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

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


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

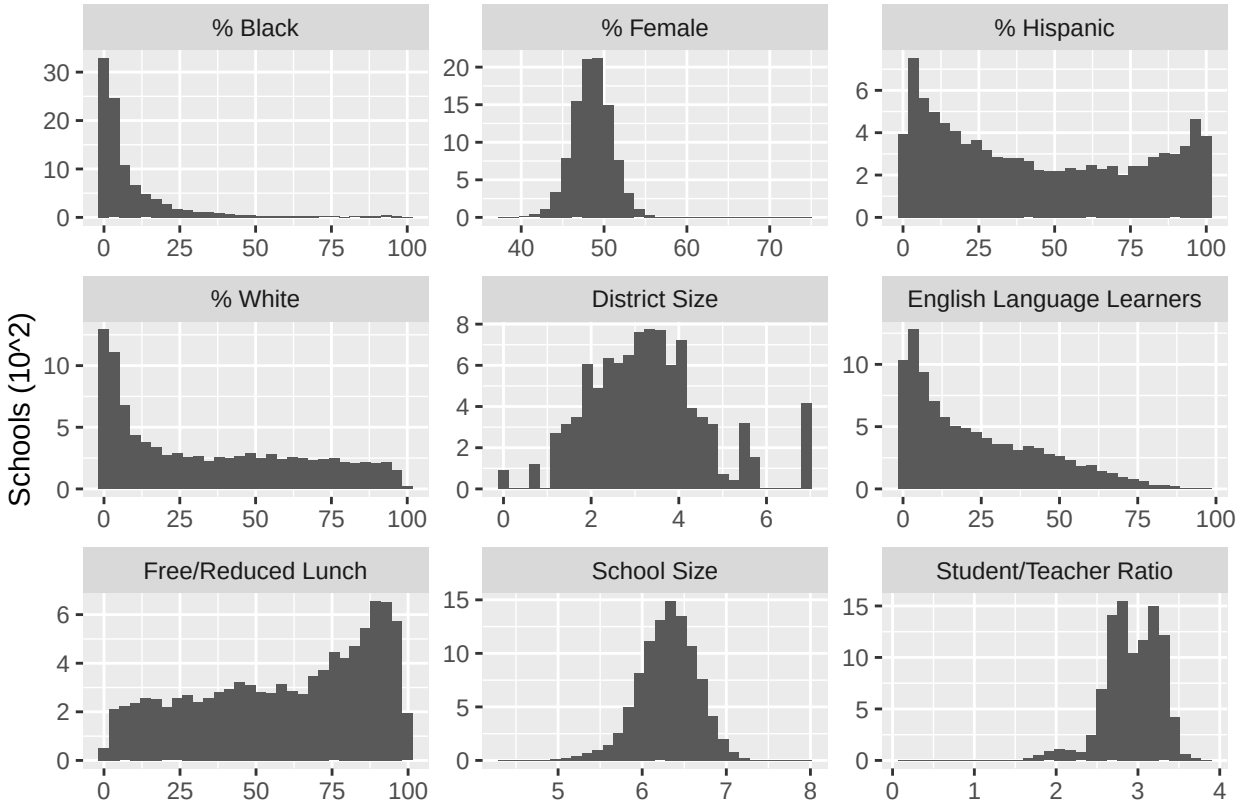
```
## Joining, by = "vnames"
```

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

Continuous variable distributions

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

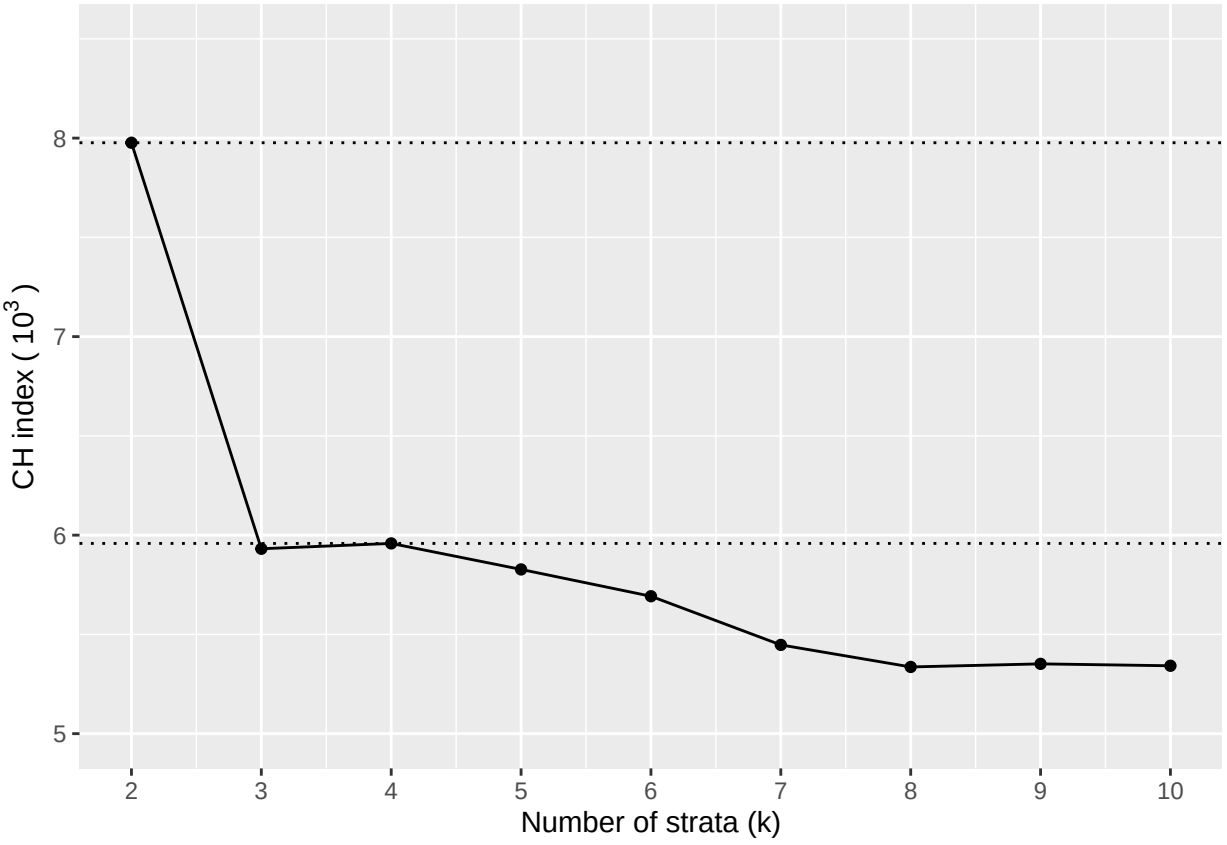
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

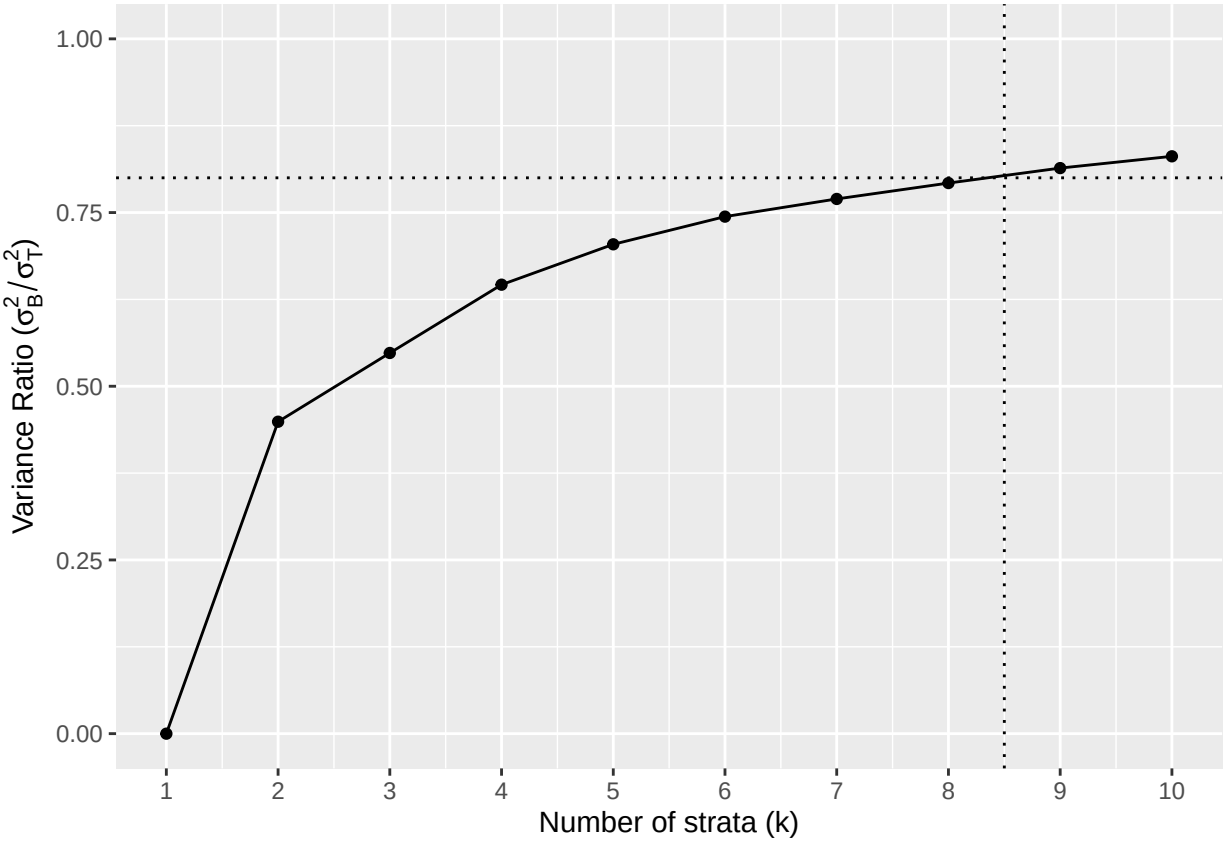


Methods Summary

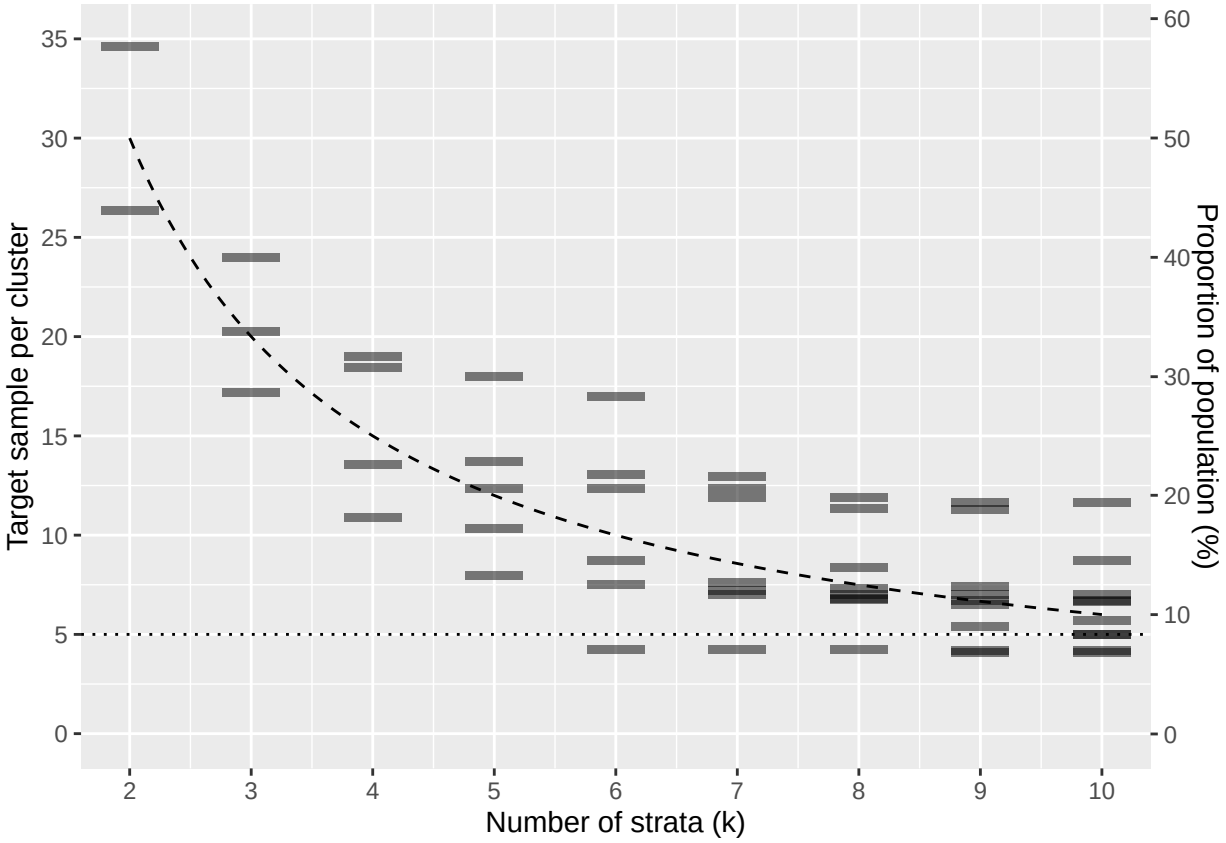
Cluster Analysis

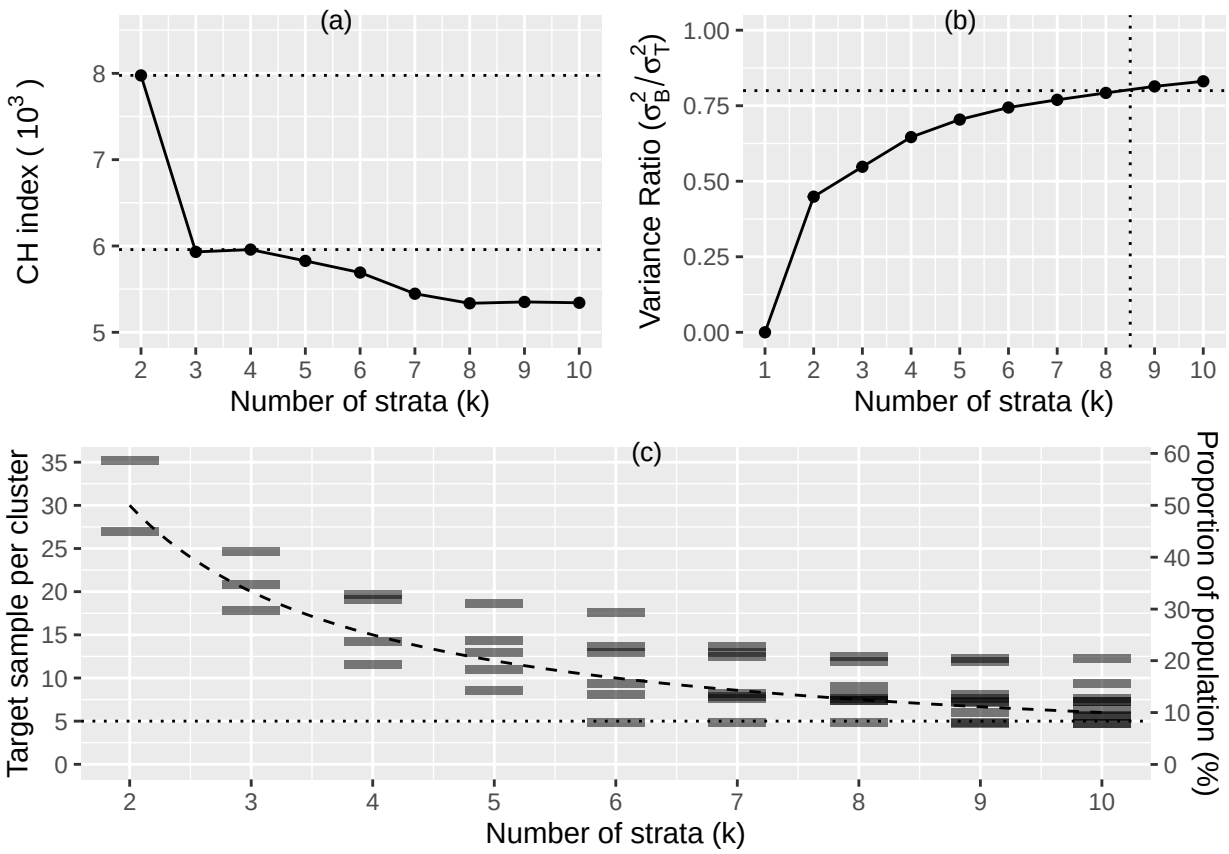
Selecting k.






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


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

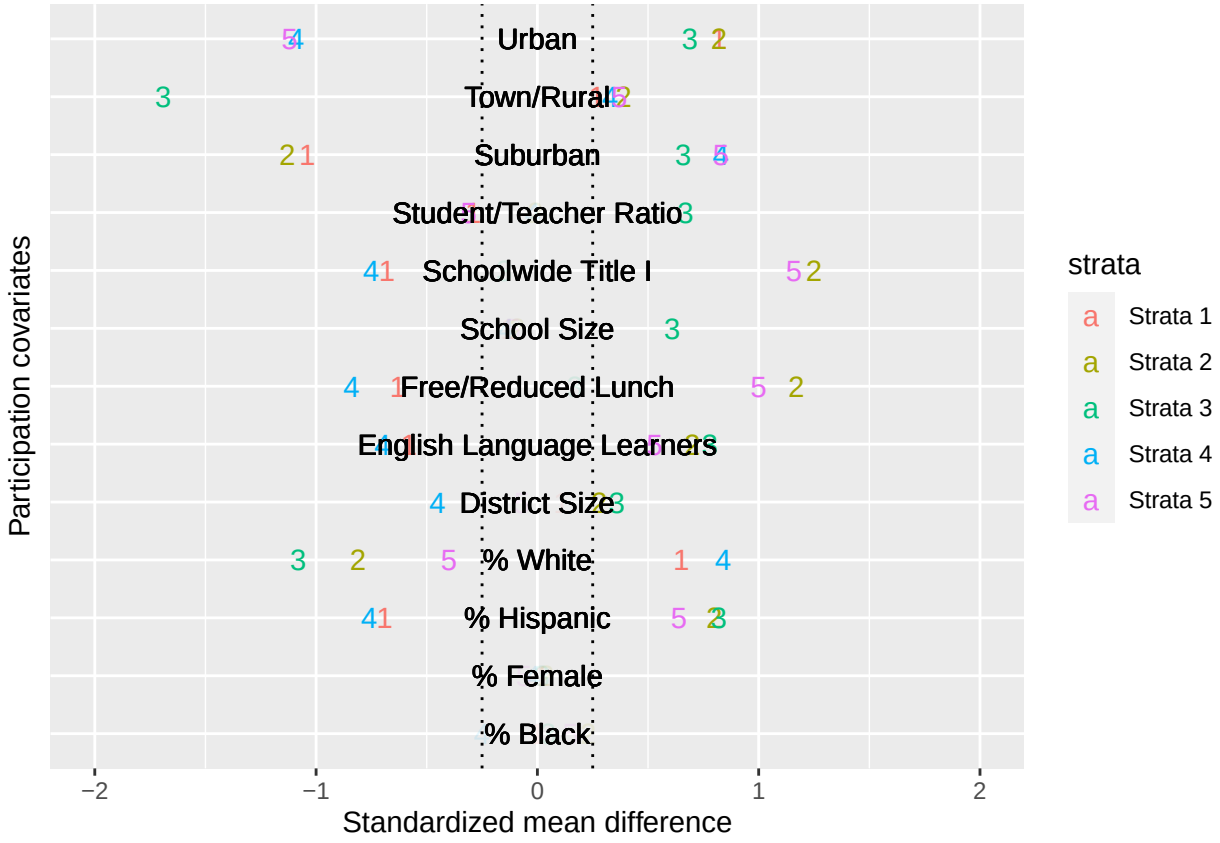



```
## `summarise()` regrouping output by 'strata' (override with `.groups` argument)
```

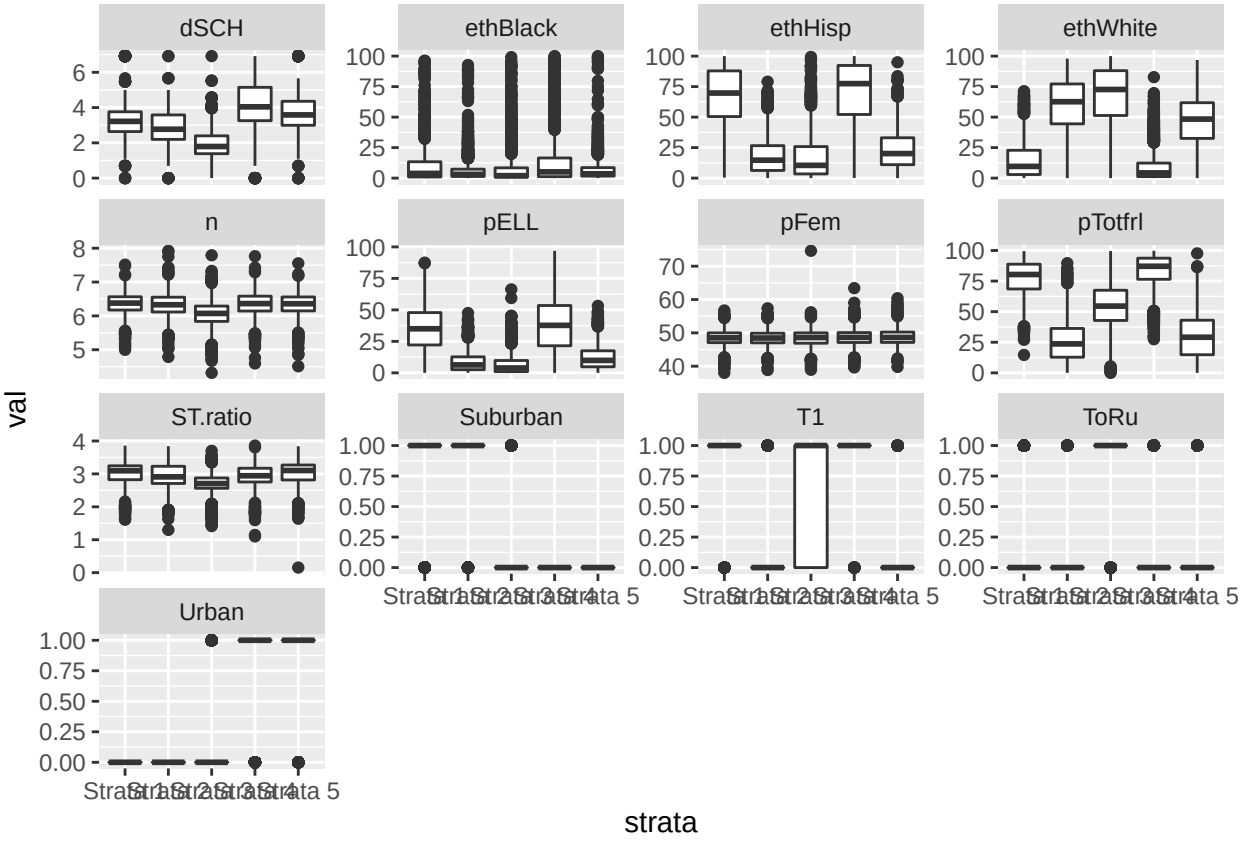
```
## Joining, by = "var"
```

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

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



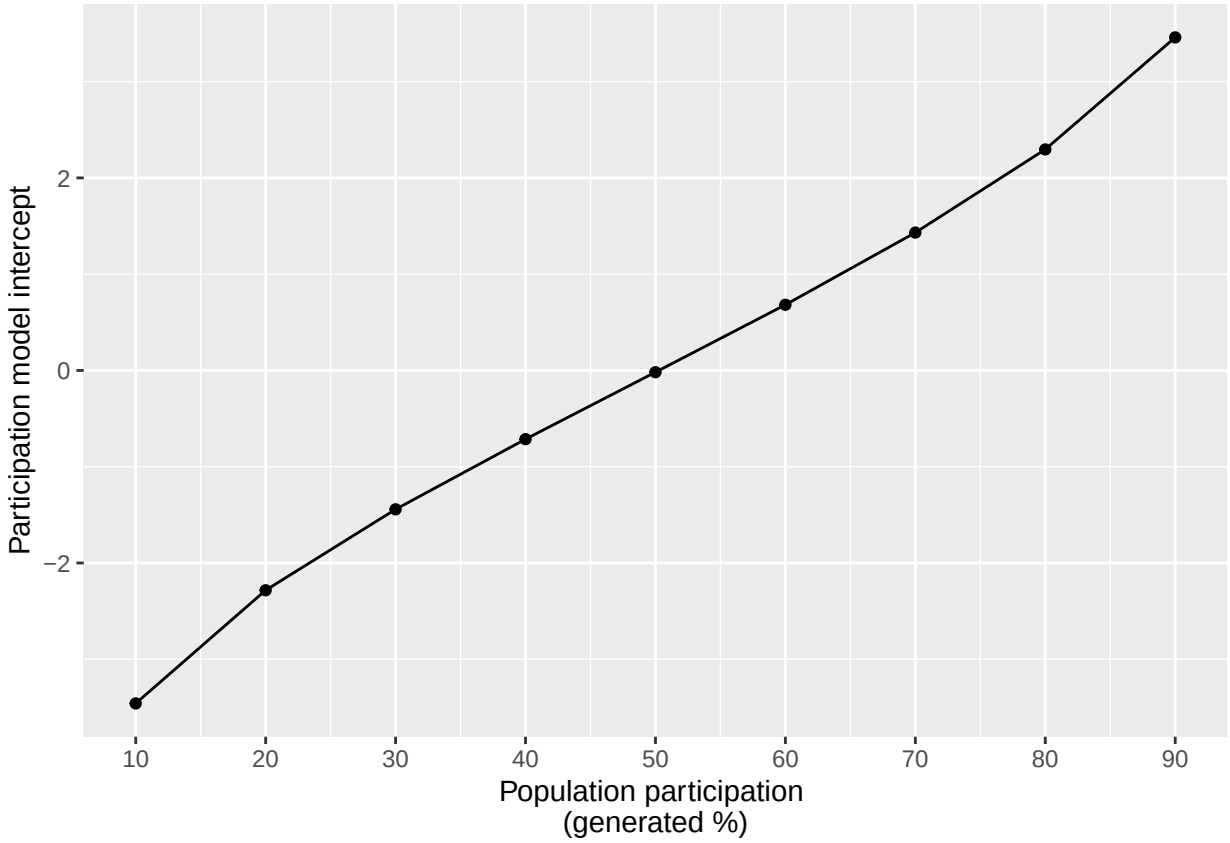

```
## Note: Using an external vector in selections is ambiguous.
## i Use ‘all_of(covariates)‘ instead of ‘covariates‘ to silence this message.
## i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
## This message is displayed once per session.
```


dSCH
ethBlack
ethHisp
ethWhite
n
pELL
pFem
pTotfrl
ST.ratio
Suburban
T1
ToRu
Urban
val
strata

Variation explained by the strata

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`.
```

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

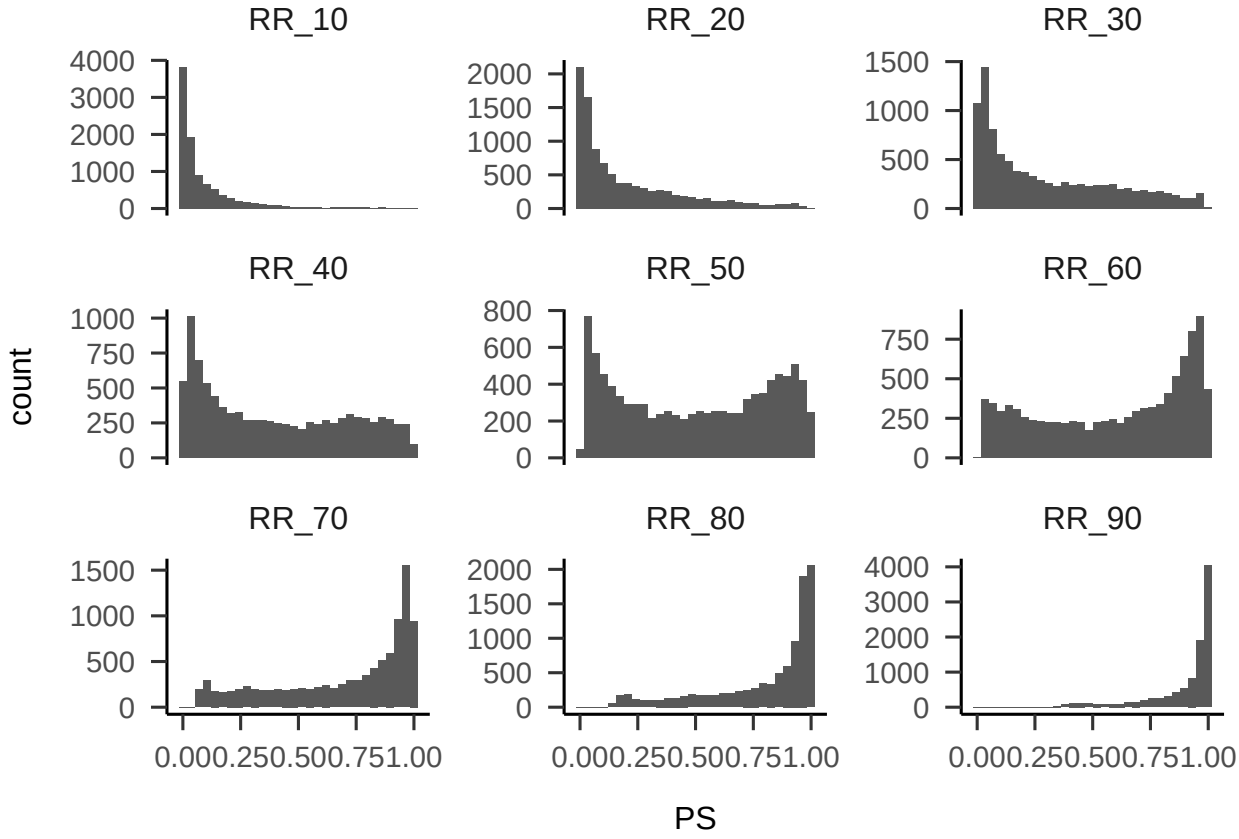
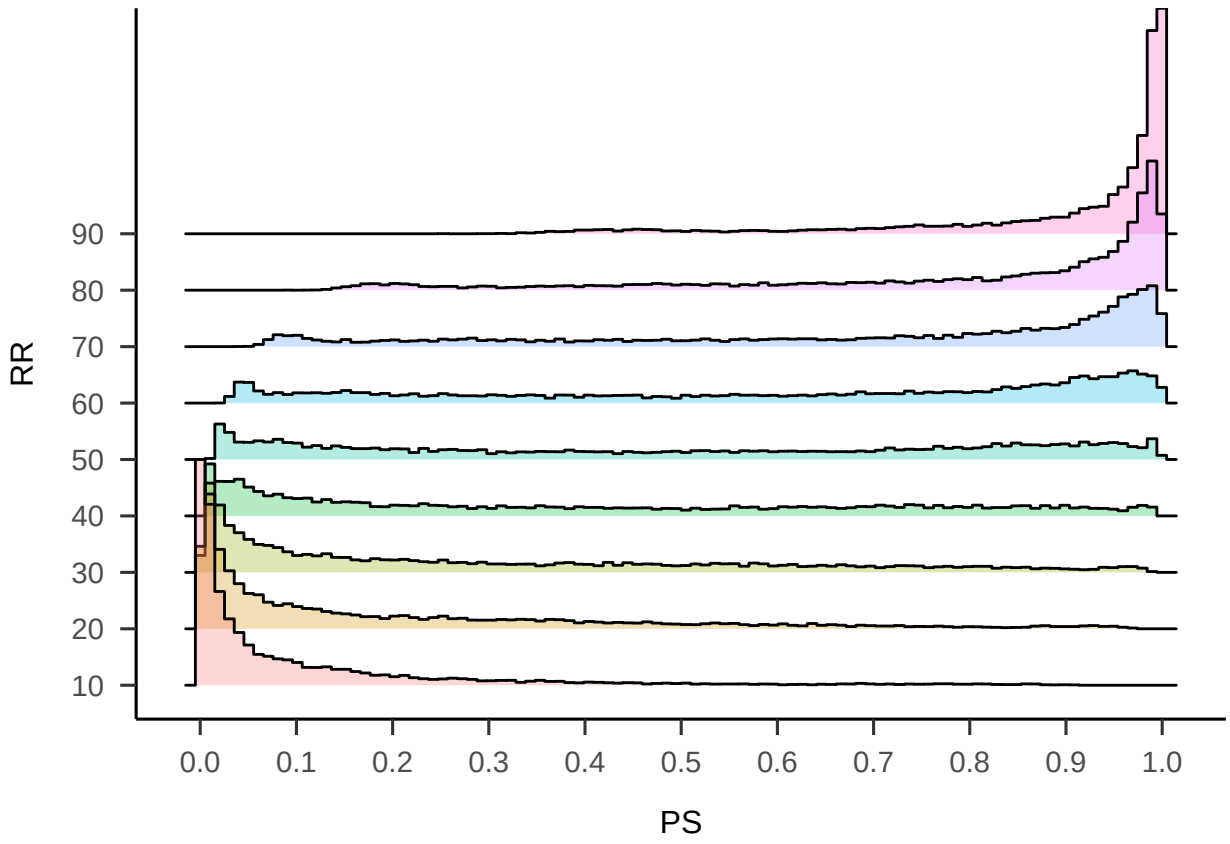


Figure 1. Distributions of Participation Propensity Scores



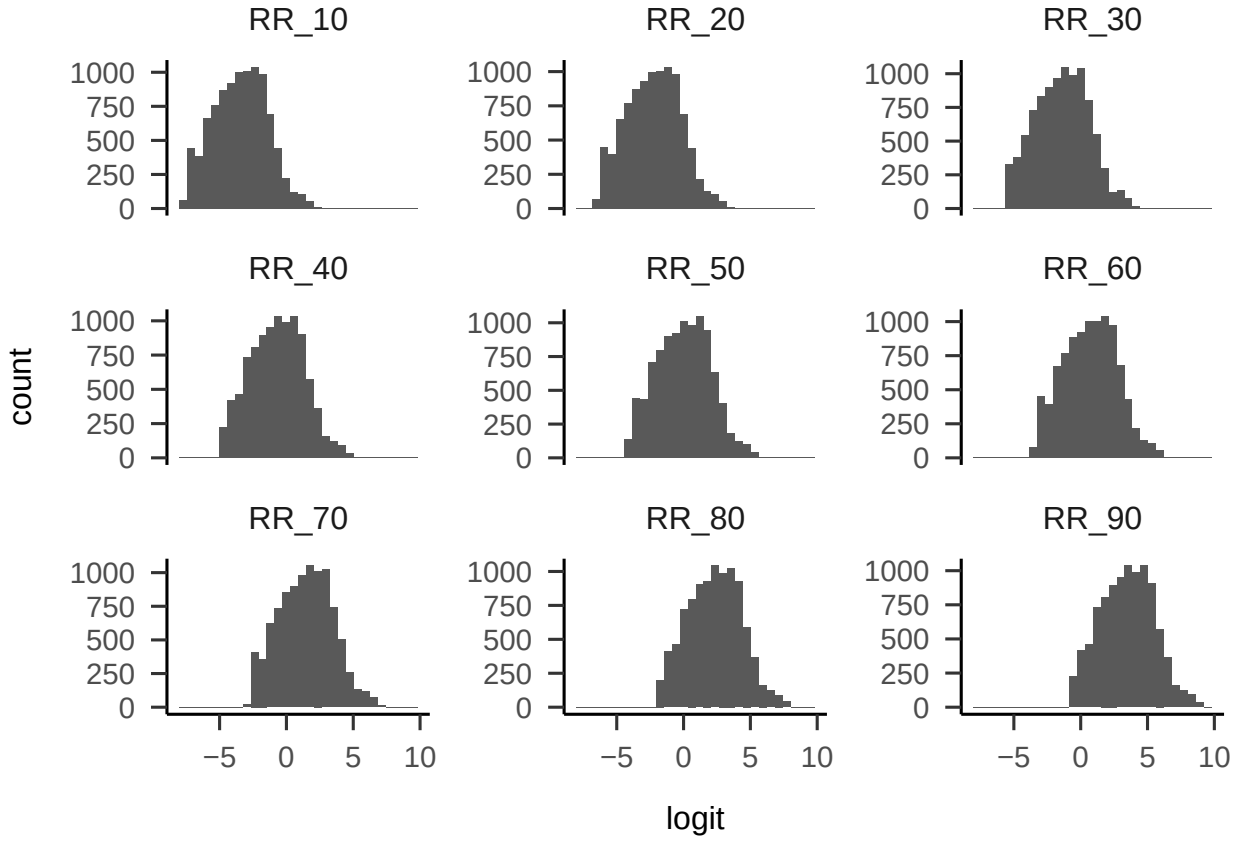
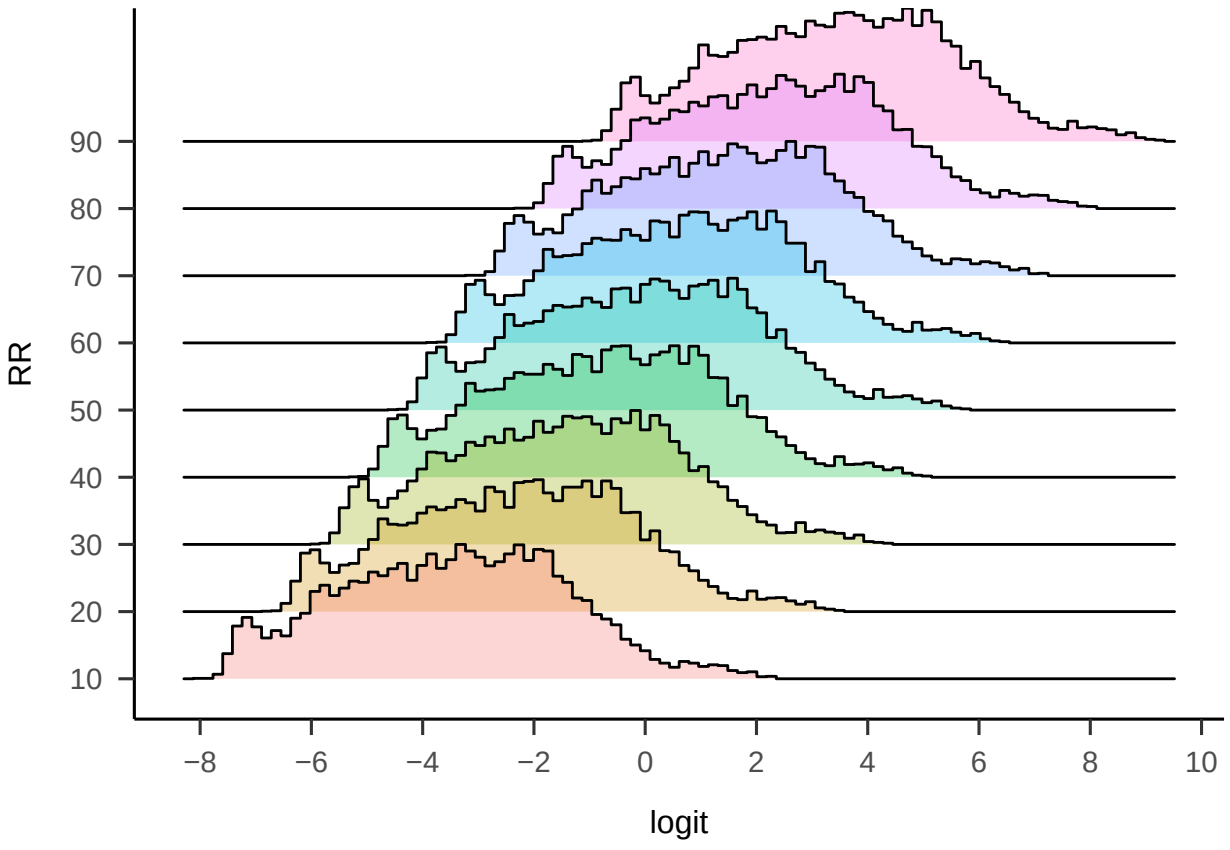
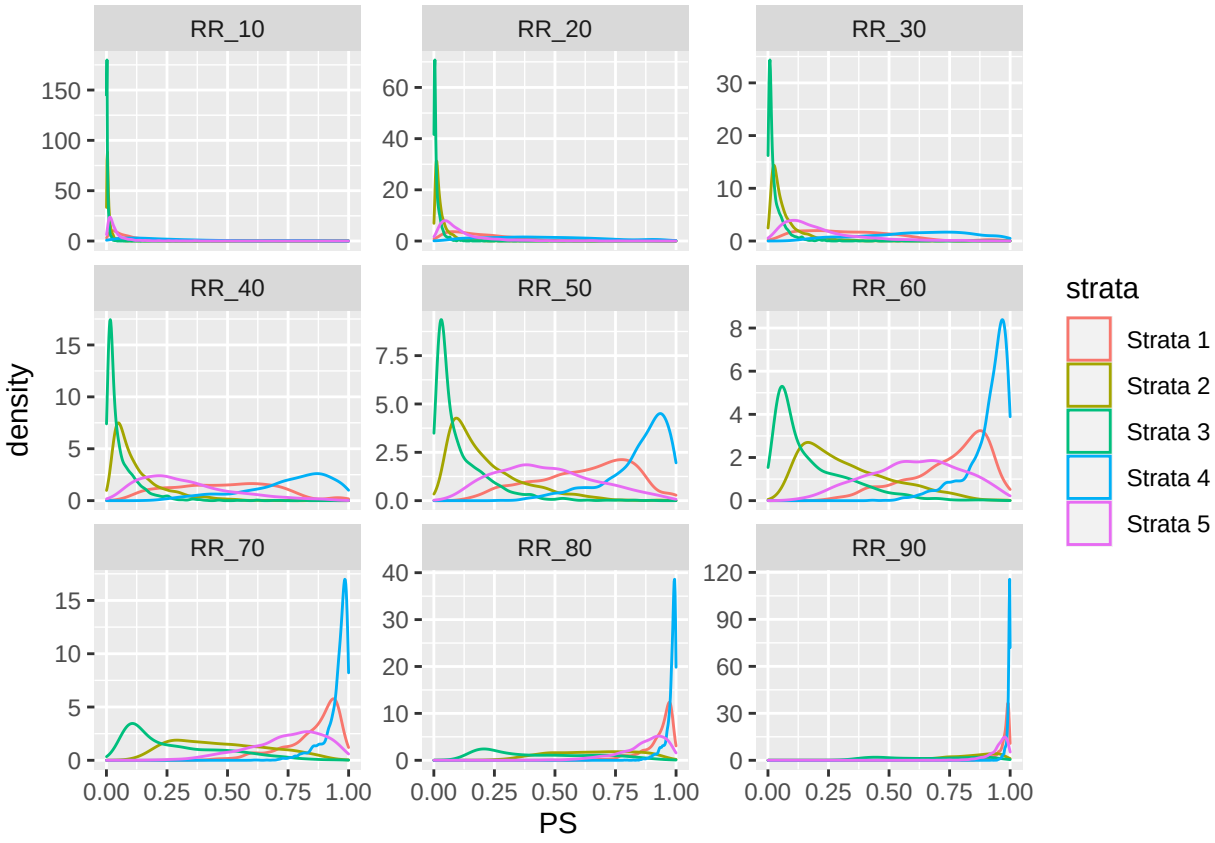


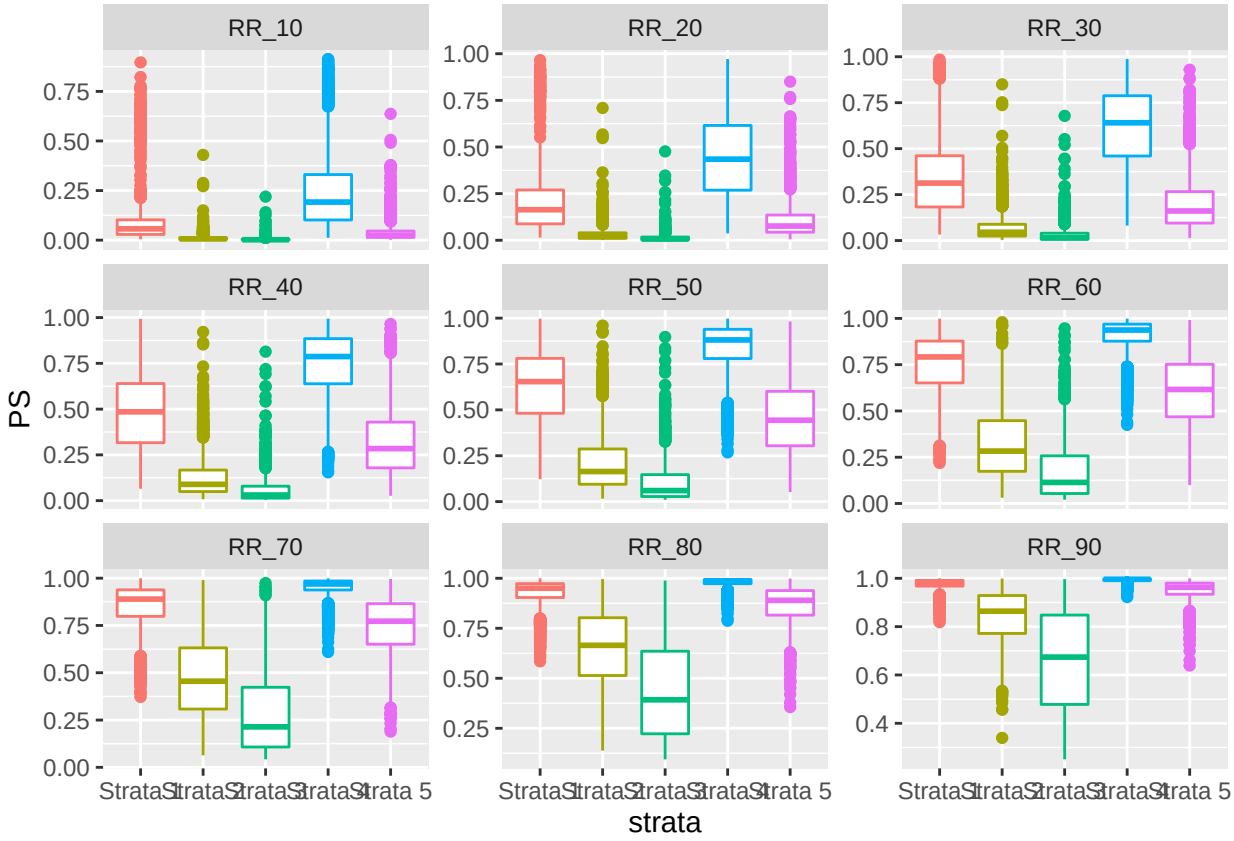
Figure 2. Distributions of Participation Propensity Scores



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




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



Results

Generalizability

B Index.

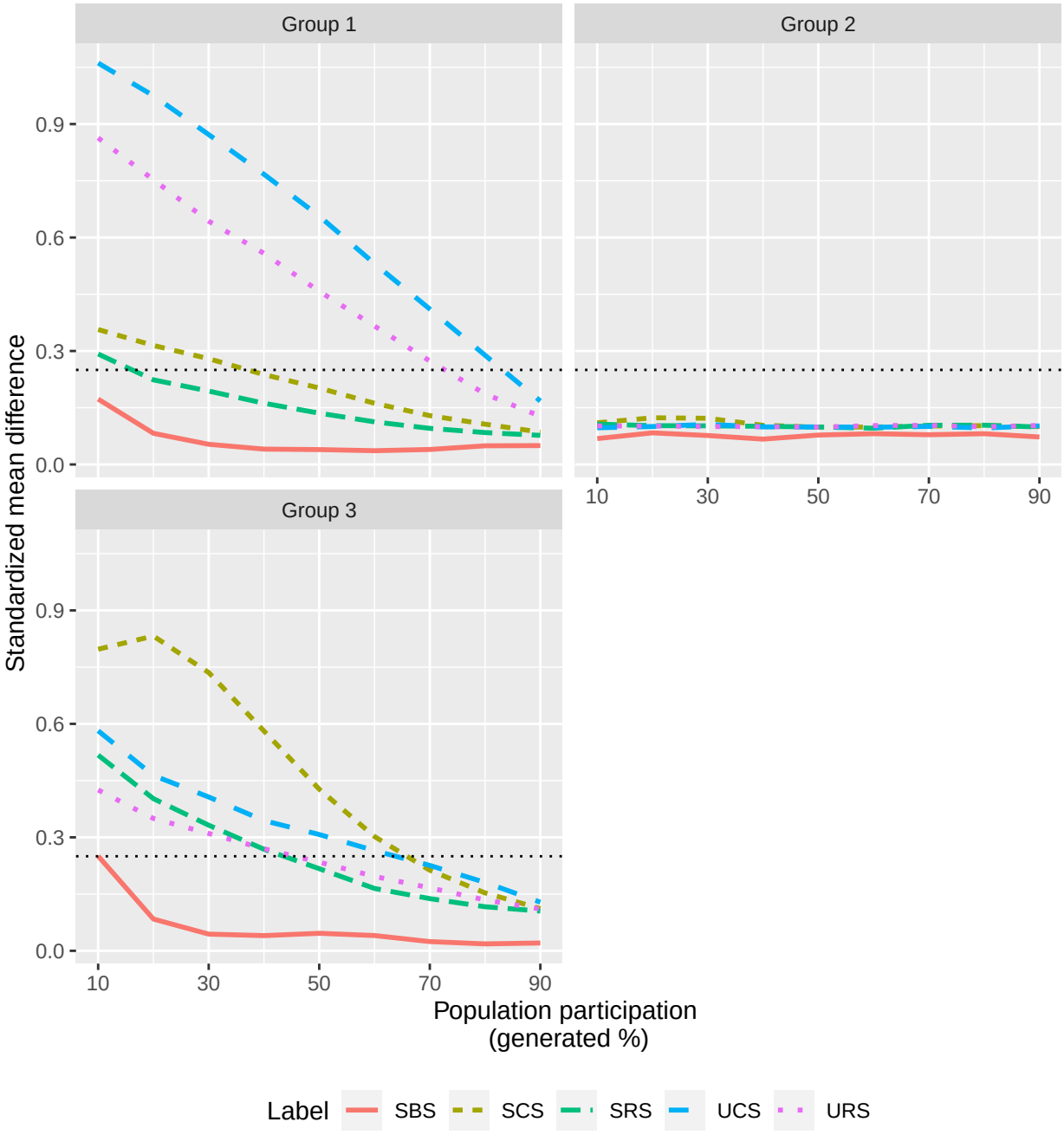
```
## `summarise()` regrouping output by 'sample_method', 'RR' (override with `.groups` arg
```

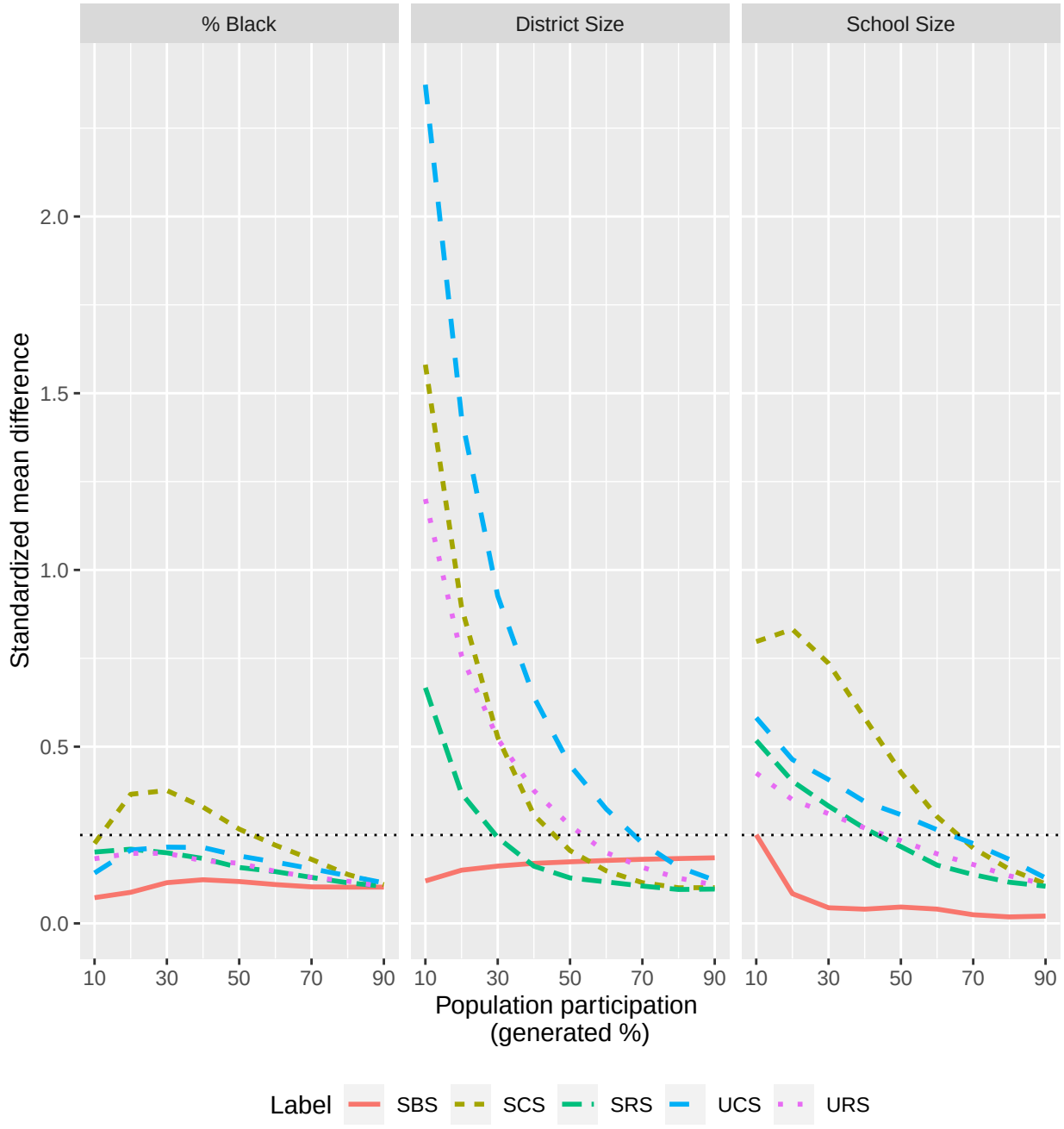
Standardized mean differences.

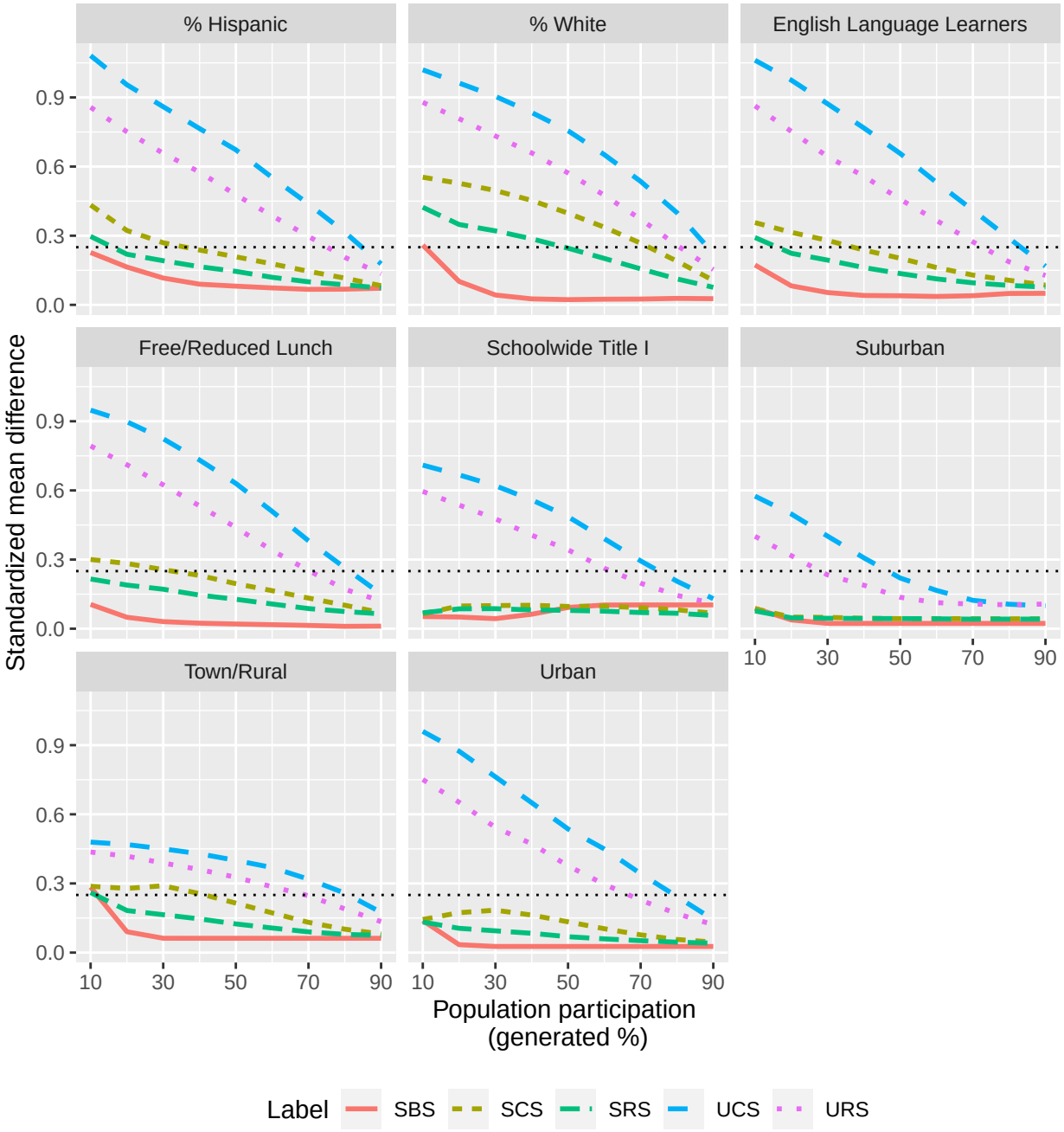
```
## Joining, by = "var"
```

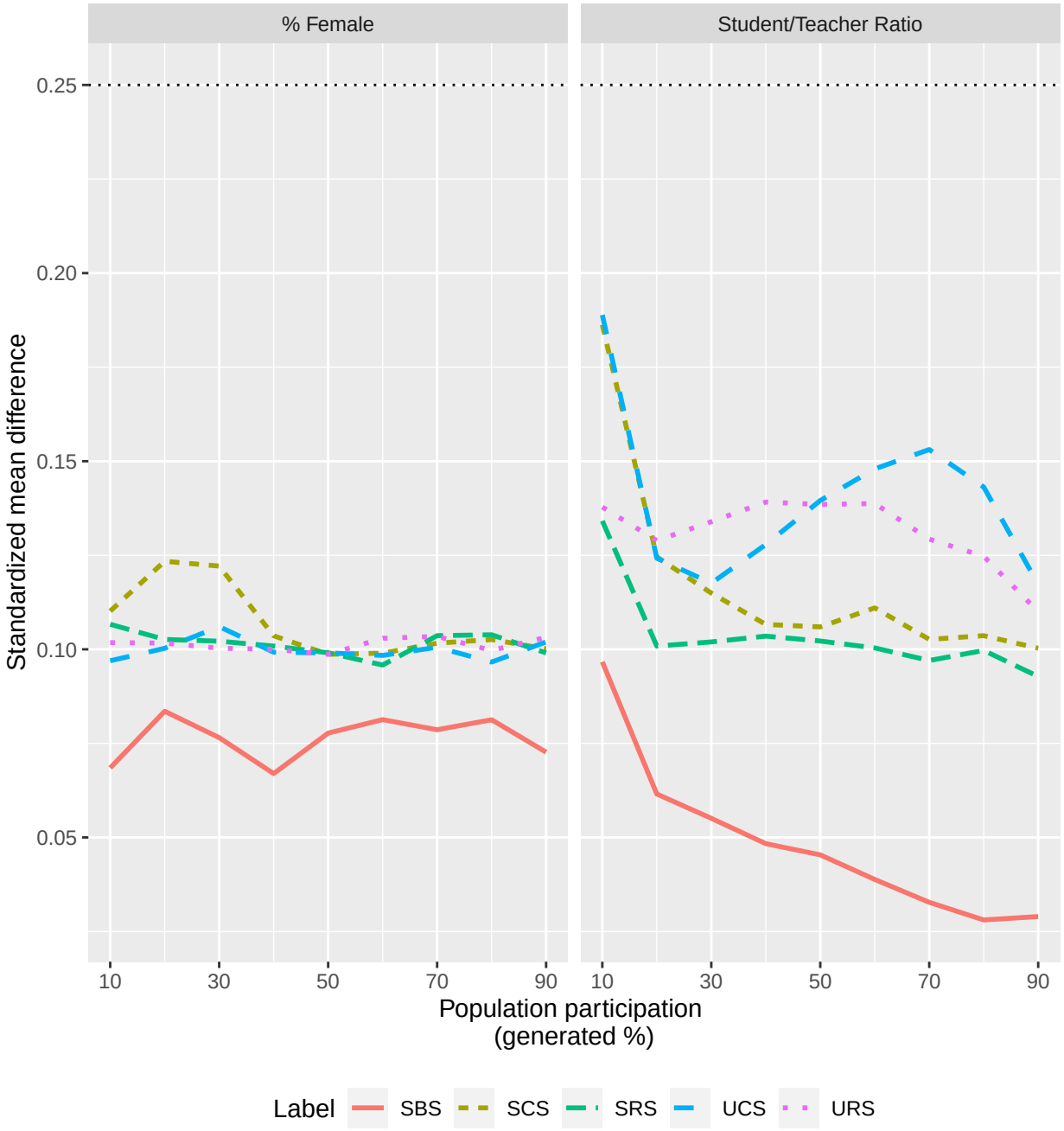
```
## `summarise()` regrouping output by 'sample_method', 'RR', 'var' (override with `.grou
```

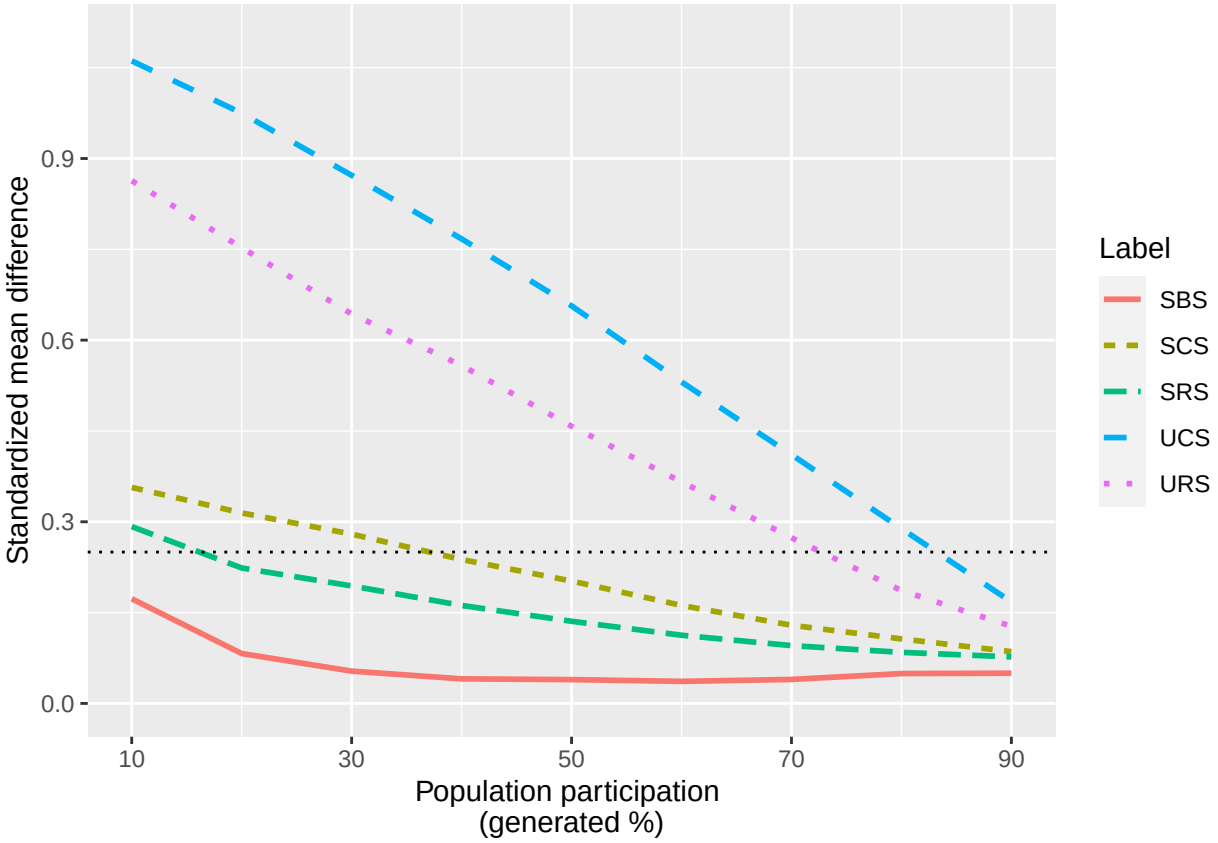
```
## Joining, by = "vnames"
## Joining, by = "vnames"
```

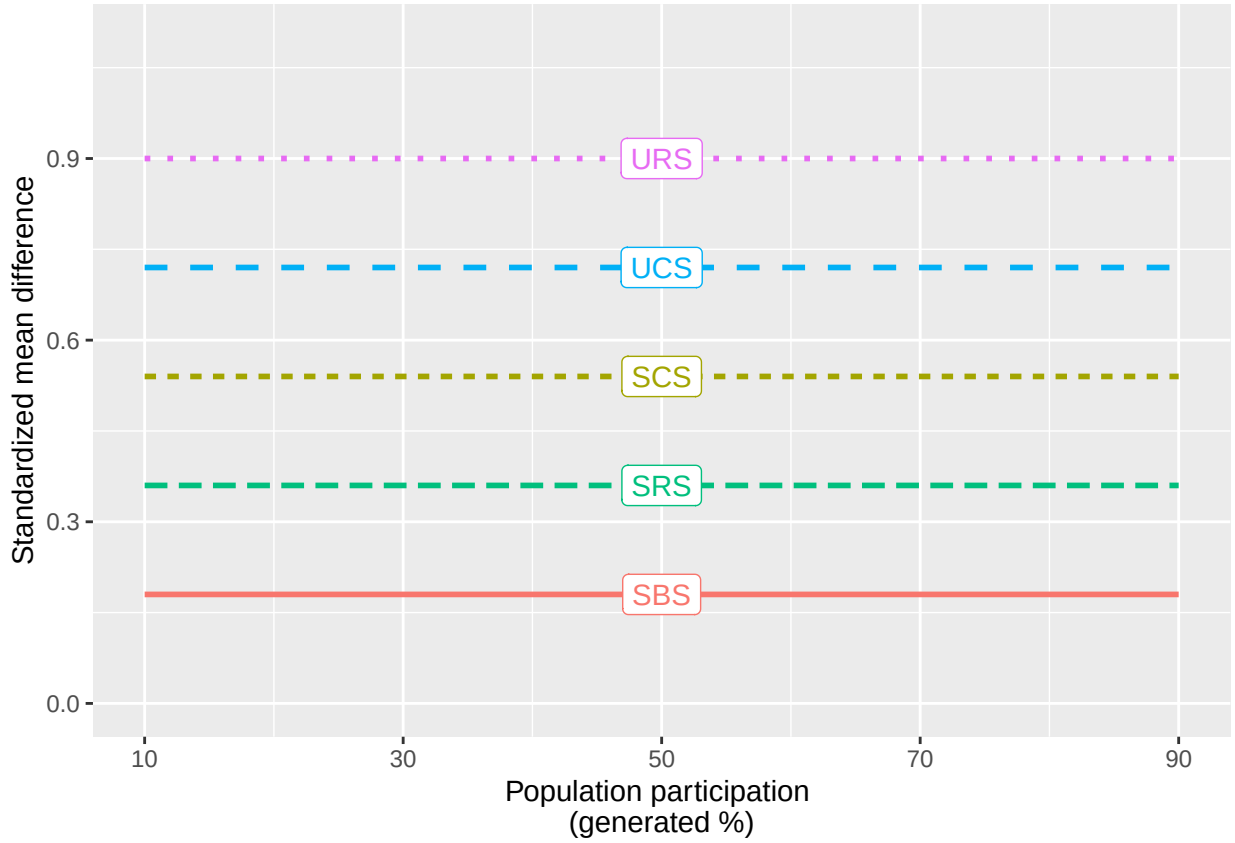

Group 1
Group 2
Group 3
Standardized mean difference
0.0
0.3
0.6
0.9
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS

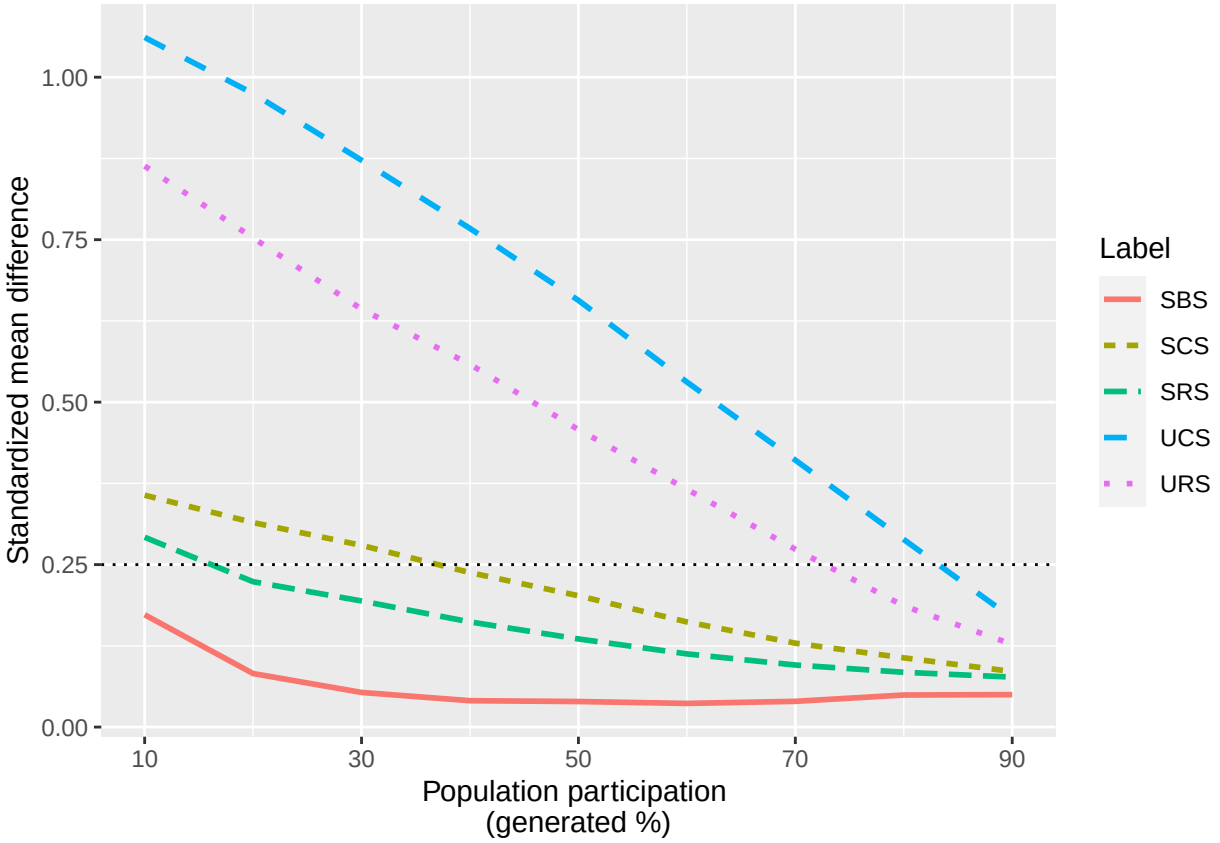

% Black
District Size
School Size
Standardized mean difference
2.0
1.5
1.0
0.5
0.0
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS

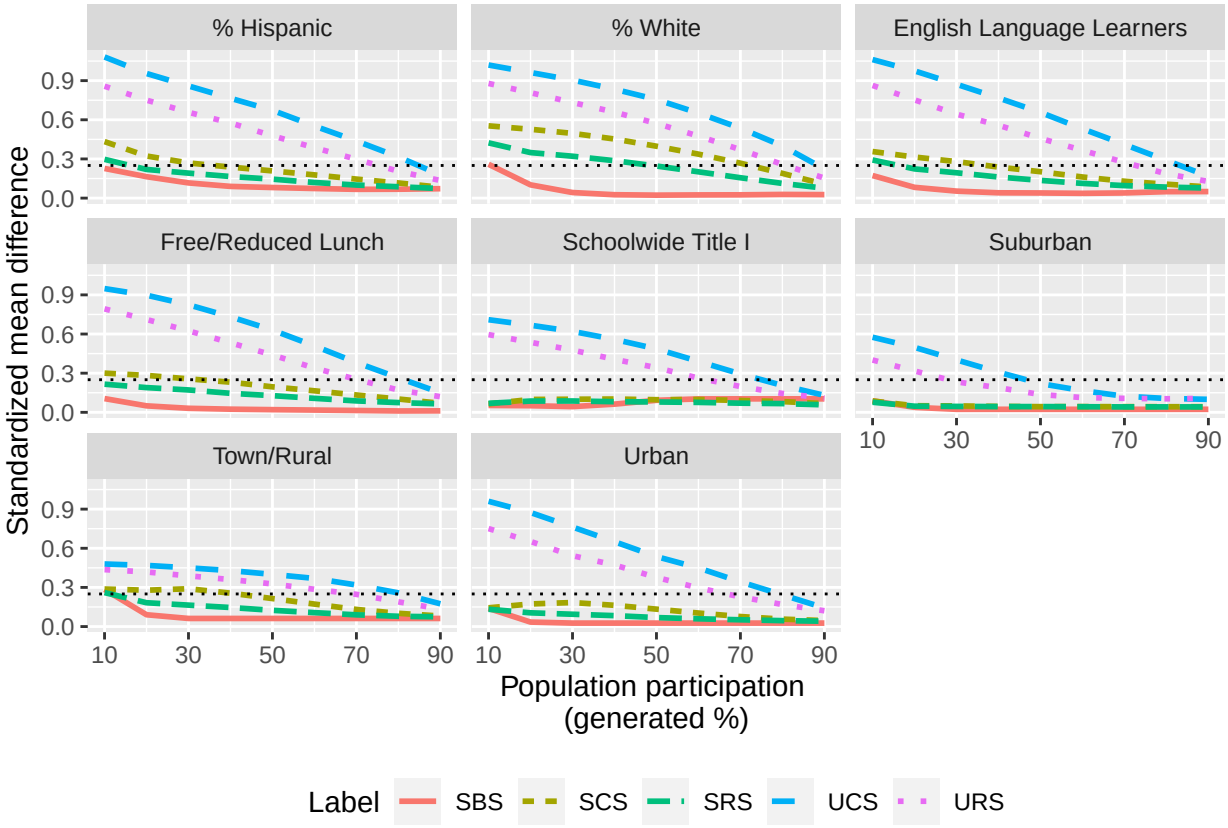

% Hispanic
% White
English Language Learners
Free/Reduced Lunch
Schoolwide Title I
Suburban
Town/Rural
Urban
Standardized mean difference
0.0
0.3
0.6
0.9
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS

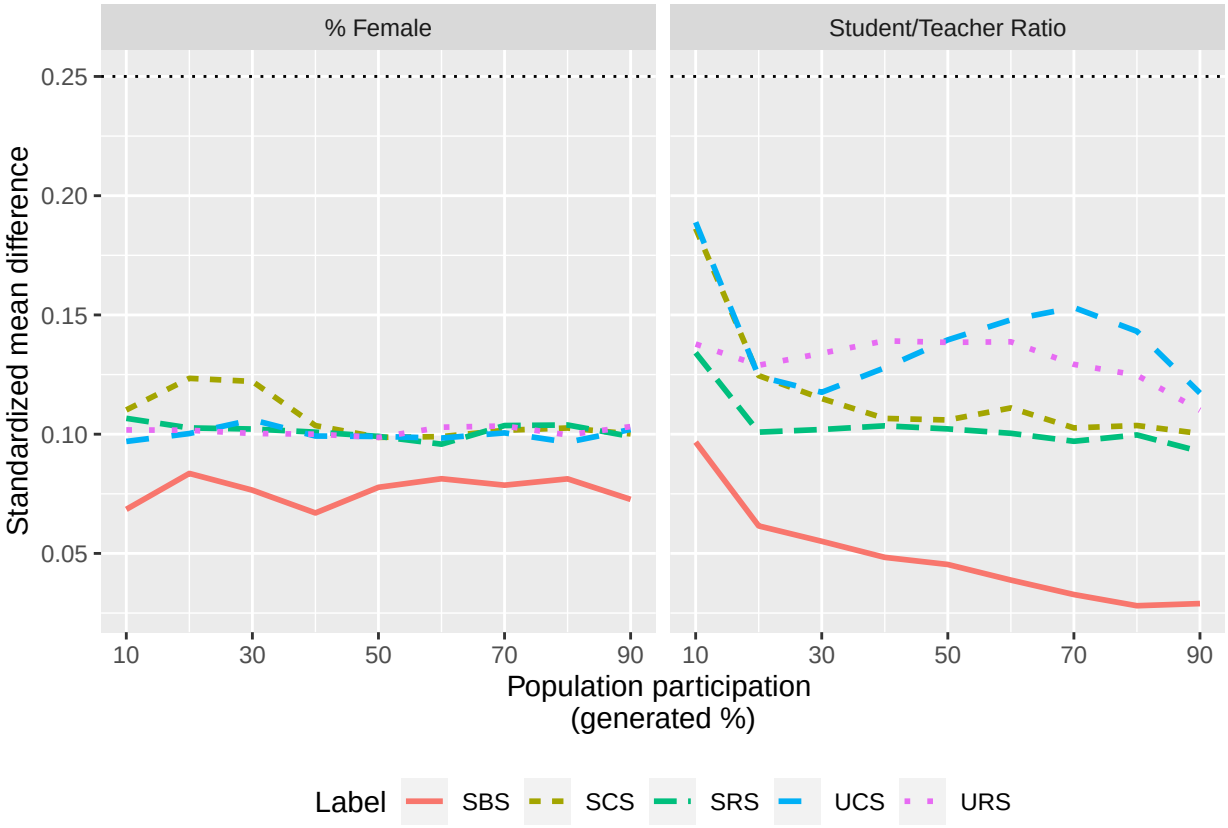

% Female
Student/Teacher Ratio
Standardized mean difference
0.25
0.20
0.15
0.10
0.05
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS


Standardized mean difference
0.9
0.6
0.3
0.0
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS


URS
UCS
SCS
SRS
SBS
0.9
0.6
0.3
0.0
Standardized mean difference
10
30
50
70
90
Population participation
(generated %)


1.00
0.75
0.50
0.25
0.00
Standardized mean difference
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS


% Hispanic
% White
English Language Learners
Free/Reduced Lunch
Schoolwide Title I
Suburban
Town/Rural
Urban
Standardized mean difference
0.0
0.3
0.6
0.9
10
30
50
70
90
Population participation
(generated %)
Label
SBS
SCS
SRS
UCS
URS



Examples for presentations

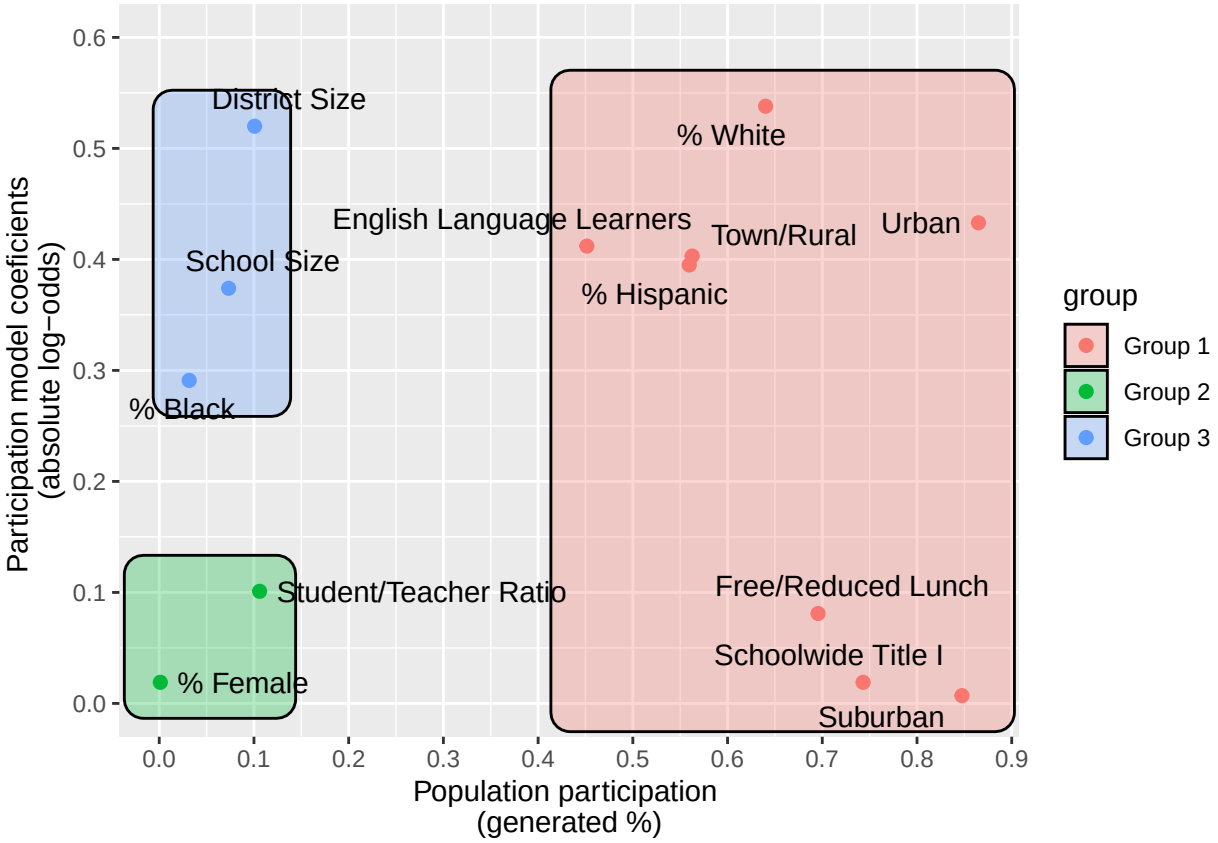
V-ratio and Log odds.

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

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

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

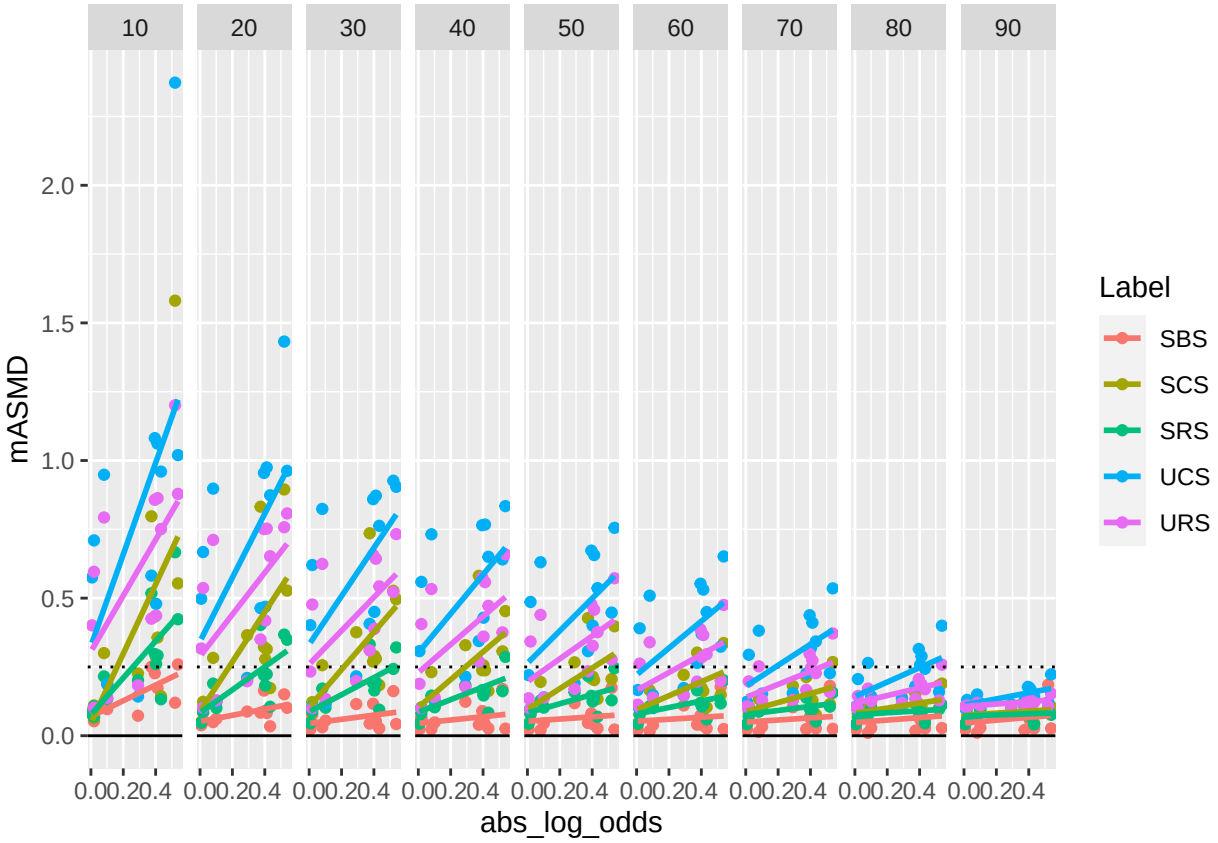
```
## Joining, by = "vnames"
```



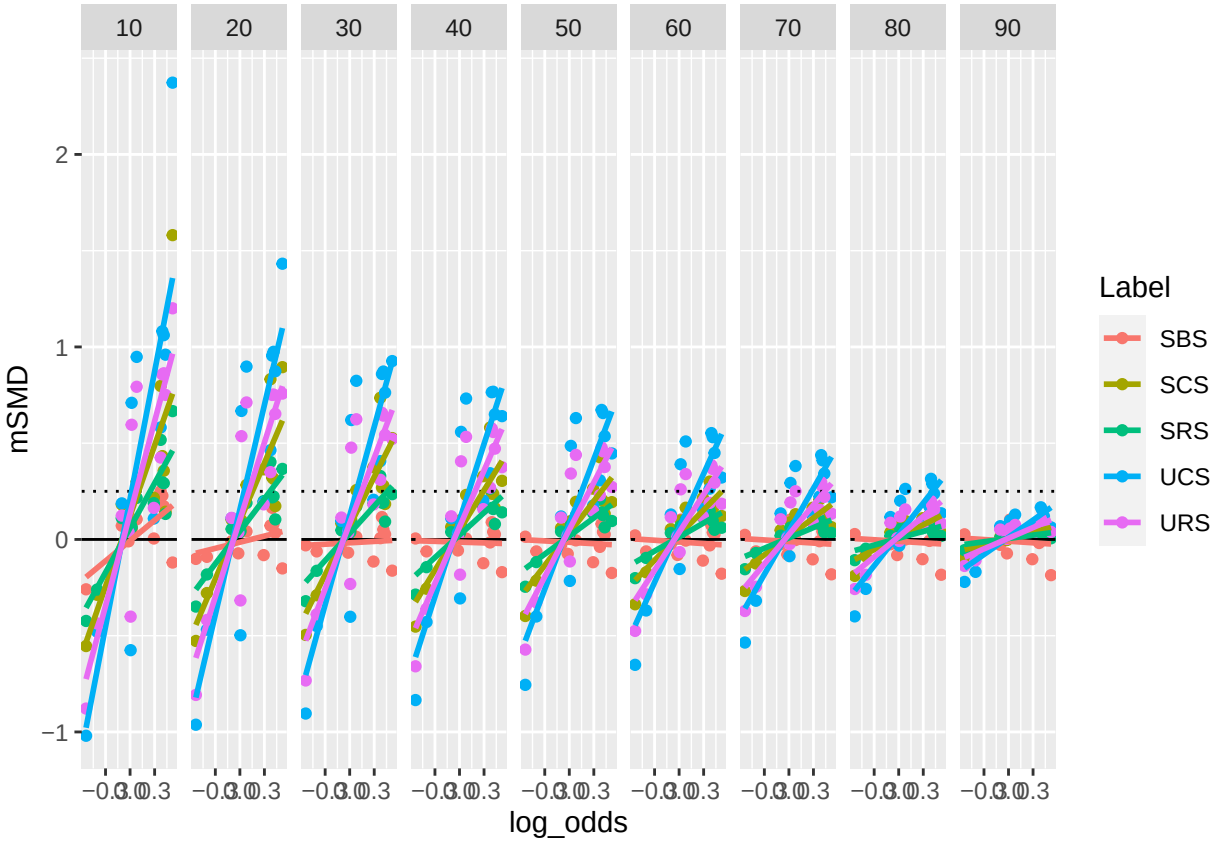
Test.

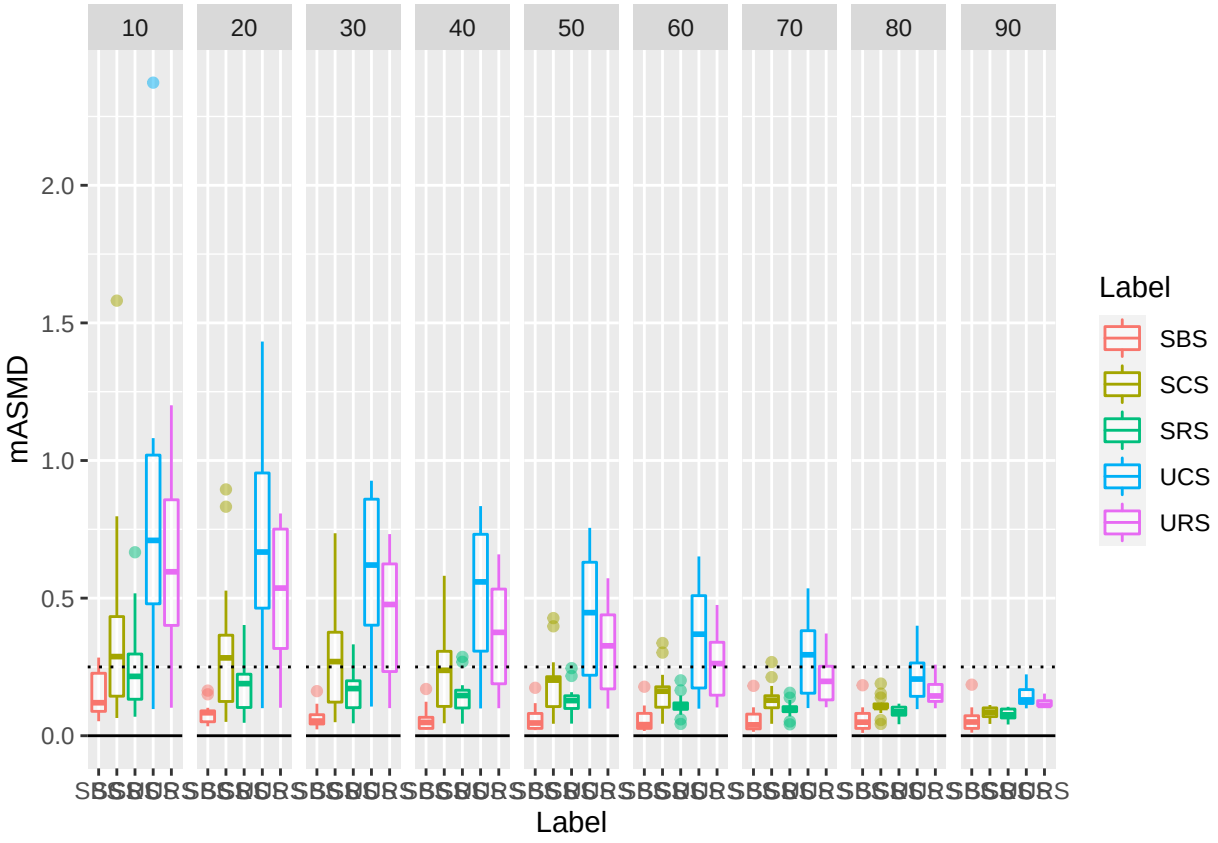
```
## Joining, by = c("vnames", "log_odds")
```

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




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


10
20
30
40
50
60
70
80
90
2
1
0
−1
mSMD
log_odds
Label
SBS
SCS
SRS
UCS
URS



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)

——————————————— R^2 0.04 0.39 0.50

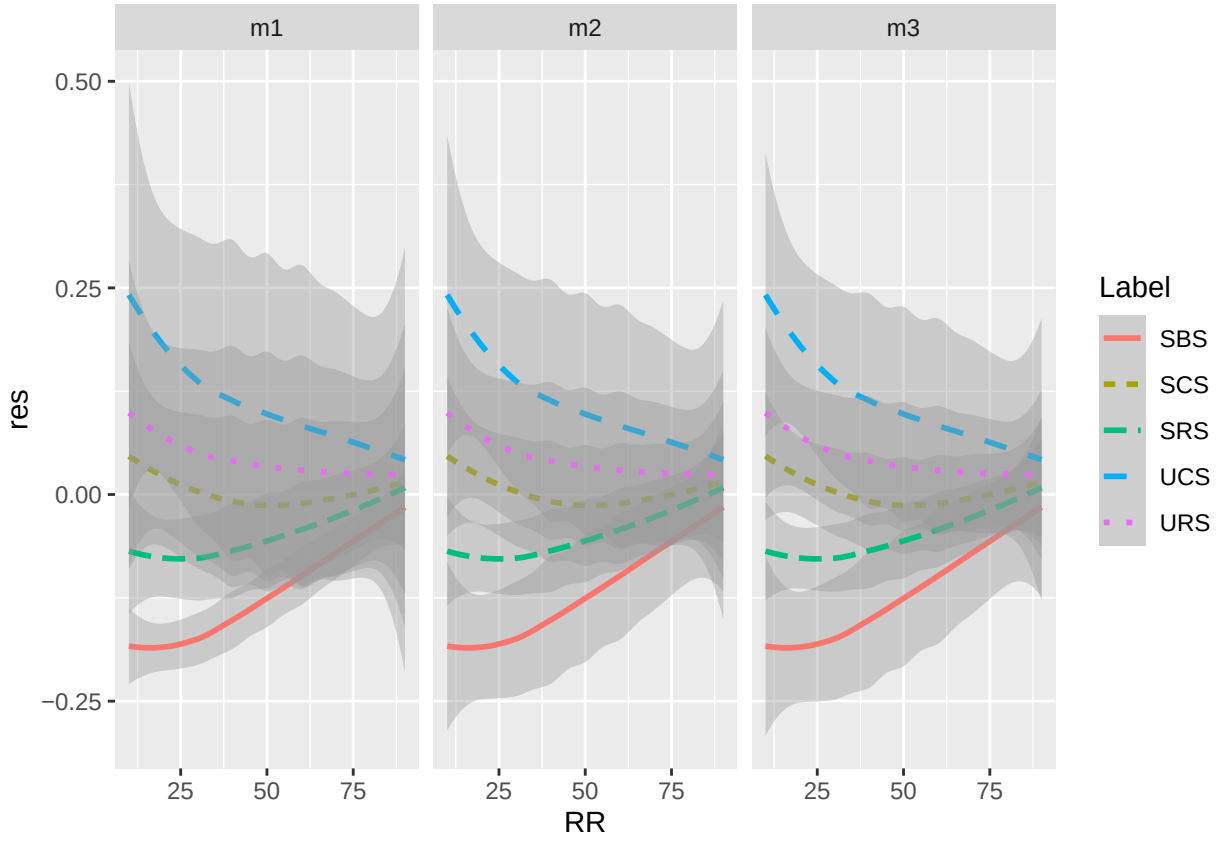
Adj. R^2 0.04 0.38 0.49

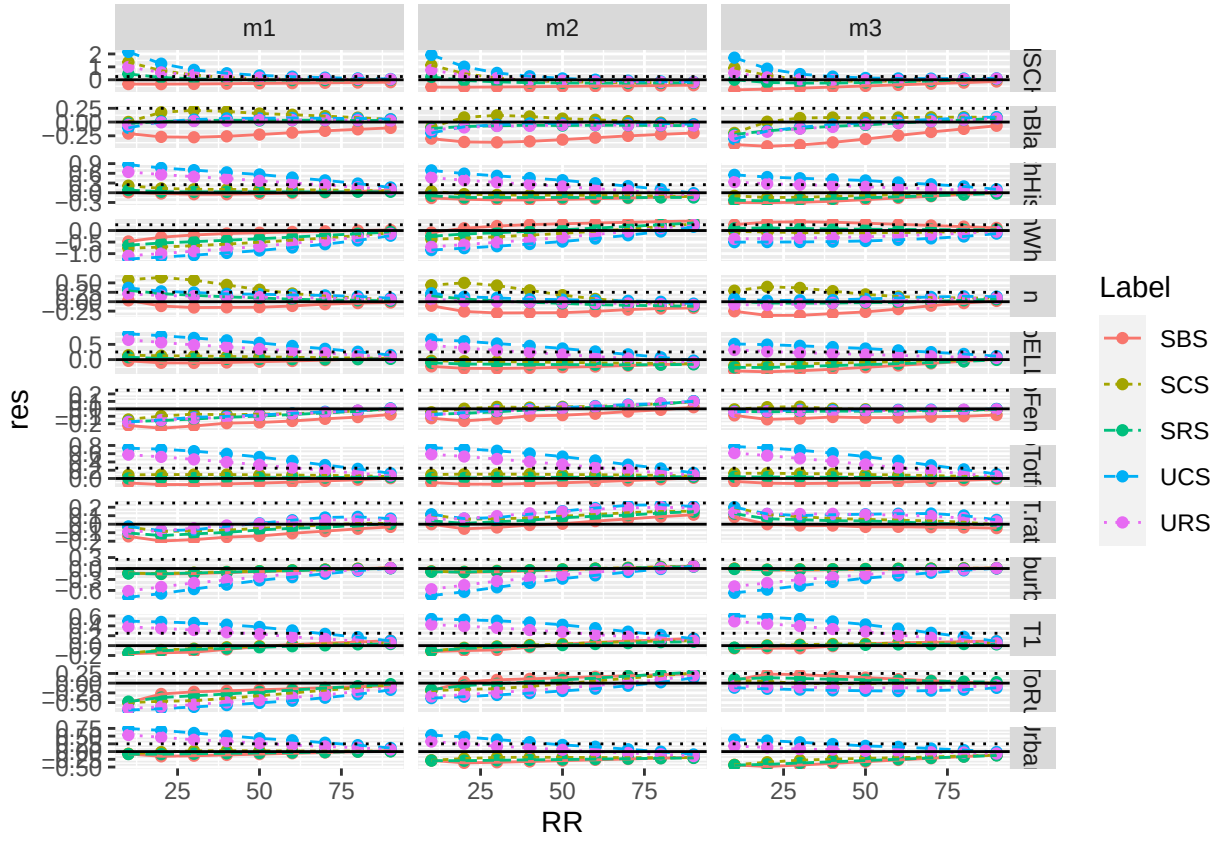
Num. obs. 585 585 585

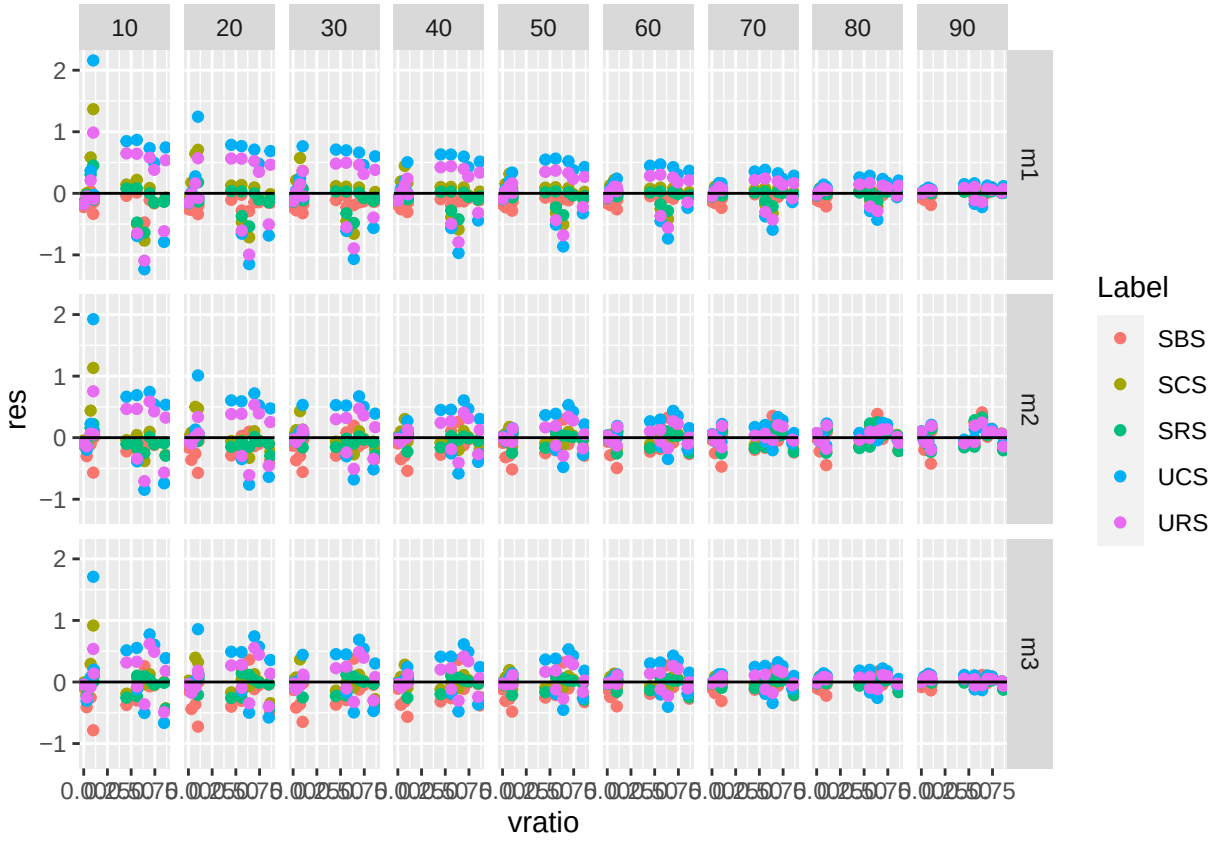
==========

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

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


m1
m2
m3
0.50
0.25
0.00
−0.25
res
25
50
75
RR
Label
SBS
SCS
SRS
UCS
URS

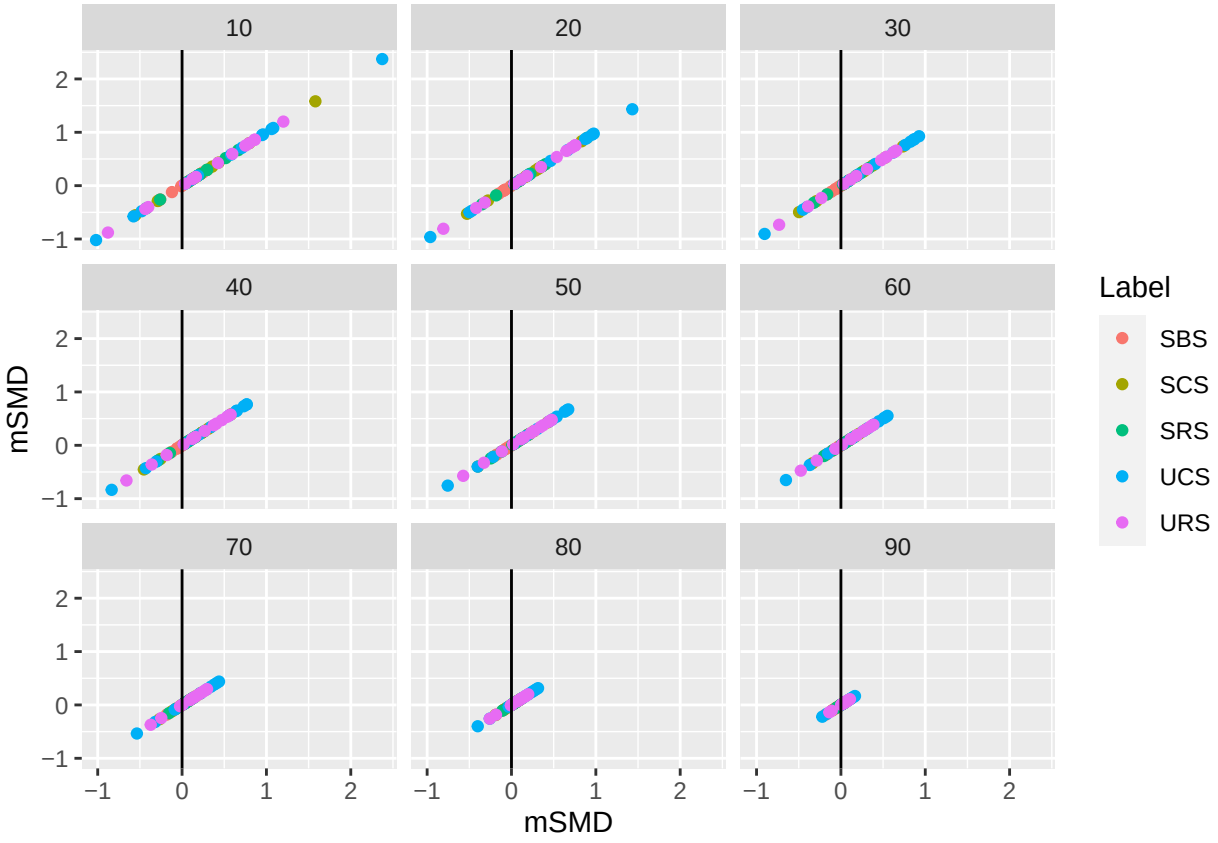

m1
m2
m3
res
RR
25
50
75
Label
SBS
SCS
SRS
UCS
URS

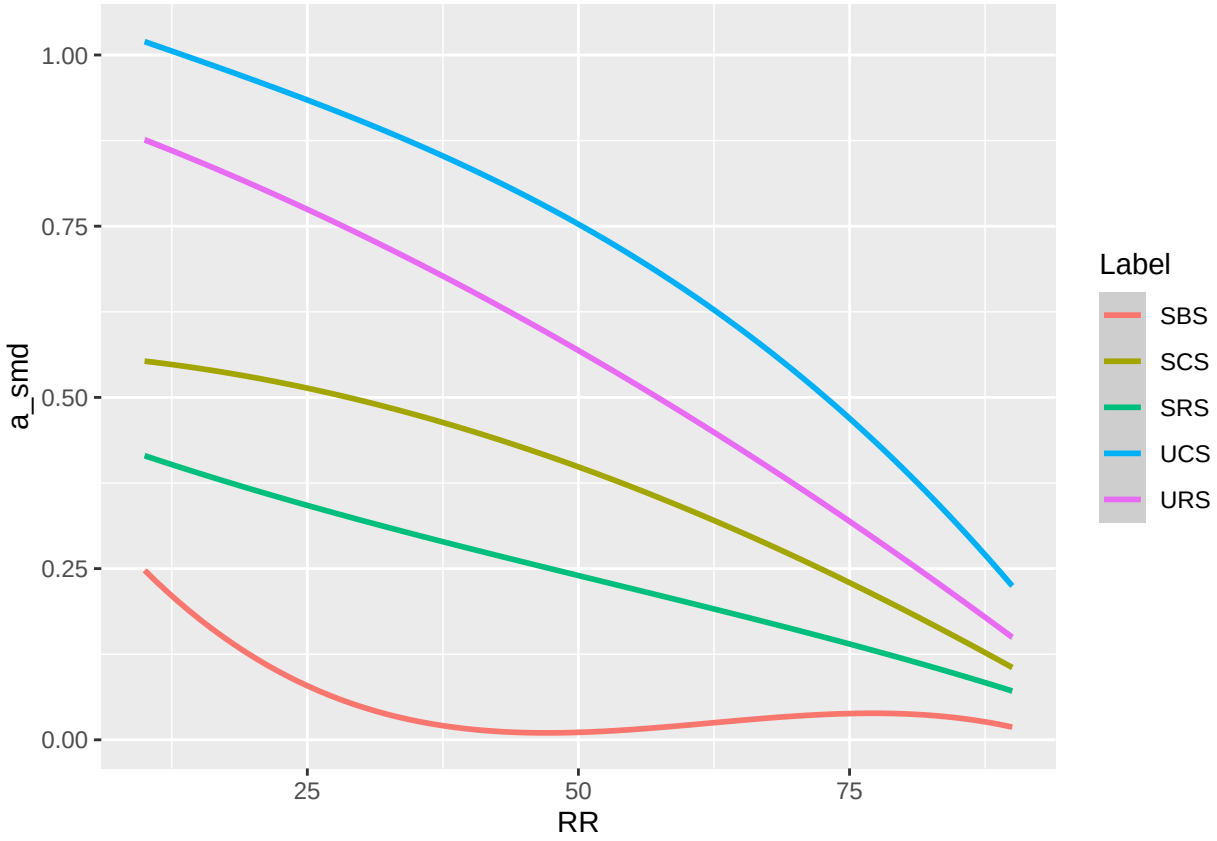


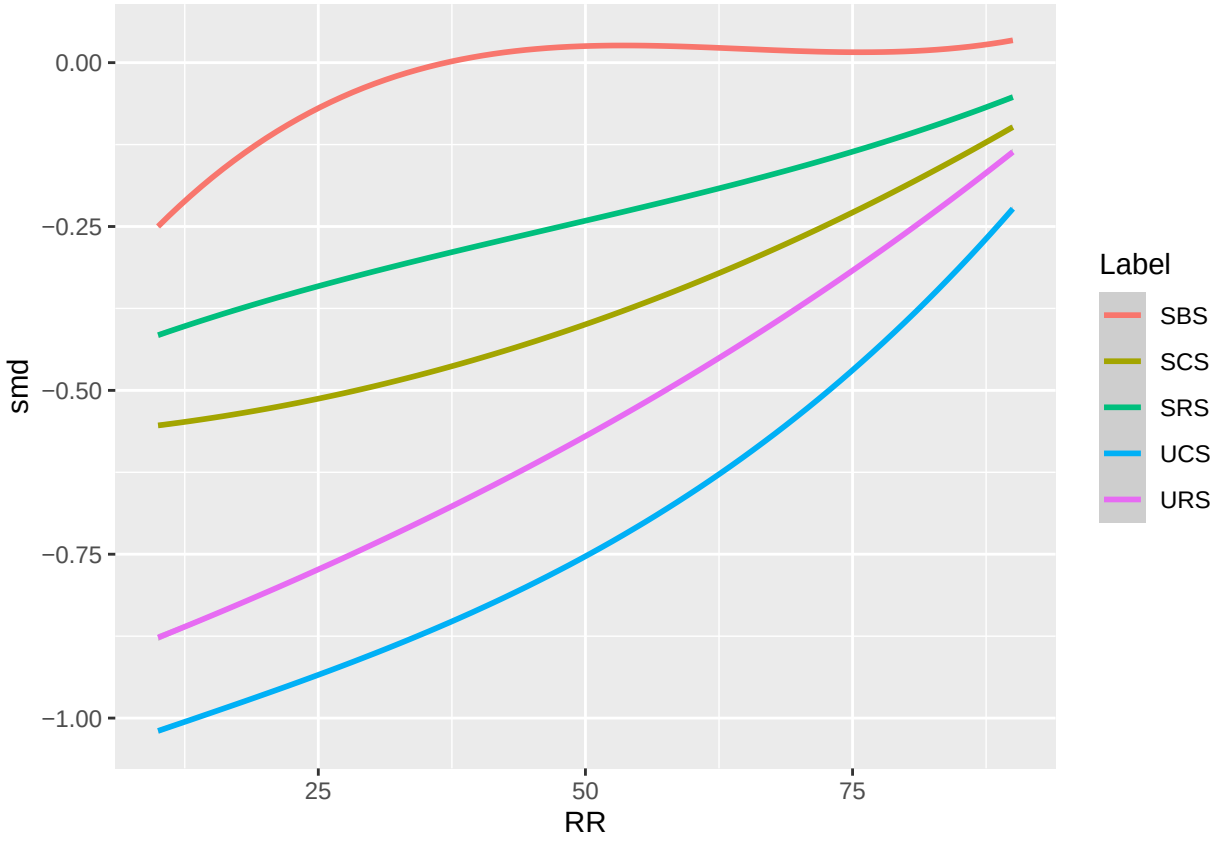
Test 2

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

```
## `summarise()` regrouping output by 'sample_method', 'RR', 'var' (override with `.grou
```


10
20
30
40
50
60
70
80
90
mSMD
mSMD
Label
SBS
SCS
SRS
UCS
URS


1.00
0.75
0.50
0.25
0.00
a_smd
25
50
75
RR
Label
SBS
SCS
SRS
UCS
URS



Feasibility

Sampling Difficulty.

```
## `summarise()` regrouping output by 'sample_method', 'RR', 'measure', 'Label' (overrid
```

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

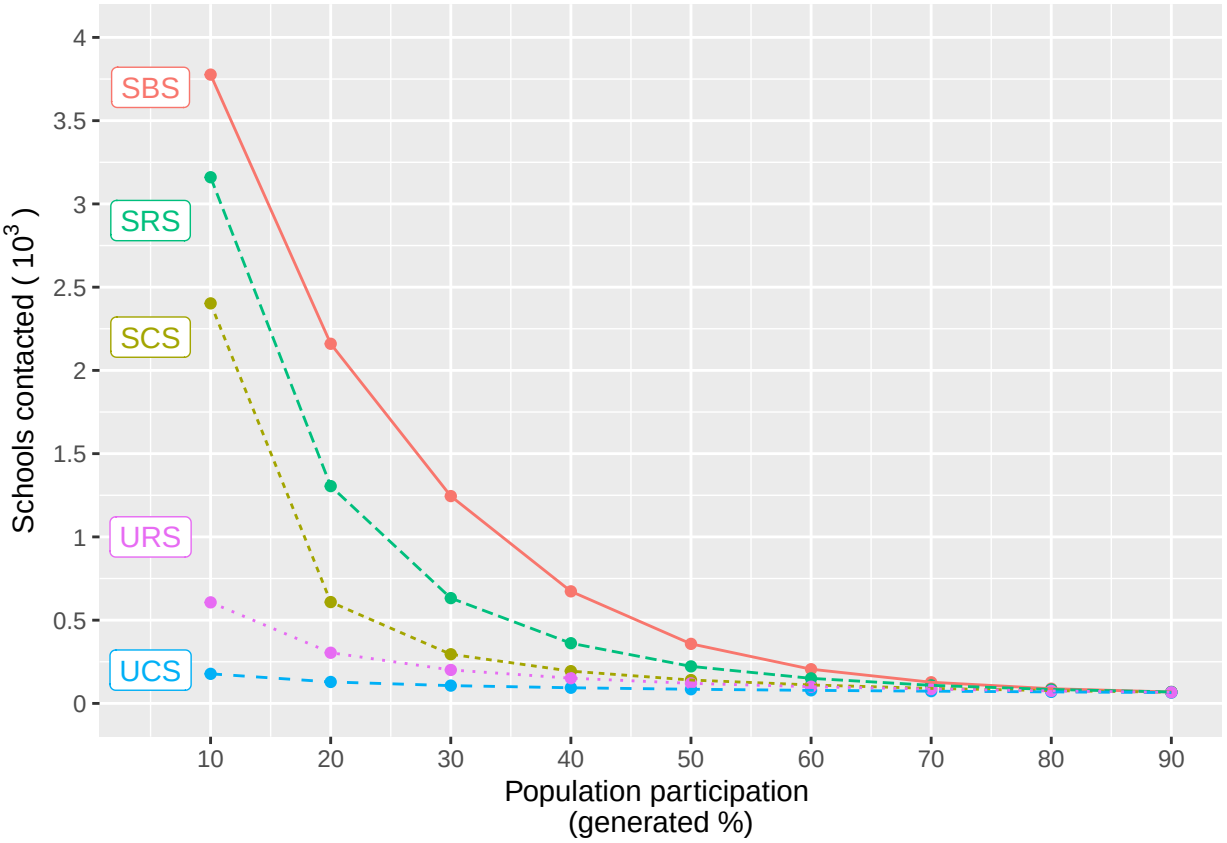
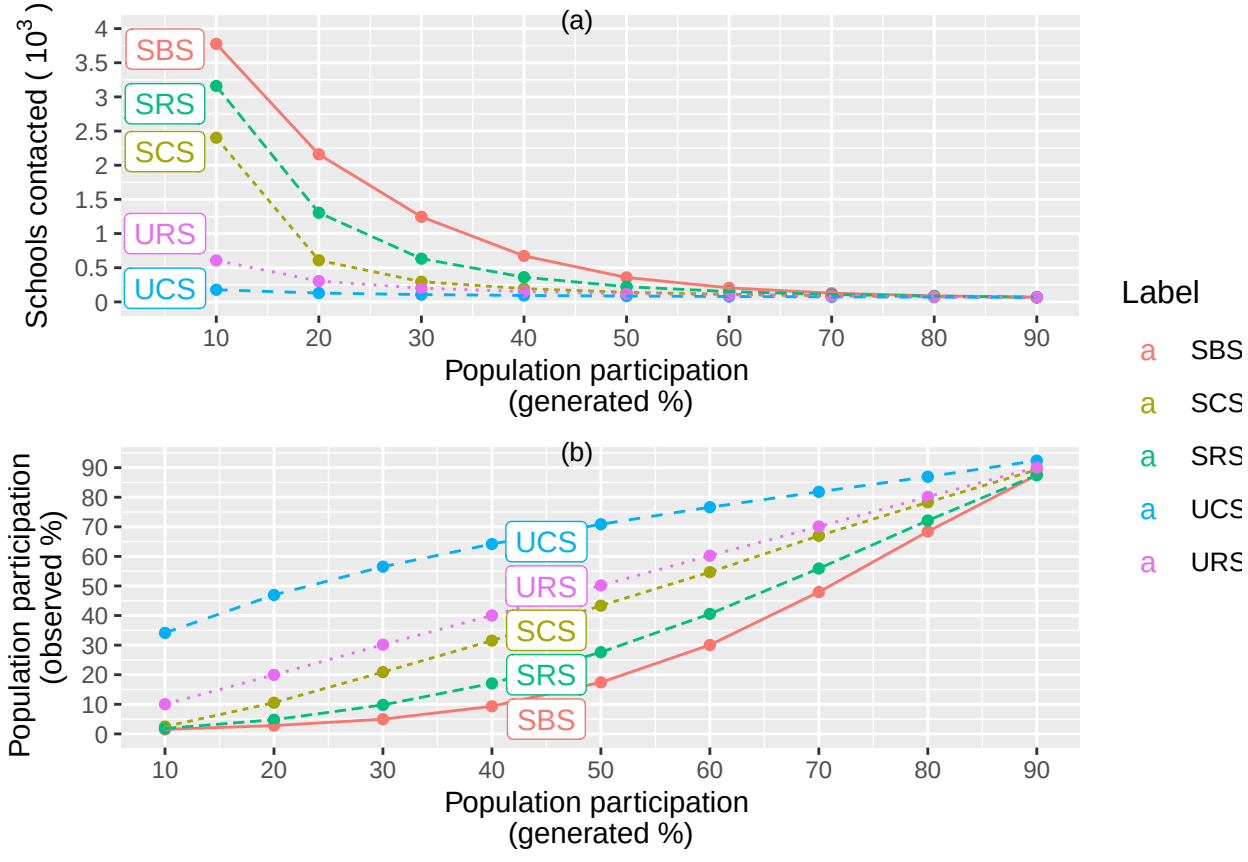


Figure 3. Schools Contacted



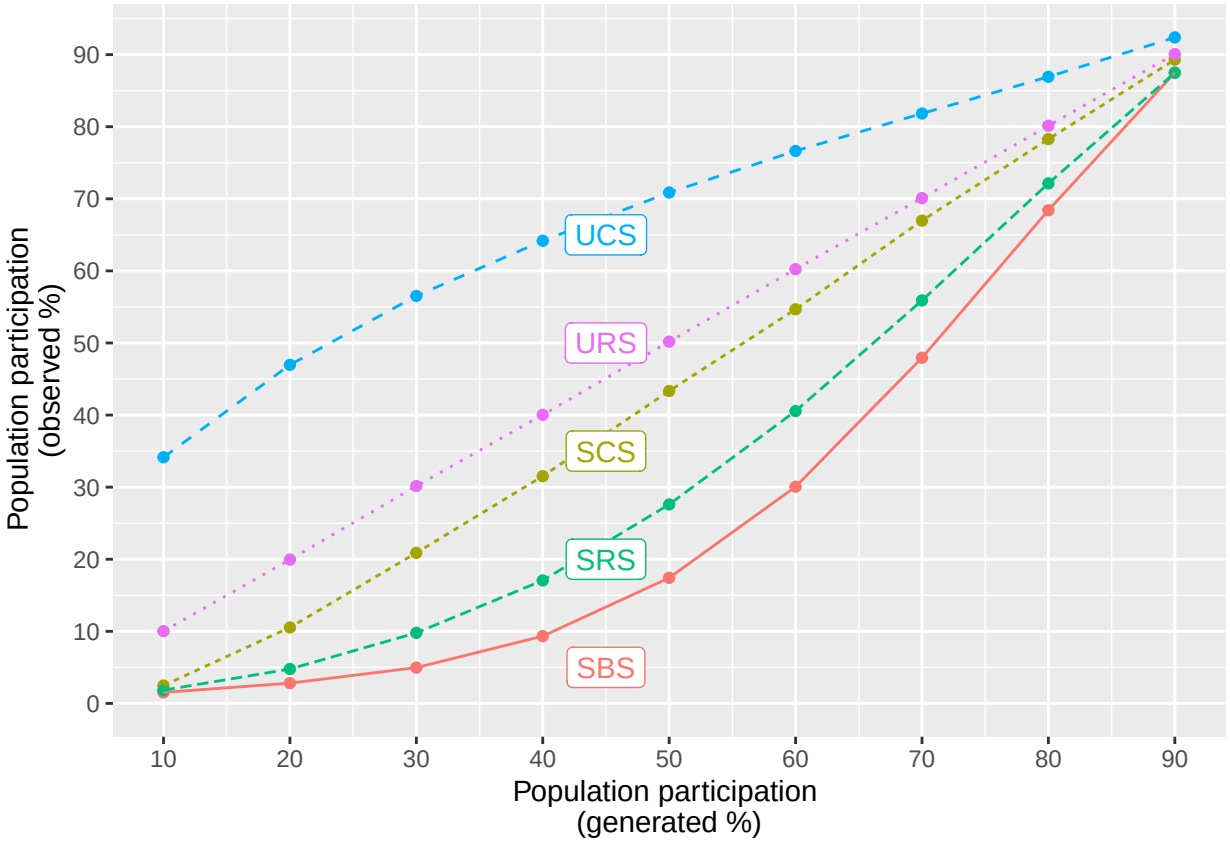
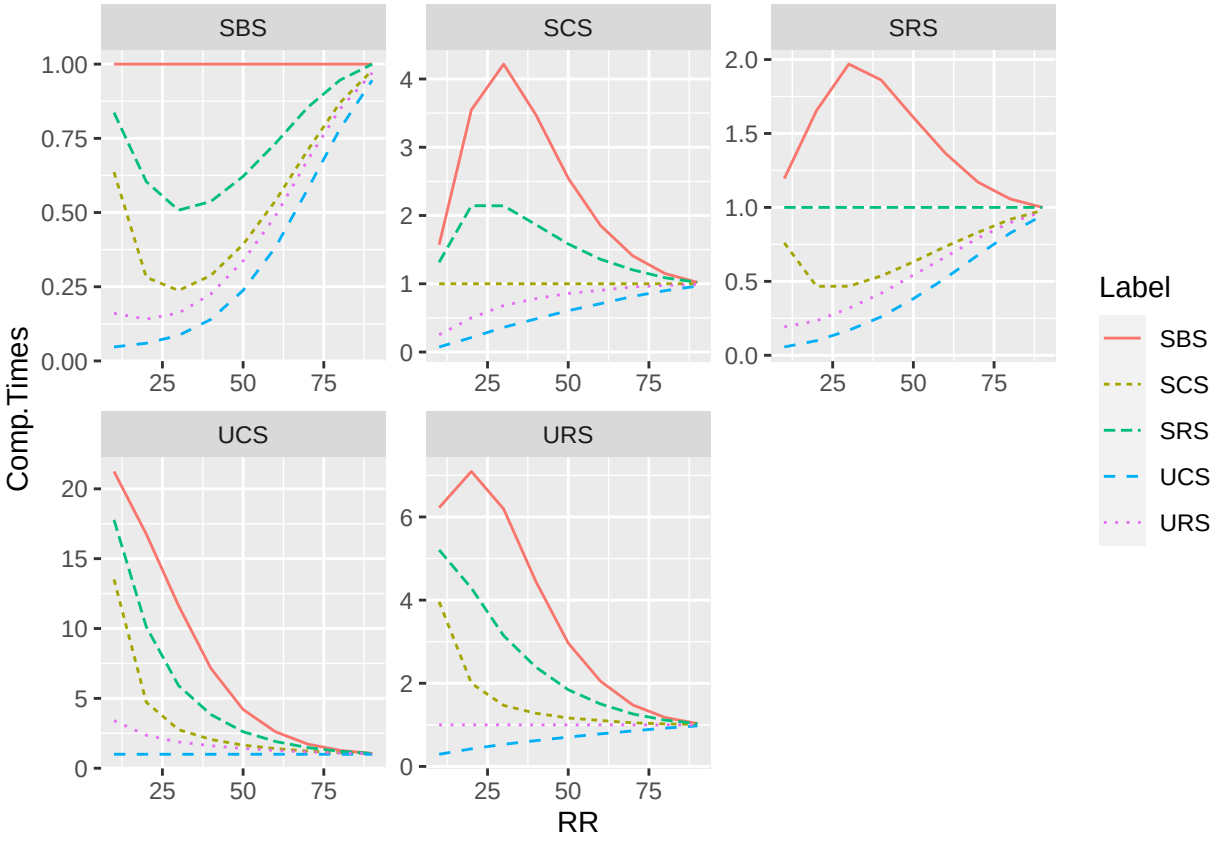


Figure 4. Sampling response rates

Relative Performance.

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



Gini Plot.

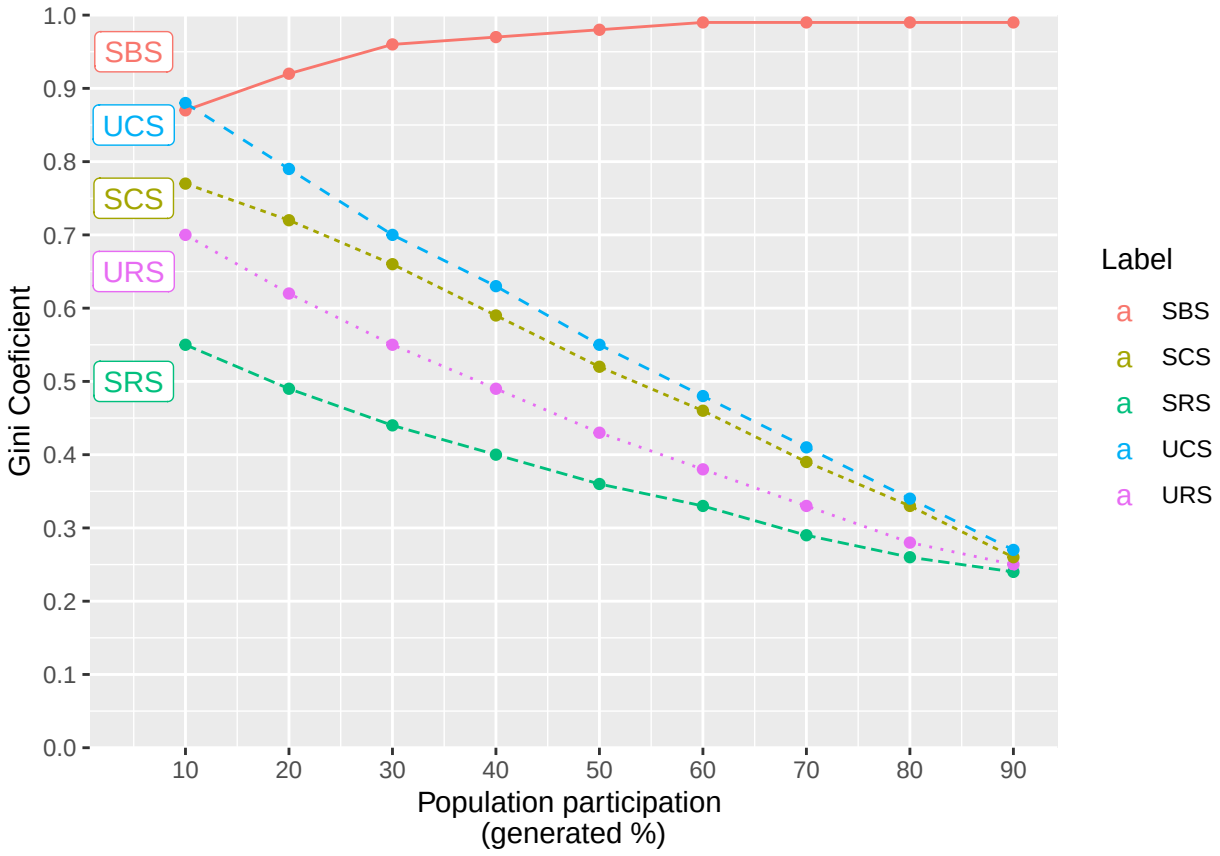
```
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
```

```
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## Joining, by = "DSID"
## 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"
## 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"
## 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’, ’Label’ (override with ‘.groups‘
```



Export Plots

APPAM PLOTS

B-index

Sample

Non-Response

Combined

Individual

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```