

Analysis 2: Exploring and Visualizing Health Care Data

In Module 3, we started to learn about data visualization. In this problem, you'll get a chance to try out some of these techniques on a dataset from the world of healthcare. Often when doing exploratory data analysis, you'll have to do various data preparation and transformation activities to get your data in a form that's amenable for analysis. This may involve creating new derived columns or even entirely new derived tables from your original dataset.

We will use a set of data from a post-anesthesia care unit (PACU) from 6/1/2010-6/30/2010. Here's what the data looks like:

	A	B	C	D	E	F	G	H	I
1	VisitNum	Severity	PatType	RecoveryIn	RecoveryOut	RecovMins	DayOfWeek	Hour	Destination
2	1	Acuity 1	I	6/3/99 19:36	6/3/99 19:55	19	5	19	4N
3	2	Acuity 2	I	6/19/99 9:40	6/19/99 12:45	185	7	9	3N
4	3	Acuity 2	I	6/15/99 0:26	6/15/99 1:15	49	3	0	4N
5	4	Acuity 2	I	6/11/99 13:56	6/11/99 15:10	74	6	13	3N
6	5	Acuity 2	I	6/8/99 13:11	6/8/99 15:30	139	3	13	3N

VisitNum	Unique patient identifier
Severity	Acuity code 1-5 (5 most acute)
PatType	I=inpatient, O=outpatient
RecoveryIn	Date and time patient enters PACU
RecoveryOut	Date and time patient leaves PACU
RecovMins	Number of minutes spent in PACU
DayOfWeek	1=Sunday, 2=Monday, ..., 7=Saturday
Hour	Hour of the day 0=12:00AM-1:00AM, ..., 23=11:00PM-12:00AM
Destination	Name of nursing unit or phase II recovery area where patient will go

Patients enter the PACU after surgery and have several defining characteristics as detailed in the table above. We are interested in using concepts from multidimensional data modeling and OLAP (online analytical processing) to do some exploratory business analysis. I have created an Excel file called **pacu.csv** for this assignment. Each record is a single patient. The key service measure included in the spreadsheet is the amount of time spent in the PACU (RecovMins).

As a prelude, I'd like you to take a look at a tutorial I wrote last year to help people get started with the open source statistical package, R. I am **NOT** asking you do anything in R. However, I'd like you to read through the tutorial to familiarize yourself with some of the types of data analysis and visualizations we might do. In Spring 2014, I'm going to offer an advanced analytics course in which we'll learn how to use tools like R (and Python) for business analytics. Here's the R version of the tutorial:

<http://hselab.org/content/getting-started-r-plyr-and-ggplot2-group-analysis>

2a – Creating new derived columns) Currently the Severity field is a string that looks like “Acuity *n*” where *n* is an integer in 1-5. Create a new column called Acuity which just contains the integer from the Severity field. Now you have a numeric acuity field that you can use if you wish. Make sure that the new column contains numeric data and NOT text that just happens to look like a number.

	A	B	C	D	E	F	G	H	I	J
1	VisitNum	Severity	PatType	RecoveryIn	RecoveryOut	RecovMins	DayOfWeek	Hour	Destination	Acuity
2	1	Acuity 1	I	6/3/2010 19:36	6/3/2010 19:55	19	5	19	4N	1
3	2	Acuity 2	I	6/19/2010 9:40	6/19/2010 12:45	185	7	9	3N	2
4	3	Acuity 2	I	6/15/2010 0:26	6/15/2010 1:15	49	3	0	4N	2

In addition, create columns containing numeric values of the month, day of month, and year based on the RecoveryIn field.

2b – Count cases by PatType) First, create a pivot table on a new sheet to count the number of cases by PatType. Use the VisitNum field as the field to count. The default pivot table will look something like this:

	A	B
1		
2		
3	Row Labels	Count of VisitNum
4	I	994
5	O	694
6	Grand Total	1688

Figure out how to modify the label for the value field and to display the actual row label field name so that it looks like this:

	A	B
1		
2		
3	PatType	Num Cases
4	I	994
5	O	694
6	Grand Total	1688

2c – Recovery minutes factors) Does PatType and/or Acuity affect recovery minutes? Create a few pivot tables in Excel that helps answer this question. Discuss your findings.

2d) Counting wizardry) Create a pivot table to count the number of cases by Hour (row label) and Acuity (column label). Now, create copies of the row and column labels as shown below. The Pivot Table is on the left. The formulas for doing the counts in the middle section use the COUNTIFS() function. You do that. The formulas on the right use array formulas relying on the SUM() function and some Boolean multiplication. That's a little trickier but well worth knowing how to do.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Uses Pivot Table									Uses COUNTIFS()							Uses array formula using SUM and boolean multiplication				
2																					
3	Count of VisitNum	Column Labels																			
4	Row Labels		1	2	3	4	5	Grand Total		1	2	3	4	5			1	2	3	4	5
5	0		10	1	1			12		0	0	10	1	0	1	0	0	10	1	0	1
6	1			3	2			5		1	0	0	3	2	0	1	0	0	3	2	0
7	2		1	4	2			7		2	1	4	0	2	0	2	1	4	0	2	0
8	3			5	1	2		8		3	0	5	1	2	0	3	0	5	1	2	0
9	4			3		1		4		4	0	3	0	0	1	4	0	3	0	0	1
10	5			5	1			6		5	0	5	1	0	0	5	0	5	1	0	0
11	6			3	2	1		6		6	0	3	2	0	1	6	0	3	2	0	1
12	7			2	1			3		7	0	2	1	0	0	7	0	2	1	0	0
13	8		14	63	4	3		84		8	14	63	4	3	0	8	14	63	4	3	0
14	9		14	100	15	16	2	147		9	14	100	15	16	2	9	14	100	15	16	2
15	10		16	123	15	19	9	182		10	16	123	15	19	9	10	16	123	15	19	9
16	11		17	119	7	25	6	174		11	17	119	7	25	6	11	17	119	7	25	6
17	12		18	94	8	24	7	151		12	18	94	8	24	7	12	18	94	8	24	7
18	13		16	104	8	29	10	167		13	16	104	8	29	10	13	16	104	8	29	10
19	14		15	89	11	29	4	148		14	15	89	11	29	4	14	15	89	11	29	4
20	15		5	93	8	28	14	148		15	5	93	8	28	14	15	5	93	8	28	14
21	16		7	68	7	31	3	116		16	7	68	7	31	3	16	7	68	7	31	3
22	17		10	57	4	17	6	94		17	10	57	4	17	6	17	10	57	4	17	6
23	18		5	42	2	13	7	69		18	5	42	2	13	7	18	5	42	2	13	7
24	19		9	32	4	8	5	58		19	9	32	4	8	5	19	9	32	4	8	5
25	20		3	20	5	6		34		20	3	20	5	6	0	20	3	20	5	6	0
26	21		6	14	5	5		30		21	6	14	0	5	5	21	6	14	0	5	5
27	22		4	13	3	1		21		22	4	13	3	1	0	22	4	13	3	1	0
28	23			9	1	3	1	14		23	0	9	1	3	1	23	0	9	1	3	1

2e – Histograms) Create three histograms of RecovMins.

Histo 1: Use the Data Analysis ToolPak

Histo 2: Use the FREQUENCY() array function to create the counts by bin and then build your own graph

Histo 3: Use the Grouping feature available in Pivot Tables to create the counts by bin.

The first two techniques are covered in the screencast on Histograms 3 Ways. The last technique is covered in the Call Center Pivot Table tutorial (hard copy and covered in screencast of same name). Make sure you use good graph design principles for your histograms.

Also, use Excel's PERCENTILE() function to compute the 95th percentile of RecovMins. Write a short cell comment explaining what this number means in plain English.

Now, something to ponder, why would it be difficult to create histograms of RecovMins for each Acuity? Also, how difficult is it to compute the 95th percentile of RecovMins by Acuity? Go back and look at the R tutorial referenced at the top of this problem. That's exactly what we did there. What makes it hard (or at least tedious) in Excel? Just comment on this in a cell comment or text box on the same sheet as your histograms.

EXTRA CREDIT: Use JMP and figure out how to do histograms and box plots of RecovMins by Acuity.

2f – Trends in cases by Acuity) Create a Pivot Chart containing a time series plot (a line chart) of the number of cases by acuity by day of month

2g – Excel Dashboard) In a new sheet, construct an Excel based “dashboard” showing a variety of relevant data visualizations to a PACU manager. In addition to some of the things we explored above, PACU managers are also interested in time of day and day of week specific volume and workload. They’d like to be able to see at a glance the percentage breakdown of cases by PatType and by Acuity.

Your dashboard must contain at least two pivot tables that are linked to the same Slicer (yes, you have to figure out what Slicers are and how to use them).

Be creative. In Module 4, we covered some dynamic charting approaches including things like in-cell graphing, sparklines, conditional formatting, dynamic charts, box plots and motion charts. Try to include a few of these.

2h – [EXTRA CREDIT] Tableau) See the **Getting Started with Tableau** tutorial available in the Module 3 book. Either download and install Tableau on your computer (instructions available in the Textbook and Software section of our Moodle site) or use Tableau Public (I’m sure you can figure it out) to create an interesting dashboard based on this healthcare data set. FYI, you can also make motion charts with Tableau. You can either submit your Tableau Workbook or the URL to your Tableau Public visualization.