The Census Bureau's New Cell Suppression System[1][2]

Philip Steel
U.S. Census Bureau

Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference

Abstract: Cell suppression is one of the oldest methodologies for protecting establishment or company based data. The U.S. Census Bureau is in the process of replacing its network flow based cell suppression system with a new, Linear Programming based system. We describe that system and compare its performance to the old system. We also look at the question of whether the LP based system is as responsive to weighting as the old network flow based system.

1 Introduction

We will begin with some background on the application: a description of the system and the Operations Research (OR) strategy we employ. In sections 3 and 4, we take a look on the relevant aspects of how we organize and process our data. The section 5 we describe our Linear Programing (LP) model. The main results are in section 6 comparing results of the network program with LP and explore how responsive the LP is to weight adjustment.

The cell suppression re-development project was prompted by aging software and coincident with the impending retirement of the last of the people who developed and ran that system. One of the goals for the project was to devise a LP system that could handle extremely large, sparse tables in a reasonable amount of time. We proceeded more or less from scratch, developing a prototype research tool then a specification for the production software, which is currently in its first version. The work presented here are part of a three-year effort by our cell suppression Research and Methodology group: Jim Fagan, Paul Massell, Richard Moore, Jr, John Slanta, the author and Bei Wang.

The main methodological difference between the old system and the new system is the adoption of LP rather than network flow. The LP solution avoids some of the problems associated with network: it is valid for higher dimensional tables and can accommodate linked tables without losing its formal guarantees for the safeness of the solution. The research goal was to avoid backtracking, which was a bookkeeping exercise to extend the network solution to higher dimensions and across linked tables. Backtracking did not have a formal guarantee of a valid solution and created very complex and error prone processing. The LP methodology does have scaling problems and the US Economic Census tables are among the largest to which cell suppression is applied.

2 OR strategy and company protection

The system is set up in three stages. The purpose of the first stage, which we call the base pass, is to find a protection pattern that guarantees that suppressed cells cannot be estimated within *p* percent of the actual value by table-non-participants. We use a symmetric protection interval[3]. We set the capacity, the upper bound on flow, to be the cell value and address protection problems arising from companies contributing to more than one cell at a later stage. The first stage applies the solution strategy to a queue of all the designated primary suppressions (Ps). After all the primaries are processed, the second stage examines the resulting pattern for aggregate protection problems and creates a new processing queue of "supercells". That is, we are looking to

[1] This article reports the results of research and analysis undertaken by U.S. Census Bureau staff. It has undergone a more limited review than official U.S. Census Bureau publications. It is released to inform interested parties of research and to encourage discussion. Any views expressed are those of the author and not necessarily those of the U.S. Census Bureau.
[2] Portions of this paper also appear in "Re-development of the Cell Suppression Methodology at the US Census Bureau" Steel, et al., presented at the UNECE Work Session in Ottawa.
[3] Allowing asymmetric protection leads to a different model, see "Models and Alogorithms for Optimizing Cell Suppression in Tabular Data with Linear Constraints" by Fischetti Salazar.

protect aggregates where the union of suppressed cells fails the p-percent rule because the selected complementary cells (Cs) from the first stage have contribution from one or both of the top two companies of the primary. In the third stage, the model is modified to accommodate the supercells and the solution strategy is applied to each supercell with an adjustment of capacity to reflect a potential suppression's ability to protect the supercell. These processes together produce a protected table.

The solution strategy is inherited from the old system. The LP problem is set up, optimized to minimize the cost times value, then re-optimized with a different cost function over the first solution. The re-optimization frequently finds a smaller solution by collapsing flows split under the initial weighting. The initial cost is a weighted value where already suppressed cells have weight 0. The second round cost is 1/(1+value) and the flow in unsuppressed first round cells is bound to 0.

The old system excelled in manipulating the costs in the first optimization, it included a weighting system that was tailored to analysts' evaluation of what was important in the tables being produced. We have retained and allowed for elaboration of that weighting. The old system also included a recalculation (on the fly) of capacity, the upper bound on how much "flow" a potential complement could carry. This was an accounting of company protection specific to each primary. We have deferred implementation of the capacity modification in the base pass due to concerns over speed and rely solely on stage 3 to apply company protection. This may lead to some oversuppression.

3 Linked data and table groups

The system addresses linked data. That is, we want to process data with a model that enforces all the additive relations (a=b+c+d) in the publication. That means that tables that share a margin must be processed together. In terms of the relations: if a pair of relations has a variable in common, then they belong in the same model. We call our processing unit, the cells and the relations for those cells, a table group. We organize our relation files by dimension; we typically have a row relation file, a column relation file and a single level relation.

The optimum table groups can be found by constructing the graph where each relation is a node and two nodes have an edge between them if they have at least one shared variable. In this graphical representation, the table groups are the cross product of the maximally connected subgraphs in each dimension. Our tables often include industry and product coding; these are hierarchies so their graphs are (connected) trees. Geographic relations are often completely connected as well. In practice, determining table groups usually means divvying up the relations in a single dimension.

If a publication being processed breaks up into distinct table groups, these groups can be run in parallel. The solutions for each are independent.

4 Duplication

One of the idiosyncrasies of our input data is a massive amount of duplication. Part of this arises from a publication convention that extends the detail of some industry codes. This in turn may be the result of having a data collection system that predates the current coding scheme (NAICS). Multiple geographic schemas also introduce duplication: a unit may appear as a county in one scheme and a metropolitan area "part" in another. Ownership of the unduplication process has been problematic … it doesn't really fit with cell suppression proper. On the other hand, it is desirable to have the prepared cell suppression input data correspond exactly to the publication data-which is duplicated. Fortunately, the unduplication process is compatible with the representation of the data as a graph and we have added it as a pre-process to cell suppression. Duplication enters the relationship files as a simple global equality.

5 The LP Model and the weighting system

We use a standard LP model with some modifications to enhance its performance on large sparse datasets. The model consists of "up" and "down" variables for each positive cell constrained by the additive relations in the table at the cell level. The variables are bounded by their capacity; usually this is the cell value. The problem of finding secondary suppressions is done sequentially for each primary suppression, adding a set of constraints for that primary that guarantee that the flow through the primary meets the protection requirement. A major innovation for the system was finding a method that determines when the existing accumulated pattern is sufficient to protect primaries that have not been processed yet. This feature is valid for strictly additive tables.

Note that we are treating negative value data as 0 (no flow variables) except to qualify the relation. This may cause an infeasibility. If the negative value requires protection it must currently be done outside the system.

Both the old system and the new system can statistically modify the pattern using a weight on the cell cost. That is, by increasing the weight on cells of a given type, the percentage of those cell used in the final pattern can be reduced. This comes at a cost elsewhere in the table. It is also effective to cheapen a class of cells (e.g. cells with high variance) to invite the pattern to make more extensive use of those cells. This increases the number of cells used but allows for a more efficient pattern on cells that are "important". This manipulation of the suppression pattern has been used very sparingly, but to great effect. Cheapening seems to be more effective than adding expense, perhaps because when one has more suppressed cells there are more patterns to choose from. One of the driving research questions has been whether the switch to LP will still allow us to produce these desirable effects.

6 LP weighting for a Service's Geographic Area Series file

We did extensive testing with the data set for disclosure group 7 sector 71. Disclosure group 7 consists of the data for the states that contribute to the Boston-Worcester-Providence Combined Statistical Area (CSA). Sector 71 data has a relatively small NAICS (North American Industry Classification System) hierarchy. The columns are geographic units that are linked and non-hierarchical. The third dimension is simple: the total is the sum of taxable and non-taxable company receipts. These data are relatively small for publications in our Economic Census' Geographic Area Series. The initial input consists of 42,198 cells, which reduces to 17,467 after unduplication. 11,037 of those are primary suppressions. The unduplicated Cartesian space it sits in has 88,908 cells. To compare patterns produced under different methodologies and, within the LP method, different weighting we look at both at complementary cell counts and the value of the complementary cells that are suppressed.

Note also that the benchmark network numbers represent a pattern that has some under-suppression. Seventeen of the primaries were inadequately protected, in the sense that one is able to estimate the suppressed value within p%. A few of those at the 1/2p level. None of the under-suppressions were exact. Five additional complementary cells are needed for the network solution to eliminate the under-suppression.

6.1 Limitations to weight adjustment.

The initial feedback on the LP rerun of the 2007 data pointed to two features: Rhode Island being over-suppressed and too many "A" cells (taxable plus non-taxable) being taken.

The RI problem was relatively simple. The CSA piece and the State level are actually the same in this instance (Providence) and we normally treat these levels very differently. CSA pieces are down-weighted in order to give a preference for the component Metro cells. That is, the weighting is used to force a complement in the total rather than the natural complement in the metro level (e.g. when the balance of CSA is sensitive). We found a hardcoded reweighting fix in the old network program that addressed this issue. The fix reset the weight for the Providence CSA piece specifically. We re-implemented that fix for LP and obtained the same effect.

The "too many A cells" problem is somewhat deeper. It reflects a subjective trade-off. We would rather see that total than a component even if it leads to a somewhat less efficient pattern. It is easy to remedy this: increase the cost of A cells. The hard part is determining where the other end of the trade-off is. The cost for A cells was variously set at 1.25, 1.5, 2, 5, and 10 times the value of the cell. This is compared with the network solution and the default weighting in tables 1a and 1b. Table 1a gives the raw number of cells suppressed and the ratio of published cells to "C cells" for 7 different suppression patterns. Table 1b repeats the comparison under the value metric rather than count and omits the raw values. The ratio statistics make it easier to compare two columns consistently; bigger is better. "Y" represents a checkbox, "yes", tax exempt.

"Too many A cells" was only evident at the state level where the proportion of A to T to Y was clearly different for the network program and the LP with "A" cell weight equal to 1 (columns c1, rc1). Even a small weight increase corrects this problem. It should be noted that the raw counts are better for the LP even where the proportion of cells is off. The question of whether the analysts prefer adjusting the ratio of A-T-Y for only state level cells as was done in the network program or the global preference we used in this study has not been addressed.

Of course, as we increase the cost of certain cells, fewer of those cells get selected. That requires that cells be taken elsewhere. Where were those cells taken and how much value was taken? We did an exhaustive set of tabulations to determine the answers to those questions. In our judgement (subjective!), the hardest hit section of the table were the industry stubs. Table 2 shows the impact of favouring the 3rd dimension's marginal total has on the "1st" dimension's higher order marginal. In the patterns without the strong emphasis on "A" cells (r1 and r125), the LP runs are suppressing noticeably fewer cells. In the patterns where the weighting strongly emphasizes the "A" cells, performance on the NAICS cells has declined. In some cases important cases, for instance the two digit industry totals for Massachusetts, the LP ratios are worse than the network numbers. The tables for the value metric are comparable and have been omitted.

6.2 Comparing LP and network

The table 3 compares five different patterns. Two provide reference to the previous set: they are network and the light (1.25) adjustment on "A" cells. The columns labeled "met" represent a run with an additional bump (1.25) on the weight of geotypes 08 & 09 (MeSA, MiSA: metropolitan and micropolitan statistical areas). The columns labeled "metb" represent a run that builds on "met" but also adds a discount (0.02) on geotype 15 (non-MSA or balance of counties). The discount was another hardcoded adjustment in the network program.

The final column (depth) represents a completely different approach. The previous weight adjustments (bump on tax total, bump on metro, discount on balance) are abandoned and instead a weight adjustment of 1+1/depth is used. Depth is a measure of a cell's distance from a grandtotal. Margins are guaranteed to have smaller depth than interior cells. The deepest cells in this set of tables had depth 11.

The metb run represents the best LP weighting to date for this data. We examined the ratio of published to complementary suppressions broken out to the 3 full margins and 2 marginal sub-tables by both metrics. The full set of 10 tables is available upon request. The selected sub-tables were state by naics detail and (table 3) geotype by tax status. That solution improves on almost all important marginal cell types, except for the count of MeSA cells. The value in MeSA cells is improved, despite the decline in cell count. The metb run suppresses 87,465,688 in value where the network solution suppresses 117,866,115. It is also worth noting that an additional 53 2-digit NAICS totals and 26 additional state totals can be shown with the LP under the metb weighting.

7 Software and hardware

The prototype is an amalgam of SAS, FORTRAN, AMPL and CPLEX. The production version is c++ interacting with CPLEX through a graph object using the Boost c++ library. They are run in a red hat linux system with eight 2.93 GHz Intel Xeon processors and 60GB memory.

8 Conclusion

Some adjustments are not that intuitive … discounting balance geographies does more for the metro cells than increasing the weight on the metro cells. Refinement of the weighting scheme is mostly a matter of trial and error. On the other hand, I see no reason why the weighting on this table would not be successful for other 3d tables of this general type. The depth based weighting looks like a decent default, if time and personnel are not available to experiment with the weighting.

We have noted the "trade off" between preserving the tax status dimension's marginal total and NAICS margin. "Trade off" may be a poor choice of words: one can move a pattern from one part of a table to another, but that exchange can only increase the sum of cells suppressed (or on rare occasion, keep it the same). That is, it is always a losing trade in terms of unweighted value. We recommend not trying to push too hard for certain cell types or for too many cells; bad effects may be difficult to identify. A good strategy may be to accompany any up-weighting with some equivalent down-weighting. Then one can verify that the attempt to direct the trade off was successful, by checking the sums in the 3 categories (up-weighted value, value, down-weighted value) against an unweighted baseline. However, there is still no guarantee that some other categorization of cells would not show some unanticipated effect. The unpredictability in adjusting the weighting applies to both LP and network.

We have produced a methodology/application that allows for the protection of large tables in a reasonable amount of time, with considerably less over-suppression than the network flow system we are replacing. The new system still allows for constructive analyst intervention through a weighting system. There is no discernible reduction in the effectiveness of that system, despite the somewhat conflicting goals of reducing over-suppression and having flexibility in manipulation of the pattern. The LP system eliminated the under-suppression that we were able to detect in patterns produced with the network methodology. The LP system reduces over-suppression, particularly for those files that caused the old system to do a lot of backtracking. For those files, the LP system produces as much as 20% more published data.

References

Federal Committee on Statistical Methodology (1994) Statistical Policy Working Paper 22 (Revised 2005)- Report on Statistical Disclosure Limitation Methodology NTIS PB94-165305.

Jewett, R. Disclosure Analysis for the 1992 Economic Census: unpublished US Census Bureau documentation 1993.

Kirkendahl, N. and Sande, G. Comparison of Systems implementing Automated Cell Suppression for Economic Statistics. Journal of Official Statistics, Vol. 14, No. 4, 1998 pp. 513-535.

Massell, P. B. Using Linear Programming for Cell Suppression in Statistical Tables: Theory and Practice. Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001.

Giessing, S. A Survey on Software Packages for Automated Secondary Cell Suppression. Joint ECE/Eurostat Work Session on Statistical Data Confidentiality, March 1999.

Giessing, S. Protection of tables with negative values. CASC report on the ESSNet SDC website, (ca 2010).

Tambay, J-L. et al., Treatment of Aggregated Sensitive Cells. Statistics Canada working paper, 2007.

Steel,P, et al., Re-development of the Cell Suppression Methodology at the US Census Bureau. UNECE work session on statistical data confidentiality, October 2013 Ottawa

Steel, P “A report on LP weighting for a Service’s Geographic Area Series file” internal census document.

Table 1a

Ratio of Published Cell Count to C Cell Count for Geotype/Margins for Network, LP and Margin Weights 1 1.25 1.5 2 5 & 10

GEOTYPE	Tax	cnet	c1	c1.25	c1.5	c2	c5	c10	rcnet	rc1	rc1.25	rc1.5	rc2	rc5	rc10
state	A	75	72	62	62	60	52	52	2.15	2.28	2.81	2.81	2.93	3.54	3.54
state	T	70	54	54	54	54	50	54	1.84	2.69	2.69	2.69	2.69	2.98	2.69
state	Y	48	34	44	44	46	46	47	1.54	2.59	1.77	1.77	1.65	1.65	1.60
county	A	384	363	360	350	340	333	335	0.82	0.93	0.94	1.00	1.06	1.10	1.09
county	T	354	322	333	339	338	346	353	0.44	0.58	0.53	0.50	0.51	0.47	0.44
county	Y	237	225	228	227	231	231	233	0.53	0.61	0.59	0.59	0.57	0.57	0.55
place	A	680	609	606	595	579	571	562	1.15	1.40	1.41	1.45	1.52	1.56	1.60
place	T	788	691	721	752	772	793	807	0.46	0.67	0.60	0.53	0.49	0.45	0.43
place	Y	174	160	158	162	172	176	184	0.64	0.78	0.80	0.76	0.66	0.62	0.55
CSA	A	64	26	24	23	25	27	20	0.42	2.50	2.79	2.96	2.64	2.37	3.55
CSA	T	61	26	24	26	26	27	27	0.25	1.92	2.17	1.92	1.92	1.81	1.81
CSA	Y	40	24	24	23	24	24	24	0.30	1.17	1.17	1.26	1.17	1.17	1.17
MeSA	A	113	132	129	127	127	135	132	1.43	1.08	1.13	1.17	1.17	1.04	1.08
MeSA	T	111	119	118	119	119	132	130	0.88	0.76	0.77	0.76	0.76	0.58	0.61
MeSA	Y	89	83	80	83	83	84	83	0.64	0.76	0.83	0.76	0.76	0.74	0.76
MiSA	A	44	53	51	51	49	42	39	1.09	0.74	0.80	0.80	0.88	1.19	1.36
MiSA	T	53	48	51	51	51	51	51	0.25	0.38	0.29	0.29	0.29	0.29	0.29
MiSA	Y	31	28	30	30	32	32	30	0.48	0.64	0.53	0.53	0.44	0.44	0.53
MD	A	69	61	64	64	64	62	59	1.29	1.59	1.47	1.47	1.47	1.55	1.68
MD	T	70	60	70	70	70	73	75	0.87	1.18	0.87	0.87	0.87	0.79	0.75
MD	Y	57	47	53	53	53	54	56	0.58	0.91	0.70	0.70	0.70	0.67	0.61
nonMSA	A	28	13	13	13	13	15	13	0.68	2.62	2.62	2.62	2.62	2.13	2.62
nonMSA	T	25	16	16	20	20	20	20	0.16	0.81	0.81	0.45	0.45	0.45	0.45
nonMSA	Y	22	16	16	18	18	18	16	0.23	0.69	0.69	0.50	0.50	0.50	0.69

Table 1b
Ratio of Published Cell Value to C Cell Value for Geotype/Margins for Network, LP and Margin Weights 1 1.25 1.5 2 5 & 10

GEOTYPE	Tax	rvnet	rv1	rv125	rv15	rv2	rv5	rv10
state	A	7.83	8.05	9.22	9.22	9.27	10.10	10.10
state	T	6.68	8.09	7.97	7.97	7.97	8.65	8.52
state	Y	4.01	15.50	12.00	12.00	11.70	11.70	11.10
county	A	1.80	2.51	2.65	2.68	2.76	2.85	2.89
county	T	1.12	1.44	1.29	1.28	1.30	1.10	1.07
county	Y	1.07	1.20	1.24	1.24	1.24	1.16	1.11
place	A	1.07	1.52	1.53	1.59	1.67	1.69	1.70
place	T	0.31	0.47	0.45	0.41	0.39	0.34	0.31
place	Y	0.44	0.51	0.57	0.54	0.51	0.44	0.38
CSA	A	0.98	4.98	4.99	5.01	4.99	4.89	5.63
CSA	T	0.90	4.76	5.00	4.76	4.76	4.61	4.61
CSA	Y	0.39	3.38	3.38	3.56	3.38	3.38	3.38
MeSA	A	4.96	4.82	5.39	4.95	4.97	7.24	7.38
MeSA	T	2.36	2.45	2.47	2.51	2.51	2.09	2.31
MeSA	Y	0.91	1.30	1.41	1.31	1.31	1.40	1.40
MiSA	A	1.15	0.82	1.23	1.23	1.28	1.68	1.75
MiSA	T	0.47	0.78	0.31	0.31	0.31	0.31	0.31
MiSA	Y	0.80	0.98	0.76	0.76	0.62	0.62	0.73
MD	A	3.61	4.78	2.84	2.84	2.84	3.52	3.57
MD	T	2.18	2.69	1.26	1.26	1.26	1.34	0.94
MD	Y	0.95	1.08	0.92	0.92	0.92	0.92	0.82
nonMSA	A	0.83	4.98	4.98	4.98	4.98	4.62	5.13
nonMSA	T	0.17	1.26	1.26	1.04	1.04	1.04	1.04
nonMSA	Y	0.33	1.21	1.21	0.76	0.76	0.76	0.90

Table 2
Ratio of Published Cell Count to C Cell Count for State/Naicslevel for Network, LP and Margin Weights 1 1.25 1.5 2 5 & 10

rstate	ndig	ccount	c1	c125	c15	c2	c5	c10	rcnet	rc1	rc125	rc15	rc2	rc5	rc10
X	2	.	.	.	.	.	2	2	.	.	.	.	.	8.00	8.00
X	3	34	23	23	23	23	21	21	0.35	1.00	1.00	1.00	1.00	1.19	1.19
X	4	62	35	35	35	34	35	37	0.31	1.31	1.31	1.31	1.38	1.31	1.19
X	5	88	73	67	69	71	72	67	0.70	1.05	1.24	1.17	1.11	1.08	1.24
X	6	92	79	73	75	77	78	73	0.72	1.00	1.16	1.11	1.05	1.03	1.16
X	7	57	49	48	50	50	51	51	0.53	0.78	0.81	0.74	0.74	0.71	0.71
MA	**2**	**169**	**150**	**155**	**157**	**158**	**167**	**169**	**2.14**	**2.53**	**2.42**	**2.38**	**2.35**	**2.17**	**2.14**
MA	3	365	336	329	342	344	336	338	0.62	0.76	0.80	0.73	0.72	0.76	0.75
MA	4	437	402	402	408	410	425	424	0.47	0.60	0.60	0.58	0.57	0.51	0.52
MA	5	358	319	334	324	329	337	337	1.03	1.28	1.18	1.24	1.21	1.16	1.16
MA	6	358	315	328	318	323	331	331	1.03	1.30	1.21	1.28	1.24	1.19	1.19
MA	7	220	203	213	220	215	205	203	0.64	0.78	0.69	0.64	0.68	0.76	0.78
NH	**2**	**68**	**42**	**55**	**57**	**61**	**64**	**64**	**1.66**	**3.31**	**2.29**	**2.18**	**1.97**	**1.83**	**1.83**
NH	3	157	145	145	143	143	144	143	0.42	0.54	0.54	0.56	0.56	0.55	0.56
NH	4	163	156	159	154	154	154	154	0.31	0.37	0.34	0.38	0.38	0.38	0.38
NH	5	100	104	105	109	109	110	111	1.01	0.93	0.91	0.84	0.84	0.83	0.81
NH	6	104	104	105	109	109	110	111	0.91	0.91	0.90	0.83	0.83	0.81	0.79
NH	7	65	67	68	68	66	68	65	0.49	0.45	0.43	0.43	0.47	0.43	0.49
RI	**2**	**40**	**40**	**41**	**34**	**36**	**39**	**49**	**1.23**	**1.23**	**1.17**	**1.62**	**1.47**	**1.28**	**0.82**
RI	3	69	65	65	72	66	63	70	0.57	0.66	0.66	0.50	0.64	0.71	0.54
RI	4	84	77	75	79	74	74	80	0.38	0.51	0.55	0.47	0.57	0.57	0.45
RI	5	71	71	69	70	71	73	77	0.94	0.94	1.00	0.97	0.94	0.89	0.79
RI	6	73	73	71	72	73	75	79	0.86	0.86	0.92	0.89	0.86	0.81	0.72
RI	7	40	37	39	39	41	40	41	0.68	0.81	0.72	0.72	0.63	0.68	0.63
VT	**2**	**68**	**39**	**39**	**39**	**39**	**41**	**47**	**1.06**	**2.59**	**2.59**	**2.59**	**2.59**	**2.41**	**1.98**
VT	3	119	91	91	94	94	87	76	0.26	0.65	0.65	0.60	0.60	0.72	0.97
VT	4	103	75	77	79	83	79	71	0.41	0.93	0.88	0.84	0.75	0.84	1.04
VT	5	49	44	48	47	45	45	44	1.04	1.27	1.08	1.13	1.22	1.22	1.27
VT	6	47	44	48	47	45	45	44	1.09	1.23	1.04	1.09	1.18	1.18	1.23
VT	7	27	24	22	23	23	23	23	0.44	0.63	0.77	0.70	0.70	0.70	0.70

Table 3
C Counts and Ratios by Geotype and Tax Status

GEOTYPE	Tax	cnet	c125	cmet	cmetb	cdepth	rcnet	rc125	rcmet	rcmetb	rcdepth
state	A	75	62	72	69	58	2.15	2.81	2.28	2.42	3.07
state	T	70	54	54	55	50	1.84	2.69	2.69	2.62	2.98
state	Y	48	44	34	43	44	1.54	1.77	2.59	1.84	1.77
county	A	384	360	363	350	369	0.82	0.94	0.93	1.00	0.89
county	T	354	333	322	342	337	0.44	0.53	0.58	0.49	0.51
county	Y	237	228	225	225	220	0.53	0.59	0.61	0.61	0.65
place	A	680	606	609	597	615	1.15	1.41	1.40	1.44	1.37
place	T	788	721	691	718	711	0.46	0.60	0.67	0.60	0.62
place	Y	174	158	160	157	161	0.64	0.80	0.78	0.82	0.77
CSA	A	64	24	26	38	26	0.42	2.79	2.50	1.39	2.50
CSA	T	61	24	26	30	27	0.25	2.17	1.92	1.53	1.81
CSA	Y	40	24	24	26	24	0.30	1.17	1.17	1.00	1.17
MeSA	A	113	129	132	136	141	1.43	1.13	1.08	1.02	0.95
MeSA	T	111	118	119	127	129	0.88	0.77	0.76	0.65	0.62
MeSA	Y	89	80	83	81	77	0.64	0.83	0.76	0.80	0.90
MiSA	A	44	51	53	38	48	1.09	0.80	0.74	1.42	0.92
MiSA	T	53	51	48	46	51	0.25	0.29	0.38	0.43	0.29
MiSA	Y	31	30	28	25	30	0.48	0.53	0.64	0.84	0.53
MD	A	69	64	61	63	64	1.29	1.47	1.59	1.51	1.47
MD	T	70	70	60	72	70	0.87	0.87	1.18	0.82	0.87
MD	Y	57	53	47	54	52	0.58	0.70	0.91	0.67	0.73
nonMSA	A	28	13	13	25	15	0.68	2.62	2.62	0.88	2.13
nonMSA	T	25	16	16	26	18	0.16	0.81	0.81	0.12	0.61
nonMSA	Y	22	16	16	20	16	0.23	0.69	0.69	0.35	0.69