

GenSamp: RESULTS

Gleb Furman¹ & James E. Pustejovsky¹

¹ University of Texas at Austin

GenSamp: RESULTS

Setup**Packages and Data****Organize Objects**

```
## Joining, by = "DSID"
```

```
## Parsed with column specification:
```

```
## cols(  
##   vnames = col_character(),  
##   Variables = col_character(),  
##   Sub = col_character(),  
##   Category = col_character(),  
##   Type = col_character()  
## )
```

```
## Parsed with column specification:
```

```
## cols(  
##   vnames = col_character(),  
##   Variables = col_character(),  
##   Sub = col_character(),  
##   Category = col_character(),  
##   Type = col_character()  
## )
```

```
## Joining, by = "vnames"
```

Data Summary

Covaraite Statistics

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

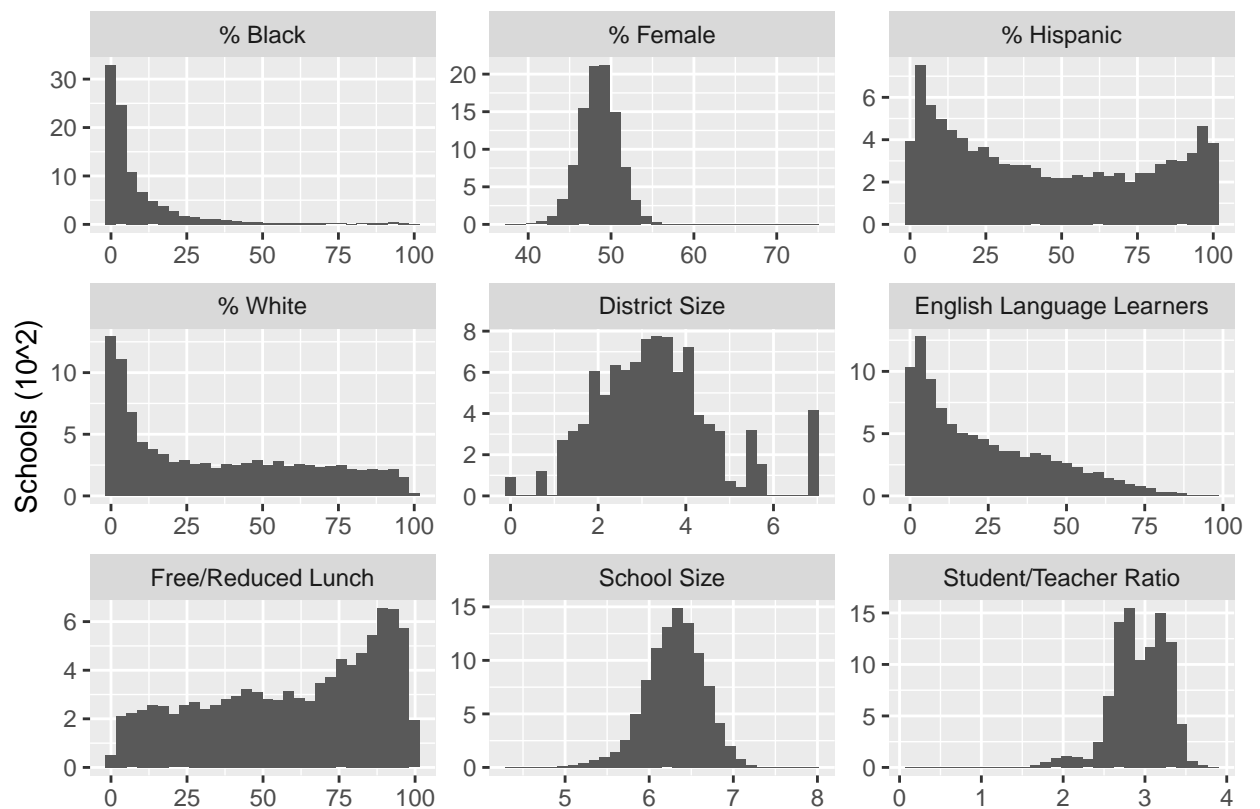
```
## Joining, by = "vnames"
```

```
## Joining, by = c("vnames", "Variables", "Sub", "Category", "Type")
```

Continuous variable distributions

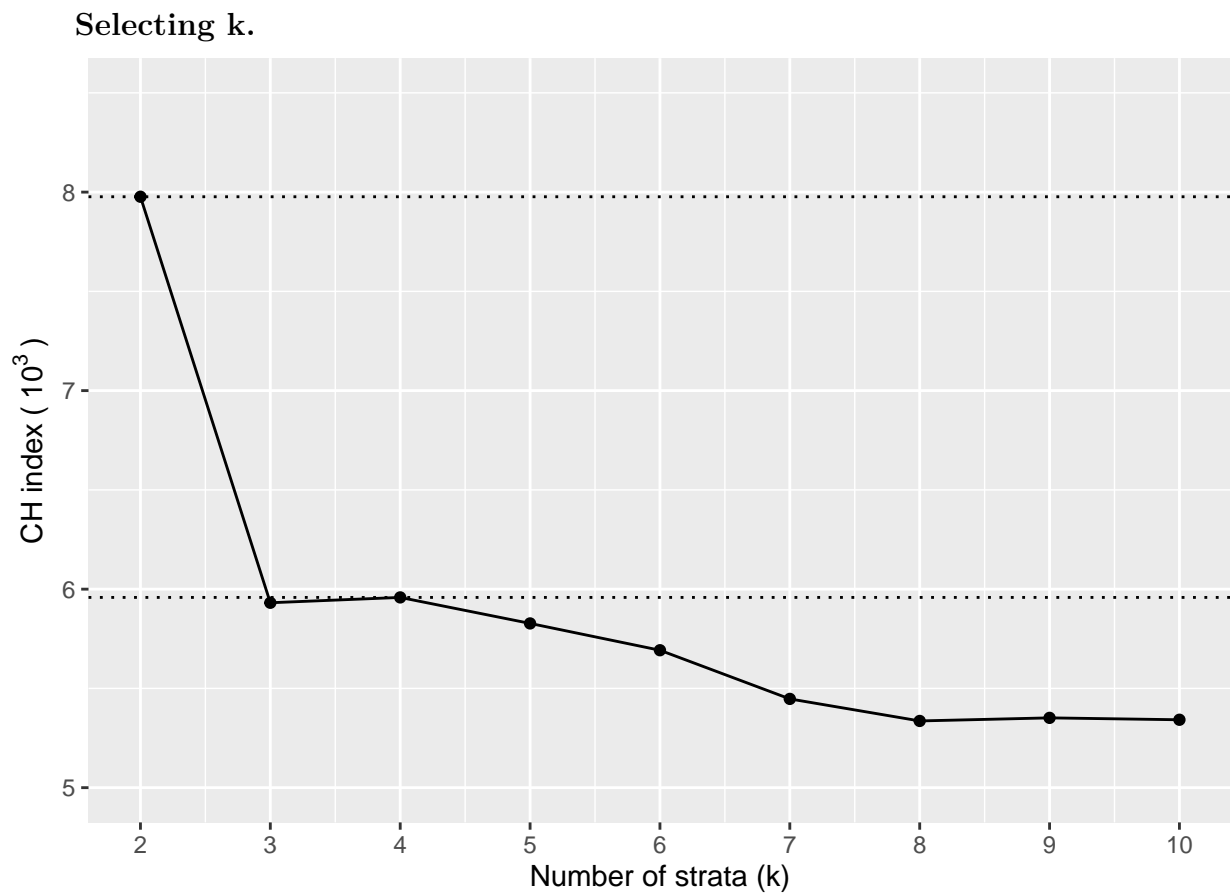
```
## Joining, by = "vnames"
```

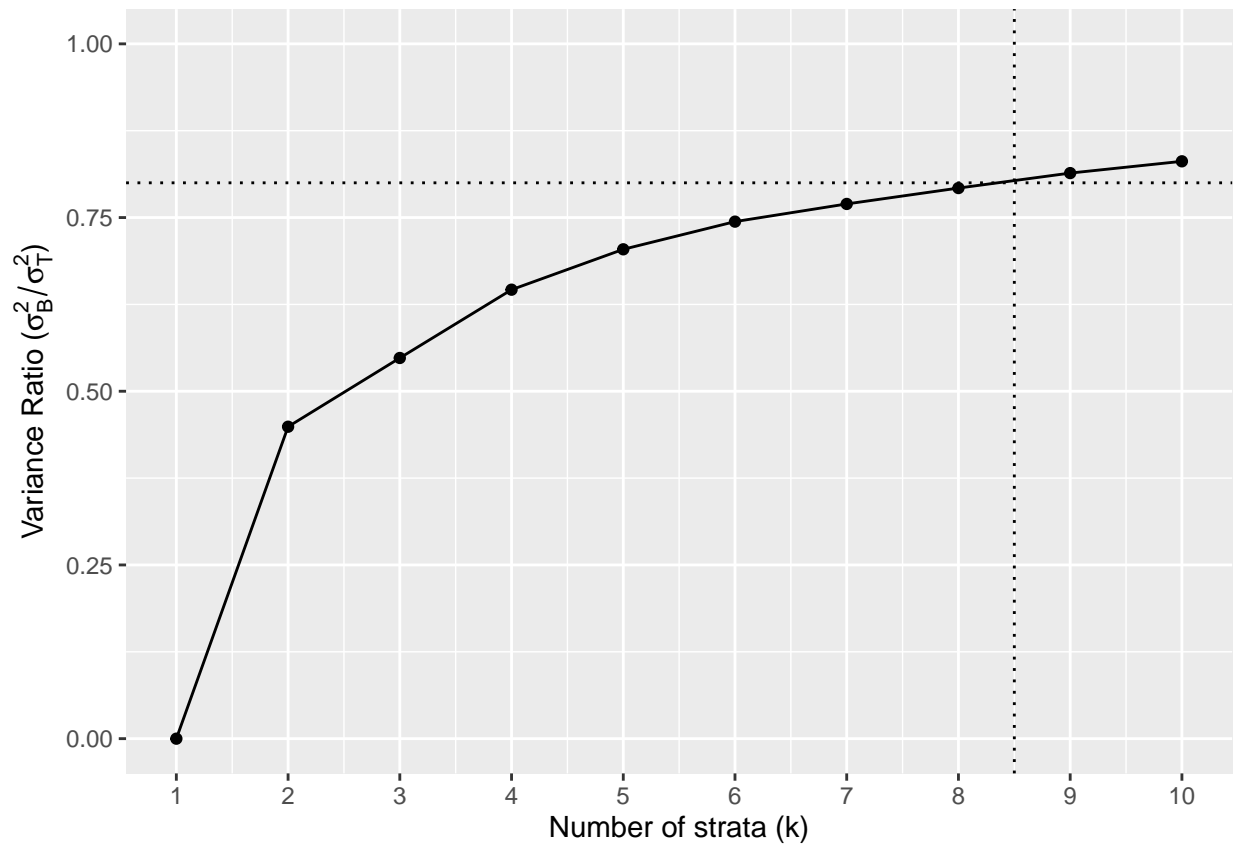
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



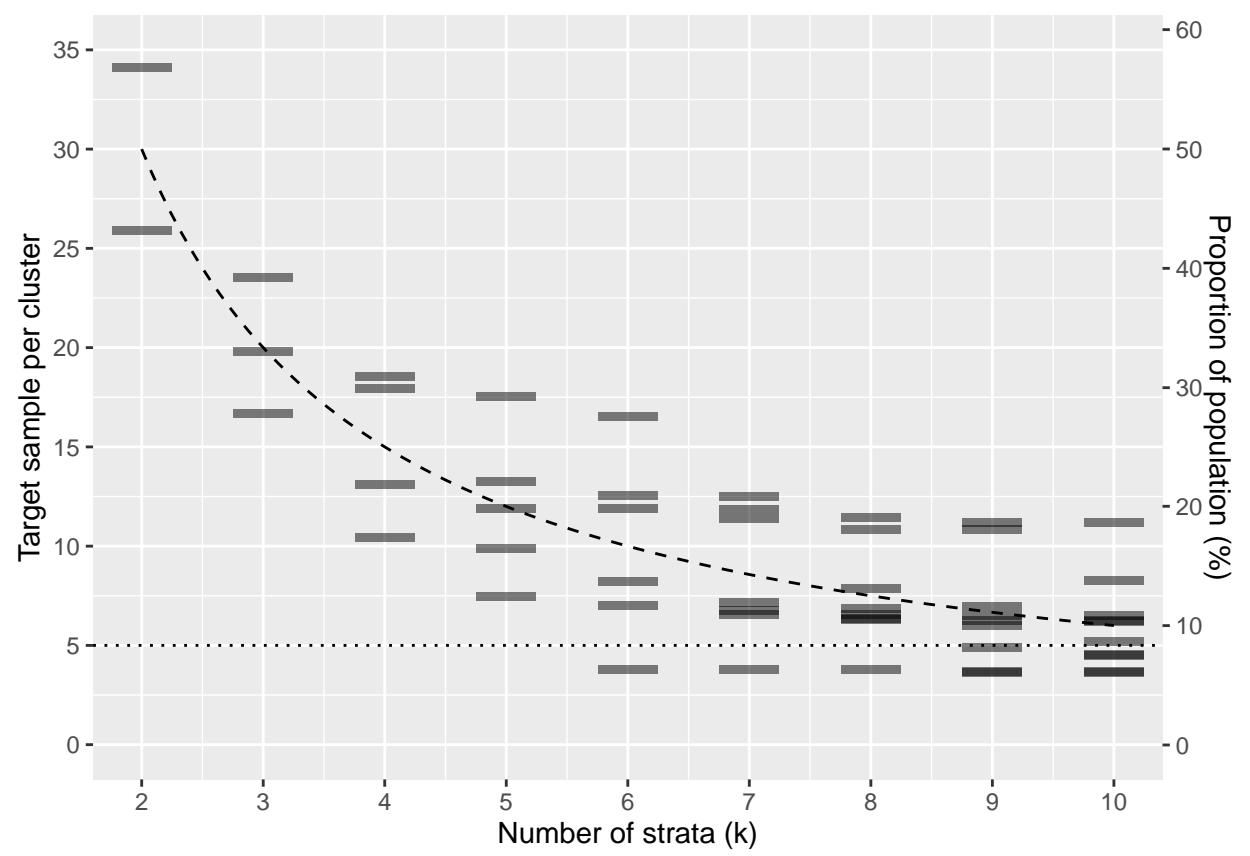
Methods Summary

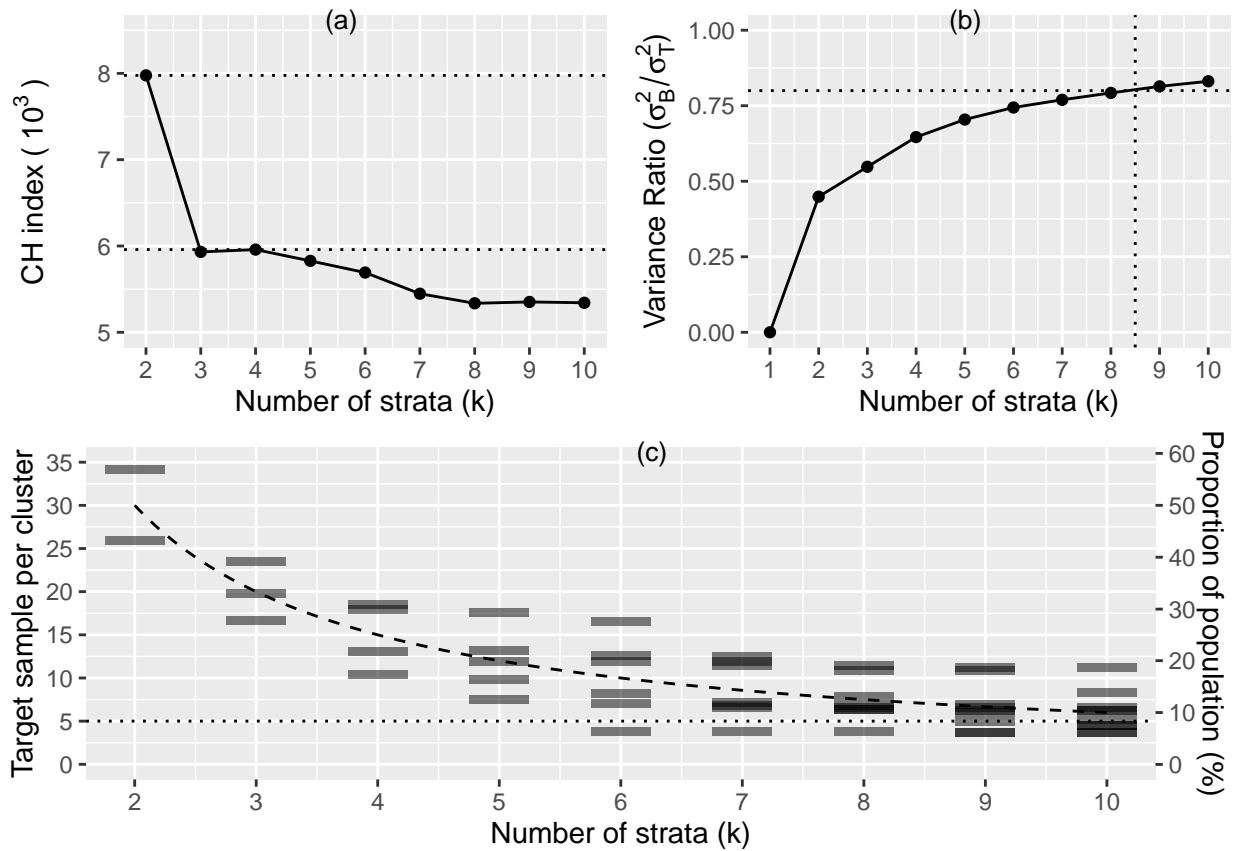
Cluster Analysis





```
## 'summarise()' regrouping output by 'k' (override with '.groups' argument)
```



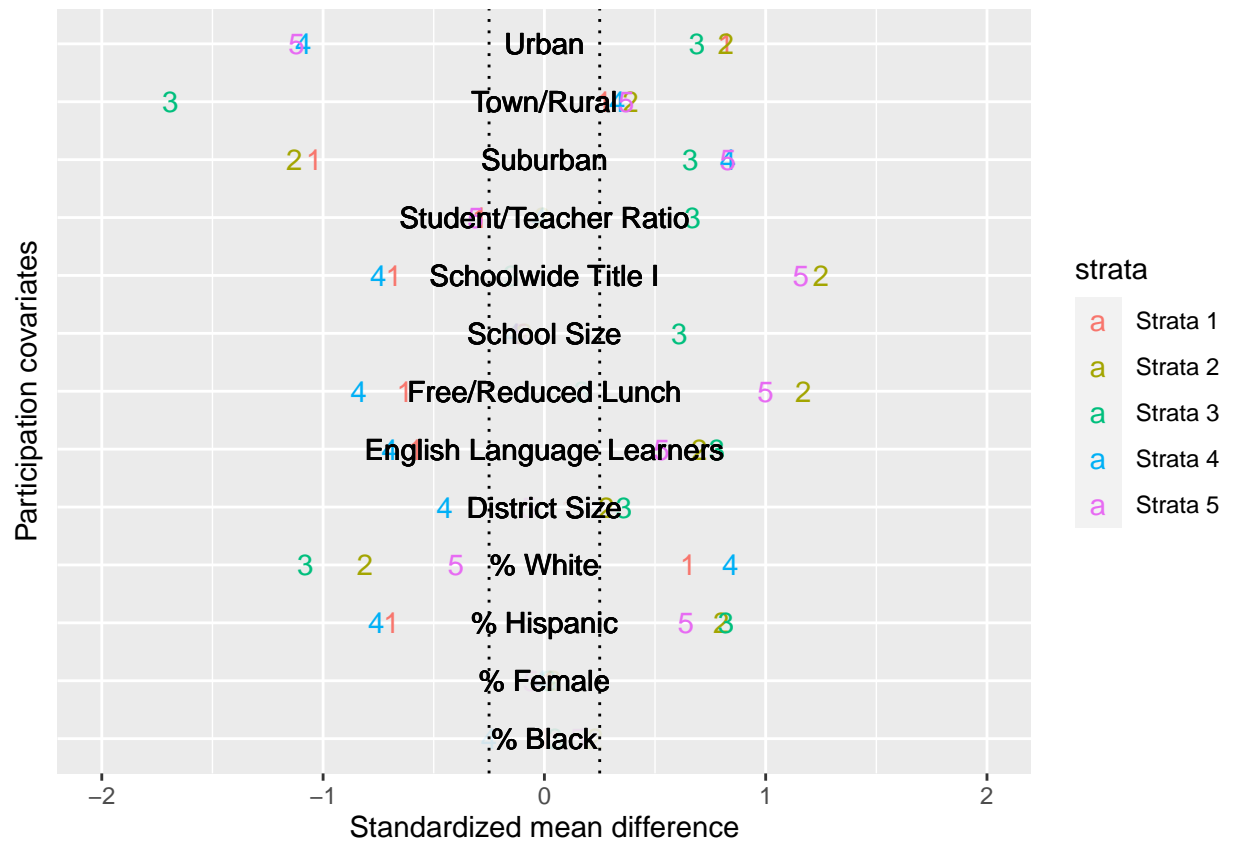


```
## 'summarise()' regrouping output by 'strata' (override with '.groups' argument)
```

```
## Joining, by = "var"
```

```
## Joining, by = "vnames"
```

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

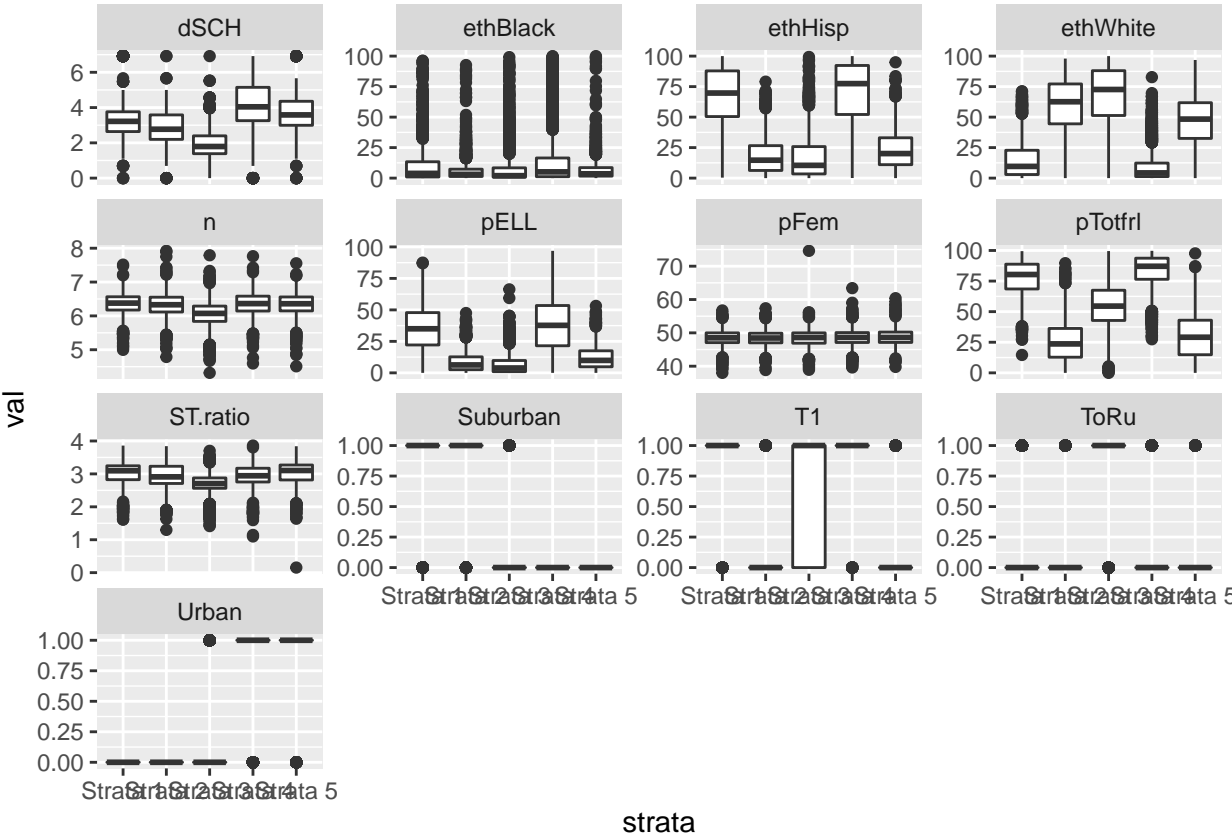


Note: Using an external vector in selections is ambiguous.

i Use 'all_of(covariates)' instead of 'covariates' to silence this message.

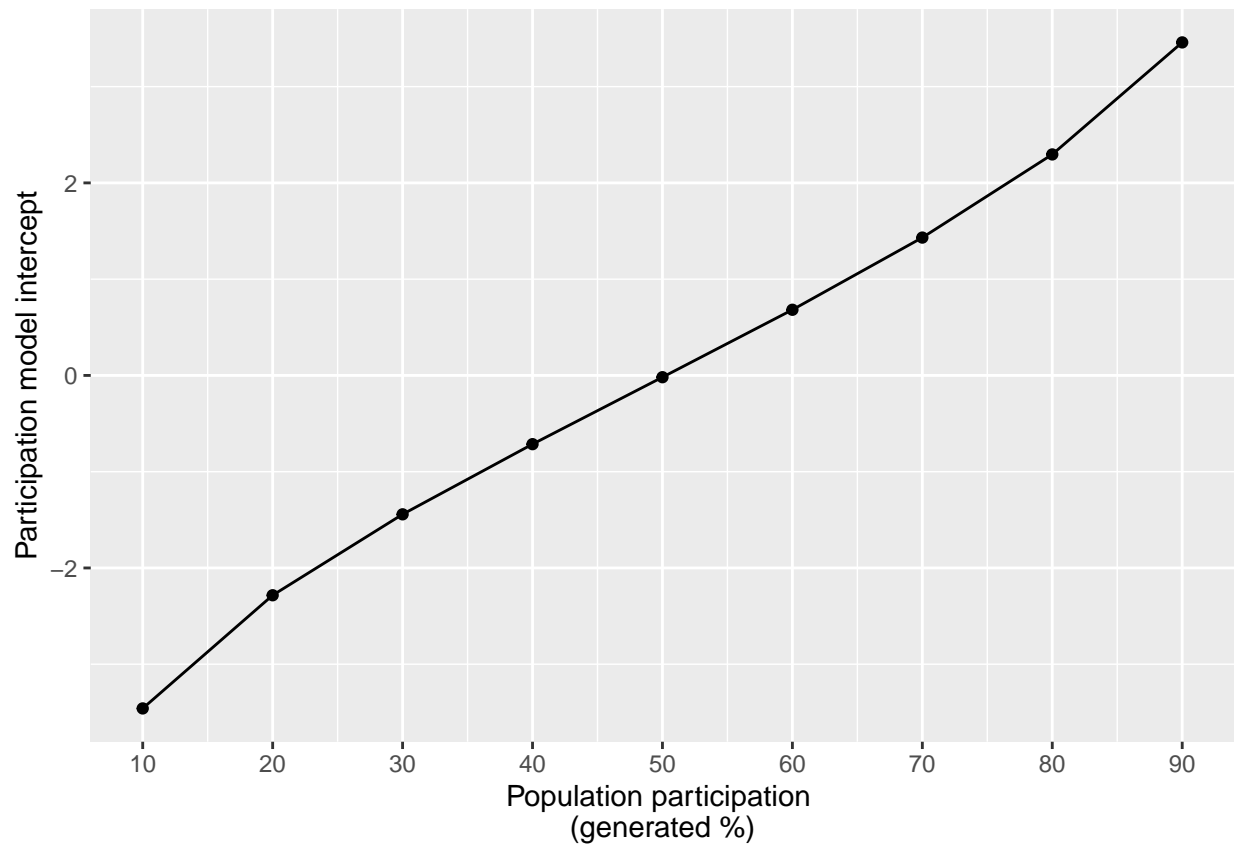
i See <<https://tidyselect.r-lib.org/reference/faq-external-vector.html>>.

This message is displayed once per session.



Participation Generating Model

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



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

```
## Joining, by = "DSID"
```

```
## Joining, by = "DSID"
```

Results

Generalizability

B Index.

```
## 'summarise()' regrouping output by 'sample_method', 'RR', 'Clustering' (override with
```

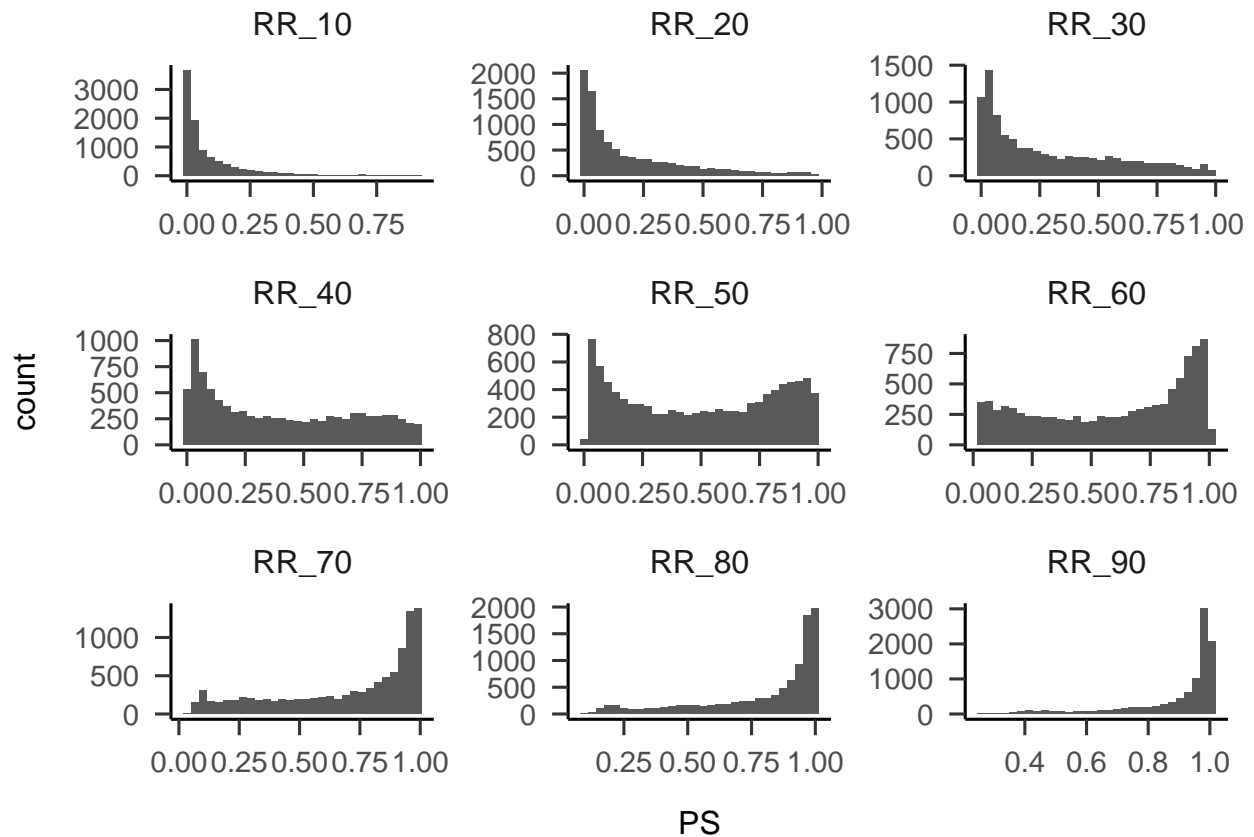


Figure 1. Distributions of Participation Propensity Scores

Standardized mean differences.

```
## Joining, by = "var"
```

```
## 'summarise()' regrouping output by 'sample_method', 'RR', 'var', 'Sampling' (override
```

```
## Joining, by = "vnames"
```

```
## Joining, by = "vnames"
```

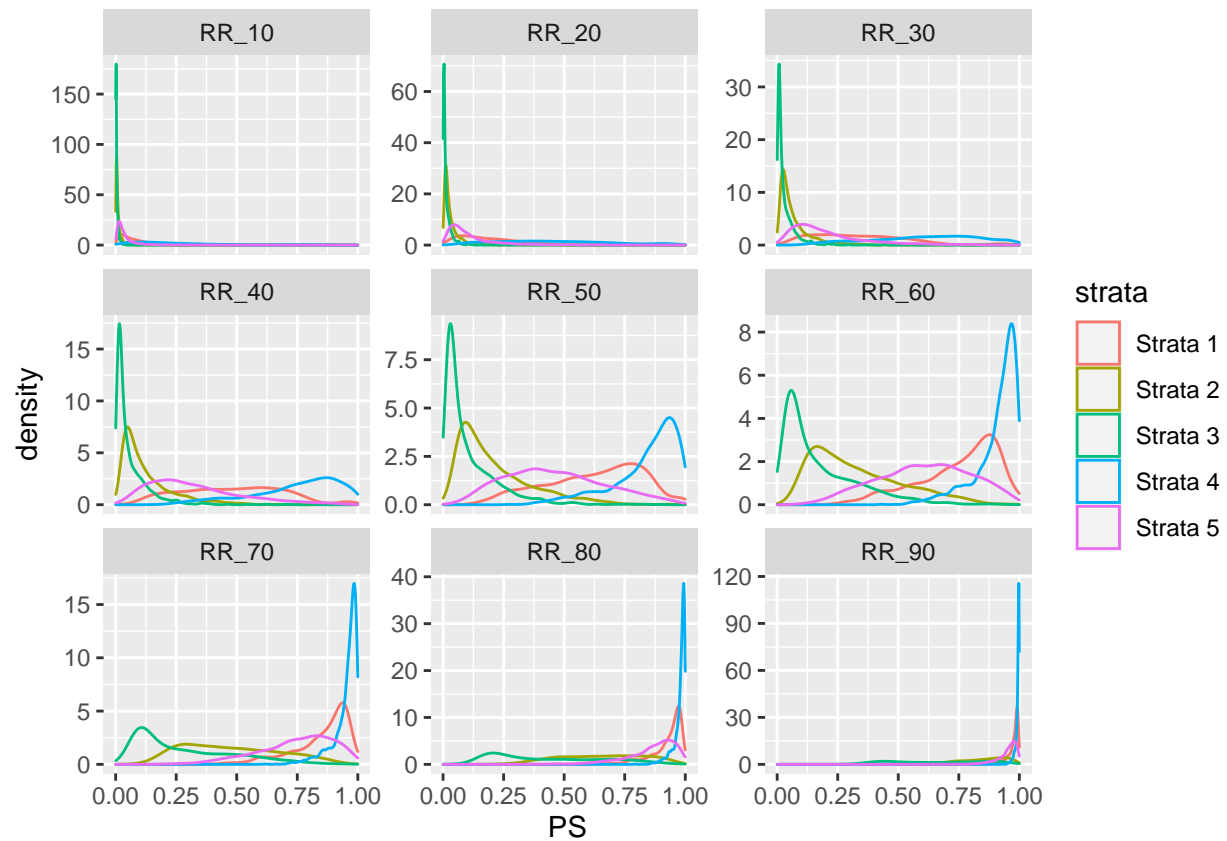
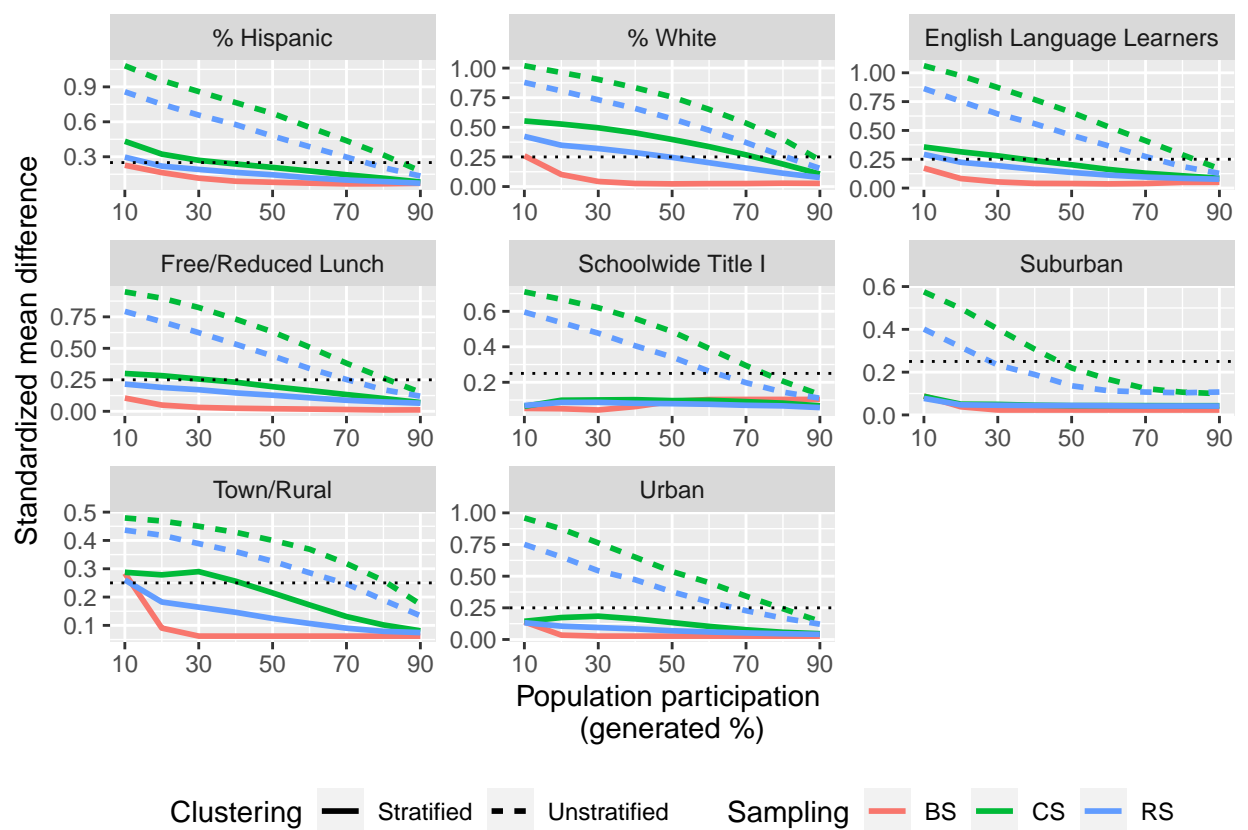


Figure 2. Distributions of Participation Propensity Scores



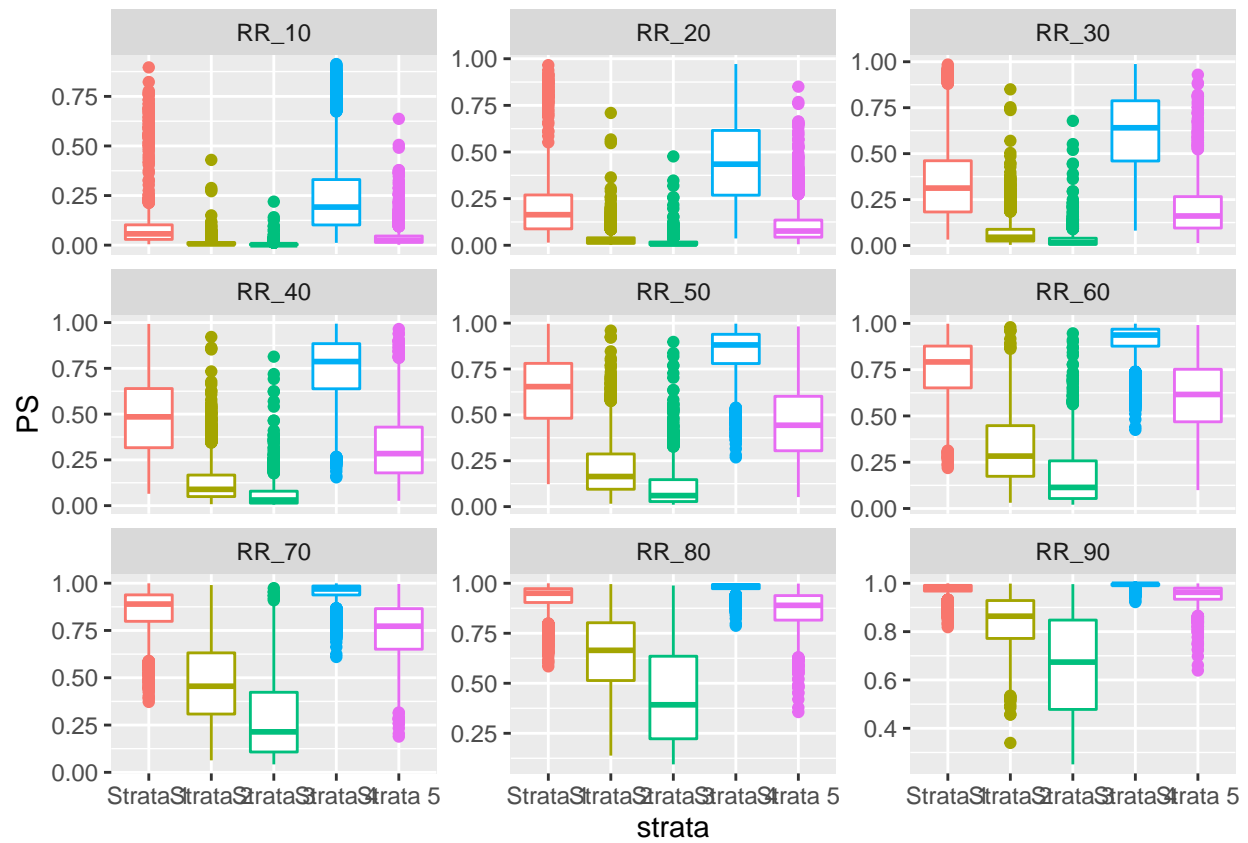
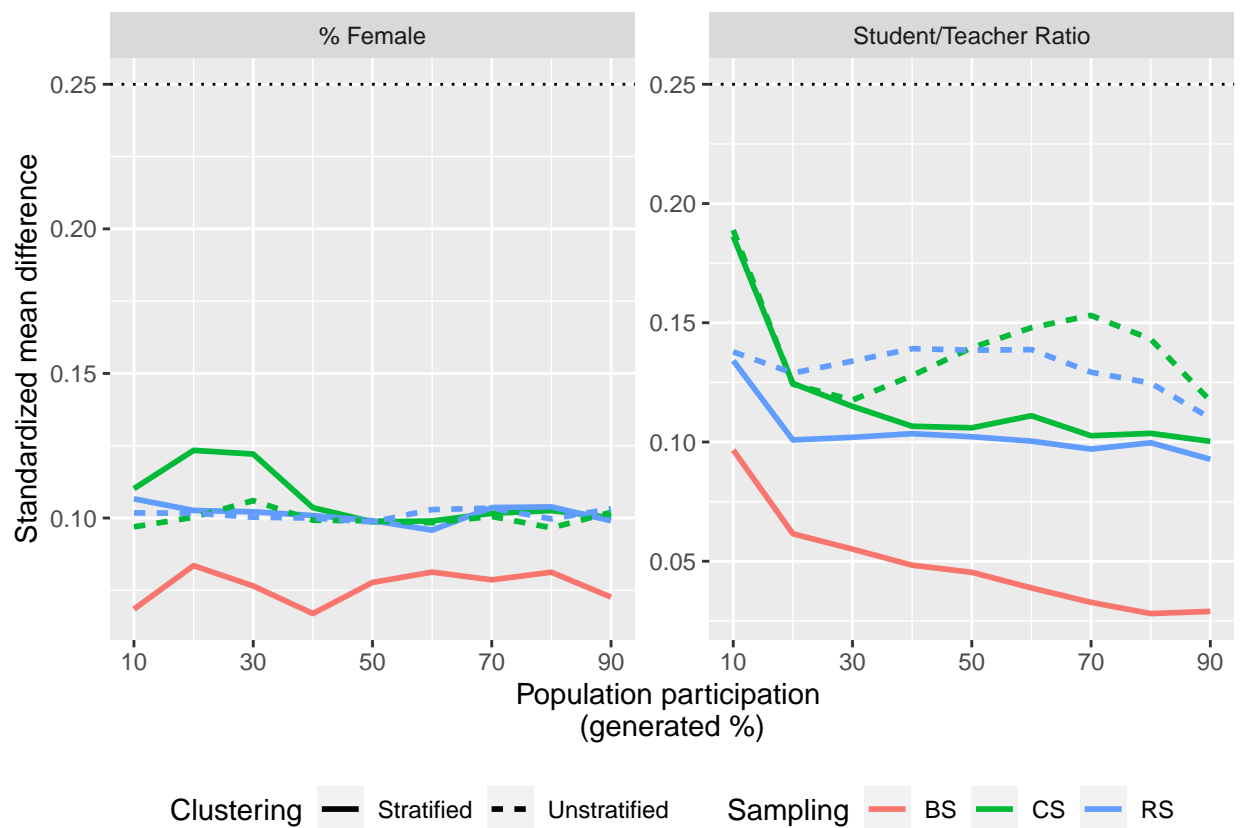


Figure 3. Distributions of Participation Propensity Scores



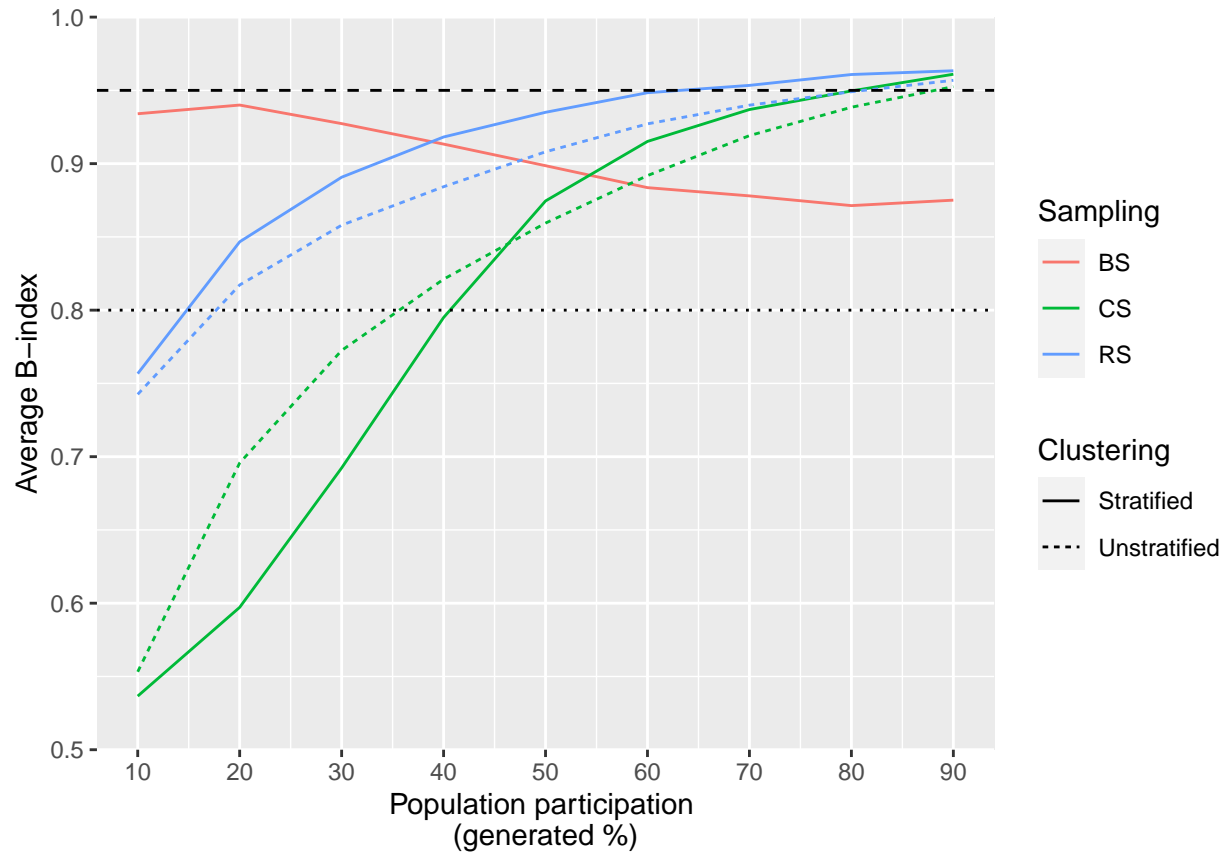


Figure 4. Average B-Index

Examples for presentations

V-ratio and Log odds.

```
## 'summarise()' regrouping output by 'vnames', 'T.SS' (override with '.groups' argument)
```

```
## 'summarise()' regrouping output by 'vnames' (override with '.groups' argument)
```

```
## Joining, by = "vnames"
```

```
## Joining, by = "vnames"
```

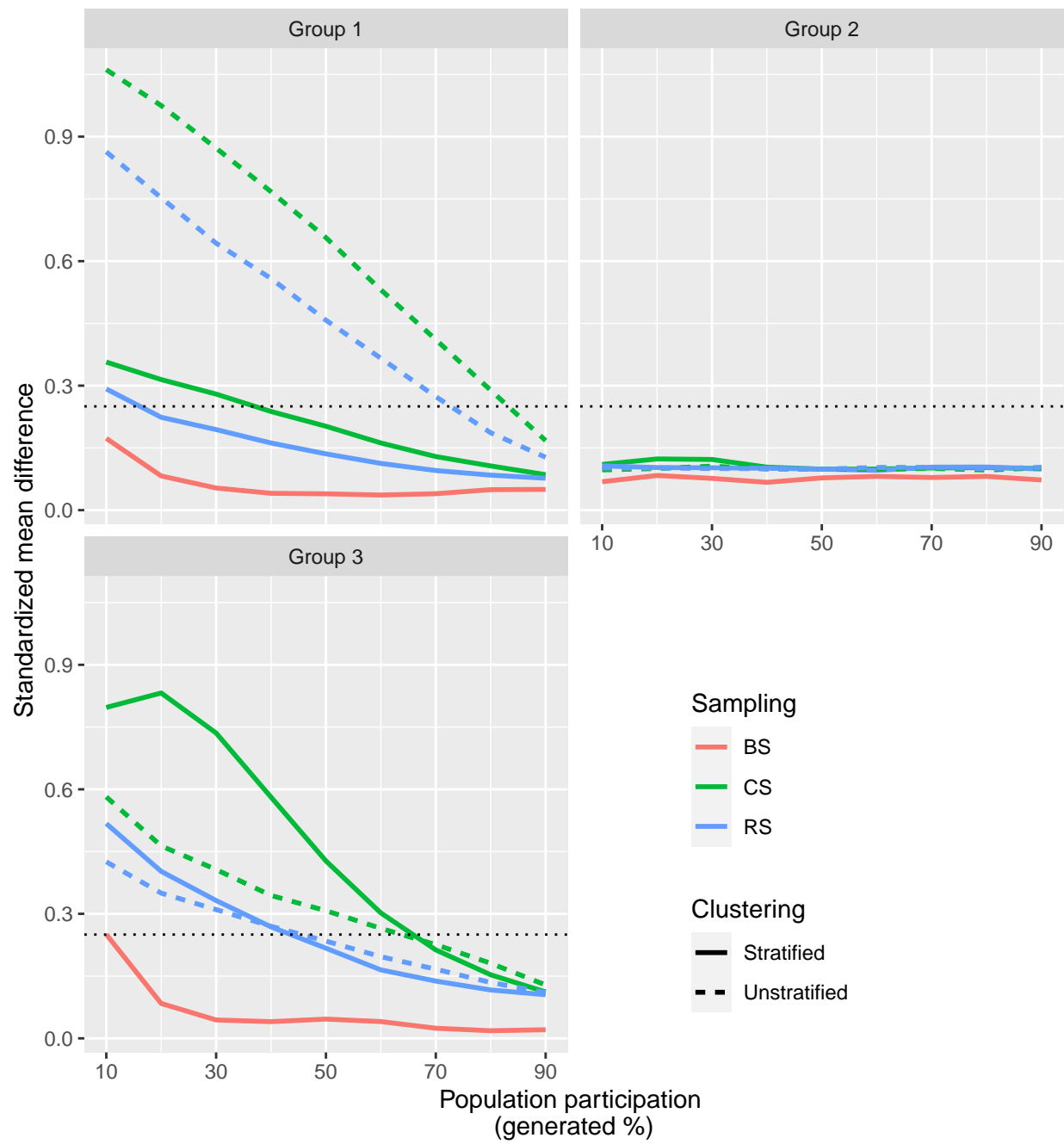
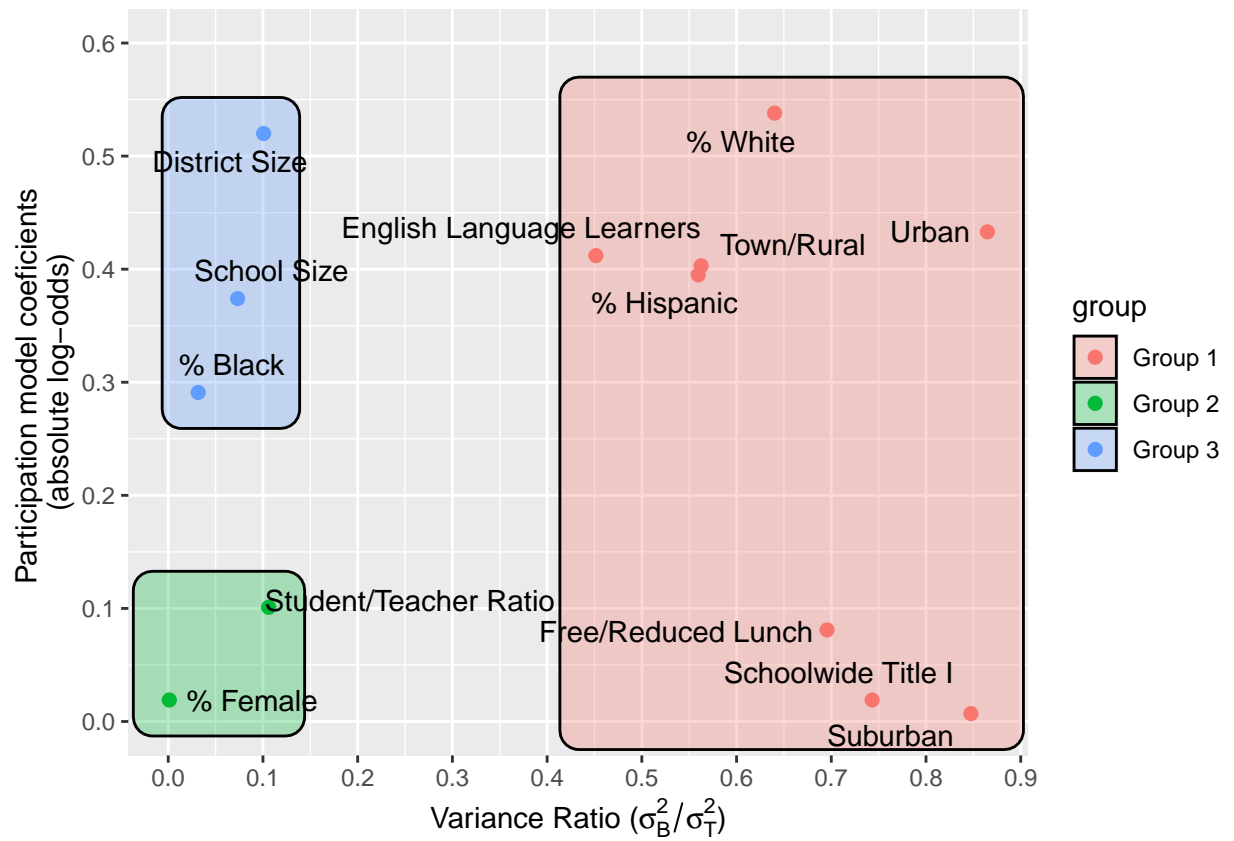


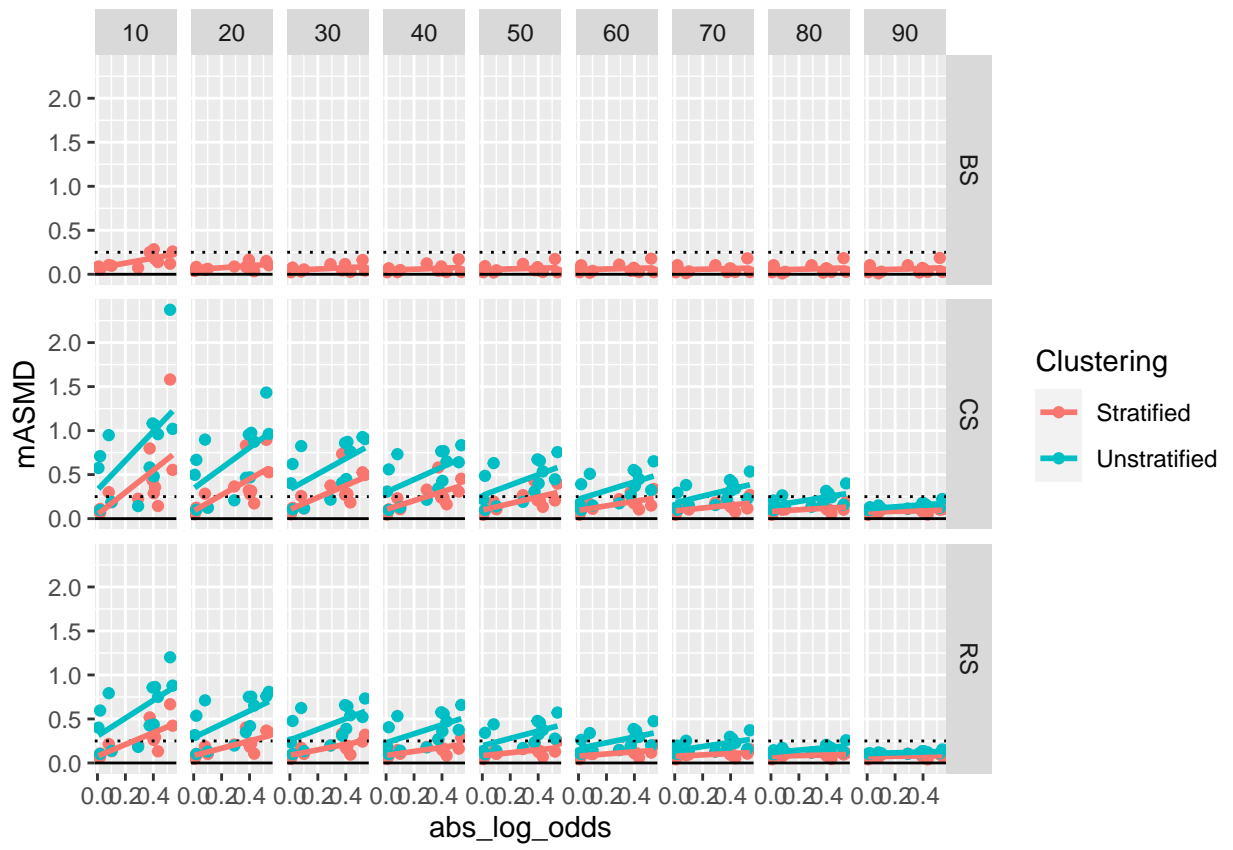
Figure 5. Average SMDs



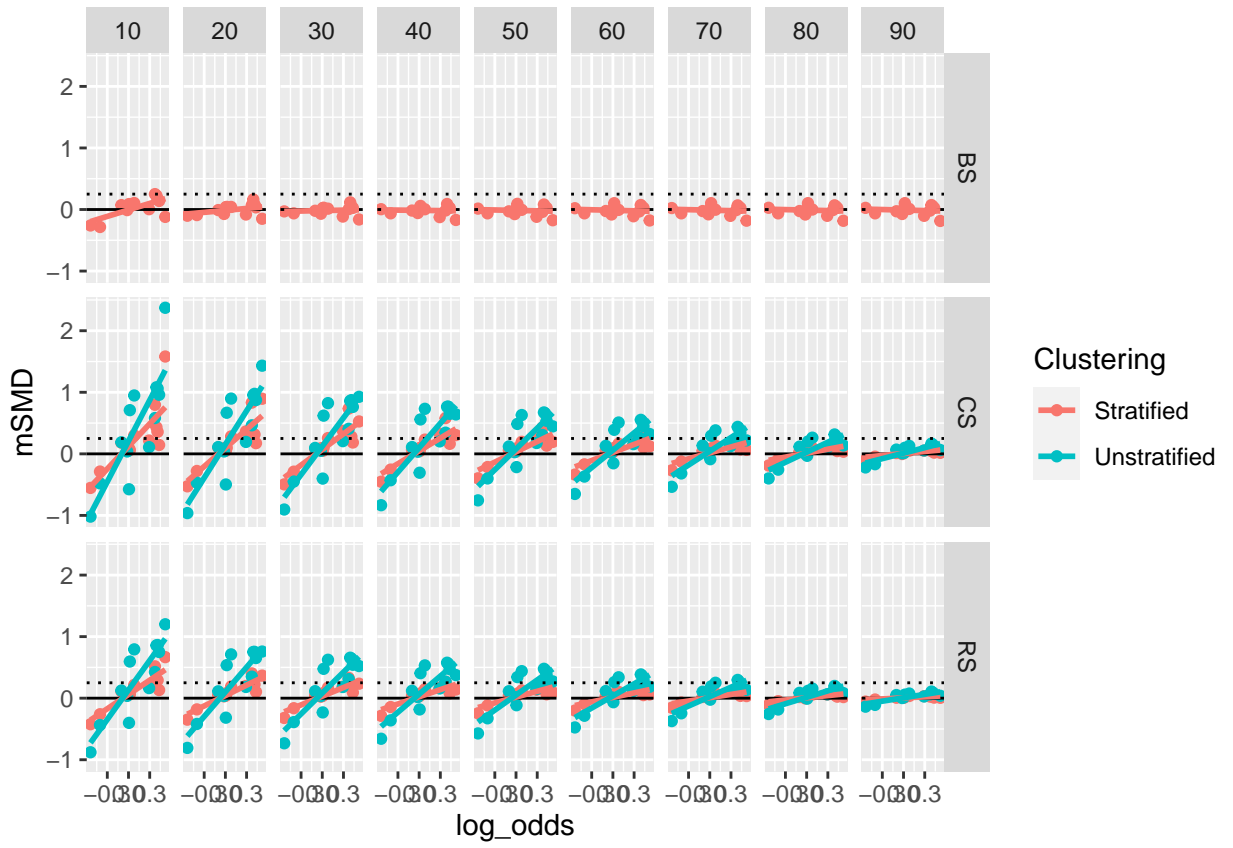
Test.

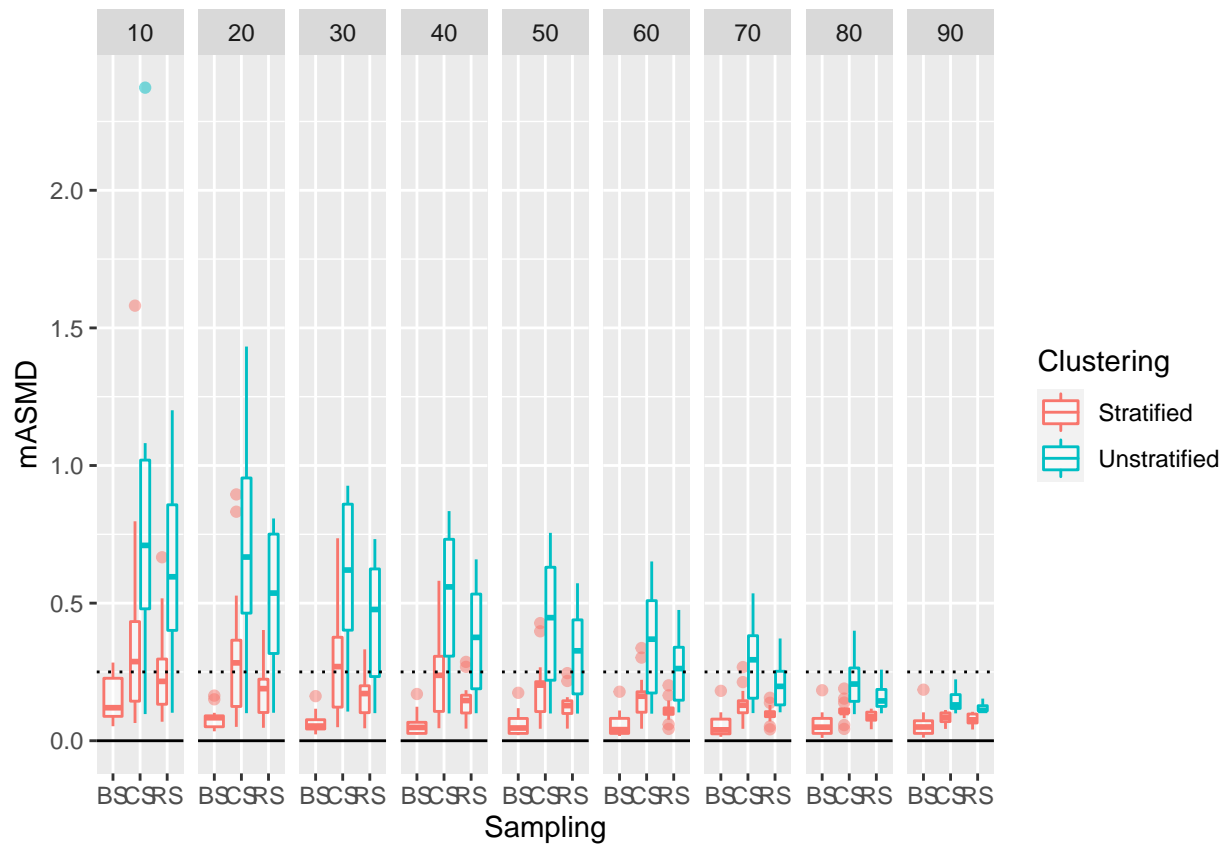
```
## Joining, by = c("vnames", "log_odds")
```

```
## 'geom_smooth()' using formula 'y ~ x'
```



```
## 'geom_smooth()' using formula 'y ~ x'
```





Analysis of Variance Table

Model 1: mSMD ~ RR Model 2: mSMD ~ log_odds + vratio + RR Model 3: mSMD									
~ log_odds * vratio * RR Res.Df RSS Df Sum of Sq F Pr(>F)									
1	583	59.336							
2	581	38.107	2	21.2295	196.574	< 2.2e-16	3	577	31.157
							4	6.9494	32.174
2.2e-16 — Signif. codes: 0 “’ 0.001 ” 0.01 ” 0.05 “.” 0.1 ’ ’ 1									

=====									
Model 1	Model 2	Model 3							
----- (Intercept) 0.24 *** 0.15 *** 0.11 *									
(0.03)	(0.03)	(0.05)							
RR	-0.00 ***	-0.00 ***	-0.00						
(0.00)	(0.00)	(0.00)							

log_odds 0.61 *** 1.24 (*0.03*) (*0.16*)

vratio 0.04 -0.02

(0.03) (0.08)

log_odds:vratio -0.00

(0.28)

log_odds:RR -0.02 (0.00)

vratio:RR 0.00

(0.00)

log_odds:vratio:RR 0.01

(0.00)

————— R² 0.04 0.39 0.50

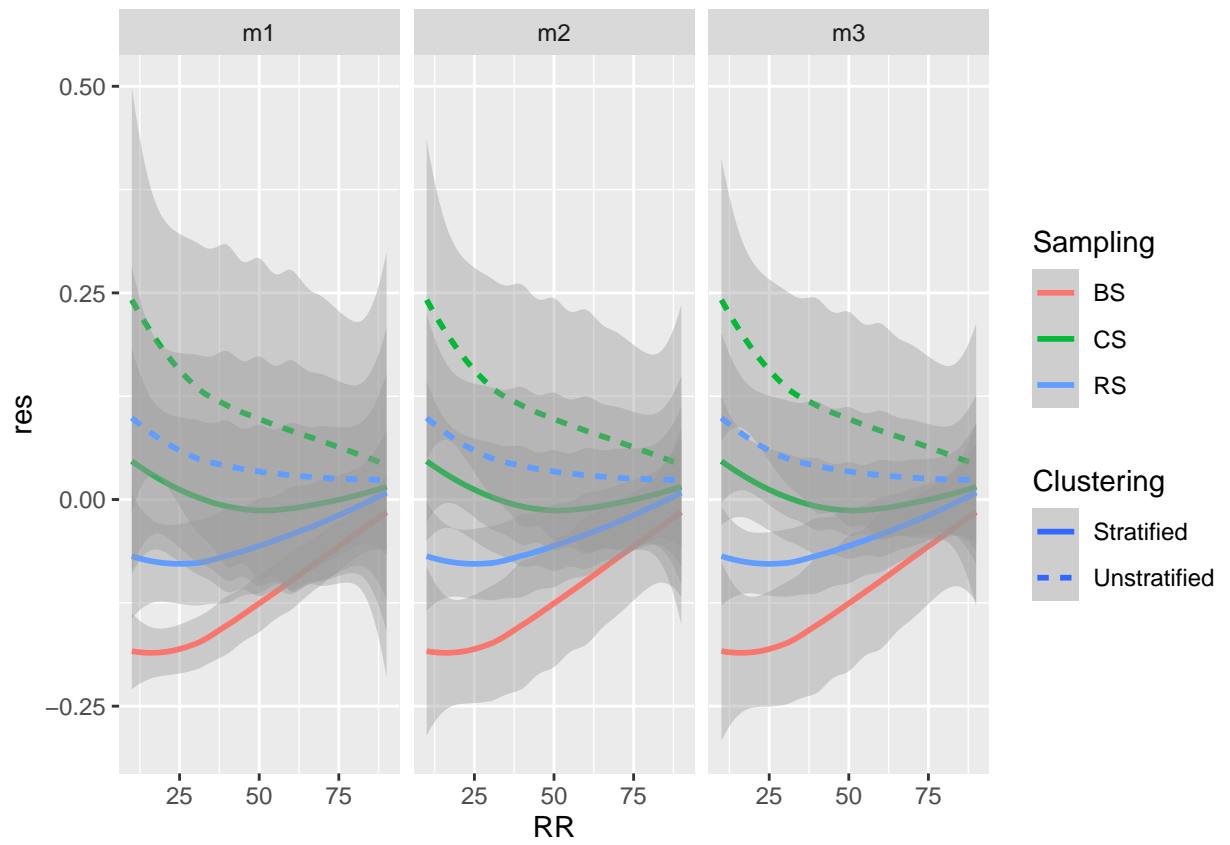
Adj. R² 0.04 0.38 0.49

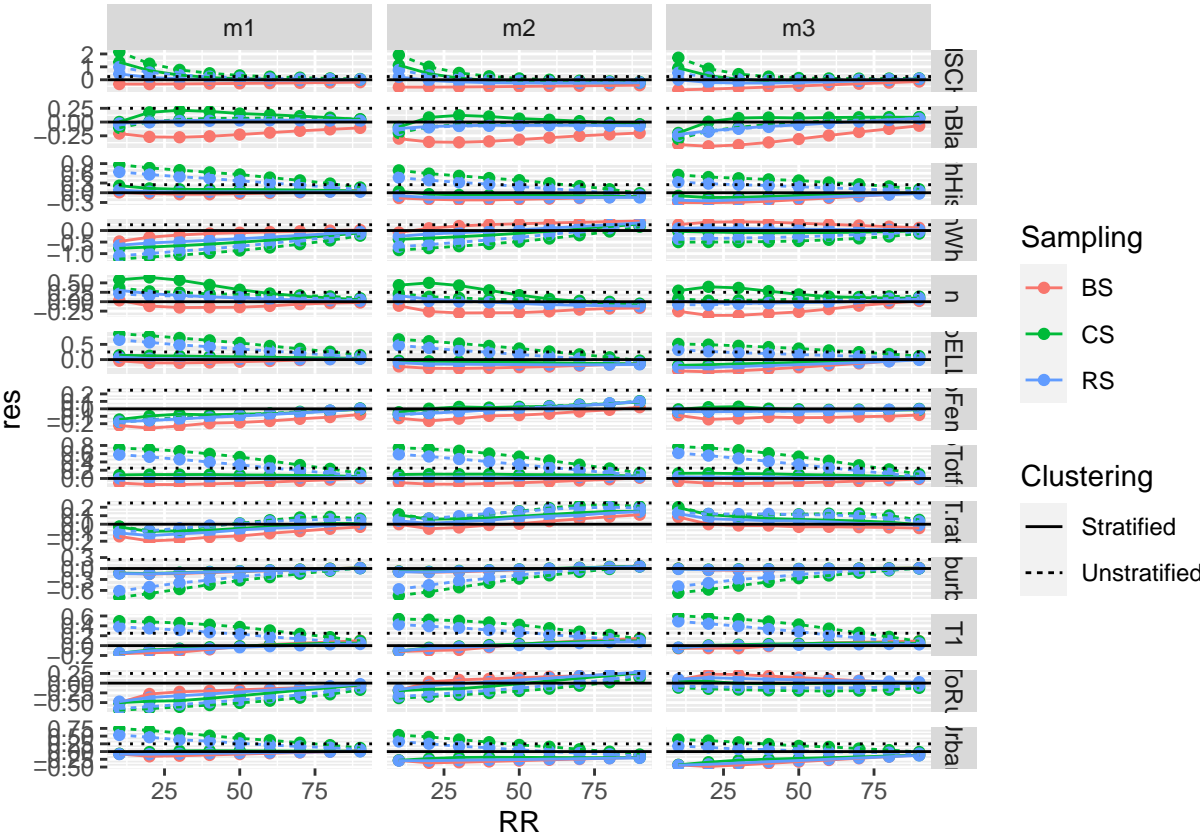
Num. obs. 585 585 585

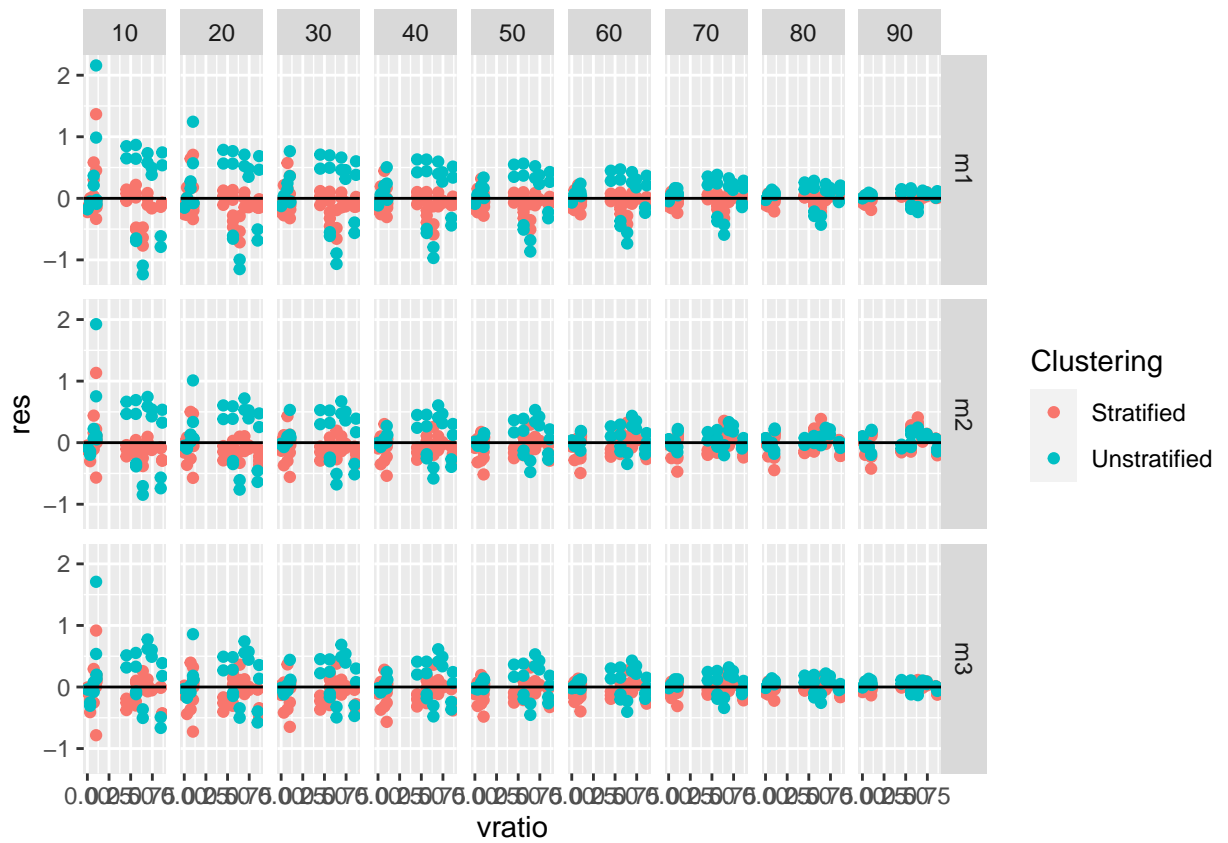
=====

*** p < 0.001; ** p < 0.01; * p < 0.05

'geom_smooth()' using method = 'loess' and formula 'y ~ x'



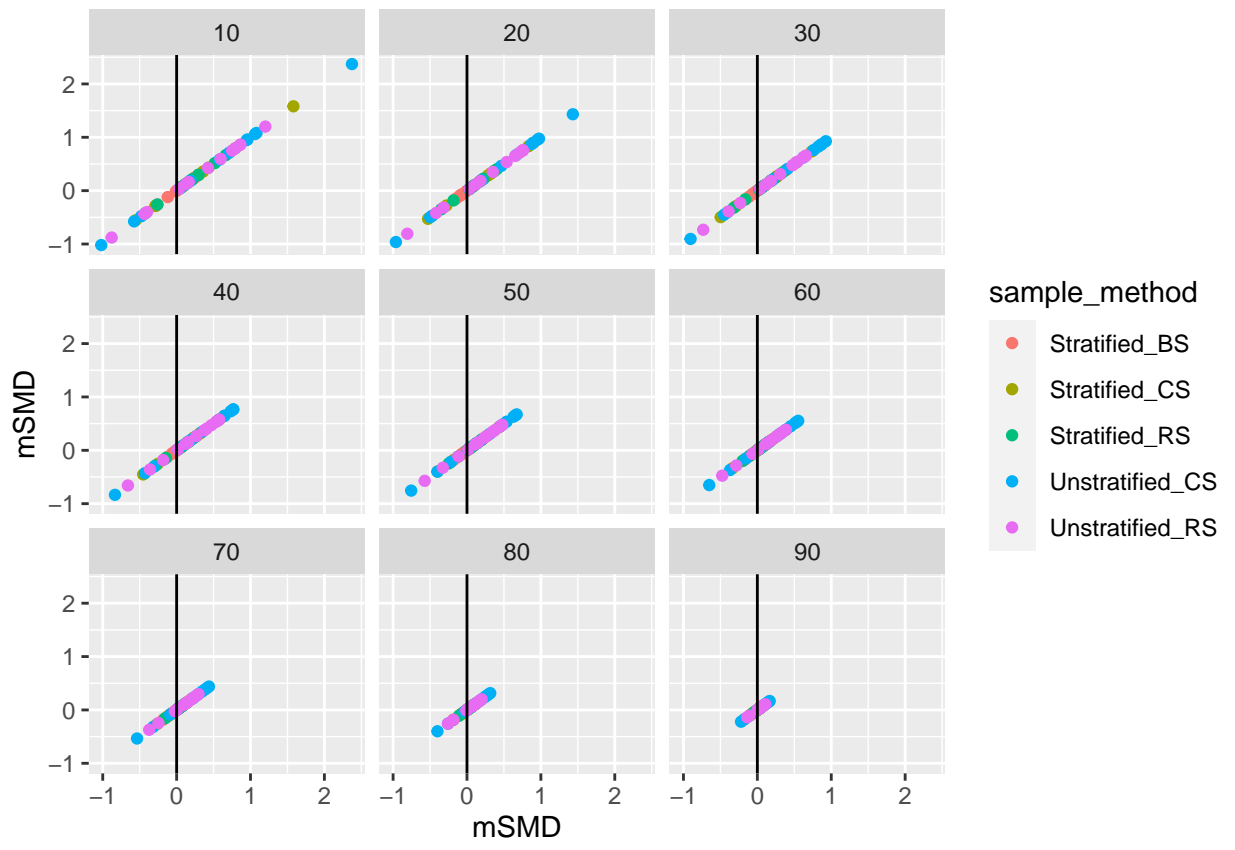


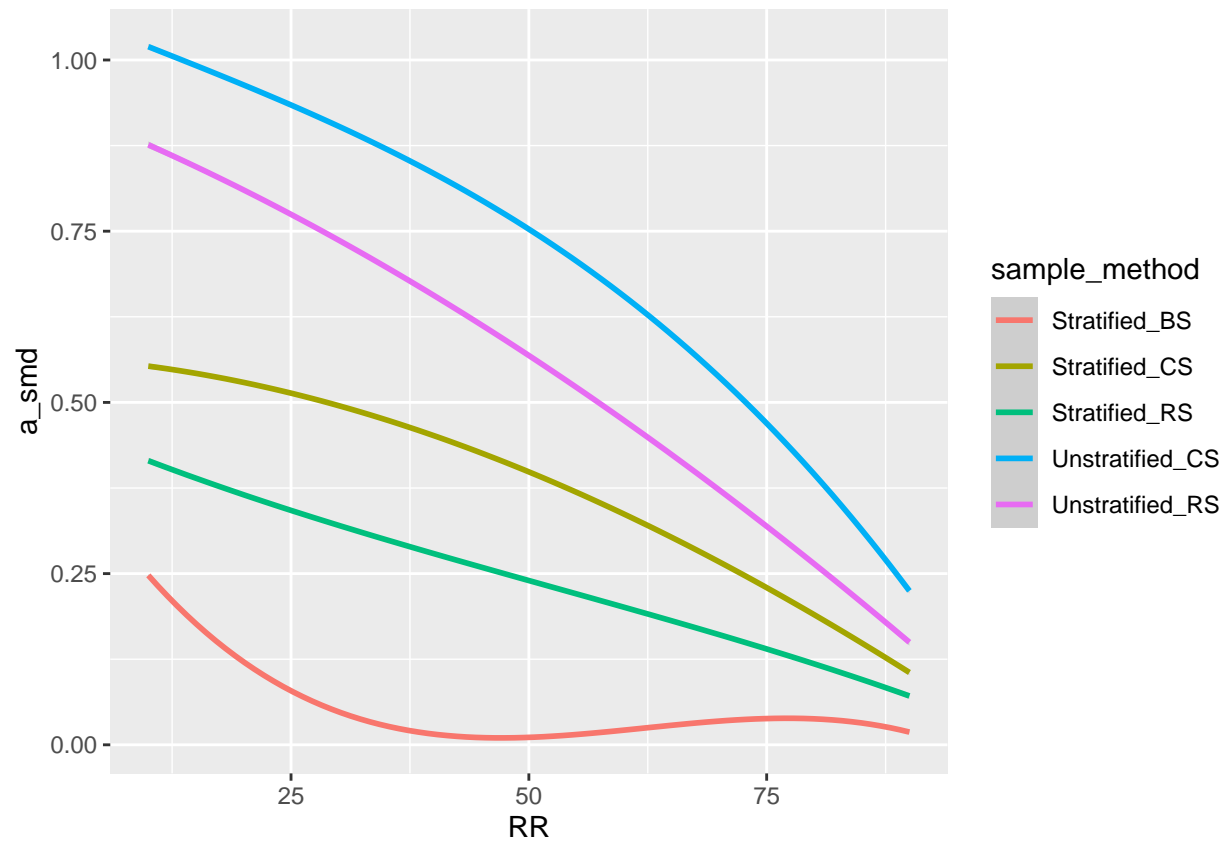


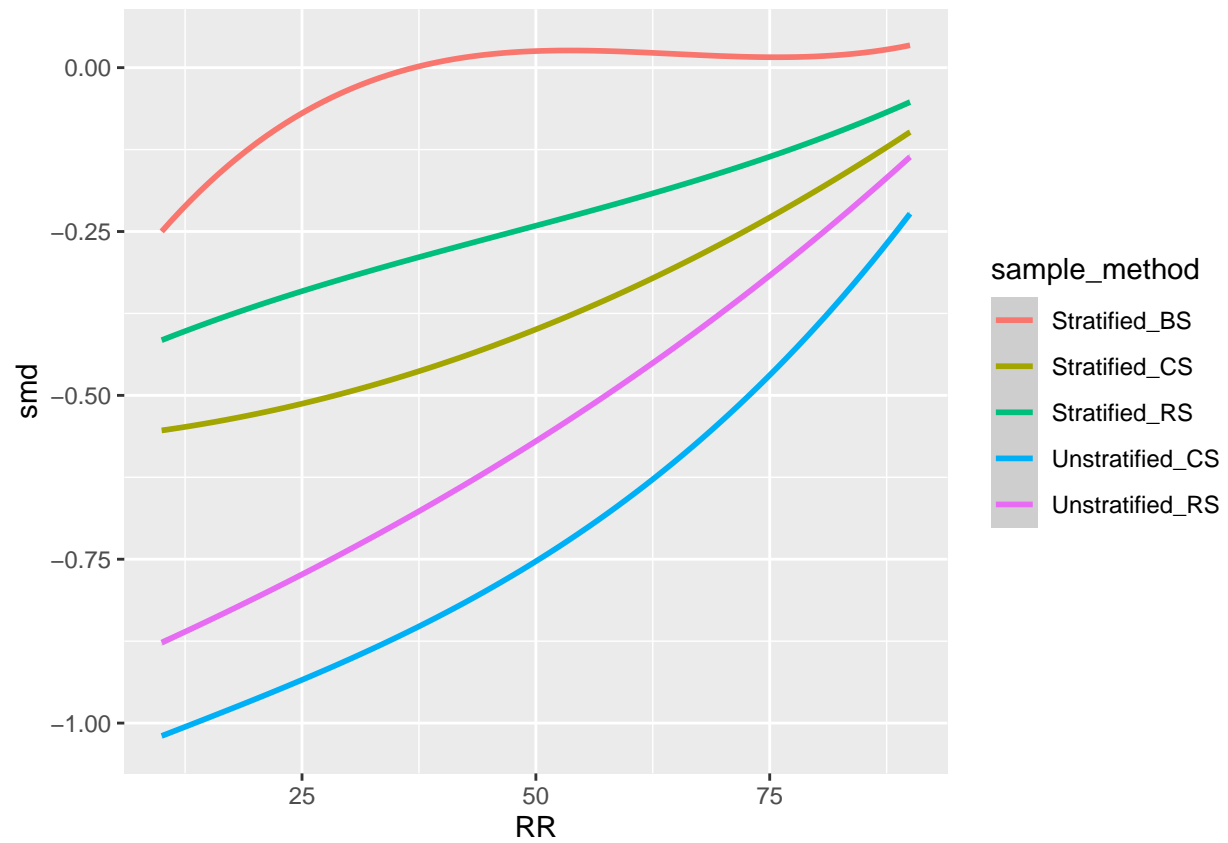
Test 2

Joining, by = "var"

'summarise()' regrouping output by 'sample_method', 'RR', 'var', 'Sampling' (override







Feasibility

Sampling Difficulty.

```
## 'summarise()' regrouping output by 'sample_method', 'RR', 'measure', 'Sampling', 'Clu
```

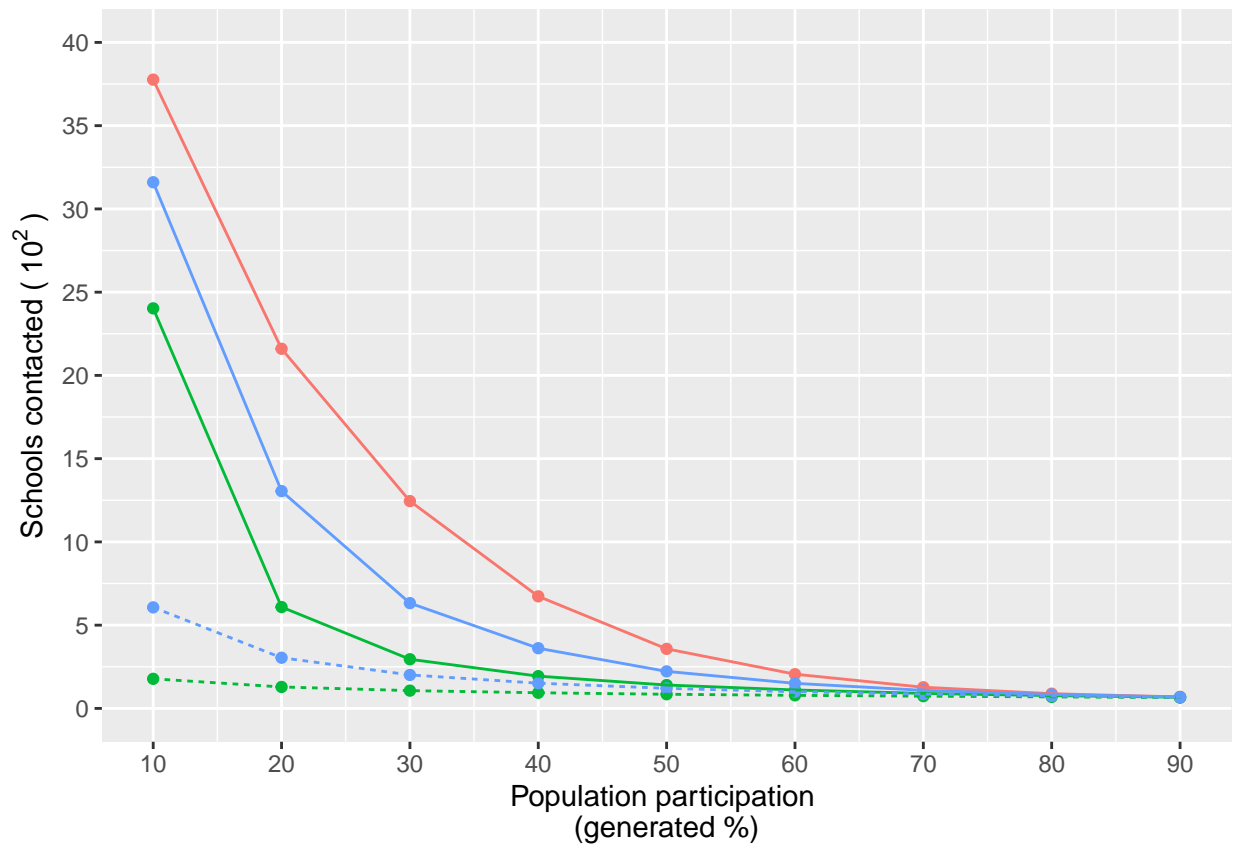
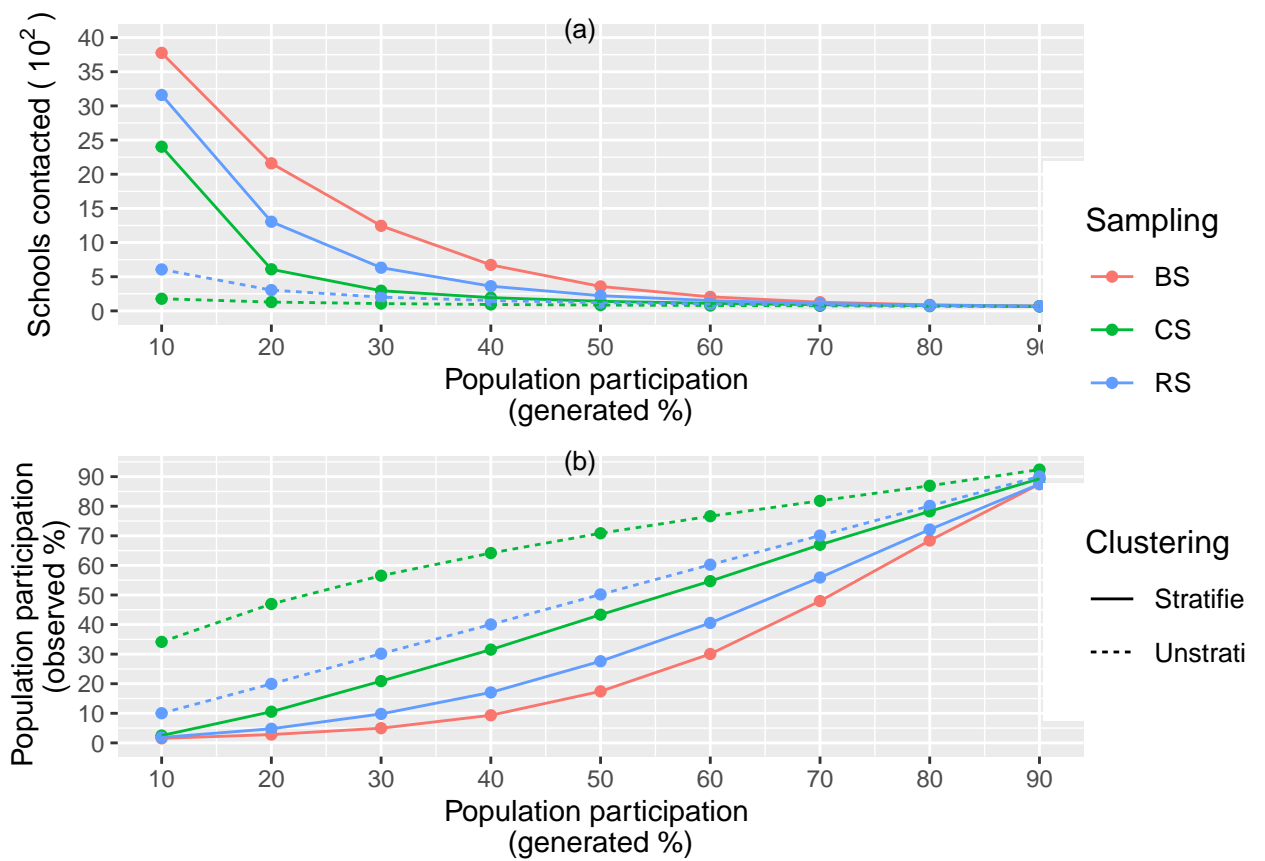


Figure 6. Schools Contacted



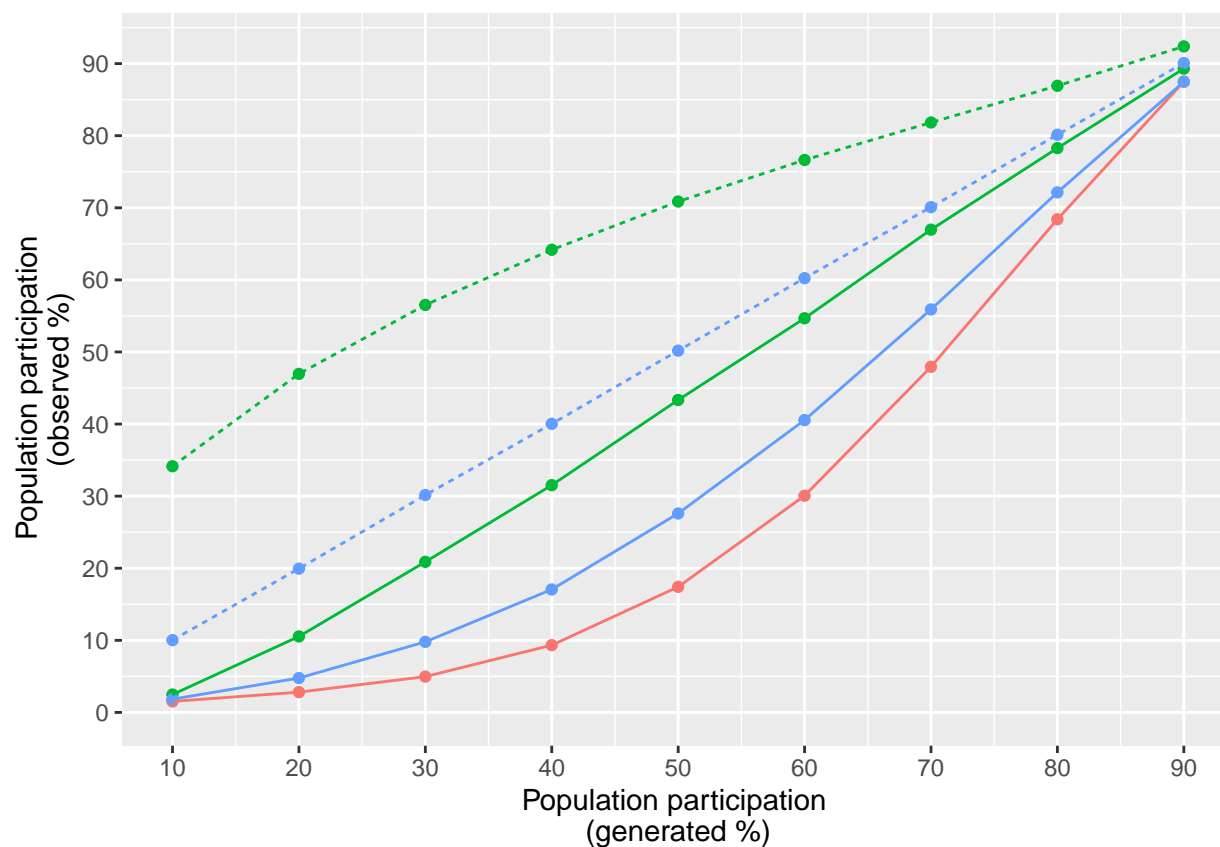


Figure 7. Sampling response rates

Relative Performance.

```
## Joining, by = c("RR", "measure", "K")
```

Gini Plot.

[illegible]

[illegible]

```
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"  
## Joining, by = "DSID"
```

```
## Warning: 'cols' is now required when using unnest().
```

```
## Please use 'cols = c(data)'
```

```
## 'summarise()' regrouping output by 'sample_method', 'Clustering', 'Sampling' (override)
```

Export Plots