UCDAVIS UNIVERSITY OF CALIFORNIA

QBE Lab

Making the Martyn Method a Reality

Judy Brown, Ray Martyn, Forrest Nutter, Jim Stack & Carla Thomas

Neil McRoberts
Plant Pathology Department
UC Davis

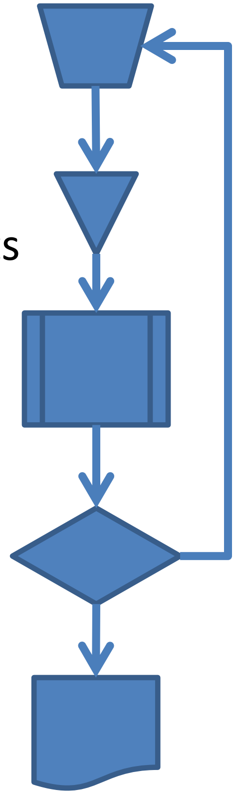
nmcroberts@ucdavis.edu

Aims

- Two inter-connected aims:
- (A) Examine whether there are generic types of disease
- (B) Examine the extent to which well-informed raters agree on the important properties of diseases
  - Note, A is an emergent property of the extent to which B is true.

Algorithm design

- A blend of human and computer components



Outlining of data and problem

Merging of individual data sets

Statistical pattern search and analysis

Deliberation

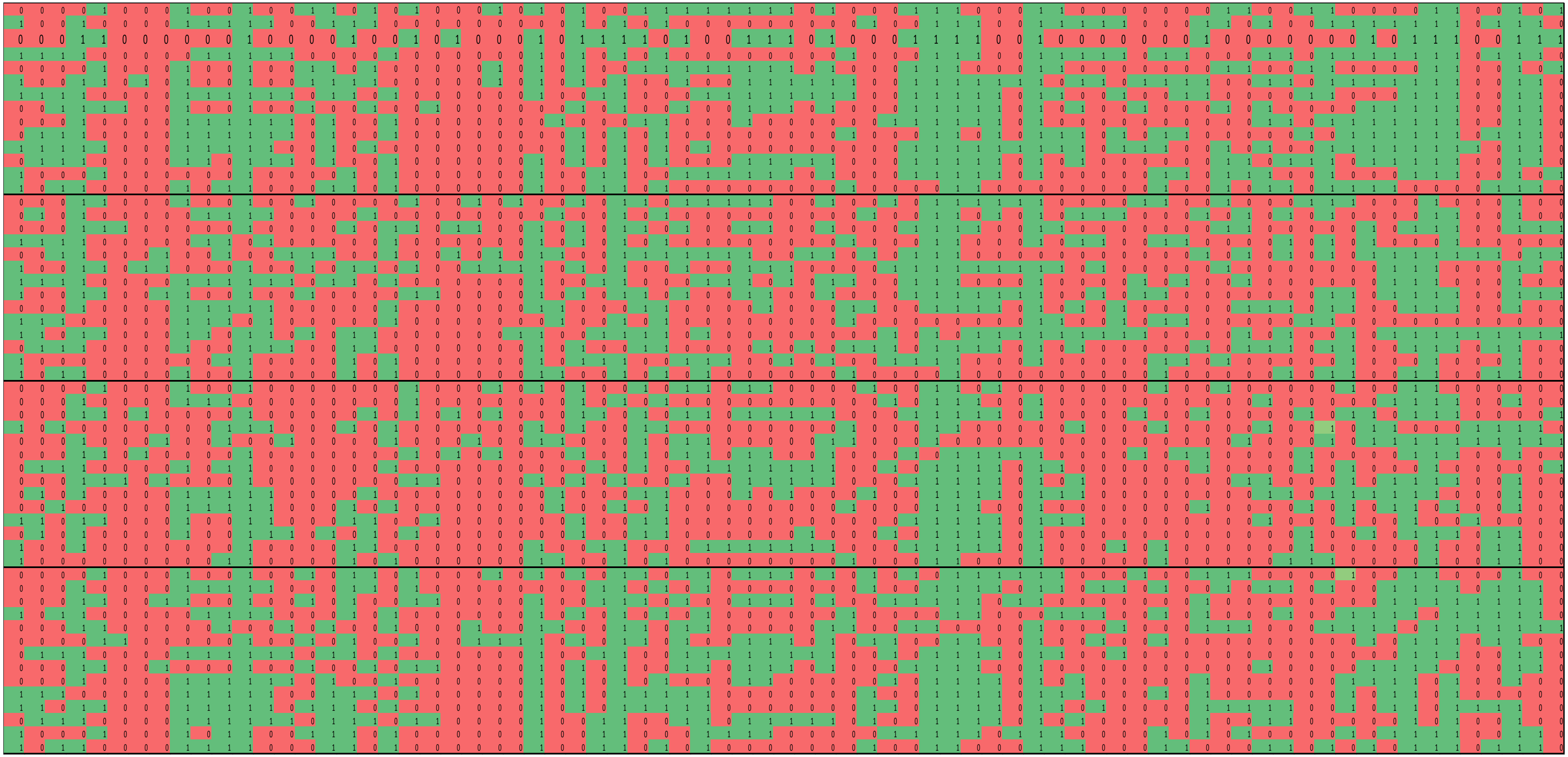
Final documentation

Double-blind approach: analyst did not know identities of raters or diseases in the first iteration

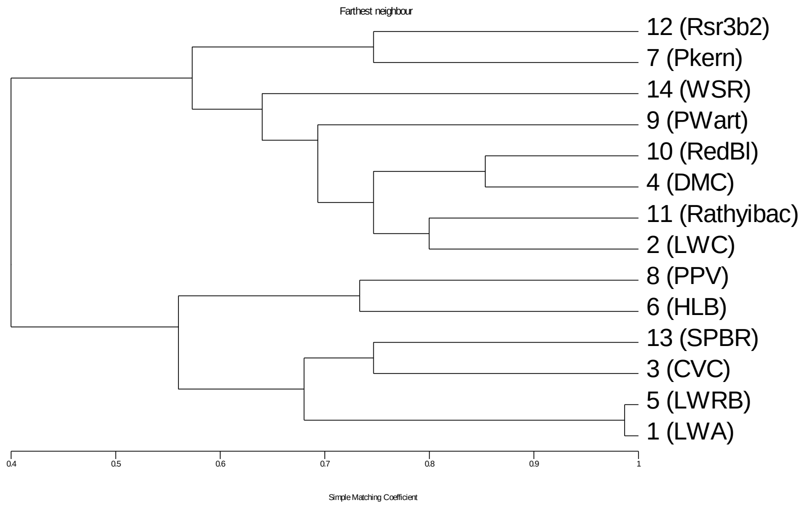
Data structure

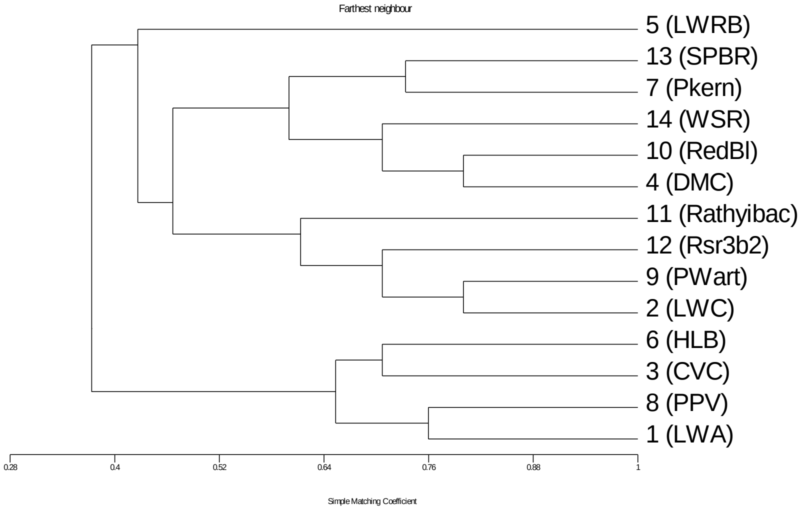
4 Raters × 14 Diseases × 75 variables (binary - 1/0)

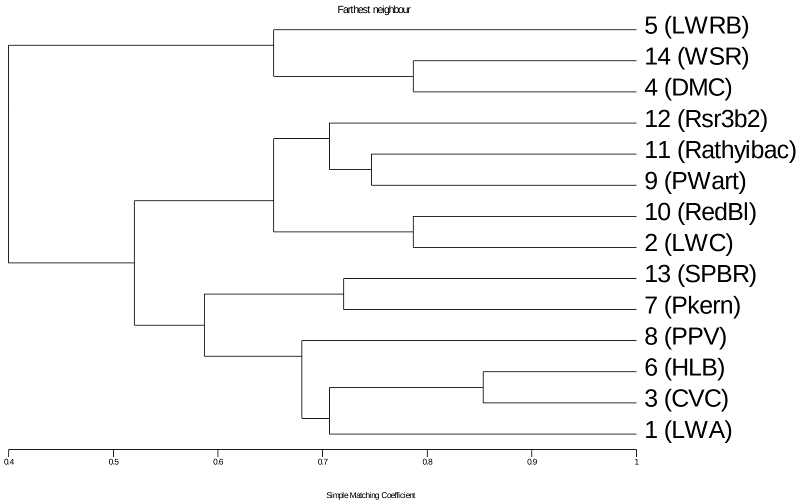
$2^{75} \approx 10^{21} – 10^{22}$ possible unique fingerprints

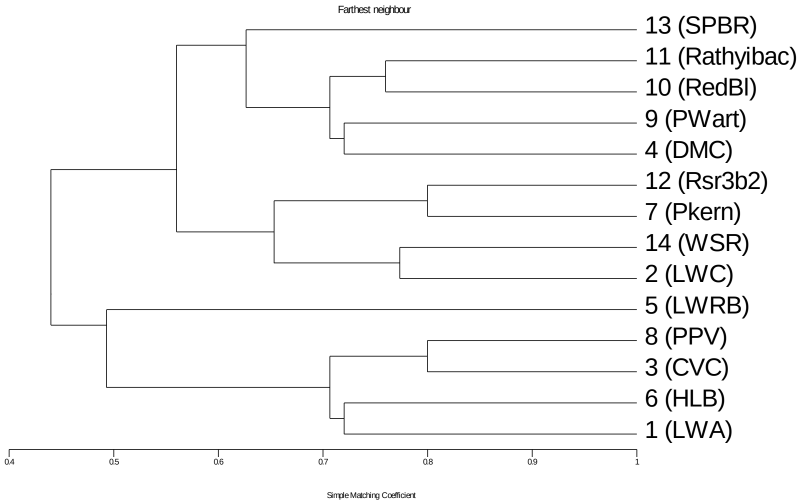


Individual rater disease dendrograms

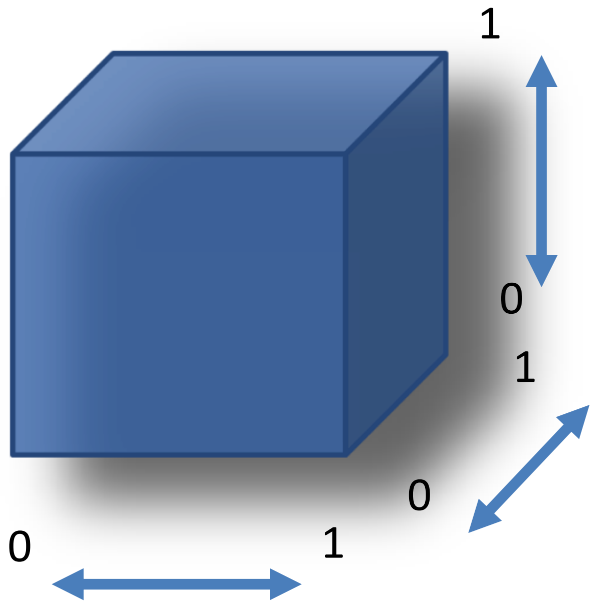








Alternative geometric interpretation

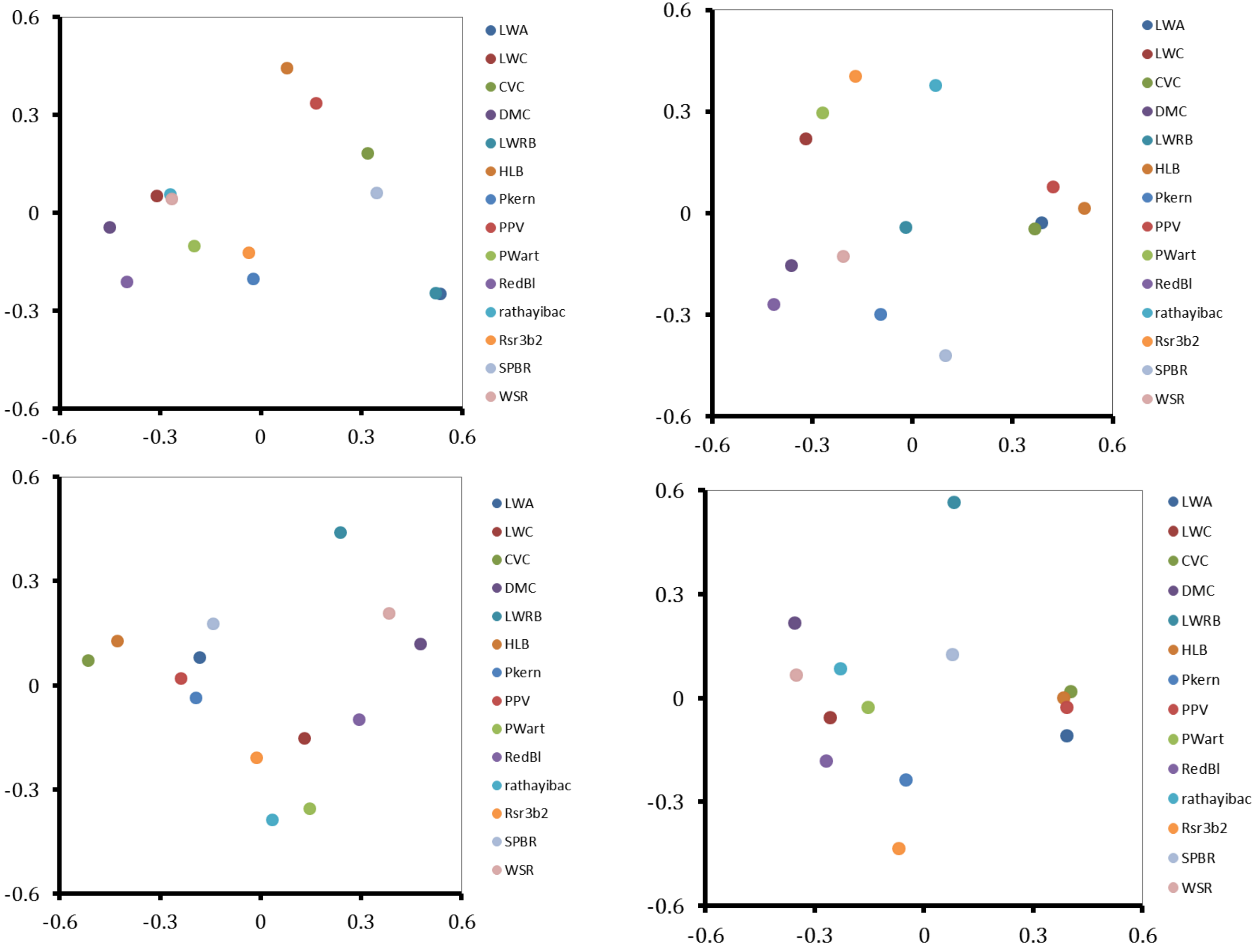


Can think of 3 variables , say, as defining 2^3 = 8 unique locations in a 3D space, or as the 8 vertices on a cube with sides of length = 1.

We have a 2^{75} hypercube.

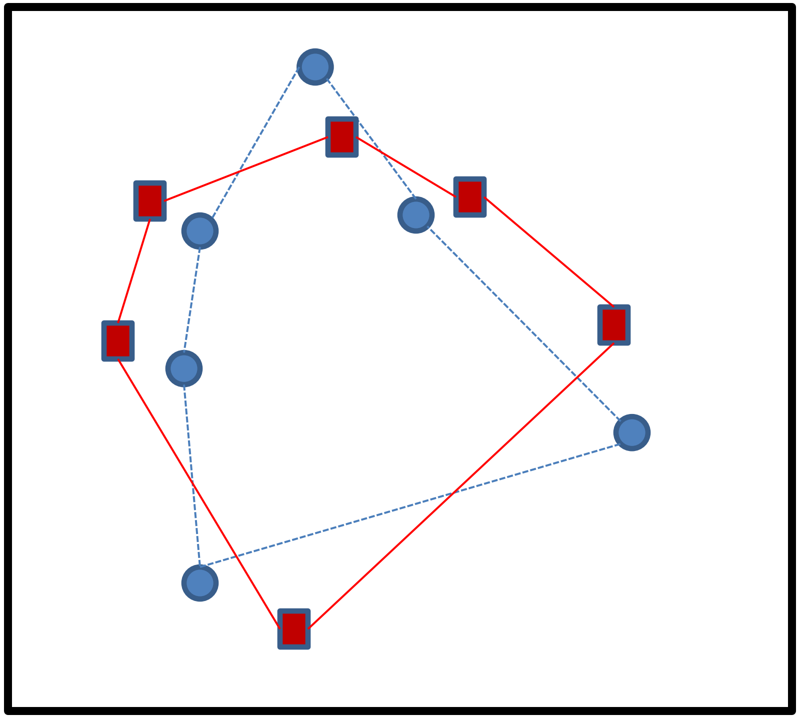
The Martyn theorem is that disease will occupy a limited set of close vertices and raters will put diseases on similar sets of vertices .

Projection of “distances” between diseases according to four rates



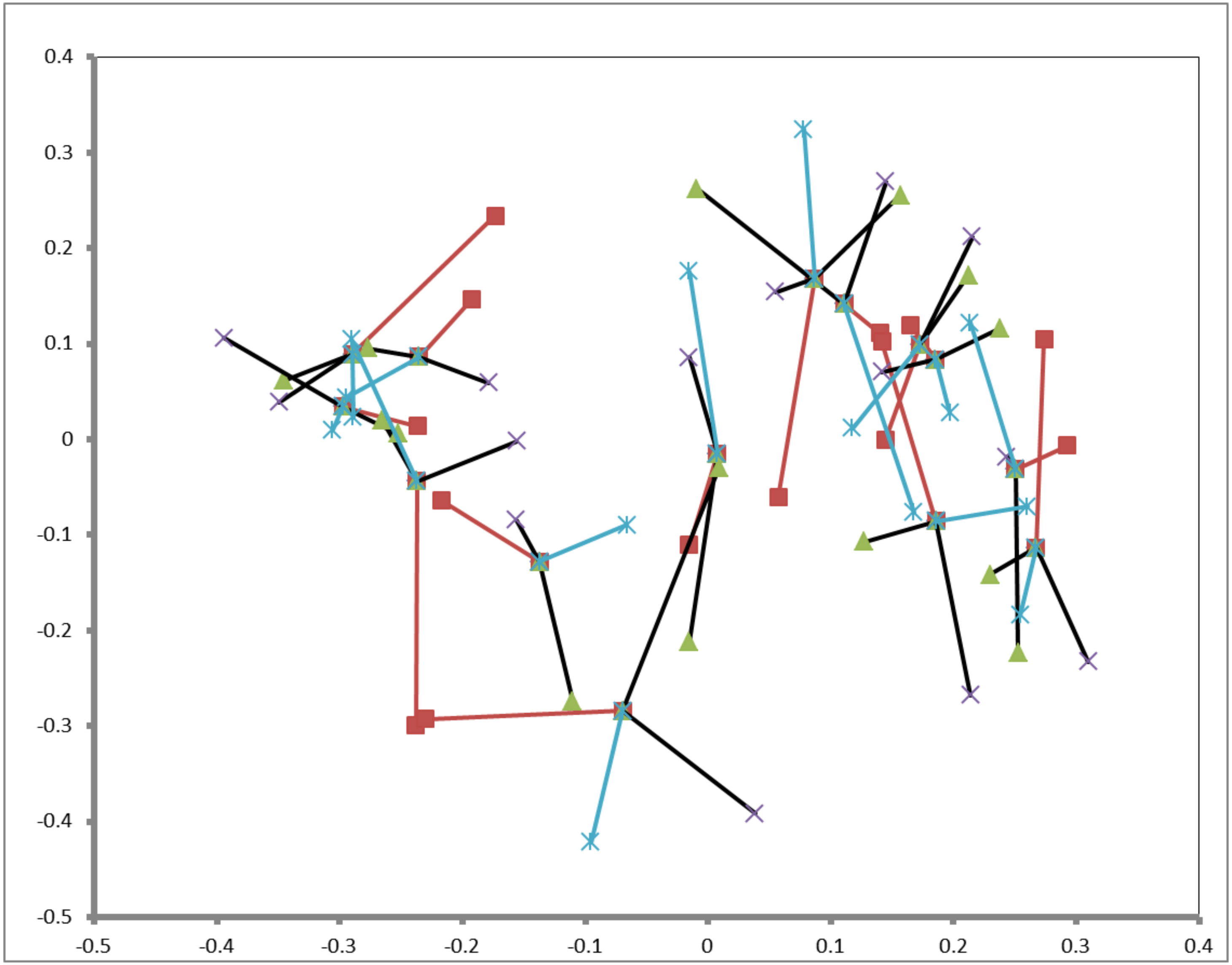
Generalized Procrustes analysis

Procrustes analysis is a technique for matching two sets of points to each another and finding which points are responsible for lack of fit. Generalized Procrustes analysis extends the approach to more than two data sets



Can calculate point by point distance once the best –possible overall fit has been found.
Distances (or residuals) calculated as "squared" values so sum of squares gives overall lack of fit and each point
can be attributed a %S.S.

Considering all raters together



ANOVA associated with GPA

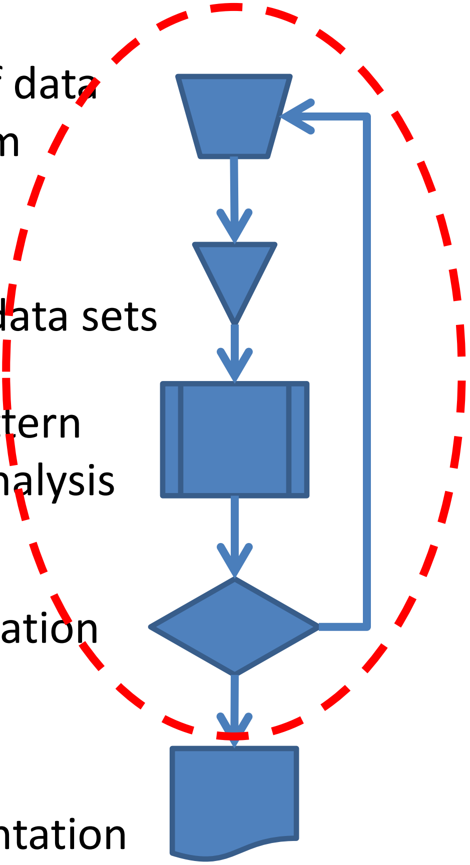
	fit	residual	total	%residual
LWA	5.817	2.578	8.396	10.844
LWC	4.134	0.274	4.409	1.154
CVC	8.957	0.554	9.511	2.330
DMC	8.415	1.782	10.198	7.497
LWRB	8.547	3.476	12.023	14.621
HLB	9.173	1.222	10.395	5.141
Pkern	0.029	2.989	3.018	12.573
PPV	6.292	0.415	6.707	1.746
PWart	3.945	1.072	5.017	4.510
RedBl	6.380	1.605	7.985	6.749
Rathyibac	3.232	2.477	5.709	10.417
Rsr3b2	3.586	2.282	5.868	9.596
SPBR	3.530	1.032	4.562	4.340
WSR	4.185	2.017	6.202	8.482
Totals	76.223	23.777	100.000	

	% fit S.S.
rater a	21.39864
rater b	24.88994
rater c	28.09269
rater d	25.61872

LWA, Laurel wilt of Avacado; **LWC**, Late wilt of corn; **CVC**, citrus variegated chlorosis; **DMC**, downy mildew of corn; **LWRB**, Laurel wilt of red Bay; **HLB**, Huanglongbing; **Pkern**, *Phytophthora kernoviae*; **PPV**, Plum Pox virus; **PWart**, Potato wart; **RedBl**, Red Blotch; **Rathayibac**, *Rathayibacter toxicus*; Rsr3b2, *Ralstonia solanacearum* race 3 biovar 2; **SPBR**, Scots Pine blister rust; **WSR**, wheat stem rust.

Algorithm design

- A blend of human and computer components



Outlining of data and problem

Merging of individual data sets

Statistical pattern search and analysis

Deliberation

Final documentation

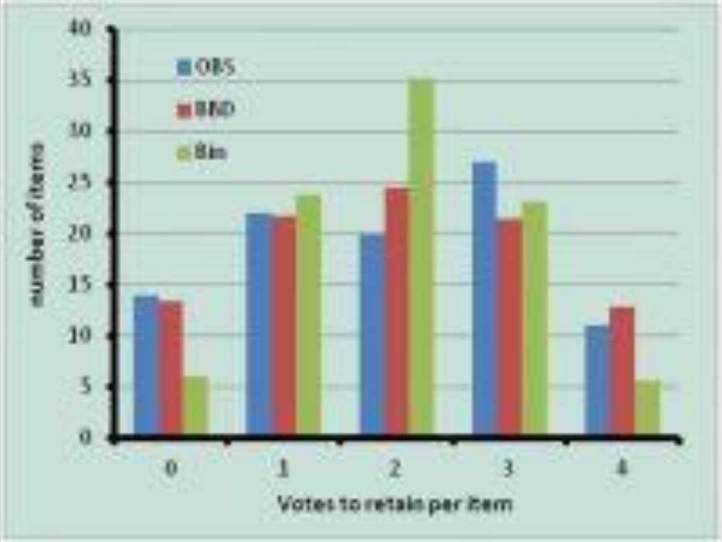
Double-blind approach: analyst did not know identities of raters or diseases in the first iteration

Second iteration

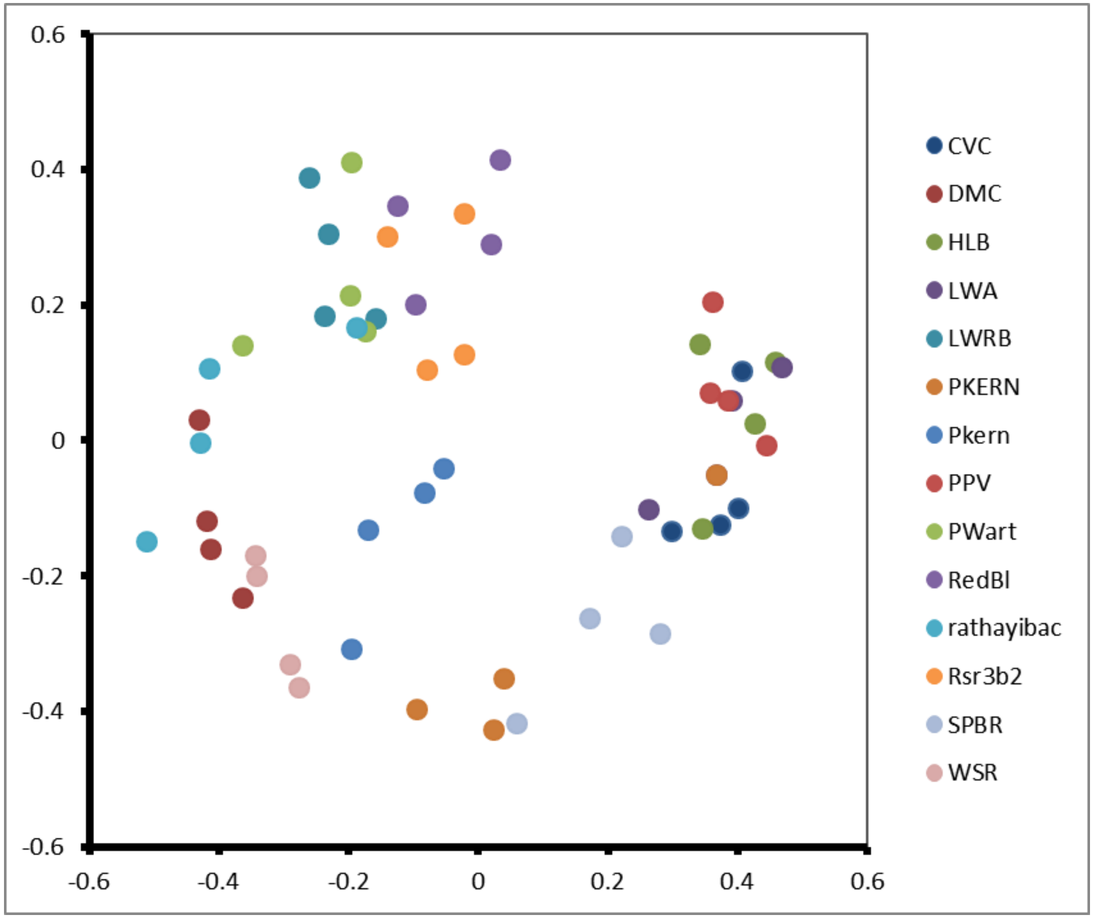
- Raters work independently to reduce number of variables to more manageable, core set
- Reduced data sets compared with each other and original analysis to determine
  - If sufficient information has been retained
  - If consensus still exists

Second round results

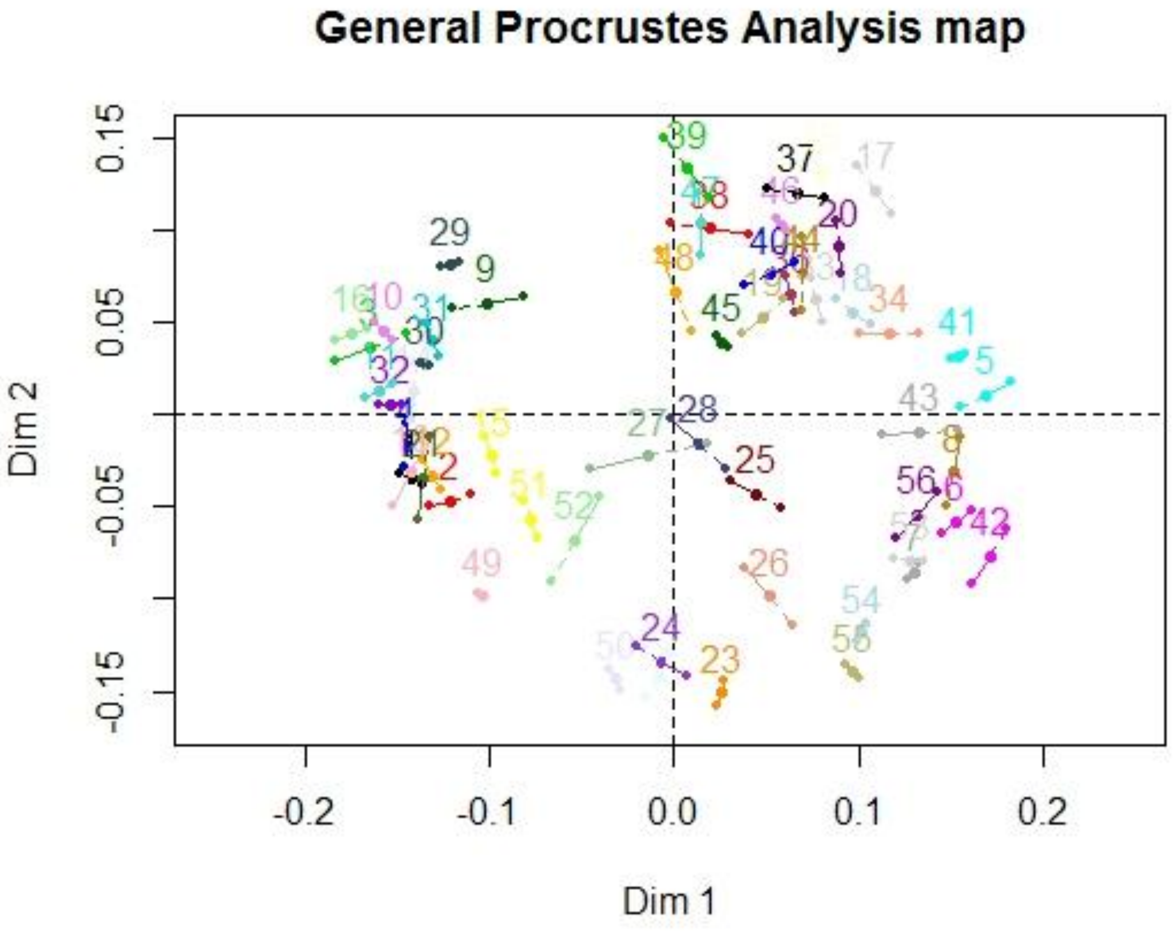
- Reduced set of 36 candidate key items suggested by combining the lists from 4 raters
- The level of agreement among raters was greater than would be expected by independent, random selection processes.
- Using the reduced data set to analyze similarity among the 14 diseases resulted in strong clustering of the 4 replicates of each disease. Evidence for the existence of generic disease types was retained in the reduced data set.
- The agreement between the whole- and reduced data disease similarities was very high (generally in the range of 95-100% fit S.S. in the GPA.



Similarity analysis based on reduced data set



Second round compared with first

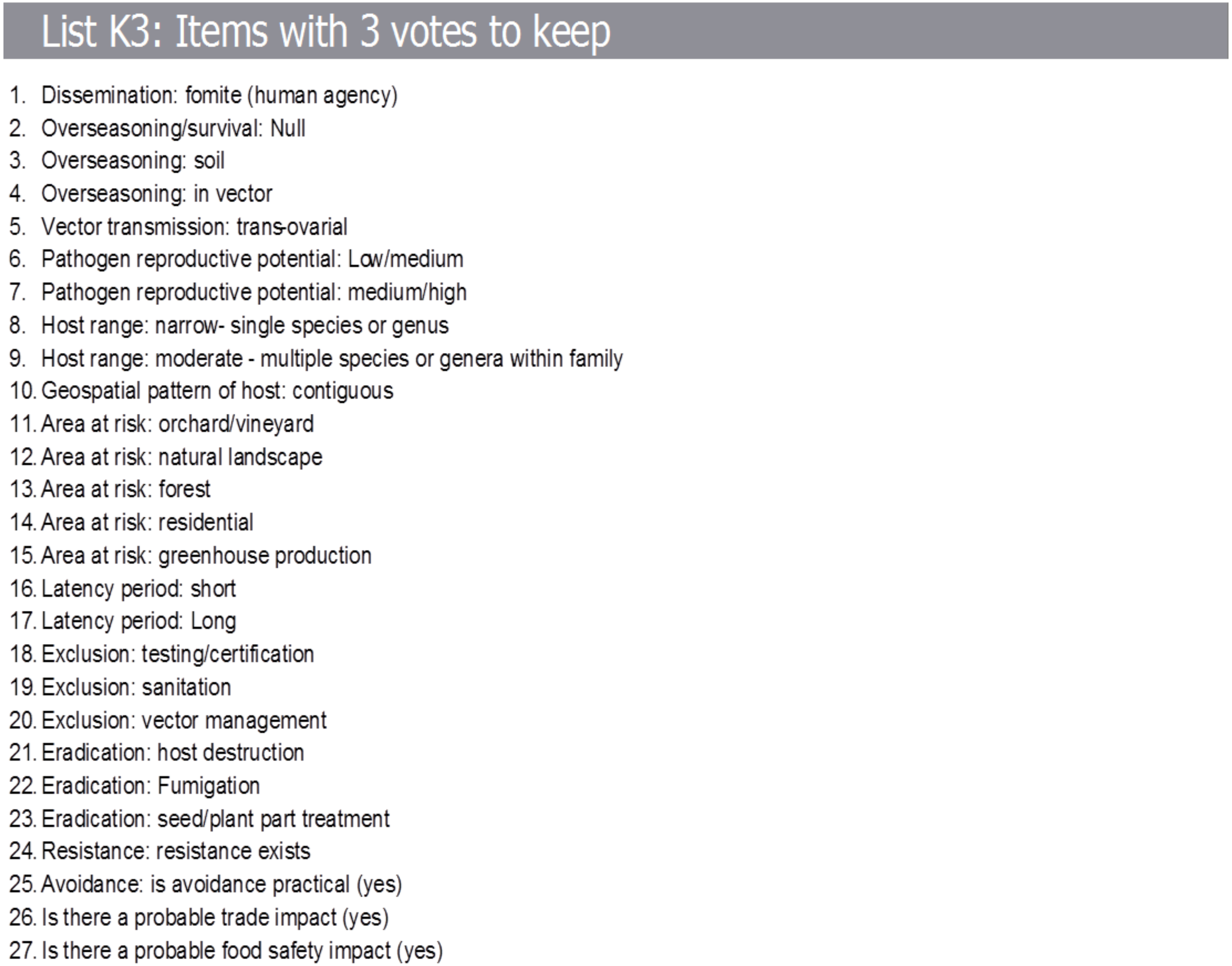


Item analysis

List D: Items to be discarded

1. Method of dissemination: Circulative in vector
2. Method of dissemination: Semi-persistent (in vector)
3. Over-seasons in soil saprophytically
4. Vector transmission: Circulative (passes through gut)
5. Exclusion: surveillance and monitoring
6. Protection: Not known if efficacious protective treatment exists
7. Protection: Efficacious treatment known but not feasible or practical
8. Resistance: begin a race-specific breeding program
9. Crop insurance available? (no)
10. Would public assistance/compensation to growers be available (no)
11. Would and emergency declaration be a likely action (no)
12. Potential private economic impact (low)
13. Potential public economic impact (low)
14. Potential social impact (low)

Item analysis



List K3: Items with 3 votes to keep

1. Dissemination: fomite (human agency)
2. Overseasoning/survival: Null
3. Overseasoning: soil
4. Overseasoning: in vector
5. Vector transmission: trans-ovarial
6. Pathogen reproductive potential: Low/medium
7. Pathogen reproductive potential: medium/high
8. Host range: narrow- single species or genus
9. Host range: moderate - multiple species or genera within family
10. Geospatial pattern of host: contiguous
11. Area at risk: orchard/vineyard
12. Area at risk: natural landscape
13. Area at risk: forest
14. Area at risk: residential
15. Area at risk: greenhouse production
16. Latency period: short
17. Latency period: Long
18. Exclusion: testing/certification
19. Exclusion: sanitation
20. Exclusion: vector management
21. Eradication: host destruction
22. Eradication: Fumigation
23. Eradication: seed/plant part treatment
24. Resistance: resistance exists
25. Avoidance: is avoidance practical (yes)
26. Is there a probable trade impact (yes)
27. Is there a probable food safety impact (yes)

Item analysis

List K4: Items with maximum votes for retention

1. Dissemination: Wind
2. Dissemination: Rain
3. Dissemination: Seed/plant parts
4. Dissemination: Vectored
5. Dissemination: Soil
6. Disease cycle: polycyclic
7. Disease cycle: monocyclic
8. Host range: broad - multiple plant families
9. Area at risk: agricultural field
10. Exclusion: quarantine
11. Protection: known efficacious treatment exists

Summary

- The approach appears to have promise
- Four raters generated reasonably robust clustering of disease recovery plans
- We were able to identify a small number of variables that differentiate among disease types
- Prototype generic plans might be made from these variables and a set of others which all plans will include.