

LCA data modeling Seth-Josh

1. Loading, setting up

```
library(dplyr)
library(purrr)
library(tidyr)
library(poLCA)

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

d <- read_csv(f)

recode_vals <- function(x) {
  if_else
}

d %>% count(`Teacher::TeacherID`, `Teacher::Condition`) %>% arrange(`Teacher::Condition`)

## # A tibble: 47 x 3
##   Teacher::TeacherID `Teacher::Condition`     n
##   <int> <chr>           <int>
## 1 1050200 1 Treatment      32
## 2 1050300 1 Treatment      39
## 3 1050400 1 Treatment      62
## 4 1050500 1 Treatment      12
## 5 1050600 1 Treatment      43
## 6 1070200 1 Treatment     153
## 7 1070300 1 Treatment     165
## 8 1080100 1 Treatment      49
## 9 1180100 1 Treatment     125
## 10 3020300 1 Treatment     92
## # ... with 37 more rows

d %>% count(SegNum, `Teacher::TeacherID`, `ClassObservation::Unit`)

## # A tibble: 2,294 x 4
##   SegNum `Teacher::TeacherID` `ClassObservation::Unit`     n
##   <int> <int>           <int> <int>
## 1 1      1020200            1     1
## 2 1      1020300            1     1
## 3 1      1020300            2     1
## 4 1      1050200            1     1
## 5 1      1050200            3     2
## 6 1      1050200            4     1
## 7 1      1050300            1     1
## 8 1      1050300            2     1
## 9 1      1050300            3     1
## 10 1     1050300            4     1
## # ... with 2,284 more rows
```

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, tConnectBigIdeas)
  map_df(replace_na, 0) %>%
  map_df(add_one)
```

3. Choosing the number of classes/profiles

Using latent class analysis through the **poLCA** R package.

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

od <- map(1:9, poLCA, formula = f, data = ds, maxiter = 5000) %>%
  map_df(broom::glance)

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.4941 0.5059
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.4817 0.5183
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.4042 0.5958
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.8208 0.1792
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.7853 0.2147
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.1895 0.8105
##
## $tConnectBigIdeas
```

```

##          Pr(1) Pr(2)
## class 1:  0.808 0.192
##
## $tConnectOthers
##          Pr(1) Pr(2)
## class 1:  0.9296 0.0704
##
## $tPressExplain
##          Pr(1) Pr(2)
## class 1:  0.4856 0.5144
##
## Estimated class population shares
## 1
##
## Predicted class memberships (by modal posterior prob.)
## 1
##
## =====
## Fit for 1 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 9
## residual degrees of freedom: 502
## maximum log-likelihood: -13987.36
##
## AIC(1): 27992.72
## BIC(1): 28046.19
## G^2(1): 4148.178 (Likelihood ratio/deviance statistic)
## X^2(1): 27801.46 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1) Pr(2)
## class 1:  0.7400 0.2600
## class 2:  0.3927 0.6073
##
## $sProcedural
##          Pr(1) Pr(2)
## class 1:  0.6299 0.3701
## class 2:  0.4206 0.5794
##
## $sConceptual
##          Pr(1) Pr(2)
## class 1:  0.8436 0.1564
## class 2:  0.2230 0.7770
##
## $tInitSelect
##          Pr(1) Pr(2)
## class 1:  0.9505 0.0495
## class 2:  0.7674 0.2326
##
## $tCompare

```

```

##          Pr(1)  Pr(2)
## class 1: 0.9863 0.0137
## class 2: 0.7024 0.2976
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.5815 0.4185
## class 2: 0.0278 0.9722
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.9977 0.0023
## class 2: 0.7298 0.2702
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.9953 0.0047
## class 2: 0.9025 0.0975
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.9935 0.0065
## class 2: 0.2762 0.7238
##
## Estimated class population shares
## 0.292 0.708
##
## Predicted class memberships (by modal posterior prob.)
## 0.2705 0.7295
##
## =====
## Fit for 2 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 19
## residual degrees of freedom: 492
## maximum log-likelihood: -12828.56
##
## AIC(2): 25695.12
## BIC(2): 25808.02
## G^2(2): 1830.582 (Likelihood ratio/deviance statistic)
## X^2(2): 5480.488 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.0366 0.9634
## class 2: 0.8345 0.1655
## class 3: 0.7292 0.2708
##
## $sProcedural
##          Pr(1)  Pr(2)

```

```

## class 1: 0.2835 0.7165
## class 2: 0.6037 0.3963
## class 3: 0.6160 0.3840
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.2729 0.7271
## class 2: 0.1600 0.8400
## class 3: 0.9097 0.0903
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.5817 0.4183
## class 2: 0.9930 0.0070
## class 3: 0.9511 0.0489
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.5643 0.4357
## class 2: 0.8868 0.1132
## class 3: 0.9793 0.0207
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.0478 0.9522
## class 2: 0.0046 0.9954
## class 3: 0.6341 0.3659
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.6765 0.3235
## class 2: 0.8126 0.1874
## class 3: 0.9951 0.0049
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.8586 0.1414
## class 2: 0.9601 0.0399
## class 3: 0.9947 0.0053
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.2956 0.7044
## class 2: 0.3065 0.6935
## class 3: 0.9937 0.0063
##
## Estimated class population shares
## 0.3913 0.3418 0.2668
##
## Predicted class memberships (by modal posterior prob.)
## 0.4042 0.3228 0.273
##
## =====
## Fit for 3 latent classes:

```

```

## =====
## number of observations: 2813
## number of estimated parameters: 29
## residual degrees of freedom: 482
## maximum log-likelihood: -12400.05
##
## AIC(3): 24858.1
## BIC(3): 25030.41
## G^2(3): 973.5587 (Likelihood ratio/deviance statistic)
## X^2(3): 2279.99 (Chi-square goodness of fit)
##
## 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: 0.0528 0.9472
## class 4: 0.8365 0.1635
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.6498 0.3502
## class 2: 0.5140 0.4860
## class 3: 0.2098 0.7902
## class 4: 0.6201 0.3799
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.9259 0.0741
## class 2: 0.0951 0.9049
## class 3: 0.3518 0.6482
## class 4: 0.1507 0.8493
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.9670 0.0330
## class 2: 0.4704 0.5296
## class 3: 0.6232 0.3768
## class 4: 0.9938 0.0062
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.9820 0.0180
## class 2: 0.1027 0.8973
## class 3: 0.7076 0.2924
## class 4: 0.8829 0.1171
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.6550 0.3450
## class 2: 0.0000 1.0000
## class 3: 0.0806 0.9194

```

```

## class 4: 0.0039 0.9961
##
## $tConnectBigIdeas
##           Pr(1)  Pr(2)
## class 1: 0.9961 0.0039
## class 2: 0.2109 0.7891
## class 3: 0.8224 0.1776
## class 4: 0.8006 0.1994
##
## $tConnectOthers
##           Pr(1)  Pr(2)
## class 1: 0.9952 0.0048
## class 2: 0.5536 0.4464
## class 3: 0.9446 0.0554
## class 4: 0.9575 0.0425
##
## $tPressExplain
##           Pr(1)  Pr(2)
## class 1: 0.9995 0.0005
## class 2: 0.1073 0.8927
## class 3: 0.3858 0.6142
## class 4: 0.3013 0.6987
##
## Estimated class population shares
## 0.2461 0.0813 0.3341 0.3384
##
## Predicted class memberships (by modal posterior prob.)
## 0.241 0.075 0.3466 0.3374
##
## =====
## 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)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##           Pr(1)  Pr(2)
## class 1: 0.1246 0.8754
## class 2: 0.0366 0.9634
## class 3: 0.0141 0.9859
## class 4: 0.9436 0.0564
## class 5: 0.8732 0.1268
##
## $sProcedural

```

```

##          Pr(1)  Pr(2)
## class 1: 0.2635 0.7365
## class 2: 0.2614 0.7386
## class 3: 0.7570 0.2430
## class 4: 0.7358 0.2642
## class 5: 0.6157 0.3843
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.7499 0.2501
## class 2: 0.2486 0.7514
## class 3: 0.0597 0.9403
## class 4: 0.9445 0.0555
## class 5: 0.1590 0.8410
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.7365 0.2635
## class 2: 0.6360 0.3640
## class 3: 0.2043 0.7957
## class 4: 0.9920 0.0080
## class 5: 0.9913 0.0087
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.9429 0.0571
## class 2: 0.5838 0.4162
## class 3: 0.0000 1.0000
## class 4: 0.9861 0.0139
## class 5: 0.8845 0.1155
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.4783 0.5217
## class 2: 0.0251 0.9749
## class 3: 0.0000 1.0000
## class 4: 0.6520 0.3480
## class 5: 0.0055 0.9945
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.9839 0.0161
## class 2: 0.6948 0.3052
## class 3: 0.1034 0.8966
## class 4: 0.9984 0.0016
## class 5: 0.8097 0.1903
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.9550 0.0450
## class 2: 0.9290 0.0710
## class 3: 0.0000 1.0000
## class 4: 0.9993 0.0007
## class 5: 0.9563 0.0437

```

```

## 
## $tPressExplain
##           Pr(1)  Pr(2)
## class 1: 0.8770 0.1230
## class 2: 0.2681 0.7319
## class 3: 0.0214 0.9786
## class 4: 1.0000 0.0000
## class 5: 0.3124 0.6876
##
## Estimated class population shares
## 0.1157 0.3382 0.0266 0.1899 0.3295
##
## Predicted class memberships (by modal posterior prob.)
## 0.1013 0.3726 0.0277 0.1927 0.3057
##
## =====
## Fit for 5 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 49
## residual degrees of freedom: 462
## maximum log-likelihood: -12169.39
##
## AIC(5): 24436.79
## BIC(5): 24727.95
## G^2(5): 512.2494 (Likelihood ratio/deviance statistic)
## X^2(5): 761.1207 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##           Pr(1)  Pr(2)
## class 1: 0.0079 0.9921
## class 2: 0.9418 0.0582
## class 3: 0.0676 0.9324
## class 4: 0.9130 0.0870
## class 5: 0.8416 0.1584
## class 6: 0.0469 0.9531
##
## $sProcedural
##           Pr(1)  Pr(2)
## class 1: 0.7544 0.2456
## class 2: 0.7605 0.2395
## class 3: 0.2656 0.7344
## class 4: 0.5734 0.4266
## class 5: 0.6528 0.3472
## class 6: 0.2562 0.7438
##
## $sConceptual
##           Pr(1)  Pr(2)
## class 1: 0.0603 0.9397
## class 2: 0.9475 0.0525
## class 3: 0.7570 0.2430

```

```

## class 4: 0.5168 0.4832
## class 5: 0.0000 1.0000
## class 6: 0.2613 0.7387
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.2026 0.7974
## class 2: 0.9901 0.0099
## class 3: 0.7305 0.2695
## class 4: 0.9852 0.0148
## class 5: 0.9992 0.0008
## class 6: 0.6346 0.3654
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.0000 1.0000
## class 2: 0.9937 0.0063
## class 3: 0.9581 0.0419
## class 4: 0.9243 0.0757
## class 5: 0.8638 0.1362
## class 6: 0.5881 0.4119
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.0000 1.0000
## class 2: 0.8108 0.1892
## class 3: 0.5138 0.4862
## class 4: 0.0180 0.9820
## class 5: 0.0037 0.9963
## class 6: 0.0283 0.9717
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.1017 0.8983
## class 2: 1.0000 0.0000
## class 3: 0.9867 0.0133
## class 4: 0.9392 0.0608
## class 5: 0.7292 0.2708
## class 6: 0.7040 0.2960
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.0000 1.0000
## class 2: 0.9974 0.0026
## class 3: 0.9573 0.0427
## class 4: 0.9883 0.0117
## class 5: 0.9349 0.0651
## class 6: 0.9302 0.0698
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.0222 0.9778
## class 2: 1.0000 0.0000
## class 3: 0.8956 0.1044

```

```

## class 4: 0.6295 0.3705
## class 5: 0.1754 0.8246
## class 6: 0.2787 0.7213
##
## Estimated class population shares
## 0.0266 0.1532 0.1004 0.1758 0.1974 0.3466
##
## Predicted class memberships (by modal posterior prob.)
## 0.0274 0.1212 0.0985 0.1817 0.1845 0.3868
##
## =====
## Fit for 6 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 59
## residual degrees of freedom: 452
## maximum log-likelihood: -12147.3
##
## AIC(6): 24412.61
## BIC(6): 24763.18
## G^2(6): 468.0677 (Likelihood ratio/deviance statistic)
## X^2(6): 769.3171 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.8345 0.1655
## class 2: 0.0399 0.9601
## class 3: 0.9220 0.0780
## class 4: 0.0093 0.9907
## class 5: 0.0000 1.0000
## class 6: 0.2745 0.7255
## class 7: 0.9386 0.0614
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.6354 0.3646
## class 2: 0.2890 0.7110
## class 3: 0.6669 0.3331
## class 4: 0.7513 0.2487
## class 5: 0.3011 0.6989
## class 6: 0.0000 1.0000
## class 7: 0.7473 0.2527
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.0000 1.0000
## class 2: 0.2107 0.7893
## class 3: 0.6265 0.3735
## class 4: 0.0681 0.9319
## class 5: 0.7487 0.2513
## class 6: 0.7188 0.2812

```

```

## class 7: 0.9374 0.0626
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.9991 0.0009
## class 2: 0.6073 0.3927
## class 3: 0.9835 0.0165
## class 4: 0.2089 0.7911
## class 5: 0.7135 0.2865
## class 6: 0.8366 0.1634
## class 7: 0.9878 0.0122
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.8651 0.1349
## class 2: 0.5547 0.4453
## class 3: 0.9468 0.0532
## class 4: 0.0000 1.0000
## class 5: 0.9644 0.0356
## class 6: 0.8596 0.1404
## class 7: 0.9900 0.0100
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.0043 0.9957
## class 2: 0.0273 0.9727
## class 3: 0.0000 1.0000
## class 4: 0.0000 1.0000
## class 5: 0.5525 0.4475
## class 6: 0.1090 0.8910
## class 7: 1.0000 0.0000
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.7425 0.2575
## class 2: 0.6842 0.3158
## class 3: 0.9657 0.0343
## class 4: 0.1063 0.8937
## class 5: 0.9956 0.0044
## class 6: 0.8736 0.1264
## class 7: 0.9964 0.0036
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.9374 0.0626
## class 2: 0.9355 0.0645
## class 3: 1.0000 0.0000
## class 4: 0.0000 1.0000
## class 5: 0.9753 0.0247
## class 6: 0.8913 0.1087
## class 7: 0.9954 0.0046
##
## $tPressExplain
##          Pr(1)  Pr(2)

```

```

## class 1: 0.1816 0.8184
## class 2: 0.2902 0.7098
## class 3: 0.7600 0.2400
## class 4: 0.0179 0.9821
## class 5: 0.9788 0.0212
## class 6: 0.2926 0.7074
## class 7: 1.0000 0.0000
##
## Estimated class population shares
## 0.2194 0.3034 0.1714 0.0268 0.0788 0.0711 0.129
##
## Predicted class memberships (by modal posterior prob.)
## 0.2119 0.3399 0.1681 0.0281 0.0931 0.037 0.1219
##
## =====
## Fit for 7 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 69
## residual degrees of freedom: 442
## maximum log-likelihood: -12127.53
##
## AIC(7): 24393.07
## BIC(7): 24803.06
## G^2(7): 428.5271 (Likelihood ratio/deviance statistic)
## X^2(7): 691.5893 (Chi-square goodness of fit)
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.2676 0.7324
## class 2: 1.0000 0.0000
## class 3: 0.8166 0.1834
## class 4: 0.2209 0.7791
## class 5: 0.0000 1.0000
## class 6: 0.0142 0.9858
## class 7: 0.2538 0.7462
## class 8: 0.0000 1.0000
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.0658 0.9342
## class 2: 0.7361 0.2639
## class 3: 0.7148 0.2852
## class 4: 0.2831 0.7169
## class 5: 0.3852 0.6148
## class 6: 0.7904 0.2096
## class 7: 0.5454 0.4546
## class 8: 0.2512 0.7488
##
## $sConceptual
##          Pr(1)  Pr(2)

```

```

## class 1: 0.4934 0.5066
## class 2: 0.9394 0.0606
## class 3: 0.1318 0.8682
## class 4: 0.1286 0.8714
## class 5: 0.8171 0.1829
## class 6: 0.0621 0.9379
## class 7: 0.2387 0.7613
## class 8: 0.2352 0.7648
##
## $tInitSelect
##           Pr(1)  Pr(2)
## class 1: 0.9011 0.0989
## class 2: 0.9887 0.0113
## class 3: 0.9875 0.0125
## class 4: 1.0000 0.0000
## class 5: 0.7722 0.2278
## class 6: 0.1503 0.8497
## class 7: 0.8790 0.1210
## class 8: 0.0000 1.0000
##
## $tCompare
##           Pr(1)  Pr(2)
## class 1: 0.8524 0.1476
## class 2: 0.9915 0.0085
## class 3: 0.9485 0.0515
## class 4: 0.4749 0.5251
## class 5: 0.9784 0.0216
## class 6: 0.0000 1.0000
## class 7: 0.0966 0.9034
## class 8: 0.5932 0.4068
##
## $tDiscussQ
##           Pr(1)  Pr(2)
## class 1: 0.0689 0.9311
## class 2: 0.6567 0.3433
## class 3: 0.0000 1.0000
## class 4: 0.0000 1.0000
## class 5: 0.6596 0.3404
## class 6: 0.0000 1.0000
## class 7: 0.0546 0.9454
## class 8: 0.0573 0.9427
##
## $tConnectBigIdeas
##           Pr(1)  Pr(2)
## class 1: 0.8971 0.1029
## class 2: 1.0000 0.0000
## class 3: 0.8345 0.1655
## class 4: 0.4782 0.5218
## class 5: 0.9920 0.0080
## class 6: 0.1119 0.8881
## class 7: 0.6332 0.3668
## class 8: 0.7340 0.2660
##
## $tConnectOthers

```

```

##          Pr(1)  Pr(2)
## class 1: 0.9347 0.0653
## class 2: 0.9961 0.0039
## class 3: 0.9636 0.0364
## class 4: 0.8900 0.1100
## class 5: 0.9823 0.0177
## class 6: 0.0000 1.0000
## class 7: 0.9707 0.0293
## class 8: 0.9218 0.0782
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.4151 0.5849
## class 2: 1.0000 0.0000
## class 3: 0.3284 0.6716
## class 4: 0.0000 1.0000
## class 5: 0.9884 0.0116
## class 6: 0.0161 0.9839
## class 7: 0.6981 0.3019
## class 8: 0.3060 0.6940
##
## Estimated class population shares
## 0.1631 0.1829 0.2822 0.1107 0.0743 0.025 0.0484 0.1134
##
## Predicted class memberships (by modal posterior prob.)
## 0.1585 0.193 0.2904 0.102 0.0732 0.0252 0.0341 0.1234
##
## =====
## Fit for 8 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 79
## residual degrees of freedom: 432
## maximum log-likelihood: -12109.41
##
## AIC(8): 24376.83
## BIC(8): 24846.25
## G^2(8): 392.2909 (Likelihood ratio/deviance statistic)
## X^2(8): 584.6844 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.0000 1.0000
## class 2: 0.0518 0.9482
## class 3: 0.0075 0.9925
## class 4: 1.0000 0.0000
## class 5: 0.3357 0.6643
## class 6: 0.8540 0.1460
## class 7: 0.1314 0.8686

```

```

## class 8: 1.0000 0.0000
## class 9: 0.0119 0.9881
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.3431 0.6569
## class 2: 0.2669 0.7331
## class 3: 0.8284 0.1716
## class 4: 0.7849 0.2151
## class 5: 0.0000 1.0000
## class 6: 0.6520 0.3480
## class 7: 0.3641 0.6359
## class 8: 0.6206 0.3794
## class 9: 0.2767 0.7233
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.8159 0.1841
## class 2: 0.1772 0.8228
## class 3: 0.0484 0.9516
## class 4: 1.0000 0.0000
## class 5: 0.6750 0.3250
## class 6: 0.1216 0.8784
## class 7: 0.5448 0.4552
## class 8: 0.6643 0.3357
## class 9: 0.2682 0.7318
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.8270 0.1730
## class 2: 0.6827 0.3173
## class 3: 0.1252 0.8748
## class 4: 0.9873 0.0127
## class 5: 1.0000 0.0000
## class 6: 0.9888 0.0112
## class 7: 0.2622 0.7378
## class 8: 0.9903 0.0097
## class 9: 0.5764 0.4236
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.9758 0.0242
## class 2: 0.4078 0.5922
## class 3: 0.0000 1.0000
## class 4: 1.0000 0.0000
## class 5: 0.8992 0.1008
## class 6: 0.8858 0.1142
## class 7: 0.5885 0.4115
## class 8: 0.9363 0.0637
## class 9: 0.7125 0.2875
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.6430 0.3570

```

```

## class 2: 0.0000 1.0000
## class 3: 0.0000 1.0000
## class 4: 0.8787 0.1213
## class 5: 0.0916 0.9084
## class 6: 0.0028 0.9972
## class 7: 0.4904 0.5096
## class 8: 0.2111 0.7889
## class 9: 0.0000 1.0000
##
## $tConnectBigIdeas
##           Pr(1)  Pr(2)
## class 1: 0.9914 0.0086
## class 2: 0.3092 0.6908
## class 3: 0.1058 0.8942
## class 4: 1.0000 0.0000
## class 5: 0.8691 0.1309
## class 6: 0.7903 0.2097
## class 7: 0.7899 0.2101
## class 8: 0.9843 0.0157
## class 9: 1.0000 0.0000
##
## $tConnectOthers
##           Pr(1)  Pr(2)
## class 1: 1.0000 0.0000
## class 2: 0.8790 0.1210
## class 3: 0.0000 1.0000
## class 4: 1.0000 0.0000
## class 5: 0.8965 0.1035
## class 6: 0.9538 0.0462
## class 7: 0.8274 0.1726
## class 8: 0.9899 0.0101
## class 9: 0.9650 0.0350
##
## $tPressExplain
##           Pr(1)  Pr(2)
## class 1: 0.9935 0.0065
## class 2: 0.2179 0.7821
## class 3: 0.0000 1.0000
## class 4: 1.0000 0.0000
## class 5: 0.2119 0.7881
## class 6: 0.2333 0.7667
## class 7: 0.5005 0.4995
## class 8: 1.0000 0.0000
## class 9: 0.3862 0.6138
##
## Estimated class population shares
## 0.075 0.1401 0.0225 0.1116 0.0523 0.2842 0.0298 0.1088 0.1756
##
## Predicted class memberships (by modal posterior prob.)
## 0.0786 0.1134 0.0242 0.1123 0.0427 0.2975 0.0124 0.0871 0.2318
##
## =====
## Fit for 9 latent classes:
## =====

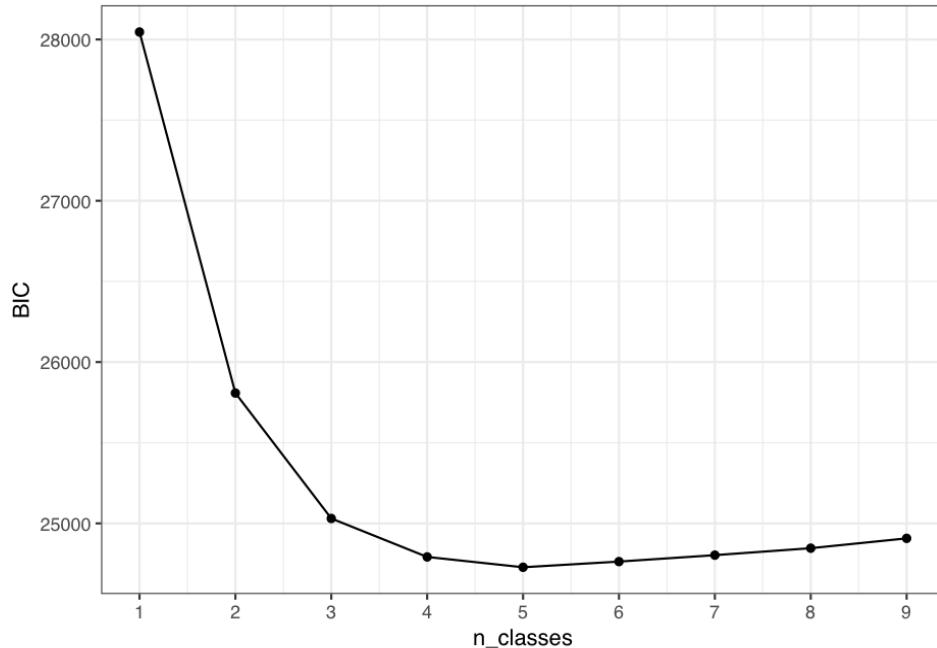
```

```

## number of observations: 2813
## number of estimated parameters: 89
## residual degrees of freedom: 422
## maximum log-likelihood: -12100.21
##
## AIC(9): 24378.41
## BIC(9): 24907.25
## G^2(9): 373.8727 (Likelihood ratio/deviance statistic)
## X^2(9): 443.6173 (Chi-square goodness of fit)
##
## ALERT: iterations finished, MAXIMUM LIKELIHOOD NOT FOUND
##

od %>%
  mutate(n_classes = 1:9) %>%
  ggplot(aes(x = n_classes, y = BIC)) +
  geom_point() +
  geom_line() +
  scale_x_continuous(breaks = 1:9, labels = 1:9) +
  theme_bw()

```

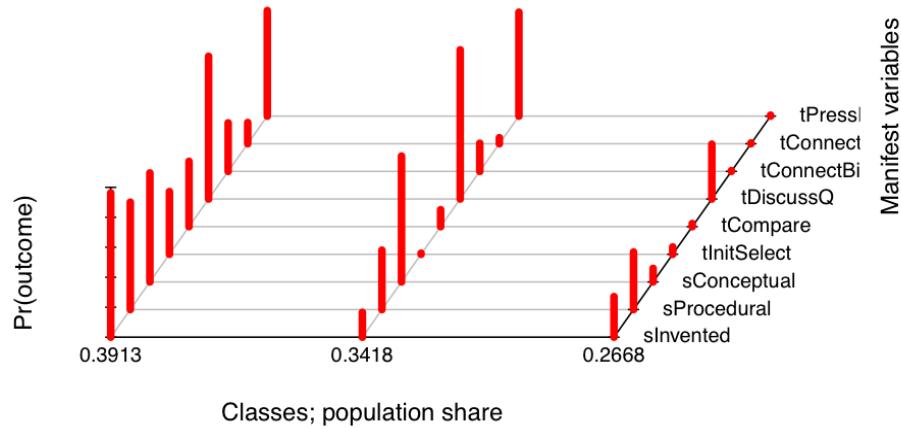


Based on this fit statistic—the Bayesian Information Criteria, which is just a transformation of the log-likelihood, and is usually recommended along with the AIC as one criterion for model selection—it looks like 3 and especially 4 or 5 class solutions seem reasonable.

4. Examining 3, 4, and 5 class solutions

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

```
m3 <- poLCA(f, ds, nclass = 3, maxiter = 5000, graphs = TRUE)
```



```
## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)  Pr(2)
## class 1: 0.0366 0.9634
## class 2: 0.8345 0.1655
## class 3: 0.7292 0.2708
##
## $sProcedural
##          Pr(1)  Pr(2)
## class 1: 0.2835 0.7165
## class 2: 0.6037 0.3963
## class 3: 0.6160 0.3840
##
## $sConceptual
##          Pr(1)  Pr(2)
## class 1: 0.2729 0.7271
```

```

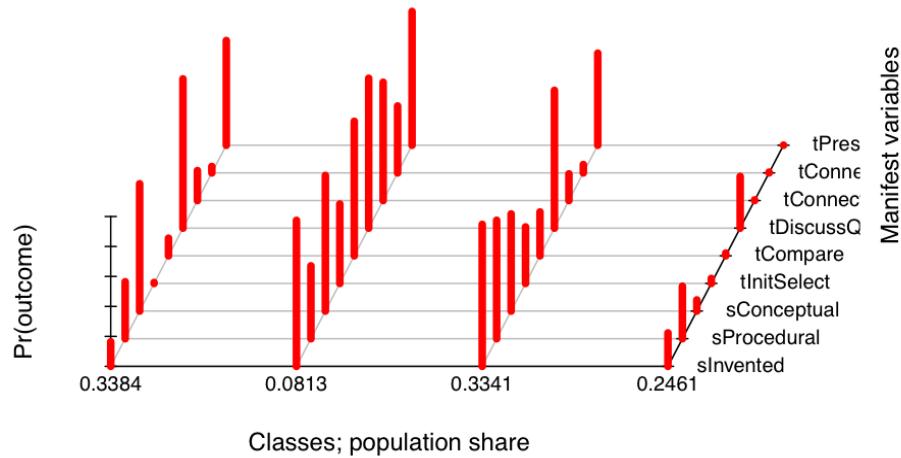
## class 2: 0.1600 0.8400
## class 3: 0.9097 0.0903
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.5817 0.4183
## class 2: 0.9930 0.0070
## class 3: 0.9511 0.0489
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.5643 0.4357
## class 2: 0.8868 0.1132
## class 3: 0.9793 0.0207
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.0478 0.9522
## class 2: 0.0046 0.9954
## class 3: 0.6341 0.3659
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.6765 0.3235
## class 2: 0.8126 0.1874
## class 3: 0.9951 0.0049
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.8586 0.1414
## class 2: 0.9601 0.0399
## class 3: 0.9947 0.0053
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.2956 0.7044
## class 2: 0.3065 0.6935
## class 3: 0.9937 0.0063
##
## Estimated class population shares
## 0.3913 0.3418 0.2668
##
## Predicted class memberships (by modal posterior prob.)
## 0.4042 0.3228 0.273
##
## =====
## Fit for 3 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 29
## residual degrees of freedom: 482
## maximum log-likelihood: -12400.05
##
## AIC(3): 24858.1

```

```

## BIC(3): 25030.41
## G^2(3): 973.5587 (Likelihood ratio/deviance statistic)
## X^2(3): 2279.99 (Chi-square goodness of fit)
##
m4 <- poLCA(f, ds, nclass = 4, maxiter = 5000, graphs = TRUE)

```



```

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

```

```

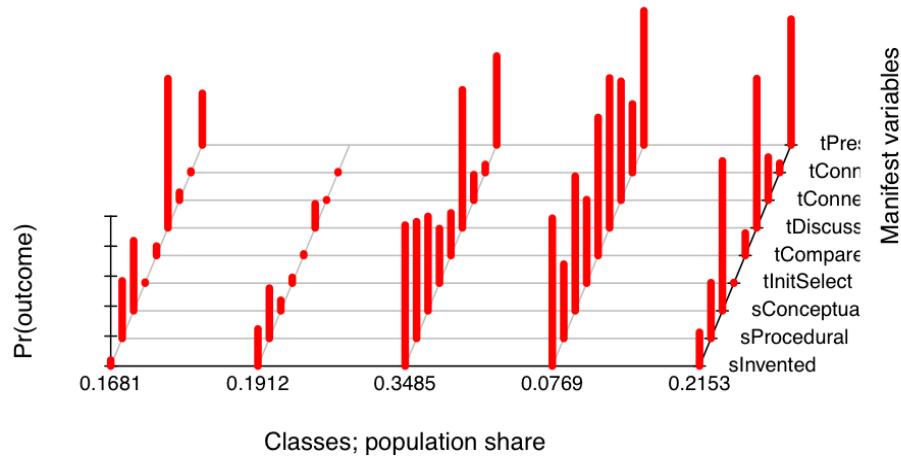
##          Pr(1)  Pr(2)
## class 1: 0.1507 0.8493
## class 2: 0.0951 0.9049
## class 3: 0.3518 0.6482
## class 4: 0.9259 0.0741
##
## $tInitSelect
##          Pr(1)  Pr(2)
## class 1: 0.9938 0.0062
## class 2: 0.4704 0.5296
## class 3: 0.6232 0.3768
## class 4: 0.9670 0.0330
##
## $tCompare
##          Pr(1)  Pr(2)
## class 1: 0.8829 0.1171
## class 2: 0.1027 0.8973
## class 3: 0.7076 0.2924
## class 4: 0.9820 0.0180
##
## $tDiscussQ
##          Pr(1)  Pr(2)
## class 1: 0.0039 0.9961
## class 2: 0.0000 1.0000
## class 3: 0.0806 0.9194
## class 4: 0.6550 0.3450
##
## $tConnectBigIdeas
##          Pr(1)  Pr(2)
## class 1: 0.8006 0.1994
## class 2: 0.2109 0.7891
## class 3: 0.8224 0.1776
## class 4: 0.9961 0.0039
##
## $tConnectOthers
##          Pr(1)  Pr(2)
## class 1: 0.9575 0.0425
## class 2: 0.5536 0.4464
## class 3: 0.9446 0.0554
## class 4: 0.9952 0.0048
##
## $tPressExplain
##          Pr(1)  Pr(2)
## class 1: 0.3013 0.6987
## class 2: 0.1073 0.8927
## class 3: 0.3858 0.6142
## class 4: 0.9995 0.0005
##
## Estimated class population shares
## 0.3384 0.0813 0.3341 0.2461
##
## Predicted class memberships (by modal posterior prob.)
## 0.3374 0.075 0.3466 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): 959.092 (Chi-square goodness of fit)
##
m5 <- poLCA(f, ds, nclass = 5, maxiter = 5000, graphs = TRUE)

```



```

## Conditional item response (column) probabilities,
## by outcome variable, for each class (row)
##
## $sInvented
##          Pr(1)   Pr(2)
## class 1: 0.9608 0.0392
## class 2: 0.7530 0.2470
## class 3: 0.0597 0.9403
## class 4: 0.0138 0.9862

```

```

## class 5: 0.7749 0.2251
##
## $sProcedural
##          Pr(1) Pr(2)
## class 1: 0.6161 0.3839
## class 2: 0.6646 0.3354
## class 3: 0.2211 0.7789
## class 4: 0.5044 0.4956
## class 5: 0.6281 0.3719
##
## $sConceptual
##          Pr(1) Pr(2)
## class 1: 0.5325 0.4675
## class 2: 0.9299 0.0701
## class 3: 0.3703 0.6297
## class 4: 0.1023 0.8977
## class 5: 0.0000 1.0000
##
## $tInitSelect
##          Pr(1) Pr(2)
## class 1: 0.9937 0.0063
## class 2: 0.9615 0.0385
## class 3: 0.6359 0.3641
## class 4: 0.4429 0.5571
## class 5: 0.9954 0.0046
##
## $tCompare
##          Pr(1) Pr(2)
## class 1: 0.9375 0.0625
## class 2: 0.9896 0.0104
## class 3: 0.7159 0.2841
## class 4: 0.0786 0.9214
## class 5: 0.8499 0.1501
##
## $tDiscussQ
##          Pr(1) Pr(2)
## class 1: 0.0043 0.9957
## class 2: 0.8397 0.1603
## class 3: 0.0788 0.9212
## class 4: 0.0000 1.0000
## class 5: 0.0036 0.9964
##
## $tConnectBigIdeas
##          Pr(1) Pr(2)
## class 1: 0.9469 0.0531
## class 2: 0.9969 0.0031
## class 3: 0.8294 0.1706
## class 4: 0.2064 0.7936
## class 5: 0.7123 0.2877
##
## $tConnectOthers
##          Pr(1) Pr(2)
## class 1: 0.9895 0.0105
## class 2: 0.9946 0.0054

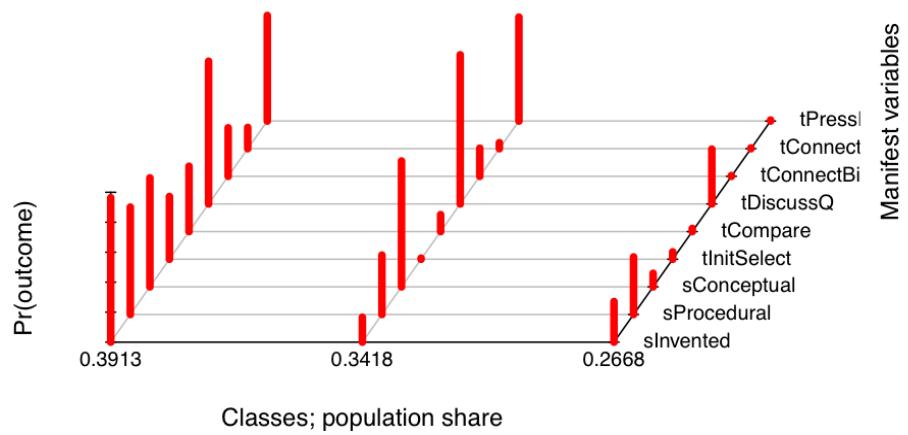
```

```

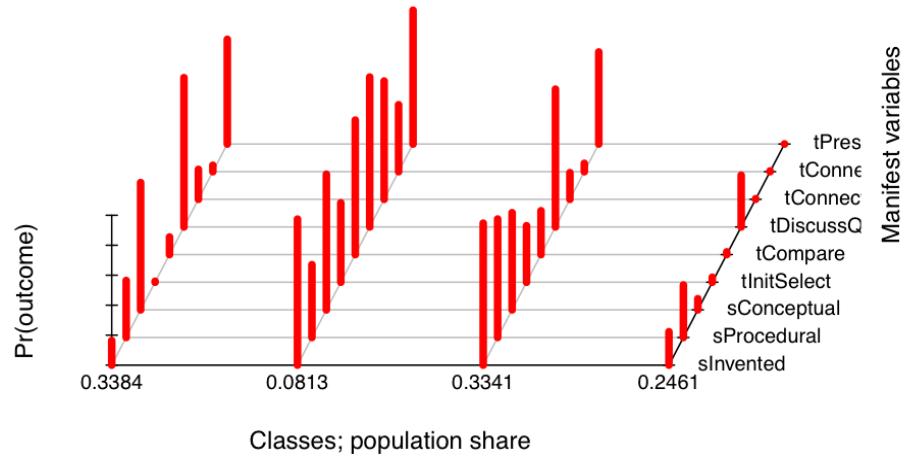
## class 3: 0.9457 0.0543
## class 4: 0.5423 0.4577
## class 5: 0.9374 0.0626
##
## $tPressExplain
##           Pr(1)  Pr(2)
## class 1: 0.6558 0.3442
## class 2: 1.0000 0.0000
## class 3: 0.4063 0.5937
## class 4: 0.1047 0.8953
## class 5: 0.1604 0.8396
##
## Estimated class population shares
## 0.1681 0.1912 0.3485 0.0769 0.2153
##
## Predicted class memberships (by modal posterior prob.)
## 0.1774 0.1696 0.3626 0.0736 0.2169
##
## =====
## Fit for 5 latent classes:
## =====
## number of observations: 2813
## number of estimated parameters: 49
## residual degrees of freedom: 462
## maximum log-likelihood: -12213.32
##
## AIC(5): 24524.64
## BIC(5): 24815.8
## G^2(5): 600.1039 (Likelihood ratio/deviance statistic)
## X^2(5): 913.0771 (Chi-square goodness of fit)
##

```

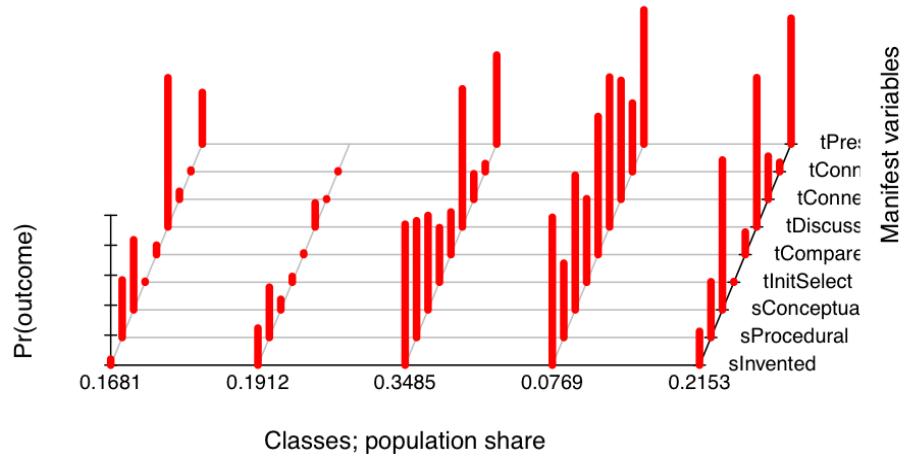
```
plot(m3)
```



```
plot(m4)
```



```
plot(m5)
```

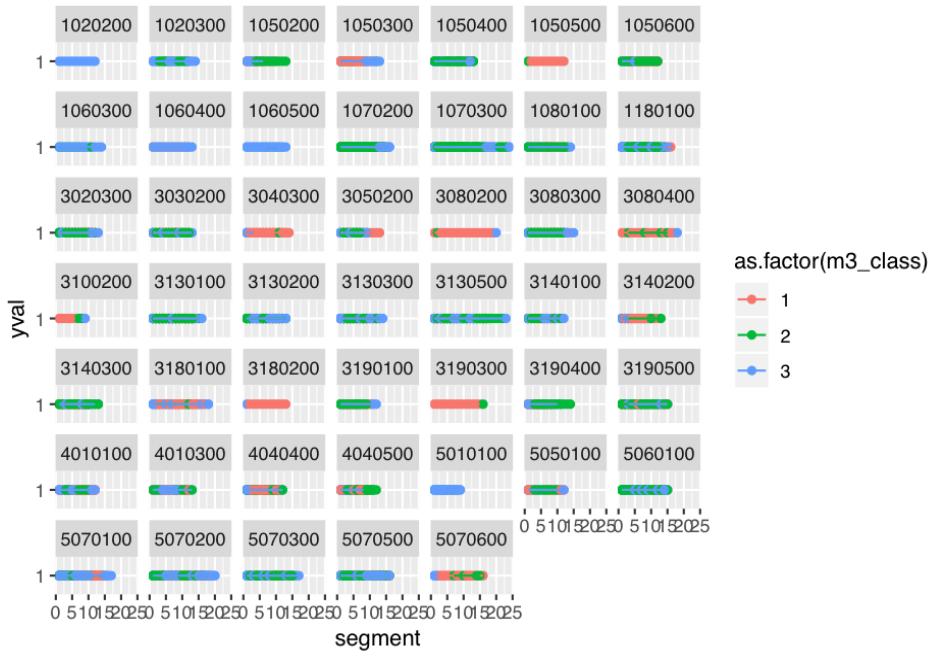


```

ds$m3_class <- m3$predclass
ds$m4_class <- m4$predclass
ds$m5_class <- m5$predclass

ds$segment <- d$SegNum
ds$teacher <- as.factor(d$`Teacher::TeacherID`)

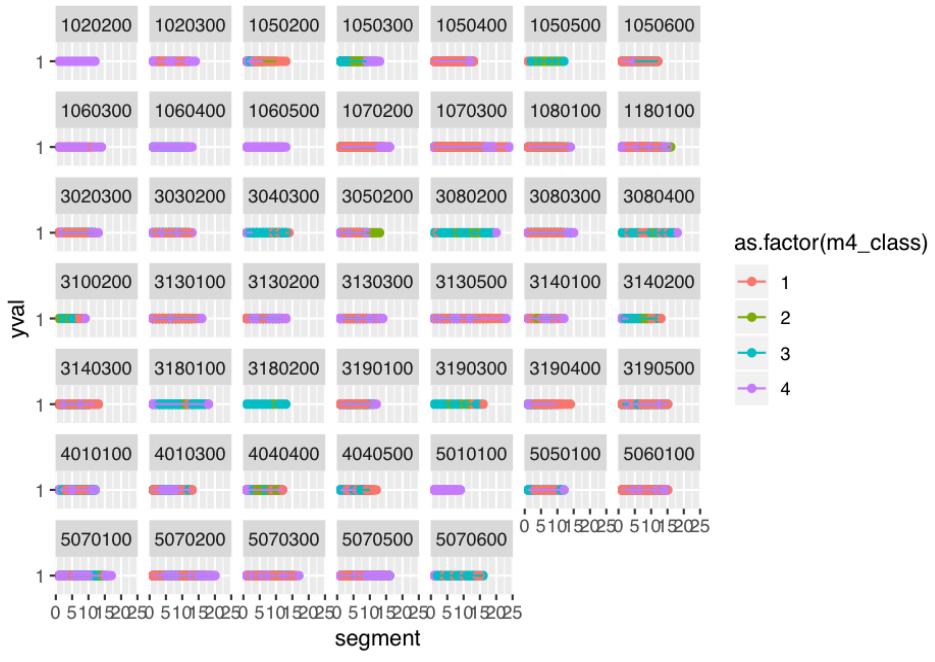
ds %>%
  mutate(yval = 1) %>%
  ggplot(aes(x = segment, y =yval, color = as.factor(m3_class))) + geom_point() + geom_line() + facet_wrap(~scale_y_continuous(labels = 1:3, breaks = 1:3)
  
```



```
ggttitle("3 class/profile solution")
```

```
## $title
## [1] "3 class/profile solution"
##
## $subtitle
## NULL
##
## attr(,"class")
## [1] "labels"

ds %>%
  mutate(yval = 1) %>%
  ggplot(aes(x = segment,y =yval, color = as.factor(m4_class))) + geom_point() + geom_line() + facet_
  scale_y_continuous(labels = 1:3, breaks = 1:3)
```



```
ggttitle("4 class/profile solution")
```

```
## $title
## [1] "4 class/profile solution"
##
## $subtitle
## NULL
##
## attr(,"class")
## [1] "labels"

ds %>%
  mutate(yval = 1) %>%
  ggplot(aes(x = segment,y =yval, color = as.factor(m5_class))) + geom_point() + geom_line() + facet_
  scale_y_continuous(labels = 1:3, breaks = 1:3)
```

