

bladderbatch_BatchEC-demo.R

hp

2021-07-15

```
#loading required libraries
remotes::install_github("jankinsan/BatchEC")

## Downloading GitHub repo jankinsan/BatchEC@HEAD

## stringi (1.6.2 -> 1.7.2) [CRAN]
## Rcpp (1.0.6 -> 1.0.7) [CRAN]
## isoband (0.2.4 -> 0.2.5) [CRAN]

## Skipping 2 packages not available: Biobase, sva

## Installing 3 packages: stringi, Rcpp, isoband

## Installing packages into 'C:/Users/hp/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)

##
##   There are binary versions available but the source versions are later:
##       binary source needs_compilation
## stringi  1.6.2  1.7.2                TRUE
## Rcpp     1.0.6  1.0.7                TRUE
## isoband  0.2.4  0.2.5                TRUE
##
##   Binaries will be installed
## package 'stringi' successfully unpacked and MD5 sums checked

## Warning: cannot remove prior installation of package 'stringi'

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:
## \Users\hp\Documents\R\win-library\4.0\00LOCK\stringi\libs\icudt69l.dat to C:
## \Users\hp\Documents\R\win-library\4.0\stringi\libs\icudt69l.dat: Invalid
## argument

## Warning in file.copy(savedcopy, lib, recursive = TRUE): problem copying C:
## \Users\hp\Documents\R\win-library\4.0\00LOCK\stringi\libs\x64\stringi.dll to C:
## \Users\hp\Documents\R\win-library\4.0\stringi\libs\x64\stringi.dll: Permission
## denied

## Warning: restored 'stringi'
```

```
## package 'Rcpp' successfully unpacked and MD5 sums checked
## package 'isoband' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\hp\AppData\Local\Temp\RtmpMtiwjn\downloaded_packages
##      checking for file 'C:\Users\hp\AppData\Local\Temp\RtmpMtiwjn\remotes2be07ccb6eb4\jankinsan-1
##      - preparing 'BatchEC':
##      checking DESCRIPTION meta-information ...      checking DESCRIPTION meta-information ... v check
##      - checking for LF line-endings in source and make files and shell scripts
##      - checking for empty or unneeded directories
##      - building 'BatchEC_0.1.0.0000.tar.gz'
##
##
```

```
## Installing package into 'C:/Users/hp/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
library(BatchEC)
library(bladderbatch)
```

```
## Loading required package: Biobase
```

```
## Warning: package 'Biobase' was built under R version 4.0.3
```

```
## Loading required package: BiocGenerics
```

```
## Warning: package 'BiocGenerics' was built under R version 4.0.5
```

```
## Loading required package: parallel
```

```
##
```

```
## Attaching package: 'BiocGenerics'
```

```
## The following objects are masked from 'package:parallel':
```

```
##
```

```
##      clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##      clusterExport, clusterMap, parApply, parCapply, parLapply,
##      parLapplyLB, parRapply, parSapply, parSapplyLB
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      IQR, mad, sd, var, xtabs
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      anyDuplicated, append, as.data.frame, basename, cbind, colnames,
##      dirname, do.call, duplicated, eval, evalq, Filter, Find, get, grep,
##      grepl, intersect, is.unsorted, lapply, Map, mapply, match, mget,
##      order, paste, pmax, pmax.int, pmin, pmin.int, Position, rank,
##      rbind, Reduce, rownames, sapply, setdiff, sort, table, tapply,
##      union, unique, unsplit, which.max, which.min
```

```

## Welcome to Bioconductor
##
## Vignettes contain introductory material; view with
## 'browseVignettes()'. To cite Bioconductor, see
## 'citation("Biobase)"', and for packages 'citation("pkgname)".

#setting the working directory
setwd("E:/ACADEMICS/Summer Training Docs/bladderbatch analysis/output files")

#loading data
data(bladderdata)
# Get the expression data
edata = exprs(bladderEset)
#Get the pheno data
pdata = pData(bladderEset)
batch.data<- cbind(rownames(pdata), pdata[,3])
colnames(batch.data)<- c("Samples", "batch")

#running BatchEC
beacon(expr1 = data.frame(edata),
       batch.info = batch.data,
       batch = "batch",
       NameString = "bladderbatch",
       discrete.batch = TRUE,
       clus.method = "km")

## [1] "Removed 0 genes with zero variance..."
## [1] "22283 genes remain..."
## [1] "=====Linear Regression Analysis=====
## [1] "Calculating Principal Components..."
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC1"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -211.93  -40.83   19.44   56.47  114.73
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      34.961     24.203   1.444  0.1546
## as.factor(pca_data$Batch)2  -75.089     30.721  -2.444  0.0179 *
## as.factor(pca_data$Batch)3  -74.231     46.870  -1.584  0.1193
## as.factor(pca_data$Batch)4 -105.431     43.296  -2.435  0.0184 *
## as.factor(pca_data$Batch)5    9.625     30.413   0.316  0.7529
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80.27 on 52 degrees of freedom
## Multiple R-squared:  0.2492, Adjusted R-squared:  0.1914
## F-statistic: 4.314 on 4 and 52 DF,  p-value: 0.004344
##

```

```
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC2"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -84.415 -26.849   5.011  19.235  69.518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.5454    10.6690   0.051 0.959429
## as.factor(pca_data$Batch)2 -47.9648    13.5421  -3.542 0.000848 ***
## as.factor(pca_data$Batch)3  17.8897    20.6604   0.866 0.390526
## as.factor(pca_data$Batch)4  62.8312    19.0853   3.292 0.001791 **
## as.factor(pca_data$Batch)5  23.5034    13.4062   1.753 0.085466 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 35.39 on 52 degrees of freedom
## Multiple R-squared:  0.5283, Adjusted R-squared:  0.492
## F-statistic: 14.56 on 4 and 52 DF,  p-value: 4.838e-08
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC3"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -81.521 -16.346   1.809  16.900  67.770
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      43.529     8.738   4.981 7.35e-06 ***
## as.factor(pca_data$Batch)2  -39.493    11.092  -3.561 0.000801 ***
## as.factor(pca_data$Batch)3  -62.993    16.922  -3.723 0.000486 ***
## as.factor(pca_data$Batch)4  -76.000    15.632  -4.862 1.11e-05 ***
## as.factor(pca_data$Batch)5  -59.912    10.980  -5.456 1.37e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.98 on 52 degrees of freedom
## Multiple R-squared:  0.4305, Adjusted R-squared:  0.3867
## F-statistic: 9.829 on 4 and 52 DF,  p-value: 5.348e-06
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC4"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
```

```

## Residuals:
##      Min       1Q   Median       3Q      Max
## -53.845 -15.540  -4.065  14.075  71.341
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.5815     8.1138  -0.072   0.943
## as.factor(pca_data$Batch)2   10.9100    10.2988   1.059   0.294
## as.factor(pca_data$Batch)3   14.5707    15.7123   0.927   0.358
## as.factor(pca_data$Batch)4   -4.2566    14.5144  -0.293   0.770
## as.factor(pca_data$Batch)5  -10.5384    10.1955  -1.034   0.306
##
## Residual standard error: 26.91 on 52 degrees of freedom
## Multiple R-squared:  0.1208, Adjusted R-squared:  0.05315
## F-statistic: 1.786 on 4 and 52 DF,  p-value: 0.1457
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC5"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.753 -12.417  -3.152  15.060  82.698
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -9.192     7.297  -1.260   0.2134
## as.factor(pca_data$Batch)2   14.906     9.263   1.609   0.1136
## as.factor(pca_data$Batch)3   -7.023    14.131  -0.497   0.6213
## as.factor(pca_data$Batch)4   -5.713    13.054  -0.438   0.6635
## as.factor(pca_data$Batch)5   16.437     9.170   1.793   0.0789 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 24.2 on 52 degrees of freedom
## Multiple R-squared:  0.1331, Adjusted R-squared:  0.06642
## F-statistic: 1.996 on 4 and 52 DF,  p-value: 0.1088
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC6"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -49.674 -11.907  -1.684  13.235  45.897
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)       2.0617     6.6496   0.310   0.758
## as.factor(pca_data$Batch)2  -0.1592     8.4403  -0.019   0.985

```

```

## as.factor(pca_data$Batch)3  19.6686    12.8768    1.527    0.133
## as.factor(pca_data$Batch)4 -13.3554    11.8951   -1.123    0.267
## as.factor(pca_data$Batch)5   -6.6604     8.3556   -0.797    0.429
##
## Residual standard error: 22.05 on 52 degrees of freedom
## Multiple R-squared:  0.1073, Adjusted R-squared:  0.03864
## F-statistic: 1.563 on 4 and 52 DF,  p-value: 0.1981
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC7"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -56.406  -9.330  -2.031   9.553  41.086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -2.422      5.998  -0.404  0.6880
## as.factor(pca_data$Batch)2   -1.316      7.613  -0.173  0.8634
## as.factor(pca_data$Batch)3  -10.113     11.615  -0.871  0.3879
## as.factor(pca_data$Batch)4  -15.323     10.729  -1.428  0.1592
## as.factor(pca_data$Batch)5   14.675      7.537   1.947  0.0569 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.89 on 52 degrees of freedom
## Multiple R-squared:  0.207, Adjusted R-squared:  0.146
## F-statistic: 3.394 on 4 and 52 DF,  p-value: 0.01534
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC8"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -70.475  -9.706  -1.736  10.443  62.320
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -10.045      6.347  -1.583  0.1196
## as.factor(pca_data$Batch)2    8.913      8.056   1.106  0.2737
## as.factor(pca_data$Batch)3   18.799     12.291   1.529  0.1322
## as.factor(pca_data$Batch)4   12.419     11.354   1.094  0.2791
## as.factor(pca_data$Batch)5   14.467      7.976   1.814  0.0755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.05 on 52 degrees of freedom
## Multiple R-squared:  0.07391, Adjusted R-squared:  0.002676

```

```

## F-statistic: 1.038 on 4 and 52 DF,  p-value: 0.3968
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC9"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -43.487  -8.974   1.668  12.150  43.823
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      12.4834     5.5394   2.254  0.0285 *
## as.factor(pca_data$Batch)2 -17.3381     7.0312  -2.466  0.0170 *
## as.factor(pca_data$Batch)3 -23.2593    10.7271  -2.168  0.0347 *
## as.factor(pca_data$Batch)4   0.9233     9.9092   0.093  0.9261
## as.factor(pca_data$Batch)5 -16.3709     6.9606  -2.352  0.0225 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 18.37 on 52 degrees of freedom
## Multiple R-squared:  0.1775, Adjusted R-squared:  0.1143
## F-statistic: 2.806 on 4 and 52 DF,  p-value: 0.03487
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC10"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -56.061  -9.511  -3.199   8.764  65.737
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)       7.712     5.787   1.333  0.1884
## as.factor(pca_data$Batch)2 -14.492     7.345  -1.973  0.0538 .
## as.factor(pca_data$Batch)3  -1.349    11.206  -0.120  0.9046
## as.factor(pca_data$Batch)4 -10.967    10.351  -1.059  0.2943
## as.factor(pca_data$Batch)5  -6.237     7.271  -0.858  0.3949
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.19 on 52 degrees of freedom
## Multiple R-squared:  0.08318, Adjusted R-squared:  0.01266
## F-statistic:  1.18 on 4 and 52 DF,  p-value: 0.3307
##
## [1] "Plotting boxplots showing how batches are associated with the first ten principal components to

## [1] "Batch is associated with the data..."
## [1] "=====Plotting PCs along with batch=====

```

```

## [1] "Calculating Principal Components..."
## [1] "PCs calculated"
## [1] "===== Performing k-means before_correction===== "

## Warning in dir.create(paste0(dir, "/", "kmeans_", when)): 'E:\ACADEMICS\Summer
## Training Docs\bladderbatch analysis\output files\kmeans_before_correction'
## already exists

## [1] "Calculating distance matrix"
## [1] "Determining the optimal number of clusters for k-means and clustering data..."
## [1] "k=3 is optimal with Average Silhouette Width = 0.168233674677971"
## [1] "Calculating Principal Components..."
## [1] "Plotting Silhouette Plot and Principal Component Analysis biplot (with batches and clustering in

## [1] "=====Batch Effects Adjustment using ComBat===== "
## [1] "Performing batch correction using ComBat..."

## Found5batches

## Adjusting for0covariate(s) or covariate level(s)

## Standardizing Data across genes

## Fitting L/S model and finding priors

## Finding parametric adjustments

## Adjusting the Data

## [1] "Batch corrected data written to file... 20210715_data_bladderbatch_batch_corrected_batch.txt"
## [1] "=====Linear Regression Analysis===== "
## [1] "Calculating Principal Components..."
## [1] "===== "
## [1] "Performing Linear Regression Analysis for PC1"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -221.13  -43.07   23.92   62.66  138.24
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      8.985     27.587   0.326   0.746
## as.factor(pca_data$Batch)2 -16.592     35.015  -0.474   0.638
## as.factor(pca_data$Batch)3 -26.792     53.421  -0.502   0.618
## as.factor(pca_data$Batch)4 -25.580     49.348  -0.518   0.606
## as.factor(pca_data$Batch)5   1.136     34.664   0.033   0.974
##
## Residual standard error: 91.49 on 52 degrees of freedom

```



```

## Multiple R-squared:  0.01476,    Adjusted R-squared:  -0.06103
## F-statistic: 0.1948 on 4 and 52 DF,  p-value: 0.9401
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC2"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -84.602 -30.131   0.771  27.212  89.261
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         4.455      12.833   0.347   0.730
## as.factor(pca_data$Batch)2  -12.974      16.288  -0.797   0.429
## as.factor(pca_data$Batch)3   -1.422      24.850  -0.057   0.955
## as.factor(pca_data$Batch)4    2.153      22.956   0.094   0.926
## as.factor(pca_data$Batch)5   -1.341      16.125  -0.083   0.934
##
## Residual standard error: 42.56 on 52 degrees of freedom
## Multiple R-squared:  0.02043,    Adjusted R-squared:  -0.05493
## F-statistic: 0.2711 on 4 and 52 DF,  p-value: 0.8953
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC3"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -99.423 -14.824   0.339  15.354  69.266
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         9.852      10.102   0.975   0.334
## as.factor(pca_data$Batch)2   -7.592      12.822  -0.592   0.556
## as.factor(pca_data$Batch)3  -21.448      19.562  -1.096   0.278
## as.factor(pca_data$Batch)4  -18.975      18.071  -1.050   0.299
## as.factor(pca_data$Batch)5  -12.854      12.694  -1.013   0.316
##
## Residual standard error: 33.5 on 52 degrees of freedom
## Multiple R-squared:  0.03767,    Adjusted R-squared:  -0.03635
## F-statistic: 0.5089 on 4 and 52 DF,  p-value: 0.7294
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC4"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:

```

```

##      Min      1Q  Median      3Q      Max
## -80.494 -18.686   0.133  18.809  63.607
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           3.018      8.451   0.357   0.722
## as.factor(pca_data$Batch)2    1.232     10.727   0.115   0.909
## as.factor(pca_data$Batch)3   -7.520     16.366  -0.460   0.648
## as.factor(pca_data$Batch)4   -8.696     15.118  -0.575   0.568
## as.factor(pca_data$Batch)5   -6.351     10.620  -0.598   0.552
##
## Residual standard error: 28.03 on 52 degrees of freedom
## Multiple R-squared:  0.02106,    Adjusted R-squared:  -0.05425
## F-statistic: 0.2796 on 4 and 52 DF,  p-value: 0.8899
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC5"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -40.759 -18.130  -5.024  13.612  69.633
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -0.9207      7.8848  -0.117   0.907
## as.factor(pca_data$Batch)2    1.5112     10.0081   0.151   0.881
## as.factor(pca_data$Batch)3   -5.7381     15.2688  -0.376   0.709
## as.factor(pca_data$Batch)4   -0.4657     14.1048  -0.033   0.974
## as.factor(pca_data$Batch)5    2.6609      9.9077   0.269   0.789
##
## Residual standard error: 26.15 on 52 degrees of freedom
## Multiple R-squared:  0.007261,    Adjusted R-squared:  -0.0691
## F-statistic: 0.09509 on 4 and 52 DF,  p-value: 0.9836
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC6"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -50.478 -14.546  -2.687  11.453  83.303
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          -1.3905      7.2948  -0.191   0.850
## as.factor(pca_data$Batch)2    2.2767      9.2593   0.246   0.807
## as.factor(pca_data$Batch)3    6.1463     14.1264   0.435   0.665
## as.factor(pca_data$Batch)4    0.1897     13.0494   0.015   0.988
## as.factor(pca_data$Batch)5    0.6709      9.1664   0.073   0.942

```

```
##
## Residual standard error: 24.19 on 52 degrees of freedom
## Multiple R-squared:  0.004673,    Adjusted R-squared:  -0.07189
## F-statistic: 0.06104 on 4 and 52 DF,  p-value: 0.9929
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC7"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -46.20 -12.62  -1.74   15.31   44.08
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1.990      6.980   0.285   0.777
## as.factor(pca_data$Batch)2    -2.374      8.859  -0.268   0.790
## as.factor(pca_data$Batch)3    -4.802     13.516  -0.355   0.724
## as.factor(pca_data$Batch)4    -5.254     12.486  -0.421   0.676
## as.factor(pca_data$Batch)5    -1.329      8.771  -0.151   0.880
##
## Residual standard error: 23.15 on 52 degrees of freedom
## Multiple R-squared:  0.004979,    Adjusted R-squared:  -0.07156
## F-statistic: 0.06505 on 4 and 52 DF,  p-value: 0.992
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC8"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -41.853 -14.593  -1.677   11.111   53.653
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         3.414      6.542   0.522   0.604
## as.factor(pca_data$Batch)2    -2.531      8.304  -0.305   0.762
## as.factor(pca_data$Batch)3    -2.561     12.669  -0.202   0.841
## as.factor(pca_data$Batch)4    -6.948     11.703  -0.594   0.555
## as.factor(pca_data$Batch)5    -5.477      8.221  -0.666   0.508
##
## Residual standard error: 21.7 on 52 degrees of freedom
## Multiple R-squared:  0.01164,    Adjusted R-squared:  -0.06438
## F-statistic: 0.1532 on 4 and 52 DF,  p-value: 0.9607
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC9"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -69.049  -9.777  -1.640   12.957   46.583
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -1.083      6.354  -0.170   0.865
## as.factor(pca_data$Batch)2    2.201      8.065   0.273   0.786
## as.factor(pca_data$Batch)3    1.361     12.305   0.111   0.912
## as.factor(pca_data$Batch)4   -3.105     11.367  -0.273   0.786
## as.factor(pca_data$Batch)5    1.695      7.984   0.212   0.833
##
## Residual standard error: 21.07 on 52 degrees of freedom
## Multiple R-squared:  0.00562,    Adjusted R-squared:  -0.07087
## F-statistic: 0.07348 on 4 and 52 DF,  p-value: 0.9899
##
## [1] "=====
## [1] "Performing Linear Regression Analysis for PC10"
##
## Call:
## lm(formula = (pca_data[, as.character(x)]) ~ as.factor(pca_data$Batch))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -47.533 -12.030    0.741    7.467   44.457
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.8051      6.1349  -0.131   0.896
## as.factor(pca_data$Batch)2    1.4432      7.7870   0.185   0.854
## as.factor(pca_data$Batch)3    5.4030     11.8802   0.455   0.651
## as.factor(pca_data$Batch)4   -1.2282     10.9744  -0.112   0.911
## as.factor(pca_data$Batch)5    0.2338      7.7089   0.030   0.976
##
## Residual standard error: 20.35 on 52 degrees of freedom
## Multiple R-squared:  0.005814,    Adjusted R-squared:  -0.07066
## F-statistic: 0.07602 on 4 and 52 DF,  p-value: 0.9892
##
## [1] "Plotting boxplots showing how batches are associated with the first ten principal components to

## [1] "=====Plotting PCs along with batch=====
## [1] "Calculating Principal Components..."
## [1] "PCs calculated"
## [1] "===== Performing k-means after_correction=====

## Warning in dir.create(paste0(dir, "/", "kmeans_", when)): 'E:\ACADEMICS\Summer
## Training Docs\bladderbatch analysis\output files\kmeans_after_correction'
## already exists

## [1] "Calculating distance matrix"
## [1] "Determining the optimal number of clusters for k-means and clustering data..."
## [1] "k=3 is optimal with Average Silhouette Width = 0.149563608081298"
```

```
## [1] "Calculating Principal Components..."
## [1] "Plotting Silhouette Plot and Principal Component Analysis biplot (with batches and clustering in

## [1] "=====Plotting PCA Proportion of Variation=====
## [1] "Calculating Principal Components..."
## [1] "Plotting Proportion of Variation"

## [1] "Plotted."
## [1] "=====Plotting scatter plot to show correlation=====
## [1] "Correlation scatter plot will be saved to: 20210715_plot_bladderbatch_batch_correlationPlot.jpeg

## Saving 6.5 x 4.5 in image

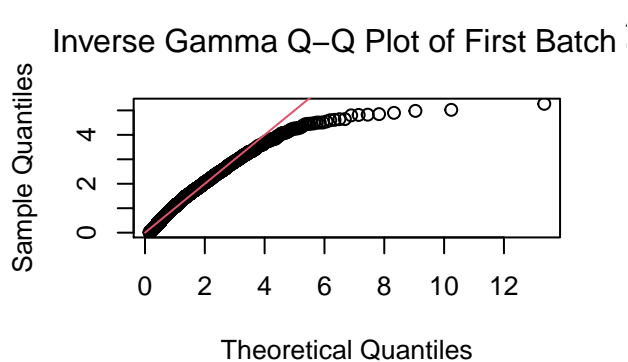
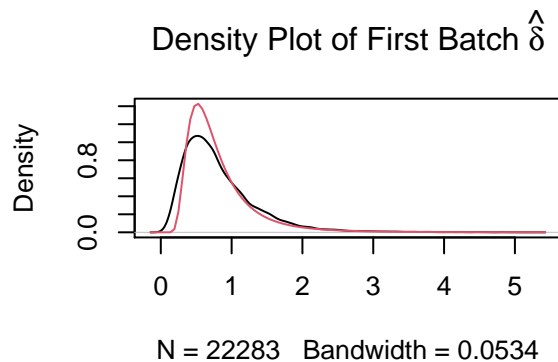
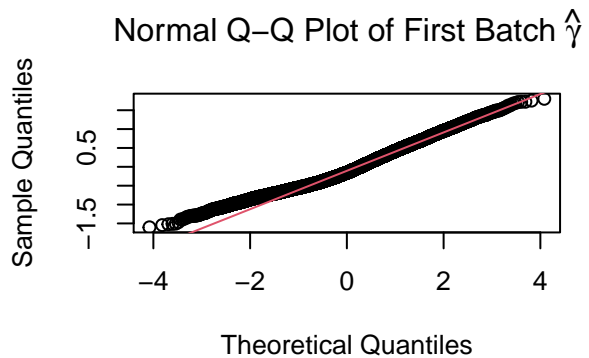
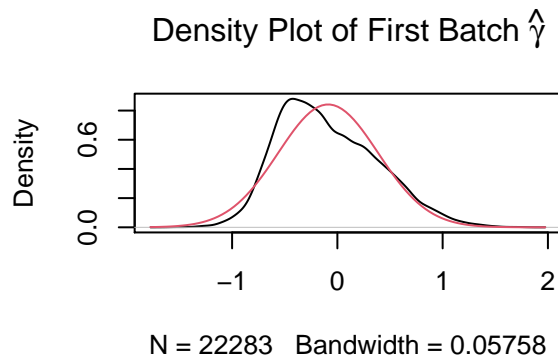
## 'geom_smooth()' using formula 'y ~ x'

## [1] "=====Plotting boxplot before=====

## No id variables; using all as measure variables

## [1] "Boxplot before Batch Correction plotted to 20210715_bladderbatch_batch_gene_expression_boxplot_1
## [1] "=====Plotting boxplot after=====

## No id variables; using all as measure variables
```



```

## [1] "Boxplot after Batch Correction plotted to 20210715_bladderbatch_batch_gene_expression_boxplot_a
## [1] "=====
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] parallel stats graphics grDevices utils datasets methods
## [8] base
##
## other attached packages:
## [1] bladderbatch_1.28.0 Biobase_2.50.0 BiocGenerics_0.36.1
## [4] BatchEC_0.1.0.0000
##
## loaded via a namespace (and not attached):
## [1] httr_1.4.2 edgeR_3.32.1 bit64_4.0.5
## [4] splines_4.0.2 assertthat_0.2.1 highr_0.9
## [7] stats4_4.0.2 blob_1.2.1 yaml_2.2.1
## [10] remotes_2.4.0 pillar_1.6.1 RSQLite_2.2.7
## [13] lattice_0.20-41 limma_3.46.0 glue_1.4.2
## [16] digest_0.6.27 colorspace_2.0-2 htmltools_0.5.1.1
## [19] Matrix_1.3-4 plyr_1.8.6 XML_3.99-0.6
## [22] pkgconfig_2.0.3 genefilter_1.72.1 purrr_0.3.4
## [25] xtable_1.8-4 scales_1.1.1 processx_3.5.2
## [28] BiocParallel_1.24.1 tibble_3.1.2 annotate_1.68.0
## [31] mgcv_1.8-36 generics_0.1.0 farver_2.1.0
## [34] IRanges_2.24.1 ggplot2_3.3.5 ellipsis_0.3.2
## [37] cachem_1.0.5 withr_2.4.2 cli_3.0.0
## [40] survival_3.1-12 magrittr_2.0.1 crayon_1.4.1
## [43] mclust_5.4.7 memoise_2.0.0 evaluate_0.14
## [46] ps_1.6.0 fansi_0.5.0 nlme_3.1-148
## [49] pkgbuild_1.2.0 tools_4.0.2 prettyunits_1.1.1
## [52] lifecycle_1.0.0 matrixStats_0.59.0 stringr_1.4.0
## [55] S4Vectors_0.28.1 locfit_1.5-9.4 munsell_0.5.0
## [58] cluster_2.1.2 AnnotationDbi_1.52.0 callr_3.7.0
## [61] compiler_4.0.2 rlang_0.4.11 grid_4.0.2
## [64] rstudioapi_0.13 labeling_0.4.2 rmarkdown_2.9
## [67] gtable_0.3.0 DBI_1.1.1 curl_4.3.2
## [70] reshape2_1.4.4 R6_2.5.0 knitr_1.33
## [73] dplyr_1.0.7 fastmap_1.1.0 bit_4.0.4
## [76] utf8_1.2.1 rprojroot_2.0.2 stringi_1.6.2
## [79] sva_3.38.0 Rcpp_1.0.6 vctrs_0.3.8
## [82] tidyselect_1.1.1 xfun_0.24

```