GenSamp: RESULTS

Gleb Furman[1], James E. Pustejovsky[2], & Elizabeth Tipton[3]

[1] University of Texas at Austin

[2] University of Wisconsin-Madison

[3] Northwestern University

GenSamp: RESULTS

Setup

Packages and Data

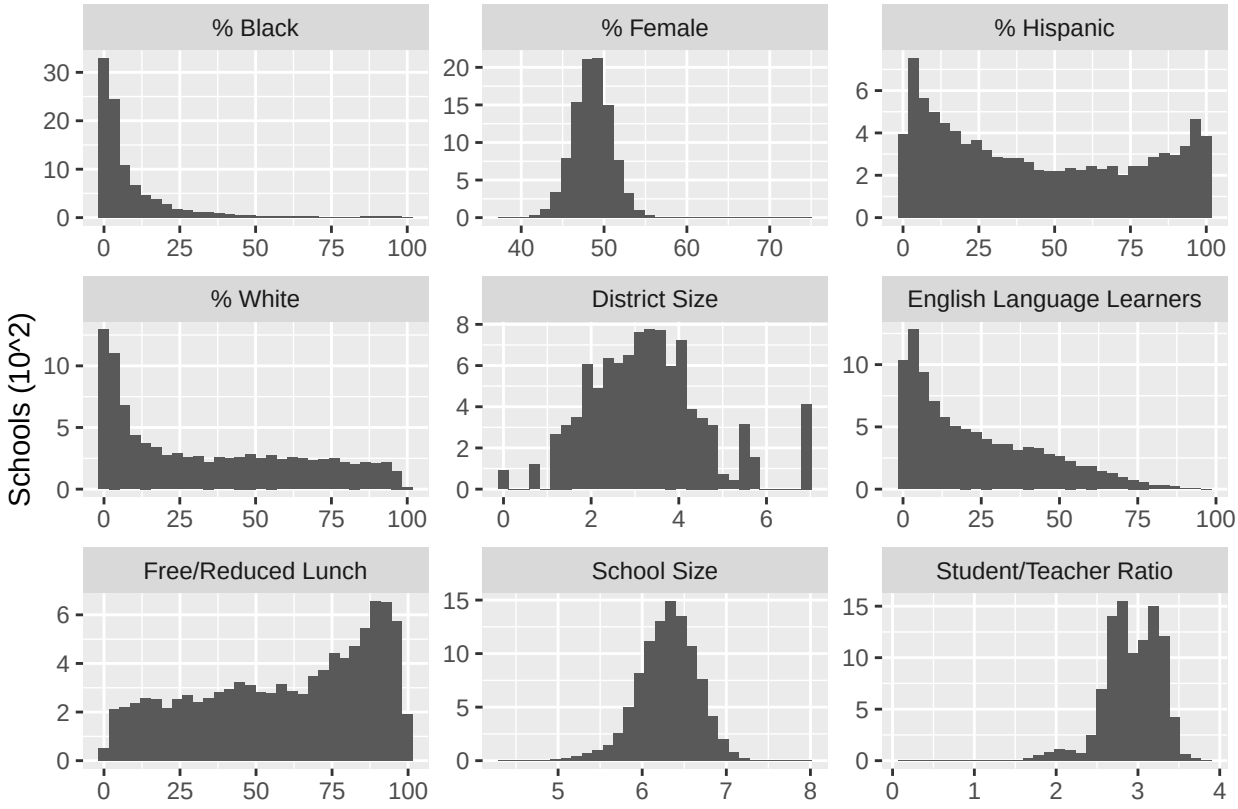
Organize Objects

Data Summary

Covaraite Statistics

Continuous variable distributions

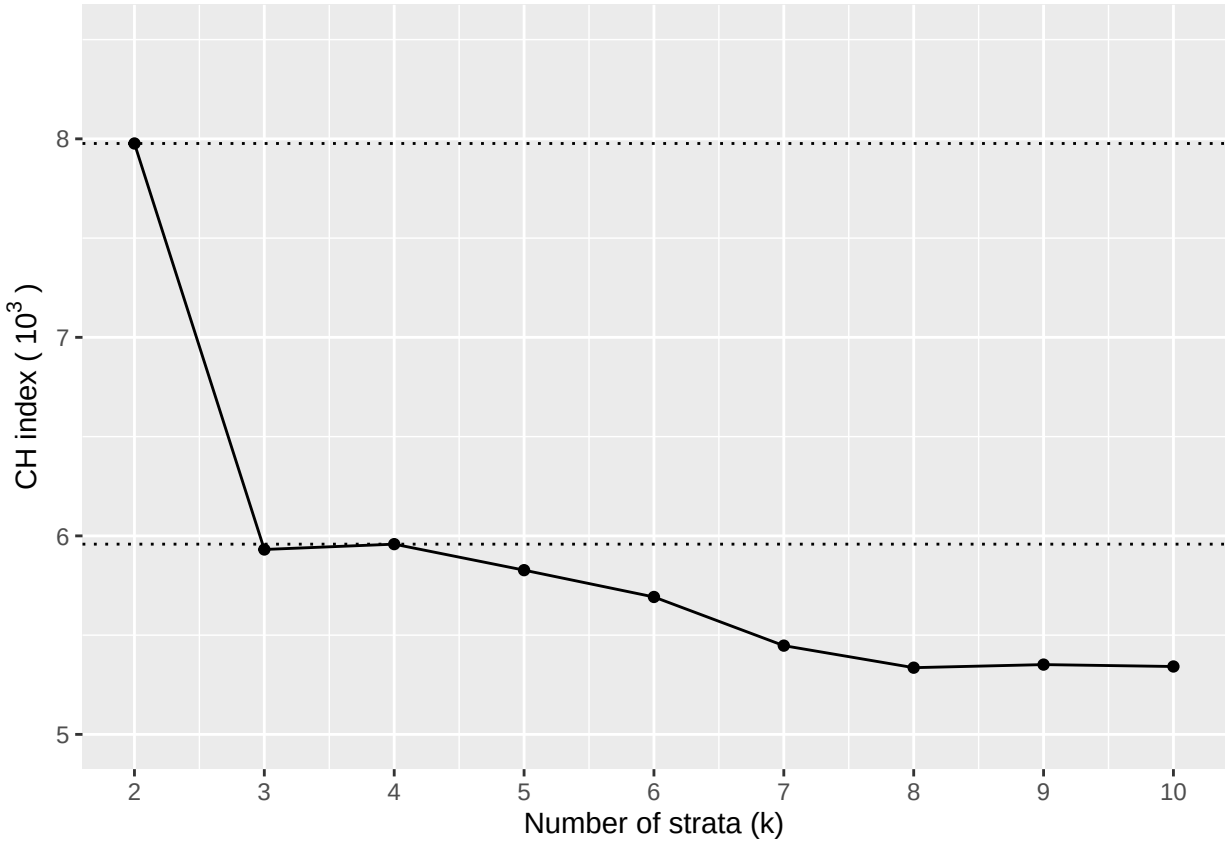
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

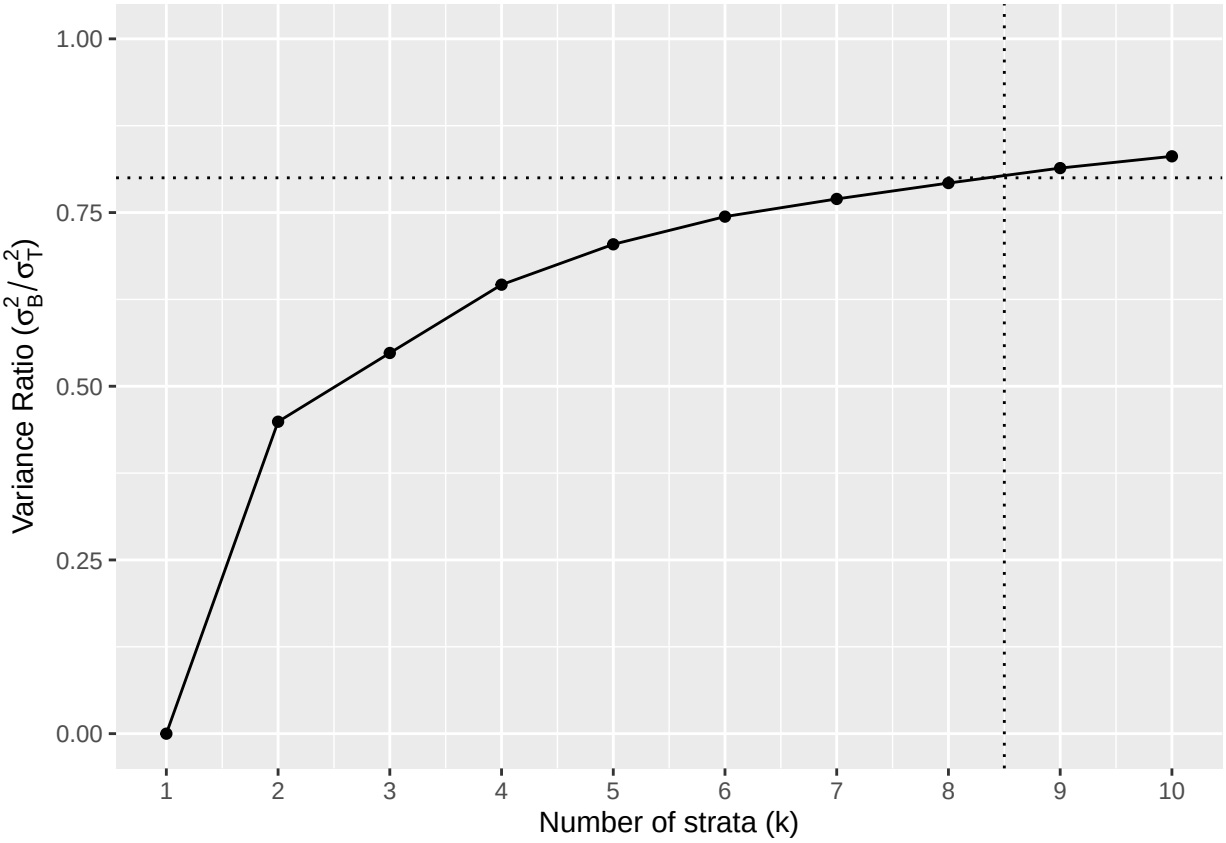


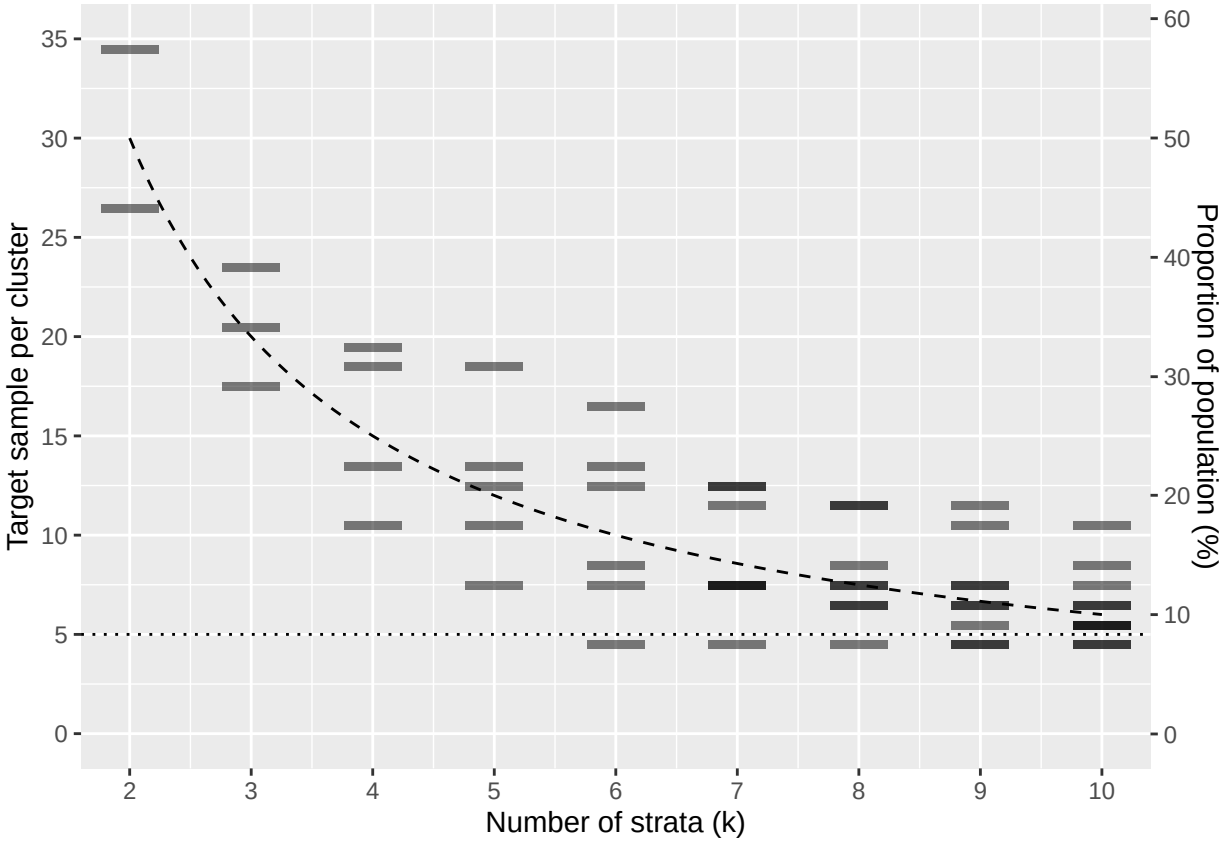
Methods Summary

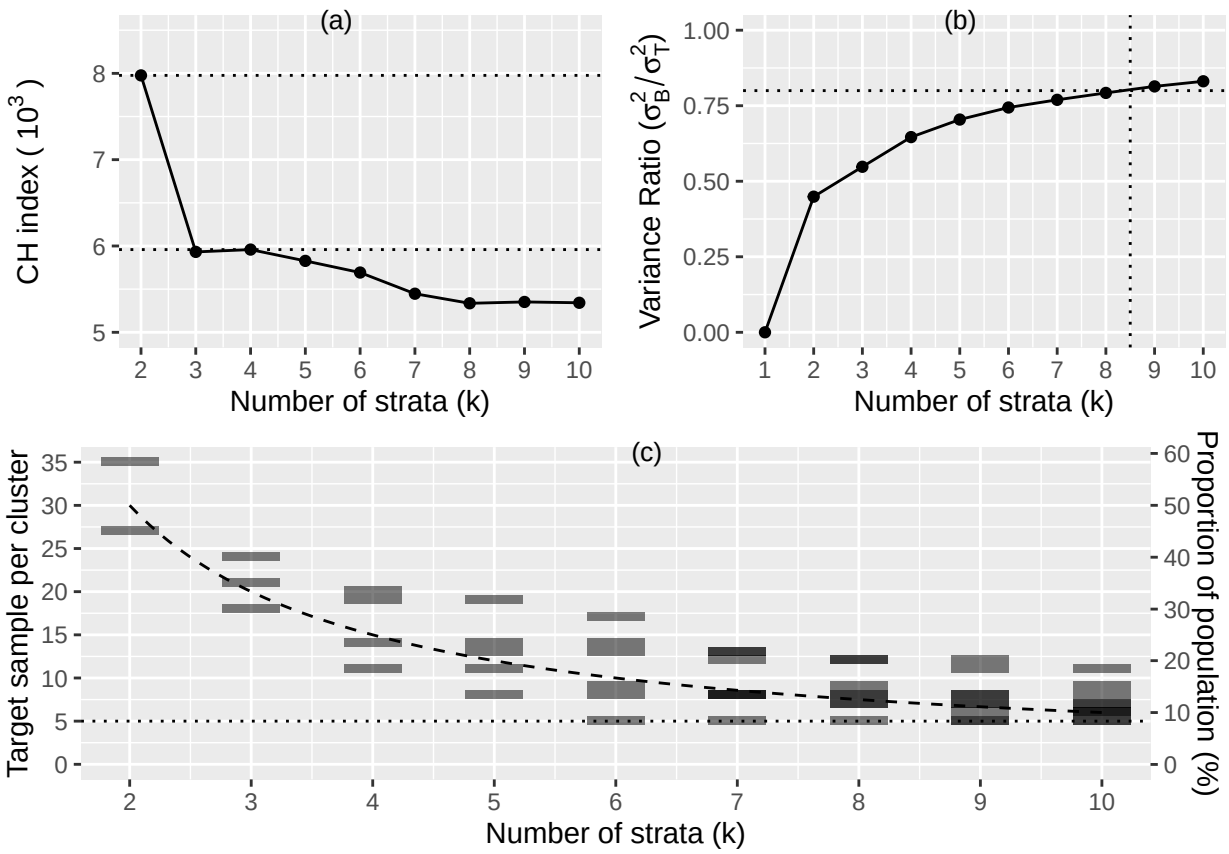
Cluster Analysis

Selecting k.

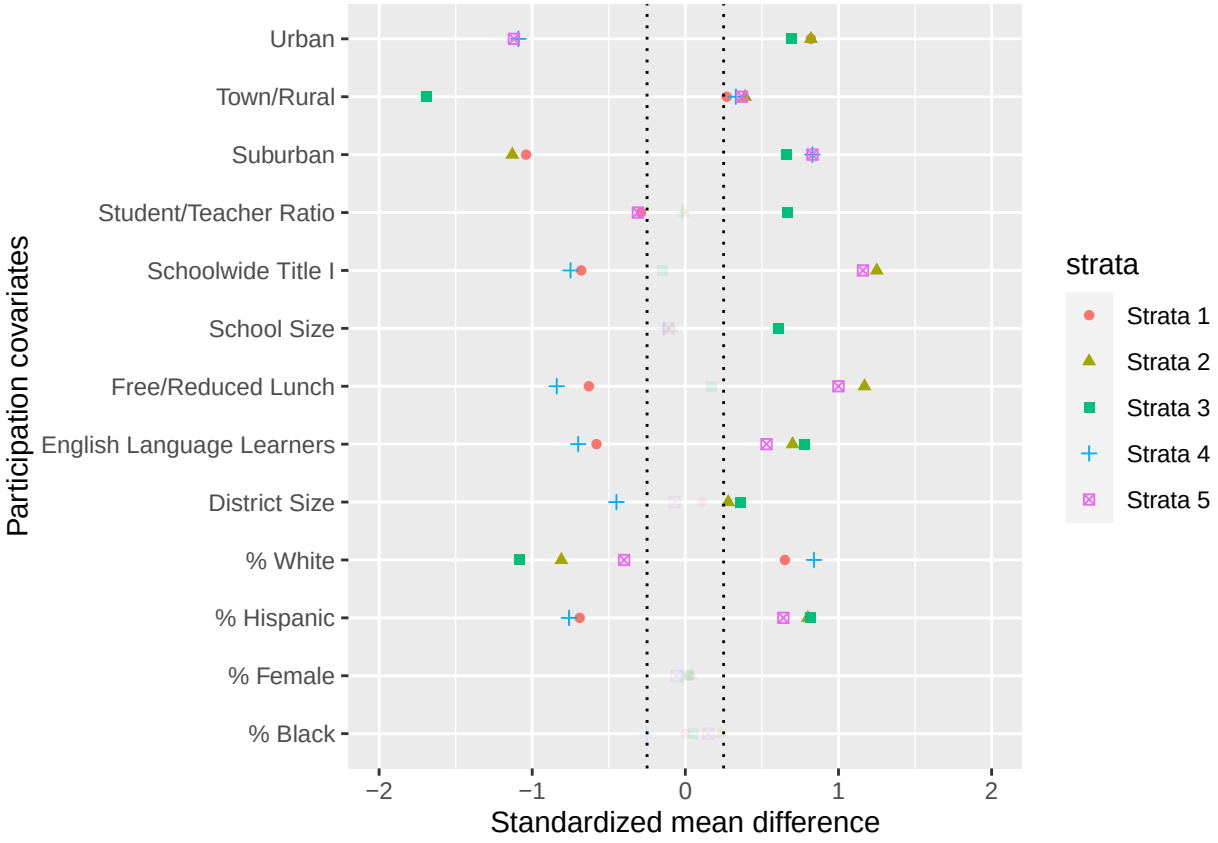



1.00
0.75
0.50
0.25
0.00
Variance Ratio (σ_B^2/σ_T^2)
1
2
3
4
5
6
7
8
9
10
Number of strata (k)


Target sample per cluster
Proportion of population (%)
Number of strata (k)
0
5
10
15
20
25
30
35
0
10
20
30
40
50
60
2
3
4
5
6
7
8
9
10



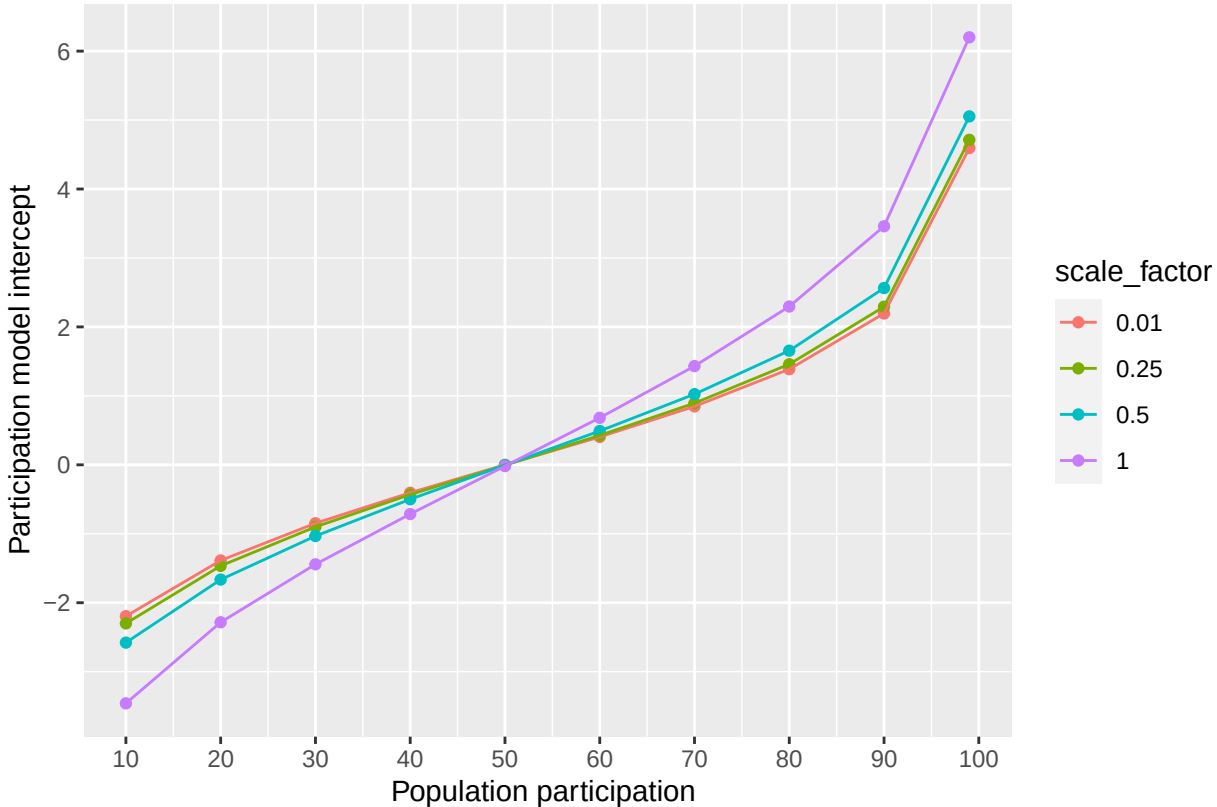

```
## Warning: Using alpha for a discrete variable is not advised.
```


Participation covariates
Urban
Town/Rural
Suburban
Student/Teacher Ratio
Schoolwide Title I
School Size
Free/Reduced Lunch
English Language Learners
District Size
% White
% Hispanic
% Female
% Black
−2
−1
0
1
2
Standardized mean difference
strata
Strata 1
Strata 2
Strata 3
Strata 4
Strata 5

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



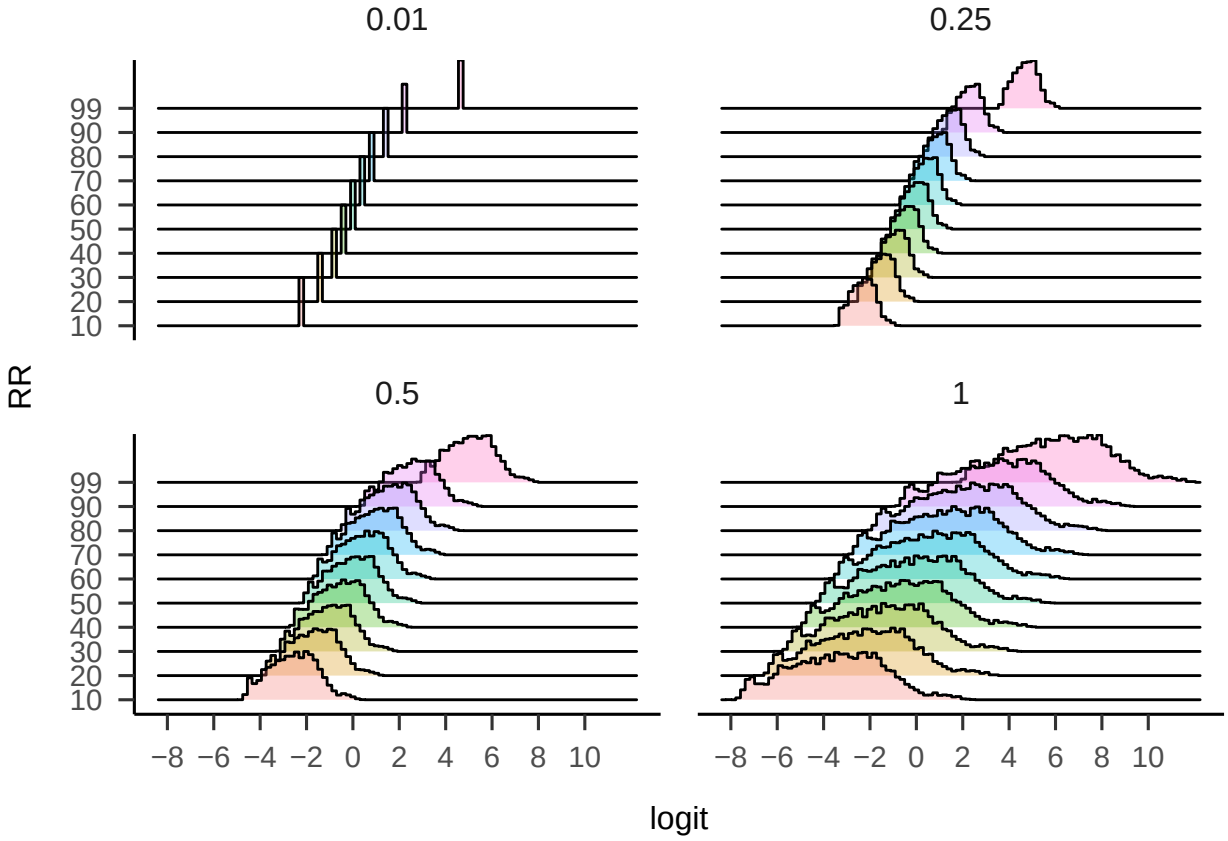


Figure 1. Distributions of Participation Logits

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>.
```

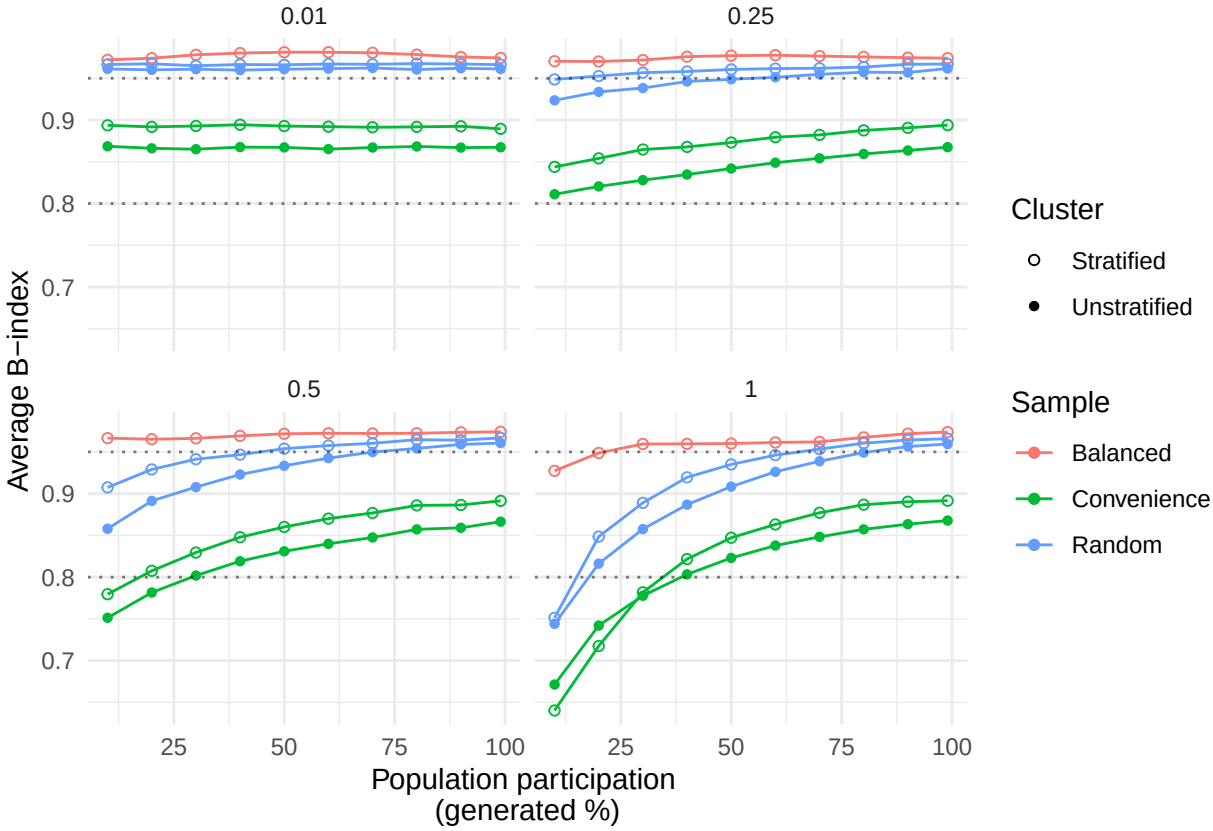
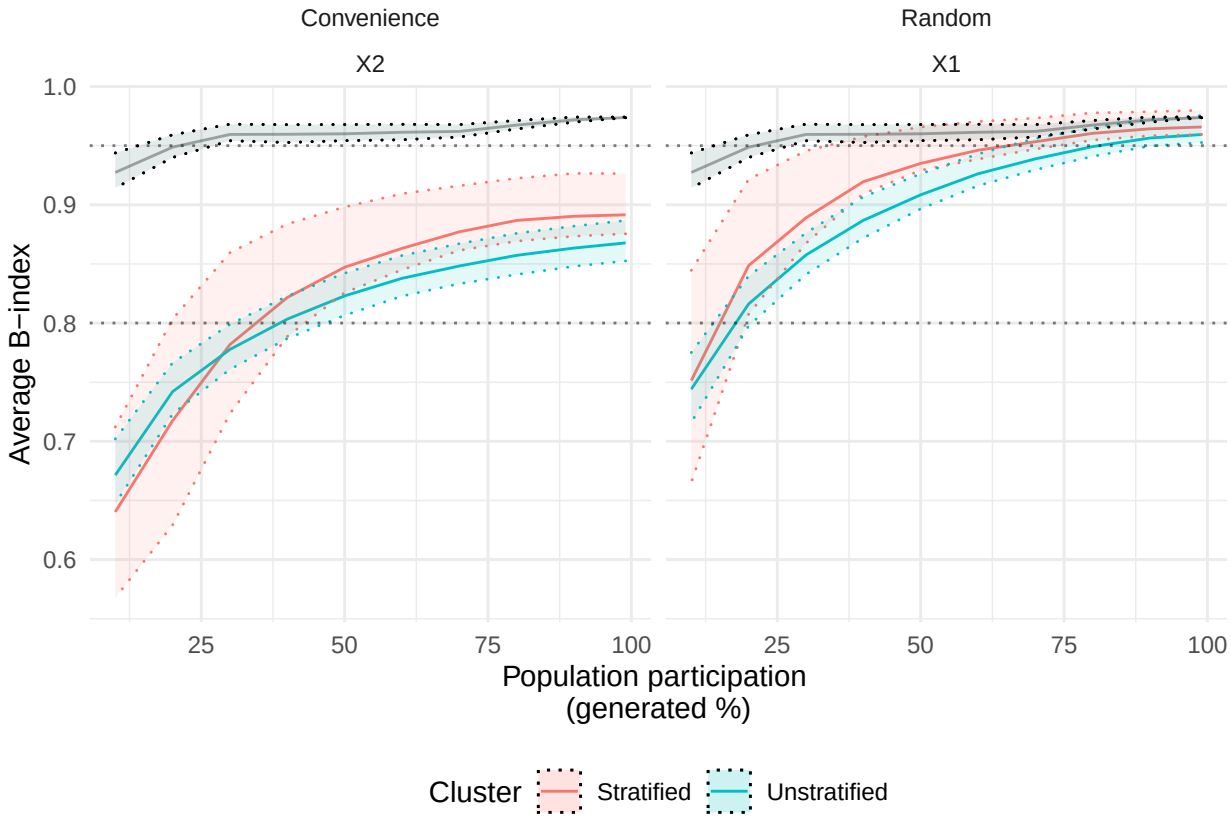


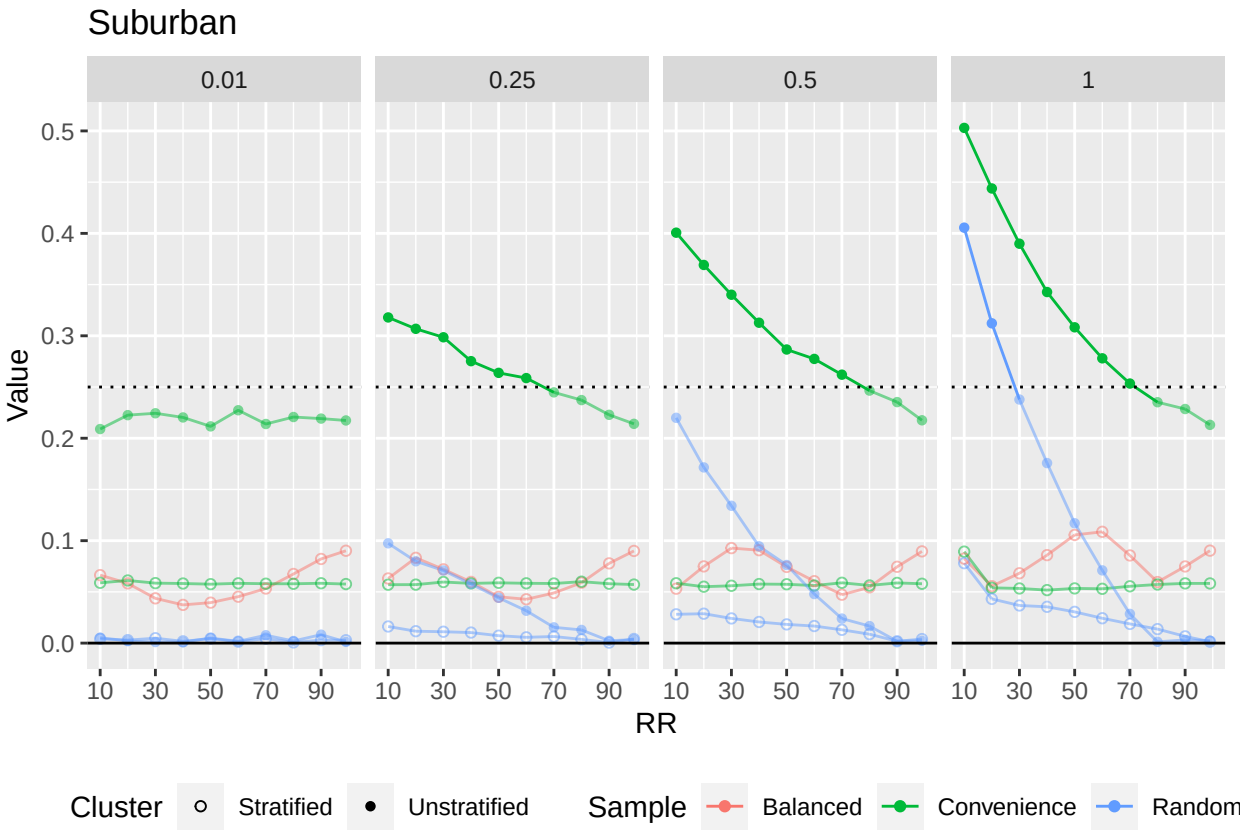
Figure 2. Averge B-Index

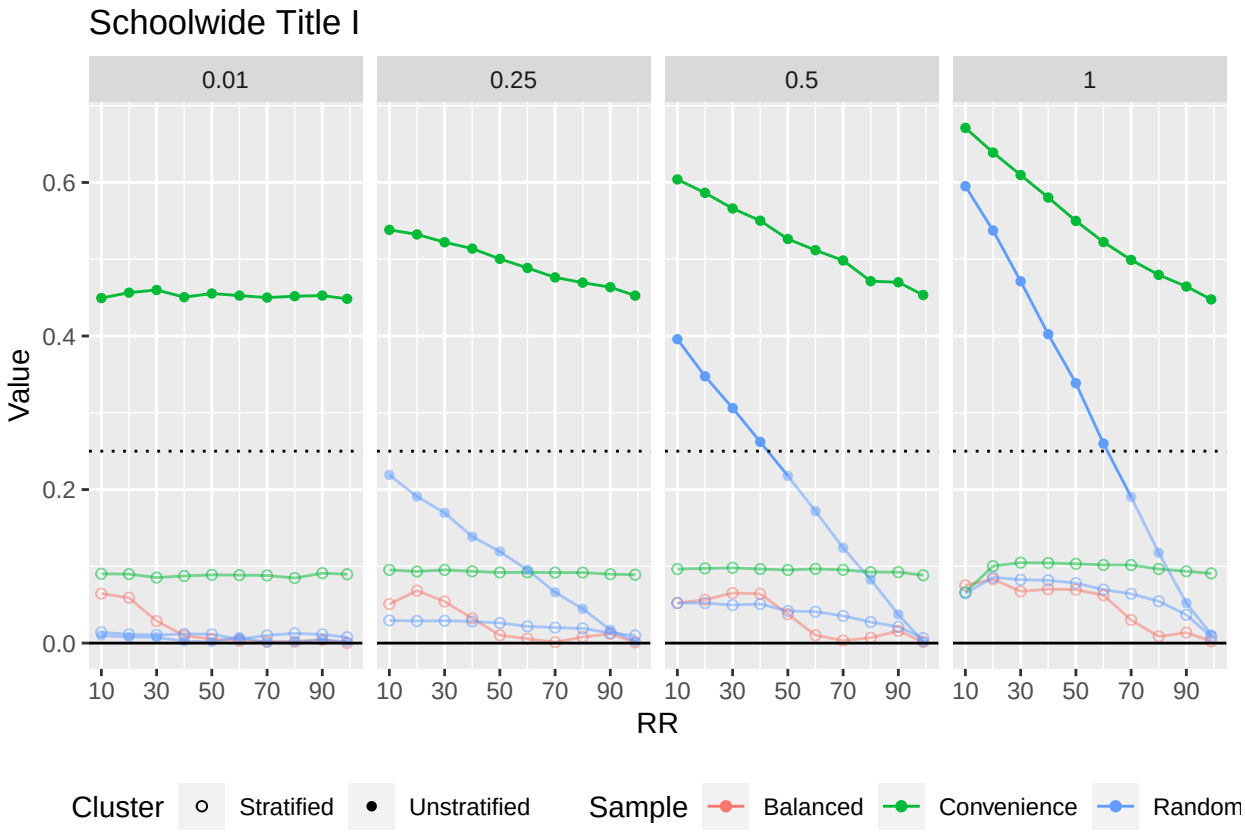
```
## This message is displayed once per session.
```

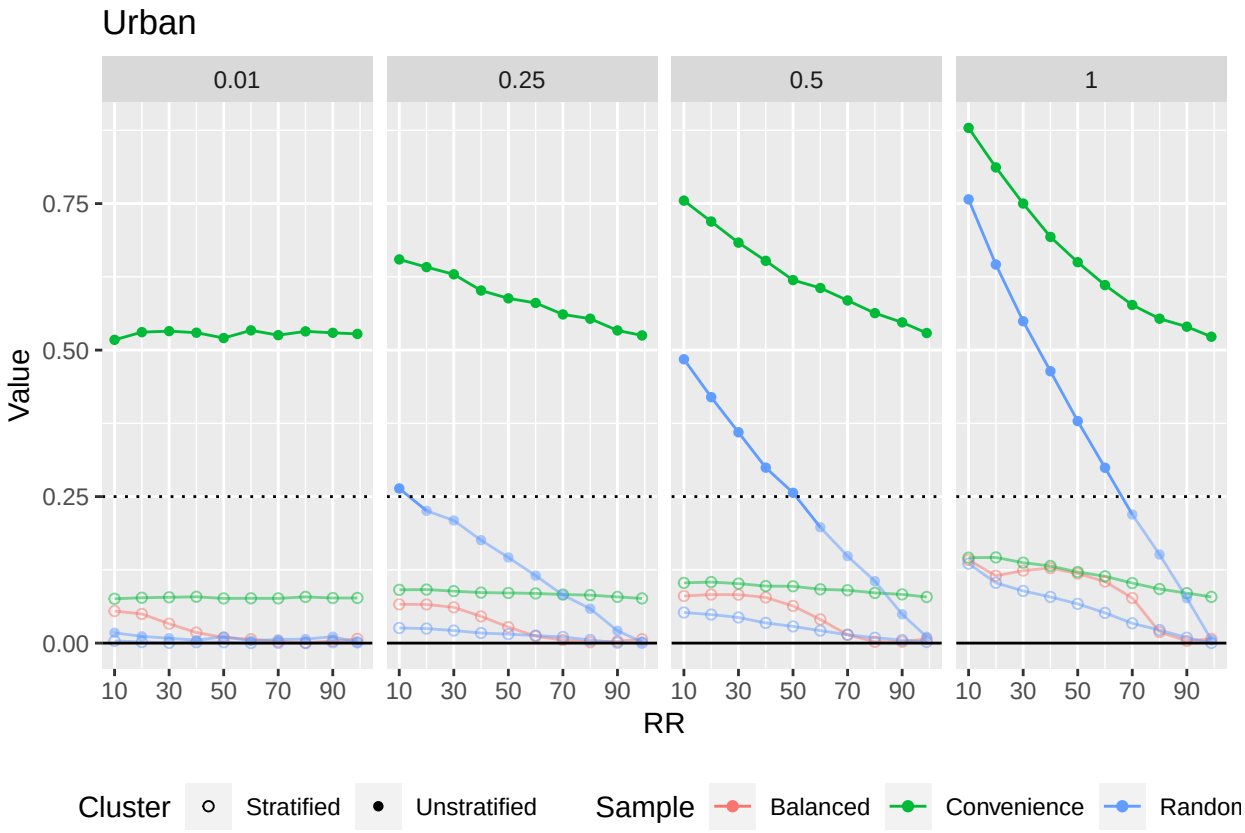


Standardized mean differences.

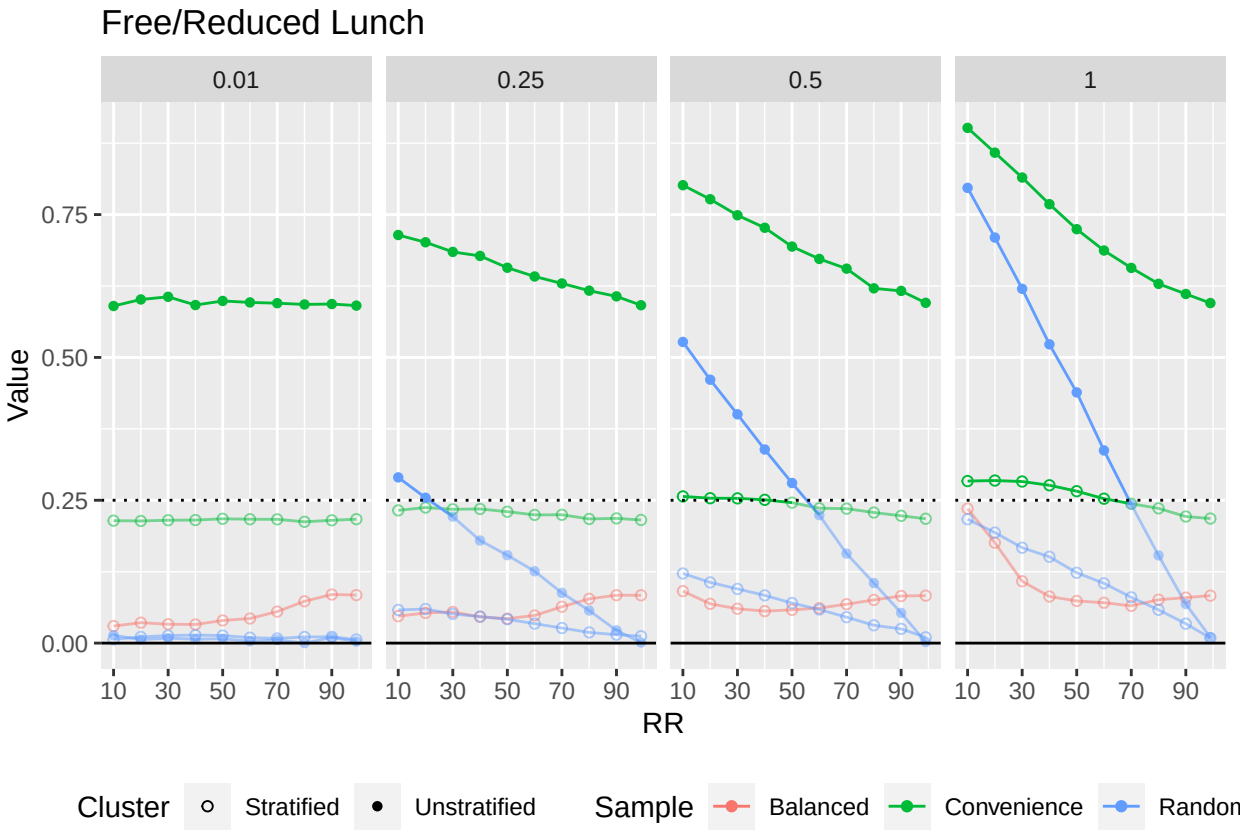
Group 1.

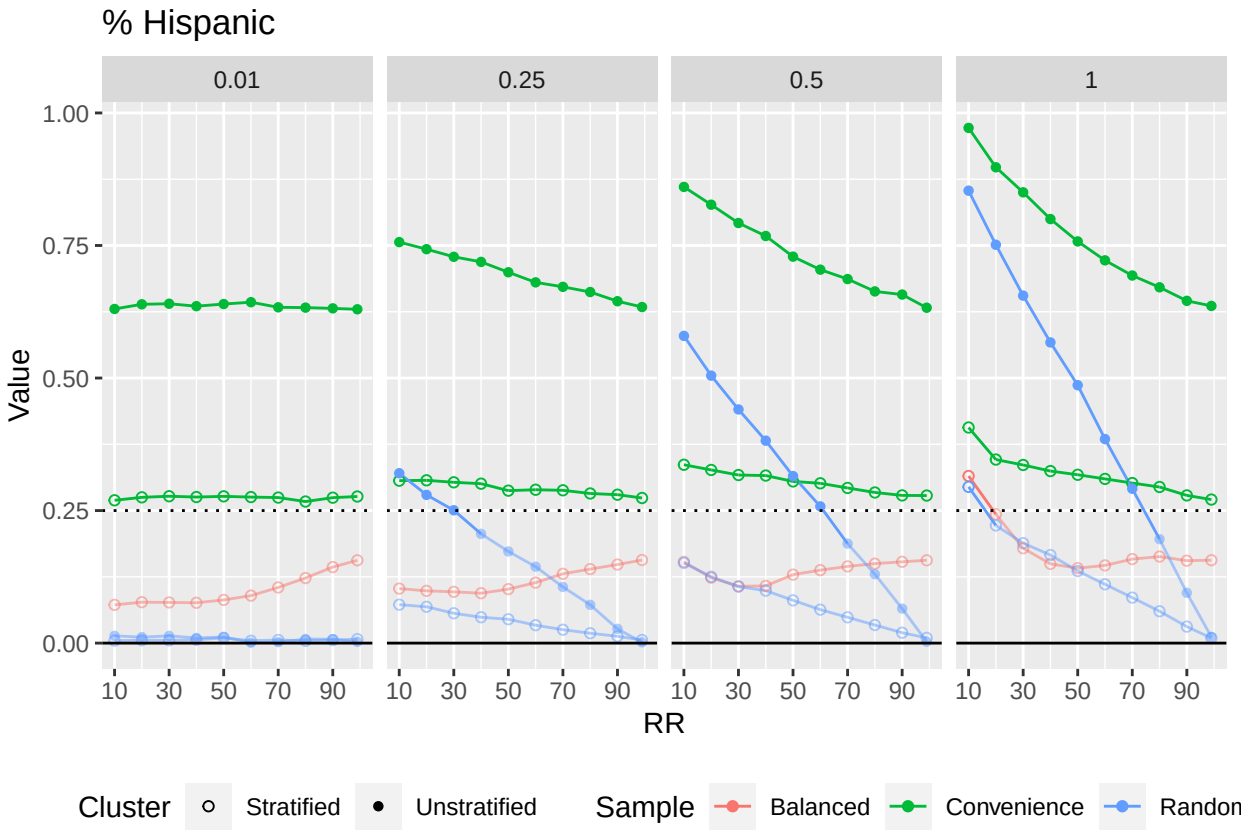

Suburban
0.01
0.25
0.5
1
Value
0.0
0.1
0.2
0.3
0.4
0.5
10
30
50
70
90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

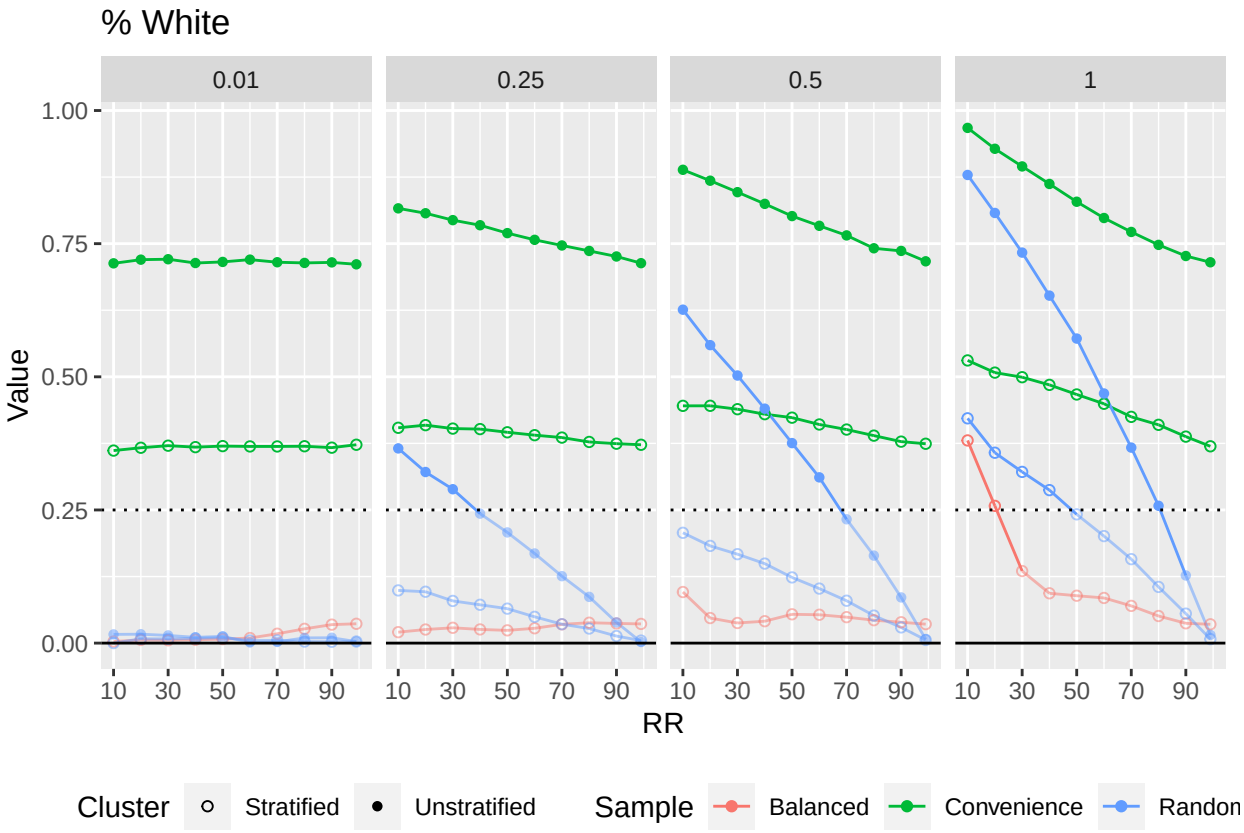

Schoolwide Title I
0.01
0.25
0.5
1
Value
0.0
0.2
0.4
0.6
10
30
50
70
90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

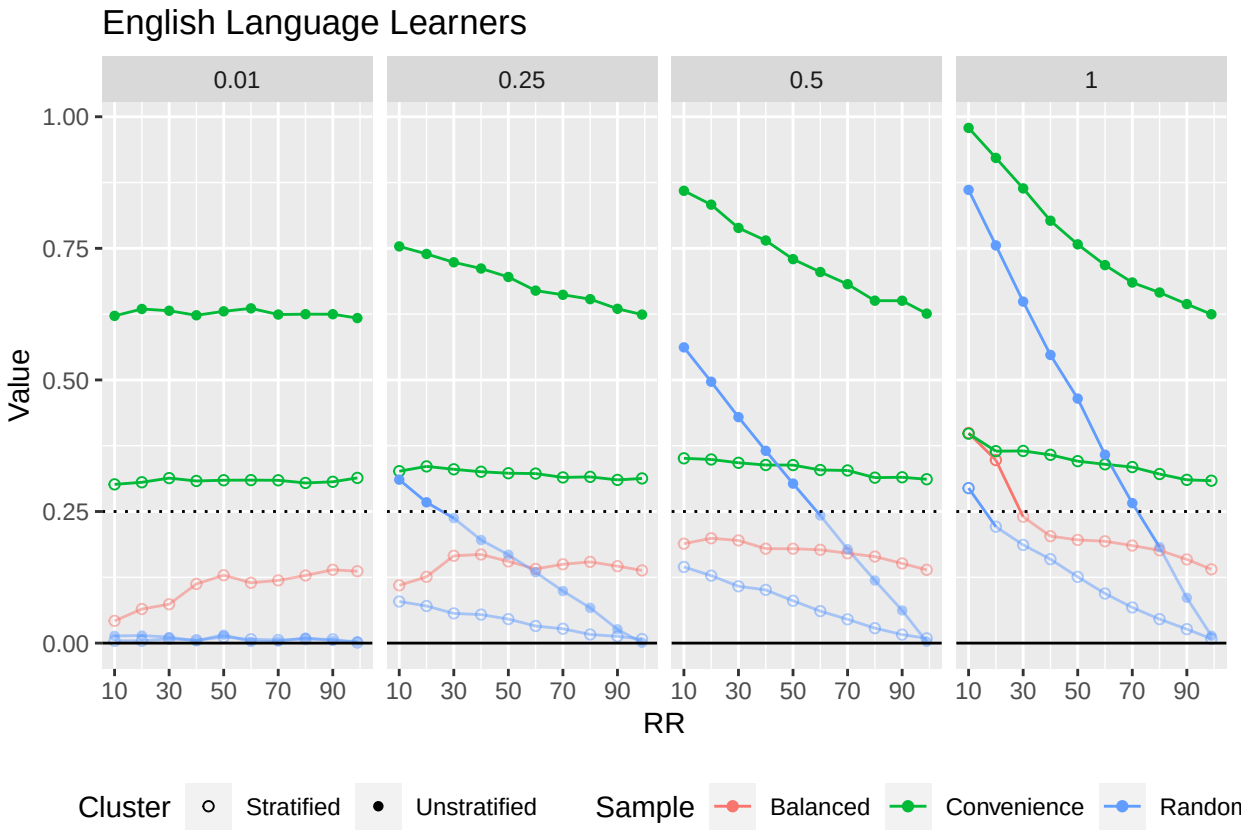

Urban
0.01
0.25
0.5
1
0.75
0.50
0.25
0.00
Value
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Randon

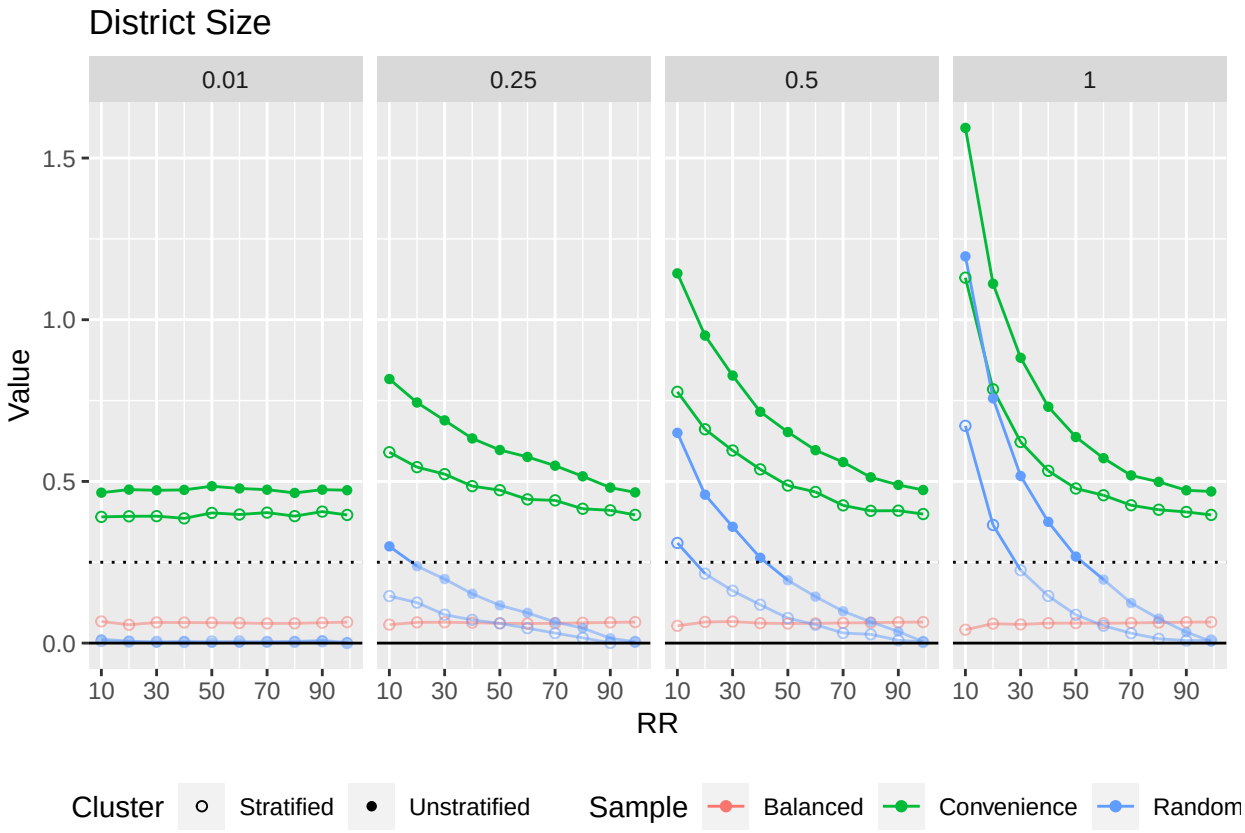
Group 2


Free/Reduced Lunch
0.01
0.25
0.5
1
Value
0.75
0.50
0.25
0.00
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

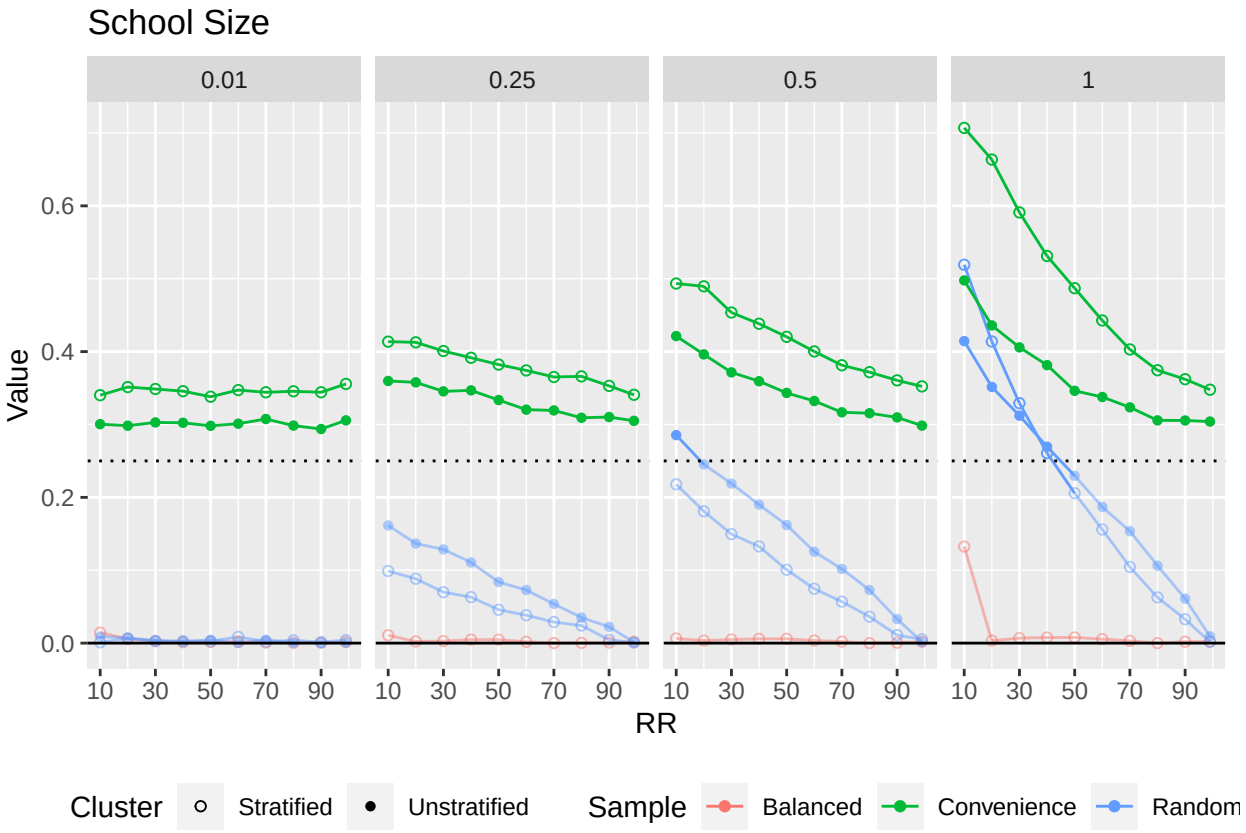

% Hispanic
0.01
0.25
0.5
1
1.00
0.75
0.50
0.25
0.00
Value
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Randon

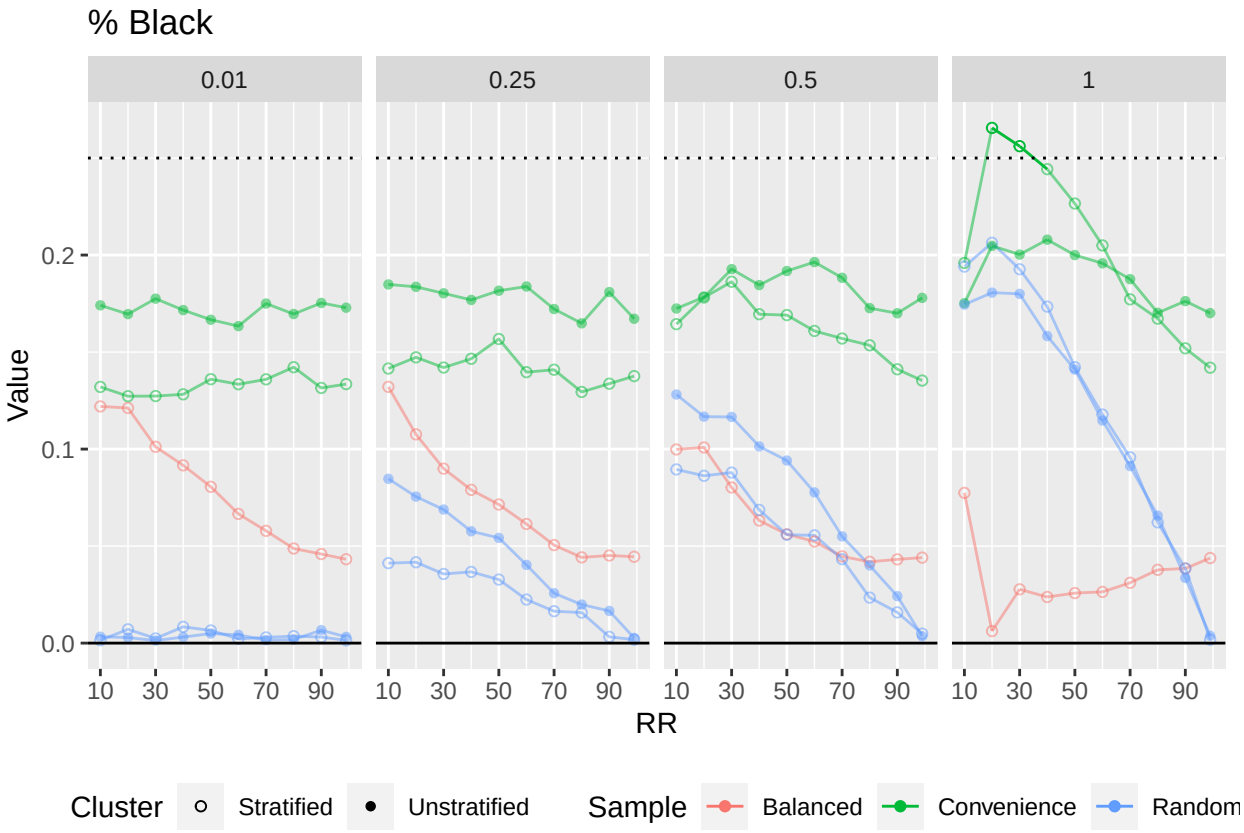

% White
0.01
0.25
0.5
1
1.00
0.75
0.50
0.25
0.00
Value
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

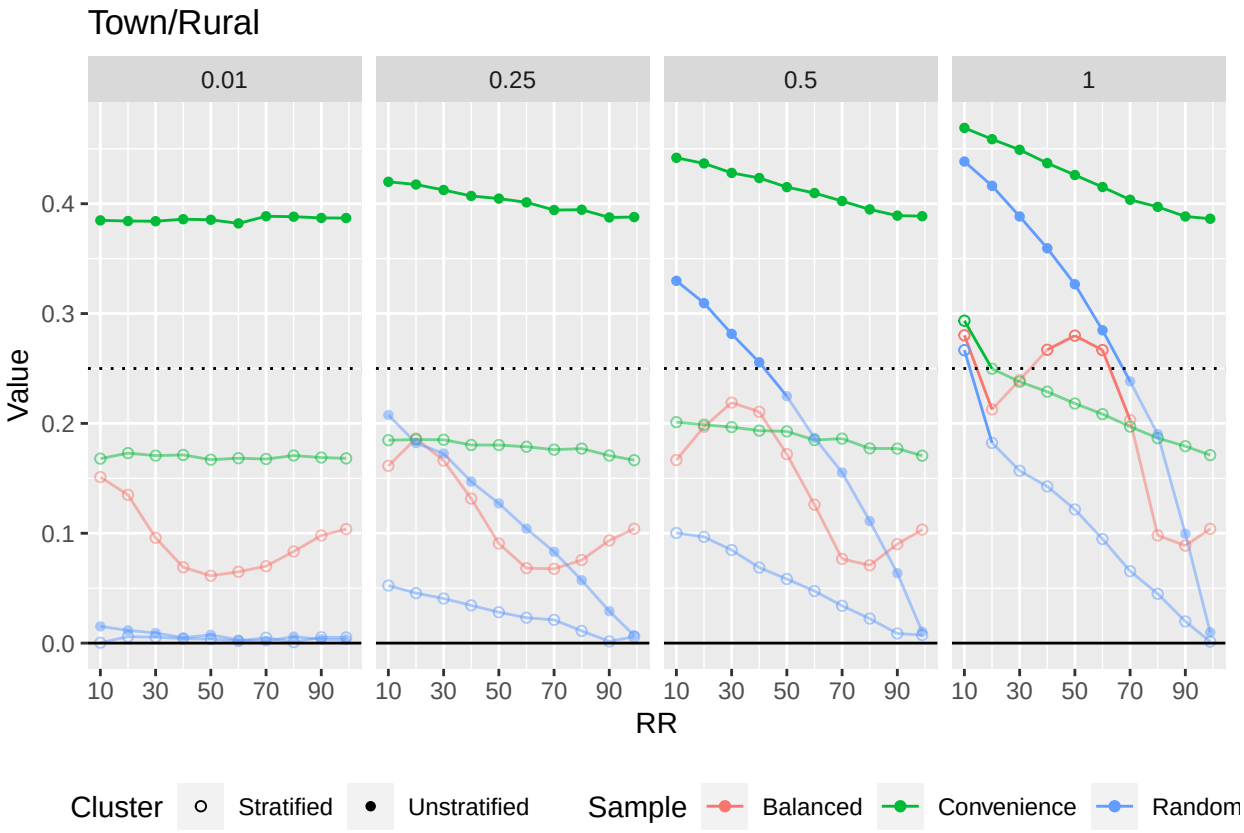

English Language Learners
0.01
0.25
0.5
1
1.00
0.75
0.50
0.25
0.00
Value
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random


District Size
0.01
0.25
0.5
1
Value
1.5
1.0
0.5
0.0
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

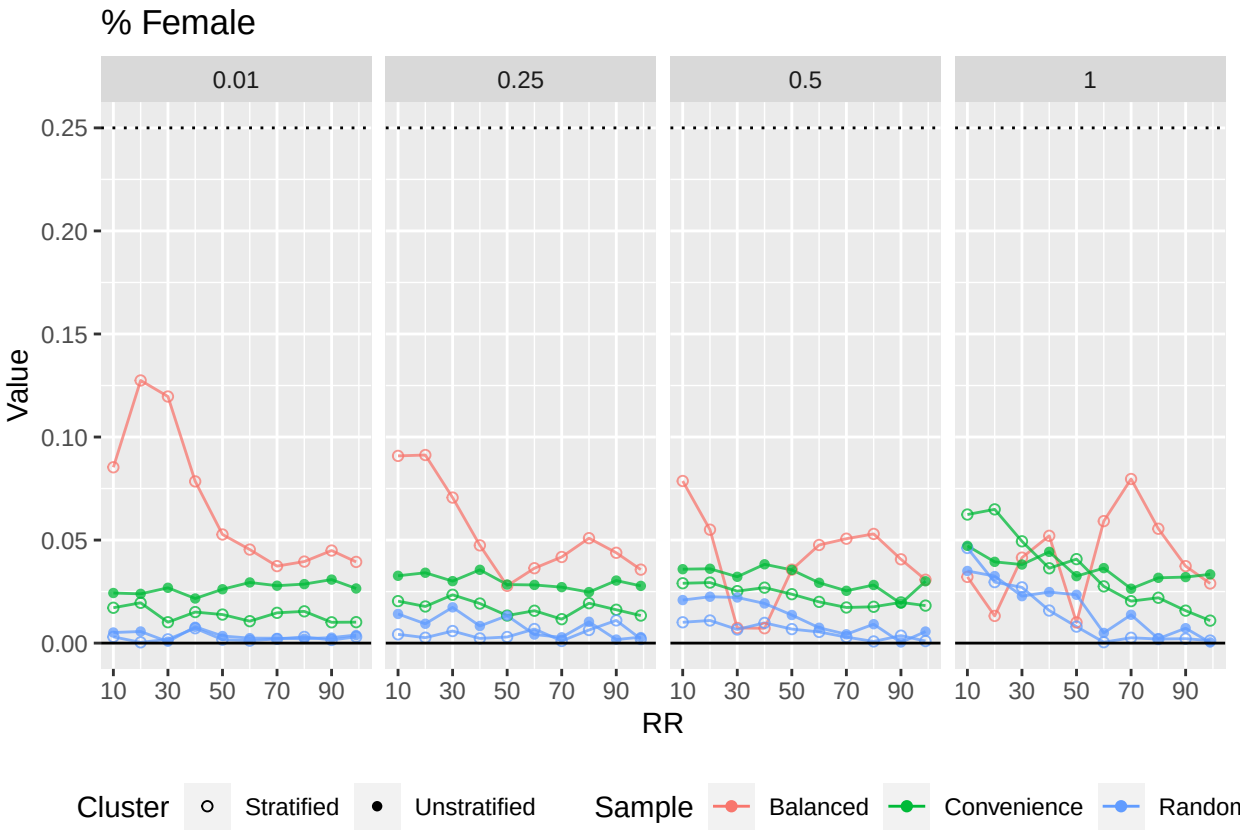
Group 3.

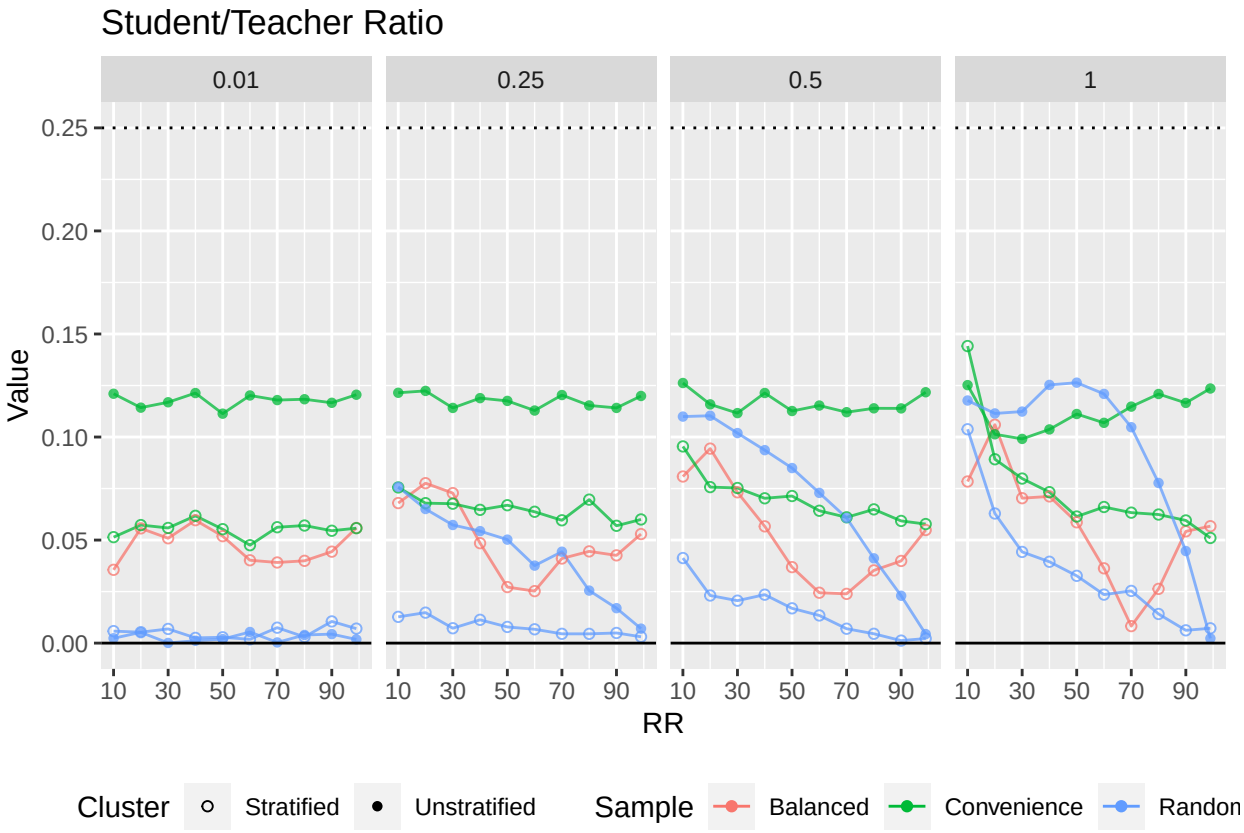

School Size
0.01
0.25
0.5
1
Value
0.0
0.2
0.4
0.6
10
30
50
70
90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

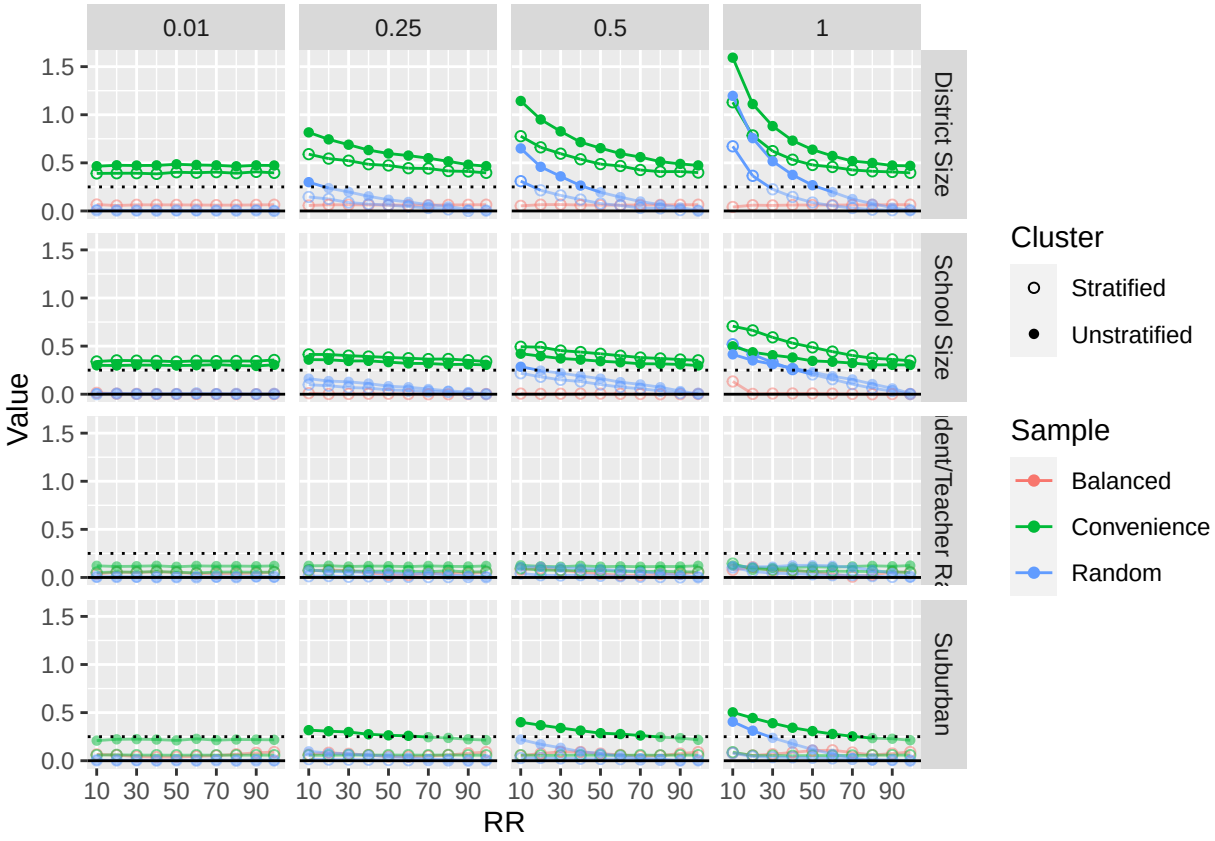

% Black
0.01
0.25
0.5
1
Value
0.2
0.1
0.0
10
30
50
70
90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

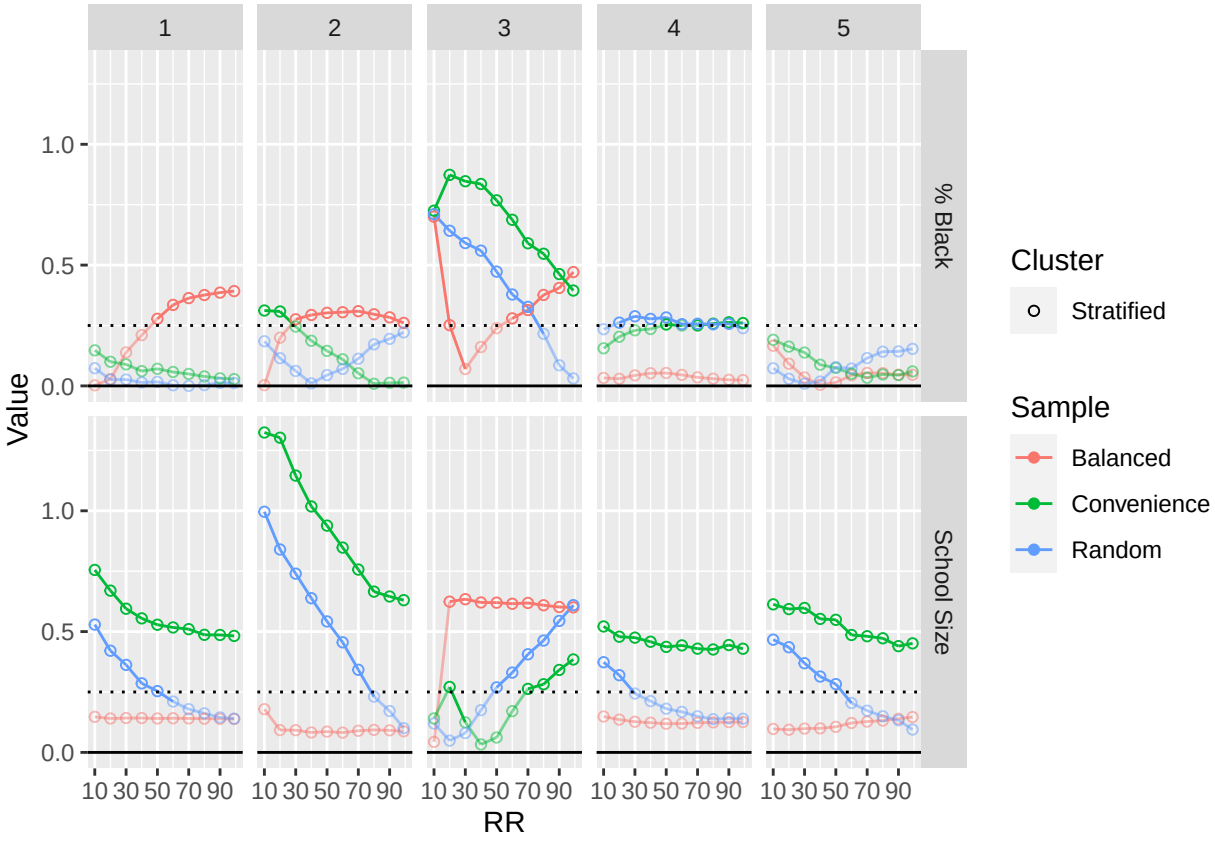

Town/Rural
0.01
0.25
0.5
1
Value
0.0
0.1
0.2
0.3
0.4
10
30
50
70
90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random

Group 4.


% Female
0.01
0.25
0.5
1
0.25
0.20
0.15
0.10
0.05
0.00
Value
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Randon


Student/Teacher Ratio
0.01
0.25
0.5
1
Value
0.25
0.20
0.15
0.10
0.05
0.00
10 30 50 70 90
RR
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Randon


0.01
0.25
0.5
1
District Size
School Size
Suburban
Value
RR
10 30 50 70 90
0.0 0.5 1.0 1.5
Cluster
Stratified
Unstratified
Sample
Balanced
Convenience
Random



Examples for presentations

V-ratio and Log odds.

Feasibility

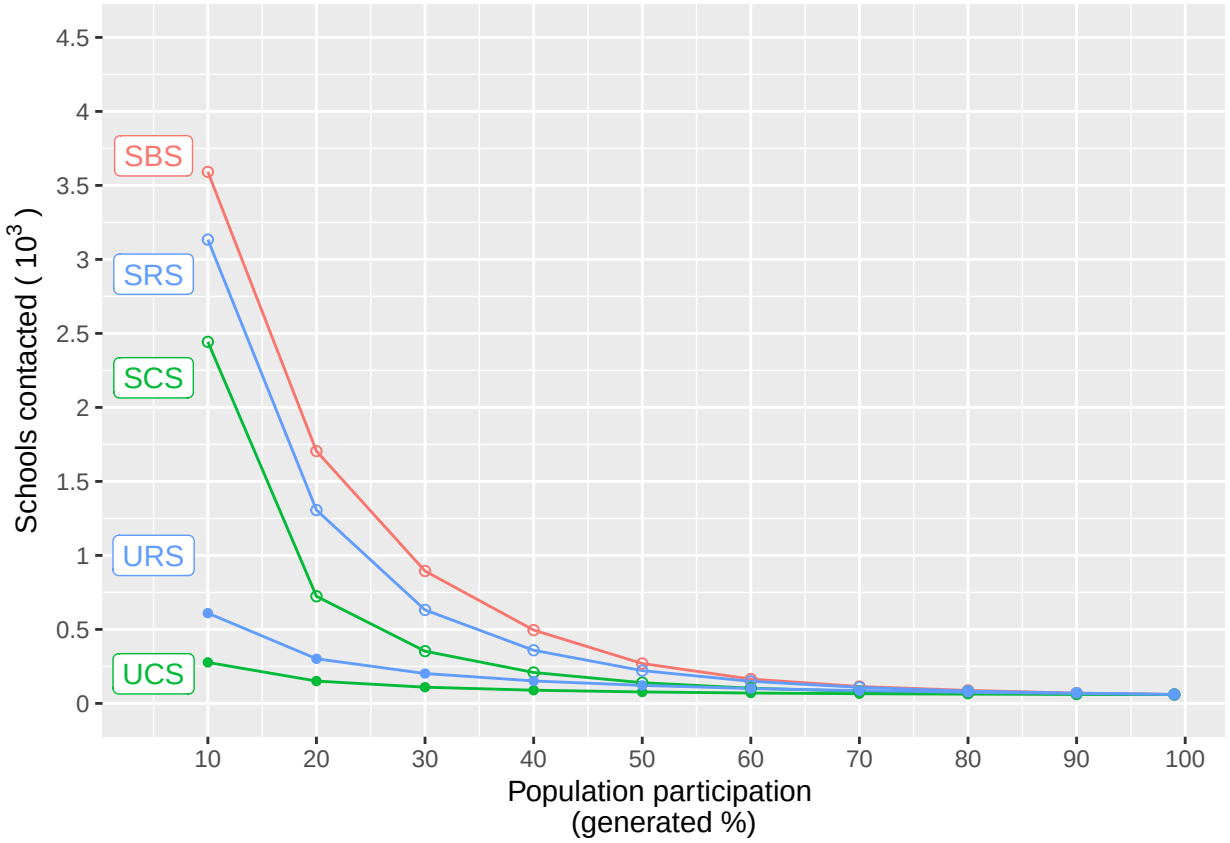
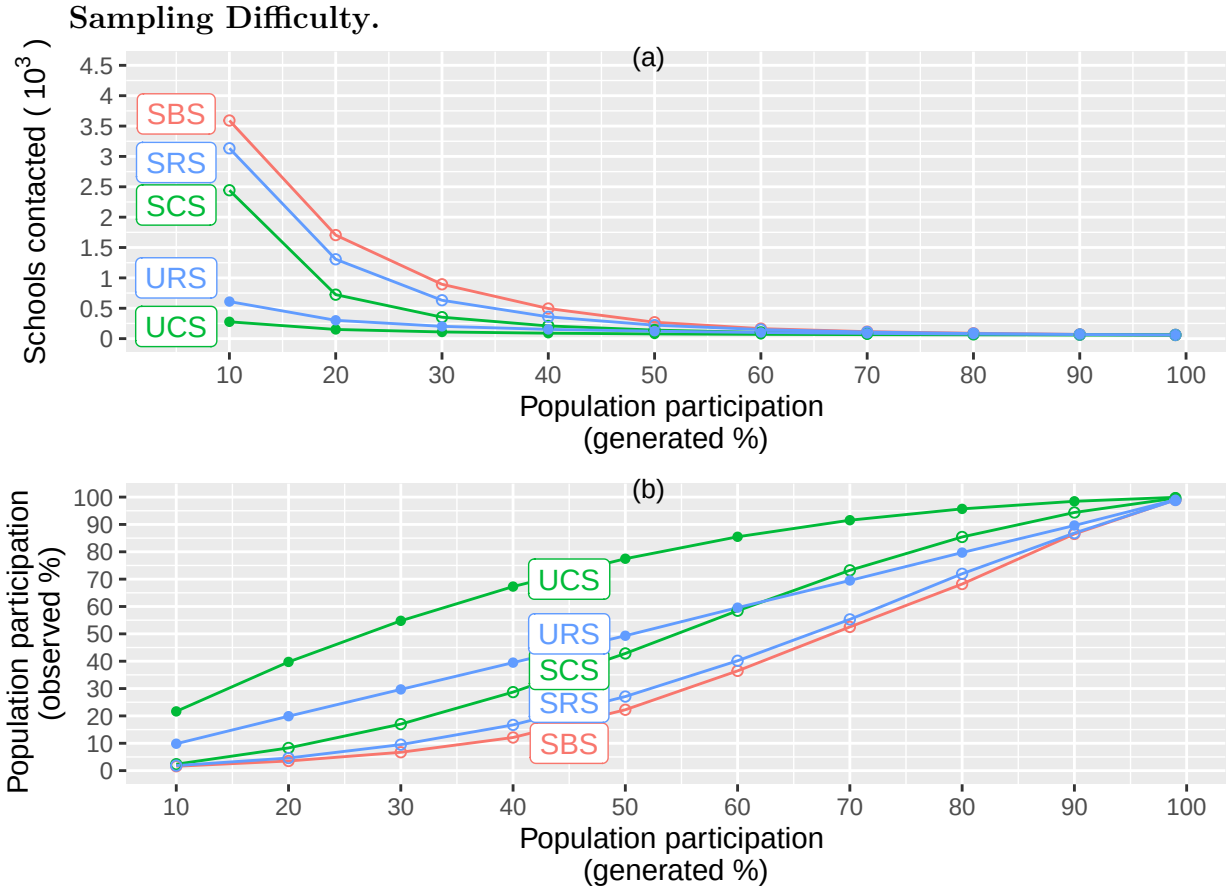


Figure 3. Schools Contacted



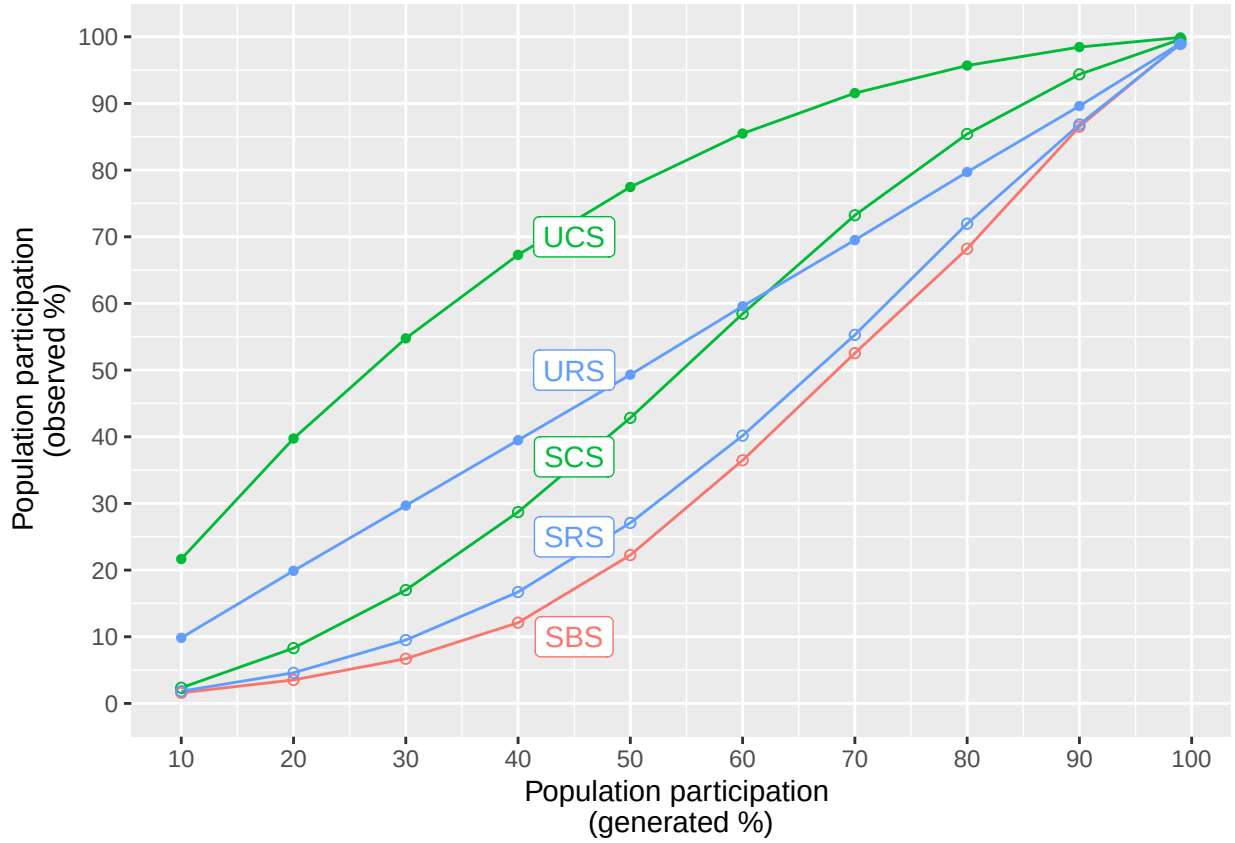
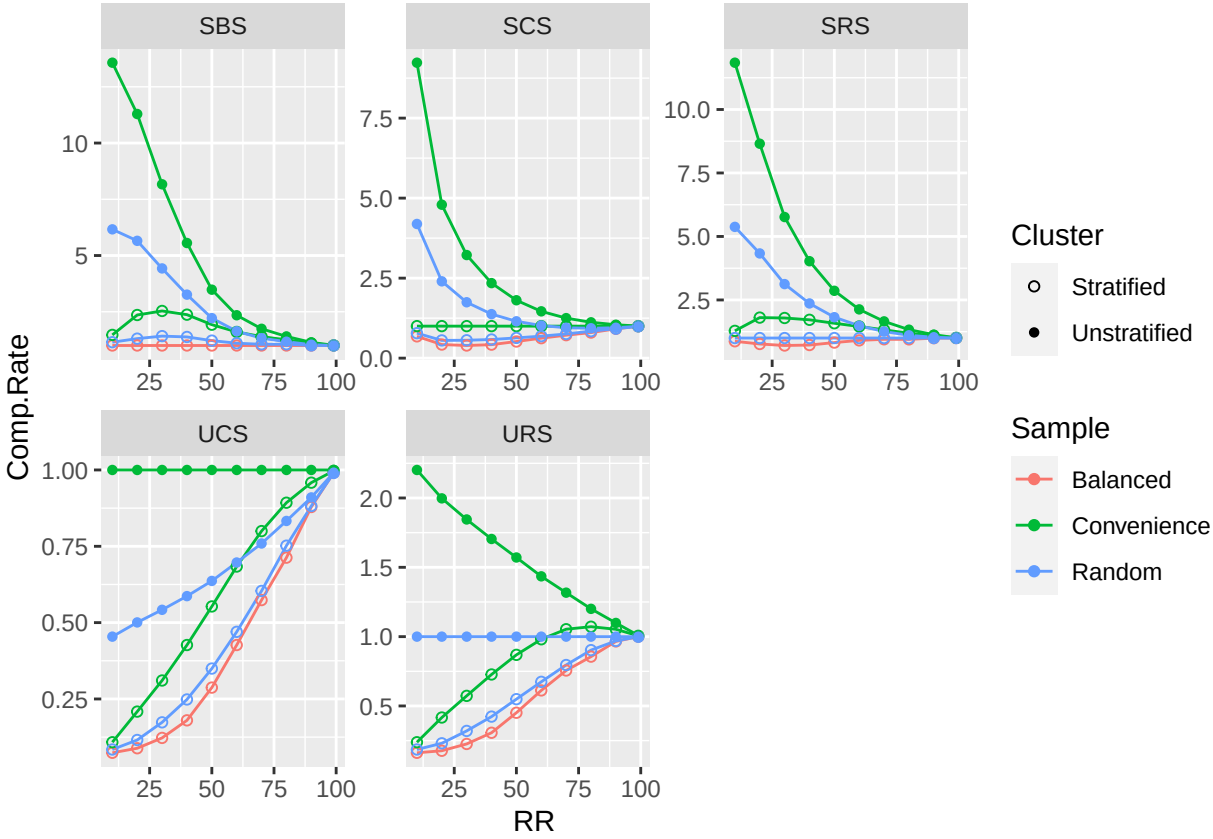
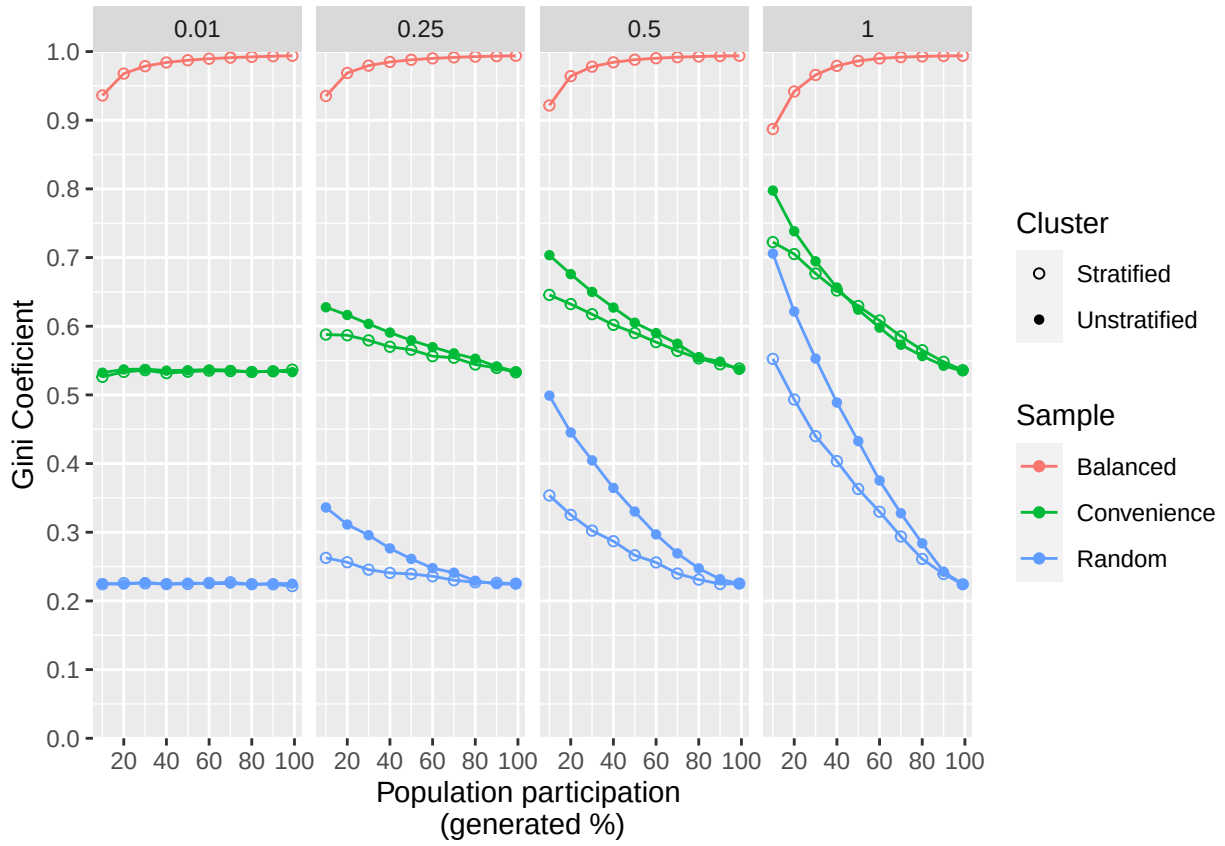


Figure 4. Sampling response rates

Relative Performance.



Gini Plot.



A tibble: 22 x 2

Vars MB

1 tabs 473.73 Mb 2 figs 301.579 Mb 3 results 84.955 Mb 4 df.list 21.411 Mb 5 df.sim 3.027 Mb

6 legend 1.161 Mb

7 smd.examples.sep 0.214 Mb

8 plot_smd 0.077 Mb

9 df.smd.ipsw 0.038 Mb

10 get_quant 0.03 Mb

... with 12 more rows