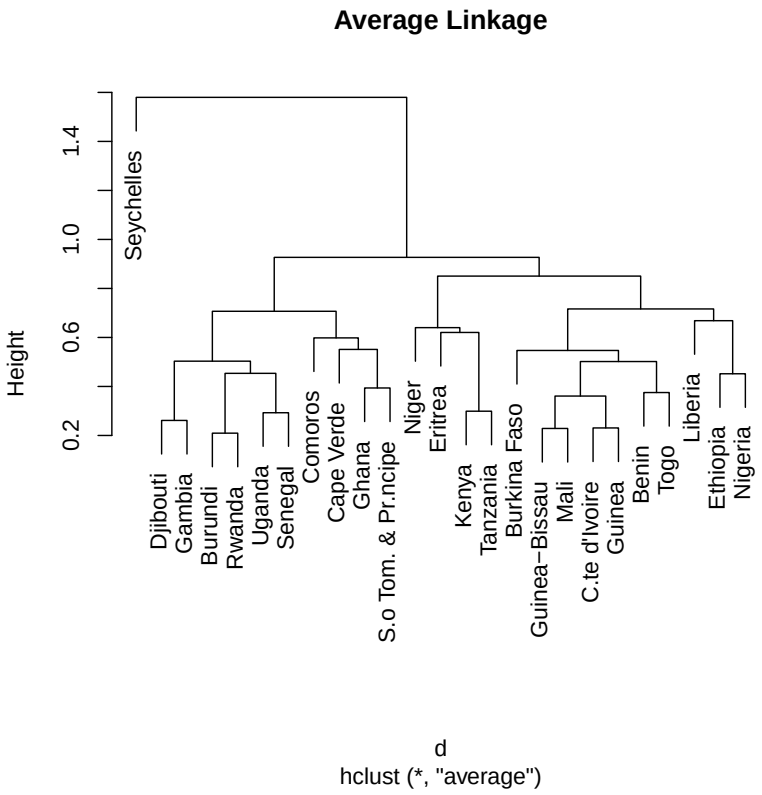
Homework 2
Hierarchichal Cluster Analysis and Latent Class Models

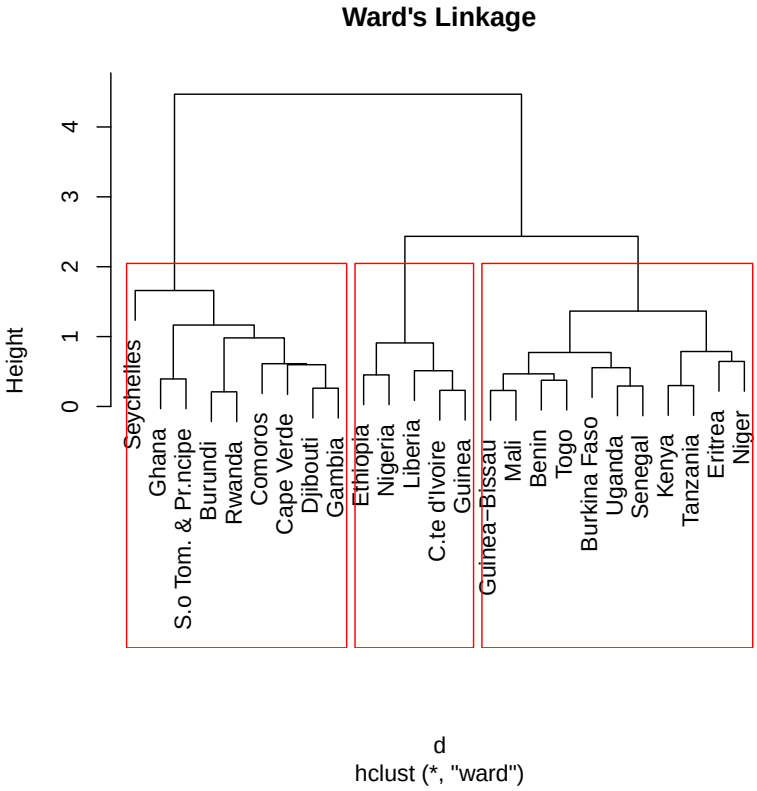
Author

March 14, 2014

1. Exercise 1

```
> d <- dist(eastwest[,c(9,10,11,12,13,14)], method="euclidean")
> fit <- hclust(d, "average")
> fit2 <- hclust(d, "ward")
```





(a) First question

```
> names(groups) <- eastwest$country
> cluster1 <- eastwest[eastwest$country %in%
+                      names(groups[groups == 1]),c(9,10,11,12,13,14)]
> stats1 <- sapply(cluster1, each(min, max, mean, sd, var))

> names(groups) <- eastwest$country
> cluster2 <- eastwest[eastwest$country %in%
+                      names(groups[groups == 2]),c(9,10,11,12,13,14)]
> stats2 <- sapply(cluster2, each(min, max, mean, sd, var))

> names(groups) <- eastwest$country
> cluster3 <- eastwest[eastwest$country %in%
+                      names(groups[groups == 3]),c(9,10,11,12,13,14)]
> stats3 <- sapply(cluster3, each(min, max, mean, sd, var))
```

	sanitation_s	water_s	measles_s	dpt_s	primary_s	teacher_s
min	0.04	0.41	0.54	0.69	0.20	0.00
max	1.00	1.00	1.00	1.00	1.00	1.00
mean	0.50	0.80	0.85	0.88	0.53	0.30
sd	0.27	0.20	0.14	0.10	0.29	0.29
var	0.08	0.04	0.02	0.01	0.08	0.09

Table 1: First Cluster Summary

	sanitation_s	water_s	measles_s	dpt_s	primary_s	teacher_s
min	0.00	0.03	0.27	0.48	0.00	0.03
max	0.48	0.65	1.00	1.00	0.60	0.23
mean	0.13	0.35	0.60	0.71	0.28	0.10
sd	0.15	0.21	0.23	0.15	0.19	0.06
var	0.02	0.04	0.05	0.02	0.04	0.00

Table 2: Second Cluster Summary

	sanitation_s	water_s	measles_s	dpt_s	primary_s	teacher_s
min	0.10	0.00	0.00	0.00	0.21	0.02
max	0.24	0.64	0.53	0.29	0.41	0.35
mean	0.15	0.39	0.25	0.13	0.29	0.14
sd	0.06	0.26	0.20	0.13	0.08	0.13
var	0.00	0.07	0.04	0.02	0.01	0.02

Table 3: Third Cluster Summary

(b) Second question

(c) Third question

```
> job <- read.csv("~/Dropbox/LSE/MY455/Week 5/EVS_job_LCM.csv")
> job <- job[,2:7]
> for (i in 1:6) {
+   job[,i] <- as.factor(job[,i])
+   levels(job[,i]) <- c("not mentioned", "mentioned")
+ }
> names(job) <- c('pay','people','security','achieve','interest','equal')
> f <- cbind(pay, people, security, achieve, interest, equal) ~ 1
> lca2 <- poLCA(f, job, nclass=2)
```

```
Conditional item response (column) probabilities,
 by outcome variable, for each class (row)

$pay
          not mentioned mentioned
class 1:        0.3939    0.6061
class 2:        0.1123    0.8877

$people
          not mentioned mentioned
class 1:        0.4742    0.5258
class 2:        0.0805    0.9195

$security
          not mentioned mentioned
class 1:        0.5328    0.4672
class 2:        0.1352    0.8648

$achieve
          not mentioned mentioned
class 1:        0.5367    0.4633
class 2:        0.1240    0.8760

$interest
          not mentioned mentioned
class 1:        0.5197    0.4803
class 2:        0.0724    0.9276

$equal
          not mentioned mentioned
class 1:        0.5892    0.4108
class 2:        0.0873    0.9127

Estimated class population shares
 0.5284 0.4716

Predicted class memberships (by modal posterior prob.)
 0.5411 0.4589

=========================================================
Fit for 2 latent classes:
```

```
============================================================
number of observations: 1532
number of estimated parameters: 13
residual degrees of freedom: 50
maximum log-likelihood: -5381.499

AIC(2): 10789
BIC(2): 10858.34
G^2(2): 128.6007 (Likelihood ratio/deviance statistic)
X^2(2): 136.2533 (Chi-square goodness of fit)

> lca3 <- poLCA(f, job, nclass=3)

Conditional item response (column) probabilities,
 by outcome variable, for each class (row)

$pay
          not mentioned mentioned
class 1:         0.6182    0.3818
class 2:         0.3063    0.6937
class 3:         0.0847    0.9153

$people
          not mentioned mentioned
class 1:         0.2438    0.7562
class 2:         0.5781    0.4219
class 3:         0.0911    0.9089

$security
          not mentioned mentioned
class 1:         0.7556    0.2444
class 2:         0.4584    0.5416
class 3:         0.0971    0.9029

$achieve
          not mentioned mentioned
class 1:         0.2186    0.7814
class 2:         0.6865    0.3135
class 3:         0.1354    0.8646

$interest
```

```
          not mentioned mentioned
class 1:        0.2915    0.7085
class 2:        0.6215    0.3785
class 3:        0.0831    0.9169

$equal
          not mentioned mentioned
class 1:        0.4968    0.5032
class 2:        0.6187    0.3813
class 3:        0.0968    0.9032

Estimated class population shares
 0.1866 0.3468 0.4666

Predicted class memberships (by modal posterior prob.)
 0.1547 0.3714 0.4739

=========================================================
Fit for 3 latent classes:
=========================================================
number of observations: 1532
number of estimated parameters: 20
residual degrees of freedom: 43
maximum log-likelihood: -5345.652

AIC(3): 10731.3
BIC(3): 10837.99
G^2(3): 56.90818 (Likelihood ratio/deviance statistic)
X^2(3): 56.04527 (Chi-square goodness of fit)
```

2. Exercise 2

	Variable	class 1:	class 2:	class 3:
1	Pay not mentioned	0.62	0.31	0.08
2	Pay mentioned	0.38	0.69	0.92
3	People not mentioned	0.24	0.58	0.09
4	People mentioned	0.76	0.42	0.91
5	Security not mentioned	0.76	0.46	0.1
6	Security mentioned	0.24	0.54	0.9
7	Achieve not mentioned	0.22	0.69	0.14
8	Achieve mentioned	0.78	0.31	0.86
9	Interest not mentioned	0.29	0.62	0.08
10	Interest mentioned	0.71	0.38	0.92
11	Equal not mentioned	0.5	0.62	0.1
12	Equal mentioned	0.5	0.38	0.9

Table 4: Summary of item response probabilities