

## Paper 232-28

**%Mondriaan: Presenting 3D information in 2D**

Erik W. Tilanus MSc., independent consultant, Vinkeveen, The Netherlands

**ABSTRACT**

The %MONDRIAAN macro creates charts that visualize 3-dimensional data in a 2-dimensional way. It uses 2 classification variables, one being a subdivision of the first (the first dimension) and 2 analysis variables, one to draw area's on the chart (the second dimension) and one to color area's (the third dimension). The macro takes care of all steps from raw data to chart.

**INTRODUCTION**

Piet Mondriaan was a Dutch painter (1872-1944) famous for his paintings consisting of coloured squares. Figure 1 presents a typical Mondriaan painting.

%MONDRIAAN honours the painter by presenting 3-dimensional data in 2D: X-axis, Y-axis and colored squares.

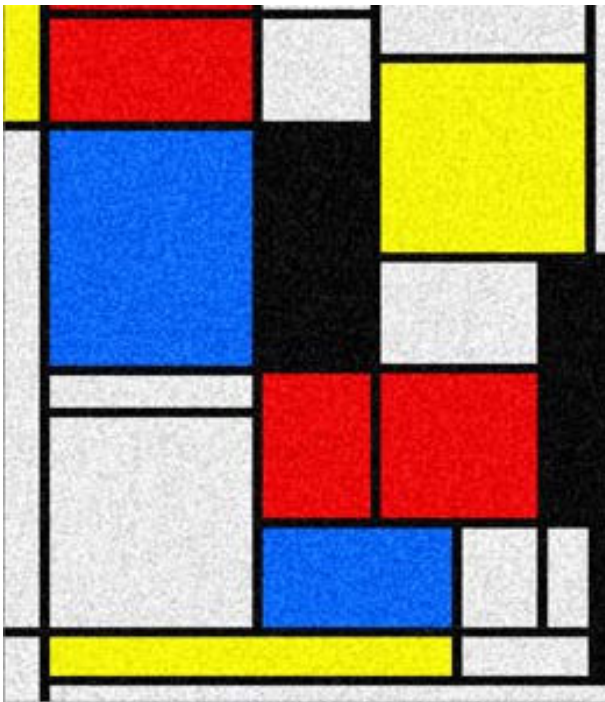


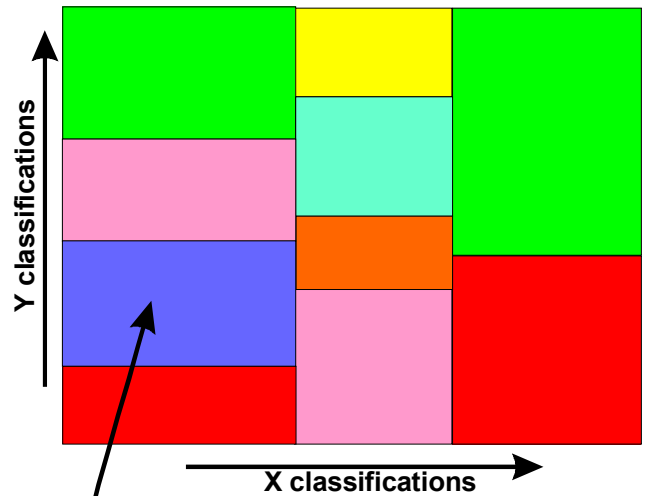
Figure 1: A typical Mondriaan painting. (1939)

**THE LAY-OUT PRINCIPLE**

The layout principle of the Mondriaan chart is given in figure 2. The chart always covers a complete page. The chart requires two classification variables, a share variable and a response variable. The X-classification values define the columns. The width of each column is based on the sum of the share variable values of that column, as a percentage of the whole chart.

Similarly each column is subdivided based on the Y-classification value. The vertical space reserved for each value is calculated based on the share variable value as a percentage of the column. As a result, each combination of an X- and Y-classification value is depicted as a square on the chart. The surface of the square represents the share of the X-Y combination as percentage of the whole chart.

The squares can be colored based on the value in the response variable.



Size: share variable

Color or pattern: response variable

Figure 2: The principle components of the MONDRIAAN chart.

**THE MACRO**

The macro takes a long list of parameters. Many of them have default values defined. Key parameters include the specification of the input data set, the X and Y classification variables, the share variable, which contains the value to construct the square and the response variable, which determines the color of the square.

The call may look as follows:

```
%MONDRIAAN (DATA= ..., XVAR= ..., YVAR= ...,
            SHARE = ..., RESPONSE= ... <other options and
            specifications>)
```

Only the XVAR=, YVAR= and SHARE= parameters are required. If DATA= is left out, the well-known value \_LAST\_ is assumed. RESPONSE= is not required, but you loose a substantial part of the value of the chart: the coloring of the squares.

The other parameters control elements like the other categories (rest levels), font and format information for the labeling, midpoint definitions for the response variable and finally, in case of WEB-publication, information regarding drill down possibilities.

**MACRO LOGIC**

In the macro four steps can be distinguished. The first step is the validation of the parameters. This step will also send notes to the SAS Log, to notify the user of all default settings that have been applied. The second step is the data preparation. In this phase the small category values (below the XREST and YREST values, default: 1%) are re-labeled into the rest category. Also the horizontal and vertical shares of each XY combination are calculated.

The third step builds the map data set, the response data set (used for the coloring) and the annotate data set (used for labeling the cells). If drill down information is present it will be integrated in the response data set.

The fourth and last step calls PROC GMAP to display the result. Following paragraphs describe the steps in more detail.

All temporary data sets that are created in the various steps follow the same naming convention: ADDMMYY, where A is a character or underscore, to recognize the file and DDMMYY is the SYSDATE value. For instance M20NOV02 is the map data set for a macro call on November 20, 2002.

## DATA PREPARATION

The actual processing starts with the calculation of the percentage share for each X classification value. This is done in a PROC FREQ.

All classification values below a threshold (determined by the XREST= parameter, default 1%) are then re-coded to dummy classification value 'FF'X, the internal code for the rest-category. With the re-coded values the percentages are recalculated. The results are possibly sorted, based on the XORDER= parameter, and then stored in a "X-data set".

Next is the calculation of the percentages in the Y direction per column.

This time PROC SUMMARY is used to calculate the total share value per column, which is then used to calculate the percentage for each X and Y classification value combination. In steps similar to the steps above, the rest values are re-coded to an internal value of 'FF'X. In this case the results are sorted according to the YORDER= parameter and stored in the "Y-data set". This data set is indexed for random access in the next step.

## CREATION OF THE MAP-, RESPONSE- AND ANNOTATE DATA SETS

PROC GMAP uses three input data sets. The MAP data set describes the complete map: all coordinates of the segments of the map. The RESPONSE data set contains the information that should be displayed at each segment. In a choropleth map, the values in the response data set determine the color of the segments. In case of drill-down facilities, the drill-down information is also included in this data set. The third data set is the ANNOTATE data set, which contains the label information.

These three data sets are created in one data step.

In the data step we read the X-data set sequentially. For each X value, we read repeatedly an observation from the Y-data set, for the given X value, using KEY access in the SET statement.

For each X-Y combination the four corner positions of the square are calculated and written to the map data set. Next the observation for the response data set is written. This contains the X and Y classification values, the response value and possible drill-down information. Finally the X classification value is written under the column and the Y classification value is written as a label inside the cell.

## PROC GMAP

The chart is basically a choropleth map, plotted by PROC GMAP. The options that are used depend on the parameters in the macro call.

## WEB OUTPUT AND DRILL-DOWN

Using ODS and the GIF driver you can add drill-down capabilities and Tool Tips to PROC GMAP output. Tool Tips display a little window when you move your mouse over a cell. Drill-down happens when you click in the cell. These Tool Tips work only in Internet Explorer, not in Netscape (not even in 7.0). The drill-down functions both in IE and in Netscape.

To make use of these capabilities, the drill-down information should be included in a data set that contains the classification variables (XVAR= and YVAR= values) and the link information in a character variable (specified in the HTMLVAR= parameter). The link information in HTMLVAR has the format of:

```
HTMLVAR="http://link-address" ALT="tool tip information"
```

You don't need to specify both, just use the part you need.

Figure 4 illustrates the effect. Moving the mouse over a cell, displays relevant detail information.

## EXAMPLE

In the following example airline booking information<sup>1</sup> is shown per region and country of origin of passengers. The flights are grouped by region, e.g. Europe, North Atlantic, Asia/Pacific etc.

The size of the square is based on the number of passengers from each country on the given region. The coloring is based on the index of the bookings compared to previous year.

Part of the input data is shown in table 1. It contains the relevant X and Y classification codes (the flight region and the country code), the number of passengers (which determines the surface of the square via the SHARE= parameter) and the index of the number of passengers in comparison to previous year.

The coding to create the result (figure 3) has been:

```
proc format;
value levels
low- 70 = '<70'
70-85 = '70-85'
85-95 = '85-95'
95-105 = '95-105'
105-115 = '105-115'
115-130 = '115-130'
130-150 = '130-150'
150-high= '>150';
run;
pattern1 v=solid c=CXff0000;
pattern2 v=solid c=CXff6666;
pattern3 v=solid c=CXffcccc;
pattern4 v=solid c=CXffffff;
pattern5 v=solid c=CXffffcc;
pattern6 v=solid c=CXffff44;
pattern7 v=solid c=CXddffdd;
pattern8 v=solid c=CX00ff00;

title h=1 'Bookings January 2003';
filename gsf'd:\sugi28\mondriaan.cgm';
goptions device=cgmo971 gsfname=gsf
ftext=hwcgm001;

%mondriaan(data=mondriaan,
xvar=region,
yvar=countrycode,
share=paxcount,
response=index,
lblfmt=6.,
xrest=2,yrest=2,
xrestlbl='other',
yrestlbl='other',
legend=y,respbl='Index:',
discrete=y,midptfmt=levels,
lblsize=0.6,loflow=N,
lblfont=hwcgm001);
```

PROC FORMAT defines the levels that are used for the coloring of the squares. The coloring is defined in the pattern statements. If the chart is to be distributed in black and white, there is also the option to define fill patterns, by changing the V parameter in the statement.

For this example I did not use ODS, rather I used the CGM driver that matches best with MS-Office, to create a file that can directly be imported in the word document.

The rest of the coding speaks for itself. The result is in figure 3.

To use the tool tip facility, you specify the required ODS information, you change the device driver into GIF and specify the input file with the drill down information. In our situation we had the information already included in the MONDRIAAN data set, in the HTMLVAR variable (the default name). This leads to the following changes to the coding:

```
filename gsf'd:\sugi28\webfiles';
goptions device=gif gsfname=gsf;

ods listing close;
ods html path='d:\sugi28\webfiles'
body='body.html';
```

<sup>1</sup> The used values are fictitious.

```
%mondriaan(data=mondriaan,
  xvar=region,
  yvar=countrycode,
  share=paxcount,
  response=index,
  lblfmt=6.,
  xrest=2,yrest=2,
  xrestlbl='other',
  yrestlbl='other',
  legend=y,resp1bl='Index:',
  discrete=y,midptfmt=levels,
  lblsize=0.6,lflow=N,
  drilldata=mondriaan);
```

```
ods html close;
ods listing;
```

The only changes are the specification of a directory in the filename statement, the GIF driver assignment, the ODS statements and the additional option DRILLDATA=.

Partial input file for %MONDRIAN 1  
09:43 Wednesday, November 20, 2002

Obs	region	countrycode	paxcount	index
1	EUROPE	AD	58	446.15
2	EUROPE	AE	264	93.29
3	EUROPE	AG	0	.
4	EUROPE	AL	0	0.00
5	EUROPE	AM	26	1300.00
6	EUROPE	AN	48	33.80
7	EUROPE	AO	14	700.00
8	EUROPE	AR	169	30.56
9	EUROPE	AT	2465	41.72
10	EUROPE	AU	592	48.37
11	EUROPE	AW	58	223.08
12	EUROPE	AZ	5	.
13	EUROPE	BA	4	400.00
14	EUROPE	BD	18	128.57
15	EUROPE	BE	7098	89.27
16	EUROPE	BF	2	.
17	EUROPE	BG	119	138.37
18	EUROPE	BH	92	180.39
19	EUROPE	BI	16	11.76
20	EUROPE	BJ	0	0.00

Table 1: Partial list of the input to %MONDRIAN. It contains the X and Y classification variables, the number of passengers that are booked and the index compared to previous year.

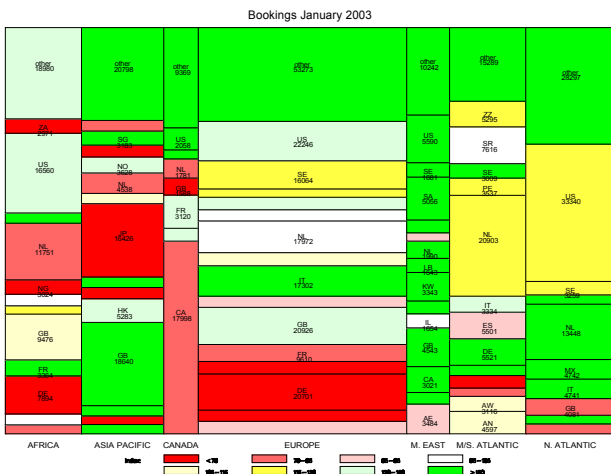


Figure 3: The Mondriaan chart resulting from the sample code and data

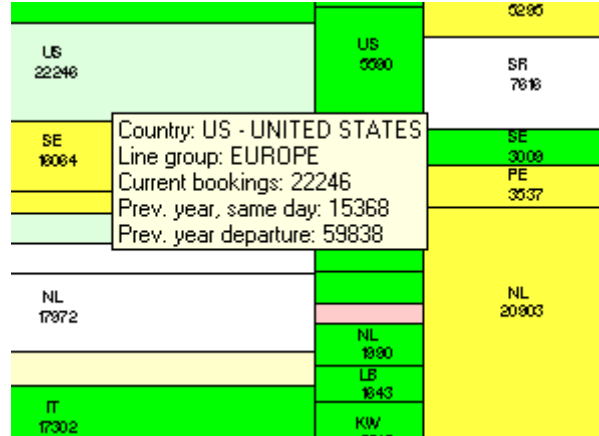


Figure 4: Detail of the chart of figure 3, with the Tool Tips window, activated for the light green US cell.

### CONCLUSION

%MONDRIAN is an easy and effective method to provide quick graphical insight in data: two classification variables, and two analysis variables. With the advent of ODS you can add drill-down and tool tip features, when using the chart in a browser environment.

### REFERENCES

- More information about the SAS macro language: *SAS Macro Language: Reference, Version 8*
- More information about PROC GMAP: *SAS/GRAPH® Software: Reference, Version 8*
- More information on using Tool Tips and Drill-Down capabilities can be found on the SAS website: [http://www.sas.com/service/techtips/ts\\_qa/0210\\_maptips.html](http://www.sas.com/service/techtips/ts_qa/0210_maptips.html)

### ACKNOWLEDGMENTS

Thanks to my former colleagues at KLM Royal Dutch Airlines for reviewing the draft of this paper.

### CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:  
 Erik W. Tilanus  
 P.O. Box 77  
 3645 ZK Vinkeveen, the Netherlands:  
 Fone: +31 297 263936  
 Fax: +31 297 261712  
 Email: erik\_tilanus@compuserve.com

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

Other brand and product names are trademarks of their respective companies.