

Cell Suppression In SAS® Visual Analytics: A Primer

Marc Flore, Institute for Health Policy and Practice, University of New Hampshire

ABSTRACT

In healthcare and other fields, the importance of cell suppression as a means to avoid unintended disclosure or identification of Protected Health Information (PHI) or any other sensitive data has grown as we move toward dynamic query systems and reports. Organizations such as the Centers for Medicare & Medicaid Services (CMS), the National Center for Health Statistics (NCHS), and the Privacy Technical Assistance Center (PTAC) have outlined best practices to help researchers, analysts, report writers, and others avoid unintended disclosure for privacy reasons and to maintain statistical validity. Cell suppression is a crucial consideration during report design and can be a substantial hurdle in the dissemination of information. Often, the goal is to display as much data as possible without enabling the identification of individuals and while maintaining statistical validity. When designing reports using SAS® Visual Analytics, achieving suppression can be handled multiple ways. One way is to suppress the data before loading it into the SAS® LASR™ Analytic server. This has the drawback that a user cannot take full advantage of the dynamic filtering and aggregation available with SAS Visual Analytics. Another method is to create formulas that govern how SAS Visual Analytics displays cells within a table (crosstab) or bars within a chart. The logic can be complex and can meet a variety of needs. This presentation walks through examples of the latter methodology, namely, the creation of suppression formulas and how to apply them to report objects.

INTRODUCTION

Report suppression is a necessity in many environments and for a variety of reasons. In other environments it is a critical consideration. Suppression protects patient and customer identity, maintains statistical validity and allows for greater dissemination of information. The purpose of this paper is to provide examples of suppression in SAS® Visual Analytics using a real-world report suite developed for a wide audience interested in viewing health care claims data.

There are many organizations (governmental and otherwise) offering suppression guidelines. Some guidelines, such as CMS' current cell size suppression policy are widely publicized while others may be internal to an organization. The examples presented in this paper may be easily modified to meet many varied needs.

SAS® Visual Analytics does not, to the author's knowledge and as of the writing of this paper, possess built-in suppression. However, there are tools available to report-writers that will allow cells and other objects (e.g. visualizations) to be suppressed. This paper will show one methodology for accomplishing suppression though there may be others. The author has tested these methods using versions 6.4 and 7.2 and has no reason to believe that other versions will function differently.

DEFINING THE PROBLEM

Using a cell suppression rule where cells less than 11 (i.e. 10 or fewer) are suppressed, row three in Table 1 would be suppressed while rows one and two would be displayed (row four represents the total). Whether row four should be displayed is an open question that depends upon the suppression rule and the preferences of the report-writer and her colleagues. Care must be taken when displaying totals and sub-totals as they may be used to reconstruct data. Let us say that only row 3 is suppressed as is the case in Table 2. It is quite easy to deduce that the contents of the suppressed cell is the number nine ($40-12-19=9$).

Row	Number of Patients	Age Group
1	12	0-17
2	19	18-64
3	9	65+
4	40	0+

Table 1. Sample Data, Unsuppressed

Row	Number of Patients	Age Group
1	12	0-17
2	19	18-64
3	*	65+
4	40	0+

Table 2. Sample Data, Suppressed

When displaying a table in SAS® Visual Analytics a report-writer may be able to browse all of the data (as is the case in Tables 1 and 2) to determine if small cells are present and take appropriate action. More often, the data is too large or complex to browse. And when employing the dynamic features of SAS® Visual Analytics such as dynamic filtering, browsing the data becomes onerous. For example, imagine that a checkbox style filter for age grouping is instituted allowing the report-viewer to combine age groupings in any combination she wishes. As the number of age groupings increases, there arises a combinatorial explosion. That is, the number of combinations of age groupings quickly becomes quite large. The number of combinations of age groupings in Table 1 is $n!=3!=6$. At four age groupings grows to $n!=4!=24$ combinations and at five age groupings $n!=5!=120$ combinations. A report-writer is unlikely to check all 120 combinations for small sample sizes.

IMPLEMENTING THE SOLUTION

The solution will depend upon the structure of the data as well as the constraints and aims of the report-writer. This paper walks through a few simplified examples to illustrate suppression in SAS® Visual Analytics. These techniques can be modified to fill many needs. Please refer to the SAS® Visual Analytics User's Guide for your version of SAS® Visual Analytics as these directions may vary for your installation.

DISAGGREGATED SUPPRESSION

If the desired report does not incorporate dynamic filtering then the task is simplified. The data may be suppressed before being loaded to the SAS® LASR™ Analytic server. This can be accomplished using base SAS®, for example and is beyond the scope of this paper. Another solution is to implement the suppression in SAS® Visual Analytics using calculated items. To create a calculated item in SAS® Visual Analytics, select Data → New Calculated Item (Figure 1). After naming the Calculated Item, create a formula by dragging and dropping Operators (Figure 2) and Data Items (Figure 3) to the scratch area (Figure 4) in the appropriate places, manually enter “10” and “.” (missing), name the Calculated Item and click “OK”.

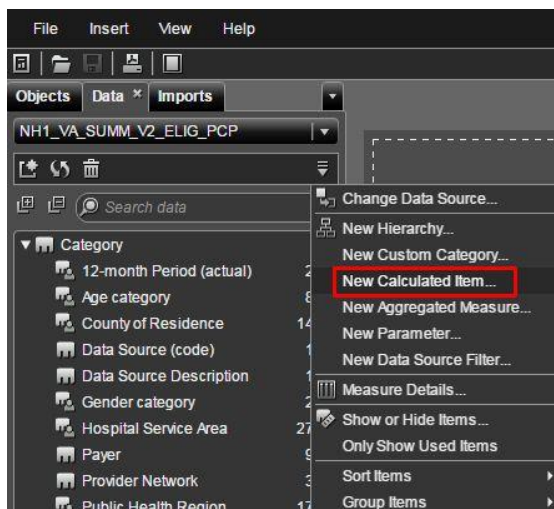


Figure 1 Creating a New Calculated Item

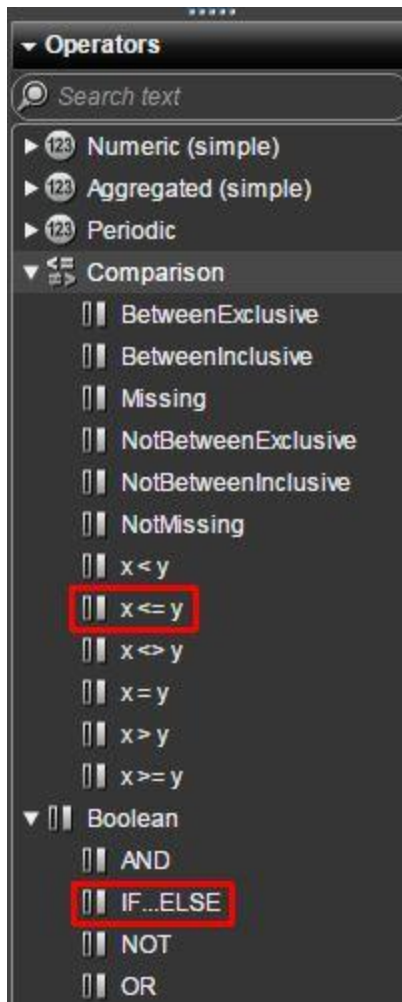


Figure 2. Operators

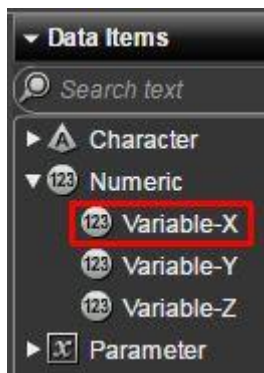


Figure 3. Data Items

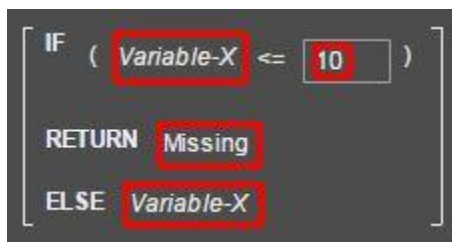


Figure 4. Calculated Item Scratch Area

At this point, the calculated item is available under the “measure” section of the data tab (Figure 5). This data item is now available to use in report-building. It may be added to tables and visualizations just as any other data item. It will suppress when the suppression criteria is triggered (in this example, when Variable-X is 10 or less).

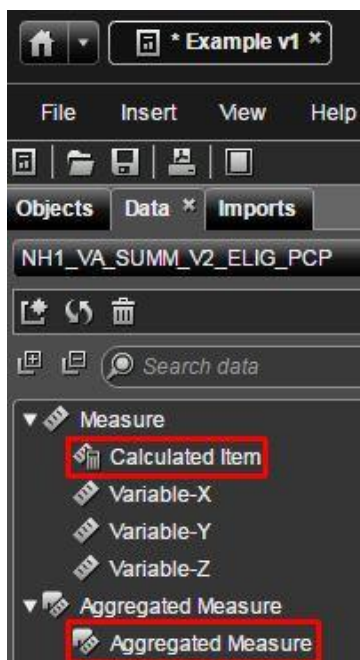


Figure 5. Measure List

AGGREGATED SUPPRESSION

If the desired report will utilize aggregation the solution is to create an Aggregated Measure. This process is very similar to that outlined above (the creation of a calculated item). Begins by selecting Data → New Aggregated Measure (Figure 6). After naming the Aggregated Measure, create a formula by dragging and dropping Operators (Figure 2) and Data Items (Figure 3) to the scratch area (Figure 4Figure 7) in the appropriate places, manually enter “10” and “.” (missing), name the Aggregated Measure and click “OK”.

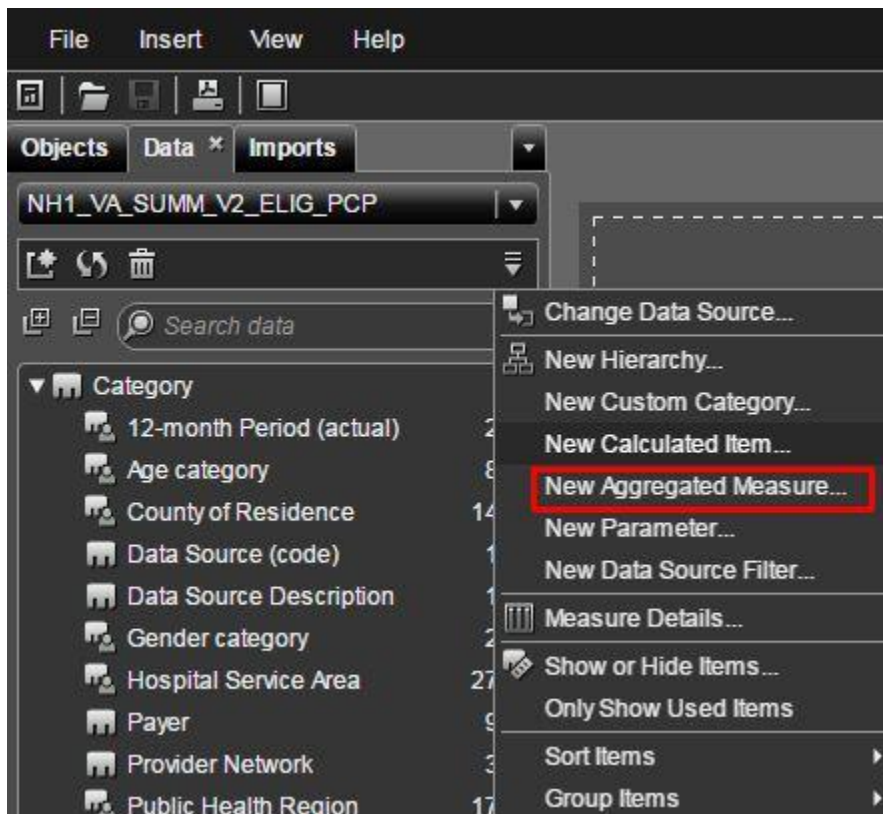


Figure 6. New Aggregated Measure



Figure 7. Aggregated Measure Scratch Area

At this point, the Aggregated Measure is available under the “Aggregated Measure” section of the data tab (Figure 5). This data item is now available to use in report-building. It may be added to tables and visualizations just as any other data item. And when coupled with dynamic filtering, it will suppress when the suppression criteria is triggered (in this example, when Variable-X is 10 or less).

MULTIPLE CONDITIONS AND INTERDEPENDENCIES

Suppression rules may be much more complicated than the above examples. The contents of one cell may be suppressed based on the contents of another and multiple conditions may be combined. Let us say that you are tasked with suppressing Variable-Z when either the minimum value of Variable-X is less than 44 or when Variable-X is greater than Variable-Y. To accomplish this, simply add the Boolean operator “OR”, to the “IF” portion of the condition, and replace the rest of the formula to match Figure 8.

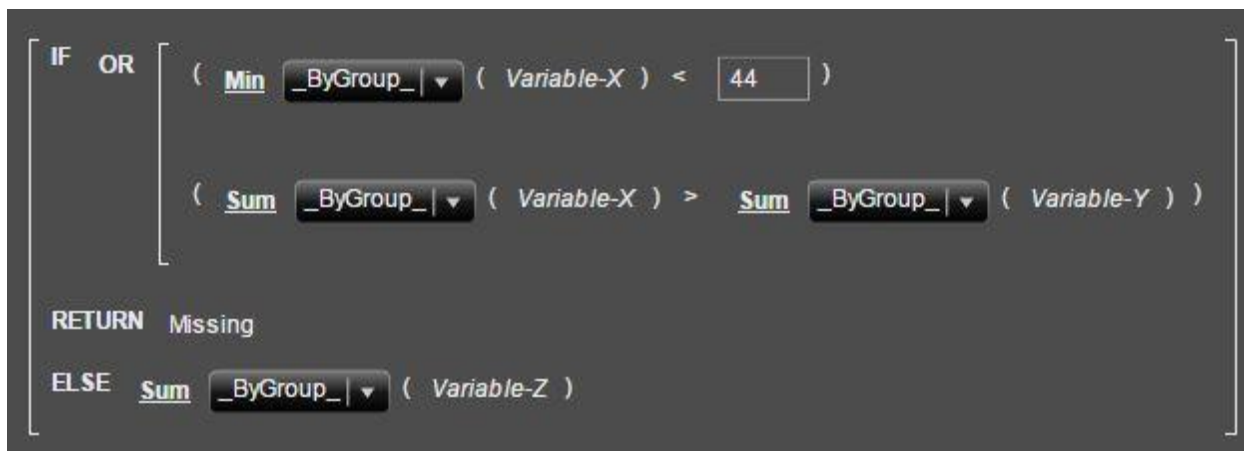


Figure 8. Multiple Conditions & Interdependencies

AGGREGATED MEASURES VS. CALCULATED ITEMS

The benefit of utilizing Aggregated Measures over Calculated Items is that Aggregated Measures trigger suppression after aggregation. If the report does not have aggregation or if it is desired that disaggregated items are suppressed then Calculated Items are appropriate.

CONCLUSION

While SAS® Visual Analytics does not feature built-in suppression, the tools available to the report-writer are flexible and powerful, allowing suppression to be achieved in a variety of contexts. This paper has provided several examples for illustrative purposes but these are only a fraction of what is possible. Try these out for yourself to become familiar with the process, then explore on your own. The author would be interested in hearing your stories, both successes and failures. You will find Contact Information below.

REFERENCES

CMS accessed March 18, 2016. <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Dataand-Systems/Privacy/Researchers.html>

ACKNOWLEDGMENTS

The author would like to acknowledge his colleagues at the Institute for Health Policy and Practice and SAS® for their help creating and testing suppression. The author would also like to thank these colleagues for their encouragement to write this paper.

RECOMMENDED READING

- SAS® *Visual Analytics User's Guide*

CONTACT INFORMATION

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

Marc Flore
 Institute for Health Policy and Practice
 University of New Hampshire
 4 Library Way
 Durham, NH 03824
 Marc.Flore@unh.edu
<http://nhaccountablecare.org/>
<http://chhs.unh.edu/faculty-member/marc-flore>

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.