LCA data modeling Seth-Josh

1. Loading, setting up

```
library(tidyverse)
library(poLCA)

f <- "obs-segment_units1-7_2013-2014.csv"

d <- read_csv(f)
```

2. Preparing data with a few teacher and student variables

None of the unit-specific variables included.

```
add_one <- function(x) {
    x + 1
}

ds <- d %>%
    dplyr::select(sInvented, sProcedural, sConceptual, tInitSelect, tCompare, tDiscussQ, tConnectBigIdea
    map_df(replace_na, 0) %>%
    map_df(add_one)
```

4A. Examining four-class solution with one random seed (1001)

```
f <- cbind(sInvented, sProcedural, sConceptual, tInitSelect, tCompare, tDiscussQ, tConnectBigIdeas, tCo

set.seed(1001)
poLCA(f, ds, nclass = 4, maxiter = 5000, graphs = TRUE, verbose = FALSE)
```


```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $sInvented
##            Pr(1)  Pr(2)
## class 1:  0.7772 0.2228
## class 2:  0.0259 0.9741
## class 3:  0.8365 0.1635
## class 4:  0.0528 0.9472
##
## $sProcedural
##            Pr(1)  Pr(2)
## class 1:  0.6498 0.3502
## class 2:  0.5140 0.4860
## class 3:  0.6201 0.3799
## class 4:  0.2098 0.7902
##
## $sConceptual
##            Pr(1)  Pr(2)
## class 1:  0.9259 0.0741
## class 2:  0.0951 0.9049
## class 3:  0.1507 0.8493
## class 4:  0.3518 0.6482
##
## $tInitSelect
##            Pr(1)  Pr(2)
```

```
## class 1:  0.9670 0.0330
## class 2:  0.4704 0.5296
## class 3:  0.9938 0.0062
## class 4:  0.6232 0.3768
##
## $tCompare
##            Pr(1)  Pr(2)
## class 1:  0.9820 0.0180
## class 2:  0.1027 0.8973
## class 3:  0.8829 0.1171
## class 4:  0.7076 0.2924
##
## $tDiscussQ
##            Pr(1)  Pr(2)
## class 1:  0.6550 0.3450
## class 2:  0.0000 1.0000
## class 3:  0.0039 0.9961
## class 4:  0.0806 0.9194
##
## $tConnectBigIdeas
##            Pr(1)  Pr(2)
## class 1:  0.9961 0.0039
## class 2:  0.2109 0.7891
## class 3:  0.8006 0.1994
## class 4:  0.8224 0.1776
##
## $tConnectOthers
##            Pr(1)  Pr(2)
## class 1:  0.9952 0.0048
## class 2:  0.5536 0.4464
## class 3:  0.9575 0.0425
## class 4:  0.9446 0.0554
##
## $tPressExplain
##            Pr(1)  Pr(2)
## class 1:  0.9995 0.0005
## class 2:  0.1073 0.8927
## class 3:  0.3013 0.6987
## class 4:  0.3858 0.6142
##
## Estimated class population shares
##  0.2461 0.0813 0.3384 0.3341
##
## Predicted class memberships (by modal posterior prob.)
##  0.241 0.075 0.3374 0.3466
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 2813
## number of estimated parameters: 39
## residual degrees of freedom: 472
## maximum log-likelihood: -12241.13
##
```

```
## AIC(4): 24560.26
## BIC(4): 24792
## G^2(4): 655.7218 (Likelihood ratio/deviance statistic)
## X^2(4): 959.092 (Chi-square goodness of fit)
##
```

4B. Examining four-class solution with one random seed (1002)

```
f <- cbind(sInvented, sProcedural, sConceptual, tInitSelect, tCompare, tDiscussQ, tConnectBigIdeas, tCo

set.seed(1002)
poLCA(f, ds, nclass = 4, maxiter = 5000, graphs = TRUE, verbose = FALSE)
```


```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $sInvented
##            Pr(1)  Pr(2)
## class 1:  0.0259 0.9741
## class 2:  0.7772 0.2228
## class 3:  0.8365 0.1635
## class 4:  0.0528 0.9472
##
```

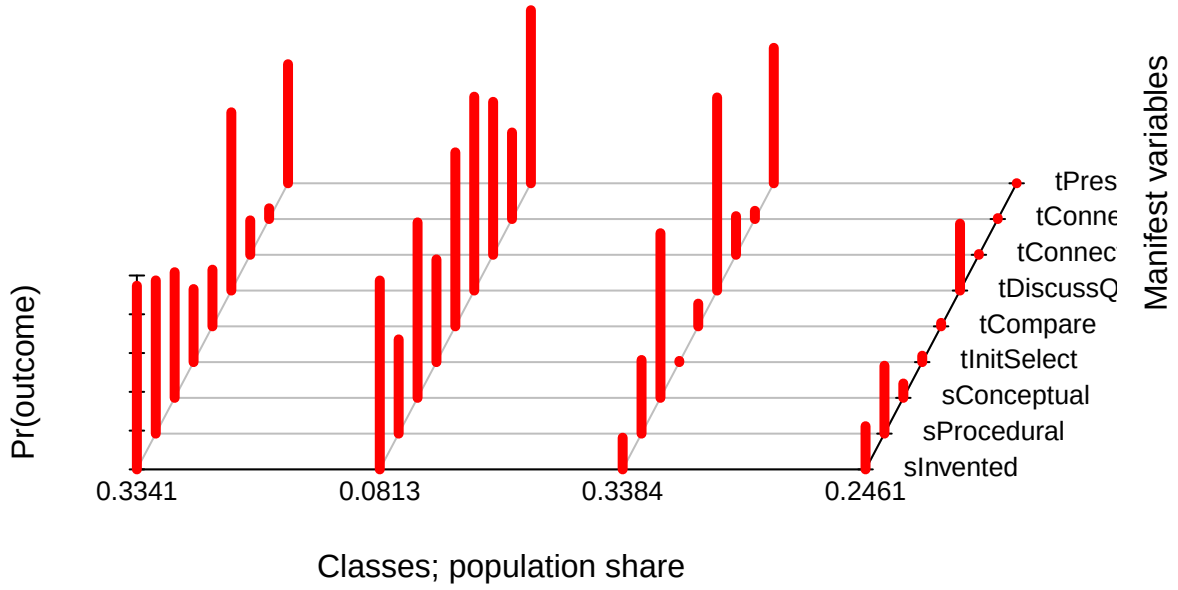
```
## $sProcedural
##            Pr(1)  Pr(2)
## class 1:  0.5140 0.4860
## class 2:  0.6498 0.3502
## class 3:  0.6201 0.3799
## class 4:  0.2098 0.7902
##
## $sConceptual
##            Pr(1)  Pr(2)
## class 1:  0.0951 0.9049
## class 2:  0.9259 0.0741
## class 3:  0.1507 0.8493
## class 4:  0.3518 0.6482
##
## $tInitSelect
##            Pr(1)  Pr(2)
## class 1:  0.4704 0.5296
## class 2:  0.9670 0.0330
## class 3:  0.9938 0.0062
## class 4:  0.6232 0.3768
##
## $tCompare
##            Pr(1)  Pr(2)
## class 1:  0.1027 0.8973
## class 2:  0.9820 0.0180
## class 3:  0.8829 0.1171
## class 4:  0.7076 0.2924
##
## $tDiscussQ
##            Pr(1)  Pr(2)
## class 1:  0.0000 1.0000
## class 2:  0.6550 0.3450
## class 3:  0.0039 0.9961
## class 4:  0.0806 0.9194
##
## $tConnectBigIdeas
##            Pr(1)  Pr(2)
## class 1:  0.2109 0.7891
## class 2:  0.9961 0.0039
## class 3:  0.8006 0.1994
## class 4:  0.8224 0.1776
##
## $tConnectOthers
##            Pr(1)  Pr(2)
## class 1:  0.5536 0.4464
## class 2:  0.9952 0.0048
## class 3:  0.9575 0.0425
## class 4:  0.9446 0.0554
##
## $tPressExplain
##            Pr(1)  Pr(2)
## class 1:  0.1073 0.8927
## class 2:  0.9995 0.0005
## class 3:  0.3013 0.6987
```

```
## class 4:  0.3858 0.6142
##
## Estimated class population shares
##  0.0813 0.2461 0.3384 0.3341
##
## Predicted class memberships (by modal posterior prob.)
##  0.075 0.241 0.3374 0.3466
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 2813
## number of estimated parameters: 39
## residual degrees of freedom: 472
## maximum log-likelihood: -12241.13
##
## AIC(4): 24560.26
## BIC(4): 24792
## G^2(4): 655.7218 (Likelihood ratio/deviance statistic)
## X^2(4): 959.092 (Chi-square goodness of fit)
##
```

4C. Examining four-class solution with one random seed (1003)

```
f <- cbind(sInvented, sProcedural, sConceptual, tInitSelect, tCompare, tDiscussQ, tConnectBigIdeas, tCo

set.seed(1003)
poLCA(f, ds, nclass = 4, maxiter = 5000, graphs = TRUE, verbose=FALSE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $sInvented
##            Pr(1)  Pr(2)
## class 1:  0.0528 0.9472
## class 2:  0.0259 0.9741
## class 3:  0.8365 0.1635
## class 4:  0.7772 0.2228
##
## $sProcedural
##            Pr(1)  Pr(2)
## class 1:  0.2098 0.7902
## class 2:  0.5140 0.4860
## class 3:  0.6201 0.3799
## class 4:  0.6498 0.3502
##
## $sConceptual
##            Pr(1)  Pr(2)
## class 1:  0.3518 0.6482
## class 2:  0.0951 0.9049
## class 3:  0.1507 0.8493
## class 4:  0.9259 0.0741
##
## $tInitSelect
##            Pr(1)  Pr(2)
```

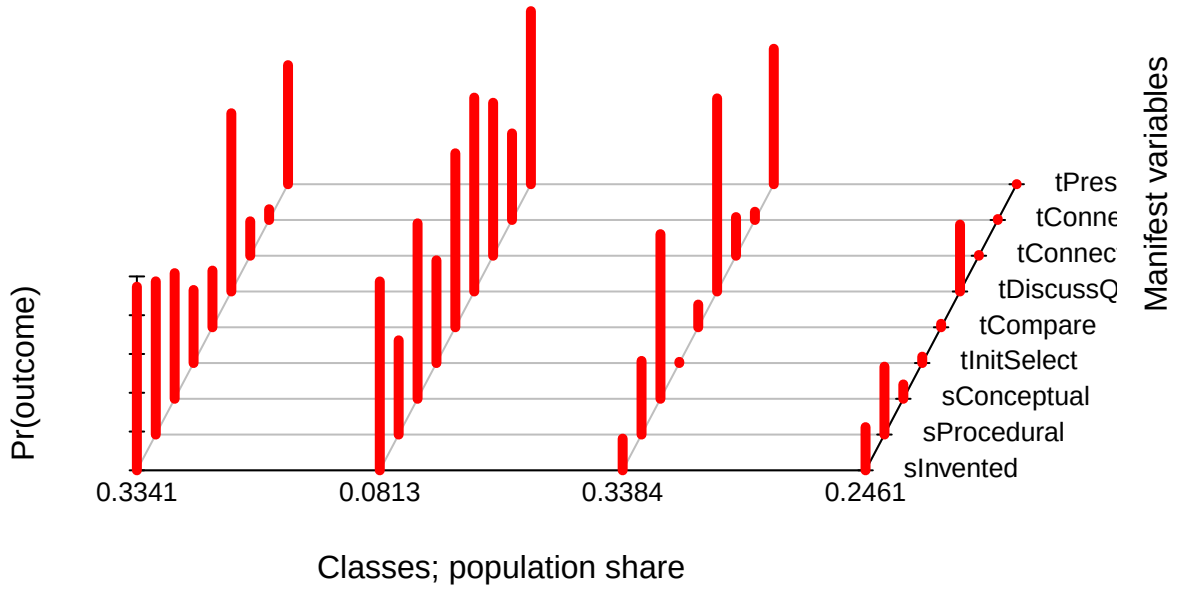
```
## class 1:  0.6232 0.3768
## class 2:  0.4704 0.5296
## class 3:  0.9938 0.0062
## class 4:  0.9670 0.0330
##
## $tCompare
##            Pr(1)  Pr(2)
## class 1:  0.7076 0.2924
## class 2:  0.1027 0.8973
## class 3:  0.8829 0.1171
## class 4:  0.9820 0.0180
##
## $tDiscussQ
##            Pr(1)  Pr(2)
## class 1:  0.0806 0.9194
## class 2:  0.0000 1.0000
## class 3:  0.0039 0.9961
## class 4:  0.6550 0.3450
##
## $tConnectBigIdeas
##            Pr(1)  Pr(2)
## class 1:  0.8224 0.1776
## class 2:  0.2109 0.7891
## class 3:  0.8006 0.1994
## class 4:  0.9961 0.0039
##
## $tConnectOthers
##            Pr(1)  Pr(2)
## class 1:  0.9446 0.0554
## class 2:  0.5536 0.4464
## class 3:  0.9575 0.0425
## class 4:  0.9952 0.0048
##
## $tPressExplain
##            Pr(1)  Pr(2)
## class 1:  0.3858 0.6142
## class 2:  0.1073 0.8927
## class 3:  0.3013 0.6987
## class 4:  0.9995 0.0005
##
## Estimated class population shares
##  0.3341 0.0813 0.3384 0.2461
##
## Predicted class memberships (by modal posterior prob.)
##  0.3466 0.075 0.3374 0.241
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 2813
## number of estimated parameters: 39
## residual degrees of freedom: 472
## maximum log-likelihood: -12241.13
##
```

```
## AIC(4): 24560.26
## BIC(4): 24792
## G^2(4): 655.7218 (Likelihood ratio/deviance statistic)
## X^2(4): 959.0919 (Chi-square goodness of fit)
##
```

5. Examining four-class solution with same random seed as for 4C (1003)

```
f <- cbind(sInvented, sProcedural, sConceptual, tInitSelect, tCompare, tDiscussQ, tConnectBigIdeas, tCo

set.seed(1003)
poLCA(f, ds, nclass = 4, maxiter = 5000, graphs = TRUE, verbose=FALSE)
```




```
## Conditional item response (column) probabilities,
##  by outcome variable, for each class (row)
##
## $sInvented
##            Pr(1)  Pr(2)
## class 1:  0.0528 0.9472
## class 2:  0.0259 0.9741
## class 3:  0.8365 0.1635
```

```
## class 4:  0.7772 0.2228
##
## $sProcedural
##            Pr(1)  Pr(2)
## class 1:  0.2098 0.7902
## class 2:  0.5140 0.4860
## class 3:  0.6201 0.3799
## class 4:  0.6498 0.3502
##
## $sConceptual
##            Pr(1)  Pr(2)
## class 1:  0.3518 0.6482
## class 2:  0.0951 0.9049
## class 3:  0.1507 0.8493
## class 4:  0.9259 0.0741
##
## $tInitSelect
##            Pr(1)  Pr(2)
## class 1:  0.6232 0.3768
## class 2:  0.4704 0.5296
## class 3:  0.9938 0.0062
## class 4:  0.9670 0.0330
##
## $tCompare
##            Pr(1)  Pr(2)
## class 1:  0.7076 0.2924
## class 2:  0.1027 0.8973
## class 3:  0.8829 0.1171
## class 4:  0.9820 0.0180
##
## $tDiscussQ
##            Pr(1)  Pr(2)
## class 1:  0.0806 0.9194
## class 2:  0.0000 1.0000
## class 3:  0.0039 0.9961
## class 4:  0.6550 0.3450
##
## $tConnectBigIdeas
##            Pr(1)  Pr(2)
## class 1:  0.8224 0.1776
## class 2:  0.2109 0.7891
## class 3:  0.8006 0.1994
## class 4:  0.9961 0.0039
##
## $tConnectOthers
##            Pr(1)  Pr(2)
## class 1:  0.9446 0.0554
## class 2:  0.5536 0.4464
## class 3:  0.9575 0.0425
## class 4:  0.9952 0.0048
##
## $tPressExplain
##            Pr(1)  Pr(2)
## class 1:  0.3858 0.6142
```

```
## class 2:  0.1073 0.8927
## class 3:  0.3013 0.6987
## class 4:  0.9995 0.0005
##
## Estimated class population shares
##  0.3341 0.0813 0.3384 0.2461
##
## Predicted class memberships (by modal posterior prob.)
##  0.3466 0.075 0.3374 0.241
##
## =========================================================
## Fit for 4 latent classes:
## =========================================================
## number of observations: 2813
## number of estimated parameters: 39
## residual degrees of freedom: 472
## maximum log-likelihood: -12241.13
##
## AIC(4): 24560.26
## BIC(4): 24792
## G^2(4): 655.7218 (Likelihood ratio/deviance statistic)
## X^2(4): 959.0919 (Chi-square goodness of fit)
##
```