r2mlm: R-squared for Multilevel Models

Mairead Shaw

Load Data and Dependencies

```
library(r2mlm) # this also loads lme4
library(lmerTest) # significance for coefficients
library(performance)
library(dplyr)
```

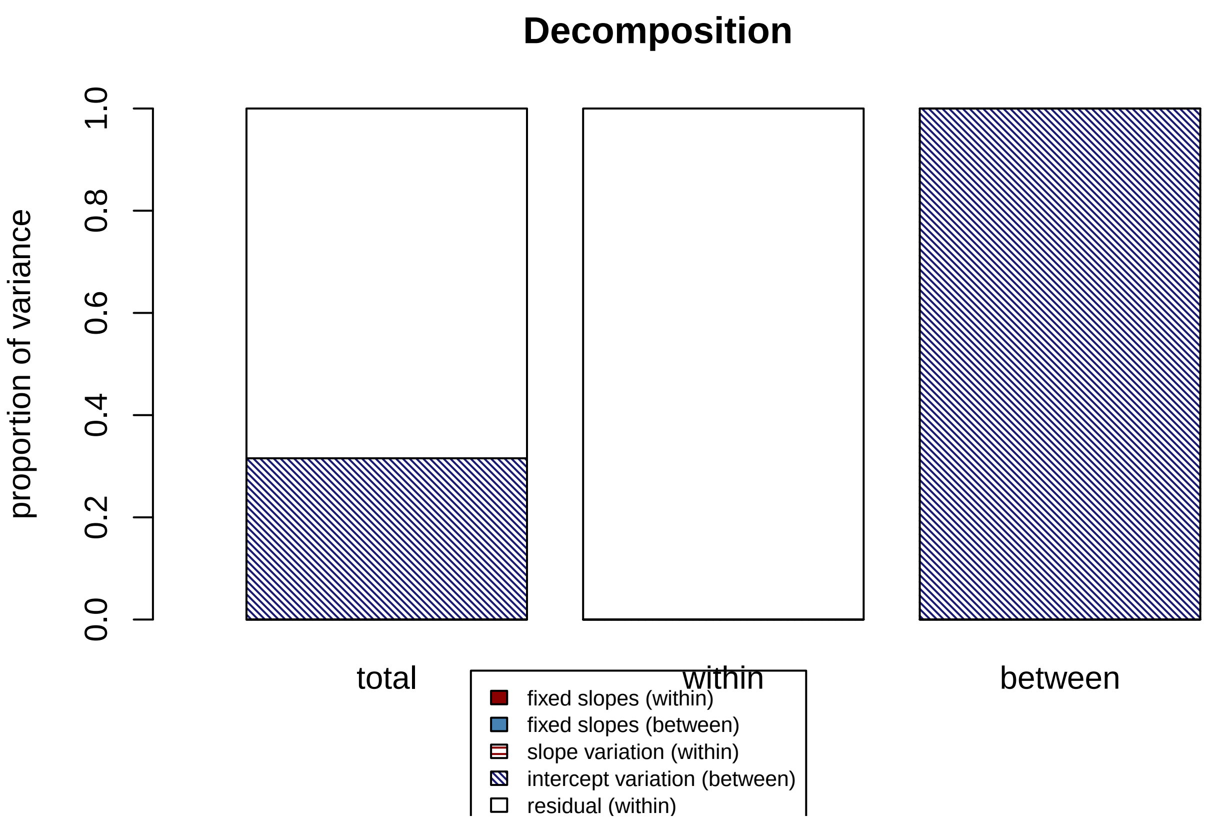
Null Model

Teachers clustered within classes:

```
null_model <- lmer(satisfaction ~ 1 + (1|schoolID), data = teachsat, REML = TRUE)
summary(null_model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: satisfaction ~ 1 + (1 | schoolID)
##    Data: teachsat
##
## REML criterion at convergence: 30098.4
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.8269 -0.6385  0.0012  0.6435  3.2874
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  schoolID (Intercept) 0.699    0.836
##  Residual             1.516    1.231
## Number of obs: 9000, groups:  schoolID, 300
##
## Fixed effects:
##              Estimate Std. Error        df t value            Pr(>|t|)
## (Intercept)   5.99677    0.04998 299.00000     120 <0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
r2mlm(null_model)
```




```
## $Decompositions
##                 total             within between
## fixed, within   0                 0      NA
## fixed, between  0                 NA     0
## slope variation 0                 0      NA
## mean variation  0.315546785367943 NA     1
## sigma2          0.684453214632058 1      NA
##
## $R2s
##     total             within between
## f1  0                 0      NA
## f2  0                 NA     0
## v   0                 0      NA
## m   0.315546785367943 NA     1
## f   0                 NA     NA
## fv  0                 0      NA
## fvm 0.315546785367943 NA     NA
```

Three sets of output: Decompositions, R2s, and Graph

- Decompositions give you the unique R-squareds
- R2s gives you the unique R-squareds and the combinations
- Graph visualizes Decompositions

My order of reading is usually: R2s, then look at the graph.

The only variance explained in job satisfaction is explained by group membership (random intercept). That's the intraclass correlation! So 31.6% of variance in job satisfaction is attributed to cluster. For comparison, we can use the performance package to calculate the ICC:

```
performance::icc(null_model)
```

```
## # Intraclass Correlation Coefficient
##
##      Adjusted ICC: 0.316
##   Conditional ICC: 0.316
```

They match, as expected!

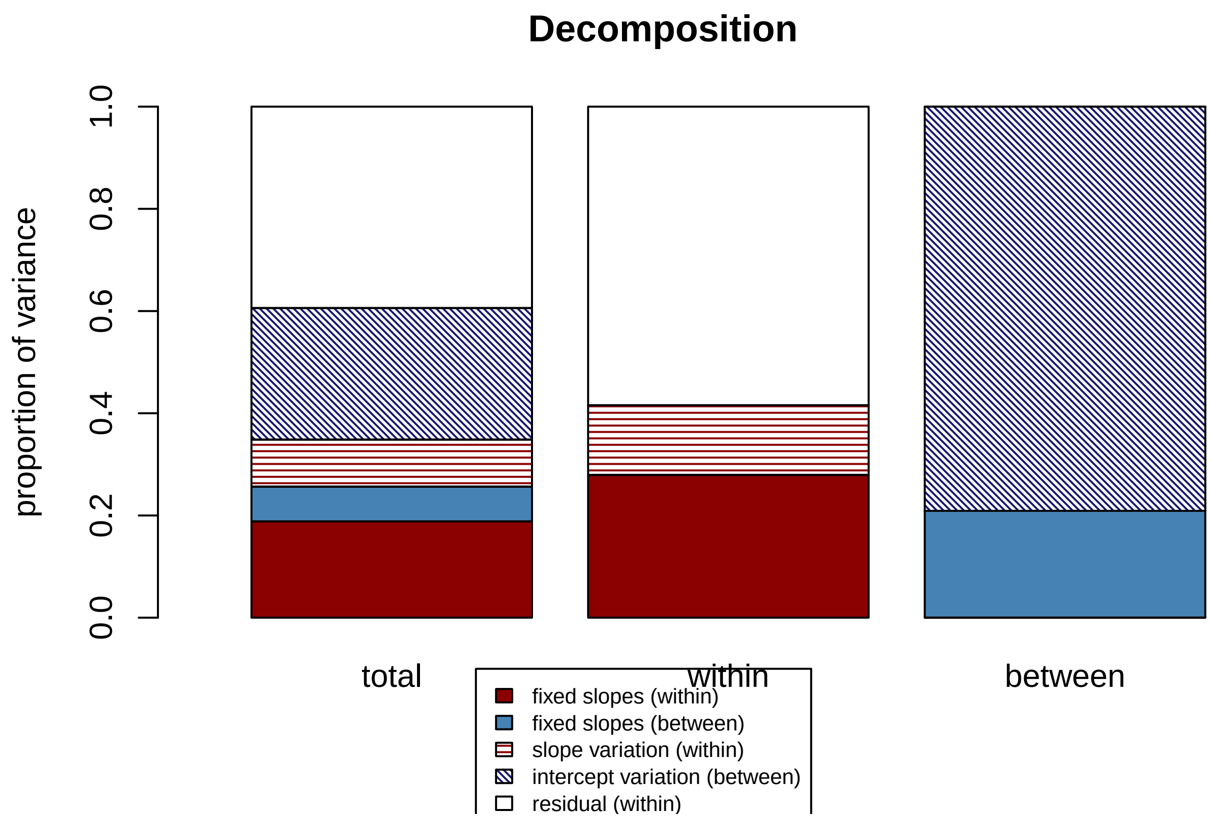
Full Model

1. Level-1: salary (centered within school)
2. Level-2: student-teacher ratio (same for all teachers within a school, differs across schools)

```
model <- lmer(satisfaction ~ 1 + salary_c + s_t_ratio + (salary_c|schoolID),
              data = teachsat,
              REML = TRUE,
              control = lmerControl(optimizer = "bobyqa")) # optimizer change to help convergence (to m
summary(model) # Instructive to look at our unstandardized results, too: coefficient values
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: satisfaction ~ 1 + salary_c + s_t_ratio + (salary_c | schoolID)
##    Data: teachsat
## Control: lmerControl(optimizer = "bobyqa")
##
## REML criterion at convergence: 25878.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5436 -0.6507 -0.0042  0.6536  3.7282
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  schoolID (Intercept) 0.571138 0.75574
##           salary_c    0.002724 0.05219  -0.01
##  Residual             0.874365 0.93508
## Number of obs: 9000, groups:  schoolID, 300
##
## Fixed effects:
##              Estimate Std. Error         df t value             Pr(>|t|)
## (Intercept)   7.189783   0.144438 298.003026  49.778 < 0.0000000000000002 ***
## salary_c      0.074648   0.003231 300.385167  23.106 < 0.0000000000000002 ***
## s_t_ratio    -0.037282   0.004292 298.000395  -8.687 0.000000000000000251 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) slry_c
## salary_c  -0.002
## s_t_ratio -0.951  0.000
```

```
r2mlm(model)
```

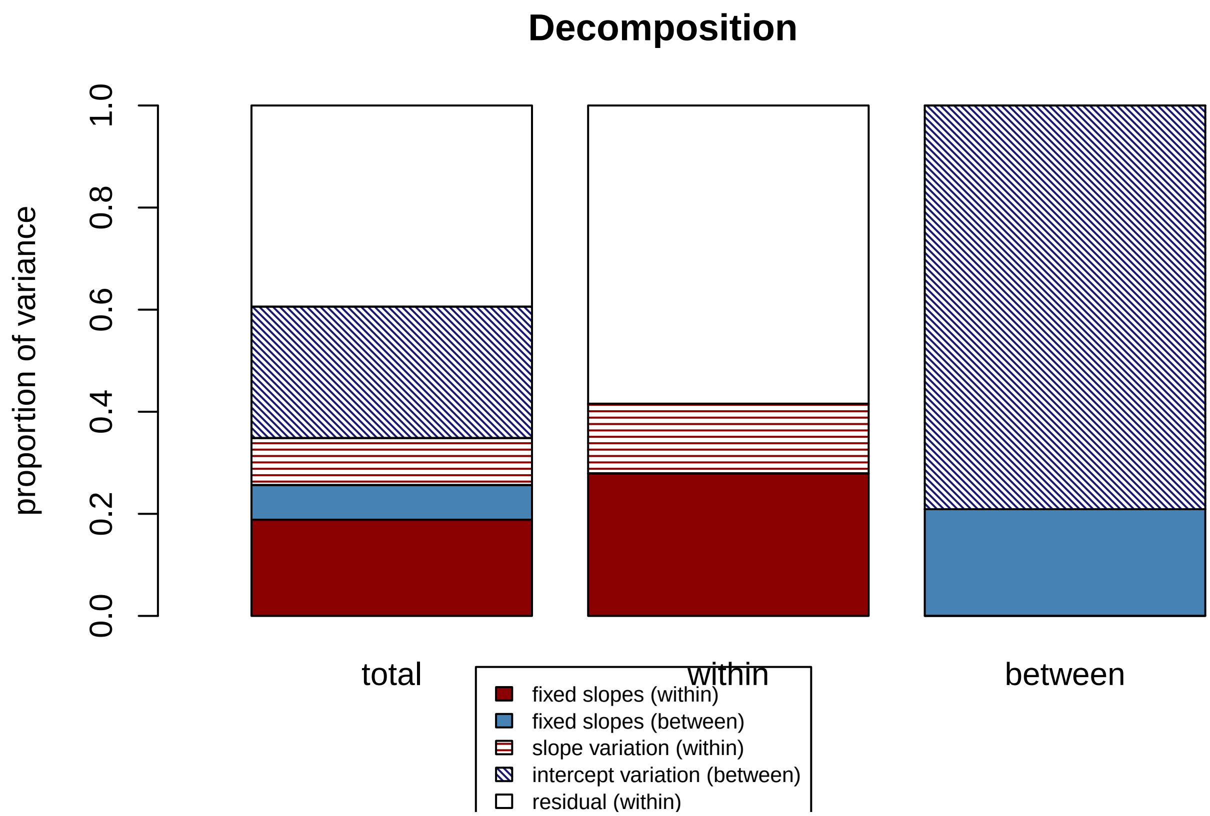



```
## $Decompositions
##                 total              within            between
## fixed, within   0.188362688254615  0.279254409150734 NA
## fixed, between  0.0680577989179737 NA                0.209099797402447
## slope variation 0.0920647772854669 0.136489318679158 NA
## mean variation  0.257422186062525  NA                0.790900202597553
## sigma2          0.394092549479419  0.584256272170108 NA
##
## $R2s
##     total              within            between
## f1  0.188362688254615  0.279254409150734 NA
## f2  0.0680577989179737 NA                0.209099797402447
## v   0.0920647772854669 0.136489318679158 NA
## m   0.257422186062525  NA                0.790900202597553
## f   0.256420487172589  NA                NA
## fv  0.348485264458056  0.415743727829892 NA
## fvm 0.605907450520581  NA                NA
```

Manual Entry

```
r2mlm_manual(
  data = teachsat,
  within_covs = c("salary_c"),
```

```
  between_covs = c("s_t_ratio"),
  random_covs = c("salary_c"),
  gamma_w = c(0.074648),
  gamma_b = c(7.189783, -0.037282),
  Tau = as.matrix(bdiag(VarCorr(model))),
  sigma2 = getME(model, "sigma")^2,
  has_intercept = TRUE,
  clustermeancentered = TRUE
)
```

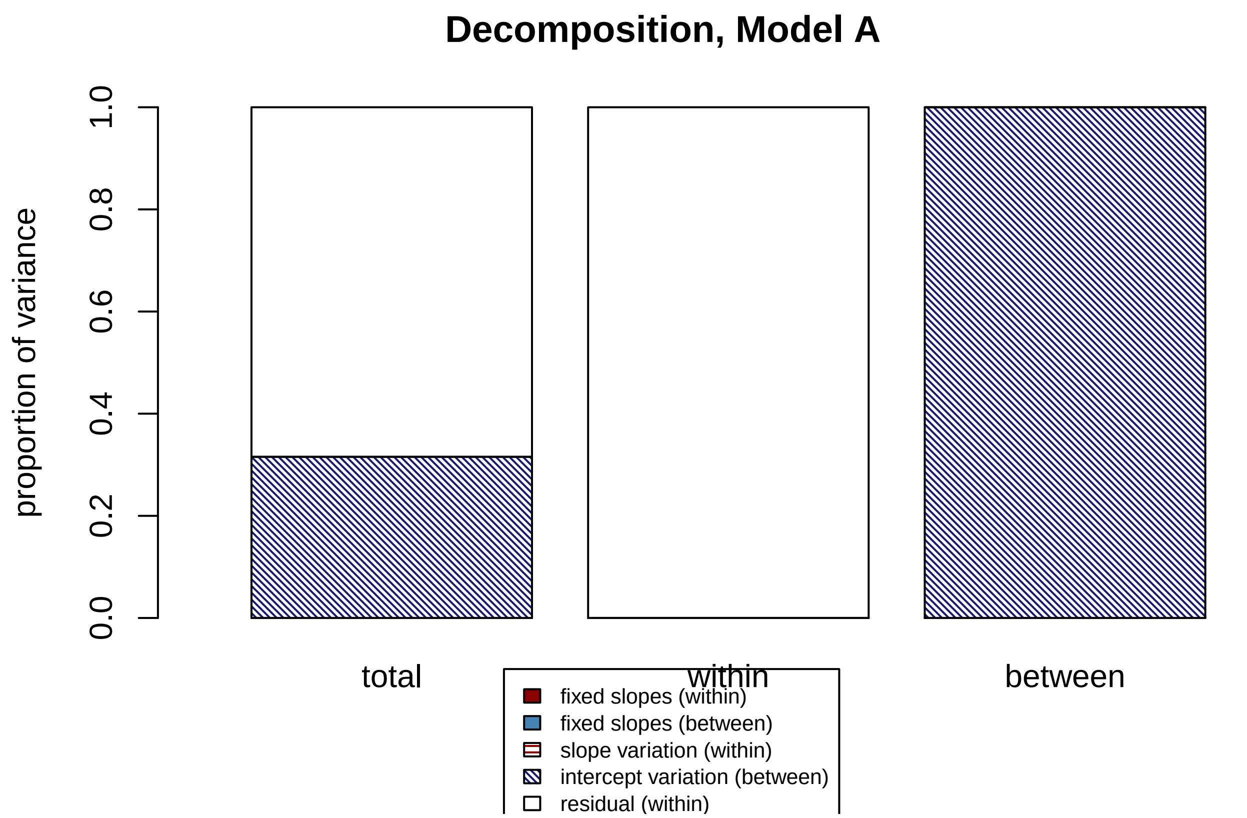


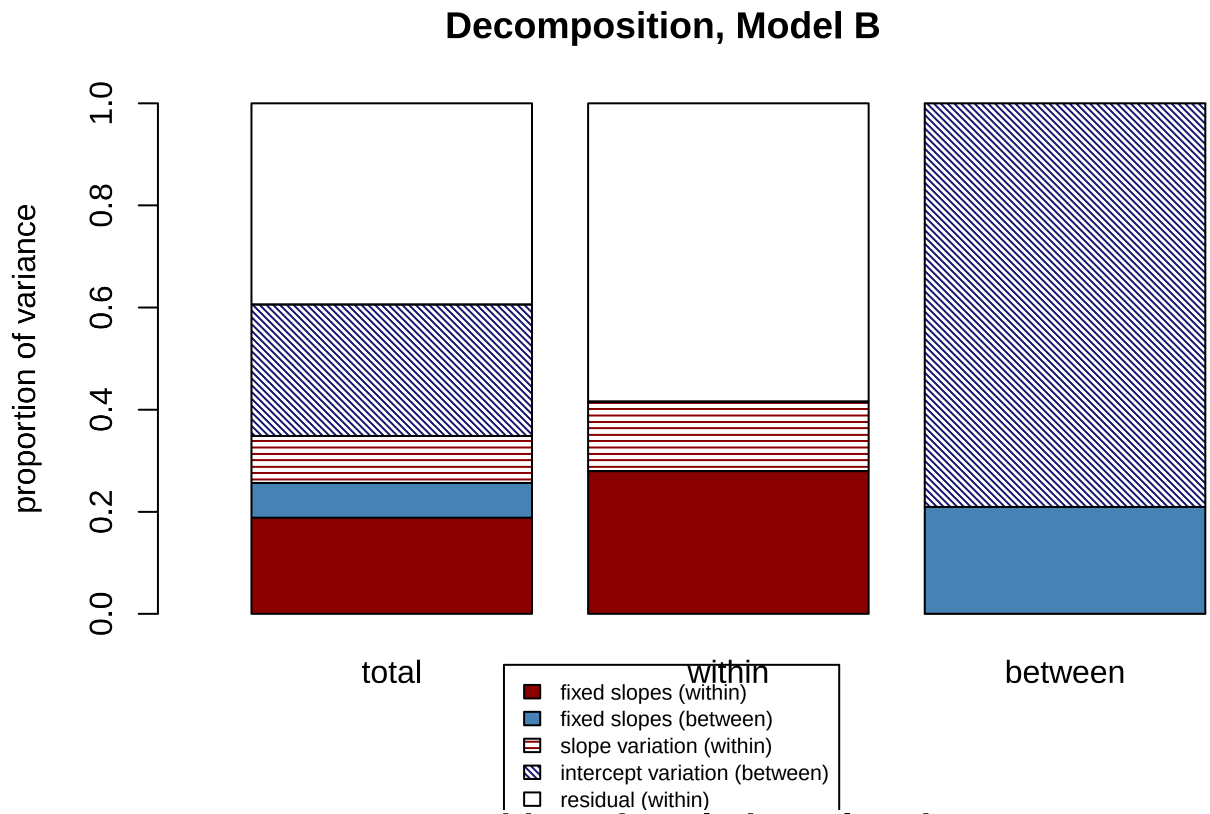

```
## $Decompositions
##                  total              within             between
## fixed, within    0.188363307468025  0.279255631555421  NA
## fixed, between   0.0680592513633171 NA                 0.209103787514305
## slope variation  0.0920645207872735 0.136489087189454  NA
## mean variation   0.25742146886831   NA                 0.790896212485695
## sigma2           0.394091451513075  0.584255281255124  NA
##
## $R2s
##     total              within             between
## f1  0.188363307468025  0.279255631555421  NA
## f2  0.0680592513633171 NA                 0.209103787514305
## v   0.0920645207872735 0.136489087189454  NA
## m   0.25742146886831   NA                 0.790896212485695
## f   0.256422558831342  NA                 NA
## fv  0.348487079618615  0.415744718744876  NA
## fvm 0.605908548486925  NA                 NA
```

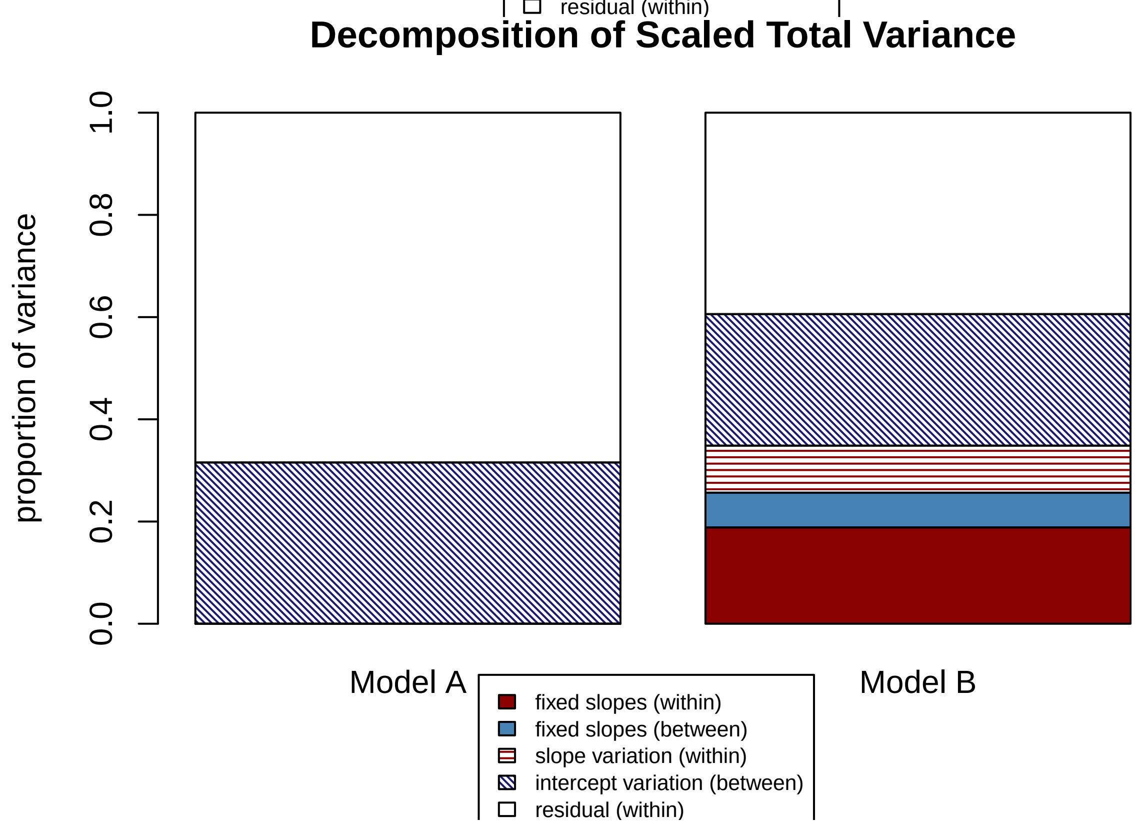
Compare to the automatic entry: they're the same!

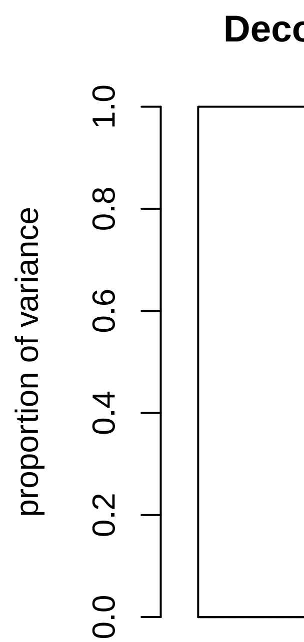
Model Comparison

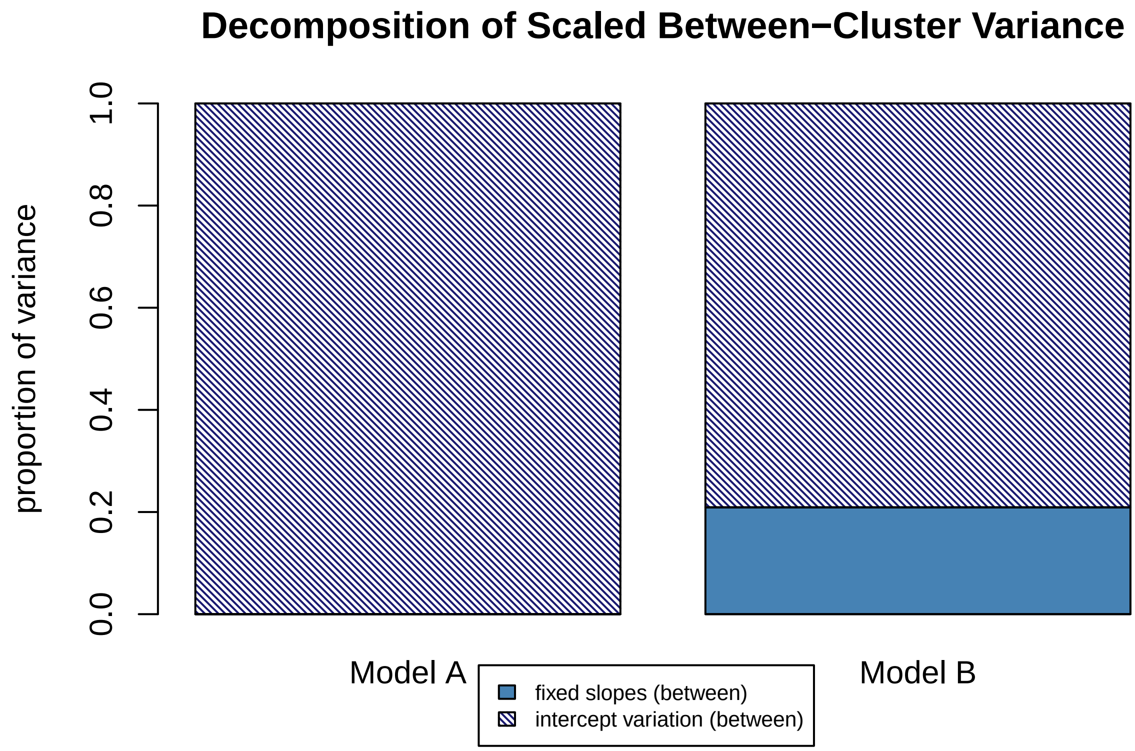
```
r2mlm_comp(null_model, model)
```




Decomposition, Model B
proportion of variance
0.0
0.2
0.4
0.6
0.8
1.0
total
within
between
fixed slopes (within)
fixed slopes (between)
slope variation (within)
intercept variation (between)
residual (within)


Decomposition of Scaled Total Variance
proportion of variance
0.0
0.2
0.4
0.6
0.8
1.0
Model A
Model B
fixed slopes (within)
fixed slopes (between)
slope variation (within)
intercept variation (between)
residual (within)


proportion of variance
0.0
0.2
0.4
0.6
0.8
1.0




```
## $`Model A R2s`
##     total             within between
## f1  0                 0      NA
## f2  0                 NA     0
## v   0                 0      NA
## m   0.315546785367943 NA     1
## f   0                 NA     NA
## fv  0                 0      NA
## fvm 0.315546785367943 NA     NA
##
## $`Model B R2s`
##     total              within            between
## f1  0.188362688254615  0.279254409150734 NA
## f2  0.0680577989179737 NA                0.209099797402447
## v   0.0920647772854669 0.136489318679158 NA
## m   0.257422186062525  NA                0.790900202597553
## f   0.256420487172589  NA                NA
## fv  0.348485264458056  0.415743727829892 NA
## fvm 0.605907450520581  NA                NA
##
## $`R2 differences, Model B - Model A`
##           total    within    between
## f1   0.18836269 0.2792544         NA
## f2   0.06805780        NA  0.2090998
## v    0.09206478 0.1364893         NA
## m   -0.05812460        NA -0.2090998
## f    0.25642049        NA         NA
## fv   0.34848526 0.4157437         NA
## fvm  0.29036067        NA         NA
```

5 graph outputs: model A and B each overall, then comparisons of within, between, and total.

There is also manual entry for model comparison, but it's verbose so we're not going to look at it here.

Non-CWC Model Options

```
# You can see the centering, each school has a mean of zero
teachsat %>%
  group_by(schoolID) %>%
  summarize(
    mean(salary_c)
  )
```

```
## # A tibble: 300 x 2
##    schoolID `mean(salary_c)`
##       <int>            <dbl>
##  1        1         1.82e-16
##  2        2        -2.66e-16
##  3        3         2.66e-16
##  4        4        -7.46e-18
##  5        5        -4.62e-17
##  6        6         1.26e-16
##  7        7         2.00e-16
##  8        8        -5.94e-17
##  9        9         8.52e-17
## 10       10         7.40e-17
## # ... with 290 more rows
```

```
# Remove centering on salary by adding a constant to each value
teachsat <- teachsat %>%
  mutate(salary = salary_c + 1)

# You can see mean is no longer zero
teachsat %>%
  group_by(schoolID) %>%
  summarize(
    mean(salary)
  )
```

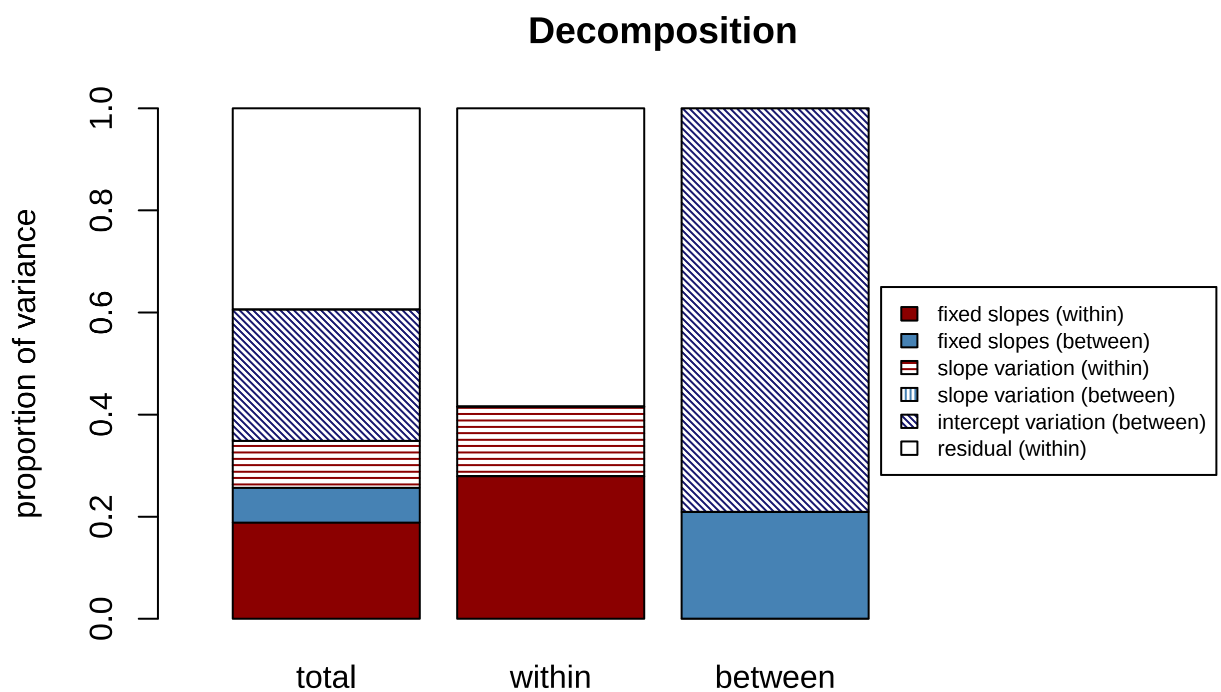
```
## # A tibble: 300 x 2
##    schoolID `mean(salary)`
##       <int>          <dbl>
##  1        1              1
##  2        2              1
##  3        3              1
##  4        4              1
##  5        5              1
##  6        6              1
##  7        7              1
##  8        8              1
##  9        9              1
## 10       10              1
## # ... with 290 more rows
```

```r
# Model with this new salary
model_uncwc <- lmer(satisfaction ~ 1 + salary + s_t_ratio + (salary|schoolID),
              data = teachsat,
              REML = TRUE,
              control = lmerControl(optimizer = "bobyqa"))
summary(model_uncwc)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: satisfaction ~ 1 + salary + s_t_ratio + (salary | schoolID)
##    Data: teachsat
## Control: lmerControl(optimizer = "bobyqa")
##
## REML criterion at convergence: 25878.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5436 -0.6507 -0.0042  0.6536  3.7282
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  schoolID (Intercept) 0.574481 0.75795
##           salary      0.002724 0.05219  -0.08
##  Residual             0.874365 0.93508
## Number of obs: 9000, groups:  schoolID, 300
##
## Fixed effects:
##              Estimate Std. Error         df t value             Pr(>|t|)
## (Intercept)   7.115136   0.144482 298.325584  49.246 < 0.0000000000000002 ***
## salary        0.074648   0.003231 300.384844  23.106 < 0.0000000000000002 ***
## s_t_ratio    -0.037282   0.004292 298.000229  -8.687 0.000000000000000251 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) salary
## salary    -0.025
## s_t_ratio -0.951  0.000
```

If you use the regular `r2mlm` function, you get just an overall breakdown. You can use `r2mlm_long` instead to get a full breakdown, though this option is only available as a manual function right now.

```r
r2mlm_long_manual(
  data = teachsat,
  covs = c("salary", "s_t_ratio"),
  random_covs = c("salary"),
  clusterID = "schoolID",
  gammas = c(0.074648, -0.037282),
  Tau = as.matrix(Matrix::bdiag(VarCorr(model_uncwc))),
  sigma = getME(model_uncwc, "sigma")^2,
  bargraph = TRUE
)
```




```
## $Decompositions
##                                total
## fixed slopes (within)          0.188363278106427
## fixed slopes (between)         0.0680592407544123
## slope variation (within)       0.0920645799175673
## slope variation (between)      0.00000000000000000000000000000000354600596230147
## intercept variation (between) 0.257421525598485
## residual (within)              0.394091375623109
##                                within
## fixed slopes (within)          0.279255607120278
## fixed slopes (between)         NA
## slope variation (within)       0.136489184184975
## slope variation (between)      NA
## intercept variation (between) NA
## residual (within)              0.584255208694747
##                                between
## fixed slopes (within)          NA
## fixed slopes (between)         0.209103725289316
## slope variation (within)       NA
## slope variation (between)      0.0000000000000000000000000000000010894671295128 9
## intercept variation (between) 0.790896274710684
## residual (within)              NA
##
## $R2s
##     total                                               within
## f1  0.188363278106427                                   0.279255607120278
## f2  0.0680592407544123                                  NA
## v1  0.0920645799175673                                  0.136489184184975
## v2  0.00000000000000000000000000000000354600596230147 NA
## m   0.257421525598485                                   NA
## f   0.256422518860839                                   NA
## fv  0.348487098778406                                   0.415744791305253
```

```
## fvm 0.605908624376891                                   NA
##     between
## f1  NA
## f2  0.209103725289316
## v1  NA
## v2  0.00000000000000000000000000000000108946712951289
## m   0.790896274710684
## f   NA
## fv  0.209103725289316
## fvm NA
```

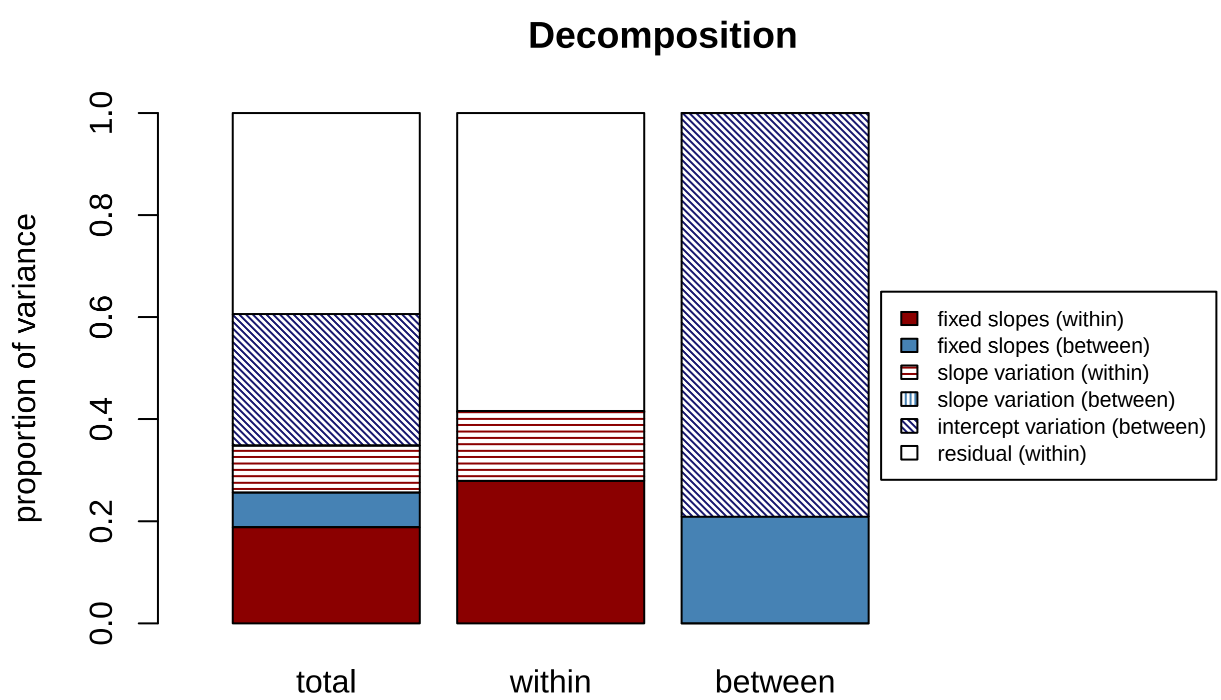
Three outputs: Decompositions, R2s, Graph. This is similar to r2mlm(), but now there are two v terms. Recall that v is variance attributed to random slopes: v1 is slope variation within a cluster; v2 is slope variation between clusters. Why two v's here? When a level-1 variable is centered within cluster, every cluster has a mean of 0 and there is no variance between clusters. All the variance is within clusters. When a level-1 variable is NOT centered within cluster, it varies both within and between clusters.

What does the `r2mlm_long` output look like with our centered model?

```
summary(model)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: satisfaction ~ 1 + salary_c + s_t_ratio + (salary_c | schoolID)
##    Data: teachsat
## Control: lmerControl(optimizer = "bobyqa")
##
## REML criterion at convergence: 25878.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5436 -0.6507 -0.0042  0.6536  3.7282
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  schoolID (Intercept) 0.571138 0.75574
##           salary_c    0.002724 0.05219  -0.01
##  Residual             0.874365 0.93508
## Number of obs: 9000, groups:  schoolID, 300
##
## Fixed effects:
##               Estimate Std. Error         df t value             Pr(>|t|)
## (Intercept)   7.189783   0.144438 298.003026  49.778 < 0.0000000000000002 ***
## salary_c      0.074648   0.003231 300.385167  23.106 < 0.0000000000000002 ***
## s_t_ratio    -0.037282   0.004292 298.000395  -8.687 0.000000000000000251 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) slry_c
## salary_c  -0.002
## s_t_ratio -0.951  0.000
```

```r
r2mlm_long_manual(
  data = teachsat,
  covs = c("salary_c", "s_t_ratio"),
  random_covs = c("salary_c"),
  clusterID = "schoolID",
  gammas = c(0.074648, -0.037282),
  Tau = as.matrix(Matrix::bdiag(VarCorr(model))),
  sigma = getME(model, "sigma")^2,
  bargraph = TRUE
)
```




```
## $Decompositions
##                                total
## fixed slopes (within)          0.188363307468025
## fixed slopes (between)         0.0680592513633171
## slope variation (within)       0.0920645207872735
## slope variation (between)      0.0000000000000000000000000000000307553254228165
## intercept variation (between) 0.25742146886831
## residual (within)              0.394091451513075
##                                within
## fixed slopes (within)          0.279255631555421
## fixed slopes (between)         NA
## slope variation (within)       0.136489087189454
## slope variation (between)      NA
## intercept variation (between) NA
## residual (within)              0.584255281255124
##                                between
## fixed slopes (within)          NA
## fixed slopes (between)         0.209103787514305
## slope variation (within)       NA
## slope variation (between)      0.0000000000000000000000000000000944920037074072
```

```
## intercept variation (between) 0.790896212485695
## residual (within)             NA
##
## $R2s
##     total                                            within
## f1  0.188363307468025                                0.279255631555421
## f2  0.0680592513633171                               NA
## v1  0.0920645207872735                               0.136489087189454
## v2  0.00000000000000000000000000000000307553254228165 NA
## m   0.25742146886831                                 NA
## f   0.256422558831342                                NA
## fv  0.348487079618615                                0.415744718744876
## fvm 0.605908548486925                                NA
##     between
## f1  NA
## f2  0.209103787514305
## v1  NA
## v2  0.00000000000000000000000000000000944920037074072
## m   0.790896212485695
## f   NA
## fv  0.209103787514305
## fvm NA
```

The same, as expected! The change: v2 is essentially zero, because there is no variance in salary_c across groups.