Summary of 0707 binarypinc int

Setup

$$\text{logit}(Pr(Y_i = 1|Z, X)) = \beta_0 + \alpha_{Za[i]} + \alpha_{Zb[i]} + \alpha_{Zc[i]} + \alpha_{X[i]}$$

$$\begin{aligned}
\alpha_{Za[i]} &= (1.37, -0.56, 0.36, 0.63, 0.40) \\
\alpha_{Zb[i]} &= (-0.11, 1.51, -0.09, 2.02, -0.06) \\
\alpha_{Zc[i]} &= (0, 0.24) \\
\alpha_{X[i]} &= (0, -1.3)
\end{aligned}$$

$$logit(P(X_i = 1)) = \Gamma_0 + \gamma_{Za[i]} + \gamma_{Zb[i]} + \gamma_{Zc[i]} + \gamma_{Za[i],Zc[i]} + \gamma_{Zb[i],Zc[i]}$$

$$\begin{aligned}
\Gamma_0 &= -0.5 \\
\gamma_{Za[i]} &= (1.7, 0.25, 0.2, -0.75, -1.7) \\
\gamma_{Zb[i]} &= (2.3, 1.5, 0.15, 0.2, 0.9) \\
\gamma_{Zc[i]} &= (0, -1) \\
\gamma_{Za[i],Zc[i]} &= (0, -0.6, 0.5, 0.35, -0.4) \\
\gamma_{Zb[i],Zc[i]} &= (0, 1.7, 0.1, 2, -0.75)
\end{aligned}$$

Subgroup definition

There are 4 subgroups. The 1st subgroup includes units from 20 of the cells in the lower 40th percentile of inclusion probabilities, the second contains those from the 20th and 60th percentiles, the third from the 40th and 80th, and the fourth from the 60th and 100th. Cell subgroup membership was prespecified via random draw while maintaining a 1:3 (or 3:1) composition of cell membership in $X = 0$ to that in $X = 1$ to ensure that we are considering circumstances in which classic MRP fails to give adequate inference.

Result tables

```r
# rMSE
round(rmse,3)
```

```
##                 Overall group4 group3 group2 group.1
## WFPBB             0.010  0.007  0.011  0.017   0.053
## WFPBB-MRP         0.007  0.007  0.011  0.011   0.019
## Multinomial-MRP   0.007  0.007  0.012  0.011   0.018
## Two-Stage MRP     0.007  0.008  0.013  0.014   0.017
```

```r
# bias
round(bias,3)
```

```
##                 Overall group4 group3 group2 group.1
## WFPBB             0.000 -0.001  0.001 -0.001   0.004
## WFPBB-MRP         0.001  0.004 -0.008 -0.005   0.008
## Multinomial-MRP   0.000  0.004 -0.008 -0.006   0.006
## Two-Stage MRP    -0.002  0.006 -0.010 -0.011   0.006
```

```r
# standardized mean difference;
#  bias / pop subgroup std dev. of Y
# round(bias/subgrpsd,3)

# standard error of estimator
round(SE,3)
```

```
##                 Overall group4 group3 group2 group.1
## WFPBB             0.010  0.007  0.011  0.017   0.052
## WFPBB-MRP         0.007  0.005  0.008  0.010   0.017
## Multinomial-MRP   0.007  0.005  0.008  0.010   0.017
## Two-Stage MRP     0.007  0.005  0.008  0.008   0.016
```

```r
# CI length
round(CI_length,3)
```

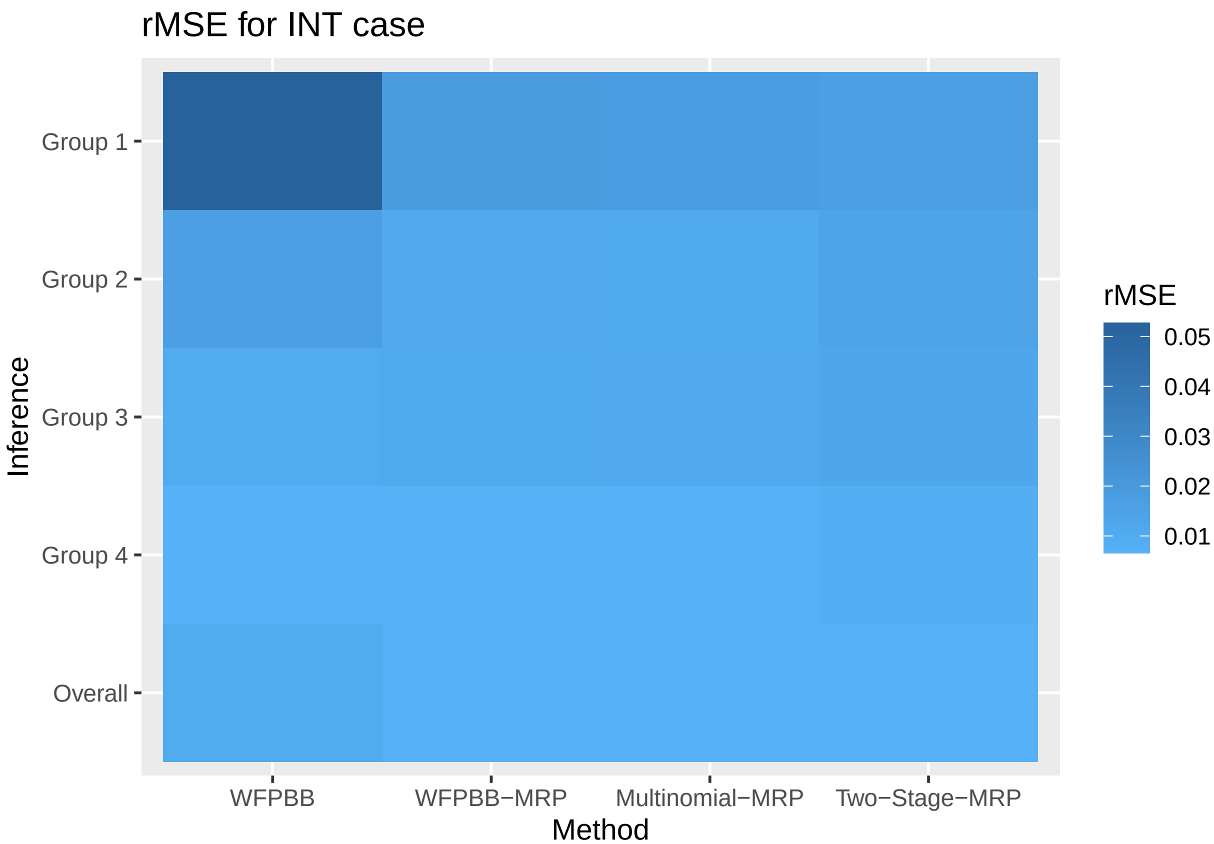
```
##                 Overall group4 group3 group2 group.1
## WFPBB             0.063  0.069  0.085  0.124   0.285
## WFPBB-MRP         0.042  0.047  0.056  0.075   0.129
## Multinomial-MRP   0.033  0.035  0.040  0.041   0.073
## Two-Stage MRP     0.033  0.035  0.041  0.041   0.073
```

```r
# coverage rate
round(coverage_rate,3)
```

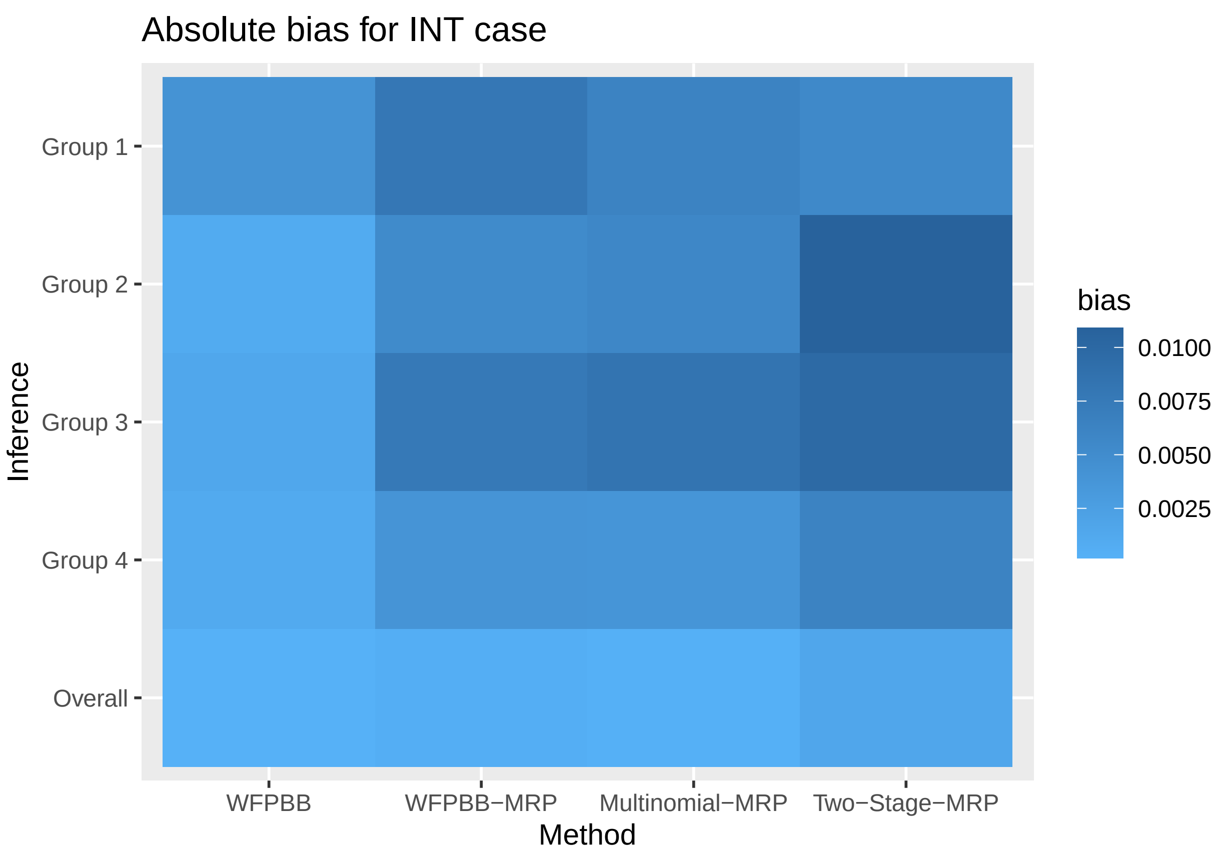
```
##                 Overall group4 group3 group2 group.1
## WFPBB             0.995   1.00  1.000   1.00   0.985
## WFPBB-MRP         1.000   1.00  0.995   1.00   1.000
## Multinomial-MRP   0.970   0.99  0.895   0.92   0.960
## Two-Stage MRP     0.975   0.97  0.895   0.88   0.970
```

Heatmap of overall summaries

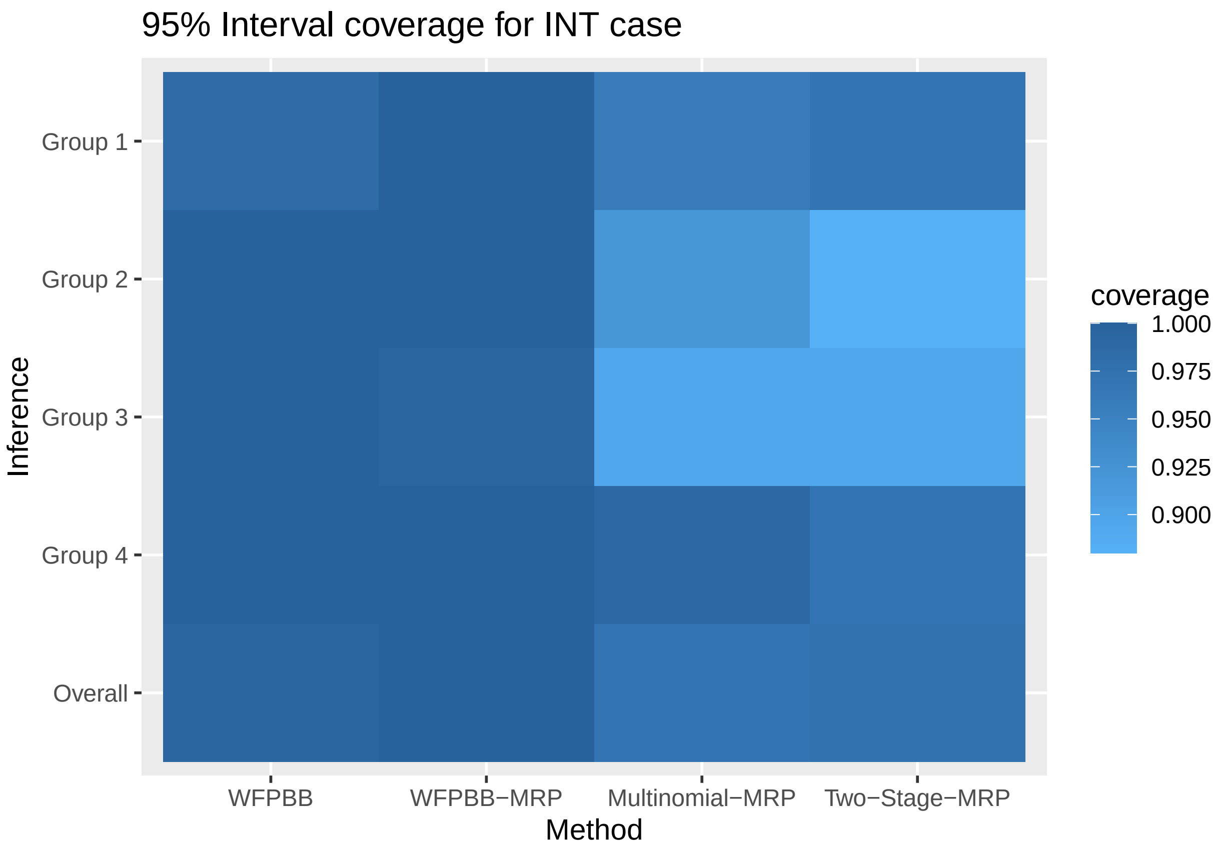
```r
#====== Overall summaries =======
## Use Bias, rMSE, coverage rate for loxz/hixy (most difference observed)
#- rMSE
  # start from overall WFPBB -> overall W-MRP -> overall M-MRP... -> group 5 WFPBB -> group 5 W-MRP...
  data$Z <- as.numeric(as.matrix(rmse))
  ggplot(data, aes(Method, Inference, fill= Z)) +
    geom_tile() + labs(title = "rMSE for INT case", fill = "rMSE")+scale_fill_continuous(high = "#28629
```



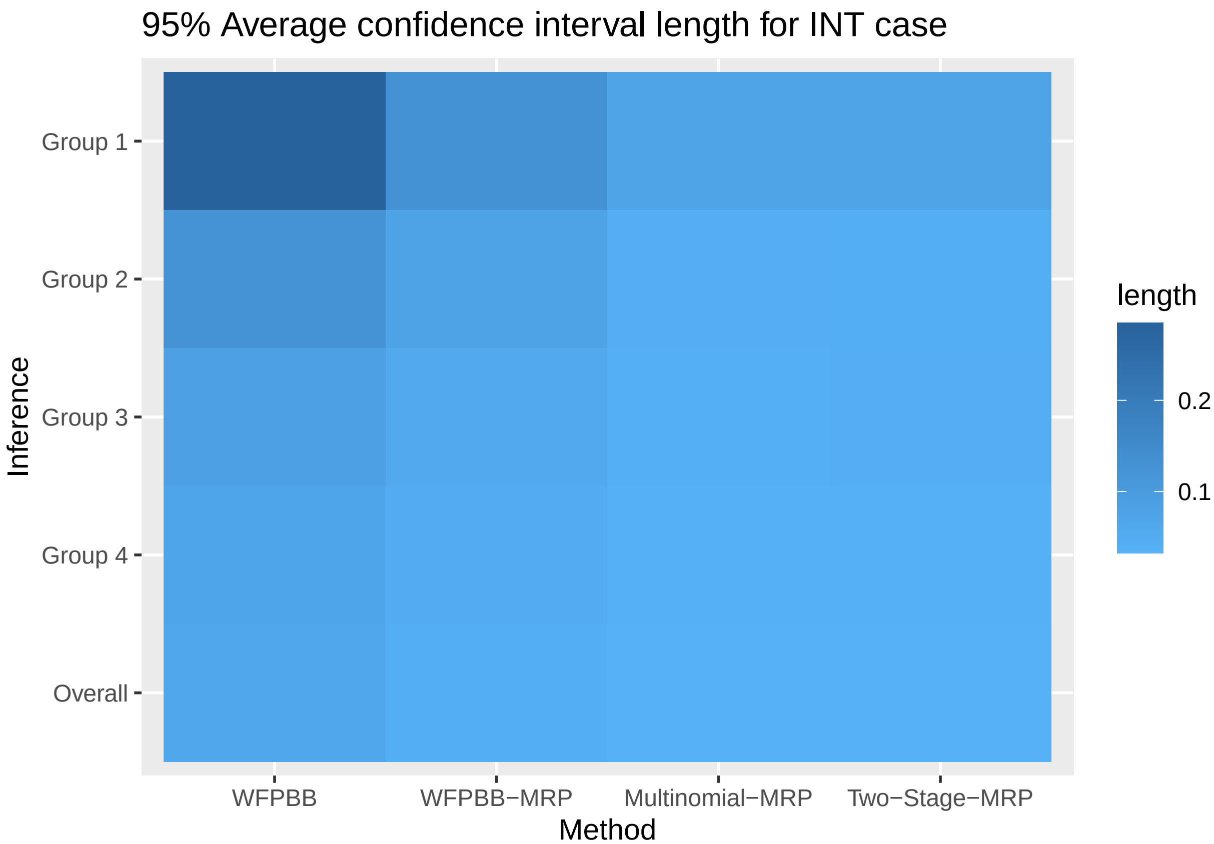

```
#- bias
  ## hi xz / hi xy
  data$Z <- as.numeric(as.matrix(bias))
  ggplot(data, aes(Method, Inference, fill= abs(Z))) +
    geom_tile() +
    labs(title = "Absolute bias for INT case", fill = "bias")+scale_fill_continuous(high = "#28629c", lo
```




```
#- coverage
  ## hi xz / hi xy
  data$Z <- as.numeric(as.matrix(coverage_rate))
  ggplot(data, aes(Method, Inference, fill= Z)) +
    geom_tile()+ labs(title = "95% Interval coverage for INT case", fill = "coverage")+scale_fill_conti
```




```
#- CI length
## hi xz / hi xy
data$Z <- as.numeric(as.matrix(CI_length))
ggplot(data, aes(Method, Inference, fill= Z)) +
  geom_tile()+ labs(title = "95% Average confidence interval length for INT case", fill = "length")+s
```

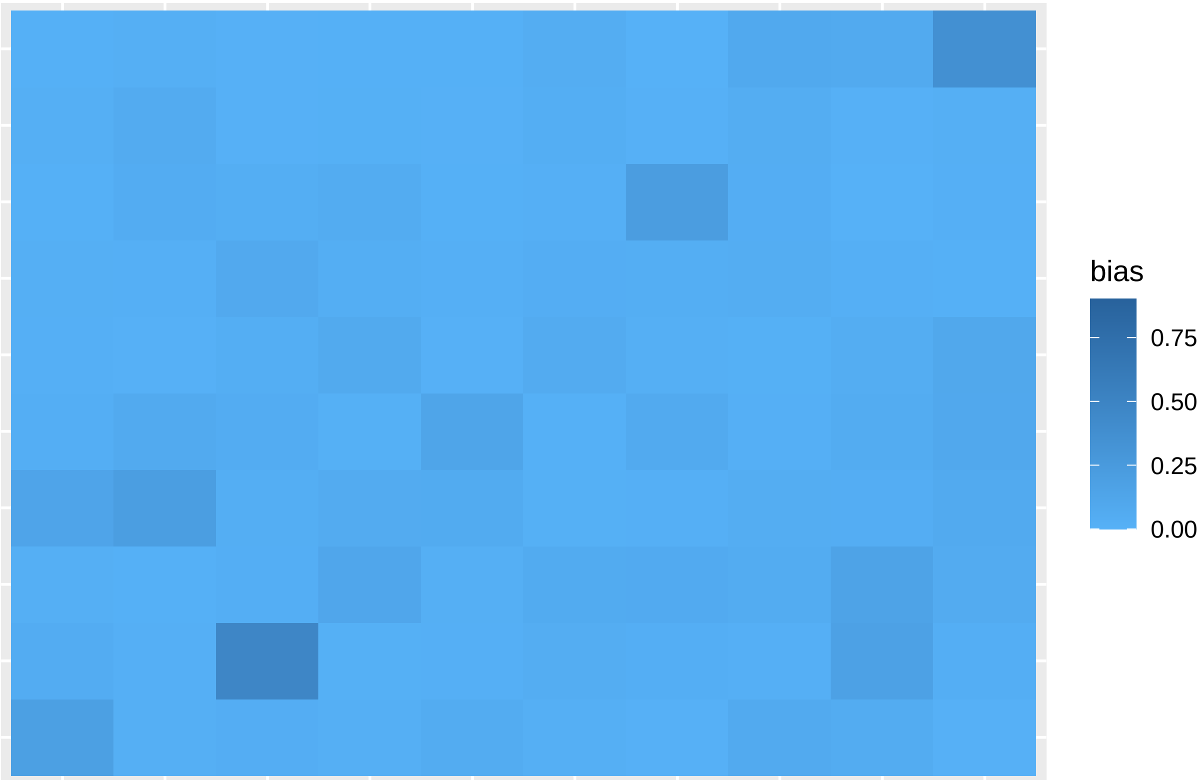


Nj summaries

Heatmaps are ordered by inclusion probability; lowest at the bottom left, highest at the top right.

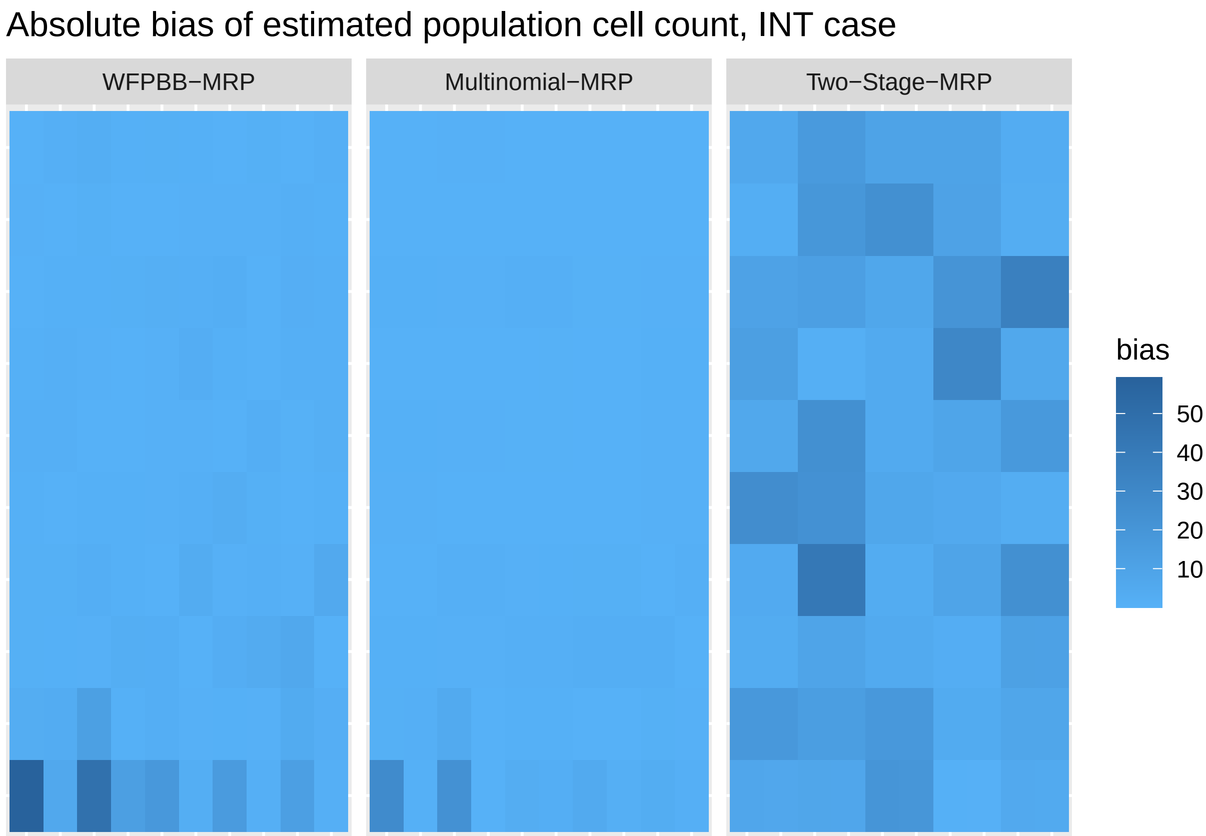
```
# ordered_cellcts_hihi <- 1:100
## == Cellmean bias ==
  x <- y <- factor(1:10)
  data <- expand.grid(X=x,Y=y)
  data$Z <- c(abs(cellmeanbias[ordered_cellcts_hihi]))
    ggplot(data, aes(x=X,y=Y, fill= Z)) +
    geom_tile()+
    theme(axis.ticks.y = element_blank()) +
    theme(axis.text.y = element_blank()) +
    theme(axis.ticks.x = element_blank()) +
    theme(axis.text.x = element_blank()) +
    theme(axis.title.x = element_blank()) +
    theme(axis.title.y = element_blank()) +
    labs(title = "Absolute bias of estimated population cell means", fill = "bias")+scale_fill_continuo
```

Absolute bias of estimated population cell means



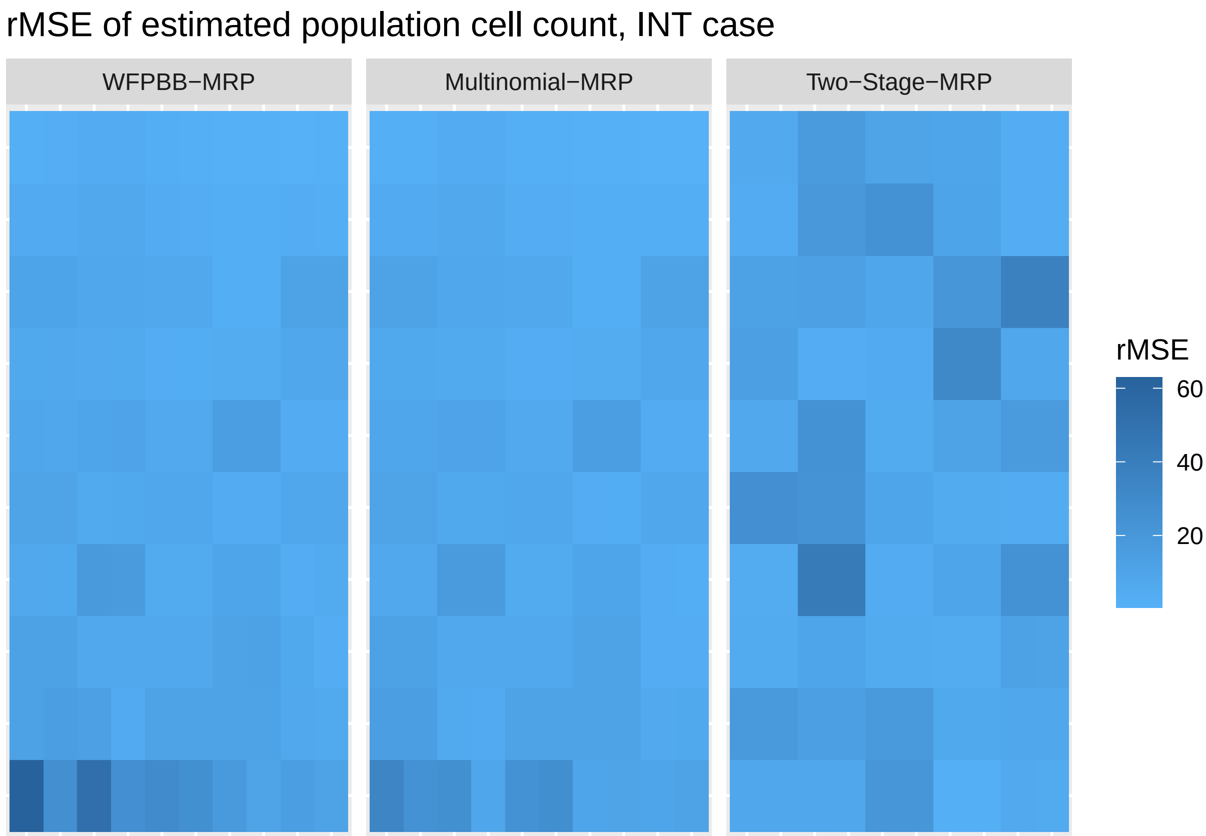

```r
## == Cellmean SE ==
  x <- y <- factor(1:10)
  data <- expand.grid(X=x,Y=y)
  data$Z <- c(abs(cellmeanse[ordered_cellcts_hihi]))
    ggplot(data, aes(x=X,y=Y, fill= Z)) +
    geom_tile()+
    theme(axis.ticks.y = element_blank()) +
    theme(axis.text.y = element_blank()) +
    theme(axis.ticks.x = element_blank()) +
    theme(axis.text.x = element_blank()) +
    theme(axis.title.x = element_blank()) +
    theme(axis.title.y = element_blank()) +
    labs(title = "SE of estimated population cell means", fill = "SE")+scale_fill_continuous(high = "#2
```


```
##======= Nj Bias =====
#--- faceted plots
x <- y <- factor(1:10)
method <- c("WFPBB-MRP", "Multinomial-MRP", "Two-Stage-MRP")
data <- expand.grid(X=x,Y=y, Method = method)
wmdat <- wmrp_njbias %>% abs()
m2dat <- mrp2_njbias%>% abs()
mmdat <- mmrp_njbias%>% abs()
wmdat <- c(wmdat[ordered_cellcts_hihi])
m2dat <- c(m2dat[ordered_cellcts_hihi])
mmdat <- c(mmdat[ordered_cellcts_hihi])
data$Z <- c(wmdat, mmdat, m2dat)
ggplot(data, aes(x=X,y=Y, fill= Z)) +
  geom_tile()+
  theme(axis.ticks.y = element_blank()) +
  theme(axis.text.y = element_blank()) +
  theme(axis.ticks.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  theme(axis.title.x = element_blank()) +
  theme(axis.title.y = element_blank()) +
  labs(title = "Absolute bias of estimated population cell count, INT case", fill = "bias")+scale_fil
```



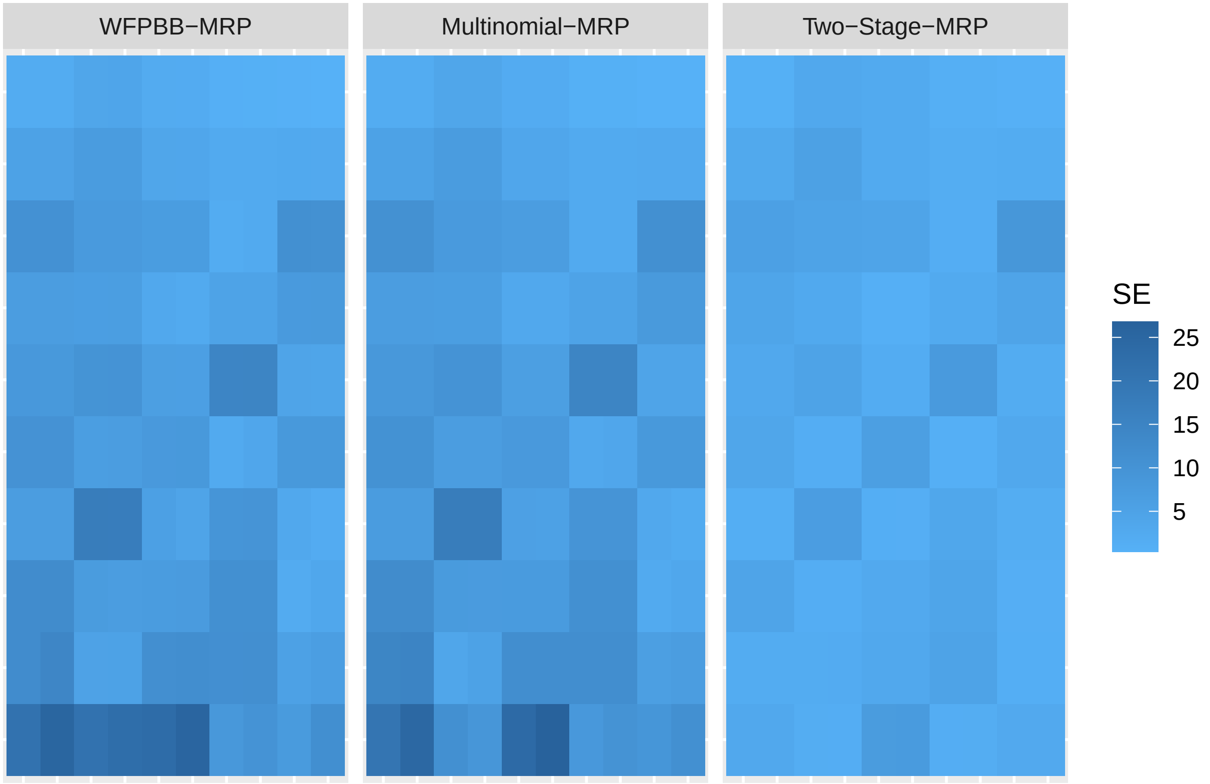

```r
##===== Nj rMSE ======

#-- faceted plots
x <- y <- factor(1:10)
method <- c("WFPBB-MRP", "Multinomial-MRP", "Two-Stage-MRP")
data <- expand.grid(X=x,Y=y, Method = method)
wmdat <- wmrp_njrmse
mmdat <- mmrp_njrmse
m2dat <- mrp2_njrmse
wmdat <- c(wmdat[ordered_cellcts_hihi])
m2dat <- c(m2dat[ordered_cellcts_hihi])
mmdat <- c(mmdat[ordered_cellcts_hihi])
data$Z <- c(wmdat, mmdat, m2dat)
ggplot(data, aes(x=X,y=Y, fill= Z)) +
  geom_tile()+
  theme(axis.ticks.y = element_blank()) +
  theme(axis.text.y = element_blank()) +
  theme(axis.ticks.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  theme(axis.title.x = element_blank()) +
  theme(axis.title.y = element_blank()) +
  labs(title = "rMSE of estimated population cell count, INT case", fill = "rMSE")+
  scale_fill_continuous(high = "#28629c", low = "#56B1F7")+
  facet_grid(. ~ Method)
```



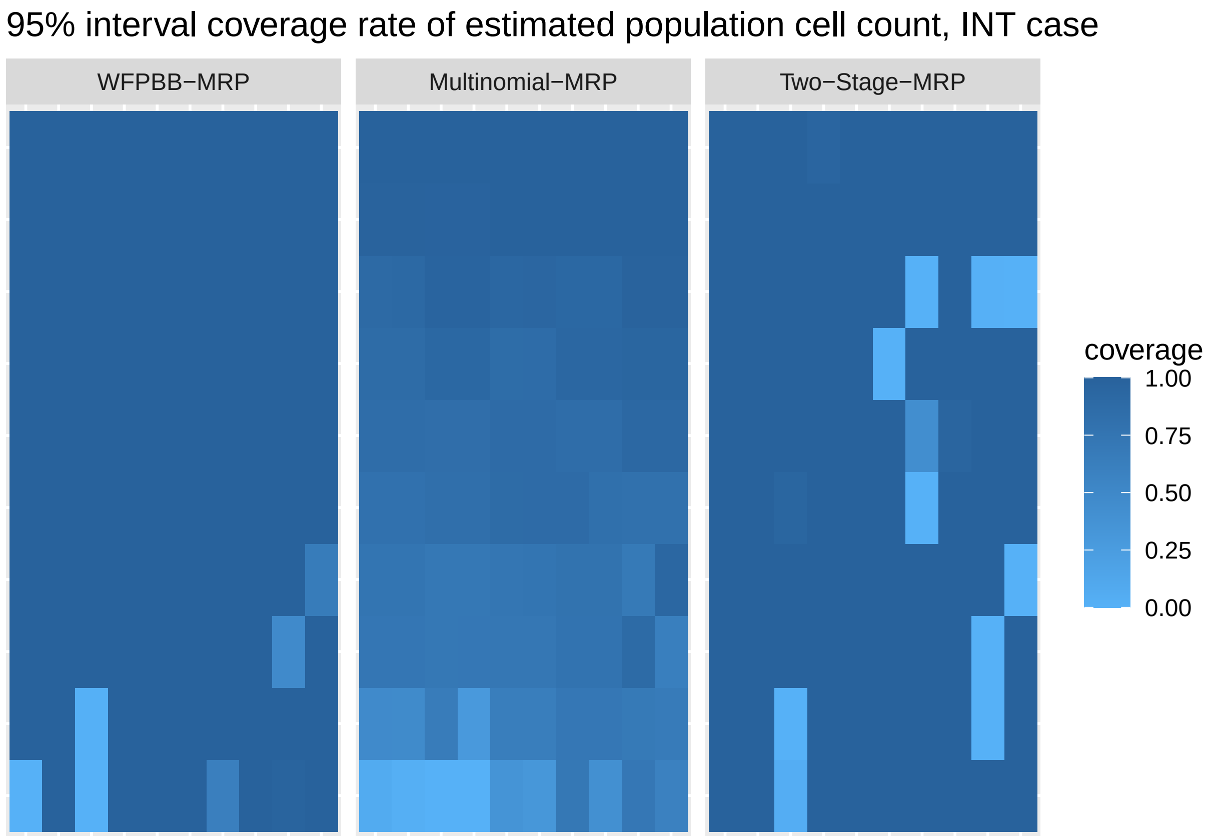

```r
##===== Nj SE ====
#-- faceted plots
x <- y <- factor(1:10)
method <- c("WFPBB-MRP", "Multinomial-MRP", "Two-Stage-MRP")
data <- expand.grid(X=x,Y=y, Method = method)
wmdat <- wmrp_njse
mmdat <- mmrp_njse
m2dat <- mrp2_njse
wmdat <- c(wmdat[ordered_cellcts_hihi])
m2dat <- c(m2dat[ordered_cellcts_hihi])
mmdat <- c(mmdat[ordered_cellcts_hihi])
data$Z <- c(wmdat, mmdat, m2dat)
ggplot(data, aes(x=X,y=Y, fill= Z)) +
  geom_tile()+
  theme(axis.ticks.y = element_blank()) +
  theme(axis.text.y = element_blank()) +
  theme(axis.ticks.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  theme(axis.title.x = element_blank()) +
  theme(axis.title.y = element_blank()) +
  labs(title = "Standard error of estimated population cell count, INT case", fill = "SE")+
  scale_fill_continuous(high = "#28629c", low = "#56B1F7")+
  facet_grid(. ~ Method)
```

Standard error of estimated population cell count, INT case




```
#==== Nj coverage rates======

  #-- faceted plots
x <- y <- factor(1:10)
method <- c("WFPBB-MRP", "Multinomial-MRP", "Two-Stage-MRP")
data <- expand.grid(X=x,Y=y, Method = method)
wmdat <- wmrp_njcov
mmdat <- mmrp_njcov
m2dat <- mrp2_njcov
wmdat <- c(wmdat[ordered_cellcts_hihi])
m2dat <- c(m2dat[ordered_cellcts_hihi])
mmdat <- c(mmdat[ordered_cellcts_hihi])
data$Z <- c(wmdat, mmdat, m2dat)
ggplot(data, aes(x=X,y=Y, fill= Z)) +
  geom_tile()+
  theme(axis.ticks.y = element_blank()) +
  theme(axis.text.y = element_blank()) +
  theme(axis.ticks.x = element_blank()) +
  theme(axis.text.x = element_blank()) +
  theme(axis.title.x = element_blank()) +
  theme(axis.title.y = element_blank()) +
  labs(title = "95% interval coverage rate of estimated population cell count, INT case", fill = "cove
  scale_fill_continuous(high = "#28629c", low = "#56B1F7")+
  facet_grid(. ~ Method)
```


95% interval coverage rate of estimated population cell count, INT case
WFPBB−MRP
Multinomial−MRP
Two−Stage−MRP
coverage
1.00
0.75
0.50
0.25
0.00