c(-2.7,-2.7), c(2.35,2.35), adj=1, cex=.9,"medium")
24 text(c(-3.7,-3.7), c(3.35,3.35), adj=1, cex=.9,"high")
25 text(c(-4.7, -4.7), c(4.35,4.35), adj=1, cex=.9,"very high")
26
27 # print labels on the southeast side of Diamond Graph
28 text(c(4.0, 4.0), c(1.8,1.8), adj=0, cex=.9,"Degree of Immune")
29 text(c(4.5, 4.5), c(1.5,1.5), adj=0, cex=.9,"Deficiency")
30 text(c(0.7, 0.7), c(.35,.35), adj=0, cex=.9,"minor")
31 text(c(1.7, 1.7), c(1.35,1.35), adj=0, cex=.9,"mild")
32 text(c(2.7, 2.7), c(2.35,2.35), adj=0, cex=.9,"moderate")
33 text(c(3.7, 3.7), c(3.35,3.35), adj=0, cex=.9,"severe")
34 text(c(4.7, 4.7), c(4.35,4.35), adj=0, cex=.9,"very severe")
35
36

```

37 3.2 Graph of Incidence Rates

38 This example of S-Plus program creates Figure 3 (a reprint of Figure 5 in Li, Buechner,
39 Tarwater, and Muñoz, 2003), which depicts the age-adjusted incidence rates of end-stage renal
40 disease due to any cause per 100,000 person-years according to six categories of both systolic and
41 diastolic blood pressure (Klag, Whelton, Randall, et al. 1996). The incidence rates to be presented
42 in Figure 3 are as following:

Incidence rates of ESRD* /100,000 person-years	Diastolic Blood Pressure (mm Hg)					
	<80	80 - 84	85 - 89	90 - 99	100 - 109	>= 110
Systolic						
>= 180	49.6	95.9	211.7	143.5	124.4	205.6
Blood						
160 - 179	90.2	68.4	57.8	62.3	49.7	88.9
Pressure						
140 - 159	22.1	27.5	22.6	27.1	35.4	62.3
(mm Hg)						
130 - 139	13.3	12.8	12.4	16.4	25.4	NA
120 - 129	6.8	6.7	8.4	11.1	40.2	NA
<120	5.3	6.0	4.9	9.8	NA	NA

43 *ESRD: end-stage renal disease

1 Below is the S-Plus program for implementing Figure 3. The program is very similar to
 2 that for Figure 2. Here, we have highlighted the differences in bold font.

```

3 #-----
4 # 1. Obtain the number of categories of the two predictors;
5 # row = number of categories of the predictor on the southwest side of Diamond Graph;
6 # col = number of categories of the predictor on the southeast side of Diamond Graph;
7 #-----
8 row <- 6
9 col <- 6
10
11 #-----
12 # 2. Select a square region for plot;
13 #-----
14 par(pty="s")
15
16 #-----
17 # 3. Assign the coordinate of the lowest point of Diamond Graph to (0,0), and define
18 # the ranges for x and y axes based on the number of categories of two predictors.
19 #-----
20 plot(0,0, xlim=c(0-row, col), ylim=c(0,row+col),axes=F, type="n", xlab="", ylab="")
21
22 #-----
23 # 4. Obtain the coordinate of the center of each square cell (cx,cy) with diagonal= 2;
24 # i.e., the distance from the center to a vertex of a square cell equals to 1.
25 #-----
26 for (i in 1:row)
27 {
28     if (i == 1){
29         cx <- seq(0,(col-1) )
30         cy <- seq(1,col)
31     }
32     else{
33         cx <- c(cx, seq( (1-i),(col-i) ) )
34         cy <- c(cy, seq(i,(col+i-1)) )
35     }
36 }
37
38 #-----
39 # 5. Read in the outcomes, which start from the value in the cell 1 (at bottom)
40 # to cell 2, to cell 3, and finally to cell 4 (at top), as depicted below in
41 # the 2x2 graph.
42 # /4\ <- top
43 # / \/\
44 # \3/\2/
45 # \1/ <- bottom
46 #
47 # Empty cells (with no subjects) are assigned -1 and will not be displayed.
48 # Only positive outcomes will be plotted as polygons in square cells.
49 #-----
50 outcome <- c( 5.3, 6.0, 4.9, 9.8, -1, -1,
51              6.8, 6.7, 8.4,11.1,40.2,-1,
52              13.3,12.8,12.4,16.4,25.4,-1,
53              22.1,27.5,22.6,27.1,35.4,62.3,
54              90.2,68.4,57.8,62.3,49.7,88.9,
55              49.6,95.9,211.7,143.5,124.4,205.6)
56
57 #-----
58 # 6. If data are not proportional, convert the outcomes to proportions -
59 # each value is divided by the highest one within that outcome.
60 # If data are proportional, no need to convert.
61 #-----

```

```

1  maxoutcome <- max(outcome)
2  p <- outcome/maxoutcome
3
4  #-----
5  # 7. Calculate the 'scale' based on the number of categories of two predictors;
6  #   We have found scaling all proportions (p) by 'scale' useful in the
7  #   graphical illustration when p is high.
8  #-----
9  scale <- 1.005 - 0.005*(row+col)
10 p[p<1] <- p[p<1] *scale
11 p[p==1] <- p[p==1]*scale+0.005
12
13 #-----
14 # 8. Draw the outline of the square cells.
15 # 1) draw 45 degree lines from southwest to northeast
16 # 2) draw -45 degree lines from southeast to northwest
17 #-----
18 for (i in 0:row) lines(c(-i,col-i),c(i,col+i),lwd=0.01)
19 for (i in 0:col) lines(c(i, -row+i), c(i, row+i),lwd=0.01)
20
21 #-----
22 # 9. Calculate the coordinates of the six vertices of inside polygons;
23 #   start clockwise from northwest vertex as (x1, y1).
24 #
25 #   x1,y1  _____  x2,y2
26 #         / \
27 #   x6,y6 /   \ x3,y3
28 #         \   /
29 #         \___/
30 #   x5,y5  _____  x4,y4
31 #-----
32 x1<- cx - (1-p)/2
33 x2<- cx + (1-p)/2
34 x3<- cx + (1+p)/2
35 x4<- cx + (1-p)/2
36 x5<- cx - (1-p)/2
37 x6<- cx - (1+p)/2
38 y1<- cy + p
39 y2<- cy + p
40 y3<- cy
41 y4<- cy - p
42 y5<- cy - p
43 y6<- cy
44
45 #-----
46 # 10. Plot a polygon inside each square cell, and then print the value of the outcome
47 #     at the center in each square cell.
48 #
49 #     Here we use cex= 0.7*scale for the labels. Other applications may
50 #     need to use a different factor so that labels can be read easily.
51 #-----
52 n <- row*col
53 for (i in 1:n) {
54   if (p[i]>0)
55     polygon(c(x1[i],x2[i],x3[i],x4[i],x5[i],x6[i]),
56             c(y1[i],y2[i],y3[i],y4[i],y5[i],y6[i])), density=-.1, border=F, col=5)
57
58   # make the outcome=6.0 shown as 6.0 not as 6;
59   outcome[outcome==6.0] <- '6.0'
60   if (outcome[i]>=0){
61     # In Windows system, S-plus needs 0.03 to be added to cy so that lable is
62     # properly centered
63     text(cx[i], cy[i]+0.03, outcome[i], cex=0.7*scale, lwd=2)
64   }

```

```

1  }
2
3  #-----
4  # 11. Below are the commands to print labels for two predictors.
5  #     It requires trial and error for the proper location;
6  #     but users can skip the following steps and add the labels in PowerPoint.
7  #-----
8
9  # print labels on the southwest side of Diamond Graph
10 text(c(-5.5,-5.5), c(1.8,1.8), adj=0.5, cex=1.0,"Systolic (mm Hg)")
11 text(c(-0.7,-0.7), c(.35,.35), adj=1, cex=.8,"<120")
12 text(c(-1.7,-1.7), c(1.35,1.35), adj=1, cex=.8,"120-129")
13 text(c(-2.7,-2.7), c(2.35,2.35), adj=1, cex=.8,"130-139")
14 text(c(-3.7,-3.7), c(3.35,3.35), adj=1, cex=.8,"140-159")
15 text(c(-4.7, -4.7), c(4.35,4.35), adj=1, cex=.8,"160-179")
16 text(c(-5.7, -5.7), c(5.35,5.35), adj=1, cex=.8,">=180")
17
18 # print labels on the southeast side of Diamond Graph
19 text(c(5.5, 5.5), c(1.8,1.8), adj=0.5, cex=1.0,"Diastolic (mm Hg)")
20 text(c(.7, .7), c(.35,.35), adj=0, cex=.8,"<80")
21 text(c(1.7, 1.7), c(1.35,1.35), adj=0, cex=.8,"80-84")
22 text(c(2.7, 2.7), c(2.35,2.35), adj=0, cex=.8,"85-89")
23 text(c(3.7, 3.7), c(3.35,3.35), adj=0, cex=.8,"90-99")
24 text(c(4.7, 4.7), c(4.35,4.35), adj=0, cex=.8,"100-109")
25 text(c(5.7, 5.7), c(5.35,5.35), adj=0, cex=.8,">=110")
26
27

```

28 3.3 Graph of Relative Risk and Excess Risk

29 In this example, we apply Diamond Graph to represent relative risk and excess risk in
30 Figure 4 (a reprint of Figure 7 in Li, Buechner, Tarwater, and Muñoz, 2003) using data of relative
31 risks of breast cancer by adult weight change and hormone use among postmenopausal women
32 (Huang, Hankinson, Colditz, et al. 1997). The left panel of Figure 4 depicts relative risks, which
33 were converted to proportions first by dividing each relative risk (rr_i) by the highest rr . The right
34 panel of Figure 4 depicts excess risks, which could be represented by polygons based on the
35 proportions $p_i = (rr_i - \text{lowest } rr) / (\text{highest } rr - \text{lowest } rr)$. Compared to the left panel showing
36 relative risk, the right panel exhibiting excess risk naturally highlights the reference category
37 because no polygon is drawn for the cell with the lowest relative risk since its $p_i = 0$. However,
38 this approach may sometimes give an exaggerated impression of the cell-to-cell differences
39 because the ratio of the areas of the two polygons is not equal to the ratio of the two
40 corresponding relative risks.

1 The data used to plot Figure 4 are listed below:

Relative risk of breast cancer among postmenopausal women		Hormone Use		
		Never	Past	Current
Adult	Loss >2.0	1.1	1.1	1.9
Weight	Loss or Gain 2.0	1	1.3	1.8
Change	Gain 2.1-10.0	1.2	1.5	2.1
(kg)	Gain 10.1-20.0	1.6	1.3	1.7
	Gain >20.0	2	1.3	1.7

2

3 The corresponding S-Plus program for plotting Figure 4 is as the following. The program
4 is very similar to those for Figures 2 and 3. Here, we have highlighted the differences in bold font.

```
5 #-----  
6 # 1. Obtain the number of categories of the two predictors;  
7 # row = number of categories of the predictor on the southwest side of Diamond Graph;  
8 # col = number of categories of the predictor on the southeast side of Diamond Graph;  
9 #-----  
10 row <- 5  
11 col <- 3  
12  
13 #-----  
14 # 2. Select a square region for plot, and put 2 plots on one page;  
15 #-----  
16 par(pty="s",mfrow=c(1,2))  
17  
18  
19 plotfigure4 <- function(panel) # Create a function to plot a panel of figure 4;  
20 {  
21 #-----  
22 # 3. Assign the coordinate of the lowest point of Diamond Graph to (0,0), and define  
23 # the ranges for x and y axes based on the number of categories of two predictors.  
24 #-----  
25 plot(0,0, xlim=c(0-row, col), ylim=c(0,row+col),axes=F, type="n", xlab="", ylab="")  
26  
27 #-----  
28 # 4. Obtain the coordinate of the center of each square cell (cx,cy) with diagonal= 2;  
29 # i.e., the distance from the center to a vertex of a square cell equals to 1.  
30 #-----  
31 for (i in 1:row)  
32 {  
33     if (i == 1){  
34         cx <- seq(0,(col-1) )  
35         cy <- seq(1,col)  
36     }  
37     else{  
38         cx <- c(cx, seq( (1-i),(col-i) ) )  
39         cy <- c(cy, seq(i,(col+i-1)) )  
40     }  
41 }  
42  
43 #-----  
44 # 5. Read in the outcomes, which start from the value in the cell 1 (at bottom)  
45 # to cell 2, to cell 3, and finally to cell 4 (at top), as depicted below in  
46 # the 2x2 graph.  
47 # /4\ <- top  
48 # / \\  
49 # \3/\2/
```

```

1 # \1/ <- bottom
2 #
3 # Empty cells (with no subjects) are assigned -1 and will not be displayed.
4 # Only positive outcomes will be plotted as polygons in square cells.
5 #-----
6 outcome <- c(2.0, 1.3, 1.7,
7             1.6, 1.3, 1.7,
8             1.2, 1.5, 2.1,
9             1.0, 1.3, 1.8,
10            1.1, 1.1, 1.9)
11
12 #-----
13 # 6.If data are not proportional, convert the outcomes to proportions;
14 # to get relative risks,each value is divided by the highest one within that outcome;
15 # to get excess risks, the difference between each value and the minimum outcome is
16 # divided by the difference between the maximum and the minimum outcome.
17 # If data are proportional, no need to convert.
18 #-----
19 maxoutcome <- max(outcome)
20 minoutcome <- min(outcome[outcome>=0])
21
22 if (panel=="Relative Risk") p<-outcome/maxoutcome #get p for left panel(relative
23 risk)
24 else p<-(outcome-minoutcome)/(maxoutcome-minoutcome)#get p for right panel(excess risk)
25
26 #-----
27 # 7. Calculate the 'scale' based on the number of categories of two predictors;
28 # We have found scaling all proportions (p) by 'scale' useful in the
29 # graphical illustration when p is high.
30 # If a page has only one graph, we use scale = 1.005 - 0.005(row+col).
31 # But figure 4 has 2 plots on one page, the actual square cell size is similar to
32 # that of a 5x6 plot. So we use scale = 1.005-0.005*(5+6) here.
33 #-----
34 scale <- 1.005 - 0.005*(5+6)
35 p[p<1] <- p[p<1] *scale
36 p[p==1] <- p[p==1]*scale+0.005
37
38 #-----
39 # 8. Draw the outline of the square cells.
40 # 1) draw 45 degree lines from southwest to northeast
41 # 2) draw -45 degree lines from southeast to northwest
42 #-----
43 for (i in 0:row) lines(c(-i,col-i),c(i,col+i),lwd=0.01)
44 for (i in 0:col) lines(c(i, -row+i), c(i, row+i),lwd=0.01)
45
46 #-----
47 # 9. Calculate the coordinates of the six vertices of inside polygons;
48 # start clockwise from northwest vertex as (x1, y1).
49 #
50 # x1,y1 _____ x2,y2
51 # / \
52 # x6,y6 / \ x3,y3
53 # \ /
54 # x5,y5 \ / x4,y4
55 #-----
56
57 x1<- cx - (1-p)/2
58 x2<- cx + (1-p)/2
59 x3<- cx + (1+p)/2
60 x4<- cx + (1-p)/2
61 x5<- cx - (1-p)/2
62 x6<- cx - (1+p)/2
63 y1<- cy + p
64 y2<- cy + p

```

```

1  y3<- cy
2  y4<- cy - p
3  y5<- cy - p
4  y6<- cy
5
6  #-----
7  # 10. Plot a polygon inside each square cell, and then print the value of the outcome
8  #    at the center in each square cell.
9  #
10 #    Here we use cex= 0.7*scale for the labels. Other applications may
11 #    need to use a different factor so that labels can be read easily.
12 #-----
13 n <- row*col
14 for (i in 1:n) {
15   if (p[i]>0)
16     polygon(c(x1[i],x2[i],x3[i],x4[i],x5[i],x6[i]),
17              c(y1[i],y2[i],y3[i],y4[i],y5[i],y6[i])), density=-.1, border=F, col=5)
18
19   # make the outcome=2.0 shown as 2.0 not as 2;
20   outcome[outcome==2.0] <- '2.0'
21   if (outcome[i]>=0){
22     # In Window system, S-plus needs 0.03 to be added to cy so that lable is properly
23     # centered
24     text(cx[i], cy[i]+0.03, outcome[i], cex=0.7*scale, lwd=2)
25   }
26 }
27
28 #-----
29 # 11. Below are the commands to print labels for two predictors.
30 #    It requires trial and error for the proper location;
31 #    but users can skip the following steps and add the labels in PowerPoint.
32 #-----
33
34 # print labels on the southwest side of Diamond Graph
35 text(c(-4.5,-4.5), c(1,1), adj=0.5, cex=0.85,"Adult")
36 text(c(-4.5,-4.5), c(0.7,0.7), adj=0.5, cex=0.85,"Weight")
37 text(c(-4.5,-4.5), c(0.4,0.4), adj=0.5, cex=0.85,"Change, kg")
38 text(c(-0.7,-0.7), c(.35,.35), adj=1, cex=.7,"Gain >20.0")
39 text(c(-1.7,-1.7), c(1.35,1.35), adj=1, cex=.7,"Gain 10.1-20.0")
40 text(c(-2.7,-2.7), c(2.35,2.35), adj=1, cex=.7,"Gain 2.1-10.0")
41 text(c(-3.7,-3.7), c(3.35,3.35), adj=1, cex=.7,"Loss or Gain 2.0")
42 text(c(-4.7,-4.7), c(4.35,4.35), adj=1, cex=.7,"Loss >2.0")
43
44 # print labels on the southeast side of Diamond Graph
45 text(c(3, 3), c(0.7,0.7), adj=0.5, cex=0.85,"Hormone")
46 text(c(3, 3), c(0.4,0.4), adj=0.5, cex=0.85,"Use")
47 text(c(.7, .7), c(.35,.35), adj=0, cex=.7,"Never")
48 text(c(1.7, 1.7), c(1.35,1.35), adj=0, cex=.7,"Past")
49 text(c(2.7, 2.7), c(2.35,2.35), adj=0, cex=.7,"Current")
50
51 title(main=panel)
52
53 } # end of function plotfigure4;
54
55 # Call the function to plot relative risk in left panel and excess risk in right panel
56 plotfigure4("Relative Risk")
57 plotfigure4("Excess Risk")

```

4. Discussion

Diamond Graph was introduced in 2003 to equally present the effects of two categorical predictors on a continuous outcome. It provides a succinct equiponderant 2-D graphical display that projects the outcome into a polygon whose area and middle vertical length represent the outcome, thus circumventing certain limitations of 3-D bar graphs and other 2-D alternatives.

Diamond Graph is suitable for depicting outcomes represented by proportions (example 3.1), as well as any outcomes represented by positive values (example 3.2 and 3.3), which are converted into proportions in the S-Plus programs by dividing each value of the outcome by the maximum one in the data set or by a predetermined maximum value. The derived proportional values are used to construct the polygons for non-proportional data, but the values of the original outcome are presented as labels in the centers of polygons to further facilitate the graphical display. Another advantage of displaying the numerical value at the center of each polygon is that it serves to distinguish a cell with an outcome of zero (represented by 0 at the center and no polygon) from a cell with no individuals (represented by a blank cell).

We developed an S-Plus program to plot the Diamond Graph. The algorithm of drawing the Diamond Graph was well documented in the S-Plus programs in Section 3 (Examples). Users can easily adapt the S-Plus programs provided in Section 3 according to their own needs. There are two major changes that users need to make - the number of categories of each of the two predictors (step 1) and values of the outcome (step 5). The last step of labeling two predictors needs to be modified if users prefer to show the labels in S-Plus plot than in PowerPoint. In addition, if there are two or more plots on one page, users may want to modify the code for calculating the 'scale', as demonstrated in step 7 of the S-Plus program in example 3.3.

A problem that we encountered when developing the Diamond Graph using S-Plus was that the four sides of the square cell occupy part of the area of the square cell so that the visible

1 area of the square cell was actually slightly less than 1. To solve this problem, we first scaled all
2 proportions when plotting the polygons, and then used thin lines (lwd=0.01 in step 8) to draw the
3 outlines of square cells. In doing so, the graphically display is enhanced, especially when p is
4 high.

5 For preparation of slides in Power Point, another issue arose when we exported Diamond
6 Graphs from S-plus to PowerPoint using the option “File” → “Create PowerPoint Presentation” in
7 S-Plus, which distorted Diamond Graphs in PowerPoint. It is because this function does not
8 preserve the square base of the Diamond Graph, which is an essential feature of the Diamond
9 graph. We recommended users use the following ways to export diamond graphs from S-Plus to
10 PowerPoint: if the S-Plus graph is saved in a file, use “Insert” “Object” option in PowerPoint to
11 insert the saved S-Plus graph file by checking “Create from file”; otherwise, copy the S-Plus
12 graph sheet, and paste it as a picture in PowerPoint. To adjust the size of Diamond Graph via
13 “Format Picture” function in PowerPoint, users should remember to check the boxes of “Lock
14 aspect ratio” and “Relative to original picture size” under the tab “size” so that the exported
15 Diamond Graphs can remain undistorted. We recommend that “height” be changed to 7” or 8”
16 depending on the room desired for a title.

17 To prepare Figures for publication purposes, if the accepted format is TIF the following
18 steps should be done: in S-plus use “File → Print graph sheet”, select printer name as “Microsoft
19 Office Document Image Writer” and click OK and save, which automatically opens Microsoft
20 Office Document Imaging and the file should be saved as .tif. If the desired format is EPS, users
21 can export S-plus graphs to “Encapsulated Post Script” format by using “File → Export Graph”
22 menu option in S-plus.

23 Diamond Graph is not an all-purpose graph. When the two predictors are continuous, the
24 contour plots can be used to present the effects of the two predictors on the outcome. On the other

1 hand, when two predictors are not equally important (e.g., exposure and confounder), Trellis
2 graphs are preferred because the exposure is depicted as the panel variable and the confounder is
3 depicted as the conditioning variable.

4 Although Diamond Graph was originally designed to display 3-D data, it can be applied to
5 depict four-dimensional data by plotting a sequence of plots. For example, to depict the likelihood
6 for developing AIDS at three, six, and nine years (Figure 2 shows data only at six years), we can
7 easily put a sequence of three graphs in a single page, which reveal how the outcomes evolve over
8 time. Creative and expansive use of Diamond Graph will improve and refine the function of
9 graphical methods in summarizing and representing the scientific information that the data
10 possess.

1 **REFERENCES**

2 Becker, R. A., Cleveland, W. S., and Shyu, M. J. (1996), "The Visual Design and Control of
3 Trellis Display," *Journal of Computational and Statistical Graphics*, 5, 123-155.

4 Cleveland, W. S., and McGill R. (1984), "Graphic Perception: Theory, Experimentation, and
5 Application to the Development of Graphical Methods," *Journal of American Statistical*
6 *Association*,79, 531-554.

7 Farahmand, B. Y., Michaëlsson, K., Baron, J. A., Persson, P. G., and Ljunghall, S. (2000), "Body
8 Size and Hip Fracture Risk.," *Epidemiology*, 11,214-219.

9 Friendly, M. (1994), "Mosaic displays for multi-way contingency tables," *Journal of American*
10 *Statistical Association*, 89, 190-200.

11 Harris, R. L. (1999), *Information Graphics: A Comprehensive Illustrated Reference*. New York,
12 NY: Oxford University Press.

13 Hartigan, J. A., and Kleiner, B. (1981), "Mosaics for Contingency Tables," in *Computer Science*
14 *and Statistics: Proceedings of the 13th Symposium on the Interface*. New York , NY: Springer
15 Verlag.

16 Hartigan, J. A., and Kleiner, B. (1984), "A Mosaics of Television ratings," *American Statistician*,
17 38, 32-35

18 Huang, Z., Hankinson, S. E., Colditz, G. A., Stamfer, M. J., Hunter, D. J., Manson, J. E.,
19 Hennekens, C. H., Rosner, B., Speizer, F. E., and Willett, W. C. (1997), "Dual Effects of
20 Weight and Weight Gain on Breast Cancer Risk," *JAMA*, 278,1407-1411.

21 Klag, M. J., Whelton, P. K., Randall, B. L., Neaton, J. D., Brancati, F. L., Ford, C. E.,
22 Shulman, N. B., and Stamler, J. (1996), "Blood Pressure and End-Stage Renal Disease in
23 Men," *New England Journal of Medicine*, 334,13-18.

1 Li, X., Buechner, J. M., Tarwater, P. M., Muñoz, A. (2003). “A Diamond-shaped Equiponderant
2 Graphical Display of the Effects of Two Categorical Predictors on Continuous Outcomes.”
3 *The American Statistician*, 57 (3), 193-199.

4 Mellors, J. W., Muñoz, A., Giorgi, J. V., Margolick, J. B., Tassoni, C. J., Gupta, P.,
5 Kingsley, L.A., Todd, J. A., Saah, A. J., Detels, R., Phair, J. P., and Rinaldo, C. R. (1997),
6 “Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection,”
7 *Annals of Internal Medicine*, 126:12, 946-954.

8 Tufte, E. R. (1983), *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics
9 Press.

10 Wilkinson, L. (1999), *The Grammar of Graphics*, New York, NY: Springer-Verlag.
11

FIGURE LEGEND

Figure 1. Representations of proportional data ($p = 0.10, 0.25, 0.50, \text{ and } 0.75$) as shaded areas that indicate the percentages (10, 25, 50, and 75) of the area of each square cell

Figure 2. Diamond Graph for depicting the likelihood of developing AIDS in six years according to the amount of HIV in the blood and the degree of immune deficiency. Cells with no individuals are blank. Reprinted with permission from *The American Statistician*. Copyright 2003 by the American Statistical Association. All rights reserved.

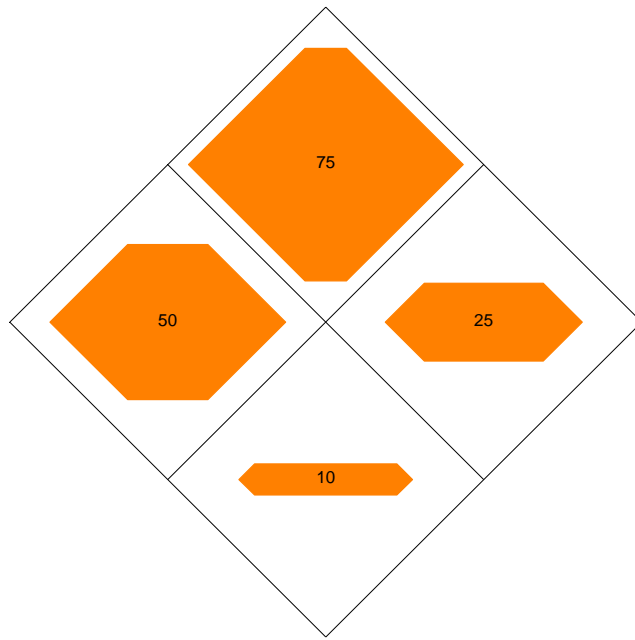
Figure 3. Diamond Graph for depicting the age-adjusted rates of end-stage renal disease due to any cause per 100,000 person-years according to systolic and diastolic blood pressures. Each shaded area is a proportion relative to the largest incidence rate with 211.7 being 1 (full shading). The exact numeric values of the rates are used as the labels. Cells with no individuals are blank. Reprinted with permission from *The American Statistician*. Copyright 2003 by the American Statistical Association. All rights reserved.

Figure 4. Diamond Graph for depicting relative risk of breast cancer by adult weight change and hormone use among postmenopausal women reported by Huang *et al.* (1997). The graph on the left panel depicts the relative risks (i.e., $p_i = rr_i / \text{highest } rr$) and the one on the right panel depicts the excess risk (i.e., $p_i = (rr_i - \text{lowest } rr) / (\text{highest } rr - \text{lowest } rr)$). Reprinted with permission from *The American Statistician*. Copyright 2003 by the American Statistical Association. All rights reserved.

1

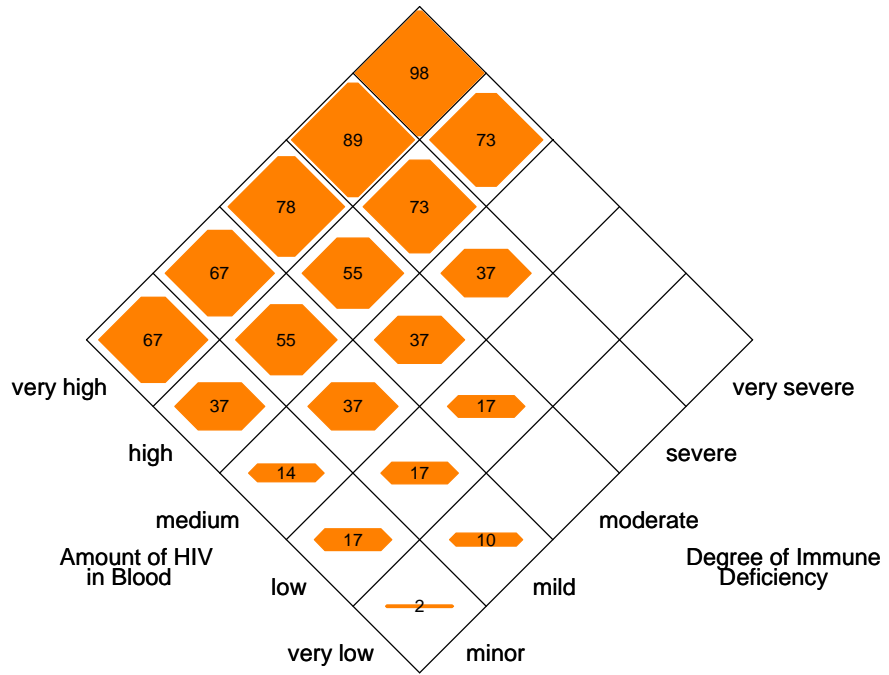
2

3 Figure



4 1

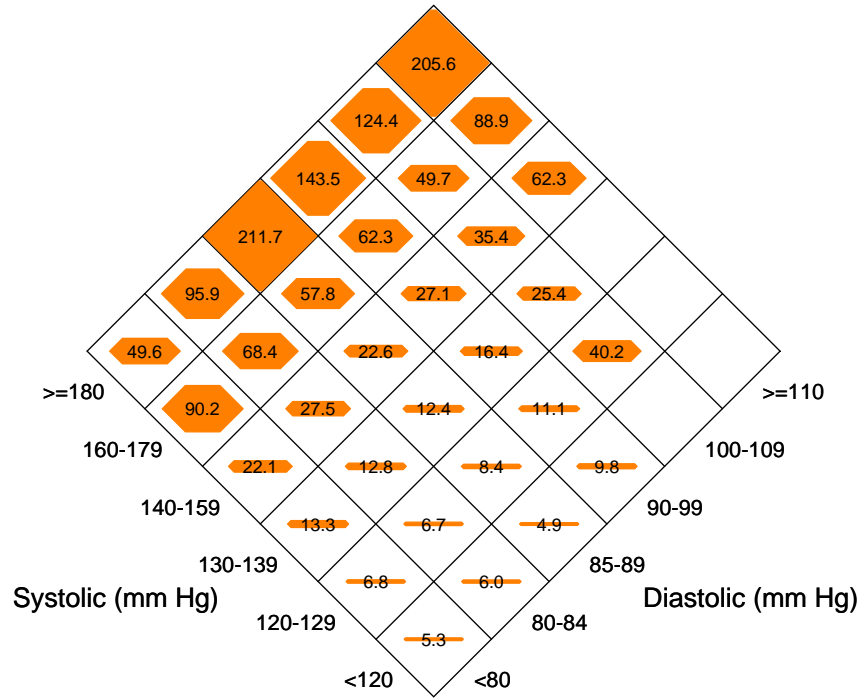
1



2

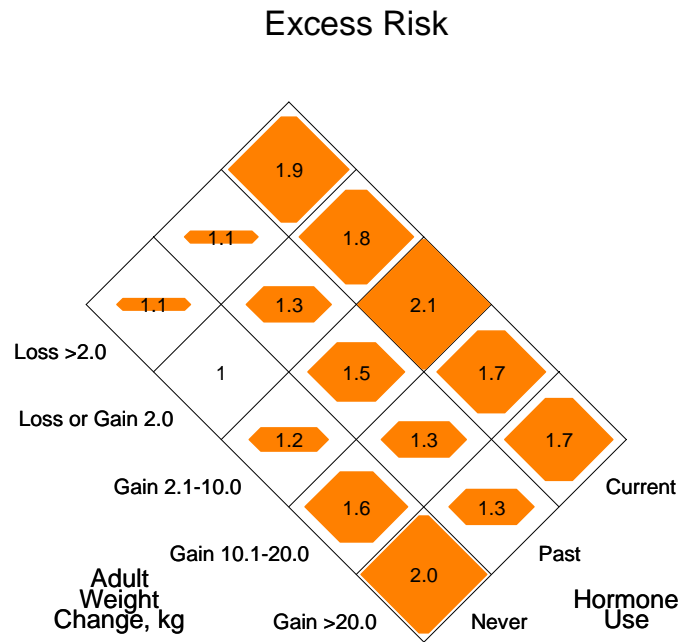
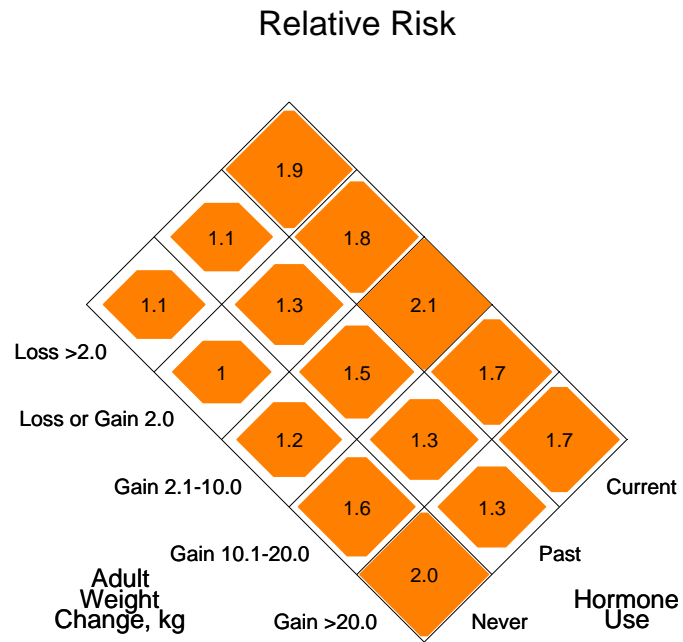
3 Figure 2

1



2

3 Figure 3



1

2 Figure 4

1 **Affiliation:**

2 Xiuhong Li

3 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health,

4 615 N. Wolfe Street, Room E7141,

5 Baltimore, MD 21205 U.S.A.

6 Email: xili@jhsph.edu

7

8 Dr. Alvaro Muñoz

9 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health,

10 615 N. Wolfe Street, Room E7648,

11 Baltimore, MD 21205 U.S.A.

12 Email: amunoz@jhsph.edu