As the NRCS National Soils Database Manager, a common soils request I receive is to aggregate soil properties to a single value at the county level. Although this seems an odd request due to the variability of soils within a county, for many who model national programs they look for information to identify gross separations within their model or to identify patterns that may exist.

The Soil Data Mart database is the official source of soils information presented using [Web Soil Survey](http://websoilsurvey.nrcs.usda.gov/app/) or via the [Soil Data Access](http://sdmdataaccess.nrcs.usda.gov/Query.aspx) web page.  The SDM database contains several 1:many relationships within the data model. These 1:many relationships complicate the aggregation of soil properties. Typically, Soil Data Mart database aggregation queries use the ‘mukey’ columns to join the ‘mupolygon’ spatial layer to the ‘mapunit’ aggregated attribute data. However, aggregating data to a county, regional, or national scale requires additional aggregations that must be taken into account.

The soil survey inventory created 'soil survey areas' to create manageable ares to map and publish the soil survey information.  A soil survey area can contain multiple counties or pieces and parts of counties.  Therefore, to aggregate data to the county, the county overlap must be used instead of the soil survey area.  Although most soil surveys east of the Rockies were created using county borders, the use of the soil survey area symbol can not be used to join to county spatial layers because not all soil survey symbols are synonymous with the county FIPS codes. To aggregate data to a county level, the query must begin by using the Legend Area Overlap table. This SDM table is used to identify the specific county(s) associated with the soil survey area.

The Legend Area Overlap table is maintained by each State Soil Scientist and should be properly populated with the map units assigned to the specific county. With ongoing update of the soil survey inventory it it is critical that the overlap table be maintained for each County or Parish. Each county will contain multiple map units. And then, each map unit contains several components and each component contains many soil horizons.  The schema would look similar to:

Soil Survey Area                                     1 survey: potential multiple counties

           County                                               1 county: many map units

                     Map units                                 1 map unit : many components

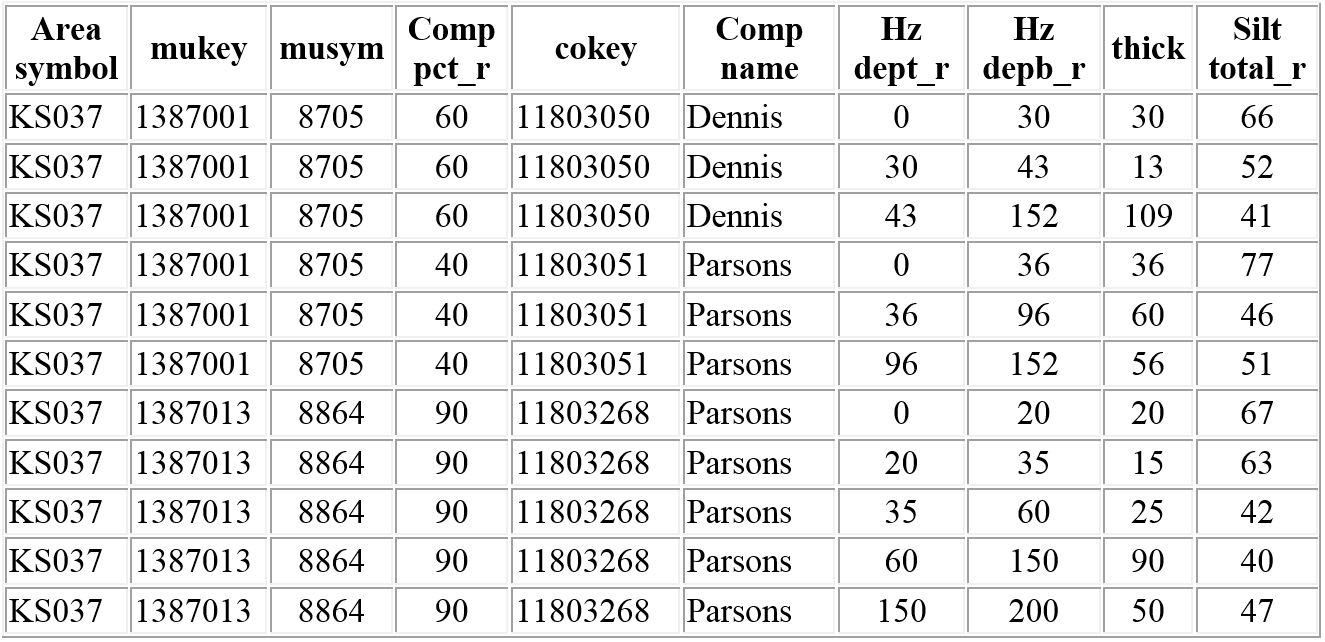
                                Components                   1 component: many horizons

                                            Horizons            1 horizon: many properties

To aggregate data, the data must include a weighted average at each level up to the county symbol. In this scenario I will use silt as an example to obtain a county silt  value.  this scenario will aggregate the data from the horizon to the county:

**Horizon Aggregation**

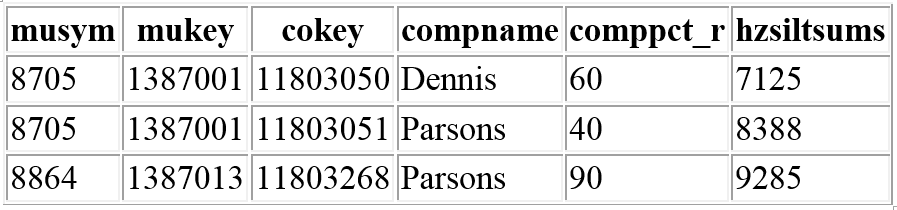
Each horizon has a top depth and a bottom depth and a thickness can be calculated (hzdepb\_r-hzdept\_r).  This image presents the county symbol, the mukey (links the spatial to attribute ), the county map unit symbol (musym), the Component percentage, the cokey (unique number for the component), the Soil name (aka Comp name), the top horizon depth, the bottom horizon depth, the determined thickness of the horizon, and the Representative Value for the Silt percentage  within that horizon.  In this image below, there are two map units (8705 and 8864) with 8705 having two components and 8864 having one component.



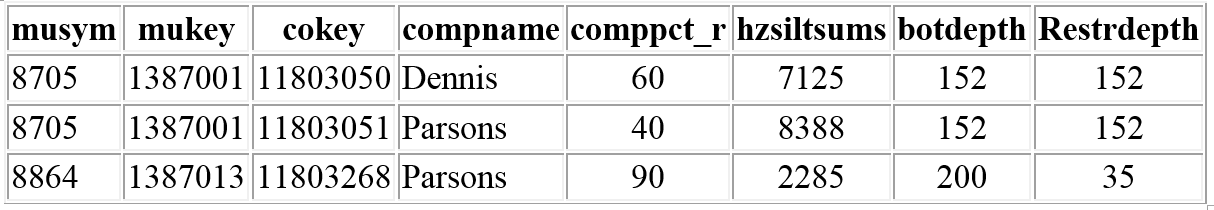
Each horizon thickness is multiplied against the silt property for the given horizon (silt total).  Notice the depths of each Horizon within the Component.



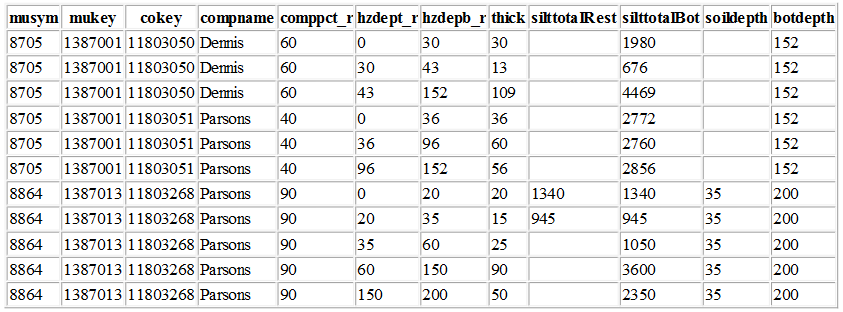
The Horizon property (silt) is then summed for the component.  Notice the silt total and the component percent of the Parsons component in each map unit.



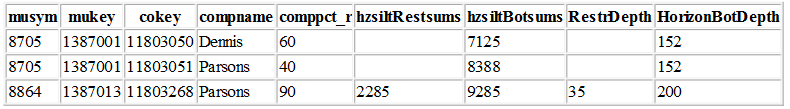
The component silt total is then weighted against the total soil thickness.  The thickness used can be dependent on the depth to a restrictive layer (Restrdepth) or bottom depth (botdepth) of the entire profile. The decision on which depth to use is based on the user criteria. In these two examples below, focus on the Parsons components and compare the ‘botdepth’ column to the hzdepb\_r column in the first two tables. The Parsons 11803051 has a soil depth and a restriction depth of 152 cm.  However, the second Parsons 11803268,  a restriction is populated at 35 cm with the bottom horizon depth set to 200 cm. Population of a restriction layer such as this makes a significant weighting discrepancy and must be taken into account depending on the needs of the interpretation.



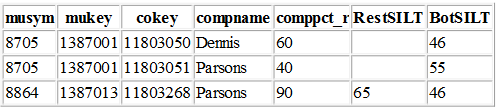
The decision has to be made to determine at what depth to aggregate the soil property.  The image below presents the differences in the numbers based on the use of the restriction depth or the bottom depth.  Notice the difference in Parsons 11803268 and the impact on the result depending on the accepted depth:



Compare the results of the two Parsons components and the result of the summed component silt property based on using the bottom horizon depth (botdepth) or the soil restriction depth (Restdepth). In the image below, notice the difference in the results for 11803268 based on the use of the restriction depth or the bottom horizon depth:

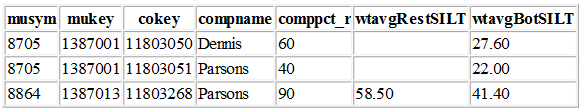


Dividing the summed silt total divided by the appropriate depth presented below will identify how the use of a Restriction depth versus a horizon bottom depth will skew the result.  Notice how the result of 11803268 can impact the result of further aggregation.



**Component Aggregation**

Each component includes a percentage in which the soil occupies within the map unit. Notice in the chart above that the two map units above each have a Parsons component, however in the first map unit (8705) the Parsons comprises 40% whereas the second map unit (8864) contains 90% Parsons. The horizon weighted silt content is then weighted against the percentage of the component as presented below:

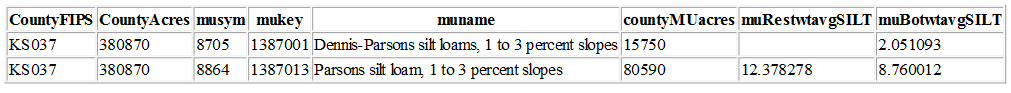


**Map unit Aggregation**

The weighted component properties are then summed to build a map unit property.



 Each map unit has a set of acres mapped with the soil survey area.  The overlap tables are used to assign the acres of that map unit within the specific county.  These acres are used to weight the map unit silt total within the given county.  In this image below, the county code, total county acres and the acres of the map unit within the county are used to weight the silt total of that map unit for that county.  Notice how the Restriction depth impacts the overall participation of the soil, and ultimately the map unit:



**County Aggregation**

Many database properties are used to aggregate data to the county level and create a county property.  They are:

* horizon thicknesses,
* the soil depth,
* the component percentage of the map unit,
* the map unit percentage of the county, and
* the total county acres

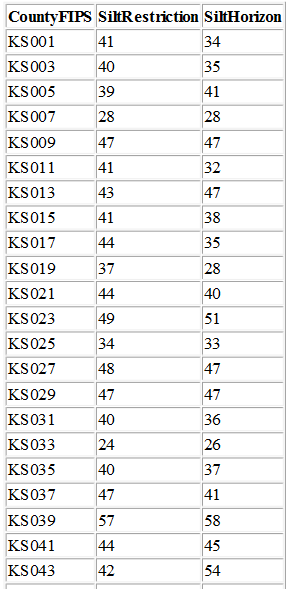
The soil property can be aggregated to a county level and used  to provide a county based soil property.  So, for this specific example, the county aggregated result for these two map units would be have a slight deviation based on the use of the restriction depth or the horizon depth:

https://media.licdn.com/mpr/mpr/AAEAAQAAAAAAAAd-AAAAJDgxNTNjYzZhLTc0Y2MtNGEwYi1hZjY0LWNhODE4M2VlODg1Yg.png

And looking at the result for all map units within the entire county:

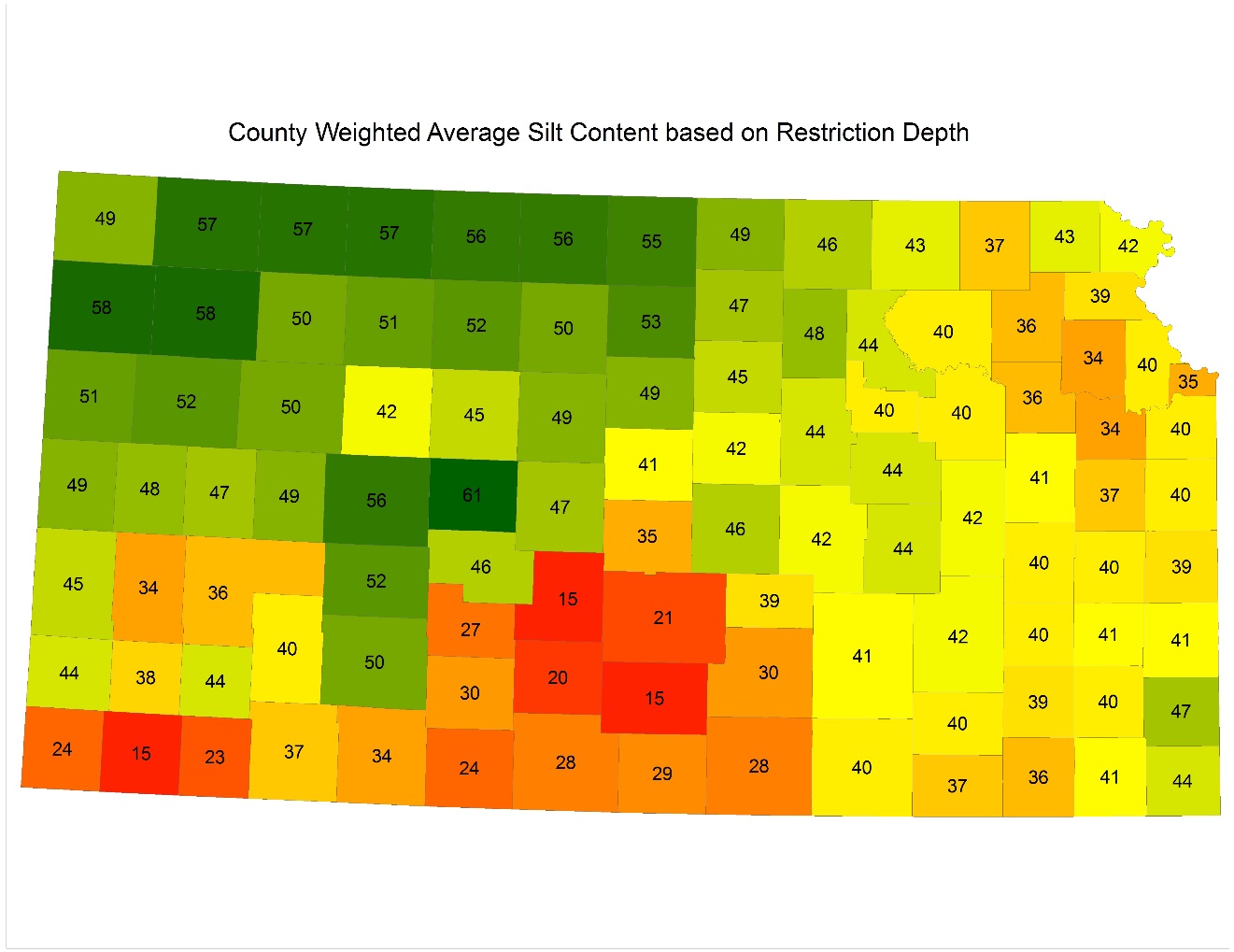
https://media.licdn.com/mpr/mpr/AAEAAQAAAAAAAAg6AAAAJGFlNjZhZjczLTUwY2YtNGU2Ni1hYzYzLWM4N2FlZjgwMDc5Zg.png

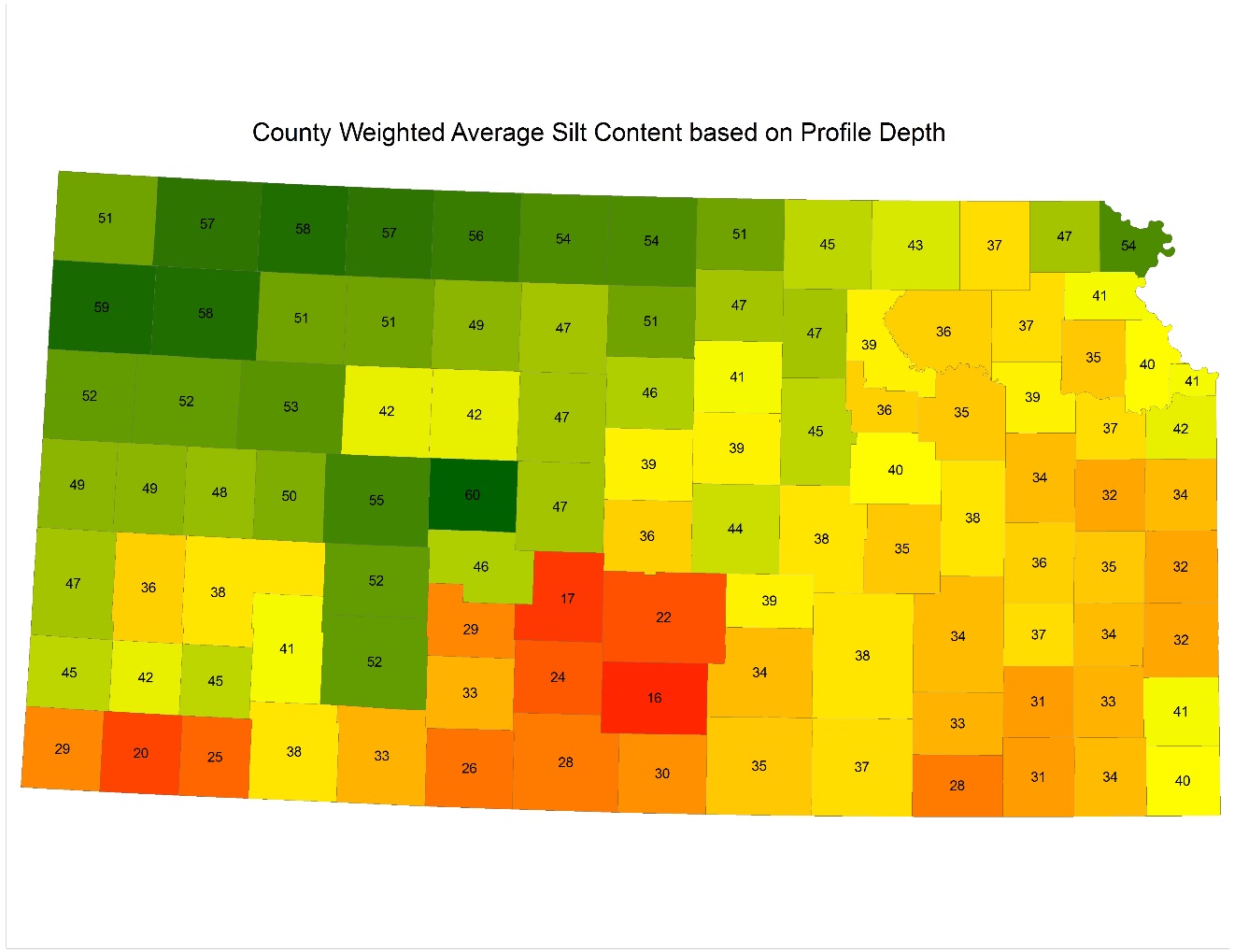
And comparing the results on how the depth determination affects the aggregation, here is a selection of the results for the state of Kansas:



**Conclusion**

Aggregation of a soil property requires careful consideration (and data population).  The manipulation of the five database properties, previously identified, used to aggregate data can greatly influence the final result.  The customer criteria must be clearly identified to properly aggregate the data to meet customer needs.   Understanding which soil depth to be used in the aggregation and how the depth choice influences the result is important to understand when aggregating the data.  And, visualizing the results between the two aggregation depths identifies the influence of soil restriction layers within the aggregated data.  For example, KS043 (42 vs 54) contains soils with considerable more restrictive layers than KS029  (47 vs 47).  Just another method of analyzing soils and just something else to consider when analyzing the data.





**SQL used for this scenario**

SELECT laoverlap.areasymbol

, l.areaacres

, m.mukey

, m.musym

, m.muname

, m.muacres

, muaoverlap.areaovacres

, ROUND((CAST(muacres\*1 as float)/areaacres),5) AS MapunitPercent

, CASE when muaoverlap.areaovacres is null then muacres else muaoverlap.areaovacres END as mapacres

INTO #Rn

FROM (mapunit m INNER JOIN ((legend l INNER JOIN laoverlap ON l.lkey = laoverlap.lkey and laoverlap.areatypename ='County Or Parish' )

INNER JOIN muaoverlap ON laoverlap.lareaovkey = muaoverlap.lareaovkey) ON m.mukey = muaoverlap.mukey and laoverlap.areasymbol like 'KS%')

--WHERE m.mukey IN (1387001, 1387013)

Order by laoverlap.areasymbol, m.mukey

SELECT areasymbol

, #Rn.mukey

, musym

, comppct\_r

, component.cokey

, compname

, h.chkey

, h.hzdept\_r

, h.hzdepb\_r

, (SELECT max(hzdepb\_r) from component c left outer join chorizon on c.cokey = chorizon.cokey where component.cokey = c.cokey group by c.cokey) as botdepth

, (SELECT CASE when min(resdept\_r) is null then '200' else cast(min(resdept\_r) as int) END from component cp left outer join corestrictions on cp.cokey = corestrictions.cokey where component.cokey = cp.cokey and reskind is not null) as soil\_depth

, case when (h.hzdepb\_r - h.hzdept\_r) = 0 then 0 else (h.hzdepb\_r - h.hzdept\_r) end AS thick

, CASE when silttotal\_r is null then '0' else silttotal\_r END as silttotal\_r

INTO #file1

FROM #Rn

INNER JOIN component on #Rn.mukey=component.mukey and majcompflag = 'yes'

INNER JOIN chorizon h on h.cokey = component.cokey

ORDER by musym, comppct\_r desc, hzdept\_r

--gather the horizon characters by thickness

SELECT

 #file1.musym

, #file1.mukey

, #file1.cokey

, #file1.compname

, comppct\_r

, hzdept\_r

, hzdepb\_r

, thick

,(SELECT thick\*silttotal\_r FROM #file1 file1 WHERE hzdepb\_r <= soil\_depth and #file1.chkey=file1.chkey and #file1.cokey=file1.cokey and #file1.mukey=file1.mukey) as silttotalRest

,(SELECT thick\*silttotal\_r FROM #file1 file2 WHERE hzdepb\_r <= botdepth and #file1.chkey=file2.chkey and #file1.cokey=file2.cokey and #file1.mukey = file2.mukey) as silttotalBot

,CASE when soil\_depth is null then botdepth else soil\_depth END as soildepth

, botdepth

INTO #file2

FROM #file1

ORDER BY #file1.mukey, #file1.cokey, hzdepb\_r

--sum the thicknesses by component

Select musym, mukey, cokey, compname, comppct\_r

, (sum(silttotalRest)) as hzsiltRestsums

, (sum(silttotalBot)) as hzsiltBotsums

, soildepth

, botdepth

INTO #file3

From #file2

Group by mukey, musym, cokey, compname, comppct\_r, soildepth, botdepth

ORDER BY mukey, cokey

--develop a weighted average for each component by using the soildepth and comppct\_r

Select distinct musym

, mukey

, cokey

, compname

, comppct\_r

, CAST((hzsiltRestsums/soildepth) as int) as RestSILT

, CAST((hzsiltBotsums/botdepth) as int) as BotSILT

INTO #file4

FROM #file3

ORDER BY mukey, cokey, comppct\_r

--develop a weighted average for each component by using the soildepth and comppct\_r

Select distinct musym

, mukey

, cokey

, compname

, comppct\_r

, (RestSILT)\*comppct\_r\*.01 as wtavgRestSILT

, (BotSILT)\*comppct\_r\*.01 as wtavgBotSILT

INTO #file5

FROM #file4

ORDER BY mukey, cokey, comppct\_r

--sum the component percent product to aggregate to the map unit

SELECT musym, mukey

, (sum(wtavgRestSILT)) as MwResttavgSILT

, (sum(wtavgBotSILT)) as MwBottavgSILT

INTO #file6

FROM #file5

Group by mukey, musym

--get the percent of the map unit

SELECT

areasymbol as CountyFIPS

, areaacres as CountyAcres

, #Rn.musym

, #file6.mukey

, muname

, areaovacres as countyMUacres

, (MwResttavgSILT\*areaovacres/areaacres) as muRestwtavgSILT

, (MwBottavgSILT\*areaovacres/areaacres) as muBotwtavgSILT

INTO #file7

FROM #Rn

INNER JOIN #file6 on #Rn.mukey=#file6.mukey

---get the county result

SELECT

CountyFIPS

, CAST(sum(muRestwtavgSILT) as int) as SiltRestriction

, CAST(sum(muBotwtavgSILT) as int) as SiltHorizon

FROM #file7

GROUP by CountyFIPS

ORDER by CountyFIPS