

GenSamp: RESULTS

Gleb Furman¹, James E. Pustejovsky², & Elizabeth Tipton³

¹ University of Texas at Austin

² University of Wisconsin-Madison

³ Northwestern University

Author Note

Correspondence concerning this article should be addressed to Gleb Furman,
University of Texas at Austin, SZB 504, 1912 Speedway, Austin, Texas 78712. E-mail:
gleb.furman@utexas.edu

GenSamp: RESULTS

Setup

Packages and Data

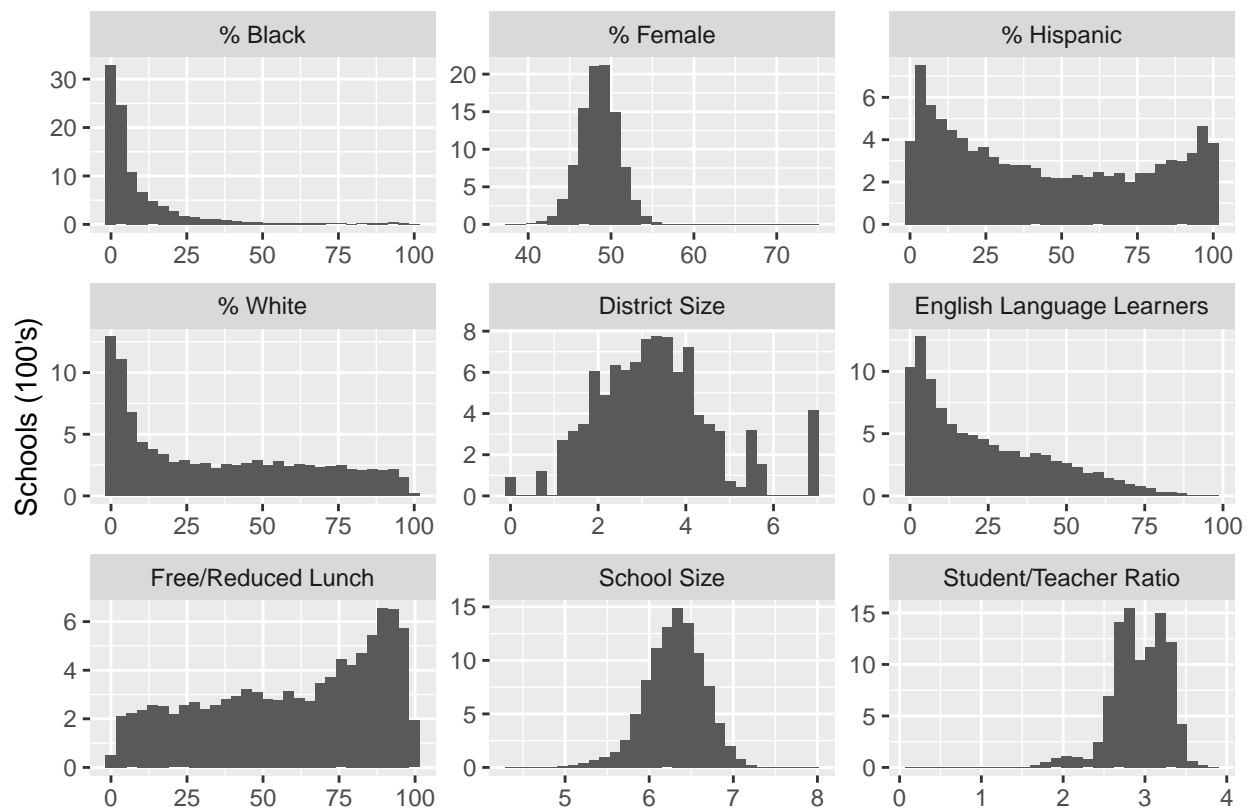
Organize Objects

Data Summary

Covariate Statistics

Continuous variable distributions

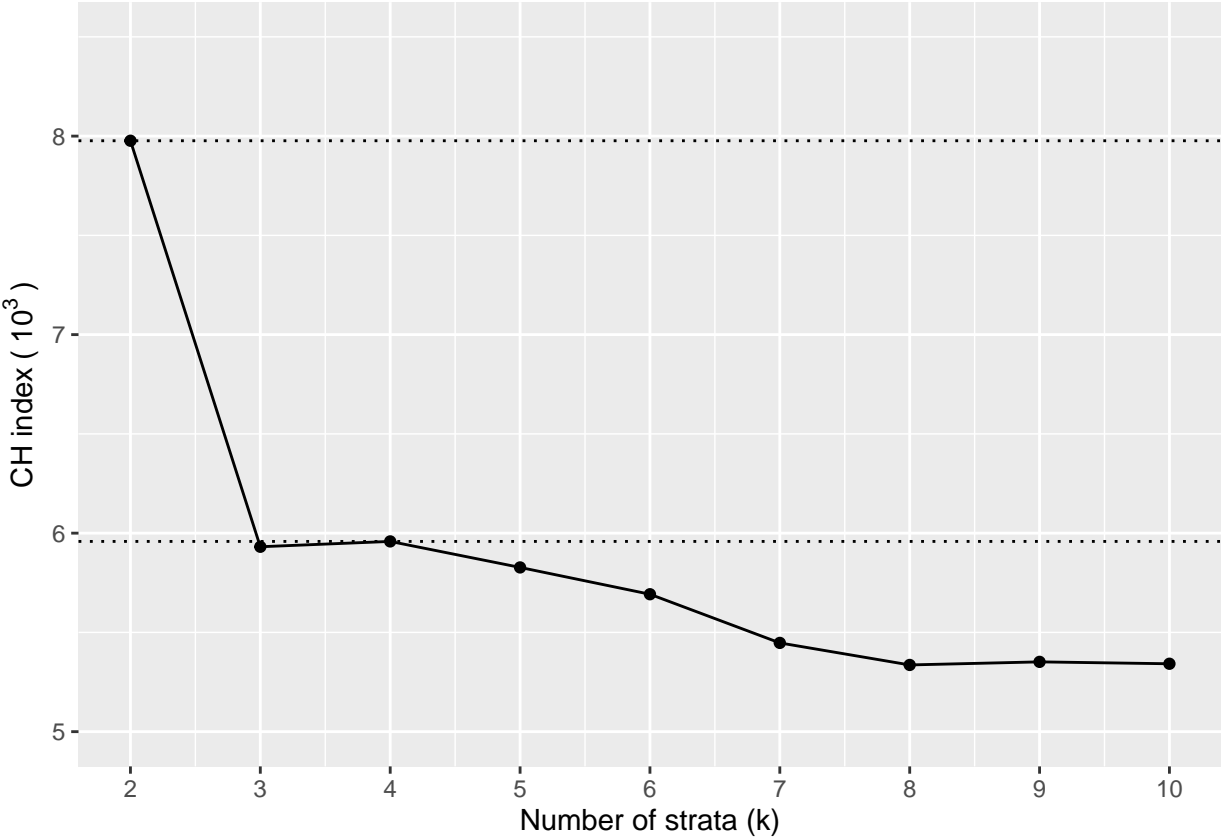
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

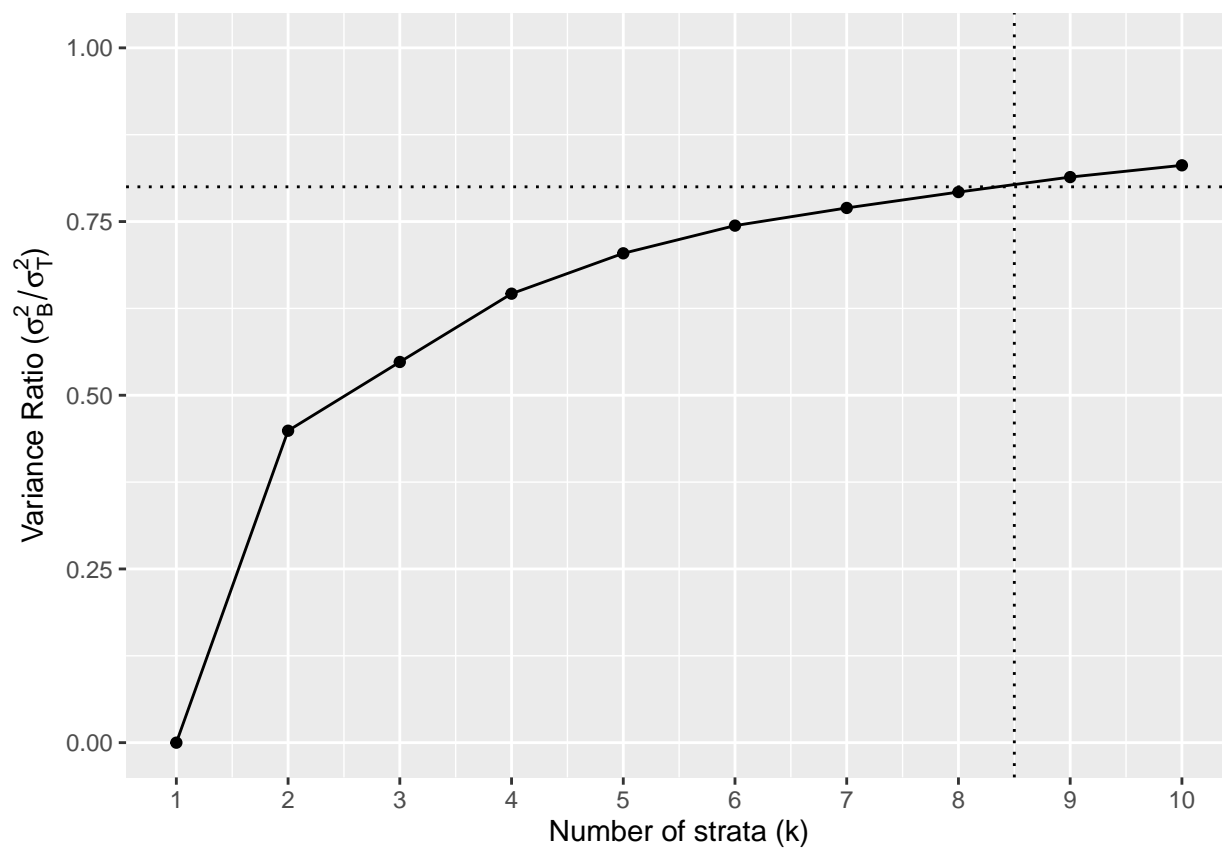


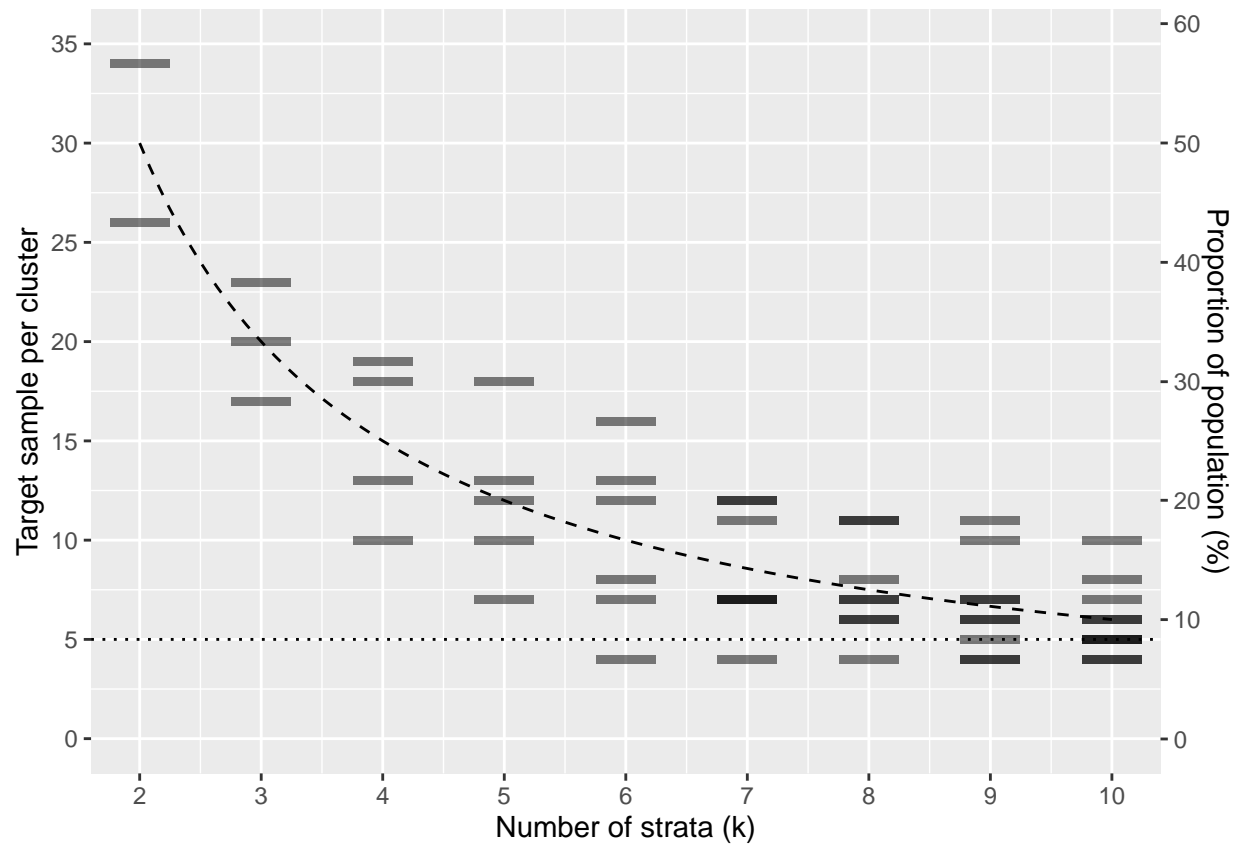
Methods Summary

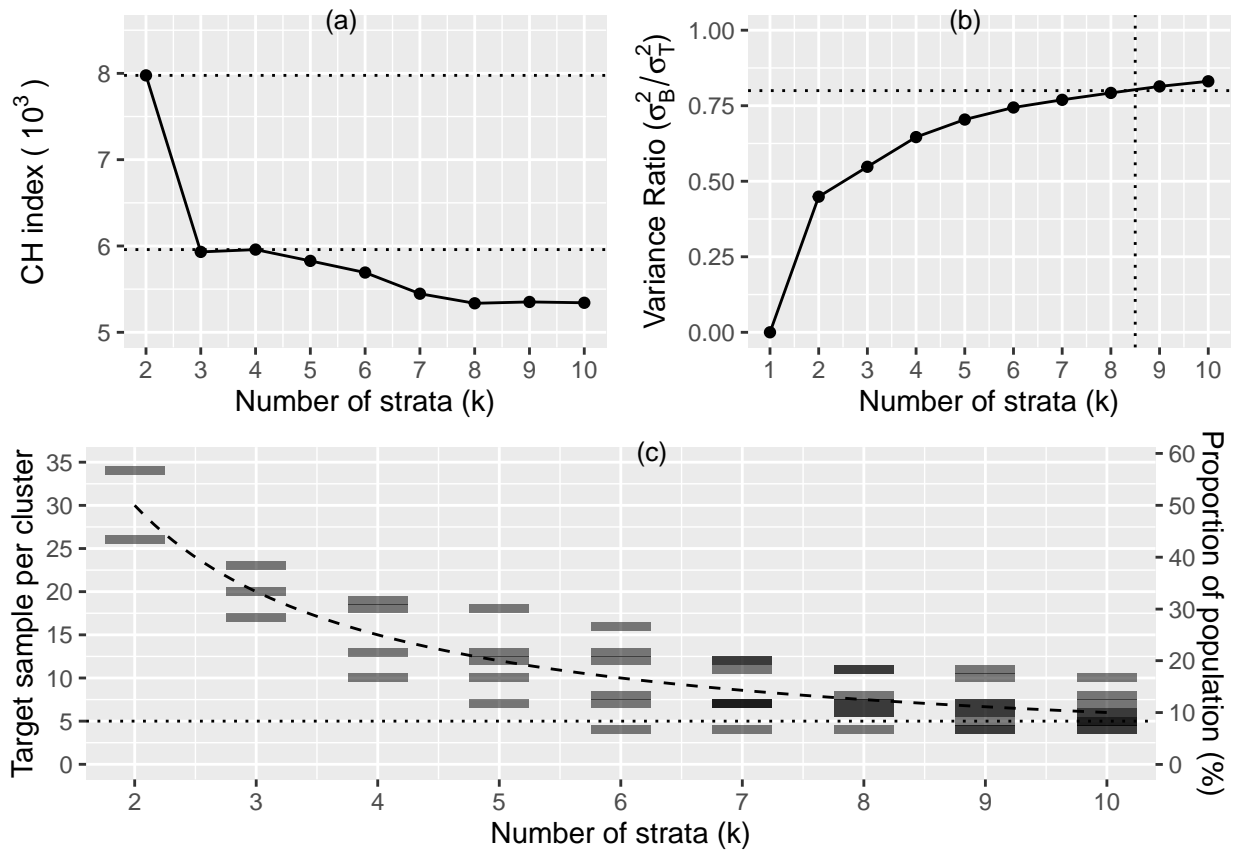
Cluster Analysis

Selecting k.

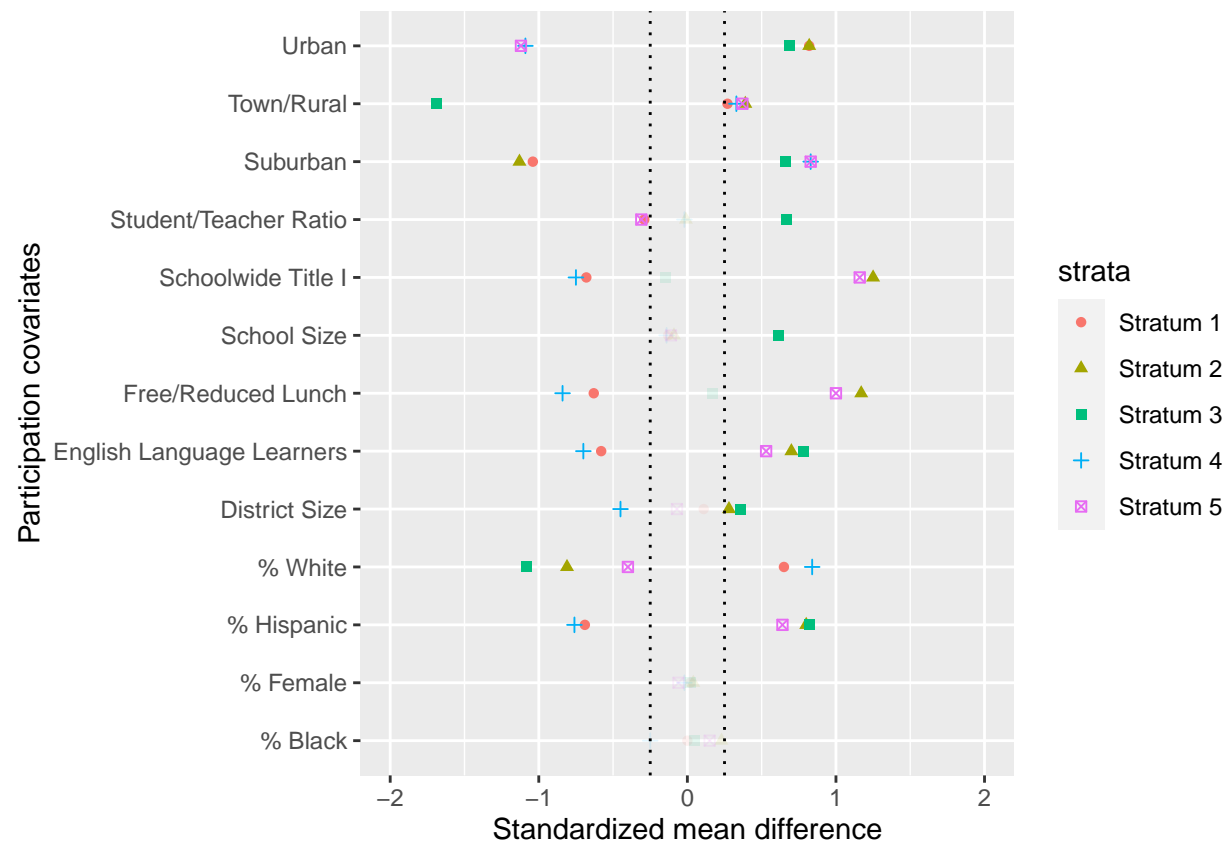








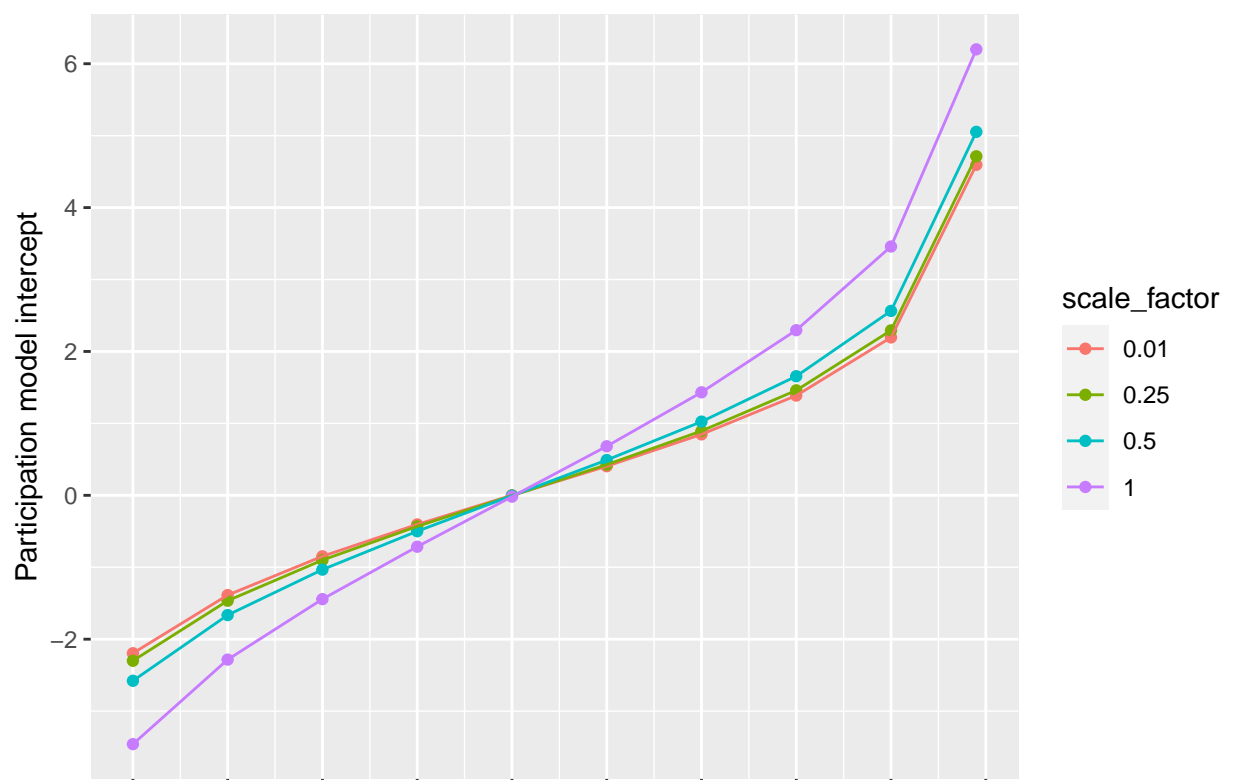
Warning: Using alpha for a discrete variable is not advised.



Sub	Category	Type	Variables	log_odds
Status	School Data	Prop	Schoolwide Title I	0.019
Enrollment	School Data	Mean	School Size	0.374
Status	Student Data	Mean	Free/Reduced Lunch	0.081
Urbanicity	School Data	Prop	Urban	0.433
Urbanicity	School Data	Prop	Suburban	0.007
Urbanicity	School Data	Prop	Town/Rural	-0.403
Ethnicity	Student Data	Mean	% White	-0.538
Ethnicity	Student Data	Mean	% Black	0.291
Ethnicity	Student Data	Mean	% Hispanic	0.395
Gender	Student Data	Mean	% Female	-0.019
Enrollment	School Data	Mean	Student/Teacher Ratio	-0.101
District	School Data	Mean	District Size	0.520
Status	Student Data	Mean	English Language Learners	0.412

Variation explained by the strata

Participation Generating Model



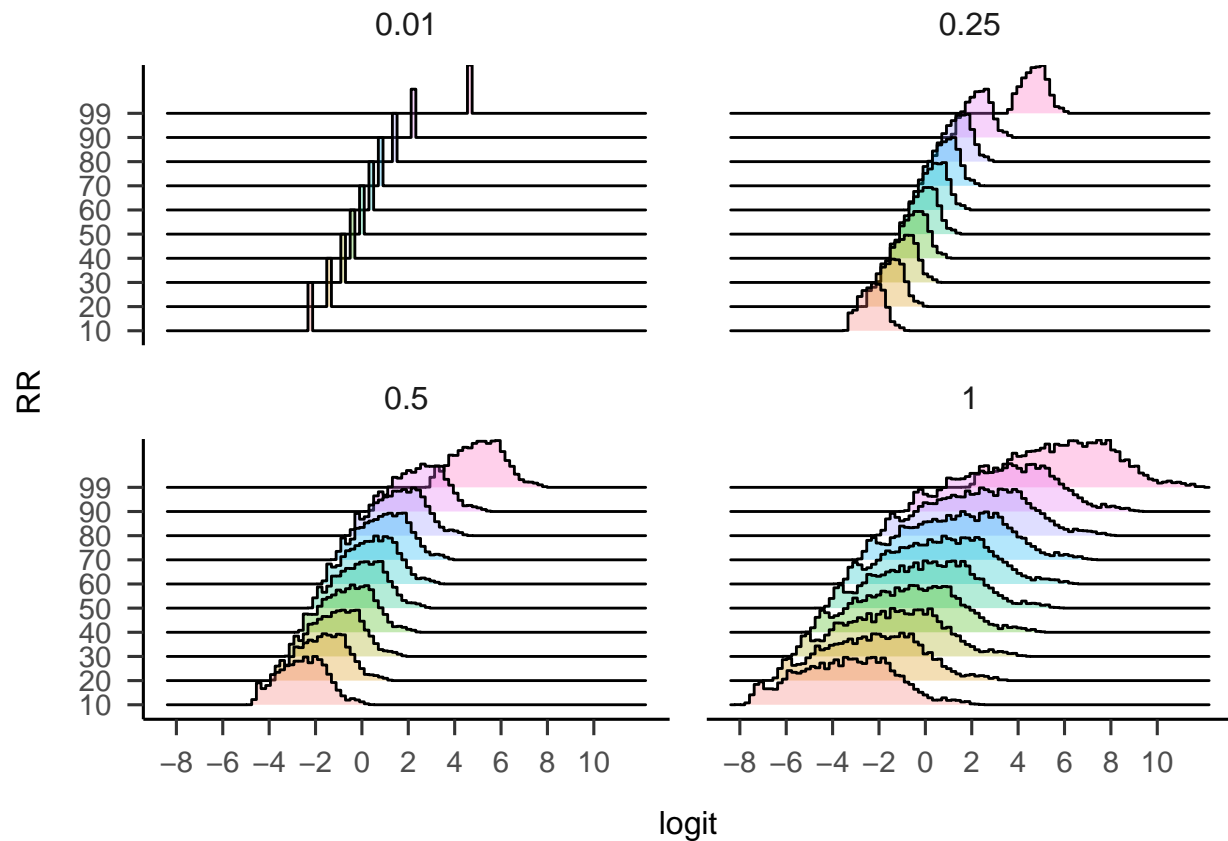


Figure 1. Distributions of Participation Logits

Results

Generalizability

B Index.

```
## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(variable)' instead of 'variable' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
## Note: Using an external vector in selections is ambiguous.
## i Use 'all_of(observations)' instead of 'observations' to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```

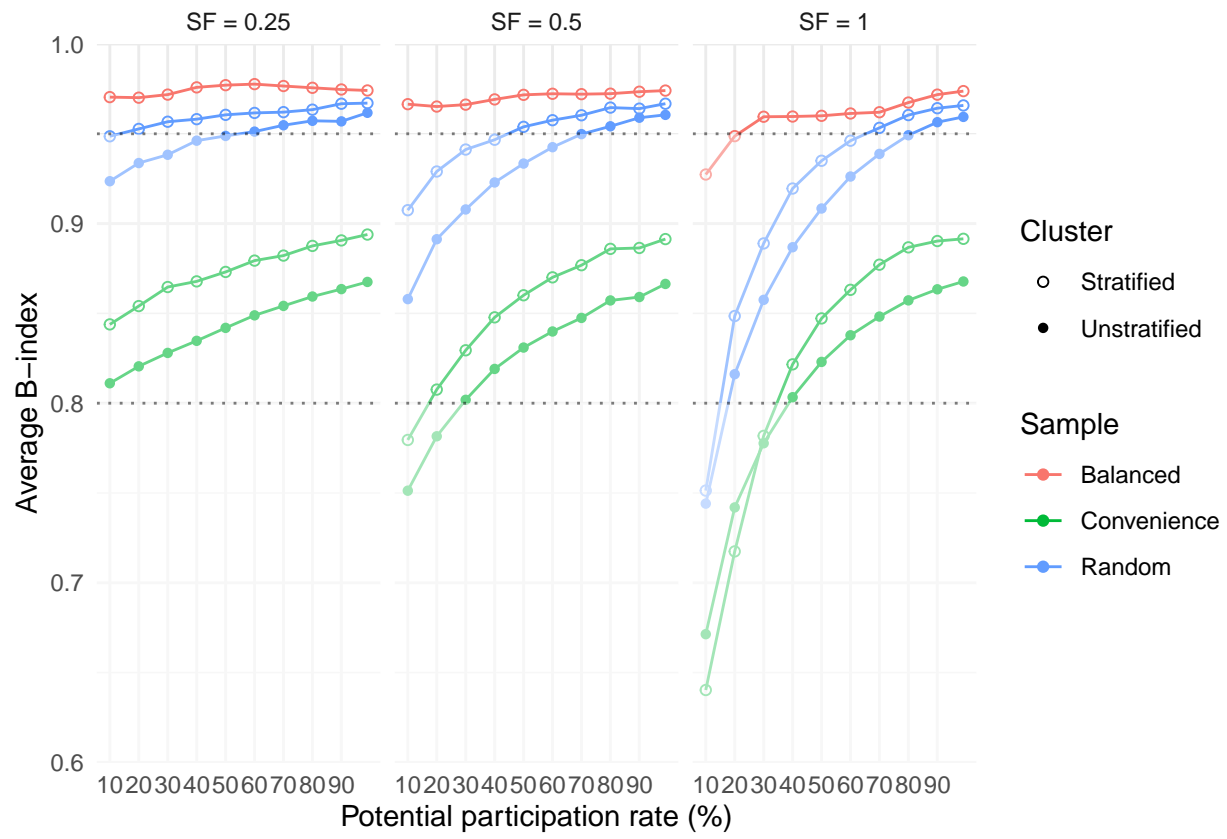
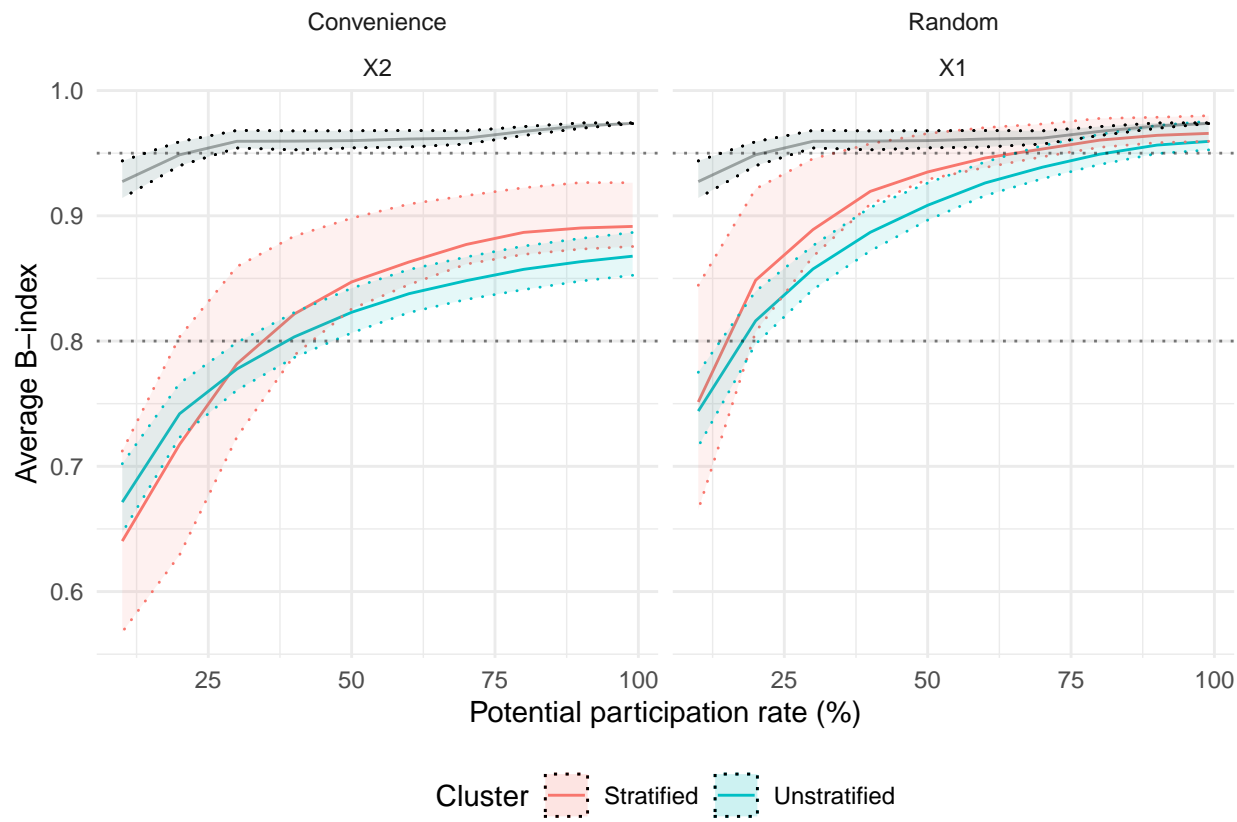
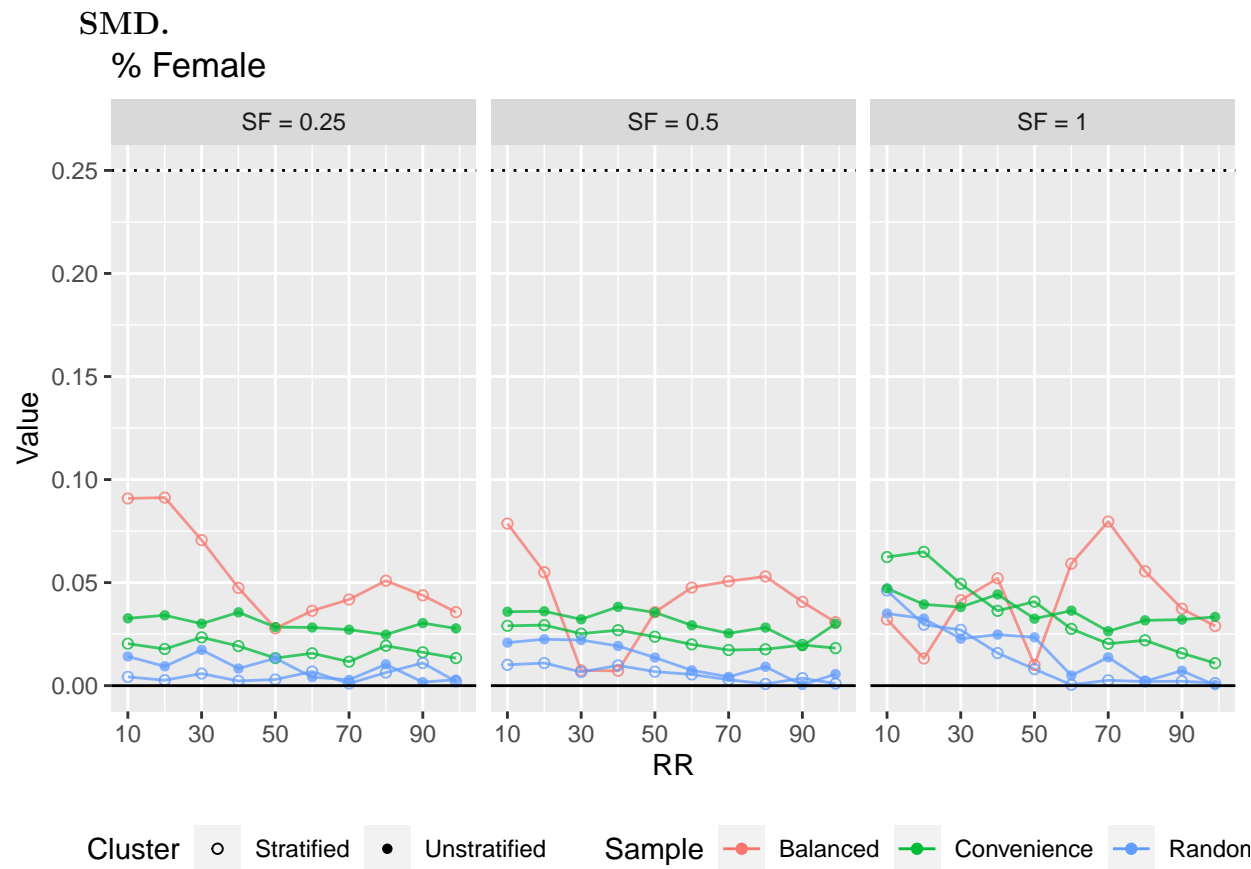
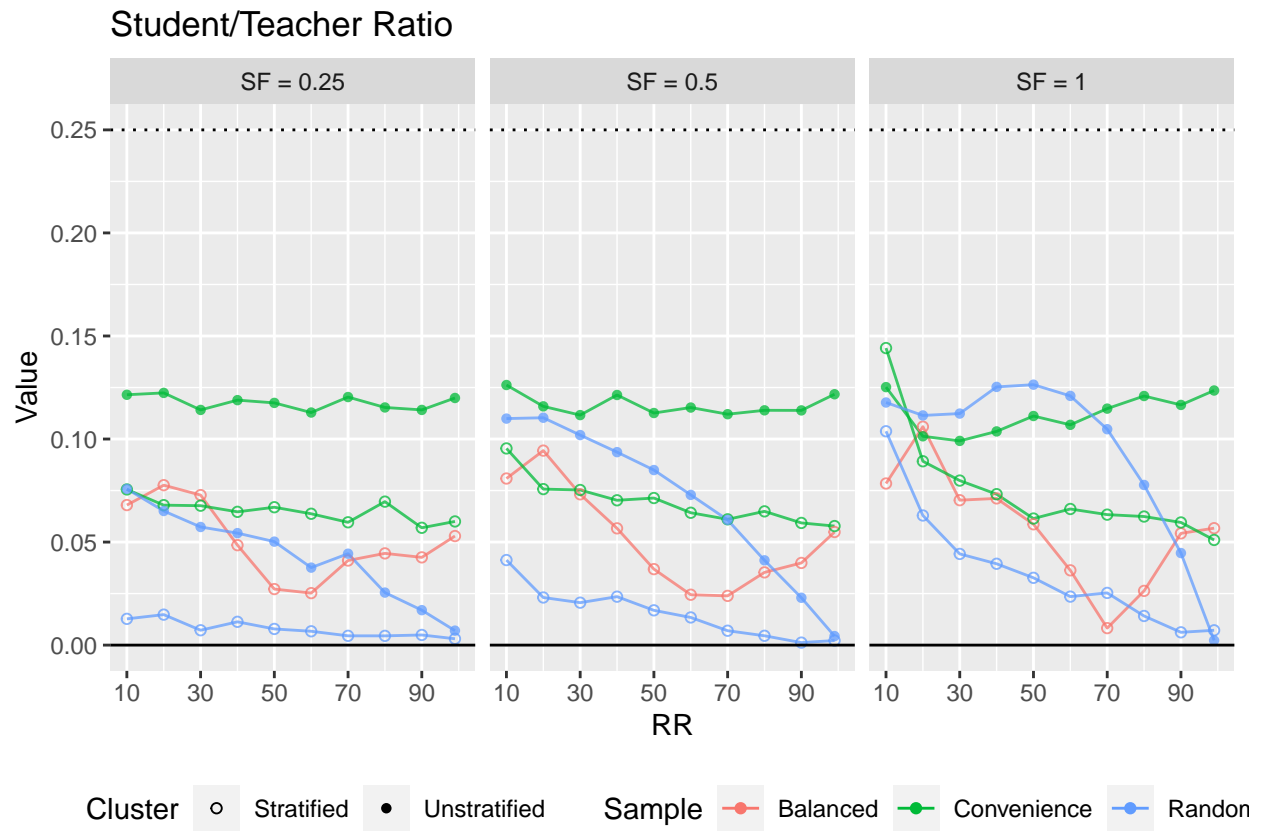


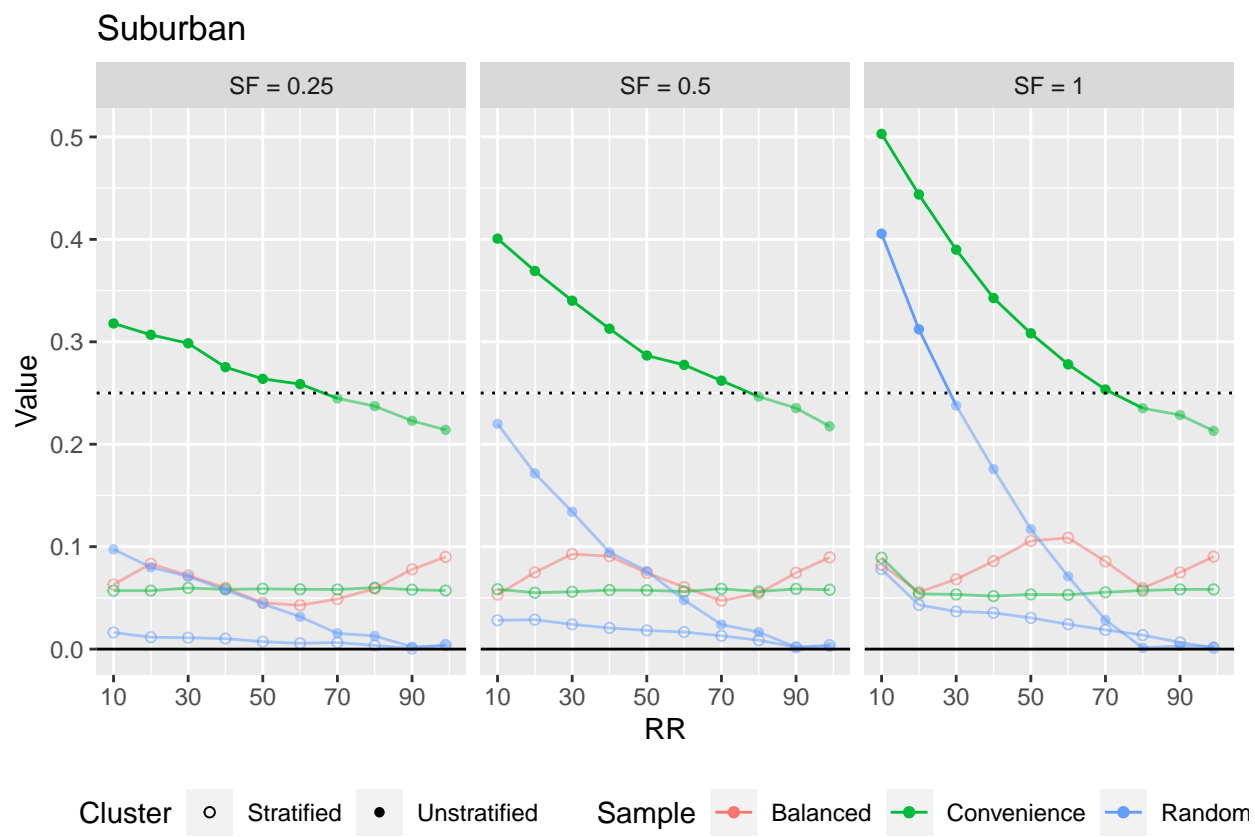
Figure 2. Average B-Index

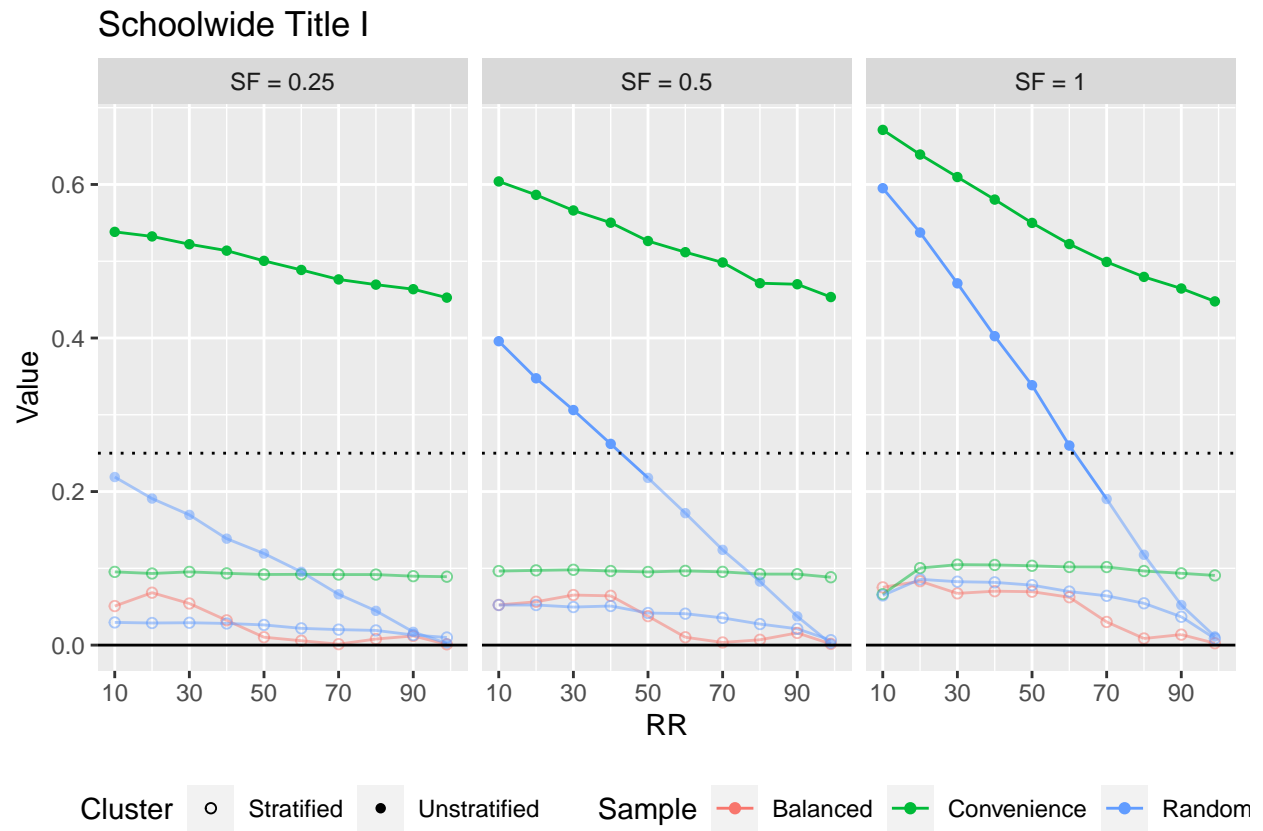
This message is displayed once per session.

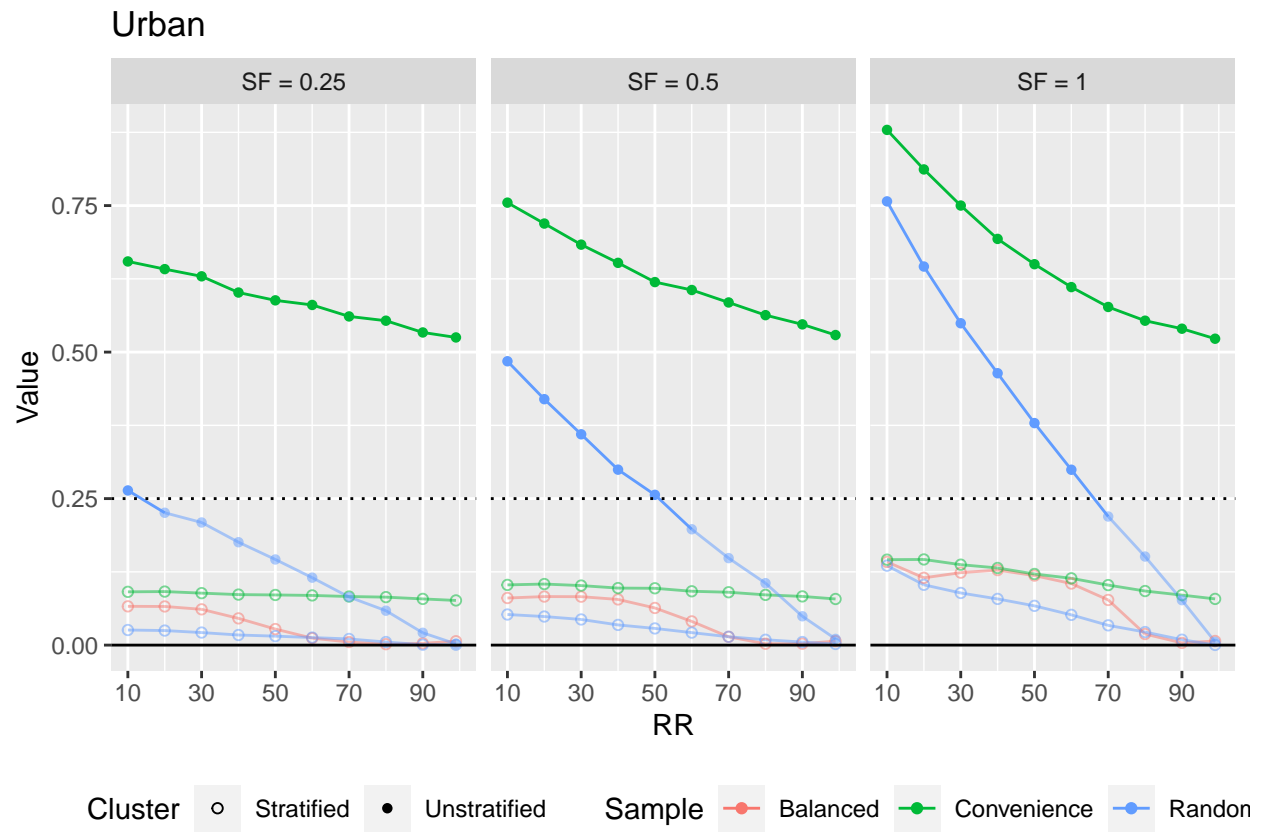


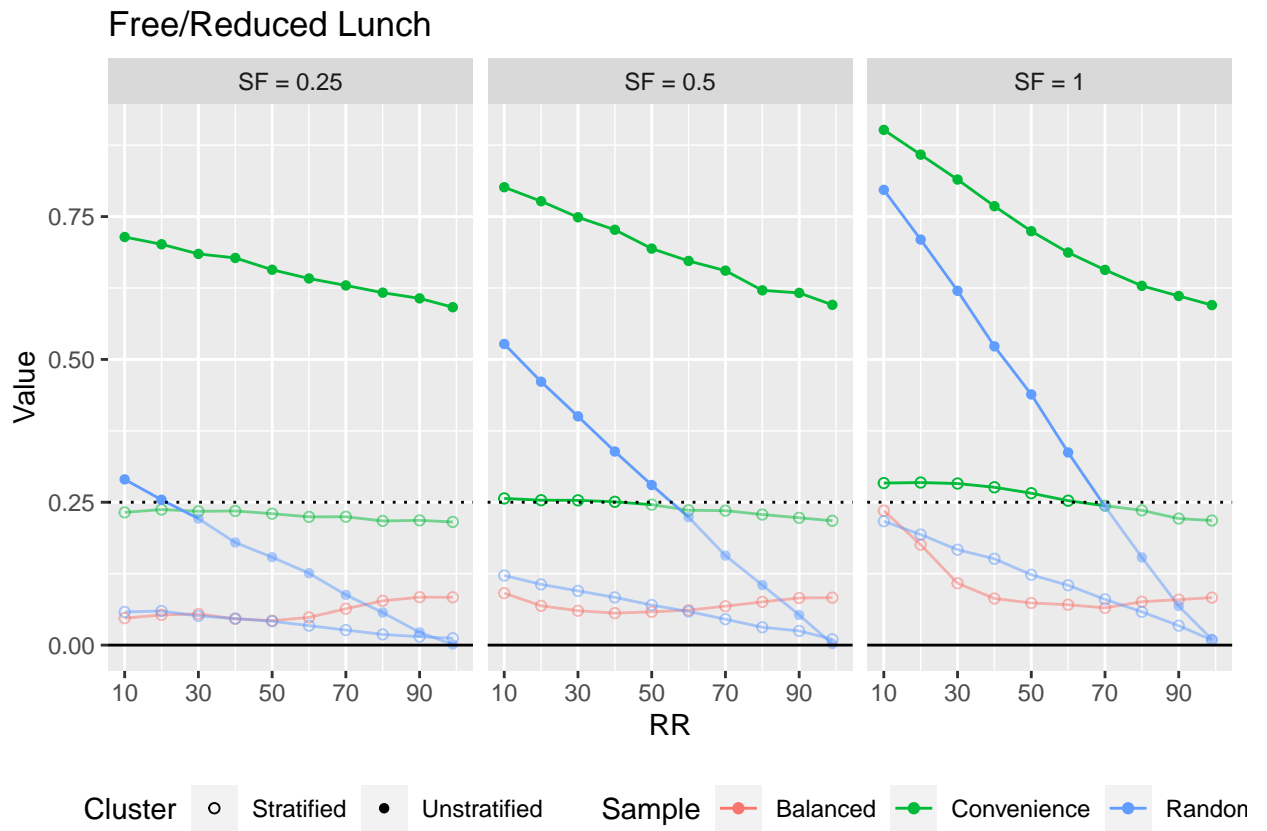


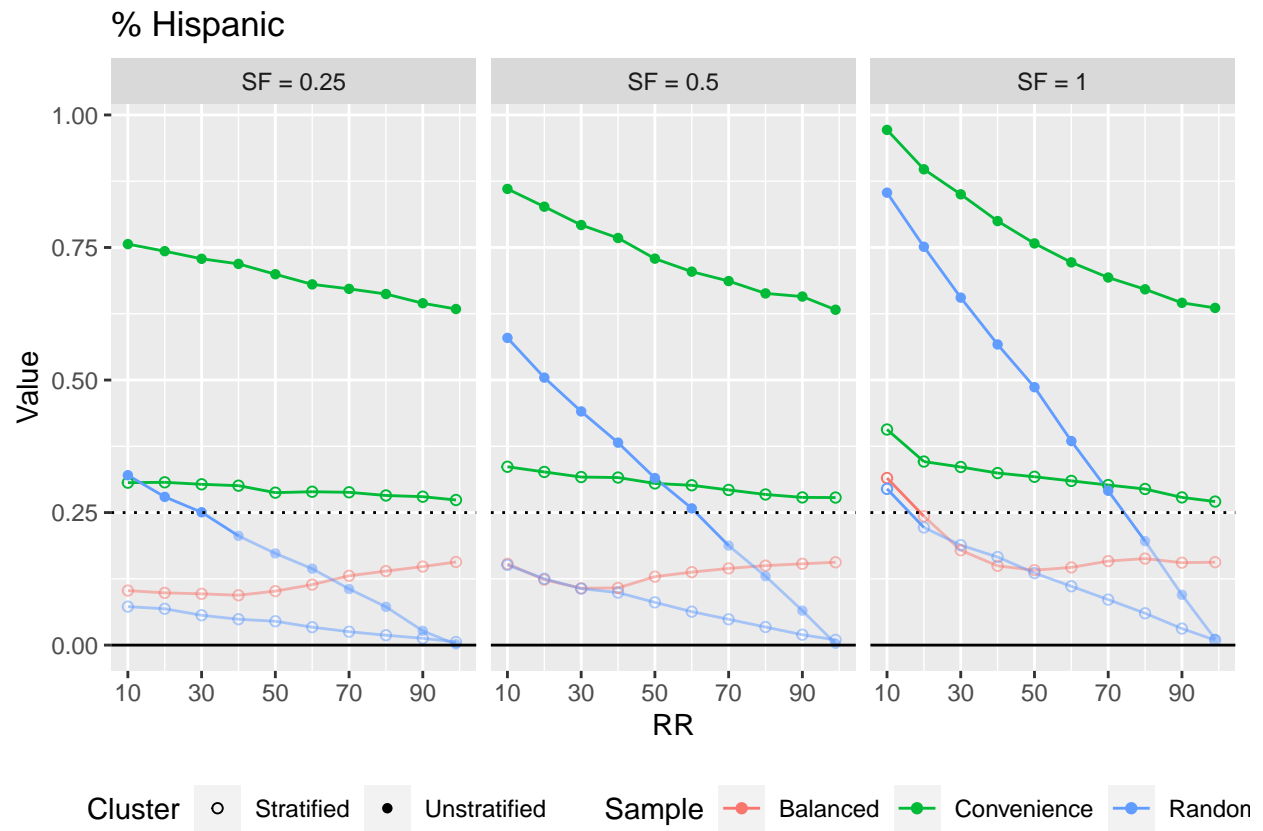


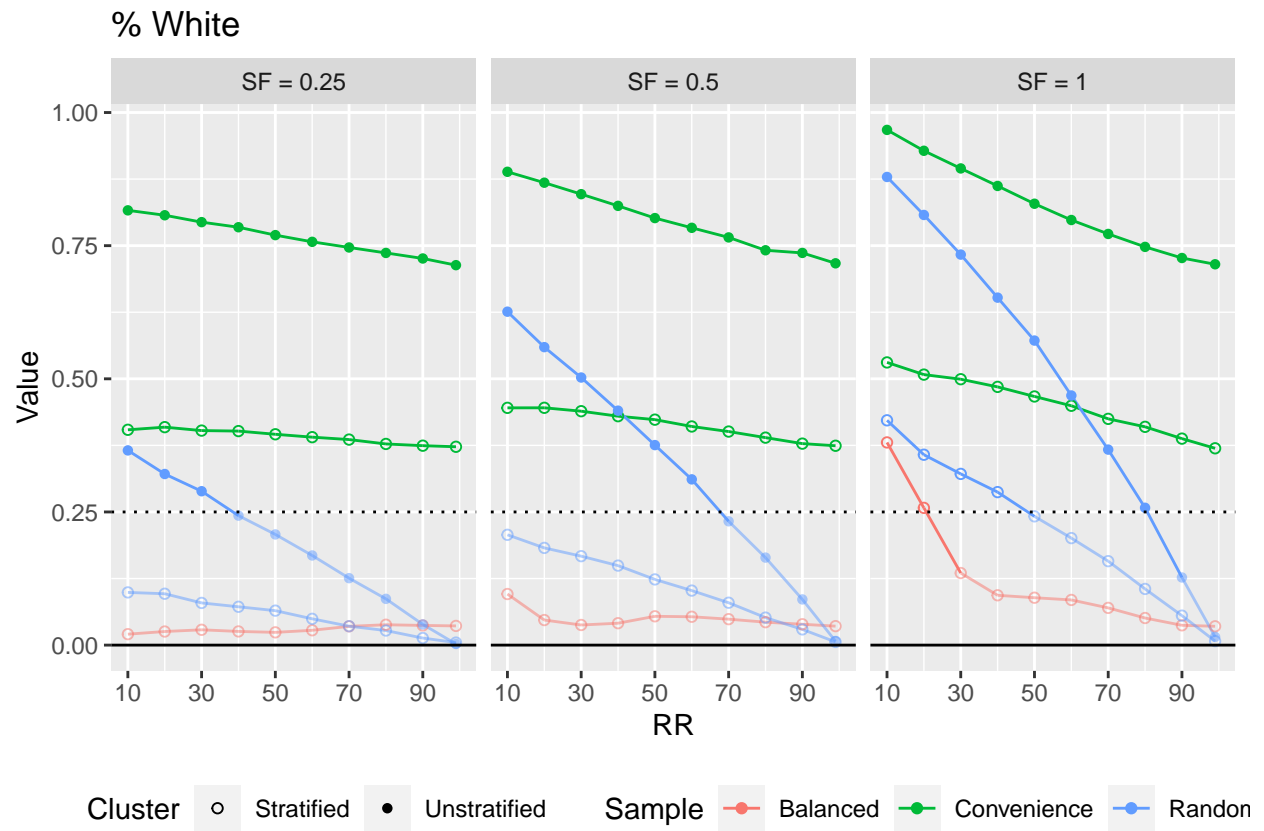


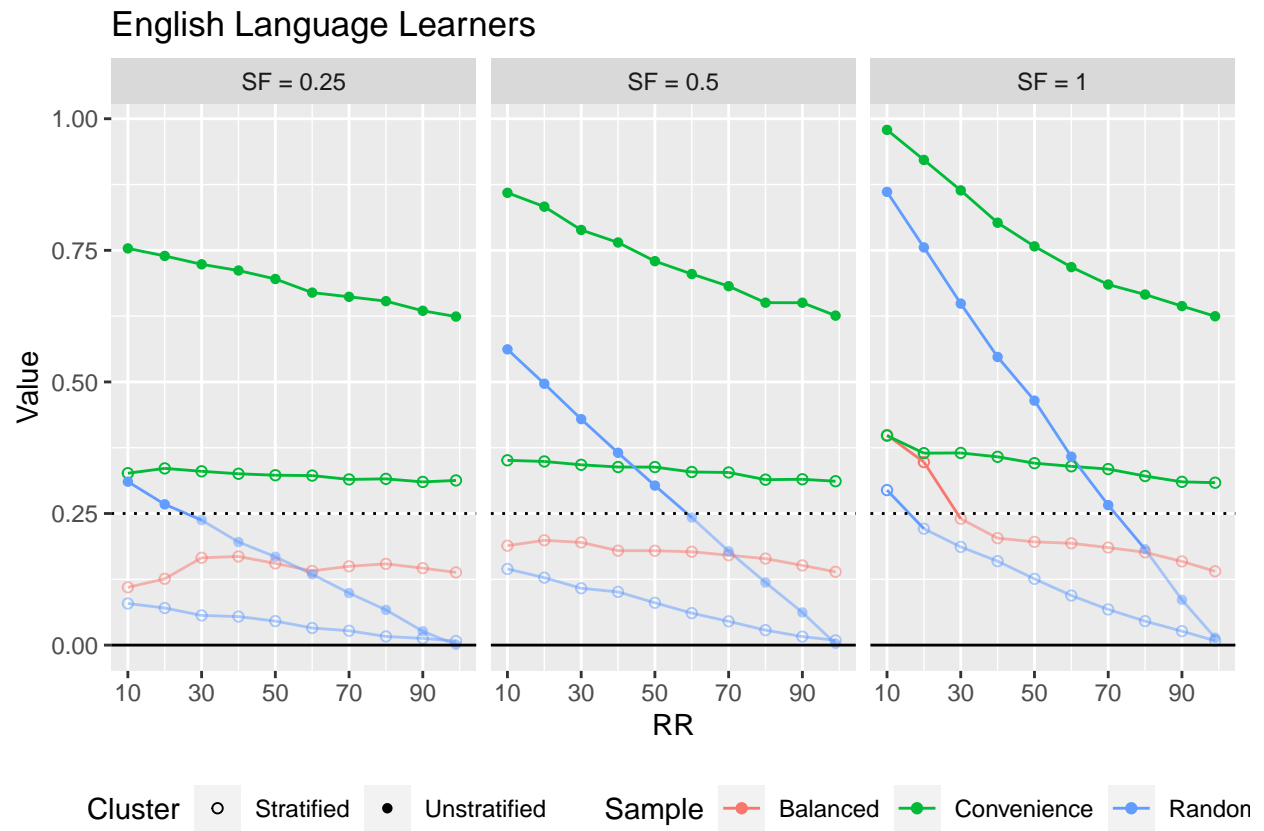


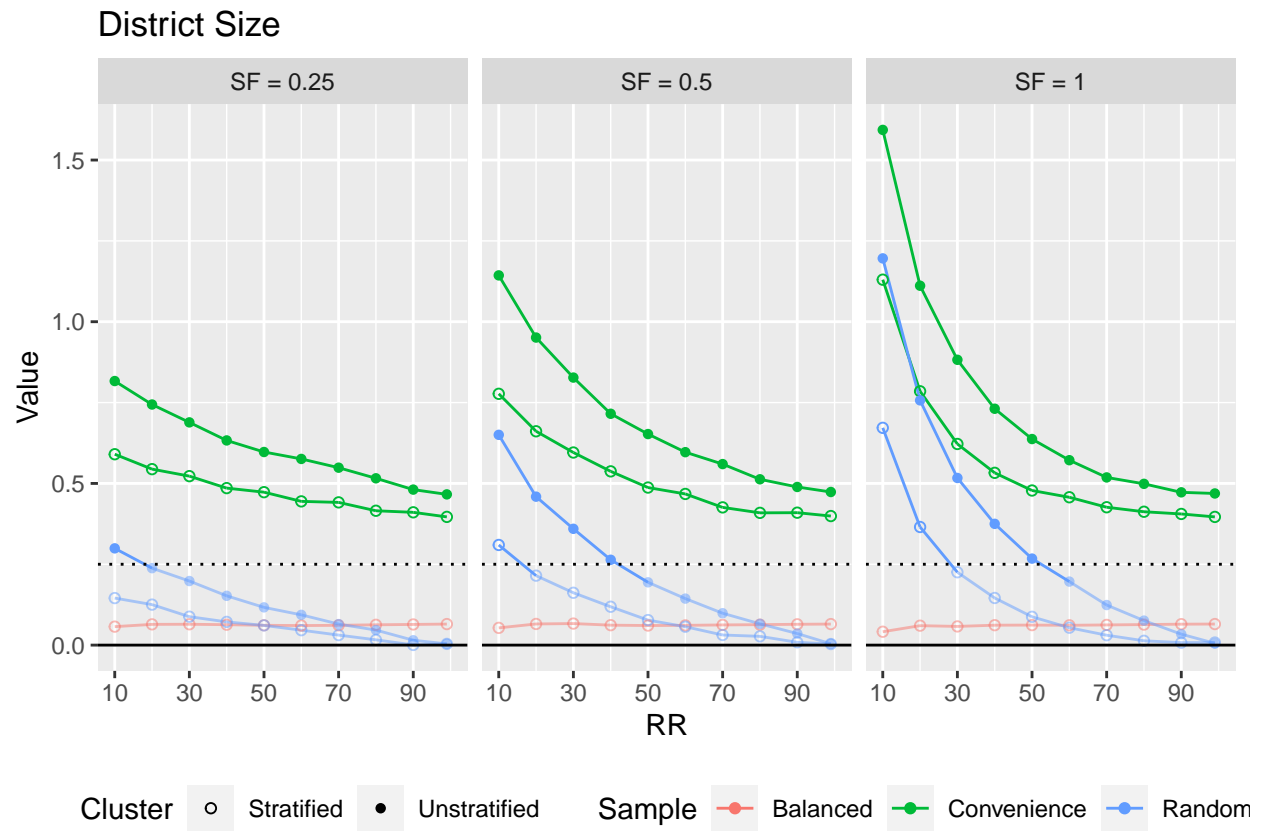


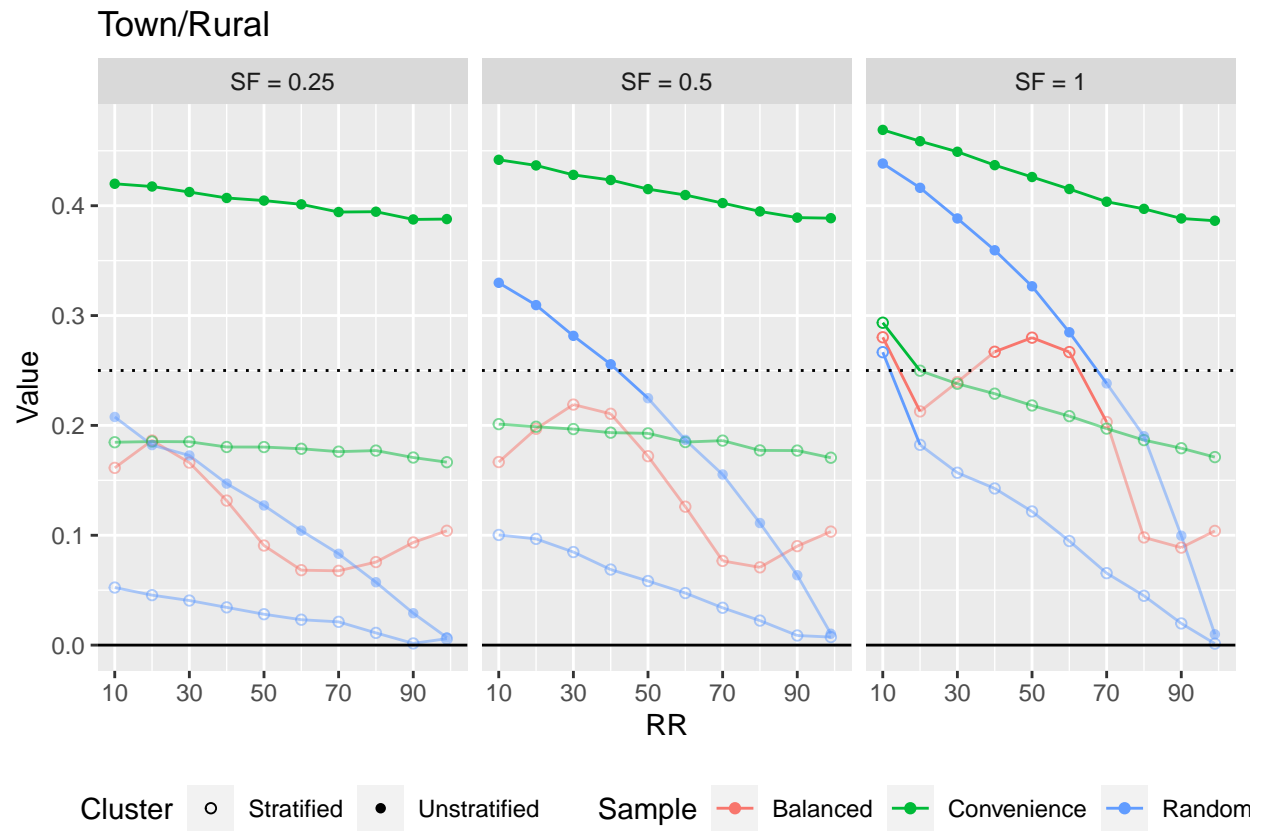


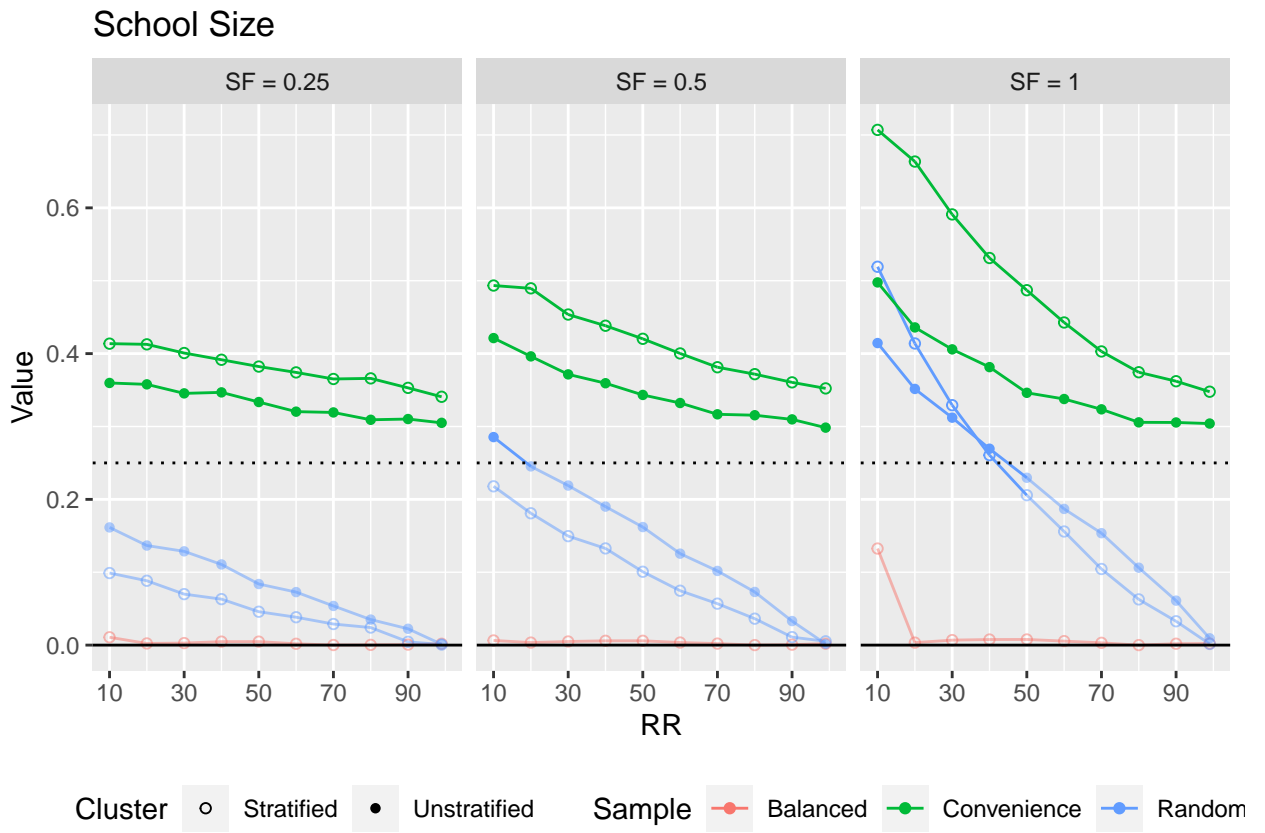


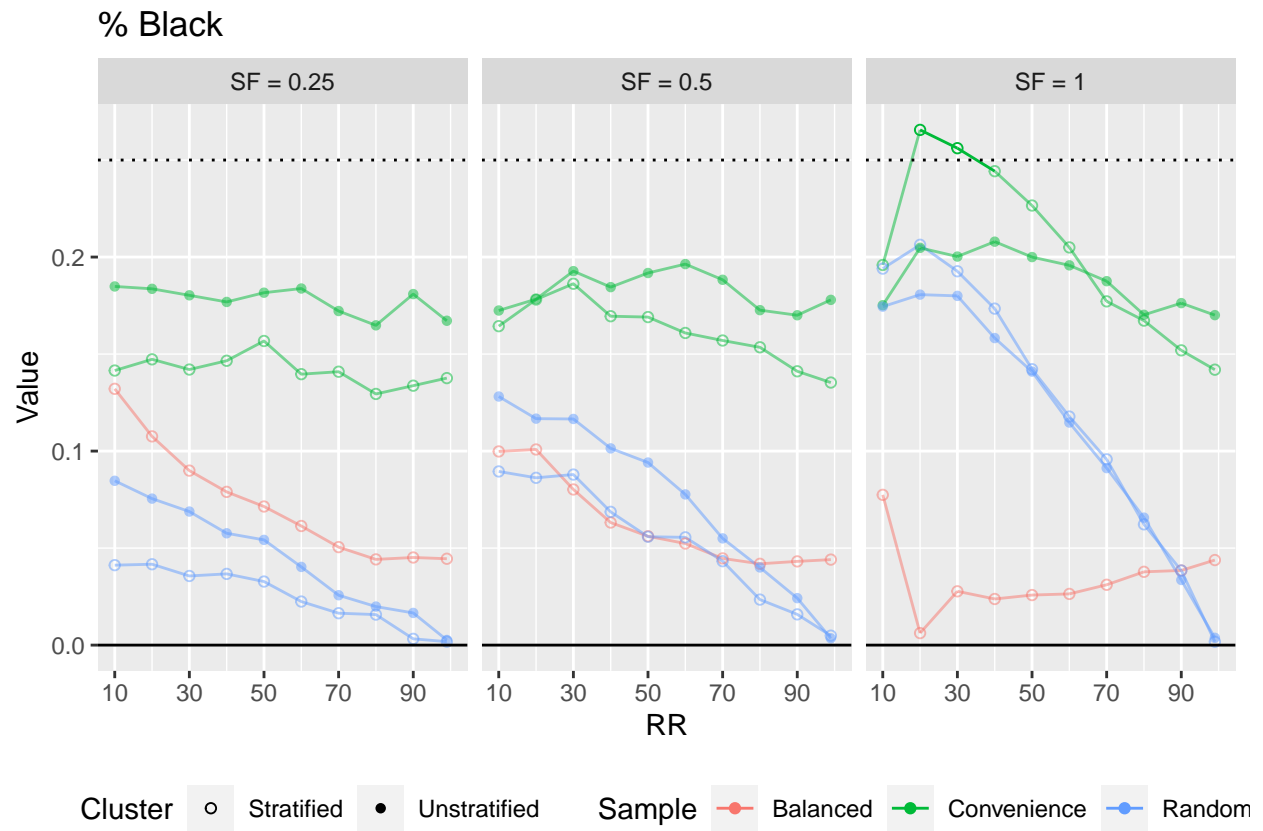




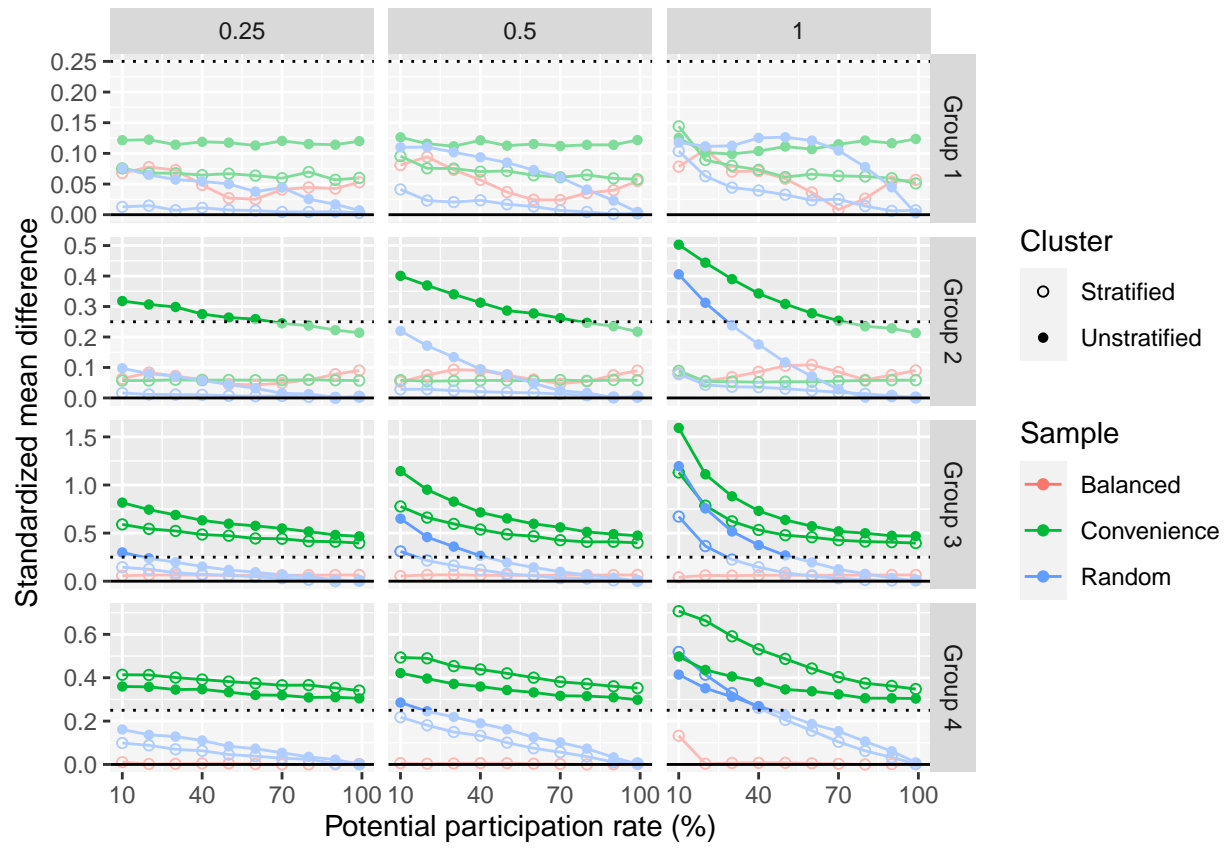


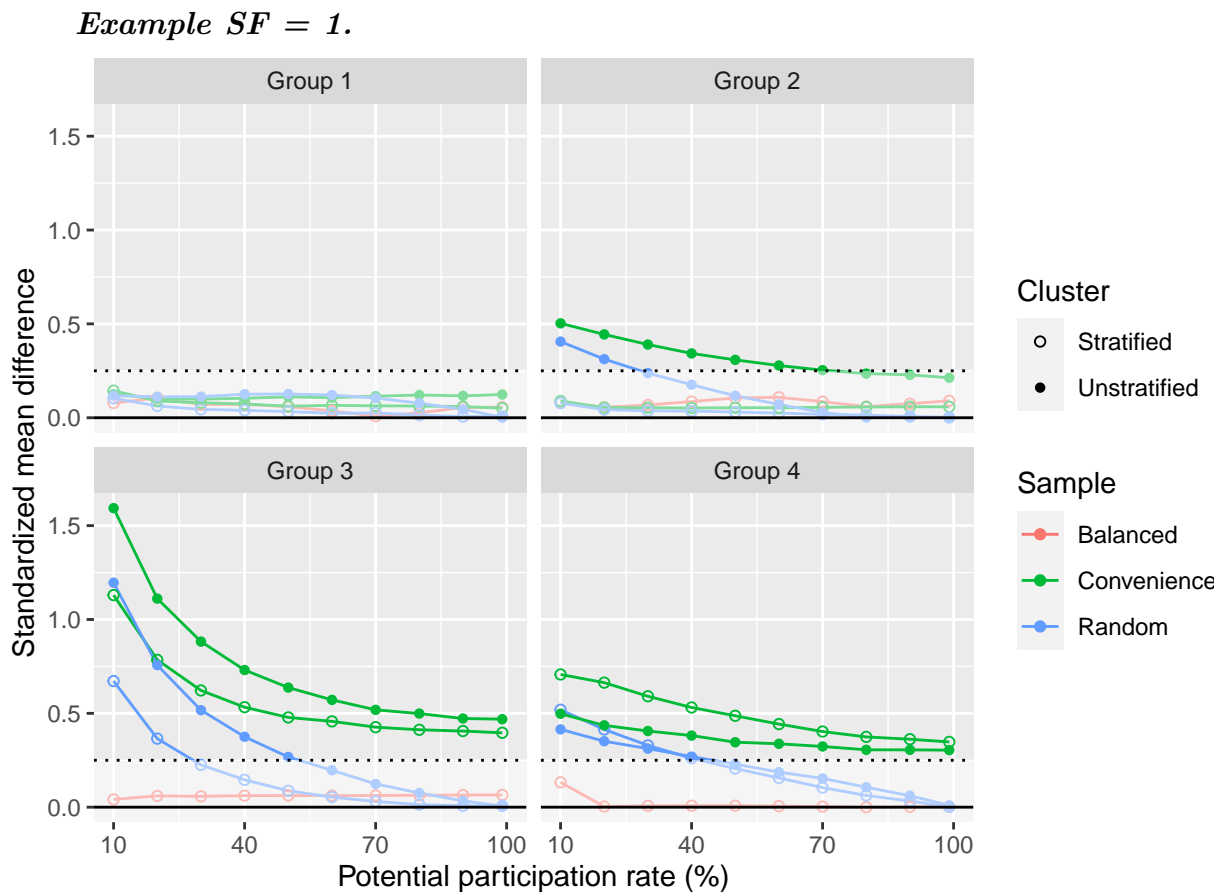






SMD Groups.

Example all SF.



Auto-Grouping.

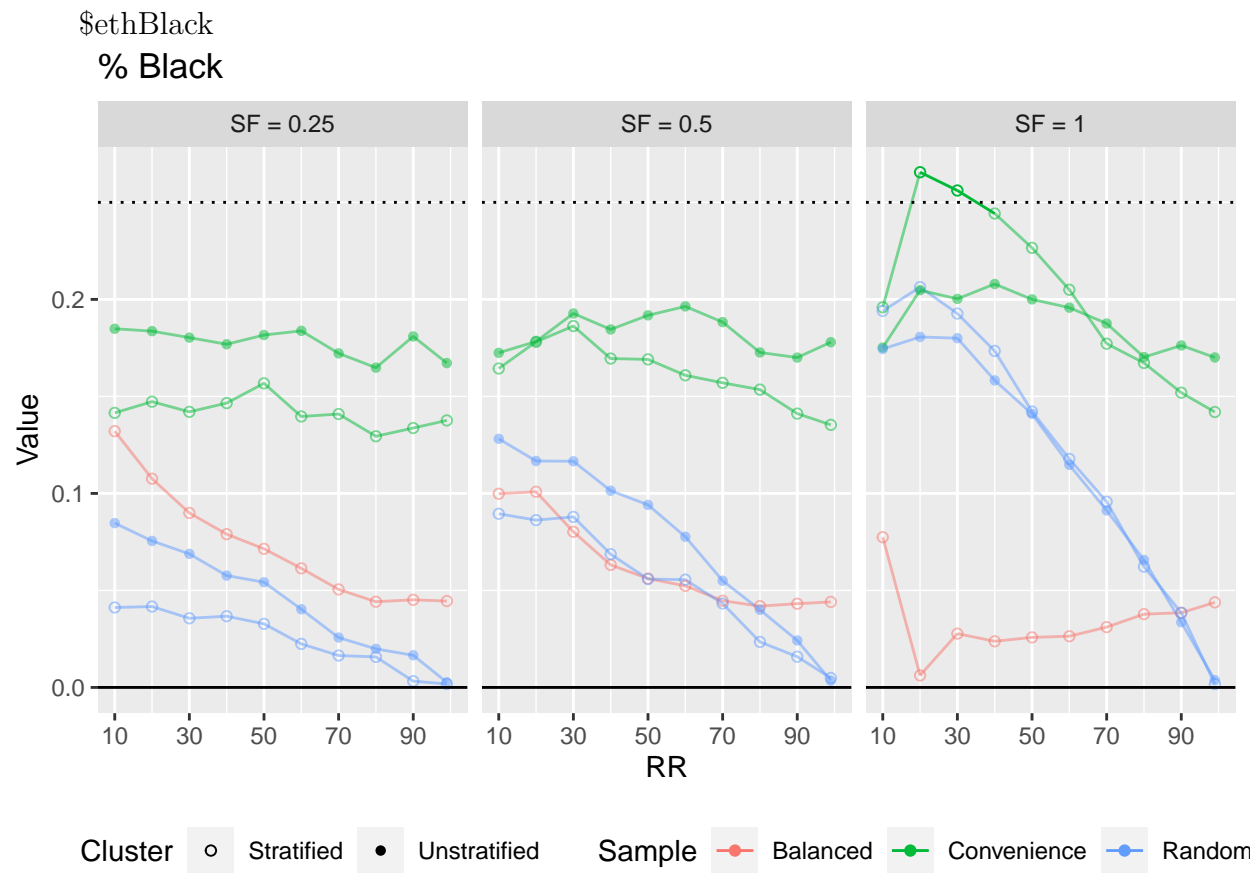
Group 1.

A tibble: 8 x 3

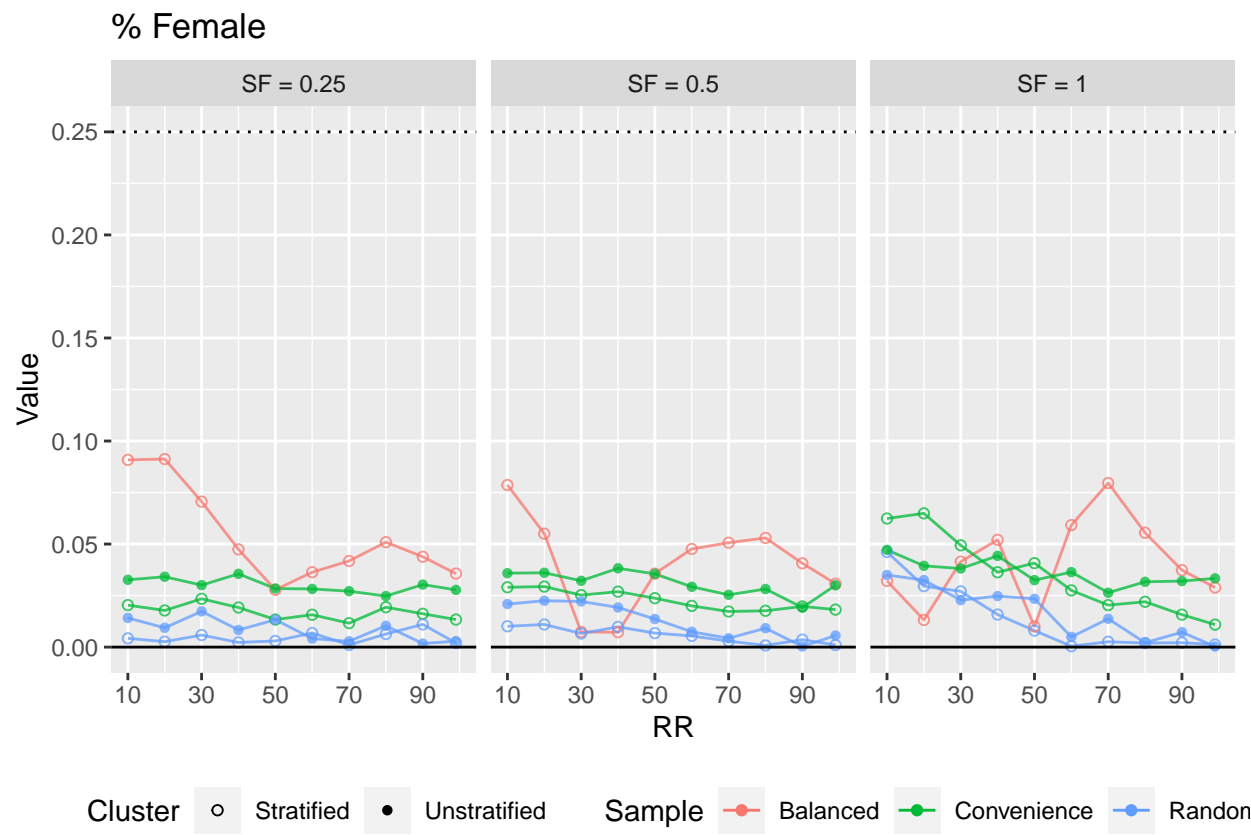
Groups: SB [3]

SB vnames Group

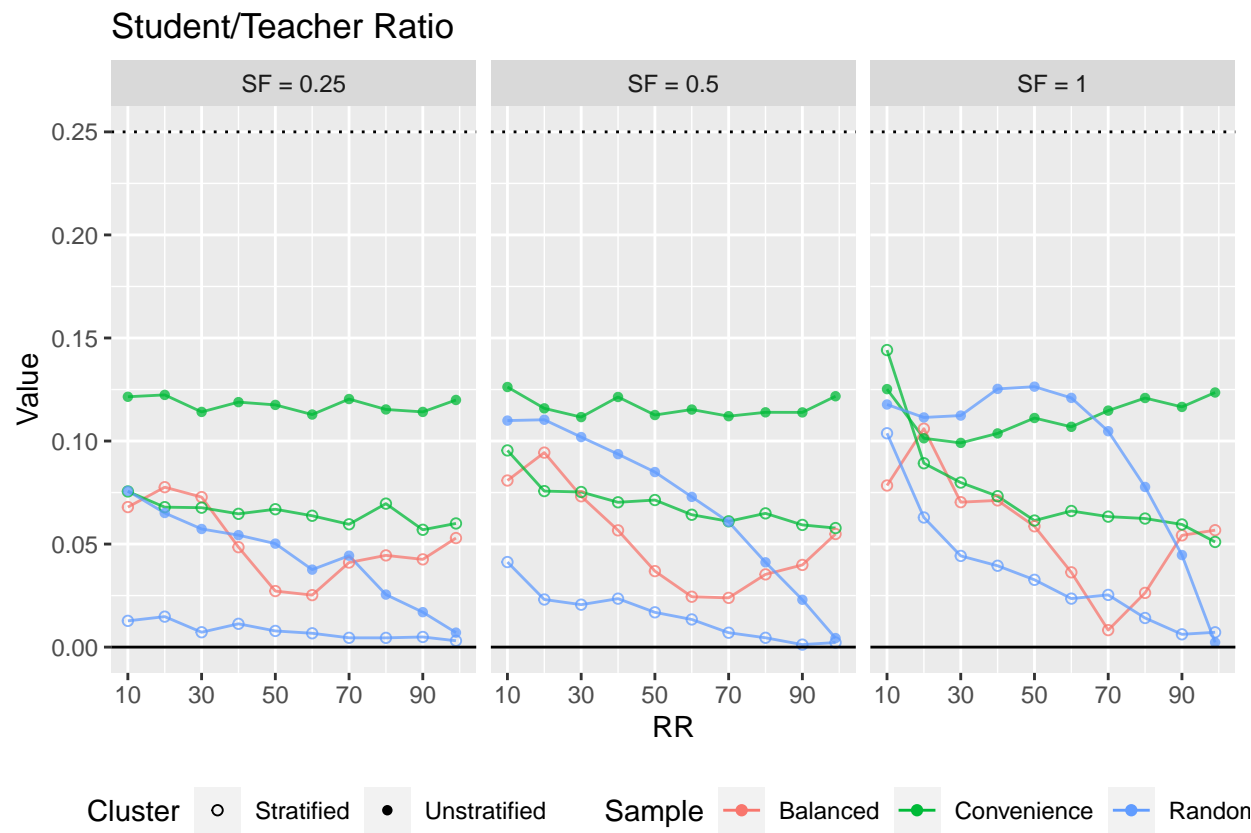
1 0.25 ethBlack Group 1 2 0.5 ethBlack Group 1 3 0.25 pFem Group 1 4 0.5 pFem
Group 1 5 1 pFem Group 1 6 0.25 ST.ratio Group 1 7 0.5 ST.ratio Group 1 8 1 ST.ratio
Group 1



\$pFem



\$ST.ratio

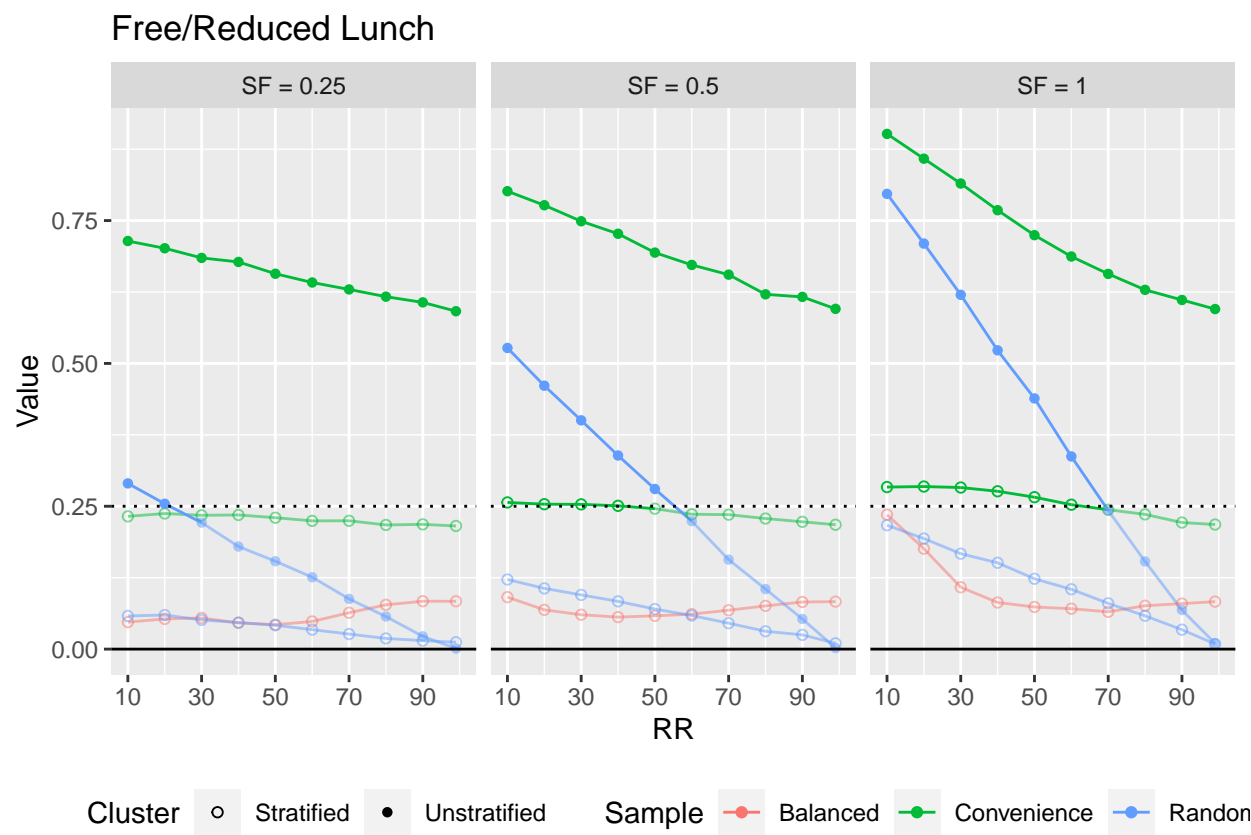


Group 2.

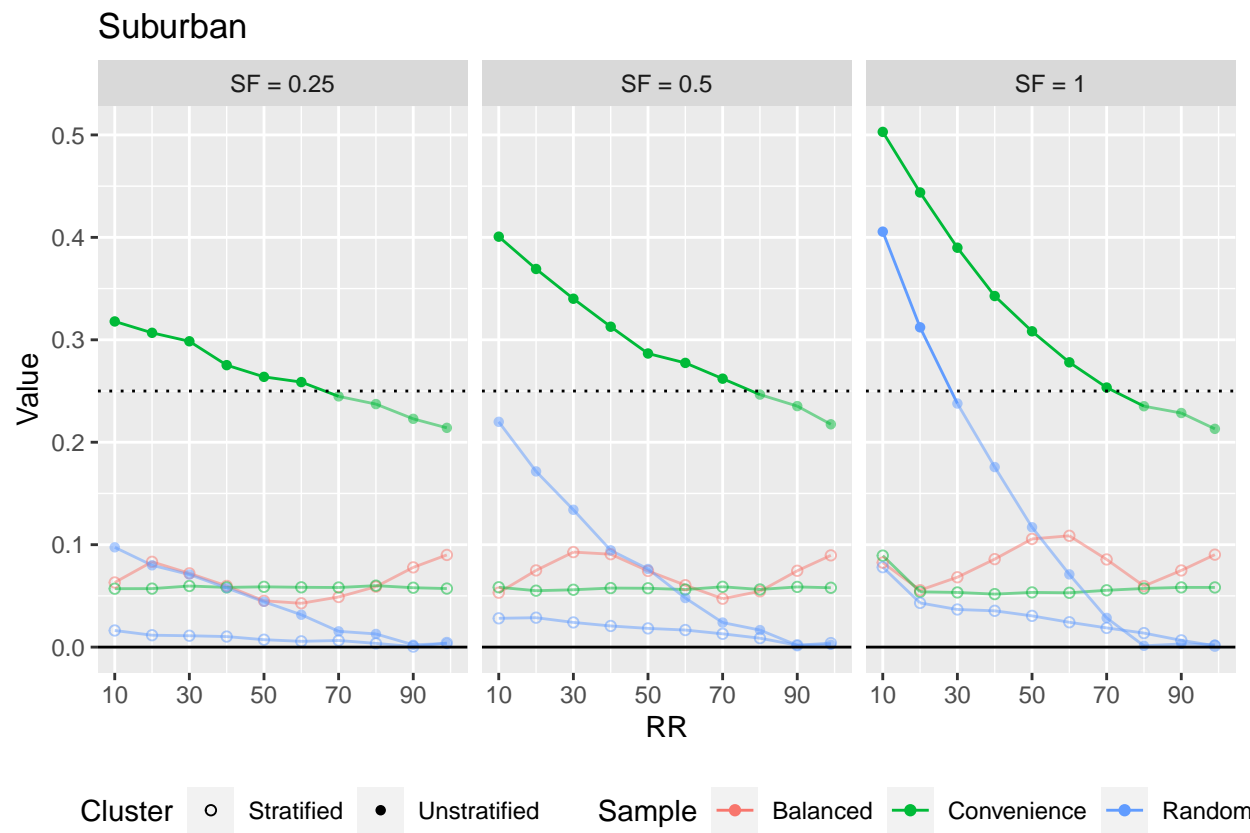
A tibble: 12 x 3

SB	vnames	Group
1	0.25 pTotfrl	Group 2
2	0.25 Suburban	Group 2
3	0.5 Suburban	Group 2
4	1 Suburban	Group 2
5	0.25 T1	Group 2
6	0.5 T1	Group 2
7	1 T1	Group 2
8	0.25 ToRu	Group 2
9	0.5 ToRu	Group 2
10	0.25 Urban	Group 2
11	0.5 Urban	Group 2
12	1 Urban	Group 2

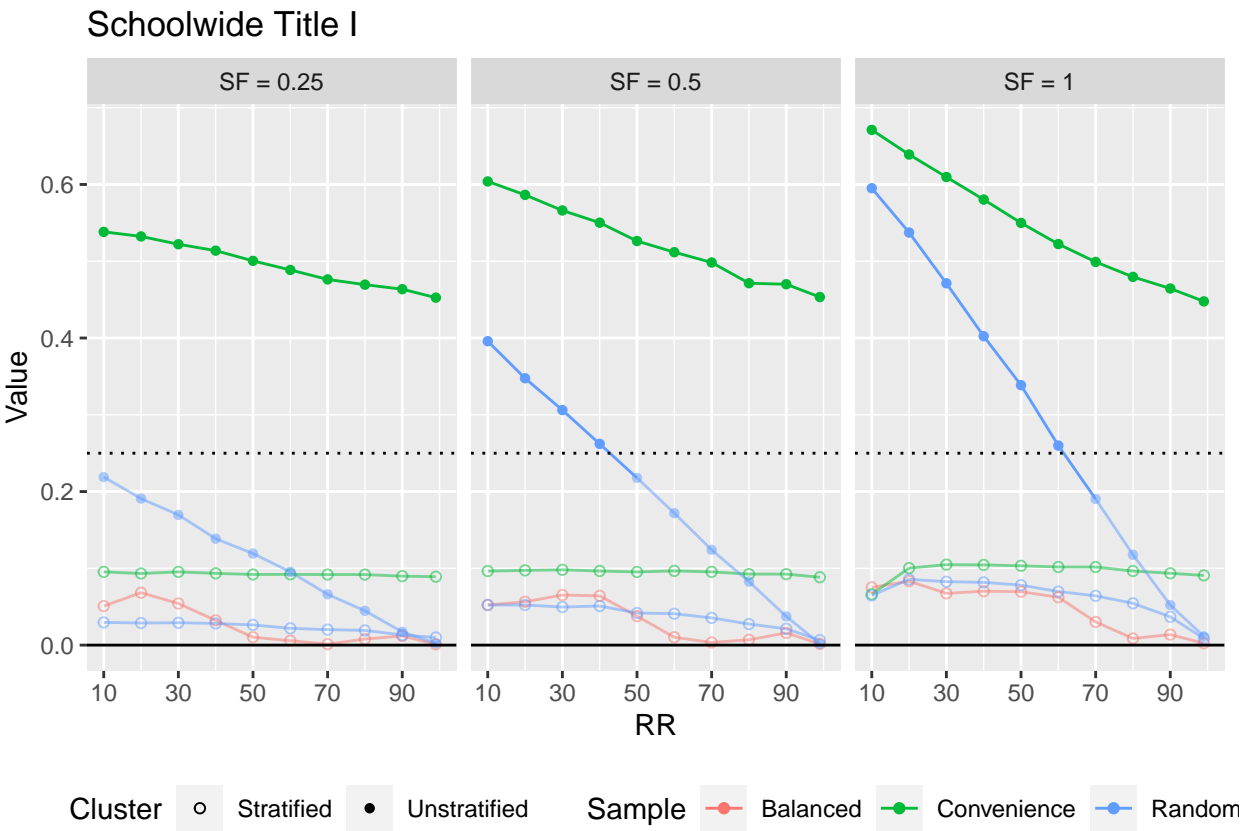
\$pTotfrl



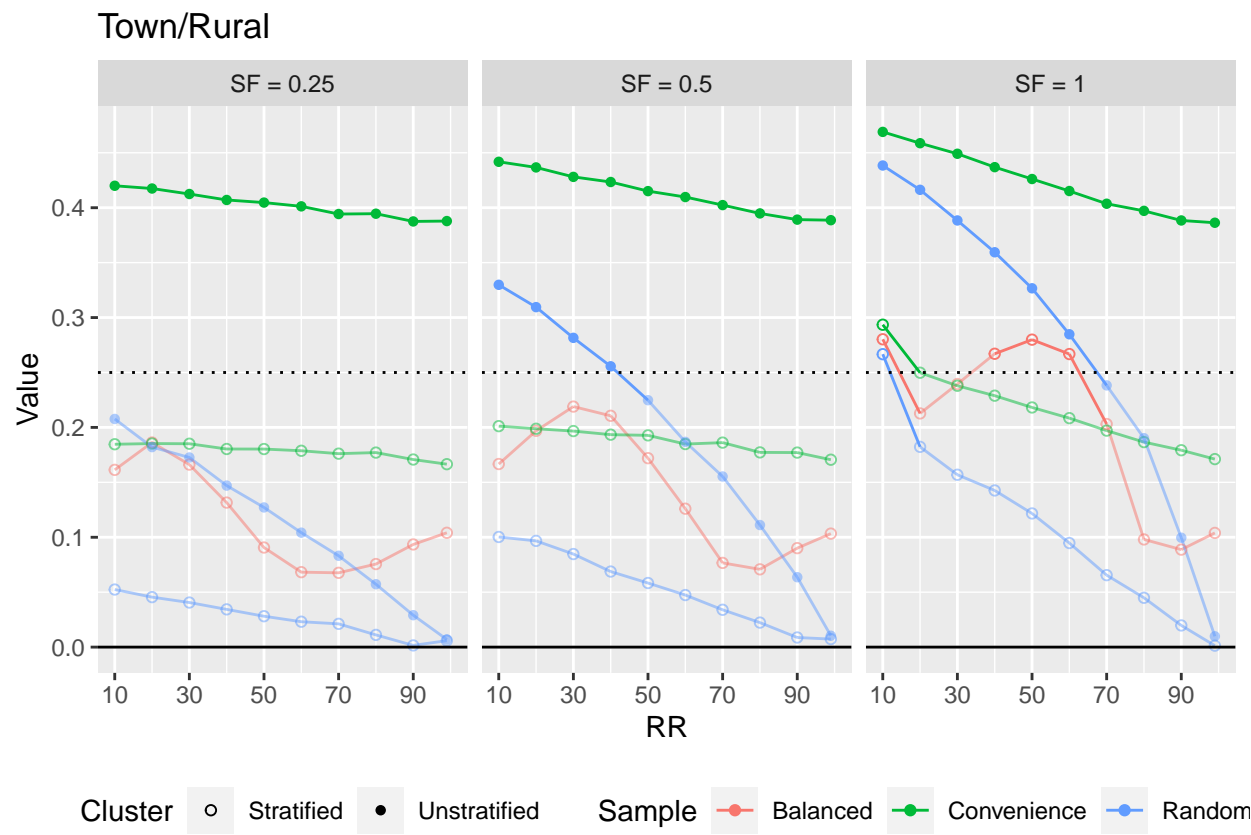
\$Suburban



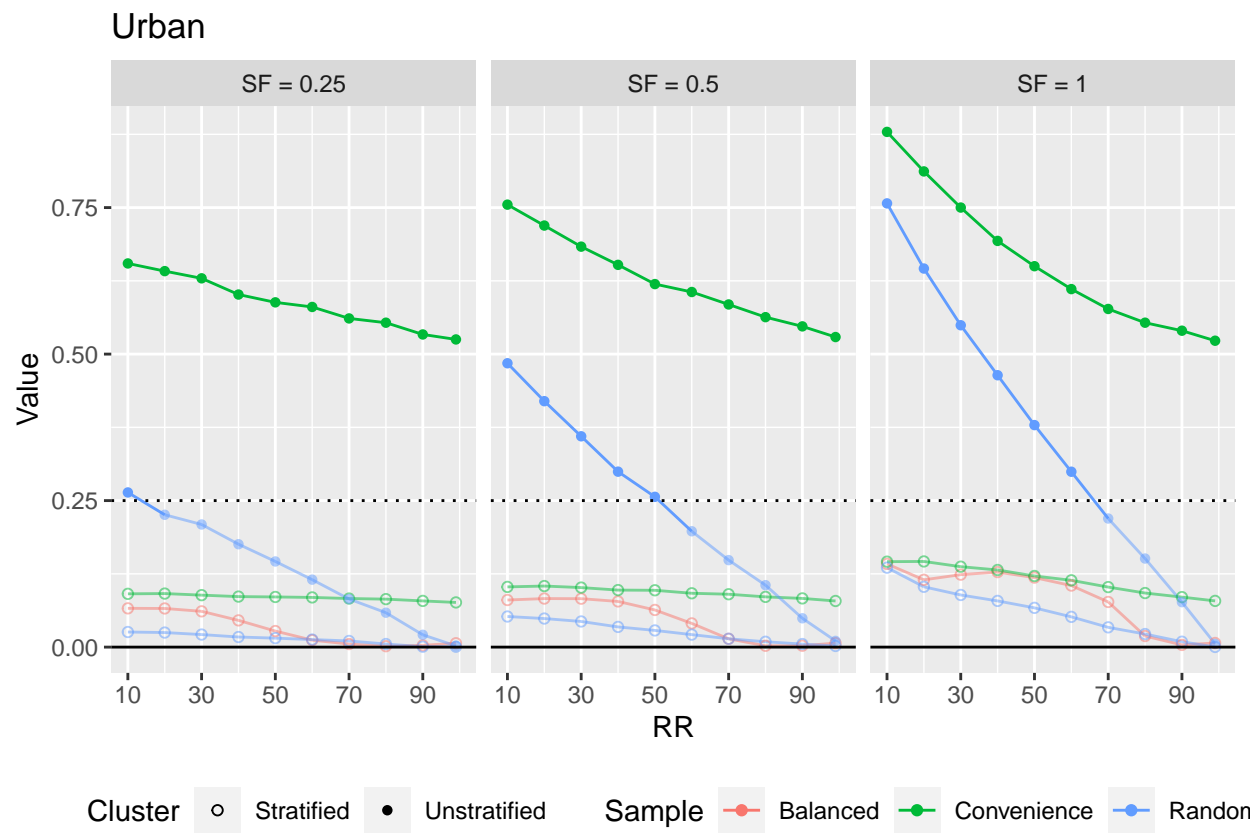
\$T1



\$ToRu



\$Urban

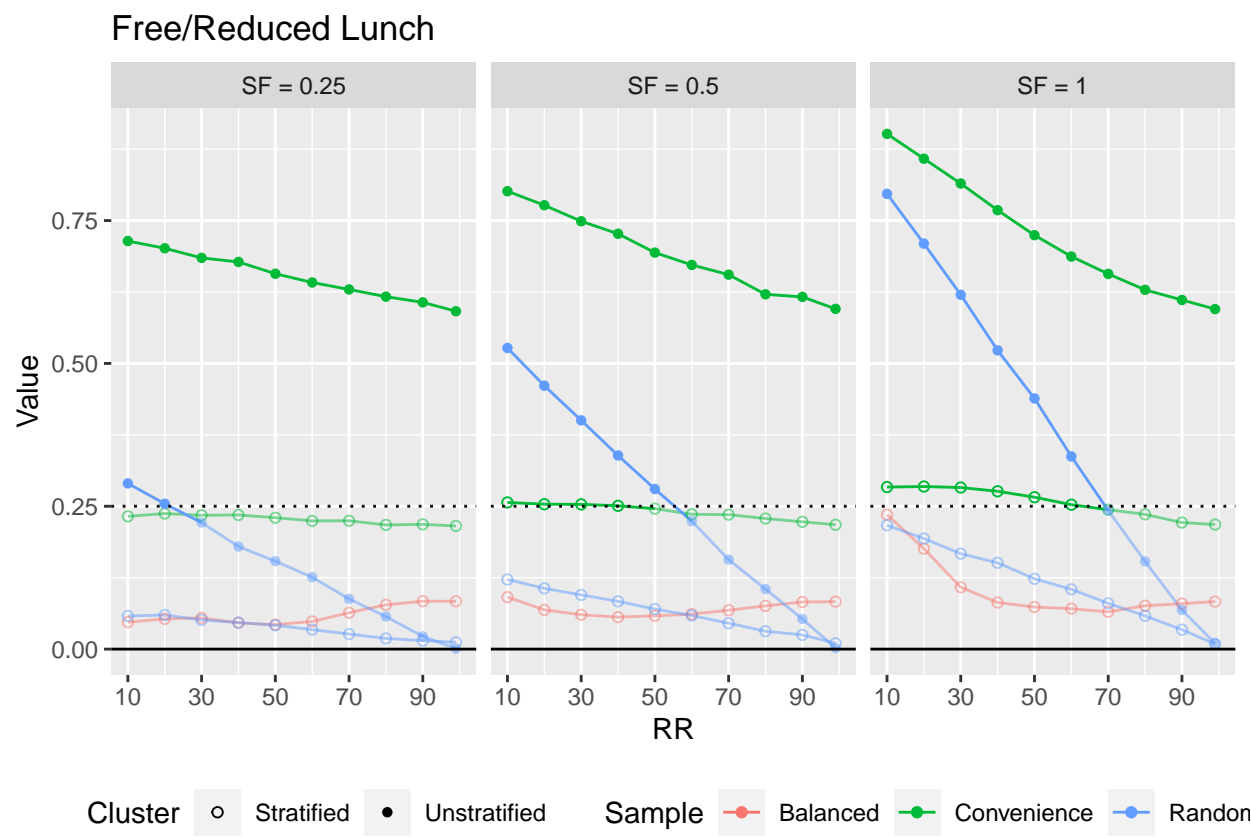


Group 3.

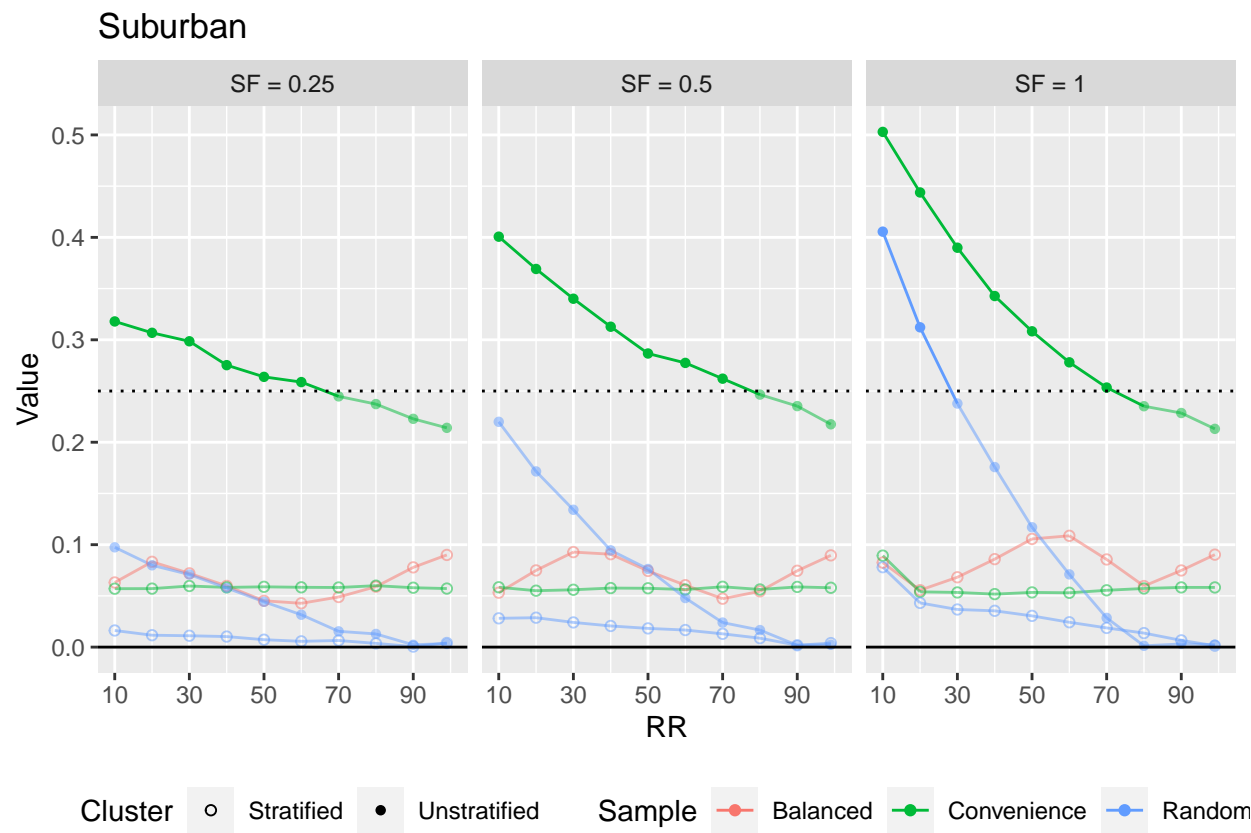
A tibble: 15 x 3

SB	vnames	Group
1	0.25 dSCH	Group 3
2	0.5 dSCH	Group 3
3	1 dSCH	Group 3
4	0.25 ethHisp	Group 3
5	0.5 ethHisp	Group 3
6	1 ethHisp	Group 3
7	0.25 ethWhite	Group 3
8	0.5 ethWhite	Group 3
9	1 ethWhite	Group 3
10	0.25 pELL	Group 3
11	0.5 pELL	Group 3
12	1 pELL	Group 3
13	0.5 pTotfrl	Group 3
14	1 pTotfrl	Group 3
15	1 ToRu	Group 3

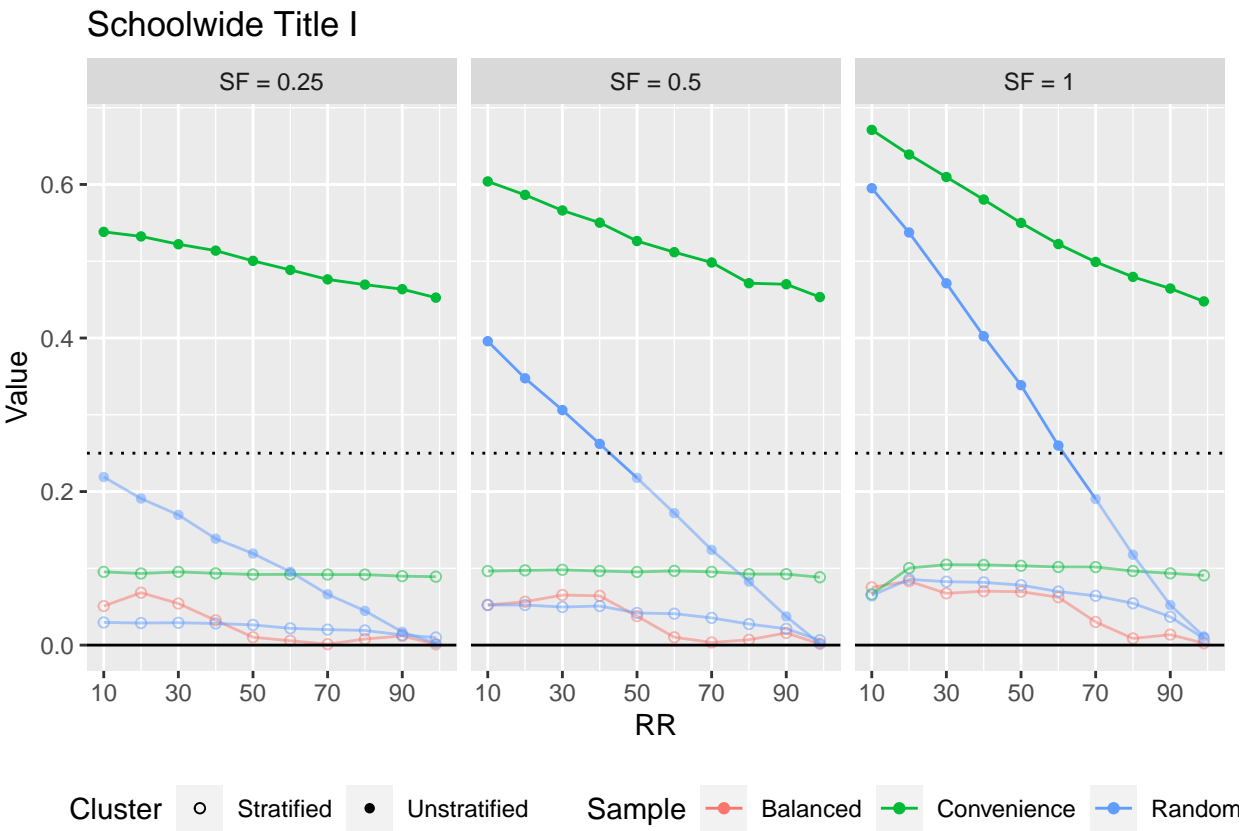
\$pTotfrl



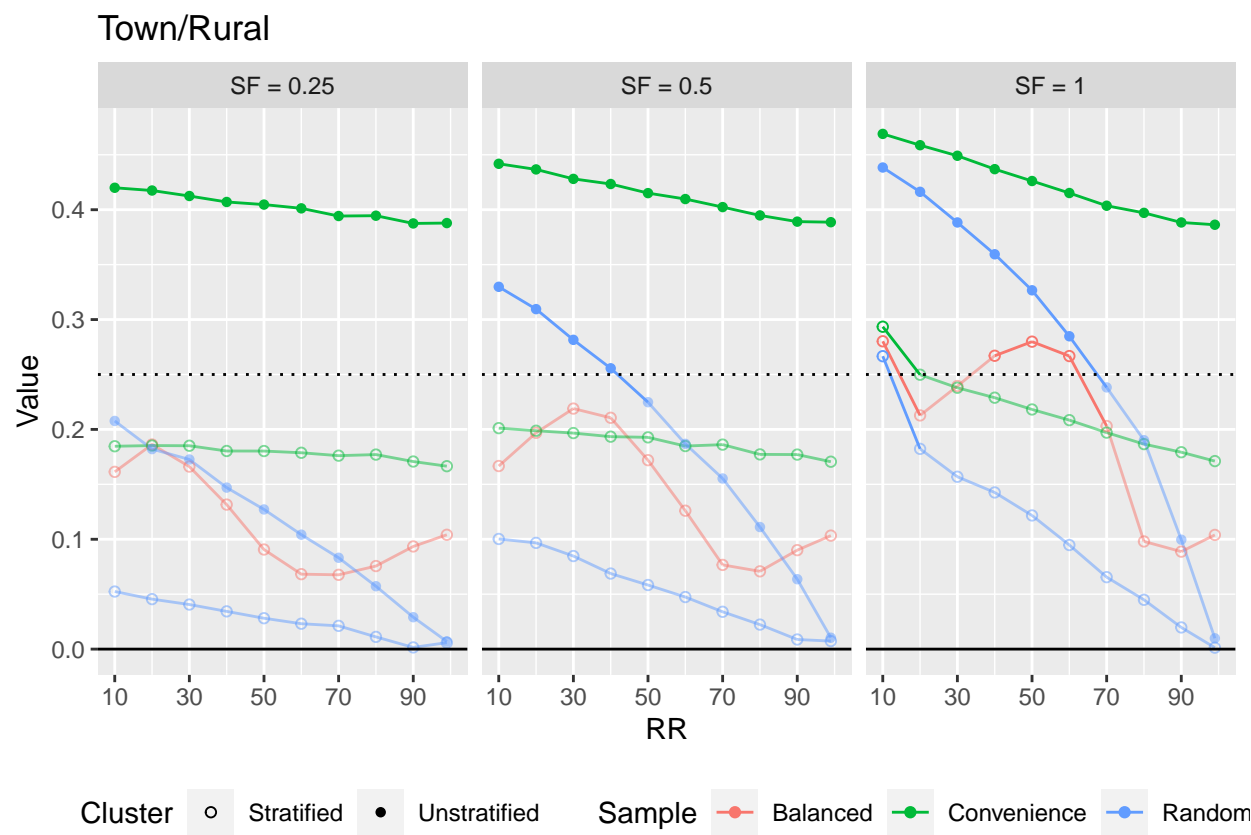
\$Suburban



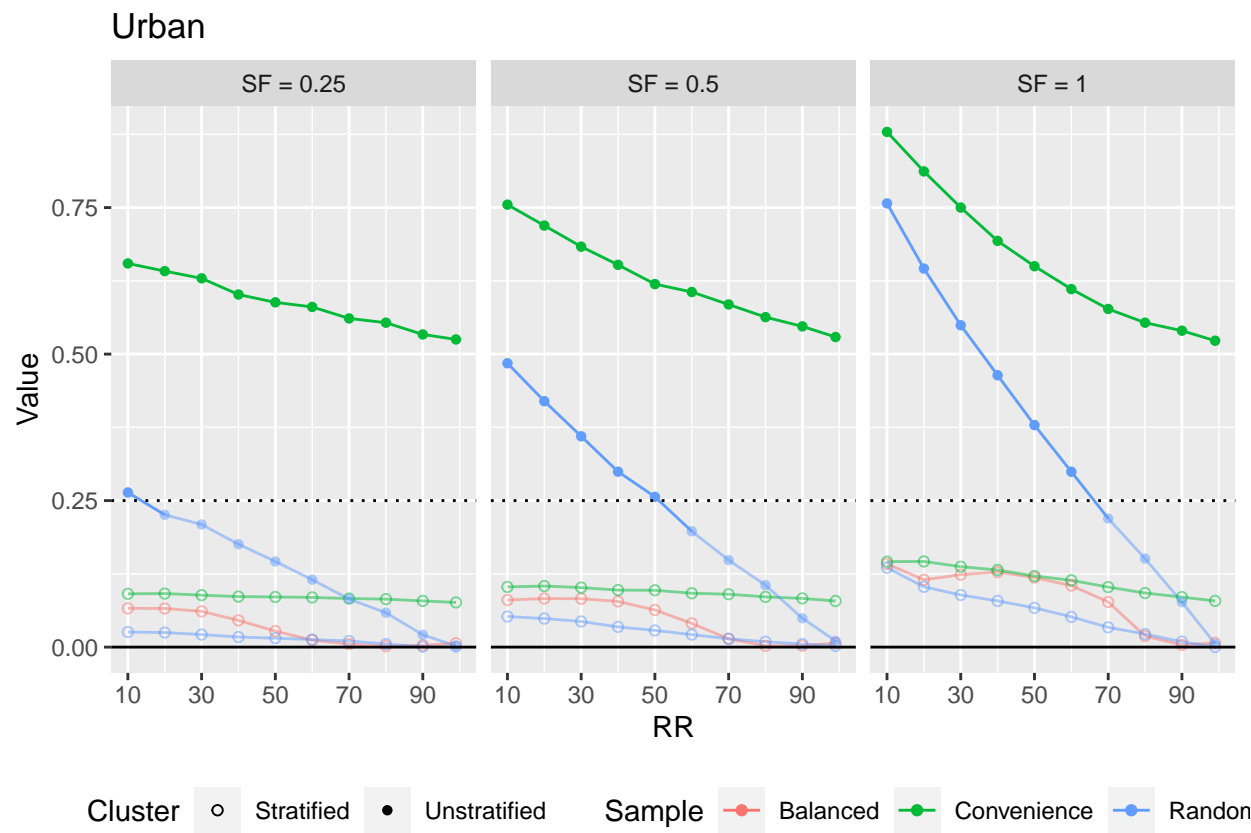
\$T1



\$ToRu



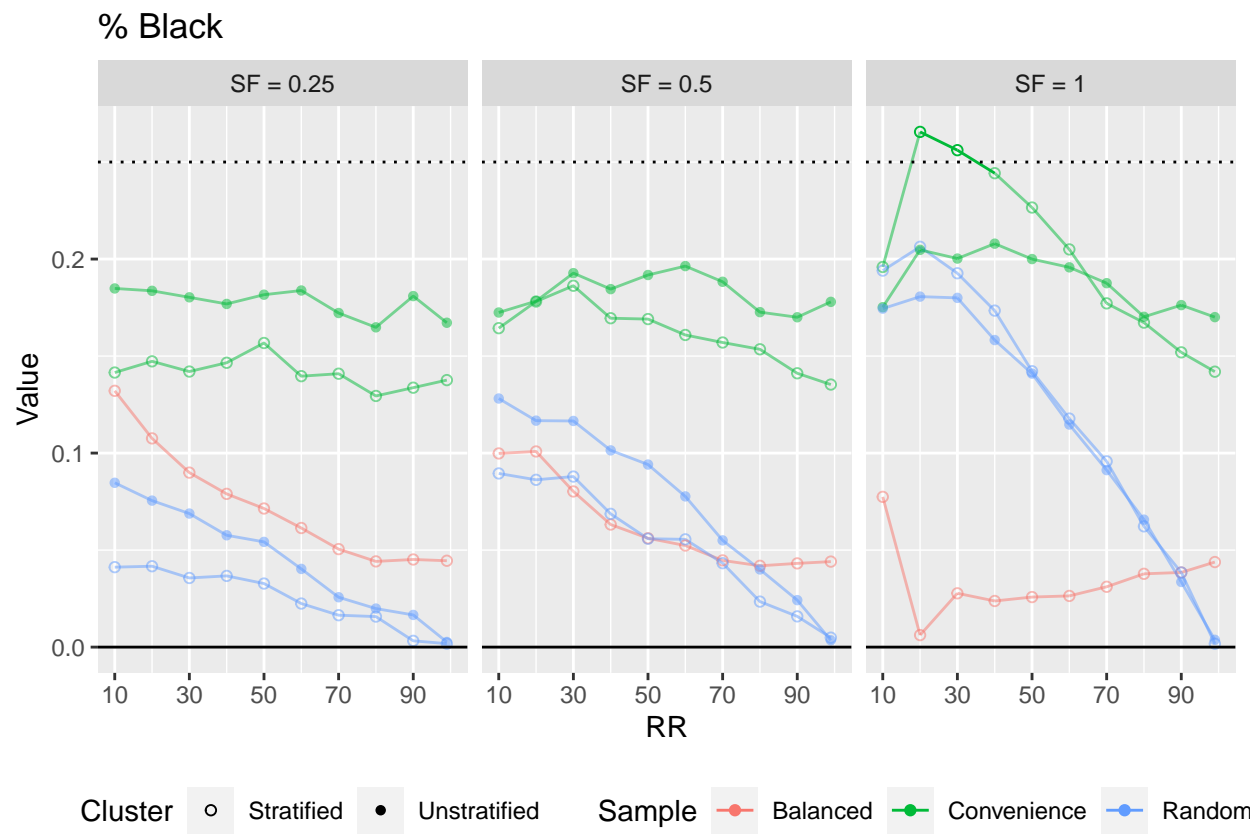
\$Urban



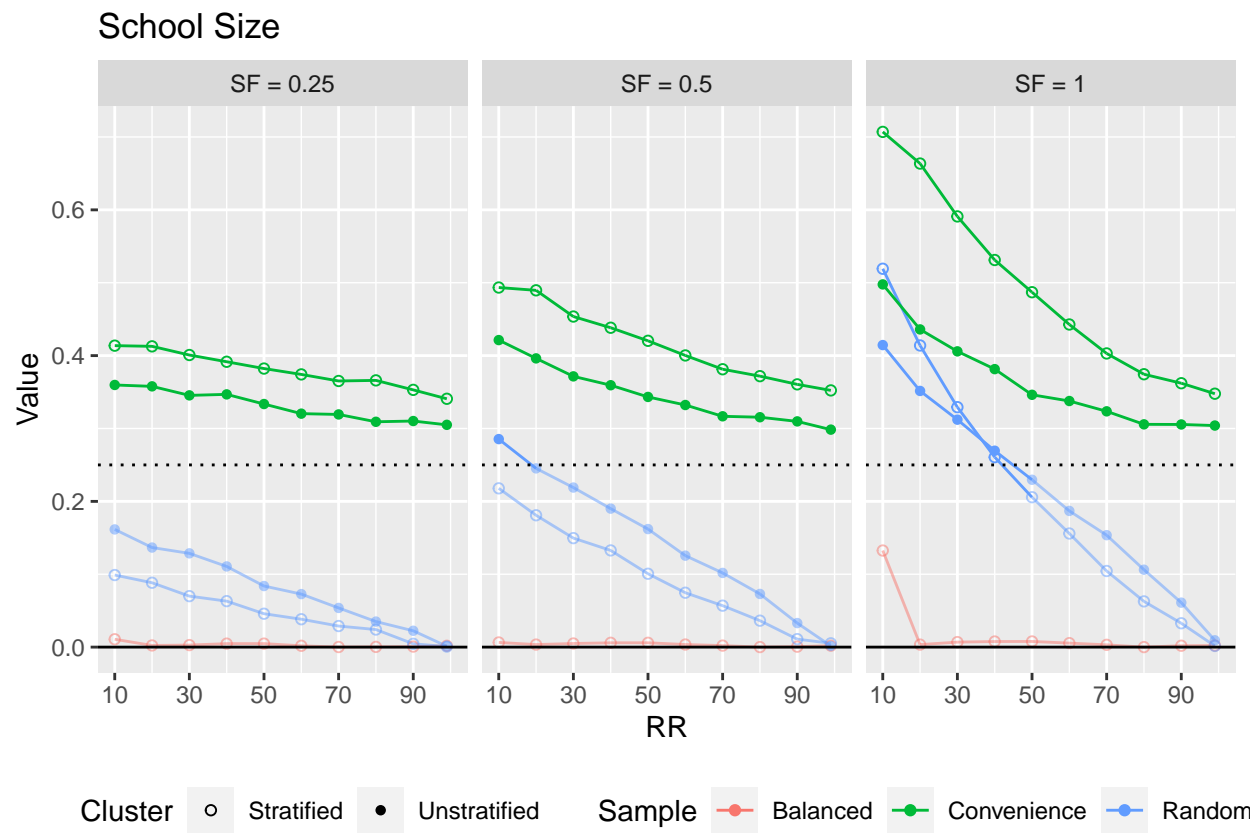
Group 4 # A tibble: 4 x 3 SB vnames Group

1 1 ethBlack Group 4 2 0.25 n Group 4 3 0.5 n Group 4 4 1 n Group 4

\$ethBlack



\$n



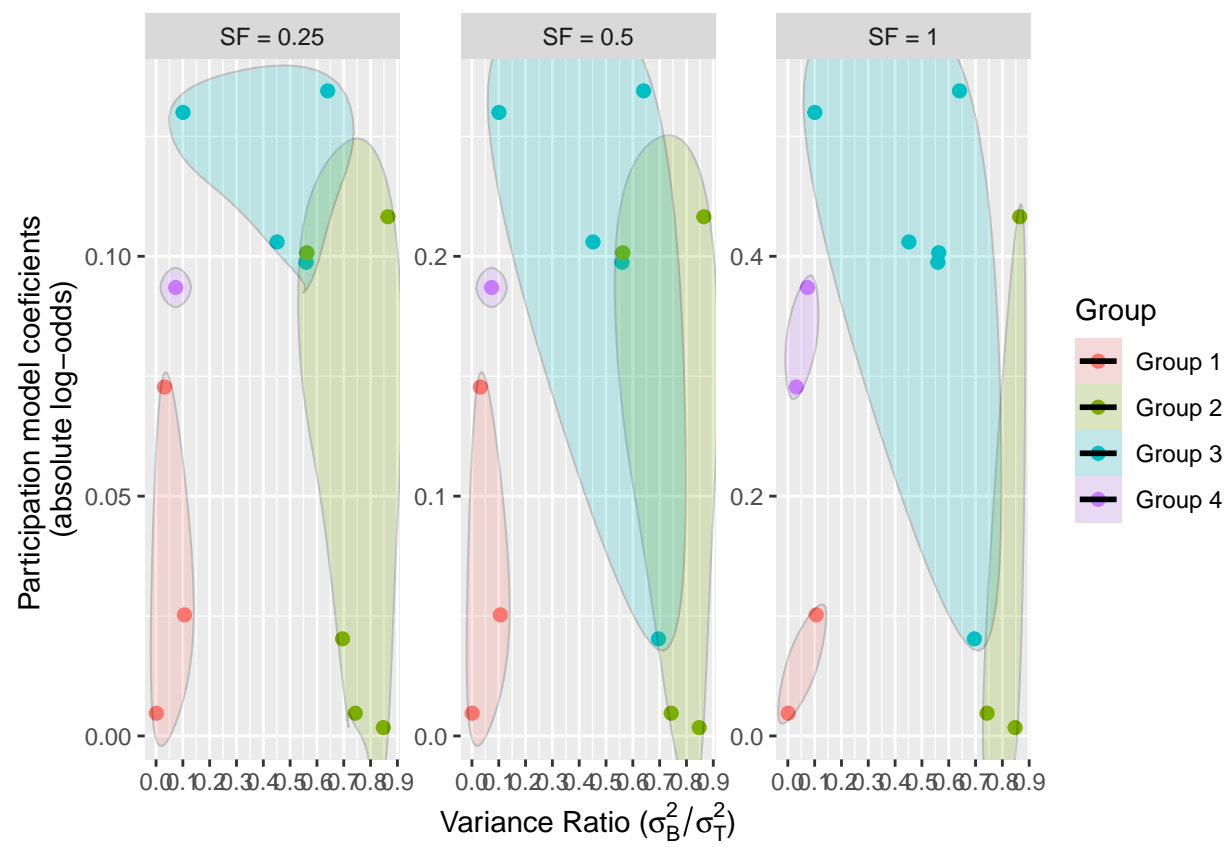
Fine Grain Groups.

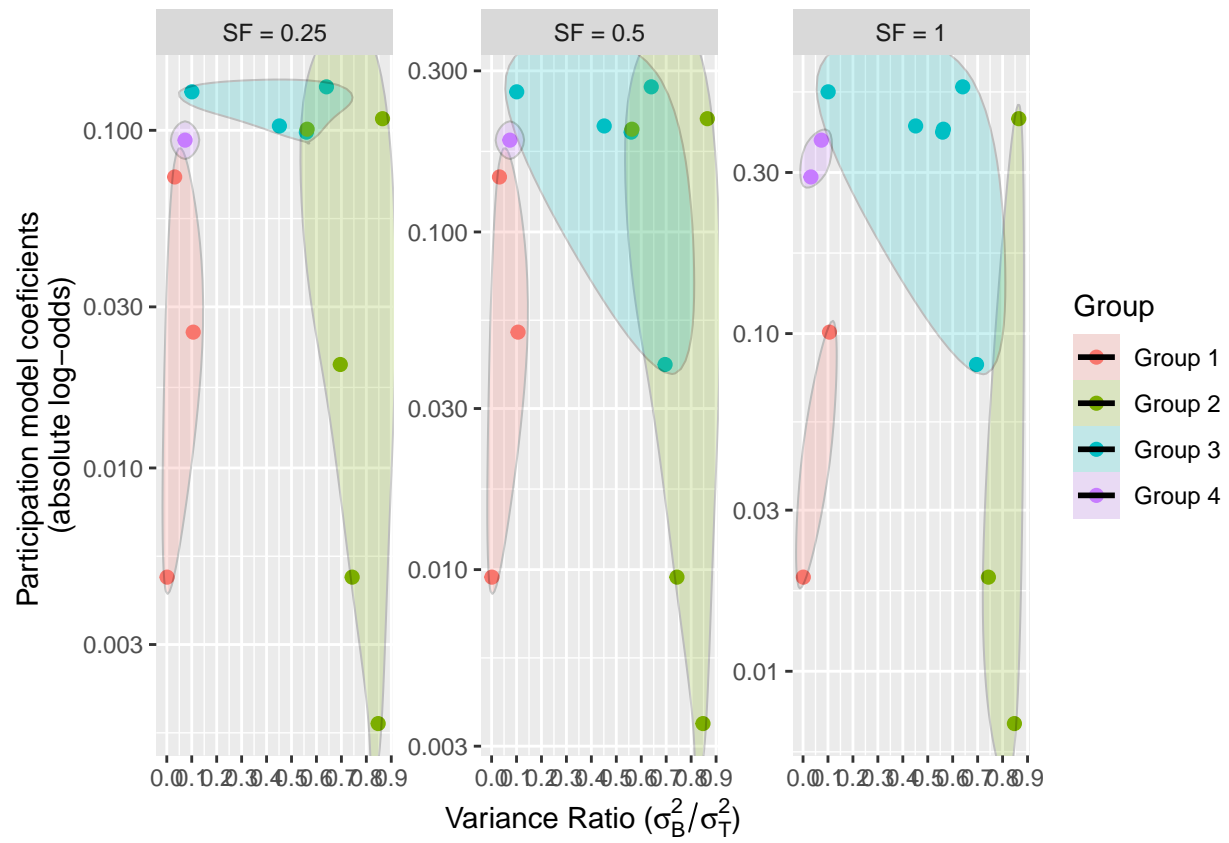
V-ratio and Log odds.

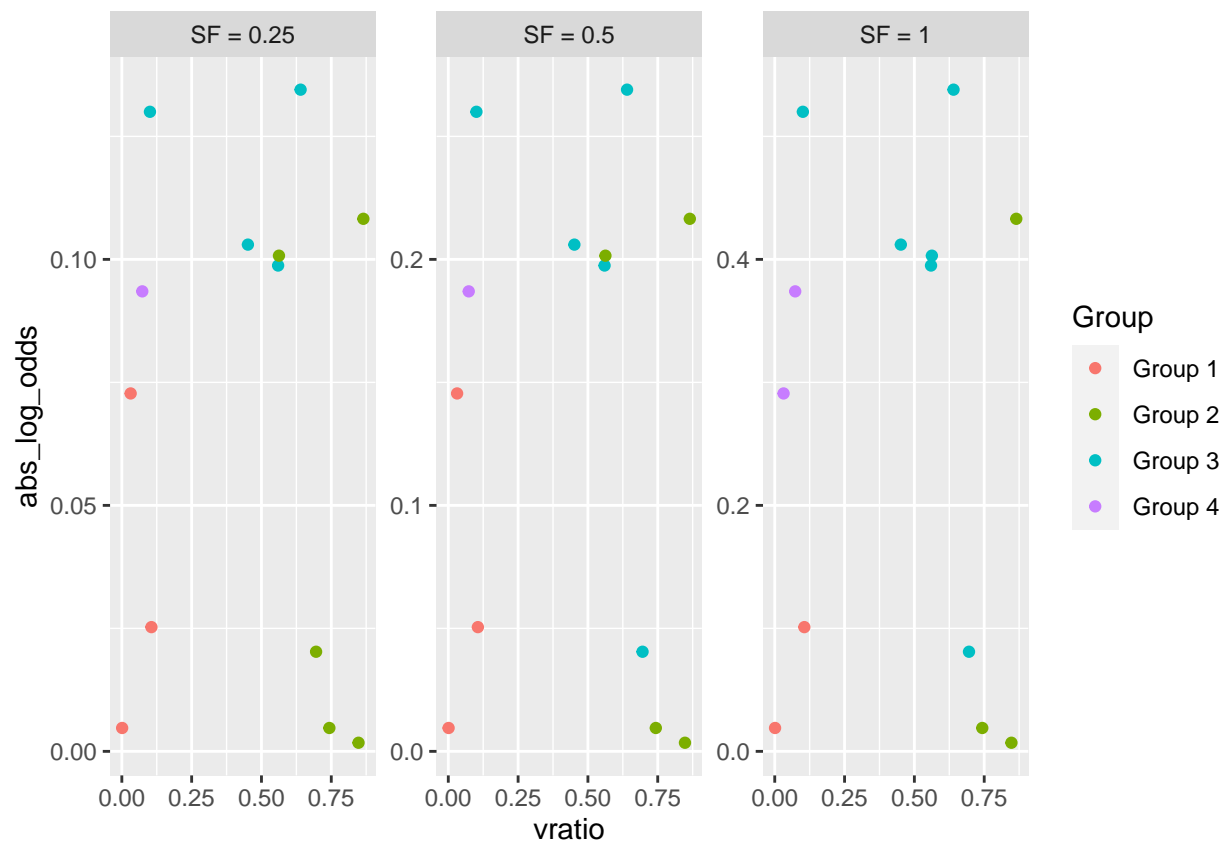
A tibble: 39 x 15

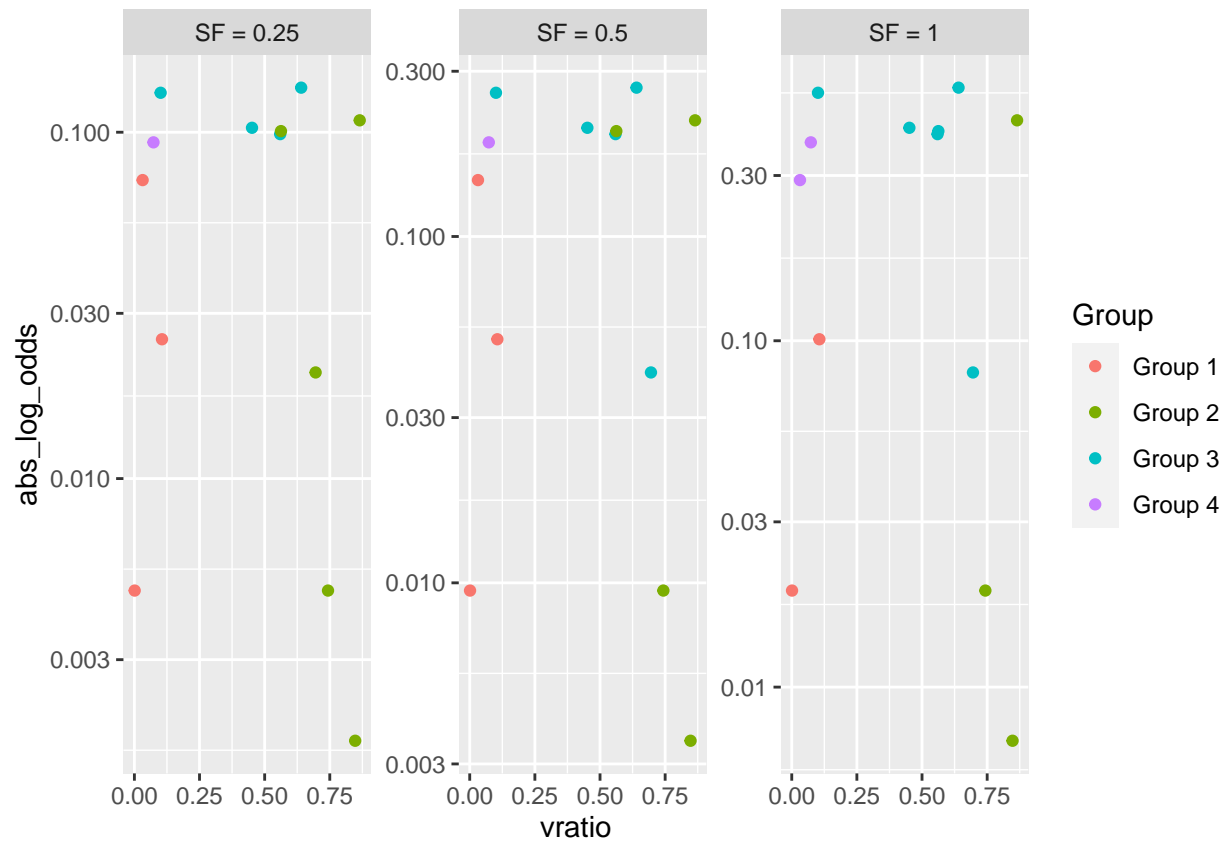
vnames	T.SS	W.SS	B.SS	vratio	Sub	Category	Type	Variables	log_odds	1	dSCH
4.02e8	3.62e8	4.05e7	0.101	Dist~	School ~	Mean	District~	0.52	2	dSCH	4.02e8 3.62e8 4.05e7
0.101	Dist~	School ~	Mean	District~	0.52	3	dSCH	4.02e8 3.62e8 4.05e7	0.101	Dist~	School ~
~	Mean	District~	0.52	4	ethBlack	2.52e6 2.44e6 7.96e4	0.0316	Ethn~	Student~	Mean %	
Black	0.291	5	ethBlack	2.52e6 2.44e6 7.96e4	0.0316	Ethn~	Student~	Mean %	Black	0.291	6
ethBlack	2.52e6 2.44e6 7.96e4	0.0316	Ethn~	Student~	Mean %	Black	0.291	7	ethHispan	1.05e7 4.63e6 5.89e6	0.560
Ethn~	Student~	Mean %	Hispan~	0.395	8	ethHispan	1.05e7 4.63e6 5.89e6	0.560	Ethn~	Student~	Mean %
Hispan~	0.395	9	ethHispan	1.05e7 4.63e6 5.89e6	0.560						

Ethn~ Student~ Mean % Hispan~ 0.395 10 ethWhite 9.10e6 3.28e6 5.83e6 0.640 Ethn~
Student~ Mean % White -0.538 # ... with 29 more rows, and 5 more variables: SB ,
Coefficient , # abs_log_odds , SF_fac , Group

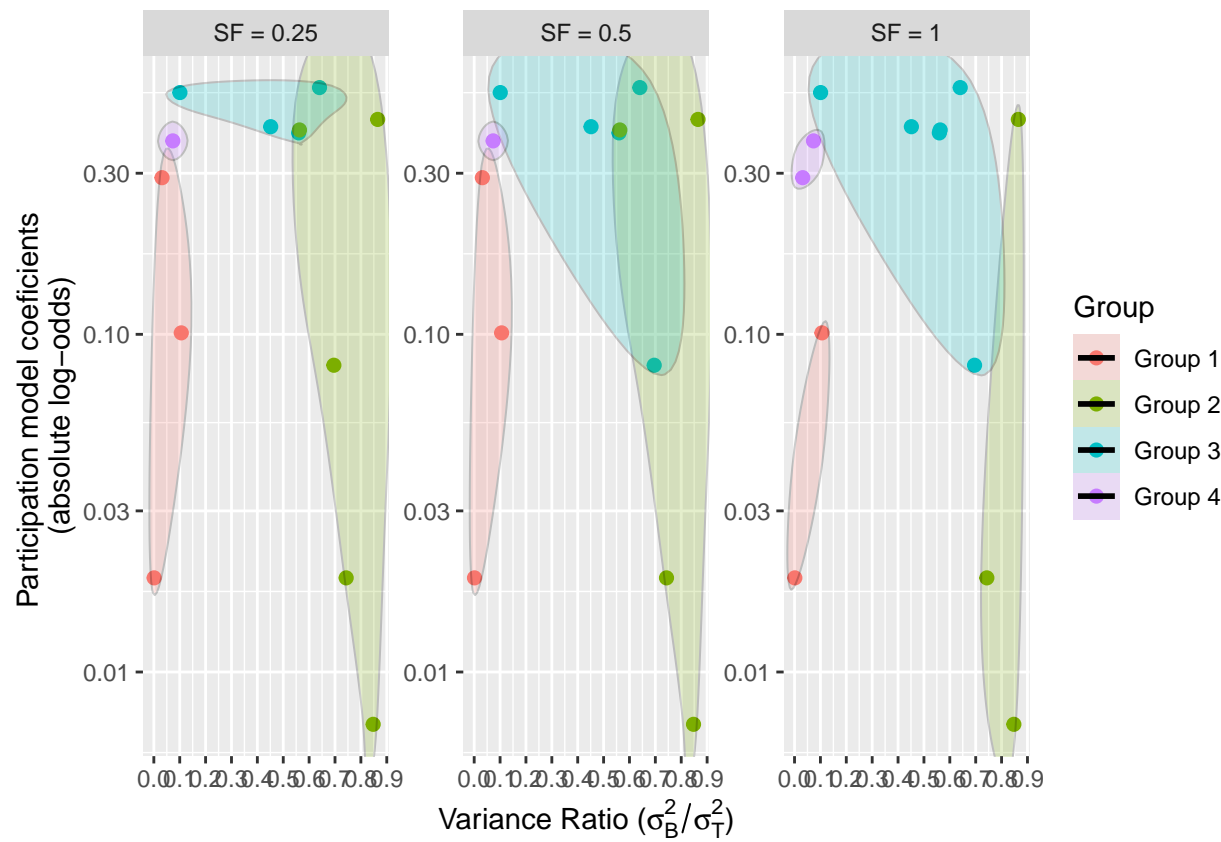




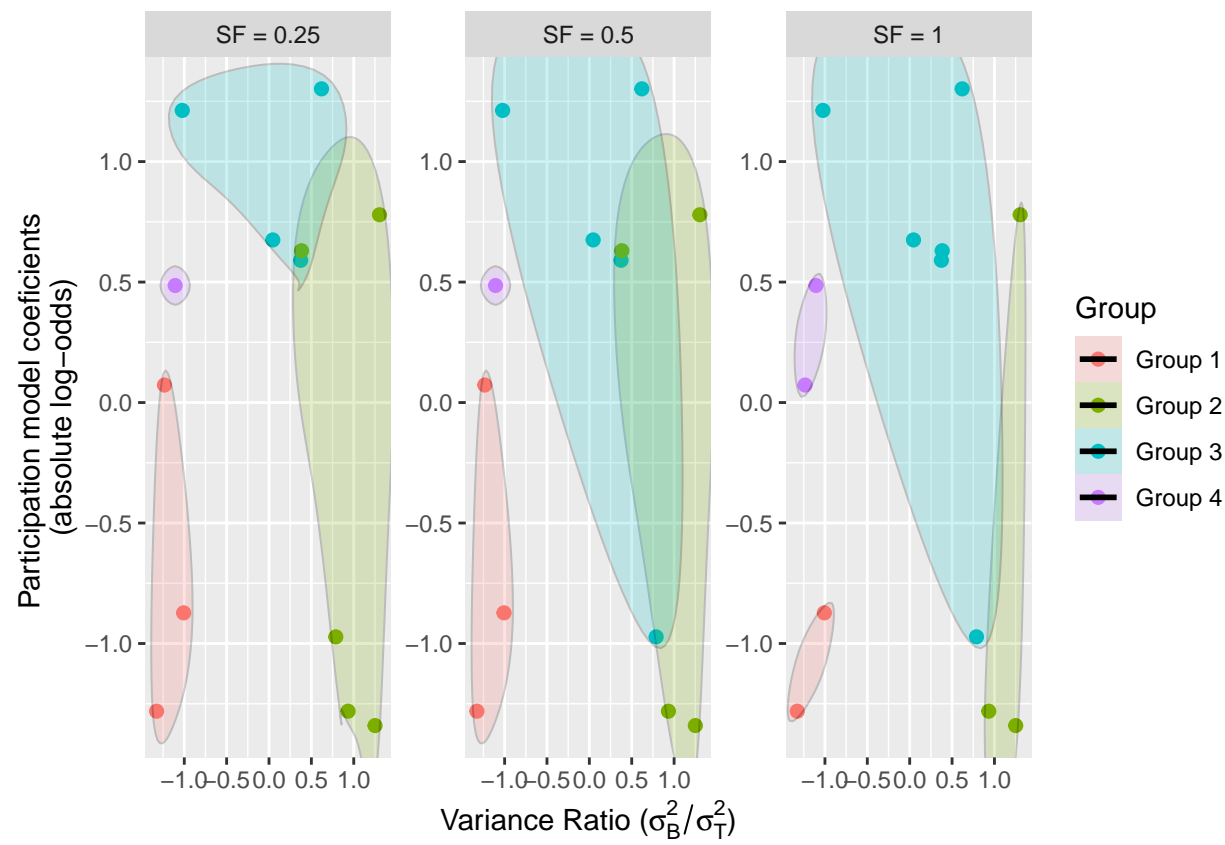




Test 1 - same abs_log_odds.



Test 2 - relative log_odds and relative vratio



Feasibility

Sampling Difficulty.

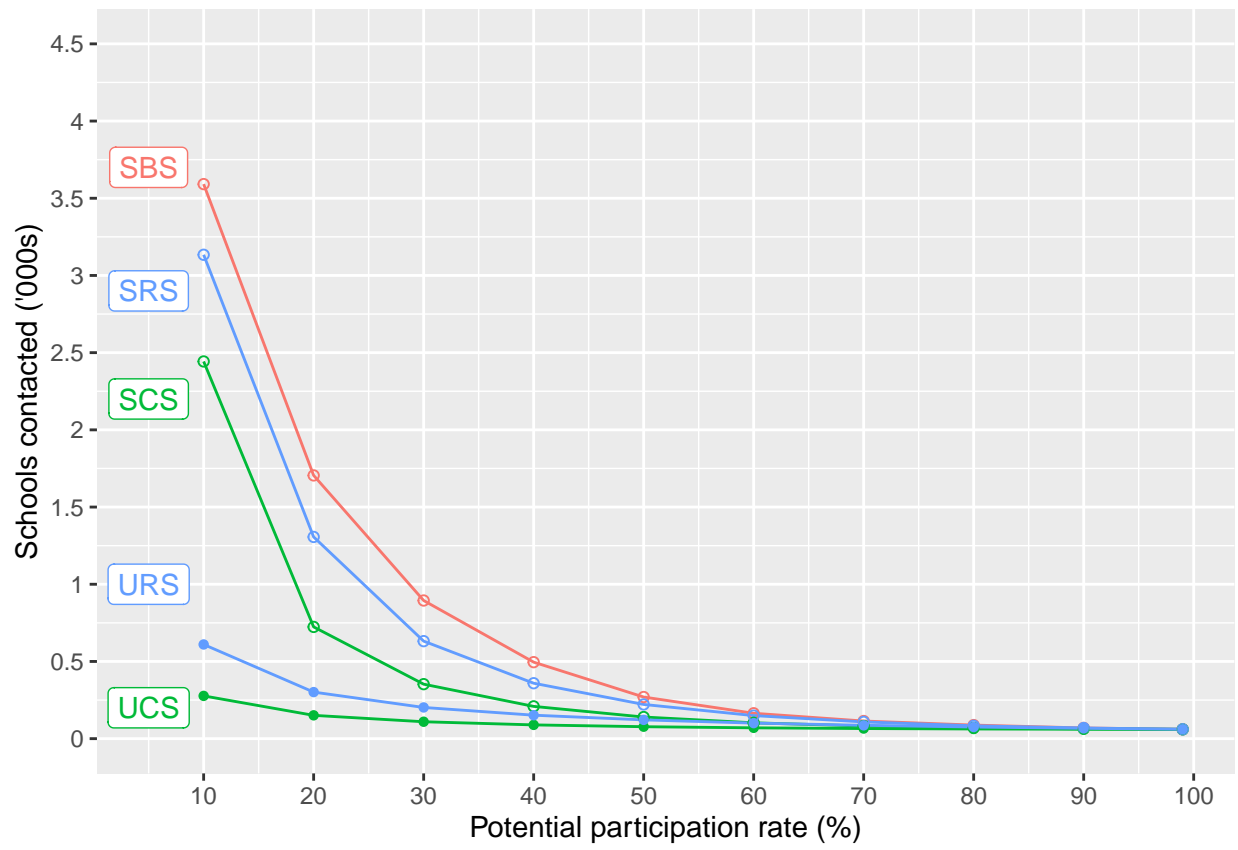
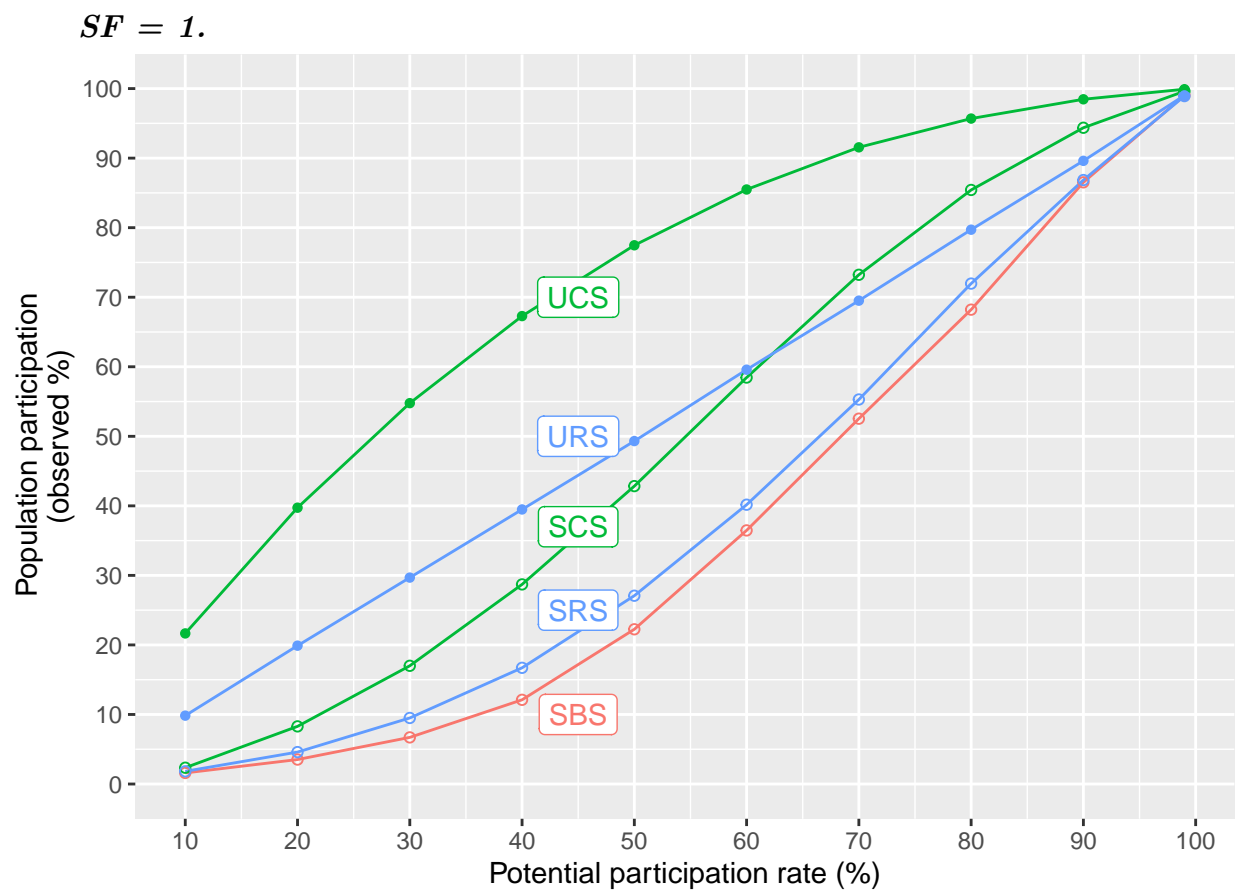
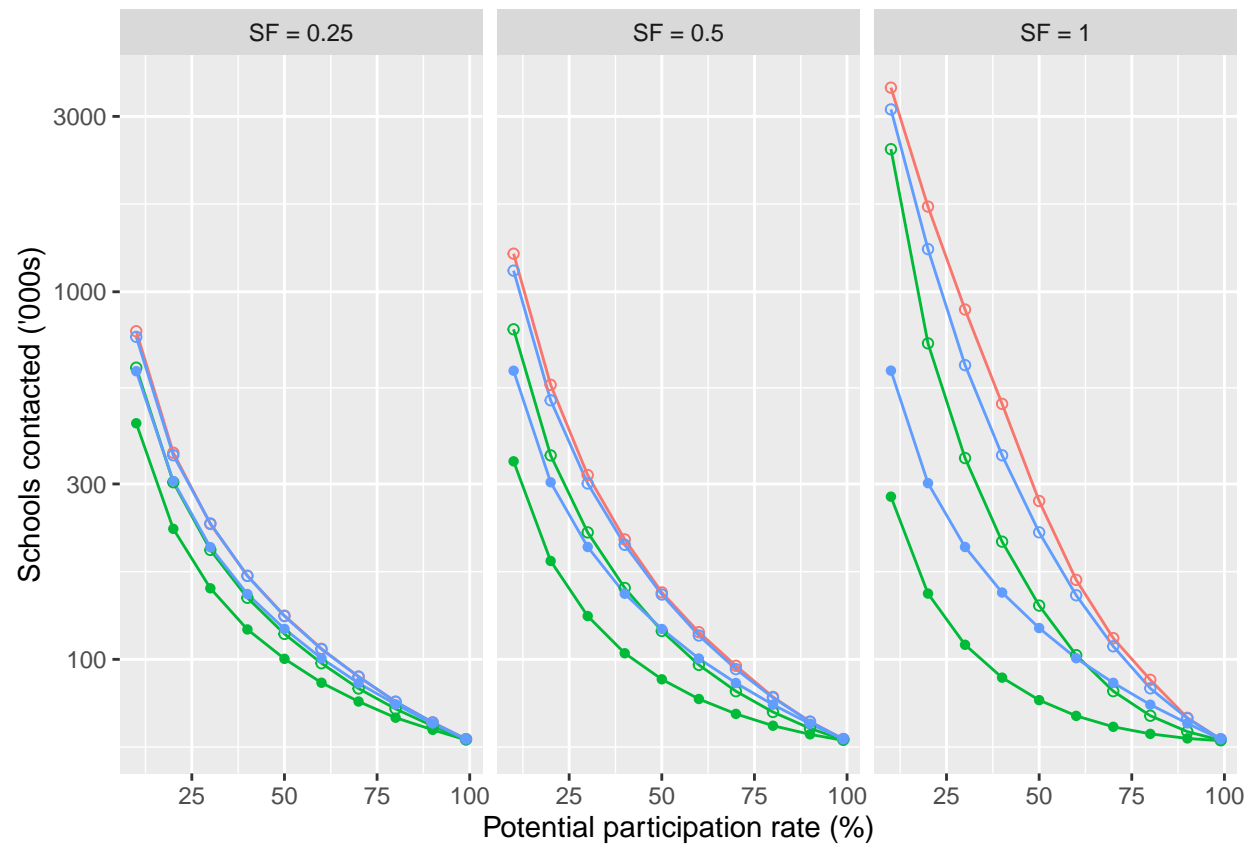
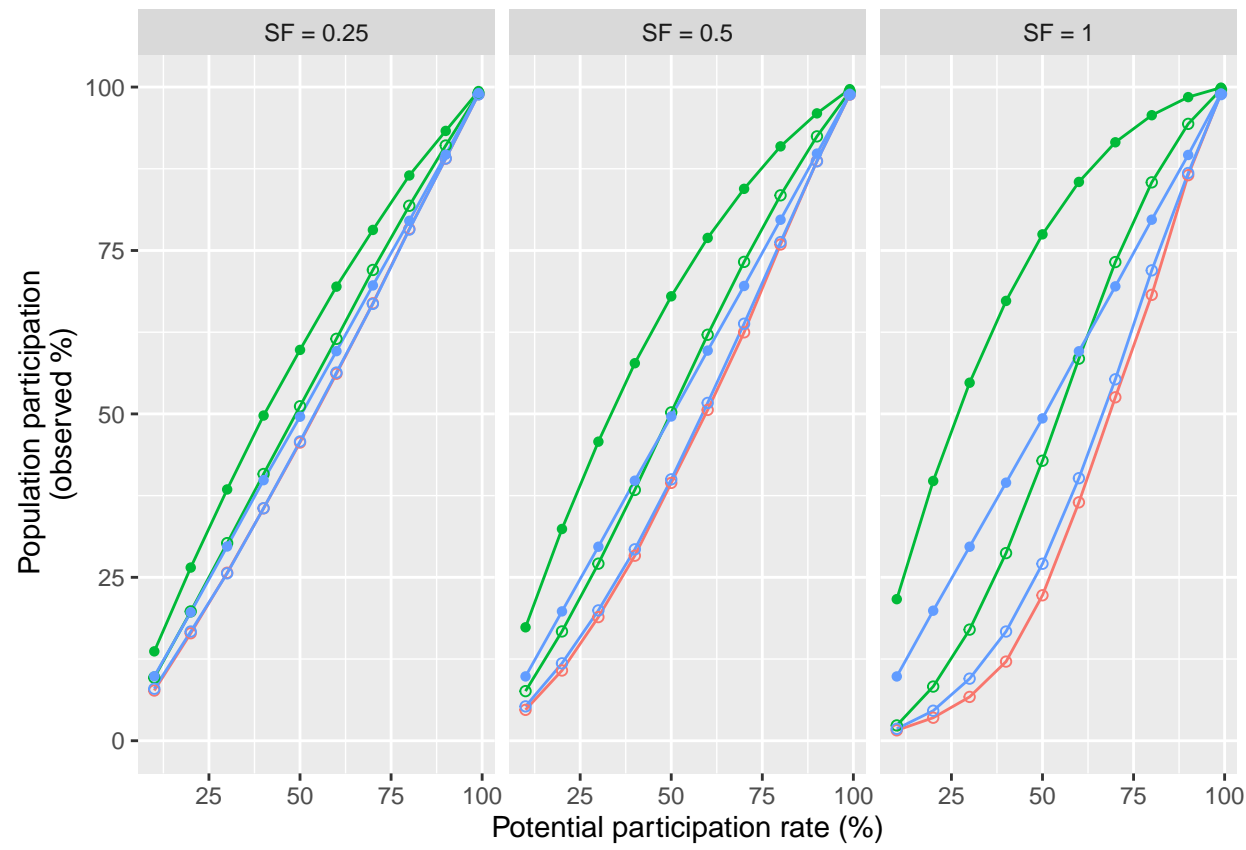


Figure 3. Schools Contacted

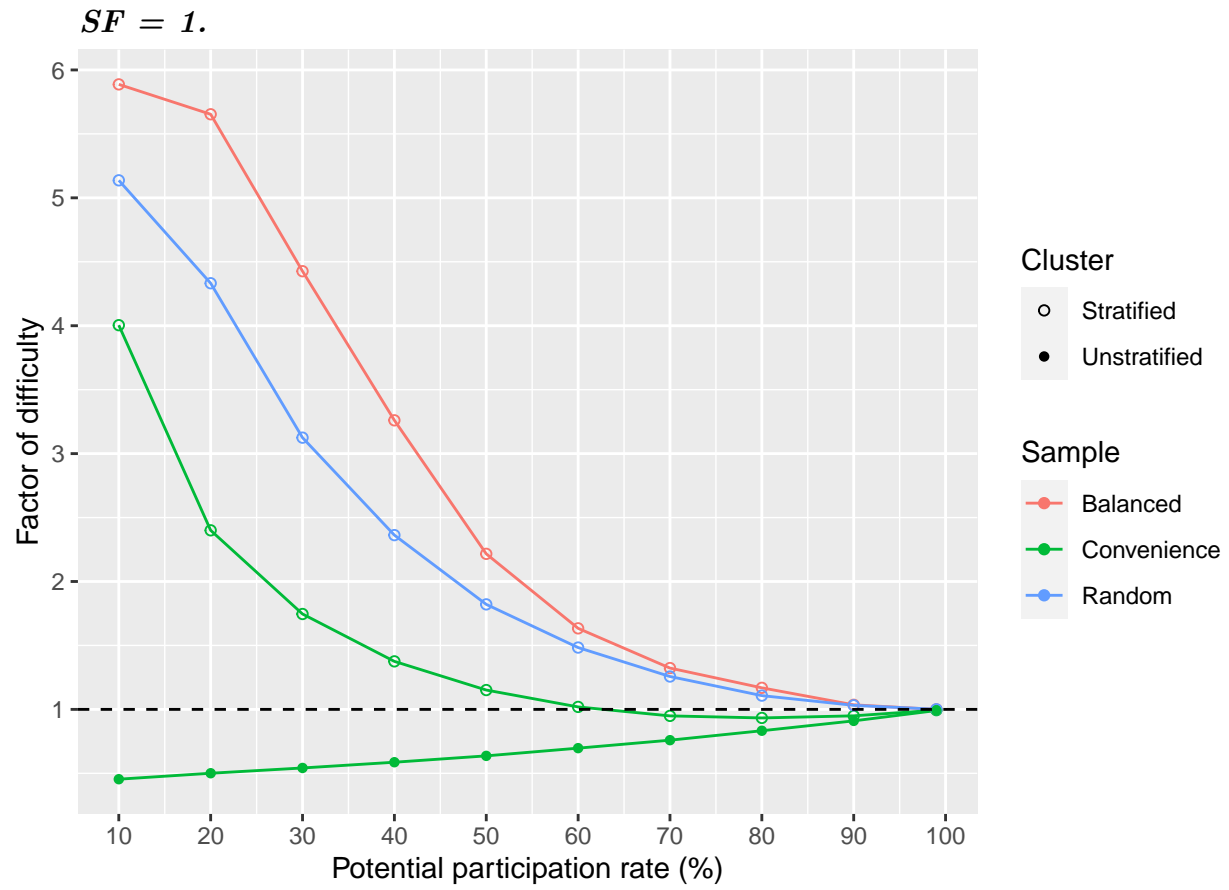


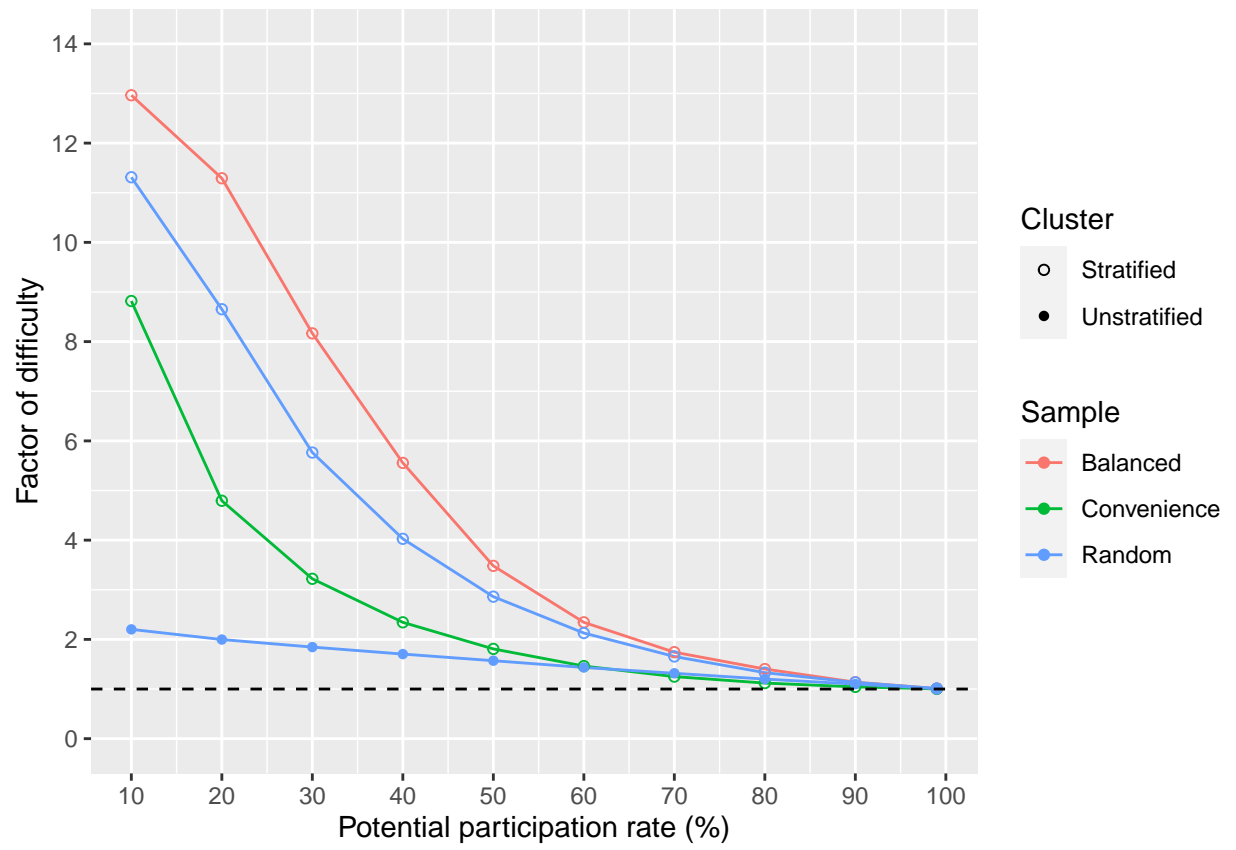
SF = [1, .5, .25, .01]

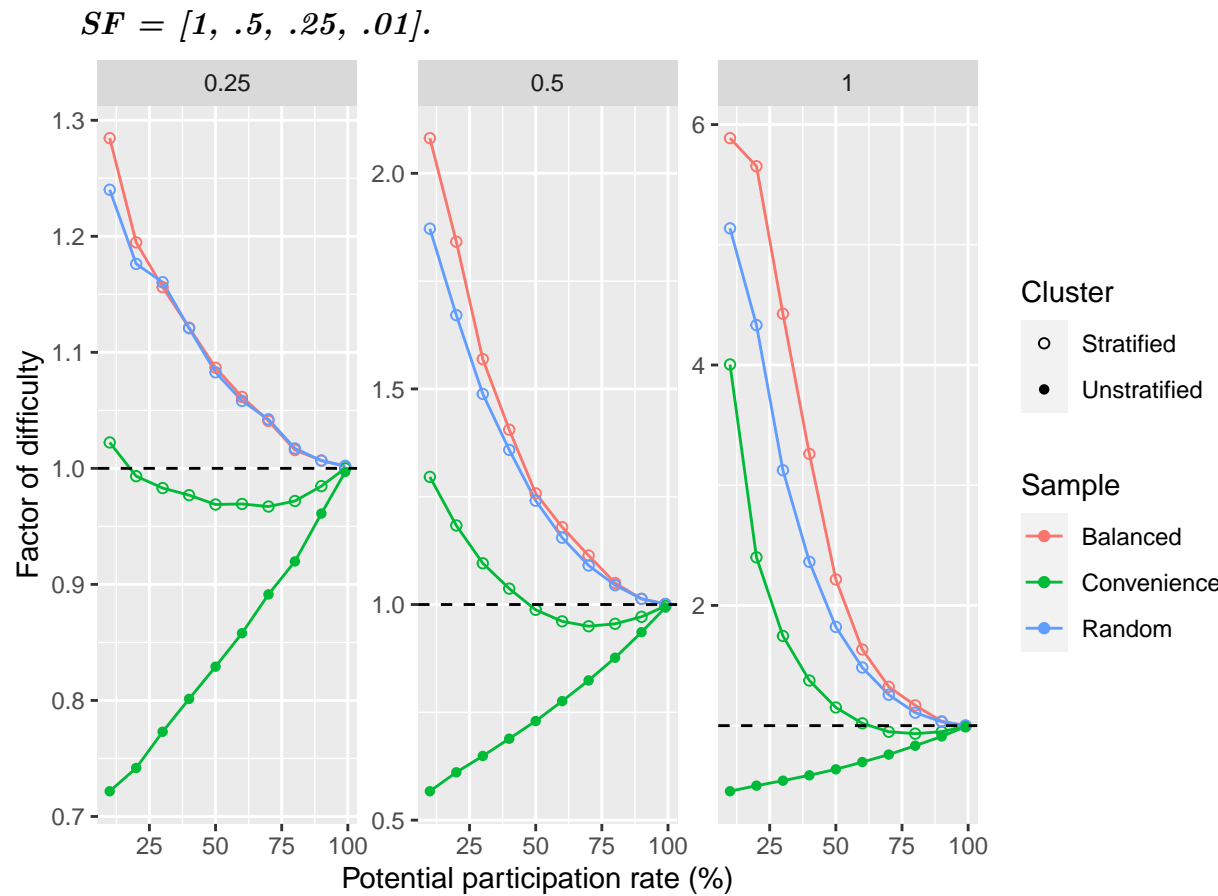




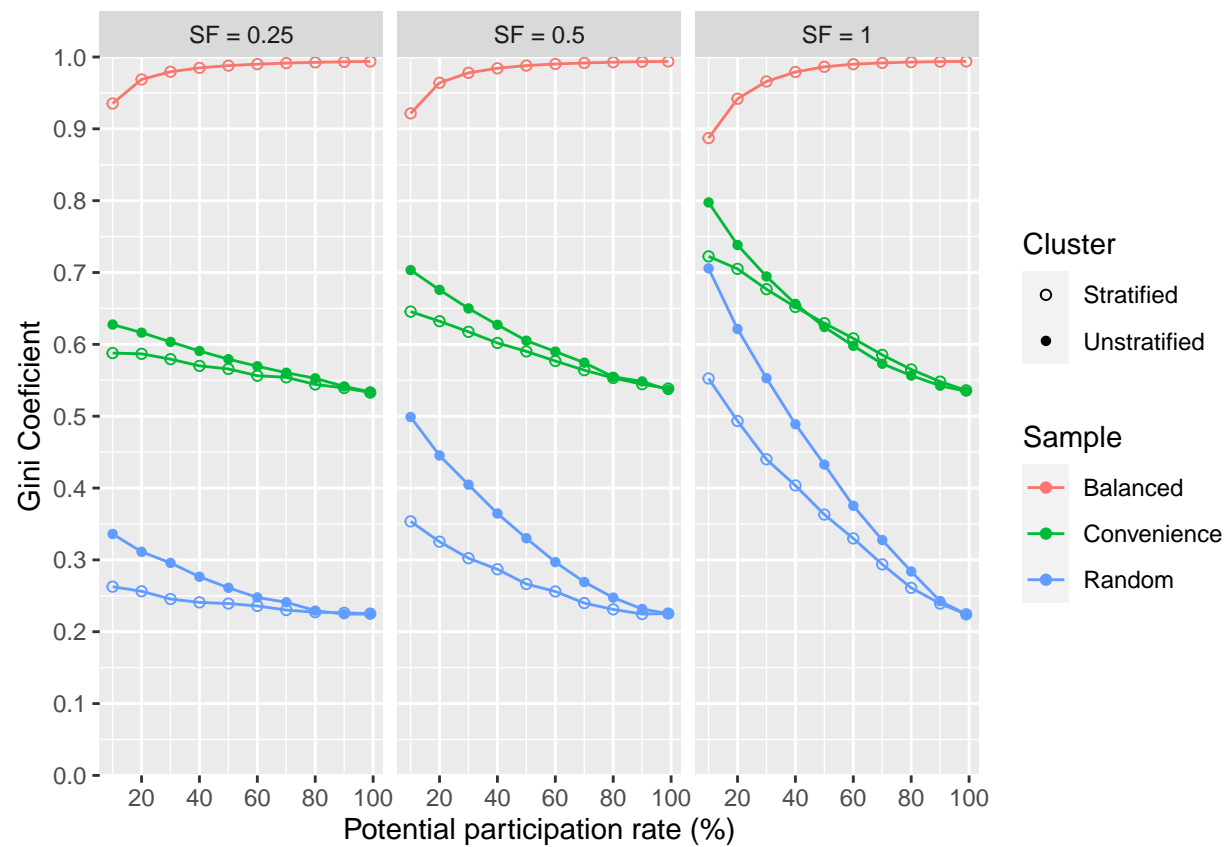
Difficulty Relative to Comparison Methods.







Gini Plot



Export Data

A tibble: 28 x 2

Vars MB

1 tabs 473.62 Mb 2 figs 301.614 Mb 3 results 84.955 Mb 4 df.list 19.796 Mb 5 df.sim

3.027 Mb

6 plot_smd 0.077 Mb

7 test.group 0.066 Mb

8 df.smd.ipsw 0.038 Mb

9 get_quant 0.03 Mb

10 test 0.013 Mb

... with 18 more rows # A tibble: 20 x 2 fig size

1 dist 263.1 Mb 2 ch 0 Mb

3 vratio 0 Mb

4 allocation 0 Mb

5 joint.cluster.stats 0 Mb

6 cluster.SMD 0.2 Mb

7 B 0.1 Mb

8 smd.by.scale 0.4 Mb

9 smd.examples 0.1 Mb

10 smd.examples.sf100 0 Mb

11 v.coef 0 Mb

12 samples.contacts.sf100 0 Mb

13 samples.response.rates.sf100 0 Mb

14 samples.contacts.sf.all 0 Mb

15 samples.response.rates.sf.all 0 Mb

16 relative.counts.URS 0 Mb

17 relative.counts.UCS 0 Mb

18 samples.relative.contacts.sf.all 0 Mb

19 gini 0 Mb

20 gini.coef 37.4 Mb # A tibble: 16 x 2 tab size

1 pop.desc 0 Mb

2 dist 263.1 Mb 3 allocation 0 Mb

4 cluster.SMD 0.2 Mb

5 strata.data 0 Mb

6 intercepts 0 Mb

7 B 0 Mb
8 smd.results 0.2 Mb
9 smd.by.scale 0.6 Mb
10 smd.examples 0.1 Mb
11 samples 0 Mb
12 samples.contacts.labels.sf100 0 Mb
13 samples.response.rates.labels.sf100 0 Mb
14 relative.counts 0.1 Mb
15 gini.coef 0 Mb
16 gini.count 209.2 Mb

Presentation Figures