Paper CC-21

Using Heatmaps and Trend Charts to Visualize Kidney Recipients' Post-Transplant Hospitalizations

Sebastian D Perez MSPH,

Department of Surgery, Emory University School of Medicine

ABSTRACT

In trying to understand the hospital resource utilization of kidney transplant patients after surgery, it is important to know how often patients have to be readmitted to the hospital, how long they spend in the hospital and what factors influence their readmission rates. As a component of exploratory analysis, data visualization is essential to elucidate patterns and relationships between study variables. This paper shows how heatmaps and trend charts, summarizing the percent of patients hospitalized, were generated to understand hospitalization rates over time.

Graphics and data manipulation were all performed using SAS® 9.2 on Windows 7 and all plots were created using the SGPLOT procedure.

INTRODUCTION

Data visualization is an essential component of exploratory data analysis, as it is essential to elucidate patterns and relationships in the data. In a study looking at the hospital resource utilization of kidney transplant patients after surgery, investigators wanted to visualize the patterns of readmission to the hospital. It was important to be able to determine when patients were being readmitted, for how long and what characteristics readmitted patients shared.

Heatmaps were developed to show every hospitalization of every patient in our data set over a 3 year period. The purpose was to see each patient's status over time: visualize the time between hospitalizations and the duration of each hospitalization. Times that a patient spent in the hospital were shown as horizontal bars so you could see how many days post-op the patient re-entered the hospital and, by the length of the bar, the length of that particular admission. The bars were color coded by number of readmission or, alternately, by a patient characteristic (eg. diabetic or not). These heatmaps were useful to get a general sense of the pattern of hospitalizations for the whole population and investigate the hospitalizations of individual patients.

Patterns of hospitalizations by *group* of patients were not immediately apparent in heatmaps, so the data was summarized by group and the trend of hospitalization was plotted over time. For each 30 day period post op, the percent of patients who were hospitalized during that time span was calculated within each group of interest. This percent was then shown on a series graph with time on the x-axis. By graphing two or more groups of patients on the same graph, the trends and differences in patterns were observed

HEATMAP OF INDIVIDUAL READMISSIONS

SAS version 9.2 does not have a specific mechanism for creating heatmaps, per se, but PROC SGPLOT graphs can be manipulated to achieve a heatmap (Fig. 1). The x-axis shows the days after the initial surgery while the y-axis lists the patients. This is achieved by using a scatter plot with rectangular symbols identifying each day in the hospital, so that consecutive days show up as a horizontal line. The bars are color coded so that the first hospitalization is a light orange, the second, a darker orange and the third or higher number hospitalizations are coded red. The code for developing this plot follows, but first, the data set has to be rearranged in the correct manner.

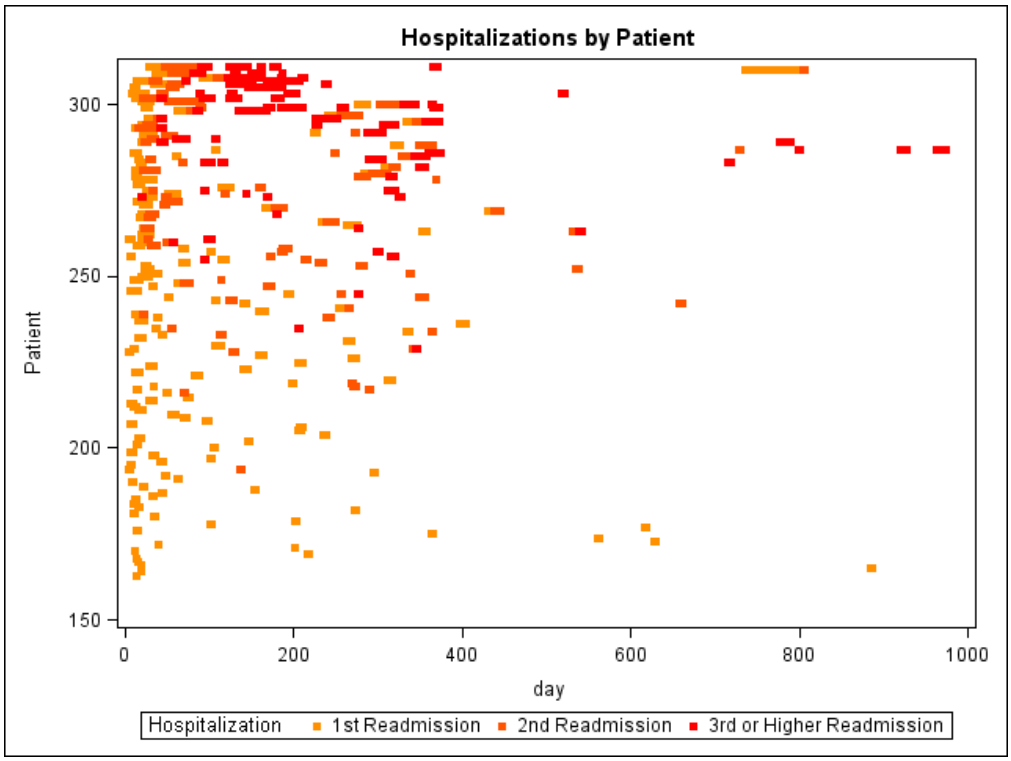

Figure 1: Heatmap of Hospitalization by Patient, Colored by Number of Readmission

PREPARING DATA FOR THE HEATMAPS

The data was provided in a *wide* format, with each patient having one line in the data set and all the dates of entry and discharge from the hospital were provided in individual variables (up to 8 admission dates and 8 discharge dates since no patient had more than 8 hospitalizations). The first step in data preparation is calculating the number of days from the original transplant date until the date of readmission and until discharge.

```
data readmit; set readmit;
Timetoreadmit1= Readmit_1_admission_date-Date_of_Transplant;
Timetodischarge1=Readmit_1_discharge_date-Date_of_Transplant;

Timetoreadmit2= Readmit_2_admission_date-Date_of_Transplant;
Timetodischarge2=Readmit_2_discharge_date-Date_of_Transplant;
   .
   .
   .
Timetoreadmit8= Readmit_8_admission_date-Date_of_Transplant;
Timetodischarge8=Readmit_8_discharge_date-Date_of_Transplant;
run;
```

From this data, it is necessary to calculate whether the patient was in the hospital or out of the hospital in each of the 1,095 days (3 years) that followed the surgery. A DO loop is used to go through each day and code it as a 1 if the patient is in the hospital for the first time, a 2 if they are hospitalized for the second time, a 3 if they are hospitalized for the 3rd or more time, and left as a missing value if the patient was not in the hospital (1095 new variables are created).

```
data map; set readmit;
array dy(1095) day1 - day1095;
do i=1 to 1095;
  if Timetoreadmit1 <= i=<Timetodischarge1 then dy(i)=1;
  else if Timetoreadmit2 <= i=<Timetodischarge2 then dy(i)=2;
```

```
  else if Timetoreadmit3 <= i=<Timetodischarge3 then dy(i)=3;
  else if Timetoreadmit4 <= i=<Timetodischarge4 then dy(i)=3;
  else if Timetoreadmit5 <= i=<Timetodischarge5 then dy(i)=3;
  else if Timetoreadmit6 <= i=<Timetodischarge6 then dy(i)=3;
  else if Timetoreadmit7 <= i=<Timetodischarge7 then dy(i)=3;
  else if Timetoreadmit8 <= i=<Timetodischarge8 then dy(i)=3;
  else dy(i)=.;
end;
run;
```

This data set is then reshaped into a *long* format (one line per day for each patient) and unnecessary variables are deleted.

```
data mapfinal; set map;
array dy(1095) day1 - day1095;
do i=1 to 1095;
 day=i;
 readmit=dy(i);
 output;
end;
label readmit='Hospitalization';
keep patient day readmit;
run;
```

CODE FOR HEATMAPS

Once the data is in the correct format, the heatmaps can be created. First, a graph style is created to define the color ramp from yellow to red using PROC TEMPLATE, and PROC FORMAT is used to label the hospitalizations correctly.

```
proc template;
define style styles.heatmap;
parent=styles.listing;
style graphcolors from graphcolors /
'gcdata1'=CXFF9000
'gcdata2'=CXFF5500
'gcdata3'=CXFF0000 ;
end;
run;
```

```
proc format;
value readmit
1='1st Readmission'
2='2nd Readmission'
3='3rd or Higher Readmission';
run;
```

Second, a scatter plot is created using SGPLOT (and the previously created style), with the patient identifier in the y-axis and the number of days from transplant in the x-axis.

```
ods graphics on;
ods listing style=heatmap;
proc sgplot data=mapfinal;
title "Hospitalizations by Patient";
where readmit>0;
scatter x=day y=patient /
markerattrs=(size=.05in symbol=squarefilled)
group=readmit;
format readmit readmit.;
run;quit;
```

The WHERE statement is used so that only days with hospitalizations are used in the plot and the MARKERATTRS statement is utilized to set the look of the plot points. Using a filled in square ensures that a hospitalization with multiple days will look like a horizontal bar. The size of the marker might need to be adjusted based on the number of patients and number of days being graphed. The marker needs to be large enough so there will not be any white space between days but not too large so as to overlap over the hospitalizations of other patients.

PANELED HEATMAPS

To compare two or more subpopulations a PROC PANEL procedure can be used to divide the heatmap by group. Below (Fig. 2) are heatmaps comparing diabetic vs. non-diabetic patients.

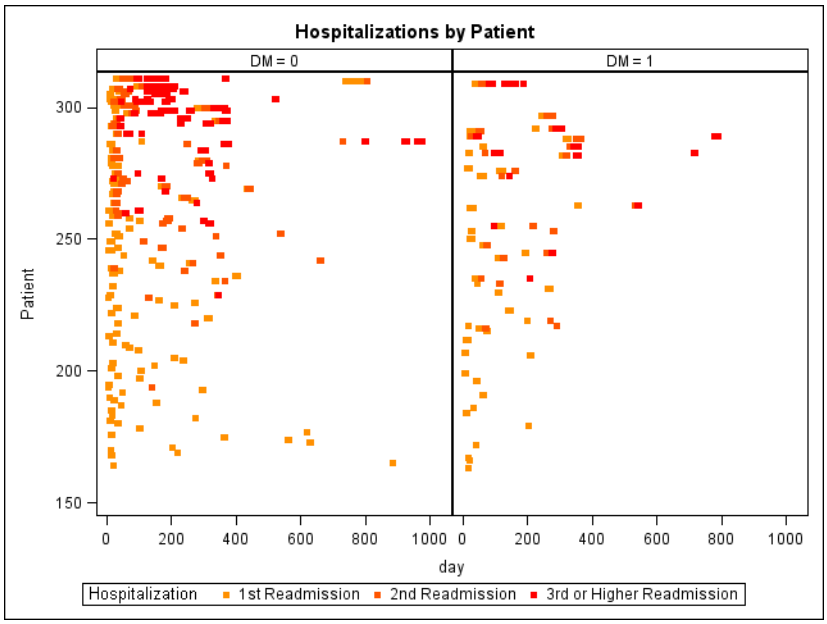

Figure 2: Paneled Heatmaps Comparing Diabetic (DM=1) and Non-Diabetic Patients

The code for the paneled plots is a simple modification of the original SGPLOT code:

```
proc sgpanel;
panelby DM;
where readmit>0;
scatter x=day y=patient /
markerattrs=(size=.05in symbol=squarefilled)
group=readmit;
format readmit readmit.;
run;
```

TREND CHARTS OF READMISSIONS OVER TIME

Trend charts (Fig. 3) summarize the percent of patients who were hospitalized at any point within a given month (or a particular 30 day period) after transplant. In this case, the trend for diabetic (DM='Yes') patients is being compared to trend for non-diabetics. In this chart, it is easy to see the overall trend for hospitalizations, (lots of early hospitalizations and then less as time goes by) and easy to identify differences in hospitalization rates between groups (in this case diabetic patients seem to have slightly higher rates at most time points).

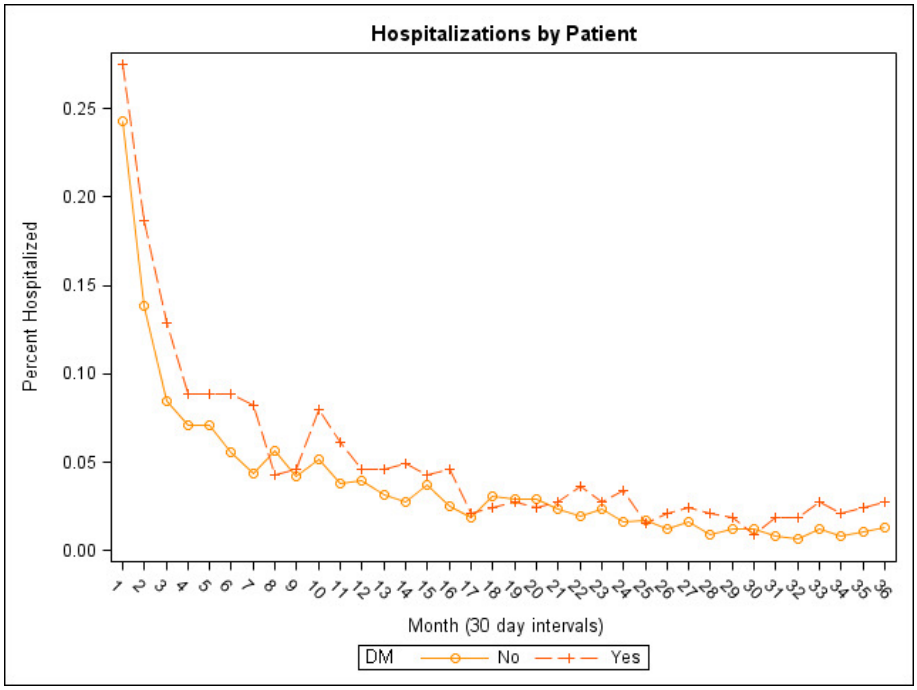

Figure 3: Hospitalization Trend Chart Comparing Diabetic (DM=1) and Non-Diabetic Patients

PREPARING DATA FOR TREND CHARTS

Setting up the data for the trend charts is slightly different from the heatmaps. It is necessary to determine whether a patient was in the hospital within a given 30 day period, not individual day (a chart by day would result in too much variation between points and make it harder to see trends). Hospitalizations have to be coded as a '1' or a '0', and then the data set is transformed into a *long* format.

First, two arrays of 8 elements are created to store the dates of admissions and dates of discharges for each hospitalization (for a maximum of 8 readmissions). These will make it easier to use DO loops to check each readmission. Another, blank, array is created to represent each 30-day period in the three years post-transplant (36 elements).

```
data test2; set test;
array AdmStart (8)
Readmit_1_admission_date Readmit_2_admission_date Readmit_3_admission_date
Readmit_4_admission_date Readmit_5_admission_date Readmit_6_admission_date
Readmit_7_admission_date Readmit_8_admission_date;
array AdmEnd (8)
Readmit_1_discharge_date Readmit_2_discharge_date Readmit_3_discharge_date
Readmit_4_discharge_date Readmit_5_discharge_date Readmit_6_discharge_date
Readmit_7_discharge_date Readmit_8_discharge_date;
array MonthN(36);
do j=1 to 36;
  MonthN(j)=0;
  do i=1 to 8;
    if (j-1)*30<(admstart(i)-date_transp)<=j*30 OR
       (j-1)*30<(admEnd(i)-date_transp)<=j*30
       then MonthN(j)=1;
    end;
  end;
output;
run;
```

Two DO loop are utilized to go through each of the 8 possible readmissions (inner loop) for each patient and check whether they fell within any of the 36 time periods (outer loop) after transplant. If so, that is indicated by a 1 in that 'month' for that patient or a zero otherwise.

The next piece of code simply puts the data into a *long* format to be used in the graph.

```
data long; set test2;
array MonthInd(36) MonthN1--MonthN36;
do i=1 to 36;
  Hospitalization=MonthInd(i);
  TimePeriod=i;
  output;
end;
keep Patient_ID DM Hospitalization TimePeriod ;
run;
```

Only the Patient ID, an indicator for diabetes, an indicator for whether the patient was hospitalized at the time period and an identifier for the time period are kept.

CODE FOR TREND CHART

The trend chart is simply a VLINE plot in SGPLOT, that plots the average value (STAT=MEAN) of 'Hospitalization' (RESPONSE) by the time period. Because hospitalizations are coded as 0s or 1s, this average is equivalent to the proportion of patients hospitalized during that period. A GROUP statement is used to divide the patients into diabetic and non-diabetic groups (the MARKERS option simply adds point markers to the line).

```
proc sgplot data=long;
vline TimePeriod/ response=Hospitalization stat=mean markers group=dm;
xaxis label='Month (30 day intervals)';
yaxis label='Percent Hospitalized';
run;
```

CONCLUSION

Both heatmaps and trend charts can be useful in visualizing hospitalization trends in patients after transplant. Heatmaps give a detailed view of each patient's hospitalization pattern but are difficult to use for detecting differences between groups of patients,. A line plot summarizing the *trend* in hospitalizations, by group, displays clearly the differences between groups. These graphs are very easy to create using SGPLOT, if the data is in the correct format. This paper gives a short summary of how to manipulate data into the correct format and produce these graphs.

REFERENCES

SAS Institute Inc. (2008). "SAS/GRAPH 9.2: Statistical Graphics Procedures Guide." http://support.sas.com/documentation/onlinedoc/graph/index.html.

Schwartz, Susan. 2008. "Butterflies, Heat Maps and More: Explore the New Power of SAS/GRAPH. *Proceedings of the 2008 SAS Global Forum.* April, 2008

CONTACT INFORMATION

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

Sebastian D. Perez MSPH
Emory University School of Medicine, Department of Surgery
101 Woodruff Cir. WMB 5008
Atlanta, GA 30322
sdperez@emory.edu