MBatch 04-04
Using MBatch Assessments: Boxplot_Group_Structures
Tod Casasent
2017-11-02-0905

Introduction

These instructions are aimed at people familiar with R and familiar with TCGA/GDC platforms and data types. They are intended to introduce the reader to producing the given assessment. These instructions will only rarely, if ever, touch on the appropriateness of the assessment algorithm or interpretation of output. See MBatch_01_InstallLinux.docx for instructions on downloading test data.

Algorithm

Boxplot_Group_Structures is a function used to perform batch effects assessments using the boxplots and violin plots algorithms. Assessment is performed on each batch type available by default.

Output

The primary output method for MBatch is to view results in the Batch Effects Website, described elsewhere. The PNG files are rough versions of the website output.

Graphical output is a set of boxplots, one for each batch, based in this case on the mean for each sample.
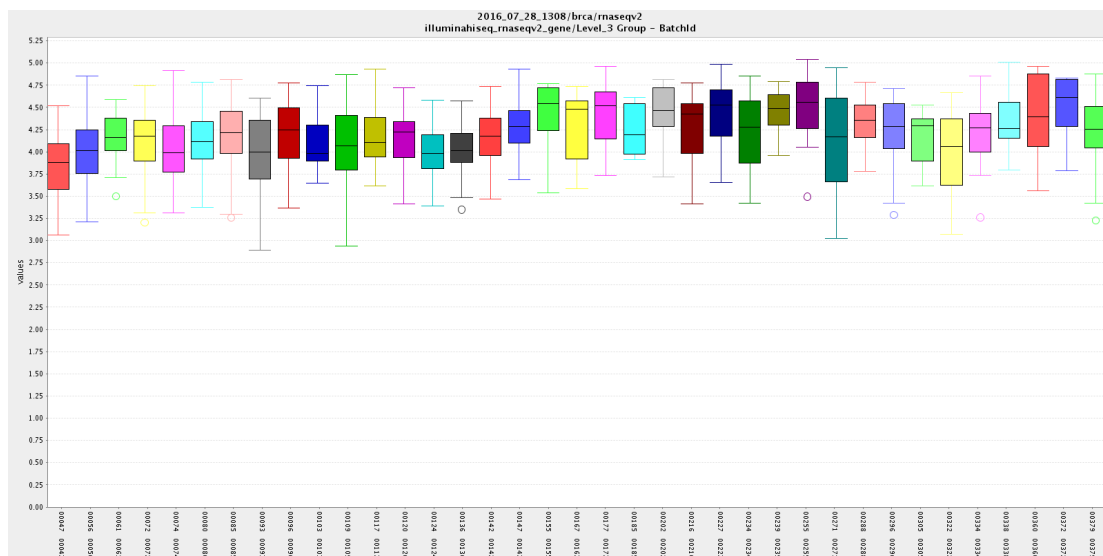
Usage

Boxplot_Group_Structures(theData, theTitle, theOutputPath,

theBatchTypeAndValuePairsToRemove, theBatchTypeAndValuePairsToKeep,

theListOfGroupBoxFunction, theListOfGroupBoxLabels,

theMaxGeneCount=20000, theJavaParameters = "-Xms8000m")

Arguments

theData An instance of BEA_DATA BEA_DATA-class. This is the MBatch Data Object (of class BEA_DATA) described in MBatch_03_UserData.docx, and returned from mbatchLoadFiles or mbatchLoadStructures.

theTitle Object of class "character". Title to use in PNG files.

theOutputPath Object of class "character". Directory in which to place output PNG files and related data files used by the Batch Effects Website.

theBatchTypeAndValuePairsToRemove Object of class "list". A list of vectors containing the batch type (or * for all types) and the value to remove. list() or NULL indicate do nothing. This type of list is described in MBatch_03_ParametersBatchTypesValues.docx.

theBatchTypeAndValuePairsToKeep Object of class "list". A list of vectors containing the batch type (or * for all types) and a vector of the value(s) to keep. list() or NULL indicate do nothing. This type of list is described in MBatch_03_ParametersBatchTypesValues.docx.

theListOfGroupBoxFunction Object of class "vector" A list of functions to use for group assessments. This is a legacy argument. Always send this an argument of "c(mean)". The length must match the length for theListOfGroupBoxLabels. A future version will replace this argument with an alternative method of selecting group functions other than mean.

theListOfGroupBoxLabels Object of class "vector" A list of strings giving strings for labels to use for group assessments. This is a legacy argument. Always send this an argument of "c("Mean")". The length must match the length for theListOfGroupBoxFunction. A future version will replace this argument with an alternative method of selecting group functions other than mean.

theMaxGeneCount Integer giving maximum number of features (genes) to keep. Default is 20000. 0 means keep all.

theJavaParameters Object of class "character" String for initializing JVM. Defaults to -Xms8000m.

Example Call

The following code performs Boxplots and is taken from the tests/Boxplot_Group_Structures.R file. Data used is from the testing data as per the MBatch_01_InstallLinux.docx document.

library(MBatch)

set the paths

theGeneFile <- "/bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv"

theBatchFile <- "/bea_testing/MATRIX_DATA/batches-Tumor.tsv"

theOutputDir <- "/bea_testing/output/Boxplot_Group_Structures"

make sure the output dir exists and is empty

unlink(theOutputDir, recursive=TRUE)

dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)

load the data and reduce the amount of data to reduce run time

myData <- mbatchLoadFiles(theGeneFile, theBatchFile)

myData@mData <- mbatchTrimData(myData@mData, 100000)

here, we take most defaults

Boxplot_Group_Structures(theData=myData,

theTitle="Test",

theOutputPath=theOutputDir,

theBatchTypeAndValuePairsToRemove=NULL,

theBatchTypeAndValuePairsToKeep=NULL,

theListOfGroupBoxFunction=c(mean),

theListOfGroupBoxLabels=c("Mean"),

theJavaParameters="-Xms8000m",

theMaxGeneCount=10000)

Command Line Output

In the future, we plan to make the output from MBatch more user friendly, but currently, this produces the following output at the command line.

> library(MBatch)

>

> # set the paths

> theGeneFile <- "/bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv"

> theBatchFile <- "/bea_testing/MATRIX_DATA/batches-Tumor.tsv"

> theOutputDir <- "/bea_testing/output/Boxplot_Group_Structures"

>

> # make sure the output dir exists and is empty

> unlink(theOutputDir, recursive=TRUE)

> dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)

>

> # load the data and reduce the amount of data to reduce run time

> myData <- mbatchLoadFiles(theGeneFile, theBatchFile)

2017 10 17 11:45:51.392 DEBUG MachineName Changing LC_COLLATE to C for duration of run

2017 10 17 11:45:51.413 INFO MachineName \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/

2017 10 17 11:45:51.413 INFO MachineName Starting mbatchLoadFiles

2017 10 17 11:45:51.414 INFO MachineName MBatch Version: 2017-09-19-1530

2017 10 17 11:45:51.415 INFO MachineName read batch file= /bea_testing/MATRIX_DATA/batches-Tumor.tsv

2017 10 17 11:45:51.432 INFO MachineName read gene file= /bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv

Read 100000 records

2017 10 17 11:46:03.431 INFO MachineName filter samples in batches using gene samples

2017 10 17 11:46:03.432 INFO MachineName sort batches by gene file samples

2017 10 17 11:46:03.526 INFO MachineName Finishing mbatchLoadFiles

2017 10 17 11:46:03.527 INFO MachineName ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

> myData@mData <- mbatchTrimData(myData@mData, 100000)

2017 10 17 11:46:03.528 DEBUG MachineName Changing LC_COLLATE to C for duration of run

2017 10 17 11:46:03.528 INFO MachineName \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/

2017 10 17 11:46:03.529 INFO MachineName mbatchTrimData Starting

2017 10 17 11:46:03.529 INFO MachineName MBatch Version: 2017-09-19-1530

2017 10 17 11:46:11.527 INFO MachineName mbatchTrimData Finishing

2017 10 17 11:46:11.528 INFO MachineName ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

>

> # here, we take most defaults

> Boxplot_Group_Structures(theData=myData,

+ theTitle="Test",

+ theOutputPath=theOutputDir,

+ theBatchTypeAndValuePairsToRemove=NULL,

+ theBatchTypeAndValuePairsToKeep=NULL,

+ theListOfGroupBoxFunction=c(mean),

+ theListOfGroupBoxLabels=c("Mean"),

+ theJavaParameters="-Xms8000m",

+ theMaxGeneCount=10000)

2017 10 17 11:46:11.534 DEBUG MachineName Changing LC_COLLATE to C for duration of run

2017 10 17 11:46:11.537 INFO MachineName \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/

2017 10 17 11:46:11.538 INFO MachineName mbatchFilterData Starting

2017 10 17 11:46:11.538 INFO MachineName MBatch Version: 2017-09-19-1530

2017 10 17 11:46:11.539 DEBUG MachineName rows pre filter 1250

2017 10 17 11:46:11.820 DEBUG MachineName rows post filter 1250

2017 10 17 11:46:11.820 DEBUG MachineName mbatchFilterData Prefilter, gene data had 1250 while post filter 1250

2017 10 17 11:46:11.822 DEBUG MachineName mbatchFilterData Prefilter, batch data had 80 while post filter 80

2017 10 17 11:46:11.822 INFO MachineName mbatchFilterData Finishing

2017 10 17 11:46:11.823 INFO MachineName ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

2017 10 17 11:46:11.838 DEBUG MachineName Changing LC_COLLATE to C for duration of run

2017 10 17 11:46:11.839 INFO MachineName \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/

2017 10 17 11:46:11.839 INFO MachineName mbatchTrimData Starting

2017 10 17 11:46:11.840 INFO MachineName MBatch Version: 2017-09-19-1530

2017 10 17 11:46:11.840 INFO MachineName mbatchTrimData Finishing

2017 10 17 11:46:11.841 INFO MachineName ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^

2017 10 17 11:46:11.842 DEBUG MachineName checkCreateDir: /bea_testing/output/Boxplot_Group_Structure

2017 10 17 11:46:11.845 DEBUG MachineName boxplotJinit - Calling .jinit /home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/jfreechart-1.0.14.jar:/home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/commons-lang3-3.3.2.jar:/home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/commons-math3-3.3.jar:/home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/BoxplotJava.jar:/home/linux/R/x86_64-pc-linux-gnu-library/3.4/MBatch/BoxplotJava/LegendJava.jar

2017 10 17 11:46:12.164 DEBUG MachineName dim(theMatrixGeneData) 1250, dim(theMatrixGeneData) 80

2017 10 17 11:46:12.165 DEBUG MachineName length(colnames(theMatrixGeneData)) 80

2017 10 17 11:46:12.165 DEBUG MachineName length(rownames(theMatrixGeneData)) 1250

2017 10 17 11:46:12.166 DEBUG MachineName dim(theDataframeBatchData) 80, dim(theDataframeBatchData) 5

2017 10 17 11:46:12.166 DEBUG MachineName length(names(theDataframeBatchData)) 5

2017_10_17_1146.12.201 DEBUG 1 BoxplotJava 2017-07-24-1300

2017_10_17_1146.12.202 DEBUG 1 Prepare data for Boxplot_GroupFunction

2017_10_17_1146.12.202 DEBUG 1 theOutputDir=/bea_testing/output/Boxplot_Group_Structures

2017_10_17_1146.12.202 DEBUG 1 theTitle=Test Group

2017_10_17_1146.12.204 DEBUG 1 ReadMatrixFile::convertToMatrix theBarcodeLength=80

2017_10_17_1146.12.204 DEBUG 1 ReadMatrixFile::convertToMatrix theGeneLength=1250

2017_10_17_1146.12.208 DEBUG 1 Before Boxplot_GroupFunction

2017_10_17_1146.12.223 DEBUG 1 BoxplotJava 2017-07-24-1300

2017_10_17_1146.12.225 DEBUG 1 Boxplot_GroupFunction::processInternal - start

2017_10_17_1146.12.225 DEBUG 1 Boxplot_GroupFunction::processInternal - batchTypeName=BatchId

2017_10_17_1146.12.225 DEBUG 1 Boxplot_GroupFunction::processInternal - list of batches

2017_10_17_1146.12.226 DEBUG 1 Boxplot_GroupFunction::processInternal - group function values for samples

2017_10_17_1146.12.280 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches

2017_10_17_1146.12.280 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.12.280 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.12.280 DEBUG 1 ================================

2017_10_17_1146.12.280 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.12.281 DEBUG 1 ================================

2017_10_17_1146.12.281 DEBUG 1 batchTypeIndex=0

2017_10_17_1146.12.281 DEBUG 1 values.length=80

2017_10_17_1146.12.281 DEBUG 1 valLabels.length=80

2017_10_17_1146.12.281 DEBUG 1 elementLabel=00304

2017_10_17_1146.12.281 DEBUG 1 groupLabel=00304

2017_10_17_1146.12.281 DEBUG 1 groupId=0

2017_10_17_1146.12.281 DEBUG 1 ================================

2017_10_17_1146.12.294 DEBUG 1 Boxplot_GroupFunction::processInternal - after iterate through batches

2017_10_17_1146.12.294 DEBUG 1 writeBoxplotFiles - boxplotPng

2017_10_17_1146.12.294 DEBUG 1 BoxplotImpl::boxplotPng theElements.size()=1

2017_10_17_1146.12.295 DEBUG 1 BoxplotImpl::boxplotPng thePNGFile=/bea_testing/output/Boxplot_Gro MEAN/BoxPlot_Group-MEAN_Diagram-BatchId.png

2017_10_17_1146.12.295 DEBUG 1 BoxplotImpl::boxplotPng theDataLabel=values

2017_10_17_1146.12.295 DEBUG 1 BoxplotImpl::boxplotPng theSampleLabel=BatchId

2017_10_17_1146.12.295 DEBUG 1 BoxplotImpl::boxplotPng theTitle=Test Group

2017_10_17_1146.12.295 DEBUG 1 BoxplotImpl::boxplotPng theGroup=BatchId

LegendJava 2013_05_03_0823

writeLegendWithSymbols theTitle = Test Group - BatchId

writeLegendWithSymbols theVersion =

writeLegendWithSymbols theFilenamePath = /bea_testing/output/Boxplot_Group_Structures/Group-MEAN/BoxPlot_Group-MEAN_Legend-BatchId.png

Colors is non-null

writeLegendWithSymbols write

writeLegendWithSymbols done

2017_10_17_1146.16.121 DEBUG 1 writeBoxplotFiles - boxplotAnnotations

2017_10_17_1146.16.170 DEBUG 1 writeBoxplotFiles - boxplotBox

2017_10_17_1146.16.171 DEBUG 1 writeBoxplotFiles - boxplotCat

2017_10_17_1146.16.172 DEBUG 1 writeBoxplotFiles - done

2017_10_17_1146.16.172 DEBUG 1 Boxplot_GroupFunction::processInternal - after files

2017_10_17_1146.16.172 DEBUG 1 Boxplot_GroupFunction::processInternal - batchTypeName=PlateId

2017_10_17_1146.16.172 DEBUG 1 Boxplot_GroupFunction::processInternal - list of batches

2017_10_17_1146.16.173 DEBUG 1 Boxplot_GroupFunction::processInternal - group function values for samples

2017_10_17_1146.16.187 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches

2017_10_17_1146.16.187 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.188 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.188 DEBUG 1 ==============================

2017_10_17_1146.16.188 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.188 DEBUG 1 ==============================

2017_10_17_1146.16.188 DEBUG 1 batchTypeIndex=1

2017_10_17_1146.16.188 DEBUG 1 values.length=80

2017_10_17_1146.16.188 DEBUG 1 valLabels.length=80

2017_10_17_1146.16.188 DEBUG 1 elementLabel=A29J

2017_10_17_1146.16.188 DEBUG 1 groupLabel=A29J

2017_10_17_1146.16.189 DEBUG 1 groupId=0

2017_10_17_1146.16.189 DEBUG 1 ==============================

2017_10_17_1146.16.189 DEBUG 1 Boxplot_GroupFunction::processInternal - after iterate through batches

2017_10_17_1146.16.189 DEBUG 1 writeBoxplotFiles - boxplotPng

2017_10_17_1146.16.189 DEBUG 1 BoxplotImpl::boxplotPng theElements.size()=1

2017_10_17_1146.16.189 DEBUG 1 BoxplotImpl::boxplotPng thePNGFile=/bea_testing/output/Boxplot_Gro MEAN/BoxPlot_Group-MEAN_Diagram-PlateId.png

2017_10_17_1146.16.190 DEBUG 1 BoxplotImpl::boxplotPng theDataLabel=values

2017_10_17_1146.16.190 DEBUG 1 BoxplotImpl::boxplotPng theSampleLabel=PlateId

2017_10_17_1146.16.190 DEBUG 1 BoxplotImpl::boxplotPng theTitle=Test Group

2017_10_17_1146.16.191 DEBUG 1 BoxplotImpl::boxplotPng theGroup=PlateId

LegendJava 2013_05_03_0823

writeLegendWithSymbols theTitle = Test Group - PlateId

writeLegendWithSymbols theVersion =

writeLegendWithSymbols theFilenamePath = /bea_testing/output/Boxplot_Group_Structures/Group-MEAN/BoxPlot_Group-MEAN_Legend-PlateId.png

Colors is non-null

writeLegendWithSymbols write

writeLegendWithSymbols done

2017_10_17_1146.16.346 DEBUG 1 writeBoxplotFiles - boxplotAnnotations

2017_10_17_1146.16.347 DEBUG 1 writeBoxplotFiles - boxplotBox

2017_10_17_1146.16.348 DEBUG 1 writeBoxplotFiles - boxplotCat

2017_10_17_1146.16.349 DEBUG 1 writeBoxplotFiles - done

2017_10_17_1146.16.349 DEBUG 1 Boxplot_GroupFunction::processInternal - after files

2017_10_17_1146.16.350 DEBUG 1 Boxplot_GroupFunction::processInternal - batchTypeName=ShipDate

2017_10_17_1146.16.350 DEBUG 1 Boxplot_GroupFunction::processInternal - list of batches

2017_10_17_1146.16.350 DEBUG 1 Boxplot_GroupFunction::processInternal - group function values for samples

2017_10_17_1146.16.360 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches

2017_10_17_1146.16.360 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.360 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.360 DEBUG 1 ===============================

2017_10_17_1146.16.361 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.361 DEBUG 1 ===============================

2017_10_17_1146.16.361 DEBUG 1 batchTypeIndex=2

2017_10_17_1146.16.361 DEBUG 1 values.length=80

2017_10_17_1146.16.361 DEBUG 1 valLabels.length=80

2017_10_17_1146.16.361 DEBUG 1 elementLabel=2013-05-08

2017_10_17_1146.16.361 DEBUG 1 groupLabel=2013-05-08

2017_10_17_1146.16.361 DEBUG 1 groupId=0

2017_10_17_1146.16.362 DEBUG 1 ===============================

2017_10_17_1146.16.362 DEBUG 1 Boxplot_GroupFunction::processInternal - after iterate through batches

2017_10_17_1146.16.362 DEBUG 1 writeBoxplotFiles - boxplotPng

2017_10_17_1146.16.362 DEBUG 1 BoxplotImpl::boxplotPng theElements.size()=1

2017_10_17_1146.16.362 DEBUG 1 BoxplotImpl::boxplotPng thePNGFile=/bea_testing/output/Boxplot_Gro
MEAN/BoxPlot_Group-MEAN_Diagram-ShipDate.png

2017_10_17_1146.16.362 DEBUG 1 BoxplotImpl::boxplotPng theDataLabel=values

2017_10_17_1146.16.362 DEBUG 1 BoxplotImpl::boxplotPng theSampleLabel=ShipDate

2017_10_17_1146.16.363 DEBUG 1 BoxplotImpl::boxplotPng theTitle=Test Group

2017_10_17_1146.16.363 DEBUG 1 BoxplotImpl::boxplotPng theGroup=ShipDate

LegendJava 2013_05_03_0823

writeLegendWithSymbols theTitle = Test Group - ShipDate

writeLegendWithSymbols theVersion =

writeLegendWithSymbols theFilenamePath = /bea_testing/output/Boxplot_Group_Structures/Group-MEAN/BoxPlot_Group-MEAN_Legend-ShipDate.png

Colors is non-null

writeLegendWithSymbols write

writeLegendWithSymbols done

2017_10_17_1146.16.482 DEBUG 1 writeBoxplotFiles - boxplotAnnotations

2017_10_17_1146.16.483 DEBUG 1 writeBoxplotFiles - boxplotBox

2017_10_17_1146.16.484 DEBUG 1 writeBoxplotFiles - boxplotCat

2017_10_17_1146.16.484 DEBUG 1 writeBoxplotFiles - done

2017_10_17_1146.16.484 DEBUG 1 Boxplot_GroupFunction::processInternal - after files

2017_10_17_1146.16.484 DEBUG 1 Boxplot_GroupFunction::processInternal - batchTypeName=TSS

2017_10_17_1146.16.484 DEBUG 1 Boxplot_GroupFunction::processInternal - list of batches

2017_10_17_1146.16.484 DEBUG 1 Boxplot_GroupFunction::processInternal - group function values for samples

2017_10_17_1146.16.492 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches

11

2017_10_17_1146.16.492 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.492 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.492 DEBUG 1 ===============================

2017_10_17_1146.16.493 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.493 DEBUG 1 ===============================

2017_10_17_1146.16.493 DEBUG 1 batchTypeIndex=3

2017_10_17_1146.16.493 DEBUG 1 values.length=72

2017_10_17_1146.16.493 DEBUG 1 valLabels.length=72

2017_10_17_1146.16.493 DEBUG 1 elementLabel=OR - University of Michigan

2017_10_17_1146.16.493 DEBUG 1 groupLabel=OR - University of Michigan

2017_10_17_1146.16.493 DEBUG 1 groupId=0

2017_10_17_1146.16.493 DEBUG 1 ===============================

2017_10_17_1146.16.493 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.493 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.493 DEBUG 1 ===============================

2017_10_17_1146.16.493 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.493 DEBUG 1 ===============================

2017_10_17_1146.16.493 DEBUG 1 batchTypeIndex=3

2017_10_17_1146.16.494 DEBUG 1 values.length=1

2017_10_17_1146.16.494 DEBUG 1 valLabels.length=1

2017_10_17_1146.16.494 DEBUG 1 elementLabel=OU - Roswell Park

2017_10_17_1146.16.494 DEBUG 1 groupLabel=OU - Roswell Park

2017_10_17_1146.16.494 DEBUG 1 groupId=1

2017_10_17_1146.16.494 DEBUG 1 ===============================

2017_10_17_1146.16.494 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.494 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.494 DEBUG 1 ===============================

2017_10_17_1146.16.494 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.494 DEBUG 1 ==============================

2017_10_17_1146.16.494 DEBUG 1 batchTypeIndex=3

2017_10_17_1146.16.494 DEBUG 1 values.length=2

2017_10_17_1146.16.494 DEBUG 1 valLabels.length=2

2017_10_17_1146.16.494 DEBUG 1 elementLabel=P6 - Translational Genomics Research Institute

2017_10_17_1146.16.494 DEBUG 1 groupLabel=P6 - Translational Genomics Research Institute

2017_10_17_1146.16.495 DEBUG 1 groupId=2

2017_10_17_1146.16.495 DEBUG 1 ==============================

2017_10_17_1146.16.495 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.495 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.495 DEBUG 1 ==============================

2017_10_17_1146.16.495 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.495 DEBUG 1 ==============================

2017_10_17_1146.16.495 DEBUG 1 batchTypeIndex=3

2017_10_17_1146.16.495 DEBUG 1 values.length=1

2017_10_17_1146.16.495 DEBUG 1 valLabels.length=1

2017_10_17_1146.16.495 DEBUG 1 elementLabel=PA - University of Minnesota

2017_10_17_1146.16.495 DEBUG 1 groupLabel=PA - University of Minnesota

2017_10_17_1146.16.495 DEBUG 1 groupId=3

2017_10_17_1146.16.495 DEBUG 1 ==============================

2017_10_17_1146.16.495 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - before inside loop

2017_10_17_1146.16.496 DEBUG 1 Boxplot_GroupFunction::processInternal - iterate through batches - after inside loop

2017_10_17_1146.16.496 DEBUG 1 ==============================

2017_10_17_1146.16.496 DEBUG 1 Boxplot_GroupFunction

2017_10_17_1146.16.496 DEBUG 1 ==============================

2017_10_17_1146.16.496 DEBUG 1 batchTypeIndex=3

2017_10_17_1146.16.496 DEBUG 1 values.length=4

2017_10_17_1146.16.496 DEBUG 1 valLabels.length=4

2017_10_17_1146.16.499 DEBUG 1 elementLabel=PK - University Health Network

2017_10_17_1146.16.499 DEBUG 1 groupLabel=PK - University Health Network

2017_10_17_1146.16.499 DEBUG 1 groupId=4

2017_10_17_1146.16.499 DEBUG 1 ==============================

2017_10_17_1146.16.499 DEBUG 1 Boxplot_GroupFunction::processInternal - after iterate through batches

2017_10_17_1146.16.499 DEBUG 1 writeBoxplotFiles - boxplotPng

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng theElements.size()=5

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng thePNGFile=/bea_testing/output/Boxplot_Gro MEAN/BoxPlot_Group-MEAN_Diagram-TSS.png

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng theDataLabel=values

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng theSampleLabel=TSS

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng theTitle=Test Group

2017_10_17_1146.16.500 DEBUG 1 BoxplotImpl::boxplotPng theGroup=TSS

LegendJava 2013_05_03_0823

writeLegendWithSymbols theTitle = Test Group - TSS

writeLegendWithSymbols theVersion =

writeLegendWithSymbols theFilenamePath = /bea_testing/output/Boxplot_Group_Structures/Group-MEAN/BoxPlot_Group-MEAN_Legend-TSS.png

Colors is non-null

writeLegendWithSymbols write

writeLegendWithSymbols done

2017_10_17_1146.16.689 DEBUG 1 writeBoxplotFiles - boxplotAnnotations

2017_10_17_1146.16.690 DEBUG 1 writeBoxplotFiles - boxplotBox

2017_10_17_1146.16.694 DEBUG 1 writeBoxplotFiles - boxplotCat

2017_10_17_1146.16.697 DEBUG 1 writeBoxplotFiles - done

2017_10_17_1146.16.697 DEBUG 1 Boxplot_GroupFunction::processInternal - after files

2017_10_17_1146.16.697 DEBUG 1 Boxplot_GroupFunction::processInternal - finished

2017_10_17_1146.16.697 DEBUG 1 After Boxplot_GroupFunction

2017 10 17 11:46:16.698 DEBUG MachineName after groupFunction call

[1] TRUE

Example File Output

The above code creates the following output files. Files are named using the following naming convention:

BoxPlot_Group-MEAN_Diagram-<BatchType>.png

BoxPlot_Group-MEAN_Legend-<BatchType>.png

The diagram file contains a Boxplot for all batch types (the columns from the batches.tsv file). The legends give the list of batches within the given batch type, and the colors used for each batch.

The Group-MEAN Boxplots plot the mean for each feature within a batch (genes or probes). So the vertical axis is based on the mean of the values of the original data and the points plotted are features. The actual meaning of the data used, such as expression, read counts, and the like, will vary based on the data being processed.

The annotation TSV files are used internally and contain the total number of data points and the number of non-NA points.

linux@MachineName:/bea_testing/output/Boxplot_Group_Structures/Group-MEAN$ ls -l

total 188

-rw-r--r-- 1 linux linux 54 Oct 17 11:46 BoxPlot_Group-MEAN_Annotations-BatchId.tsv

-rw-r--r-- 1 linux linux 53 Oct 17 11:46 BoxPlot_Group-MEAN_Annotations-PlateId.tsv

-rw-r--r-- 1 linux linux 59 Oct 17 11:46 BoxPlot_Group-MEAN_Annotations-ShipDate.tsv

-rw-r--r-- 1 linux linux 265 Oct 17 11:46 BoxPlot_Group-MEAN_Annotations-TSS.tsv

-rw-r--r-- 1 linux linux 283 Oct 17 11:46 BoxPlot_Group-MEAN_BoxData-BatchId.tsv

-rw-r--r-- 1 linux linux 282 Oct 17 11:46 BoxPlot_Group-MEAN_BoxData-PlateId.tsv

-rw-r--r-- 1 linux linux 288 Oct 17 11:46 BoxPlot_Group-MEAN_BoxData-ShipDate.tsv

-rw-r--r-- 1 linux linux 1013 Oct 17 11:46 BoxPlot_Group-MEAN_BoxData-TSS.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-BatchId-00304.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-PlateId-A29J.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-ShipDate-2013-05-08.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-TSS-OR - University of Michigan.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-TSS-OU - Roswell Park.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-TSS-P6 - Translational Genomics Research Institute.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-TSS-PA - University of Minnesota.tsv

-rw-r--r-- 1 linux linux 9 Oct 17 11:46 BoxPlot_Group-MEAN_CatData-TSS-PK - University Health Network.tsv

-rw-r--r-- 1 linux linux 23062 Oct 17 11:46 BoxPlot_Group-MEAN_Diagram-BatchId.png

-rw-r--r-- 1 linux linux 22846 Oct 17 11:46 BoxPlot_Group-MEAN_Diagram-PlateId.png

-rw-r--r-- 1 linux linux 23394 Oct 17 11:46 BoxPlot_Group-MEAN_Diagram-ShipDate.png

-rw-r--r-- 1 linux linux 24699 Oct 17 11:46 BoxPlot_Group-MEAN_Diagram-TSS.png

-rw-r--r-- 1 linux linux 2108 Oct 17 11:46 BoxPlot_Group-MEAN_Legend-BatchId.png

-rw-r--r-- 1 linux linux 2018 Oct 17 11:46 BoxPlot_Group-MEAN_Legend-PlateId.png

-rw-r--r-- 1 linux linux 2636 Oct 17 11:46 BoxPlot_Group-MEAN_Legend-ShipDate.png

-rw-r--r-- 1 linux linux 11139 Oct 17 11:46 BoxPlot_Group-MEAN_Legend-TSS.png

linux@MachineName:/bea_testing/output/Boxplot_Group_Structures/Group-MEAN$