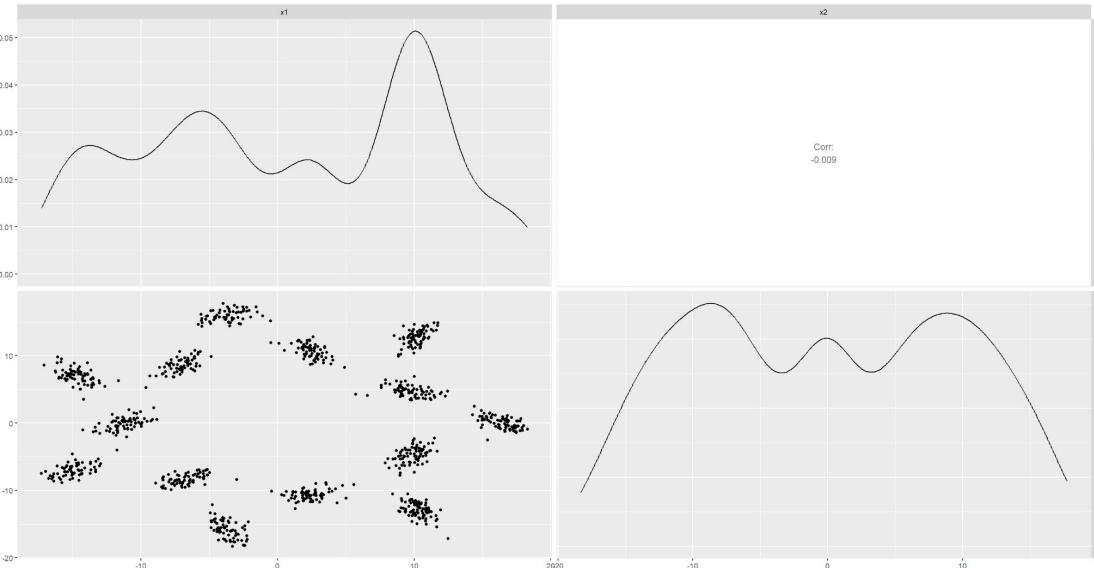


# Clustering analysis

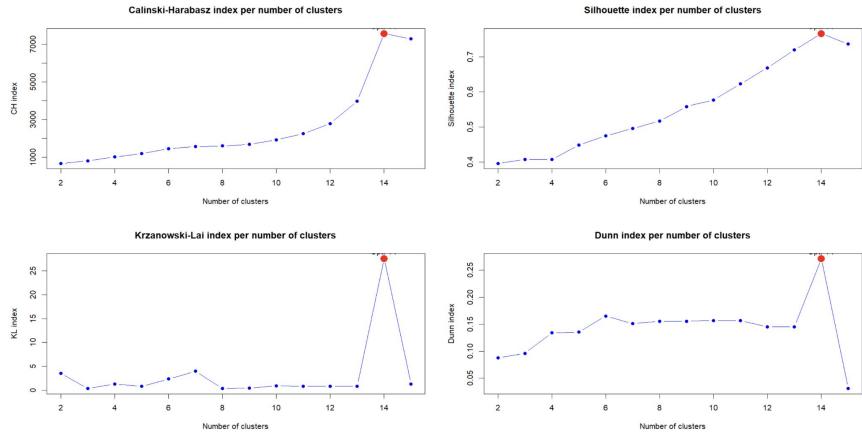
Daniele Banni

# WREATH DATASET



```
> summary(wreath)
      x1                      x2
  Min.   : -17.2744   Min.   : -18.30985
  1st Qu.: -7.4048   1st Qu.: -8.50031
  Median :  1.8301   Median : -0.06005
  Mean   :  0.5756   Mean   : -0.15089
  3rd Qu.:  9.9706   3rd Qu.:  8.47251
  Max.   : 18.3051   Max.   : 17.76186
```

# NB CLUST



```

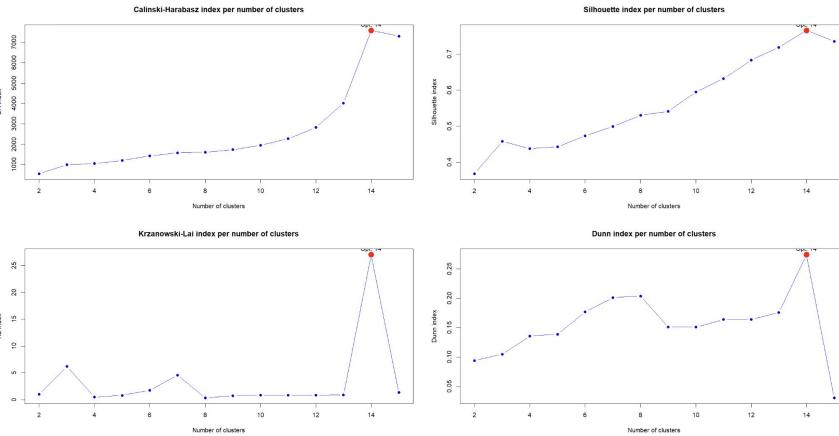
> res.wreath$All.Criticalvalues
   CritValue_Duda CritValue_PseudoT2 Fvalue_Beale
 2       0.5752      395.1145    0.2357
 3       0.5670      352.0466    0.2455
 4       0.5484      280.7600    0.1577
 5       0.5399      256.4895    0.1569
 6       0.5045      188.5638    0.1287
 7       0.4866      166.7048    0.0002
 8       0.4835      163.4595    0.0002
 9       0.5085      194.2906    0.3378
10      0.4788      158.9030    0.0002
11      0.4731      153.6773    0.0005
12      0.4693      150.4018    0.0005
13      0.4601      143.1725    0.0005
14      0.4075      114.8681    0.3762
15      0.3948      110.3626    0.3842

```

```
> res.wreath$Best.nc
          KL      CH Hartigan      CCC  Scott   Marriot    TrCovW TraceW Friedman Rubin Cindex DB
Number_clusters 14.000 14.000 14.0000 14.0000 14.0       6      3 3.00 14.0000 14.0000 13.0000 14.0000
Value_Index     27.542 7581.955 994.0461 67.1003 1096.2 1805133262 4064695761 16556.37 63.2317 -47.8779 0.1627 0.353
          Silhouette Duda PseudoT2 Beale Ratkowsky Ball PtBiserial Frey McClain Dunn Hubert SDindex
Number_clusters 14.0000 14.000 14.0000 2.0000 3.0000 3.00 4.0000 1 2.0000 14.0000 0 14.0000
Value_Index      0.7663 0.501 78.6934 1.4473 0.4491 34011.51 0.6598 NA 0.5863 0.2712 0 0.2449
          Dindex SDbw
Number_clusters 0 14.0000
Value_Index      0 0.0104
```

# NB CLUST - Mahalanobis Distance Matrix

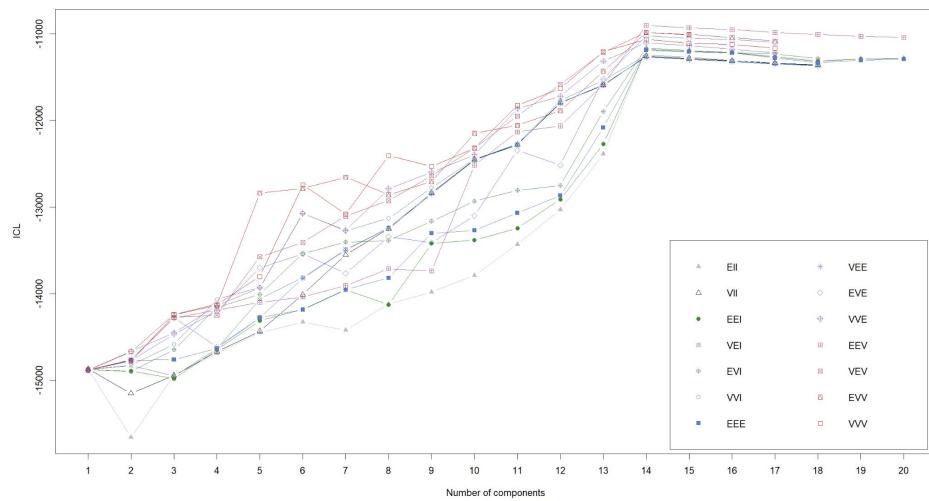
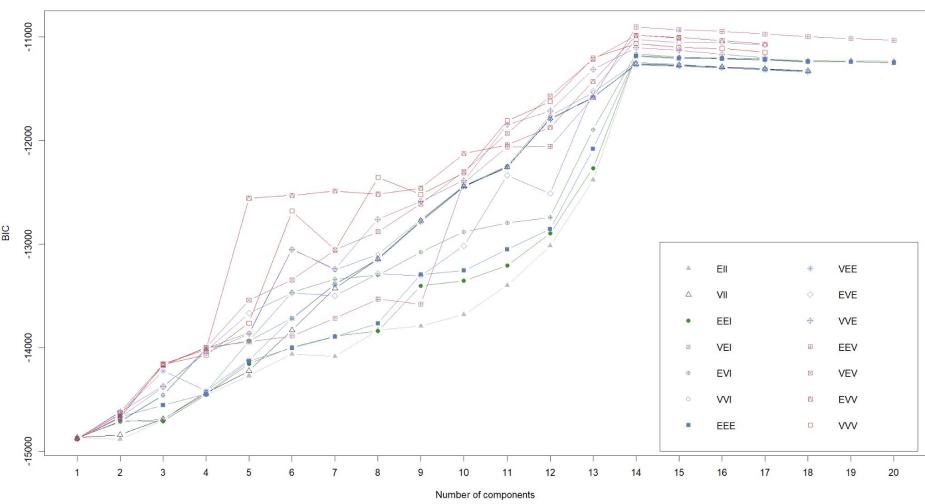
## mahalanobis



	CritValue_Duda	CritValue_PseudoT2	Fvalue_Beale
2	0.5838	451.9201	0.1886
3	0.5523	293.3888	0.1808
4	0.5481	279.5534	0.1851
5	0.5380	251.5985	0.1711
6	0.5110	198.0952	0.1161
7	0.4866	166.7048	0.0001
8	0.5106	197.4618	0.3620
9	0.4835	163.4595	0.0002
10	0.4685	149.7459	0.0002
11	0.4693	150.4018	0.0005
12	0.4669	148.4332	0.0004
13	0.4677	149.0896	0.0011
14	0.4075	114.8681	0.3762
15	0.3948	110.3626	0.3842

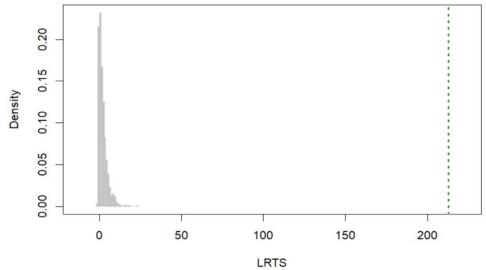
```
> results$Best.nc
      KL      CH Hartigan      CCC      Scott      Marriot      TrCovW      TraceW      Friedman      Rubin      Cindex
Number_clusters 14.0000 14.0000 14.0000 14.0000 14.0000 5          3          3.00 14.0000 14.0000 13.0000
Value_Index     27.0612 7581.955 977.3739 67.1003 1121.446 1507857388 8152774270 41686.01 71.2445 -47.4658 0.1975
      DB Silhouette      Duda PseudoT2      Beale Ratkowsky      Ball PtBiserial      Frey McClain      Dunn Hubert
Number_clusters 14.0000 14.0000 14.0000 14.0000 2.0000 3.0000 3.0000 1 2.0000 14.0000 0
Value_Index     0.353 0.7662 0.501 78.6934 1.6705 0.4712 41020.54 0.6723 NA 0.5466 0.2745 0
      SDindex      Dindex      Sdbw
Number_clusters 7.0000 0 14.0000
Value_Index     0.2404 0 0.0104
```

# McClustBIC - McClustICL

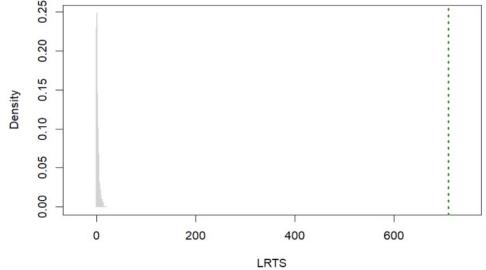


# MclustBootStrapLRT

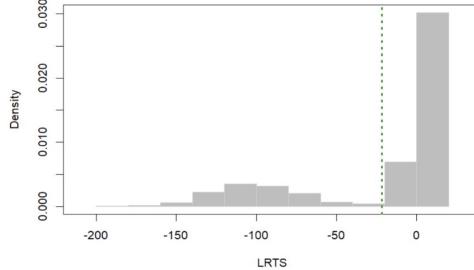
Bootstrap LRT for model EEV with 7 vs 8 components



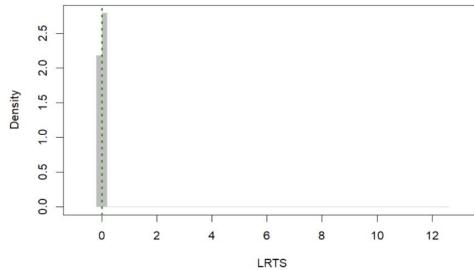
Bootstrap LRT for model EEV with 13 vs 14 components



Bootstrap LRT for model EEV with 8 vs 9 components



Bootstrap LRT for model EEV with 14 vs 15 components



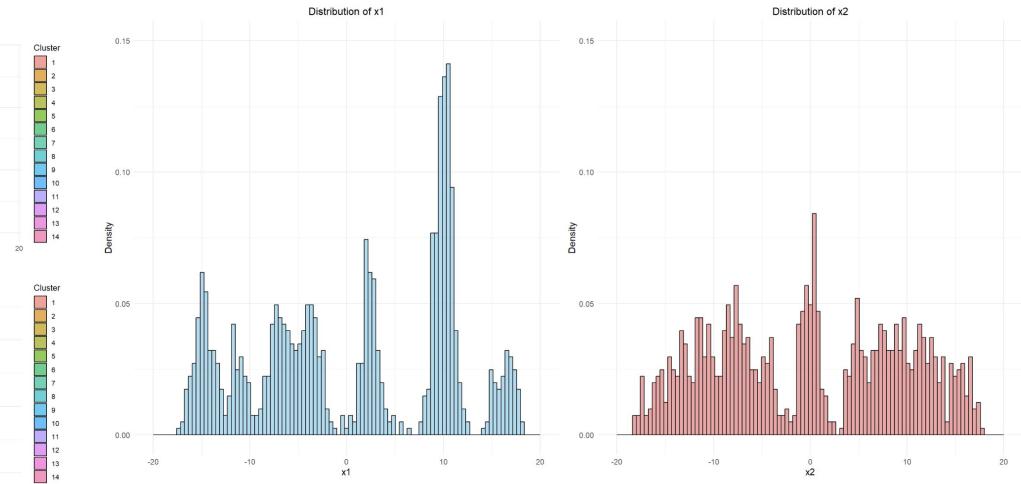
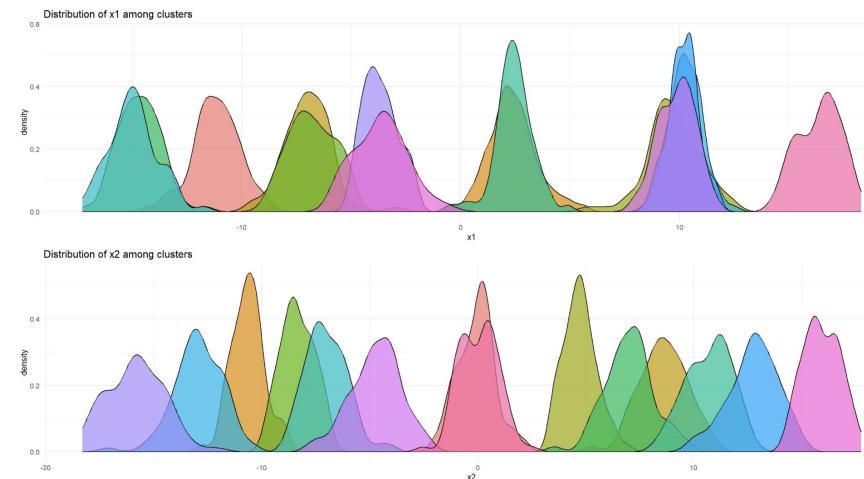
```
> LRT_clust1
```

```
Bootstrap sequential LRT for the number of mixture components
```

```
Model      = EEV  
Replications = 999
```

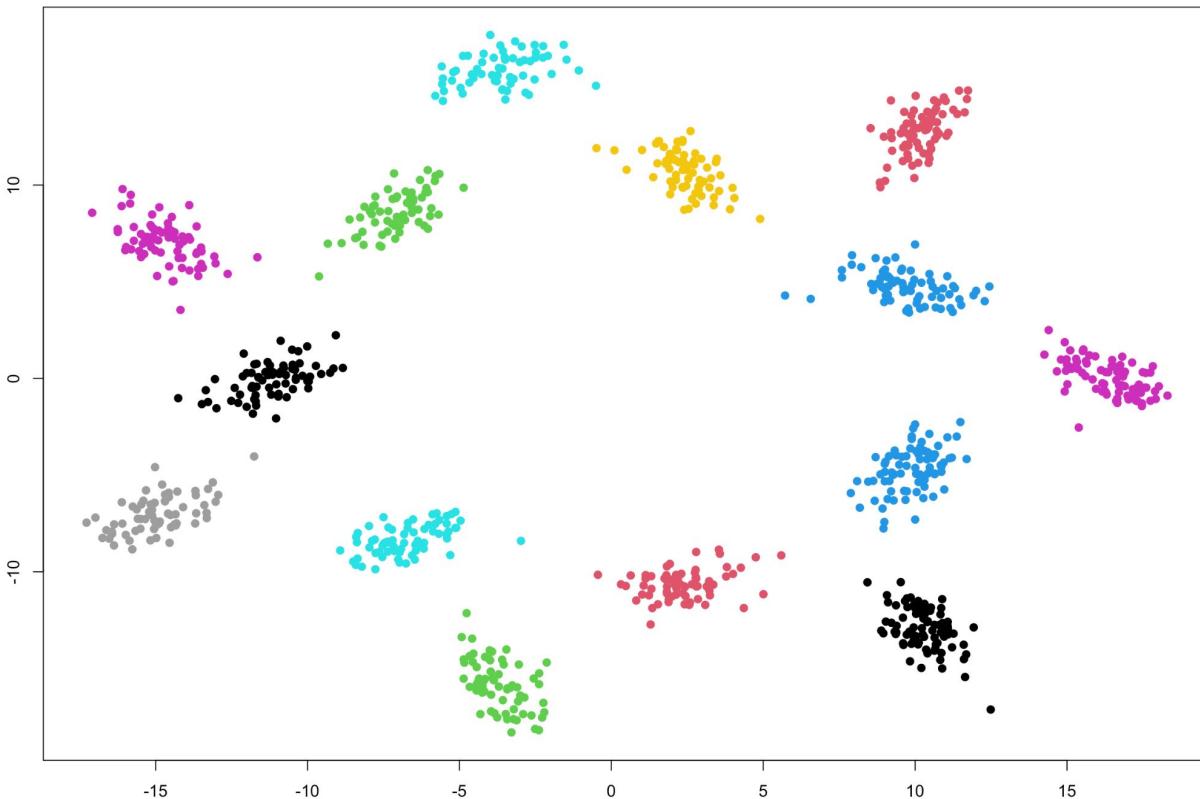
	LRTS	bootstrap	p-value
1 vs 2	2.369498e+02	0.001	
2 vs 3	5.333530e+02	0.001	
3 vs 4	1.983375e+02	0.001	
4 vs 5	8.019553e+01	0.001	
5 vs 6	8.373498e+01	0.001	
6 vs 7	1.997408e+02	0.001	
7 vs 8	2.127155e+02	0.001	
8 vs 9	-2.177318e+01	0.743	
9 vs 10	1.202919e+03	0.001	
10 vs 11	3.727638e+02	0.001	
11 vs 12	3.066054e+01	0.001	
12 vs 13	5.000733e+02	0.001	
13 vs 14	7.094307e+02	0.001	
14 vs 15	1.887003e-04	0.063	
15 vs 16	1.336872e+01	0.027	
16 vs 17	2.242939e-01	0.614	
17 vs 18	3.059050e+00	0.270	
18 vs 19	8.798749e+00	0.119	
19 vs 20	1.202281e+01	0.062	
20 vs 21	7.684063e+00	0.123	

# Distribution representation



# Mclust model

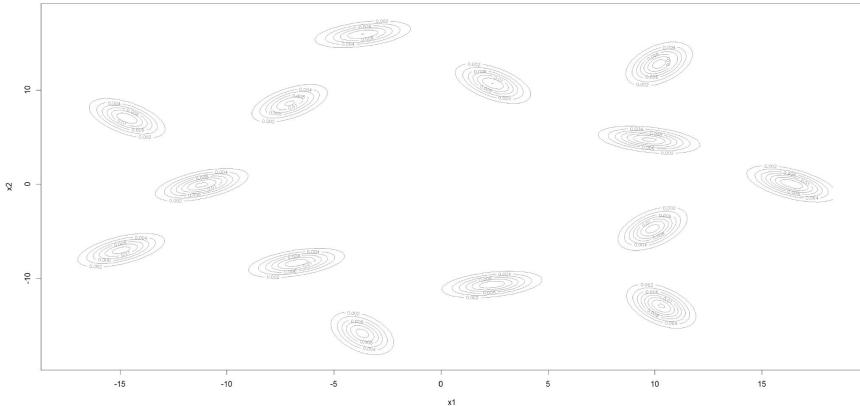
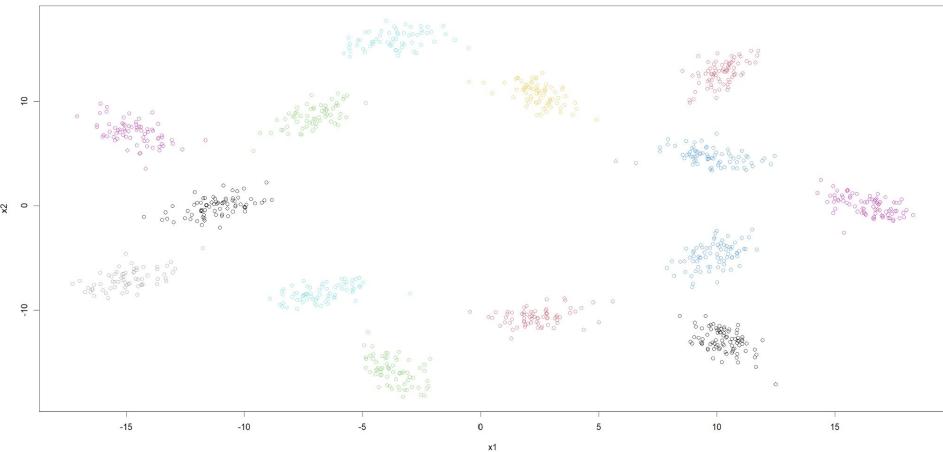
Clustering con Mclust (14 cluster)



log-likelihood	n	df	BIC	ICL									
-5254.513	1000	57	-10902.77	-10902.91									
Clustering table:													
1	2	3	4	5	6	7	8	9	10	11	12	13	14
74	69	63	74	68	70	71	66	83	77	66	77	61	81

# Mclust

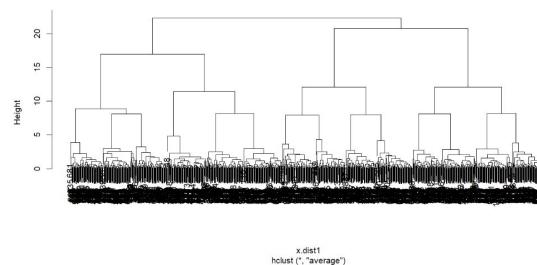
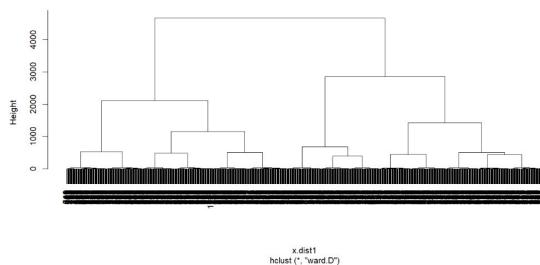
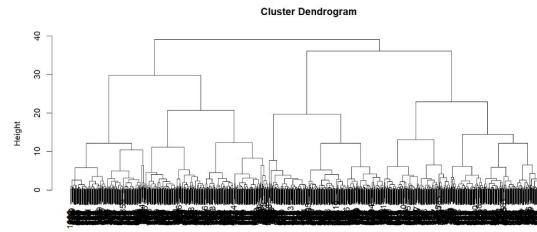
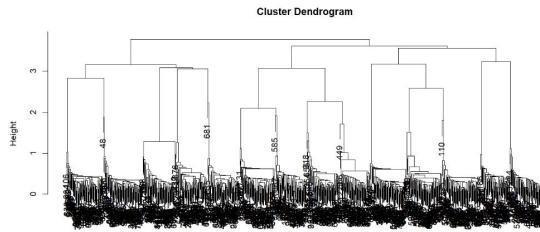
# Density Clustering



```
> table(model4$classification)
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14
74	69	63	74	68	70	71	66	83	77	66	77	61	81

# Hierarchical clustering (Euclidean)



```
> table(hcclus11)
hcclus11
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
140 69 63 74 67 70 71 83 77 66 77 61 81 1

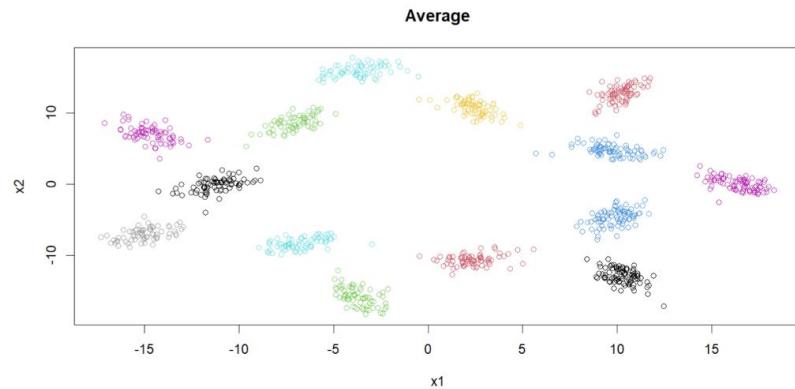
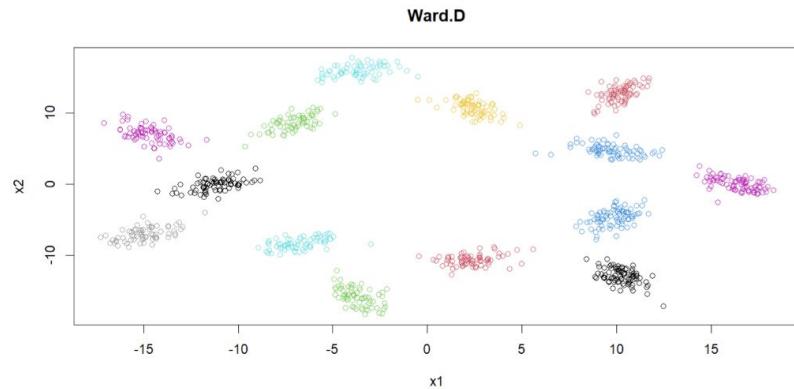
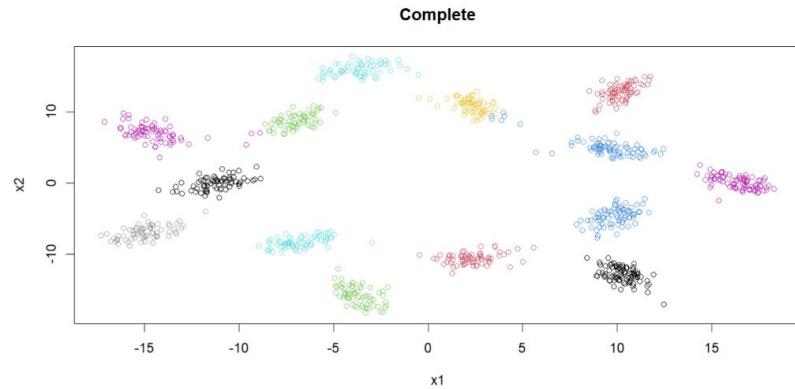
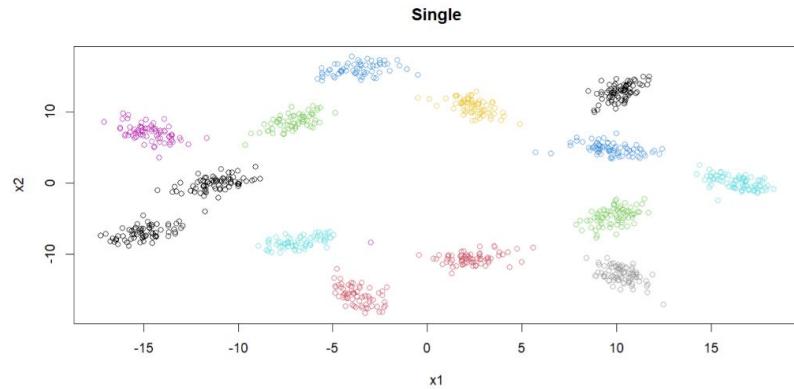
> table(hcclus12)
hcclus12
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
74 69 60 80 68 73 65 66 83 77 66 77 61 81

> table(hcclus13)
hcclus13
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
74 69 63 74 68 70 71 66 83 77 66 77 61 81

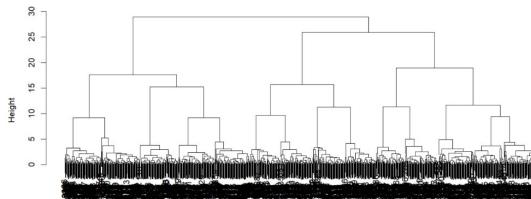
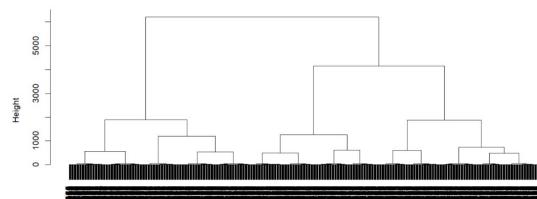
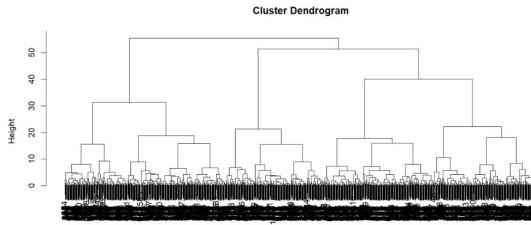
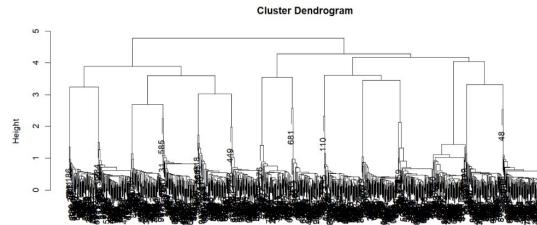
> table(hcclus14)
hcclus14
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
```

# Hierarchical Clustering (Euclidean)

---

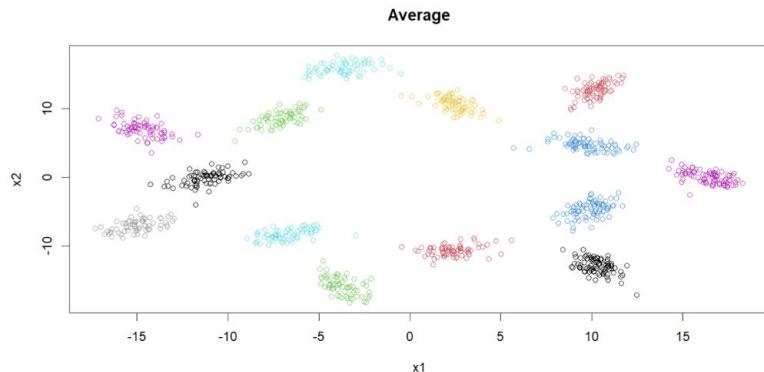
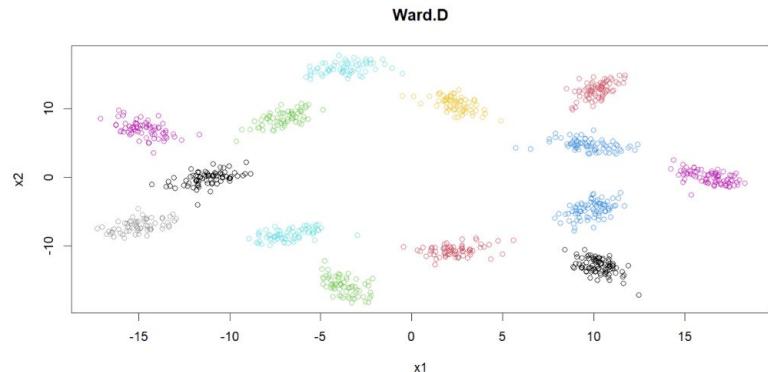
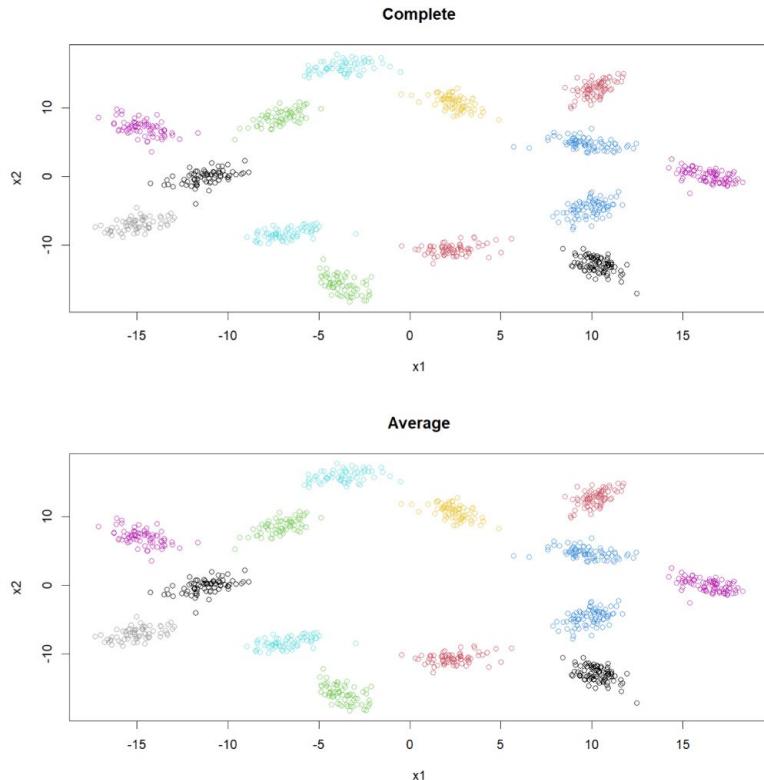
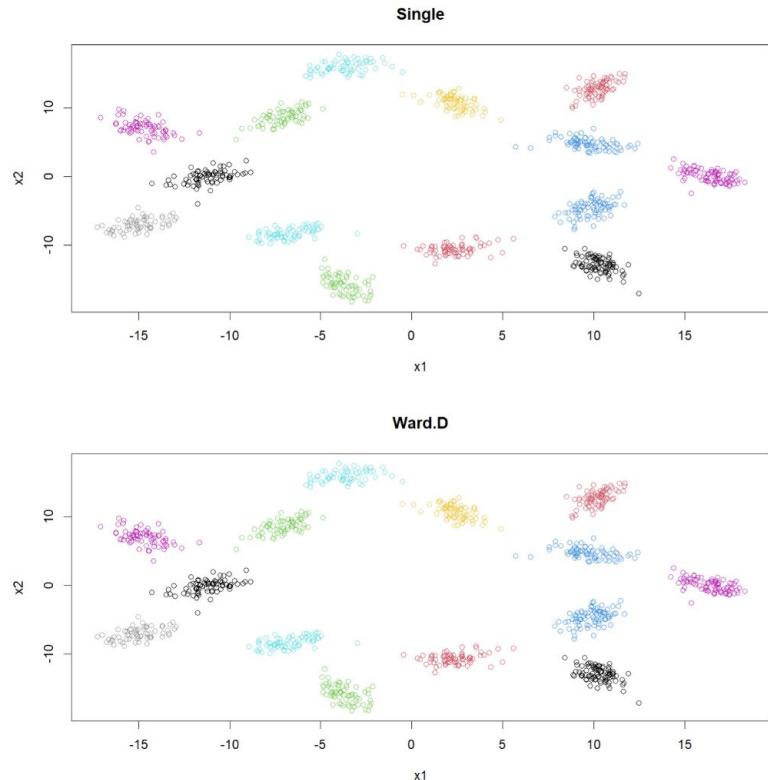


# Hierarchical Clustering (Manhattan)

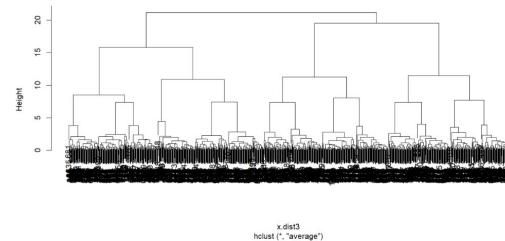
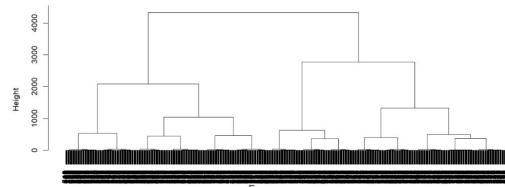
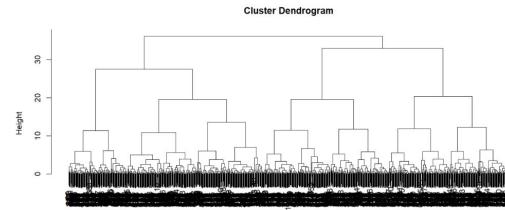
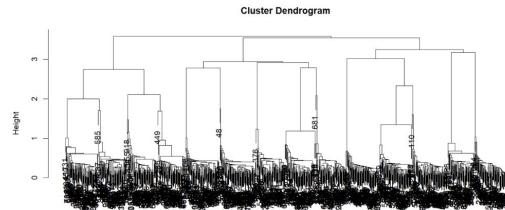


```
> table(hiclust21)
hiclust21
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
> table(hiclust22)
hiclust22
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
> table(hiclust23)
hiclust23
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
> table(hiclust24)
hiclust24
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
```

# Hierarchical Clustering (Manhattan)

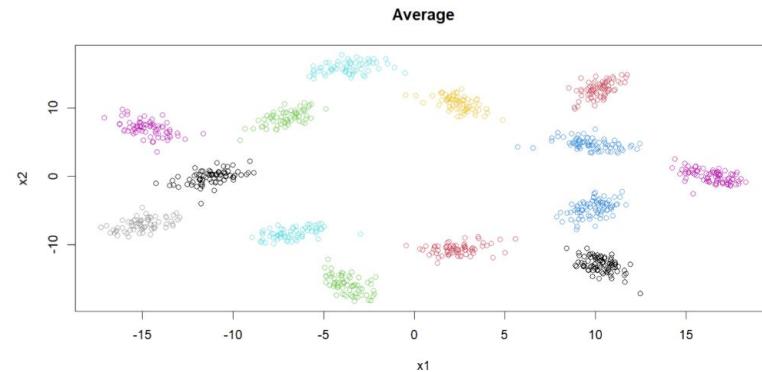
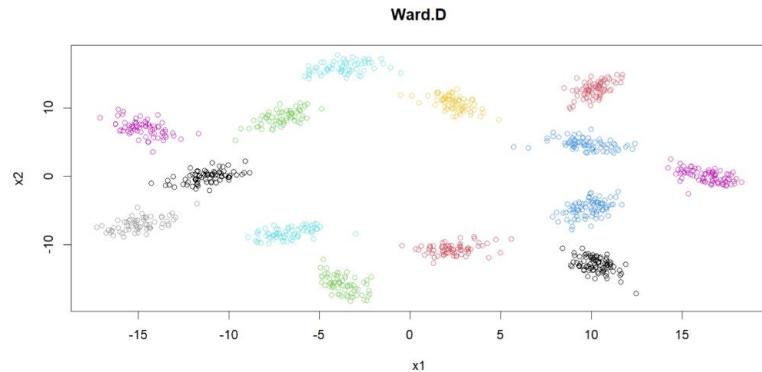
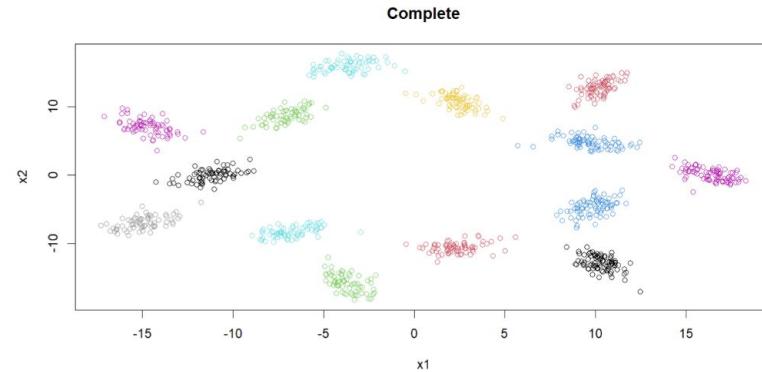
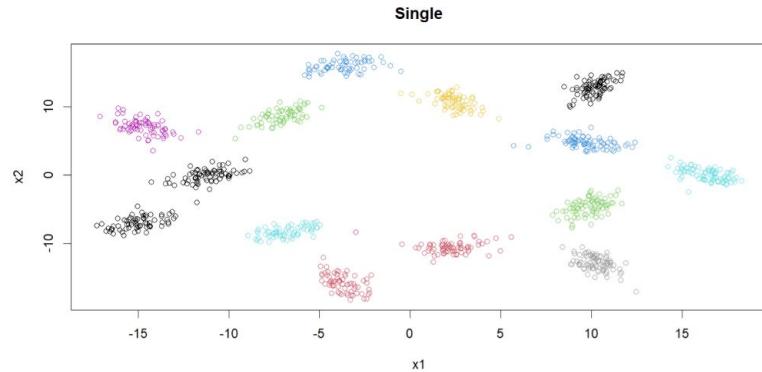


# Hierarchical Clustering (Minkowski)

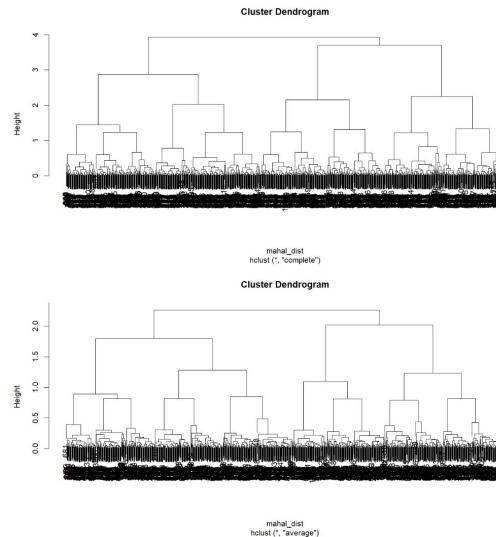
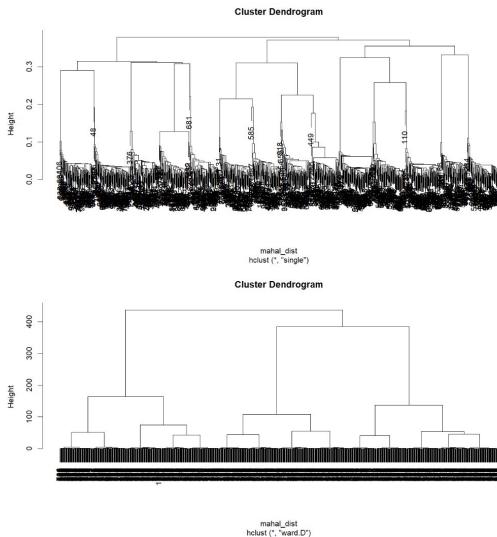


```
> table(hiclust31)
hiclust31
  1  2  3  4  5  6  7  8  9 10 11 12 13 14
140 69 63 74 67 70 71 83 77 66 77 61 81 1
> table(hiclust32)
hiclust32
  1  2  3  4  5  6  7  8  9 10 11 12 13 14
74 69 63 74 68 70 71 66 83 77 66 77 61 81
> table(hiclust33)
hiclust33
  1  2  3  4  5  6  7  8  9 10 11 12 13 14
74 69 63 74 68 70 71 66 83 77 66 77 61 81
> table(hiclust34)
hiclust34
  1  2  3  4  5  6  7  8  9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
```

# Hierarchical Clustering (Minkowski)



# Hierarchical Clustering (Mahalanobis)



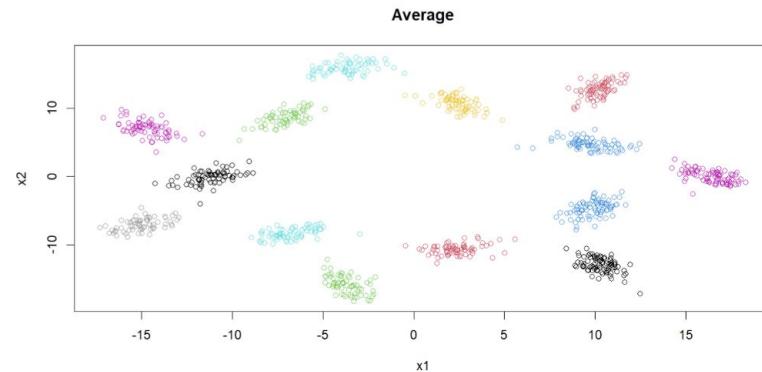
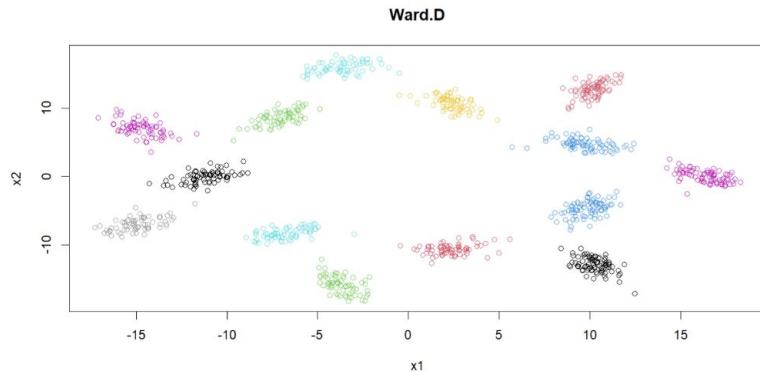
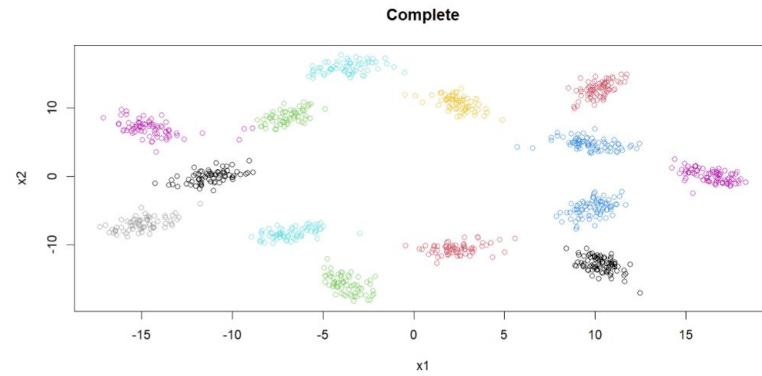
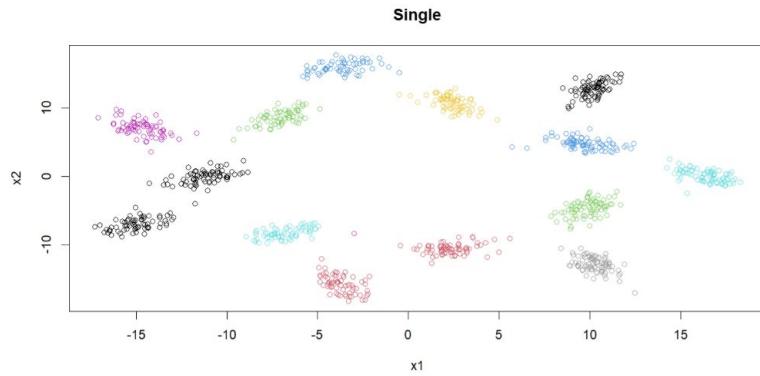
```
> table(haclust1)
haclust1
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
140 69 63 74 67 67 70 71 83 77 66 77 61 81 1

> table(haclust2)
haclust2
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
74 69 60 74 68 73 71 66 83 77 66 77 61 81

> table(haclust3)
haclust3
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
74 69 63 74 68 70 71 66 83 77 66 77 61 81

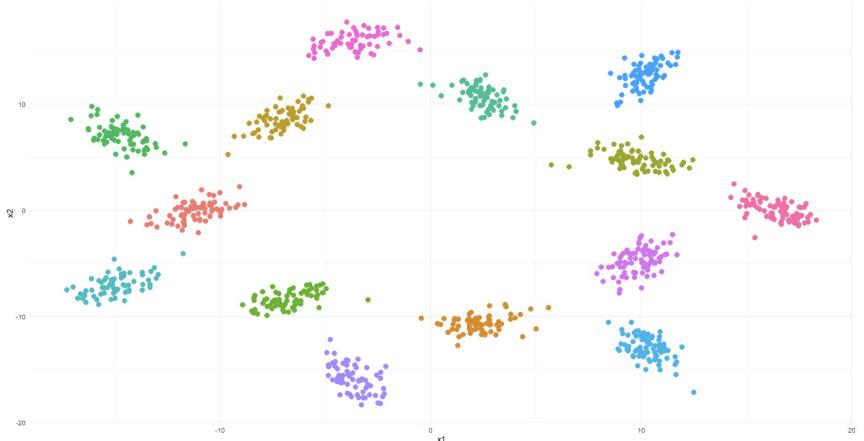
> table(haclust4)
haclust4
 1 2 3 4 5 6 7 8 9 10 11 12 13 14
75 69 63 74 68 70 71 65 83 77 66 77 61 81
```

# Hierarchical Clustering (Mahalanobis)

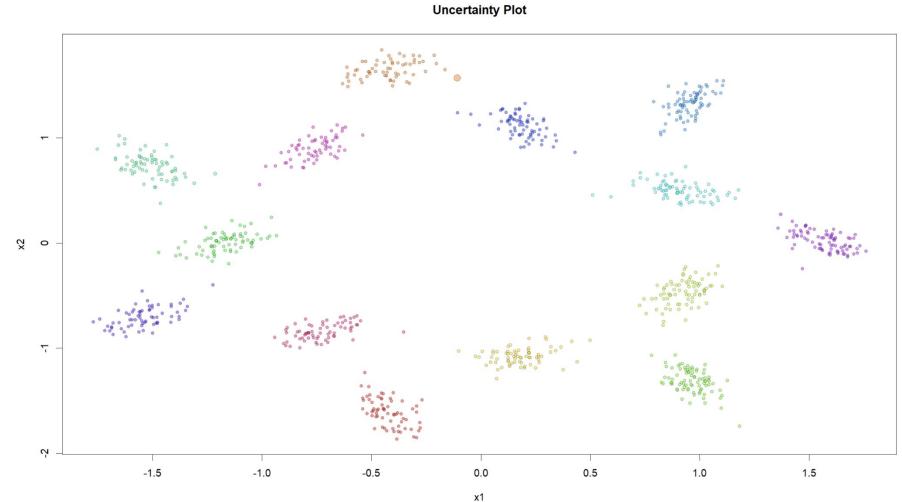


# PAM

Clustering with PAM



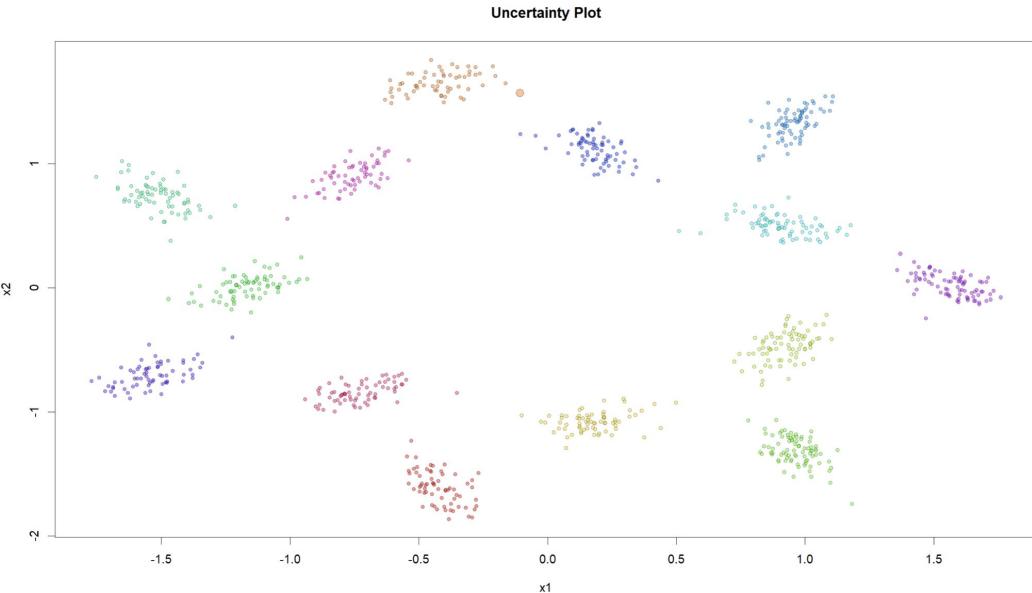
Uncertainty Plot



```
> table(pam_clustering$clustering)
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14
75	69	63	74	68	70	71	65	83	77	66	77	61	81

# TEIGEN



BIC and ICL select the same model and groups.  
The best model (BIC of -1736.07, ICL of -1736.3876) is CUCC with G=14  
> print(result)  
BIC and ICL select the same model and groups.  
The best model (BIC of -1736.07, ICL of -1736.3876) is CUCC with G=14  
> summary(result)

----- Summary for teigen -----  
----- RESULTS -----  
Loglik: -667.7114  
BIC: -1736.072  
ICL: -1736.388  
Model: CUCC  
# Groups: 14

Clustering Table:

1	2	3	4	5	6	7	8	9	10	11	12	13	14
74	69	66	81	77	63	66	74	71	68	77	70	61	83

# Metrics comparison

Hierarchical clustering Euclidean				
	Single	Complete	Ward.D	Average
CH	3554,009	6519,141	7581,955	7595,281
Average.within	2,156307	1,786128	1,721773	1,721108
Dunn	0,1705	0,0358	0,2712	0,2467
Silhouette	0,6949	0,7523	0,7663	0,7665

Hierarchical clustering Mahalanobis				
	Single	Complete	Ward.D	Average
CH	3554,009	7374,378	7581,955	7595,281
Average.within	2,156307	1,734172	1,721773	1,721108
Dunn	0,1705	0,0613	0,2712	0,2467
Silhouette	0,6949	0,763	0,7663	0,7665

Hierarchical clustering Manhattan				
	Single	Complete	Ward.D	Average
CH	7595,281	7595,281	7595,281	7595,281
Average.within	1,721108	1,721108	1,721108	1,721108
Dunn	0,2467	0,2467	0,2467	0,2467
Silhouette	0,7665	0,7665	0,7665	0,7665

Hierarchical clustering Minkowski				
	Single	Complete	Ward.D	Average
CH	3554,009	7581,955	7581,955	7595,281
Average.within	2,156307	1,721773	1,721773	1,721108
Dunn	0,1705	0,2712	0,2712	0,2467
Silhouette	0,6949	0,7663	0,7663	0,7665

	Mclust	PAM	Teigen	Mhclust
CH	7581,955	7595,281	7581,955	7595,281
Average.within	1,721773	1,721108	1,721773	1,721108
Dunn	0,2712	0,2467	0,2712	0,2467
Silhouette	0,7663	0,7665	0,7663	0,7665

# Models comparison

CH	7595,281
Average.within	1,721108
Dunn	0,2467
Silhouette	0,7665

1	2	3	4	5	6	7
75	69	63	74	68	70	71
8	9	10	11	12	13	14
65	83	77	66	77	61	81

CH	7581,955
Average.within	1,721773
Dunn	0,2712
Silhouette	0,7663

1	2	3	4	5	6	7
74	69	63	74	68	70	71
8	9	10	11	12	13	14
66	83	77	66	77	61	81

CH	7374,378
Average.within	1,734172
Dunn	0,0613
Silhouette	0,763

1	2	3	4	5	6	7
74	69	60	74	68	73	71
8	9	10	11	12	13	14
66	83	77	66	77	61	81

# Models comparison

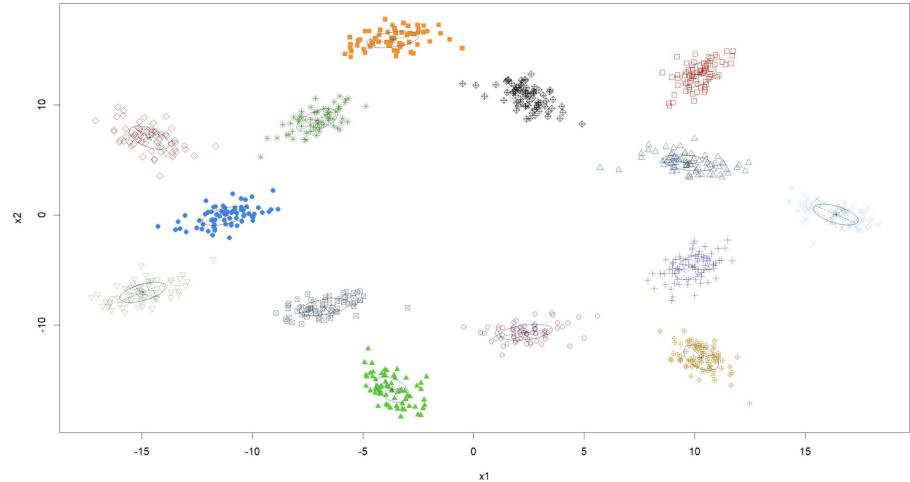
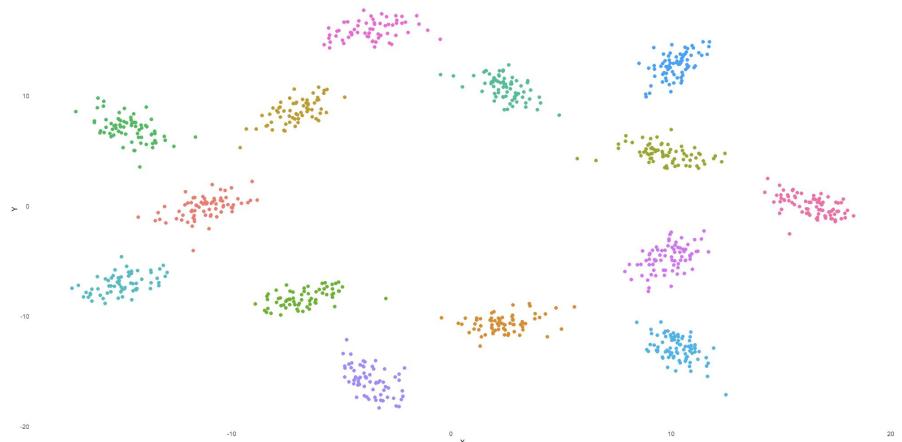
CH	6519,141
Average.within	1,786128
Dunn	0,0358
Silhouette	0,7523

1	2	3	4	5	6	7
74	69	60	80	68	73	65
8	9	10	11	12	13	14
66	83	77	66	77	61	81

CH	3554,009
Average.within	2,156307
Dunn	0,1705
Silhouette	0,6949

1	2	3	4	5	6	7
140	69	63	74	67	70	71
8	9	10	11	12	13	14
83	77	66	77	61	81	1

# Models comparison

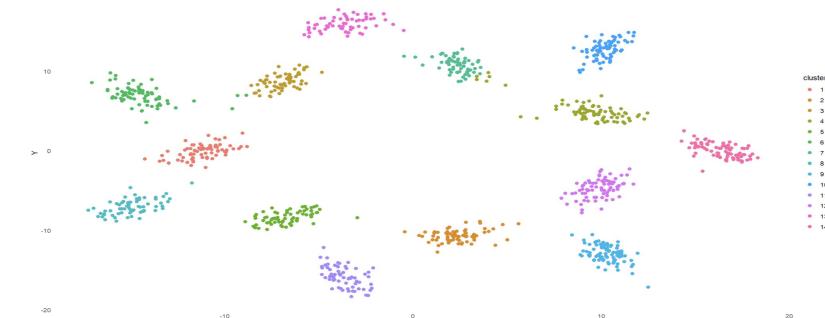


1	2	3	4	5	6	7
75	69	63	74	68	70	71
8	9	10	11	12	13	14
65	83	77	66	77	61	81

1	2	3	4	5	6	7
74	69	63	74	68	70	71
8	9	10	11	12	13	14
66	83	77	66	77	61	81



1	2	3	4	5	6	7
74	69	60	74	68	73	71
8	9	10	11	12	13	14
66	83	77	66	77	61	81



1	2	3	4	5	6	7
74	69	60	80	68	73	65
8	9	10	11	12	13	14
66	83	77	66	77	61	81



1	2	3	4	5	6	7
140	69	63	74	67	70	71
8	9	10	11	12	13	14
83	77	66	77	61	81	1