Dashboard for Statistical Visualization of Client Service Events

Alex Zolotovitski
StatVis Consulting, San Diego, CA, US. Alex.Zolot@statvis.com

Abstract - *For every business interaction with customers consists of cases and each case consists of sequence of events: First_Contact_Customer, … intermediate events, … Case_Closed. The most important characteristics are frequencies of transitions between events and mean time between events (MTBE, TBE) for each type of cases. Type of cases could be type of customer, group of products, branch of enterprise, geographical area, etc. Existed methods of visualization could not visualize two characteristics (Frequency and MTBE) simultaneously to locate business problems.*

Our dashboard combines standard SPC chart for time series representation with three new types of charts for cross-sectional representation: "matrix bar chart" for portraying types of cases, "flower bed chart" for displaying Frequencies and MTBE, and "Tower Chart" that can be element of "Flower Bed Chart" and "Matrix Bar Chart" when we need detailed visualization of distribution of TBE.

This new dashboard is applicable for any customer service – help desks, stores, doctor offices, banks and gives the user ability to identify immediately the most business important factors.

Keywords: Visualization, Data Mining, Customer Service.

1 Introduction

Improving performance of interaction with customers ("customer service") is important business task of CRM for every business. In this work we have deal with the problem of visualization for Client service events to optimize work of client service. In order to do it we have to visualize business important characteristics related to customer service. The raw data related to customer service usually has form

Table 1. Raw data

	DateTime			Case Types		
Case	**Ev1**	**Ev2**	**…**	**Type1**	**Type2**	**…**
Case1	T_{11}	T_{12}		$Type_{12}$	$Type_{11}$	
Case2	T_{21}	T_{22}		$Type_{22}$	$Type_{21}$	
Case3	T_{31}	T_{32}		$Type_{32}$	$Type_{31}$	
Case4	T_{41}	T_{42}		$Type_{42}$	$Type_{41}$	
……						

In our example of technical service center events were service cases, so variable *Case* was the foreign key identifying service case; the following columns are for DateTime stamps for service events Ev1, Ev2,.. that could be *Creation – Received – Contact_SW –Contact_HW – Pending – Closed*.

The Type columns could contain such variables as *HW_Platform*, *Product*, Geographic variables, *Customer*, *Case_Owner* and can be used for the Classification of cases. For simplicity we will show only one Type variable.

The same type of visualization can be done for analysis of events in other areas: reliability (failures), survival analysis(deceases) , transport network flow analysis, network performability analysis, cross-sell and up-sell analysis in marketing, e.g. in the last case events could be purchases of specific products by a customer. For example, we could have deal with opening a sequence of bank accounts; then instead of *Case* we have *CustomerID*, and *Event* can be *Open_Checking_Acct, Open_Saving_Acct, Open_Loan, Close_Checing_Acct* and so on; The *Type* could be *BranchID* or *Group of Clients* and can be used for the Classification of cases.

Tasks of this type are quite common in OLAP [1, 2].

2 Analysis of data

More convenient is to present the data of Table 1 in the "long" format:

Table 2. Raw data in the "long" format

	Case	DateTime	Event	Type
Case1		T_1	Ev_1	$Type_1$
Case1		T_2	Ev_3	$Type_1$
Case1		T_3	Ev_7	$Type_1$
Case2		T_4	Ev_3	$Type_1$
……				

To analyze the table we sort it by *Case, DateTime* and create variables *Previous Event* (*PrEv*) and *Time between Events (T)*:

Table 3. Sorted data

	Case	DateTime	Event	Type	Time	PrEv
Case1		T_1	Ev_1	$Type_1$	0	0
Case1		T_2	Ev_3	$Type_1$	T_2-T_1	E_1
Case1		T_3	Ev_7	$Type_1$	T_3-T_2	E_3
Case2		T_4	Ev_3	$Type_1$	0	0
……						

In terms of OLAP[1] we have multidimensional situation

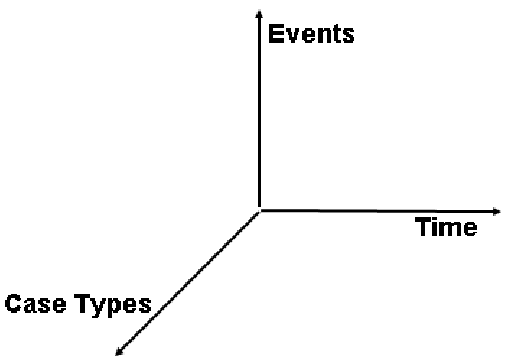


Figure 1. OLAP dimensions.

where dimension "Case Type" can also be compound of dimensions "*HW_Platform*", "*Product*", Geographic variables, "*Customer*" and so on.

2.1 Longitudinal Analysis

For longitudinal analysis we choose time between specified events , e.g. TTR = Time-To-Resolve = DateTime(Ev=CaseClosed) – DateTime(Ev=CaseCreated), choose date interval (usually one week or one month) and plot mean value of TTR by each time interval with Upper and Lower Control limits for this variable. This type of chart is standard in statistical process control (SPC) and we will not discuss it in detail:

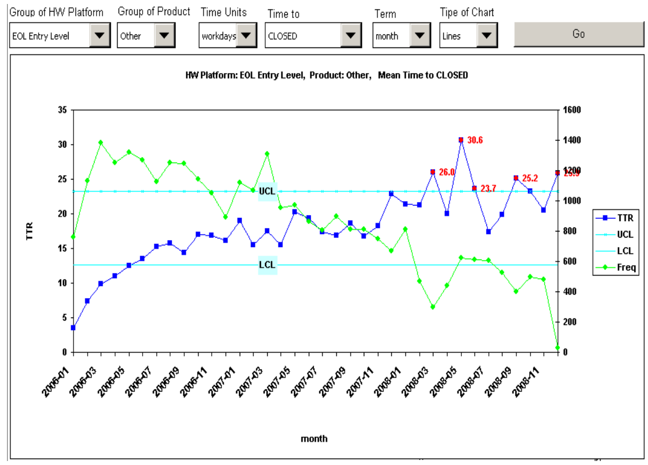


Figure 2. Screenshot of SPC Chart.

Green line (decreasing) shows Frequencies and blue line (increasing) – mean TTR. We plotted the simplest time-independent form of upper and Lower Control Limits. The constant limits are definitely not good for this specific case when we see strong time dependence of frequency that is decreasing in time because of EOL product that affects growth of volatility, but modification of SPC chart by choosing of variable control limits are out of scope of this work.

Above this chart are controls for choosing Type variables - group of HW Platform and Group of Product, Time Units – calendar or workdays, Event, term of time aggregation – week or month.

2.2 Cross-sectional Analysis for Sequence of Events

For cross-sectional analysis of quality of service we aggregate the data in Table 3 calculating count and average through Case and obtain two aggregating variables: *Frequency* (or Count) and Mean Time Between Events (MTBE, TBE) *Time* that is average of *Time* in Table 3 :

Table 4. Aggregated data

Event	PrEv	Freq	Time	Type
Ev_1	Ev_1	Fr_{11}	T_{11}	$Type_1$
Ev_1	Ev_2	Fr_{12}	T_{12}	$Type_1$
…				
Ev_2	Ev_1	Fr_{21}	T_{21}	$Type_1$
…				

During data aggregation from Table 1, instead of mean(T) we could use another aggregating function, e.g. mean($1/T$) or Scale(T) = exp(mean(ln(T))). The latter makes sense because the distribution of time between events could be Weibull rather than normal. We will discuss this choice of aggregating function later.

Now transform the Table 4 to two "wide" (or pivot) tables:

Table 5. Pivot table for *Frequency*

PrEv \ Ev	Ev_1	Ev_2	Ev_3	…
0	Fr_{01}	Fr_{02}	Fr_{03}	
Ev_1	Fr_{11}	Fr_{12}	Fr_{13}	
Ev_2				
…				

and

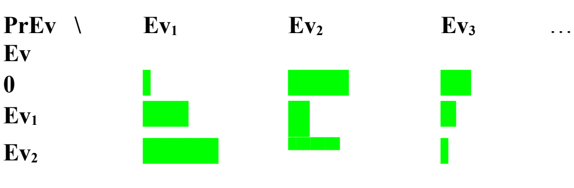
Table 6. Pivot table for Time

PrEv \ Ev	Ev_1	Ev_2	Ev_3	…
0	T_{01}	T_{02}	T_{03}	
Ev_1	T_{11}	T_{12}	T_{13}	
Ev_2				
…				

The traditional way of visualizing these two tables – "Pivot Chart" – creates two stacked bar charts, and we should match elements of these two charts to identify business important cases, because both Frequency and Time are important.

The simplest way to improve the pivot charts to visualize these two tables is to put in the cells of the table bars with width proportional to *Time* and length proportional to *Frequency*, that we named a "Matrix Bar Chart":

Table 7. Matrix Bar Chart for *Freq* and *Time* (2D)



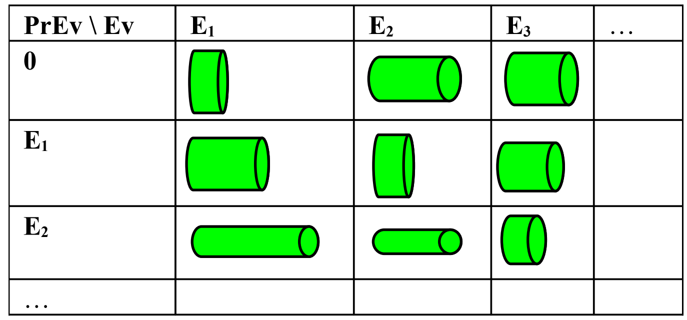
In this table the rows show frequency and average time of transactions following events PrEv and the columns show transactions that led to events Ev.

During data aggregation from Table 3 we could use the same type of chart but length of rectangle could be proportional mean(1/T) or Scale(T) =exp(mean(ln(T))). The latter makes sense because the distribution of time between events could be Weibull rather than normal.

We prefer to plot length of bars proportional mean (T) rather than scale(T) because sometimes lost for servicing company is proportional to time of service multiplied number of cases; in such situation areas of rectangles (bars) are proportional to dollar amount of loss related to these transactions, so just a short glance at the chart shows which process creates the majority of issues for the company.

Usually Frequencies are distributed in wide range of values, and more convenient to plot 3D bars with radius proportional to square root of frequency and plot the chart in 3D form:

Table 8. Matrix Bar Chart for Freq and Time (3D)



In 3D representation volume of each bar is proportional to dollar amount of loss related to these transactions.

One disadvantage of this method is that each event is presented in the table twice: in raw header as *Previous Event* (*PrEv*) and in a column header as *Event* (*Ev*).

To visualize this table without doubling the events, we present events as circles or other figures (e.g. "houses") with area proportional frequency of the events and represent frequency F_{12} and Time T_{12} as arrow (or bar or petal) from Ev_1 to Ev_2 with width proportional to F_{12} and length proportional T_{12}, color of the arrow is the same as the color of circle Ev_2:

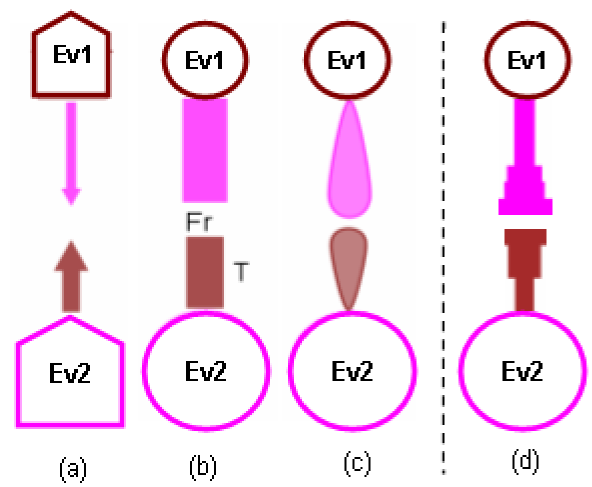


Figure 3. Four variants of visual representation: arrows, bars, petals and towers.

We can choose positions of the circles arbitrarily; the simplest case is to put it on a big circle where all event circles "can see" each other. We use the order of event circles by increasing mean time from Event 0 (so the most petals are directed clockwise):

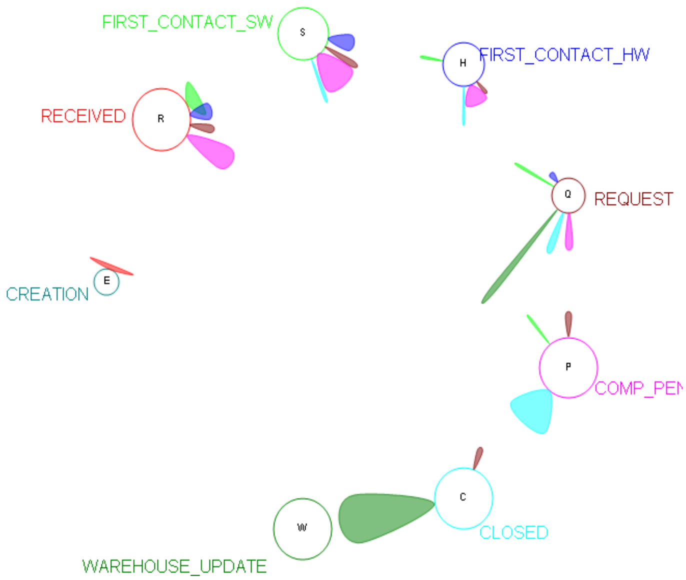


Figure 4. Screenshot of "Flower bed" chart

In the Flower Bed chart areas of petals again are proportional to dollar amount of loss related to these transactions, so just a short glance at the chart shows which process creates the majority of problems for the company: wide petals indicate business processes that happen frequently, long petals indicate business processes that take long time, and the most important business task is to optimize processes that are both long and wide.

In our special case we did not consider the possibility that an event can follow itself, which can be expected in many other real-world process-domains (for e.g., opening checking account followed by opening another checking account). The visualization technique itself has the power to show this (a purple circle can also have a purple petal that could be plotted out of center).

We named the chart "flower bed" chart.
Another alternative could be to use standard techniques for weighted multidigraph visualization [3], but we think our "flower bed" chart is easier for interpretation and visual perception.

To increase amount of information presented by the chart, instead of bars or petals we can draw more complicated figures ("Tower Charts") reflecting not only mean time between events but also distribution of the time. Usual histograms or violin plots can not be used to present distribution of time because size of the figures are not proportional to business importance ($$) .

We show creation of Tower Chart on simple example when for specific combination of (Ev, PrEv) we observe the sequence of N =5 TBE: 1, 3, 7, 1, 3 time units.

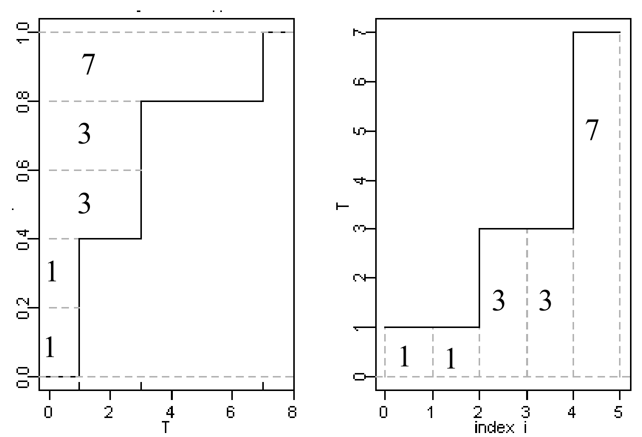


Figure 5. Empirical CDF (left) and sorted sequence (right) of TBE = {1, 1, 3, 3, 7}.

Plot of sorted TBE (right chart) is stretched empirical quantile function Q(p) that is inversed empirical CDF (ECDF):

$$f(i) = Q(i / N) = ECDF^{-1}(i / N)$$

It is obviously from comparison of area under f(i) and area left of ECDF. If 1 case * 1 day costs $1, then area under f(i) is exactly equal to business importance (dollar amount). More convenient to use symmetrical chart joining increasing and decreasing sequences of TBE:

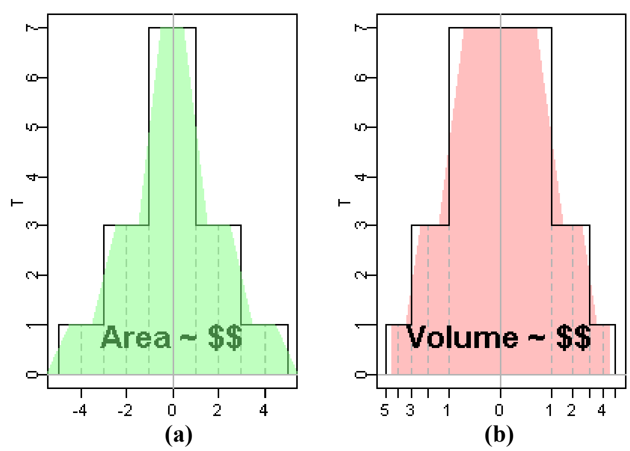


Figure 6. Tower Charts: Symmetrized sorted sequence of TBE with the same axes X as at Fig 5 (a) and with compressed axes X : $x = \sqrt{i}$ **(b) .**

We can use (a) stepwise (solid line) function, or (b) smoothed border of shaded area that is related to empirical quantile function as we mentioned above. If we rotate this lines around vertical axis Oy, then we get solid of revolution ("3D Tower"):

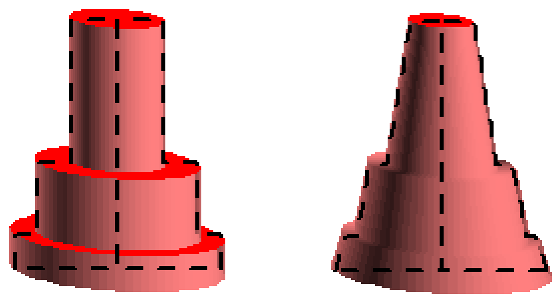


Figure 7. 3D Towers – result of revolution of stepwise 2D tower and shaded area of Fig. 6-b.

Volume of the solid of the 3D Tower again is exactly equal to business importance (dollar amount), area of base is proportional to total number of cases and height – to max(TBE). The left (a) tower consists of three cylinder rings: the internal one has height = 7 and area=1; the middle ring has height = 3 and area=2; the external ring has height = 1 and area = 2. For simplicity we will not plot on Flower Bed Chart 3D figure, but only it's section (contour) that is drawn by dashed line in Fig.7 or solid line in Fig.6:

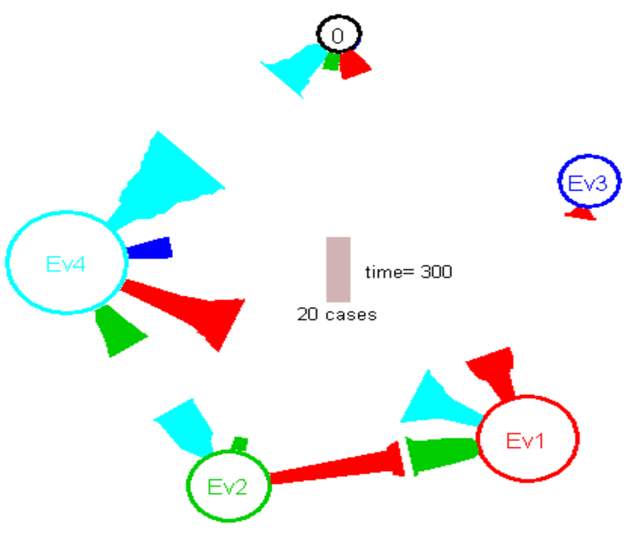


Figure 8. Flower Bed Chart with "tower" petals

In the Flower Bed Chart at Fig 6 we put in the center grey "scale bar" and used petals directed from each event circles to Event-0 and colored at the same color as the event circle to represent cases when the event was following by another event of the same type. This version of Flower Bed Chart allows easy identify outliers and other anomalies in distribution of TBE.

The Fig.5 shows evolution of chart elements to represent more information:

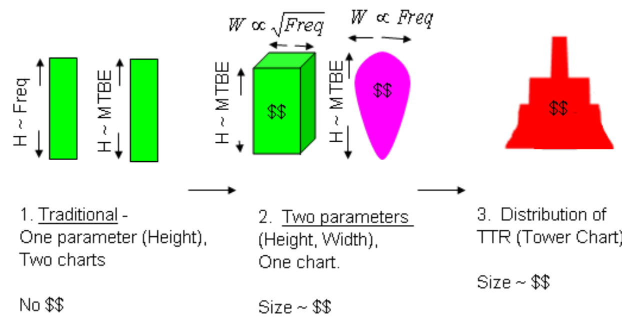


Figure 9. Evolution of chart elements.

Some additional information can be reflected by position of event circles as in widely used bubble charts.

We have to create the "flower bed" chart (Fig. 4, 8) for each Type of cases to compare quality of service between different Types.

2.3 Cross-sectional Analysis for Types of Cases.

For comparison of TTR and frequencies between Types we use the same graphic representation as in Table 4d, but instead of events rows and columns of the table can correspond to combination of two *Type* variables:

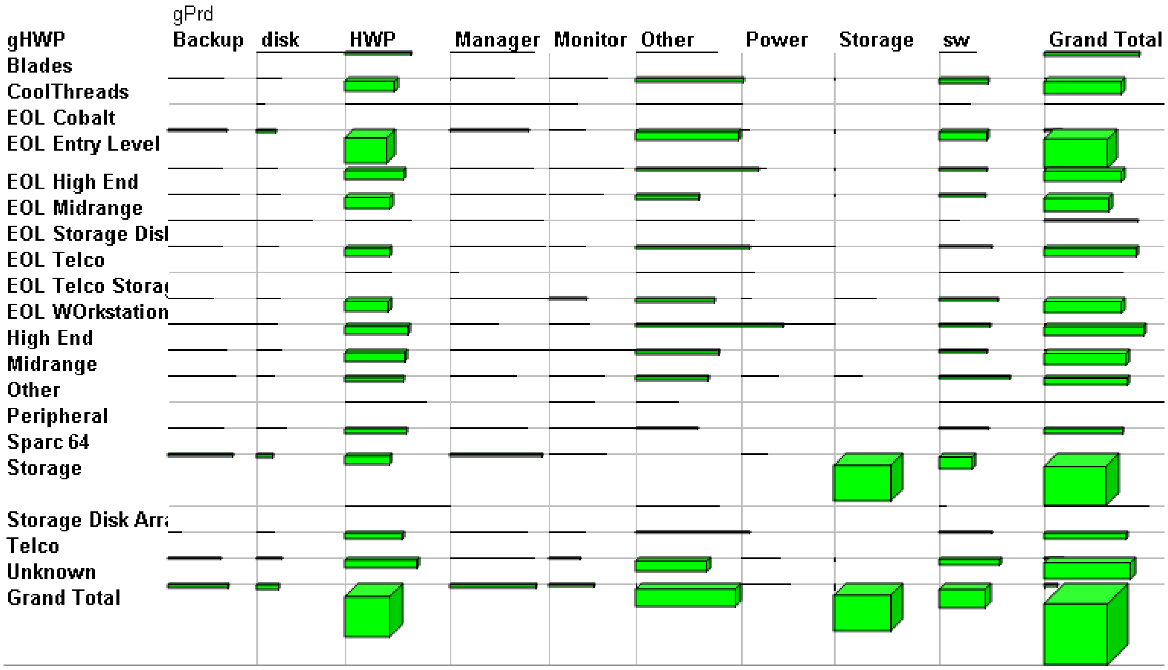


Figure 10. Matrix Bar Chart for Type variables gHWP – group of HW Platform and gPrd – Group of Product.

Again, if we suppose one case in one day costs $1, then total cost of service is proportional to volume of the bars (cuboid) in Grand Total- Grand Total cell that ids in right-down corner of the table, that equal sum of volumes (or $ amounts) of cuboids in Grand Total Row or Grand Total column that represent cost allocated to specific HWP or Product, and each of Grand total volume equal sum of cuboids volumes (or $ amounts) located in proper row or column.

More accurately, instead of "one case in one day costs $1" we could use cost matrix taking in account dependence of cost on specific HWP and Product, and plot volume of each cuboid proportional to the cost. The same approach could be applied to Flower Bed Chart.

As in case of Flower Bed Chart, if we need to reflect more information about distribution of TTR than mean and frequency, we can use tower charts instead of bars.

3 Conclusion.

The Dashboard for Visualization of Statistics Client Service Events consists from combination of these three types of charts – SPC chart for longitudinal presentation of data (Fig.), Matrix Bar Chart for Types of cases (Fig. 10), and Flower Bed Chart (Fig.4, 8), for Time between Events representation. These three types of charts correspond to tree OLAP dimensions shown on Figure 1:

Figure 11. Correspondence between three type of charts and OLAP dimensions

The Flower Bed and Matrix Bar Charts could be plotted for specific interval of time – specific week, month or year.

All Elements of these three charts are clickable and call transition from one chart to others and to rolling data down to detailed lists of individual cases. This combination gives opportunity immediately identify the most business important cases.

We illustrated our new methods of visualization using examples from a technical customer service, but these methods are applicable for any customer service – help desks, stores, doctor offices, banks, as well as for web analytics, logistic and other areas requiring analysis of statistics of events.

The R code to plot Flower Bed and Tower Charts can be downloaded from www.statvis.com or www.alex.zolot.us .

4 ACKNOWLEDGMENTS

Author would like to express his gratitude to Dr. Victor Zurkowski, Richard Castro, Dr. David Trindade, Dr. Jeff Glosup, and Dr. Paul Kuckein for fruitful discussion.

5 REFERENCES

[1] Kesaraporn Techapichetvanich and Amitava Datta, Interactive Visualization for OLAP, ICCSA 2005, www.springerlink.com/content/wxwdvwcp5p0qjlng

[2] Andreas S. Maniatis, Advanced Visualization for OLAP, Proceedings of the 6th ACM international workshop on Data warehousing and OLAP, 2003, pp: 9 - 16

[3] Graphviz - Graph Visualization Software Available: www.graphviz.org/Gallery.php