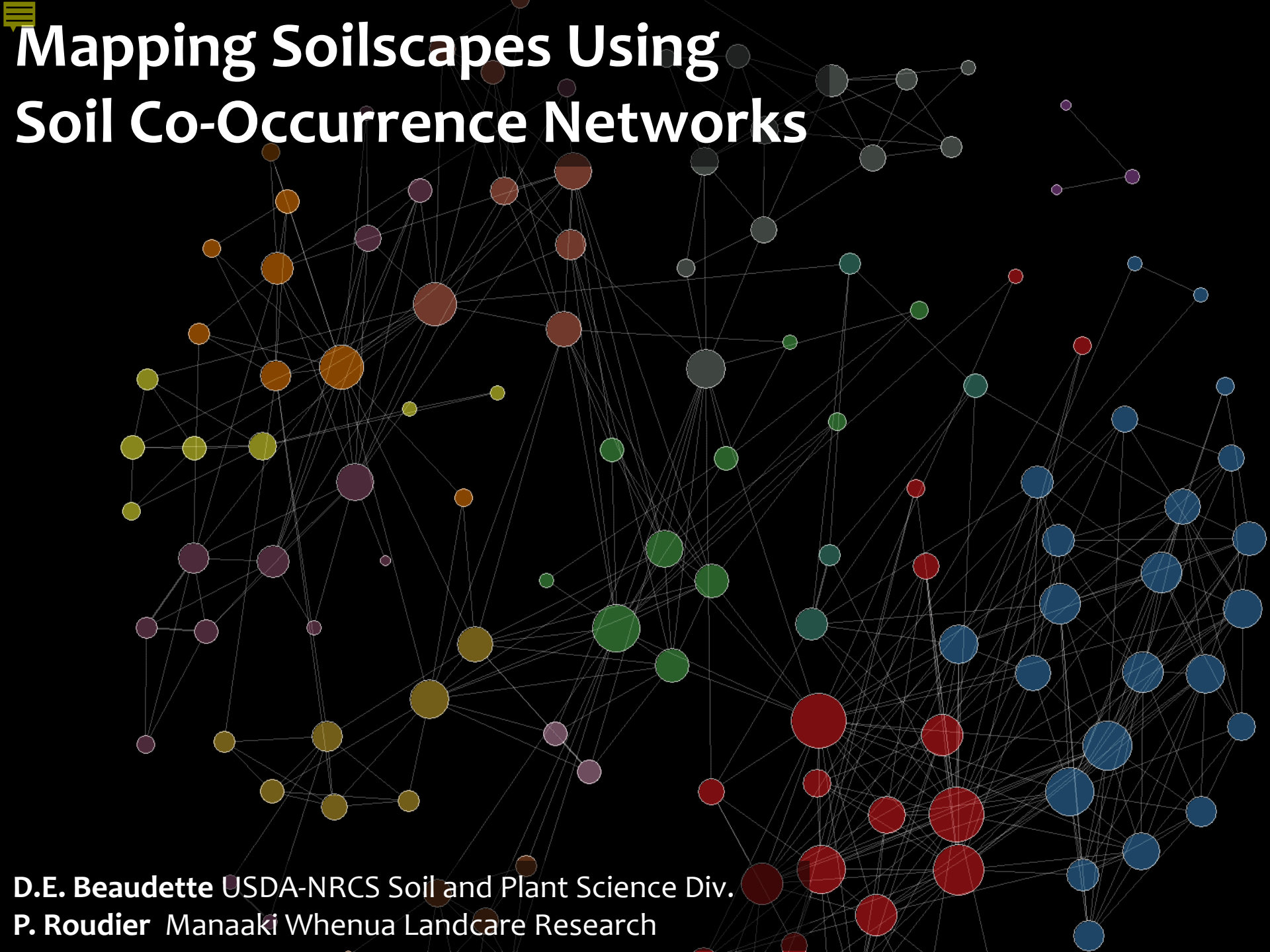
Mapping Soilscapes Using Soil Co-Occurrence Networks



D.E. Beaudette USDA-NRCS Soil and Plant Science Div.
P. Roudier Manaaki Whenua Landcare Research

Acknowledgments

- Pierre Roudier and Dion O'Neale
- A.T. O'Geen, C. Stiles, Drew Kinney
- 100+ years of soil survey effort
- **R** graph libraries (`igraph, ape`)
- **R** soil science related libraries (`soilDB, sharpshootR`)

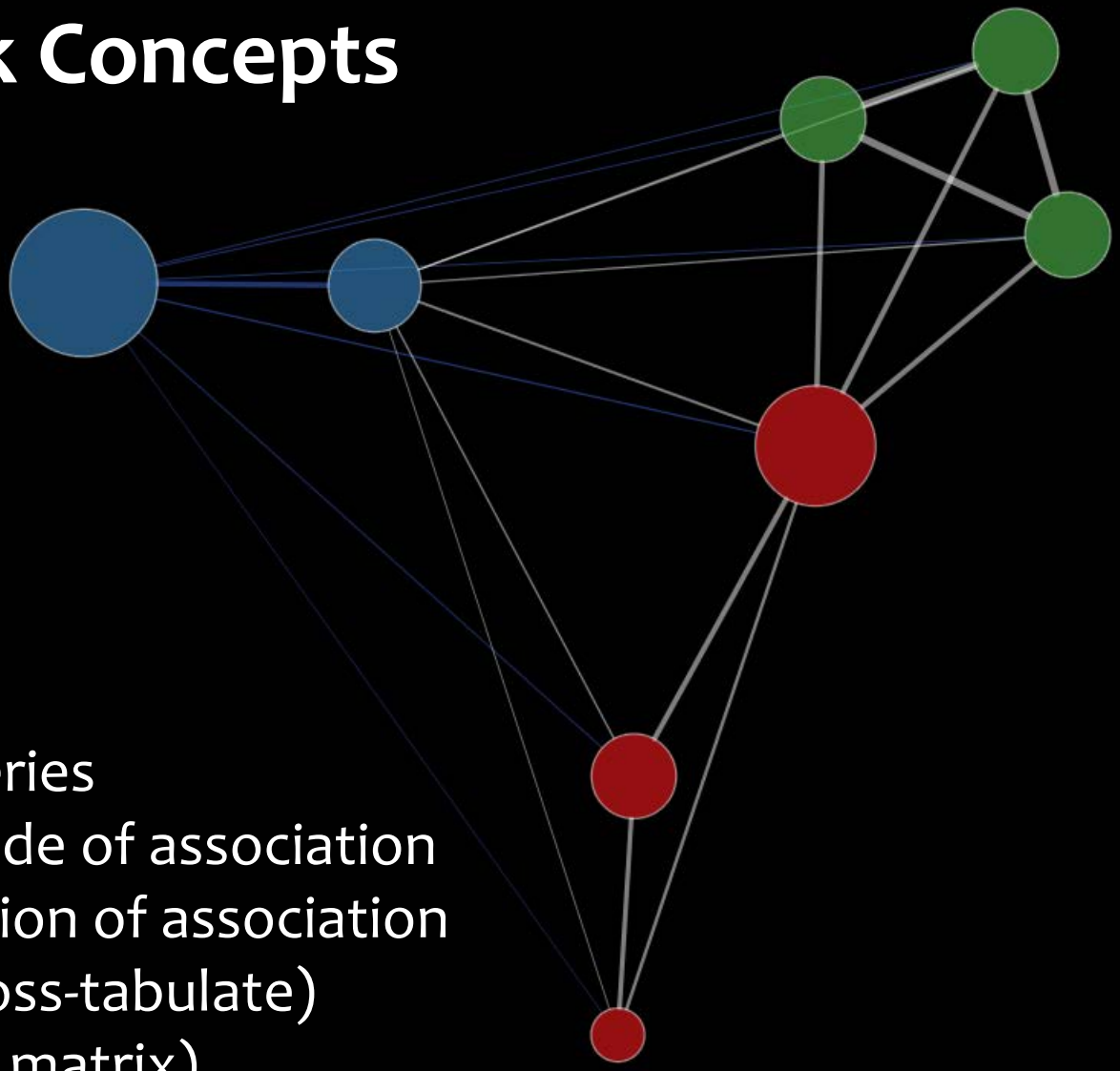
Prior Work

- 1977, 1985: F.D. Hole → soilscapes defines patterns of association
- 2001: P. Lagacherie et al. → soilscapes as landscape signature
- 2010: A.E. Hewitt et al. → soilscapes as modeling domains
- 2013/2016: J.D. Phillips → spatial adjacency of mapping units
- 2013—current: SoilWeb Series Data Explorer
- 2014/2017: D.E. Beaudette and P. Roudier → mapping co-occurrence
- 2018/2019: Odgers, Roudier, Thompson, Beaudette → soilscapes in NZ/US

This talk is about quantifying co-occurrence within tabular data

This talk is not about quantifying spatial co-occurrence

Graph / Network Concepts



- node or vertex: soil series
- edge weight: magnitude of association
- adjacency: quantification of association
  - co-occurrence (cross-tabulate)
  - distance (distance matrix)
  - similarity (restatement of distance)
- degree: number of edges / node
- community detection: clustering patterns of co-occurrence

https://casoilresource.lawr.ucdavis.edu/sde/?series=sierra

Soil Data Explorer - SIERRA

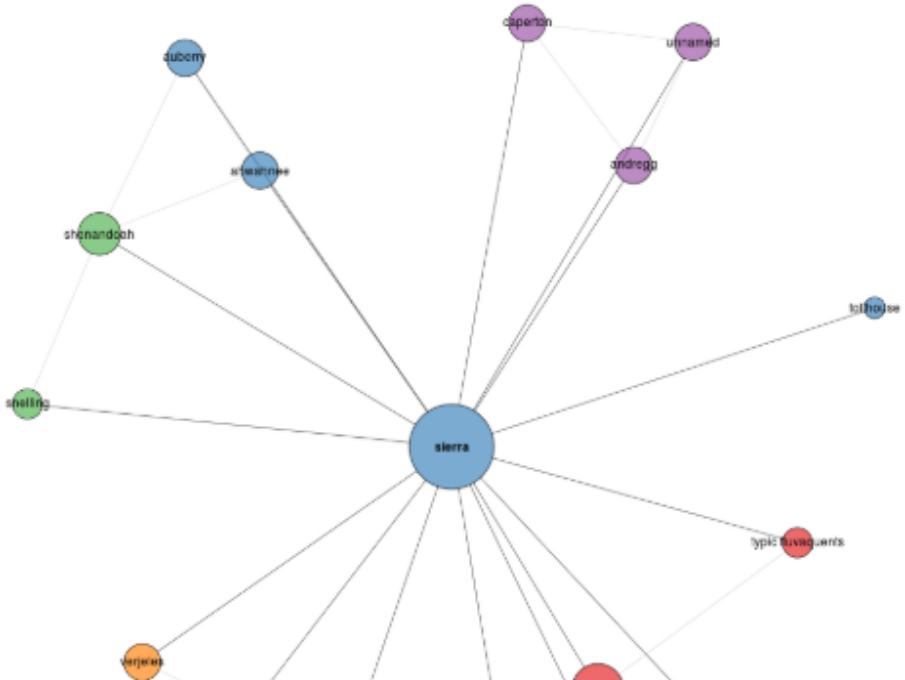
OSD | Lab Data | **Component Association** | Series Association | Block Diagrams | Map Units | Extent | Competing Ser

Component Association

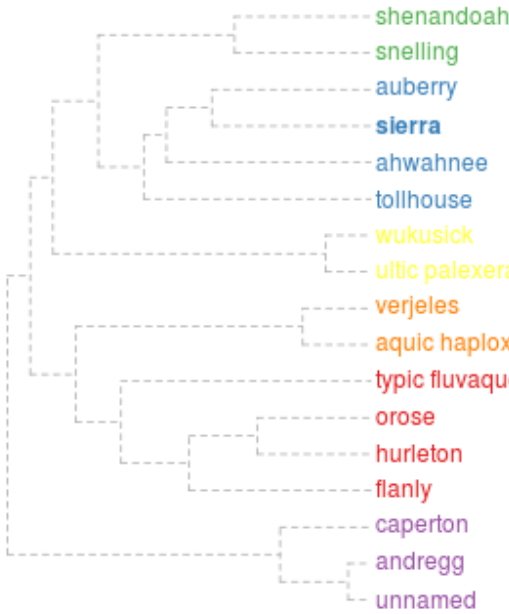
Two views depicting components that occur within map units containing the SIERRA series *as a major component*. Component association is a function of how frequently co component percentage. This summary *should* generate structures that are related to soil-landscape concepts that were used to construct map units containing the SIERRA approximations of the multi-dimensional relationship between components. Note that these diagrams give only partial information on the relationships between components SSURGO snapshot.

Graph representation: each circle (vertex) is a component, lines (edges) connect components that have been mapped together, colors define groups of components that frequently co-occur, vertex proximity is proportional to co-occurrence as weighted by component percentage. Dark lines connect vertices to the queried soil series, grey lines connect other vertices.

Click the image to view it full size.



Dendrogram representation: branch height (relative to the left-side of proportional to co-occurrence as weighted by component percentage components are highlighted in this view.



Soilscapes → Mapping Units → Soil Series

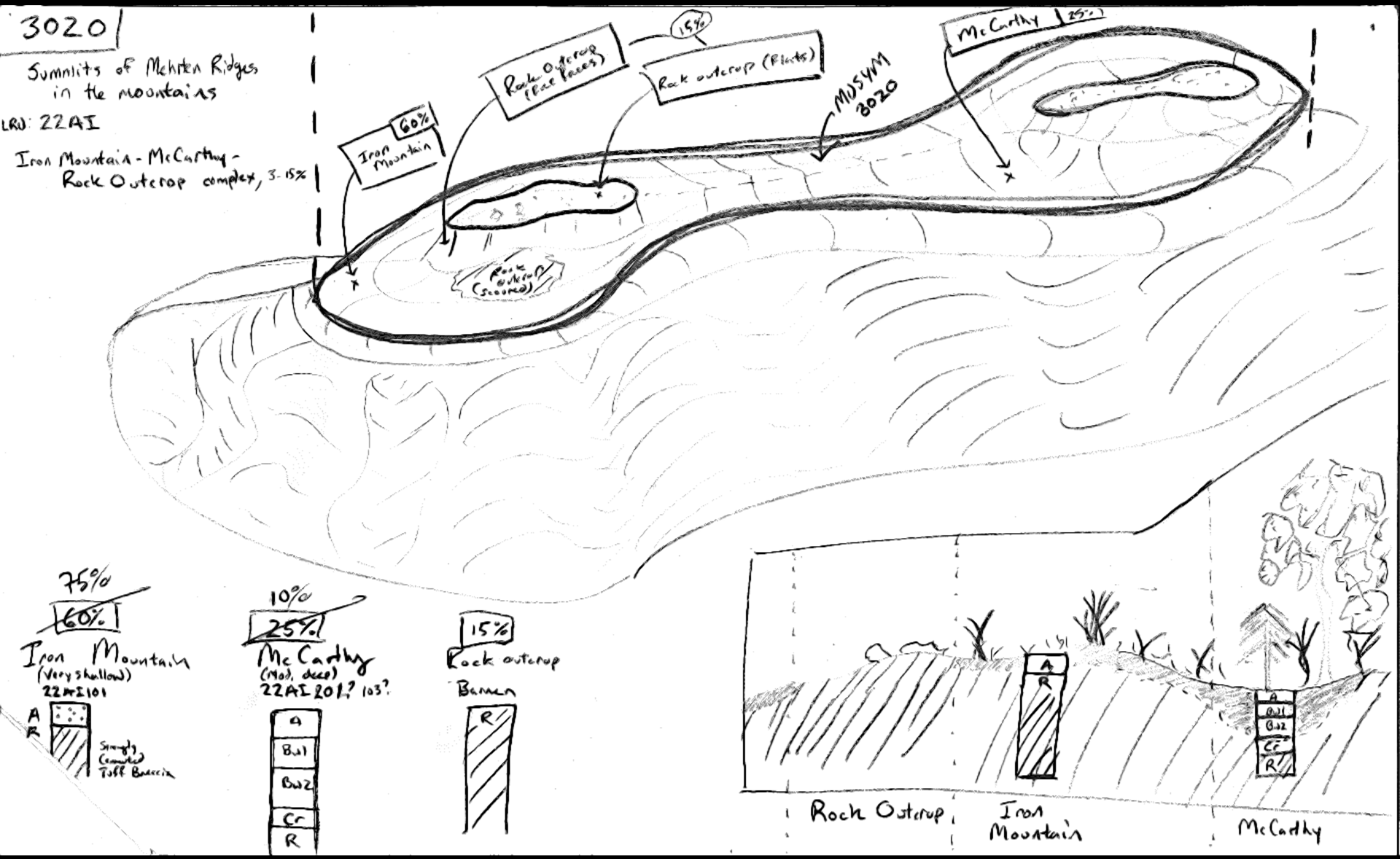
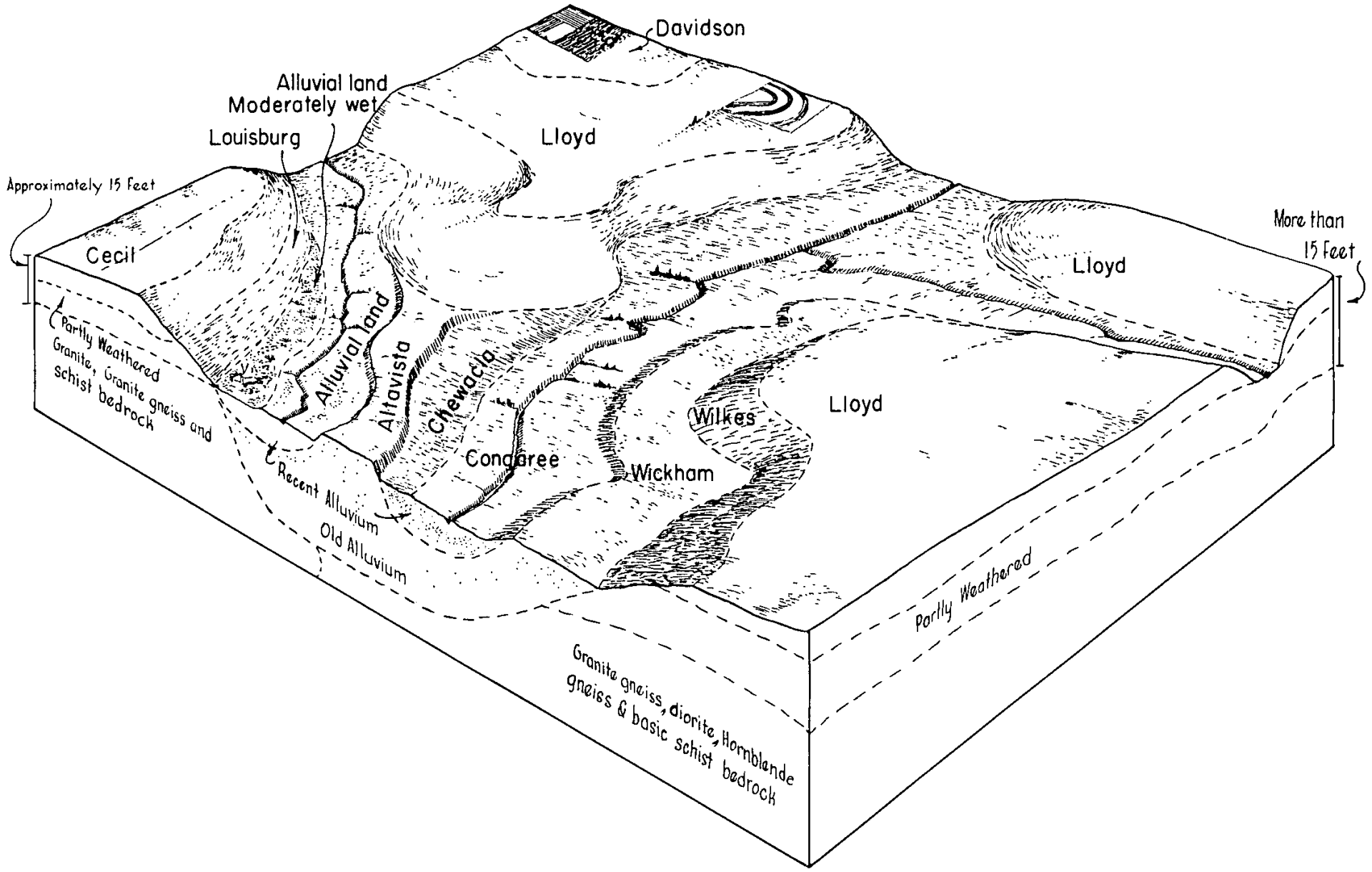


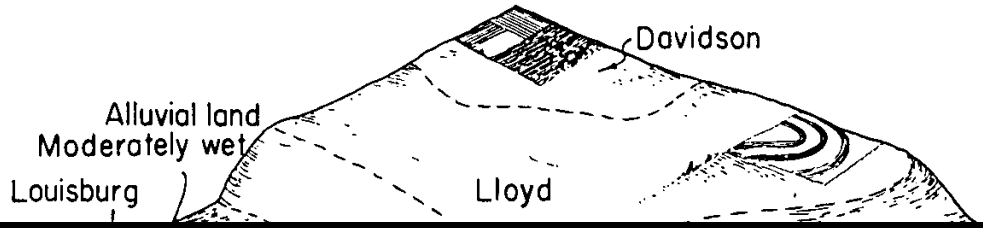
Figure by Andrew Brown, USDA-NRCS, Sonora, CA

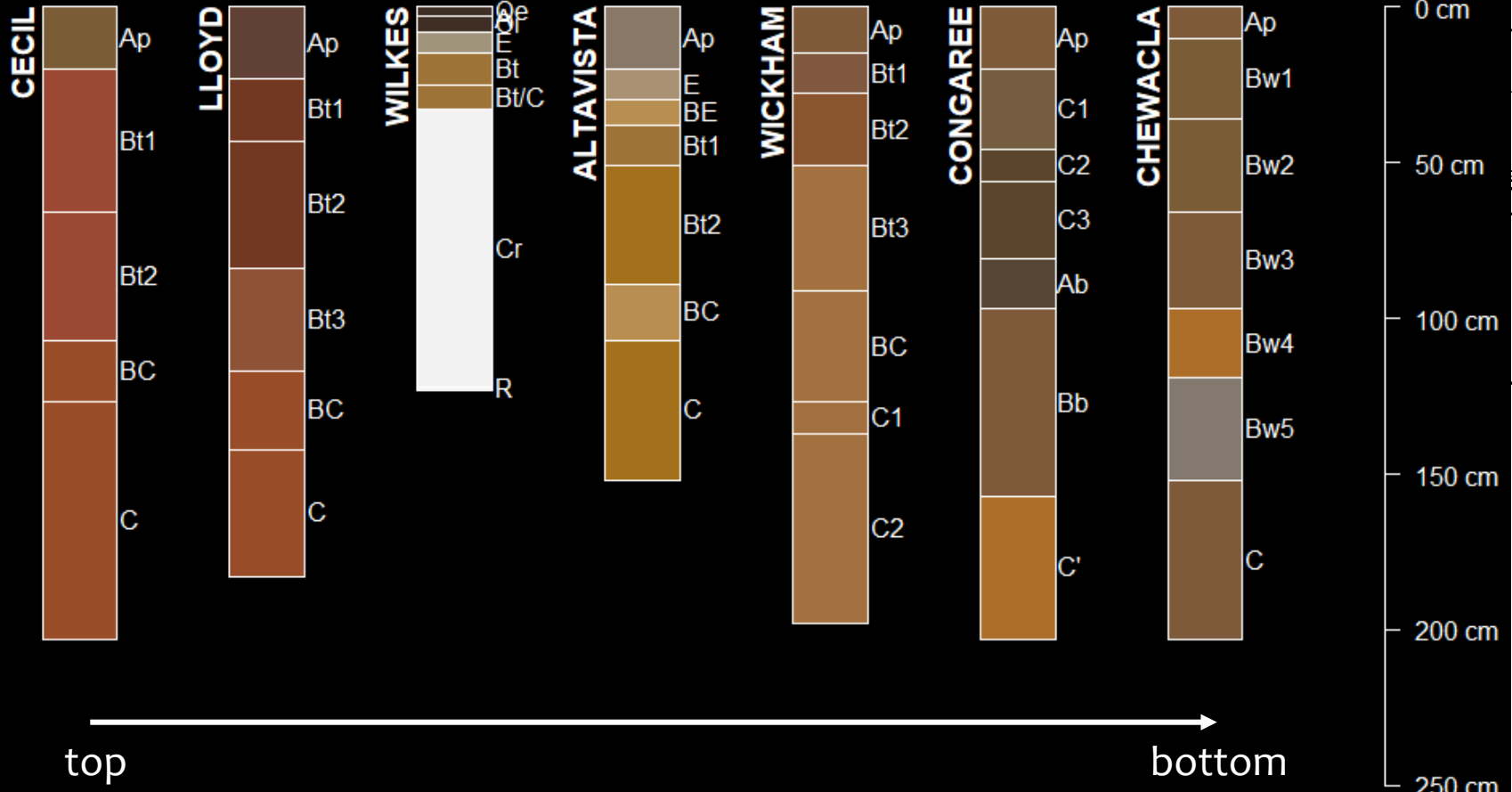
Soilscapes → Mapping Units → Soil Series

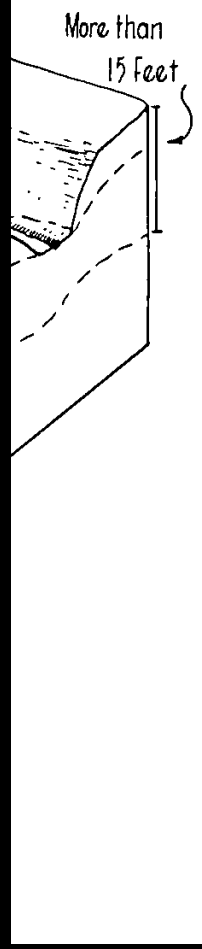


Lloyd-Davidson association (Soil Survey of Morgan County, Georgia; 1965)

Soilscapes → Mapping Units → Soil Series







Lloyd-Davidson association (Soil Survey of Morgan County, Georgia; 1965)

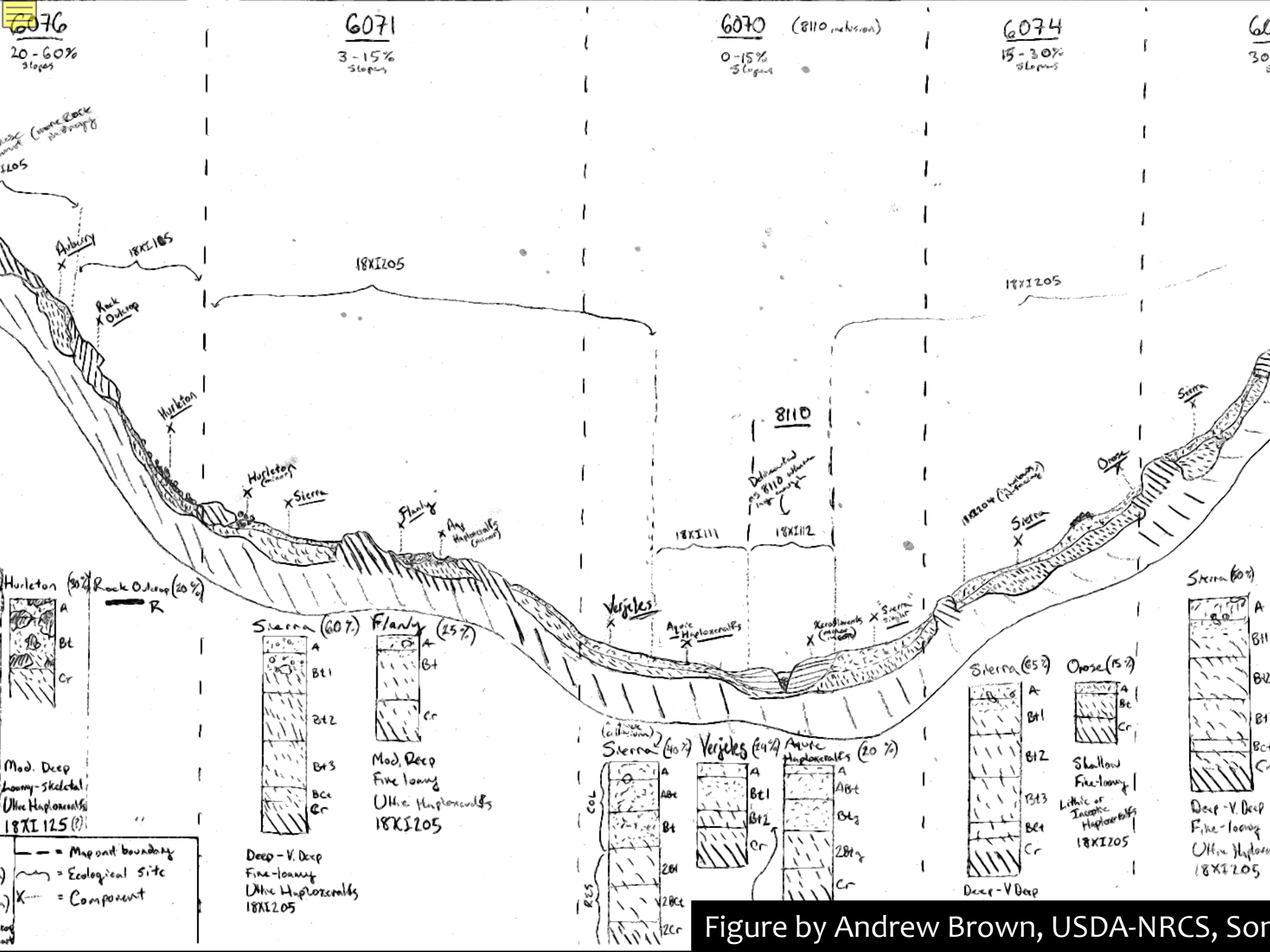
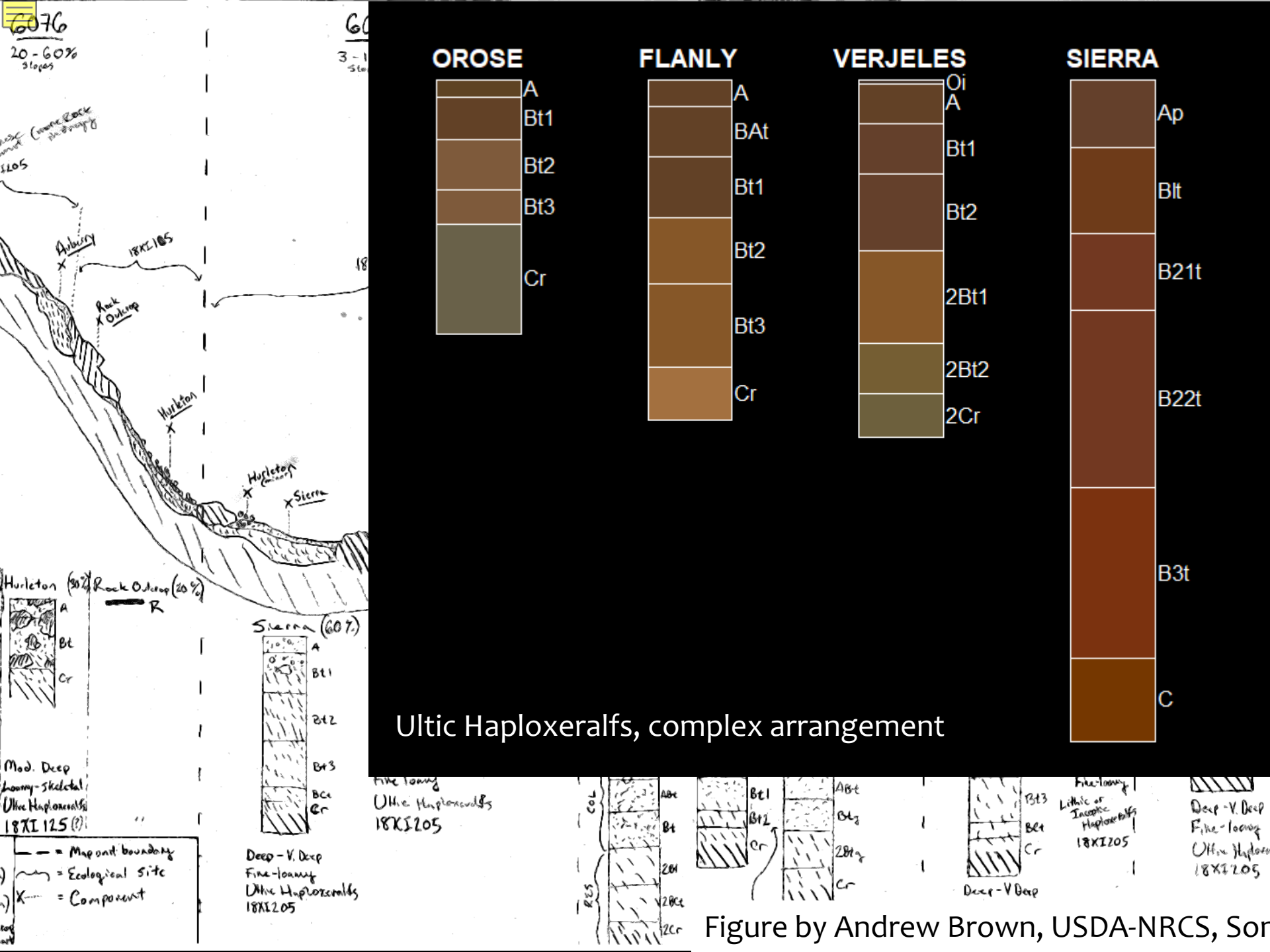
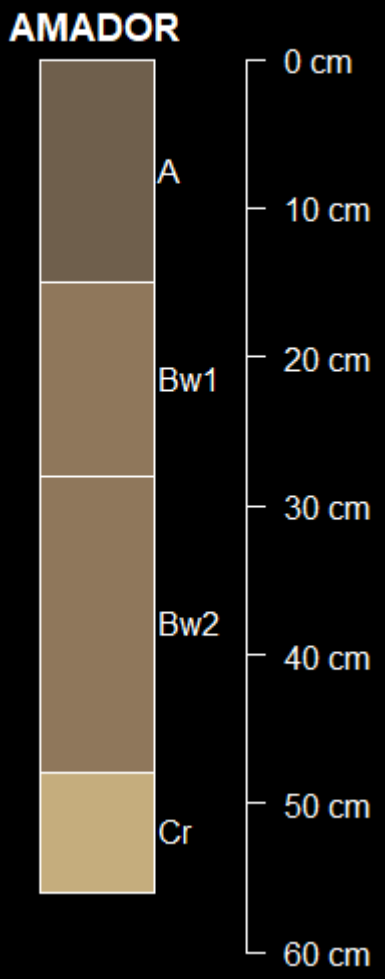


Figure by Andrew Brown, USDA-NRCS, Son


OROSE
A
Bt1
Bt2
Bt3
Cr
FLANLY
A
BAt
Bt1
Bt2
Bt3
Cr
VERJELES
Oi
A
Bt1
Bt2
2Bt1
2Bt2
2Cr
SIERRA
Ap
Blt
B21t
B22t
B3t
C
Ultic Haploxeralfs, complex arrangement
Figure by Andrew Brown, USDA-NRCS, Son

Mapping Units in SSURGO

mukey	compname	comppct_r	
1612048	**Amador**	45	complex
1612048	Gillender	40	
1612048	Ranchoseco	3	
1612048	Vleck	3	
1612048	Pardee	3	
1612048	Peters	3	
2766838	**Amador**	76	consociation
2766838	Gillender	9	
2766838	Pardee	5	
2766838	Miltonhills	5	
2766838	Redding	2	
	...		



Amador: Loamy, mixed, superactive, thermic, shallow Typic Haploxerepts

Mapping Units in SSURGO

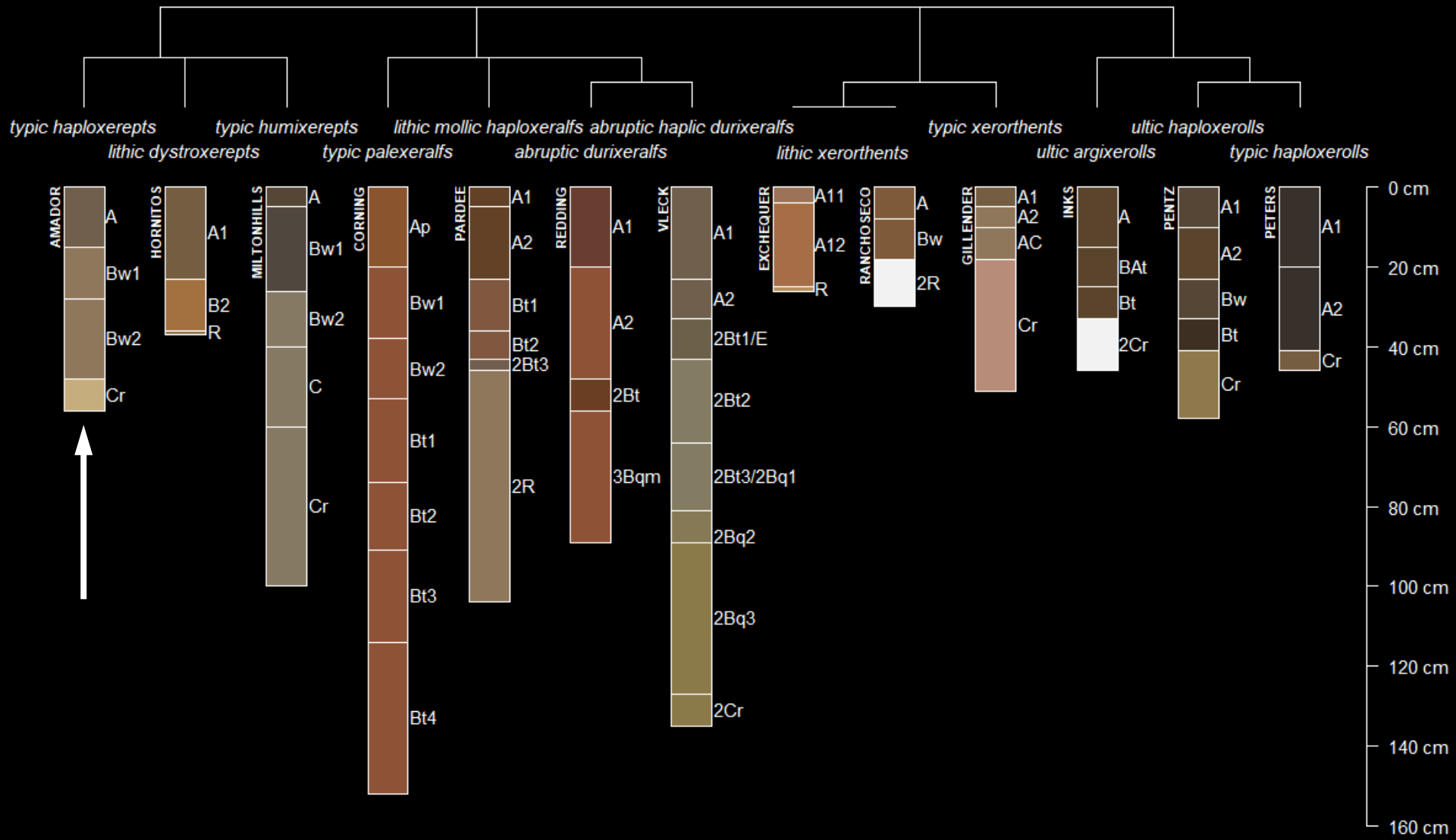
mukey	compname	comppct_r
1612048	**Amador**	45
1612048	Gillender	40

Mapping Units in SSURGO

mukey	compname	comppct_r
2766838	**Amador**	76
2766838	Gillender	9
2766838	Pardee	5
2766838	Miltonhills	5
2766838	Redding	2

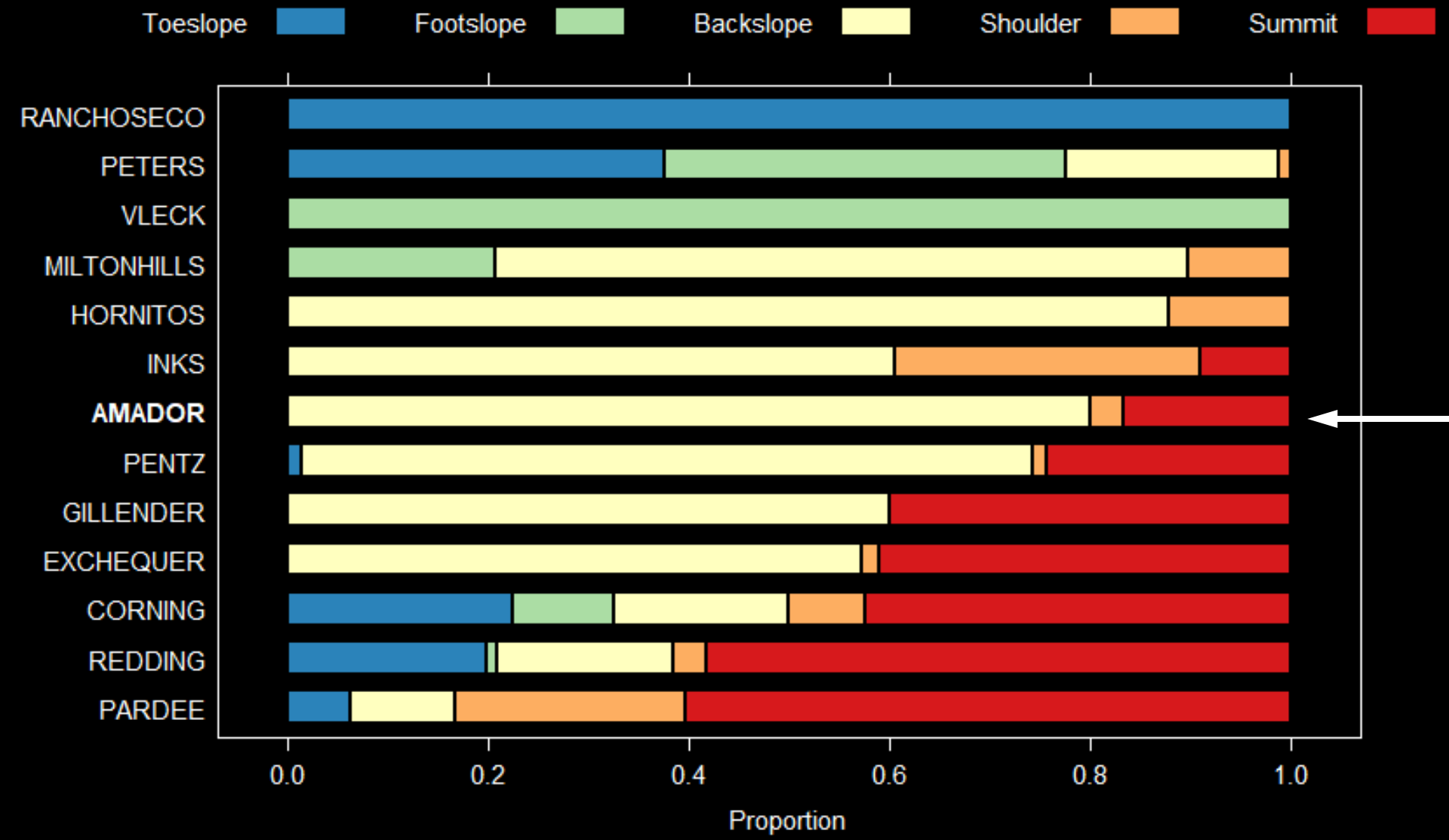
“Siblings”

Components* that co-occur with **Amador** in map units



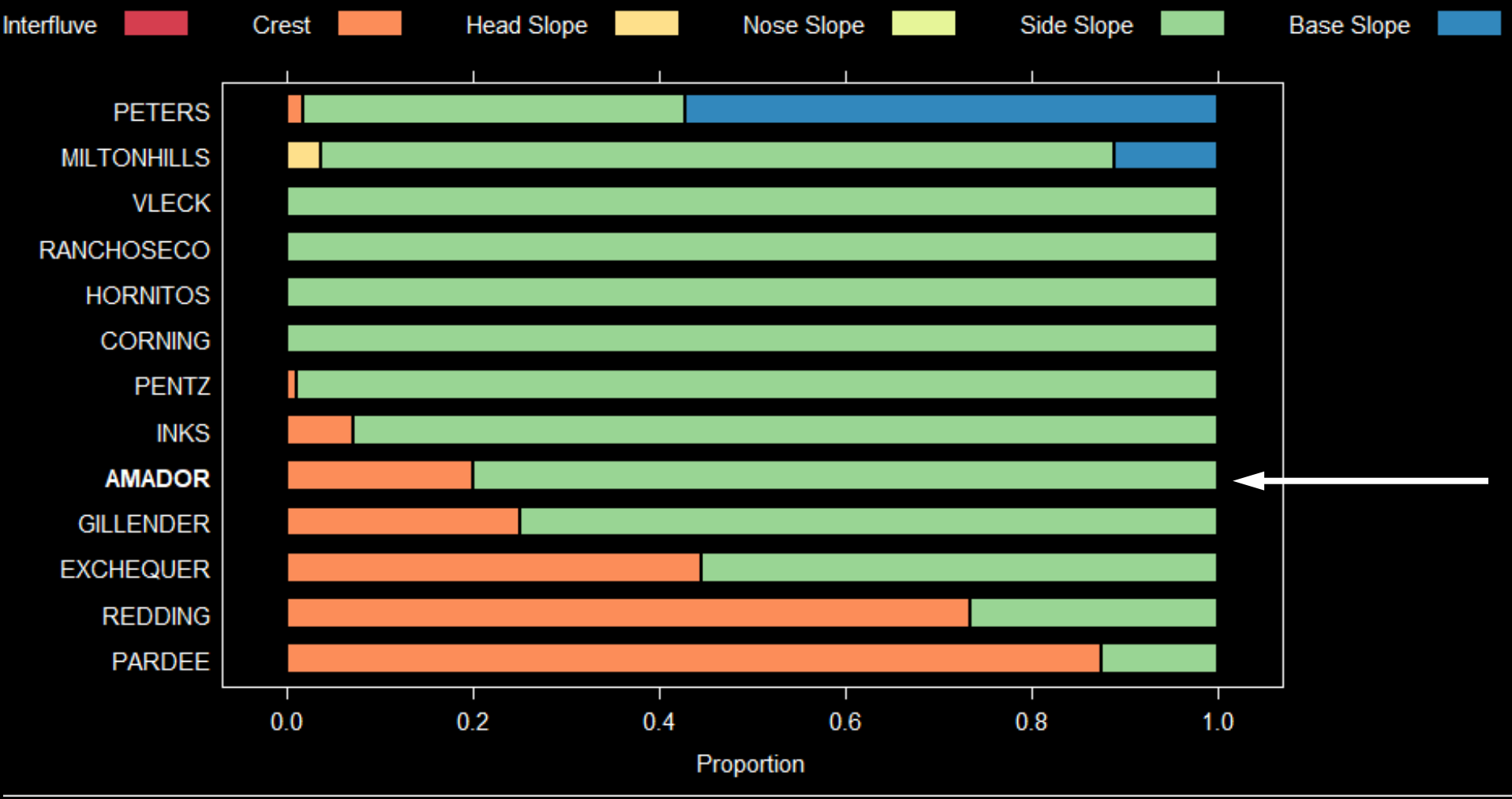
"Siblings"

Components* that co-occur with **Amador** in map units



“Siblings”

Components* that co-occur with **Amador** in map units



Quantify Co-Occurrence: Adjacency Matrix

Tabulate "co-occurrence" → *information is lost*

	Gillender	Miltonhills	Pardee	Peters	Ranchoseco	Redding	Vleck
Amador	2	1	2	1	1	1	1
Gillender		1	2	1	1	1	1
Miltonhills			1			1	
Pardee				1	1	1	1
Peters					1		1
Ranchoseco							1
Redding							
Vleck							

Compute "similarity"

1. collect mapunit / component records
2. reshape into "community matrix"
3. standardize and compute distance matrix (methods from numerical ecology)
4. convert distance matrix into similarity matrix

→ this is the adjacency matrix (details on next slide)

Why all the fuss? Component percentages (weights) matter!

Pseudo-community Matrix

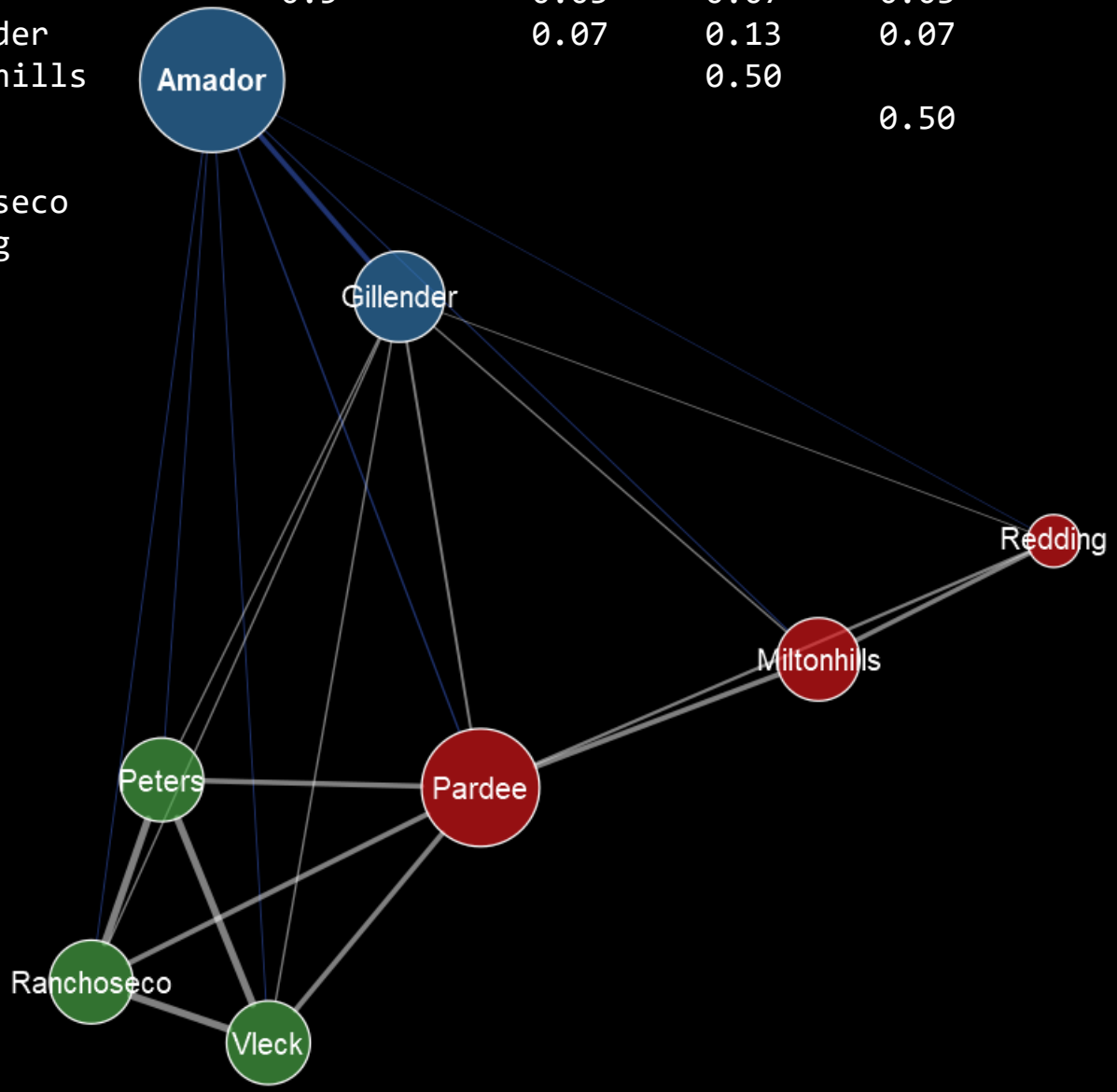
	461845	461980	462528	462529	462530	462531	462954	462955
Amador	45	25	85	85	85	85	85	85
Corning	0	5	0	0	0	0	0	0
Exchequer	0	0	5	5	5	5	0	0
Gillender	40	0	0	0	0	0	0	0
Hornitos	0	0	5	5	5	5	10	10
Inks	0	0	0	0	0	0	0	0
Miltonhills	0	0	0	0	0	0	0	0
Pardee	3	0	0	0	0	0	0	0
Pentz	0	0	5	5	5	5	5	5
Peters	3	0	0	0	0	0	0	0
Ranchoseco	3	0	0	0	0	0	0	0
Redding	0	4	0	0	0	0	0	0
Vleck	3	40	0	0	0	0	0	0

Similarity Matrix (via Jaccard index)

	Gillender	Miltonhills	Pardee	Peters	Ranchoseco	Redding	Vleck
Amador	0.5	0.03	0.07	0.03	0.03	0.01	0.03
Gillender		0.07	0.13	0.07	0.07	0.03	0.07
Miltonhills			0.50			0.40	
Pardee				0.50	0.50	0.20	0.50
Peters					1.00		1.00
Ranchoseco							1.00
Redding							
Vleck							

Adjacency Matrix → Graph

	Gillender	Miltonhills	Pardee	Peters	Ranchoseco	Redding	Vleck
Amador	0.5	0.03	0.07	0.03	0.03	0.01	0.03
Gillender		0.07	0.13	0.07	0.07	0.03	0.07
Miltonhills			0.50			0.40	
Pardee				0.50	0.50	0.20	0.50
Peters					1.00		1.00
Ranchoseco							1.00
Redding							
Vleck							



Why are we doing this?

https://xkcd.com/1838/

Why are we doing this?

Explicit, quantitative, human/machine readable encoding of historic (and future) soil knowledge.

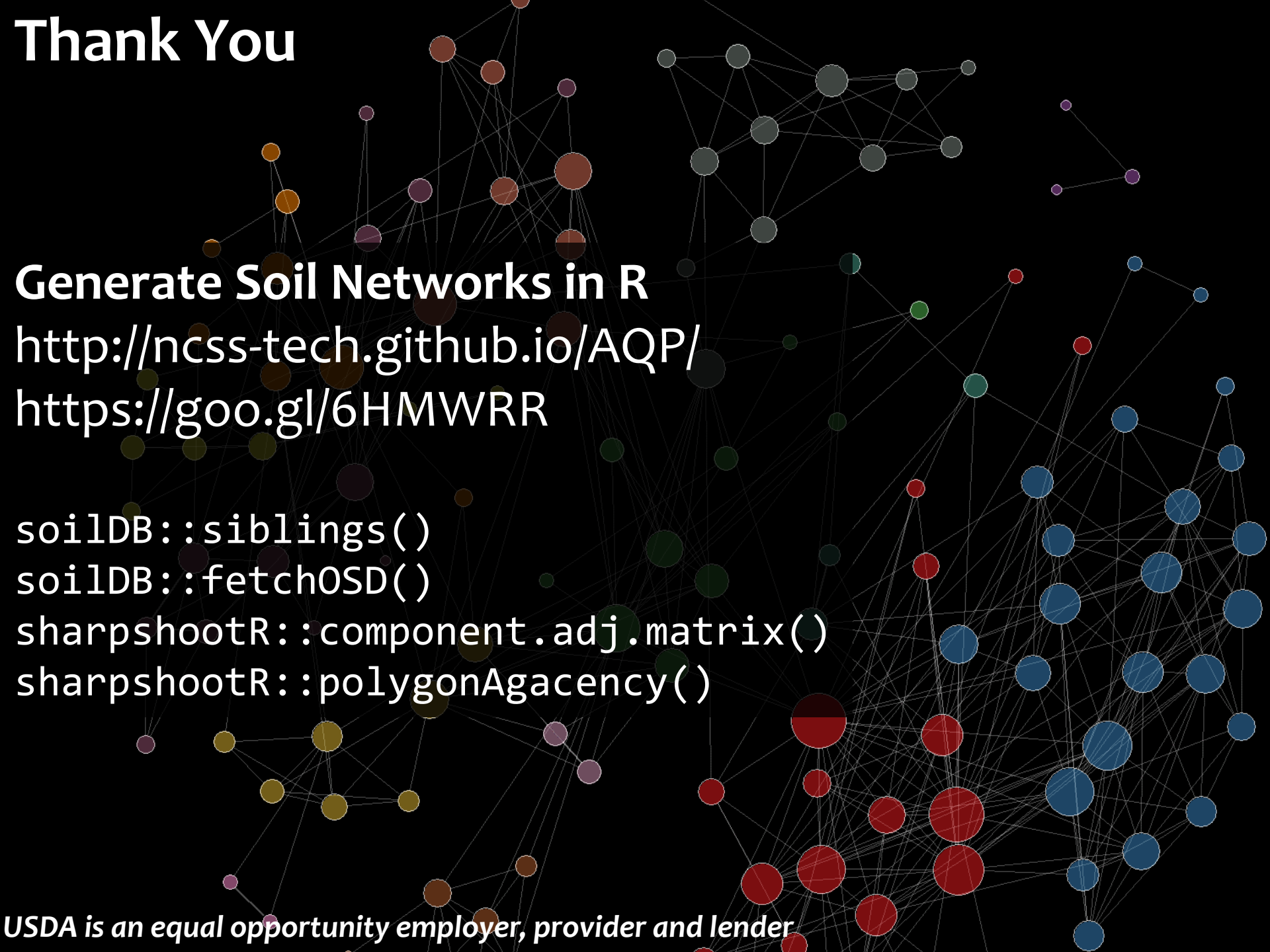
Continuous predictions of soil properties / membership are (of course) important.

Failing to integrate hard-won knowledge
(collected and synthesized in places where soils occur)
into modern efforts
would be a terrible tragedy.

Groups of co-occurring soils are useful strata that should be integrated into statistical models.

Co-occurrence networks are neat.

Thank You



Generate Soil Networks in R

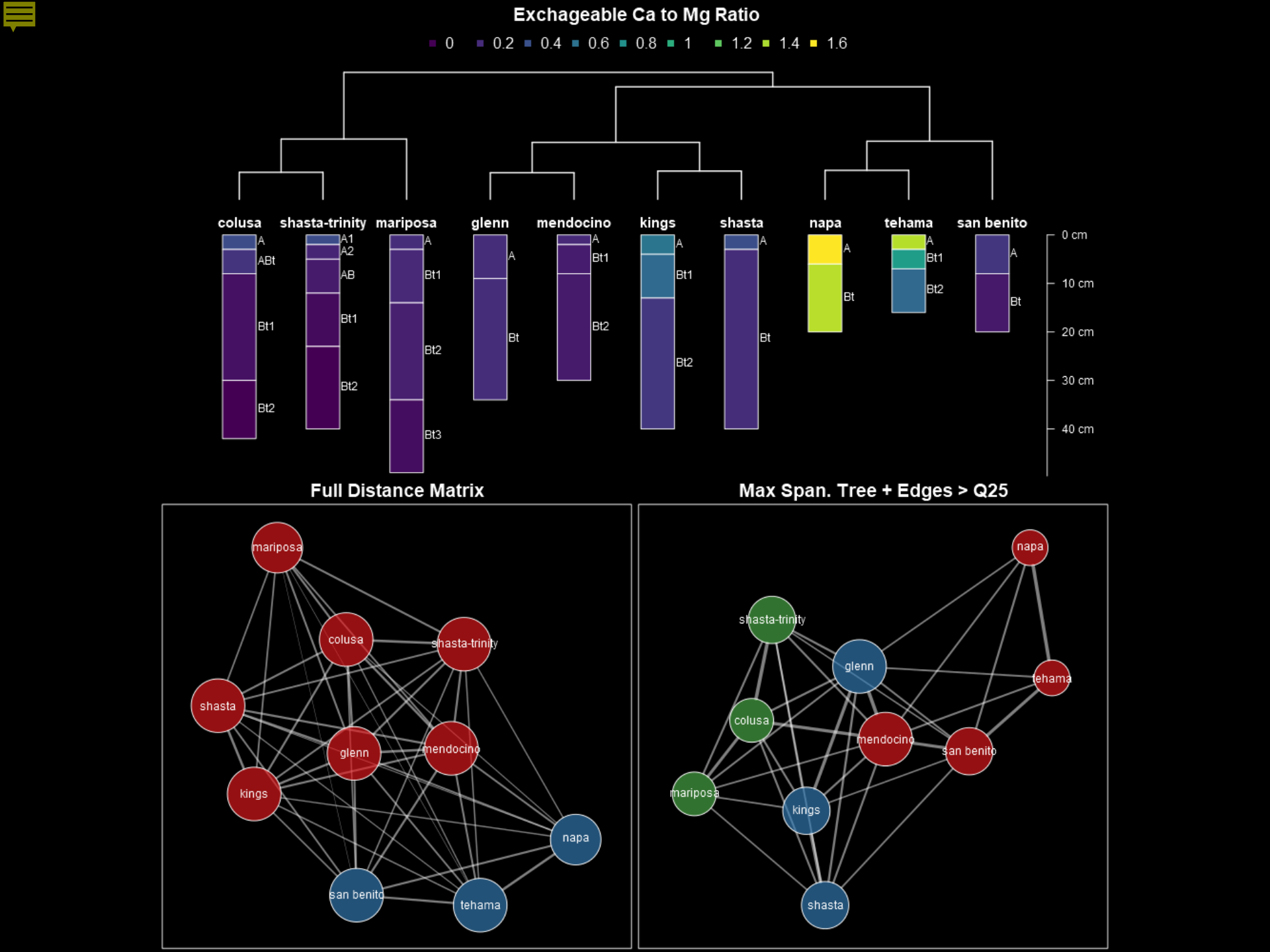
http://ncss-tech.github.io/AQP/

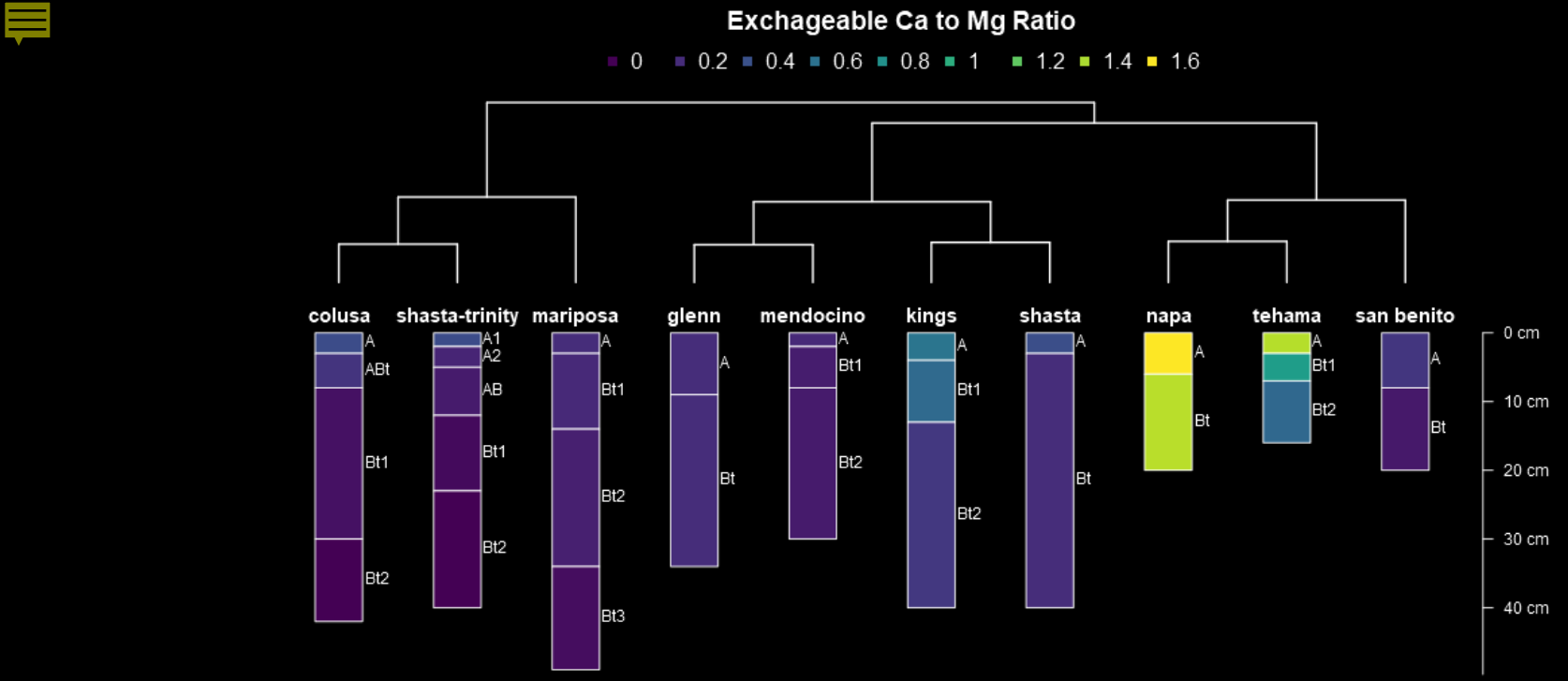
https://goo.gl/6HMWRR

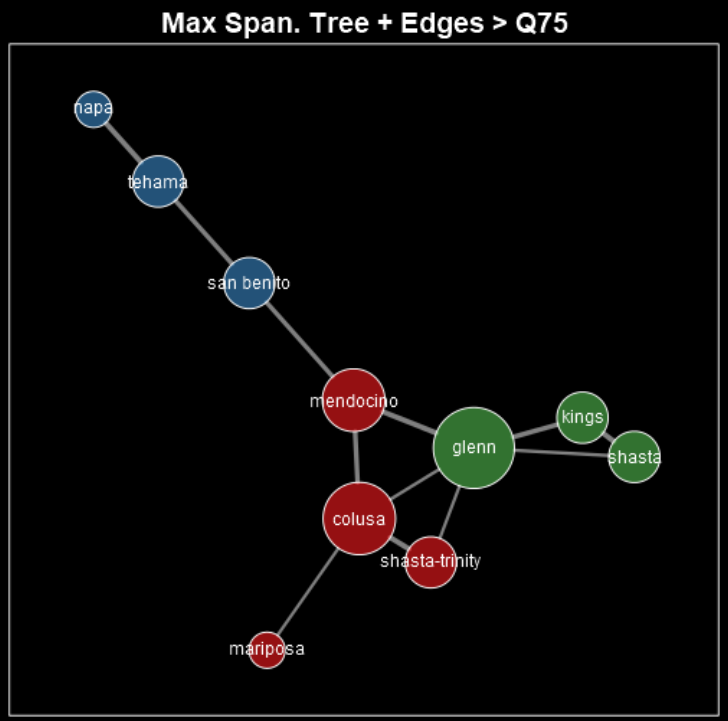
```
soilDB::siblings()
soilDB::fetchOSD()
sharpshootR::component.adj.matrix()
sharpshootR::polygonAgacency()
```

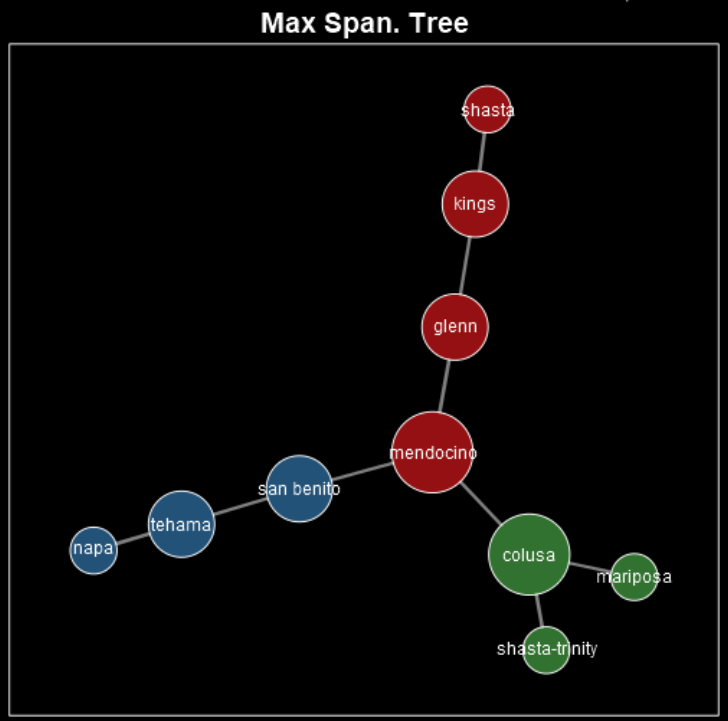
Help Test SoilWeb for Android Devices

https://goo.gl/WvrV8Y


Exchageable Ca to Mg Ratio
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
colusa
shasta-trinity
mariposa
glenn
mendocino
kings
shasta
napa
tehama
san benito
A
ABt
Bt1
Bt2
A1
A2
AB
Bt1
Bt2
A
Bt1
Bt2
Bt3
A
Bt
A
Bt1
Bt2
A
Bt1
Bt2
A
Bt
A
Bt
A
Bt1
Bt2
A
Bt
0 cm
10 cm
20 cm
30 cm
40 cm
Full Distance Matrix
mariposa
colusa
shasta-trinity
shasta
glenn
mendocino
kings
napa
san benito
tehama
Max Span. Tree + Edges > Q25
napa
shasta-trinity
glenn
tehama
colusa
mendocino
san benito
mariposa
kings
shasta


Exchageable Ca to Mg Ratio
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
colusa
shasta-trinity
mariposa
glenn
mendocino
kings
shasta
napa
tehama
san benito
A
ABt
Bt1
Bt2
A1
A2
AB
Bt1
Bt2
A
Bt1
Bt2
Bt3
A
Bt
A
Bt1
Bt2
A
Bt1
Bt2
A
Bt
A
Bt
A
Bt1
Bt2
A
Bt
0 cm
10 cm
20 cm
30 cm
40 cm


Max Span. Tree + Edges > Q75
napa
tehama
san benito
mendocino
glenn
kings
shasta
colusa
shasta-trinity
mariposa


Max Span. Tree
shasta
kings
glenn
mendocino
san benito
tehama
napa
colusa
mariposa
shasta-trinity

Adjacency Matrix ⟶ Graph

	Gillender	Miltonhills	Pardee	Peters	Ranchoseco	Redding	Vleck
Amador	2	1	2	1	1	1	1
Gillender		1	2	1	1	1	1
Miltonhills			1			1	
Pardee				1	1	1	1
Peters					1		1
Ranchoseco							1
Redding							
Vleck