Introduction to SQL II

Wide vs. Long Format

D. E. Beaudette

Dept. Land, Air and Water Resources
University of California
Davis, California

debeaudette@ucdavis.edu

Nov. 10, 2008

Data in Wide Format

Soil Solution Data

id	depth	ex_ca	ex_mg	ex_na	ex_k	water_ca	water_mg	water_na	water_k
66—CA—58—001—1	5	10.3	10.8	0.7	0.3	0.9	1.1	2.1	0.0
66—CA—58—001—2	13	13.3	14.9	1.0	0.2	1.1	1.2	2.3	0.1
66—CA—58—001—3	27	12.7	16.0	0.8	0.3	2.0	3.0	3.0	0.0
66—CA—58—005—1	10	14.4	4.5	0.4	0.3	2.9	1.6	10.7	1.7
66—CA—58—005—2	19	10.0	4.5	0.2	0.2	2.8	1.8	1.5	0.1
66—CA—58—005—3	27	10.0	6.4	0.2	0.3	1.5	1.1	1.1	0.1
66—CA—58—005—4	41	9.6	6.6	0.2	0.3	1.1	1.2	1.2	0.1
66—CA—58—005—5	54	7.8	6.1	0.2	0.2	1.5	1.8	1.2	0.1
66—CA—58—007—1	4	4.5	1.6	0.0	0.1	1.4	1.0	0.8	0.1
66—CA—58—007—2	13	6.8	2.6	0.0	0.1	2.8	1.5	0.6	0.0

[...]

Data in Long Format

Soil Solution Data

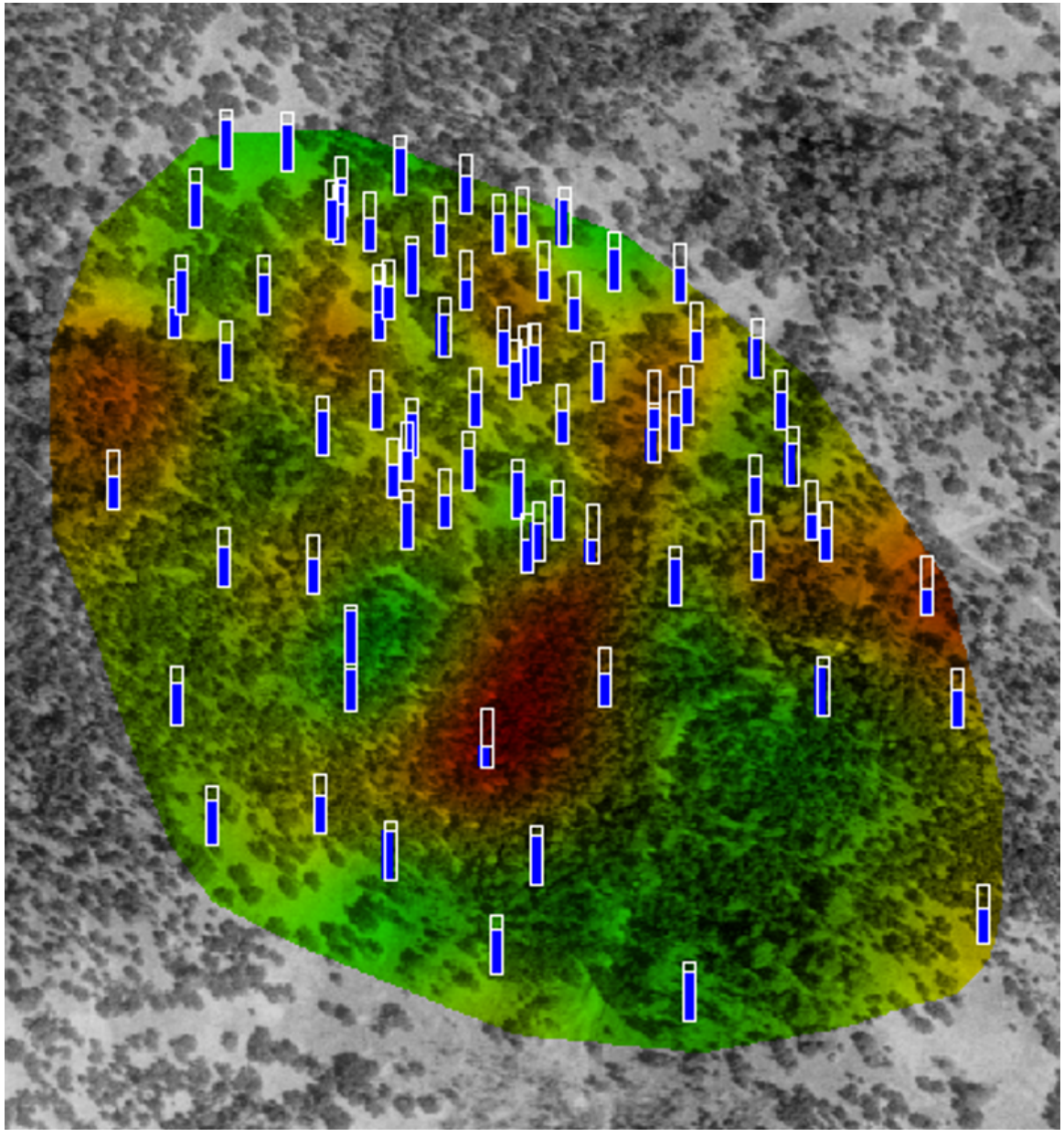
id	method	cation	depth	**value**
66—CA—58—001—1	Extractable	Na	5	0.7
66—CA—58—001—2	Extractable	Na	13	1.0
66—CA—58—001—3	Extractable	Na	27	0.8
[...]				
66—CA—58—001—1	Extractable	Ca	5	10.3
66—CA—58—001—2	Extractable	Ca	13	13.3
66—CA—58—001—3	Extractable	Ca	27	12.7
[...]				
66—CA—58—001—1	Extractable	Mg	5	10.8
66—CA—58—001—2	Extractable	Mg	13	14.9
66—CA—58—001—3	Extractable	Mg	27	16.0
[...]				
66—CA—58—001—1	Extractable	K	5	0.3
66—CA—58—001—2	Extractable	K	13	0.2
66—CA—58—001—3	Extractable	K	27	0.3
[...]				
66—CA—58—001—1	Soluble	Na	5	2.1
66—CA—58—001—2	Soluble	Na	13	2.3
66—CA—58—001—3	Soluble	Na	27	3.0
[...]				

Conversion By SQL

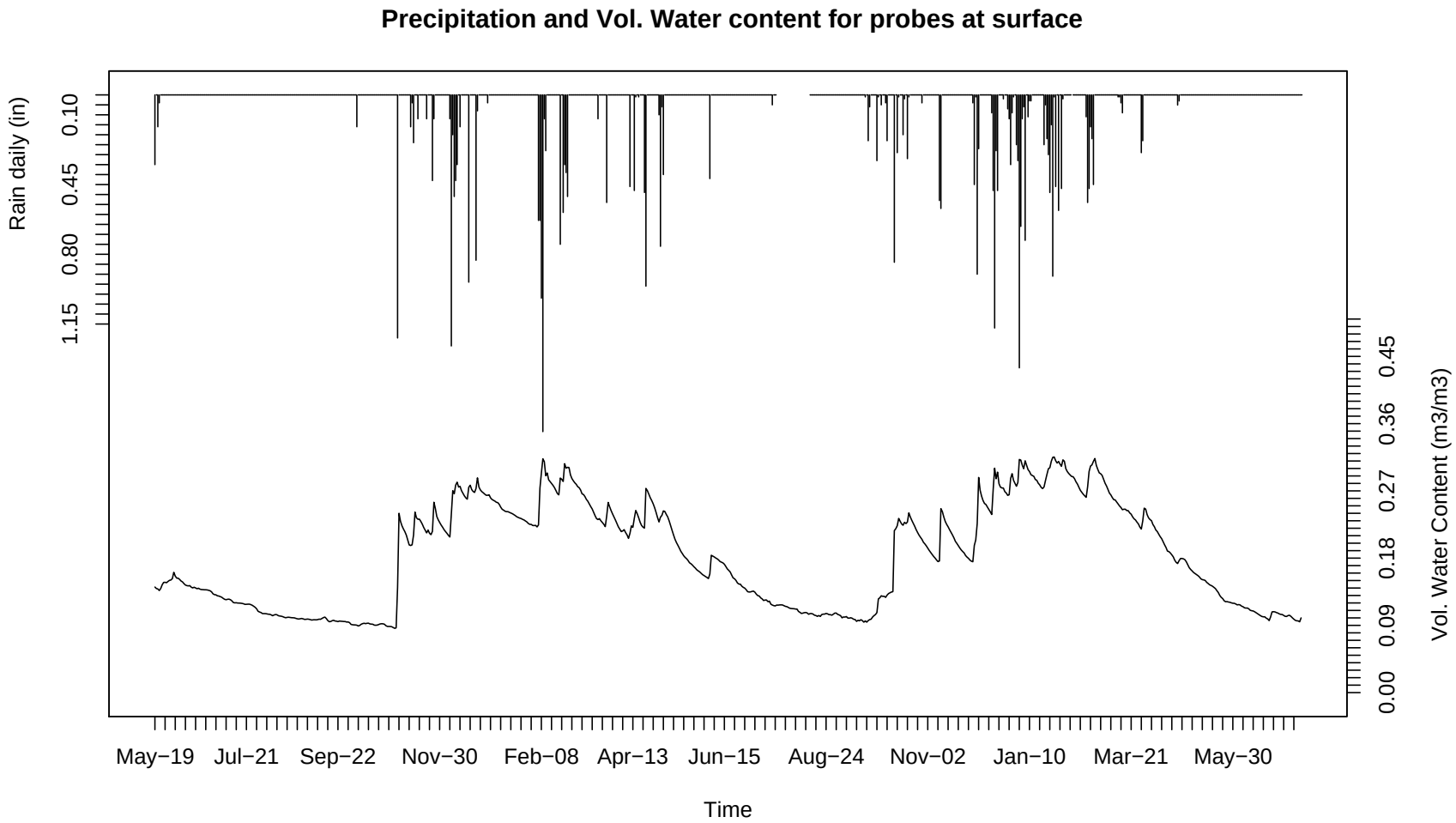
Query

```
-- start the long-version of the table
DROP TABLE cations_long ;
CREATE TABLE cations_long as
SELECT id, 'Extractable'::varchar as method, 'Na'::varchar as cation,
depth::integer, ex_na::numeric as value from cations_wide;

-- add an additional variable
INSERT INTO cations_long
SELECT id, 'Extractable'::varchar as method, 'Ca'::varchar as cation,
depth::integer, ex_ca::numeric as value from cations_wide;

-- add an additional variable
INSERT INTO cations_long
SELECT id, 'Extractable'::varchar as method, 'Mg'::varchar as cation,
depth::integer, ex_mg::numeric as value from cations_wide;

-- add an additional variable
INSERT INTO cations_long
SELECT id, 'Extractable'::varchar as method, 'K'::varchar as cation,
depth::integer, ex_k::numeric as value from cations_wide;

-- add an additional variable
INSERT INTO cations_long
SELECT id, 'Soluble'::varchar as method, 'Na'::varchar as cation,
depth::integer, water_na::numeric as value from cations_wide;

[...]
```

Sensor Installation



Raw Sensor Data



Data in Long Format

Sensor Data

date_time	probe_id	pedon_id	**value**
2007-03-01 00:00:00	m1	d4g1	
2007-03-01 00:00:00	m2	d4g1	
2007-03-01 00:00:00	m3	d4g1	
2007-03-01 00:00:00	m4	d4g1	
2007-03-01 00:00:00	m5	d4g1	
2007-03-01 01:00:00	m1	d4g1	
2007-03-01 01:00:00	m2	d4g1	
2007-03-01 01:00:00	m3	d4g1	
2007-03-01 01:00:00	m4	d4g1	
2007-03-01 01:00:00	m5	d4g1	
2007-03-01 02:00:00	m1	d4g1	0.313
2007-03-01 02:00:00	m2	d4g1	0.311
2007-03-01 02:00:00	m3	d4g1	0.394
2007-03-01 02:00:00	m4	d4g1	0.346
2007-03-01 02:00:00	m5	d4g1	0.425
2007-03-01 03:00:00	m1	d4g1	0.312
2007-03-01 03:00:00	m2	d4g1	0.309
2007-03-01 03:00:00	m3	d4g1	0.394
2007-03-01 03:00:00	m4	d4g1	0.346
2007-03-01 03:00:00	m5	d4g1	0.424

[...]

Data in Wide Format

Sensor Data

date_time	pedon_id	m1	m2	m3	m4	m5
2007-03-01 00:00:00	d4g1					
2007-03-01 01:00:00	d4g1					
2007-03-01 02:00:00	d4g1	0.313	0.311	0.394	0.346	0.425
2007-03-01 03:00:00	d4g1	0.312	0.309	0.394	0.346	0.424
2007-03-01 04:00:00	d4g1					
2007-03-01 05:00:00	d4g1	0.311	0.307	0.394	0.345	0.424
2007-03-01 06:00:00	d4g1	0.311	0.306	0.394	0.345	0.424
2007-03-01 07:00:00	d4g1	0.311	0.305	0.394	0.345	0.424
2007-03-01 08:00:00	d4g1	0.31	0.304	0.394	0.345	0.422
2007-03-01 09:00:00	d4g1	0.31	0.303	0.394	0.344	0.421
2007-03-01 10:00:00	d4g1					
2007-03-01 11:00:00	d4g1	0.309	0.302	0.393	0.345	0.421
2007-03-01 12:00:00	d4g1					
2007-03-01 13:00:00	d4g1					
2007-03-01 14:00:00	d4g1	0.309	0.301	0.394	0.344	0.421
2007-03-01 15:00:00	d4g1					
2007-03-01 16:00:00	d4g1					
2007-03-01 17:00:00	d4g1					
2007-03-01 18:00:00	d4g1					
2007-03-01 19:00:00	d4g1	0.31	0.299	0.393	0.344	0.42
2007-03-01 20:00:00	d4g1					
2007-03-01 21:00:00	d4g1	0.309	0.298	0.394	0.344	0.421
2007-03-01 22:00:00	d4g1	0.309	0.298	0.394	0.344	0.421
2007-03-01 23:00:00	d4g1	0.308	0.297	0.394	0.344	0.42

Conversion By SQL

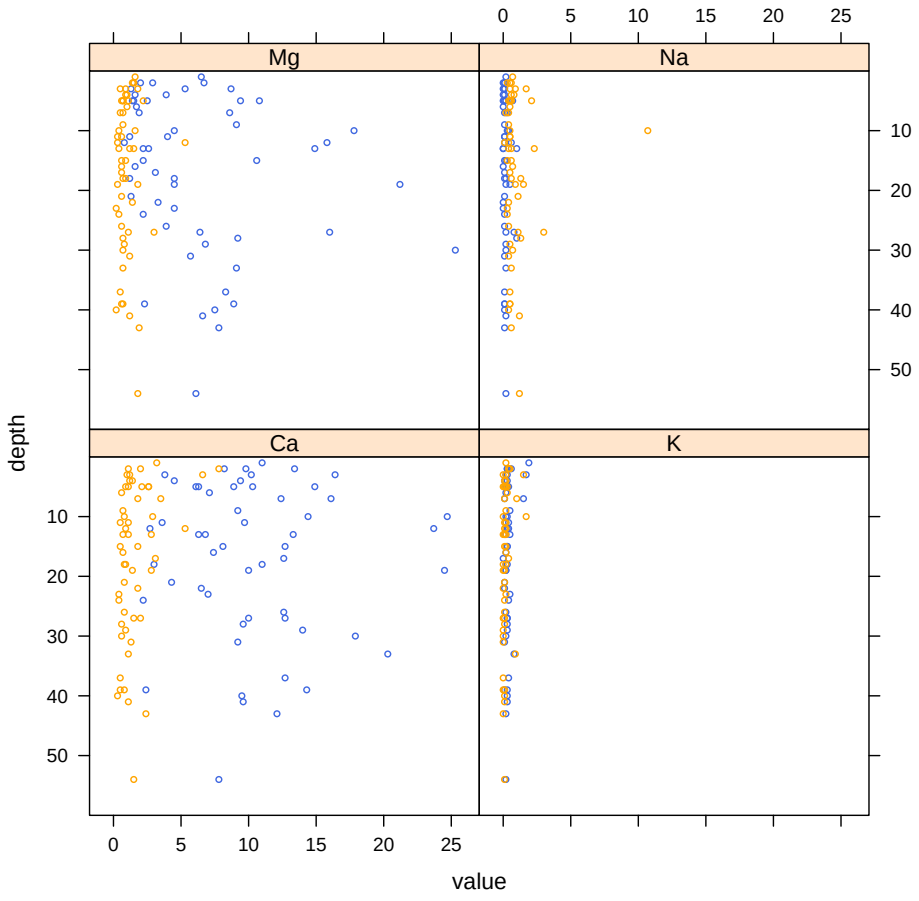
Query

```
SELECT a.date_time, a.pedon_id, m1, m2, m3, m4, m5 from
        (
        SELECT date_time, pedon_id, value as m1
        FROM smd
        WHERE probe_id = 'm1'
        ) as a
JOIN
        (
        SELECT date_time, value as m2
        FROM smd
        WHERE probe_id = 'm2'
        ) as b
ON a.date_time = b.date_time
JOIN
        (
        SELECT date_time, value as m3
        FROM smd
        WHERE probe_id = 'm3'
        ) as c
ON a.date_time = c.date_time
JOIN
        (
        SELECT date_time, value as m4
        FROM smd
        WHERE probe_id = 'm4'
        ) as d
ON a.date_time = d.date_time
JOIN
        (
        SELECT date_time, value as m5
        FROM smd
        WHERE probe_id = 'm5'
        ) as e
ON a.date_time = e.date_time ;
```

Conversion in R

```
# load libraries
library(lattice)            # comes with R
library(reshape)          # will need to install this

# load data:
wide <- read.csv('cations_wide.csv')
long <- read.csv('cations_long.csv')

# use the reshape package to convert wide to long format
wide.melted <- melt(wide, id=c('id', 'depth' ))

# summarize data:
xyplot(depth ~ value | cation, groups=method,
data=long, ylim=c(60,0), cex=0.5, col=c('RoyalBlue','Orange'))
```

Plotting Long-Format Data in R: Lattice Graphics




```
# summarize data:
xyplot(depth ~ value | cation, groups=method,
data=long, ylim=c(60,0), cex=0.5, col=c('RoyalBlue','Orange'))
```