MSKIDS: Exploratory Analysis and ComBat Harmonization with Healthy Controls

Virgilio Gonzenbach @ PennSIVE

13 November, 2020, 16:54

Batch Variables

```
##
## SIEMENSPRISMAFIT   SIEMENSTIMTRIO     SIEMENSVERIO
##               58            1192               36
```

```
##
##  CHP  HSC  PNC
##   36   65 1185
```

Scanners by sites

```
##
##                     CHP  HSC  PNC
##   SIEMENSPRISMAFIT    0   58    0
##   SIEMENSTIMTRIO      0    7 1185
##   SIEMENSVERIO       36    0    0
```

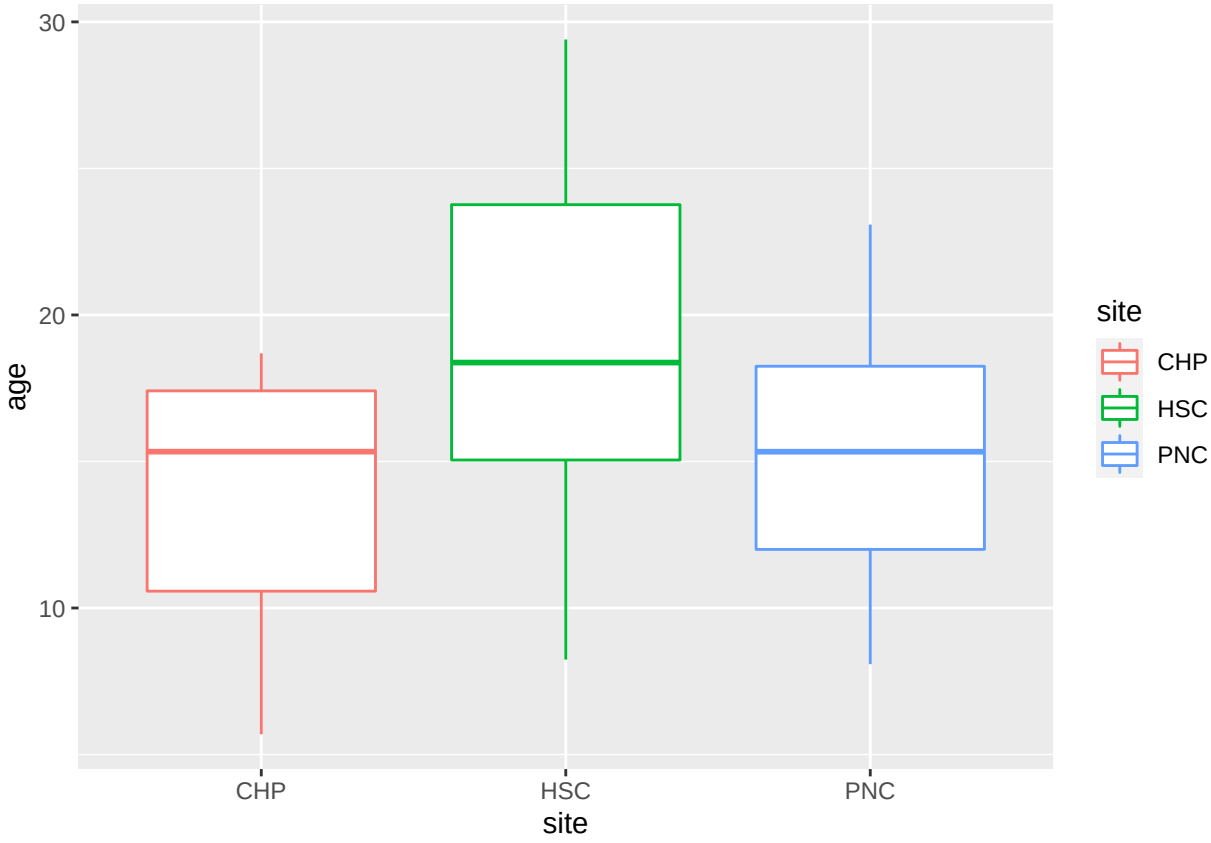
Excluding HSC-SIEMENSTIMTRIO (n=7) for now.

```
## [1] "Batches:"
```

```
## [1] "CHP-SIEMENSVERIO"     "HSC-SIEMENSPRISMAFIT" "PNC-SIEMENSTIMTRIO"
```

Confounding across sites

Age




```
## Analysis of Variance Table
##
## Response: age
##              Df  Sum Sq Mean Sq F value    Pr(>F)
## site          2  1012.5  506.27  34.606 2.313e-15 ***
## Residuals 1276 18667.2   14.63
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Age is confounded across sites.

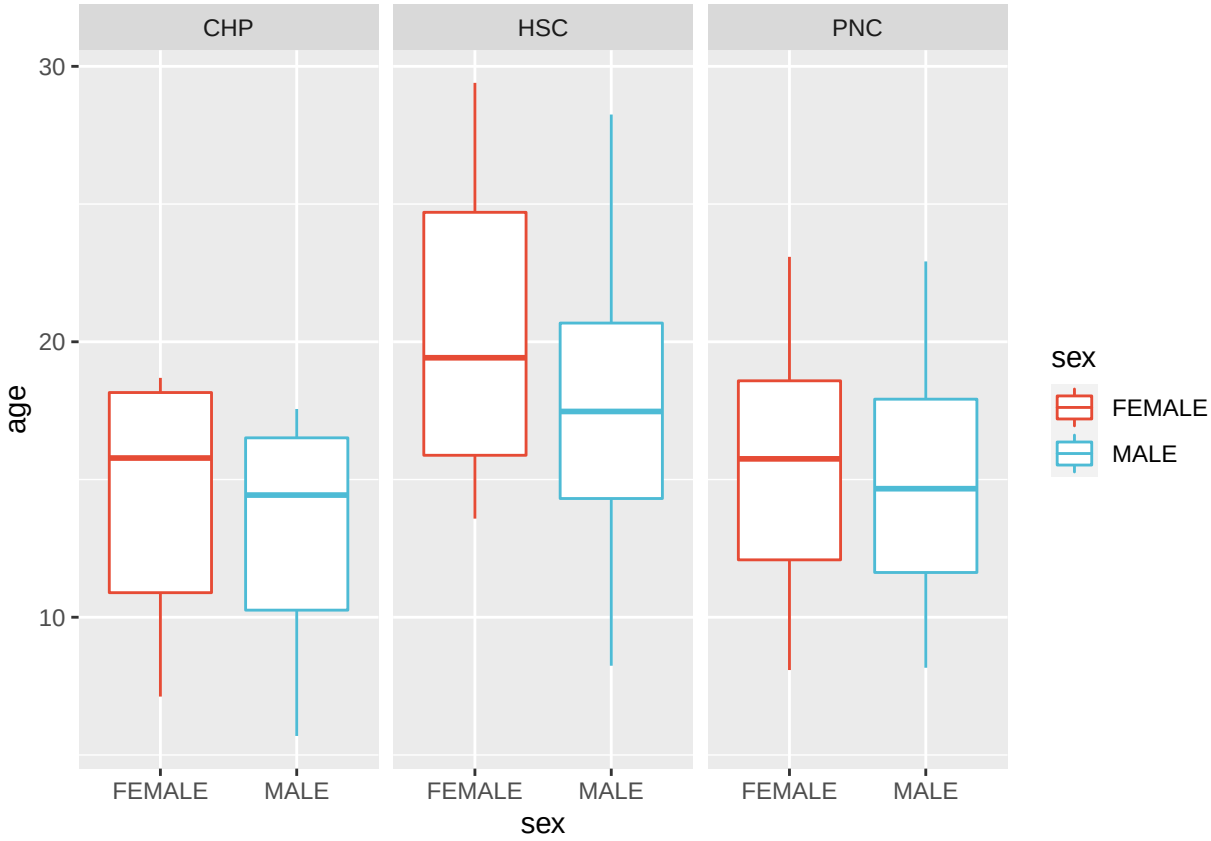
Sex

```
##
##        FEMALE MALE
##    CHP     24   12
##    HSC     37   21
##    PNC    630  555
```

```
##
##           FEMALE      MALE
##    CHP 0.6666667 0.3333333
##    HSC 0.6379310 0.3620690
##    PNC 0.5316456 0.4683544
```

```
##
##  Pearson's Chi-squared test
##
## data:  df$site and df$sex
## X-squared = 4.8976, df = 2, p-value = 0.0864
```

Sex slightly confounded.



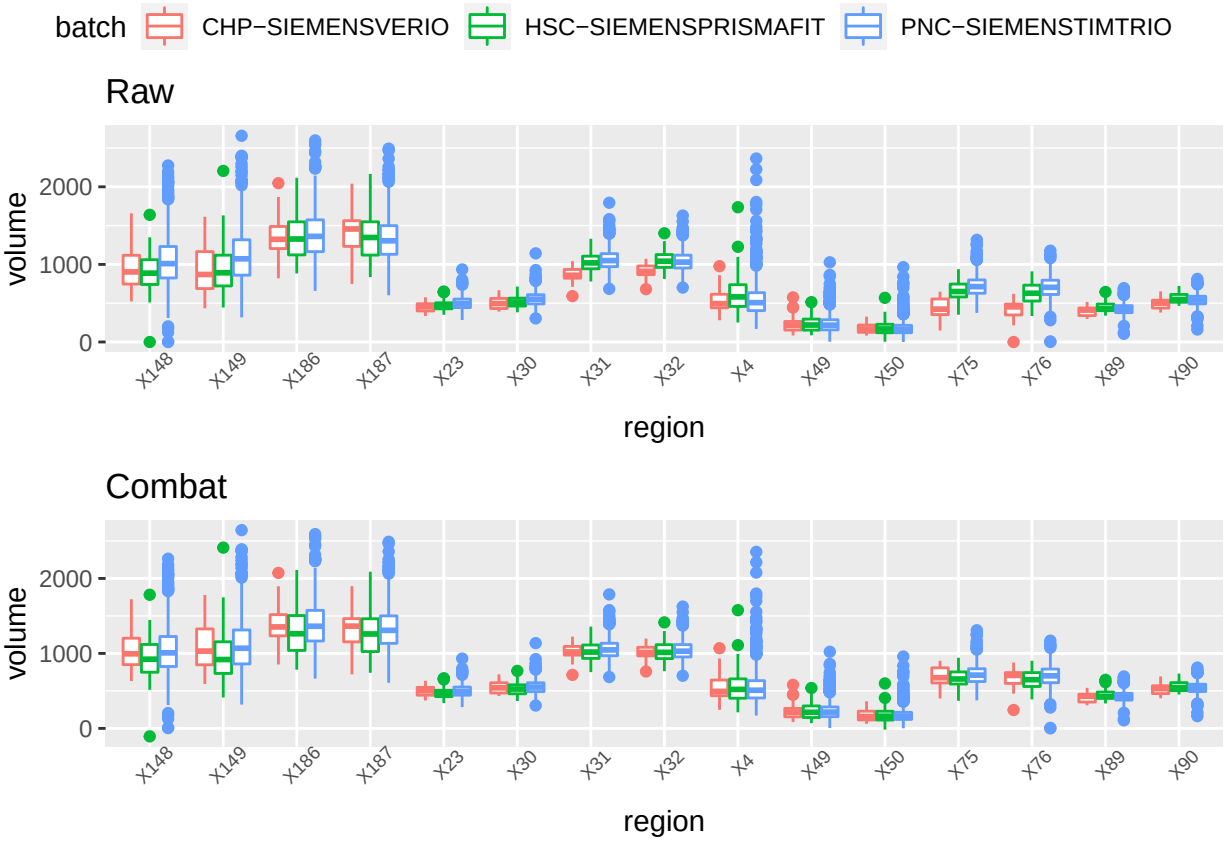
ROI volumes

Number of significant ANOVA after accounting for sites

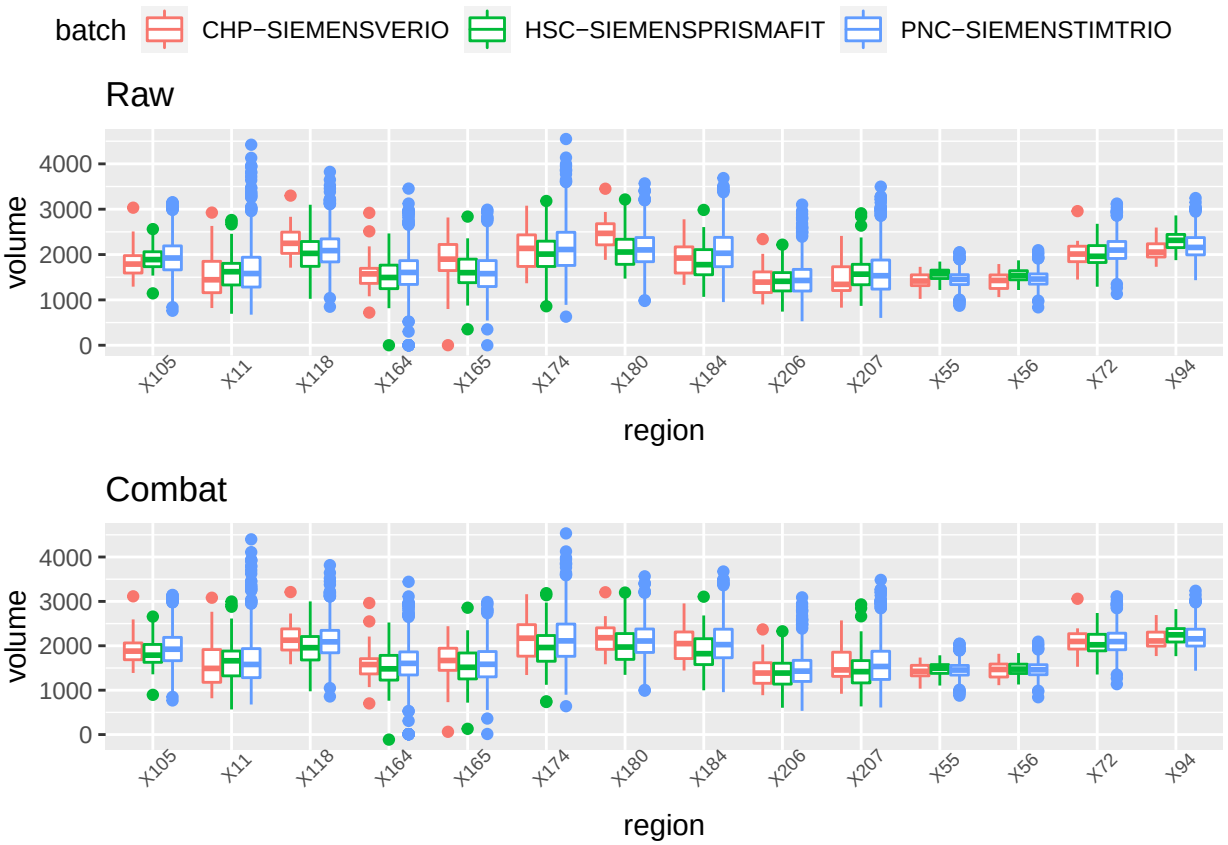
```
##    Raw Combat
##     33      0
```

ROI plots

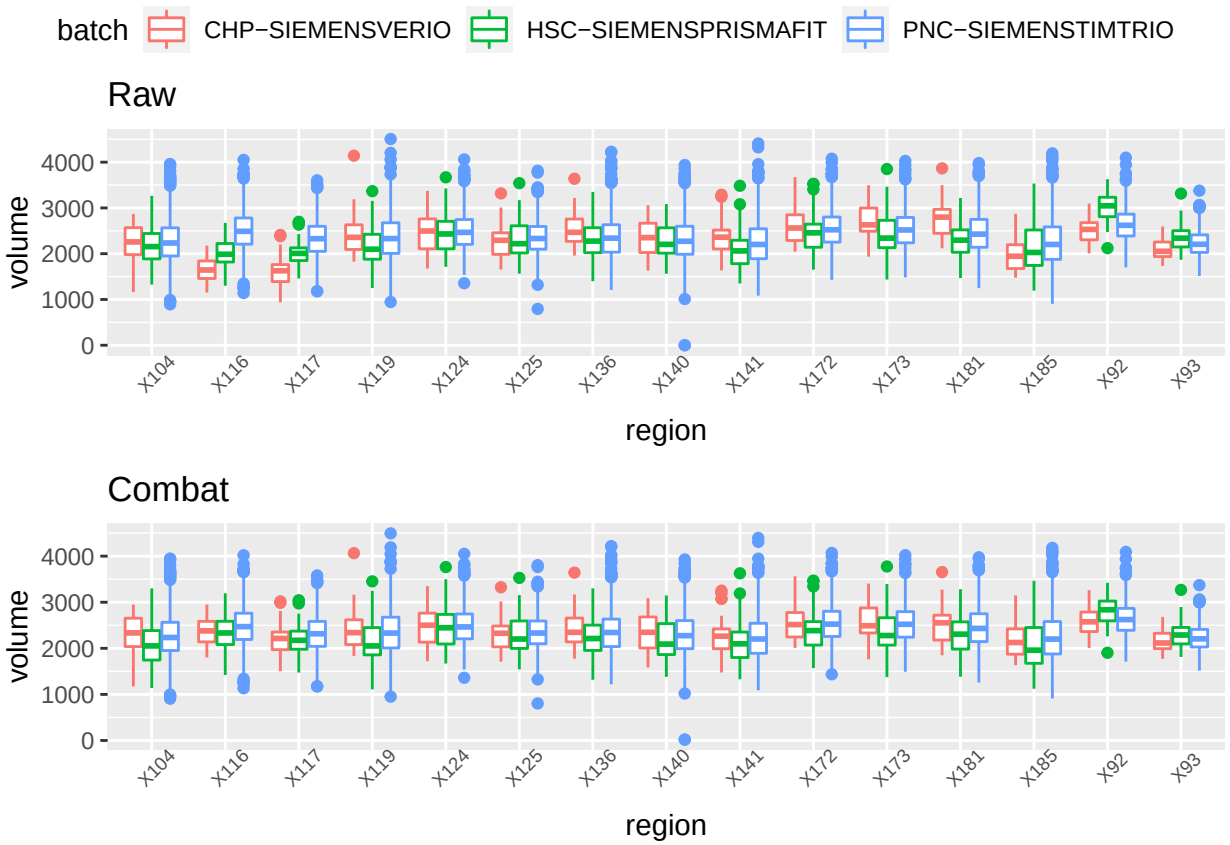
```
## [[1]]
```



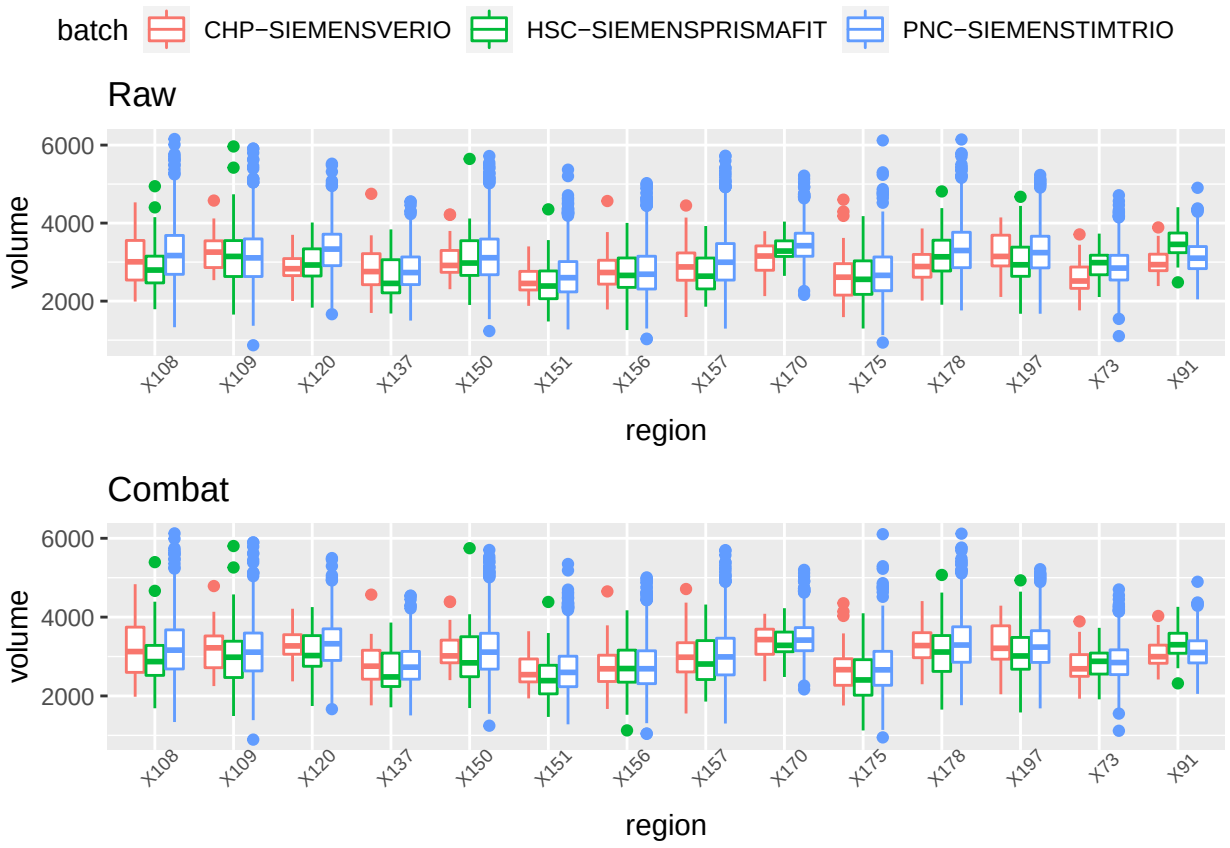

```
##
## [[2]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
4000
3000
2000
1000
0
X105
X11
X118
X164
X165
X174
X180
X184
X206
X207
X55
X56
X72
X94
region
Combat
volume
4000
3000
2000
1000
0
X105
X11
X118
X164
X165
X174
X180
X184
X206
X207
X55
X56
X72
X94
region

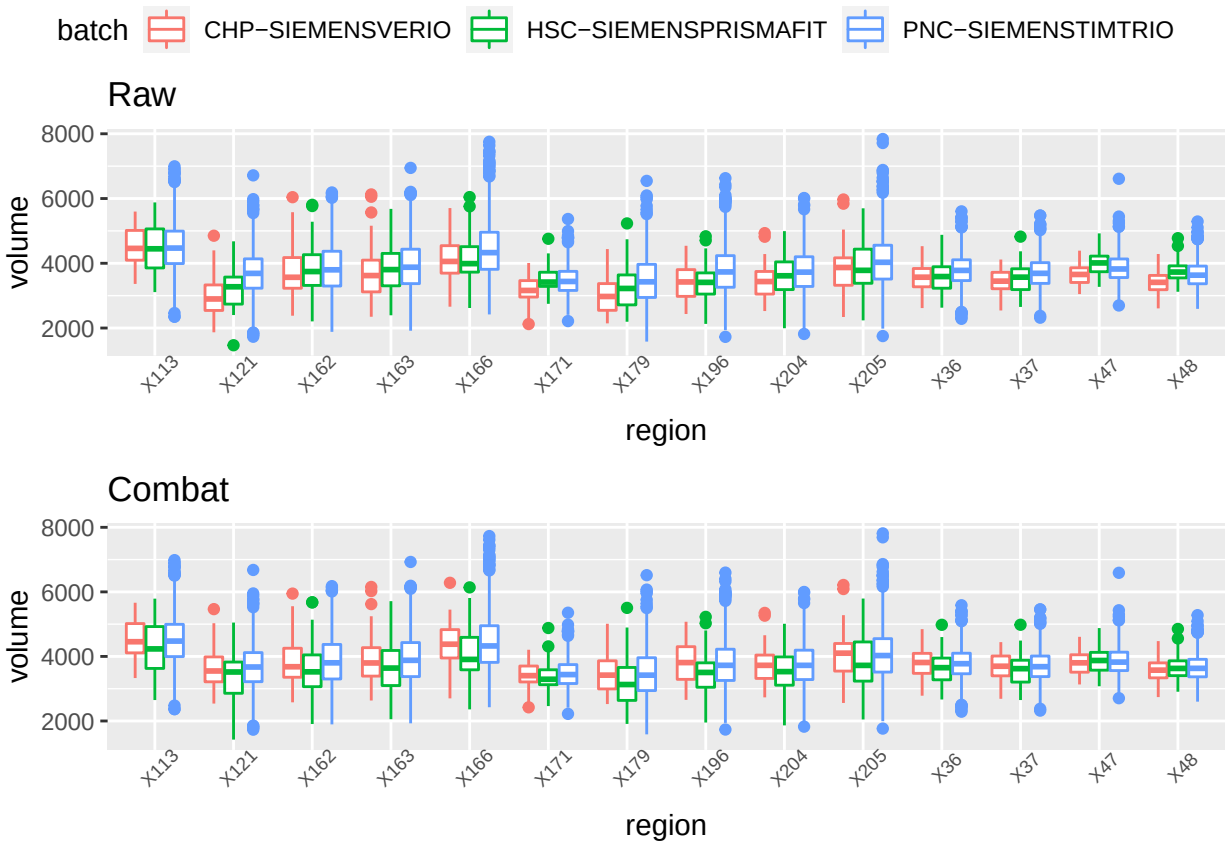
```
##
## [[3]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
4000
3000
2000
1000
0
X104
X116
X117
X119
X124
X125
X136
X140
X141
X172
X173
X181
X185
X92
X93
region
Combat
volume
4000
3000
2000
1000
0
X104
X116
X117
X119
X124
X125
X136
X140
X141
X172
X173
X181
X185
X92
X93
region

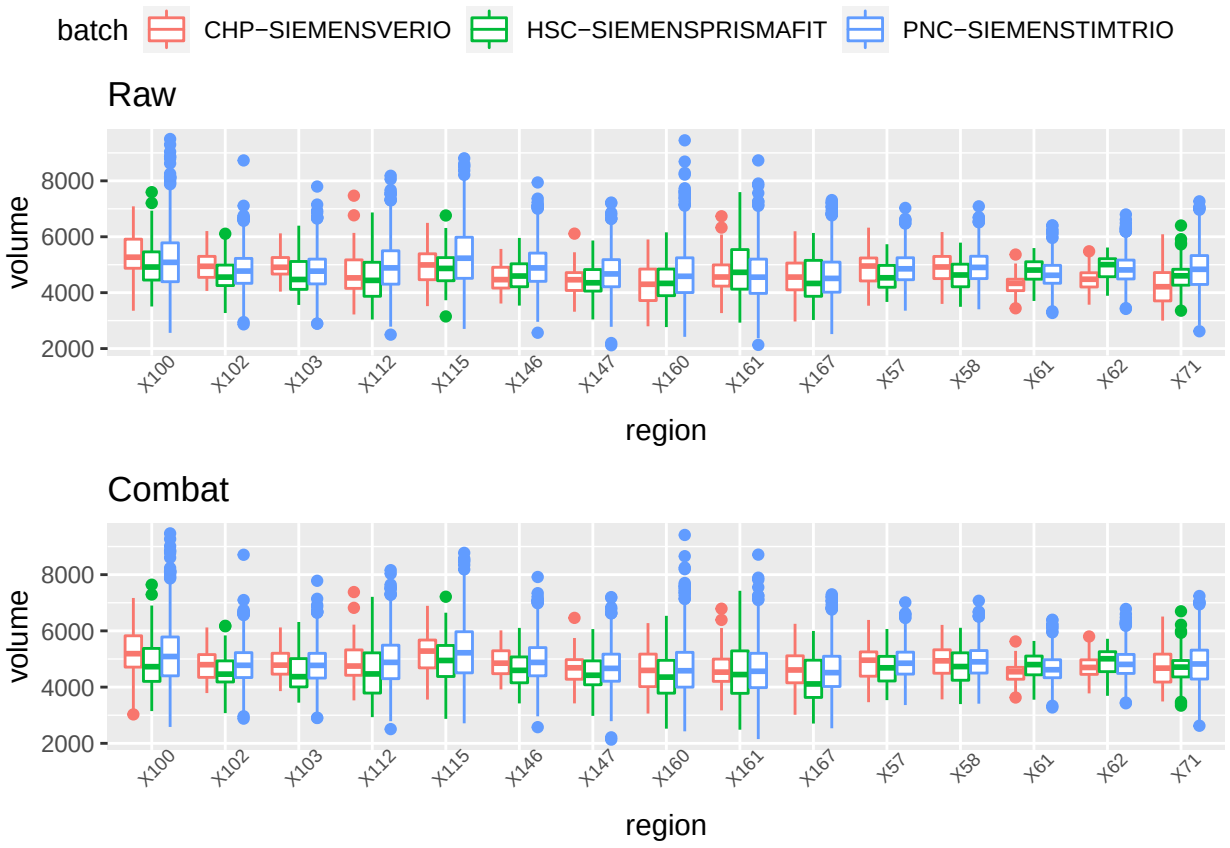
```
##
## [[4]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
6000
4000
2000
X108
X109
X120
X137
X150
X151
X156
X157
X170
X175
X178
X197
X73
X91
region
Combat
volume
6000
4000
2000
X108
X109
X120
X137
X150
X151
X156
X157
X170
X175
X178
X197
X73
X91
region

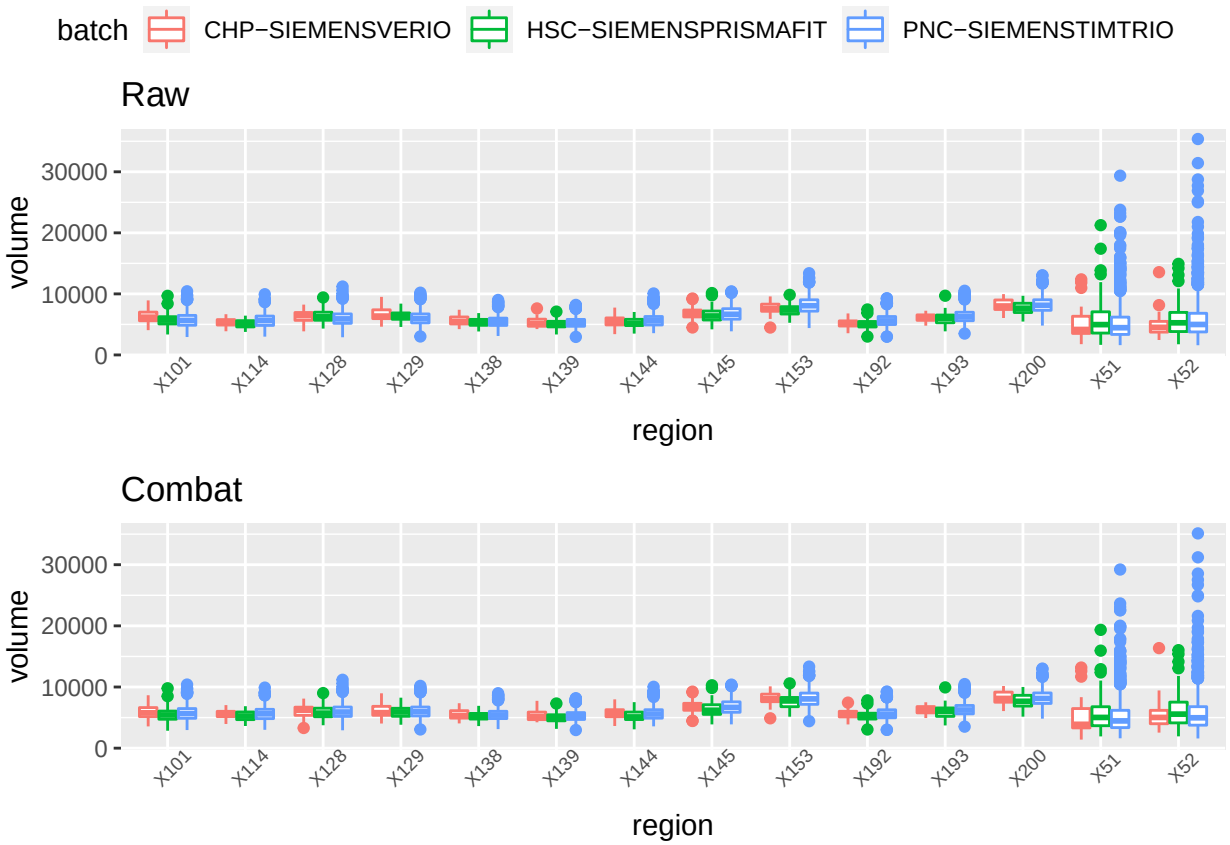
```
##
## [[5]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
8000
6000
4000
2000
X113
X121
X162
X163
X166
X171
X179
X196
X204
X205
X36
X37
X47
X48
region
Combat
volume
8000
6000
4000
2000
X113
X121
X162
X163
X166
X171
X179
X196
X204
X205
X36
X37
X47
X48
region

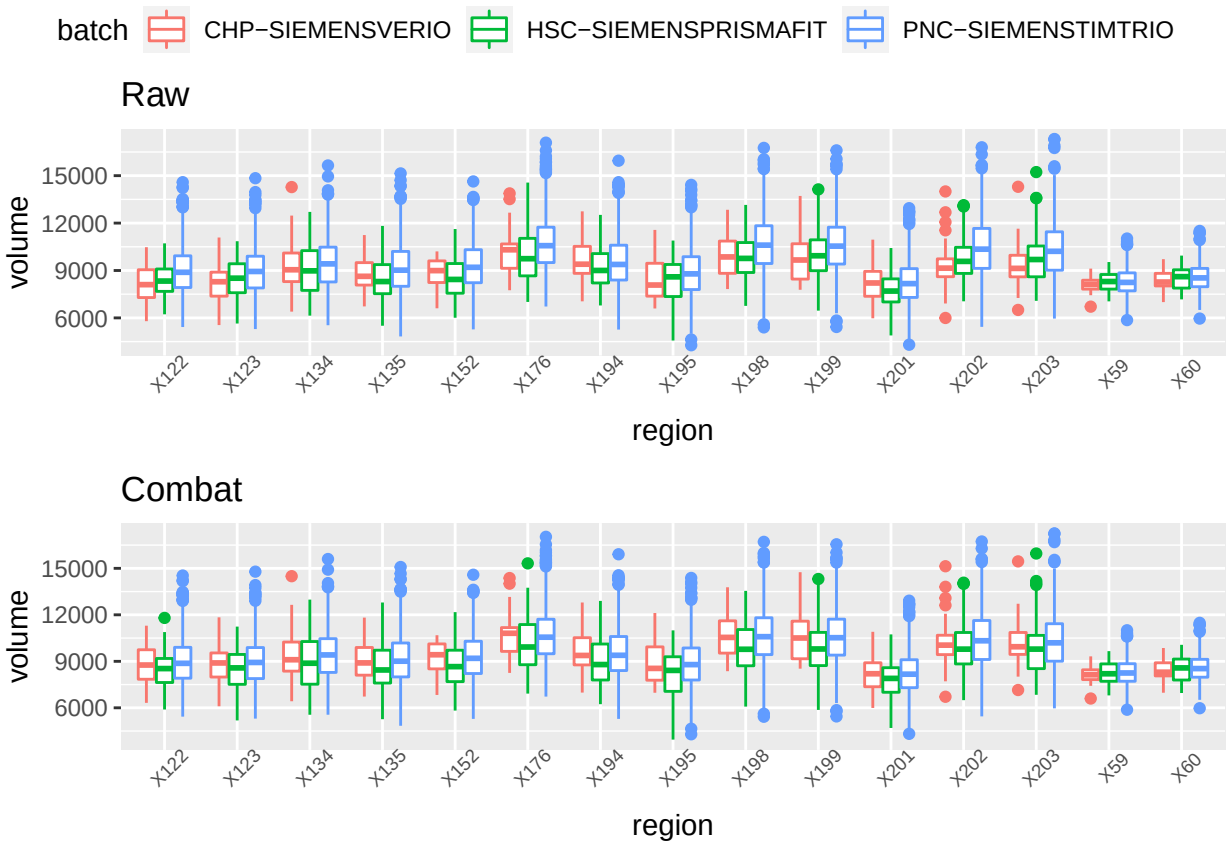
```
##
## [[6]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
8000
6000
4000
2000
X100
X102
X103
X112
X115
X146
X147
X160
X161
X167
X57
X58
X61
X62
X71
region
Combat
volume
8000
6000
4000
2000
X100
X102
X103
X112
X115
X146
X147
X160
X161
X167
X57
X58
X61
X62
X71
region

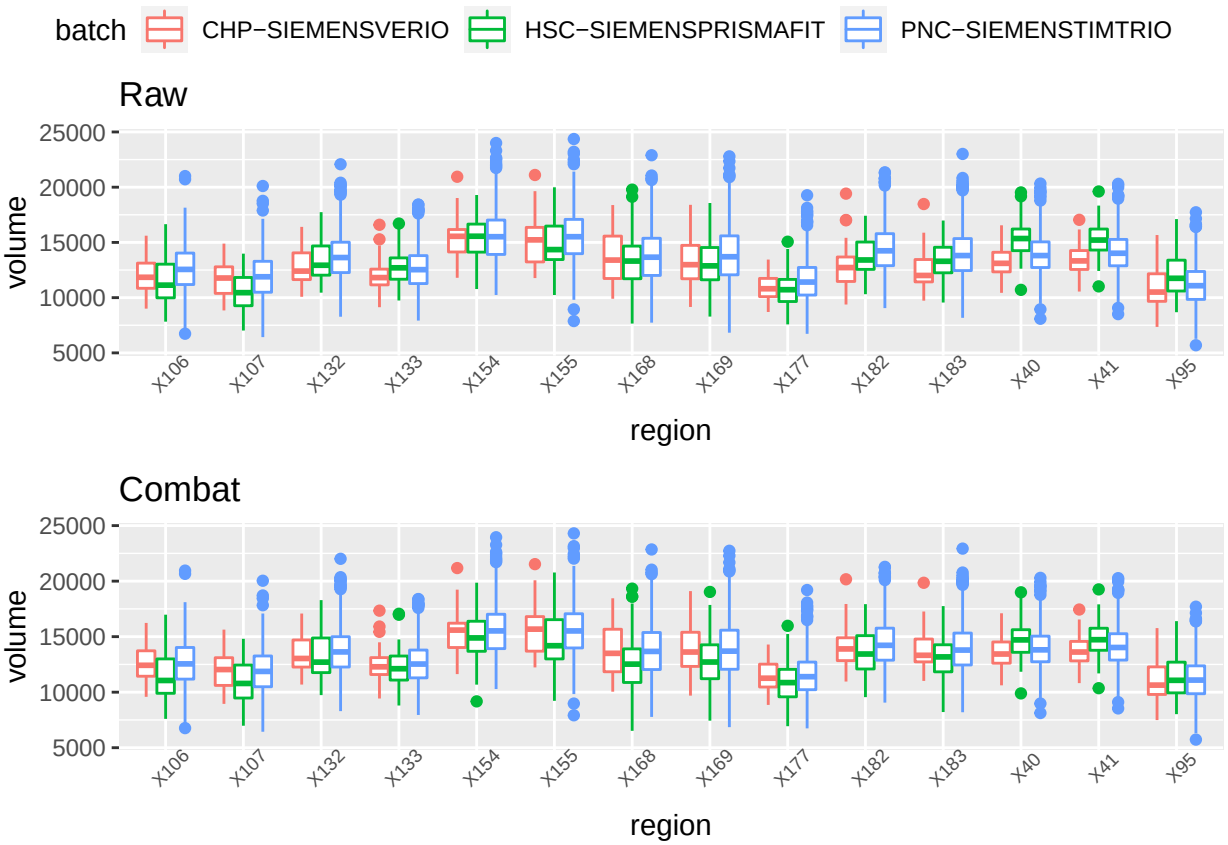
```
##
## [[7]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
30000
20000
10000
0
X101
X114
X128
X129
X138
X139
X144
X145
X153
X192
X193
X200
X51
X52
region
Combat
volume
30000
20000
10000
0
X101
X114
X128
X129
X138
X139
X144
X145
X153
X192
X193
X200
X51
X52
region

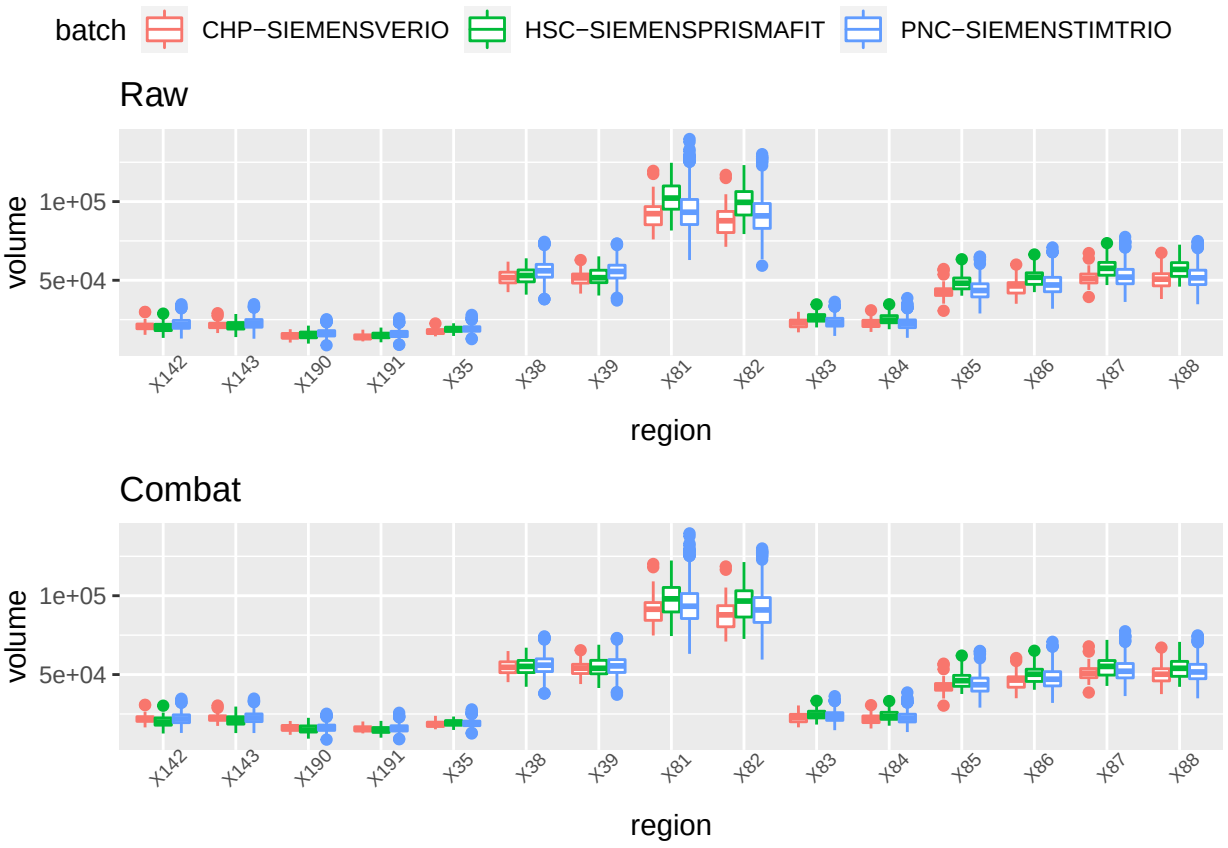
```
##
## [[8]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
15000
12000
9000
6000
X122
X123
X134
X135
X152
X176
X194
X195
X198
X199
X201
X202
X203
X59
X60
region
Combat
volume
15000
12000
9000
6000
X122
X123
X134
X135
X152
X176
X194
X195
X198
X199
X201
X202
X203
X59
X60
region

```
##
## [[9]]
```


batch
CHP−SIEMENSVERIO
HSC−SIEMENSPRISMAFIT
PNC−SIEMENSTIMTRIO
Raw
volume
25000
20000
15000
10000
5000
X106
X107
X132
X133
X154
X155
X168
X169
X177
X182
X183
X40
X41
X95
region
Combat
volume
25000
20000
15000
10000
5000
X106
X107
X132
X133
X154
X155
X168
X169
X177
X182
X183
X40
X41
X95
region

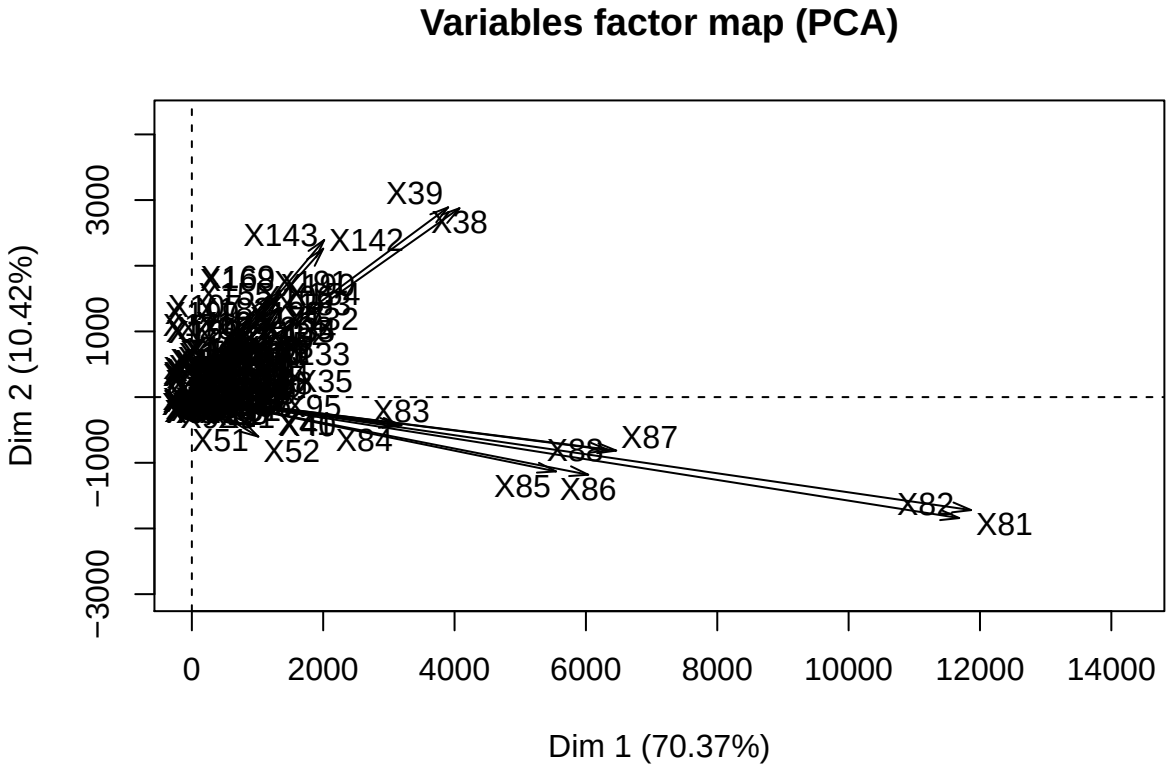
```
##
## [[10]]
```



(Covariance) PCA

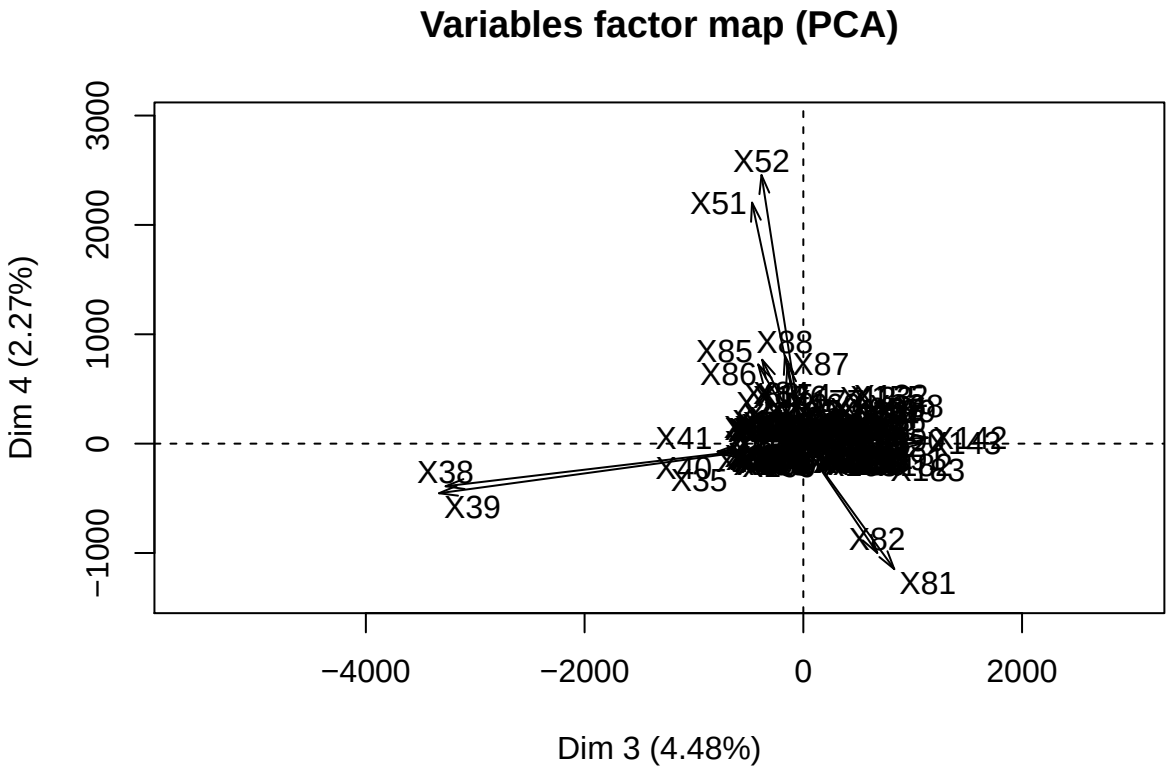
PC1: total volume (w/ higher contribution from WM areas) PC2: GM/WM PC3: Cerebellum PC4: L. Ventricles vs Frontal Lobes

```
##        eigenvalue percentage of variance cumulative percentage of variance
## comp 1  533195804              70.373731                           70.37373
## comp 2   78926919              10.417152                           80.79088
## comp 3   33940257               4.479597                           85.27048
## comp 4   17217958               2.272508                           87.54299
```




```
##        contr ROI_INDEX                ROI_NAME HEMISPHERE TISSUE_SEG
## 1  26.403469        81     frontal lobe WM right      Right         WM
## 2  25.602264        82      frontal lobe WM left       Left         WM
## 3   7.831102        87    temporal lobe WM right      Right         WM
## 4   7.559381        88     temporal lobe WM left       Left         WM
## 5   6.831837        86     parietal lobe WM left       Left         WM
## 6   5.768443        85    parietal lobe WM right      Right         WM
## 7   3.114299        38 Right Cerebellum Exterior      Right         GM
## 8   2.856952        39  Left Cerebellum Exterior       Left         GM
## 9   1.909984        83   occipital lobe WM right      Right         WM
## 10  1.846909        84    occipital lobe WM left       Left         WM
```

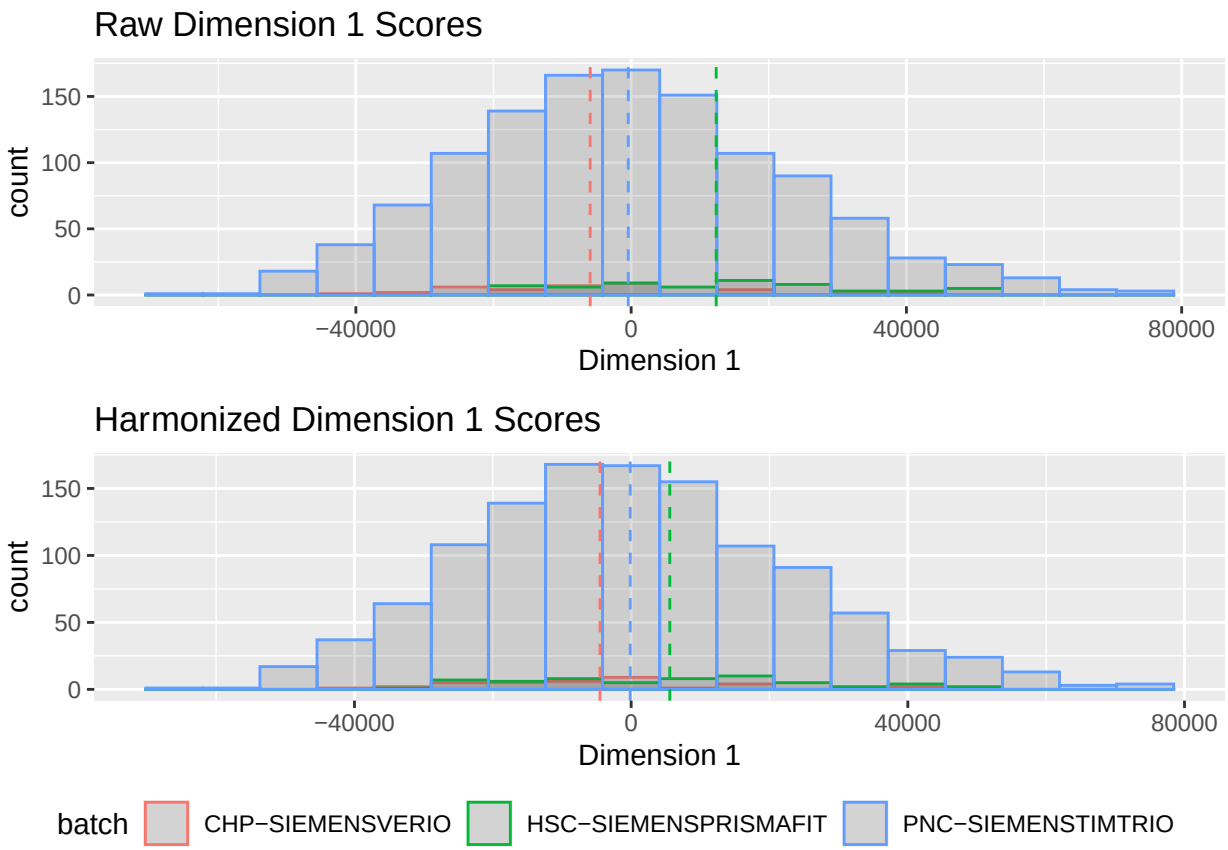
```
##        contr ROI_INDEX                          ROI_NAME HEMISPHERE TISSUE_SEG
## 1  10.582999        39           Left Cerebellum Exterior       Left         GM
## 2  10.482386        38          Right Cerebellum Exterior      Right         GM
## 3   7.251247       142   Right MFG   middle frontal gyrus      Right         GM
## 4   6.470628       143    Left MFG   middle frontal gyrus       Left         GM
## 5   4.298539        82                frontal lobe WM left       Left         WM
## 6   3.749310        81               frontal lobe WM right      Right         WM
## 7   3.326579       169                Left PCu   precuneus       Left         GM
## 8   3.108160       168               Right PCu   precuneus      Right         GM
## 9   3.016338       191  Left SFG   superior frontal gyrus       Left         GM
## 10  2.807871       190 Right SFG   superior frontal gyrus      Right         GM
```




```
##        contr ROI_INDEX                       ROI_NAME HEMISPHERE TISSUE_SEG
## 1  32.702268        39         Left Cerebellum Exterior       Left         GM
## 2  31.524971        38        Right Cerebellum Exterior      Right         GM
## 3   3.818291       142 Right MFG   middle frontal gyrus      Right         GM
## 4   3.464686       143  Left MFG   middle frontal gyrus       Left         GM
## 5   2.039264        81            frontal lobe WM right      Right         WM
## 6   1.819255        40    Right Cerebellum White Matter      Right         WM
## 7   1.814011        41     Left Cerebellum White Matter       Left         WM
## 8   1.665179       183      Left PrG   precentral gyrus       Left         GM
## 9   1.347651        82             frontal lobe WM left       Left         WM
## 10  1.226949        35                       Brain Stem       Both       NONE
```

```
##         contr ROI_INDEX                  ROI_NAME HEMISPHERE TISSUE_SEG
## 1  35.0256199        52    Left Lateral Ventricle       Left         VN
## 2  28.1972312        51   Right Lateral Ventricle      Right         VN
## 3   7.6493714        81     frontal lobe WM right      Right         WM
## 4   5.8035684        82      frontal lobe WM left       Left         WM
## 5   3.7506007        88     temporal lobe WM left       Left         WM
## 6   3.4175053        86     parietal lobe WM left       Left         WM
## 7   3.0766717        87    temporal lobe WM right      Right         WM
## 8   3.0428988        85    parietal lobe WM right      Right         WM
## 9   1.1947441        39  Left Cerebellum Exterior       Left         GM
## 10  0.8817551        38 Right Cerebellum Exterior      Right         GM
```

PC1



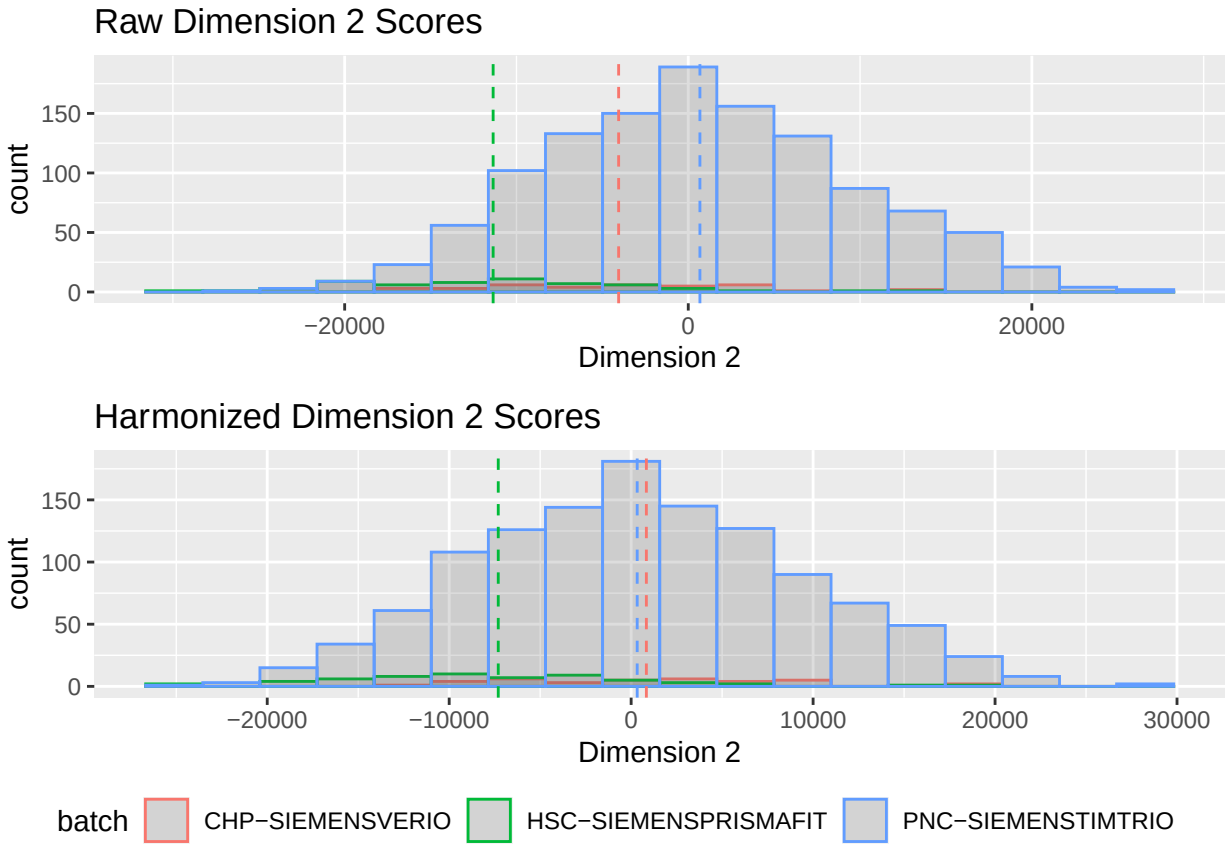

```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc1) ~ batch + age + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4608  -0.5675  -0.0262   0.5276   2.8222
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.592861   0.163874  -9.720   <2e-16 ***
## batchHSC-SIEMENSPRISMAFIT   0.373088   0.179210   2.082   0.0376 *
## batchPNC-SIEMENSTIMTRIO     0.015492   0.140852   0.110   0.9124
## age                         0.072177   0.006104  11.824   <2e-16 ***
## sexMALE                     0.998936   0.046883  21.307   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.689549)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  878.49  on 1274  degrees of freedom
```

```
## AIC: 3161.2
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc1) ~ batch + age + sex, data = dat_combat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.46176  -0.57017  -0.02811   0.53092   2.82099
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -1.528298   0.164803  -9.273   <2e-16 ***
## batchHSC-SIEMENSPRISMAFIT   0.024516   0.180226   0.136    0.892
## batchPNC-SIEMENSTIMTRIO    -0.034021   0.141651  -0.240    0.810
## age                         0.071721   0.006139  11.683   <2e-16 ***
## sexMALE                     1.007826   0.047149  21.375   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.697393)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  888.48  on 1274  degrees of freedom
## AIC: 3175.7
##
## Number of Fisher Scoring iterations: 2
```

PC2

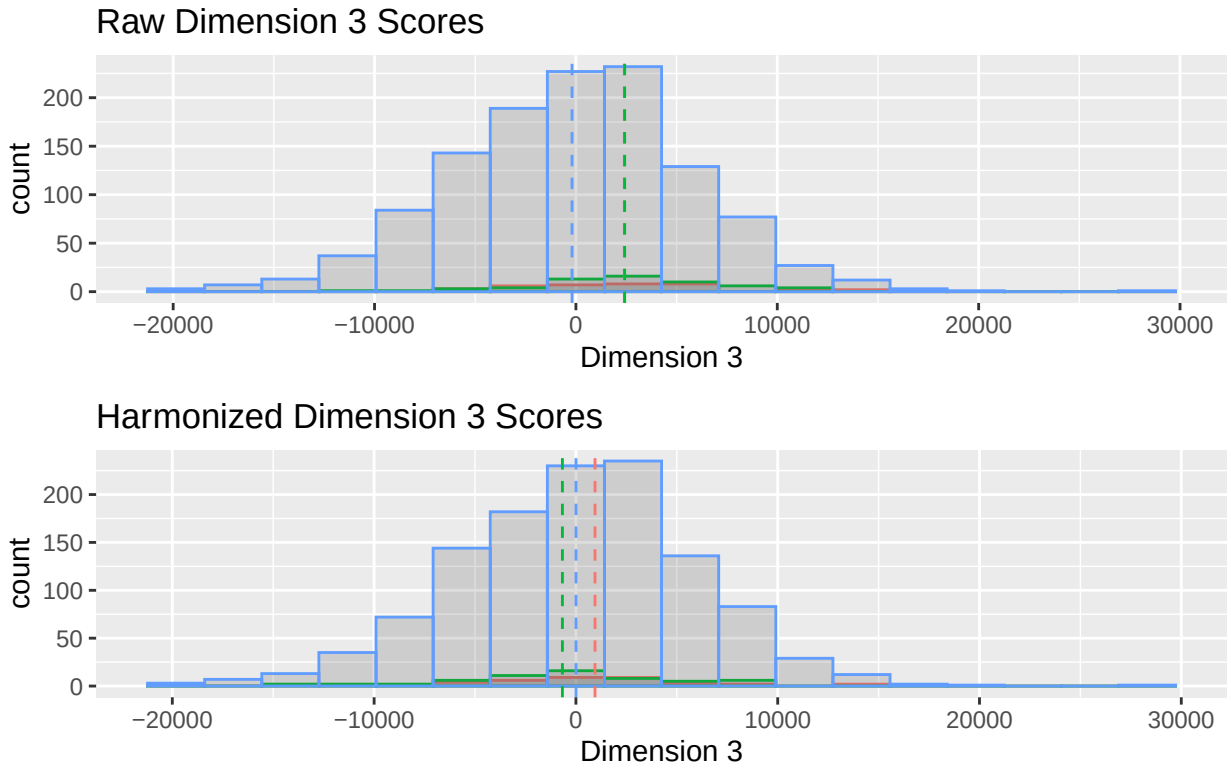



```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc2) ~ batch + age + sex, data = dat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29198  -0.48643  -0.03406   0.48108   2.05144
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.553387   0.135982  11.424  < 2e-16 ***
## batchHSC-SIEMENSPRISMAFIT  0.017259   0.148707   0.116    0.908
## batchPNC-SIEMENSTIMTRIO    0.638097   0.116879   5.459 5.73e-08 ***
## age                       -0.155728   0.005065 -30.744  < 2e-16 ***
## sexMALE                    0.506897   0.038903  13.030  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.474796)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  604.89  on 1274  degrees of freedom
```

```
## AIC: 2683.9
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc2) ~ batch + age + sex, data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3307  -0.4842  -0.0361   0.4848   2.0988
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 2.201373   0.138195  15.929   <2e-16 ***
## batchHSC-SIEMENSPRISMAFIT -0.057545   0.151128  -0.381    0.703
## batchPNC-SIEMENSTIMTRIO     0.074208   0.118781   0.625    0.532
## age                        -0.163355   0.005148 -31.733   <2e-16 ***
## sexMALE                     0.494559   0.039537  12.509   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.4903809)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  624.75  on 1274  degrees of freedom
## AIC: 2725.3
##
## Number of Fisher Scoring iterations: 2
```



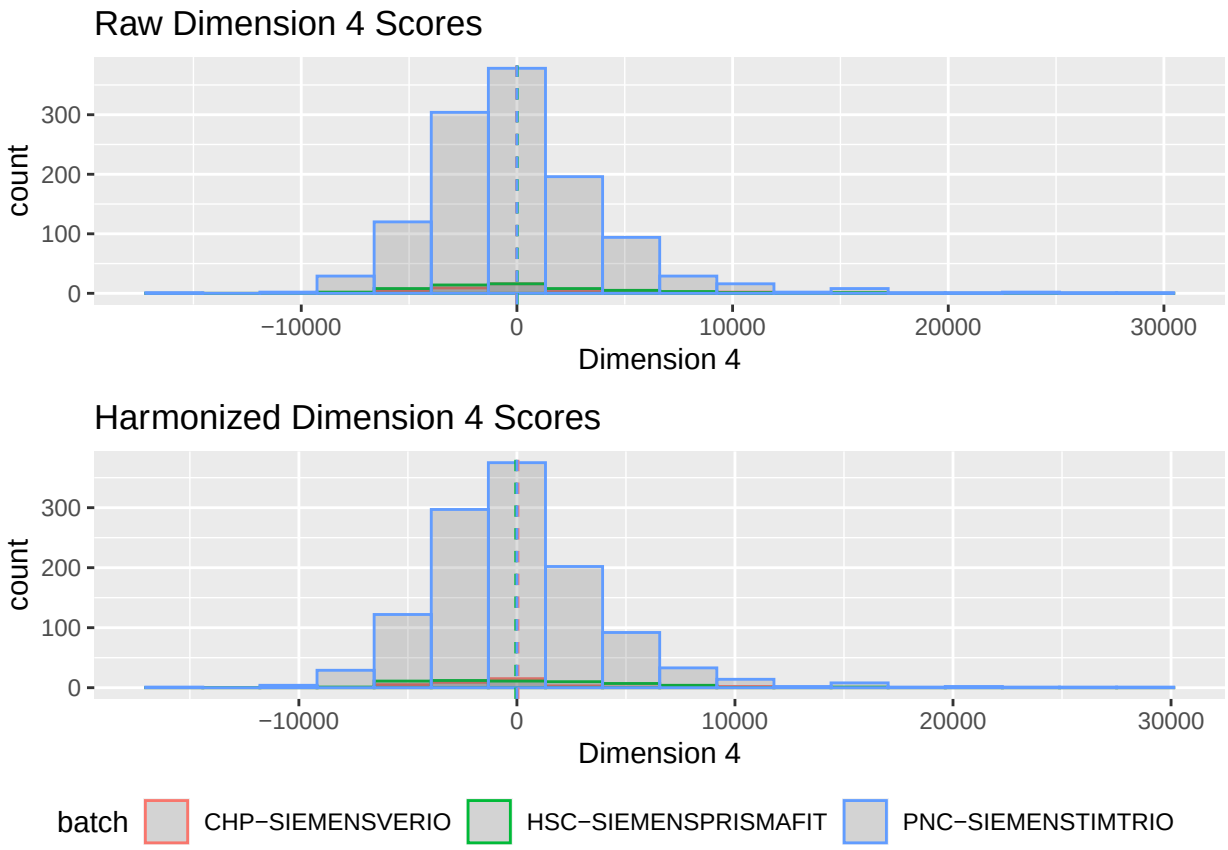
###PC3

```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc3) ~ batch + age + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2145  -0.6257   0.0497   0.6084   4.3267
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                1.547371   0.186955   8.277 3.16e-16 ***
## batchHSC-SIEMENSPRISMAFIT  0.430204   0.204451   2.104   0.0356 *
## batchPNC-SIEMENSTIMTRIO   -0.332042   0.160691  -2.066   0.0390 *
## age                       -0.079515   0.006964 -11.418  < 2e-16 ***
## sexMALE                   -0.097504   0.053486  -1.823   0.0685 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.8974707)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1143.4  on 1274  degrees of freedom
## AIC: 3498.3
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc3) ~ batch + age + sex, data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2929  -0.6488   0.0554   0.6223   4.3954
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.159474   0.190497   6.087 1.52e-09 ***
## batchHSC-SIEMENSPRISMAFIT   0.093404   0.208325   0.448   0.6540
## batchPNC-SIEMENSTIMTRIO    -0.061942   0.163736  -0.378   0.7053
## age                        -0.068752   0.007096  -9.689  < 2e-16 ***
## sexMALE                    -0.122431   0.054500  -2.246   0.0248 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9318016)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1187.1  on 1274  degrees of freedom
## AIC: 3546.3
##
## Number of Fisher Scoring iterations: 2
```

PC4




```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc4) ~ batch + age + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8037  -0.6140  -0.0957   0.4637   7.0235
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.005306   0.197461  -0.027    0.979
## batchHSC-SIEMENSPRISMAFIT   0.016352   0.215941   0.076    0.940
## batchPNC-SIEMENSTIMTRIO    -0.005198   0.169722  -0.031    0.976
## age                        -0.001978   0.007355  -0.269    0.788
## sexMALE                     0.086108   0.056492   1.524    0.128
##
## (Dispersion parameter for gaussian family taken to be 1.001179)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.5  on 1274  degrees of freedom
## AIC: 3638.1
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc4) ~ batch + age + sex, data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8461  -0.6179  -0.0937   0.4737   6.8875
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.0055162  0.1974518  -0.028    0.978
## batchHSC-SIEMENSPRISMAFIT -0.0229680  0.2159302  -0.106    0.915
## batchPNC-SIEMENSTIMTRIO   -0.0228933  0.1697135  -0.135    0.893
## age                       -0.0008873  0.0073551  -0.121    0.904
## sexMALE                    0.0898793  0.0564895   1.591    0.112
##
## (Dispersion parameter for gaussian family taken to be 1.001081)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.4  on 1274  degrees of freedom
## AIC: 3638
##
## Number of Fisher Scoring iterations: 2
```

PCs ~ Age

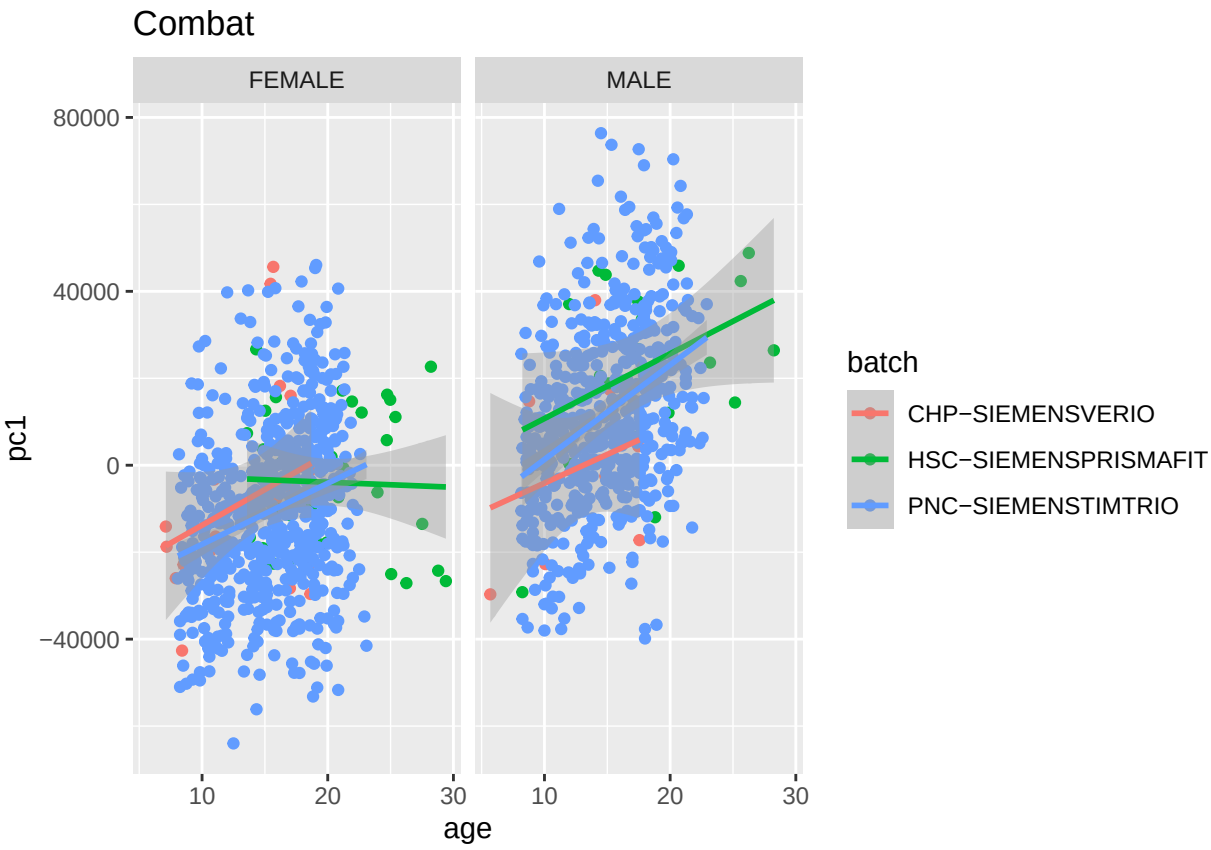
PC1

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

```
## [1] "RAW"
```

```
##
## Call:
## glm(formula = scale(pc1) ~ batch + scale(age) + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4608  -0.5675  -0.0262   0.5276   2.8222
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.49052    0.13944  -3.518  0.00045 ***
## batchHSC-SIEMENSPRISMAFIT   0.37309    0.17921   2.082  0.03756 *
## batchPNC-SIEMENSTIMTRIO     0.01549    0.14085   0.110  0.91244
## scale(age)                  0.28323    0.02395  11.824  < 2e-16 ***
## sexMALE                     0.99894    0.04688  21.307  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for gaussian family taken to be 0.689549)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  878.49  on 1274  degrees of freedom
## AIC: 3161.2
##
## Number of Fisher Scoring iterations: 2

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

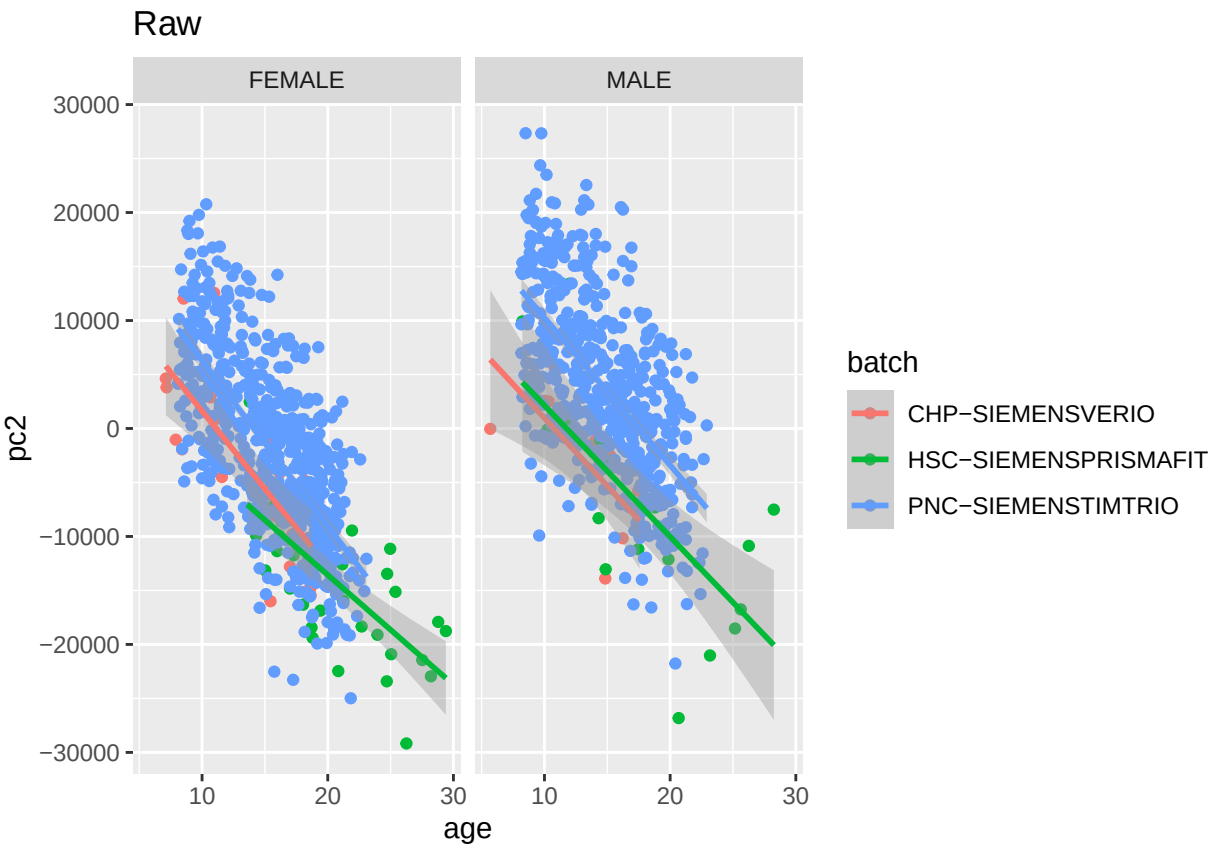



```
## [1] "COMBAT"

##
## Call:
## glm(formula = scale(pc1) ~ batch + scale(age) + sex, data = dat_combat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.46176  -0.57017  -0.02811   0.53092   2.82099
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.43292    0.14023  -3.087  0.00206 **
## batchHSC-SIEMENSPRISMAFIT   0.02452    0.18023   0.136  0.89182
## batchPNC-SIEMENSTIMTRIO    -0.03402    0.14165  -0.240  0.81023
## scale(age)                  0.28144    0.02409  11.683  < 2e-16 ***
```

```
## sexMALE                     1.00783    0.04715  21.375  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.697393)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  888.48  on 1274  degrees of freedom
## AIC: 3175.7
##
## Number of Fisher Scoring iterations: 2
```

PC2

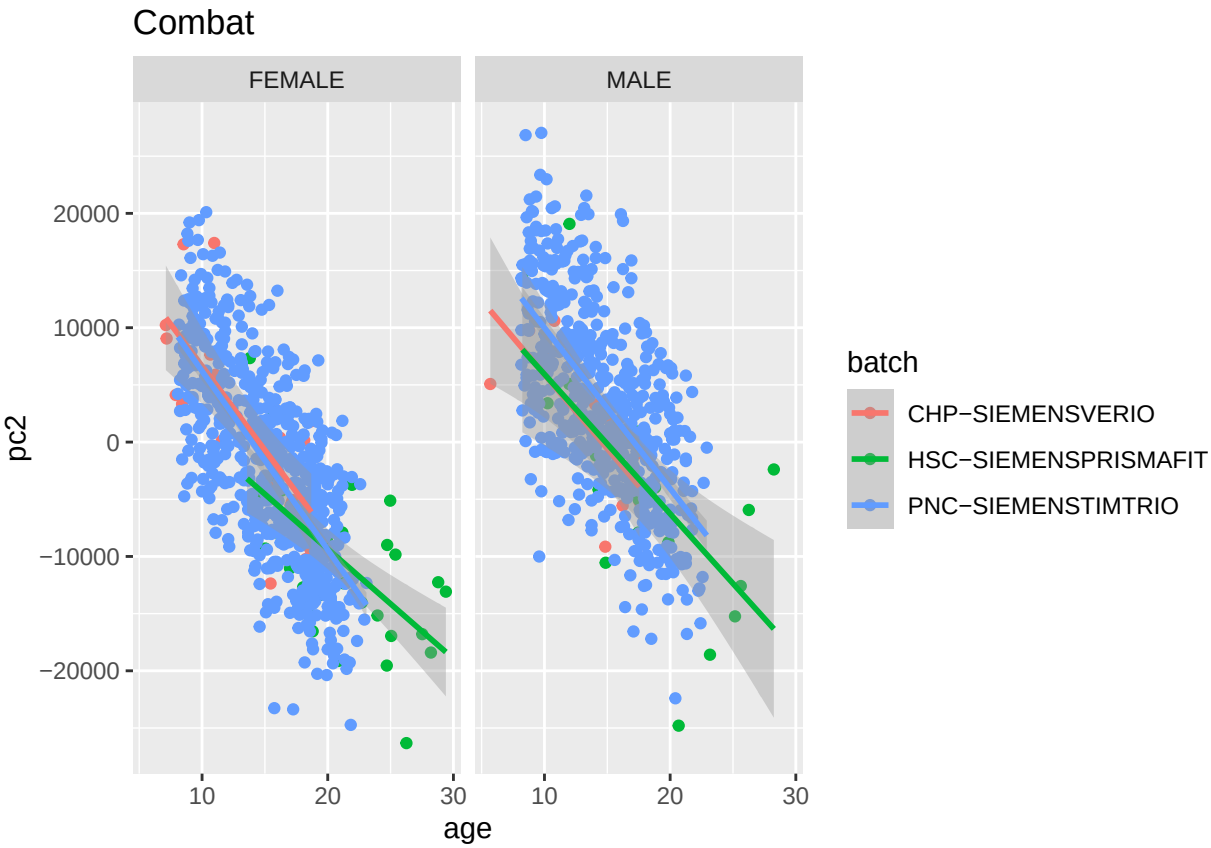
```
## `geom_smooth()` using formula 'y ~ x'
```




```
## [1] "RAW"
```

```
##
## Call:
## glm(formula = scale(pc2) ~ batch + scale(age) + sex, data = dat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29198  -0.48643  -0.03406   0.48108   2.05144
```

```
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.82502    0.11571  -7.130 1.67e-12 ***
## batchHSC-SIEMENSPRISMAFIT   0.01726    0.14871   0.116    0.908
## batchPNC-SIEMENSTIMTRIO     0.63810    0.11688   5.459 5.73e-08 ***
## scale(age)                 -0.61110    0.01988 -30.744  < 2e-16 ***
## sexMALE                     0.50690    0.03890  13.030  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.474796)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  604.89  on 1274  degrees of freedom
## AIC: 2683.9
##
## Number of Fisher Scoring iterations: 2

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

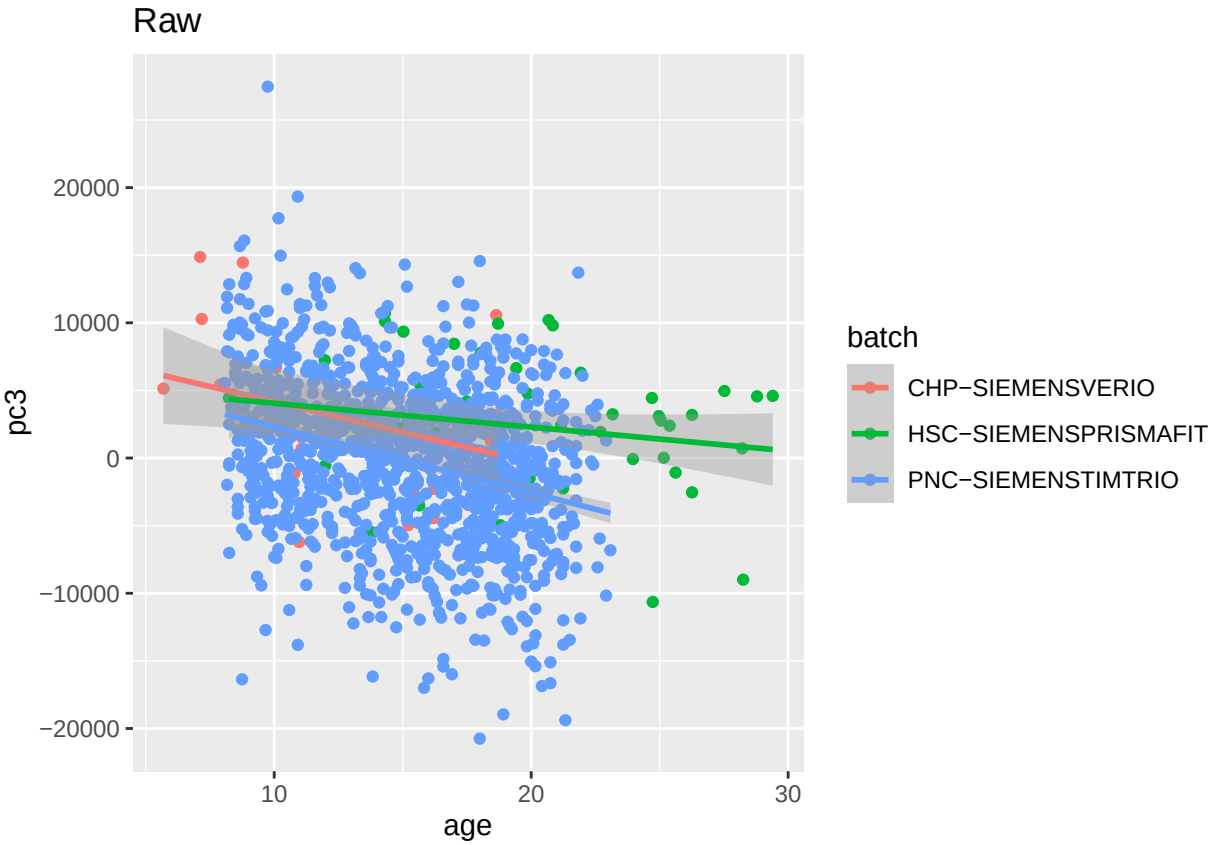



```
## [1] "COMBAT"

##
## Call:
## glm(formula = scale(pc2) ~ batch + scale(age) + sex, data = dat_combat)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3307  -0.4842  -0.0361   0.4848   2.0988
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.29351    0.11759  -2.496   0.0127 *
## batchHSC-SIEMENSPRISMAFIT -0.05754    0.15113  -0.381   0.7034
## batchPNC-SIEMENSTIMTRIO    0.07421    0.11878   0.625   0.5322
## scale(age)                 -0.64102    0.02020 -31.733   <2e-16 ***
## sexMALE                     0.49456    0.03954  12.509   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.4903809)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  624.75  on 1274  degrees of freedom
## AIC: 2725.3
##
## Number of Fisher Scoring iterations: 2
```

PC3

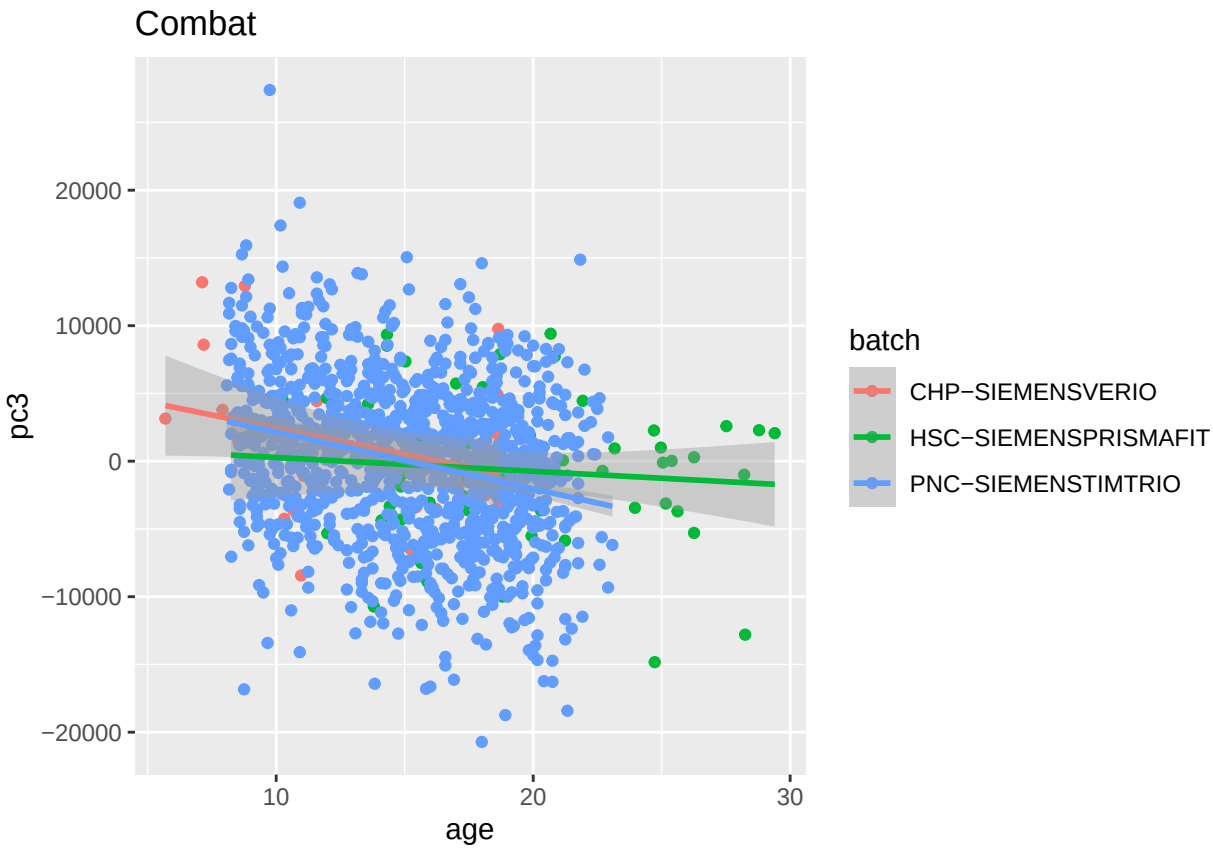
```
## `geom_smooth()` using formula 'y ~ x'
```



```
## [1] "RAW"
```

```
##
## Call:
## glm(formula = scale(pc3) ~ batch + scale(age) + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2145  -0.6257   0.0497   0.6084   4.3267
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.33296    0.15908   2.093   0.0365 *
## batchHSC-SIEMENSPRISMAFIT   0.43020    0.20445   2.104   0.0356 *
## batchPNC-SIEMENSTIMTRIO    -0.33204    0.16069  -2.066   0.0390 *
## scale(age)                 -0.31203    0.02733 -11.418   <2e-16 ***
## sexMALE                    -0.09750    0.05349  -1.823   0.0685 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.8974707)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1143.4  on 1274  degrees of freedom
## AIC: 3498.3
##
## Number of Fisher Scoring iterations: 2
```

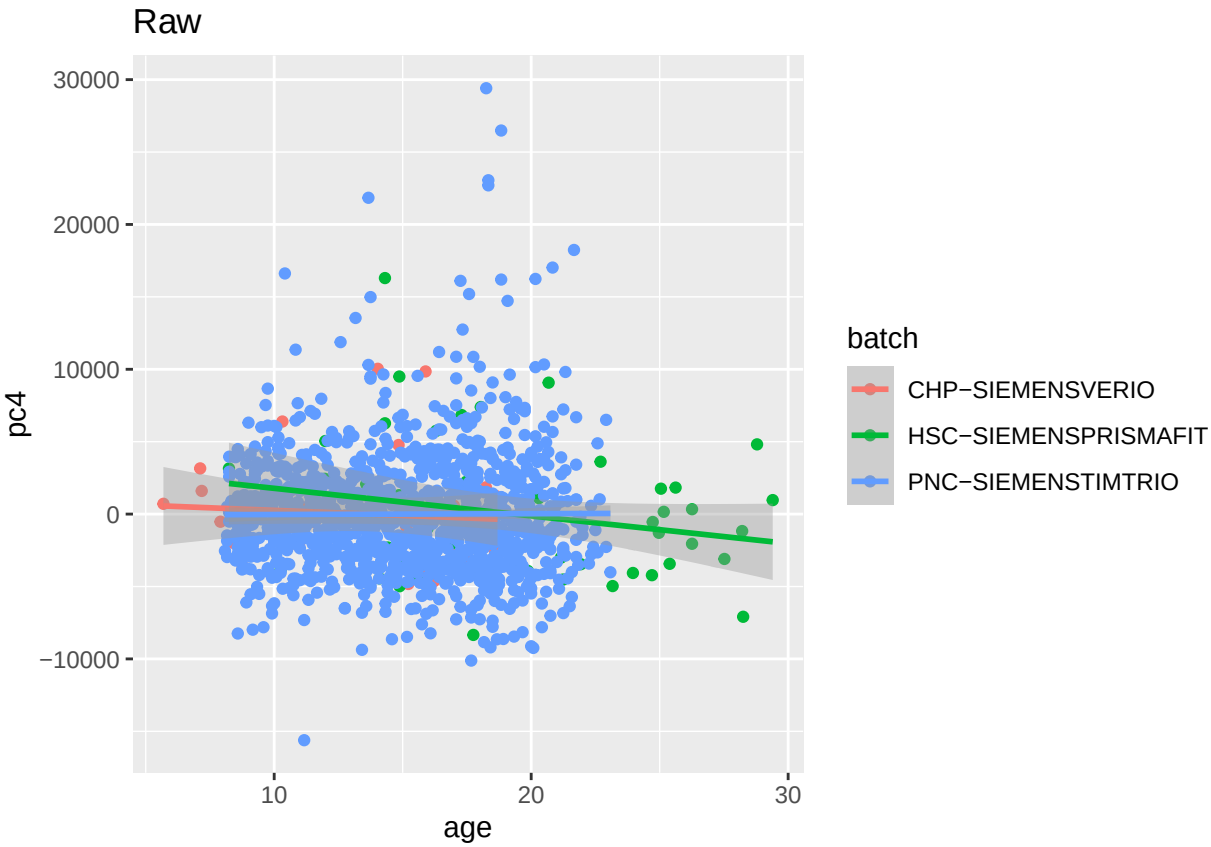
```
## `geom_smooth()` using formula 'y ~ x'
```




```
## [1] "COMBAT"

##
## Call:
## glm(formula = scale(pc3) ~ batch + scale(age) + sex, data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2929  -0.6488   0.0554   0.6223   4.3954
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.10944    0.16209   0.675   0.4997
## batchHSC-SIEMENSPRISMAFIT   0.09340    0.20832   0.448   0.6540
## batchPNC-SIEMENSTIMTRIO    -0.06194    0.16374  -0.378   0.7053
## scale(age)                 -0.26979    0.02785  -9.689   <2e-16 ***
## sexMALE                    -0.12243    0.05450  -2.246   0.0248 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9318016)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1187.1  on 1274  degrees of freedom
## AIC: 3546.3
##
## Number of Fisher Scoring iterations: 2
```

PC4

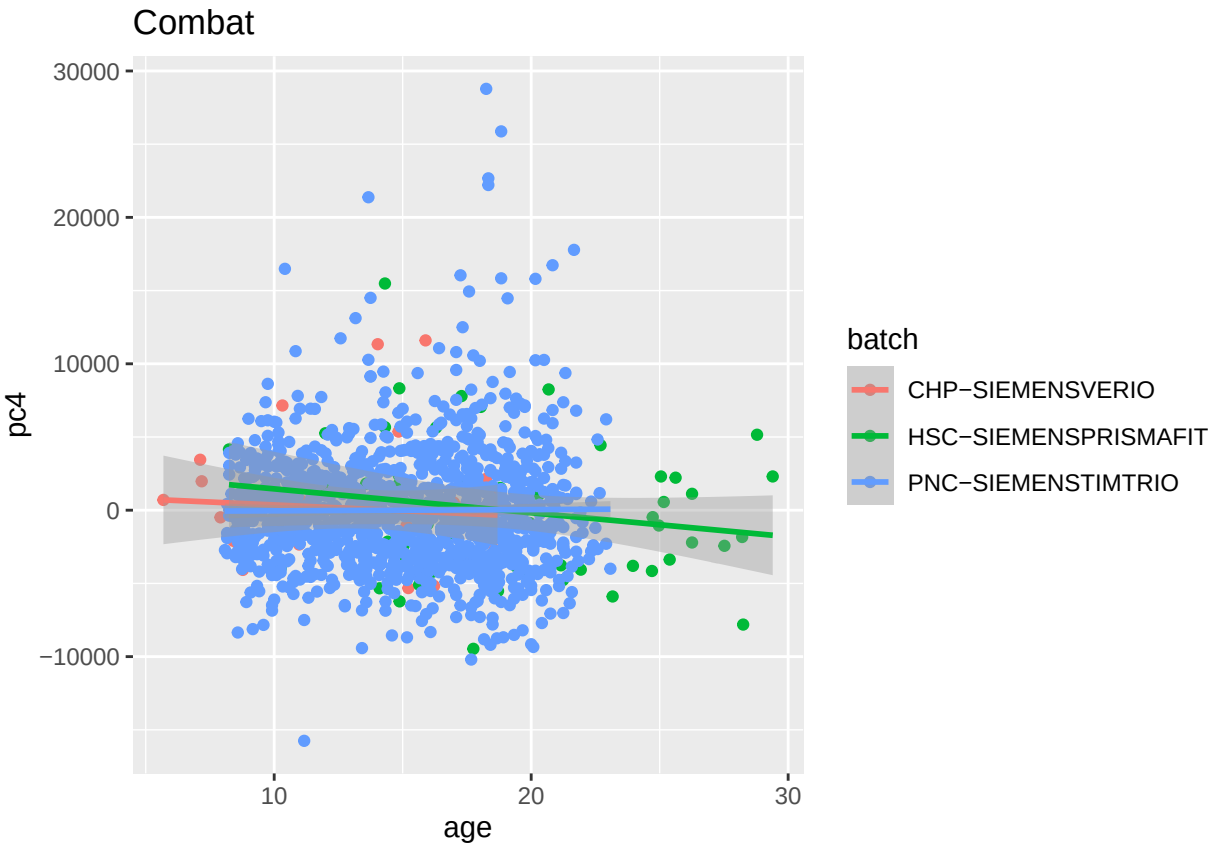
```
## `geom_smooth()` using formula 'y ~ x'
```




```
## [1] "RAW"
```

```
##
## Call:
## glm(formula = scale(pc4) ~ batch + scale(age) + sex, data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8037  -0.6140  -0.0957   0.4637   7.0235
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.035512   0.168019  -0.211    0.833
## batchHSC-SIEMENSPRISMAFIT  0.016352   0.215941   0.076    0.940
## batchPNC-SIEMENSTIMTRIO   -0.005198   0.169722  -0.031    0.976
## scale(age)                -0.007761   0.028864  -0.269    0.788
## sexMALE                    0.086108   0.056492   1.524    0.128
##
## (Dispersion parameter for gaussian family taken to be 1.001179)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.5  on 1274  degrees of freedom
```

```
## AIC: 3638.1
##
## Number of Fisher Scoring iterations: 2

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

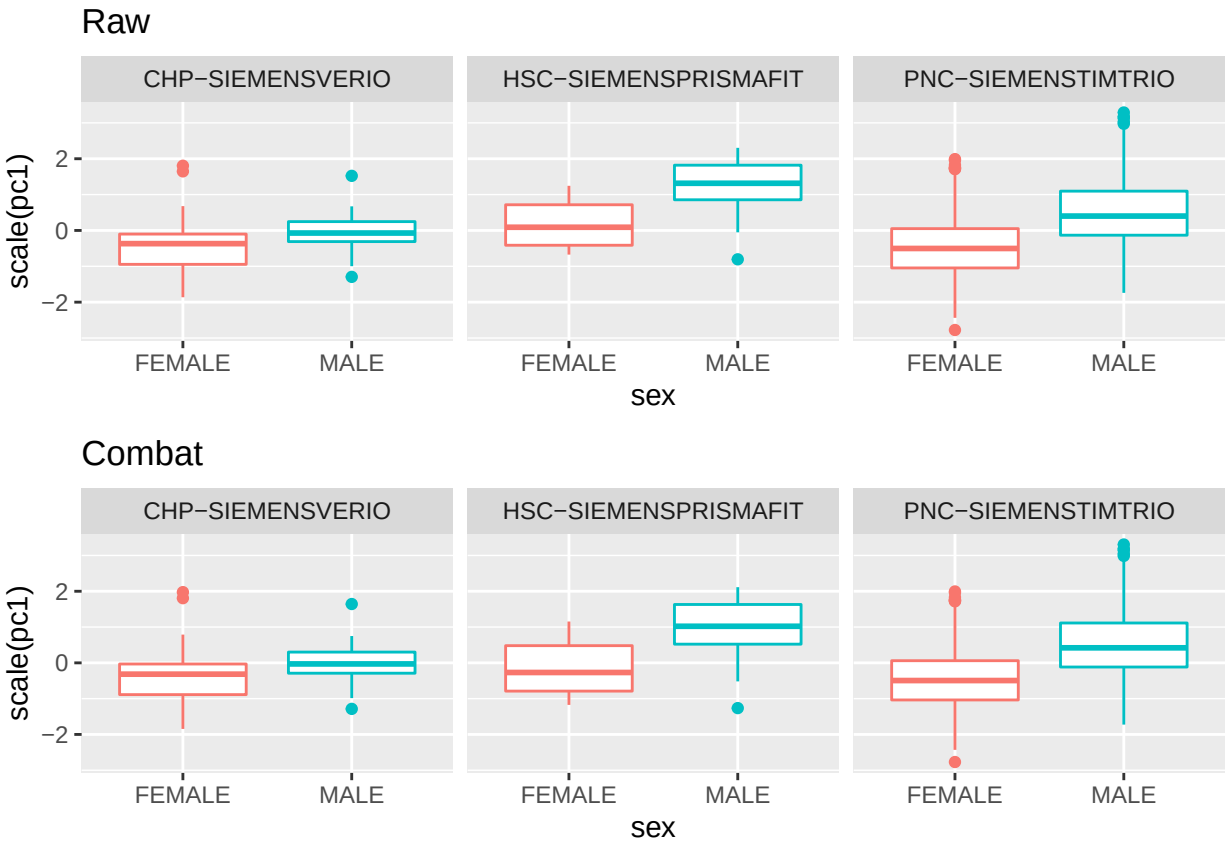



```
## [1] "COMBAT"

##
## Call:
## glm(formula = scale(pc4) ~ batch + scale(age) + sex, data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8461  -0.6179  -0.0937   0.4737   6.8875
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.019068   0.168011  -0.113    0.910
## batchHSC-SIEMENSPRISMAFIT -0.022968   0.215930  -0.106    0.915
## batchPNC-SIEMENSTIMTRIO   -0.022893   0.169714  -0.135    0.893
## scale(age)                 -0.003482   0.028862  -0.121    0.904
## sexMALE                     0.089879   0.056490   1.591    0.112
##
## (Dispersion parameter for gaussian family taken to be 1.001081)
##
```

```
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.4  on 1274  degrees of freedom
## AIC: 3638
##
## Number of Fisher Scoring iterations: 2
```

PCs ~ Sex

PC1




```
## [1] "RAW"
```

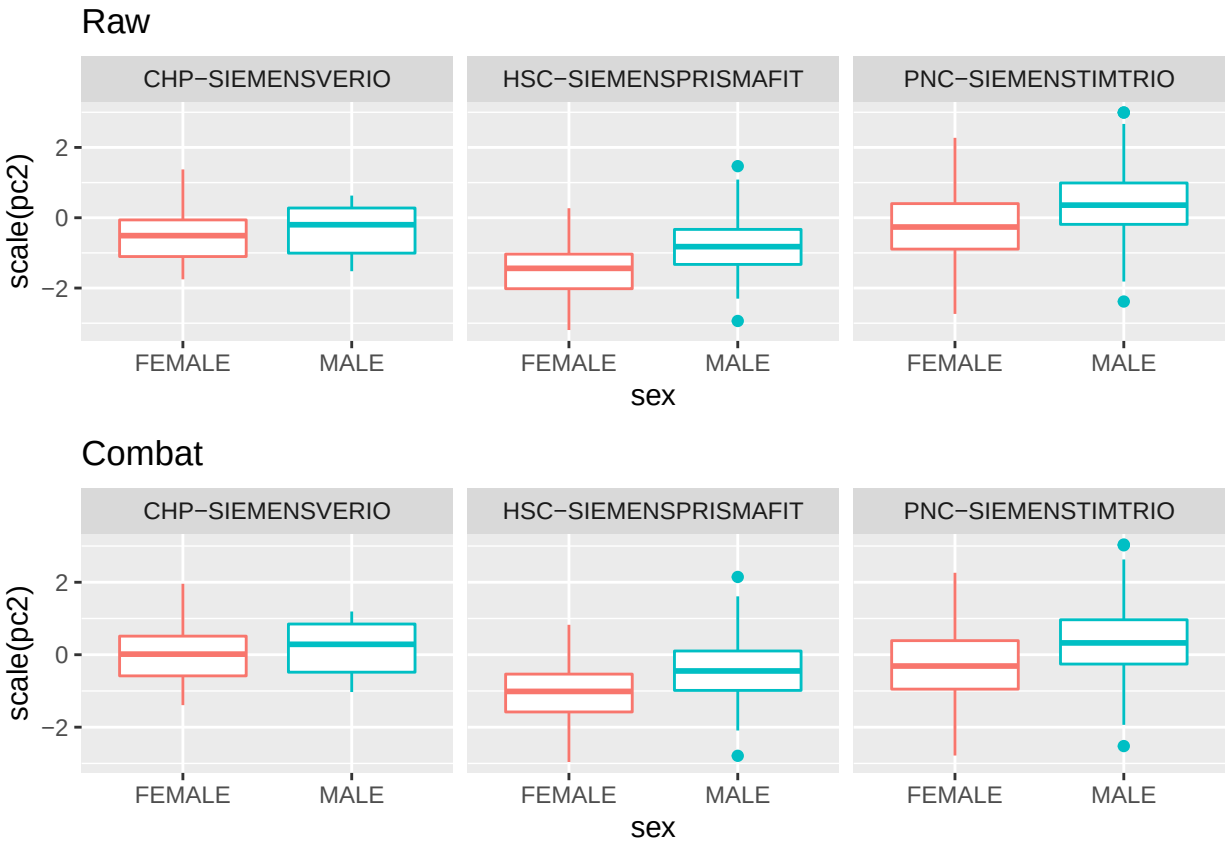
```
##
## Call:
## glm(formula = scale(pc1) ~ batch + sex + scale(age), data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4608  -0.5675  -0.0262   0.5276   2.8222
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.49052    0.13944  -3.518  0.00045 ***
## batchHSC-SIEMENSPRISMAFIT   0.37309    0.17921   2.082  0.03756 *
## batchPNC-SIEMENSTIMTRIO     0.01549    0.14085   0.110  0.91244
## sexMALE                     0.99894    0.04688  21.307  < 2e-16 ***
## scale(age)                  0.28323    0.02395  11.824  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.689549)
##
```

```
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  878.49  on 1274  degrees of freedom
## AIC: 3161.2
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc1) ~ batch + sex + scale(age), data = dat_combat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.46176  -0.57017  -0.02811   0.53092   2.82099
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.43292    0.14023  -3.087  0.00206 **
## batchHSC-SIEMENSPRISMAFIT   0.02452    0.18023   0.136  0.89182
## batchPNC-SIEMENSTIMTRIO    -0.03402    0.14165  -0.240  0.81023
## sexMALE                     1.00783    0.04715  21.375  < 2e-16 ***
## scale(age)                  0.28144    0.02409  11.683  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.697393)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  888.48  on 1274  degrees of freedom
## AIC: 3175.7
##
## Number of Fisher Scoring iterations: 2
```

PC2



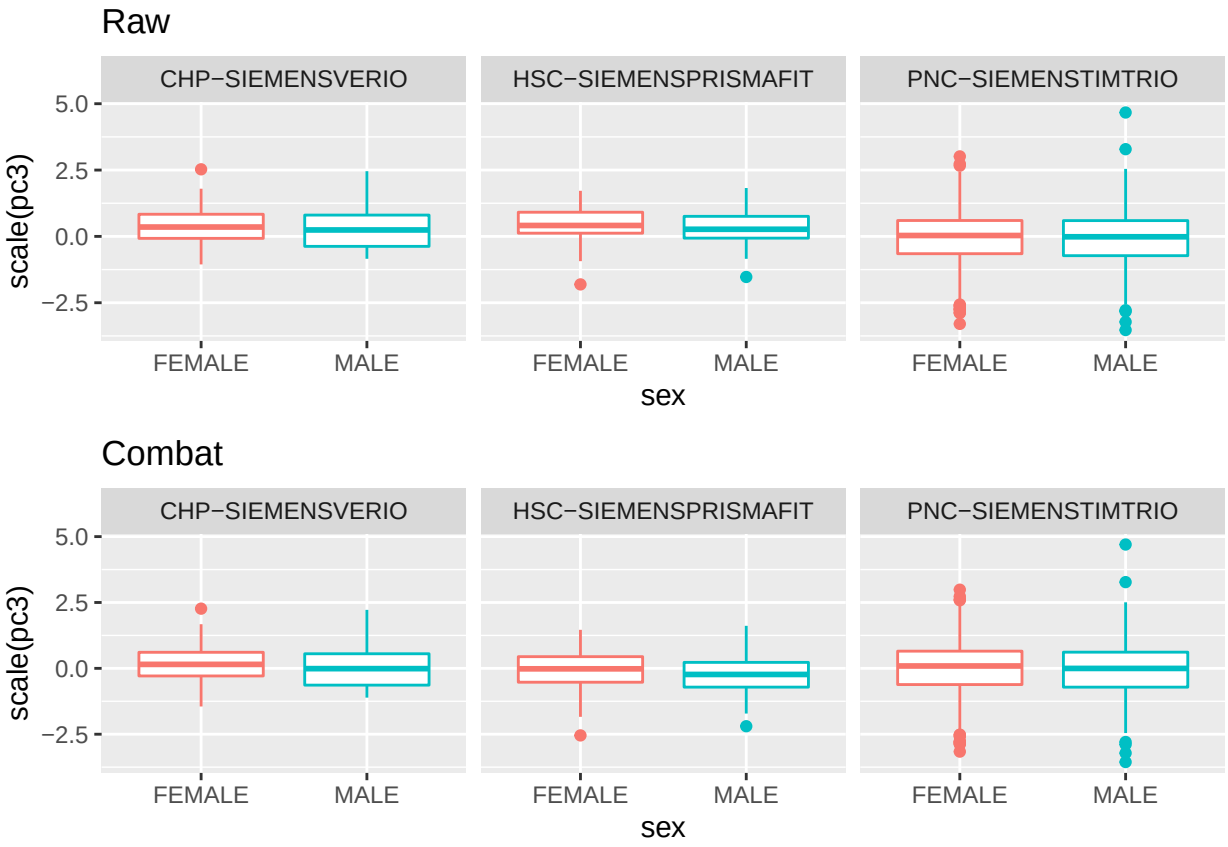

```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc2) ~ batch + sex + scale(age), data = dat)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.29198  -0.48643  -0.03406   0.48108   2.05144
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -0.82502    0.11571  -7.130 1.67e-12 ***
## batchHSC-SIEMENSPRISMAFIT  0.01726    0.14871   0.116    0.908
## batchPNC-SIEMENSTIMTRIO    0.63810    0.11688   5.459 5.73e-08 ***
## sexMALE                    0.50690    0.03890  13.030  < 2e-16 ***
## scale(age)                -0.61110    0.01988 -30.744  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.474796)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  604.89  on 1274  degrees of freedom
```

```
## AIC: 2683.9
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc2) ~ batch + sex + scale(age), data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3307  -0.4842  -0.0361   0.4848   2.0988
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.29351    0.11759  -2.496   0.0127 *
## batchHSC-SIEMENSPRISMAFIT -0.05754    0.15113  -0.381   0.7034
## batchPNC-SIEMENSTIMTRIO    0.07421    0.11878   0.625   0.5322
## sexMALE                    0.49456    0.03954  12.509   <2e-16 ***
## scale(age)                 -0.64102    0.02020 -31.733   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.4903809)
##
##     Null deviance: 1278.00  on 1278  degrees of freedom
## Residual deviance:  624.75  on 1274  degrees of freedom
## AIC: 2725.3
##
## Number of Fisher Scoring iterations: 2
```

PC3




```
## [1] "RAW"
```

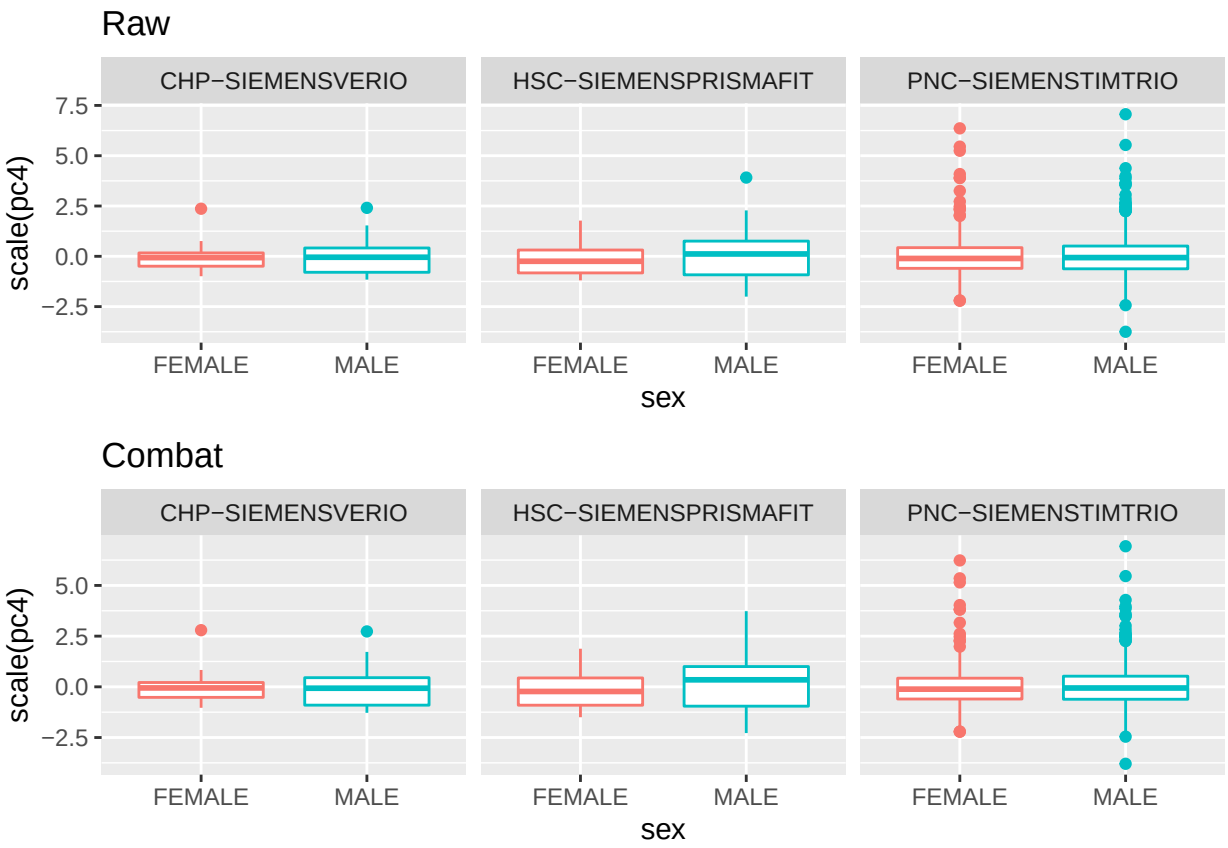
```
##
## Call:
## glm(formula = scale(pc3) ~ batch + sex + scale(age), data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2145  -0.6257   0.0497   0.6084   4.3267
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                0.33296    0.15908   2.093   0.0365 *
## batchHSC-SIEMENSPRISMAFIT  0.43020    0.20445   2.104   0.0356 *
## batchPNC-SIEMENSTIMTRIO   -0.33204    0.16069  -2.066   0.0390 *
## sexMALE                   -0.09750    0.05349  -1.823   0.0685 .
## scale(age)                -0.31203    0.02733 -11.418   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.8974707)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1143.4  on 1274  degrees of freedom
```

```
## AIC: 3498.3
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc3) ~ batch + sex + scale(age), data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2929  -0.6488   0.0554   0.6223   4.3954
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 0.10944    0.16209   0.675   0.4997
## batchHSC-SIEMENSPRISMAFIT   0.09340    0.20832   0.448   0.6540
## batchPNC-SIEMENSTIMTRIO    -0.06194    0.16374  -0.378   0.7053
## sexMALE                    -0.12243    0.05450  -2.246   0.0248 *
## scale(age)                 -0.26979    0.02785  -9.689   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.9318016)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1187.1  on 1274  degrees of freedom
## AIC: 3546.3
##
## Number of Fisher Scoring iterations: 2
```

PC4




```
## [1] "RAW"

##
## Call:
## glm(formula = scale(pc4) ~ batch + sex + scale(age), data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8037  -0.6140  -0.0957   0.4637   7.0235
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.035512   0.168019  -0.211    0.833
## batchHSC-SIEMENSPRISMAFIT   0.016352   0.215941   0.076    0.940
## batchPNC-SIEMENSTIMTRIO    -0.005198   0.169722  -0.031    0.976
## sexMALE                     0.086108   0.056492   1.524    0.128
## scale(age)                 -0.007761   0.028864  -0.269    0.788
##
## (Dispersion parameter for gaussian family taken to be 1.001179)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.5  on 1274  degrees of freedom
## AIC: 3638.1
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] "COMBAT"
```

```
##
## Call:
## glm(formula = scale(pc4) ~ batch + sex + scale(age), data = dat_combat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.8461  -0.6179  -0.0937   0.4737   6.8875
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -0.019068   0.168011  -0.113    0.910
## batchHSC-SIEMENSPRISMAFIT -0.022968   0.215930  -0.106    0.915
## batchPNC-SIEMENSTIMTRIO   -0.022893   0.169714  -0.135    0.893
## sexMALE                    0.089879   0.056490   1.591    0.112
## scale(age)                -0.003482   0.028862  -0.121    0.904
##
## (Dispersion parameter for gaussian family taken to be 1.001081)
##
##     Null deviance: 1278.0  on 1278  degrees of freedom
## Residual deviance: 1275.4  on 1274  degrees of freedom
## AIC: 3638
##
## Number of Fisher Scoring iterations: 2
```

```
```