

## Combined Line - Double Bubbles - Pie chart.

By Alexandre Zolotovitski, Sun Microsystems Inc, San Diego, CA

### ABSTRACT

This paper summarizes work done to develop a new type of chart for visualizing time series data. In the problem that led to this chart, we had to visualize monthly time series data on the production of two modifications of a product. Each of these modifications was also classified as "Good" and "Bad", and the percentage of the product categorized as bad was of most interest. We solved the problem by creation of combined line - double bubbles - pie chart in SAS, using annotated `proc gplot`. This chart combines longitudinal and cross-sectional representation of data. The program was implemented in SAS Internet server and used SAS/IntrNet and SAS/Graph.

### DATA

This is the top of simplified dataset that we had to visualize. It is monthly time series data on the production of two modifications (classes) of a specific product. Each of these modifications was also classified as "Regular" and "Bad":

Obs	ym	email	class	state	cnt
1	2008-04	ada	B	B	0
2	2008-04	ada	B	R	8
3	2008-04	ada	B	B	0
4	2008-04	ada	B	R	1
5	2008-04	ash	B	B	0
6	2008-04	ash	B	R	1
7	2008-04	bob	B	B	0
...	...	...	...	...	...

**Table 1. The top of data set to visualize.**

The first column "ym" is the string variable containing the year and the month of production, the second column "email" is the string variable containing part of e-mail address of author of the product (e.g. e-mail [ada@sun.com](mailto:ada@sun.com) is presenting as "ada"); for confidentiality we had to hide this variable and to present in all reports corresponding ID variable "emID"; to recover email from em user had to have special privileges.

the third column "class" is the string variable containing class of the product, it can get values "A" and "B", the fourth column "state" is the string variable containing state of the product, it can get values "B" for "Bad" product and "R" for "Regular" (or "Released") product, and the fifth column "cnt" is the integer variable containing number of units of the product corresponding to values in columns 1-4.

Obviously the last column is result of aggregation using `proc mean` or `proc sql` of a raw dataset that contained data about individual units of product. This aggregation is simple and is out of scope of the work.

## TASK

Our task was to visualize shown in Table 1 time series data and the percentage of the product categorized as bad was of most interest. The result of the visualization must be shown as a web page by SAS intranet server.

## CHART

To solve this task we have created a special type of chart, that we named “combined line - double bubbles - pie chart” in SAS, using annotated `proc gplot`. This chart combines longitudinal and cross-sectional representation of data.

The program was implemented in SAS Intranet server and used SAS/IntrNet and SAS/Graph modules.

The example of this chart is shown on Chart1.

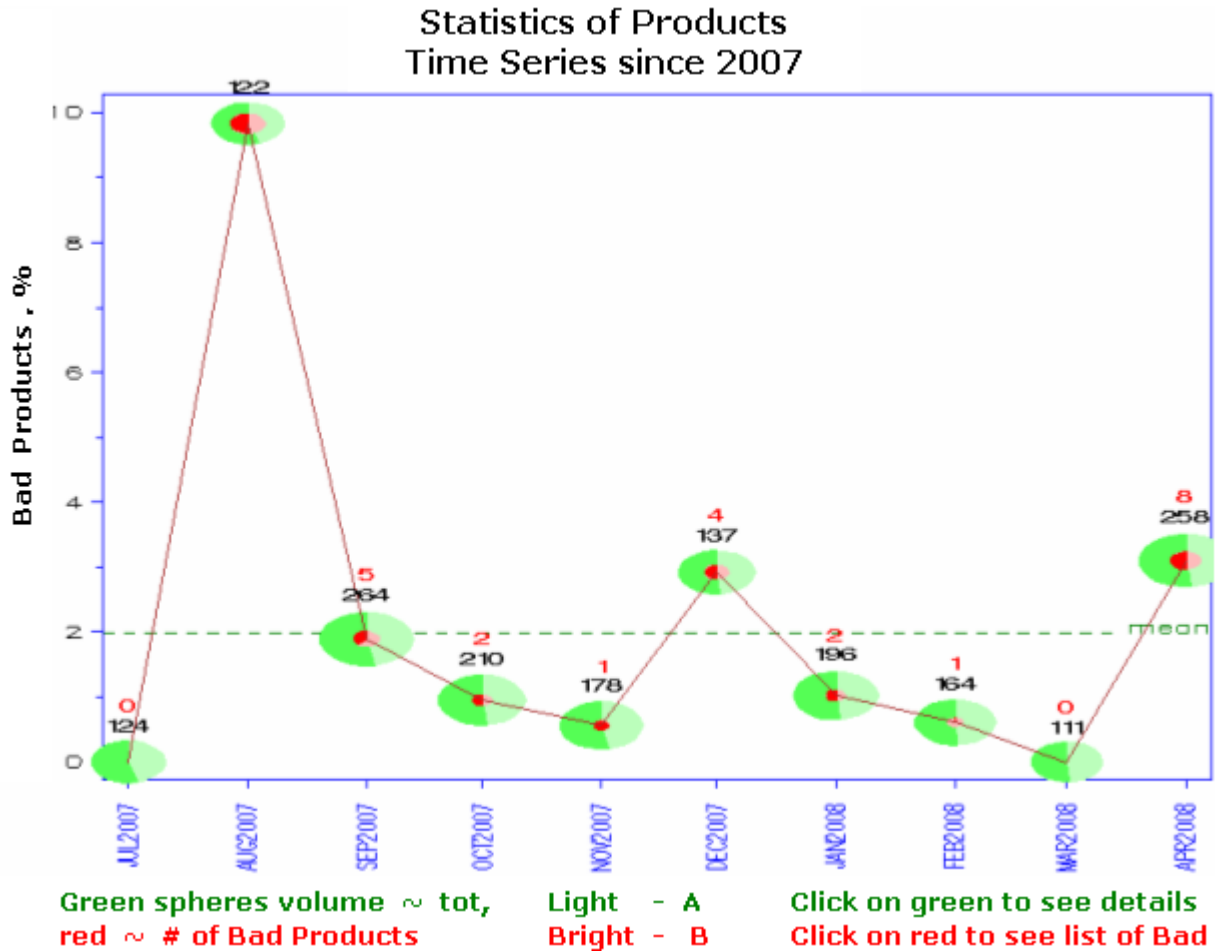


Chart 1. Statistics of Products. Time Series.

## ELEMENTS OF THE CHART

The chart has

Outside of plot area:

- 2-level header (p.1 in Code below)
- 2-level footer (p.2 in Code)
- Horizontal and vertical axes (p.1 in Code)

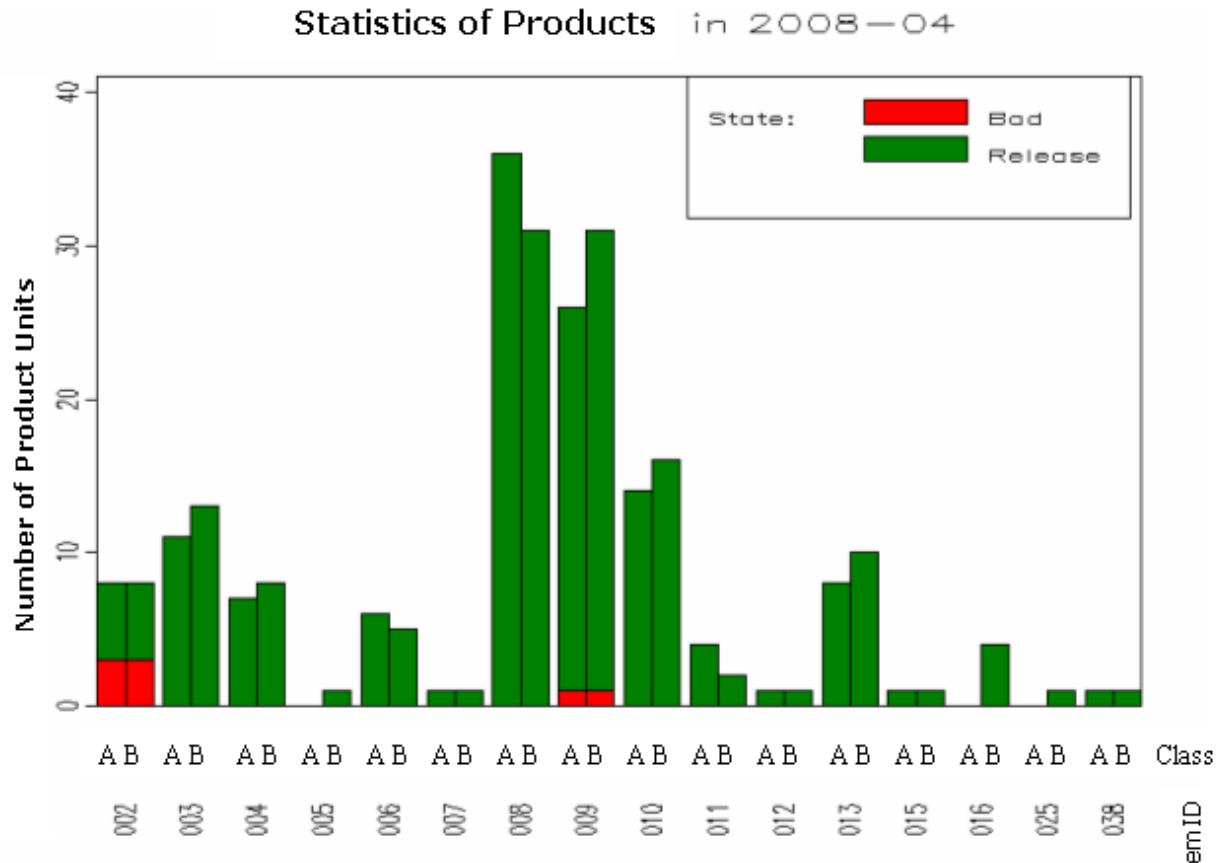
Inside plot area:

- Brown line shows percent of Bad Products vs ym (p.3 in Code)
- Light green section of spheres with volume proportional total number of units (p.4 -1 in Code)
- Dark green section of sphere segments with volume proportional number of class “A” units (p.4 -2 in Code)

- Light red section of spheres with volume proportional total number of Bad units (p.4 -3 in Code)
- Dark red section of sphere segments with volume proportional number of Bad class “A” units (p.4 -4 in Code)
- Black text label equal total number of units (p.5 in Code)
- Red text label equal total number of Bad units (p.5 in Code)
- Green dashed horizontal reference line indicating mean percent of Bad units (p.6 in Code)

We used “section of spheres with volume proportional total number of units” instead of “circle with area proportional total number of units” because typically percent of bad units is between 0 and 10, so if it is equal to 1, then in first case radius of internal sphere is about 30% of radius of external one and in the second case radius of internal sphere is about 10% of radius of external one and that is much worse for visualization.

All elements of the chart are clickable hyperlinks; click generates corresponding detailed bar chart related to pointed month for cross-sectional analysis. Example of such chart is shown on Chart 2. These charts are typical and we will not discuss it here.



**Chart 2. Statistics of Products. Cross-sectional view.**

## CODE

Below is the core part of the SAS code that generates the combined line - double bubbles - pie chart on Chart 1. To save place we use python – type indentation rather than putting “end” under corresponding “do”; we also use some elements of hard coding to reduce size of code.

In data step that prepares “anno” dataset each paragraph has “output” operator in the end.

```

%macro TimeSeries;
* the macro plot combined Line - Double Bubbles - Pie chart
* input global macrovariables:
*     outh - output directory on SAS server
*     mn_pBad - mean of percent of Bad
* input dataset out.ts

```

```

=====;
ods html body="&outh/index.htm" (notop nobot) ;

/* p.1. Titles, axes, symbol =====*/
title1 "Statistics of Products.";
title2 "Time Series since 2006.";

%let td= %sysfunc(today(), 'yy-mm-dd');
axis1 order=('01-Dec-2005'd to "&td"d by month) offset=(2,2)
      value=(a=90 r=0 f=simplex h=1.75)
      label=none color=blue
      major=(height=.75 ) minor=none;

axis2 order=(0 to 10 by 2) label=(angle=90 'Bad Products, %'
      j=center f=simplex height=.75) rellabel=(j=r c=green 'mean')
      offset=(2,2);

symbol1 interpol=none value=none; /* we plot using anno instead of symbol */

data anno(drop=k); set out.ts;
  length color style function $ 8 text $ 60 html $ 200; format date MONYY7.;
  retain color 'black' style 'swissb' xsys ysys '2'
  position 'E' size 12 hsys '3' WHEN 'a' k .6;

/* p.2. Footers =====*/
if _n_=1 then do;
  x=1; y=5; xsys='3'; ysys='3'; function='Label';
  position=6;size=2;color='green';
  text='Green spheres volume ~ tot, Light - A'; STYLE="simplex";
  html='href=Products_X_sec.htm'; output;
  x=1; y=2; xsys='3'; ysys='3'; function='Label'; position=6; size=2;
  color='red';
  text='red ~ # of Bad Products. Bright - B';
  STYLE="simplex";
  html='href=Products_X_sec.htm'; output;
  x=62; y=5; xsys='3'; ysys='3'; function='Label';size=2;color='green';
  text='Click on green to see details';STYLE="simplex";
  html='href=Products_X_sec.htm'; output;
  x=62; y=2; xsys='3'; ysys='3'; function='Label';size=2;color='red';
  text='Click on red to see list of Bad';STYLE="simplex";
  html='href=Products_X_sec.htm'; output;
end;

/* p.3. Line =====*/
if _n_>1 then do; x=date; y=pBad; xsys='2'; ysys='2'; function='draw';
  style='solid'; color='brown';size=.2; output; end;

/* p.4. Pies */
x=date; y=pBad; xsys='2'; ysys='2'; function='PIE'; ROTATE=360;
  SIZE=k*(tot)**.33; style='solid'; color="#BBFFBB"; *light green;
  html="&the_prog_html"||ym; output;
x=date; y=pBad; xsys='2'; ysys='2'; function='PIE';angle=90;
  ROTATE=360*pStot; SIZE=k*(tot)**.33; style='solid';
  color="#55FF55"; *dark green;
  html="&the_prog_html"||ym; output;
x=date; y=pBad; xsys='2'; ysys='2'; function='PIE'; ROTATE=360;
  SIZE=k*(bad)**.33; style='solid'; color="#FFBBBB"; *light red;
  html="&the_prog_html"||ym||&ListBad=1'; output;
x=date; y=pBad; xsys='2'; ysys='2'; function='PIE';angle=90;
  ROTATE=360*pSBad; SIZE=k*(bad)**.33; style='solid';
  color="#FF0000"; *dark red;
  html="&the_prog_html"||ym||&ListBad=1'; output;

```

```

/* p.5. Labels =====*/
x=date; y=pBad+ .16*k*(tot)**.33; xsys='2'; ysys='2'; function='Label';
angle=0; ROTATE=0; position= 'B'; style='swiss';
color='black';size=1.8; text=compress (put (tot,5.));output;
x=date; y=pBad+ .16*k*(tot)**.33+.3; xsys='2'; ysys='2'; function='Label';
position= 'B'; style='swiss'; color='red'; size=1.8;
text=compress (put (bad,5.));
html="&the_prog_html"||ym||'&ListBad=1'; output;

run;

proc gplot data=out.ts anno=anno;
plot pBad*date/ haxis=axis1 vaxis=axis2 vminor=1 caxis=blue
vref=&mn_pBad /* p.6. Reference line =====*/
lvref=2
cvref=green;

run;

ods html close;
%mend; *TimeSeries;

```

## CONCLUSION

We see that usage of annotated proc gplot to create a non-standard chart gives us more flexibility but requires more work than plotting standard charts using proc gplot only.

## REFERENCES

### SAS- 8

1. <http://v8doc.sas.com/sashtml/gref/index.htm> SAS/GRAPH Software: Reference
2. <http://v8doc.sas.com/sashtml/gref/zatachap.htm> The Annotate Data Set
3. <http://v8doc.sas.com/sashtml/gref/zaboutds.htm> About the Annotate Data Set – details
4. <http://v8doc.sas.com/sashtml/gref/zlotchap.htm> The GPLOT Procedure

### SAS/GRAPH(R) 9.2

5. <http://support.sas.com/documentation/cdl/en/graphref/59607/HTML/default/annodictchap.htm> Annotate Dictionary
6. <http://support.sas.com/documentation/cdl/en/graphref/59607/HTML/default/gplotchap.htm> The GPLOT Procedure

## Trademark Citation

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SASInstitute Inc. in the USA and other countries. ® indicates USA registration.

## Contact Information

The author welcomes feedback via email at [alex.zolotATsun.com](mailto:alex.zolotATsun.com) or [alex.zolotATgmail.com](mailto:alex.zolotATgmail.com)