

Johnson's Hierarchical Clustering

Steve Borgatti

Clustering Overview

- Given a 1-mode proximity matrix as input
 - Item by item matrix
- Then cluster items into groups
- Two major kinds of clustering routines
 - Process-defined
 - Outcome-defined
- Typical result is a partition (exhaustive mutually exclusive groups) or a set of hierarchically nested partitions

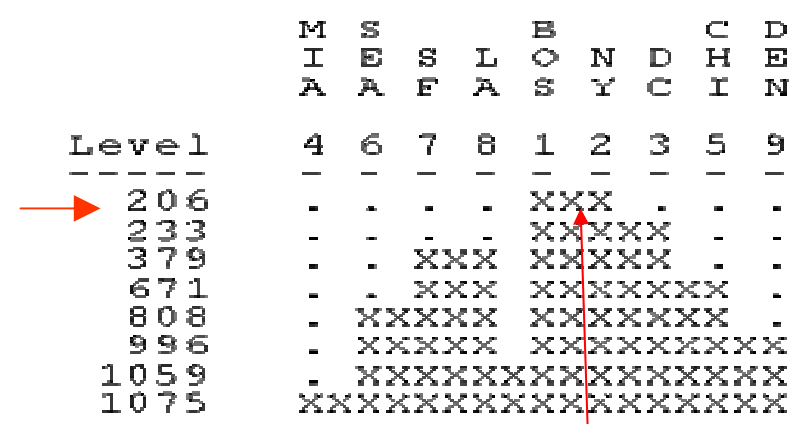
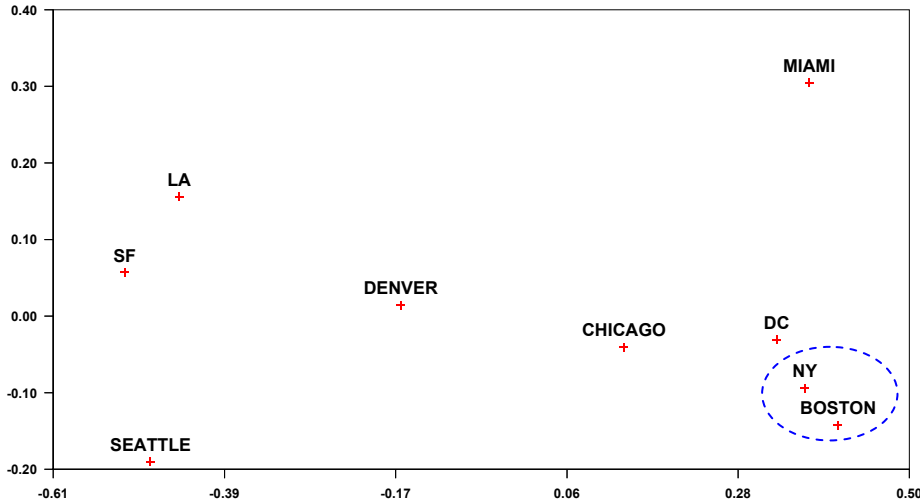
Partitions

- Partition P is just an assignment of nodes to classes
 - $P(i)$ gives the class of node i
 - Every node assigned to one & only one class
- A partition P is nested in partition M if for all nodes i and j , $P(i)=P(j)$ implies $M(i)=M(j)$
- Trivial partitions
 - Identity: $P(i) = i$ for all i
 - Complete: $P(i) = 1$ for all i

Johnson's Hierarchical Clustering

- Output is a set of nested partitions, starting with identity partition and ending with the complete partition
- Different flavors based on how distance from a point to a cluster is defined
 - Single linkage, connectedness, nearest neighbor, minimum
 - Complete linkage, diameter, farthest neighbor, maximum
 - Average, median, centroid, etc.

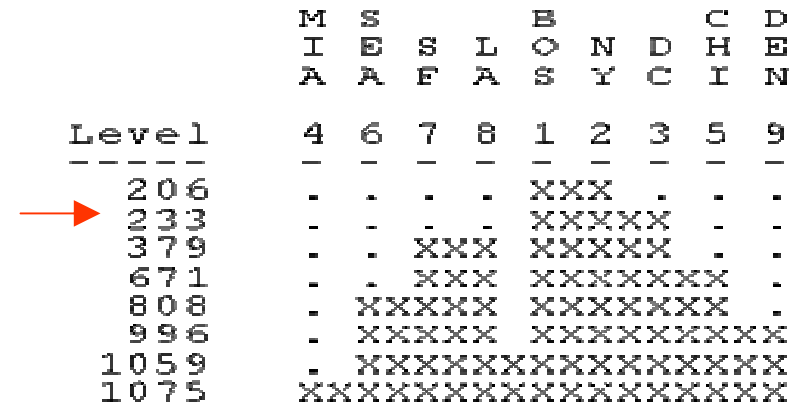
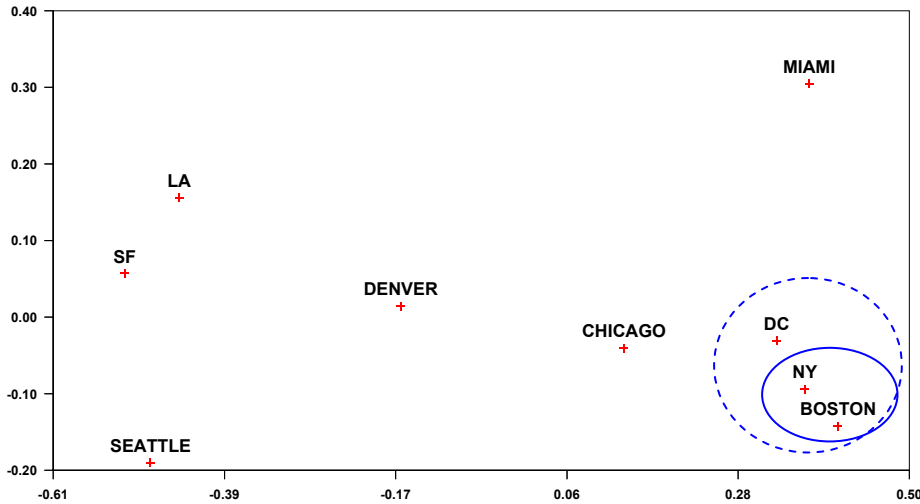
Johnson's Hierarchical Clustering via Minimum Method (single linkage)



	BOS	NY	DC	MIA	CHI	SEA	SF	LA	DEN
BOS	0	206	429	1504	963	2976	3095	2979	1949
NY		0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIA	1504	1308	1075	0	1329	3273	3053	2687	2037
CHI	963	802	671	1329	0	2013	2142	2054	996
SEA	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DEN	1949	1771	1616	2037	996	1307	1235	1059	0

Closest distance is NY-BOS = 206, so merge these.

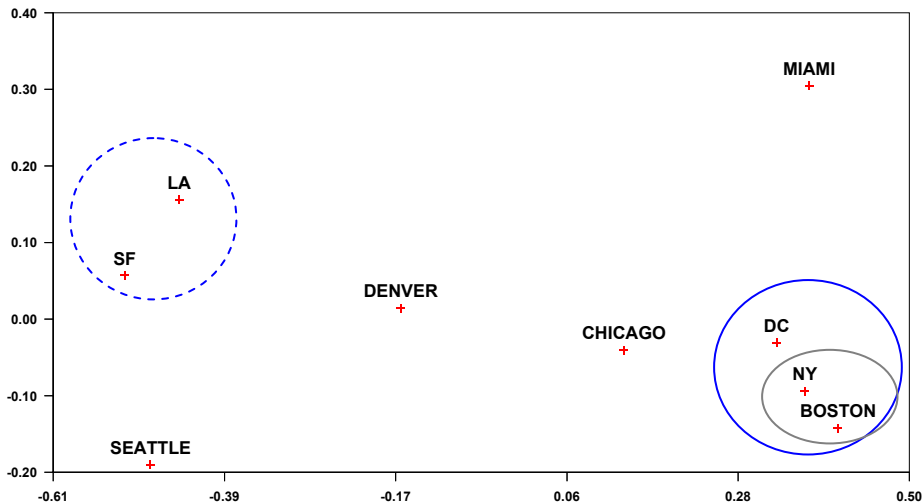
Johnson's Hierarchical Clustering via Minimum Method (single linkage)



	BOS N Y	DC	MIA	CHI	SEA	SF	LA	DEN
BOS/ NY	0	233	1308	802	2815	2934	2786	1771
DC	233	0	1075	671	2684	2799	2631	1616
MIA	1308	1075	0	1329	3273	3053	2687	2037
CHI	802	671	1329	0	2013	2142	2054	996
SEA	2815	2684	3273	2013	0	808	1131	1307
SF	2934	2799	3053	2142	808	0	379	1235
LA	2786	2631	2687	2054	1131	379	0	1059
DEN	1771	1616	2037	996	1307	1235	1059	0

Closest pair is DC to BOSNY combo @ 233. So merge these.

Johnson's Hierarchical Clustering via Minimum Method (single linkage)



Level

 206
 233
 379
 671
 808
 996
 1059
 1075

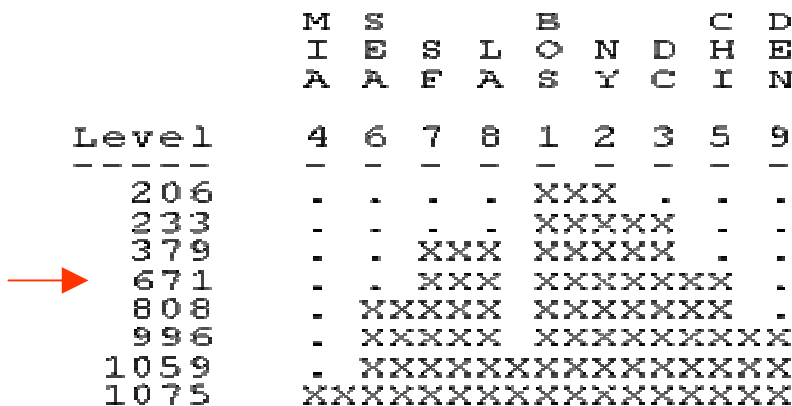
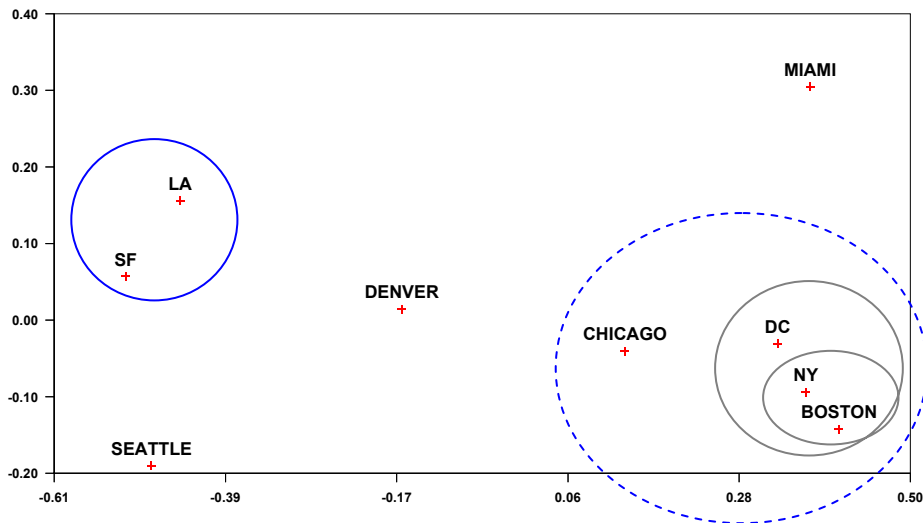


```

M S B C D
I E S L O N D H E
A A F A S Y C I N
4 6 7 8 1 2 3 5 9
- - - - - - - - -
. . . . XXX . . .
. . . . XXXXX . .
. . XXX XXXXX . .
. . XXX XXXXXXXX .
. XXXXX XXXXXXXX .
. XXXXXXX XXXXXXXXXXXX
. XXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXX
  
```

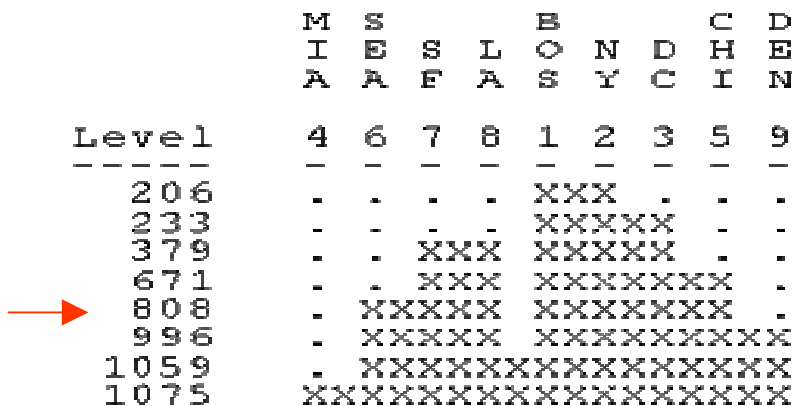
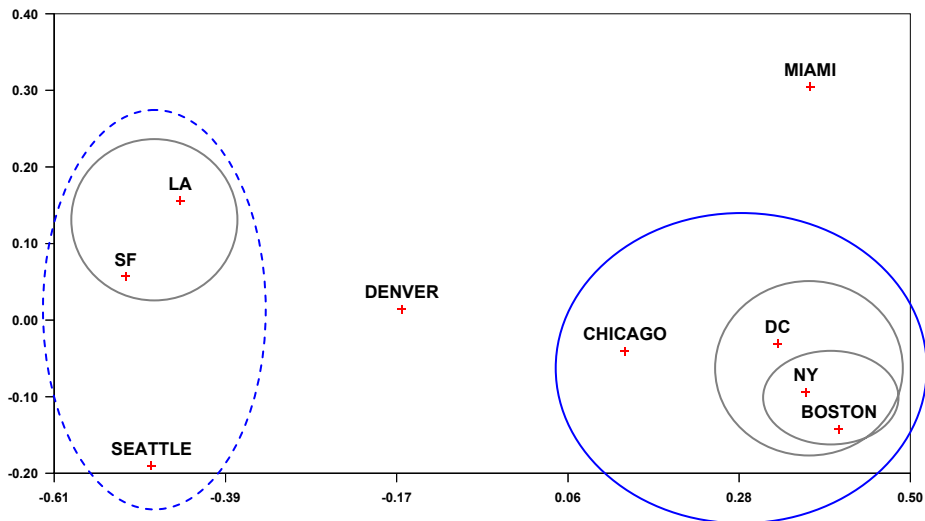
	BOS/ NY/ DC	MIA	CHI	SEA	SF	LA	DEN
BOS/NY DC	0	1075	671	2684	2799	2631	1616
MIA	1075	0	1329	3273	3053	2687	2037
CHI	671	1329	0	2013	2142	2054	996
SEA	2684	3273	2013	0	808	1131	1307
SF	2799	3053	2142	808	0	379	1235
LA	2631	2687	2054	1131	379	0	1059
DEN	1616	2037	996	1307	1235	1059	0

Johnson's Hierarchical Clustering via Minimum Method (single linkage)



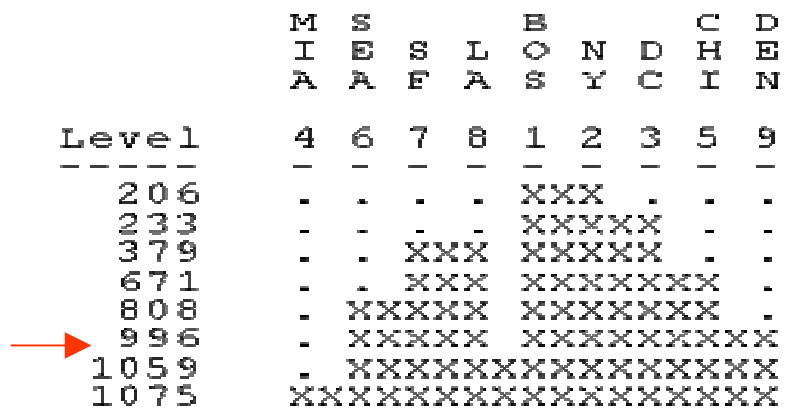
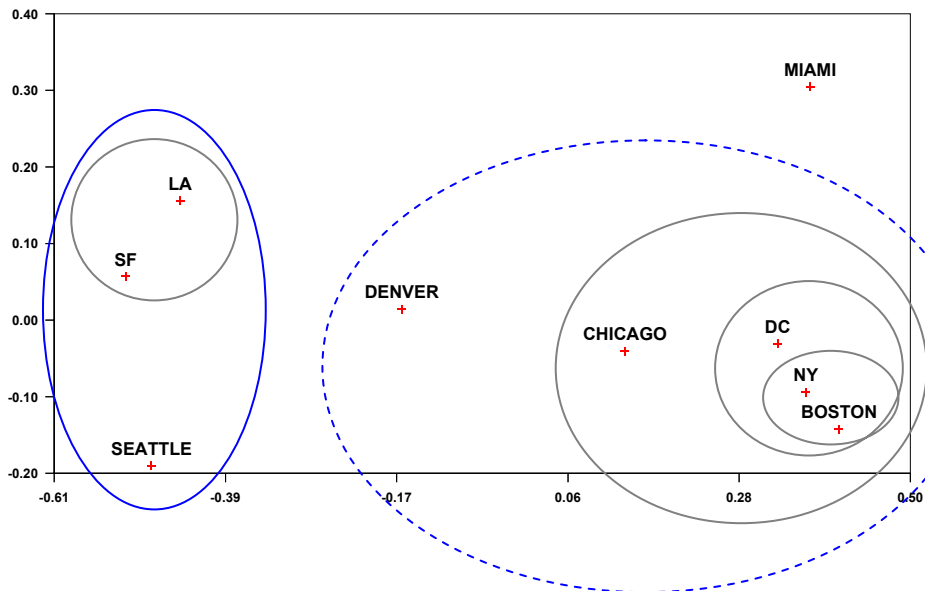
	BOS/ NY/DC	MIA	CHI	SEA	SF/LA	DEN
BOS/NY/DC	0	1075	671	2684	2631	1616
MIA	1075	0	1329	3273	2687	2037
CHI	671	1329	0	2013	2054	996
SEA	2684	3273	2013	0	808	1307
SF/LA	2631	2687	2054	808	0	1059
DEN	1616	2037	996	1307	1059	0

Johnson's Hierarchical Clustering via Minimum Method (single linkage)



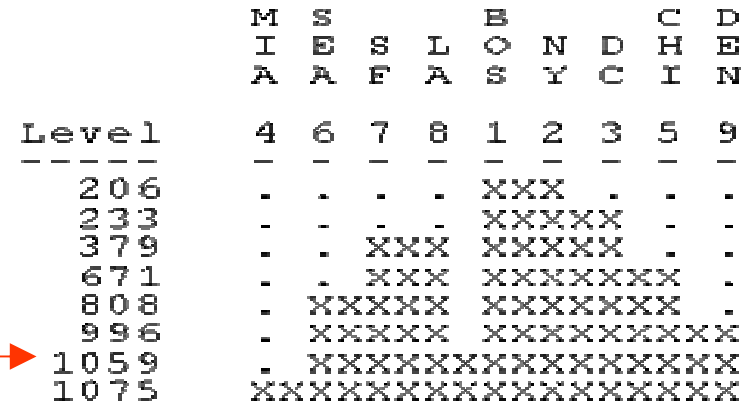
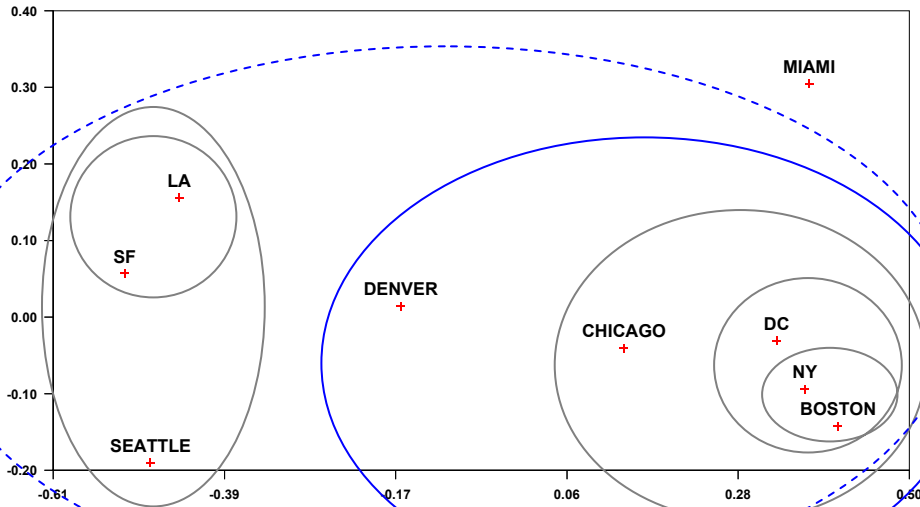
	BOS/ NY/D C/ CHI	MIA	SEA	SF/L A	DEN
BOS/NY/DC/C HI	0	1075	2013	2054	996
MIA	1075	0	3273	2687	2037
SEA	2013	3273	0	808	1307
SF/LA	2054	2687	808	0	1059
DEN	996	2037	1307	1059	0

Johnson's Hierarchical Clustering via Minimum Method (single linkage)



	BOS/ NY/D C/C HI	MIA	SF/L A/SE A	DEN
BOS/NY/DC/ CHI	0	1075	2013	996
MIA	1075	0	2687	2037
SF/LA/SEA	2054	2687	0	1059
DEN	996	2037	1059	0

Johnson's Hierarchical Clustering via Minimum Method (single linkage)

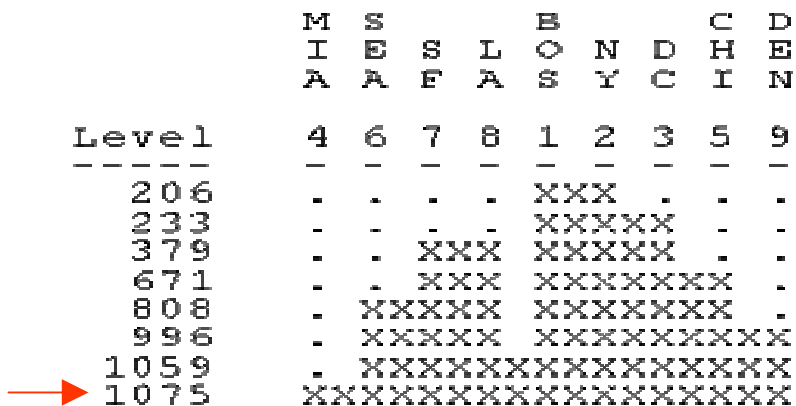
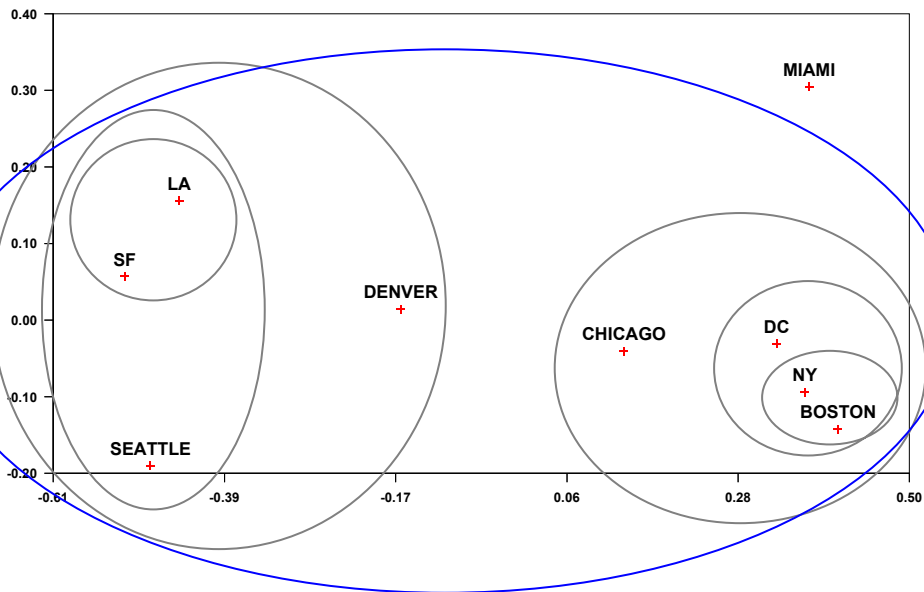


	BOS/ NY/D C/CHI /DEN	MIA	SF/LA /SEA
BOS/NY/DC/ CHI/DEN	0	1075	1059
MIA	1075	0	2687
SF/LA/SEA	1059	2687	0

Evidently, Denver is an East-coast city!

This is due to *chaining* and is a well-known issue with single-linkage clustering

Johnson's Hierarchical Clustering via Minimum Method (single linkage)



	BOS/ NY/D C/CH I/DE N/SF/ LA/S EA	MIA
BOS/NY/DC/CHI/DEN/SF/L A/SEA	0	1075
MIA	1075	0

In the final analysis, all items belong to one (very loose) cluster.

Geodesic Distances

		1 1 1 1 1 1 1 1 1																	
		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8
		H	B	C	P	P	J	P	A	M	B	L	D	J	H	G	S	B	R
		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	HOLLY	0	4	2	1	1	2	2	2	1	2	4	1	3	1	2	3	4	3
2	BRAZEY	4	0	5	5	5	6	4	5	3	4	1	4	3	4	2	1	1	2
3	CAROL	2	5	0	1	1	2	1	2	3	4	5	3	2	3	3	4	4	3
4	PAM	1	5	1	0	2	1	1	1	2	3	5	2	2	2	3	4	4	3
5	PAT	1	5	1	2	0	1	1	2	2	3	5	2	2	2	3	4	4	3
6	JENNIE	2	6	2	1	1	0	2	1	3	4	6	3	3	3	4	5	5	4
7	PAULINE	2	4	1	1	1	2	0	1	3	4	4	3	1	3	2	3	3	2
8	ANN	2	5	2	1	2	1	1	0	3	4	5	3	2	3	3	4	4	3
9	MICHAEL	1	3	3	2	2	3	3	3	0	1	3	1	2	1	1	2	3	2
10	BILL	2	4	4	3	3	4	4	4	1	0	4	1	3	1	2	3	4	3
11	LEE	4	1	5	5	5	6	4	5	3	4	0	4	3	4	2	1	1	2
12	DON	1	4	3	2	2	3	3	3	1	1	4	0	3	1	2	3	4	3
13	JOHN	3	3	2	2	2	3	1	2	2	3	3	3	0	3	1	2	2	1
14	HARRY	1	4	3	2	2	3	3	3	1	1	4	1	3	0	2	3	4	3
15	GERY	2	2	3	3	3	4	2	3	1	2	2	2	1	2	0	1	2	1
16	STEVE	3	1	4	4	4	5	3	4	2	3	1	3	2	3	1	0	1	1
17	BERT	4	1	4	4	4	5	3	4	3	4	1	4	2	4	2	1	0	1
18	RUSS	3	2	3	3	3	4	2	3	2	3	2	3	1	3	1	1	1	0

Hierarchical Clustering

```

P           M
A       J   I       B
C U   H E   C H   R S
A L   O N   B H A   B A T G J R
P R I P L N A I A R D L E Z E E O U
A O N A L I N L E R O E R E V R H S
T L E M Y E N L L Y N E T Y E Y N S

```

Level	5	3	7	4	1	6	8	0	9	4	2	1	7	2	6	5	3	8
1.000	XXXXX	XXX	XXX	XXXXXXXX	XXXXXXXX	XXXXX												
1.333	XXXXX	XXXXXXXX		XXXXXXXX	XXXXXXXX	XXXXX												
1.457	XXXXX	XXXXXXXX		XXXXXXXX	XXXXXXXX	XXXXXXXXXXXXXXXXXX												
1.481	XXXXXXXXXXXXXXXXXX			XXXXXXXX	XXXXXXXXXXXXXXXXXX													
2.723	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX			XXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXX													
3.142	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX			XXXXXXXXXXXXXXXXXX	XXXXXXXXXXXXXXXXXX													

